

Gordon L. Stüber

Principles of Mobile Communication

Fourth Edition

Principles of Mobile Communication

Gordon L. Stüber

Principles of Mobile Communication

Fourth Edition

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

ISBN 978-3-319-55614-7 ISBN 978-3-319-55615-4 (eBook)
DOI 10.1007/978-3-319-55615-4

Library of Congress Control Number: 2017935811

© Springer International Publishing AG 1996, 2000, 2011, 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*To Alana, our son Nickolas,
my father Lothar Stüber,
and my late mother Beatrice Stüber*

Preface

Principles of Mobile Communication, Fourth Edition, like its earlier editions, stresses mathematical modeling of physical layer wireless communication systems. The basic pedagogic methodology is to include fully detailed derivations from the first principles. The textbook is intended to provide enough principle material to guide the novice student, while at the same time having plenty of detailed material to satisfy graduate students inclined to pursue research in the area. The textbook stresses the *principles* of wireless communications that are applicable to a wide array of wireless systems standards. It is intended to serve as a textbook and reference for graduate students and a useful reference for practicing engineers.

Organization of the Book

Chapter 1 begins with an overview that is intended to introduce a broad array of issues relating to wireless communications. Included is a brief description of the evolution of various wireless standards from first-generation cellular standards through emerging fifth-generation cellular standards. Afterwards, the basic concepts of cellular frequency reuse, the land mobile radio propagation environment, link budget, and coverage and capacity of cellular radio systems are discussed at an introductory level.

Chapter 2 provides an extensive treatment of radio propagation, since a good understanding of the physical wireless channel is essential for the development and deployment of wireless systems. The chapter begins with a treatment of the narrowband fading envelope for conventional fixed-to-mobile channels found in cellular radio systems, and mobile-to-mobile channels found in mobile ad hoc networks. After establishing a foundation for understanding narrowband channels, the chapter next treats the statistical characterization of wideband channels. Polarization and depolarization effects are important in wireless systems, the text goes on to discuss the mechanisms and characteristics of channel depolarization. The emulation of wireless channels is essential for the development and testing of wireless systems, and the chapter provides a detailed discussion of multipath-fading channel simulation techniques. Finally, the chapter concludes with a discussion of shadowing and path loss models for land mobile radio environments, including those for small cells, and emerging mm-wave frequencies.

Chapter 3 provides a treatment of co-channel interference which, is the primary impairment in spectrally efficient cellular frequency reuse systems. Very often the receivers in such systems are affected by multiple co-channel interferers. The probability distribution of the total interfering power, and the power sum of noncoherent co-channel interferers are considered in detail. Additionally, something of interest is a certain received signal-to-interference-plus-noise ratio outage threshold performance under various large- and small-scale fading conditions.

Chapter 4 covers the various types of modulation schemes that are used in mobile communication systems along with their spectral characteristics. The chapter begins with the mathematical representation of bandpass modulated signals, along with Nyquist pulse shaping. Afterwards, a large variety of modulation schemes used in wireless systems are considered, including both single-carrier and multi-carrier modulation, and both linear and nonlinear modulation techniques. This is followed by a treatment of the power density spectrum of modulated signals. Although quite mathematical in nature, power spectrum is an important topic, since wireless systems are required to operate within a specified out-of-band emission mask. For this reason, all modulation schemes in Chap. 4 are expressed in terms of a generalized shaping function. The power spectrum is first derived for the generalized shaping function, followed by that of particular modulation schemes.

Chapter 5 discusses the error probability performance of various digital modulation schemes discussed in Chap. 4 on flat fading channels. The performance is evaluated with a variety of receiver structures, including coherent detectors that assume knowledge of the complex channel gain at the receiver and differentially coherent detectors and noncoherent detectors that do not require carrier phase information.

Chapter 6 includes a treatment of multi-antenna techniques. The chapter begins with a discussion of various receiver diversity techniques for coherent, differentially coherent, and noncoherent detection of signals on fading channels corrupted by additive white Gaussian noise. Next, optimal combining is discussed, a technique that is effective when the primary additive impairment is co-channel interference rather than noise. This is followed by a discussion of classical beam forming with uniform linear arrays. The chapter next considers multiple-input multiple-output (MIMO) channels where multiple antennas are used at both the transmitter and receiver to achieve high spectral efficiency. Next, the Alamouti transmit diversity scheme is presented, followed by a brief treatment of MIMO detection. Then, the chapter explores the concept of spatial modulation where a subset (often one) of an array of antenna elements is excited. The chapter concludes with a treatment of massive MIMO, where many base station antennas are used with one or more mobile station antennas.

Chapter 7 considers digital signaling on intersymbol interference (ISI) channels that are typical of land mobile radio systems. The chapter begins with the characterization of ISI channels and goes on to discuss techniques for combating ISI based on time domain symbol-by-symbol equalization and sequence estimation. Afterwards, error probability for maximum likelihood sequence estimation is considered. The chapter concludes with a discussion of co-channel demodulation on ISI channels and considers a receiver for ISI channels that can reject co-channel interference.

Chapter 8 covers error control coding techniques for wireless systems. The chapter begins with a discussion of basic block coding, including space-time block codes. Convolutional coding is considered next along with the Viterbi and BCJR algorithms for decoding convolutional codes, followed by trellis coded modulation. The chapter then provides a detailed discussion on the design and performance analysis of convolutional and trellis codes for AWGN channels, interleaved flat fading channels, and fading ISI channels. Afterwards, space-time trellis codes are treated, and the chapter concludes with 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 wideband channels. This is followed by a discussion of the error probability of direct sequence code division multiple access (CDMA) with a conventional correlation detector. The chapter wraps up with a discussion of CDMA multiuser detection.

Chapter 10 is devoted to multi-carrier techniques. The chapter first considers the performance of OFDM on frequency-selective channels. The effects of residual intersymbol interference (ISI) due to an insufficient guard interval are considered, followed by a remedy. Afterwards, the chapter examines single-carrier frequency domain equalization (SC-FDE) techniques. This is followed by a treatment of orthogonal frequency division multiple access (OFDMA) on both forward, and reverse links. The chapter concludes with a discussion of single-carrier frequency division multiple access (SC-FDMA).

Chapter 11 considers frequency planning techniques for cellular systems. The chapter begins with a discussion of cell sectoring, cell splitting, and reuse partitioning. Afterwards, the chapter considers radio planning for OFDMA cellular systems. This is followed by a detailed treatment of hierarchical overlay/underlay architectures based on cluster planning. Finally, the chapter wraps up with macrodiversity TDMA cellular architectures.

Chapter 12 considers CDMA cellular systems, considering topics such as capacity and power control. This is followed by a discussion of hierarchical macrodiversity CDMA architectures and their performance. Such systems allow, for example, the signals that are received at multiple base stations to be forwarded to a central point for coherent combining, rather than simple selection diversity.

Chapter 13 is devoted to cellular radio resource management. The chapter begins with an introduction to basic hard and soft handoff. Afterwards, the chapter considers the important problem of link quality evaluation, including signal strength averaging, velocity estimation, and velocity adaptive handoff algorithms. Afterwards, a detailed analysis of hard and soft handoff is provided. Finally, the chapter wraps up with methods for estimating the received carrier-to-interference-plus-noise ratio (CINR), which is an important quantity in many types of wireless systems.

Chapter 14 considers channel assignment techniques. This chapter first discusses basic channel assignment techniques and then presents the details of some channel assignment techniques including centralized and decentralized dynamic channel assignment.

Appendix A is a brief and focused tutorial discussion of probability and random processes. A good understanding of the material in Appendix A is essential, since the concepts are widely used throughout the textbook.

Using This Book for Instruction

The book has been developed from a graduate-level course on physical wireless communications that I have taught at Georgia Tech since 1993. The book contains more material than can be taught in a one-semester course. My own preference is to cover material in Chaps. 1 through 6, followed by other material as time permits. Some adjustments will have to be made to account for course prerequisites or the lack thereof.

Acknowledgments

I would like to acknowledge all those who have contributed to the preparation of this book and its earlier editions. The reviewers Prof. Vijay Bhargava, Dr. Sanjiv Nanda, Prof. Upamanyu Madhow, and Prof. Keith Chugg were highly useful for improving the first edition of this book. Prof. Mohsen Kaverhad provided good suggestions for the fourth edition. I am grateful to my doctoral students, past and present, who have contributed to this book through their thesis research. The technical 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, John (Yongchae) Kim, Alenka Zajić, and Seok-Chul Kwon are greatly appreciated.

Atlanta, GA, USA
May 2017

Gordon L. Stüber

Contents

1	Introduction	1
1.1	Brief History of Wireless Systems and Standards	1
1.1.1	First Generation (1G) Cellular Systems	1
1.1.2	Second Generation (2G) Cellular Systems	2
1.1.3	Third Generation (3G) Cellular Systems	5
1.1.4	Fourth Generation (4G) Cellular Systems	8
1.1.5	Fifth Generation (5G) Cellular and Beyond	10
1.1.6	Cordless Telephone Systems	10
1.1.7	Wireless LANs	11
1.1.8	Wireless PANs	13
1.2	Frequency Reuse and the Cellular Concept	13
1.3	Mobile Radio Propagation Environment	15
1.4	Co-Channel Interference and Noise	17
1.5	Receiver Sensitivity and Link Budget	18
1.5.1	Interference Loading	19
1.5.2	Shadow Margin and Handoff Gain	20
1.6	Coverage	23
1.7	Spectral Efficiency and Capacity	24
1.7.1	Bandwidth Efficiency	25
1.7.2	Spatial Efficiency	25
1.7.3	Trunking Efficiency	27
1.7.4	Capacity	28
	Problems	29
2	Propagation Modeling	33
2.1	Fixed-to-Mobile Channels	35
2.1.1	Envelope Correlation	37
2.1.2	Doppler Spectrum	39
2.1.3	Received Envelope and Phase Distribution	44
2.1.4	Envelope Correlation and Spectra	48
2.1.5	Level Crossing Rate and Fade Duration	53
2.1.6	Space-Time Correlation	57
2.2	Mobile-to-Mobile Channels	61
2.2.1	Mobile-to-Mobile Reference Model	62
2.3	Statistical Characterization of Multipath-Fading Channels	63
2.3.1	Statistical Channel Correlation Functions	65
2.3.2	Classification of Channels	66
2.3.3	Power-Delay Profile and Coherence Bandwidth	67
2.3.4	System Correlation Functions	69

2.4	Polarized Channel Modeling	69
2.4.1	Geometric Model for Channel Depolarization	70
2.4.2	XPD in Polarized Multipath-Fading Channels	74
2.4.3	Polarized Channel Model Application	76
2.5	Simulation of Multipath-Fading Channels	78
2.5.1	Filtered White Gaussian Noise Models	79
2.5.2	Sum of Sinusoids Methods	82
2.5.3	Multiple Uncorrelated Faded Envelopes	89
2.5.4	Wide-Band Simulation Models	95
2.5.5	Mobile-to-Mobile Simulation Models	100
2.5.6	Symbol-Spaced Models	106
2.6	Shadowing	110
2.6.1	Shadow Simulation	111
2.6.2	Composite Shadowing–Fading Distributions	112
2.7	Path Loss Models	114
2.7.1	Free Space Path Loss	114
2.7.2	Flat Earth Path Loss	115
2.7.3	Empirical Path Loss Models	117
	Appendix 2A: COST 207 Channel Models	128
	Appendix 2B: COST 259 Channel Models	130
	Appendix 2C: ITU Channel Models	131
	Appendix 2D: Derivation of Eq. (2.358)	132
	Appendix 2E: Derivation of Eq. (2.374)	132
	Problems	134
3	Co-Channel Interference	147
3.1	Multiple Log-Normal Interferers	148
3.1.1	Fenton–Wilkinson Method	149
3.1.2	Schwartz & Yeh Method	150
3.1.3	Farley’s Method	152
3.1.4	Numerical Comparisons	153
3.2	Log-Normal/Multiple Log-Normal Interferers	153
3.3	Rician/Multiple Rayleigh Interferers	156
3.4	Log-Normal Nakagami/Multiple Log-Normal Nakagami Interferers	157
3.4.1	Statistically Identical Interferers	158
3.4.2	Statistically Nonidentical Co-Channel Interferers	160
	Problems	162
4	Modulation and Power Spectrum	165
4.1	Representation of Bandpass Modulated Signals	166
4.1.1	Vector Space Representations	167
4.1.2	Gram–Schmidt Orthonormalization Procedure	167
4.1.3	Signal Energy and Correlations	172
4.2	Nyquist Pulse Shaping	173
4.2.1	Raised Cosine and Root Raised Cosine Pulse	175
4.3	Quadrature Amplitude Modulation	176
4.3.1	QAM Signal Constellations	177
4.3.2	PAM Signal Constellations	178
4.4	Phase Shift Keying	179
4.4.1	Offset QPSK (OQPSK)	180
4.4.2	$\pi/4$ -DQPSK	180
4.5	Orthogonal Modulation and Variants	182
4.5.1	Orthogonal FSK Modulation	182
4.5.2	Binary Orthogonal Codes	183
4.5.3	Biorthogonal Signals	184
4.5.4	Orthogonal Multipulse Modulation	184

4.6	Orthogonal Frequency Division Multiplexing	185
4.6.1	DFT-Based OFDM Baseband Modulator	186
4.6.2	Adaptive Bit Loading and Discrete Multitone Modulation	187
4.6.3	Multiresolution Modulation	188
4.7	Continuous Phase Modulation	188
4.7.1	Full Response CPM	190
4.8	Partial Response CPM	192
4.8.1	Phase States	194
4.8.2	Gaussian Minimum Shift Keying	195
4.8.3	Linearized GMSK (LGMSK)	198
4.8.4	Tamed Frequency Modulation	200
4.9	Power Spectrum	203
4.9.1	Psd of the Complex Envelope	204
4.9.2	Psd of QAM	209
4.9.3	Psd of PSK	210
4.9.4	Psd of OQPSK	210
4.9.5	Psd of $\pi/4$ -DQPSK	211
4.9.6	Psd of OFDM	211
4.9.7	Psd of Full Response CPM	215
4.9.8	Psd of GMSK and TFM	222
	Problems	223
5	Detection of Known Signals in Noise	231
5.1	Vector Space Representation of Received Signals	231
5.2	Detection of Known Signals in AWGN	234
5.3	Probability of Error	237
5.3.1	Pairwise Error Probability	238
5.3.2	Upper Bounds on Error Probability	239
5.3.3	Lower Bound on Error Probability	240
5.3.4	Bit Versus Symbol Error Probabilities	241
5.3.5	Rotation and Translations	242
5.4	Error Probability of PSK	243
5.4.1	Error Probability of BPSK	243
5.4.2	Error Probability of QPSK and OQPSK	244
5.4.3	Error Probability of M -PSK	245
5.4.4	Error Probability with Rayleigh Fading	246
5.4.5	Differential PSK	248
5.5	Error Probability of PAM and QAM	248
5.5.1	Error Probability of M -PAM	248
5.5.2	Error Probability of M -QAM	250
5.6	Error Probability of Orthogonal Signals	251
5.6.1	Orthogonal Signals	251
5.6.2	Biorthogonal Signals	253
5.7	Error Probability of OFDM	253
5.7.1	Interchannel Interference	255
5.8	Differential Detection	258
5.8.1	Binary DPSK	258
5.8.2	Differential Detection of $\pi/4$ -DQPSK	260
5.9	Non-coherent Detection	261
5.9.1	Error Probability of M -ary Orthogonal Signals	263
5.10	Detection of CPM Signals	265
5.10.1	Coherent Detection and Error Probability of MSK and GMSK	265
5.10.2	Coherent CPM Demodulator	266
5.10.3	Non-coherent CPM Demodulator	267
	Problems	268

6	Multi-Antenna Techniques	273
6.1	Diversity Combining	274
6.2	Selective Combining	275
6.3	Maximal Ratio Combining	277
6.3.1	Granlund Combiner	280
6.4	Equal Gain Combining	283
6.5	Switched Combining	285
6.6	Differential Detection with Equal Gain Combining	286
6.7	Non-coherent Square-Law Combining	289
6.8	Optimum Combining	291
6.8.1	Optimum Combining Performance	295
6.8.2	Distributed Antenna Systems	301
6.9	Classical Beam-Forming	303
6.10	MIMO Channels	306
6.10.1	Analytical MIMO Channel Models	308
6.11	Transmitter Diversity	310
6.11.1	Alamouti's Transmit Diversity Scheme	310
6.12	MIMO Detection	313
6.12.1	Maximum Likelihood Detection	314
6.12.2	Minimum Mean-Square-Error Detector	314
6.12.3	Inverse Channel Detector	315
6.12.4	Known Channel at the Transmitter and Receiver	315
6.13	Spatial Modulation	316
6.14	Massive MIMO	318
6.14.1	Massive MIMO System Model	319
6.14.2	Reverse Link Pilots	320
6.14.3	Reverse Link Data	321
6.14.4	Forward Link Data	323
6.14.5	Favorable Propagation Conditions	324
	Problems	329
7	Time Domain Equalization and Interference Cancellation	333
7.1	Modeling of ISI Channels	333
7.1.1	Vector Representation of Received Signals	334
7.2	Maximum Likelihood Receiver for ISI Channels with AWGN	335
7.2.1	Discrete-Time White Noise Channel Model	336
7.3	Symbol-by-Symbol Equalizers	343
7.3.1	Linear Equalizer	344
7.3.2	Decision Feedback Equalizer	352
7.4	Sequence Estimation	355
7.4.1	Maximum Likelihood Sequence Estimation	355
7.4.2	Delayed Decision-Feedback Sequence Estimation	360
7.4.3	Reduced-State Sequence Estimation	362
7.5	Error Probability for MLSE on ISI Channels	363
7.5.1	Static ISI Channels	365
7.5.2	Fading ISI Channels	367
7.6	Error Probability for $T/2$ -Spaced MLSE Receiver	370
7.6.1	T -Spaced MLSE Receiver	370
7.6.2	$T/2$ -Spaced MLSE Receiver	371
7.6.3	Timing Phase Sensitivity	372
7.6.4	Practical $T/2$ -Spaced MLSE Receiver	373
7.7	Co-channel Demodulation	375
7.7.1	System and Channel Model	375
7.7.2	Joint Maximum Likelihood Co-channel Receiver	375
7.7.3	Discrete-Time MIMO Channel Model	378

7.7.4	The Viterbi Algorithm	380
7.7.5	Pairwise Error Probability	381
7.7.6	$T/2$ -Spaced MIMO J-MLSE Receiver	382
7.7.7	Interference Rejection Combining MLSE Receiver	387
7.7.8	Examples	390
	Appendix 7A: Derivation of Eq. (7.202)	391
	Problems	392
8	Error Control Coding	397
8.1	Block Codes	399
8.1.1	Binary Block Codes	399
8.1.2	Space-Time Block Codes	403
8.2	Convolutional Codes	407
8.2.1	Encoder Description	407
8.2.2	State and Trellis Diagrams, and Weight Distribution	409
8.2.3	Recursive Systematic Convolutional Codes	412
8.2.4	Maximum Likelihood Decoding	414
8.2.5	BCJR Algorithm	416
8.3	Trellis Coded Modulation	420
8.3.1	Encoder Description	420
8.3.2	Mapping by Set Partitioning	421
8.4	Code Performance on AWGN Channels	423
8.4.1	Union Bound for Convolutional Codes	424
8.5	Interleaving	427
8.5.1	Block Interleaving	428
8.5.2	Convolutional Interleaving	429
8.6	Code Performance on Interleaved Flat Fading Channels	429
8.6.1	TCM with Symbol Interleaving	429
8.6.2	Bit Interleaved Coded Modulation	433
8.7	Performance of Space-Time Codes	434
8.7.1	Space-Time Trellis Codes	436
8.8	Turbo Codes	438
8.8.1	PCCC Encoder	439
8.8.2	PCCC Decoder	440
8.8.3	SCCC Encoder and Decoder	441
8.8.4	Weight Distribution	442
	Problems	445
9	Spread Spectrum Techniques	449
9.1	Basic Principles of Spread Spectrum	450
9.1.1	Direct Sequence Spread Spectrum	450
9.1.2	Frequency Hop Spread Spectrum	453
9.2	Spreading Sequences	454
9.2.1	Spreading Waveforms	455
9.2.2	m -sequences	456
9.2.3	Gold Sequences	458
9.2.4	Kasami Sequences	459
9.2.5	Barker Sequences	459
9.2.6	Walsh-Hadamard Sequences	460
9.2.7	Variable Length Orthogonal Codes	461
9.2.8	Complementary Code Keying	462
9.2.9	Zadoff–Chu Sequences	462
9.3	Power Spectral Density of DS Spread Spectrum Signals	464

9.4	Performance of DS/QPSK in CW Tone Interference.....	467
9.4.1	Short Code	470
9.4.2	Short Code Design	475
9.4.3	Long Code.....	477
9.5	DS Spread Spectrum on Frequency-Selective Fading Channels.....	478
9.5.1	RAKE Receiver	481
9.5.2	Error Probability of DS/BPSK with a RAKE Receiver	481
9.6	Error Probability for DS CDMA on AWGN Channels	485
9.6.1	Standard Gaussian Approximation.....	488
9.6.2	Improved Gaussian Approximation	490
9.7	CDMA Multiuser Detection	491
9.7.1	Optimum CDMA Multiuser Detection	492
9.7.2	Decorrelating Detector	494
9.7.3	Minimum Mean Square Error Detector.....	494
	Problems	495
10	Multi-Carrier Techniques	501
10.1	Orthogonal Frequency Division Multiplexing	502
10.1.1	Performance of OFDM on ISI Channels.....	504
10.1.2	Residual ISI Cancellation	508
10.1.3	Performance of the RISIC Algorithm.....	509
10.2	Single-Carrier Frequency Domain Equalization	513
10.2.1	ZF and MMSE SC-FDE	515
10.3	Orthogonal Frequency Division Multiple Access	517
10.3.1	OFDMA Forward Link	518
10.3.2	OFDMA Reverse Link	522
10.3.3	Peak-to-Average Power Ratio	523
10.4	Single-Carrier Frequency Division Multiple Access	523
10.4.1	Peak-to-Average Power Ratio	526
	Problems	528
11	Frequency Planning Techniques	529
11.1	Cell Sectoring	529
11.1.1	Cell Sectoring with Wide-Beam Directional Antennas	529
11.2	Conventional Cell Splitting	531
11.2.1	Reuse Partitioning	533
11.3	OFDMA/SC-FDMA Radio Planning	534
11.4	Cluster-Planned Hierarchical Architecture	536
11.4.1	System Architecture	536
11.4.2	Underlaid Microcell Planning Algorithm.....	538
11.4.3	Performance Analysis of Cluster-Planned Architecture	540
11.5	Macrodiversity Architectures.....	554
11.5.1	Co-channel Interference Outage	554
11.5.2	Shadow Correlation.....	555
11.5.3	Numerical Examples.....	557
	Problems	558
12	CDMA Cellular Systems	563
12.1	CDMA Power Control	563
12.2	Capacity of Cellular CDMA.....	565
12.2.1	Reverse Link Capacity.....	565
12.2.2	Forward Link Capacity	571
12.2.3	Imperfect Power Control	572

12.3	Hierarchical Macrodiversity CDMA Cellular Architectures	574
12.3.1	System Model	575
12.3.2	Reverse Link Analysis	576
12.3.3	Forward Link Analysis	581
	Appendix 12A: Derivation of Eq. (12.78)	588
	Appendix 12B: Derivation of Eq. (12.79)	590
	Problems	591
13	Radio Resource Management	593
13.1	Signal Strength Based Hard Handoff Algorithms	596
13.2	Pilot-to-Interference Ratio Based Soft Handoff Algorithms	597
13.3	Signal Strength Averaging	598
13.3.1	Choosing the Proper Window Length	599
13.3.2	Choosing the Proper Number of Samples to Average	600
13.4	Velocity Estimation	601
13.4.1	Level Crossing Rate Estimators	603
13.4.2	Covariance Approximation Methods	606
13.4.3	Velocity Estimator Sensitivity	607
13.5	Velocity Adaptive Hard Handoff Algorithms	611
13.5.1	Effect of N_λ	612
13.5.2	Corner Effects and Sensitivity to a and W_l	613
13.5.3	Velocity Adaptive Handoff Performance	613
13.6	Hard Handoff Analysis	614
13.6.1	Simulation Results	618
13.7	CDMA Soft Handoff Analysis	618
13.7.1	System Model and Analysis	619
13.7.2	Numerical Results	625
13.8	CINR Based Link Quality Measurements	632
13.8.1	Discrete-Time Model for Signal Quality Estimation	633
13.8.2	Training Sequence Based $C/(I + N)$ Estimation	636
	Appendix 13A: Derivation of Equations (13.43) and (13.58)	639
	Problems	640
14	Channel Assignment Techniques	643
14.1	Centralized DCA	646
14.1.1	Maximum Packing	646
14.1.2	MAXMIN Scheme	647
14.2	Decentralized DCA	648
14.2.1	First Available and Nearest Neighbor	648
14.2.2	Dynamic Resource Acquisition	648
14.3	Fully Decentralized DCA	649
14.3.1	Channel Segregation	649
14.3.2	Channel Segregation with Variable Threshold	649
14.3.3	Minimum Interference Schemes	650
14.3.4	Aggressive and Timid DCA Strategies	651
14.4	Hybrid FCA/DCA Schemes	652
14.5	Borrowing Schemes	652
14.5.1	Borrowing with Channel Ordering	653
14.5.2	Borrowing with Directional Locking	654
14.5.3	Borrowing Without Locking	654
14.5.4	Compact Pattern Based DCA	655
14.6	Directed Retry and Directed Handoff	656
14.7	Moving Direction Strategies	656
14.8	Reduced Transceiver Coverage	656
14.8.1	Reuse Partitioning	657
14.9	Handoff Priority	657

14.10	Example DCA Schemes for TDMA Systems	657
14.10.1	The Simple DCA (SDCA) Strategy	659
14.10.2	A Queueing DCA Strategy	660
14.10.3	An Aggressive DCA Strategy	662
14.10.4	Simulation Model, Results, and Discussion	664
14.11	Concluding Remarks	669
	Problems	669
A	Probability and Random Processes	671
A.1	Conditional Probability and Bayes' Theorem	671
A.1.1	Total Probability	672
A.1.2	Bayes' Theorem	672
A.2	Means, Moments, and Moment Generating Functions	673
A.2.1	Bivariate Random Variables	674
A.3	Some Useful Probability Distributions	675
A.3.1	Discrete Distributions	675
A.3.2	Continuous Distributions	676
A.4	Upper Bounds on the cdfc	682
A.4.1	Chebyshev Bound	682
A.4.2	Chernoff Bound	683
A.5	Random Processes	685
A.5.1	Moments and Correlation Functions	686
A.5.2	Cross-correlation and Crosscovariance	690
A.5.3	Complex-Valued Random Processes	692
A.5.4	Power Spectral Density	693
A.5.5	Random Processes Filtered by Linear Systems	694
A.5.6	Cyclostationary Random Processes	696
A.5.7	Discrete-Time Random Processes	697
	References	699
	Index	709

Acronyms

AAoA	Azimuth angle of arrival
AAoD	Azimuth angle of departure
ADC	Analog to digital converter
ADPCM	Adaptive differential pulse-code modulation
AGC	Automatic gain control
AMPS	Advanced Mobile Phone Service
AoA	Angle of arrival
AoD	Angle of departure
APP	A posteriori probability
ARQ	Automatic repeat request
ASELP	Algebraic code excitation linear prediction
AWGN	Additive white Gaussian noise
BCH	Bose-Chaudhuri-Hocquenghem
BCJR	Bahl-Cocke-Jelinek-Raviv
BER	Bit error rate
BICM	Bit interleaved coded modulation
BPSK	Binary phase shift keying
BS	Base station
BSC	Binary symmetric channel
BU	Bad urban
CC	Convolutional code
CCI	Co-channel interference
CCIR	Comité International des Radio-Communication
CCK	Complementary code keying
CC-OFDMA	Clustered-carrier orthogonal frequency division multiple access
cdf	Cumulative distribution function
cdfc	Complementary distribution function
CDMA	Code division multiple access
CIR	Carrier-to-interference ratio
CINR	Carrier-to-interference-plus-noise ratio
CNR	Carrier-to-noise ratio
COV	Covariance
CPFSK	Continuous phase frequency shift keying
CPM	Continuous phase modulation
CSI	Channel state information
CW	Continuous wave
CWEF	Conditional weight enumerating function
DAC	Digital to analog converter
DCA	Dynamic channel assignment
DDFSE	Delayed decision feedback sequence estimation
DECT	Digital enhanced cordless telephone

DFE	Decision feedback equalizer
DFT	Discrete Fourier transform
DMT	Discrete multitone modulation
DPSK	Differential phase shift keying
DQPSK	Differential quadrature phase shift keying
DS	Direct sequence
DTFT	Discrete-time Fourier transform
EAoA	Elevation angle of arrival
EAoD	Elevation angle of departure
EDGE	Enhanced data rates for GSM evolution
EGC	Equal gain combining
EFR	Enhanced full rate
EIRP	Effective isotropic radiated power
EVRC	Enhanced variable rate codec
FEPL	Flat Earth path loss
FDD	Frequency division duplexing
FDE	Frequency-domain equalization
FFH	Fast frequency hopping
FFT	Fast Fourier transform
FDMA	Frequency division multiple access
FER	Frame error rate
FH	Frequency hop
FIR	Finite impulse response
FSK	Frequency shift keying
FSM	Finite-state machine
FSPL	Free-space path loss
GFSK	Gaussian frequency shift keying
GMSK	Gaussian minimum shift keying
GPRS	General Packet Radio Service
GPS	Global positioning system
GSM	Global System for Mobile Communication
GTFM	Generalized tamed frequency modulation
H-ARQ	Hybrid-automatic repeat request
HF	High frequency
HMRC	Hierarchical maximum ratio combining
HSDPA	High-speed downlink packet access
HSPA	High-speed packet access
HSUPA	High-speed uplink packet access
HT	Hilly terrain
ICI	Interchannel interference
IDFT	Inverse discrete Fourier transform
I-FDMA	Interleaved frequency division multiple access
IFFT	Inverse fast Fourier transform
IIR	Infinite impulse response
IoT	Internet of things
IRC	Interference rejection combining
ISI	Intersymbol interference
ITU	International Telecommunications Union
ITU-R	International Telecommunications Union-Radio Communications
ITU-T	International Telecommunications Union-Telecommunications
J-MLSE	Joint maximum likelihood sequence estimation
LCR	Level crossing rate
L-FDMA	Localized frequency division multiple access
LGMSK	Linearized Gaussian minimum shift keying
LLR	Log-likelihood ratio

LMS	Least mean square
LoS	Line-of-sight
LSFR	Linear feedback shift register
LTE	Long-Term Evolution
LTE-A	Long-Term Evolution-Advanced
MAHO	Mobile-assisted handoff
MAP	Maximum a posteriori probability
MCHO	Mobile-controlled handoff
MDS	Maximum distance separable
MEDS	Method of exact Doppler spreads
MIMO	Multiple-input multiple-output
ML	Maximum likelihood
MLSE	Maximum likelihood sequence estimation
MMRC	Macrodiversity maximal ratio combining
MMSE	Minimum mean square error
MPSD	Minimum product squared Euclidean distance
MRC	Maximal ratio combining
MRM	Multi-resolution modulation
MS	Mobile station
MSE	Mean square error
MSK	Minimum shift keying
MSQ	Mean square
MTD	Minimum built-in time diversity
NCHO	Network-controlled handoff
NLoS	Non-line-of-sight
NMT	Nordic Mobile Telephone
NPP	Neighboring-cell pilot power
NTT	Nippon Telephone and Telegraph
OC	Optimum combining
OFDM	Orthogonal frequency division multiplexing
OFDMA	Orthogonal frequency division multiple access
OQASK	Offset quadrature amplitude shift keying
OQPSK	Offset quadrature phase shift keying
PDC	Personal digital cellular
PAM	Pulse amplitude modulation
PAPR	Peak-to-average power ratio
PCCC	Parallel concatenated convolutional code
pdf	Probability density function
PHS	Personal handy-phone system
PIR	Pilot-to-interference ratio
PLL	Phase locked loop
PN	Pseudo-noise
psd	Power spectral density
PSK	Phase shift keying
PSI-CELP	Pitch synchronous innovation-code excited linear prediction
QAM	Quadrature amplitude modulation
QCELP	Qualcomm code excited linear prediction
QPSK	Quadrature phase shift keying
QoS	Quality of service
RA	Rural area
RCELP	Relaxed code excited linear prediction
RI-OFDMA	Random interleaving orthogonal frequency division multiple access
RLS	Recursive least squares
rms	Root mean square
RPE-LTP	Regular pulse excitation-long-term prediction

RSC	Recursive systematic convolutional
RSSE	Reduced state sequence estimation
SC	Selective combining
SCCC	Serial concatenated convolutional code
SC-FDE	Single-carrier frequency domain equalization
SC-FDMA	Single-carrier frequency division multiple access
SC-OFDMA	Spaced-carrier orthogonal frequency division multiple access
SFH	Slow frequency hopping
SIMO	Single-input multiple-output
SINR	Signal-to-interference-plus-noise ratio
SISO	Single-input single-output
SM	Spatial modulation
SNR	Signal-to-noise ratio
SoS	Sum-of-sinusoids
SSC	Switch and stay combining
SSK	Space shift keying
STTC	Space-time trellis code
TCM	Trellis coded modulation
TDAP	Time domain aligned pilots
TDD	Time division duplexing
TDM	Time division multiplexing
TDMA	Time division multiple access
TDSP	Time domain staggered pilots
TDTD	Time division transmit diversity
TFM	Tamed frequency modulation
TSTD	Time switched transmit diversity
TU	Typical urban
ULA	Uniform linear array
UMTS	Universal mobile telecommunications system
US	Uncorrelated scattering
VHF	Very high frequency
VSELP	Vector sum excited linear prediction
VoIP	Voice over Internet Protocol
WCDMA	Wideband code division multiple access
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless local area network
WPAN	Wireless personal area network
WSS	Wide sense stationary
WSSUS	Wide sense stationary uncorrelated scattering
ZCR	Zero crossing rate
ZF	Zero-forcing

Chapter 1

Introduction

Commercial wireless systems and services have undergone rapid development and deployment, since first generation cellular telephone systems were introduced in the early 1980s. These first generation (1G) cellular telephone systems were based on analog frequency modulation (FM) technology and designed to carry narrow-band circuit-switched voice services. The first generation cellular service providers seen an exponential growth rate in their subscriptions, and by the late 1980s capacity limits were already reached in the largest markets with 1G cellular systems. In response to such heavy demand, second generation (2G) digital cellular systems were developed and introduced in the early 1990s and their evolutions are still in operation today. These 2G cellular systems were/are based on either time division multiple access (TDMA) or code division multiple access (CDMA) technologies, and were initially designed to carry circuit-switched voice and data. During the 1990s, these 2G systems were enhanced to provide packet-switched data in addition to circuit-switched voice. These transitional 2G cellular systems with their enhanced data transmission capabilities later became known as 2.5G systems. Third generation (3G) cellular systems were introduced after the year 2000 that allowed simultaneous use of speech and data services and still higher data rates. These higher data rate capabilities were supplemented by geolocation information, giving rise to location dependent services. Currently, fourth generation (4G) cellular systems are deployed that use voice over Internet Protocol (VoIP) and multimedia applications with broadband access. These 4G systems are based on multi-carrier modulation/multiplexing techniques such as orthogonal frequency division multiple access (OFDMA), or advanced single-carrier modulation/multiplexing techniques such as single-carrier frequency division multiple access (SC-FDMA). Fifth generation (5G) wireless systems are currently under research and development.

1.1 Brief History of Wireless Systems and Standards

Although this textbook is intended to address the fundamentals of wireless communications, it is nevertheless useful to have some basic knowledge of the history and evolution of wireless systems and standards. The following gives a brief description of the major standards that have been developed or are under development for cellular radio systems, cordless phone systems, and wireless local and personal area networks.

1.1.1 First Generation (1G) 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 incompatible first generation (1G) cellular systems based on frequency division multiple access (FDMA) and analog FM technology. With FDMA there is a single traffic channel per radio frequency carrier. When a user accesses the network 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, an arrangement of filters that isolates the forward and reverse link channels, thus preventing a radio transceiver from jamming itself.

Table 1.1 First generation (1G) cellular standards

Feature	NTT	NMT	AMPS
Frequency band	925–940/870–885	890–915/917–950	824–849/869–894
RL/FL ^a (MHz)	915–918.5/860–863.5 922–925/867–870		
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

^aRL = reverse link, FL = forward link

^bfrequency interleaving using overlapping channels, where the channel spacing is half the nominal channel bandwidth

Table 1.2 Second generation (2G) digital cellular standards, GSM and IS-54/136

Feature	GSM/DCS1800/PCS1900	IS-54/136
Frequency band	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

^aRL = reverse link, FL = forward link

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, IL. Many 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. All 1G cellular systems are now extinct.

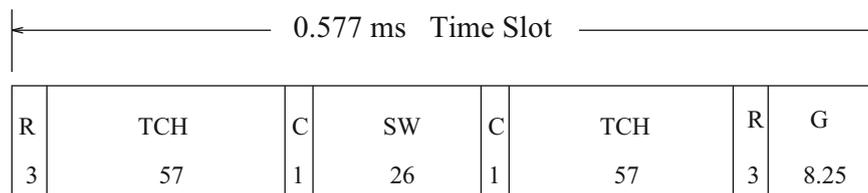
1.1.2 Second Generation (2G) Cellular Systems

Second generation (2G) digital cellular systems were developed in the 1980s and early 1990s, and widely deployed throughout the world in the 1990s. These included the GSM/DCS1800/PCS1900 standard in Europe, the Personal Digital Cellular (PDC) standard in Japan, and the IS-54/136 and IS-95 standards in the USA. Major parameters of the air interface specifications of these standards are summarized in Tables 1.2 and 1.3, and a very brief description of each is provided below.

Table 1.3 Second generation (2G) digital cellular standards, PDC and IS-95

Feature	PDC	IS-95
Frequency band	810–826/	824–829/
RL/FL ^a	940–956	869–894
(MHz)	1429–1453/	1930–1990/
	1477–1501	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)	QCELP (8,4,2,1)
	VSELP (FR 6.7)	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

^aRL = reverse link, FL = forward link



- R Guard time for burst transient response (Ramp time)
- TCH Traffic channel
- C Control bit
- SW Synchronization word
- G Guard bits

Fig. 1.1 Time slot format for GSM. Units are in bits

1.1.2.1 GSM

European countries seen the deployment of incompatible 1G cellular systems that did not admit roaming throughout Europe. As a result, the Conference of European Postal and Telecommunications Administrations (CEPT) established Groupe Spécial Mobile (GSM) in 1982 with the mandate of defining standards for future Pan-European cellular radio systems. The GSM 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.

GSM can support full rate (8 slots/carrier) and half rate (16 slots/carrier) voice operation, and provide various synchronous and asynchronous data services at 2.4, 4.8, and 9.6 kb/s. 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. 1.1. Variants of GSM have also been developed to operate in the 900 MHz and 1800 MHz bands in Europe, and the 850 MHz and 1900 MHz bands in North America.

GSM Release ’97 added packet data capabilities by aggregating all time slots together for a single user. This enhancement provides data rates up to 140 kb/s and is called General Packet Radio Service (GPRS). GSM Release ’99 introduced higher speed data transmission by using a higher-level 8-PSK modulation format (up to 473.6 kb/s with uncoded 8-PSK). This enhancement is called Enhanced Data Rates for GSM Evolution (EDGE). Some parameters of the EDGE standard are shown in Table 1.4. EDGE was deployed worldwide except for Japan and South Korea. GPRS and EDGE are generally branded as 2.5G systems.

Table 1.4 Parameters of the EDGE standard

Multiple access	TDMA
Duplexing	FDD
Carrier spacing (kHz)	200
Modulation	8-PSK/GMSK
Frame length (ms)	4.615
Slots/Frame	8
Maximum bit rate (kb/s)	473.6 (8-PSK)/140.8 (GMSK)

GSM has been a phenomenal success and is one of the most widely deployed cellular standards. By late 1993, over a million subscribers were using GSM phone networks. By 2013, GSM had over 6.5 billion subscribers across 219 countries and territories with roaming everywhere except for Korea and Japan. However, some cellular operators discontinued GSM service starting in January 2017.

1.1.2.2 IS-54/136 and IS-95

In North America the primary driver for second generation systems was the capacity crunch felt by some AMPS operators in the largest US markets by the late 1980s. One of the key objectives established by the Cellular Telephone Industry Association (CTIA) at the time was a ten-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 led to the development of dual-mode cellular standards in North America.

While Europe seen a convergence to the GSM standard based on TDMA technology, North America saw a divergence to two second generation digital cellular standards, IS-54/136 and IS-95, based on TDMA and CDMA technology, respectively. The IS-54 standard, adopted in 1990, was based on TDMA with 30 kHz carrier spacings (the same as AMPS) and $\pi/4$ phase-shifted quadrature differential phase shift keyed ($\pi/4$ -DQPSK) modulation with a raw bit rate of 48.6 kb/s [104]. IS-54 and IS-136 differed in the control channel; IS-54 used an analog control channel, whereas IS-136 used a digital control channel. The IS-54/136 air interface specified 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. 1.2. IS-54/136 was once deployed widely in the USA and Canada during the 1990s, but its use was discontinued in the 2007–2009 time frame in favor of GSM/GPRS/EDGE.

Just after the CTIA adopted IS-54 in 1990, another second generation digital cellular standard was proposed by Qualcomm based on CDMA technology. The introduction of IS-95 proposal saw considerable debate and spirited exchanges over the relative capacity and merits of TDMA and CDMA cellular systems at the time. Initial capacity claims for IS-95 were 40 times AMPS. In March 1992, CDMA was adopted as the IS-95 standard [105]. Initial commercial deployments in Korea showed a capacity gain of 6–10 times AMPS. The introduction of IS-95 CDMA cellular was of historical significance, because 3G cellular systems are based on CDMA technology.

With IS-95, the basic user data rate is 9.6 kb/s for Rate Set 1 (RS1) and 14.4 kb/s for Rate Set 2 (RS2), which is spread by using a pseudo-noise (PN) sequence with a chip rate of 1.2288 Mchips/s. The forward channel supports coherent detection by using an unmodulated pilot channel for channel estimation. Information on the forward link is encoded by using a rate-1/2 convolutional code, interleaved, spread by using one of the 64 orthogonal Walsh codes, and transmitted in 20 ms bursts. Each mobile station (MS) in a cell is assigned a different Walsh code, thus providing complete orthogonality under ideal channel conditions. Final spreading with a PN code of length 2^{15} , having a phase offset that depends on each base station (BS), is used to mitigate the multiple access interference to and from other cells.

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 function properly, all signals to be recovered must be received with (nearly) the same power. 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 the radio propagation environment. The information on the IS-95 reverse link is encoded by using a rate-1/3 convolutional code, interleaved, and mapped onto one of the 64 Walsh codes. Unlike the IS-95 forward channel that uses Walsh codes for spreading, the reverse link uses Walsh codes for 64-ary orthogonal modulation. The BS receiver uses non-coherent detection, since no pilot signal is transmitted on the reverse link. Both the BSs and the MSs use RAKE receivers to provide multipath diversity. To ensure that the power control algorithm is stable, CDMA cellular systems must use soft handoff, where the MS maintains a radio link with multiple BSs when traversing between cells and softer handoff when traversing between sectors of the same cell.

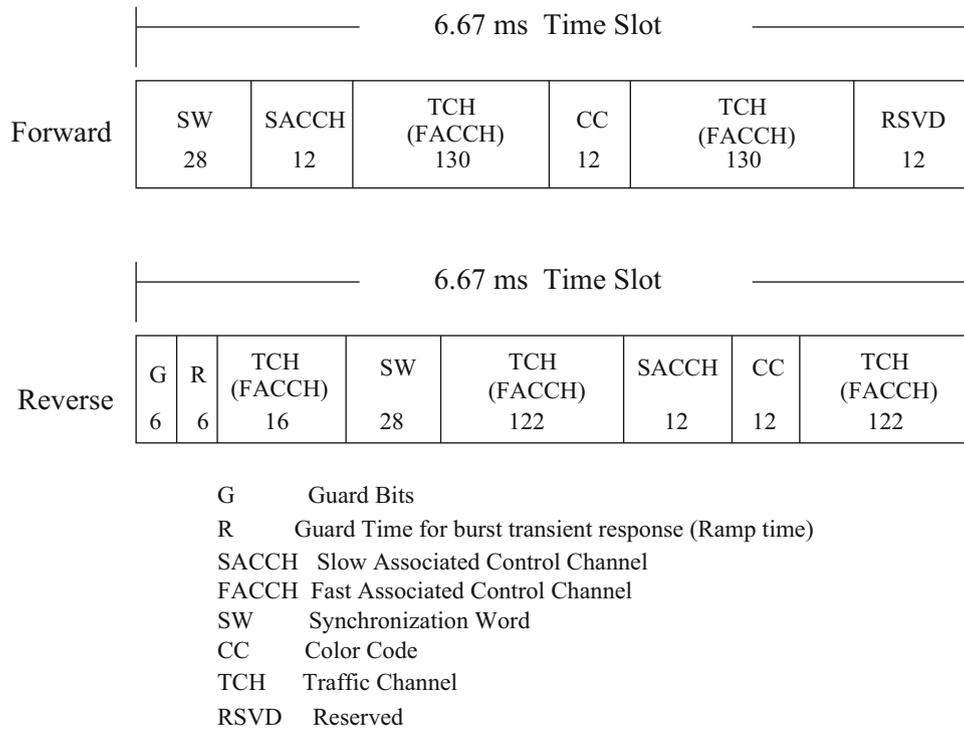


Fig. 1.2 Burst format for IS-54/136 traffic channel. Units are in bits

In 1998, the IS-95B standard was approved to support packet switched data with rates up to 115.2 kbps by using multi-code CDMA, where up to 8 Walsh codes are aggregated and assigned to a single user in a dynamic and scheduled manner.

1.1.2.3 PDC

In 1991, the Japanese Ministry of Posts and Telecommunications standardized Personal Digital Cellular (PDC). The air interface of PDC is similar 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. 1.3. 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. 1.2. 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 provides spatial diversity under these conditions. More details on the PDC system can be found in the complete standard [282]. PDC reached a peak of nearly 80 million subscribers, but the standard became extinct in 2012 in favor of 3G cellular technologies.

1.1.3 Third Generation (3G) 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 was developed by the International Telecommunications Union Radio Communications (ITU-R) and Telecommunications (ITU-T) sectors. Various standards bodies around the world provided inputs to the IMT-2000 standard definition. IMT-2000 was envisioned as an ubiquitous wireless system that could support voice, multimedia, and high speed data communication. The ITU provided no clear definition of the minimum or average rates users could expect from 3G equipment or providers. However, it was generally expected at the time that 3G

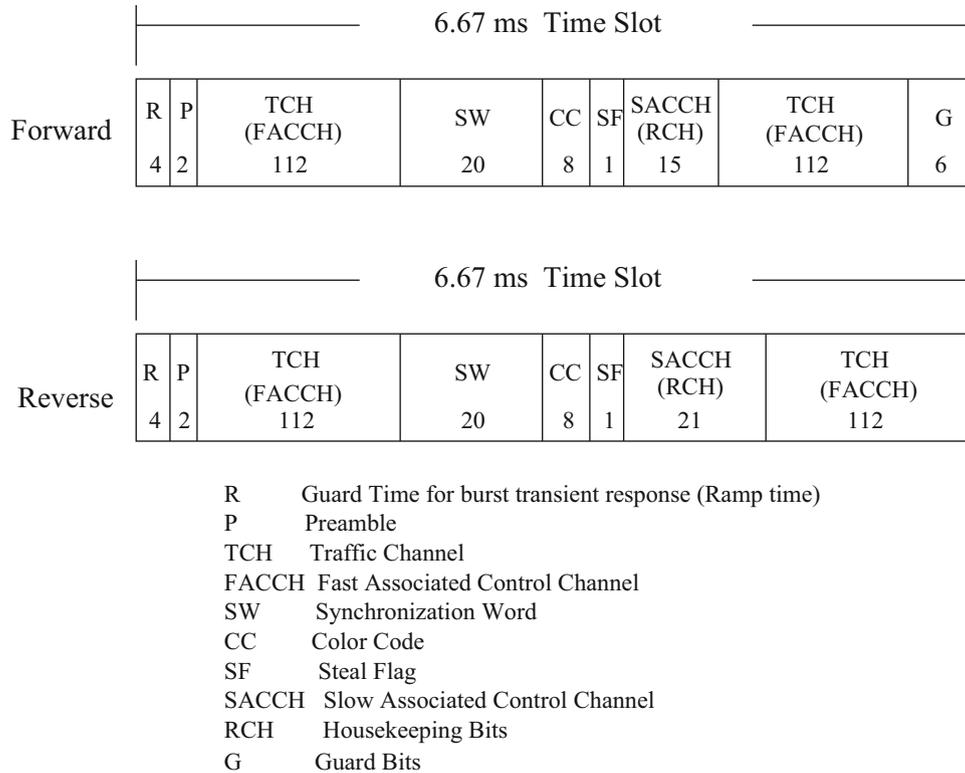


Fig. 1.3 Time slot format for Japanese PDC. Units are in bits

networks would provide a minimum downlink peak data rate of 2 Mbit/s for stationary or walking users, and 384 kbit/s in a moving vehicle.

IMT-2000 was actually a family of standards. Two of the standards were based on TDMA approaches, namely EDGE and Digital Enhanced Cordless Telephone (DECT). While the EDGE standard fulfilled the requirements for IMT-2000, EDGE networks are typically branded as 2.5G networks rather than 3G networks. The most predominant forms of IMT-2000 are cdma2000 developed by 3GPP2 and the Universal Mobile Telecommunications System (UMTS) family of standards, which includes Wideband Code Division Multiple Access (WCDMA), developed by 3GPP. Sometimes WCDMA is used synonymously with UMTS. Mobile WiMAX (Worldwide Interoperability for Microwave Access), developed by the IEEE802.16 working group, is also included under the IMT-2000 umbrella as a 3.5G standard, although it is sometimes referred to as a 4G standard for branding purposes. WiMAX is a multi-carrier scheme based on orthogonal frequency division multiple access (OFDMA).

Table 1.5 summarizes the main parameters for WCDMA and cdma2000. The common attributes of WCDMA and cdma2000 include the following:

- Provision of multi-rate services
- Packet data services
- Complex spreading
- A coherent uplink using a user dedicated pilot
- Additional pilot channel in the downlink for beam forming
- Seamless inter-frequency handoff
- Fast forward link power control
- Optional multi-user detection.

The major differences between WCDMA and cdma2000 center around the chip rate that is used, and synchronous (cdma2000) vs. asynchronous (WCDMA) network operation. Synchronous operation with cdma2000 is achieved by using a global positioning system (GPS) clock reference.

Table 1.5 Parameters for WCDMA and cdma2000

Feature	WCDMA	cdma2000
Multiple access	DS-CDMA	DS-CDMA
Chip rate (Mcps)	3.84	1.2288
Carrier spacing (MHz)	5	1.25
Frame length (ms)	10	5/20
Modulation	FL: QPSK	FL: BPSK/QPSK
	RL: BPSK	RL: BPSK
Coding		64-ary orthogonal
	Rate-1/2, 1/3	Rate-1/2, 1/3, 1/4,
	$K = 9$ CC	1/6 $K = 9$ CC
	Rate-1/3	Rate-1/2, 1/3, 1/4,
	$K = 4$ turbo code	1/5, $K = 4$ turbo code
Interleaving	Inter/Intraframe	Intraframe
Spreading	FL: BPSK	Complex
	RL: QPSK	
Inter BS synchronization	Asynchronous	Synchronous

1.1.3.1 cdma2000

The cdma2000 family of standards developed by 3GPP2 evolved from IS-95A/B and included cdma2000 1x, cdma2000EV-DO Rev. 0, cdma2000 EV-DO Rev. A, and cdma2000 EV-DO Rev. B. cdma2000 1x, also known as 1x and 1xRTT, is the core cdma2000 wireless air interface standard, and was recognized by the ITU as an IMT-2000 standard in November 1999. It was the first IMT-2000 technology deployed worldwide, in October 2000. The designation “1x” stands for 1 times Radio Transmission Technology, and means that the system has the same RF bandwidth as IS-95: a duplex pair of 1.25 MHz radio channels. cdma2000 1x almost doubles the capacity of IS-95 by adding 64 more traffic channels to the forward link, orthogonal to (in quadrature with) the original set of 64 forward channels. The cdma2000 1x Release 0 standard supports bi-directional peak data rates of up to 153 kbps and Release 1 can achieve peak data rates of up to 307 kbps in a single 1.25 MHz channel. cdma2000 1x supports a variety of applications including circuit-switched voice, short messaging service (SMS), multimedia messaging service (MMS), games, GPS-based location services, music and video downloads.

EV-DO which now stands for “Evolution-Data Optimized” was initially developed by Qualcomm in 1999 to meet the IMT-2000 requirements for a minimum 2-Mbit/s downlink speed for stationary or walking users, and is sometimes referred to as IS-856. EV-DO uses a mixture of CDMA and TDMA to maximize both individual user’s throughput and the overall system throughput. EV-DO was typically used for broadband Internet access and offered speeds of 2.4 Mbits/s downlink and 153 kbit/s uplink. EV-DO Rev A was introduced to offer higher peak data rates of 3.1 Mbits/s downlink and 1.8 Mbit/s uplink. EV-DO Rev A implemented multi-flow packet application (MFPA), where multiple application data flows can be supported that are assigned different priorities according to their associated quality of service (QoS) profiles. EV-DO Rev A has been adopted by many service providers in the USA, Canada, Mexico, Europe, Asia, Russia, Brazil, and Australia. As of 2014, there were 175 EV-DO Rev A networks.

1.1.3.2 UMTS

UMTS was developed by 3GPP and is part of the global ITU IMT-2000 standard. The most common form of UMTS uses WCDMA (IMT Direct Spread) as the underlying air interface. However, UMTS also includes the TD-CDMA and TD-SCDMA (both IMT CDMA TDD) air interfaces. The first deployment of UMTS was the Release’99 architecture. UMTS required new base stations and new frequency allocation over 2G cellular systems. However, it is closely related to GSM/EDGE as it borrows and builds upon concepts from GSM. Further, most UMTS handsets also support GSM, allowing seamless dual-mode operation. WCDMA uses paired 5 MHz channels and can support peak data rates of up to 384 kbit/s, both uplink and downlink, for Release’99 handsets. WCDMA systems were criticized for their large (5 MHz) bandwidth requirement, which delayed deployment in countries (such as the USA) that acted relatively slowly in allocating new frequencies specifically for 3G services.

Starting in 2006, UMTS networks in many countries were upgraded to include High Speed Packet Access (HSPA), sometimes known as 3.5G. High Speed Downlink Packet Access (HSDPA) enables peak downlink transfer speeds of 14 Mbit/s and High Speed Uplink Packet Access (HSUPA) has peak data rates of 5.8 Mbit/s in the uplink, although most initial HSDPA network deployments had peak downlink speeds of 7.2 Mbps. As of 2014, HSPA was commercially deployed by 550 operators in 203 countries. Evolved HSPA (also known as HSPA+) was defined in 3GPP Release 7 and 8 of the WCDMA specification. In WCDMA Release 11, HSPA+ can theoretically reach up to 337 Mbit/s in the downlink and 34 Mbit/s in the uplink by using multiple-input multiple-output (MIMO) technologies and high-order modulation schemes. However, these data speeds are rarely realized and most HSPA+ networks provide for 168 Mbit/s in the downlink and 22 Mbit/s in the uplink. As of 2014, there were 365 HSPA+ networks in 157 countries.

1.1.4 Fourth Generation (4G) Cellular Systems

Fourth generation cellular systems have capabilities defined by the ITU in IMT-Advanced. IMT-Advanced specifies peak data rates of 100 Mbps in high mobility applications and 1 Gbps in stationary or low mobility applications. IMT-Advanced has the following requirements:

- Flexible channel bandwidth, between 5 and 20 MHz, optionally up to 40 MHz.
- A nominal peak data rate of 100 Mbps in high mobility, and 1 Gbps for stationary environments.
- A data rate of at least 100 Mbps between any two points in the world.
- Bandwidth efficiency of up to 15 bit/s/Hz in the downlink, and 6.75 bit/s/Hz in the uplink.
- Spectral efficiency of up to 3 bit/s/Hz/cell in the downlink.
- Smooth handoff across heterogeneous networks.
- Seamless connectivity and global roaming across multiple networks.
- High quality of service (QoS) for next generation multimedia support.
- Backward compatibility with existing wireless standards.
- All Internet Protocol (IP) packet-switched network.

Two candidate 4G systems have been standardized and deployed commercially, (1) Long Term Evolution (LTE) as standardized by 3GPP, and (2) IEEE 802.16e (Mobile WiMAX) as standardized by the IEEE. However, it is debatable as to whether or not these first versions should be considered 4G standards at all, since they do not meet the IMT-Advanced requirements. Nevertheless, LTE and Mobile WiMAX are often branded as being 4G systems as they are technical precursors to their more advanced versions Long Term Evolution Advanced (LTE-A) and Mobile WiMAX Release 2 (IEEE 802-2012). Unlike the 3G cellular systems that use CDMA technology, the 4G cellular system proposals use multi-carrier methods, namely orthogonal frequency division multiple access (OFDMA) and single carrier frequency division multiple access (SC-FDMA).

Long Term Evolution (LTE) and Long Term Evolution—Advanced (LTE-A) have been commercially deployed, and LTE-A Pro has seen initial deployment. LTE-A Pro is a new radio interface for low data rate narrow-band Internet of Things (NB-IoT). As of 2016Q2, a total of 521 operators have commercially launched LTE, LTE-A, or LTE-A Pro networks in 170 countries. Of these, 147 mobile operators have launched LTE-Advanced or LTE-Advanced Pro networks in 69 countries with 9 LTE-Advanced Pro networks deployed. During 2016 Q2, the number of LTE subscribers worldwide doubled to 1.29 billion. By the end of 2016, it is expected that there will be 560 LTE, LTE-A, or LTE-A Pro networks. It is also expected that 1-Gbps LTE services could be introduced before the end of 2016. Operators in all regions are also ramping up the deployment of voice-over-LTE (VoLTE) to support voice-over-Internet (VoIP). A total of 82 operators in 43 countries have so far launched commercial VoLTE services. As many LTE operators have relied on circuit switch fall back to GSM to accommodate voice calls, LTE operators will start discontinuing GSM as VoLTE is deployed.

WiMAX has also been commercially deployed. In 2014, there were 580 commercial WiMAX networks in 149 countries, with 33.4 million subscribers worldwide. However, since 2015 the number of WiMAX subscribers has been declining worldwide while the number of LTE subscribers is increasing rapidly. Some operators are ending their WiMAX service and deploying LTE, LTE-A, or LTE-A Pro networks instead.

Table 1.6 Basic parameters for LTE and LTE-A

Feature	LTE	LTE-A
Multiple access	FL: OFDMA	FL: OFDMA
	RL: SC-FDMA	RL: single-carrier property
		not preserved for SC-FDMA uplink
Carrier spacing (MHz)	1.4, 3, 5, 10, 15, 20	Additionally supports up to 100 MHz downlink and 40 MHz uplink with carrier aggregation
Duplexing	FDD, TDD, half-duplex FDD	FDD, TDD, half-duplex FDD
Mobility (km/hr)	500	500
Frame length (ms)	1	1
Modulation	FL: QPSK, 16QAM, 64QAM	FL: QPSK, 16QAM, 64QAM, 256QAM
	RL: QPSK, 16QAM, 64QAM(optional)	RL: QPSK, 16QAM, 64QAM(optional)
Coding	Turbo code	Turbo code
MIMO	FL: 2×2 , 4×2 , 4×4	FL: up to 8×8
	RL: 1×2 , 1×4	RL: up to 4×4
Peak data rate	FL: 150 Mbps(2×2 MIMO, 20 MHz), 300 Mbps(4×4 MIMO, 20 MHz)	FL: 3 Gbps (8×8 MIMO, 100 MHz)
	RL: 75 Mbps(20 MHz)	RL: 500 Mbps (4×4 MIMO, 40 MHz)
Latency (ms)	≈ 10	< 5

1.1.4.1 LTE, LTE-A, and LTE-A Pro

LTE is an all IP technology that uses OFDMA on the downlink and SC-FDMA on the uplink. LTE is compatible with previous mobile technologies including GSM, GPRS, UMTS, EDGE, WCDMA, HSPA, cdmaOne, cdma2000, and EV-DO. LTE-A builds on LTE to further increase data rate as defined in 3GPP Releases 10 and 11. There are five major features: carrier aggregation, increased MIMO, coordinated multipoint transmission or cooperative MIMO, heterogeneous network (HetNet) support, and relays. Carrier aggregation combines up to five 20 MHz channels into one to increase data speed. The channels can be contiguous or non-contiguous. With a maximum 8×8 MIMO configuration, 64QAM, and 100 MHz bandwidth, a peak downlink data rate of 1 Gbit/s is possible. LTE-A and LTE-A Pro are the only truly 4G technologies. The basic parameters of LTE, LTE-A are summarized in Table 1.6. LTE-A Pro will have the following key attributes: (1) Data speeds in excess of 3 Gbps (LTE-A: 1 Gbps), (2) 640 MHz of carrier bandwidth (LTE-A: 100 MHz), (3) Latency: 2 ms (LTE-A: 10 ms).

1.1.4.2 WiMAX

WiMAX (Worldwide Interoperability for Microwave Access) is a telecommunications protocol that provides fixed and fully mobile Internet access. There are several versions of the WiMAX standard. IEEE 802.16-2004, also known as 802.16d, is sometimes referred to as “Fixed WiMAX,” since it does not support mobility. IEEE 802.16e-2005, often abbreviated as 802.16e, includes support for mobility among other things and is commonly known as “Mobile WiMAX.” Mobile WiMAX can deliver mobile broadband services, with peak data rates up to 40 Mbit/s, at vehicular speeds greater than 120 km/hr while maintaining a quality of service (QoS) comparable to broadband wireline access.

Some of the key features and attributes of WiMAX include the following:

- Tolerance to delay spread and multi-user interference due to orthogonality of OFDMA sub-carriers in both the downlink (DL) and uplink (UL) directions.
- Scalable channel bandwidths ranging from 1.25 to 20 MHz through adjustment of the Fast Fourier Transform (FFT) size in the baseband modulator/demodulator. Supported FFT sizes are 128, 256, 512, 1024, 2048.
- Hybrid-Automatic Repeat Request (H-ARQ) to provide robustness in high mobility environments.

- Adaptive sub-carrier allocation (in time and frequency) to optimize connection quality based on relative signal strengths on a connection-by-connection basis.
- Advanced modulation and coding schemes that use BPSK, QPSK, 16-QAM, 64-QAM together with convolutional and turbo coding.
- Power management to ensure power efficient operation of mobile and portable devices in sleep and idle modes.
- Network-optimized hard handoff to minimize overhead and achieve a handoff delay of less than 50 ms.
- Advanced antenna systems including MIMO, beam forming, space-time coding, and spatial multiplexing.
- Fractional frequency reuse to achieve high spectral efficiency.

1.1.5 Fifth Generation (5G) Cellular and Beyond

5G wireless networks are currently being researched and developed, and are expected to be deployed between 2020 and 2030. The objectives of 5G cellular systems are varied. With the proliferation of wireless data and the Internet of Things (IoT), 5G wireless systems are sought that will support 100–1000-fold gains in network sum capacity, provide connections for at least 100 billion devices, low-data-rate machine-to-machine communication, offer 10 Gbps peak data rates with sub-millisecond latencies and response times, and provide for very high reliability. Low latency and extremely high reliability will be essential for mobile industrial automation, vehicular connectivity and automated driving, and other IoT applications. 5G radio access will be built upon evolved existing wireless technologies, particularly LTE-A, LTE-A Pro, and evolved WiFi, while also introducing new radio access means, particularly mm-wave cellular. The basic premises for 5G cellular are massive connectivity and massive capacity, a growing variety of low and high data rate applications with grossly different QoS requirements, and radio interfaces that simultaneously operate in different frequency bands. Finally, energy-per-bit usage should be reduced by a factor of 10–100 to improve upon wireless device battery life.

The demanding requirements of 5G systems will require innovations in several key technology areas relating to future wireless communication, including but not limited to

- Ultra dense networks using hierarchical cellular architectures, simultaneous network participation
- Ultra reliable and low latency communication
- Operation at higher frequencies, particularly at mm-wave frequencies
- Advanced coding, modulation, and multi-access techniques
- Advanced radio resource management protocols for interference management
- Moving networks, such as on trains, planes, and buses
- Massive MIMO
- Device-to-device communication
- Massive machine communications in IoT
- Low power communications
- Software defined networking and cloud-based systems
- Robust security.

1.1.6 Cordless Telephone Systems

Similar to cellular telephones, first generation cordless telephones were based on analog frequency modulation technology. After their introduction, cordless telephones gained high popularity, which made them victims of their own success; the voice quality was often unacceptable due to uncoordinated deployment and resulting mutual interference between cordless phones operating on the same frequency. This led to the development of digital cordless telephones. The most predominant cordless phone standard is DECT. DECT originated in Europe, where it is the universal standard, replacing earlier cordless phone standards. It has been adopted by Australia, and most countries in Asia and South America. Adoption in the USA was delayed due to radio licensing regulations, and earlier technologies are still competitive. In the USA, DECT operates in the 1920–1930 MHz, or 1.9 GHz band, and is branded as DECT 6.0. DECT is recognized by the ITU as fulfilling the IMT-2000 requirements and, thus, it actually qualifies as a 3G system. The major technical properties of DECT are described in Table 1.7.

Table 1.7 Cordless telephone standards

Feature	DECT	PHS
Frequency band (MHz)	1880–1900 (Europe)	1895–1918
	1900–1920 (China)	
	1910–1930 (Latin America)	
	1920–1930 (United States, Canada)	
Multiple access	F/TDMA	F/TDMA
Duplexing	TDD	TDD
Carrier spacing (kHz)	1728	300
Modulation	GFSK	$\pi/4$ -DQPSK
Number of carriers	10 (Europe)	77
Number of carriers	5 (United States)	
Channels/Carrier	12	4
Bit rate (kb/s)	1152	384
Speech coding	ADPCM	ADPCM
	32 kb/s	32 kb/s
Frame size (ms)	10	5
Mean TX power (mW)	10 (Europe)	10
	4 (United States)	
Peak TX power (mW)	250 (Europe)	80
	100 (United States)	

Table 1.8 2.4 and 5 GHz bands for license exempt use. $B = -26$ dB emission bandwidth in MHz

Location	Frequency	Maximum output
	Range (GHz)	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 $+10\log_{10}B$

The Personal Handy-phone System (PHS) is a mobile network system developed by NTT Laboratory in Japan in 1989. It is deployed mainly in Japan, China, Taiwan, and some other Asian countries. PHS is a cordless telephone like DECT, with the capability to handover from one cell to the next. PHS operates in the 1880–1930 MHz frequency band, and is far simpler to implement and deploy than cellular systems like PDC or GSM. However, the PHS cells are small due to a maximum base station transmit power of 500 mW, and cell radii are typically in the order of tens or at most hundreds of meters in non-line-of-sight conditions. While the small cell size makes PHS suitable for dense urban areas, it is impractical for rural areas, and the small cell size also makes it difficult if not impossible to make calls from moving vehicles. For this reason PHS has seen declining popularity in Japan. Nevertheless, PHS has seen a resurgence in markets like China, Taiwan, Vietnam, Bangladesh, Nigeria, Mali, Tanzania, and Honduras where its low deployment costs make it attractive to operators. The major technical properties of PHS are described in Table 1.7.

1.1.7 Wireless LANs

A variety of wireless local area network (WLAN) standards have been developed to operate in unlicensed bands. Table 1.8 lists some of the unlicensed bands that are used in various regions of the world.

Table 1.9 Key parameters of the IEEE 802.11a OFDM standard, from [328]

Bandwidth	20 MHz
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 sub-carriers	52
Number of pilots	4
OFDM symbol duration	4 μ s
Guard interval	800 ns
Sub-carrier spacing	312.5 kHz
3 dB bandwidth	16.56 MHz

Table 1.10 Key parameters of the IEEE 802.11ac OFDM standard

Bandwidth	Mandatory 20, 40, and 80 MHz channel bandwidth, 160 MHz (optional)
Data Rate	86.7–6933 Mb/s depending on configuration
Modulation	BPSK, QPSK, 16-QAM, 64-QAM, 256-QAM
Coding	Convolutional code 1/2, 2/3, 3/4, 5/6, or LDPC (optional)
MIMO	up to 8×8 MIMO providing 8 spatial streams single or multi-user MIMO
FFT size	64, 128, 256, and 512 (optional)
Data sub-carriers/Pilots	52/4, 108/6, 234/8, and 468/16 (optional)
OFDM symbol duration	4 μ s, 3.6 μ s (optional)
Guard interval	800 ns and 400 ns (optional)
Sub-carrier spacing	312.5 kHz

In 1997, the IEEE 802.11 standards group established the first WLAN standard to provide 1 and 2 Mb/s aggregate data rates, commonly known as WiFi. IEEE 802.11 uses direct sequence spread spectrum with a length-11 Barker sequence for spreading, and either BPSK (1 Mbps) or QPSK (2 Mbps) peak data speeds. Barker sequences are discussed in Chap. 9. In 1998, the IEEE 802.11b standard was defined to provide 5.5 and 11 Mbps aggregate data rates. IEEE 802.11b uses complementary code keying (CCK), which is also described in Chap. 9. In 1998, the IEEE 802.11a standard was defined for operation in the newly unlicensed 5 GHz UNII bands in the USA. IEEE 802.11a is based on a combination of orthogonal frequency division (OFDM) and time division multiplexing (TDM), and can provide a range of aggregate data rates ranging from 6 to 54 Mbps. The parameters of the IEEE 802.11a OFDM standard are summarized in Table 1.9. The same network has been adopted as IEEE 802.11g for operation in the 2.4 GHz unlicensed band.

The IEEE802.11n standard was released in 2009 to provide up to 4 MIMO streams. Similar to IEEE802.11a/g, the IEEE802.11n standard uses OFDM and operates in either the 2.4 GHz or 5 GHz unlicensed bands. However, IEEE802.11n may use a bandwidths of either 20 MHz or 40 MHz. The standard allows up to 4 MIMO streams providing a data rate ranging from 54 Mbit/s to 600 Mbit/s.

The IEEE802.11ac standard was released in 2013 and builds on IEEE802.11n. Changes include the use of wider channel bandwidths 80 or 160 MHz versus 40 MHz, however, operating only in the 5 GHz band. IEEE802.11ac also added more spatial MIMO streams (up to eight vs. four), higher-order modulation (up to 256-QAM vs. 64-QAM), and the addition of multi-user MIMO (MU-MIMO). Data rates with IEEE802.11ac depend on the configuration and range from 7.2 Mbps to 6.9333 Gbps in the maximum configuration with 160 MHz bandwidth, 8 MIMO streams, and 256QAM modulation. Key parameters of the IEEE802.11ac standard are listed in Table 1.10.

The IEEE802.11ad standard is similar to IEEE802.11ac, except for operation in the 60 GHz mm-wave band. Radio propagation at 60 GHz suffers from higher path loss attenuation than at 5 GHz, so the range of IEEE802.11ad is roughly 10 m as compared to 100 m for IEEE802.11ac. The IEEE802.11ad standard is branded as WiGig.

IEEE 802.11af, branded as Super WiFi, was standardized in February 2014 to allow operation in TV white space in the VHF and UHF bands between 54 and 790 MHz. Due to the low path loss attenuation of VHF frequencies, the range of IEEE 802.11af can extend up to 1000 m. So as to coexist with commercial TV broadcasting and wireless microphones, IEEE 802.11af access points and stations determine their position using global positioning system (GPS) and query a geolocation database (GDB) over the Internet to discover what frequency channels are available for use at a given time and position. The physical layer is based on 802.11ac, however, the frequency channels are 6–8 MHz wide depending on the regulatory domain. Up to four TV channels may be bonded in either one or two contiguous blocks. MIMO operation is possible with

up to four streams used for either space-time block code (STBC) or multi-user (MU) operation. The achievable data rate per spatial stream is 26.7 Mbit/s for 6 and 7 MHz channels, and 35.6 Mbit/s for 8 MHz channels. With four spatial streams and four bonded channels, the maximum data rate is 426.7 Mbit/s for 6 and 7 MHz channels and 568.9 Mbit/s for 8 MHz channels.

1.1.8 Wireless PANs

A variety of wireless personal area network (WPAN) standards have been developed to operate in unlicensed bands. A WPAN is used for communication among devices such as computers and telephones that are in close proximity. The physical size of a WPAN is typically less than 10 m. WPANs can be used for communication among the devices themselves, or to connect to the Internet. A key feature WPAN technology is the ability to “plug-in” devices, such that when any two WPAN-equipped devices are in close proximity, they can communicate with each other. Another important feature is the ability of each device to lock out other devices selectively, thus preventing unauthorized access.

In 1999, the IEEE802.15 Working Group was created to define Wireless Personal Area Network (WPAN) standards. The Bluetooth v1.1 specification [153] was adopted as the IEEE 802.15.1-2002 standard and was later published as IEEE 802.15.1-2005 based upon the additions incorporated into Bluetooth v1.2. A Bluetooth WPAN is also called a piconet, and consists of up to 8 active devices connected in a master-slave configuration (others maybe in idle mode). The first Bluetooth device in the piconet is the master, and all other devices are slaves that communicate with the master. Bluetooth uses Frequency Hop CDMA (FH-CDMA) with a set of 79 hop carriers with a spacing of 1 MHz and a hop dwell time of 625 μ s. Classical Bluetooth uses Gaussian frequency shift keying (GFSK) with a modulation index of 0.3 and either a very simple rate-1/3 3-bit repetition code or a simple rate-2/3 shortened Hamming code. Classical Bluetooth supports a data rate of 1 Mbps. Extended data rate Bluetooth systems are available that use $\pi/4$ -DQPSK and 8-DPSK, giving 2 and 3 Mbps, respectively.

Other than Bluetooth, two approaches have been pursued in IEEE802.15 regarding personal area networks. These include IEEE 802.15.4: Low Rate WPAN and IEEE 802.15.3: High Rate WPAN. High rate WPAN is dormant, and only Low Rate WPAN is available in various open standard and propriety networks.

1.2 Frequency Reuse and the Cellular Concept

A cellular land mobile radio network is a collection of individual cells that are served by cell sites or base stations (BSs). Each BS covers a defined geographical area called a cell. By integrating the coverage of a plurality of cells, a cellular network provides radio coverage over a large geographic area. A group of BSs is sometimes called a location area, or a routing area. A cellular land mobile radio system has two basic functions; it must locate and track both active and idle subscribers or mobile stations (MSs), and it must always attempt to connect the MSs to the best available BSs. The former task is the subject of mobility management, and requires a location update procedure which allows a MS to inform the cellular network, whenever it moves from one location area to the next. The latter task is the subject of radio resource management and requires the continuous evaluation of the radio link quality with the serving BS(s), and the radio link qualities of alternate BSs. This monitoring is performed by a base station controller (BSC) or mobile switching center (MSC) that uses knowledge of the link quality evaluations on the forward and reverse channels, in addition to the system topology and traffic flow, to decide upon the best BS(s) to serve a particular MS.

A cellular land mobile radio system uses low power radio communication between a MS and a grid of BSs. Movement of the MS, however, leads to highly erratic radio link conditions, and careful monitoring and control are required to keep the radio link quality 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 prior 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 changes in link quality and compensate by changing the set of serving BSs and the allocation of BS and MS 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 link quality measurement and network control essentially determine overall link quality and the size and distribution of cells in 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 deployment of small cells known as microcells and picocells. However, the microcellular and picocellular propagation environment is also highly erratic and the radio links are more difficult to control due to the combination of small cell sizes and mobility. Distributed radio resource management algorithms are typically used to maintain acceptable link quality and high spectral efficiency.

Cellular systems are designed to have high spectral efficiency and offer ubiquitous service coverage. These systems require (1) effective cellular architectures, (2) fast and accurate link quality measurements, (3) rapid control in all types of environments, (4) installation of BSs to provide radio coverage virtually everywhere, and (5) power and bandwidth efficient air interface schemes that can mitigate the harsh effects of the propagation environment and tolerate high levels of interference. Since the radio links in high capacity cellular systems will interfere with each other due to frequency reuse, it is always desirable to maintain each radio link at a target QoS while using the lowest possible transmit power. This means that radio links should not significantly exceed their target QoS since doing so will cause unnecessary interference to other radio links.

Cellular mobile radio systems rely upon frequency reuse, where users in geographically separated cells simultaneously use the same carrier frequency and/or time slot. The cellular layout of a conventional macrocellular system is very often described by a grid of hexagonal cells or radio coverage zones. In reality the cells are neither hexagonal or circular, but instead are irregular and overlapping regions. The hexagon is a common choice for representing macrocellular coverage areas, because it closely approximates a circle and offers a wide range of tessellating frequency reuse cluster sizes. Tessellating frequency reuse clusters are those that will fit together without leaving any gaps. A tessellating reuse cluster of size N can be constructed if [256]

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

where i and j are non-negative integers, and $i \geq j$. It follows that the allowable hexagonal cluster sizes are $N = 1, 3, 4, 7, 9, 12, \dots$. Examples of 3-, 4-, and 7-cell reuse clusters are shown in Fig. 1.4. The reuse clusters are tessellated to form a frequency plan. A simplified 7-cell frequency reuse plan is shown in Fig. 1.5, where cells marked with the same letter label use identical sets of carrier frequencies.

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 cell radii R , where R is the distance from the center to the corner of a cell. For regular hexagonal cells, the reuse cluster size N and co-channel reuse factor D/R are related by (see Problem 1.2)

$$D/R = \sqrt{3N}. \quad (1.2)$$

For microcellular systems with lower BS antenna heights, regular hexagons may no longer be appropriate for approximating the radio coverage areas. 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 depends greatly upon the local topography. For example, the linear cells shown in Fig. 1.6 can be used

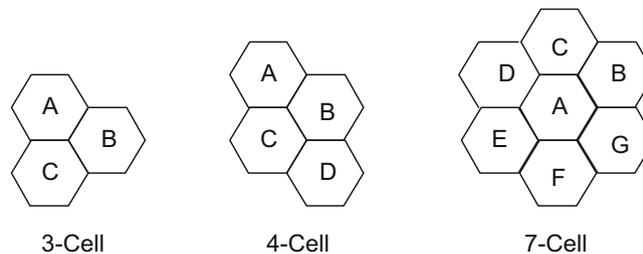


Fig. 1.4 Commonly used cellular reuse clusters

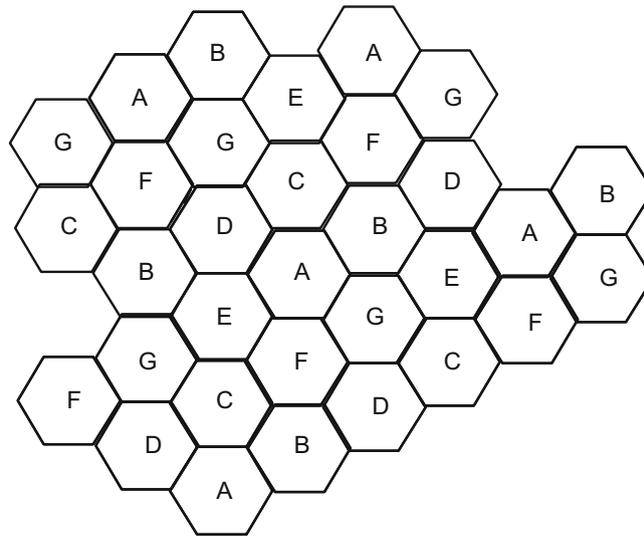


Fig. 1.5 Macrocellular deployment using 7-cell reuse pattern

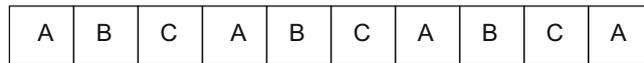
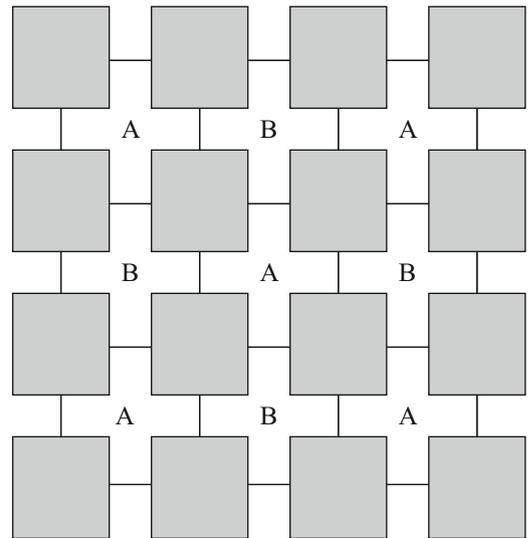


Fig. 1.6 Microcellular deployment along a highway with a 3-cell reuse pattern

Fig. 1.7 Microcellular deployment in a dense urban area, with a rectangular grid of streets. Base stations are deployed at every intersection with a 2-cell reuse pattern



to model microcells that are deployed along a highway with directional antennas. In a dense metropolitan area with urban canyons, the buildings act as wave guides to channel the radio waves along the street corridors. Figure 1.7 shows a typical “Manhattan” street cell deployment that can be used to model microcells that are deployed in city centers.

1.3 Mobile Radio Propagation Environment

Radio signals in cellular land mobile radio systems generally propagate according to the three mechanisms: reflection, diffraction, and scattering. Reflections arise when radio waves are incident upon a surface with dimensions that are large compared to the wavelength. The radio wave reflects off the surface with an angle of departure equal to the angle of incidence, while the amplitude and phase of the reflected wave depend on the surface characteristics. Diffraction occurs according to

the Huygens–Fresnel 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 radio 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. A good example occurs at millimeter wave frequencies, where rain drops cause scattering that manifests itself in a phenomenon called rain attenuation. The relative importance of these three propagation mechanisms depends on the particular propagation scenario.

As a result of the above three mechanisms, cellular land mobile radio propagation can be roughly characterized by three nearly independent phenomenon: path loss with distance, shadowing, and multipath-fading. Each of these phenomenon is caused by a different underlying physical principle. 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 a receiver antenna. Typically, the amplitude of a narrow-band received envelope can vary by as much as 30–40 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 transmission paths of different lengths and, therefore, reach the receiver antenna with different time delays. Time dispersion can be combatted and exploited by using time domain equalization in TDMA systems, RAKE receivers in CDMA systems, and frequency domain equalization in OFDM systems. Multipath-fading is discussed in detail in Chap. 2.

Free space propagation does not apply in a land mobile radio environment and the propagation path loss depends not only on the distance and frequency, but also on the antenna heights and topography. The simplest empirical land mobile radio propagation model assumes that the average received power at distance d from an emitter is

$$\Omega_p(d) = \frac{\mu_{\Omega_p(d_o)}\epsilon}{(d/d_o)^\beta}, \quad (1.3)$$

or after converting to decibel units

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

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 on the order of 1.6 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 describes the slope of the path loss characteristic (in dB) as a function of distance. The path loss 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. As mentioned earlier, path loss exponents are often determined by curve fitting to measured data. The above and other path loss models are the subject of Sect. 2.7.

The parameter ϵ is modeled a log-normally distributed random variable representing the effects of large scale fading otherwise known as shadowing. Shadowing is caused by large terrain features such as buildings, hills, and valleys. The random variable $\epsilon_{\text{(dB)}} = 10\log_{10}(\epsilon)$ in (1.4) is a zero-mean Gaussian random variable, and represents the error between the actual and predicted path loss at a given distance. For log-normal shadowing, the probability density function of $\Omega_{p \text{ (dBm)}}(d)$ has the normal distribution

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

where

$$\mu_{\Omega_{p \text{ (dBm)}}}(d) = \mu_{\Omega_{p \text{ (dBm)}}}(d_o) - 10\beta \log_{10}\{d/d_o\} \text{ (dBm)}, \quad (1.6)$$

is the average received signal power.

The parameter σ_Ω is called the shadow standard deviation. A more accurate path loss model will result in a smaller σ_Ω . However, there will always be some residual error between the actual and predicted path loss due to terrain irregularities. For macrocells σ_Ω typically ranges from 5 to 12 dB, with $\sigma_\Omega = 8$ dB being a commonly used 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. Figure 1.8 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 macrocellular environment. Shadowing is discussed in further detail in Sect. 2.6.

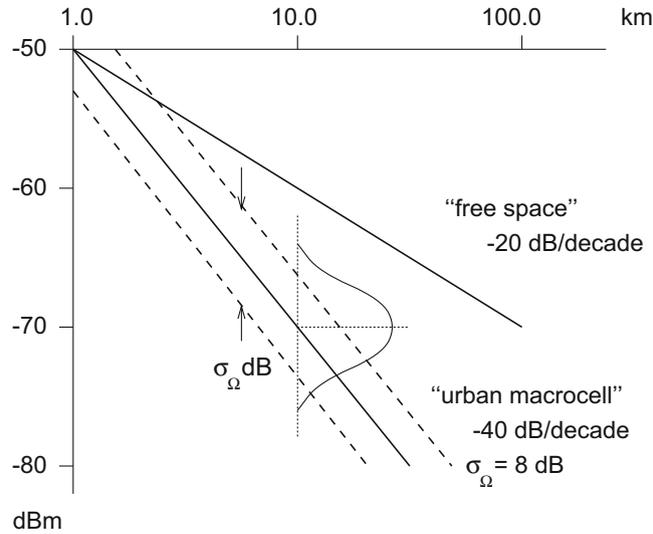


Fig. 1.8 Path loss in free space and typical macrocellular environments; $\beta = 4$, $\sigma_{\Omega} = 8$ dB. The received signal strength (in dBm units) at a distance of 10 km is Gaussian distributed with a mean of -70 dBm and a variance of σ_{Ω}^2 dB

1.4 Co-Channel Interference and Noise

Frequency reuse in cellular systems introduces co-channel interference (CCI), which is the primary factor that limits cellular spectral efficiency. CCI arises when the same carrier frequency is used in different cells and/or cell sectors. In this case, the power density spectra of the desired and interfering signals 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 case the power density spectrum of the desired and interfering signals partially overlap. However, since adjacent frequencies are used at close distances, the interference can still be significant. Consequently, the transmit power is regulated to fit within a regulatory or standard emission mask. Modulation schemes along with their power spectra are the subject of Chap. 4.

Radio links exhibit a link performance threshold according to a defined QoS criterion. For large cells carrying light traffic, the links are noise limited such that QoS is acceptable if received carrier-to-noise ratio, Γ , exceeds a certain defined threshold Γ_{th} . Otherwise, the QoS is unacceptable and a service outage occurs. The probability of outage due to thermal noise is defined as

$$O_N = P[\Gamma < \Gamma_{th}]. \quad (1.7)$$

For dense cellular deployments carrying heavy traffic, the links are typically interference limited such that the QoS is acceptable if the received carrier-to-interference ratio, Λ , exceeds a certain defined threshold Λ_{th} , and a service outage occurs otherwise. The CCI outage is defined as

$$O_I = P[\Lambda < \Lambda_{th}]. \quad (1.8)$$

CCI outage is the subject of Chap. 3.

The thresholds Γ_{th} and Λ_{th} depend on many parameters of the radio link, including the modulation and coding scheme that is employed, and the particular receiver processing algorithms that are implemented. Often the thresholds exhibit a cliff effect in terms of QoS, especially when powerful modulation and coding techniques are used.

1.5 Receiver Sensitivity and Link Budget

Receiver sensitivity refers to the ability of the receiver to detect signals in the presence of noise. The ratio of the desired carrier power to 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. A simplified link budget for cellular radio systems can be expressed in terms of the following parameters:

- Ω_t = transmit carrier power
- G_T = transmit antenna gain
- L_p = path loss
- G_R = receiver antenna gain
- Ω_p = receive signal power
- E_c = receive energy per modulated symbol
- T_o = receiving system noise temperature in degrees Kelvin
- B_w = receiver noise equivalent bandwidth
- N_o = thermal noise power spectral density
- R_c = modulation symbol rate
- $k = 1.38 \times 10^{-23}$ Ws/K = Boltzmann's constant
- F = noise figure, typically 5–6 dB
- L_{RX} = receiver implementation losses
- L_I = interference margin
- M_{shad} = shadow margin
- G_{HO} = handoff gain
- Ω_{th} = receiver sensitivity

Many other parameters may be included in a detailed link budget, such as cable losses, but they are not included here.

The average received carrier power (or local mean) can be expressed as

$$\Omega_p = \frac{\Omega_t G_T G_R}{L_{RX} L_p}, \quad (1.9)$$

where Ω_t is the transmitted power, and G_T and G_R are the transmitter and receiver antenna gains. Receiver implementation losses are included in the denominator of (1.9), since imperfect receiver implementation often results in a loss of effective received signal power. Sometimes it also increases the effective noise power as well.

The total input thermal noise power to the detector is [123]

$$N = kT_o B_w F. \quad (1.10)$$

The value of kT_o at a room temperature of 17 °C (290 °K) is $kT_o = -174$ dBm/Hz. The noise figure F is a degradation caused by elements in a radio frequency signal chain introducing noise. The noise figure of a system, for example an amplifier, is defined by the ratio of the input signal-to-noise ratio to output signal-to-noise ratio. The noise figure is always greater than or equal to unity, and a lower noise figure indicates better performance.

The received carrier-to-noise ratio, Γ , defines the link budget, where

$$\Gamma = \frac{\Omega_p}{N} = \frac{\Omega_t G_T G_R}{kT_o B_w F L_{RX} L_p}. \quad (1.11)$$

The carrier-to-noise ratio, Γ , and modulated symbol energy-to-noise ratio, E_s/N_o , are related as follows [123]:

$$\frac{E_s}{N_o} = \Gamma \times \frac{B_w}{R_s}. \quad (1.12)$$

Hence, the link budget can be written as

$$\frac{E_s}{N_o} = \frac{\Omega_t G_T G_R}{k T_o R_s F L_{R_X} L_p}. \quad (1.13)$$

Converting to decibel units gives

$$E_s/N_{o(\text{dB})} = \Omega_t (\text{dBm}) + G_T (\text{dB}) + G_R (\text{dB}) - k T_o (\text{dBm/Hz}) - R_s (\text{dBHz}) - F (\text{dB}) - L_{R_X} (\text{dB}) - L_p (\text{dB}). \quad (1.14)$$

The receiver sensitivity is defined as

$$\Omega_{\text{th}} = L_{R_X} k T_o F (E_s/N_o) R_s \quad (1.15)$$

or converting to decibel units

$$\Omega_{\text{th}} (\text{dBm}) = L_{R_X} (\text{dB}) + k T_o (\text{dBm/Hz}) + F (\text{dB}) + E_s/N_{o(\text{dB})} + R_s (\text{dBHz}). \quad (1.16)$$

In (1.16), all parameters are usually fixed except for $E_s/N_{o(\text{dB})}$. To determine the receiver sensitivity, the minimum $E_s/N_{o(\text{dB})}$ is first found that will yield an acceptable link QoS, and this value is then substituted into (1.16). Then by substituting the resulting value for $\Omega_{\text{th}} (\text{dBm})$ into (1.14) and solving for $L_p (\text{dB})$, the maximum allowable path loss that the link can tolerate is obtained as

$$L_{\text{max}} (\text{dB}) = \Omega_t (\text{dBm}) + G_T (\text{dB}) + G_R (\text{dB}) - \Omega_{\text{th}} (\text{dBm}). \quad (1.17)$$

Once the maximum allowable path loss is known, a path loss model can be applied to determine the maximum radio path length which is equal to the cell radii. For cellular radio systems, there are three other relevant link budget parameters: (1) the margin for system loading or interference loading, (2) the shadow margin, and (3) the handoff gain. The first two quantities will reduce the maximum allowable path loss, while the third increases it.

1.5.1 Interference Loading

Frequency reuse results in co-channel and adjacent channel interference. As the system load increases the level of interference will also increase. This increase in interference will cause the cell radii to shrink since the radio receivers will be subjected to interference in addition to the thermal noise. Once the cell radii shrink, MSs that are located near the edges of the cells will experience an unacceptably low QoS. This will result in some links being dropped. However, as the connections are dropped the level of interference will decrease. This in turn will expand the cell radii and MSs located near the edges of the cells will again be able to establish links. This will once again increase the system load, and the entire process will repeat itself. This phenomenon of shrinking and expanding cell boundaries due to interference is sometimes called cell breathing.

To ensure coverage and prevent dropped links as the system load increases, an interference margin must be included in the link budget. Note that the received carrier-to-interference-plus-noise ratio is

$$\Gamma_{IN} = \frac{\Omega_p}{I + N} = \frac{\Omega_p/N}{1 + I/N}, \quad (1.18)$$

where I is the total interference power. A key parameter in this equation is the interference-to-noise ratio, I/N . The net effect of such interference is to reduce the carrier-to-noise ratio Ω_p/N by the factor $L_I = (1 + I/N)$. To allow for system loading, the maximum allowable path loss in (1.17) must be reduced by an amount equal to $L_I (\text{dB})$, otherwise known as the interference margin. The required $L_I (\text{dB})$ depends on the particular cellular system under consideration and the traffic load.

The interference margin can be quite difficult to derive since it depends not only on the parameters of the radio link but also on the detailed resource management algorithms being used for power control and handoff. CDMA systems typically require a higher interference margin than TDMA systems due to their universal frequency reuse. With universal frequency reuse, every cell and cell sector in the network can reuse the same set of carrier frequencies, and the emitter will act as a source of interference to a relatively large number of receivers (compared to larger reuse clusters) due to the small reuse distance. In any case, comparisons between different systems should always be made using the same total system bandwidth and the same level of traffic loading.

1.5.2 Shadow Margin and Handoff Gain

Suppose that an outage due to thermal noise occurs whenever the received carrier-to-noise ratio at distance d , $\Gamma = \Omega_p(d)/N$ drops below some threshold Γ_{th} . Once the noise power N in (1.10) is specified, an outage will occur when the local mean drops below the receiver sensitivity, i.e., $\Omega_p \text{ (dBm)}(d) < \Omega_{\text{th}} \text{ (dBm)}$. The outage probability due to noise on the cell boundary is defined as the probability that $\Omega_p \text{ (dBm)}(R) < \Omega_{\text{th}} \text{ (dBm)}$, where $d = R$ for a MS located on the cell edge. The area averaged noise outage probability is defined as the probability that $\Omega_p \text{ (dBm)}(d) < \Omega_{\text{th}} \text{ (dBm)}$, where the average is taken over the random location of a MS in the entire cell area. If the spatial distribution of MSs unknown, then the MSs may be assumed to be uniformly distributed throughout the cell area. To ensure a specified edge or area averaged outage probability, a shadow margin, M_{shad} , must be included in the link budget. Finally, the outage probability will depend on the transmit power. In cellular systems, the BSs and MSs are power controlled so an outage is generally calculated under the condition of maximum allowable transmit power.

The outage probability due to noise on the cell edge is

$$\begin{aligned} O_N(R) &= \text{P}[\Omega_p \text{ (dBm)}(R) < \Omega_{\text{th}} \text{ (dBm)}] \\ &= \int_{-\infty}^{\Omega_{\text{th}} \text{ (dBm)}} \frac{1}{\sqrt{2\pi}\sigma_{\Omega}} \exp\left\{-\frac{(x - \mu_{\Omega_p \text{ (dBm)}}(R))^2}{2\sigma_{\Omega}^2}\right\} dx \\ &= Q\left(\frac{M_{\text{shad}}}{\sigma_{\Omega}}\right), \end{aligned} \quad (1.19)$$

where

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

and

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

is the shadow margin. The outage probability, $O_N(R)$ is plotted against M_{shad} in Fig. 1.9 for various shadow standard deviations.

Example 1.1. Suppose it is desirable to have $O_N(R) = 0.1$. To determine the required shadow margin, M_{shad} is chosen so that the shaded area under the Gaussian density function in Fig. 1.10 is equal to 0.1. Thus

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

(continued)

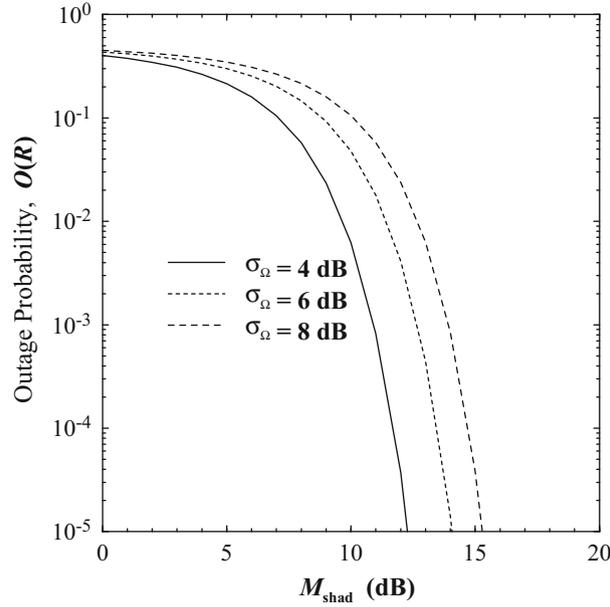


Fig. 1.9 Outage probability due to noise on the cell edge, $O_N(R)$, against the shadow margin M_{shad}

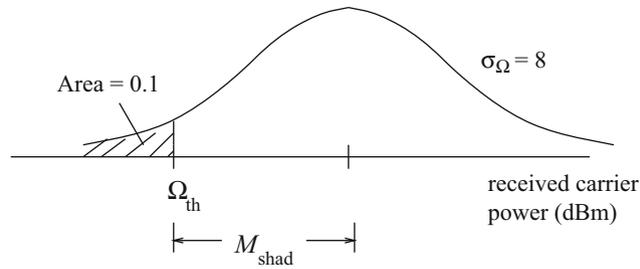


Fig. 1.10 Determining the required shadow margin

Example 1.1 (continued)

Hence,

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

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

$$M_{\text{shad}} = 1.28 \times 8 = 10.24 \text{ dB}. \quad (1.24)$$

To obtain a relationship between the edge and area averaged noise outage probabilities, models are needed for the propagation path loss and spatial distribution of MSs. It is common to assume that the MSs are uniformly distributed throughout each cell area. This assumption along with the path loss model in (1.6) yields an area averaged noise outage probability [122]

$$\begin{aligned} O_N &= \frac{1}{\pi R^2} \int_0^R O_N(r) 2\pi r dr \\ &= Q(X) - e^{XY+Y^2/2} Q(X+Y) \end{aligned} \quad (1.25)$$

where

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

where $\xi = \ln(10)/10$. The first term of this expression is equal to the noise outage probability on the cell edge, $O_N(R)$, while the second term is a correction factor. This obviously means that the edge outage probability is higher than the area averaged outage probability.

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 complicated. As a MS moves from one cell to the next, handoffs or handovers will be executed to maintain service 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 be in an outage condition, the link to an alternate BS may at the same time provide an acceptable link quality. This is due to the fact that different shadowing conditions are usually encountered on links with different BSs. Hence, at the boundary area between two cells, a macrodiversity gain is obtained. The word macrodiversity is used to describe the case where the multiple receiver antennas are located in different base stations, as opposed to microdiversity where the multiple antennas are collocated in the same base station. Handoffs take advantage of macrodiversity, and they will increase the maximum allowable path loss 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-95A/B, cdma2000, and WCDMA use soft handoff, while TDMA cellular systems such as GSM/GPRS/EDGE and OFDMA cellular systems such as LTE/LTE-A and WiMAX 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 adjacent cells. Although the MS is located in the center cell, it is possible that the MS could be connected to any one of the 7 BSs. The area averaged noise outage probability for the target cell is to be calculated, assuming that the MS location is uniform distributed over the target cell area. This can be done quite effectively by using Monte Carlo approaches with a large number of trials. Our results assume that the links to the serving BS and the six neighboring BSs experience correlated log-normal shadowing. The shadow gain at BS_{*i*} is obtained as

$$\epsilon_i \text{ (dB)} = a\zeta + b\zeta_i, \quad (1.27)$$

where

$$a^2 + b^2 = 1,$$

and ζ and ζ_i are generated once each simulation trial, and constitute independent Gaussian random variables with zero mean and variance σ_{Ω}^2 . It follows that the shadow gains (in decibel units) have the correlation

$$E[\epsilon_i \text{ (dB)} \epsilon_j \text{ (dB)}] = a^2 \sigma_{\Omega}^2 = \rho \sigma_{\Omega}^2, \quad i \neq j. \quad (1.28)$$

where $\rho = a^2$ is the correlation coefficient. Here we assume that $\rho = 0.5$.

Let $\Omega_{p,k} \text{ (dBm)}$, $k = 0, \dots, 6$ denote the received signal strength associated with the target BS ($k = 0$) and each of the six neighboring BSs ($k = 1, \dots, 6$). Three cases are considered: a single isolated cell to compare with earlier results, soft handoffs, and hard handoffs. For the case of a single isolated cell, no handoffs are used and the outage probability is identical to that obtained in (1.19) or (1.25). With a soft handoff algorithm, the BS that provides the largest instantaneous received signal strength is always selected as the serving BS. The instantaneous received signal strength is affected by not only the path loss and shadowing variations, but envelope fading as well. However, for the present purpose envelope fading is not considered.¹ If any BS has an associated received signal power that is above the receiver sensitivity, $\Omega_{\text{th}} \text{ (dBm)}$, then the link quality is acceptable; otherwise an outage will occur. Other more sophisticated and effective soft handoff and power control strategies are the subject of Chaps. 12 and 13.

With a hard handoff algorithm, the received signal power at the target MS from the serving BS is denoted as $\Omega_{p,0} \text{ (dBm)}$. If this value exceeds the receiver sensitivity, $\Omega_{\text{th}} \text{ (dBm)}$, then the link quality is acceptable. Otherwise, the six surrounding BSs are evaluated for handoff candidacy by using a mobile assisted handoff (MAHO) algorithm. BS k is included in the neighbor

¹As discussed in Sect. 2.6.2.1, a composite Nakagami-log-normal random variable can be approximated by a purely log-normal random variable with an appropriate area mean and shadow variance.

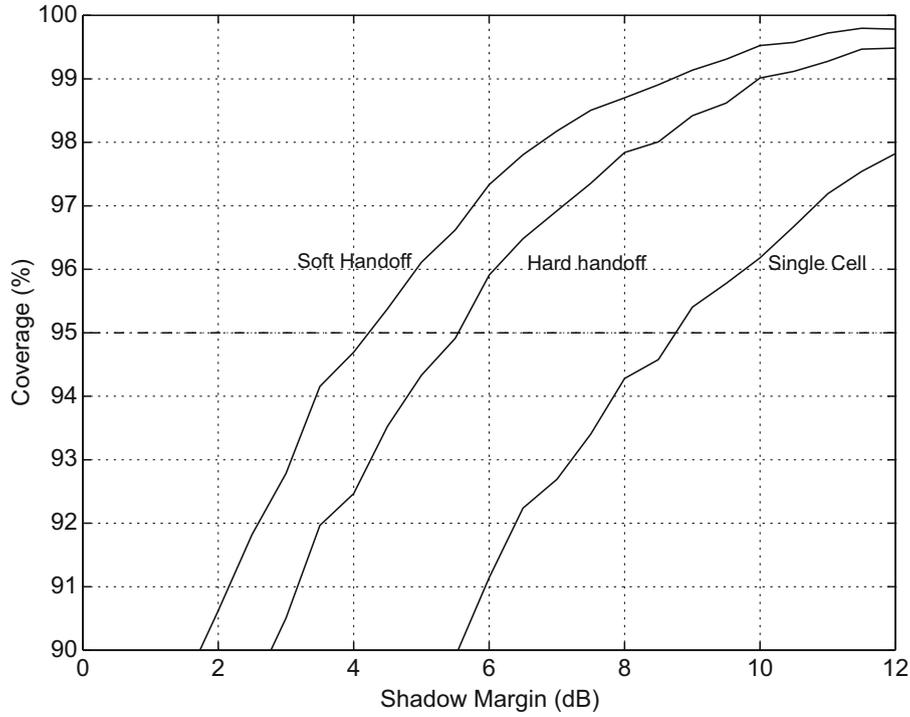


Fig. 1.11 Required shadow margin with hard and soft handoffs and 95% coverage; handoff hysteresis $H_{(\text{dB})} = 6$ dB

set if $\Omega_{p,k} (\text{dBm}) - \Omega_{p,0} (\text{dBm}) \geq H_{(\text{dB})}$, where $H_{(\text{dB})}$ is the handoff hysteresis. If the received signal power for any member of the neighbor set is above the receiver sensitivity, $\Omega_{\text{th}} (\text{dBm})$, then link quality is acceptable; otherwise an outage occurs. A more detailed description and analysis of hard handoff is provided in Chap. 13.

Figure 1.11 compares hard and soft handoffs, for $H_{(\text{dB})} = 6$ dB. Note that a 10% area noise outage probability (90% coverage) requires a shadow margin of 5.6 dB for a single isolated cell (no handoff). With soft handoffs, the required shadow margin is only 1.8 dB. The difference of 3.8 dB between the two is the soft handoff gain. The corresponding hard handoff gain is about 2.8 dB. Note that the soft handoff gain is always greater than the hard handoff gain. However, the true relative advantage of soft handoff over hard handoff depends on many factors, such as hard handoff delay, and is difficult to predict.

The maximum allowable path loss with the inclusion of the margins for shadowing, interference loading, and handoff gain is

$$L_{\text{max}} (\text{dB}) = \Omega_r (\text{dBm}) + G_T (\text{dB}) + G_R (\text{dB}) - \Omega_{\text{th}} (\text{dBm}) - M_{\text{shad}} (\text{dB}) - L_I (\text{dB}) + G_{\text{HO}} (\text{dB}). \quad (1.29)$$

1.6 Coverage

Coverage refers to the number of base stations or cell sites that are required to cover or provide service to a given geographical area with an acceptable quality of service. This is an important consideration when a new cellular network is deployed. Clearly the choice of cellular system that requires the fewest number of cell sites to cover a given geographic area has an infrastructure cost advantage. However, it is always important to include interference margin into the coverage calculation to allow for system loading and this is often where the difficulty lays in comparing the different options. First, the traffic loads must be the same to allow for a fair comparison. Second, the function relating the required interference margin, $L_I (\text{dB})$, to the system load can be quite complicated, especially for CDMA cellular systems with universal frequency reuse.

The number of cell sites that are required to cover a given area is determined from knowledge of the maximum allowable path loss and the path loss characteristic with distance. To compare the coverage of different cellular systems, the maximum allowable path loss is first determined for the different systems at the same QoS. From (1.6), it is apparent that

$$L_{\text{max}} (\text{dB}) = C + 10\beta \log_{10}\{d_{\text{max}}\} (\text{dB}) \quad (1.30)$$

where d_{\max} is the distance corresponding to the maximum allowable path loss and C is some constant that depends on factors that are common to different cellular system alternatives, such as the antenna heights and operating frequency. 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. A variety of theoretical and empirical path loss models will be considered in detail in Sect. 2.7.

Once L_{\max} has been determined for the various cellular systems to be evaluated, the relative coverage of the different systems can be compared, all other factors being equal. For example, suppose that System 1 has $L_{\max(\text{dB})} = L_1$ and System 2 has $L_{\max(\text{dB})} = L_2$, 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:

$$\begin{aligned} L_1 - L_2 &= 10\beta (\log_{10}\{d_1\} - \log_{10}\{d_2\}) \\ &= 10\beta \log_{10} \left\{ \frac{d_1}{d_2} \right\} \end{aligned} \quad (1.31)$$

Looking at things another way

$$\frac{d_1}{d_2} = 10^{(L_1 - L_2)/(10\beta)}. \quad (1.32)$$

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

$$\frac{A_1}{A_2} = \frac{d_1^2}{d_2^2} = \left(\frac{d_1}{d_2} \right)^2 \quad (1.33)$$

and, hence,

$$\frac{A_1}{A_2} = 10^{2(L_1 - L_2)/(10\beta)}. \quad (1.34)$$

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

$$\frac{N_1}{N_2} = \frac{A_{\text{tot}}/A_1}{A_{\text{tot}}/A_2} = \frac{A_2}{A_1} = 10^{-2(L_1 - L_2)/(10\beta)}. \quad (1.35)$$

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 substantial difference in infrastructure cost. Since parameters such as the $E_c/N_o(\text{dB})$ required, handoff gain and interference margin can each vary considerably from one system to the next, careful consideration is required when comparing different cellular system alternatives.

1.7 Spectral Efficiency and Capacity

Spectral efficiency is of primary 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 with circuit-switched voice traffic can be expressed in terms of the following parameters:

- G_c = offered traffic per channel (Erlangs/channel)
- N_{slot} = number of channels per RF carrier (channels/carrier)
- N_c = number of RF carriers per cell (carriers/cell)
- W_{sys} = bandwidth per reuse cluster (Hz/reuse cluster)

W_c = bandwidth per channel (Hz/channel)

A = area per cell (m^2/cell)

N = number of cells per reuse cluster (cells/reuse cluster).

One Erlang is the traffic intensity in a channel that is continuously occupied, so that a channel occupied for $x\%$ of the time carries $x/100$ Erlangs. Adjustment of this parameter controls the system loading and it is important to compare systems at the same traffic load level. For an N -cell reuse cluster, the spectral efficiency can be defined as follows:

$$\eta_s = \frac{N_c N N_{\text{slot}} G_c}{W_{\text{sys}} A} \text{ Erlangs/m}^2/\text{Hz}. \quad (1.36)$$

Recognizing that the bandwidth per channel, W_c , is $W_{\text{sys}}/(N N_c N_{\text{slot}})$, the spectral efficiency can be written as the product of three efficiencies, viz.,

$$\eta_s = \frac{1}{W_c} \cdot \frac{1}{A} \cdot G_c \quad (1.37)$$

$$= \eta_M \cdot \eta_C \cdot \eta_T, \quad (1.38)$$

where

η_M = modulation efficiency

η_C = spatial efficiency

η_T = trunking efficiency

Unfortunately, these efficiencies are not mutually independent so the optimization of spectral efficiency can be quite complicated.

For cellular systems, the cell capacity, defined as number of channels per cell, is sometimes used instead of the Erlang capacity. The cell capacity is

$$N_c N_{\text{slot}} = \frac{W_{\text{sys}}}{W_c \cdot N} \quad (1.39)$$

where, again, W_c is the bandwidth per channel and N_{slot} is the number of traffic channels multiplexed on each RF carrier. Note that cell capacity does not account for trunking efficiency.

1.7.1 Bandwidth Efficiency

Bandwidth efficiency is measured in bits per second per unit bandwidth (b/s/Hz). High bandwidth efficiency can be achieved by using bandwidth efficient modulation and coding techniques, along with effective receiver signal processing techniques that produce radio links that are tolerant to interference. Bandwidth efficiency (bits/s/Hz) is equal to modulation efficiency (channels/Hz) defined above multiplied by the bit rate per channel (bits/s/channel).

1.7.2 Spatial Efficiency

High spatial efficiency can be achieved by (1) minimizing the area per cell, and (2) minimizing the co-channel reuse distance. The first of these explains the intense interest in microcell and picocell technologies, where cell radii on the order of 50–500 m are used. The co-channel reuse distance D/R is minimized by (1) controlling the generation of CCI, and (2) mitigating the effect of any CCI that is present. The generated levels of CCI can be controlled by using techniques such as cell sectoring, smart antennas, power control, scheduling, effective handoff algorithms, macroscopic BS diversity, and a whole host of other

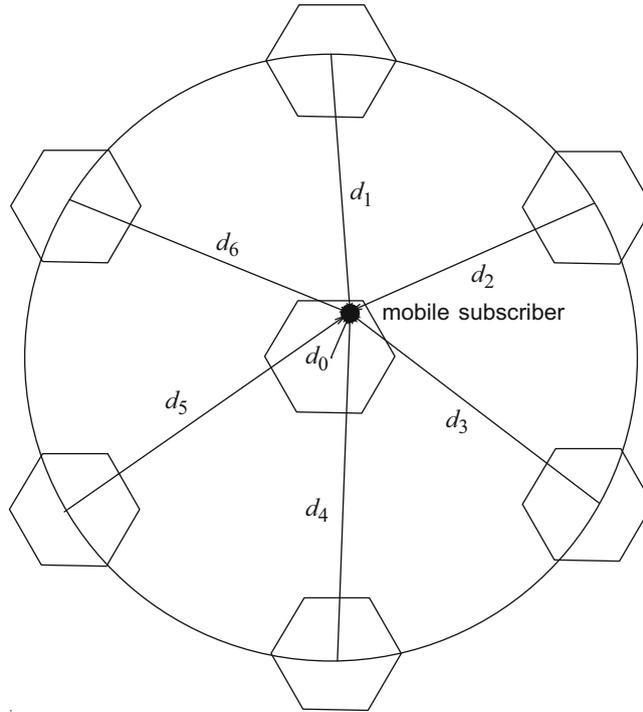


Fig. 1.12 Co-channel interference on the forward channel at a desired MS. There are six first-tier interfering BSs

techniques. The impact of CCI on the radio link can be mitigated by using techniques such as optimum combining, single antenna interference cancelation, equalization, antenna diversity, and others.

Consider the situation shown in Fig. 1.12, depicting a simplified worst case forward link channel CCI condition with isotropic BS antennas where the maximum six first-tier co-channel interfering channels are present. The MS is located at distance d_0 from the serving BS and at distances $d_k, k = 1, 2, \dots, N_I$ from the first tier of $N_I = 6$ interfering co-channel BSs. If $\mathbf{d} = (d_0, d_1, \dots, d_{N_I})$ denotes the vector of distances at a particular MS location, then the forward link carrier-to-interference ratio as a function of the distance vector \mathbf{d} is

$$A_{(\text{dB})}(\mathbf{d}) = \Omega_{p(\text{dBm})}(d_0) - 10 \log_{10} \left\{ \sum_{k=1}^{N_I} 10^{\Omega_{p(\text{dBm})}(d_k)/10} \right\}. \quad (1.40)$$

At this point, the handoff gain must be accounted for. Consider, for example, the case of soft handoff. Let $\Lambda_k(\text{dB})(\mathbf{d}), k = 0, \dots, M$ denote the carrier-to-interference ratio for serving BS and M surrounding BSs. Note that the vector \mathbf{d} is different for each candidate BS. With soft handoff, the BS that provides the most robust link is always selected such that the resulting carrier-to-interference ratio is

$$\Lambda_{(\text{dB})} = \max\{\Lambda_0(\text{dB})(\mathbf{d}), \Lambda_1(\text{dB})(\mathbf{d}), \dots, \Lambda_M(\text{dB})(\mathbf{d})\}. \quad (1.41)$$

The area averaged probability co-channel interference outage is

$$O_I = \text{P}[\Lambda_{(\text{dB})} < \Lambda_{\text{th}}(\text{dB})], \quad (1.42)$$

where the calculation is performed by averaging the probability of outage over the random location of a MS within the reference cell. Finally, the outage depends on the number of interferers, N_I that are present. Due to the statistical nature of the user traffic, the number of interferers present is random. In the case of Fig. 1.12, it ranges from 0 to 6.

Finally, Fig. 1.13 depicts the worst case co-channel interference on the reverse link at the serving BS, considering first-tier co-channel interferers. Note that the co-channel interference conditions are not the same on the forward and reverse links, because the worst case vector \mathbf{d} is different in each direction. This phenomenon is known as link imbalance.

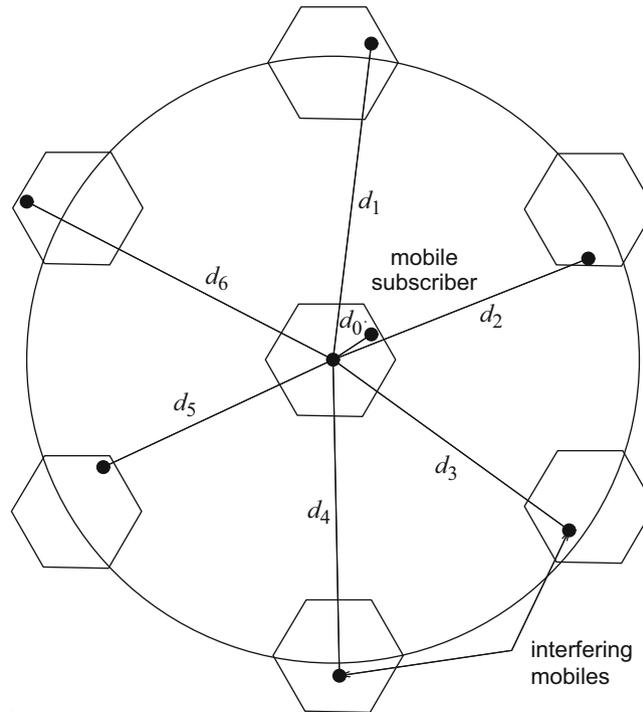


Fig. 1.13 Co-channel interference on the reverse channel at a desired BS. There are six first-tier interfering MSs

1.7.3 Trunking Efficiency

High trunking efficiency can be achieved by using channel assignment schemes that maximize channel utilization. There is usually a trade-off between trunking efficiency (or offered traffic per channel) and grade of service in terms of new call and handoff blocking probabilities. Various fundamental formulas were developed by Erlang almost a century ago that laid the foundation of modern teletraffic theory. One of Erlang's most famous results is the Erlang-B formula, first derived in 1917, that gives the probability that a newly arriving call will not find any available channel in a trunk of m channels and is blocked. Sometimes this policy is called the blocked calls cleared queueing discipline, meaning that blocked calls are not buffered or queued, and if no free channels are available they are dropped. The Erlang-B formula is not entirely applicable to cellular systems, because it does not account for handoff calls. Furthermore, the total offered traffic per cell may be time-varying due to the spatial movement of the subscribers, whereas the offered traffic in the Erlang-B formula is assumed to be constant. The Erlang-B formula is

$$B(\rho_T, m) = \frac{\rho_T^m}{m! \sum_{k=0}^m \frac{\rho_T^k}{k!}}, \quad (1.43)$$

where m is the total number of channels per cell and $\rho_T = \lambda_a \mu_d$ is the total offered traffic per cell (λ_a is the call arrival rate and μ_d is the mean call duration). The Erlang-B formula is derived under the so-called standard Markovian assumptions, including an infinite population of users, Poisson call arrivals with rate λ_a calls/s, and exponentially distributed call durations with a mean call duration μ_d s/call. Note that the total offered traffic ρ_T is a dimensionless quantity, but the quantity is expressed as Erlangs.

Figure 1.14 plots the blocking probability $B(\rho_T, m)$ as a function of the offered traffic per channel $G_c = \rho_T/m$. The benefit from trunking is obvious, since the offered traffic per channel, G_c , increases as the number of trunked channels increases, at any blocking probability. Note that diminishing returns are obtained as the number of trunked channels becomes larger. Finally, it is important to realize that doubling the number of channels in a cell or cell sector will double the cell or sector capacity. However, the Erlang capacity will more than double due to trunking efficiency.

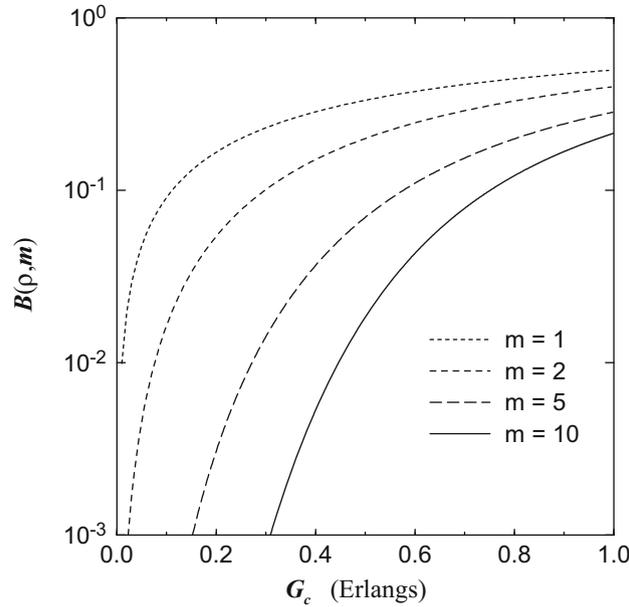


Fig. 1.14 Erlang-B blocking probability $B(\rho_T, m)$ vs. offered traffic per channel $G_c = \rho_T/m$. Trunking is shown to improve the spectral efficiency

1.7.4 Capacity

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

1. The cell capacity or sector capacity equal to the number of available traffic channels per cell or cell sector.
2. The cell Erlang capacity equal to the traffic carrying capacity of a cell (in Erlangs) for a specified call blocking probability.

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

Comparing the spectral efficiency of different cellular systems can be difficult, because the various systems may be in different evolutionary stages. However, a fair comparison between suitably optimized digital cellular systems without deployment constraints will probably show roughly equal spectral efficiency. Indeed this is what is seen with EV-DO Rev B., HSPA+, LTE-A. HSPA+ achieves peak downlink speeds of 21 Mbps in 5 MHz, while LTE-A achieves 22 Mbps in the same bandwidth. Both systems are highly optimized. Recall that LTE-A downlink uses OFDMA, while the EV-DO downlink uses TDM/CDMA, which are two entirely different air interfaces.

1.7.4.1 GSM Cell Capacity

GSM systems may or may not use frequency hopping, where the frames are hopped over a predefined set of carrier frequencies and in a predefined order so as to average the effect of interference. A 3/9 (3-cell/9-sector) reuse pattern is achievable for most GSM systems that employ frequency hopping; without frequency hopping, a 4/12 reuse pattern may be possible. A capacity gain is achieved with frequency hopping, since the co-channel interference is averaged over the set of hop frequencies. GSM has 8 logical channels that are time division multiplexed onto a single radio frequency carrier, and the carriers are spaced 200 kHz apart. Therefore, the bandwidth per logical channel is roughly 25 kHz, which was common in first generation European analog mobile phone systems. In a nominal bandwidth of 1.25 MHz (uplink or downlink) there are $1250/200 = 6.25$ carriers spaced 200 kHz apart.² Hence, there are $6.25/9 \approx 0.694$ carriers per sector or $6.25/3 = 2.083$ carriers/cell. Each logical channel commonly carries half rate traffic, such that there are 16 traffic channels/carrier. Hence, the 3/9 reuse system has a sector capacity of 1.11 traffic channels/sector or a cell capacity of 33.33 traffic channels/cell in 1.25 MHz.

²Fractional carriers are used for the capacity calculation but can be eliminated in practice by using a larger nominal bandwidth.

1.7.4.2 IS-95 Cell Capacity

The cell capacity of IS-95 was a topic of historical debate. CDMA systems were initially reported to achieve 40 times AMPS cell capacity, which made GSM appear to have a rather meager capacity. The key attribute of CDMA systems that gives it high capacity is universal frequency reuse, where all cells and cell sectors reuse the same carrier frequency. Universal frequency reuse means that every transmitter interferes with every receiver within radio range. The complicated interference environment combined with the features of soft handoff and power control make the determination of cell capacity is quite elusive even for the basic IS-95A system.

To illustrate the difficulty in evaluating CDMA cell capacity, consider the following simplified analysis. Suppose there are N users in a cell: one desired user and $N - 1$ interfering users. For the time being, ignore the interference from surrounding cells. Consider the reverse link, and assume perfectly power controlled MS transmissions that arrive chip and phase asynchronously at the BS receiver. Treating the $N - 1$ co-channel signals as uncorrelated the effective carrier-to-noise ratio is

$$\Gamma = \frac{3}{N-1}, \quad (1.44)$$

and the effective received bit energy-to-noise ratio is

$$\begin{aligned} \frac{E_b}{N_o} &= \Gamma \times \frac{B_w}{R_b} \\ &= \frac{3G}{N-1} \approx \frac{3G}{N}, \end{aligned}$$

where $G = B_w/R_b$. The factor of 3 in the numerator of (1.44) arises from the assumption of randomly generated spreading sequences and the signals arriving at the receiver antenna in a chip and phase asynchronous fashion [230]. For a required E_b/N_o , $(E_b/N_o)_{\text{req}}$, the cell capacity is

$$N \approx \frac{3G}{(E_b/N_o)_{\text{req}}}.$$

Suppose that 1.25 MHz of spectrum is available and the source coder operates at $R_b = 4$ kbps. Then $G = 1250/4 = 312.5$. If $(E_b/N_o)_{\text{req}} = 6$ dB (a typical IS-95 value), then the cell capacity is roughly $N = 3 \cdot 312.5/4 \approx 234$ channels per cell. This is roughly 7 times the cell capacity of GSM. However, this rudimentary analysis did not include out-of-cell interference which is typically 50–60% of the in-cell interference. This will result in a reduction of cell capacity by a factor of 1.5 and 1.6, respectively. Also, with CDMA receivers, great gains can be obtained by improving receiver sensitivity. For example, if $(E_b/N_o)_{\text{req}}$ can be reduced by 1 dB, then the cell capacity N increases by a factor of 1.26. Finally, CDMA systems are known to be sensitive to power control errors. An rms power control error of 2 dB will reduce the capacity by roughly a factor of 2 as discussed in Chap. 12.

Problems

- 1.1. Show that the area averaged outage probability due to thermal noise is given by (1.25).
- 1.2. By using geometric arguments, show that the co-channel reuse factor, D/R , for cellular deployments based on hexagonal cells is given by $D/R = \sqrt{3N}$.
- 1.3. A brief measurement campaign indicates that the median propagation loss at 420 MHz in a mid-size North American city can be modeled by the following path loss equation:

$$L_p = 25 \text{ dB} + 10 \log_{10}\{d^{2.8}\},$$

where d is in units of meters, i.e., the path loss exponent is $\beta = 2.8$ and there is a 25 dB fixed loss.

- (a) Assuming a cell phone receiver sensitivity of -95 dBm, what transmitter power is required to service a circular area of radius 10 km?
- (b) Suppose the measurements were optimistic and $\beta = 3.1$ is more appropriate. What is the corresponding increase in transmit power (in decibels) that would be required?
- (c) If log-normal shadowing is present with $\sigma_{\Omega} = 8$ dB, how much additional transmit power is required to ensure 10% thermal noise outage at a distance of 10 km?

1.4. A receiver in an urban cellular radio system detects a 1 mW signal at $d_o = 1$ m from the transmitter. In order to mitigate co-channel interference, it is required that the co-channel interference power that is received from any co-channel base station be no more than -100 dBm. A measurement team has determined that the average path loss exponent in the system is $\beta = 3$.

- (a) Determine the radius R of each cell if a 7-cell reuse pattern is used.
- (b) What is the radius R if a 4-cell reuse pattern is used?

1.5. Consider a cellular system that uses a 7-cell hexagonal reuse cluster. The base stations employ 120° wide-beam directional antennas and they all have the same antenna height and transmit with the same power level. Consider the forward channel (base-to-mobile). Ignore shadowing and envelope fading and consider only the path loss. A mobile station will experience the lowest co-channel interference ratio, Λ , when it is located in the corner of a cell.

- (a) Considering only the first tier of co-channel base stations, what is the worst case Λ with a path loss exponent of 4?
- (b) Considering the first and second tiers of co-channel base stations, what is the worst case Λ with path loss exponent of 4?
- (c) From parts a) and b) what conclusions can you make about the effect of the second-tier co-channel base stations? What happens if the path loss exponent is equal to 3?

Note: In this problem you should use exact radio path distances, expressed in terms of the cell radius R . You can obtain these distances by laying out a 7-cell deployment on hex paper and using geometric arguments.

1.6. Consider the worst case forward channel co-channel interference situation shown in Fig. 1.15. The path loss is described by the following model:

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

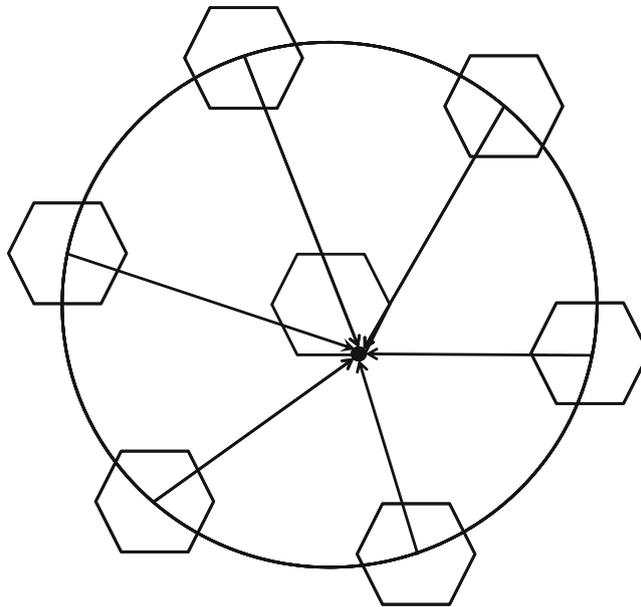


Fig. 1.15 Worst case co-channel interference on the forward channel

where

μ_{Ω_p} = average received power

Ω_t = transmitted power

h_b = base station antenna height

h_m = mobile station antenna height

d = radio path length

- Assume that $h_b = 30$ m, $h_m = 1.5$ m, and Ω_t is the same for all BSs. What is the worst case carrier-to-interference ratio Λ for a cluster size $N = 4$?
- 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 it is desired to maintain the same worst case Λ value obtained in part a), what is the new radius of the center cell?
- 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 enlarging the co-channel cell and distorting the cell boundaries. Assuming, again, it is desired to maintain the same worst case Λ value obtained in part a), what are the new boundaries of the center cell?

1.7. Suppose that a cellular frequency reuse environment is characterized by path loss and log-normal shadowing, such that the local mean at distance d is given by (1.4), where $\mu_{\Omega_p \text{ (dBm)}}(d_o) = -80$ dBm at $d_o = 1.6$ km, $\beta = 3.68$, and $\epsilon_{\text{(dB)}} \sim \mathcal{N}(0, \sigma_{\Omega}^2)$, where $\sigma_{\Omega} = 8$ dB. Use the above model in the questions below.

- Suppose that a mobile station is at distances 2 and 2.2 km from two different base stations, where the link to each of the two base stations is affected by independent log-normal shadowing. If the base station that provides the largest local mean $\Omega_{p,i \text{ (dBm)}}$, $i = 1, 2$, is selected to serve the mobile station, what is the probability density function of the observed local mean $\Omega_p^o \text{ (dBm)} = \max\{\Omega_{p,1 \text{ (dBm)}}, \Omega_{p,2 \text{ (dBm)}}\}$ at the mobile station.
- As a follow-up to part (a), suppose that an outage occurs if the observed local mean $\Omega_p^o \text{ (dBm)}$ is less than -100 dBm. What is the probability of outage?

1.8. A TDMA cellular system consists of a deployment of uniform radii hexagonal cells with a 9-cell reuse pattern. The cell diameter (corner-to-corner) is equal to 8 km. The system has a total bandwidth of 12.5 MHz (for both uplink and downlink). The channels have a channel spacing of 30 kHz. Calculate the following:

- Number of traffic channels/cell.
- Number of cells required to cover a total area of 3600 km^2 . In this problem use the exact area of the hexagon cell rather than approximating the hexagon cell by a circle cell with the same cell radius.
- Co-channel reuse distance D .

1.9. Whenever a mobile station crosses a cell boundary a handoff occurs to the target cell. However, sometimes a handoff will occur 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 available channel in the target cell is allowed to remain in a queue for t_q seconds and is dropped from the queue, i.e., it will 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

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

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

$$P[W > t_q] = \exp\left\{-\frac{(m - \rho_T)t_q}{\mu}\right\},$$

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 $G_c = \rho_T/m$, for $m = 5, 10$, and 15 . Comment on your results.

1.10. A GSM cellular service provider uses base station receivers that have a carrier-to-interference ratio threshold $A_{th} = 9$ dB.

(a) Find the optimal cluster size N for the following cases:

- (i) omnidirectional antennas
- (ii) 120° sectoring
- (iii) 60° sectoring

Ignore shadowing and use path loss model in (1.6) with 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. The average call duration is equal to 120 s. Determine the offered traffic load (per cell) in units of *Erlangs* and *calls per hour* for each of the cases in part (a).

1.11. 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, $B(\rho_T, m)$, for each system versus ρ_T .

(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.

1.12. A service area is covered by a cellular radio system with 84 cells and a cluster size N . A total of 300 voice channels are available for the system. Users are uniformly distributed over the service area, and the offered traffic per user is 0.04 Erlang. Assume a blocked calls cleared queueing discipline, and the designated blocking probability from the Erlang-B formula is $B = 1\%$.

(a) Determine the carried traffic per cell if cluster size $N = 4$ is used. Repeat for cluster sizes $N = 3, 7$, and 12 .

(b) Determine the number of users that can be served by the system for a blocking probability of 1% and cluster size $N = 4$. Repeat for cluster sizes $N = 7$ and 12 .

In this question, the offered traffic per user is 0.04 Erlang. However,

$$\rho_T = K\rho_u,$$

where

$$\rho_u = \text{offered traffic per user}$$

$$K = \text{number of users}$$

Note that ρ_T in this case is the total offered traffic per cell and K is the number of users per cell.

Chapter 2

Propagation Modeling

The design of spectrally efficient wireless communication systems requires a thorough understanding of the radio propagation channel. The characteristics of the radio channel will vary greatly with the operating frequency, and the propagation environment, e.g., line-of-sight (LoS) versus non-line-of-sight (NLoS), stationary versus mobile transmitters and receivers, and other factors. This chapter emphasizes land mobile radio channels, including those found in cellular land mobile radio systems and other types of vehicular radio systems. However, many of the concepts are of a fundamental nature and will apply to other types of radio channels as well.

A typical cellular land mobile radio system consists of a collection of fixed base stations (BSs) that define radio coverage areas known as cells. The height and placement of the BS antennas affects the proximity of local scatterers at the BSs. In a macrocellular environment where the cell radii are large, the BS antennas are well elevated above the local terrain and are free of local scatterers. Mobile stations (MSs), on the other hand, tend to be surrounded by local scatterers due to their low elevation antennas. Sometimes a line-of-sight (LoS) condition will exist between a BS and a MS, for example in a rural (or open) environment. However, in an urban environment, a non-line-of-sight (NLoS) condition typically exists between the BSs and MSs. As a consequence, the radio waves must propagate between the BSs and MSs via reflections, diffraction, and scattering. Due to the typically large distance between the BSs and MSs in macrocellular systems, radio propagation is often assumed to occur in a two-dimensional (2-D) plane. As shown in Fig. 2.1, multiple plane waves will arrive at the MS (or BS) receiver antenna(s) from different directions, with each having a distinct polarization, amplitude, phase, and delay. This phenomenon is called multipath propagation. The multiple plane waves combine vectorially at each MS (or BS) receiver antenna to produce a composite received signal.

Commercial cellular land mobile radio systems operate at UHF frequencies in bands located at 700/800/900 MHz and 1800/1900 MHz. At these frequencies, the carrier wavelength, λ_c , is approximately 15 cm and 30 cm, respectively, using the relationship $c = f_c \lambda_c$, where f_c is the carrier frequency and c is the speed of light. Therefore, small changes in the propagation delays of the individual multipath components due to MS mobility on the order of a few centimeters will cause a large change in the relative phases of the plane wave components arriving at the MS (or BS) receiver antennas. Hence, when the arriving plane waves combine vectorially at the receiver antenna(s), they will experience constructive and destructive addition depending on the physical location of the MS. If the MS is moving or there are changes in the location of the scatterers, then these spatial variations will manifest themselves as time variations in the amplitude and phase of the composite signal received at each MS (or BS) antenna, a phenomenon known as envelope fading. If the propagation environment is such that no individual multipath component is dominant, such as when NLoS conditions exist between the BS and MS, then the composite receive envelope under narrowband propagation conditions is often modeled as being Rayleigh distributed at any time. Such a channel is said to exhibit Rayleigh fading. However, if a dominant multipath component exists, such as when a LoS or specular condition exists between the BS and MS, then the envelope is often modeled as being Ricean distributed at any time. Such a channel is said to exhibit Ricean fading.

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 NLoS macrocellular environment is usually surrounded by local scatterers, so that the plane waves will arrive at the MS antenna from many different directions, as shown in Fig. 2.1. Two-dimensional (2-D) isotropic scattering, where the plane waves arrive from all azimuth directions with equal probability and with equal strength, is a very commonly used reference model to describe the signals received at a MS in this case. The BSs, on the other hand, are relatively free from local scatterers. Hence, plane waves tend to arrive at a BS with a small azimuth angle of arrival (AoA) spread as shown in Fig. 2.1. It will be shown later in this chapter that these differences in the scattering environment for the forward and reverse links

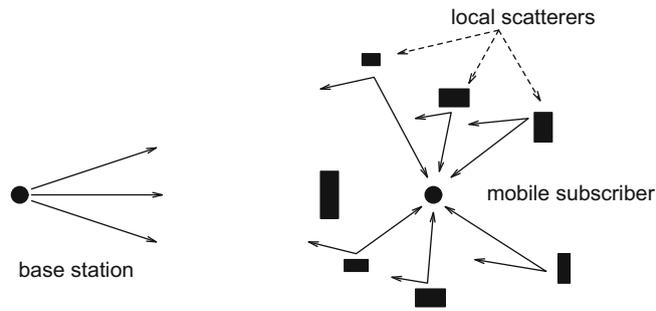


Fig. 2.1 Typical macrocellular radio propagation environment

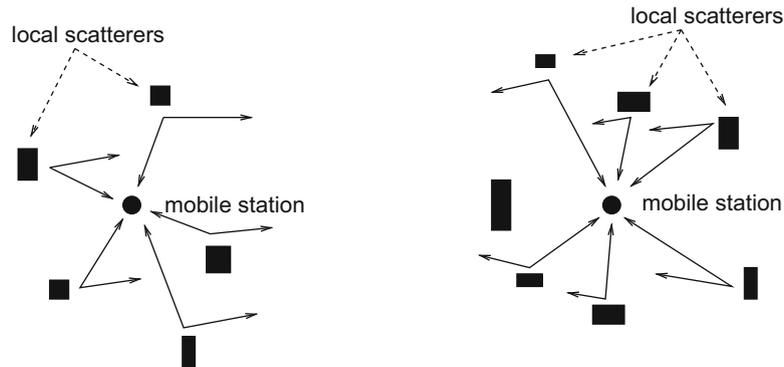


Fig. 2.2 Typical mobile-to-mobile radio propagation environment

will cause significant and important differences in the spatial correlation properties of the respective faded envelopes at the MSs and BSs.

Some types of land mobile radio systems, such as mobile ad hoc networks, consist of vehicle-to-vehicle (or mobile-to-mobile) and vehicle-to-infrastructure links. In this book vehicle-to-vehicle links are referred to as mobile-to-mobile links, because the user terminals are not necessarily vehicular mounted. A typical mobile-to-mobile radio propagation environment is depicted in Fig. 2.2. Such mobile-to-mobile communication systems differ from conventional cellular radio systems where one end of the link (the BS) is stationary and free of scattering, and only the MS is moving. Though the received signal envelope of mobile-to-mobile links is still Rayleigh faded under NLoS narrowband propagation conditions, the mobility of both the transmitters and receivers, and scattering at both ends of the links, causes these links to exhibit much different statistical properties than cellular land mobile radio channels.

If the received envelope or squared-envelope is averaged over a spatial distance of 20–30 wavelengths, an estimated of the mean envelope or mean squared-envelope can be obtained. Sometimes, this quantity is called the local mean because it corresponds to a particular locality. The local mean will experience slow variations over distances of several tens of wavelengths due to the presence of large terrain features such as buildings, hills, and valleys. This phenomenon is known as shadow fading or shadowing. Experimental observations have confirmed that the local mean signal strength follows a log-normal distribution. This log-normal distribution applies to both macrocellular [171, 194] and microcellular environments [149, 226, 228].

If the local mean is averaged over a sufficiently large spatial distance (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 locations that lie at the same distance from the BS. The area mean is inversely proportional to the path loss, which describes how the area mean decreases with the distance between the BS and MS. Early studies by Okumura [250] and Hata [161] yielded an empirical path loss model for macrocellular radio systems operating in urban, suburban, and rural areas. The Okumura–Hata model is accurate to within 1 dB for distances ranging from 1 to 20 km and carrier frequencies between 150 and 1000 MHz, and was adopted in the COST207 study [78]. The Okumura–Hata model is only valid for carrier frequencies less than 1000 MHz. Consequently, when additional spectrum were made available in the 1990s for cellular systems operating in the 1800/1900 MHz band, new path loss models were needed. The COST231 study [81] resulted in the COST231-Hata model that extended the Okumura–Hata model to the 1800/1900 MHz band, and provided the Walfish–Ikegami model for path loss prediction in LoS and NLoS urban microcells. More recent models include the 3GPP models [1] and their extension to mm-wave frequencies [158].

The remainder of this chapter presents the fundamentals of radio propagation modeling, analysis and simulation. Section 2.1 considers conventional narrow-band fixed-to-mobile channels, and various properties of the faded envelope are considered. Section 2.2 considers mobile-to-mobile channels. Section 2.3 treats the statistical characterization of wide-band multipath-fading channels. Simulation models for fading channels are covered in Sect. 2.5. Shadowing models and simulation techniques are discussed in Sect. 2.6. Finally, Sect. 2.7 treats theoretical and empirical models for path loss in macrocellular and microcellular systems.

2.1 Fixed-to-Mobile Channels

For land mobile radio applications, the signals from the BSs are usually transmitted with vertical polarization, meaning that the electric field is perpendicular to the Earth's surface. At VHF frequencies, vertical polarization produces a higher field strength close to the ground than horizontal polarization. Likewise, the MS antennas are also vertically polarized, although tilting of the MS antenna will result in a polarization mismatch. Even if the signals are transmitted with vertical polarization, reflections and diffractions from objects will cause the signals to undergo depolarization. This effect can be exploited by using polarization diverse antennas. For example, cross-polarized antennas, where two antennas having $+45^\circ$, and -45° polarizations from vertical are sometimes used at the BSs.

In cellular land mobile radio systems, the radio signals will propagate in three dimensions. However, if the distance between the BS and MS is sufficiently large, the radio propagation environment is often modeled as occurring in a two-dimensional (2-D) plane. Figure 2.3 depicts a horizontal $x - y$ plane, where a MS is moving in the direction of the positive x -axis with velocity v . The BS is assumed stationary. With vertical polarization, the electric field vector is aligned with the z -axis. The n th 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

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

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 right half plane will experience a positive Doppler shift, while those arriving from the left half plane will experience a negative Doppler shift.

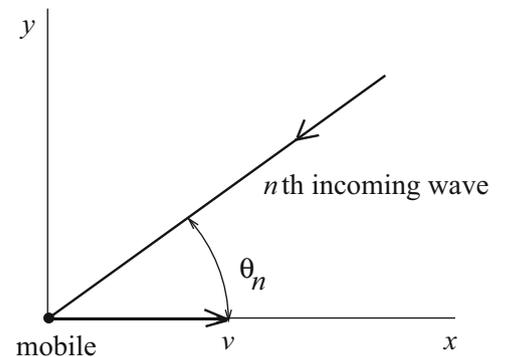
Consider the transmission of the bandpass signal

$$s(t) = \text{Re} \{ \tilde{s}(t) e^{j2\pi f_c t} \}. \quad (2.2)$$

where $\tilde{s}(t)$ is the complex envelope of the bandpass 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 bandpass waveform is

$$r(t) = \text{Re} \left\{ \sum_{n=1}^N C_n e^{j\phi_n - j2\pi c\tau_n/\lambda_c + j2\pi(f_c + f_{D,n})t} \tilde{s}(t - \tau_n) \right\}, \quad (2.3)$$

Fig. 2.3 A typical plane wave incident on a MS receiver



where C_n , ϕ_n , $f_{D,n}$, and τ_n are the amplitude, phase, Doppler shift, and time delay, respectively, associated with the n th propagation path, and $c = f_c \lambda_c$ is the speed of light. The magnitude C_n depends on the cross sectional area of the n th reflecting surface or the length of the n th diffracting edge. The phase ϕ_n is randomly introduced by the n th scatterer and can be assumed to be uniformly distributed on $[-\pi, \pi)$. The delay $\tau_n = d_n/c$ is the propagation delay associated with the n th propagation path, where d_n is the length of the path. The set of path lengths, $\{d_n\}$, will depend on the physical scattering geometry which has not been specified at this point. The Doppler shift $f_{D,n}$ of the n th propagation path is as discussed previously.

Similar to (2.2), the received bandpass signal $r(t)$ has the representation

$$r(t) = \text{Re} \{ \tilde{r}(t) e^{j2\pi f_c t} \}, \quad (2.4)$$

where the received complex envelope is

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

and

$$\phi_n(t) = \phi_n - 2\pi c \tau_n / \lambda_c + 2\pi f_{D,n} t \quad (2.6)$$

is the time-variant phase associated with the n th path. The complex envelope at the input of the channel, $\tilde{s}(t)$, and the complex envelope at the output of the channel, $\tilde{r}(t)$, are related through the convolution integral

$$\tilde{r}(t) = \int_0^t g(t, \tau) \tilde{s}(t - \tau) d\tau \quad (2.7)$$

where $g(t, \tau)$ is the time-variant complex impulse response of the channel. From (2.5), it follows that

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

where $\delta(\cdot)$ is the Dirac delta function or unit impulse function.

From (2.5) and (2.6), several interesting observations can be made. Since the carrier wavelength λ_c is small (approximately 30 cm at 1 GHz), even small changes in the path delays $d_n = c\tau_n$ will cause large changes in the phases $\phi_n(t)$, due to the term $2\pi c \tau_n / \lambda_c$. For a straight line trajectory $2\pi c \tau_n / \lambda_c = 2\pi d_n / \lambda_c$. Also, due to the Doppler frequency $f_{D,n}$, the phases $\phi_n(t)$ vary with time. Hence, at any given point in space-time, the phases $\phi_n(t)$ will result in the constructive or destructive addition of the N multipath components, a phenomenon known as fading.

If the differential path delays $\tau_i - \tau_j$ for all i, j are very small compared to the modulation symbol period, T , then the τ_n that appear in the argument of Dirac delta function in (2.8) can be approximated by their average value $\bar{\tau}$, i.e., $\tau_n \approx \bar{\tau}$. In this case, the complex channel impulse response has the form

$$g(t, \tau) = g(t) \delta(\tau - \hat{\tau}), \quad (2.9)$$

where

$$g(t) = \sum_{n=1}^N C_n e^{j\phi_n(t)}. \quad (2.10)$$

Note that $\phi_n(t)$ remains as defined with the τ_n in (2.6), since the approximation $f_c \tau_n \approx f_c \bar{\tau}$ cannot be made when f_c is large. Therefore, the received complex envelope is

$$\tilde{r}(t) = g(t) \tilde{s}(t - \hat{\tau}) \quad (2.11)$$

which experiences fading due to the time-varying complex channel gain $g(t)$. In the frequency domain, the received complex envelope is

$$\tilde{R}(f) = G(f) * \left(\tilde{S}(f) e^{-j2\pi f \hat{t}} \right). \quad (2.12)$$

Since the channel changes with time, $G(f)$ has a finite non-zero width in the frequency domain. Due to the convolution operation, the output spectrum $\tilde{R}(f)$ will be larger than the input spectrum $\tilde{S}(f)$. This broadening of the transmitted signal spectrum is caused by the channel time variations and is called frequency spreading or Doppler spreading.

The time-variant channel transfer function can be obtained by taking the Fourier transform of (2.9) with respect to the τ variable, giving

$$T(f, t) = \mathcal{F}\{g(t, \tau)\} = g(t) e^{-j2\pi f \hat{t}}. \quad (2.13)$$

The time-variant channel magnitude response is $|T(f, t)| = |g(t)|$. Note that all frequency components in the received signal are scaled by the same time-variant magnitude $|g(t)|$. In this case the received signal is said to exhibit frequency flat fading, because the magnitude of the time-variant channel transfer function is constant (or flat) with respect to frequency variable f .

If the differential path delays $\tau_i - \tau_j$ for some i, j are sufficiently large compared to the modulation symbol period T , then the magnitude response $|T(f, t)|$ is no longer flat and the channel exhibits frequency-selective fading. Sometimes frequency-selective fading channels are called wide-band channels. A simplified concentric-ellipses model for frequency-selective fading channels is depicted in Fig. 2.4, where the transmitter and receiver are located at the foci of the ellipses. Considering only single bounce reflections between the transmitter and receiver, all paths that are associated with scatterers on the n th elliptical contour will have the same propagation delay τ_n . Frequency-selective channels have strong scatterers that are located on several ellipses such that the corresponding differential path delays $\tau_i - \tau_j$ for some i, j , are significant compared to the modulation symbol period T .

2.1.1 Envelope Correlation

A flat fading channel can be characterized by assuming the transmission of a continuous wave sinusoid at frequency f_c , because the channel magnitude response is flat. For convenience, the transmitted sinusoid is assumed to have complex envelope $\tilde{s}(t) = 1$ in (2.5). From (2.4) and (2.5), the corresponding received bandpass signal can be expressed in the quadrature form

$$r(t) = g_I(t) \cos(2\pi f_c t) - g_Q(t) \sin(2\pi f_c t), \quad (2.14)$$

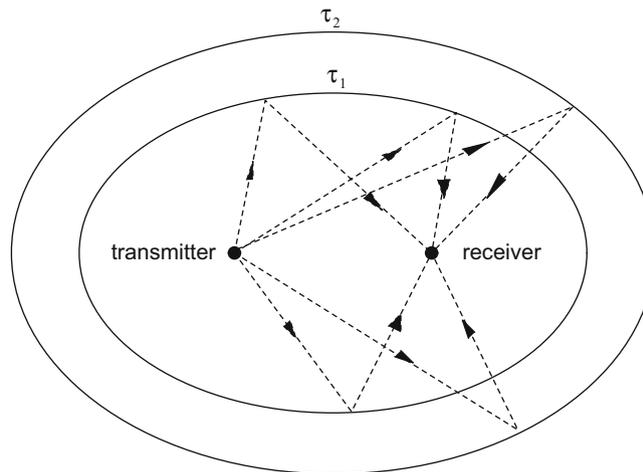


Fig. 2.4 Concentric-ellipses model for frequency-selective fading channels

where

$$g_I(t) = \sum_{n=1}^N C_n \cos(\phi_n(t)) \quad (2.15)$$

$$g_Q(t) = \sum_{n=1}^N C_n \sin(\phi_n(t)) \quad (2.16)$$

are the in-phase and quadrature components of the received bandpass signal. Assuming that the bandpass random process $r(t)$ is wide-sense stationary, the autocorrelation function of $r(t)$ is

$$\begin{aligned} \phi_{rr}(\tau) &= E[r(t)r(t + \tau)] \\ &= \phi_{g_I g_I}(\tau) \cos(2\pi f_c \tau) - \phi_{g_I g_Q}(\tau) \sin(2\pi f_c \tau), \end{aligned} \quad (2.17)$$

where $E[\cdot]$ is the ensemble average operator, and

$$\phi_{g_I g_I}(\tau) \triangleq E[g_I(t)g_I(t + \tau)] \quad (2.18)$$

$$\phi_{g_I g_Q}(\tau) \triangleq E[g_I(t)g_Q(t + \tau)]. \quad (2.19)$$

Problem 2.1 shows that the wide-sense stationarity of the bandpass random process $r(t)$ imposes the condition

$$\phi_{g_I g_I}(\tau) = \phi_{g_Q g_Q}(\tau) \quad (2.20)$$

$$\phi_{g_I g_Q}(\tau) = -\phi_{g_Q g_I}(\tau). \quad (2.21)$$

The expectations in (2.18) and (2.19) are now calculated.

It is safe to assume that the phases $\phi_n(t)$ are statistically independent random variables at any time t , since the path delays τ_n are all independent due to the random placement of scatterers and the phases ϕ_n are also independent. Furthermore, the phases $\phi_n(t)$ at any time t can be treated as being uniformly distributed over the interval $[-\pi, \pi)$. The azimuth angles of arrival, θ_n are all independent due to the random placement of scatterers. In the limit as $N \rightarrow \infty$, the central limit theorem can be invoked and $g_I(t)$ and $g_Q(t)$ can be treated as Gaussian random processes. Also, in the limit as $N \rightarrow \infty$, the discrete azimuth angles of arrival θ_n can be replaced by a continuous random variable θ having the probability density function $p(\theta)$. By using these statistical properties, the autocorrelation function $\phi_{g_I g_I}(\tau)$ can be obtained from (2.15), (2.6), and (2.1) as follows:

$$\begin{aligned} \phi_{g_I g_I}(\tau) &= \lim_{N \rightarrow \infty} E_{\tau, \theta, \phi} [g_I(t)g_I(t + \tau)] \\ &= \frac{\Omega_p}{2} E_{\theta} [\cos(2\pi f_m \tau \cos(\theta))] \end{aligned} \quad (2.22)$$

where

$$\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_N) \quad (2.23)$$

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_N) \quad (2.24)$$

$$\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_N) \quad (2.25)$$

$$\Omega_p = E[|g(t)|^2] = E[g_I^2(t)] + E[g_Q^2(t)] = \sum_{n=1}^N C_n^2 \quad (2.26)$$

and Ω_p can be interpreted as the received envelope power, while $\Omega_p/2$ is the power received in the corresponding bandpass waveform in (2.14) as given by $\phi_{rr}(0)$ in (2.17).

Likewise, the cross-correlation function $\phi_{g_I g_Q}(\tau)$ can be derived as

$$\begin{aligned}\phi_{g_I g_Q}(\tau) &= \lim_{N \rightarrow \infty} E_{\tau, \theta, \phi} [g_I(t) g_Q(t + \tau)] \\ &= \frac{\Omega_p}{2} E_{\theta} [\sin(2\pi f_m \tau \cos(\theta))].\end{aligned}\quad (2.27)$$

Evaluation of the expectations in (2.22) and (2.27) requires the fraction of arriving power $p(\theta)$, and the antenna azimuth gain pattern $G(\theta)$, as a function of the azimuth angle θ . Here, the overall power angle distribution $p(\theta)G(\theta)$ is a probability density function that integrates to unity over all arrival angles, such that the total received envelope power is given by

$$\int_{-\pi}^{\pi} G(\theta)p(\theta)d\theta = 1. \quad (2.28)$$

One simple and commonly used model assumes that the plane waves arrive at the receiver antenna from all azimuth directions with equal probability, i.e., $p(\theta) = 1/(2\pi)$, $\theta \in [-\pi, \pi)$, and an isotropic antenna is used such that the antenna gain is uniform for all azimuth arrival angles, i.e., $G(\theta) = 1$, $\theta \in [-\pi, \pi)$. This model was first suggested by Clarke [74], and is commonly referred to as Clarke's 2-D isotropic scattering model. The expectation in (2.22) becomes

$$\begin{aligned}\phi_{g_I g_I}(\tau) &= \frac{\Omega_p}{2} \int_{-\pi}^{\pi} \cos(2\pi f_m \tau \cos(\theta)) p(\theta) G(\theta) d\theta \\ &= \frac{\Omega_p}{2} \frac{1}{\pi} \int_0^{\pi} \cos(2\pi f_m \tau \cos(\theta)) d\theta \\ &= \frac{\Omega_p}{2} J_0(2\pi f_m \tau),\end{aligned}\quad (2.29)$$

where

$$J_0(x) = \frac{1}{\pi} \int_0^{\pi} \cos(x \cos(\theta)) d\theta \quad (2.30)$$

is the zero-order Bessel function of the first kind. The normalized autocorrelation function $\phi_{g_I g_I}(\tau)/(\Omega_p/2)$ in (2.29) is plotted against the normalized time delay $f_m \tau$ in Fig. 2.5.

Likewise, for 2-D isotropic scattering and an isotropic receiver antenna, the cross-correlation function in (2.27) becomes

$$\phi_{g_I g_Q}(\tau) = \frac{\Omega_p}{4\pi} \int_{-\pi}^{\pi} \sin(2\pi f_m \tau \cos(\theta)) d\theta = 0. \quad (2.31)$$

A cross-correlation of zero 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 consequence of the symmetry of the 2-D isotropic scattering environment and the isotropic receiver antenna. In general, $g_I(t)$ and $g_Q(t)$ are correlated random processes for non-isotropic scattering environments and/or a non-isotropic receiver antenna.

2.1.2 Doppler Spectrum

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

$$\begin{aligned}\phi_{gg}(\tau) &= \frac{1}{2} E[g^*(t)g(t + \tau)] \\ &= \phi_{g_I g_I}(\tau) + j\phi_{g_I g_Q}(\tau),\end{aligned}\quad (2.32)$$

as derived in greater detail at the beginning of Sect. 4.9. The Doppler spectrum and autocorrelation function of the channel complex envelope are Fourier transform pairs, that is,

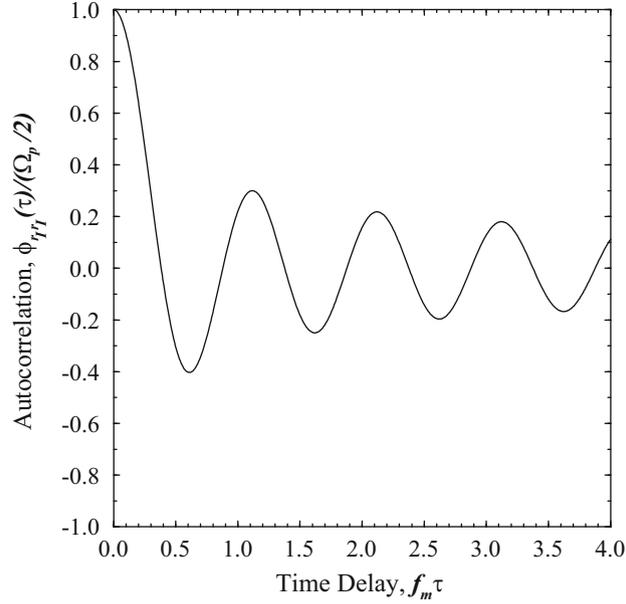


Fig. 2.5 Normalized autocorrelation function of the quadrature components of the received complex envelope with 2-D isotropic scattering and an isotropic receiver antenna

$$S_{gg}(f) = \int_{-\infty}^{\infty} \phi_{gg}(\tau) e^{-j2\pi f\tau} d\tau$$

$$\phi_{gg}(\tau) = \int_{-\infty}^{\infty} S_{gg}(f) e^{j2\pi f\tau} df,$$

where

$$S_{gg}(f) = S_{g_I g_I}(f) + jS_{g_I g_Q}(f), \quad (2.33)$$

is the Doppler spectrum. From (2.33), it may appear that the Doppler spectrum is complex valued. However, the autocorrelation function $\phi_{g_I g_I}(\tau)$, as with any autocorrelation function, is an even valued function such that $\phi_{g_I g_I}(\tau) = \phi_{g_I g_I}(-\tau)$. Hence, its Fourier transform $S_{g_I g_I}(f)$ is real valued. Moreover, the cross-correlation function $\phi_{g_Q g_I}(\tau)$ in (2.32), as with any cross-correlation function, satisfies the property

$$\phi_{g_I g_Q}(\tau) = \phi_{g_Q g_I}(-\tau). \quad (2.34)$$

Combining (2.34) with (2.21) yields the property

$$\phi_{g_I g_Q}(\tau) = -\phi_{g_I g_Q}(-\tau). \quad (2.35)$$

This means that the cross-correlation function $\phi_{g_I g_Q}(\tau)$ is an odd function of τ and, hence, its Fourier transform $S_{g_I g_Q}(f)$ is purely imaginary-valued. Therefore, the Doppler spectrum in (2.33) is always real valued.

For the autocorrelation function in (2.29), the corresponding psd is [147, 6.671.7]

$$S_{g_I g_I}(f) = \mathcal{F}[\phi_{g_I g_I}(\tau)]$$

$$= \begin{cases} \frac{\Omega_p}{2\pi f_m} \frac{1}{\sqrt{1-(f/f_m)^2}} & |f| \leq f_m \\ 0 & \text{otherwise} \end{cases}. \quad (2.36)$$

For the case of 2-D isotropic scattering and an isotropic receiver antenna $S_{g_Q g_I}(f) = 0$ so that $S_{gg}(f) = S_{g_I g_I}(f)$.

The power spectrum of the channel complex envelope $g(t)$ can also be related to that of the received bandpass process $r(t)$ in (2.14). From (2.17),

$$\phi_{rr}(\tau) = \text{Re} \{ \phi_{gg}(\tau) e^{j2\pi f_c \tau} \}. \quad (2.37)$$

As shown in Sect. 4.9, by using the identity

$$\text{Re} \{ z \} = \frac{z + z^*}{2} \quad (2.38)$$

and the property $\phi_{gg}(\tau) = \phi_{gg}^*(-\tau)$, the bandpass Doppler spectrum is

$$S_{rr}(f) = \frac{1}{2} (S_{gg}(f - f_c) + S_{gg}(-f - f_c)). \quad (2.39)$$

The Doppler spectrum $S_{gg}(f)$ is always a real-valued function of frequency, but not necessarily even. The corresponding bandpass Doppler spectrum $S_{rr}(f)$ is always real-valued and even.

2.1.2.1 Doppler Spectrum Alternate Approach

The Doppler spectrum can be derived by using a different approach that is sometimes useful because it can avoid the need to evaluate integrals. As mentioned earlier, the incident power on the receiver antenna is a function of the azimuth angle θ , such that the fraction of the total received power arriving between angles θ and $\theta + d\theta$ is $p(\theta)d\theta$. If the antenna has an azimuth gain of $G(\theta)$, then the fraction of the total power received between angles θ and $\theta + d\theta$ is $G(\theta)p(\theta)d\theta$. From (2.1), the Doppler frequency $f = f_m \cos(\theta)$ is an even function of the arrival angle θ . Therefore, the Doppler spectrum can be expressed as [171]

$$S_{gg}(f)|df| = \frac{\Omega_p}{2} (G(\theta)p(\theta) + G(-\theta)p(-\theta))|d\theta|. \quad (2.40)$$

From (2.1),

$$|df| = f_m |\sin(\theta)d\theta| = \sqrt{f_m^2 - f^2} |d\theta|. \quad (2.41)$$

Therefore,

$$S_{gg}(f) = \frac{\Omega_p}{2\sqrt{f_m^2 - f^2}} (G(\theta)p(\theta) + G(-\theta)p(-\theta)), \quad (2.42)$$

where

$$\theta = \cos^{-1}(f/f_m). \quad (2.43)$$

Hence, if $p(\theta)$ and $G(\theta)$ are known, the Doppler spectrum can be easily calculated. Once again, for 2-D isotropic scattering and an isotropic antenna $G(\theta)p(\theta) = 1/(2\pi)$, so that

$$S_{gg}(f) = \begin{cases} \frac{\Omega_p}{2\pi f_m} \frac{1}{\sqrt{1-(f/f_m)^2}} & |f| \leq f_m \\ 0 & \text{otherwise} \end{cases}. \quad (2.44)$$

The same result was obtained from the autocorrelation function in (2.29), but it required the evaluation of a Fourier transform integral.

The normalized Doppler spectrum $S_{gg}(f)/(\Omega_p/2\pi f_m)$ in (2.44) is plotted against the normalized Doppler frequency f/f_m in Fig. 2.6. Notice that $S_{gg}(f)$ is limited to the range of frequencies $0 \leq |f| \leq f_m$ and $S_{gg}(f) = \infty$ at $f = \pm f_m$. In reality, the Doppler spectrum is bounded, and the singular behavior at $f = \pm f_m$ is due to the assumption of 2-D plane wave propagation.

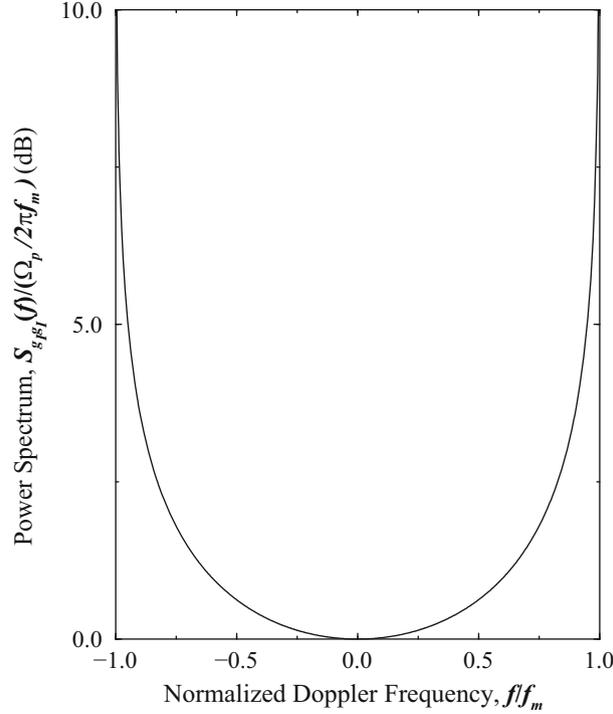


Fig. 2.6 Normalized psd of the quadrature components of the received complex envelope with 2-D isotropic scattering channel and an isotropic receiver antenna

Aulin [22] modified Clarke's 2-D model to yield a 3-D model that accounts for both azimuth and elevation angles-of-arrival. The resulting Doppler spectrum has the general same U-shape as Fig. 2.6, but does not have singularities at frequencies $f = \pm f_m$.

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

$$p(\theta) = \frac{1}{K+1} \hat{p}(\theta) + \frac{K}{K+1} \delta(\theta - \theta_0), \quad (2.45)$$

where $\hat{p}(\theta)$ is the continuous distribution of the *scatter* component, θ_0 is the angle-of-arrival (AoA) of the specular or LoS component, and K is the ratio of the received specular to scattered power. Figure 2.7 shows a plot of $p(\theta)$ for one such scattering environment, where the scatter component is characterized by 2-D isotropic scattering, i.e., $\hat{p}(\theta) = 1/(2\pi)$, $\theta \in [-\pi, \pi]$. The correlation functions $\phi_{gI}(\tau)$ and $\phi_{gQ}(\tau)$ corresponding to (2.45) can be readily obtained from (2.22) and (2.27) as

$$\phi_{gI}(\tau) = \frac{1}{K+1} \frac{\Omega_p}{2} J_0(2\pi f_m \tau) + \frac{K}{K+1} \frac{\Omega_p}{2} \cos(2\pi f_m \tau \cos(\theta_0)) \quad (2.46)$$

$$\phi_{gQ}(\tau) = \frac{K}{K+1} \frac{\Omega_p}{2} \sin(2\pi f_m \tau \cos(\theta_0)). \quad (2.47)$$

The azimuth distribution in (2.45) yields a complex envelope having a Doppler spectrum of the form

$$S_{gg}(f) = \frac{1}{K+1} S_{gg}^c(f) + \frac{K}{K+1} S_{gg}^d(f) \quad (2.48)$$

where $S_{gg}^d(f)$ is the discrete portion of the Doppler spectrum due to the specular component and $S_{gg}^c(f)$ is the continuous portion of the Doppler spectrum due to the scatter component. For the case when $\hat{p}(\theta) = 1/(2\pi)$, $\theta \in [-\pi, \pi]$, the correlation functions in (2.46) and (2.47) are obtained and the corresponding Doppler spectrum can be obtained, by using the cross-correlation property in (2.34) along with (2.33), as

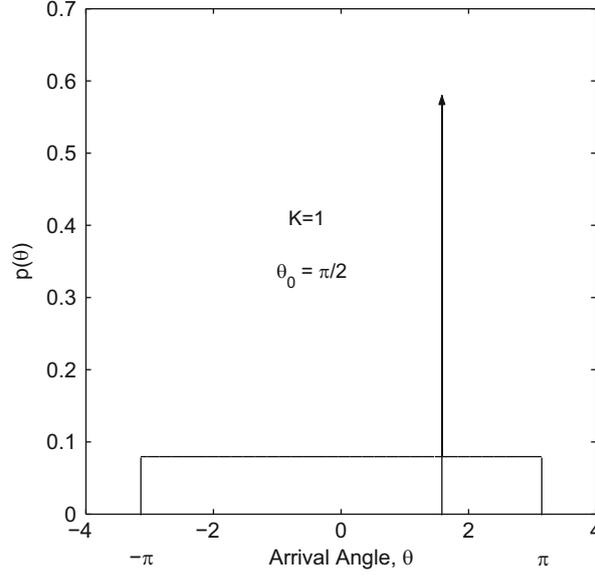


Fig. 2.7 Plot of $p(\theta)$ vs. θ with 2-D isotropic scattering plus a LoS or specular component arriving at angle $\theta_0 = \pi/2$

$$S_{gg}(f) = \begin{cases} \frac{1}{K+1} \cdot \frac{\Omega_p}{2\pi f_m} \frac{1}{\sqrt{1-(f/f_m)^2}} + \frac{K}{K+1} \frac{\Omega_p}{2} \delta(f - f_m \cos(\theta_0)) & 0 \leq |f| \leq f_m \\ 0 & \text{otherwise} \end{cases} \quad (2.49)$$

Note that the Doppler spectrum in (2.49) has the same shape as Fig. 2.6, except for the discrete spectral tone at frequency $f = f_m \cos(\theta_0)$.

Sometimes the azimuth distribution $p(\theta)$ may not be uniform, a condition commonly called non-isotropic scattering. Several distributions have been suggested to model non-isotropic scattering. One possibility is the Gaussian distribution

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left\{-\frac{(\theta - \mu)^2}{2\sigma_s^2}\right\}, \quad (2.50)$$

where μ is the mean AoA, and σ_s is the root mean square (rms) AoA spread. Another possibility is the von Mises distribution

$$p(\theta) = \frac{1}{2\pi I_0(k)} \exp\{k \cos(\theta - \mu)\}, \quad (2.51)$$

where $\theta \in [-\pi, \pi)$, $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind, $\mu \in [-\pi, \pi)$ is the mean AoA, and k controls the spread of scatterers around the mean. When $k = 0$, the von Mises distribution reduces to $p(\theta) = 1/(2\pi)$, yielding 2-D isotropic scattering. As k increases, the scatterers become more clustered around the mean AoA μ and the scattering becomes increasingly non-isotropic as shown in Fig. 2.8. Still another possibility is the cosine distribution

$$p(\theta) = \begin{cases} \frac{\pi}{4\theta_{\max}} \cos\left(\frac{\pi}{2} \frac{\theta}{\theta_{\max}}\right), & |\theta| \leq \theta_{\max} \leq \frac{\pi}{2} \\ 0 & \text{elsewhere} \end{cases} \quad (2.52)$$

The parameter θ_{\max} controls the AoA spread of the incoming waves. Figure 2.9 shows a plot of $p(\theta)$ for $\theta_{\max} = 30^\circ$, 60° , and 90° . Note that the distribution is symmetric about $\theta = 0$. Therefore, this azimuth distribution is less flexible than either the Gaussian or von Mises distributions. The density in (2.52) is sometimes used to model the *elevation* AoA distribution of scatterers in 3-D propagation models, where the mean elevation AoA is zero degrees [361].

Fig. 2.8 Plot of $p(\theta)$ vs. θ for the von Mises distribution with a mean angle-to-arrival $\mu = \pi/2$

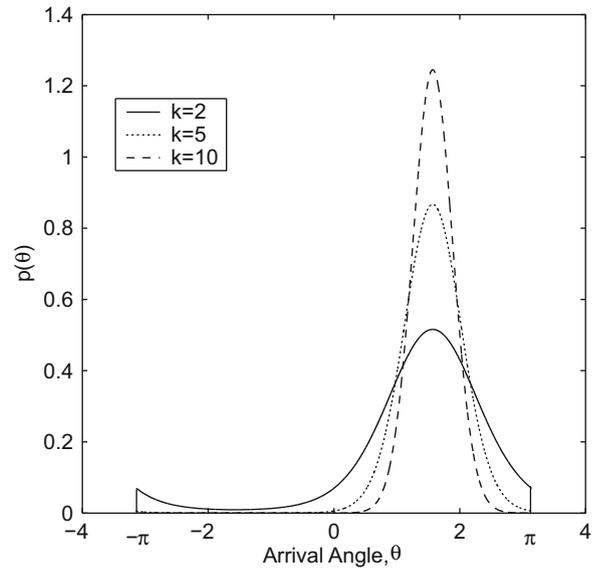
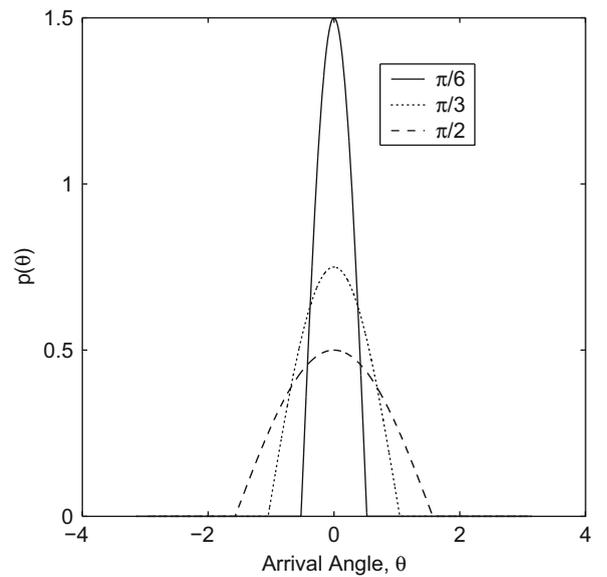


Fig. 2.9 Plot of $p(\theta)$ vs. θ for the angle-of-arrival pdf in (2.52) for various θ_{\max} . The mean angle of arrival is $\theta = 0$



Once the azimuth distribution is specified, 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 (2.22) and (2.27), respectively, using, for example, the densities in (2.50), (2.51), or (2.52). The Doppler spectrum, $S_{gg}(f)$, can be obtained by taking the Fourier transform of $\phi_{gg}(\tau)$ in (2.33) or, alternatively, by substituting the azimuth distribution $p(\theta)$ directly into (2.42).

2.1.3 Received Envelope and Phase Distribution

2.1.3.1 Rayleigh Fading

When the composite received signal consists of a large number of sinusoidal components, the received complex envelope $g(t) = g_I(t) + jg_Q(t)$ can be treated as a complex Gaussian random process. For some types of scattering environments, e.g., 2-D isotropic scattering, $g_I(t)$ and $g_Q(t)$ at any time t_1 are independent identically distributed Gaussian random variables with zero mean and variance $b_0 = E[g_I^2(t_1)] = E[g_Q^2(t_1)]$. Under these conditions the magnitude of the received complex envelope

$$\alpha \triangleq |g(t_1)| = \sqrt{g_I^2(t_1) + g_Q^2(t_1)} \quad (2.53)$$

has a Rayleigh distribution at any time t_1 , as shown in (A.26), i.e.,

$$p_\alpha(x) = \frac{x}{b_0} \exp\left\{-\frac{x^2}{2b_0}\right\}, \quad x \geq 0, \quad (2.54)$$

The received envelope power is equal to the second moment of α , $\Omega_p = E[\alpha^2] = 2b_0$, so that

$$p_\alpha(x) = \frac{2x}{\Omega_p} \exp\left\{-\frac{x^2}{\Omega_p}\right\}, \quad x \geq 0. \quad (2.55)$$

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

$$p_{\alpha^2}(x) = \frac{1}{\Omega_p} \exp\left\{-\frac{x}{\Omega_p}\right\}, \quad x \geq 0. \quad (2.56)$$

The squared-envelope is important for the performance analysis of digital communication systems because it is proportional to the received signal power and, hence, the received signal-to-noise ratio. The performance of digital communication links on flat fading channels will be discussed in more detail in Chap. 5.

2.1.3.2 Ricean Fading

As mentioned earlier, 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 it is again assumed that $g_I(t)$ and $g_Q(t)$ 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 $\alpha = |g(t_1)|$ at any time t_1 has a Ricean distribution as shown in (A.61), i.e.,

$$p_\alpha(x) = \frac{x}{b_0} \exp\left\{-\frac{x^2 + s^2}{2b_0}\right\} I_0\left(\frac{xs}{b_0}\right) \quad x \geq 0, \quad (2.57)$$

where

$$s^2 = m_I^2(t) + m_Q^2(t) \quad (2.58)$$

is called the non-centrality parameter. This type of fading is called Ricean fading and is often used to describe fading in environments where a LoS or strong specular path exists between the transmitter and receiver.

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 it lacks any physical meaning. A better model has been suggested by Aulin [22], such that the azimuth distribution $p(\theta)$ is defined in (2.45) and shown in Fig. 2.7. In this case, the LoS or specular component determines the means $m_I(t)$ and $m_Q(t)$ of the in-phase and quadrature components, respectively, as follows:

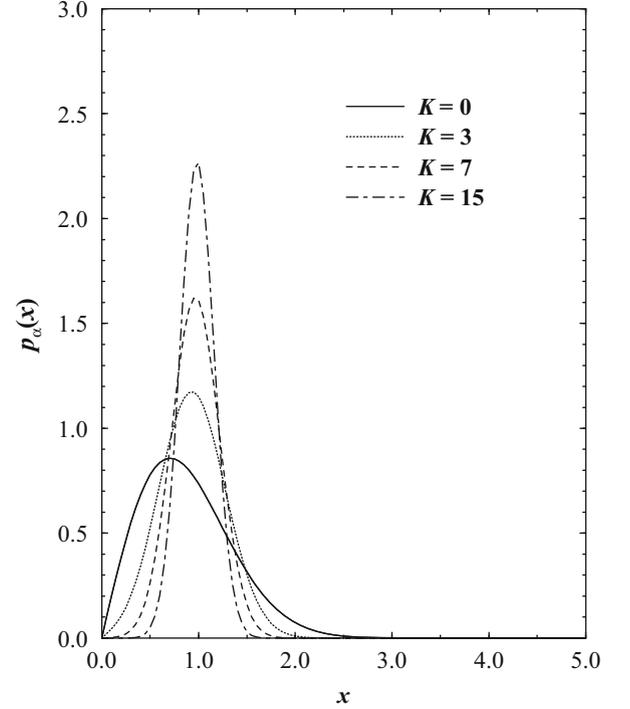
$$m_I(t) = s \cdot \cos(2\pi f_m \cos(\theta_0)t + \phi_0) \quad (2.59)$$

$$m_Q(t) = s \cdot \sin(2\pi f_m \cos(\theta_0)t + \phi_0), \quad (2.60)$$

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

The Rice factor, K , is defined as the ratio of the LoS or specular power s^2 to scattered power $2b_0$, i.e., $K = s^2/2b_0$. When $K = 0$, there is no LoS or specular component and the envelope exhibits Rayleigh fading. When $K = \infty$, there is no scatter component and 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 $\Omega_p = E[\alpha^2] = s^2 + 2b_0$ by first noting that

Fig. 2.10 The Rice pdf for several values of K with $\Omega_p = 1$



$$s^2 = \frac{K\Omega_p}{K+1}, \quad b_0 = \frac{\Omega_p}{2(K+1)}. \quad (2.61)$$

Substituting s^2 and b_0 into (2.57) yields

$$p_\alpha(x) = \frac{2(K+1)x}{\Omega_p} \exp\left\{-K - \frac{(K+1)x^2}{\Omega_p}\right\} I_0\left(2x\sqrt{\frac{K(K+1)}{\Omega_p}}\right), \quad x \geq 0. \quad (2.62)$$

Figure 2.10 shows the Rice pdf for several values of K . The curve for $K = 0$ is the Rayleigh pdf.

The squared-envelope $\alpha^2 = |g(t_1)|^2$ at any time t_1 has the following non-central chi-square distribution with two degrees of freedom:

$$p_{\alpha^2}(x) = \frac{(K+1)}{\Omega_p} \exp\left\{-K - \frac{(K+1)x}{\Omega_p}\right\} I_0\left(2\sqrt{\frac{K(K+1)x}{\Omega_p}}\right), \quad x \geq 0. \quad (2.63)$$

2.1.3.3 Nakagami Fading

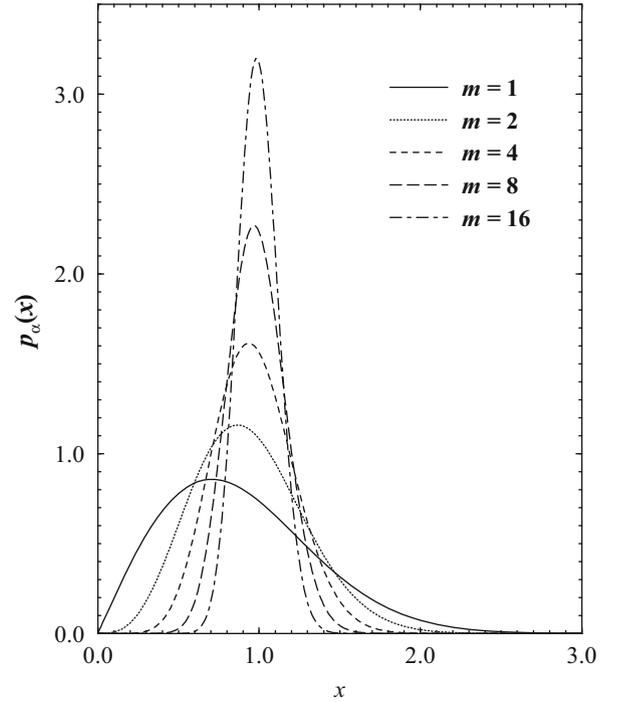
The Nakagami distribution was introduced by Nakagami in the early 1940s to characterize rapid fading in long distance HF channels [238]. The Nakagami distribution was selected to fit empirical data, and is known to provide a closer match to some measurement data than either the Rayleigh, Ricean, or log-normal distributions [57].

Nakagami fading describes the magnitude of the received complex envelope by the distribution

$$p_\alpha(x) = 2 \left(\frac{m}{\Omega_p}\right)^m \frac{x^{2m-1}}{\Gamma(m)} \exp\left\{-\frac{mx^2}{\Omega_p}\right\} \quad m \geq \frac{1}{2}, \quad (2.64)$$

where $\Omega_p = E[\alpha^2]$. Figure 2.11 shows the Nakagami distribution for several values of the shape factor, m . Beyond its empirical justification, the Nakagami distribution is often used 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 $m \rightarrow \infty$ the

Fig. 2.11 The Nakagami pdf for several values of m with $\Omega_p = 1$



distribution approaches 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 [238];

$$K \approx \sqrt{m^2 - m} + m - 1 \quad (2.65)$$

$$m \approx \frac{(K + 1)^2}{(2K + 1)}. \quad (2.66)$$

Finally, 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 may otherwise be intractable.

With Nakagami fading, the squared-envelope has the Gamma distribution

$$p_{\alpha^2}(x) = \left(\frac{m}{\Omega_p} \right)^m \frac{x^{m-1}}{\Gamma(m)} \exp \left\{ -\frac{mx}{\Omega_p} \right\}. \quad (2.67)$$

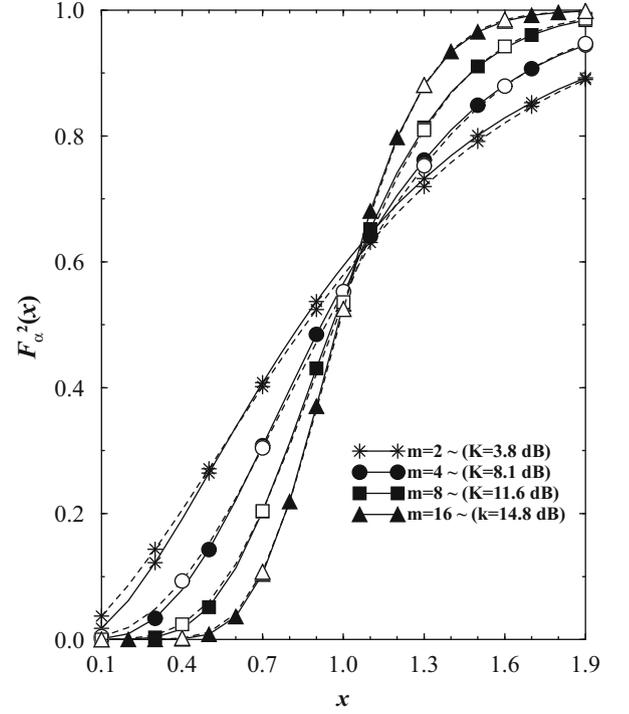
By using the relationship between the K factor and the shape factor m in (2.65), the cumulative distribution function (cdf), $F_{\alpha^2}(x) = P(\alpha^2 \leq x)$ of the squared-envelope with Nakagami and Ricean fading is plotted in Fig. 2.12. It is apparent from Fig. 2.12 that a Gamma distribution can approximate a non-central chi-square distribution to a reasonable degree of accuracy. However, the reader is cautioned that the tails of the pdf are often the most important. The reason being that the bit errors in a communication link tend occur during deep fades, which correspond to small values of x in the cdf. Figure 2.12 does not show how well the tails of a non-central chi-square pdf are approximated by a Gamma pdf.

2.1.3.4 Envelope Phase

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

$$\phi(t) = \text{Tan}^{-1} \left(\frac{g_Q(t)}{g_I(t)} \right). \quad (2.68)$$

Fig. 2.12 Comparison of the cdf of the squared-envelope with Ricean and Nakagami fading



For Rayleigh fading, $g_I(t_1)$ and $g_Q(t_1)$ are independent identically distributed zero-mean Gaussian random variables at any time t_1 . It follows (Appendix A.3.2.4) that the phase $\phi \equiv \phi(t_1)$ at any time t_1 is uniformly distributed over the interval $[-\pi, \pi)$, i.e.,

$$p_\phi(x) = \frac{1}{2\pi}, \quad -\pi \leq x \leq \pi. \quad (2.69)$$

For Ricean fading channels, the phase ϕ is not uniformly distributed and takes on a more complicated integral form, see Problem 2.9.

2.1.4 Envelope Correlation and Spectra

The autocorrelation of the envelope $\alpha(t) = |g(t)|$ of a complex Gaussian random process can be expressed in terms of the hypergeometric function $F(\cdot, \cdot; \cdot, \cdot)$ as [89]

$$\begin{aligned} \phi_{\alpha\alpha}(\tau) &= E[\alpha(t)\alpha(t+\tau)] \\ &= \frac{\pi}{2} |\phi_{gg}(0)| F\left(-\frac{1}{2}, -\frac{1}{2}; 1, \frac{|\phi_{gg}(\tau)|^2}{|\phi_{gg}(0)|^2}\right), \end{aligned} \quad (2.70)$$

where

$$|\phi_{gg}(\tau)|^2 = \phi_{g_I g_I}^2(\tau) + \phi_{g_Q g_Q}^2(\tau). \quad (2.71)$$

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

$$F\left(-\frac{1}{2}, -\frac{1}{2}; 1, x\right) = 1 + \frac{1}{4}x + \frac{1}{64}x^2 + \dots \quad (2.72)$$

Neglecting the terms beyond second order yields the approximation

$$\phi_{\alpha\alpha}(\tau) \doteq \frac{\pi}{2} |\phi_{gg}(0)| \left(1 + \frac{1}{4} \frac{|\phi_{gg}(\tau)|^2}{|\phi_{gg}(0)|^2} \right). \quad (2.73)$$

At $\tau = 0$, the approximation gives $\phi_{\alpha\alpha}(0) = 5\pi\Omega_p/8$, 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 $\lambda_{\alpha\alpha}(\tau)$ is of interest, where

$$\begin{aligned} \lambda_{\alpha\alpha}(\tau) &= E[\alpha(t)\alpha(t+\tau)] - E[\alpha(t)]E[\alpha(t+\tau)] \\ &= \frac{\pi}{2} |\phi_{gg}(0)| \left(1 + \frac{1}{4} \frac{|\phi_{gg}(\tau)|^2}{|\phi_{gg}(0)|^2} \right) - \frac{\pi}{2} |\phi_{gg}(0)| \\ &= \frac{\pi}{8|\phi_{gg}(0)|} |\phi_{gg}(\tau)|^2. \end{aligned} \quad (2.74)$$

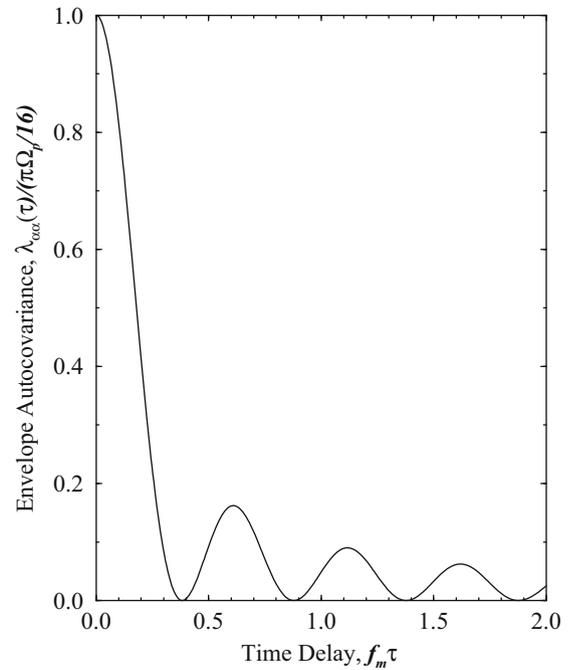
For 2-D isotropic scattering and an isotropic receiver antenna $|\phi_{gg}(\tau)|^2 = \phi_{gigt}^2(\tau)$ and, therefore,

$$\lambda_{\alpha\alpha}(\tau) = \frac{\pi\Omega_p}{16} J_0^2(2\pi f_m \tau). \quad (2.75)$$

Figure 2.13 plots the normalized envelope autocovariance $\lambda_{\alpha\alpha}(\tau)/(\pi\Omega_p/16)$ against the normalized time delay $f_m\tau$ for the case of 2-D isotropic scattering, $p(\theta) = 1/(2\pi)$, $-\pi \leq \theta \leq \pi$ and an isotropic receiver antenna $G(\theta) = 1$.

The Fourier transform of $\lambda_{\alpha\alpha}(\tau)$ yields the continuous portion $S_{\alpha\alpha}^c(f)$ of the envelope psd $S_{\alpha\alpha}(f)$, and 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

Fig. 2.13 Normalized envelope autocovariance against the normalized time delay $f_m\tau$ for a 2-D isotropic scattering with an isotropic receiver antenna



$$\begin{aligned}
S_{\alpha\alpha}^c(f) &= \frac{\pi}{8|\phi_{gg}(0)|} S_{gg}(f) * S_{gg}(f) \\
&= \frac{\pi}{8|\phi_{gg}(0)|} \int_{-\infty}^{\infty} S_{gg}(x) S_{gg}(x-f) dx \\
&= \frac{\pi}{8|\phi_{gg}(0)|} \int_{-f_m}^{f_m-|f|} S_{gg}(x) S_{gg}(x+|f|) dx, \quad 0 \leq |f| \leq 2f_m.
\end{aligned} \tag{2.76}$$

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, $S_{gg}(f)$ needs to be specified. With 2-D isotropic scattering and an isotropic receiver antenna $S_{gg}(f) = S_{gigl}(f)$, where $S_{gigl}(f)$ is given by (2.36). The result from evaluating (2.76) is (see Problem 2.10)

$$S_{\alpha\alpha}^c(f) = \frac{\Omega_p}{16\pi} \frac{1}{f_m} K \left(\sqrt{1 - \left(\frac{f}{2f_m} \right)^2} \right) \quad 0 \leq |f| \leq 2f_m, \tag{2.77}$$

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

$$K(\gamma) = \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-\gamma^2x^2)}}. \tag{2.78}$$

The continuous portion of the normalized envelope psd $S_{\alpha\alpha}(f)/(\Omega_p/16\pi f_m)$ is plotted against the normalized frequency f/f_m in Fig. 2.14.

The psd of the received envelope $\alpha(t)$ for a non-isotropic scattering channel can be obtained using the above procedure. For example, consider the particular scattering environment shown in Fig. 2.7 with the associated Doppler spectrum in (2.49). To obtain the continuous portion of the psd of the envelope $\alpha(t)$, substitute (2.49) into (2.76) to obtain (see Problem 2.13)

Fig. 2.14 Continuous portion of the normalized envelope psd against the normalized frequency f/f_m for a 2-D isotropic scattering channel with an isotropic receiver antenna

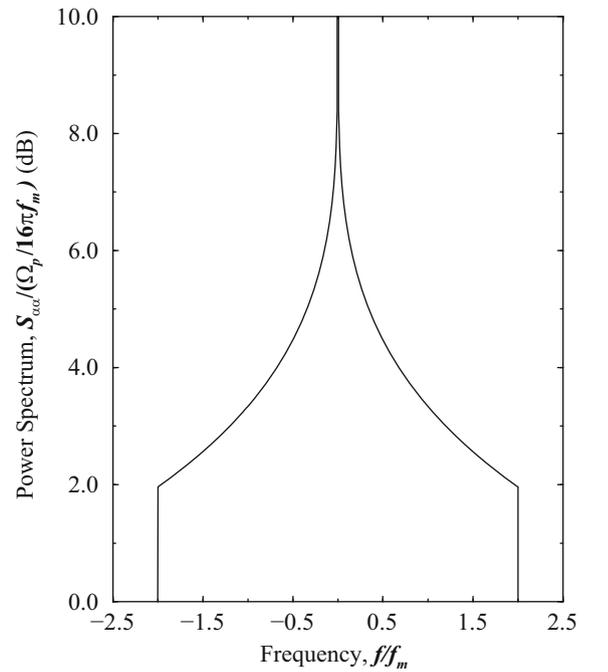
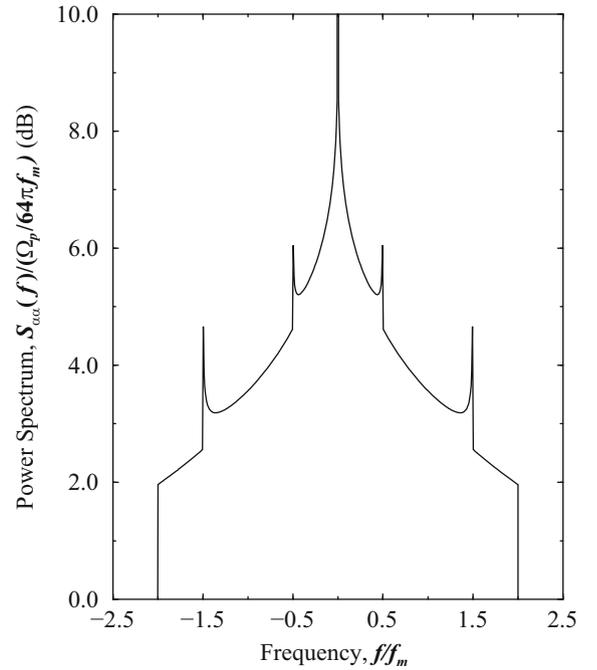


Fig. 2.15 Continuous portion of the normalized envelope psd against the normalized frequency f/f_m for the scattering environment shown in Fig. 2.7; $K = 10$ and $\theta_0 = \pi/3$



$$S_{\alpha\alpha}^c(f) = \left(\frac{1}{K+1}\right)^2 \frac{\Omega_p}{16\pi f_m} \left\{ K \left(\sqrt{1 - \left(\frac{f}{2f_m}\right)^2} \right) + \frac{K\pi}{\sqrt{1 - (f/f_m + \cos(\theta_0))^2}} + \frac{K\pi}{\sqrt{1 - (f/f_m - \cos(\theta_0))^2}} + K^2\pi^2 f_m \delta(f) \right\}. \quad (2.79)$$

Figure 2.15 shows a plot of the continuous portion 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 for $K = 10$ and $\theta_0 = \pi/3$.

2.1.4.1 Squared-Envelope Correlation and Spectrum

The autocorrelation of the squared-envelope is

$$\phi_{\alpha^2\alpha^2}(\tau) = E[\alpha^2(t)\alpha^2(t+\tau)]. \quad (2.80)$$

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

$$\phi_{\alpha^2\alpha^2}(\tau) = E[g_I^2(t)g_I^2(t+\tau)] + E[g_Q^2(t)g_Q^2(t+\tau)] + E[g_I^2(t)g_Q^2(t+\tau)] + E[g_Q^2(t)g_I^2(t+\tau)]. \quad (2.81)$$

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 2.14)

$$\begin{aligned} \phi_{\alpha^2\alpha^2}(\tau) &= 4\phi_{g_I g_I}^2(0) + 4\phi_{g_I g_I}^2(\tau) + 4\phi_{g_I g_Q}^2(\tau) \\ &= 4\phi_{g_I g_I}^2(0) + 4|\phi_{g g}(\tau)|^2. \end{aligned} \quad (2.82)$$

Finally, the squared-envelope autocovariance is

$$\begin{aligned} \lambda_{\alpha^2\alpha^2}(\tau) &= \phi_{\alpha^2\alpha^2}(\tau) - E[\alpha^2(t)] \\ &= 4|\phi_{g g}(\tau)|^2. \end{aligned} \quad (2.83)$$

With 2-D isotropic scattering and an isotropic receiver antenna, the above expression reduces to

$$\lambda_{\alpha^2\alpha^2}(\tau) = \Omega_p^2 J_0^2(2\pi f_m \tau). \quad (2.84)$$

By comparing (2.74) and (2.83), observe that the *approximate* autocorrelation function of the envelope and the *exact* autocorrelation function 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 has a more complicated form. Let

$$g_I(t) = \hat{g}_I(t) + m_I(t) \quad (2.85)$$

$$g_Q(t) = \hat{g}_Q(t) + m_Q(t), \quad (2.86)$$

where $m_I(t)$ and $m_Q(t)$ are the means of $g_I(t)$ and $g_Q(t)$, respectively. From Problem 2.15,

$$\begin{aligned} \phi_{\alpha^2\alpha^2}(\tau) = & 4|\phi_{\hat{g}\hat{g}}(\tau)|^2 + 4\phi_{\hat{g}_I\hat{g}_I}^2(0) + 4\text{Re} \{m(t)m^*(t+\tau)\phi_{\hat{g}\hat{g}}(\tau)\} \\ & + 2(|m(t)|^2 + |m(t+\tau)|^2)\phi_{\hat{g}_I\hat{g}_I}(0) + |m(t)|^2|m(t+\tau)|^2, \end{aligned} \quad (2.87)$$

where

$$m(t) = m_I(t) + jm_Q(t) \quad (2.88)$$

$$m(t+\tau) = m_I(t+\tau) + jm_Q(t+\tau). \quad (2.89)$$

The squared-envelope autocovariance is

$$\lambda_{\alpha^2\alpha^2}(\tau) = 4|\phi_{\hat{g}\hat{g}}(\tau)|^2 + 4\text{Re} \{m(t)m^*(t+\tau)\phi_{\hat{g}\hat{g}}(\tau)\}. \quad (2.90)$$

Consider the scattering environment shown in Fig. 2.7. The corresponding correlation functions $\phi_{g_I g_I}(\tau)$ and $\phi_{g_I g_Q}(\tau)$ are given by (2.46) and (2.47), respectively, and the means $m_I(t)$ and $m_Q(t)$ are defined in (2.59) and (2.60). It can be shown that

$$\phi_{\hat{g}_I\hat{g}_I}(\tau) = \frac{1}{K+1} \frac{\Omega_p}{2} J_0(2\pi f_m \tau) \quad (2.91)$$

$$\phi_{\hat{g}_I\hat{g}_Q}(\tau) = 0 \quad (2.92)$$

and

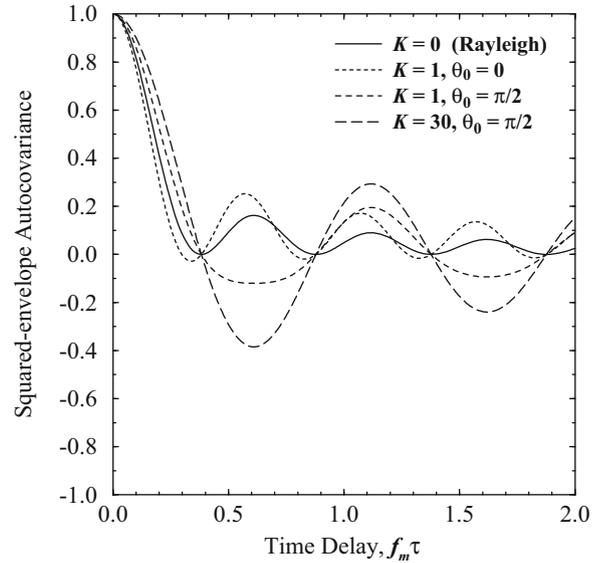
$$\begin{aligned} m(t)m^*(t+\tau) = & s^2 \left(\cos(2\pi f_m \tau \cos(\theta_0)) - j \sin(2\pi f_m \tau \cos(\theta_0)) \right) \\ = & \frac{K\Omega_p}{K+1} \left(\cos(2\pi f_m \tau \cos(\theta_0)) - j \sin(2\pi f_m \tau \cos(\theta_0)) \right), \end{aligned} \quad (2.93)$$

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 (2.90) gives

$$\lambda_{\alpha^2\alpha^2}(\tau) = \left(\frac{\Omega_p}{K+1} \right)^2 J_0(2\pi f_m \tau) \left(J_0(2\pi f_m \tau) + 2K \cos(2\pi f_m \tau \cos(\theta_0)) \right). \quad (2.94)$$

The normalized squared-envelope autocovariance is plotted in Fig. 2.16 as a function of the normalized time delay $f_m \tau$ for various values of K and θ_0 .

Fig. 2.16 Squared-envelope autocovariance against the normalized time delay $f_m \tau$ for the scattering environment shown in Fig. 2.7



2.1.5 Level Crossing Rate and Fade Duration

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 on average below a specified level). Also of interest is the zero crossing rate of the real and imaginary components of the complex faded envelope. These quantities are affected not only by the scattering environment but also by the velocity of the MS.

2.1.5.1 Envelope Level Crossing Rate

The envelope level crossing rate at a specified envelope level R , L_R , is defined as the rate (in crossings per second) at which the envelope α crosses the 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 $\alpha = |g(t_1)|$ and the envelope slope $\dot{\alpha} = d|g(t_1)|/dt$ at any time instant t_1 .¹ In terms of the joint pdf $p(\alpha, \dot{\alpha})$, the expected amount of time the envelope lies in the interval $(R, R + d\alpha)$ for a given envelope slope $\dot{\alpha}$ and time increment dt is

$$p(R, \dot{\alpha}) d\alpha d\dot{\alpha} dt. \quad (2.95)$$

The time required for the envelope α to traverse the interval $(R, R + d\alpha)$ *once* for a given envelope slope $\dot{\alpha}$ is

$$d\alpha / \dot{\alpha}. \quad (2.96)$$

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 increment dt , viz.

$$\dot{\alpha} p(R, \dot{\alpha}) d\dot{\alpha} dt. \quad (2.97)$$

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

$$\int_0^T \dot{\alpha} p(R, \dot{\alpha}) d\dot{\alpha} dt = \dot{\alpha} p(R, \dot{\alpha}) d\dot{\alpha} T. \quad (2.98)$$

¹For simplicity of notation, the time variable is suppressed as $\alpha \equiv \alpha(t_1)$ and $\dot{\alpha} \equiv \dot{\alpha}(t_1)$.

The expected number of crossings of the envelope level R with a positive slope in the time interval T is

$$N_R = T \int_0^\infty \dot{\alpha} p(R, \dot{\alpha}) d\dot{\alpha}. \quad (2.99)$$

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

$$L_R = \int_0^\infty \dot{\alpha} p(R, \dot{\alpha}) d\dot{\alpha}. \quad (2.100)$$

This is actually a general result that applies to any random process characterized by the joint pdf $p(\alpha, \dot{\alpha})$.

Rice has derived the joint pdf $p(\alpha, \dot{\alpha})$ for a sine wave in bandpass Gaussian noise [283]. A Ricean fading channel similarly consists of LoS or specular (sine wave) component plus a scatter (Gaussian noise) component. For the case of a Ricean fading channel,

$$p(\alpha, \dot{\alpha}) = \frac{\alpha(2\pi)^{-3/2}}{\sqrt{Bb_0}} \int_{-\pi}^{\pi} \exp \left\{ -\frac{1}{2Bb_0} \left[B(\alpha^2 - 2\alpha s \cos(\theta) + s^2) + (b_0\dot{\alpha} + b_1 s \sin(\theta))^2 \right] \right\} d\theta \quad (2.101)$$

where s is the non-centrality parameter in the Rice distribution, and $B = b_0 b_2 - b_1^2$, and where b_0 , b_1 , and b_2 are constants that are derived from the power spectrum of the scatter component. For the scattering environment described by (2.45) and Fig. 2.7, the sine wave corresponds to the specular component arriving at angle θ_0 , while the Gaussian noise is due to the scatter component with azimuth distribution $\hat{p}(\theta) = 1/(2\pi)$, $-\pi \leq \theta \leq \pi$. Note that the joint pdf derived by Rice in (2.101) is general enough to apply to scattering environments described by other $\hat{p}(\theta)$ as well.

Suppose that the specular or LoS component of the complex envelope $g(t)$ has a Doppler frequency equal $f_q = f_m \cos(\theta_0)$, where $0 \leq |f_q| \leq f_m$. In this case [283]

$$b_n = (2\pi)^n \int_{-f_m}^{f_m} S_{gg}^c(f) (f - f_q)^n df \quad (2.102)$$

$$= (2\pi)^n b_0 \int_0^{2\pi} \hat{p}(\theta) G(\theta) (f_m \cos(\theta) - f_q)^n d\theta, \quad (2.103)$$

where $\hat{p}(\theta)$ is the azimuth distribution of the *scatter* component, $G(\theta)$ is the antenna gain pattern, and $S_{gg}^c(f)$ is the corresponding continuous portion of the Doppler power spectrum. Equivalence between (2.102) and (2.103) can be established by using (2.42). Note that $S_{gg}^c(f)$ is given by the Fourier transform of

$$\phi_{gg}^c(\tau) = \phi_{g_{gl}}^c(\tau) - j\phi_{g_{lq}}^c(\tau), \quad (2.104)$$

where

$$\phi_{g_{gl}}^c(\tau) = \frac{\Omega_p}{2} \int_0^{2\pi} \cos(2\pi f_m \tau \cos(\theta)) \hat{p}(\theta) G(\theta) d\theta \quad (2.105)$$

$$\phi_{g_{lq}}^c(\tau) = \frac{\Omega_p}{2} \int_0^{2\pi} \sin(2\pi f_m \tau \cos(\theta)) \hat{p}(\theta) G(\theta) d\theta. \quad (2.106)$$

In some special cases, the psd $S_{gg}^c(f)$ is symmetrical about the frequency $f_q = f_m \cos(\theta_0)$. This condition occurs, for example, when $f_q = 0$ ($\theta_0 = 90^\circ$), $\hat{p}(\theta) = 1/(2\pi)$, $-\pi \leq \theta \leq \pi$, and $G(\theta) = 1$. In this case, $b_n = 0$ for all odd values of n (and in particular $b_1 = 0$) so that (2.101) reduces to the convenient product form

$$\begin{aligned} p(\alpha, \dot{\alpha}) &= \sqrt{\frac{1}{2\pi b_2}} \exp \left\{ -\frac{\dot{\alpha}^2}{2b_2} \right\} \cdot \frac{\alpha}{b_0} \exp \left\{ -\frac{(\alpha^2 + s^2)}{2b_0} \right\} I_0 \left(\frac{\alpha s}{b_0} \right) \\ &= p(\dot{\alpha}) \cdot p(\alpha). \end{aligned} \quad (2.107)$$

Since $p(\alpha, \dot{\alpha}) = p(\dot{\alpha}) \cdot p(\alpha)$ in (2.107), it follows that α and $\dot{\alpha}$ are statistically independent. When $f_q = 0$ and $\hat{p}(\theta) = 1/(2\pi)$, a closed form expression can be obtained for the envelope level crossing rate. Substituting (2.44) into (2.102) gives

$$b_n = \begin{cases} b_0 (2\pi f_m)^n \frac{1 \cdot 3 \cdot 5 \cdots (n-1)}{2 \cdot 4 \cdot 6 \cdots n} & n \text{ even} \\ 0 & n \text{ odd} \end{cases}. \quad (2.108)$$

Therefore, $b_1 = 0$ and $b_2 = b_0(2\pi f_m)^2/2$. By substituting the joint density in (2.107) into (2.100) and using the expression for b_0 in (2.61), the envelope level crossing rate is obtained as

$$L_R = \sqrt{2\pi(K+1)} f_m \rho e^{-K-(K+1)\rho^2} I_0\left(2\rho\sqrt{K(K+1)}\right), \quad (2.109)$$

where

$$\rho = \frac{R}{\sqrt{\Omega_p}} = \frac{R}{R_{\text{rms}}} \quad (2.110)$$

and $R_{\text{rms}} \triangleq \sqrt{E[\alpha^2]}$ is the rms envelope level. Under the further condition that $K = 0$ (Rayleigh fading), the above expression simplifies to

$$L_R = \sqrt{2\pi} f_m \rho e^{-\rho^2}. \quad (2.111)$$

Notice that the level crossing rate is directly proportional to the maximum Doppler frequency f_m and, hence, the MS speed $v = f_m \lambda_c$. The normalized level crossing rate L_R/f_m in (2.109) is plotted in Fig. 2.17 as a function of ρ and K .

2.1.5.2 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 or LoS component, then $g_I(t)$ and $g_Q(t)$ have mean values $m_I(t)$ and $m_Q(t)$, respectively. Of interest, is the zero crossing rate of the zero-mean Gaussian random processes $\hat{g}_I(t) = g_I(t) - m_I(t)$ and $\hat{g}_Q(t) = g_Q(t) - m_Q(t)$. Rice [283] has derived this zero crossing rate as

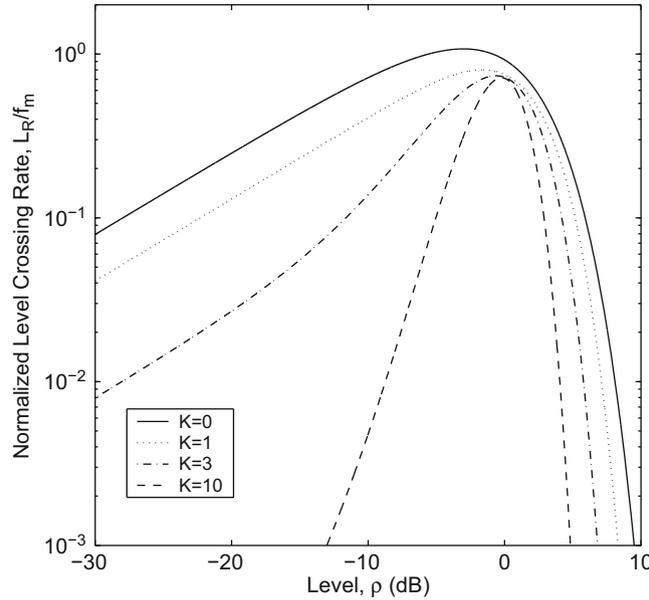


Fig. 2.17 Normalized level crossing rate for the scattering environment shown in Fig. 2.7 with $\theta_0 = 90^\circ$

$$L_Z = \frac{1}{\pi} \sqrt{\frac{b_2}{b_0}}. \quad (2.112)$$

When the scatter component has the azimuth distribution $\hat{p}(\theta) = 1/(2\pi)$, $-\pi \leq \theta \leq \pi$, the zero crossing rate is

$$L_Z = \sqrt{2}f_m. \quad (2.113)$$

Similar to the level crossing rate, the zero crossing rate is directly proportional to the maximum Doppler frequency f_m .

2.1.5.3 Average Fade Duration

Another quantity of interest is the average duration that the envelope 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 i th fade below the level R . The probability that the received envelope is less than R is

$$P[\alpha \leq R] = \frac{1}{T} \sum_i t_i. \quad (2.114)$$

The average fade duration is equal to

$$\bar{t} = \frac{\sum_i t_i}{TL_R} = \frac{P[\alpha \leq R]}{L_R}. \quad (2.115)$$

If the envelope is Ricean distributed as in (2.62), then

$$P(\alpha \leq R) = \int_0^R p(\alpha) d\alpha = 1 - Q\left(\sqrt{2K}, \sqrt{2(K+1)\rho^2}\right), \quad (2.116)$$

where $Q(a, b)$ is the Marcum Q function. Moreover, if it is again assumed that $f_q = 0$ and $\hat{p}(\theta) = 1/(2\pi)$, then

$$\bar{t} = \frac{1 - Q\left(\sqrt{2K}, \sqrt{2(K+1)\rho^2}\right)}{\sqrt{2\pi(K+1)}f_m\rho e^{-K-(K+1)\rho^2} I_0\left(2\rho\sqrt{K(K+1)}\right)}. \quad (2.117)$$

If $K = 0$ (Rayleigh fading), then

$$P[\alpha \leq R] = \int_0^R p(\alpha) d\alpha = 1 - e^{-\rho^2} \quad (2.118)$$

and

$$\bar{t} = \frac{e^{\rho^2} - 1}{\rho f_m \sqrt{2\pi}}. \quad (2.119)$$

The normalized average fade duration $\bar{t}f_m$ in (2.117) is plotted in Fig. 2.18 as a function of ρ .

The level crossing rate, zero crossing rate, and average fade duration all depend on the velocity of the MS, since $f_m = v/\lambda_c$. 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 $f_m = 80$ Hz. Therefore, with 2-D isotropic scattering and Rayleigh fading ($K = 0$) there are $L_R = 74$ fades/s at $\rho = 0$ dB with an average fade duration of 8.5 ms. However, at $\rho = -20$ dB there are only 20 fades/s with an average fade duration of 0.5 ms. Note that since ρ represents a normalized envelope (magnitude) level, $\rho_{(\text{dB})} = 20 \log_{10} \rho$. Also observe from Fig. 2.17 that for small $\rho_{(\text{dB})}$, the level crossing rate decreases with an increasing Rice factor K . Furthermore, from Fig. 2.18, for small $\rho_{(\text{dB})}$ the average fade duration increases with increasing K . Hence, as K increases the fades occur less frequently, but last longer when they do occur.

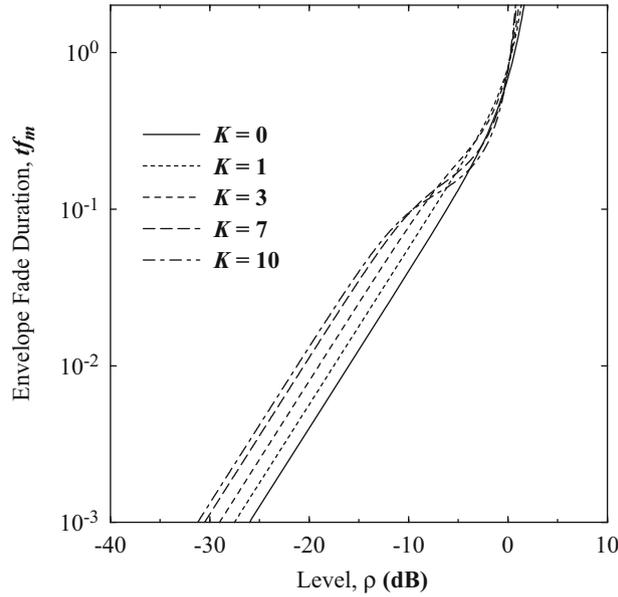


Fig. 2.18 Normalized average fade duration for the scattering environment shown in Fig. 2.7 with $\theta_0 = 90^\circ$

2.1.6 Space-Time Correlation

Many mobile radio systems employ receiver antenna diversity, where spatially separated receiver antennas are used to provide multiple faded replicas of the same information bearing signal. In order for such diversity systems to provide the maximum diversity gain, it is desirable that the multiple faded replicas experience uncorrelated fading. The spatial correlation characteristics are needed for determining the required spatial separation between antenna elements so that they are sufficiently decorrelated. Moreover, it is sometimes desirable to simultaneously characterize both the spatial and temporal channel correlation characteristics. For this purpose, the space-time correlation function is useful. To obtain the space-time correlation functions, it is necessary to specify the scattering geometry. One possibility for NLoS conditions is the single-ring model shown in Fig. 2.19, where the BS and MS are located at O_B and O_M , respectively, and separated by distance D , and the scatterers are assumed to be located on a ring of radius R centered around the MS. Very often, a local scattering assumption is made such that $D \gg R$, in order to simplify and obtain closed form expressions for the various space-time correlation functions. For short distances between the transmitter and receiver, or large scattering radii, the local scattering assumption is not applicable.

2.1.6.1 Space-Time Correlation at the Mobile Station

Consider two MS antennas, $A_M^{(q)}$, $q = 1, 2$, separated by distance δ_M . The MS antenna array is oriented with angle θ_M with respect to the x -axis, and the MS moves with velocity v and angle γ_M with respect to the x -axis. For the environment shown in Fig. 2.19, the channel from O_B to $A_M^{(q)}$ has the complex envelope

$$g_q(t) = \sum_{n=1}^N C_n e^{j\phi_n - j2\pi(\epsilon_n + \epsilon_{nq})/\lambda_c} e^{j2\pi f_m t \cos(\alpha_M^{(n)} - \gamma_M)}, \quad q = 1, 2, \quad (2.120)$$

where ϵ_n and ϵ_{nq} denote the distances $O_B - S_n$ and $S_M^{(n)} - A_M^{(q)}$, $q = 1, 2$, respectively. From the law of cosines, the distances ϵ_n and ϵ_{nq} can be expressed as a function of the angle-of-arrival $\alpha_M^{(n)}$ as follows:

$$\epsilon_n^2 = D^2 + R^2 + 2DR \cos(\alpha_M^{(n)}) \quad (2.121)$$

$$\epsilon_{nq}^2 = ((1.5 - q)\delta_M)^2 + R^2 - 2(1.5 - q)\delta_M R \cos(\alpha_M^{(n)} - \theta_M), \quad q = 1, 2. \quad (2.122)$$

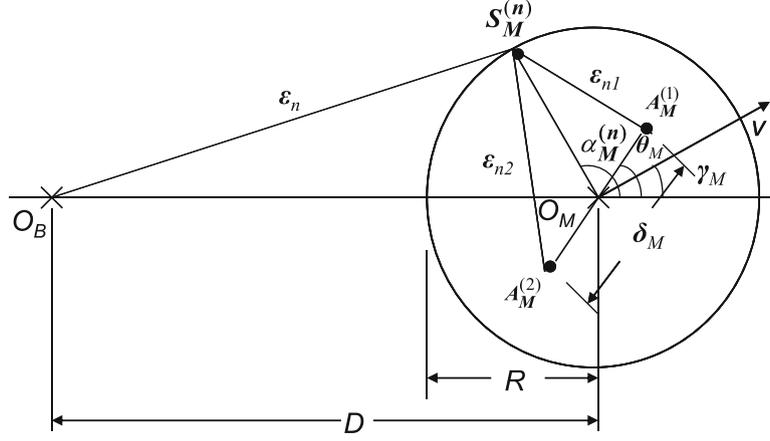


Fig. 2.19 Single-ring scattering model for NLoS propagation on the forward link of a cellular system. The MS is surrounded by a scattering ring of radius R and is at distance D from the BS, where $R \ll D$

Assuming that $R/D \ll 1$, $\delta_M \ll R$ and using the approximation $\sqrt{1 \pm x} \approx 1 \pm x/2$ for small x , gives

$$\epsilon_n \approx D + R \cos(\alpha_M^{(n)}) \quad (2.123)$$

$$\epsilon_{nq} \approx R - (1.5 - q)\delta_M \cos(\alpha_M^{(n)} - \theta_M), \quad q = 1, 2. \quad (2.124)$$

Substituting (2.123) and (2.124) into (2.120) gives

$$g_q(t) = \sum_{n=1}^N C_n e^{j\phi_n - j2\pi(D + R \cos(\alpha_M^{(n)} + R - (1.5 - q)\delta_M \cos(\alpha_M^{(n)} - \theta_M))/\lambda_c + j2\pi f_m t \cos(\alpha_M^{(n)} - \gamma_M)}, \quad q = 1, 2. \quad (2.125)$$

The space-time correlation function between the two complex faded envelopes $g_1(t)$ and $g_2(t)$ is

$$\phi_{g_1, g_2}(\delta_M, \tau) = \frac{1}{2} \mathbb{E} [g_1^*(t) g_2(t + \tau)]. \quad (2.126)$$

Using (2.125) and (2.126), the space-time correlation function between $g_1(t)$ and $g_2(t)$ can be written as

$$\phi_{g_1, g_2}(\delta_M, \tau) = \frac{\Omega_p}{2N} \sum_{n=1}^N \mathbb{E} \left[e^{-j2\pi(\delta_M/\lambda_c) \cos(\alpha_M^{(n)} - \theta_M) + j2\pi f_m \tau \cos(\alpha_M^{(n)} - \gamma_M)} \right]. \quad (2.127)$$

Since the number of scatters is infinite, the discrete angles-of-arrival $\alpha_M^{(n)}$ can be replaced with a continuous random variable α_M with probability density function $p(\alpha_M)$. Hence, the space-time correlation function becomes

$$\phi_{g_1, g_2}(\delta_M, \tau) = \frac{\Omega_p}{2} \int_0^{2\pi} e^{-jb \cos(\alpha_M - \theta_M)} e^{ja \cos(\alpha_M - \gamma_M)} p(\alpha_M) d\alpha_M, \quad (2.128)$$

where $a = 2\pi f_m \tau$ and $b = 2\pi \delta_M / \lambda_c$.

For the case of 2-D isotropic scattering with isotropic MS antenna, $p(\alpha_M) = 1/(2\pi)$, $-\pi \leq \alpha_M \leq \pi$, and the space-time correlation function becomes

$$\phi_{g_1, g_2}(\delta_M, \tau) = \frac{\Omega_p}{2} J_0 \left(\sqrt{a^2 + b^2 - 2ab \cos(\theta_M - \gamma_M)} \right). \quad (2.129)$$

The spatial and temporal correlation functions can be obtained by setting $\tau = 0$ and $\delta_M = 0$, respectively. This gives $\phi_{g_1, g_2}(\delta_M) = \phi_{g_1, g_2}(\delta_M, 0) = \frac{\Omega_p}{2} J_0(2\pi \delta_M / \lambda_c)$ and $\phi_{gg}(\tau) = \phi_{g_1, g_2}(0, \tau) = \frac{\Omega_p}{2} J_0(2\pi f_m \tau)$, which matches our earlier result

in (2.29) as expected. For the case of 2-D isotropic scattering and an isotropic MS antenna, it follows that Fig. 2.13 also plots the normalized envelope spatial autocovariance function $\lambda_{\alpha\alpha}(\ell)/(\pi\Omega_p/16)$ against the normalized spatial separation $f_m\tau = \delta_M/\lambda_c$. The spatial autocovariance function is zero at $\delta_M/\lambda_c = 0.38$ and is less than 0.3 for $\delta_M/\lambda_c > 0.38$. The implication is that, under conditions of 2-D isotropic scattering and isotropic MS antennas, sufficient spatial decorrelation can be obtained by spacing the MS antenna elements roughly a half-wavelength apart.

2.1.6.2 Space-Time Correlation at the Base Station

Radio channels are reciprocal in the sense that if a propagation path exists, it will carry energy equally well in either the uplink or downlink directions. That is, the plane waves in either direction will propagate by exactly the same set of scatterers. Therefore, the temporal autocorrelation functions and Doppler spectra are expected to be the same for both the uplink and downlink directions. However, for cellular land mobile radio applications, most of the scatters are in the vicinity of the MS, while the BS antennas are elevated and free of local scatters. Consequently, the plane waves will arrive at the BS antennas with a narrow angle-of-arrival spread, whereas they arrive with a large angle-of-arrival spread at the MS. This will cause significant differences in the spatial correlation properties of the uplink and downlink. It will be shown that a much larger spatial distance is required to obtain a given degree of spatial decorrelation at the BS as compared to the MS.

To obtain the space-time correlation functions, it is once again necessary to specify the scattering geometry. One possibility for NLoS conditions is the single-ring model shown in Fig. 2.20, where the MS and BS are located at O_M and O_B , respectively, and separated by distance D , and the scatterers are assumed to be located on a ring of radius R centered around the MS. Once again, a local scattering assumption is often made such that $D \gg R$.

Consider two BS antennas, $A_B^{(q)}$, $q = 1, 2$, separated by distance δ_B . The BS antenna array is oriented with angle θ_B with respect to the x -axis, and the MS moves with velocity v and angle γ_M with respect to the x -axis. For the environment shown in Fig. 2.20, the channel from O_M to $A_B^{(q)}$ has the complex envelope

$$g_q(t) = \sum_{m=1}^N C_m e^{j\phi_m - j2\pi(R + \epsilon_{mq})/\lambda_c} e^{j2\pi f_m t \cos(\alpha_M^{(m)} - \gamma_M)}, \quad q = 1, 2, \quad (2.130)$$

where ϵ_{mq} denotes the distance $S_M^{(m)} - A_B^{(q)}$, $q = 1, 2$. To proceed further, ϵ_{mq} needs to be expressed as a function of $\alpha_M^{(m)}$.

Applying the cosine law to the triangle $\triangle S_M^{(m)} O_B A_B^{(q)}$, the distance ϵ_{mq} can be expressed as a function of the angle $\theta_B^{(m)} - \theta_B$ as follows:

$$\epsilon_{mq}^2 = ((1.5 - q)\delta_B)^2 + \epsilon_m^2 - 2(1.5 - q)\delta_B\epsilon_m \cos(\theta_B^{(m)} - \theta_B), \quad q = 1, 2, \quad (2.131)$$

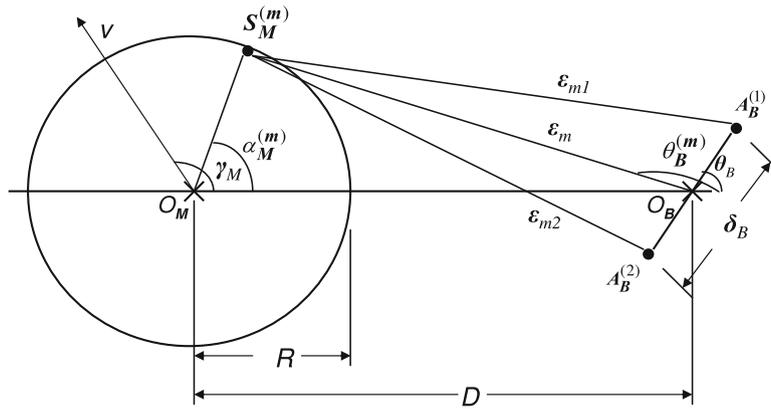


Fig. 2.20 Single-ring scattering model for NLoS propagation on the reverse link of a cellular system. The MS is surrounded by a scattering ring of radius R and is at distance D from the BS, where $R \ll D$

where ϵ_m is the distance $S_M^{(m)} - O_B$. By applying the sine law to the triangle $\triangle O_M S_M^{(m)} O_B$, the following identity is obtained

$$\frac{\epsilon_m}{\sin(\alpha_M^{(m)})} = \frac{R}{\sin(\pi - \theta_B^{(m)})} = \frac{D}{\sin(\pi - \alpha_M^{(m)} - (\pi - \theta_B^{(m)}))}. \quad (2.132)$$

Since the angle $\pi - \theta_B^{(m)}$ is small, the small angle approximations $\sin x \approx x$ and $\cos x \approx 1$ for small x can be applied to the second equality in the above identity. This gives

$$\frac{R}{(\pi - \theta_B^{(m)})} \approx \frac{D}{\sin(\pi - \alpha_M^{(m)})} \quad (2.133)$$

or

$$(\pi - \theta_B^{(m)}) \approx (R/D) \sin(\pi - \alpha_M^{(m)}). \quad (2.134)$$

It follows that the cosine term in (2.131) becomes

$$\begin{aligned} \cos(\theta_B^{(m)} - \theta_B) &= \cos(\pi - \theta_B - (\pi - \theta_B^{(m)})) \\ &= \cos(\pi - \theta_B) \cos(\pi - \theta_B^{(m)}) + \sin(\pi - \theta_B) \sin(\pi - \theta_B^{(m)}) \\ &\approx \cos(\pi - \theta_B) + \sin(\pi - \theta_B)(R/D) \sin(\pi - \alpha_M^{(m)}) \\ &= -\cos(\theta_B) + (R/D) \sin(\theta_B) \sin(\alpha_M^{(m)}). \end{aligned} \quad (2.135)$$

Substituting the approximation in (2.135) into (2.131), along with $\delta_B/\epsilon_m \ll 1$, gives

$$\epsilon_{mq}^2 \approx \epsilon_m^2 \left(1 - 2(1.5 - q) \frac{\delta_B}{\epsilon_m} \left((R/D) \sin(\theta_B) \sin(\alpha_M^{(m)}) - \cos(\theta_B) \right) \right). \quad (2.136)$$

Applying the approximation $\sqrt{1 \pm x} \approx 1 \pm x/2$ for small x gives

$$\epsilon_{mq} \approx \epsilon_m - (1.5 - q) \delta_B \left((R/D) \sin(\theta_B) \sin(\alpha_M^{(m)}) - \cos(\theta_B) \right). \quad (2.137)$$

Applying the cosine law to the triangle $\triangle O_M S_M^{(m)} O_B$ gives

$$\begin{aligned} \epsilon_m^2 &= D^2 + R^2 - 2DR \cos(\alpha_M^{(m)}) \\ &\approx D^2 \left(1 - 2(R/D) \cos(\alpha_M^{(m)}) \right), \end{aligned} \quad (2.138)$$

and again using the approximation $\sqrt{1 \pm x} \approx 1 \pm x/2$ for small x gives

$$\epsilon_m \approx D - R \cos(\alpha_M^{(m)}). \quad (2.139)$$

Finally, using (2.139) in (2.131) gives

$$\epsilon_{mq} \approx D - R \cos(\alpha_M^{(m)}) - (1.5 - q) \delta_B \left((R/D) \sin(\theta_B) \sin(\alpha_M^{(m)}) - \cos(\theta_B) \right). \quad (2.140)$$

Substituting (2.140) into (2.130) gives the result

$$g_q(t) = \sum_{m=1}^N C_m e^{j\phi_m + j2\pi f_m t \cos(\alpha_M^{(m)} - \gamma_M) - j2\pi \left(R + D - R \cos(\alpha_M^{(m)}) - (1.5 - q) \delta_B \left((R/D) \sin(\theta_B) \sin(\alpha_M^{(m)}) - \cos(\theta_B) \right) \right) / \lambda_c}, \quad (2.141)$$

which is a function of the azimuth angle of departure $\alpha_M^{(m)}$.

The space-time correlation function between the two complex faded envelopes $g_1(t)$ and $g_2(t)$ at the BS is once again given by (2.126). Using (2.141) and (2.126), the space-time correlation function between $g_1(t)$ and $g_2(t)$ can be written as

$$\phi_{g_1, g_2}(\delta_B, \tau) = \frac{\Omega_p}{2N} \sum_{m=1}^N \mathbb{E} \left[e^{-j2\pi(\delta_B/\lambda_c)((R/D) \sin(\theta_B) \sin(\alpha_M^{(m)}) - \cos(\theta_B)) + j2\pi f_m \tau \cos(\alpha_M^{(m)} - \gamma_M)} \right]. \quad (2.142)$$

Since the number of scatters around the MS is infinite, the discrete angles-of-arrival $\alpha_M^{(m)}$ can be replaced with a continuous random variable α_M with probability density function $p(\alpha_M)$. Hence, the space-time correlation function becomes

$$\phi_{g_1, g_2}(\delta_B, \tau) = \frac{\Omega_p}{2} \int_{-\pi}^{\pi} e^{ja \cos(\alpha_M - \gamma_M)} e^{-jb((R/D) \sin(\theta_B) \sin(\alpha_M) - \cos(\theta_B))} p(\alpha_M) d\alpha_M, \quad (2.143)$$

where $a = 2\pi f_m \tau$ and $b = 2\pi \delta_B / \lambda_c$.

For the case of 2-D isotropic scattering with an isotropic MS transmit antenna, $p(\alpha_M) = 1/(2\pi)$, $-\pi \leq \alpha_M \leq \pi$, and the space-time correlation function becomes

$$\phi_{g_1, g_2}(\delta_B, \tau) = \frac{\Omega_p}{2} e^{jb \cos(\theta_B)} J_0 \left(\sqrt{a^2 + b^2 (R/D)^2 \sin^2(\theta_B) - 2ab(R/D) \sin(\theta_B) \sin(\gamma_M)} \right). \quad (2.144)$$

The spatial and temporal correlation functions can be obtained by setting $\tau = 0$ and $\delta_B = 0$, respectively. For the temporal correlation function $\phi_{gg}(\tau) = \phi_{g_1, g_2}(0, \tau) = \frac{\Omega_p}{2} J_0(2\pi f_m \tau)$ which matches our earlier result in (2.29) as expected. The spatial correlation function is

$$\phi_{g_1, g_2}(\delta_B) = \phi_{g_1, g_2}(\delta_B, 0) = \frac{\Omega_p}{2} e^{jb \cos(\theta_B)} J_0(b(R/D) \sin(\theta_B)). \quad (2.145)$$

Figure 2.21 plots the magnitude of the normalized spatial envelope crosscovariance function, $|\phi_{g_1, g_2}(\delta_B)|/(\Omega_p/2)$, for $R = 60$ m and various BS array orientation angles θ_B . Likewise, Fig. 2.22 plots $|\phi_{g_1, g_2}(\delta_B)|/(\Omega_p/2)$ for $\theta_B = \pi/3$ and various scattering radii, R . In general, a much greater spatial separation is required to achieve a given degree of envelope decorrelation at the BS as compared to the MS. This can be readily seen by the term $R/D \ll 1$ in the argument of the Bessel function in (2.145). Also, the spatial correlation increases as the angle θ_B and scattering radii R decrease. BS antenna arrays that are broadside with the MS direction will experience the lowest correlation, while those that are in-line with the MS direction will experience the highest correlation. In fact, for in-line BS antennas $|\phi_{g_1, g_2}(\delta_B)| = 1$ and, therefore, the faded envelopes on the different BS antennas are perfectly correlated.

2.2 Mobile-to-Mobile Channels

Mobile-to-mobile (M-to-M) communication channels arise when both the transmitter and receiver are in motion and are equipped with low elevation antennas that are surrounded by local scatterers. The statistical properties of M-to-M channels differ significantly from those of conventional fixed-to-mobile (F-to-M) cellular land mobile radio channels, where the mobile station is surrounded by local scatterers and the base station is stationary, elevated, and relatively free of local scattering. Akki and Haber [13, 14] were the first to propose a mathematical reference model for M-to-M flat fading channels under NLoS propagation conditions. The Akki and Haber model was extended by Vatalaro and Forcella [329] to account for 3-D scattering, and by Linnartz and Fiesta [209] to include LoS propagation conditions. Channel measurements for outdoor-to-outdoor, narrow-band outdoor-to-indoor, and wide-band mobile-to-mobile communications have been reported in [9, 182, 188, 222], and methods for simulating M-to-M channels have been proposed in [259, 345, 371].

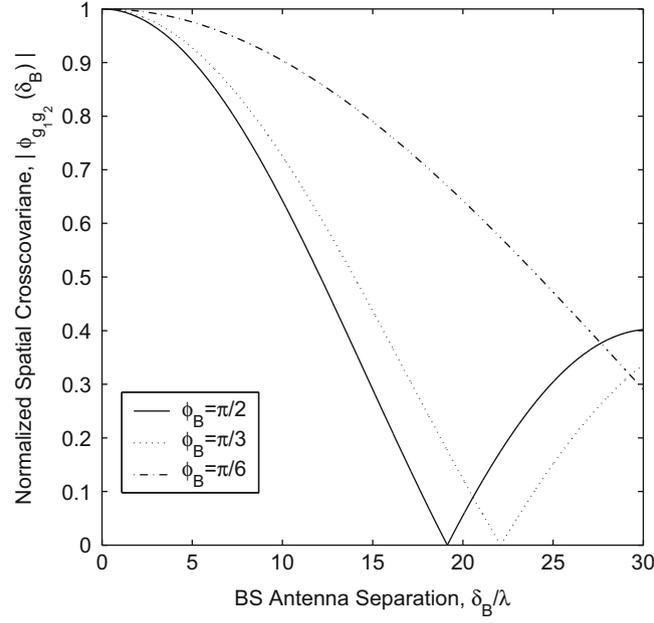


Fig. 2.21 Envelope cross-correlation magnitude at the base station for $R = 60$ m and various arrival angles, θ_B ; $D = 3000$ m

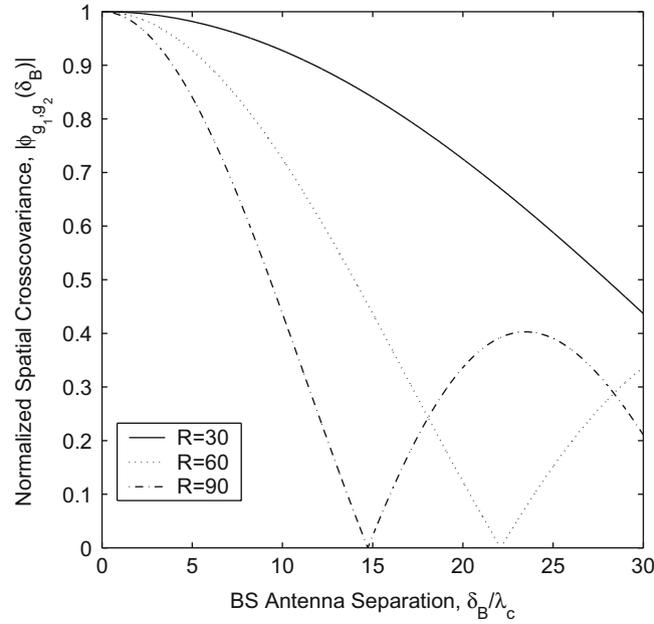


Fig. 2.22 Envelope cross-correlation magnitude at the base station for $\theta_B = \pi/3$ and various scattering radii, R ; $D = 3000$ m

2.2.1 Mobile-to-Mobile Reference Model

Akki and Haber's mathematical reference model for M-to-M flat fading channels gives the complex faded envelope as [14]

$$g(t) = \sqrt{\frac{1}{N}} \sum_{n=1}^N e^{j2\pi(f_m^T \cos(\alpha_T^{(n)}) + f_m^R \cos(\alpha_R^{(n)}))t + j\phi_n}, \quad (2.146)$$

where N is the number of propagation paths; f_m^T and f_m^R are the maximum Doppler frequencies due to the motion of the transmitter and receiver, respectively; $\alpha_T^{(n)}$ is the random angle of departure and $\alpha_R^{(n)}$ is the random angle of arrival of the n th propagation path measured with respect to the transmitter and receiver velocity vectors, respectively; ϕ_n is a random phase uniformly distributed on $[-\pi, \pi)$ independent of $\alpha_T^{(n)}$ and $\alpha_R^{(n)}$ for all n .

In the above model, the Doppler frequency experienced by each propagation path is the sum of the individual Doppler frequencies induced by the motion of the transmitter and receiver. Once again, for sufficiently large N , the central limit theorem can be invoked with the result that $g_I(t)$ and $g_Q(t)$ can be treated as zero-mean Gaussian random processes. If isotropic transmitter and receiver antennas and 2-D isotropic scattering around both the transmitter and receiver are assumed, then the envelope $|g(t)|$ is Rayleigh distributed at any time t . This is similar to the case of conventional F-to-M cellular land mobile radio channels with 2-D isotropic scattering and an isotropic antenna at the MS. However, the ensemble averaged temporal correlation function of the faded envelope of M-to-M channels is quite different from that of F-to-M channels, and can be derived as follows [13, 14]:

$$\begin{aligned}\phi_{g_I g_I}(\tau) &= \frac{1}{2} J_0(2\pi f_m^T \tau) J_0(2\pi a f_m^T \tau) \\ \phi_{g_Q g_Q}(\tau) &= \frac{1}{2} J_0(2\pi f_m^T \tau) J_0(2\pi a f_m^T \tau) \\ \phi_{g_I g_Q}(\tau) &= \phi_{g_Q g_I}(\tau) = 0 \\ \phi_{g g}(\tau) &= \frac{1}{2} J_0(2\pi f_m^T \tau) J_0(2\pi a f_m^T \tau),\end{aligned}\tag{2.147}$$

where $a = f_m^R/f_m^T$ is the ratio of the two maximum Doppler frequencies (or speeds) of the receiver and transmitter, and $0 \leq a \leq 1$ assuming $f_m^R \leq f_m^T$. Observe that the temporal correlation functions of M-to-M channels involve a product of two Bessel functions in contrast to the single Bessel function found in F-to-M channels. Also, $a = 0$ yields the temporal correlation functions for F-to-M channels as expected. The corresponding Doppler spectrum, obtained by taking the Fourier transform of (2.147) is

$$S_{g g}(f) = \frac{1}{\pi^2 f_m^T \sqrt{a}} K \left(\frac{1+a}{2\sqrt{a}} \sqrt{1 - \left(\frac{f}{(1+a)f_m^T} \right)^2} \right),\tag{2.148}$$

where $K(\cdot)$ is the complete elliptic integral of the first kind. The Doppler spectrum of M-to-M channels in (2.148) differs from the classical spectrum of F-to-M cellular land mobile radio channels, as illustrated in Fig. 2.23, which shows the Doppler spectrum for different values of a .

2.3 Statistical Characterization of Multipath-Fading Channels

Multipath-fading channels can be modeled as randomly time-variant linear filters, whose inputs and outputs can be described in both the time and frequency domains. This leads to four possible transmission functions [35]; the time-variant impulse response $g(\tau, t)$, the output Doppler-spread function $H(f, \nu)$, the time-variant transfer function $T(f, t)$, and the delay Doppler-spread function $S(\tau, \nu)$. The time-variant impulse response relates the complex low-pass input and output time waveforms, $\tilde{s}(t)$ and $\tilde{r}(t)$, respectively, through the convolution integral

$$\tilde{r}(t) = \int_0^t g(t, \tau) \tilde{s}(t - \tau) d\tau.\tag{2.149}$$

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 is causal, $g(t, \tau) = 0$ for $\tau < 0$ and, therefore, the lower limit of integration in (2.149) is zero. If the convolution in (2.149) is approximated as a discrete sum, then

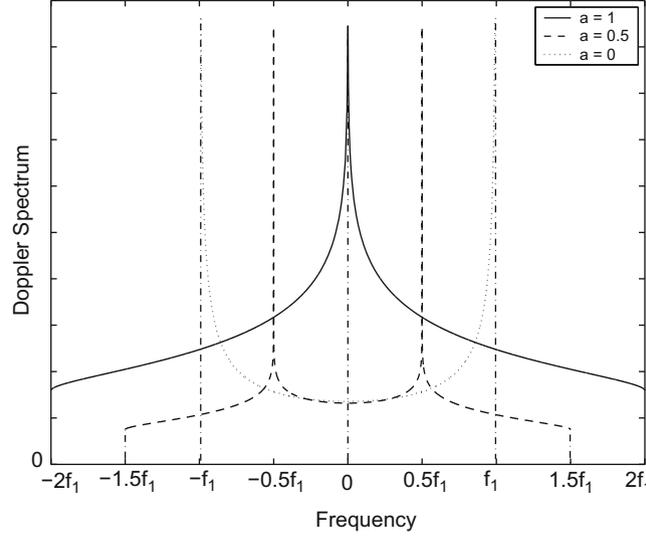


Fig. 2.23 Doppler spectrum for M-to-M and F-to-M channels

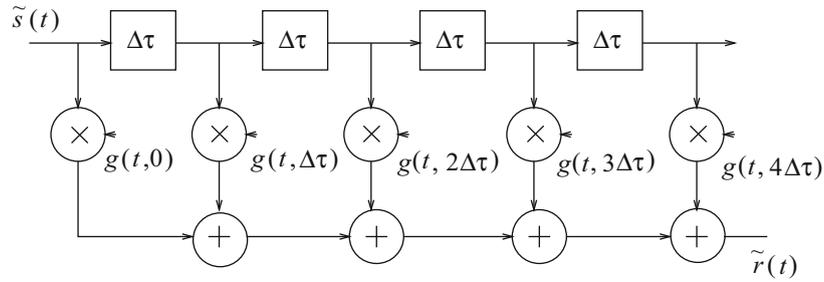


Fig. 2.24 Discrete-time tapped delay line model for a multipath-fading channel

$$\tilde{r}(t) = \sum_{m=0}^n g(t, m\Delta\tau) \tilde{s}(t - m\Delta\tau) \Delta\tau. \quad (2.150)$$

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. 2.24.

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

$$\tilde{R}(f) = \int_{-\infty}^{\infty} H(f - \nu, \nu) \tilde{S}(f - \nu) d\nu. \quad (2.151)$$

Bello called the function $H(f, \nu)$ the output Doppler-spread function [35]. 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 (2.151) can be approximated by the discrete sum

$$\tilde{R}(f) = \sum_{m=1}^n H(f - m\Delta\nu, m\Delta\nu) \tilde{S}(f - m\Delta\nu) \Delta\nu. \quad (2.152)$$

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 dense frequency conversion chain with tap spacing $\Delta\nu$ that produces the Doppler shifts as shown in Fig. 2.25.

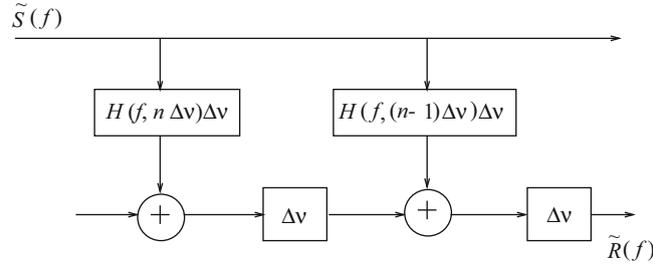
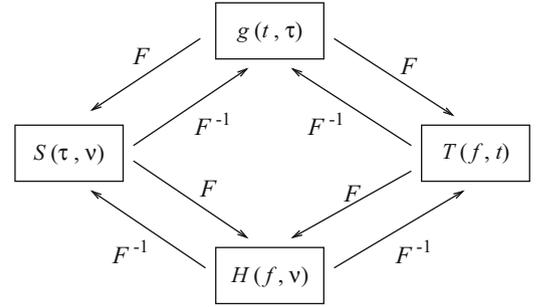


Fig. 2.25 Frequency conversion model for a multipath-fading channel

Fig. 2.26 Fourier transform relations between the transmission functions



The third transmission function $T(f, t)$ is the time-variant transfer function, which relates the output time waveform to the input spectrum as follows:

$$\tilde{r}(t) = \int_{-\infty}^{\infty} \tilde{S}(f)T(f, t)e^{j2\pi ft}df. \quad (2.153)$$

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

$$\tilde{r}(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(\tau, \nu)\tilde{s}(t - \tau)e^{j2\pi f\tau}d\nu d\tau. \quad (2.154)$$

The function $S(\tau, \nu)$ is called the delay Doppler-spread function [35], 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. 2.26. In each Fourier transform pair there is always a fixed variable, so that the transform involves the other two variables.

2.3.1 Statistical Channel Correlation Functions

Similar to flat fading channels, the channel impulse response $g(t, \tau) = g_I(t, \tau) + jg_Q(t, \tau)$ of a frequency-selective fading channel can be modeled as a complex Gaussian random process, where the quadrature components $g_I(t, \tau)$ and $g_Q(t, \tau)$ are real Gaussian random processes. Hence, all of the transmission functions defined in the last section are themselves random processes. Since the underlying process is Gaussian, a complete statistical description of these transmission functions is provided by their means and autocorrelation functions. In the following discussion, zero-mean Gaussian random processes are assumed for simplicity so that only the autocorrelation functions are of interest. Since there are four transmission functions, four autocorrelation functions can be defined as follows [255, 272]:

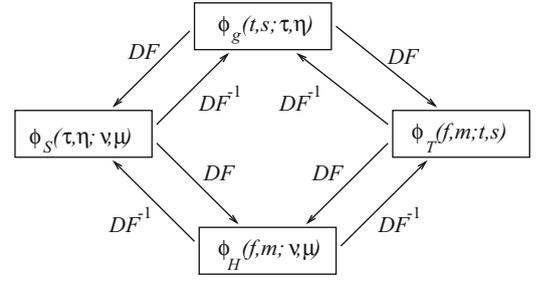
$$\phi_g(t, s; \tau, \eta) = E[g^*(t, \tau)g(s, \eta)] \quad (2.155)$$

$$\phi_T(f, m; t, s) = E[T^*(f, t)T(m, s)] \quad (2.156)$$

$$\phi_H(f, m; \nu, \mu) = E[H^*(f, \nu)H(m, \mu)] \quad (2.157)$$

$$\phi_S(\tau, \eta; \nu, \mu) = E[S^*(\tau, \nu)S(\eta, \mu)]. \quad (2.158)$$

Fig. 2.27 Double Fourier transform relations between the channel autocorrelation functions



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

$$\phi_S(\tau, \eta; \nu, \mu) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_g(t, s; \tau, \eta) e^{-j2\pi(\nu t - \mu s)} dt ds \quad (2.159)$$

$$\phi_g(t, s; \tau, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_S(\tau, \eta; \nu, \mu) e^{j2\pi(\nu t - \mu s)} d\nu d\mu. \quad (2.160)$$

The complete set of such relationships is summarized in Fig. 2.27.

2.3.2 Classification of Channels

Wide-sense stationary (WSS) channels have fading statistics that remain constant over small periods of time or short spatial distances. 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 2.23) that WSS channels give rise to scattering with uncorrelated Doppler shifts. This behavior suggests that the attenuations and phase shifts associated with multipath components having different Doppler shifts are uncorrelated. This makes sense, because multipath components with different Doppler shifts arrive from different directions and, hence, propagate via different sets of scatterers. For WSS channels the correlation functions become

$$\phi_g(t, t + \Delta_t; \tau, \eta) = \phi_g(\Delta_t; \tau, \eta) \quad (2.161)$$

$$\phi_T(f, m; t, t + \Delta_t) = \phi_T(f, m; \Delta_t) \quad (2.162)$$

$$\phi_H(f, m; \nu, \mu) = \psi_H(f, m; \nu) \delta(\nu - \mu) \quad (2.163)$$

$$\phi_S(\tau, \eta; \nu, \mu) = \psi_S(\tau, \eta; \nu) \delta(\nu - \mu), \quad (2.164)$$

where

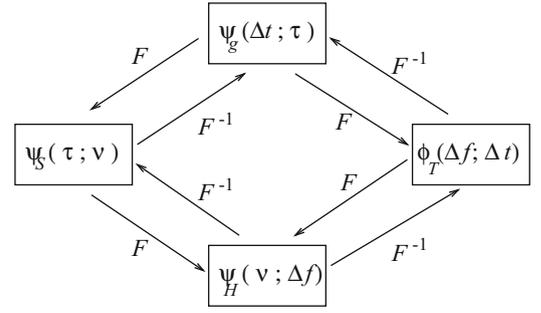
$$\psi_H(f, m; \nu) = \int_{-\infty}^{\infty} \phi_T(f, m; \Delta_t) e^{-j2\pi\nu\Delta_t} d\Delta_t \quad (2.165)$$

$$\psi_S(\tau, \eta; \nu) = \int_{-\infty}^{\infty} \phi_g(\Delta_t; \tau, \eta) e^{-j2\pi\nu\Delta_t} d\Delta_t \quad (2.166)$$

are Fourier transform pairs.

Uncorrelated scattering (US) channels are characterized by paths of different delays having uncorrelated complex gains. 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 $\Delta f = m - f$ [35]. Analogous to (2.163) and (2.164), the channel correlation functions are singular in the time-delay variable (see Problem 2.24). Again, this makes sense because multipath components arriving with different delays propagate via different sets of scatterers. For US channels, the channel correlation functions become

Fig. 2.28 Fourier transform relations between the channel correlation functions for WSSUS channels



$$\phi_g(t, s; \tau, \eta) = \psi_g(t, s; \tau)\delta(\eta - \tau) \quad (2.167)$$

$$\phi_T(f, f + \Delta_f; t, s) = \phi_T(\Delta_f; t, s) \quad (2.168)$$

$$\phi_H(f, f + \Delta_f; \nu, \mu) = \phi_H(\Delta_f; \nu, \mu) \quad (2.169)$$

$$\phi_S(\tau, \eta; \nu, \mu) = \psi_S(\tau; \nu, \mu)\delta(\eta - \tau), \quad (2.170)$$

where

$$\psi_g(t, s; \tau) = \int_{-\infty}^{\infty} \phi_T(\Delta_f; t, s) e^{j2\pi\Delta_f\tau} d\Delta_f \quad (2.171)$$

$$\psi_S(\tau; \nu, \mu) = \int_{-\infty}^{\infty} \phi_H(\Delta_f; \nu, \mu) e^{j2\pi\Delta_f\tau} d\Delta_f \quad (2.172)$$

are Fourier transform pairs.

Wide-sense stationary uncorrelated scattering (WSSUS) channels exhibit uncorrelated scattering in both the time-delay and Doppler shift domains. Fortunately, many radio channels can be effectively 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:

$$\phi_g(t, t + \Delta_t; \tau, \eta) = \psi_g(\Delta_t; \tau)\delta(\eta - \tau) \quad (2.173)$$

$$\phi_T(f, f + \Delta_f; t, t + \Delta_t) = \phi_T(\Delta_f; \Delta_t) \quad (2.174)$$

$$\phi_H(f, f + \Delta_f; \nu, \mu) = \psi_H(\Delta_f; \nu)\delta(\nu - \mu) \quad (2.175)$$

$$\phi_S(\tau, \eta; \nu, \mu) = \psi_S(\tau, \nu)\delta(\eta - \tau)\delta(\nu - \mu). \quad (2.176)$$

The correlation functions for WSSUS channels are related through the Fourier transform pairs shown in Fig. 2.28.

2.3.3 Power-Delay Profile and Coherence Bandwidth

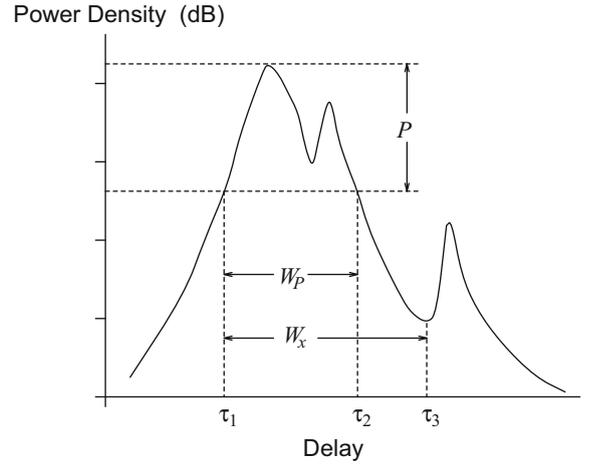
The function $\psi_g(0; \tau) \equiv \psi_g(\tau)$ is called the power-delay profile and gives the average power at the channel output as a function of the time delay τ . A typical power-delay profile is shown in Fig. 2.29. One quantity of interest is the average delay, defined as

$$\mu_\tau = \frac{\int_0^\infty \tau \psi_g(\tau) d\tau}{\int_0^\infty \psi_g(\tau) d\tau}. \quad (2.177)$$

Note that the normalization $\int_0^\infty \psi_g(\tau) d\tau$ is applied because $\psi_g(\tau)$ does not necessarily integrate to unity like a probability density function. Another quantity of interest is the rms delay spread, defined as

$$\sigma_\tau = \sqrt{\frac{\int_0^\infty (\tau - \mu_\tau)^2 \psi_g(\tau) d\tau}{\int_0^\infty \psi_g(\tau) d\tau}}. \quad (2.178)$$

Fig. 2.29 A typical power-delay profile



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 a fraction x of the total power in the profile. Referring to Fig. 2.29

$$W_x = \tau_3 - \tau_1 \quad (2.179)$$

where τ_1 and τ_3 are chosen so that

$$\int_0^{\tau_1} \psi_g(\tau) d\tau = \int_{\tau_3}^{\infty} \psi_g(\tau) d\tau \quad (2.180)$$

and

$$\int_{\tau_1}^{\tau_3} \psi_g(\tau) d\tau = x \int_0^{\infty} \psi_g(\tau) d\tau. \quad (2.181)$$

Another quantity is the difference in delays where the power-delay profile rises to a value P dB below its maximum value for the first time and where the power-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. 2.29, where $W_P = \tau_2 - \tau_1$. 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 1–10 μ s, the delay spreads in a typical microcellular applications are much less. Delay spreads for indoor channels can range 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 $\phi_T(\Delta_f; \Delta_f)$ is called the spaced-time spaced-frequency 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 Δ_f for which $\phi_T(\Delta_f) = x\phi_T(0)$ for some suitably small value of x , $0 < x < 1$. 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.,

$$B_c \propto \frac{1}{\mu_\tau} \quad \text{or} \quad B_c \propto \frac{1}{\sigma_\tau}. \quad (2.182)$$

Wideband fading channels can be classified according to the relationship between the transmitted signal bandwidth W_s and the coherence bandwidth B_c . If $B_c < W_s$, the channel is said to exhibit frequency-selective fading which introduces intersymbol interference (ISI) into the received signal. If $B_c \gg W_s$, the channel is said to exhibit flat fading, and very little ISI is introduced into the received signal.

The function $\psi_H(\nu; 0) \equiv \psi_H(\nu)$ is identical to the Doppler spectrum $S_{gg}(f)$ in (2.30), i.e., $\psi_H(\nu) \equiv S_{gg}(f)$, and gives the average power at the channel output as a function of the Doppler frequency ν . The range of values over which $\psi_H(\nu)$ is significant is called the Doppler spread and is denoted by B_d . Since $\psi_H(\nu)$ and $\phi_T(\Delta_f)$ are a Fourier transform pair, it follows that the inverse of the Doppler spread gives a measure of the channel coherence time, T_c , i.e.,

$$T_c \approx \frac{1}{B_d}. \quad (2.183)$$

Wideband fading channels can also be classified according to the relationship between the transmitted symbol duration, T , and the coherence time T_c . If $T_c < T$, the channel is said to exhibit fast fading which introduces severe frequency dispersion into the received signal. If $T_c \gg T$, the channel is said to exhibit slow fading which introduces very little frequency dispersion into the received signal.

Finally, 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 . Plots of the scattering function are often used to provide a concise statistical description of a multipath-fading channel from measurement data.

2.3.4 System Correlation Functions

The astute reader may wonder why the factor of 1/2 does not appear in the definition of the channel correlation functions in (2.158). The multipath-fading channel can be treated as a time-variant linear filter, such that the complex envelopes of the channel input and output are related through the convolution integral in (2.149). By using (2.149), the autocorrelation functions of the output waveform $\tilde{r}(t)$ and of the input waveform $\tilde{s}(t)$ can be related. Assuming that the channel is WSSUS, the time autocorrelation function of the channel output $\tilde{r}(t)$ is

$$\begin{aligned}
\phi_{\tilde{r}\tilde{r}}(t, t + \Delta_t) &= \frac{1}{2} \text{E} [\tilde{r}^*(t) \tilde{r}(t + \Delta_t)] \\
&= \frac{1}{2} \text{E} \left[\int_0^t g^*(t, \alpha) \tilde{s}^*(t - \alpha) d\alpha \times \int_0^{t+\Delta_t} g(t + \Delta_t, \beta) \tilde{s}(t + \Delta_t - \beta) d\beta \right] \\
&= \int_0^t \int_0^{t+\Delta_t} \text{E} [g^*(t, \alpha) g(t + \Delta_t, \beta)] \frac{1}{2} \text{E} [\tilde{s}^*(t - \alpha) \tilde{s}(t + \Delta_t - \beta)] d\alpha d\beta \\
&= \int_0^t \int_0^{t+\Delta_t} \psi_g(\Delta_t; \alpha) \delta(\beta - \alpha) \frac{1}{2} \text{E} [\tilde{s}^*(t - \alpha) \tilde{s}(t + \Delta_t - \beta)] d\alpha d\beta \\
&= \int_0^t \psi_g(\Delta_t; \alpha) \frac{1}{2} \text{E} [\tilde{s}^*(t - \alpha) \tilde{s}(t + \Delta_t - \alpha)] d\alpha \\
&= \int_0^t \psi_g(\Delta_t; \alpha) \phi_{\tilde{s}\tilde{s}}(t - \alpha, t - \alpha + \Delta_t) d\alpha \\
&= \psi_g(\Delta_t; t) * \phi_{\tilde{s}\tilde{s}}(t, t + \Delta_t)
\end{aligned} \tag{2.184}$$

where

$$\phi_{\tilde{s}\tilde{s}}(t, t + \Delta_t) = \frac{1}{2} \text{E} [\tilde{s}^*(t) \tilde{s}(t + \Delta_t)].$$

Thus, the output autocorrelation function is the convolution of the channel correlation function $\psi_g(\Delta_t; t)$ and the correlation function of the input waveform. Note that the factor of 1/2 in the definition of $\phi_{\tilde{r}\tilde{r}}(t, t + \Delta_t)$ is absorbed into the similar definition of $\phi_{\tilde{s}\tilde{s}}(t, t + \Delta_t)$. Consequently, a factor of 1/2 does not appear in the definition of the channel (or system) correlation functions in (2.158). For digital modulation, the input waveform $\tilde{s}(t)$ is a cyclostationary random process as described in Sect. 4.9.1 and, hence, the channel output waveform will be cyclostationary as well.

2.4 Polarized Channel Modeling

Multi-polarized antennas have attracted considerable attention as a means for increasing channel capacity and reducing the required antenna spacing through the principle of polarization diversity [19, 184, 191, 198, 245, 299, 330]. Such antennas exploit channel depolarization and have the potential to double or even triple the channel capacity that is achievable with single-polarized antennas [19, 110, 111, 245, 299, 306]. In practice, cellular land mobile radio systems have employed slant polarized array antennas at the base stations to good benefit.

Depolarization can occur as a result of two factors. One factor arises from the transmit and receive antennas themselves. Any practical antenna gain pattern can be described in terms of two orthogonal antenna polarization functions. For a vertically polarized antenna these polarization functions are f_{VV}^a and f_{HV}^a , while for a horizontally polarized antenna they are f_{HH}^a and f_{VH}^a , where H and V refer to horizontal and vertical polarization, respectively. For an ideal linear vertically or horizontally polarized antenna only f_{VV}^a or f_{HH}^a is non-zero, respectively. However, for practical antennas, the cross-polarization functions f_{VH}^a and f_{HV}^a will be non-zero as well due to antenna depolarization.

The other factor causing depolarization is the radio channel propagation environment. Similar to antenna depolarization, the radio channel can be characterized in terms of four orthogonal channel polarization functions identified as f_{VV} , f_{HV} , f_{VH} , and f_{HH} , where again H and V refer to horizontal and vertical polarization, respectively. The antenna and channel depolarization can be decoupled, and in order to isolate the effects of channel depolarization, ideal linear polarized antennas are assumed.

The degree of depolarization is often measured in terms of the cross polarization discrimination (XPD), defined as the ratio of the co-polarized average received power to the cross-polarized average received power. For a vertically polarized transmit antenna the XPD can be defined as

$$\text{XPD} = \frac{E[f_{VV}^2]}{E[f_{HV}^2]}. \quad (2.185)$$

In conventional XPD modeling, it is assumed that $E[f_{VV}^2] = E[f_{HH}^2] = 1 - \zeta$, and $E[f_{HV}^2] = E[f_{VH}^2] = \zeta$, resulting in

$$\text{XPD} = \frac{1 - \zeta}{\zeta}, \quad (2.186)$$

where ζ is a parameter that having the range $0 \leq \zeta \leq 1$. Such a model ensures a conservation of power or energy, where the total received power is split between the co-polarized and cross-polarized components. When $\zeta = 0$, only the co-polarized component exists and $\text{XPD} = \infty$, and when $\zeta = 1$, only the cross-polarized component exists and $\text{XPD} = 0$. The polarized channel is conventionally treated as a 2×2 MIMO channel, such that

$$\mathbf{G} = \mathbf{Z} \otimes \mathbf{G}_{\text{iid}} \quad (2.187)$$

where \mathbf{G}_{iid} is a 2×2 matrix of zero-mean complex i.i.d. Gaussian random variables,

$$\mathbf{Z} = \begin{bmatrix} \sqrt{1-\zeta} & \sqrt{\zeta} \\ \sqrt{\zeta} & \sqrt{1-\zeta} \end{bmatrix}, \quad (2.188)$$

and \otimes denotes the Hamadard product or Schur product (element-by-element matrix multiplication).

Despite its simplicity, the conventional model for XPD provides little insight into the mechanisms of channel depolarization and how the parameter ζ behaves. Typically, a value or ranges of values for ζ are merely assumed. The remainder of this section considers a geometrical theoretical model for channel depolarization in wireless fixed-to-mobile channels, characterized by isotropic or non-isotropic azimuthal scattering and with either non-line-of-sight (NLoS) or line-of-sight (LoS) conditions between the transmitter and receiver. Predictions from the model are compared with previously reported measurement data [110, 191, 306] and shown to have close agreement. Finally, the theoretical geometry-based XPD model is used to explain the origin of some phenomena observed in these empirical results.

2.4.1 Geometric Model for Channel Depolarization

Consider the 3-D cylindrical model shown in Fig. 2.30, where scatterers are distributed on a cylindrical surface in the vicinity of the mobile station.² These scatterers result in single-bounced propagation between a BS and a MS. The high-elevation BS antenna is assumed to be free of local scattering, which is typical of cellular land mobile radio systems. A LoS component may or may not exist between the BS and MS. The cylindrical scattering surface is characterized by its radius R_r , and the

²Here the term ‘‘scatterer’’ refers to a plane boundary that is typically much larger than a wavelength.

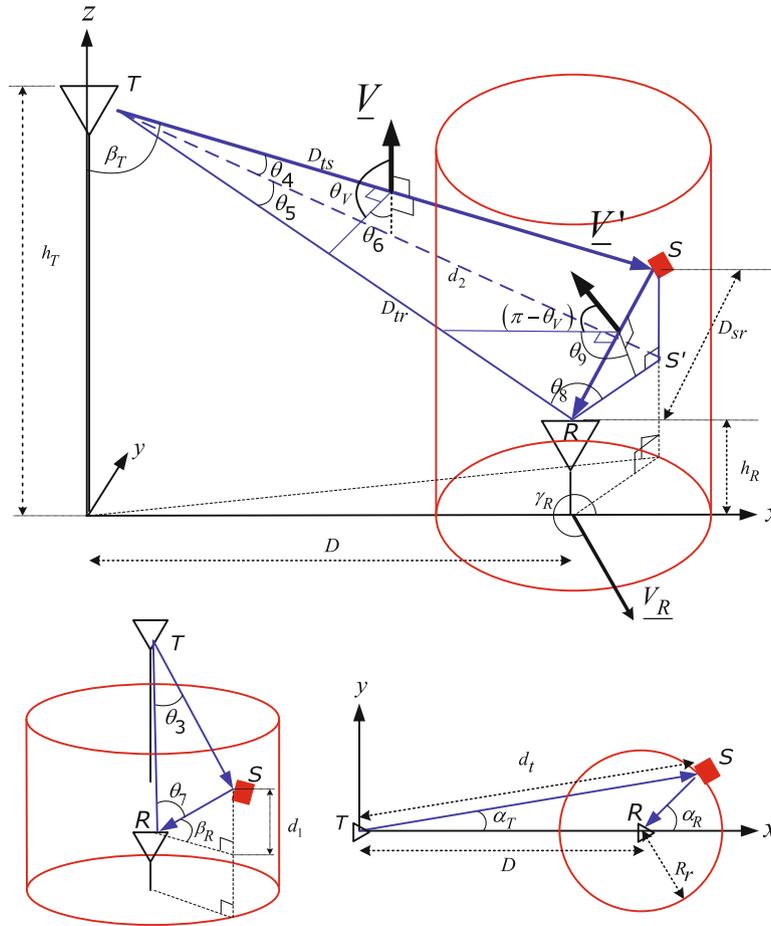


Fig. 2.30 3-D cylindrical scattering model for fixed-to-mobile channels

ray incident on the MS antenna from each scatterer on the cylinder has an associated azimuth angle of arrival (AAoA), α_R , and elevation angle of arrival (EAoA), β_R . As will be explained later, a 3-D scattering model is necessary to explain the mechanism of channel depolarization.

The phenomenon of channel depolarization between a MS and MS can be explained by the NLoS plane waves that arrive at the MS via scattering objects. Figure 2.30 presents a scenario where a vertically polarized plane wave emitted from the BS can result in a horizontally polarized plane wave component being received at the MS. The vector \underline{V} in Fig. 2.30 is the vertical polarization vector of a plane wave from the BS, and \underline{V}' is the corresponding polarization vector of the plane wave after reflection from a scatterer. By convention the vertical and horizontal polarization vectors are parallel to \underline{e}_θ and \underline{e}_ϕ , respectively, when describing the plane wave's direction of propagation at the BS or MS in a spherical coordinate system. It is clear that \underline{V}' is not entirely vertically polarized at the MS, but has a cross-polarization component as well.

Channel depolarization is affected by the scattering geometry as illustrated in Fig. 2.30. For the plane wave that is emitted from the BS and received at the MS via a scatterer, the plane that is defined by the BS, MS, and the scatterer must be taken into account. The left side of Fig. 2.31 depicts one such plane (corresponding to the page in this book). The circled dot and arrow represent the vectors of the perpendicular and parallel polarization components associated with the plane, whose directions are from the backside to the front side of the figure and the arrow's direction, respectively [59]. If each scatterer is modeled as a planar boundary, then the directions of the vectors of the perpendicular and parallel polarization components as viewed from the BS and MS will be the same within the plane. Since any polarization vector can be expressed as a linear combination of the perpendicular and parallel polarization component vectors, as shown in the right side of Fig. 2.31, the polarization angle viewed at the BS will be same as that observed at the MS [59]. For this reason, each BS-scatterer-MS plane in Figs. 2.30 and 2.31 is called a *conservation-of-polarization plane* (CoP-plane).

The polarization angle associated with a CoP-plane, denoted by θ_V in Fig. 2.30, is defined as the angle between \underline{V} and the line that includes the projection of \underline{V} onto the CoP plane. Due to the conservation of polarization principle, the angle between

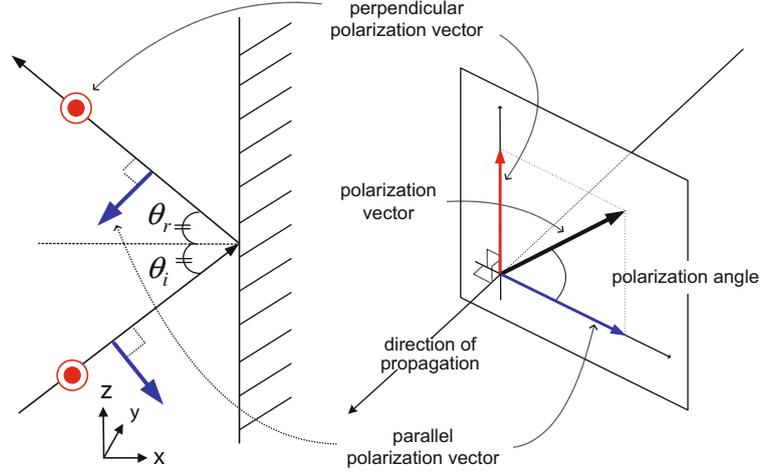


Fig. 2.31 Conservation of polarization (CoP) plane and polarization angle

\underline{V}' and the line that includes the projection of \underline{V}' onto the same CoP plane in Fig. 2.30 is equal to $\pi - \theta_V$ [59]. Thus, by using the polarization angle with respect to the CoP-plane, the polarization vector, \underline{V}' or \underline{H}' on its CoP-plane can be decomposed into vertical and horizontal polarization components with respect to the plane wave's direction of propagation at the MS.

Consider a vertically polarized plane wave from the BS, and let A_{VV} (A_{HV}) be the amplitude of the received plane wave at a vertically (horizontally) polarized MS antenna via a single scatterer. The geometry in Fig. 2.30 implies that A_{VV} and A_{HV} can be expressed as a function of D , D_{ts} , D_{sr} , R_r , α_R , and β_R as follows:

$$\begin{bmatrix} A_{VV} \\ A_{HV} \end{bmatrix} = A_V \begin{bmatrix} f_{VV}(D, D_{ts}, D_{sr}, R_r, \alpha_R, \beta_R) \\ f_{HV}(D, D_{ts}, D_{sr}, R_r, \alpha_R, \beta_R) \end{bmatrix}, \quad (2.189)$$

where A_V is the amplitude of \underline{V}' at the MS. The azimuth and elevation angles of departure (AAoDs and EAoDs) from the BS are dependent upon and are fixed by the azimuth and elevation angles of arrival (AAoAs and EAoAs) at the MS and, thus, the channel polarization functions f_{VV} and f_{HV} in (2.189) do not depend on α_T and β_T . Similarly, for the case of a horizontally polarized plane wave from the BS:

$$\begin{bmatrix} A_{VH} \\ A_{HH} \end{bmatrix} = A_H \begin{bmatrix} f_{VH}(D, D_{ts}, D_{sr}, R_r, \alpha_R, \beta_R) \\ f_{HH}(D, D_{ts}, D_{sr}, R_r, \alpha_R, \beta_R) \end{bmatrix}, \quad (2.190)$$

where A_H is the amplitude of \underline{H}' at the MS. From the geometry in Fig. 2.30, the distances are

$$d_t = \sqrt{(D + R_r \cos \alpha_R)^2 + (R_r \sin \alpha_R)^2}, \quad (2.191)$$

$$D_{ts} = \sqrt{d_t^2 + (\Delta_h - R_r \tan \beta_R)^2}, \quad (2.192)$$

$$D_{sr} = R_r / \cos \beta_R, d_1 = R_r \tan \beta_R, \quad (2.193)$$

$$D_{tr} = \sqrt{D^2 + \Delta_h^2}, d_2 = \sqrt{d_t^2 + \Delta_h^2}, \quad (2.194)$$

where $\Delta_h = (h_T - h_R)$. Based on (2.191)–(2.194) and the law of cosines, the angles in Fig. 2.30 can be written as

$$\cos \theta_3 = \frac{D_{ts}^2 + D_{tr}^2 - D_{sr}^2}{2D_{ts}D_{tr}} \frac{a_2 - a_3 + a_4}{\sqrt{(a_1^2 + 2a_2 - 2a_3 + a_4) a_4}}, \quad (2.195)$$

$$\cos \theta_4 = \frac{D_{ts}^2 + d_2^2 - d_1^2}{2D_{ts}d_2} \frac{(R_r/D)^2 + 2a_2 - a_3 + a_4}{\sqrt{(a_1^2 + 2a_2 - 2a_3 + a_4) (2a_2 + a_4 + a_5^2)}}, \quad (2.196)$$

$$\cos \theta_5 = \frac{d_2^2 + D_{tr}^2 - R_r^2}{2d_2 D_{tr}} = \frac{a_2 + a_4}{\sqrt{(2a_2 + a_4 + a_5^2) a_4}}, \quad (2.197)$$

$$\cos \theta_6 = \frac{\cos \theta_5 - \cos \theta_3 \cos \theta_4}{\sqrt{(1 - \cos^2 \theta_3) (1 - \cos^2 \theta_4)}}, \quad (2.198)$$

$$\cos \theta_7 = \frac{D_{sr}^2 + D_{tr}^2 - D_{ts}^2}{2D_{sr} D_{tr}} = \frac{(-a_2 + a_3) \cos \beta_R}{a_5 \sqrt{a_4}}, \quad (2.199)$$

$$\cos \theta_8 = \frac{D_{tr}^2 + R_r^2 - d_2^2}{2D_{tr} R_r} = -\frac{\cos \alpha_R}{\sqrt{a_4}}, \quad (2.200)$$

$$\cos \theta_9 = \frac{\cos \theta_8 - \cos \theta_7 \cos \beta_R}{\sqrt{(1 - \cos^2 \theta_7) (1 - \cos^2 \beta_R)}}, \quad (2.201)$$

$$\theta_V = \pi - \theta_6 \quad (2.202)$$

where

$$\begin{aligned} a_1 &= \frac{R_r/D}{\cos \beta_R}, a_2 = \frac{R_r}{D} \cos \alpha_R, a_3 = \frac{R_r \Delta_h}{D^2} \tan \beta_R, \\ a_4 &= 1 + \left(\frac{\Delta_h}{D}\right)^2, a_5 = \frac{R_r}{D}. \end{aligned} \quad (2.203)$$

The angles $\cos \theta_6$ and $\cos \theta_9$ in (2.198) and (2.201), respectively, can be used to derive the channel polarization functions f_{VV} , f_{HV} , f_{HH} , and f_{VH} in (2.189) and (2.190) as is now described.

From the geometry in Fig. 2.30, the amplitudes of the vertical and horizontal polarization components that are received from a vertically polarized transmitted plane wave are, respectively,

$$\begin{aligned} A_{VV} &= A_V |\cos(\theta_9 + (\pi - \theta_V) - \pi)| \\ &= A_V \left| \cos \theta_6 \cos \theta_9 - \sqrt{1 - \cos^2 \theta_6} \sqrt{1 - \cos^2 \theta_9} \right|, \end{aligned} \quad (2.204)$$

$$\begin{aligned} A_{HV} &= A_V |\sin(\theta_9 + (\pi - \theta_V) - \pi)| \\ &= A_V \left| \cos \theta_6 \sqrt{1 - \cos^2 \theta_9} + \cos \theta_9 \sqrt{1 - \cos^2 \theta_6} \right|, \end{aligned} \quad (2.205)$$

where $\cos \theta_6$ and $\cos \theta_9$ in (2.204) and (2.205), respectively, are functions of D , R_r , α_R , β_R as is evident from (2.195)–(2.201). Comparing (2.204) and (2.205) with (2.189), the channel polarization functions f_{VV} and f_{HV} can be written as:

$$\begin{aligned} &f_{VV}(D, D_{ts}, D_{sr}, R_r, \alpha_R, \beta_R) \\ &\equiv f_{VV}(D, R_r, \alpha_R, \beta_R) \\ &= \left| \cos \theta_6 \cos \theta_9 - \sqrt{1 - \cos^2 \theta_6} \sqrt{1 - \cos^2 \theta_9} \right|, \end{aligned} \quad (2.206)$$

$$\begin{aligned} &f_{HV}(D, D_{ts}, D_{sr}, R_r, \alpha_R, \beta_R) \\ &\equiv f_{HV}(D, R_r, \alpha_R, \beta_R) \\ &= \left| \cos \theta_6 \sqrt{1 - \cos^2 \theta_9} + \cos \theta_9 \sqrt{1 - \cos^2 \theta_6} \right|. \end{aligned} \quad (2.207)$$

In a similar fashion, the channel polarization functions f_{HH} and f_{VH} can be written as:

$$\begin{aligned} &f_{VH}(D, D_{ts}, D_{sr}, R_r, \alpha_R, \beta_R) \\ &\equiv f_{VH}(D, R_r, \alpha_R, \beta_R) \end{aligned}$$

$$= \left| \cos \theta_6 \sqrt{1 - \cos^2 \theta_9} + \cos \theta_9 \sqrt{1 - \cos^2 \theta_6} \right|, \quad (2.208)$$

$$\begin{aligned} & f_{\text{HH}}(D, D_{\text{ts}}, D_{\text{sr}}, R_r, \alpha_R, \beta_R) \\ & \equiv f_{\text{HH}}(D, R_r, \alpha_R, \beta_R) \\ & = \left| \cos \theta_6 \cos \theta_9 - \sqrt{1 - \cos^2 \theta_6} \sqrt{1 - \cos^2 \theta_9} \right|. \end{aligned} \quad (2.209)$$

Note that a symmetry exists in the co- and cross-polarization functions such that

$$f_{\text{VV}} = f_{\text{HH}}, f_{\text{HV}} = f_{\text{VH}}. \quad (2.210)$$

This model may be enhanced so as to include reflection coefficients following the methodology in [189].

2.4.2 XPD in Polarized Multipath-Fading Channels

The polarized channels in Fig. 2.30 can be modeled by combining the channel polarization functions with the conventional co-polarized channel impulse response functions derived for the exact same 3-D single cylinder scattering geometry shown in Fig. 2.30. First consider the case of vertically polarized BS and MS antennas (VV-channel). The complex low-pass channel impulse response for the VV-channel can be expressed as

$$g_{\text{VV}}(t) = g_{\text{VV}}^{\text{SBR}}(t) + g_{\text{VV}}^{\text{LoS}}(t), \quad (2.211)$$

where

$$\begin{aligned} g_{\text{VV}}^{\text{SBR}}(t) &= \sum_{n=1}^N A_{V,n} f_{\text{VV},n} e^{j\phi_n(t)}, \\ g_{\text{VV}}^{\text{LoS}}(t) &= A_V^{\text{LoS}} e^{j\phi^{\text{LoS}}(t)}, \end{aligned} \quad (2.212)$$

are the diffuse single-bounce receive-side (SBR) and LoS components, respectively, and

$$A_{V,n} = A_V(\alpha_{R,n}, \beta_{R,n}), \quad (2.213)$$

$$f_{\text{VV},n} = f_{\text{VV}}(D, R_r, \alpha_{R,n}, \beta_{R,n}), \quad (2.214)$$

$$\phi_n(t) = 2\pi(f_c + f_{D,n}) \left(t - (D_{\text{ts},n} + D_{\text{sr},n}) / c_0 \right) + \phi_n, \quad (2.215)$$

$$\phi^{\text{LoS}}(t) = 2\pi(f_c + f_D^{\text{LoS}}) \left(t - D_{\text{tr}} / c_0 \right), \quad (2.216)$$

$$f_c = \frac{c_0}{\lambda}, \quad f_{D,n} = \left(\frac{v_R}{\lambda} \right) \cos \beta_{R,n} \cos(\gamma_R - \alpha_{R,n}), \quad (2.217)$$

$$f_D^{\text{LoS}} = \frac{D}{\sqrt{D^2 + \Delta_n^2}} \left(\frac{v_R}{\lambda} \right) \cos(\gamma_R - \pi), \quad (2.218)$$

where c_0 , λ , v_R , and γ_R are the velocity of light, the carrier wavelength, the velocity of the MS, and the angle of the MS's moving direction with respect to the positive x -axis, respectively. Further, f_c and $f_{D,n}$ are the carrier and Doppler frequency associated with the n th arriving plane wave, respectively. Note that the MS is considered to be in the center of the scattering cylinder regardless of the physical location and velocity of the MS, which is a typical assumption for geometrical channel modeling. Of course the exact location of the scatterers on the cylinder will change with the physical location of the MS, due to site specific propagation conditions. In this case, the channel is modeled as being locally wide-sense stationary, i.e., the locations of the scatterers with respect to the MS will stay the same if the MS moves over small distances.

It is assumed at this stage that the phases $\{\phi_n\}$ in (2.215) are independent and identically distributed (i.i.d.) uniform random variables on the interval $[-\pi, \pi)$, and the $\{\phi_n\}$ are independent of the AAoAs $\{\alpha_{R,n}\}$, EAoAs $\{\beta_{R,n}\}$ and the radius of the scattering cylinder. For large N , the central limit theorem can be invoked and $g_{VV}(t)$ can be treated as a complex Gaussian random process. By using the complex low-pass channel impulse response and the aforementioned properties, it is straightforward to obtain the received power through the VV-channel at the Rx as follows:

$$\Omega_{VV} = \Omega_{VV}^{\text{SBR}} + \Omega_{VV}^{\text{LoS}} = (1 + K) \Omega_{VV}^{\text{SBR}}, \quad (2.219)$$

$$\begin{aligned} \Omega_{VV}^{\text{SBR}} &= \text{E} \left[|g_{VV}^{\text{SBR}}(t)|^2 \right] \\ &= \text{E} \left[\left(\sum_{n=1}^N A_{V,n} f_{VV,n} e^{j\phi_n(t)} \right) \left(\sum_{m=1}^N A_{V,m} f_{VV,m} e^{-j\phi_m(t)} \right) \right] \\ &= \text{E} \left[\sum_{n=1}^N A_{V,n}^2 f_{VV,n}^2 \right], \end{aligned} \quad (2.220)$$

$$\Omega_{VV}^{\text{LoS}} = |A_V^{\text{LoS}}|^2 = K \Omega_{VV}^{\text{SBR}}, \quad (2.221)$$

where K denotes the Ricean K-factor, and the third equality in (2.220) is due to the fact that the phases $\{\phi_n\}$ in (2.215) are independent and identically distributed (i.i.d.) uniform random variables on the interval $[-\pi, \pi)$. Further, as $N \rightarrow \infty$, the discrete AAoAs, $\alpha_{R,n}$, and discrete EAoAs, $\beta_{R,n}$, can be replaced with continuous random variables α_R and β_R having the joint probability density function (pdf), $f(\alpha_R, \beta_R)$. Here, the azimuth and elevation angles of arrival are assumed independent of each other and, thus, the joint pdf $f(\alpha_R, \beta_R)$ can be decomposed as $f(\alpha_R)f(\beta_R)$. It follows that

$$\Omega_{VV}^{\text{SBR}} = A_V^2 \int_{-\pi}^{\pi} \int_{-\beta_{R,\text{Max}}}^{\beta_{R,\text{Max}}} f_{VV}^2(D, R, \alpha_R, \beta_R) p_r(\beta_R) p_r(\alpha_R) d\beta_R d\alpha_R, \quad (2.222)$$

where $\beta_{R,\text{Max}}$ is the absolute value of the maximum EAoA.

Several different distributions have been suggested in the literature to characterize the AAoA distribution, such as uniform, Gaussian, and Laplacian. Later, the von Mises distribution in (2.51) will be used such that

$$p_r(\alpha_R) = \frac{\exp[k_R \cos(\alpha_R - \mu_R)]}{2\pi I_0(k_R)}, \quad k_R \geq 0, \quad (2.223)$$

where $\alpha_R \in [-\pi, \pi)$, $I_0(\cdot)$ is the zero-order modified Bessel function of the first kind, $\mu_R \in [-\pi, \pi)$ is the mean AAoA of the scatterers in the x - y plane with respect to the positive x axis, and k_R controls the spread of the scatterers around the mean AAoA μ_R . Prior work has also used several different distributions to characterize the EAoA distribution. Later, the cosine pdf in (2.52) will be used such that

$$p_r(\beta_R) = \frac{\pi}{4\beta_{R,\text{Max}}} \cos\left(\frac{\pi}{2} \frac{\beta_R}{\beta_{R,\text{Max}}}\right), \quad |\beta_R| \leq \beta_{R,\text{Max}} \leq \pi/2, \quad (2.224)$$

where $\beta_{R,\text{Max}}$ typically lies in the range $20^\circ < \beta_{R,\text{Max}} < 45^\circ$ for fixed-to-mobile land mobile radio channels [185, 191, 257]. Of course, other suitable distributions for $p_r(\alpha_R)$ and $p_r(\beta_R)$ can be used as well at this stage.

Similar to (2.222), the SBR power at the MS with a horizontally polarized MS antenna and a vertically polarized BS antenna (HV-channel), a vertically polarized MS antenna and a horizontally polarized BS antenna (VH-channel), and horizontally polarized MS and BS antennas (HH-channel), denoted by P_{HV} , P_{VH} , and P_{HH} , respectively, can be expressed as

$$\Omega_{\text{HV}}^{\text{SBR}} = A_V^2 \int_{-\pi}^{\pi} \int_{-\beta_{R,\text{Max}}}^{\beta_{R,\text{Max}}} \left(f_{\text{HV}}(D, R, \alpha_R, \beta_R) \right)^2 p_r(\beta_R) p_r(\alpha_R) d\beta_R d\alpha_R, \quad (2.225)$$

$$\Omega_{\text{VH}}^{\text{SBR}} = A_H^2 \int_{-\pi}^{\pi} \int_{-\beta_{R,\text{Max}}}^{\beta_{R,\text{Max}}} \left(f_{\text{VH}}(D, R_r, \alpha_R, \beta_R) \right)^2 p_r(\beta_R) p_r(\alpha_R) d\beta_R d\alpha_R, \quad (2.226)$$

$$\Omega_{\text{HH}}^{\text{SBR}} = A_H^2 \int_{-\pi}^{\pi} \int_{-\beta_{R,\text{Max}}}^{\beta_{R,\text{Max}}} \left(f_{\text{HH}}(D, R_r, \alpha_R, \beta_R) \right)^2 p_r(\beta_R) p_r(\alpha_R) d\beta_R d\alpha_R. \quad (2.227)$$

Notice that by taking into account the statistical distribution of scatterers, the total power of the vertical and horizontal polarization components can be derived via the superposition of the CoP-plane components by averaging over the pdfs of the AAoA and EAoA, $p(\alpha_R)$ and $p(\beta_R)$, respectively. Finally, the average XPD values are obtained from (2.222) and (2.225)–(2.227) as

$$\text{XPD}_{\text{VV}/\text{HV}} = \frac{\Omega_{\text{VV}}}{\Omega_{\text{HV}}}, \text{XPD}_{\text{HH}/\text{VH}} = \frac{\Omega_{\text{HH}}}{\Omega_{\text{VH}}}. \quad (2.228)$$

2.4.3 Polarized Channel Model Application

This section compares the analytical XPD predicted from the geometrical depolarization model developed in Sects. 2.4.1 and 2.4.2 with some previously reported XPD results obtained from empirical data [110, 306]. The empirical XPD data reported in [110, 306] were based on narrow-band fixed-to-mobile channel measurements taken in a suburban area at 2.48 GHz, where the corresponding h_T and h_R in Fig. 2.30 were 20 m and 3 m, respectively.

The parameters in (2.223)–(2.224) are set as $\mu_R = 180^\circ$ and $\beta_{R,\text{Max}} = 40^\circ$, unless otherwise stated, particularly when comparing our analytical XPD results with the XPD measurement data in [110, 306]. It is stated in [110, 306] that the direction of the strongest signal was the Tx–Rx bore sight path in most cases, corresponding to $\mu_R = 180^\circ$ in the model described in Sects. 2.4.1 and 2.4.2. The spread constant of the AAoA distribution, k_R , in (2.223) can exhibit substantial variation from one location to the next due to variations in the local scattering environment around the Rx. Finally, R_r in Fig. 2.30 is set to 30 m.

2.4.3.1 Suburban Area in San Jose

Figures 2.32 and 2.33 portray the analytical mean XPD curves of F2M fading channels as k_R varies from 0 to 9 and from 10 to 700, respectively. The same antenna configuration as that in [110, 306] is utilized in the analysis to allow for comparison with the measurement data illustrated in Fig. 2.32, which is a $\pm 45^\circ$ dual-polarization antenna at both the BS and MS. Note that a directional MS antenna with the azimuthal beamwidth of 90° and a gain of 12 dBi was used in the measurements in [110, 306]. Figure 2.32 shows that the XPD curve tends to decrease from roughly 6.5–0 dB as k_R increases from 0 to 9. On the other hand, the XPD curve increases from 0 dB to roughly 7.5 dB as k_R increases from 10 to 700 in Fig. 2.33. The XPD curve for $0 \leq k_R \leq 700$ at a fixed BS–MS distance, $D = 1$ km, is also illustrated in the insert of Fig. 2.33 to further illustrate the effect of k_R on the XPD shown in Figs. 2.32 and 2.33. The reason for the strong dependency of the mean XPD on k_R will be explained later when discussing Fig. 2.34.

When k_R is in the range 7–10, the XPD stays at around 0 dB regardless of the distance between the BS and MS, and it can be shown that the AAoA spread in (2.223) is roughly 90° when k_R is 7–10 [3]. As the distance between the BS and MS, D , increases it is expected that k_R will tend to decrease since the plane waves at the MS for large distances and NLoS conditions will come from many angles due to local scattering and, hence, the azimuthal scattering will become more isotropic; $k_R = 0$ corresponds to the case of isotropic azimuthal scattering. However, in the case of isotropic scattering the observed AAoA spread will be determined by the beam pattern of the 90° directional Rx antenna that was used to make the measurements. Thus, while the actual scattering environment may be near isotropic, i.e., $k_R \approx 0$, the observed value of k_R will lay in the range 7–10. For this reason, the empirical XPD tends to converge toward 0 dB as D increases, and it is in close agreement with the analytical XPD results with k_R in the range 7–10, as previously stated.

The AAoA spread factor, k_R , is expected to increase as the horizontal BS–MS distance, D , decreases. As the MS moves closer to the BS most of the scatters will tend to be in the direction of the Tx so that k_R will tend to increase and the scattering will be non-isotropic. If the AAoA distribution is modeled using the von Mises pdf in (2.223) with $\mu_R = 180^\circ$,

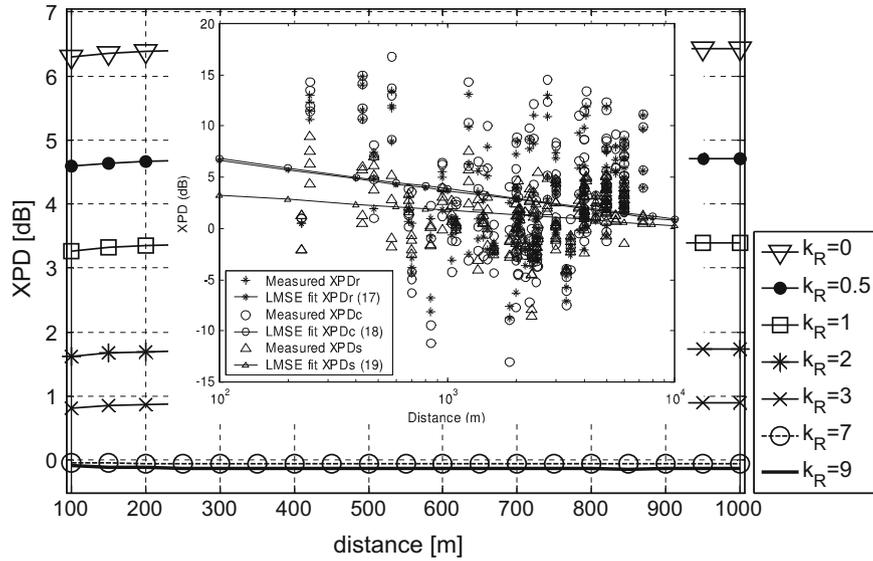


Fig. 2.32 $\overline{\text{XPD}}^{F2M}$ curves for a variety of k_R (0–9) with BS–MS distance in the range 100 m to 1 km, and the measurement data in [306]

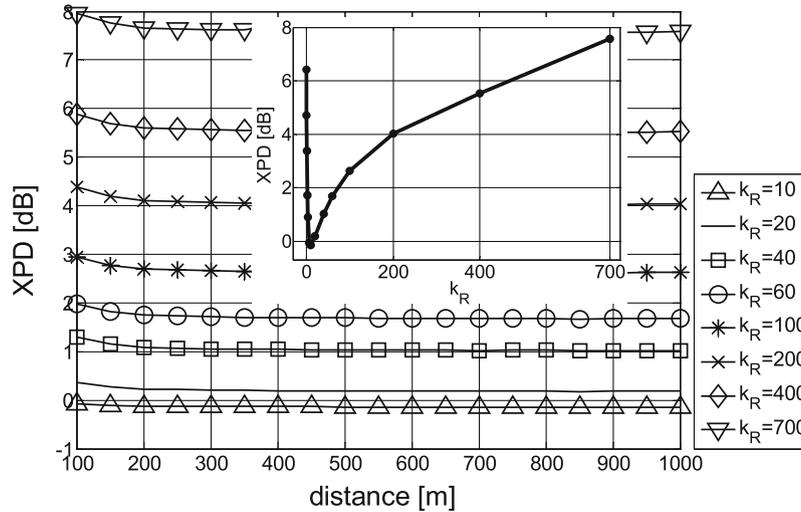


Fig. 2.33 $\overline{\text{XPD}}^{F2M}$ curves for a variety of k_R (10–700) with BS–MS distance in the range 100 m to 1 km

then the effective value of k_R will be greater than 7, since a directional antenna with an azimuthal beamwidth of 90° was used in the measurements. Notice that as k_R increases, the azimuthal angular spread becomes narrower. When the actual AAoA spread factor, k_R , is equal to 3 for example, $p_r(\alpha_R = 180^\circ)|_{k_R=3} \simeq 0.66$, and the effective value of k_R due to the directional Rx antenna gain of 12 dBi (15.8 in linear scale) in the direction of BS–MS boresight ($\alpha_R = 180^\circ$) is roughly $0.66 \times 15.8 \simeq 10.43$, which is close to $p_r(\alpha_R = 180^\circ)|_{k_R=700} \simeq 10.55$. Thus, while the actual value of the AAoA spread factor may be in the neighborhood of $k_R = 3$, the observed value may be considerably higher and in the range of $k_R = 700$ due to the gain of the directional MS measurement antenna.

The empirical XPD data from [110, 306] are illustrated in Fig. 2.32, where XPD_r , XPD_c , and XPD_s represent the XPD of the total received signal, the XPD of the time-invariant portion of the received signal, and the XPD of the time-varying portion of the received signal. The empirical data in [110, 306] shows that the LMSE curve fit for the XPD of total received signal is almost same as the XPD for only the time-invariant portion of the received signal. This is expected, since the measurements in [110, 306] were made with a stationary MS. Most of the surrounding scatterers will be stationary and contribute to XPD_c , while only the moving scatterers such as nearby moving vehicles (which are not included in the model) will contribute to XPD_s . It is also shown that the LMSE curve fit of XPD_r tends to decrease from 6.5 to 0 dB as D increases from 100 m to 10 km. This phenomenon can be explained by the XPD curve for varying k_R inserted in Fig. 2.33, such that a

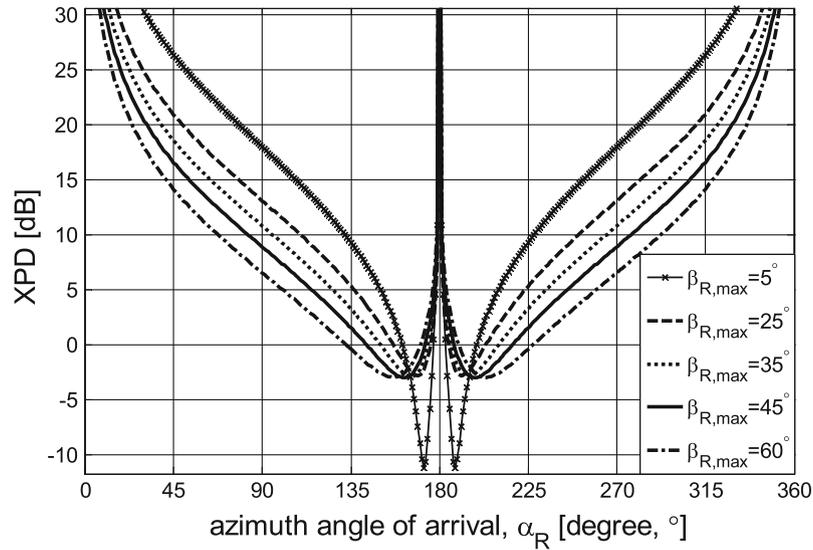


Fig. 2.34 $\overline{\text{XPD}}^{F2M}$ curves for varying α_R and $\beta_{\text{Max},R}$

decrease in the effective value of k_R from 550 to 7 corresponds to a decrease in the XPD from 6.5 to 0, as D increases from 100 m to 10 km. Even though an LMSE curve fit can be obtained from the measurement data, the measured XPD values show substantial variation even at the same D as portrayed in Figs. 2.32 and 2.33. The reason is that the AAoA spread factor k_R varies depending on the site specific local scattering environment, even at the same D . The XPD is not directly dependent on D , but k_R is expected to have a tendency to decrease with increasing D as mentioned earlier, which is reflected on the LMSE curve in the illustrated empirical results. Finally, the analytical XPD curves in the range 1–10 km are omitted from Figs. 2.32 and 2.33, since the XPD remains constant after 1 km.

The AAoA α_R has a significant impact on the XPD as shown in Fig. 2.34, and computing the XPD as a function of the AAoA α_R provides insight for the mean XPD that is obtained when averaging over the AAoA and EAoA distributions. Figure 2.34 also shows the XPD averaged over the EAoA distribution only at each AAoA, α_R . For AAoAs equal to 180° , 0° and 360° , which represent vertical CoP planes, the XPD goes infinity, meaning that there are no cross-polarization components at those angles. This makes sense given the geometry in Fig. 2.30. Remarkably, each XPD curve has a minimum point when the AAoA α_R lays in the range $155\text{--}170^\circ$. The XPD at this minimum point is below 0 dB, meaning that the power of the cross-polarization component is stronger than that of the co-polarization component. Figure 2.34 also shows that a larger maximum EAoA will result in a smaller mean XPD, since the XPD is below 0 dB over a larger AAoA range. Considering the XPD averaged over AAoA distribution with $\mu_R = 180^\circ$ and k_R varying from 0 to 9, it can be recognized that as k_R increases the mean XPD decreases, since the impact of the small XPDs for AAoAs around the minimum point will dominate the high XPDs for AAoAs that are far from the minimum point. Furthermore, small changes in the AAoA distribution near 180° (bore sight) due to site specific local scattering will cause significant changes in the mean XPD which explains the high variance of the empirical XPD values observed in [110, 306].

2.5 Simulation of Multipath-Fading Channels

A channel simulator is an essential component for the development and testing of wireless systems. Simulation of mobile radio channels is commonly used as opposed to field trials, because it allows for less expensive and more reproducible system tests and evaluations. For this purpose it is desirable to generate complex faded envelopes that match the statistical characteristics of a reference model while at the same time having low complexity. For example, the reference model might be a 2-D isotropic scattering channel with an isotropic receiving antenna.

Two fundamentally different approaches to channel simulation models are the filtered white Gaussian noise models and the sum-of-sinusoids models. The basic idea of the filtered white noise approach is to shape the power spectrum of a white Gaussian noise process by using a filter whose magnitude response is the square root of the desired Doppler spectrum. The sum of sinusoids approach, on the other hand, approximates the complex fading envelope by the superposition of a finite number of sinusoids with properly chosen amplitudes, frequencies, and phases to yield the desired Doppler spectrum.

2.5.1 Filtered White Gaussian Noise Models

Recall that the faded envelope $g(t) = g_I(t) + jg_Q(t)$ is a complex Gaussian random process. A straightforward method for constructing a fading simulator is to low-pass filter two independent white Gaussian noise processes, as shown in Fig. 2.35. Since the input processes to the low-pass filters are independent, the random processes $g_I(t)$ and $g_Q(t)$ are also independent and have power spectral densities that are shaped by the squared magnitude response of the low-pass filters. If the white Gaussian noise sources have power spectral densities equal to $\Omega_p/2$ watts/Hz and the low-pass filters have transfer function $H(f)$, then

$$S_{g_I g_I}(f) = S_{g_Q g_Q}(f) = \frac{\Omega_p}{2} |H(f)|^2$$

$$S_{g_I g_Q}(f) = 0, \tag{2.229}$$

where the normalization $\int_{-\infty}^{\infty} |H(f)|^2 df = 1$ is assumed so that the envelope power is equal to Ω_p . The processes $g_I(t)$ and $g_Q(t)$ are independent zero-mean Gaussian random process, such that the envelope $\alpha \equiv |g(t_1)|$ is Rayleigh distributed at any time t_1 . The problem is to construct the low-pass filter $H(f)$ to give the desired Doppler spectrum.

2.5.1.1 IDFT Method

One approach for generating the faded envelope using the filtering method has been suggested by Young and Beaulieu [369] and is based on an inverse discrete Fourier transform (IDFT) as shown in Fig. 2.36. The input sequences $\{A[k]\}$ and $\{B[k]\}$ are first generated, each consisting of N i.i.d. real zero-mean Gaussian random variables with variance $N\Omega_p/2$. These samples are then applied to a filter with frequency response $H[k]$, $k = 0, \dots, N - 1$, followed by an N -point IDFT to generate the time-domain samples of the complex faded envelope as

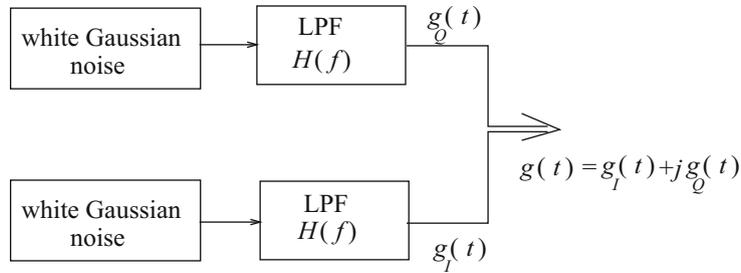


Fig. 2.35 Fading simulator that uses low-pass filtered white Gaussian noise

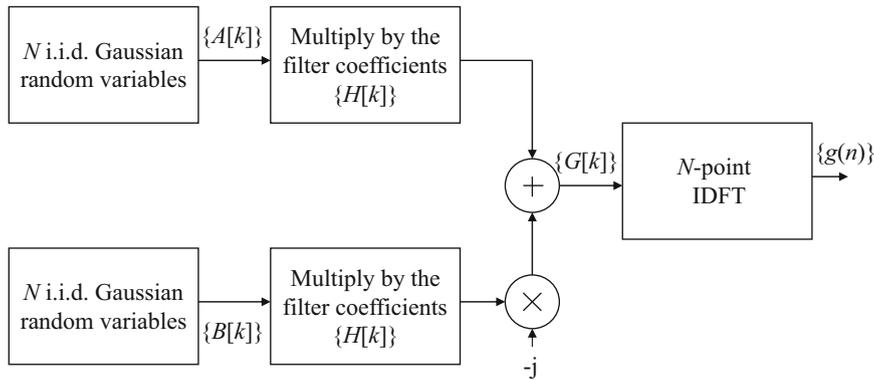


Fig. 2.36 IDFT-based fading simulator

$$\begin{aligned}
g[n] &= \text{IDFT} \{A[k]H[k] - jB[k]H[k]\}_{n=0}^{N-1} \\
&= \frac{1}{N} \sum_{k=0}^{N-1} (A[k]H[k] - jB[k]H[k]) e^{j2\pi kn/N}, \quad n = 0, \dots, N-1.
\end{aligned} \tag{2.230}$$

Due to linearity of the IDFT operation, the discrete-time autocorrelation function of the output sample sequence $\{g[n]\} = \{g_I[n]\} + j\{g_Q[n]\}$ is [369]

$$\begin{aligned}
\phi_{gg}[n] &= \phi_{g_I g_I}[n] + j\phi_{g_I g_Q}[n] \\
&= \frac{\Omega_p}{2} (a_I[n] + ja_Q[n]),
\end{aligned} \tag{2.231}$$

where $\{a[n]\} = \{a_I[n]\} + j\{a_Q[n]\}$ is given by the IDFT of the sequence $\{(H[k])^2\}$, i.e.,

$$a[n] = \frac{1}{N} \sum_{k=0}^{N-1} (H[k])^2 e^{j2\pi kn/N}, \quad n = 0, \dots, N-1. \tag{2.232}$$

The only problem remaining is to design the filter, $H[k]$. To do so, the argument by Young and Beaulieu [369] is followed.

A sequence $\{A_{CS}[k]\}$ that satisfies the property $A_{CS}[k] = A_{CS}^*[N-k]$ is called a conjugate-symmetric sequence, while a sequence $\{A_{CAS}[k]\}$ that satisfies the property $A_{CAS}[k] = -A_{CAS}^*[N-k]$ is called a conjugate-antisymmetric sequence. The IDFT of a conjugate-symmetric sequence is real-valued, while the IDFT of a conjugate-antisymmetric sequence is imaginary-valued. If the sequence $\{a[n]\}$ has discrete Fourier transform (DFT) $\{A[k]\}$, then

$$\begin{aligned}
a[n] &= a_I[n] + ja_Q[n] = \text{IDFT}\{A[k]\} \\
&= \text{IDFT}\{A_{CS}[k]\} + \text{IDFT}\{A_{CAS}[k]\}.
\end{aligned} \tag{2.233}$$

Also, note that

$$A_{CS}[k] = \frac{1}{2}A[k] + \frac{1}{2}A^*[N-k] \tag{2.234}$$

and

$$A_{CAS}[k] = \frac{1}{2}A[k] - \frac{1}{2}A^*[N-k] \tag{2.235}$$

for any sequence $\{A[k]\}$, where $A[k] = A_{CS}[k] + A_{CAS}[k]$. Thus, the sequence $\{A[k]\}$ can always be decomposed into the sum of conjugate-symmetric and conjugate-antisymmetric components. To ensure that the quadrature components of the faded envelope, $\{g_I[n]\}$ and $\{g_Q[n]\}$, are uncorrelated, which is the case for 2-D isotropic scattering environment and an isotropic receive antenna, it is required that $A_{CAS}[k] = 0$ for all k . Using (2.234), the filter should have the following structure:

$$H[k] = \begin{cases} \sqrt{H_s[0]} & , \quad k = 0 \\ \sqrt{\frac{H_s[k]}{2}} & , \quad k = 1, 2, \dots, (\frac{N}{2} - 1) \\ \sqrt{H_s[k]} & , \quad k = \frac{N}{2} \\ \sqrt{\frac{H_s[N-k]}{2}} & , \quad k = (\frac{N}{2} + 1), \dots, (N-1) \end{cases}, \tag{2.236}$$

where the filter $H_s[k]$ can be obtained from the sampled Doppler spectrum.

To obtain the required filter $H[k]$, assume 2-D isotropic scattering with an isotropic receiver antenna. The required normalized autocorrelation function and normalized Doppler spectrum are, respectively,

$$\phi_{gg}(\tau) = J_0(2\pi f_m \tau) \tag{2.237}$$

and

$$S_{gg}(f) = \begin{cases} \frac{1}{\pi f_m} \frac{1}{\sqrt{1-(f/f_m)^2}} & |f| \leq f_m \\ 0 & \text{otherwise} \end{cases}. \quad (2.238)$$

Ideally, the generated sequence $g[n]$ should have the normalized autocorrelation function

$$\phi_{gg}[n] = J_0(2\pi \hat{f}_m |n|), \quad (2.239)$$

where $\hat{f}_m = f_m T_s$ is the maximum normalized Doppler frequency and T_s is the sampling period. However, an exact realization of this autocorrelation function is not possible, because the time-domain sequence is truncated to N samples. Looked at another way, while the theoretical Doppler spectrum is bandlimited, the truncation to N samples in the time-domain means that the realized Doppler spectrum is not bandlimited.

To obtain the required filter, the continuous spectrum in (2.238) is sampled at frequencies $f_k = k/(NT_s)$, $k = 0, \dots, N-1$. Special treatment is given to the frequency domain coefficients at two points. The first is at zero frequency, where $H_s[0] = 0$ to ensure that the generated time-domain sequence always has zero mean regardless of the particular values assumed by the length- N input sequences $\{A[k]\}$ and $\{B[k]\}$. The second is at the index k_m that is at, or just below, the maximum Doppler frequency, i.e.,

$$k_m = \lfloor f_m NT_s \rfloor = \lfloor \hat{f}_m N \rfloor, \quad (2.240)$$

where $\lfloor x \rfloor$ is the largest integer contained in x . The realized maximum Doppler frequency is $k_m/(NT_s)$ Hz. The area under the Doppler spectrum curve in (2.238) from zero to frequency f is [147, 2.271.4]

$$C(f) = \frac{k_m}{NT_s} \arcsin(fNT_s/k_m), \quad 0 \leq f \leq k_m/(NT_s). \quad (2.241)$$

The area under the Doppler spectrum between the frequencies represented by the samples $(k_m - 1)$ and k_m is equal to $C(k_m/(NT_s)) - C((k_m - 1)/(NT_s))$. Approximating this area by a rectangle of height $(H_s[k_m])^2$ and width $1/(NT_s)$ gives

$$H_s[k_m] = \sqrt{k_m \left[\frac{\pi}{2} - \arctan \left(\frac{k_m - 1}{\sqrt{2k_m - 1}} \right) \right]}. \quad (2.242)$$

The complete filter $H[k]$ can now be specified as follows:

$$H[k] = \begin{cases} 0 & , k = 0 \\ \sqrt{\frac{1}{2\pi f_m \sqrt{1-(k/(N\hat{f}_m))^2}}} & , k = 1, 2, \dots, k_m - 1 \\ \sqrt{k_m \left[\frac{\pi}{2} - \arctan \left(\frac{k_m - 1}{\sqrt{2k_m - 1}} \right) \right]} & , k = k_m \\ 0 & , k = k_m + 1, \dots, N - k_m - 1 \\ \sqrt{k_m \left[\frac{\pi}{2} - \arctan \left(\frac{k_m - 1}{\sqrt{2k_m - 1}} \right) \right]} & , k = N - k_m \\ \sqrt{\frac{1}{2\pi f_m \sqrt{1-(N-k)/(N\hat{f}_m)^2}}} & , N - k_m + 1, \dots, N - 1 \end{cases} \quad (2.243)$$

The IDFT method will generate a Rayleigh faded envelope with an autocorrelation function that closely matches the reference model for large N . However, the main limitation of the IDFT approach arises from the block-oriented structure which precludes continuous transmission in that there is a discontinuity in the time series (i.e., the faded envelope) from one block of N samples to the next.

2.5.1.2 IIR Filtering Method

Another approach implements the filters in the time domain as finite impulse response (FIR) or infinite impulse response (IIR) filters. There are two main challenges with this approach. The first challenge arises from the fact that the sampled channel waveform is bandlimited to a discrete frequency $\hat{f}_m = f_m T_s$, where T_s is the sample period. Consider, for example, a cellular system operating at a carrier frequency of $f_c = 1800$ MHz with a maximum MS speed of 300 km/h. In this case, the maximum Doppler frequency is $f_m = f_c(v/\lambda_c) = 500$ Hz, where c is the speed of light. If the signal is sampled at rate $R_s = 1/T_s = 1$ MHz, then the normalized Doppler frequency is $\hat{f}_m = f_m T_s = 0.0005$. If implemented as a finite impulse response (FIR) filter, such an extremely narrowband filter would require an impractically high filter order. Fortunately, this can be satisfied by using a filter designed at a lower sampling frequency followed by an interpolator to increase the sampling frequency. For example, the filter could be designed at a sampling frequency of 2 kHz, which is 2 times the Nyquist frequency. Afterwards interpolate by a factor of $I = 500$ to obtain the desired sampling frequency of 1 MHz.

The second main challenge is that the square-root of the target Doppler spectrum for 2-D isotropic scattering and an isotropic receive antenna in (2.238) is irrational and, therefore, none of the straightforward filter design methods can be applied. However, an approach developed by Steiglitz [307] allows the design of an IIR filter with an arbitrary magnitude response. Another possibility is to use the MATLAB function `iirlpnorm` [221].

Consider an IIR filter of order $2K$ that is synthesized as the cascade of K Direct-Form II second-order (two poles and two zeroes) sections (biquads) having the form

$$H(z) = A \prod_{k=1}^K \frac{1 + a_k z^{-1} + b_k z^{-2}}{1 + c_k z^{-1} + d_k z^{-2}}. \quad (2.244)$$

For example, for $f_m T_s = 0.4$, $K = 5$, and an ellipsoidal accuracy of 0.01, the filter design procedure described by Kominakis [181] results in the coefficients tabulated in Table 2.1. Figure 2.37 plots the magnitude response of the designed filter, which is shown to closely match that of the theoretical ideal filter.

2.5.2 Sum of Sinusoids Methods

Sum-of-sinusoids (SoS) channel models attempt to simulate the channel as a stationary complex Gaussian random process, formed by the sum of multiple sinusoids having appropriately selected frequencies, amplitudes, and phases. The objective is to generate a faded envelope having statistical properties that are as close as possible to a specified reference model, while at the same time minimizing the number of sinusoids that are required to achieve a given degree of modeling accuracy. SoS models are broadly categorized as either *deterministic* or *statistical*. Deterministic SoS models use fixed frequencies, amplitudes, and phases for the sinusoidal components. Therefore, the statistical properties of the faded envelope are deterministic for all simulation trials. Such models are useful for simulations that require continuous transmission over

Table 2.1 Coefficients for $K = 5$ biquad stage elliptical filter

Stage	Filter coefficients			
k	a_k	b_k	c_k	d_k
1	1.5806655278853	0.99720549234156	-0.64808639835819	0.88900798545419
2	0.19859624284546	0.99283177405702	-0.62521063559242	0.97280125737779
3	-0.60387555371625	0.9999939585621	-0.62031415619505	0.99996628706514
4	-0.56105447536557	0.9997677910713	-0.79222029531477	0.2514924845181
5	-0.39828788982331	0.99957862369507	-0.71405064745976	0.64701702807931
A	0.020939537466725			

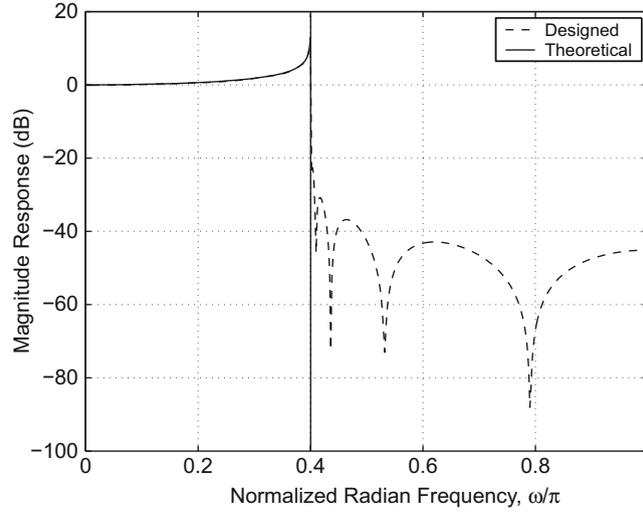


Fig. 2.37 Magnitude response of the designed shaping filter with $K = 5$ biquad sections

a long time interval, such as a real-time hardware channel simulator. In contrast, the statistical SoS models require multiple simulation trials, where one or more of the parameter sets (frequencies, amplitudes, or phases) are randomly selected for each simulation trial. As a result, the simulated channels have statistical properties that vary for each simulation trial, but they converge to the desired statistical properties when averaged over a sufficiently large number of simulation trials. Since a statistical model requires multiple simulation trials, it cannot be used in cases where continuous transmission is required. An ergodic statistical SoS model is one whose statistical properties converge to the desired properties in a single simulation trial. For this reason an ergodic statistical model is essentially a deterministic model, but differs in the sense that one of the parameter sets, usually the phases, is randomly generated when the simulator is initialized.

2.5.2.1 Clarke's Model

Clarke derived a statistical SoS simulation model that begins with (2.10) and (2.6) and assumes equal strength multipath components ($C_n = \sqrt{1/N}$). The received complex envelope has the form

$$g(t) = \sqrt{\frac{1}{N}} \sum_{n=1}^N e^{j(2\pi f_m t \cos(\theta_n) + \hat{\phi}_n)}, \quad (2.245)$$

where N is the number of sinusoids and the phases $\hat{\phi}_n$ are independent identically distributed (i.i.d.) uniform random variables on $[-\pi, \pi)$.

If a reference model having 2-D isotropic scattering and an isotropic antenna is assumed, then the θ_n are also i.i.d. uniform random variables on $[-\pi, \pi)$, and are independent of the $\hat{\phi}_n$. Based on the above assumptions concerning the C_n , $\hat{\phi}_n$, and θ_n , the ensemble averaged correlation functions of Clarke's model in (2.245) for a *finite* N are³

$$\phi_{g_I g_I}(\tau) = \phi_{g_Q g_Q}(\tau) = \frac{1}{2} J_0(2\pi f_m \tau) \quad (2.246)$$

$$\phi_{g_I g_Q}(\tau) = \phi_{g_Q g_I}(\tau) = 0 \quad (2.247)$$

$$\phi_{gg}(\tau) = \frac{1}{2} J_0(2\pi f_m \tau) \quad (2.248)$$

$$\phi_{\alpha^2 \alpha^2}(\tau) = 1 + \frac{N-1}{N} J_0^2(2\pi f_m \tau). \quad (2.249)$$

³Note that $\Omega_p = \sum_{n=1}^N C_n^2 = 1$ in this case; other values of Ω_p can be obtained by straightforward scaling.

Note that for finite N , the auto- and cross-correlation functions of $g_I(t)$ and $g_Q(t)$ match those of the reference model in (2.29) and (2.31), while the squared-envelope autocorrelation function reaches the desired form $1 + J_0^2(2\pi f_m \tau)$ in (2.82) asymptotically as $N \rightarrow \infty$.

2.5.2.2 Jakes' Model

Jakes [171] derived a deterministic SoS simulation model that is perhaps the most widely cited fading model in literature. Jakes' model begins by choosing the N sinusoidal components to be uniformly distributed in angle, i.e.,

$$\theta_n = \frac{2\pi n}{N}, \quad n = 1, 2, \dots, N. \quad (2.250)$$

By choosing $N/2$ to be an odd integer, the sum in (2.245) can be rearranged into the form

$$g(t) = \sqrt{\frac{1}{N}} \left\{ \sum_{n=1}^{N/2-1} \left[e^{-j(2\pi f_m t \cos(\theta_n) + \hat{\phi}_{-n})} + e^{j(2\pi f_m t \cos(\theta_n) + \hat{\phi}_n)} \right] + e^{-j(2\pi f_m t + \hat{\phi}_{-N})} + e^{j(2\pi f_m t + \hat{\phi}_N)} \right\}, \quad (2.251)$$

where the phase indices have been relabeled. 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 they progress from $+2\pi f_m \cos(2\pi/N)$ to $-2\pi f_m \cos(2\pi/N)$ in the second sum. Therefore, the frequencies in these terms overlap. To reduce the number of sinusoidal components used in the model, Jakes uses non-overlapping frequencies to write $g(t)$ as

$$g(t) = \sqrt{\frac{1}{N}} \left\{ \sqrt{2} \sum_{n=1}^M \left[e^{-j(2\pi f_m t \cos(\theta_n) + \hat{\phi}_{-n})} + e^{j(2\pi f_m t \cos(\theta_n) + \hat{\phi}_n)} \right] + e^{-j(2\pi f_m t + \hat{\phi}_{-N})} + e^{j(2\pi f_m t + \hat{\phi}_N)} \right\} \quad (2.252)$$

where

$$M = \frac{1}{2} \left(\frac{N}{2} - 1 \right) \quad (2.253)$$

and the factor $\sqrt{2}$ is included so that the total power remains unchanged. Note that (2.251) and (2.252) are not equal. In (2.251) all phases are independent. However, (2.252) implies that $\hat{\phi}_n = -\hat{\phi}_{-N/2+n}$ and $\hat{\phi}_{-n} = -\hat{\phi}_{N/2-n}$ for $n = 1, \dots, M$. If the phases are further constrained such that $\hat{\phi}_n = -\hat{\phi}_{-n}$, $n = 1, \dots, M$ and $\hat{\phi}_N = -\hat{\phi}_{-N}$, then (2.252) can be rewritten in the form $g(t) = g_I(t) + jg_Q(t)$, where

$$g_I(t) = \sqrt{\frac{2}{N}} \left(2 \sum_{n=1}^M \cos(\beta_n) \cos(2\pi f_n t) + \sqrt{2} \cos(\alpha) \cos(2\pi f_m t) \right) \quad (2.254)$$

$$g_Q(t) = \sqrt{\frac{2}{N}} \left(2 \sum_{n=1}^M \sin(\beta_n) \cos(2\pi f_n t) + \sqrt{2} \sin(\alpha) \cos(2\pi f_m t) \right), \quad (2.255)$$

with $\alpha = \hat{\phi}_N$ and $\beta_n = \hat{\phi}_n$.

Jakes chooses the phases α and β_n with the objective of making $\langle g_I^2(t) \rangle = \langle g_Q^2(t) \rangle$ and $\langle g_I(t)g_Q(t) \rangle = 0$, where $\langle \cdot \rangle$ denotes time averaging. From (2.254) and (2.255),

$$\begin{aligned} \langle g_I^2(t) \rangle &= \frac{2}{N} \left[2 \sum_{n=1}^M \cos^2(\beta_n) + \cos^2(\alpha) \right] \\ &= \frac{2}{N} \left[M + \cos^2(\alpha) + \sum_{n=1}^M \cos(2\beta_n) \right], \end{aligned} \quad (2.256)$$

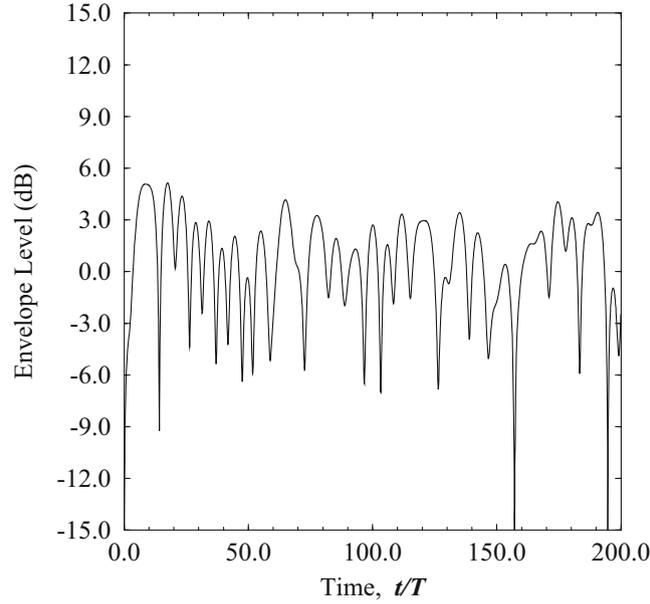


Fig. 2.38 Faded envelope generated by using Jakes' method with $M = 8$ oscillators; $f_m T = 0.1$

$$\begin{aligned} \langle g_Q^2(t) \rangle &= \frac{2}{N} \left[2 \sum_{n=1}^M \sin^2(\beta_n) + \sin^2(\alpha) \right] \\ &= \frac{2}{N} \left[M + \sin^2(\alpha) - \sum_{n=1}^M \cos(2\beta_n) \right], \end{aligned} \quad (2.257)$$

and

$$\langle g_I(t)g_Q(t) \rangle = \frac{2}{N} \left[2 \sum_{n=1}^M \sin(\beta_n) \cos(\beta_n) + \sin(\alpha) \cos(\alpha) \right]. \quad (2.258)$$

Choosing $\alpha = 0$ and $\beta_n = \pi n/M$ gives $\langle g_Q^2(t) \rangle = M/(2M + 1)$, $\langle g_I^2(t) \rangle = (M + 1)/(2M + 1)$ and $\langle g_I(t)g_Q(t) \rangle = 0$. Note that there is a small imbalance in the values of $\langle g_Q^2(t) \rangle$ and $\langle g_I^2(t) \rangle$. Finally, $\langle g_I^2(t) \rangle + \langle g_Q^2(t) \rangle = \Omega_p = 1$. The envelope power $\langle g_I^2(t) \rangle + \langle g_Q^2(t) \rangle$ can be easily scaled to any other desired value. A typical faded envelope, obtained by using Jakes' method with $N = 34$ ($M = 8$) is shown in Fig. 2.38.

The accuracy of Jakes' simulator can be evaluated by comparing the autocorrelation functions of the complex envelope with those of the 2-D isotropic scattering reference model. Suppose Clarke's method is modified by imposing the additional restrictions of even N and $\theta_n = 2\pi n/N$. The resulting simulation model is a statistical model, because the phases $\hat{\phi}_n$ are i.i.d. uniform random variable on $[-\pi, \pi)$. A faded envelope is obtained having the ensemble averaged autocorrelation function

$$\phi_{gg}(\tau) = \frac{1}{2N} \sum_{n=1}^N \cos \left(2\pi f_m \tau \cos \left(\frac{2\pi n}{N} \right) \right), \quad (2.259)$$

where ensemble averaging is performed over the random phases $\hat{\phi}_n$. Note that this autocorrelation tends to a Bessel function as $N \rightarrow \infty$. The normalized autocorrelation functions in (2.248) and (2.259) with $N = 8$ are plotted against the normalized time delay $f_m \tau$ in Fig. 2.39. Observe that Clark's model with $\theta_n = 2\pi n/N$ yields an autocorrelation function that deviates 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. 2.40 shows the autocorrelation function when the number of oscillators is doubled from $N = 8$ to $N = 16$.

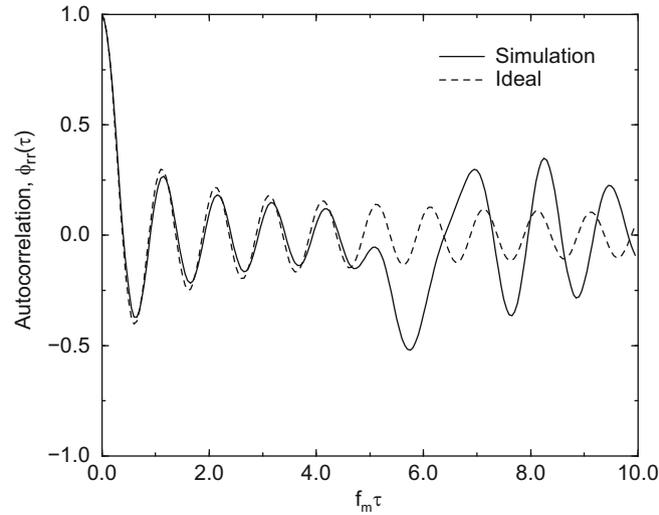


Fig. 2.39 Autocorrelation of in-phase and quadrature components obtained by using Clarke's method with $\theta_n = 2\pi n/N$ and $N = 8$ oscillators

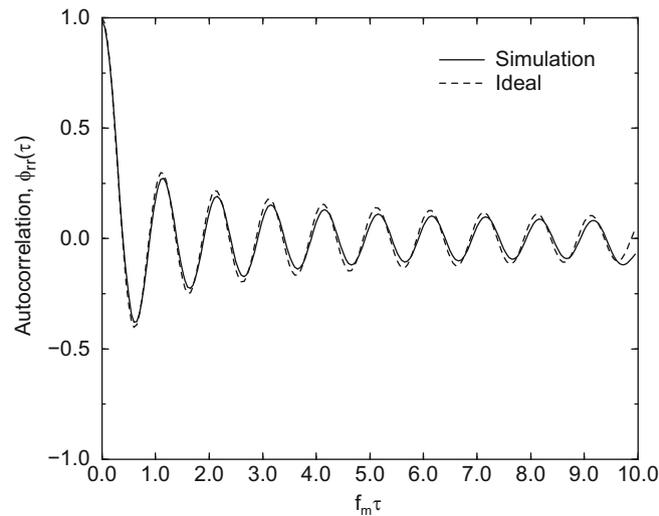


Fig. 2.40 Autocorrelation of in-phase and quadrature components obtained by using Clarke's method with $\theta_n = 2\pi n/N$ and $N = 16$ oscillators

The autocorrelation function in (2.259) was originally reported in Jakes' book [171] and it may be confused with the autocorrelation function of the faded envelope generated by (2.254) and (2.255). However, Jakes' method as described in (2.254) and (2.255) is a deterministic simulation model with no random parameters and, therefore, only the time averaged autocorrelation function exists. This time averaged autocorrelation function has been derived by Pop and Beaulieu as [264]

$$\hat{\phi}_{gg}(t, t + \tau) = \frac{1}{2N} (\cos(2\pi f_m \tau) + \cos(2\pi f_m (2t + \tau))) + \frac{1}{N} \sum_{n=1}^M (\cos(2\pi f_n \tau) + \cos(2\pi f_n (2t + \tau))). \quad (2.260)$$

From (2.260), it is apparent that the time averaged autocorrelation function obtained with the Jakes' method, $\hat{\phi}_{gg}(t, t + \tau)$, depends on the time variable t . Hence, Jakes' method yields a faded envelope that is not a stationary or even a wide-sense stationary random process. The root cause of this non-stationary behavior is the correlation that is introduced into the phases when using Jakes' method, i.e., the restrictions that $\hat{\phi}_n = -\hat{\phi}_{-N/2+n}$, $\hat{\phi}_{-n} = -\hat{\phi}_{N/2-n}$, $\hat{\phi}_n = -\hat{\phi}_{-n}$, $n = 1, \dots, M$, and $\hat{\phi}_N = -\hat{\phi}_{-N}$. Due to the non-stationary behavior of the original Jakes simulator, and also due to the fact that the Jakes' method yields only a single faded envelope, a number of subsequent sum-of-sinusoids models have been proposed in the literature.

2.5.2.3 Method of Exact Doppler Spreads

Pätzold et al. [261] proposed a deterministic simulation model, called the method of exact Doppler spreads (MEDS). The method is derived by starting with an integral representation for the zero-order Bessel function of the first kind:

$$J_0(x) = \frac{2}{\pi} \int_0^{\pi/2} \cos(x \sin \theta) d\theta \quad (2.261)$$

and replacing the integral by a series expansion as follows:

$$J_0(x) = \lim_{N_I \rightarrow \infty} \frac{2}{\pi} \sum_{n=1}^{N_I} \cos(x \sin \alpha_n) \Delta_\alpha, \quad (2.262)$$

where $\alpha_n \triangleq \pi(2n-1)/(4N_I)$ and $\Delta_\alpha \triangleq \pi/(2N_I)$. Hence,

$$\phi_{g_I g_I}(\tau) = \frac{1}{2} J_0(2\pi f_m \tau) = \lim_{N_I \rightarrow \infty} \frac{1}{2N_I} \sum_{n=1}^{N_I} \cos(2\pi f_{n,I} \tau) \quad (2.263)$$

where

$$f_{n,I} = f_m \sin \left[\frac{\pi}{2N_I} \left(n - \frac{1}{2} \right) \right]. \quad (2.264)$$

With finite N_I , the autocorrelation is

$$\phi_{g_I g_I}(\tau) = \frac{1}{2N_I} \sum_{n=1}^{N_I} \cos(2\pi f_{n,I} \tau). \quad (2.265)$$

Using the above result, the complex fading envelope, $g(t) = g_I(t) + jg_Q(t)$, is generated as

$$g_{I/Q}(t) = \sqrt{\frac{1}{N_{I/Q}}} \sum_{n=1}^{N_{I/Q}} \cos(2\pi f_{I/Q,n} t + \phi_{I/Q,n}). \quad (2.266)$$

where the phases $\phi_{I/Q,n}$ are chosen as arbitrary realizations of uniform random variables on the interval $[-\pi, \pi)$. Since the phases only need to be generated once, the MEDS model is an ergodic statistical (deterministic) model.

With the MEDS model, the autocorrelation functions of $g_I(t)$ and $g_Q(t)$ for infinite $N_{I/Q}$ match those of the 2-D isotropic scattering reference model. The quadrature components $g_I(t)$ and $g_Q(t)$ will also be uncorrelated if they do not share common frequencies. This condition can be achieved by choosing $N_Q = N_I + 1$.

2.5.2.4 Zheng & Xiao Statistical Model

Zheng and Xiao proposed a number of statistical simulation models [359, 360, 376, 377]. These models differ from one another in terms of the model parameters and, consequently, their statistical properties. Patel et al. [259] have shown that the model in [376] has the best overall performance, so the discussion here is limited to that model. The complex fading envelope, $g(t) = g_I(t) + jg_Q(t)$, is generated as

$$g_I(t) = \sqrt{\frac{1}{N}} \sum_{n=1}^N \cos[2\pi f_m t \cos(\theta_n) + \phi_{I,n}] \quad (2.267)$$

$$g_Q(t) = \sqrt{\frac{1}{N}} \sum_{n=1}^N \cos[2\pi f_m t \sin(\theta_n) + \phi_{Q,n}], \quad (2.268)$$

where the angles θ_n are

$$\theta_n = \frac{2\pi n - \pi + \theta}{4N}, \quad n = 1, 2, \dots, N, \quad (2.269)$$

and where θ , $\phi_{I,n}$, and $\phi_{Q,n}$ are all mutually independent uniform random variables on the interval $[-\pi, \pi)$.

The statistical correlation functions of the quadrature components have been computed by Zheng and Xiao [376] and match the desired functions in (2.246)–(2.248). The autocorrelation function of the squared-envelope is [259]

$$\phi_{\alpha^2\alpha^2}(\tau) = 1 + \frac{J_0(4\pi f_m \tau)}{4N} + \frac{1}{N^2} \sum_{n=1}^N \sum_{\substack{m=1 \\ n \neq m}}^N E\{\cos[2\pi f_m \tau \cos \theta_n] \cos[2\pi f_m \tau \cos \theta_m]\}. \quad (2.270)$$

Unfortunately, there is no closed form expressions for the squared-envelope autocorrelation function due to the presence of the double summation in (2.270). The 2-D isotropic scattering reference model requires that

$$\phi_{\alpha^2\alpha^2}(\tau) = 1 + J_0^2(2\pi f_m \tau) \quad (2.271)$$

Clearly, as N increases the second term in (2.270) diminishes with $1/N$, while the third term should approach $J_0^2(2\pi f_m \tau)$.

2.5.2.5 Modified Hoehner Model

The modified Hoehner model is a statistical simulation model. The complex faded envelope, $g(t) = g_I(t) + jg_Q(t)$, is generated as [260]

$$g_I(t) = \sqrt{\frac{1}{N}} \sum_{n=1}^{N_I} \cos(2\pi f_{I,n} t + \phi_{I,n}) \quad (2.272)$$

$$g_Q(t) = \sqrt{\frac{1}{N}} \sum_{m=1}^{N_Q} \cos(2\pi f_{Q,m} t + \phi_{Q,m}), \quad (2.273)$$

where

$$f_{I/Q,n/m} = f_m \sin\left(\frac{\pi}{2} u_{I/Q,n/m}\right). \quad (2.274)$$

The Doppler frequencies $f_{I/Q,n/m}$ for the I and Q components are determined by the random variables $u_{I/Q,n/m}$, where the $u_{I/Q,n/m}$ are uniform on $(0, 1]$ and are mutually independent for all n and m . The random phases $\phi_{I/Q,n/m}$ are uniform on $[-\pi, \pi)$, are mutually independent for all n and m , and are also independent of the $u_{I/Q,n/m}$. For convenience, the number of sinusoids in the quadrature components are usually set equal, i.e., $N_I = N_Q = N$. This model is based on Hoehner's model in [165], but differs by considering only the positive Doppler frequencies. Hence, it is referred to here as the modified Hoehner model.

Observe that the MEDS and modified Hoehner models are the same except for the arguments of the sine function in (2.264) and (2.274), respectively. The set of numbers $\{(n - 1/2)/N_i, i = 1, \dots, N_i\}$ are uniformly spaced on the interval $(0, 1]$ in the MEDS model, while the $u_{I/Q,n/m}$ are uniformly distributed on the interval $(0, 1]$ in the modified Hoehner model.

The statistical correlation functions for the modified Hoehner model match those of the 2-D isotropic scattering reference model. The squared-envelope autocorrelation function for the modified Hoehner model is [259]

$$\phi_{\alpha^2\alpha^2}(\tau) = 1 + \frac{1}{4N} J_0(4\pi f_m \tau) + \frac{N-1}{N} J_0^2(2\pi f_m \tau) \quad (2.275)$$

which differs from the reference model in (2.82) for finite N .

The time-averaged correlation functions of the modified Hoehner model can be derived as follows [259]:

$$\begin{aligned}\hat{\phi}_{g_I g_I}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T g_I(t) g_I(t + \tau) dt \\ &= \frac{1}{2N} \sum_{n=1}^N \cos(2\pi f_{I,n} \tau)\end{aligned}\quad (2.276)$$

$$\hat{\phi}_{g_Q g_Q}(\tau) = \frac{1}{2N} \sum_{n=1}^N \cos(2\pi f_{Q,n} \tau) \quad (2.277)$$

$$\hat{\phi}_{g_I g_Q}(\tau) = \hat{\phi}_{g_Q g_I}(\tau) = 0 \quad (2.278)$$

$$\begin{aligned}\hat{\phi}_{g g}(\tau) &= \frac{1}{2} \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T g^*(t) g(t + \tau) dt \\ &= \frac{1}{4N} \sum_{n=1}^N \cos(2\pi f_{I,n} \tau) + \cos(2\pi f_{Q,n} \tau).\end{aligned}\quad (2.279)$$

The variances of the time-averaged autocorrelation and cross-correlation of the quadrature components, and the variance of the autocorrelation of the complex envelope $g(t)$ are [259]

$$\text{Var}[\hat{\phi}_{g_I g_I}(\tau)] = \text{Var}[\hat{\phi}_{g_Q g_Q}(\tau)] = \frac{1 + J_0(4\pi f_m \tau) - 2J_0^2(2\pi f_m \tau)}{8N} \quad (2.280)$$

$$\text{Var}[\hat{\phi}_{g_I g_Q}(\tau)] = \text{Var}[\hat{\phi}_{g_Q g_I}(\tau)] = 0 \quad (2.281)$$

$$\text{Var}[\hat{\phi}_{g g}(\tau)] = \frac{1 + J_0(4\pi f_m \tau) - 2J_0^2(2\pi \tau)}{16N}. \quad (2.282)$$

2.5.3 Multiple Uncorrelated Faded Envelopes

In many cases it is necessary to generate multiple uncorrelated faded envelopes. Jakes [171] suggested a method to modify the method in Sect. 2.5.2.2 to generate up to M fading envelopes. However, the method yields faded envelopes that exhibit very large cross-correlations at some non-zero lags and, therefore, the method is not recommended. Dent et al. [93] suggested a modification to Jakes' approach that uses orthogonal Walsh-Hadamard codewords to decorrelate the multiple faded envelopes. However, the approach generates multiple faded envelopes that still have significant cross-correlation functions at large lags. Several simulation models to generate multiple faded envelopes are now discussed that have low cross-correlations.

2.5.3.1 Zheng and Xiao Statistical Model

The Zheng and Xiao statistical method can be easily extended to generate multiple faded envelopes. The k th complex faded envelope, $g_k(t) = g_{I,k}(t) + jg_{Q,k}(t)$, is generated as

$$g_{I,k}(t) = \sqrt{\frac{1}{N}} \sum_{n=1}^N \cos[2\pi f_m t \cos(\theta_{n,k}) + \phi_{I,n,k}] \quad (2.283)$$

$$g_{Q,k}(t) = \sqrt{\frac{1}{N}} \sum_{n=1}^N \cos[2\pi f_m t \sin(\theta_{n,k}) + \phi_{Q,n,k}], \quad (2.284)$$

where

$$\theta_{n,k} = \frac{2\pi n - \pi + \theta_k}{4N}, \quad n = 1, 2, \dots, N \quad (2.285)$$

and where θ_k , $\phi_{I,n,k}$, and $\phi_{Q,n,k}$ are all uniform on $[-\pi, \pi)$, and all values are mutually independent. The $g_k(t)$ are all uncorrelated.

2.5.3.2 Li & Huang Model

Li and Huang proposed a deterministic fading model that can generate multiple uncorrelated faded envelopes. They assume that P uncorrelated fading envelopes are required, each of which is composed of N sinusoids. The k th complex faded envelope, $g_k(t) = g_{I,k}(t) + jg_{Q,k}(t)$, is generated as [202]

$$g_{I,k}(t) = \sqrt{\frac{1}{N}} \sum_{n=0}^{N-1} \cos(2\pi f_m \cos \theta_{n,k} t + \phi_{I,n,k}) \quad (2.286)$$

$$g_{Q,k}(t) = \sqrt{\frac{1}{N}} \sum_{n=0}^{N-1} \sin(2\pi f_m \sin \theta_{n,k} t + \phi_{Q,n,k}). \quad (2.287)$$

The phases $\phi_{I,n,k}$ and $\phi_{Q,n,k}$ arbitrary realizations of independent random variables uniform on $[-\pi, \pi)$, and

$$\theta_{n,k} = \frac{2\pi n}{N} + \frac{2\pi k}{PN} + \theta_{00}, \quad n = 0, \dots, N, k = 0, \dots, P-1, \quad (2.288)$$

where θ_{00} is an initial arrival angle chosen to satisfy $0 < \theta_{00} < 2\pi/PN$ and $\theta_{00} \neq \pi/PN$. Although the Li & Huang model generates uncorrelated faded envelopes, it fails to satisfy Eqs. (2.29) and (2.82) of the reference model [371].

2.5.3.3 Zajić & Stüber Deterministic Model

With the Zajić & Stüber deterministic model the k th faded envelope, $g_k(t) = g_{I,k}(t) + jg_{Q,k}(t)$, is generated as [372]

$$g_{I,k}(t) = \sqrt{\frac{2}{N}} \sum_{n=0}^M a_n \cos(2\pi f_m t \cos(\theta_{n,k}) + \phi_{n,k}) \quad (2.289)$$

$$g_{Q,k}(t) = \sqrt{\frac{2}{N}} \sum_{n=0}^M b_n \sin(2\pi f_m t \cos(\theta_{n,k}) + \phi_{n,k}), \quad (2.290)$$

where

$$M = \frac{1}{2} \left(\frac{N}{2} - 1 \right), \quad (2.291)$$

and

$$a_n = \begin{cases} 2 \cos(\beta_n), & n = 1, \dots, M \\ \sqrt{2} \cos(\beta_n), & n = 0 \end{cases} \quad (2.292)$$

$$b_n = \begin{cases} 2 \sin(\beta_n), & n = 1, \dots, M \\ \sqrt{2} \sin(\beta_n), & n = 0 \end{cases} \quad (2.293)$$

$$\beta_n = \frac{\pi n}{M}, \quad n = 0, \dots, M \quad (2.294)$$

$$\theta_{n,k} = \frac{2\pi n}{N} + \frac{2\pi k}{PN} + \theta_{0,0}, \quad n = 0, \dots, M, \quad k = 0, \dots, P-1. \quad (2.295)$$

This method will generate P faded envelopes, where $1 \leq P \leq M$.

The angle $\theta_{n,k}$ is chosen so that the arrival angles associated with the k th faded envelope are obtained by rotating the arrival angles of the $(k-1)$ th faded envelope by $2\pi/PN$. This ensures an asymmetrical arrangement of arrival angles, which minimizes the correlation between the multiple faded envelopes. The initial arrival angle, $\theta_{0,0}$, can be optimized to minimize the correlation between the quadrature components of each faded envelope. This results in the choice $\theta_{0,0} = 0.2\pi/PN$, a value optimized by experimentation. Finally, the phases $\phi_{n,k}$ are chosen as arbitrary realizations of uniform random variables on the interval $[0, 2\pi)$.

For the Zajić & Stüber deterministic model, the auto- and cross-correlation functions of the quadrature components, the auto- and cross-correlation functions of the multiple faded envelopes, and the squared-envelope autocorrelation are, respectively, [372]

$$\begin{aligned} \lim_{N \rightarrow \infty} \phi_{g_{I,k}g_{I,k}}(\tau) &= \lim_{N \rightarrow \infty} \frac{2}{N} \sum_{n=0}^M \frac{a_n^2}{2} \cos(2\pi f_m \tau \cos(\theta_{n,k})) \\ &= \frac{1}{2} J_0(2\pi f_m \tau) + \frac{1}{2} J_4(2\pi f_m \tau), \end{aligned} \quad (2.296)$$

$$\begin{aligned} \lim_{N \rightarrow \infty} \phi_{g_{Q,k}g_{Q,k}}(\tau) &= \lim_{N \rightarrow \infty} \frac{2}{N} \sum_{n=0}^M \frac{b_n^2}{2} \cos(2\pi f_m \tau \sin(\theta_{n,k})) \\ &= \frac{1}{2} J_0(2\pi f_m \tau) - \frac{1}{2} J_4(2\pi f_m \tau), \end{aligned} \quad (2.297)$$

$$\phi_{g_{I,k}g_{Q,k}}(\tau) = \phi_{g_{Q,k}g_{I,k}}(\tau) = \phi_{g_{k \neq l}g_{l \neq k}}(\tau) = 0, \quad (2.298)$$

$$\begin{aligned} \lim_{N \rightarrow \infty} \phi_{g_{k}g_{k}}(\tau) &= \lim_{N \rightarrow \infty} \frac{2}{N} \sum_{n=0}^M \frac{a_n^2}{2} \cos(2\pi f_m \tau \cos(\theta_{n,k})) \\ &\quad + \lim_{N \rightarrow \infty} \frac{2}{N} \sum_{n=0}^M \frac{b_n^2}{2} \cos(2\pi f_m \tau \sin(\theta_{n,k})) \\ &= \frac{1}{2} J_0(2\pi f_m \tau), \end{aligned} \quad (2.299)$$

$$\begin{aligned} \phi_{\alpha_k^2 \alpha_k^2}(\tau) &= \frac{2}{N^2} \sum_{n=0}^M a_n^4 + \frac{2}{N^2} \sum_{n=0}^M b_n^4 + \frac{1}{2} \phi_{g_{I,k}g_{I,k}}^2(\tau) \\ &\quad + \frac{1}{2} \phi_{g_{Q,k}g_{Q,k}}^2(\tau) + \phi_{g_{I,k}g_{Q,k}}^2(\tau) \\ &= \frac{1}{4} J_0^2(\omega_m \tau), \end{aligned} \quad (2.300)$$

where $J_0(\cdot)$ is the zero-order Bessel function of the first kind and $J_4(\cdot)$ is the fourth-order Bessel function of the first kind.

Figures 2.41 and 2.42 confirm that, for $M = 8$, the auto- and cross-correlations of the quadrature components and the auto- and cross-correlation of the multiple faded envelopes approach values given by (2.296)–(2.299), respectively. The model satisfies (2.247) and (2.248) of the reference model. However, the auto-correlations of the quadrature components and the auto-correlation of the squared-envelope are close to, but do not perfectly match, (2.246) and (2.249).

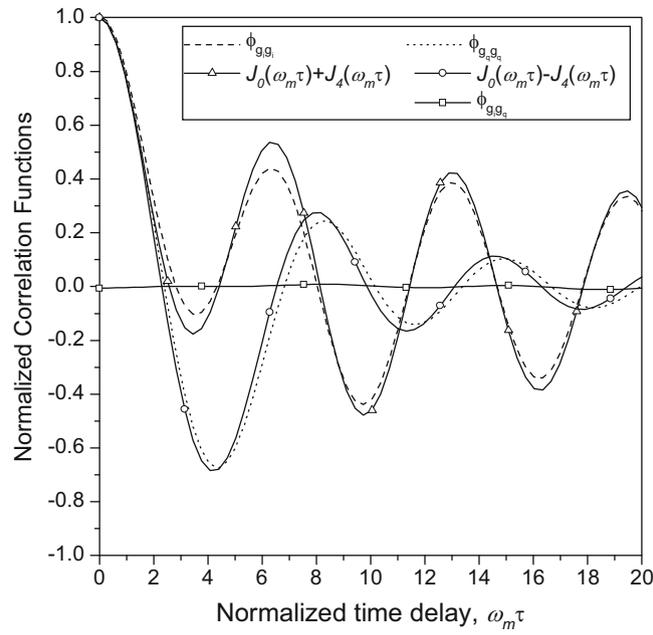


Fig. 2.41 Theoretical and simulated auto-correlation functions and the cross-correlation function of the in-phase and quadrature components of the Zajić and Stüber deterministic model

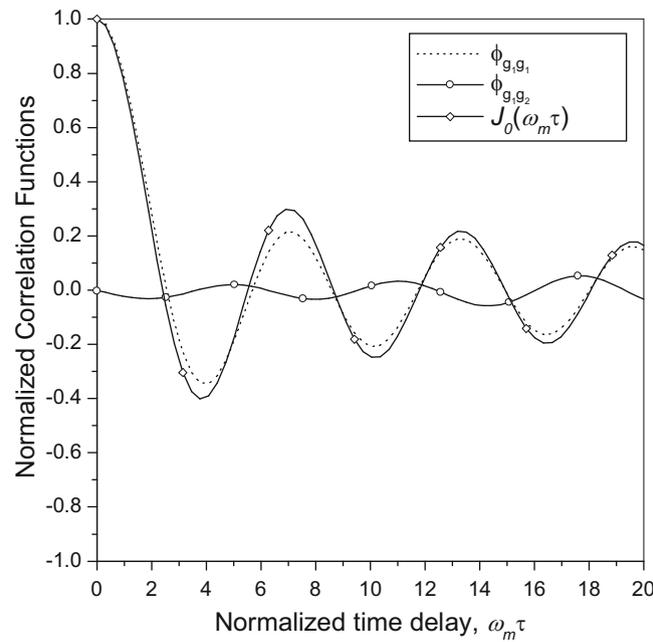


Fig. 2.42 Theoretical and simulated auto-correlation functions and the cross-correlation function of the first and the second complex envelope of the Zajić and Stüber deterministic model

2.5.3.4 Zajić & Stüber Statistical Model

The Zajić & Stüber deterministic model can be modified to better match the statistical properties of the reference model by introducing randomness into the model. To do so, the k th faded envelope, $g_k(t) = g_{I,k}(t) + jg_{Q,k}(t)$, is generated as

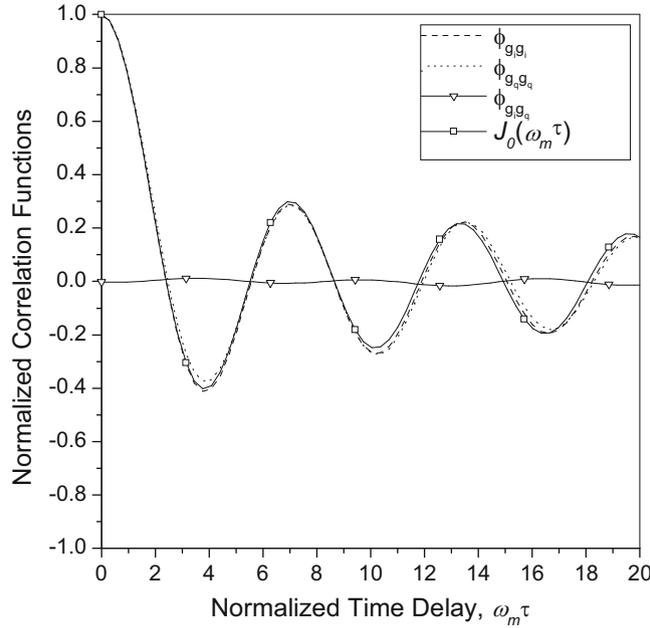


Fig. 2.43 Theoretical and simulated autocorrelation functions and the cross-correlation function of the quadrature components of the Zajić and Stüber statistical model

$$g_{I,k}(t) = \sqrt{\frac{2}{N}} \sum_{n=0}^M \cos(\beta_{n,k}) \cos(2\pi f_m t \cos(\theta_{n,k}) + \phi_{n,k}) \quad (2.301)$$

$$g_{Q,k}(t) = \sqrt{\frac{2}{N}} \sum_{n=0}^M \sin(\beta_{n,k}) \sin(2\pi f_m t \cos(\theta_{n,k}) + \phi_{n,k}), \quad (2.302)$$

where $M = N/4$,

$$\theta_{n,k} = \frac{2\pi n}{N} + \frac{2\pi k}{PN} + \frac{\alpha - \pi}{N}, \quad n = 1, \dots, M, \quad k = 0, \dots, P-1, \quad (2.303)$$

and $\phi_{n,k}$, $\beta_{n,k}$, and α are independent uniform random variables on the interval $[-\pi, \pi)$. The parameter $\theta_{n,k}$ is chosen in the following manner: the arrival angles of the k th faded envelope are obtained by rotating the arrival angles of the $(k-1)$ th faded envelope by $\pi/2MN$. Randomness in the Doppler frequencies is introduced by random variable α . Again, this method will generate P faded envelopes, where $1 \leq P \leq M$.

Figures 2.43 and 2.44 show that, for $M = P = 8$ and $N_{\text{stat}} = 30$ simulation trials, the auto- and cross-correlations of the quadrature components, and the auto- and cross-correlations of the complex faded envelopes approach those of the reference model in (2.246) and (2.248). Although not shown, the squared-envelope autocorrelation also approaches that of the reference model in (2.249).

2.5.3.5 SoS Simulation Model Comparisons

This section compares the various sum-of-sinusoids simulation models for fixed-to-mobile channels. A more detailed comparison is available in [259, 371]. In all simulations, a normalized sampling period $f_m T_s = 0.05$ and $M = P = 8$ is used. However, for the MEDS model $N_I = 8$ and $N_Q = 9$ are used to obtain uncorrelated quadrature components in the faded envelope. Note that in the Li & Huang, MEDS and Zajić & Stüber deterministic models, the random phases associated with the sinusoidal components are computed before the actual simulation starts, because an ergodic statistical (“deterministic”)

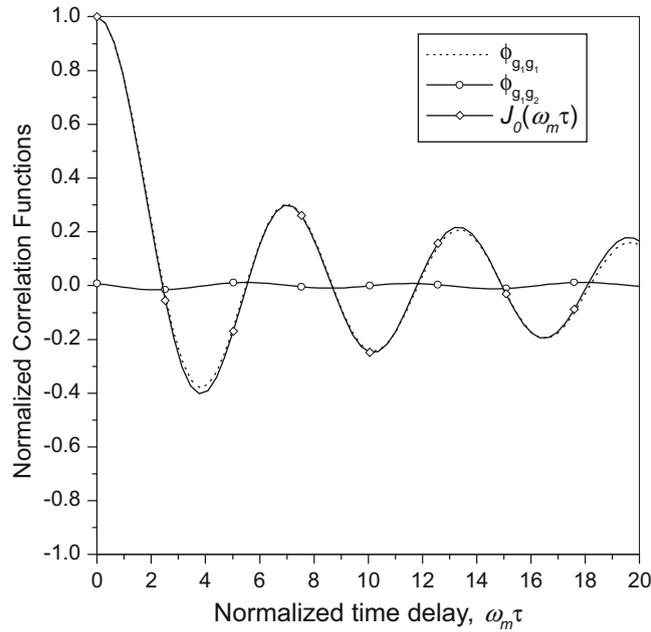


Fig. 2.44 Theoretical and simulated autocorrelation functions and the cross-correlation function of the first and the second complex envelopes of the Zajić and Stüber statistical model

simulator needs only one simulation trial. During the simulations, all parameters are kept constant to provide simulation results that are always the same, i.e., deterministic.

In the Zajić & Stüber deterministic model, the following set of uniformly distributed random phases (in radians) are used: $\phi_{n,0} = [4.0387, 1.7624, 2.7844, 1.5590, 0.9523, 1.2972, 5.7420, 3.6592, 4.3548]$ and $\phi_{n,1} = [5.3798, 3.0556, 2.1528, 2.6296, 0.7457, 3.2572, 6.1027, 2.0670, 2.1304]$. The same set of phases is used in MEDS model, for $g_Q(t)$, while for $g_I(t)$ the following set is used $\phi_{I,n,0} = [2.2107, 5.3033, 2.4634, 1.0679, 2.2818, 4.6113, 0.7513, 0.7383]$ and $\phi_{I,n,1} = [3.2627, 4.7036, 0.5824, 2.1097, 4.6264, 5.4790, 0.9391, 0.2017]$. Li & Huang's model uses the same set of phases as the MEDS model for $g_I(t)$, while for $g_Q(t)$ the following set is used: $\phi_{Q,n,0} = [2.6372, 4.7339, 4.9865, 5.7784, 5.3059, 2.3099, 3.8994, 4.5933]$, and $\phi_{Q,n,1} = [4.1175, 2.4616, 3.9403, 4.3911, 2.4948, 2.5981, 4.115, 5.2610]$.

Figure 2.45 compares the cross-correlation functions of the quadrature components obtained with the various simulation models. Since the Zheng & Xiao model is a statistical model, the plotted correlation functions have been averaged over $N_{\text{stat}} = 50$ trials. Figure 2.45 shows that the Zajić & Stüber deterministic model yields a lower cross-correlation between the I and Q components of the faded envelope. Figure 2.46 compares the cross-correlation functions of two faded envelopes. Again, since the Zheng & Xiao model is a statistical model, the plotted cross-correlation function has been averaged over $N_{\text{stat}} = 50$ trials. The Li & Huang, Zheng & Xiao, and Zajić & Stüber deterministic models all yield a low cross-correlation between different faded envelopes.

Figures 2.47 and 2.48 compare the cross-correlation functions of the quadrature components and the cross-correlation functions of two faded envelopes, respectively, for the various statistical models. For the Zheng & Xiao model averaging is done over $N_{\text{stat}} = 100$ trials, while for the Zajić & Stüber statistical model averaging is done over $N_{\text{stat}} = 30$ and $N_{\text{stat}} = 50$ trials. From Fig. 2.47, the Zajić & Stüber statistical model with $N_{\text{stat}} = 30$ performs similar when compared to the Zheng & Xiao model with $N_{\text{stat}} = 100$. Increasing the number of trials in the Zajić & Stüber statistical model to $N_{\text{stat}} = 50$ yields a significantly lower cross-correlation between the quadrature components of the complex faded envelope. Furthermore, with $N_{\text{stat}} = 50$ trials, the Zajić & Stüber statistical model achieves a larger de-correlation between different complex envelopes than the Zheng & Xiao model with $N_{\text{stat}} = 100$ trials. Figures 2.47, 2.48, and 2.49 show that with the Zajić & Stüber statistical model, adequate statistics can be achieved with $N_{\text{stat}} = 30$ trials.

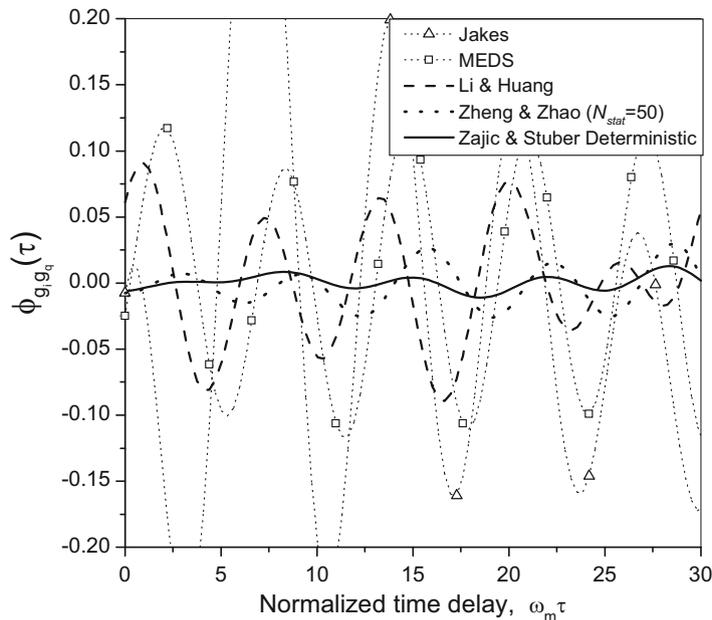


Fig. 2.45 Normalized cross-correlation function between the quadrature components of the faded envelope for various simulation models

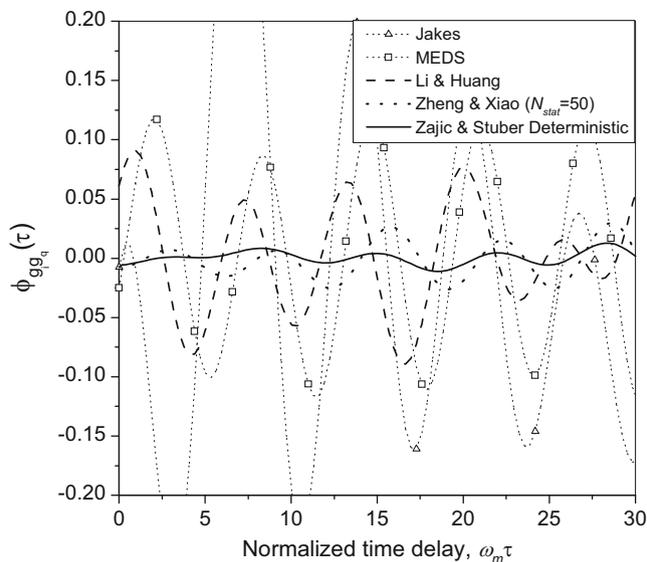


Fig. 2.46 Normalized cross-correlation function between the first and second faded envelopes for various simulation models

2.5.4 Wide-Band Simulation Models

Wide-band channels can be modeled by a tapped delay line with irregularly spaced tap delays. Each channel tap is the superposition of a large number of scattered plane waves that arrive with approximately the same delay and, therefore, the channel taps will undergo fading. The wide-band channel has the time-variant impulse response

$$g(t, \tau) = \sum_{i=1}^{\ell} g_i(t) \delta(\tau - \tau_i), \tag{2.304}$$

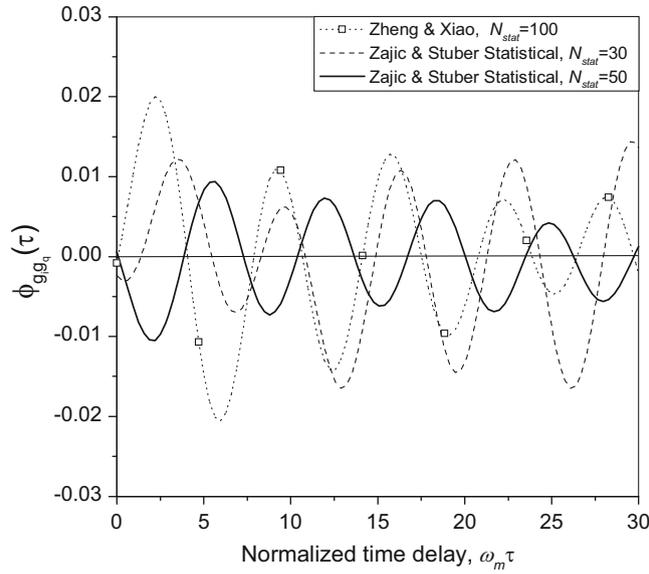


Fig. 2.47 Normalized cross-correlation function of the in-phase and quadrature components for various statistical simulation models

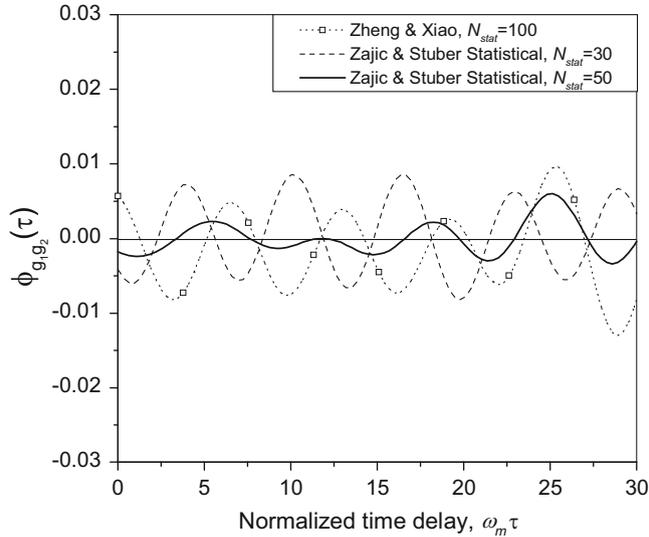


Fig. 2.48 Normalized cross-correlation function of the first and the second complex envelope for various statistical simulation models

where ℓ is the number of channel taps, and the $\{g_i(t)\}$ and $\{\tau_i\}$ are the complex gains and path delays associated with the channel taps. The corresponding time-variant transfer function is

$$T(f, t) = \sum_{i=1}^{\ell} g_i(t) e^{j2\pi f \tau_i}. \quad (2.305)$$

Usually, the tap delays are multiples of some very small delay τ , such that $\tau_i = K_i \tau$, $i = 1, \dots, \ell$, which yields the sparse tapped delay line channel model shown in Fig. 2.50, sometimes called the “ τ -spaced” model. Many of the tap coefficients in the tapped delay line are zero, reflecting the fact that no energy is received at these delays. Also, the K_i should be chosen to be relatively prime so as to prevent any periodicity in the channel. Assuming a WSSUS channel, the autocorrelation function of the time-variant channel impulse response in (2.304) is

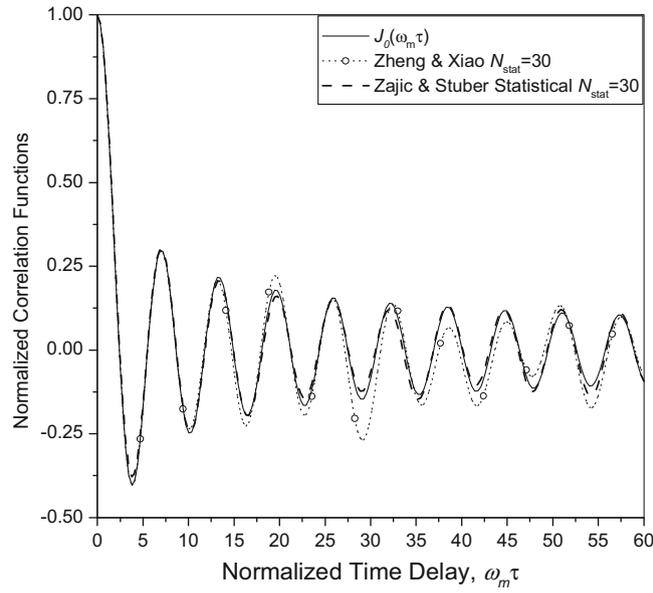


Fig. 2.49 Theoretical and simulated normalized auto-correlation functions for the various statistical simulation models

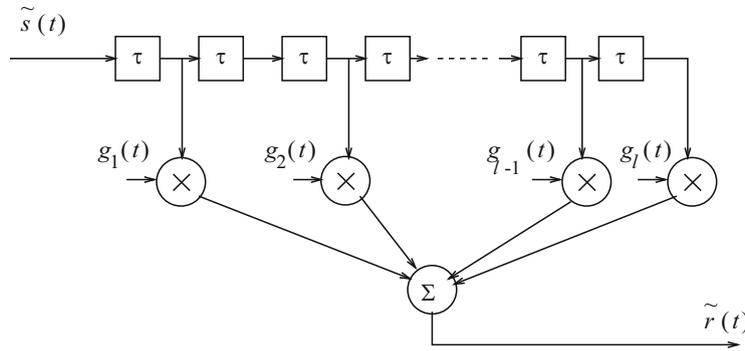


Fig. 2.50 Sparse tapped delay line model for wide-band multipath-fading channels

$$\phi_g(t, s; \tau, \eta) = E[g(t, \tau)g^*(s, \eta)] \quad (2.306)$$

$$= \sum_{i=1}^{\ell} \phi_{g_i g_i}(\Delta t) \delta(\tau - \tau_i) \delta(\eta - \tau) \quad (2.307)$$

It follows that the channel correlation function is

$$\psi_g(\Delta t; \tau) = \sum_{i=1}^{\ell} \phi_{g_i g_i}(\Delta t) \delta(\tau - \tau_i) \quad (2.308)$$

and the power-delay profile is

$$\psi_g(\tau) = \psi_g(0; \tau) = \sum_{i=1}^{\ell} \Omega_i \delta(\tau - \tau_i), \quad (2.309)$$

where $\Omega_i = \phi_{g_i g_i}(0)$ is the envelope power of the i th channel tap, and the total envelope power is

$$\Omega_p = \sum_{k=1}^{\ell} \Omega_k. \quad (2.310)$$

Hence, the channel can be described by the power profile

$$\mathbf{\Omega} = (\Omega_0, \Omega_1, \dots, \Omega_\ell). \quad (2.311)$$

and the delay profile

$$\mathbf{\tau} = (\tau_1, \tau_2, \dots, \tau_\ell). \quad (2.312)$$

Taking the Fourier transform of (2.308) with respect to the τ variable yields the spaced-time spaced-frequency correlation function

$$\phi_T(\Delta f; \Delta t) = \sum_{i=1}^{\ell} \phi_{g_i g_i}(\Delta t) e^{-j2\pi \Delta f \tau_i} \quad (2.313)$$

Sometimes the channel taps are assumed to have the same time correlation function, such that

$$\phi_{g_i g_i}(\Delta t) = \Omega_i \phi_t(\Delta t), \quad i = 1, \dots, \ell. \quad (2.314)$$

For example, if each tap is characterized by 2-D isotropic scattering, then $\phi_t(\Delta t) = \frac{1}{2} J_0(2\pi f_m \Delta t)$. In this case the spaced-time spaced-frequency correlation function has the separable form

$$\phi_T(\Delta f; \Delta t) = \phi_t(\Delta t) \phi_f(\Delta f), \quad (2.315)$$

where

$$\phi_f(\Delta f) = \sum_{i=1}^{\ell} \Omega_i e^{-j2\pi \Delta f \tau_i}. \quad (2.316)$$

2.5.4.1 COST 207 Models

The COST 207 models were developed for the GSM cellular system. Four different Doppler spectra, $S_{gg}(f)$ are specified in the COST 207 models [79]. First define the function

$$G(A, f_1, f_2) = A \exp \left\{ -\frac{(f - f_1)^2}{2f_2^2} \right\} \quad (2.317)$$

which has the shape of a Gaussian probability density function with mean f_1 and variance f_2^2 . COST 207 specifies the following Doppler spectra:

a) CLASS is the classical Doppler spectrum, and is used for excess path delays not exceeding 500 ns ($\tau_i \leq 500$ ns);

$$\text{(CLASS)} \quad S_{gg}(f) = \frac{A}{\sqrt{1 - (f/f_m)^2}}, \quad |f| \leq f_m \quad (2.318)$$

Multipath components arriving with short path delays are those that experience local scattering around the MS. In the COST 207 models, the local scattering is modeled as 2-D isotropic scattering and the MS is assumed to have an isotropic antenna.

b) GAUS1 is the sum of two Gaussian functions, and is used for excess path delays ranging from 500 ns to 2 μ s; (500 ns $\leq \tau_i \leq 2$ μ s)

$$\text{(GAUS1)} \quad S_{gg}(f) = G(A, -0.8f_m, 0.05f_m) + G(A_1, 0.4f_m, 0.1f_m), \quad (2.319)$$

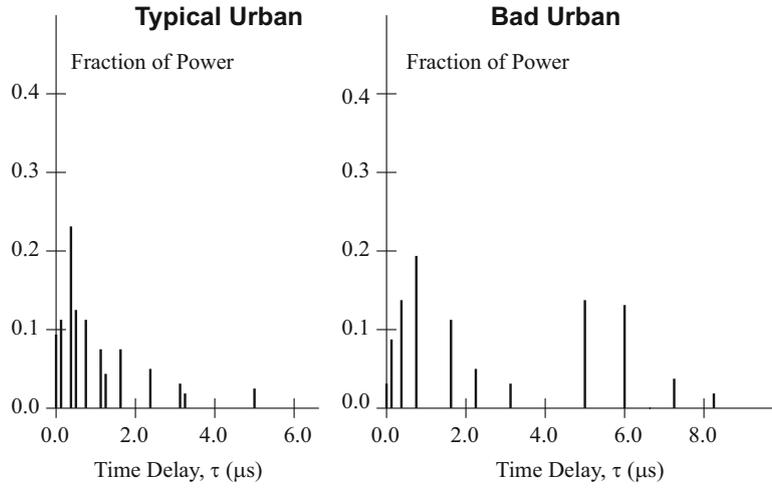


Fig. 2.51 COST 207 typical urban (TU) and bad urban (BU) power-delay profiles

where A_1 is 10 dB below A . The GAUS1 Doppler spectra implies two clusters of scattering objects. The larger cluster is located behind the MS such that the mean angle-of-arrival is $\theta = \cos^{-1}(-0.8) = \pm 143^\circ$. A smaller cluster is located in front of the MS such that the mean angle-of-arrival is $\theta = \cos^{-1}(0.4) = \pm 66^\circ$.

c) GAUS2 is the sum of two Gaussian functions, used for excess path delays exceeding $2 \mu\text{s}$; ($\tau_i > 500 \text{ ns}$)

$$\text{(GAUS2)} \quad S_{gg}(f) = G(B, 0.7f_m, 0.1f_m) + G(B_1, -0.4f_m, 0.15f_m), \quad (2.320)$$

where B_1 is 15 dB below B . Like GAUS1, GAUS2 implies two clusters of scattering objects such that the mean angles-of-arrival are $\theta = \cos^{-1}(0.7) = \pm 45^\circ$ and $\theta = \cos^{-1}(-0.4) = \pm 113^\circ$.

d) RICE is a combination of 2-D isotropic scattering and a line-of-sight component;

$$\text{(RICE)} \quad S_{gg}(f) = \frac{0.41}{2\pi f_m \sqrt{1 - (f/f_m)^2}} + 0.91\delta(f - 0.7f_m), \quad |f| \leq f_m \quad (2.321)$$

The RICE scattering environment is depicted in Fig. 2.7, where $\theta_0 = \cos^{-1}(0.7) = 45^\circ$.

Several power-delay profiles have been defined in the COST 207 study for different propagation environments [79]. Typical urban (TU) (non-hilly) and bad urban (BU) (hilly) discrete power-delay profiles are shown in Fig. 2.51 and Table 2.7 of Appendix 2A. In these figures and tables, the fractional power sums to unity, i.e., the total envelope power Ω_p is normalized to unity. Also notice that the discrete tap delays in Table 2.7 are chosen to avoid a regular spacing between taps so as to avoid any periodicities in the time-variant transfer function. Sometimes it is desirable to use a smaller number of taps to reduce the computational requirements of computer simulations. Figure 2.52 and Table 2.8 of Appendix 2A show the 6-ray reduced typical urban and reduced bad urban channel. Also provided in Appendix 2A are discrete power-delay profiles for rural (non-hilly) areas (RA) in Table 2.9, typical hilly terrain (HT) in Table 2.10, and reduced hilly terrain (HT) in Table 2.11.

2.5.4.2 COST 259 Models

The COST 207 models were developed for the GSM system having a channel bandwidth of 200 kHz. However, the COST 207 models are not appropriate for wideband CDMA (WCDMA) systems having 5, 10, and 20 MHz channel bandwidths. Similar to the COST 207 models, typical realizations have been developed by 3GPP for the COST 259 models [113]. These are tabulated in Tables 2.12, 2.13, and 2.14 of Appendix 2B for typical urban (TU), rural area (RA), and hilly terrain (HT). Notice that the tap delays in the 3GPP realizations of the COST 259 models have a resolution of $0.001 \mu\text{s}$ as compared to the $0.1 \mu\text{s}$ resolution used in the COST 207 models. Moreover, a large number of paths (20) are included to ensure that the frequency domain correlation properties are realistic, which is important for wide-band systems.

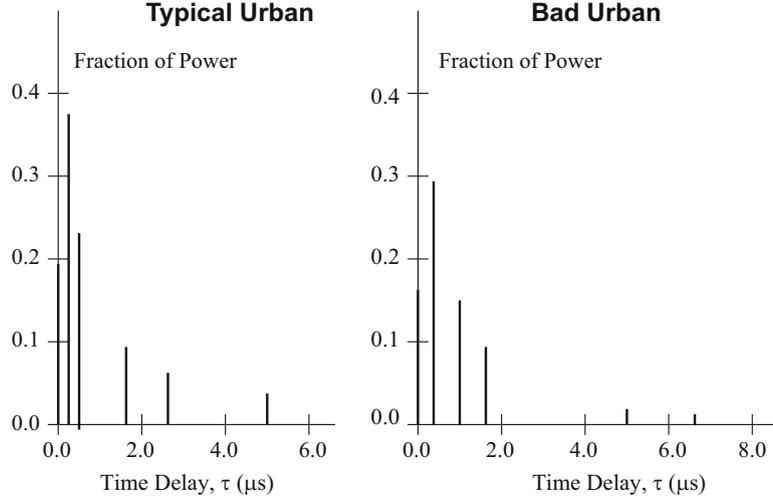


Fig. 2.52 COST 207 reduced typical urban (TU) and bad urban (BU) power-delay profiles

2.5.4.3 ITU Models

Several models have been developed by the International Telecommunications Union (ITU) for indoor office, outdoor to indoor and pedestrian, and vehicular—high antenna [276]. These models are summarized in Tables 2.15 and 2.16 of Appendix 2C.

2.5.5 Mobile-to-Mobile Simulation Models

An important distinction between F-to-M cellular land mobile radio channels and M-to-M channels arises due to the scattering and mobility at both ends of the link. In contrast to cellular F-to-M channels, it is natural for both the transmitter and the receiver to have low elevation antennas in M-to-M applications. This results in local scattering around both the transmitter and receiver antennas which led to the 2-D isotropic scattering assumption in the Akki and Haber reference model [14].

2.5.5.1 Akki & Haber Model

An obvious M-to-M channel simulation model follows from (2.146), where N sinusoidal components are used with randomly generated angles of departure and arrival, and random phases. This model is called the Akki & Haber (A & H) model, since they originally proposed the model in (2.146). Although the ensemble averaged statistical correlations of this model, given in (2.147), do not depend on N , the time averaged correlations functions [denoted by $\hat{\phi}(\cdot)$] do depend on N [259]. For example, the time average auto-correlation of the complex envelope is

$$\begin{aligned}\hat{\phi}_{gg}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \frac{1}{2} g^*(t) g(t + \tau) dt \\ &= \frac{1}{2N} \sum_{n=1}^N e^{j(2\pi f_m^T \cos(\alpha_n^T) \tau + 2\pi f_m^R \cos(\alpha_n^R) \tau)}.\end{aligned}\quad (2.322)$$

The time average correlations are random due to the random angles of departure and arrival and, hence, they will vary with each simulation trial. The variance $\text{Var}[\hat{\phi}(\cdot)] = E[|\hat{\phi}(\cdot) - \lim_{N \rightarrow \infty} \phi(\cdot)|^2]$ provides a measure of the usefulness of the model with finite N . A lower variance means that a smaller number of simulation trials are needed to achieve the desired statistical properties to a given degree of accuracy and, hence, the corresponding simulation model is better. The variance of

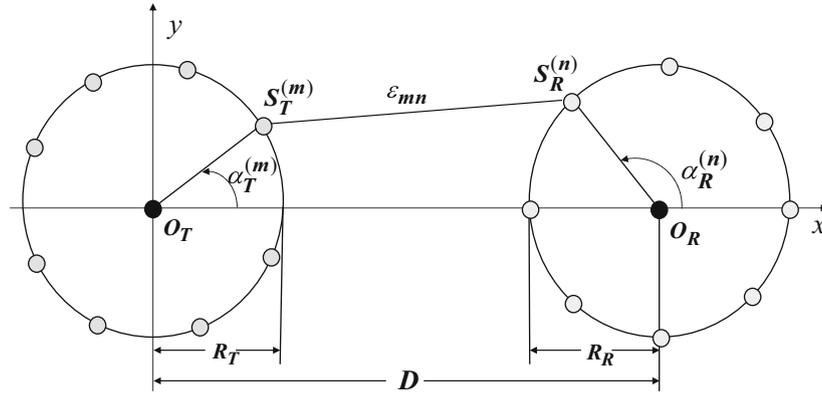


Fig. 2.53 Double-ring model for M-to-M radio propagation with isotropic scattering at the transmitter and receiver

the time averaged auto-correlation and the cross-correlation of the quadrature components of the complex envelope are as follows [259]:

$$\begin{aligned} \text{Var}[\hat{\phi}_{g_I g_I}(\tau)] &= \text{Var}[\hat{\phi}_{g_Q g_Q}(\tau)] & (2.323) \\ &= \frac{1 + J_0(4\pi f_m^T \tau) J_0(4\pi a f_m^T \tau) - 2J_0^2(2\pi f_m^T \tau) J_0^2(2\pi a f_m^T \tau)}{8N} \end{aligned}$$

$$\text{Var}[\hat{\phi}_{g_I g_Q}(\tau)] = \text{Var}[\hat{\phi}_{g_Q g_I}(\tau)] = \frac{1 - J_0(4\pi f_m^T \tau) J_0(4\pi a f_m^T \tau)}{8N} \quad (2.324)$$

$$\text{Var}[\hat{\phi}_{g_g}(\tau)] = \frac{1 - J_0^2(2\pi f_m^T \tau) J_0^2(2\pi a f_m^T \tau)}{4N}. \quad (2.325)$$

2.5.5.2 Double Ring Simulation Models

To develop other M-to-M channel simulation models, a double-ring concept is applied that defines two rings of isotropic scatterers, one placed around the transmitter and another placed around the receiver as shown in Fig. 2.53. Assuming omnidirectional antennas at both ends, the waves from the transmitter antenna first arrive at the scatterers located on the transmitter ring. Considering these fixed scatterers as “virtual base stations (VBS),” the communication link from each VBS to the receiver is treated like an F-to-M link. The signals from each VBS arrive at the receiver antenna uniformly from all directions in the horizontal plane due to isotropic scatterers located on the receiver end ring. It should be noted here that the double-ring model has been often used in various forms, for example in [50] and the references therein for the study and simulation of the spatial correlations and capacity in multiple-input multiple-output (MIMO) systems, where multiple antennas are used at both the transmitter and receiver. Here the double-ring model is applied to M-to-M channel modeling. The use of a double-ring model has a strong rationale. First and foremost, it is a mathematically convenient reference model. Second, considering the lack of detailed and standardized M-to-M channel models, the model provides a generic scenario with isotropic scattering at both ends of the communication link. Finally, empirical Doppler spectrum measurements for vehicle-to-vehicle communication links [9, 222] more or less match those predicted by the double-ring model.

Using the double-ring model in Fig. 2.53, the complex faded envelope can be written as

$$g(t) = \sqrt{\frac{2}{NM}} \sum_{m=1}^N \sum_{n=1}^M e^{-j2\pi \epsilon_{mn} / \lambda_c} e^{j2\pi t \{f_m^T \cos(\alpha_T^{(m)}) + f_m^R \cos(\alpha_R^{(n)})\} + j\phi_{n,m}}, \quad (2.326)$$

where the index “ m ” refers to the paths traveling from the transmitter to the N scatterers located on the transmitter end ring, the index “ n ” refers to the paths traveling from the M scatterers on the receiver end ring to the receiver. The angle $\alpha_T^{(m)}$ is the random angle of departure and $\alpha_R^{(n)}$ is the random angle of arrival of the $\{m, n\}$ th propagation path measured with respect to the x -axis, respectively. The phases $\phi_{n,m}$ are uniformly distributed on $[-\pi, \pi)$ and are independent for all pairs $\{n, m\}$. Note that the single summation in (2.146) is replaced with a double summation, because each plane wave on its way from the transmitter to the receiver is double bounced. The temporal correlation characteristics remain the same as those of the model

in (2.146), because each path will undergo a Doppler shift due to the motion of both the transmitter and receiver. Finally, the temporal correlation properties do not depend on the distance ϵ_{mn} in Fig. 2.53. Although not discussed here, the spatial correlation properties will depend on ϵ_{mn} and, through ϵ_{mn} , on the scattering radii R_T and R_R , and the distance D .

2.5.5.3 Patel & Stüber Deterministic (P & S Det.) Model

An ergodic statistical (deterministic) simulation model is now suggested. By choosing only the phases to be random variables, the statistical correlation properties of this model will converge to those of the reference model in a single simulation trial. The complex faded envelope, $g(t) = g_I(t) + jg_Q(t)$, is generated as

$$g_I(t) = \sqrt{\frac{1}{N_I M_I}} \sum_{m=1}^{M_I} \sum_{n=1}^{N_I} \cos(2\pi f_m^T \cos(\alpha_{I,n}^T) t + 2\pi f_m^R \cos(\alpha_{I,m}^R) t + \phi_{I,n,m}), \quad (2.327)$$

$$g_Q(t) = \sqrt{\frac{1}{N_Q M_Q}} \sum_{m=1}^{M_Q} \sum_{n=1}^{N_Q} \cos(2\pi f_m^T \cos(\alpha_{Q,n}^T) t + 2\pi f_m^R \cos(\alpha_{Q,m}^R) t + \phi_{Q,n,m}), \quad (2.328)$$

where $\phi_{I,n,m}$ and $\phi_{Q,n,m}$ are all independent random phases uniformly distributed on $[-\pi, \pi)$. The n th angle of departure is equal to $\alpha_{I/Q,n}^T = \pi(n-0.5)/(2N_{I/Q})$, for $n = 1, 2, \dots, N_{I/Q}$. The m th angle of arrival is equal to $\alpha_{I/Q,m}^R = \pi(m-0.5)/(M_{I/Q})$, for $m = 1, 2, \dots, M_{I/Q}$.

There are two remarks about this model. First, the angles $\alpha_{I/Q,n}^T$ are essentially the same as those in the MEDS (method of exact Doppler spreads) model [260], while the angles $\alpha_{I/Q,m}^R$ are chosen to reproduce desired statistical properties for M-to-M channels. Second, to make the time average correlation functions deterministic and independent of the random phases $\phi_{I/Q,n,m}$, the frequencies $f_m^T \cos(\alpha_{I,n}^T) + f_m^R \cos(\alpha_{I,m}^R)$ in $g_I(t)$ and $f_m^T \cos(\alpha_{Q,n}^T) + f_m^R \cos(\alpha_{Q,m}^R)$ in $g_Q(t)$ must all be distinct and, moreover, they must also be distinct among themselves for all pairs (n, m) and (k, l) . Although it is difficult to establish a general rule to meet this criterion, choosing $N_I = M_I$ and $N_Q = M_Q = N_I + 1$ yields distinct Doppler frequencies for practical ranges of N_I varying from 5 to 60 and for different Doppler frequency ratios, i.e., different values of $a = f_m^R/f_m^T$. This rule is similar to the one used in the MEDS model [260]. Under these assumptions, it can be shown that the time average correlation functions are

$$\hat{\phi}_{g_I g_I}(\tau) = \frac{1}{2N_I^2} \sum_{n,m=1}^{N_I, N_I} \cos(2\pi f_m^T \tau \cos(\alpha_{I,n}^T) + 2\pi f_m^R \tau \cos(\alpha_{I,m}^R)) \quad (2.329)$$

$$\hat{\phi}_{g_Q g_Q}(\tau) = \frac{1}{2(N_I + 1)^2} \sum_{n,m=1}^{(N_I+1), (N_I+1)} \cos(2\pi f_m^T \tau \cos(\alpha_{Q,n}^T) + 2\pi f_m^R \tau \cos(\alpha_{Q,m}^R)) \quad (2.330)$$

$$\hat{\phi}_{g_I g_Q}(\tau) = \hat{\phi}_{g_Q g_I}(\tau) = 0. \quad (2.331)$$

The P & S Det. model has the disadvantage that the time averaged correlation functions match those of the reference model only for a small range of normalized time delays ($0 \leq f_m^T \tau \leq 5$). To improve upon the statistical properties, two alternative statistical SoS models are suggested. By allowing all three parameter sets (amplitudes, phases, and Doppler frequencies) to be random variables for each simulation trial, the statistical properties of these models will vary for each simulation trial, but they will converge to those of the reference model when averaged over a sufficient number of simulation trials.

2.5.5.4 Patel & Stüber Statistical (P & S Stat.) Model

The P & S Stat. model modifies the model proposed by Zheng et al. [376] for F-to-M cellular land mobile radio channels. The complex faded envelope, $g(t) = g_I(t) + jg_Q(t)$, is generated as

$$g_I(t) = \sqrt{\frac{1}{N_0 M}} \sum_{n=1}^{N_0} \sum_{m=1}^M \cos(2\pi f_m^T \cos(\alpha_n^T) t + 2\pi f_m^R \cos(\alpha_m^R) t + \phi_{I,n,m}) \quad (2.332)$$

$$g_Q(t) = \sqrt{\frac{1}{N_0 M}} \sum_{n=1}^{N_0} \sum_{m=1}^M \cos(2\pi f_m^T \sin(\alpha_n^T)t + 2\pi f_m^R \cos(\alpha_m^R)t + \phi_{Q,n,m}), \quad (2.333)$$

where $\phi_{I,n,m}$ and $\phi_{Q,n,m}$ are all independent random phases uniformly distributed on $[-\pi, \pi)$. The P & S Stat. model assumes N_0 scatterers located on the transmitter ring and M scatterers located on the receiver ring. The n th angle of departure is $\alpha_n^T = (2\pi n - \pi + \theta)/(4N_0)$, where θ is an independent uniform random variable on $[-\pi, \pi)$. The m th angle of arrival is equal to $\alpha_m^R = (2\pi m - \pi + \psi)/(2M)$, where ψ is an independent uniform random variable on $[-\pi, \pi)$.

The ensemble averaged statistical properties of the P & S Stat. model match those of the reference model in (2.147) [259]. The time averaged correlation functions can be derived as:

$$\hat{\phi}_{g_I g_I}(\tau) = \frac{1}{2NM} \sum_{n,m=1}^{N,M} \cos(2\pi f_m^T \cos(\alpha_n^T)\tau + 2\pi f_m^R \cos(\alpha_m^R)\tau) \quad (2.334)$$

$$\hat{\phi}_{g_Q g_Q}(\tau) = \frac{1}{2NM} \sum_{n,m=1}^{N,M} \cos(2\pi f_m^T \sin(\alpha_n^T)\tau + 2\pi f_m^R \cos(\alpha_m^R)\tau) \quad (2.335)$$

$$\hat{\phi}_{g_I g_Q}(\tau) = \hat{\phi}_{g_Q g_I}(\tau) = 0 \quad (2.336)$$

$$\hat{\phi}_{g_g}(\tau) = \frac{1}{2NM} \sum_{n,m=1}^{N,M} [\cos(2\pi f_m^T \cos(\alpha_n^T)\tau + 2\pi f_m^R \cos(\alpha_m^R)\tau) + \cos(2\pi f_m^T \sin(\alpha_n^T)\tau + 2\pi f_m^R \cos(\alpha_m^R)\tau)]. \quad (2.337)$$

The P & S Stat. model requires a fairly large number of simulation trials (at least 50) to obtain adequate statistical properties. Being a statistical model, its time average correlations are random and depend on the random Doppler frequencies. However, for the same complexity, the model performs better than the A & H Model in terms of the variance of these correlation functions [259]. These variances are not given here, since they do not exist in closed form.

2.5.5.5 Zajić & Stüber Statistical (Z & S Stat.) Model

The P & S models have difficulty in producing time averaged auto- and cross-correlation functions that match those of the reference model. Here another statistical model is suggested where orthogonal functions are chosen for the quadrature components of the complex faded envelope. This model is also able to generate multiple uncorrelated complex envelopes, something that the other models cannot do.

The following function is considered as the k th complex faded envelope:

$$g_k(t) = \sqrt{\frac{1}{NM}} \sum_{n=1}^N \sum_{m=1}^M e^{j(2\pi f_m^T \cos(\alpha_{k,n}^T)t + 2\pi f_m^R \cos(\alpha_{k,m}^R)t + \phi_{k,n,m})}, \quad (2.338)$$

where $f_m^T, f_m^R, \alpha_{k,n}^T, \alpha_{k,m}^R$, and $\phi_{k,n,m}$ are the maximum Doppler frequencies, the random angle of departure, the random angle of arrival, and the random phase, respectively. It is assumed that P independent complex faded envelopes are required ($k = 0, \dots, P-1$) each consisting of NM sinusoidal components.

The number of sinusoidal components needed for simulation can be reduced by choosing $N_0 = N/4$ to be an integer, by taking into account shifts of the angles $\alpha_{k,n}^T$ and $\alpha_{k,m}^R$, and by splitting the sum in (2.338) into four terms, viz.,

$$g_k(t) = \sqrt{\frac{1}{NM}} \sum_{m=1}^M e^{j2\pi f_m^R t \cos(\alpha_{k,m}^R)} \left[\sum_{n=1}^{N_0} \left\{ e^{j(2\pi f_m^T t \cos(\alpha_{k,n}^T) + \phi_{k,n,m})} + e^{j(2\pi f_m^T t \cos(\alpha_{k,n}^T + \pi/2) + \phi_{k,n,m} + \pi/2)} \right\} \right] \\ + \sqrt{\frac{1}{NM}} \sum_{m=1}^M e^{j2\pi f_m^R t \cos(\alpha_{k,m}^R)} \left[\sum_{n=1}^{N_0} \left\{ e^{j(2\pi f_m^T t \cos(\alpha_{k,n}^T + \pi) + \phi_{k,n,m} + \pi)} + e^{j(2\pi f_m^T t \cos(\alpha_{k,n}^T + 3\pi/2) + \phi_{k,n,m} + 3\pi/2)} \right\} \right]. \quad (2.339)$$

Equation (2.339) simplifies as

$$\begin{aligned}
g_k(t) &= \sqrt{\frac{1}{N_0 M}} \sum_{n=1}^{N_0} \sum_{m=1}^M \cos(2\pi f_m^R t \cos(\alpha_{k,m}^R)) \cos(2\pi f_m^T t \cos(\alpha_{k,n}^T) + \phi_{k,n,m}) \\
&\quad + j \sqrt{\frac{1}{N_0 M}} \sum_{n=1}^{N_0} \sum_{m=1}^M \sin(2\pi f_m^R t \cos(\alpha_{k,m}^R)) \sin(2\pi f_m^T t \sin(\alpha_{k,n}^T) + \phi_{k,n,m}). \tag{2.340}
\end{aligned}$$

Based on $g_k(t)$ in (2.340), the Z & S statistical simulation model can be defined. The k th complex faded envelope, $g_k(t) = g_{I,k}(t) + jg_{Q,k}(t)$, is generated as

$$g_{I,k}(t) = \frac{1}{\sqrt{N_0 M}} \sum_{n=1}^{N_0} \sum_{m=1}^M \cos(2\pi f_m^R \cos(\alpha_{k,m}^R) t) \cos(2\pi f_m^T \cos(\alpha_{k,n}^T) t + \phi_{k,n,m}) \tag{2.341}$$

$$g_{Q,k}(t) = \frac{1}{\sqrt{N_0 M}} \sum_{n=1}^{N_0} \sum_{m=1}^M \sin(2\pi f_m^R \cos(\alpha_{k,m}^R) t) \sin(2\pi f_m^T \sin(\alpha_{k,n}^T) t + \phi_{k,n,m}). \tag{2.342}$$

It is assumed that P independent complex envelopes are desired ($k = 0, \dots, P-1$), each having MN_0 sinusoidal terms in the I and Q components. The angles of departures and the angles of arrivals are chosen as follows:

$$\alpha_{k,n}^T = \frac{2\pi n}{4N_0} + \frac{2\pi k}{4PN_0} + \frac{\theta - \pi}{4N_0}, \tag{2.343}$$

$$\alpha_{k,m}^R = 0.5 \left(\frac{2\pi m}{M} + \frac{2\pi k}{PM} + \frac{\psi - \pi}{M} \right), \tag{2.344}$$

for $n = 1, \dots, N_0$, $m = 1, \dots, M$, $k = 0, \dots, P-1$. The angles of departure and the angles of arrival in the k th complex faded envelope are obtained by rotating the angles of departure and the angles of arrival in the $(k-1)$ th complex envelope by $(2\pi)/(4PN_0)$ and $(2\pi)/(2PM)$, respectively. The parameters $\phi_{k,n,m}$, θ , and ψ are independent uniform random variables on the interval $[-\pi, \pi)$.

The ensemble averaged statistical correlation functions of the Z & S Stat. model match those of the reference model [371]. For brevity, only the derivation of the auto-correlation function of the in-phase component is presented. Other properties can be derived in an analogous fashion. The auto-correlation function of the in-phase component of the k th complex faded envelope is

$$\phi_{g_{k,I}g_{k,I}}(\tau) = E[g_{k,I}(t)g_{k,I}(t+\tau)] \tag{2.345}$$

$$\begin{aligned}
&= \frac{4}{N_0 M} \sum_{n,m=1}^{N_0, M} \sum_{p,r=1}^{N_0, M} E[\cos(2\pi f_m^R(t+\tau) \cos(\alpha_{k,m}^R)) \\
&\quad \times \cos(2\pi f_m^R t \cos(\alpha_{k,p}^R)) \cos(2\pi f_m^T(t+\tau) \cos(\alpha_{k,n}^T) + \phi_{k,n,m}) \\
&\quad \times \cos(2\pi f_m^T t \cos(\alpha_{k,r}^T) + \phi_{k,p,r})] \\
&= \frac{1}{N_0 M} \sum_{n=1}^{N_0} \sum_{m=1}^M E[\cos(2\pi f_m^R \tau \cos(\alpha_{k,m}^R)) \cos(2\pi f_m^T \tau \cos(\alpha_{k,n}^T))] \\
&= \frac{1}{M} \sum_{m=1}^M \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos\left(2\pi f_m^R \tau \cos\left(\frac{2\pi m}{2M} + \frac{2\pi k}{2PM} + \frac{\psi - \pi}{2M}\right)\right) d\psi \\
&\quad \times \frac{1}{N_0} \sum_{n=1}^{N_0} \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos\left(2\pi f_m^T \tau \cos\left(\frac{2\pi n}{4N_0} + \frac{2\pi k}{4PN_0} + \frac{\theta - \pi}{4N_0}\right)\right) d\theta \tag{2.346}
\end{aligned}$$

Table 2.2 Mean square error, maximum deviation and variations, for various mobile-to-mobile channel simulators

Simulator	P & S Det.	P & S Stat.	P & S Stat.	Z & S Stat.	Z & S Stat.	Z & S Stat.
		$N_{\text{stat}} = 1$	$N_{\text{stat}} = 50$	$N_{\text{stat}} = 1$	$N_{\text{stat}} = 30$	$N_{\text{stat}} = 50$
$\text{mse}(\phi_{g_I g_Q})$	$1.8 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$3.16 \cdot 10^{-5}$	$8.81 \cdot 10^{-4}$	$4.39 \cdot 10^{-5}$	$2.05 \cdot 10^{-5}$
$\text{Max}(\phi_{g_I g_Q})$	$9.61 \cdot 10^{-2}$	$5.91 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$	$6.84 \cdot 10^{-2}$	$1.28 \cdot 10^{-2}$	$7.0 \cdot 10^{-3}$
$\text{Var}(\phi_{g_I g_Q})$	$2.74 \cdot 10^{-6}$	$4.59 \cdot 10^{-6}$	$6.67 \cdot 10^{-9}$	$3.11 \cdot 10^{-6}$	$5.69 \cdot 10^{-9}$	$2.51 \cdot 10^{-9}$
$\text{mse}(\phi_{g_1 g_2})$	$5.8 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$1.31 \cdot 10^{-5}$	$7.02 \cdot 10^{-4}$	$3.59 \cdot 10^{-5}$	$1.26 \cdot 10^{-5}$
$\text{Max}(\phi_{g_1 g_2})$	$1.59 \cdot 10^{-1}$	$6.34 \cdot 10^{-2}$	$0.63 \cdot 10^{-2}$	$6.57 \cdot 10^{-2}$	$9.6 \cdot 10^{-3}$	$5.6 \cdot 10^{-3}$
$\text{Var}(\phi_{g_1 g_2})$	$1.24 \cdot 10^{-4}$	$1.77 \cdot 10^{-6}$	$7.38 \cdot 10^{-10}$	$1.22 \cdot 10^{-6}$	$3.0 \cdot 10^{-9}$	$4.39 \cdot 10^{-10}$

As in [360], the derivation can be completed by replacing the variables of integration, θ and ψ , with $\gamma_{k,n} = (2\pi n)/(4N_0) + (2\pi k)/(4PN_0) + (\theta - \pi)/(4N_0)$ and $\delta_{k,m} = (2\pi m)/(2M) + (2\pi k)/(2PM) + (\psi - \pi)/(2M)$, respectively. Finally,

$$\lim_{N_0, M \rightarrow \infty} \phi_{g_{k,l} g_{k,l}}(\tau) = J_0(2\pi f_m^T \tau) J_0(2\pi f_m^R \tau). \quad (2.347)$$

The performance of the various models is now compared. All simulations use a normalized sampling rate of $f_m T_s = .01$ ($f_m = f_m^T = f_m^R$ are the maximum Doppler frequencies) and $M = N_0 = P = 8$. For the P & S Det. model, $N_I = M_I = 8$ and $N_Q = M_Q = 9$ are used to obtain a complex envelope with uncorrelated quadrature components. Using these parameters, the mean square error (mse) and maximum deviations (Max) from the theoretical value (zero) have been calculated for the normalized cross-correlations of the I and Q components, and for the normalized cross-correlations of the first and the second faded envelopes. The results are shown in Table 2.2. Note that different simulation trials yield slightly different simulation results. To estimate these differences, the variances are computed by averaging over 1000 simulation trials. The variances of the normalized cross-correlations of the I and Q components, and the variances of the normalized cross-correlations of the first and the second faded envelopes are also shown in Table 2.2. From Table 2.2, the Z & S Stat. model with $N_{\text{stat}} = 1$ has cross-correlations similar to the P & S Det. model and the P & S Stat. model with $N_{\text{stat}} = 1$. The Z & S Stat. model with $N_{\text{stat}} = 30$ performs similar to the P & S Stat. model with $N_{\text{stat}} = 50$ and significantly better than the P & S Det. model. Increasing the number of simulation trials to $N_{\text{stat}} = 50$ yields a lower cross-correlation between the I and Q components of the complex faded envelope.

Figure 2.54 shows that, for $N_0 = M = P = 8$ and 30 simulation trials, the auto- and cross-correlations of the complex faded envelopes produced by the Z & S Stat. model approach those of the reference model.

Figures 2.55 and 2.56 compare the variance of the time averaged auto-correlation functions of the quadrature components averaged over 1000 simulation trials. For the statistical models, the variance is defined as $\text{Var}[\hat{\phi}(\cdot)] = E[|\hat{\phi}(\cdot) - \lim_{N \rightarrow \infty} \phi(\cdot)|^2]$, where $\hat{\phi}(\cdot)$ denotes the time averaged correlation functions and $\phi(\cdot)$ denotes the statistical correlation functions. For the P & S Det. model, the relevant quantity is the squared error $|\hat{\phi}(\cdot) - \phi(\cdot)|^2$. The variance for the A & H model is obtained using (2.323). The variance provides a measure of the usefulness of the model in simulating the desired channel with a finite N . A lower variance means that a smaller number of simulation trials are needed to achieve the desired statistical properties and, hence, the corresponding model is better. Since the A & H model does not exploit the symmetry of the double-ring model, as the other models do, for fair comparison $N = 4N_0 \times 2M = 512$ sinusoids are used when using the A & H model. From Fig. 2.55, it can be concluded that the Z & S Stat. model with one simulation trial has a variance similar to the P & S Det. model and a lower variance than the P & S Stat. model with one simulation trial. A single simulation trial is chosen for this comparison because the P & S Det. model is an ergodic statistical simulation model that requires only a single simulation trial. In any case, Fig. 2.55 shows that all three models do not perform as well as the A & H model for a single simulation trial. Figure 2.56 shows that increasing the number of simulation trials to 30 in the P & S and Z & S Stat. models yields a significantly lower variance of the auto-correlation function of the quadrature components. Figure 2.56 shows that the Z & S Stat. model with 30 simulation trials outperforms the P & S Stat. model with 30 simulation trials, and outperforms P & S Det. model and the A & H model each with one simulation trial.

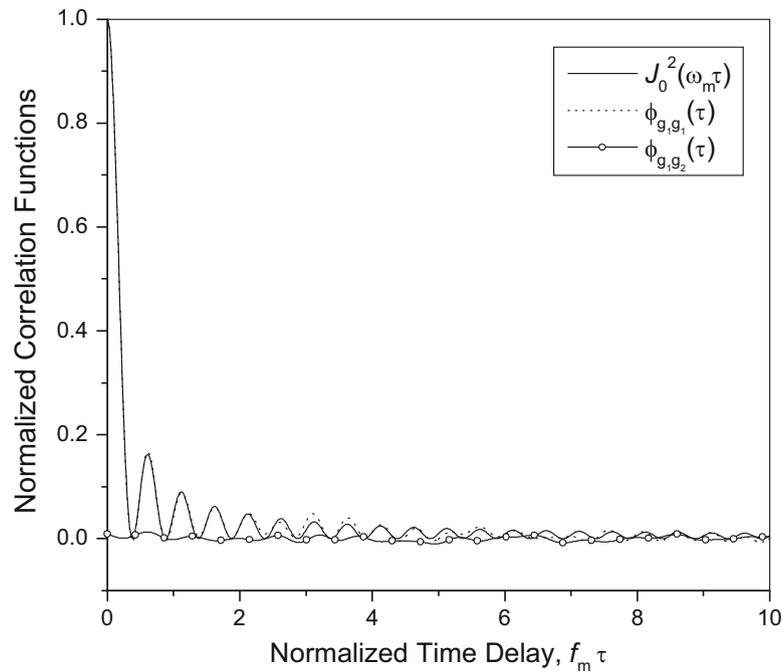


Fig. 2.54 Theoretical and simulated auto-correlation functions and cross-correlation function of the first and the second faded envelopes produced by the Z & S Stat. model with 30 simulation trials

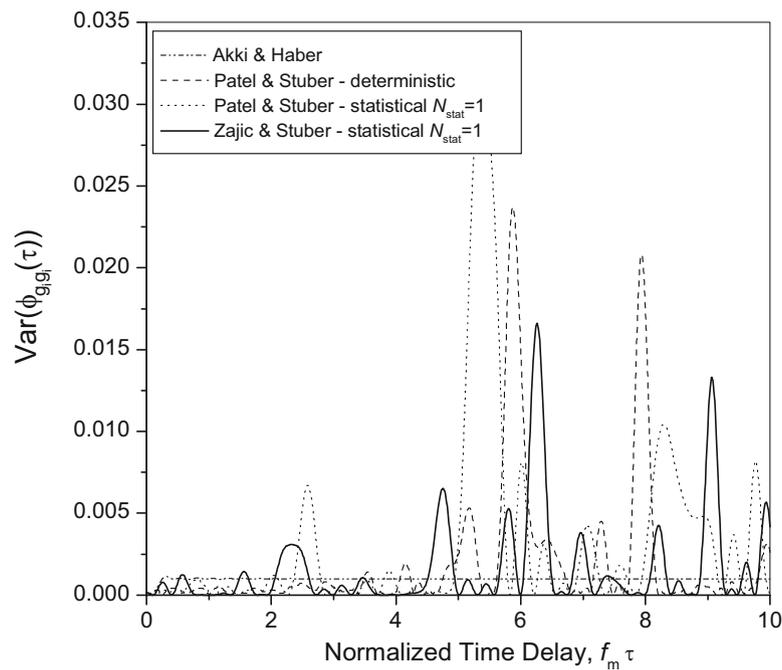


Fig. 2.55 Variance of the auto-correlation function of the quadrature components with a single simulation trial ($N_{\text{stat}} = 1$)

2.5.6 Symbol-Spaced Models

The baseband representation of a typical digital communication system consists of the concatenation of an up-sampler, a discrete-time transmit filter or pulse shaping filter, digital-to-analog converter (DAC), waveform channel, analog-to-digital converter (ADC), and discrete-time receiver filter or matched filter, and down-sampler as shown in Fig. 2.57. The channel

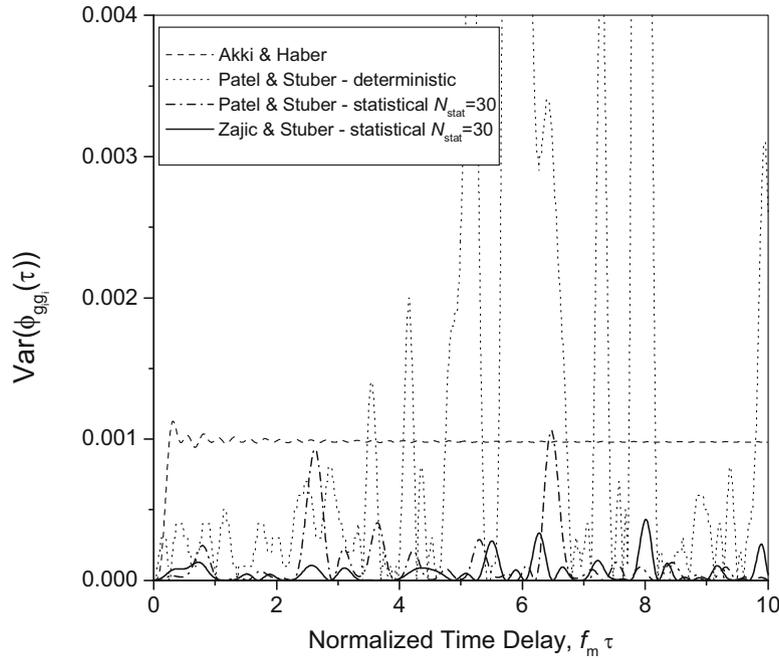


Fig. 2.56 Variance of the auto-correlation function of the quadrature components with 30 simulation trials ($N_{\text{stat}} = 30$)

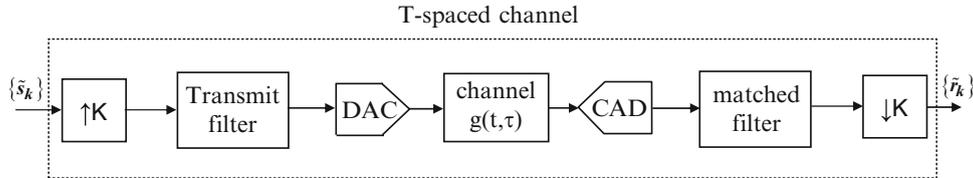


Fig. 2.57 Baseband model for a typical digital communication system

$g(t, \tau)$ is assumed to have the structure in (2.304) or Fig. 2.50. Data symbols are input to the up-sampler, one every T seconds, where T is the baud period. Usually, the bandwidth of the transmitted signal W will exceed the Nyquist frequency $1/2T$; for example, when root-raised cosine pulse shaping is used. Therefore, up-sampling is required at the transmitter so that the DAC operates with sample period T/K , where K is the up-sampling factor. At the receiver, the ADC also operates with a sampling period T/K . The up-sampling factor K must be chosen to at least satisfy the sampling theorem at the transmit side, i.e., $K/T \geq 2W$, but often K is made larger to facilitate timing synchronization in the receiver and to accommodate the Doppler spreading that is introduced by the channel. Once the correct sample timing phase is determined, the sample sequence at the output of the receiver filter can be down-sampled for further processing. For the purpose of illustration, down-sampling by factor of K is assumed, so that symbol-spaced samples are taken at the output of the receiver filter.⁴

From the above discussion, it is apparent that the overall channel from the input to the transmitter filter to the output of the receiver filter can be modeled as a finite impulse response (FIR) filter as shown in Fig. 2.58, where the $\{g_n^T[k]\}$ are the tap gains. While it is true that the channel taps, $\{g_i(t)\}$ in the underlying waveform channel in Fig. 2.50 are uncorrelated for WSSUS channels, the same cannot be said of the taps $\{g_n^T[k]\}$ in the FIR filter of Fig. 2.58. The tap correlations in the symbol-spaced model often leads to analytical intractability when evaluating the theoretical performance of digital communication systems that operate on these channels. This difficulty is often overcome by assuming that the taps $\{g_n^T[k]\}$ are uncorrelated [90, 107, 140, 192, 206, 301, 320], when in fact they are not. However, when the same systems are evaluated by software simulation, such modeling simplifications are unnecessary and in fact undesirable. A method is now described for generating the tap coefficients $\{g_n^T[k]\}$ with the proper cross-correlations in the case where linear modulation schemes are used. The procedure can be readily extended to generate the required tap coefficients when fractional sampling is used, i.e., if $K > 1$.

⁴In practice, $T/2$ -spaced samples at the output of the receiver filter are often used for further processing, such as equalization.

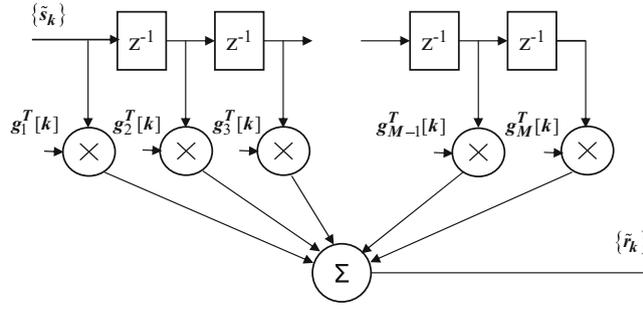


Fig. 2.58 Symbol-spaced tapped delay line model for wide-band multipath-fading channels

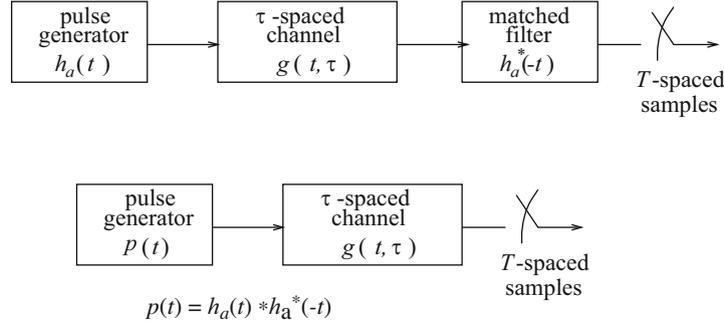


Fig. 2.59 Method for generating correlated tap coefficients in a T -spaced channel model

Consider the arrangement shown in Fig. 2.59, where the equivalent analog representation of the transmit filter, $h_a(t)$, and receiver matched filter $h_a^*(-t)$ are assumed. Pulses that are shaped by the transmit filter $h_a(t)$ are transmitted through the channel $g(t, \tau)$ and the receiver matched filter $h_a^*(-t)$. The output of the receiver matched filter is sampled at symbol-spaced intervals. Since the filters $h_a(t)$, $g(t, \tau)$ and $h_a^*(-t)$ are linear, their order may be exchanged as shown in Fig. 2.59, where the overall pulse is equal to $p(t) = h_a(t) * h_a^*(-t)$. The overall pulse $p(t)$ is usually chosen to be a Nyquist pulse. For example, $p(t)$ might be a raised cosine pulse, such that $h_a(t)$ and $h_a^*(-t)$ are root-raised cosine pulses. To obtain the symbol-spaced channel tap coefficients, the pulse $p(t)$ is passed through the channel $g(t, \tau)$ and symbol-spaced samples are extracted at the output. Assuming that the channel has the form in (2.304), the symbol-spaced samples are now shown to be a linear combination of the elements of the tap gain vector

$$\mathbf{g}(t) = (g_1(t), g_2(t), \dots, g_\ell(t))^T. \quad (2.348)$$

Suppose that a vector of M , symbol-spaced, tap coefficients

$$\mathbf{g}^T(t) = (g_1^T(t), g_2^T(t), \dots, g_M^T(t))^T \quad (2.349)$$

is to be generated. Then $\mathbf{g}^T(t) = \mathbf{A}\mathbf{g}(t)$, where $\mathbf{g}(t)$ is defined in (2.348), and \mathbf{A} is an $M \times \ell$ real matrix. As shown in Example 2.1 below, the entries of the matrix \mathbf{A} are determined by the overall pulse $p(t)$, the delay profile τ in (2.312), and the timing phase of the sampler that extracts the symbol-spaced samples. The matrix \mathbf{A} must be generated each time the delay profile and/or the sampler timing phase changes. For systems where timing information is derived from a training sequence or synchronization word that is inserted into every transmitted slot or burst, the sampler timing phase is usually adjusted on a burst-by-burst basis and, consequently, the matrix \mathbf{A} must be computed on a burst-by-burst basis as well.

The autocovariance matrix of the symbol-spaced tap gain vector $\mathbf{g}^T(t)$ is

$$\begin{aligned}\Phi_{\mathbf{g}^T}(\tau) &= \frac{1}{2} \mathbb{E} [\mathbf{g}^T(t) \mathbf{g}^T H(t + \tau)] \\ &= \frac{1}{2} \mathbb{E} [\mathbf{A} \mathbf{g}(t) \mathbf{g}^H(t + \tau) \mathbf{A}^T] \\ &= \mathbf{A} \frac{1}{2} \mathbb{E} [\mathbf{g}(t) \mathbf{g}^H(t + \tau)] \mathbf{A}^T \\ &= \mathbf{A} \Phi_{\mathbf{g}}(\tau) \mathbf{A}^T.\end{aligned}$$

For a WSSUS channel and 2-D isotropic scattering on each of the channel taps $g_i(t)$,

$$\Phi_{\mathbf{g}}(\tau) = \frac{1}{2} \text{diag}[\Omega_1, \Omega_2, \dots, \Omega_\ell] J_0(2\pi f_m \tau). \quad (2.350)$$

Example 2.1. Suppose that the channel $g(t, \tau)$ consists of two taps having the spacing $\tau_s = |\tau_1(t) - \tau_0(t)|$. In this example, the two main taps in the symbol-spaced channel model, $g_0^T(t)$ and $g_1^T(t)$, are generated. Let

$$\begin{aligned}\mathbf{g}(t) &= (g_0(t), g_1(t))^T \\ \mathbf{g}^T(t) &= (g_0^T(t), g_1^T(t))^T\end{aligned}$$

and

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

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 taps just happen to have equal strength, $|g_0(t)|^2 = |g_1(t)|^2$, when the sampler timing is being determined, i.e., during the training sequence or synchronization word. Furthermore, for the purpose of illustration, suppose that the result of the sampler timing phase adjustment is such that the symbol-spaced taps have equal strength as well, i.e., $|g_0^T(t)|^2 = |g_1^T(t)|^2$. Figure 2.60 illustrates this situation. The entries of matrix \mathbf{A} can be obtained by writing

$$\begin{aligned}g_0^T(t) &= p(\tau_s/2 - T/2)g_0(t) + p(-\tau_s/2 - T/2)g_1(t) \\ g_1^T(t) &= p(\tau_s/2 + T/2)g_0(t) + p(-\tau_s/2 + T/2)g_1(t).\end{aligned}$$

Hence,

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

Now suppose that the combination of the transmitter and receiver filter is a raised cosine pulse⁵

$$p(t) = \text{sinc}(t/T) \cdot \frac{\cos(\pi\beta t/T)}{1 - 4\beta^2 t^2/T^2}, \quad (2.351)$$

(continued)

⁵See Chap. 4 for a discussion of raised cosine pulse shaping.

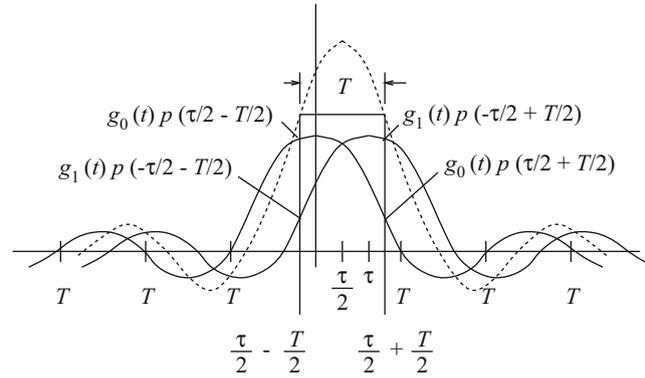


Fig. 2.60 Generation symbol-spaced channel taps

Example 2.1 (continued)

with roll-off factor $\beta = 0.35$, and $\tau_s = T/4$. Then

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

2.6 Shadowing

It was shown earlier in (2.26) that the received envelope power is

$$\Omega_p = E[|g(t)|^2] = \sum_{n=1}^N C_n^2, \quad (2.352)$$

where C_n depends on the cross sectional area of the n th local scatterer. In practice, the local mean is calculated by computing the time average

$$\hat{\Omega}_p = \frac{1}{T} \int_T |g(t)|^2 dt, \quad (2.353)$$

where the time interval T is chosen to correspond to a spatial averaging interval that is large enough to average over the envelope fades. In practice, this corresponds to a spatial averaging distance of about 20 wavelengths. Note that the required averaging interval T will depend on velocity. The averaging interval must be small enough so that the $\{C_n\}$ do not change over the averaging interval. The location area is defined as the largest volume of space where this condition will hold true. Sometimes Ω_p is called the local mean because it is computed within a location area. If the receiver moves outside the location area, the $\{C_n\}$ will change due to the presence of large terrain features like hills, valley, and buildings. Therefore, the local mean Ω_p (or $\hat{\Omega}_p$) changes with location in a process known as shadowing. The same statements can also be made for the mean envelope $\Omega_v = E[|g(t)|]$ and its time average

$$\hat{\Omega}_v = \frac{1}{T} \int_T |g(t)| dt. \quad (2.354)$$

Empirical studies have shown that Ω_p has the log-normal distribution

$$p_{\Omega_p}(x) = \frac{1}{x\sigma_{\Omega}\xi\sqrt{2\pi}} \exp\left\{-\frac{(10\log_{10}\{x\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_{\Omega}^2}\right\}, \quad (2.355)$$

where

$$\mu_{\Omega_p \text{ (dBm)}} = 10E[\log_{10}\{\Omega_p\}], \quad (2.356)$$

σ_{Ω} is the shadow standard deviation in decibel units, and $\xi = \ln(10)/10$. Note that since the local mean is in units of dBm (decibels with respect to 1 mW), the units of Ω_p in (2.355) are milliwatts (mW).

The mean value $\mu_{\Omega_p \text{ (dBm)}}$ is sometimes called the area mean because it corresponds to an extended location area defined as the largest volume of space where the $\{C_n\}$ can be characterized as stationary random variables and N is a constant. Within an extended location area, the mean envelope power $\mu_{\Omega_p} = E[\Omega_p]$ is constant. The mean envelope power depends on the propagation path loss between the transmitter and receiver, such that the path loss is constant within the extended location area.

By using a transformation of random variables, it can be shown that $\Omega_p \text{ (dBm)} = 10\log_{10}\{\Omega_p\}$ has the Gaussian density

$$p_{\Omega_p \text{ (dBm)}}(x) = \frac{1}{\sqrt{2\pi}\sigma_{\Omega}} \exp\left\{-\frac{(x - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_{\Omega}^2}\right\}. \quad (2.357)$$

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

Some confusion may arise in the literature because some authors [127, 128, 232] treat the mean envelope Ω_v as being log-normally distributed with shadow standard deviation σ_{Ω} , while other authors [208, 227, 267] treat the mean square-envelope Ω_p as being log-normally distributed with the same value of σ_{Ω} . Clearly, the mean envelope and mean square-envelope are not the same, and one may wonder if the same shadow standard deviation should be used in each case. It is shown in Appendix 2D that the shadow standard deviation σ_{Ω} is indeed the same in each case. However, with Ricean fading the means differ by

$$\mu_{\Omega_p \text{ (dBm)}} = \mu_{\Omega_v \text{ (dBm)}} + 10\log_{10}\{C(K)\}, \quad (2.358)$$

where

$$C(K) = \frac{4e^{2K}(K+1)}{\pi {}_1F_1^2(3/2, 1; K)} \quad (2.359)$$

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 [227]. The shadow standard deviation that is observed in microcells varies between 4 and 13 dB [40, 142, 143, 216, 228, 280]. Mogensen [228] has reported $\sigma_{\Omega} = 6.5\text{--}8.2$ dB at 900 MHz in urban areas, while Mockford et al. [227] report a value of 4.5 dB for urban areas. Berg [40] and Goldsmith and Greenstein [142] 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. [227] suggest that σ_{Ω} is 1.3–1.8 dB higher in a suburban environment than in an urban environment.

2.6.1 Shadow Simulation

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 [151, 152, 161, 170, 216]. 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

$$\Omega_{k+1} \text{ (dBm)} = \zeta\Omega_k \text{ (dBm)} + (1 - \zeta)v_k, \quad (2.360)$$

where Ω_k (dBm) is the mean envelope or mean squared-envelope, expressed in decibels, that is experienced at index k , ζ is a parameter that controls the spatial correlation of the shadows, and v_k is a zero-mean Gaussian random variable with $\phi_{v_v}(n) = \tilde{\sigma}^2 \delta(n)$. It can be shown that the spatial autocorrelation function of Ω_k (dBm) as generated by (2.360) is

$$\phi_{\Omega_{(\text{dBm})}\Omega_{(\text{dBm})}}(k) = \frac{1 - \zeta}{1 + \zeta} \tilde{\sigma}^2 \zeta^{|k|}. \quad (2.361)$$

Since the shadow variance is

$$\sigma_{\Omega}^2 = \phi_{\Omega_{(\text{dBm})}\Omega_{(\text{dBm})}}(0) = \frac{1 - \zeta}{1 + \zeta} \tilde{\sigma}^2 \quad (2.362)$$

the autocorrelation of Ω_k can be expressed as

$$\phi_{\Omega_{(\text{dBm})}\Omega_{(\text{dBm})}}(k) = \sigma_{\Omega}^2 \zeta^{|k|}. \quad (2.363)$$

This approach generates shadows that decorrelate exponentially with distance. 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 anything better, Gudmundson's model in (2.360) is still useful and effective.

While shadows decorrelate spatially, simulations are usually conducted in discrete-time. Therefore, to use the simulator in (2.360), the spatial decorrelation parameter ζ must be related to the simulation index k . Suppose the shadows that are experienced by a MS that is traveling with velocity v are to be modeled. 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 distance of D m. Then the time autocorrelation of shadowing is

$$\phi_{\Omega_{(\text{dBm})}\Omega_{(\text{dBm})}}(k) \equiv \phi_{\Omega_{(\text{dBm})}\Omega_{(\text{dBm})}}(kT) = \sigma_{\Omega}^2 \zeta_D^{(vT/D)|k|}. \quad (2.364)$$

Comparing (2.363) and (2.364), observe 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 and $\zeta_{100} = 0.82$. For typical microcellular propagation at 1700 MHz, Gudmundson has also reported $\sigma_{\Omega} = 4.3$ dB and $\zeta_{10} = 0.3$.

2.6.2 Composite Shadowing–Fading Distributions

Sometimes it is desirable to obtain the composite envelope distribution due to shadowing and multipath-fading. Such a composite distribution is relevant in cases where the MSs are slowly moving or stationary. In this case, the fading rate may be so slow that entire codewords are either faded completely or not at all regardless of the interleaving depth that is used, and the code fails. In this case, the composite shadow-fading distribution is useful for evaluating system coverage, i.e., the fraction of the service area having an acceptable quality of service.

Two different approaches have been suggested in the literature for obtaining the composite shadow-fading distribution. The first approach is to express the squared-envelope as a conditional density, conditioned on Ω_p , and then integrate over the density of Ω_p to obtain the composite distribution. This results in the composite shadow-fading distribution

$$p_{\alpha_c^2}(x) = \int_0^{\infty} p_{\alpha^2 | \Omega_p}(x|w) p_{\Omega_p}(w) dw. \quad (2.365)$$

For the case of Rayleigh fading, at any time instant t_1

$$\Omega_p = E[\alpha^2(t_1)] = 2b_0 \quad (2.366)$$

and, hence,

$$p_{\alpha^2 | \Omega_p}(x|w) = \frac{x}{w} e^{-x/w}. \quad (2.367)$$

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

$$p_{\alpha_c^2}(x) = \int_0^\infty \frac{1}{w} e^{-x/w} \frac{1}{w\sigma_\Omega \xi \sqrt{2\pi}} \exp\left\{-\frac{(10\log_{10}\{w\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} dw. \quad (2.368)$$

where $\xi = \ln(10)/10$. Unfortunately, this distribution does not exist in closed form, but can be efficiently evaluated using Gauss–Hermite quadrature integration.

The second approach, originally suggested by Lee and Yeh [199], expresses the composite squared-envelope as the product of the squared-envelope due to multipath-fading and shadow fading. Using this approach,

$$\hat{\alpha}_c^2(t) = \alpha^2(t) \cdot \Omega_p(t). \quad (2.369)$$

Under the assumption that the fading and shadowing are independent random processes, both approaches lead to identical results as is now shown. The composite squared-envelope in (2.369) is the product of two random variables at any time instant t_1 . Hence, the corresponding density function of the squared-envelope can be obtained by using a bivariate transformation of random variables and then integrating to obtain the marginal density. This leads to the composite density function

$$p_{\hat{\alpha}_c^2}(x) = \int_0^\infty \frac{1}{w} p_{\alpha^2}\left(\frac{x}{w}\right) p_{\Omega_p}(w) dw. \quad (2.370)$$

Again, consider the case of log-normal shadowing and Rayleigh fading. Using (2.54) and (2.355) gives

$$p_{\hat{\alpha}_c^2}(x) = \int_0^\infty \frac{1}{\Omega_p w} \exp\left\{-\frac{x}{\Omega_p w}\right\} \frac{1}{w\sigma_\Omega \xi \sqrt{2\pi}} \exp\left\{-\frac{(10\log_{10}\{w\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} dw. \quad (2.371)$$

Observe that (2.368) and (2.371) are related by

$$p_{\alpha_c}(x) = \Omega_p p_{\hat{\alpha}_c}(\Omega_p x). \quad (2.372)$$

Hence, if the faded envelope $\alpha(t)$ in the second approach is assumed to have $E[\alpha(t)^2] = \Omega_p = 1$, then α_c^2 and $\hat{\alpha}_c^2$ will have the exact same composite distribution. Although the above result may not matter much for analysis, it does have implications for software simulation. When simulating the combined effects of fading and shadowing, the composite squared-envelope can be obtained by generating the fading and shadowing processes separately and multiplying them together as shown in (2.369). To do so, the faded envelope should be generated such that $\Omega_p = 1$.

2.6.2.1 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 Rice distribution. The composite distribution of the squared-envelope due to Nakagami fading and log-normal shadowing has the Gamma-log-normal density function

$$p_{\alpha_c^2}(x) = \int_0^\infty \left(\frac{m}{w}\right)^m \frac{x^{m-1}}{\Gamma(m)} \exp\left\{-\frac{mx}{w}\right\} \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega w} \exp\left\{-\frac{(10\log_{10}\{w\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} dw. \quad (2.373)$$

where $\xi = \ln(10)/10$. As shown in Appendix 2E, the composite Gamma-log-normal distribution in (2.373) can be closely approximated by a log-normal distribution with parameters

$$\begin{aligned} \mu_{\text{(dBm)}} &= \xi^{-1} (\psi(m) - \ln(m)) + \mu_{\Omega_p \text{ (dBm)}} \\ \sigma^2 &= \xi^{-2} \zeta(2, m) + \sigma_\Omega^2 \end{aligned} \quad (2.374)$$

where $\psi(\cdot)$ is the Euler psi function and $\zeta(\cdot, \cdot)$ is Riemann's zeta function as defined in Appendix 2E. When $m = 1$ the approximation is valid for $\sigma_\Omega > 6$ dB; when $m = 2$, $\sigma_\Omega > 6$ must be greater than or equal to 4 dB; when $m \geq 4$, the approximation is good for all σ_Ω [164].

The effect of Nakagami fading in (2.374) is to decrease the mean and increase the variance of the composite squared-envelope. However, this affect decreases as the shape factor m increases (corresponding to less severe fading). For example, with $m = 1$ (Rayleigh fading), $\mu_{(\text{dBm})} = \mu_{\Omega_p (\text{dBm})} - 2.50675$ and $\sigma^2 = \sigma_\Omega^2 + 31.0215$ while, with $m = 8$, $\mu_{(\text{dBm})} = \mu_{\Omega_p (\text{dBm})} - 0.277$ and $\sigma^2 = \sigma_\Omega^2 + 2.50972$.

2.7 Path Loss Models

Path loss is the largest and most variable quantity in a communication link budget. It depends on frequency, antenna heights, and distance and topography. A variety of theoretical and empirical path loss models exist in the literature. Our discussion starts with a discussion of theoretical models, followed by empirical models.

2.7.1 Free Space Path Loss

Free-space path loss (FSPL) is proportional to the square of the distance between the transmitter and receiver, and also proportional to the square of the frequency f_c of the radio signal. The FSPL equation is

$$L_{\text{FS}} = \left(\frac{4\pi d}{\lambda_c} \right)^2 = \left(\frac{4\pi d f_c}{c} \right)^2. \quad (2.375)$$

FSPL is a combination of two effects: First, the intensity of an electromagnetic wave in free space decays with the square of the radio path length, d , such that the received power per unit area or power spatial density (in watts per meter-squared) at distance d is

$$\Omega_r(d) = \Omega_t \frac{1}{4\pi d^2}, \quad (2.376)$$

where Ω_t is the total transmit power in watts. Note that this term is not frequency dependent.

The second effect is due to aperture, which determines how well an antenna picks up power from an incoming electromagnetic wave. For an isotropic antenna,

$$\Omega_p(d) = \Omega_r(d) \frac{\lambda_c^2}{4\pi}, \quad (2.377)$$

where $\Omega_p(d)$ is the received power. Note that this is entirely dependent on wavelength, λ_c , which is how the frequency-dependent behavior arises.

Using (2.376) and (2.377) gives the free space propagation path loss as

$$\begin{aligned} L_{\text{FS}} (\text{dB}) &= 10 \log_{10} \left\{ \frac{\Omega_t}{\Omega_p(d)} \right\} = 10 \log_{10} \left\{ \left(\frac{4\pi d}{\lambda_c} \right)^2 \right\} \\ &= 10 \log_{10} \left\{ \left(\frac{4\pi d f_c}{c} \right)^2 \right\} \\ &= 20 \log_{10} \{f_c\} + 20 \log_{10} \{d\} - 147.55 \text{ dB}. \end{aligned} \quad (2.378)$$

2.7.2 Flat Earth Path Loss

The signals in land mobile radio environments do not experience free space propagation. A more appropriate theoretical path loss model is one that assumes propagation over a flat reflecting surface, the Earth, as shown in Fig. 2.61. The length of the direct path is

$$d_1 = \sqrt{d^2 + (h_b - h_m)^2} \quad (2.379)$$

and the length of the reflected path is

$$d_2 = \sqrt{d^2 + (h_b + h_m)^2} \quad (2.380)$$

Given that $d \gg h_b h_m$, it follows that $d_1 \approx d$ and $d_2 \approx d$. However, since the wavelength is small, the direct and reflected paths may add constructively or destructively over small distances. The carrier phase difference between the direct and reflected paths is

$$\phi_2 - \phi_1 = \frac{2\pi}{\lambda_c} (d_2 - d_1) \quad (2.381)$$

Taking into account the phase difference, the received envelope power is

$$\mu_{\Omega_p} = \Omega_t \left(\frac{\lambda_c}{4\pi d} \right)^2 \left| 1 + a e^{-jb} e^{j(\phi_2 - \phi_1)} \right|^2, \quad (2.382)$$

where a and b are the amplitude attenuation and phase change introduced by the flat reflecting surface. If a perfect specular reflection is assumed, then $a = 1$ and $b = \pi$ for small θ and

$$\begin{aligned} \mu_{\Omega_p} &= \Omega_t \left(\frac{\lambda_c}{4\pi d} \right)^2 \left| 1 - e^{j\left(\frac{2\pi}{\lambda_c} \Delta_d\right)} \right|^2 \\ &= 4\Omega_t \left(\frac{\lambda_c}{4\pi d} \right)^2 \sin^2 \left(\frac{\pi}{\lambda_c} \Delta_d \right), \end{aligned} \quad (2.383)$$

where $\Delta_d = (d_2 - d_1)$. Given that $d \gg h_b$ and $d \gg h_m$, and applying the approximation $\sqrt{1+x} \approx 1 + x/2$ for small x to (2.379) and (2.380),

$$\Delta_d \approx \frac{2h_b h_m}{d}. \quad (2.384)$$

Finally, the received envelope power is

$$\mu_{\Omega_p} \approx 4\Omega_t \left(\frac{\lambda_c}{4\pi d} \right)^2 \sin^2 \left(\frac{2\pi h_b h_m}{\lambda_c d} \right), \quad (2.385)$$

where h_b and h_m are the heights of the BS and MS antennas, respectively. Under the condition that $d \gg h_b h_m$, (2.385) reduces to

$$\mu_{\Omega_p} \approx \Omega_t \left(\frac{h_b h_m}{d^2} \right)^2, \quad (2.386)$$

where the small angle approximation $\sin x \approx x$ for small x has been invoked. 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 is inversely proportional to the fourth power of the distance rather than the square of the distance. Finally, the model in (2.386) shows that the received signal power is proportional to the product of the square of the BS and MS antenna heights.

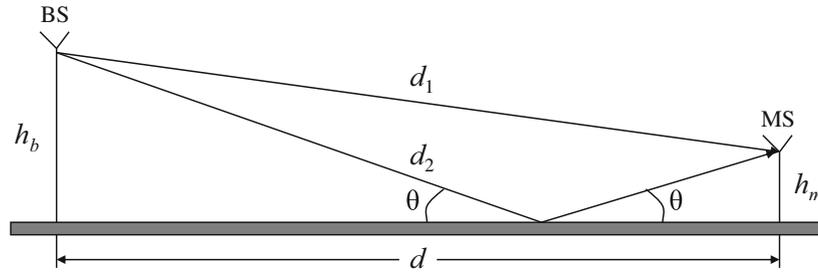


Fig. 2.61 Radio propagation over a flat reflecting surface

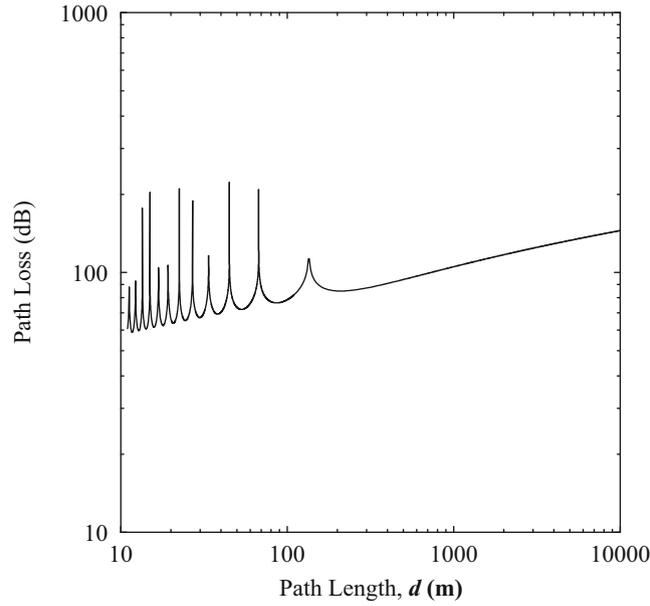


Fig. 2.62 Propagation path loss with distance over a flat reflecting surface; $h_b = 7.5$ m, $h_m = 1.5$ m, $f_c = 1800$ MHz

Figure 2.62 plots the flat Earth path loss (FEPL)

$$\begin{aligned}
 L_{FE} \text{ (dB)} &= 10 \log_{10} \left\{ \frac{\Omega_t}{\mu_{\Omega_p}} \right\} \\
 &= -10 \log_{10} \left\{ 4 \left(\frac{\lambda_c}{4\pi d} \right)^2 \sin^2 \left(\frac{2\pi h_b h_m}{\lambda_c d} \right) \right\} \text{ dB}
 \end{aligned} \tag{2.387}$$

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 propagation property has been noted in experiments by Milstein et al. [225]. The last local maxima in the path loss occurs when

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

giving the break-point distance

$$d_{BP} = \frac{4h_b h_m}{\lambda_c}. \tag{2.388}$$

2.7.3 Empirical Path Loss Models

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 [250] based on Okumura's prediction method [161] and Lee's model [196].

2.7.3.1 Okumura–Hata and CCIR Models

Hata's empirical model [161] is simple to use, and distinguishes between various degrees of urbanization. The empirical data for this model was collected by Okumura [250] in the city of Tokyo. The Okumura–Hata 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. Note the units of the parameters that are used in the model. The model is known to match the experimental data from which is formed to within 1 dB for distances ranging from 1 to 20 km. With the Okumura–Hata model, the path loss between two isotropic BS and MS antennas is:

$$L_p \text{ (dB)} = \begin{cases} A + B \log_{10}\{d\} & \text{for urban area} \\ A + B \log_{10}\{d\} - C & \text{for suburban area} \\ A + B \log_{10}\{d\} - D & \text{for open area} \end{cases} \quad (2.389)$$

where

$$A = 69.55 + 26.16 \log_{10}\{f_c\} - 13.82 \log_{10}\{h_b\} - a(h_m) \quad (2.390)$$

$$B = 44.9 - 6.55 \log_{10}\{h_b\}$$

$$C = 5.4 + 2 (\log_{10}\{f_c/28\})^2$$

$$D = 40.94 + 4.78 (\log_{10}\{f_c\})^2 - 18.33 \log_{10}\{f_c\}$$

and

$$a(h_m) = \begin{cases} (1.1 \log_{10}\{f_c\} - 0.7) h_m - (1.56 \log_{10}\{f_c\} - 0.8) & \text{for medium or small city} \\ \begin{cases} 8.28 (\log_{10}\{1.54 h_m\})^2 - 1.1 & \text{for } f_c \leq 200 \text{ MHz} \\ 3.2 (\log_{10}\{11.75 h_m\})^2 - 4.97 & \text{for } f_c \geq 400 \text{ MHz} \end{cases} & \text{for large city} \end{cases} \quad (2.391)$$

Typical values from the Okumura–Hata “large city” model are plotted in Fig. 2.63, for a BS height of 70 m, a MS antenna height of 1.5 m, and a carrier frequency of 900 MHz. The reader is cautioned that, due to a lesser degree of urbanization, 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. Also, the North American urban areas have path losses more like the Japanese suburban areas.

To account for varying degrees of urbanization, the CCIR (Comité International des Radio-Communication, now ITU-R) developed an empirical model for the path loss as:

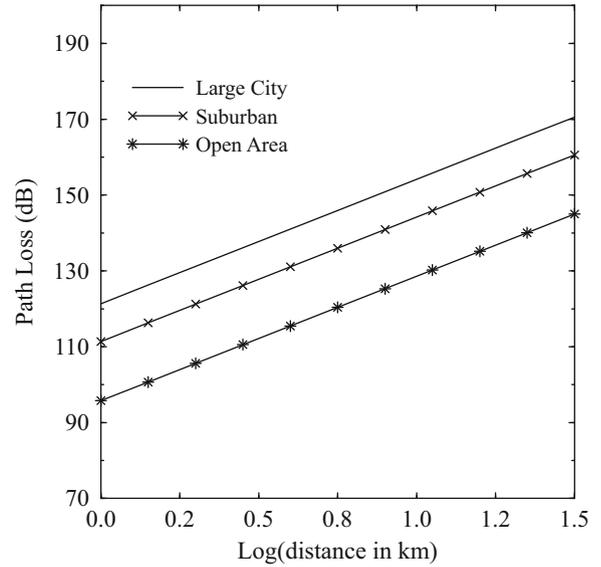
$$L_p \text{ (dB)} = A + B \log_{10}\{d\} - E, \quad (2.392)$$

where A and B are defined in (2.390) with $a(h_m)$ being the medium or small city value in (2.391). The parameter E accounts for the degree of urbanization and is given by

$$E = 30 - 25 \log_{10}\{\% \text{ of area covered by buildings}\}, \quad (2.393)$$

where $E = 0$ when the area is covered by approximately 16% buildings.

Fig. 2.63 Path loss obtained from the Okumura–Hata model; $h_b = 70$ m, $h_m = 1.5$ m, $f_c = 900$ MHz



2.7.3.2 Lee's Area-to-Area Model

William C. Y. Lee's area-to-area model [196] can be 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, and other prediction models can be used [196]. Two parameters are required for Lee's area-to-area model; the received power at a 1 mile (1.6 km) point of interception, $\mu_{\Omega_p}(d_o)$, and the path-loss exponent, β . The received signal power at distance d can be expressed as

$$\mu_{\Omega_p}(d) = \mu_{\Omega_p}(d_o) \left(\frac{d}{d_o}\right)^{-\beta} \left(\frac{f}{f_o}\right)^{-n} \alpha_0 \quad (2.394)$$

or in decibel units

$$\mu_{\Omega_p}(\text{dBm})(d) = \mu_{\Omega_p}(\text{dBm})(d_o) - 10\beta \log_{10} \left\{ \frac{d}{d_o} \right\} - 10n \log_{10} \left\{ \frac{f}{f_o} \right\} + 10 \log_{10} \{ \alpha_0 \}, \quad (2.395)$$

where d is in units of 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 $f_o = 900$ 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 the following parameters are computed:

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

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

$$\alpha_3 = \frac{\text{transmitter power}}{10 \text{ watts}}$$

Table 2.3 Parameters for Lee's area-to-area model in various propagation environments, from [196]

Terrain	$\mu_{\Omega_p} \text{ (dBm)}(d_o)$	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

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

$$\alpha_5 = \text{different antenna-gain correction factor at the MS} \quad (2.396)$$

From these parameters, the correction factor α_0 is

$$\alpha_0 = \alpha_1 \cdot \alpha_2 \cdot \alpha_3 \cdot \alpha_4 \cdot \alpha_5. \quad (2.397)$$

The parameters β and $\mu_{\Omega_p}(d_o)$ have been found from empirical measurements, and are listed in Table 2.3.

Experimental data suggest that n in (2.395) 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 (2.396), also determined from empirical data, is

$$\kappa = \begin{cases} 2 & \text{for a MS antenna height} > 10 \text{ m} \\ 3 & \text{for a MS antenna height} < 3 \text{ m} \end{cases}. \quad (2.398)$$

The path loss $L_p \text{ (dB)}$ is the difference between the transmitted and received envelope power, $L_p \text{ (dB)} = \mu_{\Omega_p} \text{ (dBm)}(d) - \mu_{\Omega_r} \text{ (dBm)}$. To compare directly with the Okumura–Hata model in Fig. 2.63, we assume an isotropic BS antenna with 0 dB gain, such that $\alpha_4 = -6$ dB. Then by using the same parameters as in Fig. 2.63, $h_b = 70$ m, $h_m = 1.5$ m, $f_c = 900$ MHz, a nominal BS transmitter power of 40 dBm (10 watts), and the parameters in Table 2.3 for $\mu_{\Omega_p} \text{ (dBm)}(d_o)$ and β , the following path losses are obtained:

$$L_p \text{ (dB)} = \begin{cases} 85.74 + 20.0 \log_{10}\{d\} & \text{Free Space} \\ 84.94 + 43.5 \log_{10}\{d\} & \text{Open Area} \\ 98.68 + 38.4 \log_{10}\{d\} & \text{Suburban} \\ 107.31 + 36.8 \log_{10}\{d\} & \text{Philadelphia} \\ 100.02 + 43.1 \log_{10}\{d\} & \text{Newark} \\ 122.59 + 30.5 \log_{10}\{d\} & \text{Tokyo} \end{cases} \quad (2.399)$$

These typical values from Lee's area-to-area model are plotted in Fig. 2.64.

2.7.3.3 COST231-Hata Model

The COST231-Hata model is based on the proposal by Mogensen et al. [228] to extend the Okumura–Hata model for use in the 1500–2000 MHz frequency range, where it is known that the Okumura–Hata model underestimates 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). Note again that the parameters must be used in the model with their specified units. The path loss as predicted by the COST231-Hata model is [80]

$$L_p \text{ (dB)} = A + B \log_{10}\{d\} + C, \quad (2.400)$$

where

$$A = 46.3 + 33.9 \log_{10}\{f_c\} - 13.82 \log_{10}\{h_b\} - a(h_m)$$

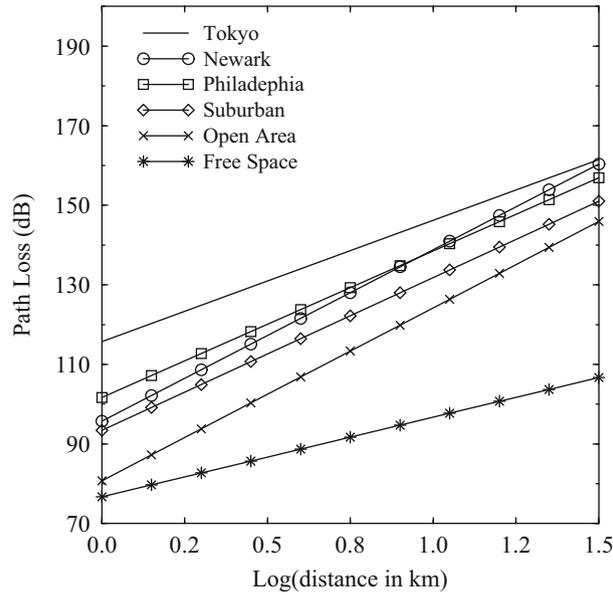


Fig. 2.64 Path loss obtained by using Lee's method; $h_b = 70$ m, $h_m = 1.5$ m, $f_c = 900$ MHz, and an isotropic BS antenna

$$B = 44.9 - 6.55 \log_{10}\{h_b\}$$

$$C = \begin{cases} 0 & \text{medium city and suburban areas} \\ & \text{with moderate tree density} \\ 3 & \text{for metropolitan centers} \end{cases}$$

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. They should not be used for smaller ranges, where path loss becomes highly dependent upon the local topography. The COST231-Hata model is good down to a path length of 1 km.

2.7.3.4 COST231-Walfish-Ikegami Model

The COST231-Walfish-Ikegami model was developed for microcellular systems and 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

$$L_p \text{ (dB)} = 42.6 + 26 \log_{10}\{d\} + 20 \log_{10}\{f_c\}, \quad d \geq 20 \text{ m}, \quad (2.401)$$

where the first constant is chosen so that L_p is equal to the FSPL 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. 2.65, the path loss for non-line-of-sight (NLoS) propagation is expressed in terms of the following parameters:

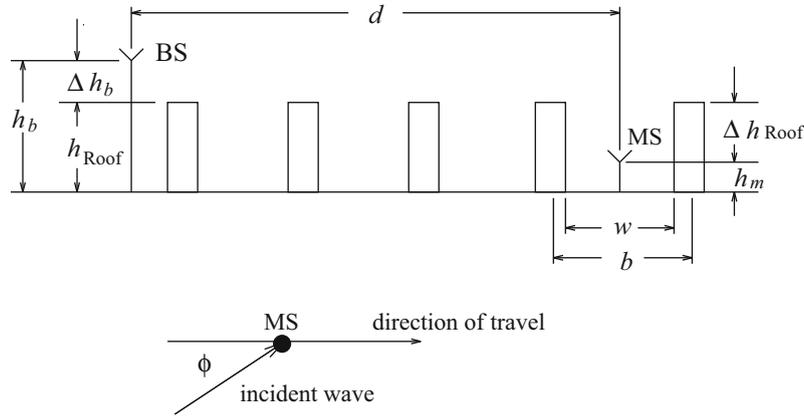


Fig. 2.65 Definition of parameters used in the COST231-Walfish-Ikegami model

d = distance (m)

h_b = BS antenna height over street level, $4 \leq h_b \leq 50$ (m)

h_m = MS antenna height over street level, $1 \leq h_m \leq 3$ (m)

h_{Roof} = nominal roof height of buildings (m)

$\Delta h_b = h_b - h_{\text{Roof}}$ = height of BS relative to rooftops (m)

$\Delta h_m = h_{\text{Roof}} - h_m$ = height of MS relative to rooftops (m)

w = width of streets (m)

b = building separation (m)

ϕ = angle of incident wave with respect to street (degrees)

If no data on the structure of the buildings and roads are available, the following default values are recommended, $b = 20 \dots 50$ (m), $w = b/2$ (m), $\phi = 90^\circ$, and $h_{\text{Roof}} = 3 \times \text{number of floors} + \text{roof}$ (m), where roof = 3 (m) pitched and 0 (m) flat.

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

$$L_p \text{ (dB)} = \begin{cases} L_o + L_{\text{rts}} + L_{\text{msd}} & , \text{ for } L_{\text{rts}} + L_{\text{msd}} \geq 0 \\ L_o & , \text{ for } L_{\text{rts}} + L_{\text{msd}} < 0 \end{cases} \quad (2.402)$$

where

$$L_o = \text{free-space loss} = 32.4 + 20\log_{10}\{d\} + 20\log_{10}\{f_c\}$$

$$L_{\text{rts}} = \text{roof-top-to-street diffraction and scatter loss}$$

$$L_{\text{msd}} = \text{multi-screen diffraction loss.}$$

Note that the expression for free-space loss differs from (2.378) because here the units of d are in kilometers and the units of f_c are in megahertz. The roof-top-to-street diffraction and scatter loss represents the coupling of the wave propagation along the multi-screen path into the street where the MS is located, and is given by

$$L_{\text{rts}} = -16.9 - 10\log_{10}\{w\} + 10\log_{10}\{f_c\} + 20\log_{10}\{\Delta h_m\} + L_{\text{ori}}, \quad (2.403)$$

where

$$L_{\text{ori}} = \begin{cases} -10 + 0.354(\phi) & , 0 \leq \phi \leq 35^\circ \\ 2.5 + 0.075(\phi - 35^\circ) & , 35^\circ \leq \phi \leq 55^\circ \\ 4.0 - 0.114(\phi - 55^\circ) & , 55^\circ \leq \phi \leq 90^\circ \end{cases} \quad (2.404)$$

is a street orientation loss.

The multi-screen diffraction loss is

$$L_{\text{msd}} = L_{\text{bsh}} + k_a + k_d \log_{10}\{d\} + k_f \log_{10}\{f_c\} - 9 \log_{10}\{b\}, \quad (2.405)$$

where

$$L_{\text{bsh}} = \begin{cases} -18 \log_{10}\{1 + \Delta h_b\} & , h_b > h_{\text{Roof}} \\ 0 & , h_b \leq h_{\text{Roof}} \end{cases} \quad (2.406)$$

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

$$k_a = \begin{cases} 54, & h_b > h_{\text{Roof}} \\ 54 - 0.8\Delta h_b, & d \geq 0.5 \text{ km and } h_b \leq h_{\text{Roof}} \\ 54 - 0.8\Delta h_b d / 0.5, & d < 0.5 \text{ km and } h_b \leq h_{\text{Roof}} \end{cases} \quad (2.407)$$

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

$$k_d = \begin{cases} 18, & h_b > h_{\text{Roof}} \\ 18 - 15\Delta h_b / h_{\text{Roof}}, & h_b \leq h_{\text{Roof}} \end{cases} \quad (2.408)$$

$$k_f = -4 + \begin{cases} 0.7(f_c/925 - 1), & \text{medium city and suburban} \\ 1.5(f_c/925 - 1), & \text{metropolitan area} \end{cases} \quad (2.409)$$

The COST231-Walfish-Ikegami model works best for $h_b \gg h_{\text{Roof}}$. Large prediction errors can be expected for $h_b \approx h_{\text{Roof}}$. The model is poor for $h_b \ll h_{\text{Roof}}$ because the terms in (2.407) do not consider wave guiding in street canyons and diffraction at street corners.

2.7.3.5 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 [149, 160, 173, 267, 338, 351]

$$\mu_{\Omega_p} = \frac{k\Omega_t}{d^a(1 + d/g)^b}, \quad (2.410)$$

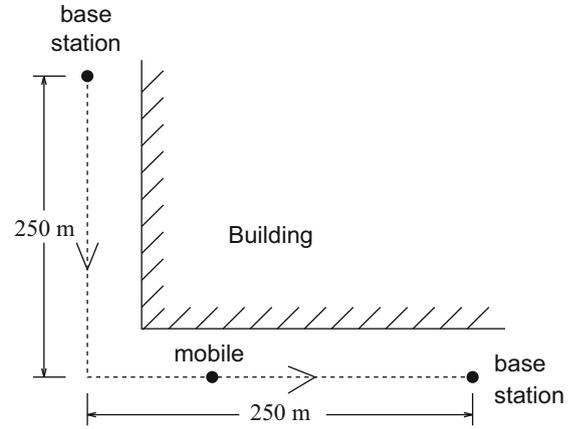
where Ω_t is the transmitted power, k is a constant of proportionality, and d (m) is the distance. For small path length distances, free space propagation will prevail so that $a = 2$. The parameter g is called the break-point and ranges from 150 to 300 m [149, 160, 173, 351]. At larger distances, an inverse-fourth to -eighth power law is experienced so that b ranges from 2 to 6 [160]. The model parameters that were obtained by Harley are listed in Table 2.4. Xia [358] has demonstrated that the break-point occurs where the Fresnel zone between the transmit and receive antennas just touches the ground assuming a flat surface. This distance is

$$g = \frac{1}{\lambda_c} \sqrt{(\Sigma^2 - \Delta^2)^2 - 2(\Sigma^2 + \Delta^2) \left(\frac{\lambda_c}{2}\right)^2 + \left(\frac{\lambda_c}{2}\right)^4}, \quad (2.411)$$

Table 2.4 Two-slope path loss parameters obtained by Harley, from [160]

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

Fig. 2.66 The corner effect in a street microcell environment



where $\Sigma = h_b + h_m$ and $\Delta = h_b - h_m$. For high frequencies this distance can be approximated as $g = 4h_b h_m / \lambda_c$, which is the same distance as the last local maxima in the flat reflecting surface model in Sect. 2.7.2. 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. 2.66. In this case, the average received signal strength can drop by as much as 25–30 dB over distances as small as 10 m for low antenna heights in an area with multi-story buildings [61, 210, 233, 286, 324], 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 known as 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 by assuming LoS propagation from a virtual 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 is given by

$$\mu_{\Omega_p} = \begin{cases} \frac{k\Omega_t}{d^a(1+d/g)^b} & d \leq d_c \\ \frac{k\Omega_t}{d_c^a(1+d_c/g)^b} \cdot \frac{1}{(d-d_c)^a(1+(d-d_c)/g)^b} & d > d_c \end{cases} \quad (2.412)$$

where d_c (m) is the distance between the serving BS and the corner. For the scenario depicted in Fig. 2.66, the received signal strength with this model is shown in Fig. 2.67. The heavy curves show the average received signal strength from the two BSs as the MS traverses the dashed path shown in Fig. 2.66. These curves were obtained by using $a = 2$, $b = 2$, $g = 150$ m, and $d_c = 250$ m in (2.412), and assuming that $\mu_{\Omega_p} = 1$ dBm at $d = 1$ m. The dotted curves superimposed on the heavy lines in Fig. 2.67 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 Sects. 2.6.1 and 2.5.2.2.

2.7.3.6 3GPP 3-D Path Loss Models

The 3GPP path loss models as described in [1] are valid from 2 to 6 GHz for different BS and MS antenna heights. The distance definitions are defined in Fig. 2.68 for outdoor scenarios and Fig. 2.69 for outdoor–indoor scenarios. The 3GPP path loss models are categorized into urban macrocell (UMa) and urban microcell (UMi) cases, corresponding to BS antenna heights of 25 m or less and 25 m or more, respectively. The UMa and UMi cases are further categorized into LoS, NLoS, and outdoor-to-indoor scenarios.

Fig. 2.67 Received signal strength for the street microcell environment in Fig. 2.66. Solid lines show the area mean signal strength, while the dashed lines account for shadowing and fading as well. For this latter case, $\sigma_{\Omega} = 6$ dB and $\phi_{\Omega_p, \text{dBm}} \Omega_p, \text{dBm}(d) = 0.1\sigma_{\Omega}^2$ at $d = 30$ m. For each 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. 2.66

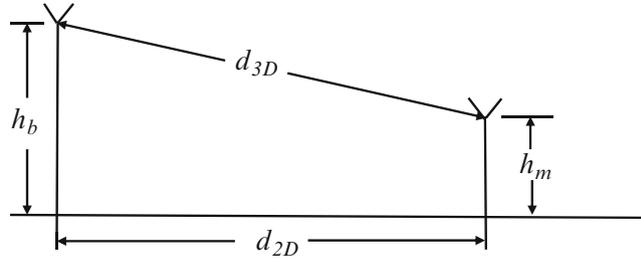
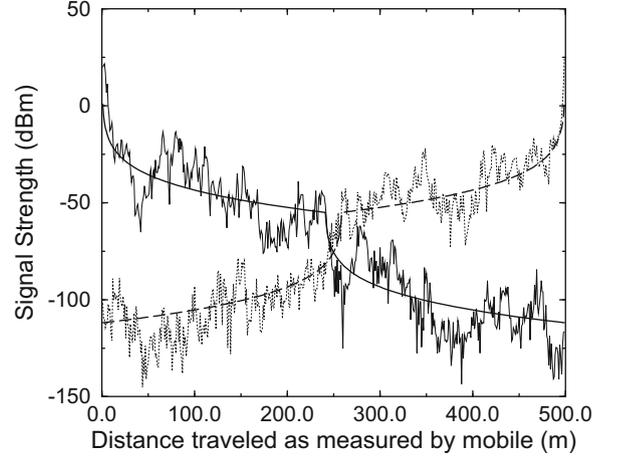


Fig. 2.68 Definition of d_{2D} , and d_{3D} for outdoor mobile stations, from [1]

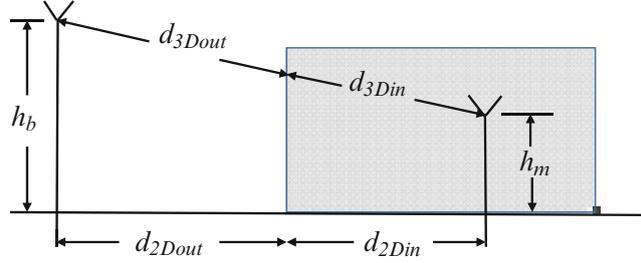


Fig. 2.69 Definition of d_{2Dout} , d_{2Din} , d_{3Dout} and d_{3Din} for indoor mobile stations, from [1]

LoS Probability

The various 3GPP path loss models make use of the probability of LoS condition. For microcells and outdoor MSs, the probability of LoS is [1]

$$P_{\text{LoS}} = \min(18/d_{2D}, 1)(1 - e^{-d_{2D}/36}) + e^{-d_{2D}/36} \quad (2.413)$$

For microcells and indoor MSs, the above formula is used with d_{2D} replaced by d_{2Dout} . For macrocells and outdoor MSs, the probability of LoS is [1]

$$P_{\text{LoS}} = (\min(18/d_{2D}, 1)(1 - e^{-d_{2D}/63}) + e^{-d_{2D}/63})(1 + C(d_{2D}, h_m)) \quad (2.414)$$

where

$$C(d_{2D}, h_m) = \begin{cases} 0 & , h_m < 13 \text{ m} \\ \left(\frac{h_m - 13}{10}\right)^{1.5} g(d_{2D}) & , 13 \text{ m} \leq h_m \leq 23 \text{ m} \end{cases} \quad (2.415)$$

and

$$g(d_{2D}) = \begin{cases} (1.25e^{-6})d_{2D}^2 e^{d_{2D}/150} & , d_{2D} > 18 \text{ m} \\ 0 & , \text{ otherwise} \end{cases} \quad (2.416)$$

For macrocells and indoor MSs, the above formulas are used with d_{2D} replaced by $d_{2D\text{out}}$. For frequencies above 6 GHz, including mm-wave frequencies, the study in [158] recommends to use the above 3GPP LoS models.

Using the above LoS probability models, the 3GPP 3-D path loss models are defined below for various topographic scenarios, as described in [1].

3D-UMa LoS

For macrocells with LoS conditions

$$L_{\text{UMaLoS (dB)}} = 22.0 \log_{10}(d_{3D}) + 28.0 + 20 \log_{10}(f_c) , \quad 10 \text{ m} < d_{2D} < d_{BP} \quad (2.417)$$

$$L_{\text{UMaLoS (dB)}} = 40 \log_{10}(d_{3D}) + 28.0 + 20 \log_{10}(f_c) - 9 \log_{10}(d_{BP}^2 + (h_b - h_m)^2) , \quad d_{BP} < d_{2D} < 5000 \text{ m} \\ h_b = 25 \text{ m}; \quad 1.5 \text{ m} \leq h_m \leq 22.5 \text{ m} \quad (2.418)$$

The break-point distance is given by $d_{BP} = 4h_b h_m f_c / c$ corresponding to the last local maxima in the flat earth model (2.388). In the 3D-UMa scenario the effective antenna heights h_b and h_m are computed as follows: $h_b = \hat{h}_b - h_E$, $h_m = \hat{h}_m - h_E$, where \hat{h}_b and \hat{h}_m are the actual antenna heights, and the effective environment height h_E depends on the link between a BS and a MS. For LoS links, $h_E = 1 \text{ m}$ with probability $1/(1 + C(d_{2D}, h_m))$, where the function $C(d_{2D}, h_m)$ is defined in (2.415). Otherwise, h_E is chosen from a discrete uniform distribution on the set $\{12, 15, \dots, (h_m - 1.5)\}$.

The shadow standard deviation is $\sigma_\Omega = 4 \text{ dB}$.

3D-UMa NLoS

For macrocells with NLoS conditions

$$L_{\text{UMaNLoS (dB)}} = \max \{L_{\text{UMaNLoS (dB)}}, L_{\text{UMaLoS (dB)}}\}, \quad (2.419)$$

where

$$L_{\text{UMaNLoS (dB)}} = 161.04 - 7.1 \log_{10}(W) + 7.5 \log_{10}(h_{\text{build}}) \\ - (24.37 - 3.7(h_{\text{build}}/h_b)^2) \log_{10}(h_b) \\ + (43.42 - 3.1 \log_{10}(h_b)) (\log_{10}(d_{3D}) - 3) \\ + 20 \log_{10}(f_c) - (3.2(\log_{10}(17.625))^2 - 4.97) - 0.6(h_m - 1.5) \quad (2.420)$$

and

$$10 \text{ m} < d_{2D} < 5000 \text{ m}$$

h_{build} = average building height

W = street width

$$h_b = 25 \text{ m}, \quad 1.5 \text{ m} \leq h_m \leq 22.5 \text{ m}, \quad W = 20 \text{ m}, \quad h_{\text{build}} = 20 \text{ m}$$

Applicable ranges:

$$5 \text{ m} < h_{\text{build}} < 50 \text{ m}$$

$$5 \text{ m} < W < 50 \text{ m}$$

$$10 \text{ m} < h_b < 150 \text{ m}$$

$$1.5 \text{ m} \leq h_m \leq 22.5 \text{ m}$$

The shadow standard deviation is $\sigma_\Omega = 6 \text{ dB}$.

3D-UMa O-to-I

For macrocells with outdoor-to-indoor conditions

$$L_{\text{UMaO-to-I (dB)}} = L_b \text{ (dB)} + L_{\text{tw (dB)}} + L_{\text{in (dB)}} \quad (2.421)$$

For a hexagonal cell layout:

$$L_b \text{ (dB)} = L_{\text{UMa (dB)}}(d_{3D\text{-out}} + d_{3D\text{-in}})$$

$$L_{\text{tw (dB)}} = 20 \text{ (loss through wall)}$$

$$L_{\text{in (dB)}} = 0.5d_{2D\text{-in}} \text{ (inside loss)}$$

where

$$10 \text{ m} < d_{2D\text{-out}} + d_{2D\text{-in}} < 1000 \text{ m}$$

$$0 \text{ m} < d_{2D\text{-in}} < 25 \text{ m}$$

$$h_b = 25 \text{ m}, h_m = 3(n_{fl} - 1) + 1.5, \quad n_{fl} = 1, 2, 3, 4, 5, 6, 7, 8$$

$d_{2D\text{-in}}$ is assumed uniformly distributed between 0 and 25.

The shadow standard deviation is $\sigma_\Omega = 7 \text{ dB}$.

The building penetration loss (BPL) or “loss through wall” in the 3GPP 3D-UMa O-to-I model is 20 dB. However, this will vary greatly depending on the building. Moreover, the building penetration loss increases with frequency. An empirical BPL model was suggested in [158] as

$$\text{BPL}_{\text{(dB)}} = 10 \log_{10} (A + Bf_c^2), \quad (2.422)$$

where f_c is the frequency in GHz, $A = 5$ and $B = 0.03$ for low loss buildings and $A = 10$ and $B = 5$ for high loss buildings.

3D-UMi LoS

The microcell LoS path loss is the same as the macrocell LoS path loss $L_{\text{UMaLoS (dB)}}$, except that $h_E = 1 \text{ m}$ with probability one and the shadow standard deviation is $\sigma_\Omega = 3 \text{ dB}$.

3D-UMi NLoS

For a hexagonal cell layout

$$L_{\text{UMiNLoS (dB)}} = \max \{L_{\text{UMiNLoS (dB)}}, L_{\text{UMiLoS (dB)}}\}, \quad (2.423)$$

where

$$L_{\text{UMiNLoS (dB)}} = 36.7 \log_{10}(d_{3D}) + 22.7 + 26 \log_{10}(f_c) - 0.3(h_m - 1.5) \quad (2.424)$$

$$10 \text{ m} < d_{2D} < 2000 \text{ m}$$

$$h_b = 10 \text{ m}$$

$$1.5 \text{ m} \leq h_m \leq 22.5 \text{ m}$$

The shadow standard deviation is $\sigma_\Omega = 4 \text{ dB}$.

3D-UMi O-to-I

The microcell outdoor-to-indoor path loss is the same as the macrocell outdoor-to-indoor path loss $L_{\text{UMaO-to-I}} \text{ (dB)}$, except that $h_b = 10 \text{ m}$ instead of $h_b = 25 \text{ m}$. The shadow standard deviation remains at $\sigma_\Omega = 7 \text{ dB}$.

2.7.3.7 mm-Wave Path Loss Models

Next generation or 5G wireless systems will use frequencies from 600 MHz up to 100 GHz. Existing path loss models were developed for frequencies up to 6 GHz. For frequencies above 6 GHz, new path loss models are required for both LoS and NLoS environments.

Several models have been proposed for mm-wave path loss including the close-in (CI) path loss with a free-space reference distance, the close-in path loss model with a frequency dependent path loss exponent (CIF), and the Alpha-Beta-Gamma (ABG) path loss model, as described in [158].

The CI path loss model is defined by

$$L_{\text{CI}} \text{ (dB)} = L_{\text{FS}} \text{ (dB)}(f_c, 1 \text{ m}) + 10\beta \log_{10} \left(\frac{d}{1 \text{ m}} \right) \quad (2.425)$$

where $L_{\text{FS}} \text{ (dB)}(f_c, 1 \text{ m})$ represents the FSPL at a distance of 1 m, and f_c is the frequency in Hz. From (2.378), the FSPL at 1 m at frequency f_c is given by

$$L_{\text{FS}} \text{ (dB)}(f_c, 1 \text{ m}) = 10 \log_{10} \left\{ \left(\frac{4\pi f_c}{c} \right)^2 \right\} \quad (2.426)$$

The CI path loss model requires only the path loss exponent β , and ties the path loss to the FSPL at a distance of 1 m and frequency f_c .

The CIF model is an extension of the CI LoS path loss model that captures the frequency dependency of the path loss across a range of operating frequencies. This is particularly important for wideband mm-wave systems, where the operating bandwidth may be tens of Gigahertz.

$$L_{\text{CIF}} \text{ (dB)} = L_{\text{FS}} \text{ (dB)}(f_c, 1 \text{ m}) + 10\beta \left(1 + b \left(\frac{f_c - f_o}{f_o} \right) \right) \log_{10} \left(\frac{d}{1 \text{ m}} \right) \quad (2.427)$$

where β is again the path loss exponent, while b is an optimization parameter that captures the frequency dependency of the path loss that balances at centroid frequency f_o . The path loss increases with frequency when b is a positive value. When $b = 0$ or $f_c = f_o$, the CIF path loss model reduces to the CI path loss model. The centroid frequency is determined according to a weighted average of empirical data as

$$f_o = \frac{\sum_{k=1}^K f_k N_k}{\sum_{k=1}^K N_k} \quad (2.428)$$

where N_k is the number of path loss measurements at frequency f_k .

The ABG path loss model is similar to Lee's Area-to-Area path loss model in Sect. 2.7.3.2. The ABG path loss is given by

$$L_{\text{ABG}} \text{ (dB)} = \alpha_{\text{(dB)}} + 10\beta \log_{10}(d) + 10\gamma \log_{10}(f_c) \quad (2.429)$$

Table 2.5 Parameters for CI and ABG mm-wave path loss models, where SC stands for street canyon and OS stands for open square, from [158]

Scenario	CI model parameters	ABG model parameters
UMa-LoS	$\beta = 2.00, \sigma_{\Omega} = 4.1$ dB	N/A
UMa-NLoS	$\beta = 3.00, \sigma_{\Omega} = 6.8$ dB	$\alpha = 19.20, \beta = 3.40, \gamma = 2.30, \sigma_{\Omega} = 6.5$ dB
UMi-SC-LoS	$\beta = 1.98, \sigma_{\Omega} = 3.1$ dB	N/A
UMi-SC-NLoS	$\beta = 3.19, \sigma_{\Omega} = 8.2$ dB	$\alpha = 21.02, \beta = 3.48, \gamma = 2.34, \sigma_{\Omega} = 7.8$ dB
UMi-OS-LoS	$\beta = 1.85, \sigma_{\Omega} = 4.2$ dB	N/A
UMi-OS-NLoS	$\beta = 2.89, \sigma_{\Omega} = 7.1$ dB	$\alpha = 3.66, \beta = 4.14, \gamma = 2.43, \sigma_{\Omega} = 7.0$ dB

Table 2.6 Path loss exponents and shadow standard deviations for several different types of buildings, from [18]

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

where $\alpha_{(\text{dB})}$ is a floating offset value in dB units, and γ captures the frequency dependency of the model.

The usage of each path loss model will depend on the particular scenario. For outdoor mm-wave systems, two 3GPP models have been proposed [158]. The first applies to BS antenna heights of 25 m or less, and is called the urban microcell (UMi) model. The second applies to BS antenna heights of 25 m or more, and is called the urban macrocell (UMa) model. In each case, there are LoS and NLoS scenarios. Table 2.5 shows the parameters of the CI and ABG mm-wave path loss models for different environments.

2.7.3.8 Path Loss in Indoor Microcells

The path loss and shadowing characteristics for indoor microcells vary greatly from one building to the next. Typical path loss exponents and shadow standard deviations are provided in Table 2.6 for several different types of buildings.

For multistory buildings, the signal attenuation between floors is important. Measurements have shown that the greatest floor loss occurs when the transmitter and receiver are separated by a single floor. Typically, the floor loss is 15–20 dB for one floor and an additional 6–10 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 decibels 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, 896.5, 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 9–15 floors and then increases again [339]. 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 2A: COST 207 Channel Models

The COST 207 study has specified typical realizations for the power-delay profile in the following environments; Typical Urban (TU), Bad Urban (BA), Reduced TU, Reduced BU, Rural Area (RA), and Hilly Terrain (HT) [79]. The models below are identical to the COST 207 models, except that fractional powers have been normalized so as to sum to unity, i.e., the envelope power is normalized to unity.

Table 2.7 COST 207 Typical Urban (TU) ($\sigma_\tau = 1.0 \mu\text{s}$) and Bad Urban (BU) ($\sigma_\tau = 2.5 \mu\text{s}$) power-delay profiles, from [79]

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 2.8 COST 207 Reduced Typical Urban (TU) ($\sigma_\tau = 1.0 \mu\text{s}$) and Reduced Bad Urban (BU) ($\sigma_\tau = 2.5 \mu\text{s}$) power-delay profiles, from [79]

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 2.9 COST 207 Typical Rural (non-hilly) Area (RA) power-delay profile ($\sigma_\tau = 0.1 \mu\text{s}$), from [79]

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 2.10 COST 207 Typical Hilly Terrain (HT) power-delay profile ($\sigma_\tau = 5.0 \mu\text{s}$), from [79]

Delay μs	Fractional power	Doppler category
0.0	0.026	CLASS
0.1	0.042	CLASS
0.3	0.066	CLASS
0.5	0.105	CLASS
0.7	0.263	GAUS1
1.0	0.263	GAUS1
1.3	0.105	GAUS1
15.0	0.042	GAUS2
15.2	0.034	GAUS2
15.7	0.026	GAUS2
17.2	0.016	GAUS2
20.0	0.011	GAUS2

Table 2.11 COST 207 Reduced Hilly Terrain (HT) power-delay profile ($\sigma_\tau = 5.0 \mu\text{s}$), from [79]

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

Appendix 2B: COST 259 Channel Models

The 3GPP standards group has defined three typical realizations for the COST 259 models; Typical Urban (TU x), Rural Area (RA x), and Hilly Terrain (HT x), where x is the MS speed in km/h, [113]. Default speeds are 3, 50, and 120 km/h for the TU x model, 120 and 250 km/h for the RA x model, and 120 km/h for the HT x model.

Table 2.12 COST 259 Typical Urban (TU x) channel model, from [113]

Delay μs	Fractional power	Doppler category
0.000	0.26915	CLASS
0.217	0.17378	CLASS
0.512	0.09772	CLASS
0.514	0.09550	CLASS
0.517	0.09550	CLASS
0.674	0.07079	CLASS
0.882	0.04571	CLASS
1.230	0.02344	CLASS
1.287	0.02042	CLASS
1.311	0.01950	CLASS
1.349	0.01820	CLASS
1.533	0.01259	CLASS
1.535	0.01259	CLASS
1.622	0.01047	CLASS
1.818	0.00708	CLASS
1.836	0.00692	CLASS
1.884	0.00617	CLASS
1.943	0.00550	CLASS
2.048	0.00447	CLASS
2.140	0.00372	CLASS

Table 2.13 COST 259 Rural Area (RA x) channel model, from [113]

Delay μs	Fractional power	Doppler category
0.000	0.30200	Direct path, $f_0 = 0.7f_m$
0.042	0.22909	CLASS
0.101	0.14454	CLASS
0.129	0.11749	CLASS
0.149	0.10000	CLASS
0.245	0.04898	CLASS
0.312	0.02951	CLASS
0.410	0.01413	CLASS
0.469	0.00912	CLASS
0.528	0.00575	CLASS

Table 2.14 COST 259 Hilly Terrain (HTx) channel model, from [113]

Delay μ s	Fractional power	Doppler category
0.000	0.43652	CLASS
0.356	0.12882	CLASS
0.441	0.09550	CLASS
0.528	0.07079	CLASS
0.546	0.06607	CLASS
0.609	0.05370	CLASS
0.625	0.05012	CLASS
0.842	0.02399	CLASS
0.916	0.01862	CLASS
0.941	0.01698	CLASS
15.000	0.01738	CLASS
16.172	0.00537	CLASS
16.492	0.00389	CLASS
16.876	0.00263	CLASS
16.882	0.00263	CLASS
16.978	0.00240	CLASS
17.615	0.00126	CLASS
17.827	0.00102	CLASS
17.849	0.00100	CLASS
18.016	0.00085	CLASS

Appendix 2C: ITU Channel Models

ITU models have been developed indoor office, outdoor to indoor and pedestrian, and vehicular-high antenna [276]. The models below are identical to the ITU models, except that fractional powers have been normalized so as to sum to unity, i.e., the envelope power is normalized to unity.

Table 2.15 ITU indoor office environment power-delay profiles, from [276]

Channel A			Channel B		
Delay ns	Fractional power	Doppler category	Delay ns	Fractional power	Doppler category
0	0.61722	FLAT	0	0.57833	FLAT
50	0.30934	FLAT	100	0.25245	FLAT
110	0.06172	FLAT	200	0.11020	FLAT
170	0.00978	FLAT	300	0.04811	FLAT
290	0.00155	FLAT	500	0.00917	FLAT
310	0.00039	FLAT	700	0.00175	FLAT

Table 2.16 ITU outdoor to indoor and pedestrian power-delay profiles, from [276]

Channel A			Channel B		
Delay ns	Fractional power	Doppler category	Delay ns	Fractional power	Doppler category
0	0.88935	CLASSIC	0	0.40569	CLASSIC
110	0.09529	CLASSIC	200	0.32976	CLASSIC
190	0.01069	CLASSIC	800	0.13128	CLASSIC
410	0.00467	CLASSIC	1200	0.06430	CLASSIC
			2300	0.06733	CLASSIC
			3700	0.00165	CLASSIC

Table 2.17 ITU vehicular power-delay profiles, from [276]

Channel A			Channel B		
Delay ns	Fractional power	Doppler category	Delay ns	Fractional power	Doppler category
0	0.48500	CLASSIC	0	0.34020	CLASSIC
310	0.38525	CLASSIC	300	0.60498	CLASSIC
710	0.06106	CLASSIC	8900	0.03175	CLASSIC
1090	0.04850	CLASSIC	12,900	0.00605	CLASSIC
1730	0.01534	CLASSIC	17,100	0.00183	CLASSIC
2510	0.00485	CLASSIC	20,000	0.01520	CLASSIC

Appendix 2D: Derivation of Eq. (2.358)

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, cf., (2.57)

$$p_X(x) = \frac{x}{b_0} \exp\left\{-\frac{x^2 + s^2}{2b_0}\right\} I_0\left(\frac{xs}{b_0}\right) \quad x > 0 \quad (2D.1)$$

and moments [272]

$$E[X^n] = (2b_0)^{\frac{n}{2}} \exp\left\{-\frac{s^2}{2b_0}\right\} \Gamma\left((2+n)/2\right) {}_1F_1\left(\frac{n+2}{2}, 1; \frac{s^2}{2b_0}\right) \quad (2D.2)$$

where $\Gamma(\cdot)$ is the Gamma function, and ${}_1F_1(a, b; x)$ is the confluent hypergeometric function. The first moment of X is

$$E[X] \equiv \Omega_v = (2b_0)^{\frac{1}{2}} e^{-K} \frac{\sqrt{\pi}}{2} {}_1F_1(3/2, 1; K), \quad (2D.3)$$

where $K = s^2/2b_0$ is the Rice factor. The second moment of X is

$$\begin{aligned} E[X^2] \equiv \Omega_p &= 2b_0 e^{-K} {}_1F_1(2, 1; K) \\ &= 2b_0(K + 1). \end{aligned} \quad (2D.4)$$

Substituting $2b_0$ from (2D.3) into (2D.4) gives

$$\Omega_p = \frac{4e^{2K}(K+1)}{\pi {}_1F_1^2(3/2, 1; K)} \Omega_v^2 = C(K) \Omega_v^2. \quad (2D.5)$$

Note that $C(0) = 4/\pi$, $C(\infty) = 1$, and $4/\pi \leq C(K) \leq 1$ for $0 \leq K \leq \infty$.

Appendix 2E: Derivation of Eq. (2.374)

From (2.373), the composite distribution for the squared-envelope, α_c^2 , is

$$p_{\alpha_c^2}(x) = \int_0^\infty \left(\frac{m}{w}\right)^m \frac{x^{m-1}}{\Gamma(m)} \exp\left\{-\frac{mx}{w}\right\} \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega w} \exp\left\{-\frac{(10\log_{10}\{w\} - \mu_{\Omega_p}(\text{dB}))^2}{2\sigma_\Omega^2}\right\} dw, \quad (2E.1)$$

where $\xi = \ln(10)/10$. The mean of the approximate log-normal distribution is

$$\begin{aligned}\mu_{(\text{dBm})} &= E[10\log_{10}\{\alpha_c^2\}] \\ &= \int_0^\infty \int_0^\infty 10\log_{10}\{x\} \left(\frac{m}{w}\right)^m \frac{x^{m-1}}{\Gamma(m)} \exp\left\{-\frac{mx}{w}\right\} \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega w} \exp\left\{-\frac{(10\log_{10}\{w\} - \mu_{\Omega_p(\text{dB})})^2}{2\sigma_\Omega^2}\right\} dw dx \\ &= \frac{10m^m}{\sqrt{2\pi}\xi\sigma_\Omega\Gamma(m)} \int_0^\infty \frac{1}{w^{m+1}} \exp\left\{-\frac{(10\log_{10}\{w\} - \mu_{\Omega_p(\text{dB})})^2}{2\sigma_\Omega^2}\right\} \int_0^\infty \log_{10}\{x\} x^{m-1} \exp\left\{-\frac{mx}{w}\right\} dx dw.\end{aligned}\quad (2E.2)$$

Assuming that m is an integer, the inner integral becomes [147, 4.352.2]

$$\int_0^\infty \log_{10}\{x\} x^{m-1} \exp\left\{-\frac{mx}{w}\right\} dx = \frac{\Gamma(m)w^m}{m^m \ln 10} (\psi(m) - \ln(m/w)). \quad (2E.3)$$

Then by using the change of variables $x = 10\log_{10}\{w\}$,

$$\mu_{(\text{dBm})} = \xi^{-1} (\psi(m) - \ln(m)) + \mu_{\Omega_p(\text{dB})}, \quad (2E.4)$$

where $\psi(\cdot)$ is the Euler psi function, and

$$\psi(m) = -C + \sum_{k=1}^{m-1} \frac{1}{k} \quad (2E.5)$$

and $C \simeq 0.5772$ is Euler's constant. Likewise, the second moment of the approximate log-normal distribution is

$$\begin{aligned}E[(10\log_{10}\{\alpha_c^2\})^2] &= \int_0^\infty \int_0^\infty (10\log_{10}\{x\})^2 \left(\frac{m}{w}\right)^m \frac{x^{m-1}}{\Gamma(m)} \exp\left\{-\frac{mx}{w}\right\} \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega w} \exp\left\{-\frac{(10\log_{10}\{w\} - \mu_{\Omega_p(\text{dB})})^2}{2\sigma_\Omega^2}\right\} dw dx \\ &= \frac{m^m}{\sqrt{2\pi}\xi\Gamma(m)} \int_0^\infty \frac{1}{w^{m+1}} \exp\left\{-\frac{(10\log_{10}\{w\} - \mu_{\Omega_p(\text{dB})})^2}{2\sigma_\Omega^2}\right\} \int_0^\infty (10\log_{10}\{x\})^2 x^{m-1} \exp\left\{-\frac{mx}{w}\right\} dx dw.\end{aligned}\quad (2E.6)$$

Assuming again that m is an integer, the inner integral is [147, 4.358.2]

$$\int_0^\infty (10\log_{10}(x))^2 x^{m-1} \exp\left\{-\frac{mx}{w}\right\} dx = \frac{(m-1)!w^m}{m^m \ln 10} \left((\psi(m) - \ln(m/w))^2 + \zeta(2, m) \right)$$

leading to

$$E[(10\log_{10}\{\alpha_c^2\})^2] = \zeta^2 \left((\psi(m) - \ln(m))^2 \mu_{\Omega_p(\text{dB})}^2 + \zeta(2, m) \right) + 2\zeta (\psi(m) - \ln(m)) \mu_{\Omega_p(\text{dB})} + \sigma^2 + \mu_{\Omega_p(\text{dB})}^2, \quad (2E.7)$$

where

$$\zeta(2, m) = \sum_{k=0}^{\infty} \frac{1}{(m+k)^2} \quad (2E.8)$$

is Reimann's zeta function. Finally, the variance of the approximate log-normal distribution is

$$\begin{aligned}\sigma^2 &= E[(10\log_{10}\{\alpha_c^2\})^2] - E^2[10\log_{10}\{\alpha_c^2\}] \\ &= \xi^{-2}\zeta(2, m) + \sigma_\Omega^2.\end{aligned}\quad (2E.9)$$

Problems

2.1. Suppose that $r(t)$ is a wide-sense stationary (WSS) bandpass random process, such that

$$r(t) = g_I(t) \cos(2\pi f_c t) - g_Q(t) \sin(2\pi f_c t)$$

(a) Show that the auto- and cross-correlations of $g_I(t)$ and $g_Q(t)$ must satisfy the following conditions:

$$\begin{aligned}\phi_{g_I g_I}(\tau) &= \phi_{g_Q g_Q}(\tau) \\ \phi_{g_I g_Q}(\tau) &= -\phi_{g_Q g_I}(\tau)\end{aligned}$$

(b) Under the conditions in part a) show that the autocorrelation of $r(t)$ is

$$E[r(t)r(t + \tau)] = \phi_{g_I g_I}(\tau) \cos(2\pi f_c \tau) - \phi_{g_I g_Q}(\tau) \sin(2\pi f_c \tau).$$

2.2. What is the maximum Doppler shift for the GSM mobile cellular system on the “downlink” from the base station to the mobile unit (935–960 MHz RF band)? What is it on the “uplink” direction, or mobile to base (890–915 MHz RF band)? Assume a high-speed train traveling at a speed of $v = 250$ km/h.

2.3. This problem considers two ray channels exhibiting either frequency selective or time selective behavior.

(a) Consider the transmission of a bandpass signal having complex envelope $\tilde{s}(t)$ on a channel such that the received complex envelope is

$$\tilde{r}(t) = \alpha \tilde{s}(t) + \beta \tilde{s}(t - \tau_1),$$

where α and β are real valued.

- i) Find the channel impulse response $g(t, \tau)$.
- ii) Find the channel magnitude response $|G(t, f)|$.
- iii) Find the channel phase response $\angle G(t, f)$.

(b) Consider the transmission of a bandpass signal having complex envelope $\tilde{s}(t)$ on a channel such that the received complex envelope is

$$\tilde{r}(t) = \alpha \tilde{s}(t) + \beta \tilde{s}(t) e^{j2\pi f_d t},$$

where α and β are real valued.

- i) Find the channel impulse response $g(t, \tau)$.
- ii) Find the channel magnitude response $|G(t, f)|$.
- iii) Find the channel phase response $\angle G(t, f)$.

2.4. A wireless channel is characterized by the time-variant impulse response

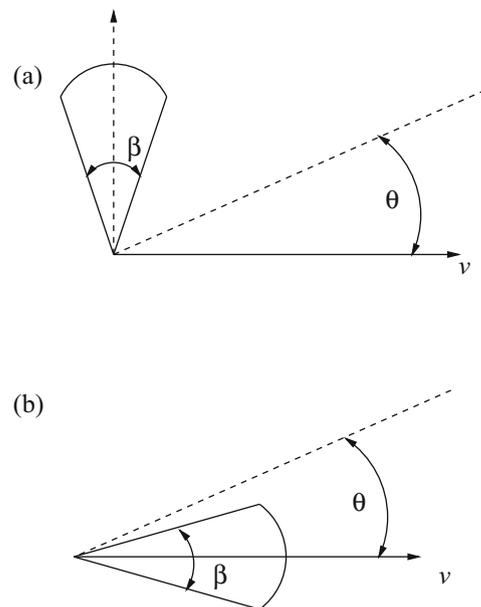
$$g(t, \tau) = \left(1 - \frac{\tau}{T}\right) \cos(\Omega t + \phi_0), \quad 0 \leq \tau \leq T,$$

where $T = 0.05$ ms, $\Omega = 10\pi$, and $\phi_0 \in (-\pi, +\pi]$ is a constant.

- (a) Determine the channel time-variant transfer function.
- (b) Given an input signal having the complex envelope

$$\tilde{s}(t) = \begin{cases} 1, & 0 \leq t \leq T_s \\ 0, & \text{otherwise} \end{cases},$$

Fig. 2.70 Scenario for Problem 2.7 parts (a) and (b)



determine the complex envelope of the signal at the output of the channel, $\tilde{r}(t)$. Make sure to consider cases when $0 < T_s < T$ and $0 < T < T_s$, separately.

(c) Consider digital modulation scheme with a modulated symbol interval T_s . If the channel fading is frequency selective, specify the relation between T_s and T .

2.5. Suppose that an omnidirectional antenna is used and the azimuth angle-of-arrival distribution, $p(\theta)$, is given by (2.52). Find the Doppler power spectrum $S_{gg}(f)$.

2.6. A very useful model for a non-isotropic scattering environment assumes that the azimuth angle-of-arrival distribution is described by the von Mises pdf in (2.51).

(a) Assuming an isotropic receiver antenna, calculate the received Doppler power spectrum, $S_{gg}(f)$.

(b) Under what conditions are the quadrature components $g_I(t)$ and $g_Q(t)$ uncorrelated?

2.7. Determine and plot the (normalized) power spectral densities $S_{gg}(f)$ 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. 2.70a)

$$G(\theta) = \begin{cases} G_0, & |\frac{\pi}{2} - \theta| < \beta/2 \\ 0, & \text{otherwise} \end{cases}.$$

(d) A directional antenna of beamwidth β directed along vehicle motion with (see Fig. 2.70b)

$$G(\theta) = \begin{cases} G_0, & |\theta| < \beta/2 \\ 0, & \text{otherwise} \end{cases}.$$

2.8. Consider a narrow-band channel with a 700 MHz carrier frequency. The complex channel gain at a mobile station is $g(t) = g_I(t) + jg_Q(t)$, such that

$$S_{g_I g_I}(f) = \begin{cases} \text{rect}\left(\frac{f}{200}\right), & |f| \leq 100 \text{ Hz} \\ 0, & \text{elsewhere} \end{cases}$$

$$S_{g_I g_Q}(f) = 0.$$

- (a) What is the speed of the mobile station?
 (b) What is the cross-correlation function $\phi_{g_I g_Q}(\tau)$ of the I and Q components of the faded envelope?
 (c) If antenna diversity is deployed at the mobile station, what are the possible spatial separations between the antenna elements such that the corresponding faded envelopes will be uncorrelated?
 (d) Write down an expression for $\phi_{g_Q g_Q}(\tau)$.

2.9. Consider a Ricean fading channel with Rice factor K and average envelope power Ω_p . Assume that the means $m_I(t)$ and $m_Q(t)$ of the in-phase and quadrature components are given by (2.59) and (2.60), respectively. Derive an integral expression for the probability density function of the envelope phase in terms of K and Ω_p .

2.10. Consider a 2-D isotropic scattering channel. Show that the psd of the received envelope $\alpha(t) = |g(t)|$ is given by (2.77).

2.11. Suppose that the Doppler spectrum is given by

$$S_{gg}(f) = A \text{rect}\left(\frac{f + f_1}{W_d}\right) + A \text{rect}\left(\frac{f - f_1}{W_d}\right)$$

where

$$\text{rect}\left(\frac{f}{W_d}\right) = \begin{cases} 1, & |f| \leq W_d/2 \\ 0, & \text{elsewhere} \end{cases}$$

and A, f_1 , and W_d are constants.

- (a) Sketch the Doppler spectrum.
 (b) Find the envelope correlation function

$$\phi_{gg}(\tau) = \phi_{g_I g_I}(\tau) + j\phi_{g_I g_Q}(\tau)$$

- (c) For which values of τ are $g(t)$ and $g(t + \tau)$ uncorrelated?

2.12. Suppose that the Doppler power spectrum is given by the following function:

$$S_{gg}(f) = A \cos\left(\frac{\pi f}{2f_m}\right) \text{rect}\left(\frac{f}{2f_m}\right)$$

- (a) Find the corresponding envelope autocorrelation function $\phi_{gg}(\tau)$.
 (b) For what values of τ are $g_I(t)$ and $g_Q(t + \tau)$ uncorrelated?
 (c) Given that

$$\Omega_p = \phi_{gg}(0) = \int_{-f_m}^{f_m} S_{gg}(f) df$$

find the value of A in terms of Ω_p .

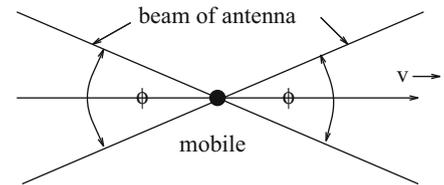
2.13. Consider the non-isotropic scattering environment shown in Fig. 2.7. Show that the continuous portion of the psd of the received envelope $\alpha(t) = |g(t)|$ is given by (2.79).

2.14. 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 (2.82) and (2.83), respectively.

2.15. Consider a wide-sense stationary non-zero-mean complex Gaussian random process $g(t) = g_I(t) + jg_Q(t)$, where

$$\begin{aligned} g_I(t) &= \hat{g}_I(t) + m_I(t) \\ g_Q(t) &= \hat{g}_Q(t) + m_Q(t) \end{aligned}$$

Fig. 2.71 Mobile with directional antenna for Problem 2.19



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 (2.87) and (2.90), respectively.

2.16. Establish the equivalence between (2.102) and (2.103).

2.17. A flat Rayleigh fading signal at 6 GHz is received by a vehicle traveling at 80 km/hr.

- Determine the number of positive-slope zero crossings of the rms envelope level that occur over a 5 s interval.
- Determine the average duration of a fade below the rms envelope level.
- Determine the average duration of a fade at a level of 20 dB below the rms envelope level.

2.18. 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

$$\rho = \frac{R}{\sqrt{\Omega_p}} = \frac{R}{\sqrt{2b_0}}.$$

and the b_i are defined in (2.102) with $f_q = 0$.

2.19. Consider the situation in Fig. 2.71, where the mobile station employs a directional antenna with a beam width of ϕ° . Assume a 2-D isotropic scattering environment.

- In receiving a radio transmission at 850 MHz, a Doppler frequency of 20–60 Hz is observed. What is the beam width of the mobile station antenna, and how fast is the mobile station traveling?
- Suppose that the mobile station antenna has a beam width of 13° . What is the level crossing rate with respect to the rms envelope level, assuming that the mobile station is traveling at a speed of 30 km/h?

2.20. Consider a cellular radio system with fixed base stations and moving mobile stations. The channel is characterized by flat Rayleigh fading channel with two-dimensional isotropic scattering. The mobile station employs omni-directional antennas and the system operates at an RF carrier frequency of 900 MHz.

- Determine the positive going level crossing rate at the normalized envelope level $\rho = 1$, when the maximum Doppler frequency is $f_m = 20$ Hz. Compute the velocity of the mobile station.
- Now suppose that the mobile station is travelling at a speed of 24 km/h. Calculate the average fade duration (AFD) and level crossing rate (LCR) at the normalized envelope level $\rho = 0.294$.

2.21. A vehicle experiences 2-D isotropic scattering and receives a Rayleigh faded 900 MHz signal while traveling at a constant velocity for 10 s. Assume that the local mean remains constant during travel, and 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?

2.22. 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.

- 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 with probability one-half. Assuming that the data is transmitted in 10,000-bit packets, how many bits errors (on the average) will each transmitted packet contain?

2.23. Show that for wide-sense stationary (WSS) channels

$$\begin{aligned}\phi_H(f, m; \nu, \mu) &= \psi_H(f, m; \nu)\delta(\nu - \mu) \\ \phi_S(\tau, \eta; \nu, \mu) &= \psi_S(\tau, \eta; \nu)\delta(\nu - \mu).\end{aligned}$$

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?

2.24. Show that for uncorrelated scattering (US) channels

$$\begin{aligned}\phi_g(t, s; \tau, \eta) &= \psi_g(t, s; \tau)\delta(\eta - \tau) \\ \phi_S(\tau, \eta; \nu, \mu) &= \psi_S(\tau; \nu, \mu)\delta(\eta - \tau).\end{aligned}$$

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?

2.25. Given the channel input signal $\tilde{s}(t)$ and the channel delay-Doppler spread function $S(\tau, \nu)$, show that the channel output signal is

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

How do you interpret the channel function $S(\tau, \nu)$?

2.26. Suppose that the spaced-time spaced-frequency correlation function of a WSSUS channel has the following form

$$\phi_T(\Delta f; \Delta t) = \exp\{-b|\Delta t|\} \frac{1}{a + j2\pi\Delta f}.$$

- Find the corresponding channel correlation function $\psi_g(\Delta t; \tau)$.
- Find the corresponding scattering function $\psi_S(\nu; \tau)$.
- What is the average delay spread, μ_τ , and rms delay spread σ_τ ?

2.27. The scattering function for a WSSUS scattering channel is given by

$$\psi_S(\tau, \nu) = \Omega_p \cdot \frac{2a}{a^2 + (2\pi\nu)^2} \cdot be^{-b\tau} u(\tau).$$

- What is the spaced-time spaced-frequency correlation function?
- What is the average delay?
- What is the rms delay spread?

2.28. The scattering function for a WSSUS channel is given by

$$\psi_S(\tau, \nu) = \begin{cases} \frac{\Omega_p}{100} be^{-b\tau}, & 0 \leq |\nu| \leq 50 \text{ Hz}, 0 \leq \tau \leq \infty \\ 0, & \text{elsewhere} \end{cases}$$

- What is the speed of the mobile station?
- What is the channel correlation function $\psi_g(\Delta t; \tau)$?

- (c) If the faded envelope is sampled, $g_k = g(kT_s)$, what sample spacings T_s will yield uncorrelated samples, g_k ?
 (d) What is the envelope level crossing rate, L_R ?

2.29. The frequency correlation function of a channel is defined as

$$\phi_T(\Delta_f) \triangleq \phi_T(\Delta_f, \Delta_t = 0)$$

where $\phi_T(\Delta_f, \Delta_t)$ is the spaced-time, spaced-frequency correlation function. Suppose that

$$\phi_T(\Delta_f) = \frac{A}{1 + j2\pi\Delta_f/f_o}$$

- (a) Find the power-delay profile of the channel $\psi_g(\tau)$.
 (b) Find the average delay μ_τ of the channel.

2.30. Consider a linear time-invariant channel consisting of two equal rays

$$g(t, \tau) = \delta(\tau) + \delta(\tau - \tau_1).$$

- (a) Derive an expression for magnitude response of the channel $|T(f, t)|$ and sketch showing all important points.
 (b) Repeat for the phase response of the channel $\angle T(f, t)$.

2.31. Consider a linear time-invariant channel having the impulse response

$$g(t, \tau) = \delta(\tau) + 2\delta(\tau - \tau_1) + \delta(\tau - 2\tau_1).$$

- (a) Derive a closed-form expression for magnitude response of the channel $|T(f, t)|$ and sketch showing all important points.
 (b) Repeat part a) for the phase response of the channel $\angle T(f, t)$.
 (c) What is the mean delay and rms delay spread of the channel.

2.32. Consider a linear time-invariant channel having the impulse response

$$g(t, \tau) = \frac{1}{\sqrt{L}} \sum_{k=1}^L \delta(\tau - (k-1)\Delta_\tau)$$

- (a) Find the magnitude response $|T(f)|$ and phase response $\angle T(f)$ of the channel, where $T(f)$ is the time-invariant transfer function of the channel. Simplify your expressions as much as possible. Plot $|T(f)|$ and $\angle T(f)$.
 (b) What is the mean delay μ_τ and rms delay spread σ_τ of this channel? *The following may be useful:*

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}$$

$$\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_{k=0}^n k^3 = \frac{n^2(n+1)^2}{4}$$

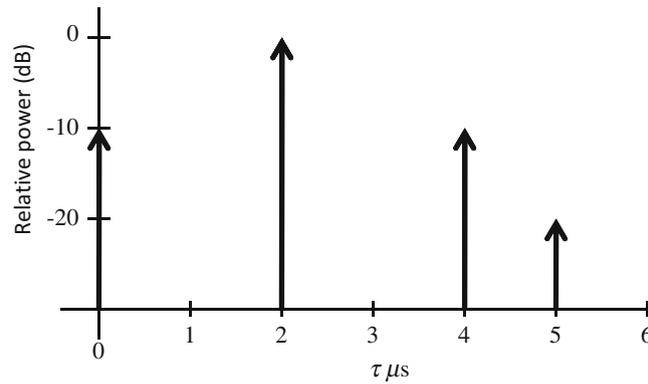


Fig. 2.72 Power-delay profile for Problem 2.35

2.33. The power-delay profile of a WSSUS channel is given by

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

Assuming that $T = 10 \mu\text{s}$, determine

- the average delay
- the rms delay spread
- the coherence bandwidth of the channel.

2.34. The power-delay profile of a WSSUS channel is given by

$$\psi_g(\tau) = \begin{cases} 0.5[1 + \cos(2\pi\tau/T)], & 0 \leq \tau \leq T/2 \\ 0, & \text{otherwise} \end{cases}.$$

- Find the channel frequency correlation function.
- Calculate the mean delay, rms delay spread, and the coherence bandwidth.
- If $T = 0.1 \text{ ms}$, determine whether the channel exhibits frequency-selective fading to a GSM cellular system.

2.35. Consider the power-delay profile shown in Fig. 2.72. Calculate the following:

- mean delay
- rms delay spread
- If the modulated symbol duration is $40 \mu\text{s}$, is the channel frequency selective? Why?

2.36. The power-delay profile in Fig. 2.73 is observed for a multipath-fading channel in hilly terrain.

- Compute the mean delay.
- Compute the rms delay spread.
- What is the frequency correlation function of the channel?

2.37. Consider a WSSUS channel with scattering function

$$\psi_s(\tau, \nu) = \psi_1(\tau) \cdot \psi_2(\nu)$$

where

$$\psi_1(\tau) = \begin{cases} 1, & 0 \leq \tau \leq 100 \text{ ms} \\ 0, & \text{otherwise} \end{cases}$$

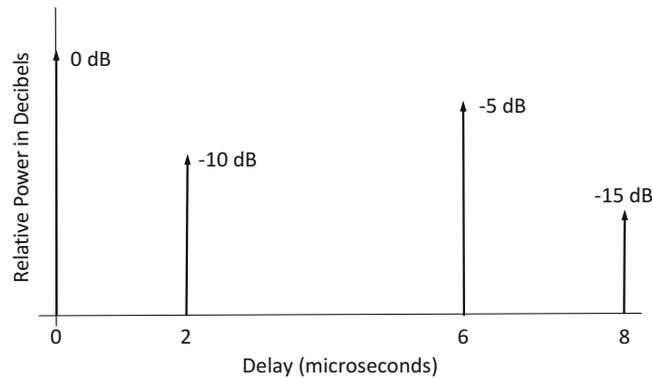


Fig. 2.73 Power-delay profile for Problem 2.36

$$\psi_2(\nu) = \begin{cases} \frac{1}{f_m} [1 - (\nu/f_m)^2], & 0 \leq |\nu| \leq f_m \\ 0, & \text{otherwise} \end{cases}$$

Assume $f_m = 10$ Hz. Find

- the power-delay profile.
- the Doppler power spectrum.
- the mean delay and the rms delay spread.
- the maximum Doppler frequency, the mean Doppler frequency, and the rms Doppler frequency.
- the coherence bandwidth and the coherence time of the channel.

2.38. Consider the COST-207 typical urban (TU) and bad urban (BU) power-delay profiles shown in Fig. 2.51 of the text with delays and fractional powers given in Table 2.7.

- Calculate the average delay, μ_τ .
- Calculate the rms delay spread, σ_τ .
- Calculate the approximate values of W_{50} and W_{90} .
- If the channel is to be used with a modulation scheme 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.

2.39. The scattering function $\psi_S(\tau, \nu)$ for a multipath-fading channel is non-zero for the range of values $0 \leq \tau \leq 1 \mu\text{s}$ and $-40 \leq \nu \leq 40$ Hz. Furthermore, $\psi_S(\tau, \nu)$ is uniform in the two variables τ and ν .

- Find numerical values for the following parameters;
 - the average delay, μ_τ , and rms delay spread, σ_τ .
 - the rms Doppler spread, B_d
 - the approximate coherence time, T_c
 - the approximate coherence bandwidth, B_c
- Given the answers in part a), what does it mean when the channel is
 - frequency-nonsselective
 - slowly fading
 - frequency-selective

2.40. The scattering function $\psi_S(\tau, \nu)$ for a multipath-fading channel is non-zero for the range of values $0 \leq \tau \leq 1 \mu\text{s}$ and $-40 \leq \nu \leq 40$ Hz. Assume that the scattering function is uniform in the two variables τ and ν with a value equal $A = 0.00125$.

- What is the average delay and rms delay spread of the channel?
- Find the spaced-time spaced-frequency correlation function $\psi_T(\Delta f, \Delta t)$ of the channel.
- What is the total envelope power? Express your answer in dBm units.

- (d) If the scattering function describes a conventional fixed-to-mobile cellular land mobile radio channel, and a carrier frequency of 900 MHz is used with an isotropic receive antenna, how fast is the mobile station moving?

2.41. Suppose that the Doppler spectrum is given by

$$S_{gg}(f) = G(A, 0.5f_m, 0.1f_m)$$

where

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

- (a) Sketch the Doppler spectrum.
 (b) Find the envelope correlation function

$$\phi_{gg}(\tau) = \phi_{g_I g_I}(\tau) + j\phi_{g_I g_Q}(\tau)$$

- (c) For which values of τ are $g_I(t)$ and $g_Q(t + \tau)$ uncorrelated?

2.42. Suppose that a fading simulator is constructed by low-pass filtering white Gaussian noise as shown in Fig. 2.35. Assume that the white Gaussian noise generators that produce $g_I(t)$ and $g_Q(t)$ are uncorrelated, and have power density spectrum $\Omega_p/2$ watts/Hz. The low-pass filters that are employed have the transfer function

$$H(f) = \frac{A}{1 + j2\pi\beta f}.$$

- (a) What is the Doppler power spectrum $S_{gg}(f)$ and autocorrelation function $\phi_{gg}(\tau)$?
 (b) Find A such that the envelope power is equal to Ω_p .
 (c) What is the joint probability density function of the output $g(t)$ and $g(t + \tau)$?

2.43. Suppose that a fading simulator is constructed using low-pass filtered white Gaussian noise as shown in Fig. 2.35. Assume that the white Gaussian noise generators used to produce $g_I(t)$ and $g_Q(t)$ are uncorrelated. The low-pass filters that are employed have the transfer function

$$H(f) = \frac{1}{\sqrt{B}} \text{rect} \left(\frac{f}{B} \right).$$

- (a) What is the Doppler power spectrum $S_{gg}(f)$?
 (b) For the $S_{gg}(f)$ in part (a), derive an expression for the envelope level crossing rate.

2.44. Consider Jakes' method in (2.254) and (2.255).

- (a) With the choice that $\alpha = 0$ and $\beta_n = \pi n/(M + 1)$ show that

$$\begin{aligned} \langle g_I(t)g_Q(t) \rangle &= 0 \\ \langle g_Q^2(t) \rangle &= (M + 1)/(2M + 1) \\ \langle g_I^2(t) \rangle &= M/(2M + 1) \end{aligned}$$

- (b) Rederive the time averages in part a) for the choice $\alpha = 0$ and $\beta_n = \pi n/M$.

2.45. (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.

Note that Jakes' simulator is non-stationary as shown in (2.260). Therefore, you may not necessary get a plot that is identical to Fig. 2.38. In fact, it would be good if you could observe the non-stationary behavior of the simulator, i.e., the pdf of the envelope distribution changes with time.

2.46. (Computer exercise) In this problem you are to generate a Ricean faded envelope $\hat{g}(t) = \hat{g}_I(t) + j\hat{g}_Q(t)$ by making appropriate modifications to Jakes' method such that

$$\begin{aligned}\hat{g}_I(t) &= m_I(t) + g_I(t) \\ \hat{g}_Q(t) &= m_Q(t) + g_Q(t),\end{aligned}$$

where $g_I(t)$ and $g_Q(t)$ are defined in (2.254) and (2.255), respectively. Assume that the means $m_I(t)$ and $m_Q(t)$ are generated according to Aulin's model in (2.59) and (2.60). For $f_m T = 0.1$, $\Omega_p = 1$ and $K = 0, 4, 7$, and 16, plot the following:

- (a) The envelope $\hat{\alpha}(t) = \sqrt{\hat{g}_I^2(t) + \hat{g}_Q^2(t)}$.
 (b) The wrapped phase $\phi(t) = \text{Tan}^{-1}(\hat{g}_Q(t)/\hat{g}_I(t))$, mod 2π .

2.47. (Computer exercise) This problem uses the fading simulator developed in Problem 2.46. The objective is to compute an estimate of the mean-squared-envelope $\Omega_p = E[\hat{\alpha}^2(t)]$ from samples of $\hat{g}_I(kT)$ and $\hat{g}_Q(kT)$, where T is the sample spacing in seconds. The estimate is computed by forming the empirical average

$$\hat{\Omega}_p = \frac{1}{N} \sum_{i=1}^N (\hat{g}_I^2(iT) + \hat{g}_Q^2(iT)),$$

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

$$\frac{d}{\lambda_c} = Nf_m T.$$

- (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 that estimates of the local mean Ω_p are used in resource management algorithms such as handoff algorithms.

2.48. Suppose that two complex faded envelopes $g_i(t) = g_{I,i}(t) + jg_{Q,i}(t)$, $i = 1, 2$ are available, such that

$$\begin{aligned}\mu_{\mathbf{g}} &= E[\mathbf{g}(t)] = \mathbf{0} \\ \Phi_{\mathbf{g}}(\tau) &= \frac{1}{2} E[\mathbf{g}(t)\mathbf{g}^H(t + \tau)] = \mathbf{\Omega} J_o(2\pi f_m \tau),\end{aligned}$$

where

$$\begin{aligned}\mathbf{g}(t) &= (g_1(t), g_2(t)) \\ \mathbf{\Omega} &= \begin{bmatrix} \Omega_1 & 0 \\ 0 & \Omega_2 \end{bmatrix}.\end{aligned}$$

A third faded envelope $g_3(t)$ is now generated that is correlated with $g_1(t)$ and $g_2(t)$ according to

$$g_3(t) = \alpha g_1(t) + (1 - \alpha)g_2(t), \quad 0 \leq \alpha \leq 1.$$

(a) Compute the values of

$$\phi_{g_1 g_3}(\tau) = \frac{1}{2} \text{E}[g_1^*(t) g_3(t + \tau)]$$

$$\phi_{g_2 g_3}(\tau) = \frac{1}{2} \text{E}[g_2^*(t) g_3(t + \tau)]$$

$$\phi_{g_3 g_3}(\tau) = \frac{1}{2} \text{E}[g_3^*(t) g_3(t + \tau)].$$

(b) What is the envelope distribution of $g_3(t)$?

2.49. Suppose that the two $\tau = T/4$ spaced taps in Example 2.1 do not have equal magnitude. In particular, suppose that $|g_0(t)|^2 = |g_1(t)|^2/2$. Once again, a T -spaced channel model is to be generated such that the two T -spaced taps capture the maximum possible total energy.

(a) Find the optimal sampler timing instant.

(b) Determine the corresponding matrix \mathbf{A} for $\beta = 0.35$.

2.50. By starting with the Gaussian density for the local mean envelope power in (2.357) derive the log-normal density in (2.355).

2.51. One simple model for shadow simulation is to model log-normal shadowing as a Gaussian white noise process that is filtered by a first-order low-pass filter. With this model

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

where $\Omega_k \text{ (dBm)}$ is the mean envelope or mean squared-envelope, expressed in dBm units, that is experienced at *spatial* index k , ζ is a parameter that controls the spatial correlation of the shadows, and v_k is a zero-mean Gaussian random variable with $\phi_{v_v}(n) = \sigma^2 \delta(n)$.

(a) Show that the resulting spatial autocorrelation function of $\Omega_k \text{ (dBm)}$ is

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

(b) What is the mean and variance of $\Omega_k \text{ (dBm)}$ at any spatial index k ?

2.52. (computer exercise) The purpose of this problem is to generate variations in the local mean Ω_p due to shadowing. The shadows are generated according to the state equation in (2.360).

(a) Suppose that the simulation step size is $T = 0.1$ s and the mobile station velocity is $v = 30$ km/h. It is desired to have a shadow decorrelation of 0.1 at a distance of 30 m. Find ζ .

(b) Using the value of ζ obtained in part a) and a shadow standard deviation of $\sigma_{\Omega} = 8$ dB, plot a graph of $\Omega_p \text{ (dB)}$ against the *distance* traveled. Scale your plot so the distance traveled goes from 0 to 100 m.

2.53. Plot and compare the path loss (dB) for the free-space and flat specular surface models at 800 MHz versus distance on a log-scale for distances from 1 m to 40 km. Assume that the antennas are isotropic and have a height of 10 m.

2.54. 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:

- base station antenna height $h_b = 30$ m
- mobile station antenna height $h_m = 3$ m
- carrier frequency $f_c = 1000$ MHz
- isotropic base station and mobile station antennas.

Compare the measured path loss with the predicted path loss from Okumura and Hata's model and Lee's area-to-area model. *If any model parameters are undefined, then use the default values.*

2.55. Suppose that the received power from a transmitter at the input to a receiver is 1 mW at a distance of 1 km. Find the predicted power at the input to the same receiver (in dBm) at distances of 2, 3, and 5 km from the transmitter for the following path loss models:

- (a) Free space.
- (b) 2-ray ground reflection.
- (c) Model described by

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

where $d_o = 1 \text{ km}$ and $\beta = 3.5$.

- (d) COST231-Hata model (medium city).

In all cases assume that $f_c = 1800 \text{ MHz}$, $h_b = 40 \text{ m}$, $h_m = 3 \text{ m}$, $G_T = G_R = 0 \text{ dB}$. Tabulate your results.

2.56. Consider Fig. 2.74 and the following data

- The symbol transmission rate is 24,300 symbols/s with 2 bits/symbol
- The channel bandwidth is 30 kHz
- The propagation environment is characterized by an rms delay spread of $\sigma_\tau = 1 \text{ ns}$

A mobile station 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 mobile station from (BSA) and (BSC), as the mobile station moves from BSA to BSB. Clearly state your assumptions and explain the relationship between the propagation characteristics in your model.

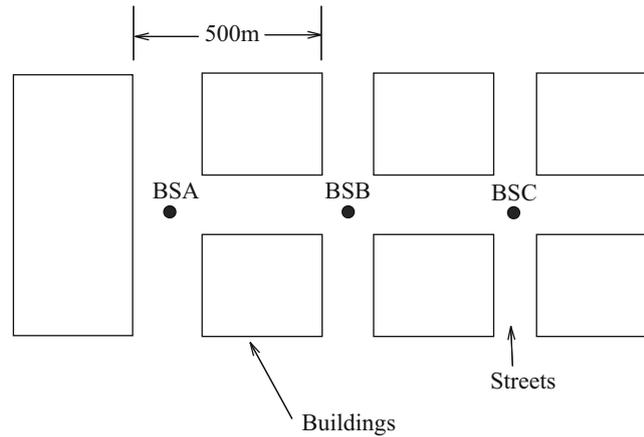


Fig. 2.74 Base station and street layout for Problem 2.56

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 link outage due to co-channel interference (CCI), O_I , is of primary concern. For the remainder of this chapter, the probability of outage refers to the probability of link outage due to CCI. The definition of the outage probability depends on the assumptions made about the radio receiver and propagation environment. One extreme occurs with fast moving mobile stations (MSs), i.e., high Doppler conditions, where the radio receiver can average over the fast envelope variations by using variety of 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 critical receiver threshold A_{th} . The receiver threshold A_{th} is determined by the performance of the radio link in the presence of envelope fading and CCI. Once A_{th} has been determined, the variations in A due to path loss and shadowing will determine the outage probability. Another extreme occurs with stationary or very slowly moving MSs, i.e., low Doppler conditions, where the radio receiver cannot average over the fast envelope variations because the required coding and interleaving depth is too large and will result in excessive transmission delay. Such delays may be acceptable for non-real-time services, but they are unacceptable for real-time services such as voice and streaming video. 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 The threshold λ_{th} is determined by the performance of the radio link in the presence of CCI under the condition that the link does not experience fading. Once λ_{th} has been determined, variations in λ due to path loss, shadowing, and *envelope fading* will determine the outage probability. Sometimes the MSs will move with moderate velocities and the performance will lie somewhere between these two extreme cases.

The effect of CCI on the radio link performance depends on the ability of the radio receiver to reject CCI. Some of the more advanced receivers incorporate sophisticated signal processing techniques to reject or cancel the CCI, e.g., single antenna interference cancellation techniques, or optimum combining with multiple antennas. In this case, the radio receiver is more tolerant to CCI and the receiver thresholds A_{th} and λ_{th} are generally reduced. This will reduce the outage probability or, conversely, improve the coverage probability.

Evaluating the outage probability for the log-normally shadowed signals that are typically found in land mobile radio systems requires the probability distribution of the total interference power that is accumulated from the sum of multiple log-normally shadowed interferers. 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 in the literature. All of these approaches approximate the sum of log-normal random variables by another log-normal random variable. One such method that matches the first two moments of the approximation was developed by Fenton [124]. Sometimes Wilkinson is credited with this method, as in [296]. Here, it is called the Fenton–Wilkinson method. Schwartz and Yeh developed another log-normal approximation that is based on the exact first two moments for the sum of two log-normal random variables [296]. The Schwartz & Yeh method generally provides a more accurate approximation than the Fenton–Wilkinson method but it is more difficult to use. Prasad and Arnbak [265] have corrected some errors in the equations found in Schwartz and Yeh’s original paper, but the equations for the Schwartz & Yeh method in their paper also have errors. Another log-normal approximation is the cumulants matching approach suggested by Schleher [294]. 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 & Yeh, Farley, and Schleher has been undertaken by Beaulieu et al. [34].

¹Note that A_{th} and λ_{th} are not the same.

The above log-normal approximations have been extensively applied to the calculation of the probability of outage in cellular systems. For example, the Fenton–Wilkinson method has been applied by Nagata and Akaiwa [237], Cox [84], Muammar and Gupta [232], and Daikoku and Ohdate [88]. Likewise, the Schwartz & Yeh approach has been applied by Yeh and Schwartz [365], Prasad and Arnbak [265], and Prasad et al. [269].

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 [362], Muammar [231], and Prasad and Kegel [268]. Section 3.3 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 [208], Nakagami fading by Ho and Stüber [164], and Ricean fading by Austin and Stüber [27]. Section 3.4 shows 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 Nakagami- m fading, the probability of outage is sensitive to the shape factor m of the desired signal but is insensitive to the shape 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 receiver or decision feedback equalizer.

Most of the literature dealing with the probability of outage assumes that the interfering co-channel signals add non-coherently. The probability of outage has also been evaluated by Prasad and Kegel [266, 268] 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 [266] and Linnartz [208], it is more realistic to assume non-coherent addition of co-channel interferers in mobile radio environments. Coherently addition of co-channel interferers generally leads to pessimistic predictions of the probability of outage.

The remainder of this chapter begins with approximations for the sum of multiple log-normally shadowed interferers in Sect. 3.1. The various approximations are compared in terms of their accuracy. Section 3.2 derives the probability of outage with log-normal/multiple log-normal (desired/interfering) signals. Section 3.3 considers the outage probability for Ricean/multiple Rayleigh signals without shadowing. Section 3.4 does the same for log-normal Nakagami/multiple log-normal Nakagami signals.

3.1 Multiple Log-Normal Interferers

Consider the sum of N_I log-normal random variables

$$I = \sum_{k=1}^{N_I} \Omega_k = \sum_{k=1}^{N_I} 10^{\Omega_k(\text{dBm})/10}, \quad (3.1)$$

where the Ω_k (dBm) are Gaussian random variables with means $\mu_{\Omega_k(\text{dBm})}$ and variances $\sigma_{\Omega_k}^2$, and the $\Omega_k = 10^{\Omega_k(\text{dBm})/10}$ are 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. One may think to apply the central limit theorem, provided that N_I is large enough, and approximate I as a Gaussian random variable. However, since I represents a power sum it cannot assume negative values so that the resulting approximation is invalid. Moreover, the value of N_I will be small in the case of a few dominant co-channel interferers so the central limit theorem will not apply anyway. 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,

$$I = \sum_{k=1}^{N_I} 10^{\Omega_k(\text{dBm})/10} \approx 10^{Z(\text{dBm})/10} = \hat{I}, \quad (3.2)$$

where $Z(\text{dBm})$ is a Gaussian random variable with mean μ_Z (dBm) and variance σ_Z^2 . The problem is to determine μ_Z (dBm) and σ_Z^2 in terms of the $\mu_{\Omega_k(\text{dBm})}$ and $\sigma_{\Omega_k}^2$, $k = 1, \dots, N_I$. Several methods have been suggested in the literature to solve this problem including those by Fenton [124], Schwartz and Yeh [296], and Farley [296]. 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 interferers N_I .

3.1.1 Fenton–Wilkinson Method

With the Fenton–Wilkinson method, the mean μ_Z (dBm) and variance σ_Z^2 of $Z_{(\text{dBm})}$ are obtained by matching the first two moments of the sum I with the first two moments of the approximation \hat{I} . To derive these moments, it is convenient to use natural logarithms. Hence,

$$\Omega_k = 10^{\Omega_k (\text{dBm})/10} = e^{\xi \Omega_k (\text{dBm})} = e^{\hat{\Omega}_k}, \quad (3.3)$$

where $\xi = (\ln 10)/10 = 0.23026$ and $\hat{\Omega}_k = \xi \Omega_k (\text{dBm})$. Note that $\mu_{\hat{\Omega}_k} = \xi \mu_{\Omega_k (\text{dBm})}$ and $\sigma_{\hat{\Omega}_k}^2 = \xi^2 \sigma_{\Omega_k}^2$. The n th 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

$$\text{E}[\Omega_k^n] = \text{E}[e^{n\hat{\Omega}_k}] = e^{n\mu_{\hat{\Omega}_k} + (1/2)n^2\sigma_{\hat{\Omega}_k}^2}. \quad (3.4)$$

To find the required moments for the log-normal approximation, (3.4) can be used by equating the first two moments on both sides of the approximation

$$I = \sum_{k=1}^{N_I} e^{\hat{\Omega}_k} \approx e^{\hat{Z}} = \hat{I}, \quad (3.5)$$

where $\hat{Z} = \xi Z_{(\text{dBm})}$. For example, suppose that the $\hat{\Omega}_k, k = 1, \dots, N_I$ have means $\mu_{\hat{\Omega}_k}, k = 1, \dots, N_I$ and identical variances $\sigma_{\hat{\Omega}_k}^2$. Identical variances are often assumed for the sum of log-normal interferers because the standard deviation of log-normal shadowing is largely independent of the radio path length [194, 196]. Equating the means on both sides of (3.5) gives

$$\mu_I = \text{E}[I] = \sum_{k=1}^{N_I} \text{E}[e^{\hat{\Omega}_k}] = \text{E}[e^{\hat{Z}}] = \text{E}[\hat{I}] = \mu_{\hat{I}} \quad (3.6)$$

and substituting (3.4) with $n = 1$ into (3.6) gives the result

$$\left(\sum_{k=1}^{N_I} e^{\mu_{\hat{\Omega}_k}} \right) e^{(1/2)\sigma_{\hat{\Omega}}^2} = e^{\mu_{\hat{Z}} + (1/2)\sigma_{\hat{Z}}^2}. \quad (3.7)$$

Likewise, the variances on both sides of (3.5) can be equated, i.e.,

$$\sigma_I^2 = \text{E}[I^2] - \mu_I^2 = \text{E}[\hat{I}^2] - \mu_{\hat{I}}^2 = \sigma_{\hat{I}}^2. \quad (3.8)$$

Under the assumption that the $\hat{\Omega}_k, k = 1, \dots, N_I$ are independent random variables when calculating the second moments in the above equation, this gives the result

$$\left(\sum_{k=1}^{N_I} e^{2\mu_{\hat{\Omega}_k}} \right) e^{\sigma_{\hat{\Omega}}^2} (e^{\sigma_{\hat{\Omega}}^2} - 1) = e^{2\mu_{\hat{Z}}} e^{\sigma_{\hat{Z}}^2} (e^{\sigma_{\hat{Z}}^2} - 1). \quad (3.9)$$

By squaring each side of the equality in (3.7) and dividing each side of resulting equation by the respective sides of the equality in (3.9), $\sigma_{\hat{Z}}^2$ can be solved in terms of the known values of $\mu_{\hat{\Omega}_k}, k = 1, \dots, N_I$ and $\sigma_{\hat{\Omega}_k}^2$. Afterwards, $\mu_{\hat{Z}}$ can be obtained by substituting the obtained expression for $\sigma_{\hat{Z}}^2$ into (3.7). This procedure yields the following solution:

$$\mu_{\hat{Z}} = \frac{\sigma_{\hat{\Omega}}^2 - \sigma_{\hat{Z}}^2}{2} + \ln \left(\sum_{k=1}^{N_I} e^{\mu_{\hat{\Omega}_k}} \right) \quad (3.10)$$

$$\sigma_{\hat{Z}}^2 = \ln \left((e^{\sigma_{\hat{\Omega}}^2} - 1) \frac{\sum_{k=1}^{N_I} e^{2\mu_{\hat{\Omega}_k}}}{\left(\sum_{k=1}^{N_I} e^{\mu_{\hat{\Omega}_k}}\right)^2} + 1 \right). \quad (3.11)$$

Finally, conversion back to base 10 logarithms is done by scaling, such that $\mu_Z \text{ (dBm)} = \xi^{-1} \mu_{\hat{Z}}$ 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 $I_{\text{(dB)}} = 10 \log_{10} I$ are estimated, and how well the cumulative distribution function (cdf) of $I_{\text{(dB)}}$ is described by a Gaussian cdf. It has been reported in [296] that the Fenton–Wilkinson method breaks down in the accuracy of the values obtained for $\mu_Z \text{ (dBm)}$ and σ_Z^2 when $\sigma_{\Omega} > 4$ dB. For cellular radio applications σ_{Ω} typically ranges from 6 to 12 dB and the Fenton–Wilkinson method has often been discredited in the literature on that basis. However, as pointed out in [34], the Fenton–Wilkinson method breaks down only if one considers the application of the Fenton–Wilkinson method for the prediction of the first two moments of $I_{\text{(dB)}}$. Moreover, in problems relating to the probability of CCI outage in cellular radio systems, the tails of the complementary distribution function (cdfc) $F_I^c(x) = \text{P}[I \geq x]$ and the cdf $F_I(x) = 1 - F_I^c(x) = \text{P}[I < x]$ are of interest. In this case, the accuracy of the approximation

$$F_I^c(x) \approx \text{P}[e^{\hat{Z}} \geq x] = Q \left(\frac{\ln x - \mu_{\hat{Z}}}{\sigma_{\hat{Z}}} \right) \quad (3.12)$$

is of interest, and $F_I(x) = 1 - F_I^c(x)$ for large and small values of x , respectively. It will be shown later that the Fenton–Wilkinson method approximates the tails of the cdf and cdfc functions with good accuracy, a result that was reported in [34].

3.1.2 Schwartz & Yeh Method

The Schwartz & Yeh method [296] calculates exact values for the first two moments of the sum of two independent 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 independent 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 first computed. Afterwards, 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, the Schwartz & Yeh method only needs to be detailed for the case when $N_I = 2$, i.e.,

$$I = e^{\hat{\Omega}_1} + e^{\hat{\Omega}_2} \approx e^{\hat{Z}} = \hat{I} \quad (3.13)$$

or

$$\hat{Z} \approx \ln \left(e^{\hat{\Omega}_1} + e^{\hat{\Omega}_2} \right), \quad (3.14)$$

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 $\sigma_{\hat{\Omega}_1}^2$ and $\sigma_{\hat{\Omega}_2}^2$, respectively.

Define the Gaussian random variable $\hat{\Omega}_d \triangleq \hat{\Omega}_2 - \hat{\Omega}_1$ so that

$$\mu_{\hat{\Omega}_d} = \mu_{\hat{\Omega}_2} - \mu_{\hat{\Omega}_1} \quad (3.15)$$

$$\sigma_{\hat{\Omega}_d}^2 = \sigma_{\hat{\Omega}_1}^2 + \sigma_{\hat{\Omega}_2}^2. \quad (3.16)$$

Taking the expectation of both sides of (3.14) and assuming that the approximation holds with equality gives

$$\begin{aligned} \mu_{\hat{Z}} &= \text{E} \left[\ln \left(e^{\hat{\Omega}_2} + e^{\hat{\Omega}_1} \right) \right] \\ &= \text{E} \left[\ln \left(e^{\hat{\Omega}_1} \left(1 + e^{\hat{\Omega}_2 - \hat{\Omega}_1} \right) \right) \right] \\ &= \text{E} \left[\hat{\Omega}_1 \right] + \text{E} \left[\ln \left(1 + e^{\hat{\Omega}_d} \right) \right]. \end{aligned} \quad (3.17)$$

The second term in (3.17) is

$$\mathbb{E} \left[\ln \left(1 + e^{\hat{\Omega}_d} \right) \right] = \int_{-\infty}^{\infty} (\ln(1 + e^x)) p_{\hat{\Omega}_d}(x) dx. \quad (3.18)$$

Now invoke the power series expansion

$$\ln(1 + x) = \sum_{k=1}^{\infty} c_k x^k, \quad c_k = \frac{(-1)^{k+1}}{k}, \quad (3.19)$$

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 in the above equation was obtained by using the identity

$$\begin{aligned} \ln(1 + e^x) &= \ln((e^{-x} + 1)e^x) \\ &= \ln(1 + e^{-x}) + \ln(e^x) \\ &= \ln(1 + e^{-x}) + x. \end{aligned} \quad (3.21)$$

After a very long derivation that is detailed in [296],

$$\mu_{\hat{Z}} = \mu_{\hat{\Omega}_1} + G_1, \quad (3.22)$$

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

$$T_1 = e^{-k\mu_{\hat{\Omega}_d}} \Phi \left(\frac{\mu_{\hat{\Omega}_d} - k\sigma_{\hat{\Omega}_d}^2}{\sigma_{\hat{\Omega}_d}} \right), \quad (3.24)$$

and where $\Phi(x)$ is the complementary distribution function (cdfc) of a standard normal random variable, defined as

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \quad (3.25)$$

The variance of \hat{Z} can be computed in a similar fashion, resulting in the expression [296]

$$\sigma_{\hat{Z}}^2 = \sigma_{\hat{\Omega}_1}^2 - G_1^2 - 2\sigma_{\hat{\Omega}_1}^2 G_3 + G_2, \quad (3.26)$$

where

$$\begin{aligned} G_2 &= \sum_{k=1}^{\infty} b_k T_2 + \left(1 - \Phi \left(-\frac{\mu_{\hat{\Omega}_d}}{\sigma_{\hat{\Omega}_d}} \right) \right) (\mu_{\hat{\Omega}_d}^2 + \sigma_{\hat{\Omega}_d}^2) + \frac{\mu_{\hat{\Omega}_d} \sigma_{\hat{\Omega}_d}}{\sqrt{2\pi}} e^{-\mu_{\hat{\Omega}_d}^2 / (2\sigma_{\hat{\Omega}_d}^2)} \\ &\quad + \sum_{k=1}^{\infty} b_k e^{-(k+1)\mu_{\hat{\Omega}_d} + (k+1)^2 \sigma_{\hat{\Omega}_d}^2 / 2} \Phi \left(\frac{\mu_{\hat{\Omega}_d} - \sigma_{\hat{\Omega}_d}^2 (k+1)}{\sigma_{\hat{\Omega}_d}} \right) \end{aligned} \quad (3.27)$$

$$\begin{aligned}
& -2 \sum_{k=1}^{\infty} c_k e^{-k\mu_{\hat{\Omega}_d} + k^2\sigma_{\hat{\Omega}_d}^2/2} \left(\xi_{\hat{\Omega}_k} \Phi \left(-\frac{\xi_{\hat{\Omega}_k}}{\sigma_{\hat{\Omega}_d}} \right) - \frac{\sigma_{\hat{\Omega}_d}}{\sqrt{2\pi}} e^{-\xi_{\hat{\Omega}_k}^2/(2\sigma_{\hat{\Omega}_d}^2)} \right) \\
G_3 = & \sum_{k=0}^{\infty} (-1)^k e^{k^2\sigma_{\hat{\Omega}_d}^2/2} T_1 + \sum_{k=0}^{\infty} (-1)^k T_2 \tag{3.28}
\end{aligned}$$

with

$$T_2 = e^{\mu_{\hat{\Omega}_d}(k+1) + (k+1)^2\sigma_{\hat{\Omega}_d}^2/2} \Phi \left(\frac{-\mu_{\hat{\Omega}_d} - (k+1)\sigma_{\hat{\Omega}_d}^2}{\sigma_{\hat{\Omega}_d}} \right) \tag{3.29}$$

and

$$b_k = \frac{2(-1)^{k+1}}{k+1} \sum_{n=1}^k \frac{1}{n} \tag{3.30}$$

$$\xi_{\hat{\Omega}_k} = -\mu_{\hat{\Omega}_d} + k\sigma_{\hat{\Omega}_d}^2. \tag{3.31}$$

It has been reported in [296] that approximately 40 terms are required in the infinite summations for G_1 , G_2 , and G_3 to achieve four significant digits of accuracy in the moments of \hat{Z} . On the next step of the recursion it is important to let $\sigma_{\hat{\Omega}_1}^2 = \sigma_{\hat{Z}}^2$ and $\mu_{\hat{\Omega}_1} = \mu_{\hat{Z}}$; otherwise, the recursive procedure will fail to converge.

3.1.3 Farley's Method

Consider N_I independent identically distributed (i.i.d.) normal random variables $\hat{\Omega}_k$ each with mean $\mu_{\hat{\Omega}}$ and variance $\sigma_{\hat{\Omega}}^2$. Farley approximated the cdf of the sum

$$I = \sum_{k=1}^{N_I} \Omega_k = \sum_{k=1}^{N_I} e^{\hat{\Omega}_k} \tag{3.32}$$

as [296]

$$\text{P}[I > x] \approx 1 - \left(1 - Q \left(\frac{\ln x - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) \right)^{N_I}. \tag{3.33}$$

As shown in [34], Farley's approximation is actually a strict lower bound on the cdf. To obtain this result let

$$F_I^c(x) = \text{P}[\Omega_1 + \Omega_2 + \dots + \Omega_{N_I} > x] \tag{3.34}$$

and define the two events

$$\begin{aligned}
A &= \{\text{at least one } \Omega_i > x\} \\
B &= A^c, \text{ the complement of event } A.
\end{aligned} \tag{3.35}$$

Events A and B are mutually exclusive and partition the sample space. Therefore,

$$\begin{aligned}
\text{P}[I > x] &= \text{P}[I > x \cap A] + \text{P}[I > x \cap B] \\
&= \text{P}[A] + \text{P}[I > x \cap B].
\end{aligned} \tag{3.36}$$

The second term in (3.36) is positive for continuous pdfs such as the log-normal pdf. For example, the event

$$C = \left\{ \bigcap_{i=1}^{N_I} x/N_I < \Omega_i < x \right\} \quad (3.37)$$

is a subset of the event B . Under the assumption that the Ω_i are i.i.d., event C occurs with non-zero probability because

$$P[C] = \left(Q \left(\frac{\ln(x/N_I) - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) - Q \left(\frac{\ln x - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) \right)^{N_I} > 0. \quad (3.38)$$

Therefore, $P[I > x] > P[A]$. Since the Ω_i are i.i.d.,

$$\begin{aligned} P[A] &= 1 - \prod_{i=1}^{N_I} P[\Omega_i \leq x] \\ &= 1 - \left(1 - Q \left(\frac{\ln x - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) \right)^{N_I}. \end{aligned} \quad (3.39)$$

This finally yields the lower bound on the cdfc

$$P[I > x] > 1 - \left(1 - Q \left(\frac{\ln x - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) \right)^{N_I} \quad (3.40)$$

or, equivalently, the upper bound on the cdf

$$P[I \leq x] > \left(1 - Q \left(\frac{\ln x - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) \right)^{N_I}. \quad (3.41)$$

3.1.4 Numerical Comparisons

Figure 3.1 compares the cdf for the sum of $N_I = 2$ and $N_I = 6$ log-normal random variables, obtained with the various approximations. Likewise, Figs. 3.2, 3.3, and 3.4 compare the cdfc, obtained with the various approximations. 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 argument of the cdfc. The cdf is approximated less accurately, especially for $N_I = 6$ log-normal random variables.

3.2 Log-Normal/Multiple Log-Normal Interferers

Consider the situation shown in Fig. 1.15, where a mobile station (MS) is at distance d_0 from the serving base station (BS) and at distances $d_k, k = 1, 2, \dots, N_I$ from the first tier of N_I co-channel BSs. Define the vector $\mathbf{d} = (d_0, d_1, \dots, d_{N_I})$ as the set of distances of a particular MS from the serving BS and surrounding BSs. The average received carrier-to-interference ratio as a function of the vector \mathbf{d} is

$$A_{(\text{dB})}(\mathbf{d}) = \Omega_{(\text{dBm})}(d_0) - 10 \log_{10} \sum_{k=1}^{N_I} 10^{\Omega_{(\text{dBm})}(d_k)/10}. \quad (3.42)$$

For the case of a single interferer ($N_I = 1$) the sum on the right side of (3.42) has only one term. In this case, $A_{(\text{dB})}(\mathbf{d})$ is Gaussian distributed with mean $\mu_{\Omega_{(\text{dBm})}(d_0)} - \mu_{\Omega_{(\text{dBm})}(d_1)}$ and variance $2\sigma_{\Omega}^2$. For the case of multiple log-normal interferers, the second term of (3.42) is approximated as a normal random variable $Z_{(\text{dBm})}$ with mean μ_Z and variance σ_Z^2 using the techniques discussed in Sect. 3.1. Then

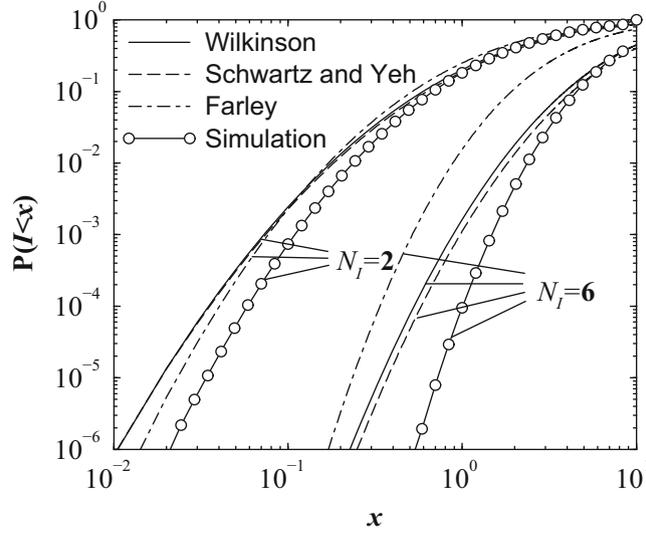


Fig. 3.1 Comparison of the cdf for the sum of two and six log-normal random variables with various approximations; $\mu_{\Omega_k \text{ (dB)}} = 0 \text{ dB}$, $\sigma_{\Omega} = 6 \text{ dB}$

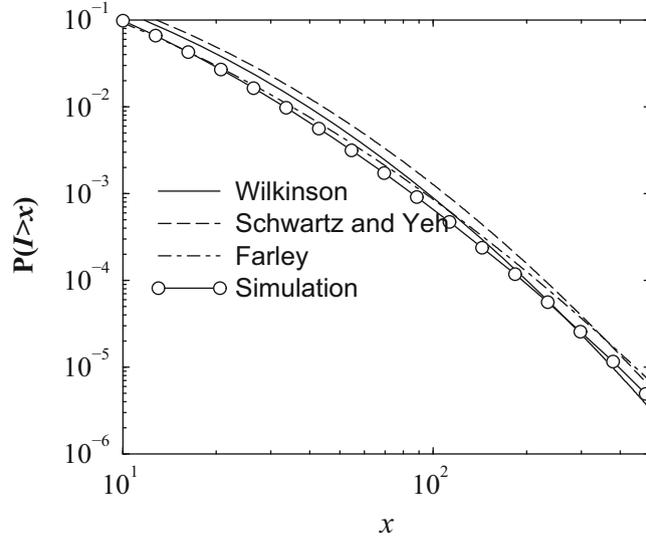


Fig. 3.2 Comparison of the cdfc for the sum of two log-normal random variables with various approximations; $\mu_{\Omega_k \text{ (dB)}} = 0 \text{ dB}$, $\sigma_{\Omega} = 6 \text{ dB}$

$$\Lambda_{\text{(dB)}}(\mathbf{d}) = \Omega_{\text{(dBm)}}(d_0) - Z_{\text{(dBm)}}(d_1, d_2, \dots, d_{N_I}), \quad (3.43)$$

where we, again, show the dependency of the CCI on the set of distances. Note that $\Lambda_{\text{(dB)}}(\mathbf{d})$ has mean and variance

$$\mu_{\Lambda_{\text{(dB)}}(\mathbf{d})} = \mu_{\Omega_{\text{(dBm)}}(d_0)} - \mu_Z \text{ (dBm)} \quad (3.44)$$

$$\sigma_{\Lambda_{\text{(dB)}}(\mathbf{d})}^2 = \sigma_{\Omega}^2 + \sigma_Z^2. \quad (3.45)$$

If there were only one possible choice of serving BS, then the probability of outage at a particular MS location is

$$O_I(\mathbf{d}) = Q \left(\frac{\mu_{\Omega_{\text{(dBm)}}(d_0)} - \mu_Z \text{ (dBm)} - \Lambda_{\text{th}} \text{ (dB)}}{\sqrt{\sigma_{\Omega}^2 + \sigma_Z^2}} \right). \quad (3.46)$$

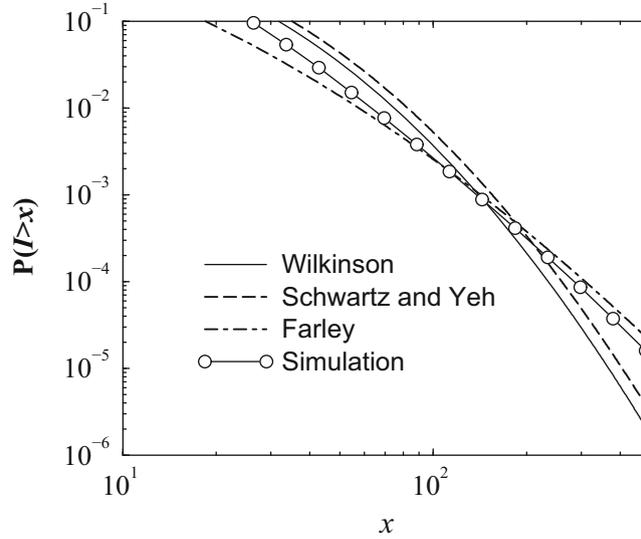


Fig. 3.3 Comparison of the cdfc for the sum of six log-normal random variables with various approximations; $\mu_{\Omega_k \text{ (dB)}} = 0 \text{ dB}$, $\sigma_{\Omega} = 6 \text{ dB}$

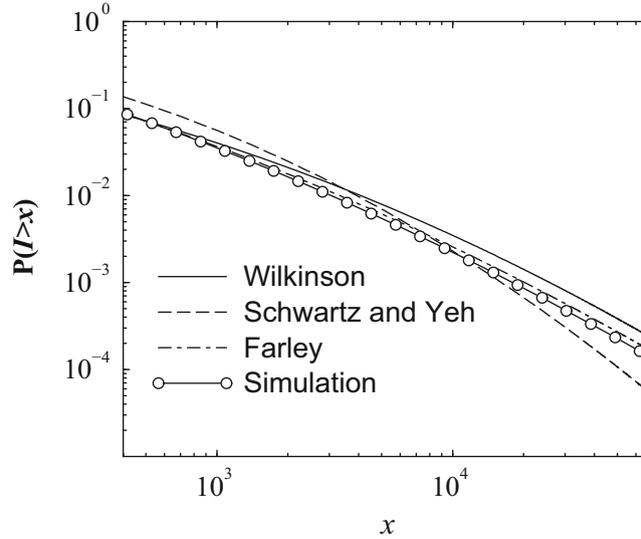


Fig. 3.4 Comparison of the cdfc for the sum of six log-normal random variables with various approximations; $\mu_{\Omega_k \text{ (dB)}} = 0 \text{ dB}$, $\sigma_{\Omega} = 12 \text{ dB}$

If handoffs are allowed, then 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, soft handoffs can be considered where the BS that provides the best 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 the receiver threshold Λ_{th} . In this case, the probability of outage at a particular location is

$$O_I(\mathbf{d}) = \prod_{k=0}^M Q \left(\frac{\mu_{\Omega_k \text{ (dBm)}}(d_0) - \mu_{Z_k \text{ (dBm)}} - \Lambda_{\text{th}}(\text{dB})}{\sqrt{\sigma_{\Omega}^2 + (\sigma_{Z_k})^2}} \right). \quad (3.47)$$

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 the reference cell.

3.3 Rician/Multiple Rayleigh Interferers

Sometimes propagation conditions may exist such that the received signals experience envelope fading, but not shadowing. This section calculates the outage probability for the case of envelope fading only. The case of combined shadowing and envelope fading is deferred until the next section. In the case of envelope fading only, the received desired signal may consist of a direct line of sight (LoS) component, or perhaps a specular component, accompanied by a diffuse component. The envelope of the received desired signal experiences Ricean fading. The co-channel interferers are often assumed to be Rayleigh faded, because a direct LoS condition is unlikely to exist between the co-channel interferers and target receiver due to their large physical separation. Let the instantaneous power in the desired signal and the N_I interfering signals be denoted by s_0 and s_k , $k = 1, \dots, N_I$, respectively. Note that $s_i = \alpha_i^2$, where α_i^2 is the squared-envelope. The carrier-to-interference ratio is defined as

$$\lambda \triangleq \frac{s_0}{\sum_{k=1}^{N_I} s_k}. \quad (3.48)$$

For a specified receiver threshold λ_{th} , the outage probability is

$$O_I = \text{P}[\lambda < \lambda_{\text{th}}]. \quad (3.49)$$

The instantaneous received power of the desired signal, s_0 , has the non-central chi-square distribution (Ricean fading) in (2.63), while the instantaneous power of each interferer, s_k , has the exponential distribution (Rayleigh fading) in (2.56).

For the case of a single interferer, the outage probability reduces to the simple closed form expression [362]

$$O_I = \frac{\lambda_{\text{th}}}{\lambda_{\text{th}} + A_1} \exp\left\{-\frac{KA_1}{\lambda_{\text{th}} + A_1}\right\}, \quad (3.50)$$

where K is the Rice factor of the desired signal, $A_1 = \Omega_0/(K+1)\Omega_1$, and $\Omega_k = \text{E}[s_k]$. Note that A_1 can be interpreted as the ratio of the average desired signal power to the total interfering power. If the desired signal is Rayleigh faded, then the outage probability can be obtained by setting $K = 0$ in (3.50). For the case of multiple interferers, each with mean power Ω_k , the outage probability has the closed form expression [362]

$$O_I = 1 - \sum_{k=1}^{N_I} \left(1 - \frac{\lambda_{\text{th}}}{\lambda_{\text{th}} + A_k} \exp\left\{-\frac{KA_k}{\lambda_{\text{th}} + A_k}\right\}\right) \prod_{\substack{j=1 \\ j \neq k}}^{N_I} \frac{A_j}{A_j - A_k}, \quad (3.51)$$

where $A_k = \Omega_0/(K+1)\Omega_k$. This expression is valid only if $\Omega_i \neq \Omega_j$ when $i \neq j$, i.e., the different interferers are received with distinct mean power levels. If some of the interfering signals are received with the same mean power, then an appropriate expression for the outage probability can be derived in straightforward manner. If all the interferers are received with the same mean power, then the total interference power $s_M = \sum_{k=1}^{N_I} s_k$ has the Gamma pdf

$$p_{s_M}(x) = \frac{x^{N_I-1}}{\Omega_1^{N_I} (N_I-1)!} \exp\left\{-\frac{x}{\Omega_1}\right\}. \quad (3.52)$$

The outage probability can be derived as [362]

$$O_I = \frac{\lambda_{\text{th}}}{\lambda_{\text{th}} + A_1} \exp\left\{-\frac{KA_1}{\lambda_{\text{th}} + A_1}\right\} \sum_{k=0}^{N_I-1} \left(\frac{A_1}{\lambda_{\text{th}} + A_1}\right)^k \sum_{m=0}^k \binom{k}{m} \frac{1}{m!} \left(\frac{K\lambda_{\text{th}}}{\lambda_{\text{th}} + A_1}\right)^m. \quad (3.53)$$

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) as appropriate. In Fig. 3.5, the outage probability is plotted as a function of the average received carrier-to-average-interference ratio

$$\Lambda = \frac{\Omega_0}{N_I \Omega_1} \quad (3.54)$$

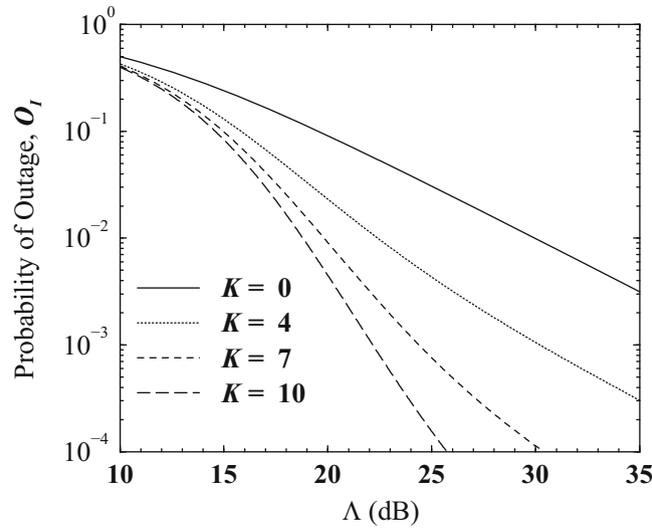


Fig. 3.5 Probability of CCI outage with a single interferer. The desired signal is Ricean faded with various Rice factors, while the interfering signal is Rayleigh faded; $\lambda_{\text{th}} = 10.0$ dB

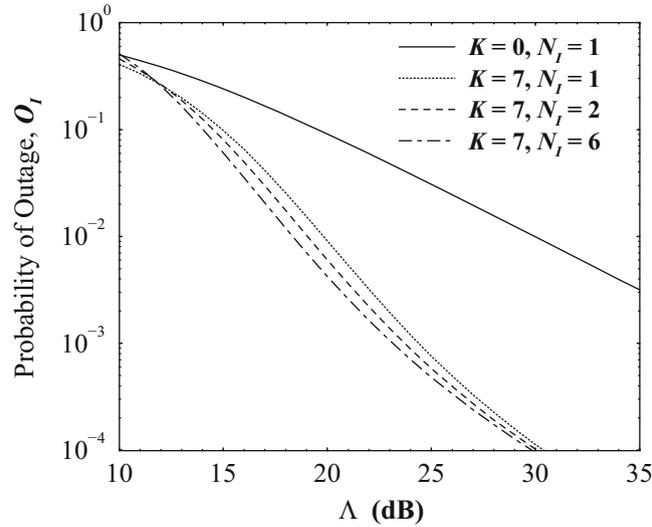


Fig. 3.6 Probability of CCI outage with multiple interferers. The desired signal is Ricean faded with various Rice factors, while the interfering signals are Rayleigh faded and of equal power; $\lambda_{\text{th}} = 10.0$ dB

for various Rice factors and a single interferer. Observe that the Rice factor of the desired signal has a significant effect on the outage probability. Figure 3.6 plots the outage probability for $K = 0$ and 7 and a varying number of interferers. Meanwhile, the number of interferers does not affect the probability of outage as much as the Rice factor of the desired signal, provided that the total interfering power remains constant.

3.4 Log-Normal Nakagami/Multiple Log-Normal Nakagami Interferers

The probability of outage has been evaluated in the literature for a single Nakagami interferer [356] and multiple Nakagami interferers [6, 363], in the absence of shadowing. Here, the probability of outage is analytically formulated for a log-normal Nakagami faded desired signal with multiple log-normal Nakagami interferers. For the case when the interfering signals have the same shadowing and fading statistics, an exact mathematical expression is derived 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 , $k = 1, \dots, N_I$, respectively. The instantaneous carrier-to-interference ratio is $\lambda = s_0 / \sum_{k=1}^{N_I} s_k$. For a specified receiver threshold λ_{th} , the probability of outage is

$$O_I = \text{P}[\lambda < \lambda_{\text{th}}] \equiv \text{P}\left[s_0 < \lambda_{\text{th}} \sum_{k=1}^{N_I} s_k\right]. \quad (3.55)$$

The k th interfering signal, s_k , $k = 1, \dots, N_I$, is affected by log-normal shadowing and Nakagami fading with the composite pdf c.f. (2.373)

$$p_{s_k}(x) = \int_0^\infty \left(\frac{m_k}{w}\right)^{m_k} \frac{x^{m_k-1}}{\Gamma(m_k)} \exp\left\{-\frac{m_k x}{w}\right\} \frac{1}{\sqrt{2\pi\xi\sigma_\Omega w}} \exp\left\{-\frac{(10\log_{10}\{w\} - \mu_{\Omega_k \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} dw. \quad (3.56)$$

Let $W = \sum_{k=1}^{N_I} s_k$ be the total instantaneous received power from the N_I interfering signals such that $\lambda = s_0/W$, and define the auxiliary random variable $Y = W$. Then by using a bivariate transformation of random variables, the joint pdf of λ and Y is $p_{\lambda,Y}(x, y) = yp_{s_0,W}(xy, y) = yp_{s_0}(xy)p_W(y)$ and

$$p_\lambda(x) = \int_0^\infty yp_{s_0}(xy)p_W(y)dy, \quad (3.57)$$

where the fact that s_0 and W are statistically independent random variables has been used. It follows that the outage probability is

$$\begin{aligned} O_I &= \text{P}[\lambda < \lambda_{\text{th}}] \\ &= 1 - \int_{\lambda_{\text{th}}}^\infty \int_0^\infty yp_{s_0}(xy)p_W(y)dydx. \end{aligned} \quad (3.58)$$

Suppose for the time being that the desired signal is affected by Nakagami fading only, i.e., there is no shadowing and $\Omega_0 = \text{E}[s_0]$ is fixed. Then s_0 has the Gamma distribution in (2.67) and

$$\int_{\lambda_{\text{th}}}^\infty p_{s_0}(xy)dx = \sum_{h=0}^{m_0-1} \left(\frac{m_0}{\Omega_0}\right)^h \frac{y^{h-1} \lambda_{\text{th}}^h}{h!} \exp\left\{-\frac{m_0 \lambda_{\text{th}} y}{\Omega_0}\right\}. \quad (3.59)$$

Hence, the conditional outage probability in (3.58) becomes

$$\text{P}(\lambda < \lambda_{\text{th}} | \Omega_0) = 1 - \sum_{h=0}^{m_0-1} \left(\frac{m_0 \lambda_{\text{th}}}{\Omega_0}\right)^h \frac{1}{h!} \int_0^\infty \exp\left\{-\frac{m_0 \lambda_{\text{th}} y}{\Omega_0}\right\} y^h p_W(y) dy. \quad (3.60)$$

3.4.1 Statistically Identical Interferers

Here statistically identical interferers are assumed so that $m_k = m_I$ and $\mu_{\Omega_k \text{ (dBm)}} = \mu_{\Omega_I \text{ (dBm)}}$, $i = 1, \dots, N_I$. Following Linnartz [208], the integral in (3.60) can be obtained by using Laplace transform techniques. The Laplace transform of the pdf $p_W(y)$ is

$$\mathcal{L}_W(s) = \int_0^\infty e^{-sy} p_W(y) dy. \quad (3.61)$$

The integral in (3.60) is then equal to the h th derivative of $\mathcal{L}_W(s)$ with respect to s evaluated at the point $s = (m_0 \lambda_{\text{th}}) / \Omega_0$. That is,

$$\begin{aligned} \int_0^\infty e^{-sy} y^h p_W(y) dy &= (-1)^h \mathcal{L}_W^{(h)}(s) \\ &= (-1)^h \frac{d^h}{ds^h} \left\{ \prod_{k=1}^{N_I} \int_0^\infty e^{-sy_k} p_{s_k}(y_k) dy_k \right\}, \end{aligned} \quad (3.62)$$

where the last line follows under the assumption of statistically independent interferers. By using the composite distribution in (3.56) with $m_k = m_I$ and $\mu_{\Omega_k \text{ (dBm)}} = \mu_{\Omega_I \text{ (dBm)}}$, $i = 1, \dots, N_I$, it can be shown that

$$\int_0^\infty e^{-sy_k} p_{s_k}(y_k) dy_k = \frac{m_I^{m_I}}{\sqrt{\pi}} \int_{-\infty}^\infty \frac{e^{-x^2}}{\left(10^{(\mu_{\Omega_I \text{ (dBm)}} + \sqrt{2}\sigma_{\Omega,x})/10} s + m_I\right)^{m_I}} dx. \quad (3.63)$$

Using this result and averaging over the log-normal shadowing distribution of the desired signal gives the final result

$$\begin{aligned} O_I &= 1 - \int_0^\infty \left(\sum_{h=0}^{m_0-1} \left(-\frac{m_0 \lambda_{\text{th}}}{\Omega_0} \right)^h \frac{1}{h!} \right. \\ &\quad \times \left. \frac{d^h}{ds^h} \left\{ \left(\int_{-\infty}^\infty \frac{m_I^{m_I} e^{-x^2} dx}{\sqrt{\pi} \left(10^{(\mu_{\Omega_k \text{ (dBm)}} + \sqrt{2}\sigma_{\Omega,x})/10} s + m_I\right)^{m_I}} \right)^{N_I} \right\} \Big|_{s=\frac{m_0 \lambda_{\text{th}}}{\Omega_0}} \right) \\ &\quad \times \frac{1}{\sqrt{2\pi} \xi \sigma_\Omega \Omega_0} \exp \left\{ -\frac{(10 \log_{10} \{\Omega_0\} - \mu_{\Omega_0 \text{ (dBm)}})^2}{2\sigma_\Omega^2} \right\} d\Omega_0. \end{aligned} \quad (3.64)$$

Equation (3.64) is an exact expression for log-normal Nakagami fading channels. When $m_0 = m_I = 1$, it reduces to the simpler expression obtained by Linnartz [208] for log-normal Rayleigh fading channels. In (3.64), let

$$F(s) = \int_{-\infty}^\infty \frac{e^{-x^2}}{\left(10^{(\mu_{\Omega_I \text{ (dBm)}} + \sqrt{2}\sigma_{\Omega,x})/10} s + m_I\right)^{m_I}} dx \quad (3.65)$$

and use the identity [147]

$$\begin{aligned} G(s) &= \frac{d^h}{ds^h} (F(s))^{N_I} \\ &= N_I \binom{h - N_I}{N_I} \sum_{i=1}^h (-1)^i \binom{h}{i} \frac{(F(s))^{N_I - i}}{N_I - i} \frac{d^h}{ds^h} (F(s))^i. \end{aligned} \quad (3.66)$$

Observe that $G(s)$ is a function of the derivatives of $F(s)$ only, and

$$\begin{aligned} \frac{d^h F(s)}{ds^h} &= \frac{d^h}{ds^h} \left\{ \int_{-\infty}^\infty \frac{e^{-x^2}}{\left(10^{(\mu_{\Omega_I \text{ (dBm)}} + \sqrt{2}\sigma_{\Omega,x})/10} s + m_I\right)^{m_I}} dx \right\} \\ &= (-1)^h \frac{(m_I + h - 1)!}{(m_I - 1)!} \int_{-\infty}^\infty \frac{\left(10^{(\mu_{\Omega_I \text{ (dBm)}} + \sqrt{2}\sigma_{\Omega,x})/10}\right)^h e^{-x^2}}{\left(10^{(\mu_{\Omega_I \text{ (dBm)}} + \sqrt{2}\sigma_{\Omega,x})/10} s + m_I\right)^{m_I + h}} dx. \end{aligned} \quad (3.67)$$

$G(s)$ can be obtained from (3.66) and (3.67), and substituted into (3.64). Then by using the following change in the variable of integration

Table 3.1 Zeros and weight factors of 16 order Hermite polynomials [4]

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	$2.654807474011 \times 10^{-10}$

$$x = \frac{10 \log_{10} \{ \Omega_0 \} - \mu_{\Omega_0} \text{ (dBm)}}{\sqrt{2} \sigma_{\Omega}} \quad (3.68)$$

the outage probability in (3.64) becomes

$$O_I = 1 - \sum_{h=0}^{m_0-1} \left(-m_0 \lambda_{\text{th}} 10^{-\mu_{\Omega_0} \text{ (dBm)}/10} \right)^h \frac{m_I^{m_I N_I}}{\sqrt{\pi}^{N_I+1} h!} \int_{-\infty}^{\infty} 10^{-\sqrt{2} \sigma_{\Omega} x h / 10} e^{-x^2} G \left(m_0 \lambda_{\text{th}} 10^{-(\mu_{\Omega_0} \text{ (dBm)} + \sqrt{2} \sigma_{\Omega} x h) / 10} \right) dx. \quad (3.69)$$

The integrals in (3.67) and (3.69) can be efficiently computed by using Gauss–Hermite quadrature integration. Applying the Gauss–Hermite quadrature formula to (3.67) gives

$$\frac{d^h F(s)}{ds^h} = (-1)^h \frac{(m_I + h - 1)!}{(m_I - 1)!} \sum_{i=1}^{N_p} H_{x_i} \frac{10^{(\mu_{\Omega_I} \text{ (dBm)} + \sqrt{2} \sigma_{\Omega} x_i) h / 10}}{\left(10^{(\mu_{\Omega_I} \text{ (dBm)} + \sqrt{2} \sigma_{\Omega} x_i) / 10} s + m_I \right)^{m_I + h}} \quad (3.70)$$

where H_{x_i} are weight factors, x_i are the zeros of the Hermite polynomial $H_p(x)$, and N_p is the order of the Hermite polynomial. When obtaining numerical results, a Hermite polynomial of order 16 resulted in sufficient accuracy and the corresponding values for H_{x_i} and x_i are listed in Table 3.1. Likewise, for (3.69),

$$O_I = 1 - \sum_{h=0}^{m_0-1} \left(-m_0 \lambda_{\text{th}} 10^{\mu_{\Omega_0} \text{ (dBm)}/10} \right)^h \frac{m_I^{m_I N_I}}{\sqrt{\pi}^{N_I+1} h!} \sum_{\ell=1}^{N_p} H_{x_{\ell}} 10^{-\sqrt{2} \sigma_{\Omega} x_{\ell} h / 10} G \left(m_0 \lambda_{\text{th}} 10^{-(\mu_{\Omega_0} \text{ (dBm)} + \sqrt{2} \sigma_{\Omega} x_{\ell} h) / 10} \right) dx. \quad (3.71)$$

Figure 3.7 shows the probability of outage as a function of the average-carrier-to-average-interference ratio

$$\Lambda = \frac{\mu_{\Omega_0}}{N_I \mu_{\Omega_I}}. \quad (3.72)$$

Results are plotted for $N_I = 6$ interfering signals and varying degrees of fading on the desired and interfering signals. Observe that the outage probability is insensitive to changes in the Nakagami shape factor, m , for interfering signals. This phenomenon demonstrates that the probability of CCI outage is dominated by the fading of the desired signal rather than fading of the interfering signals. Figure 3.8 shows the outage probability for different values of the shadow standard deviation σ_{Ω} . The shadow standard deviation has a significant effect on the outage probability.

3.4.2 Statistically Nonidentical Co-Channel Interferers

If the interferers are statistically nonidentical, then Eq. (3.64) still applies with m_I replaced by m_k . Since the product in (3.64) does not reduce to taking the n th power, the numerical evaluation is difficult. This difficulty can be overcome by using approximations. Section 2.6.2.1 showed that the composite distribution of the squared-envelope due to Nakagami fading and log-normal shadowing can be approximated by a log-normal distribution with the parameters in (2.374). Moreover, the sum of log-normal random variables can be approximated by still another log-normal random variable by using either the

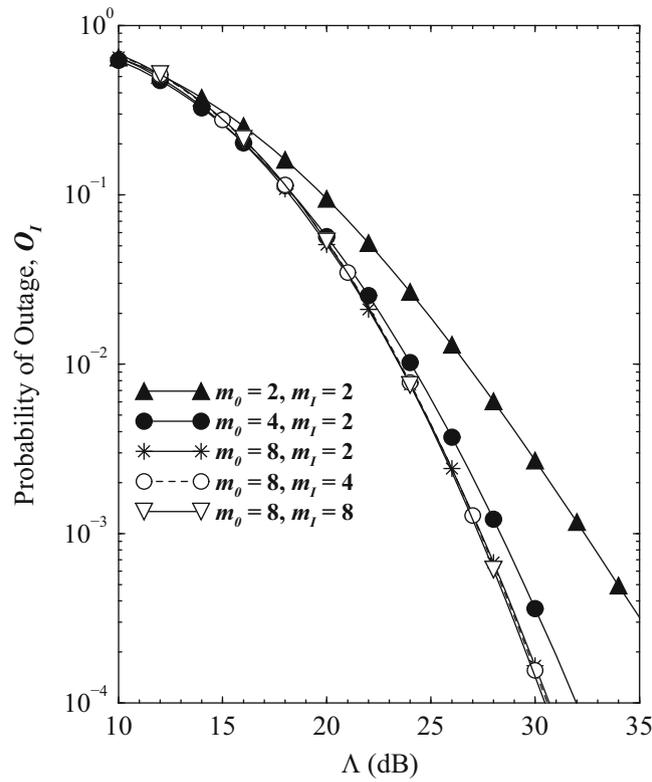


Fig. 3.7 Probability of CCI outage when the desired and interfering signals are Nakagami faded. Results are shown for various fading distribution parameters; $\sigma_\Omega = 6$ dB, $\lambda_{th} = 10.0$ dB

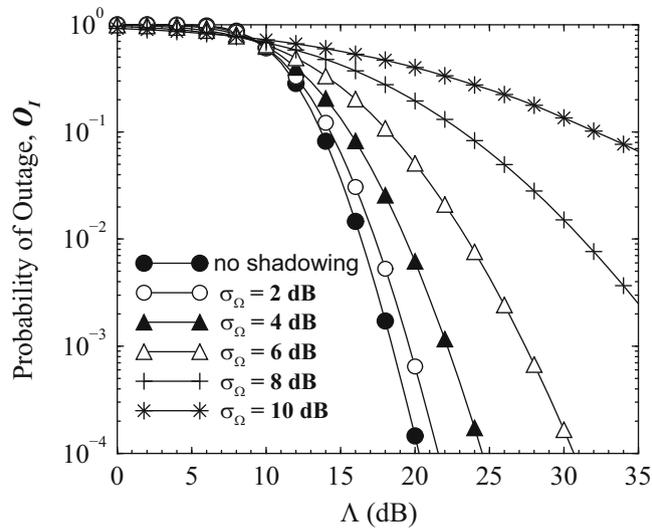
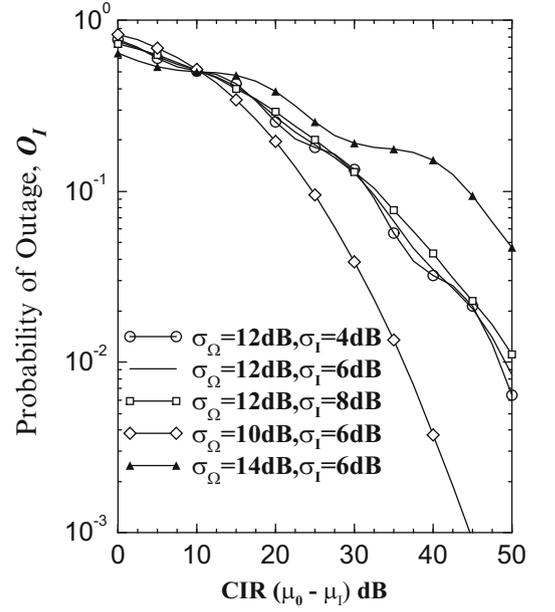


Fig. 3.8 Probability of CCI outage when the desired and interfering signals are Nakagami faded. Results are shown for various shadow standard deviations; $m_0 = 8$, $m_I = 2$, $\lambda_{th} = 10.0$ dB

Fenton–Wilkinson method in Sect. 3.1.1 or the Schwartz & Yeh method in Sect. 3.1.2. Hence, (2.374) can be used to find individual approximated log-normal distribution for each of the interfering signals, then Schwartz & Yeh’s method or the Fenton–Wilkinson method is applied to find a pure log-normal distribution for the total interference power s_I . This results in the density

Fig. 3.9 Probability of CCI outage for different dB spreads and statistically nonidentical interferers; $m_0 = 4$, $\lambda_{th} = 10.0$ dB



$$p_{s_I}(x) = \frac{1}{x\sigma_I\xi\sqrt{2\pi}} \exp \left\{ -\frac{(10\log_{10}\{x\} - \mu_{s_I} \text{ (dBm)})^2}{2\sigma_I^2} \right\}. \quad (3.73)$$

To maintain accuracy, the desired signal is still treated as a composite Nakagami log-normal signal with the *pdf* in (3.56).

$$O_I = 1 - \sum_{k=0}^{m_0-1} \frac{\left(m_0\lambda_{th}10^{(\mu_{s_I} \text{ (dBm)} - \mu_{\Omega_0} \text{ (dBm)})/10}\right)^k}{k!\pi} \int_{-\infty}^{\infty} e^{-y^2} e^{\sqrt{2}\sigma_I\xi ky} \int_{-\infty}^{\infty} e^{-x^2} e^{-\sqrt{2}\sigma_{\Omega}\xi kx} \\ \times \exp \left\{ -m_0\lambda_{th}e^{\sqrt{2}\xi(\sigma_I y - \sigma_{\Omega} x) + \xi(\mu_{s_I} \text{ (dBm)} - \mu_{\Omega_0} \text{ (dBm)})} \right\} dx dy. \quad (3.74)$$

When the number of interferers increases, σ_I decreases, while μ_I increases. For a fixed μ_d the CIR will be reduced when the number of interferers is increased. Once again (3.74) can be evaluated by using double Gauss–Hermite quadrature integration. Figure 3.9 shows the probability of co-channel interference for interferers with various statistics. Observe that the number of interferers and shadowing are the dominant factors in determining the probability of CCI outage.

Problems

3.1. A receiver is affected by 3 log-normally shadowed co-channel signals having the power sum

$$I = \sum_{k=1}^3 I_k,$$

where

$$\begin{aligned} I_{1(\text{dB})} &\sim \mathcal{N}(-10 \text{ dBm}, \sigma_{\Omega}^2) \\ I_{2(\text{dB})} &\sim \mathcal{N}(-15 \text{ dBm}, \sigma_{\Omega}^2) \\ I_{3(\text{dB})} &\sim \mathcal{N}(-20 \text{ dBm}, \sigma_{\Omega}^2), \end{aligned}$$

and where $\sigma_\Omega = 8$ dB, and $\mathcal{N}(\mu, \sigma_\Omega^2)$ refers a Gaussian random variable with mean μ and variance σ_Ω^2 . The sum I is to be approximated as another log-normal random variable, Z , by using the Fenton–Wilkinson method.

- (a) Find the mean and variance of $Z_{(\text{dB})}$.
- (b) Suppose that the received carrier power $C_{(\text{dB})}$ has the distribution

$$C_{(\text{dB})} \sim N(0 \text{ dBm}, \sigma_\Omega^2),$$

and $\sigma_\Omega = 8$ dB. Using your result from part a), what is the distribution of the carrier-to-interference ratio $A_{(\text{dB})} \equiv (C/I)_{(\text{dB})}$?

3.2. Consider the system shown in Fig. 3.10 below. A mobile station lies at a distance of 5, 10, and 15 km from three base stations, BS_i , $i = 1, 2, 3$. BS_2 is the serving base station, while BS_1 and BS_3 are co-channel base stations (co-channel interferers). The propagation path loss follows the model

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

where $\beta = 3.5$, and $\mu_{\Omega_p}(d_o) = 1$ mW at $d_o = 1$ km. The radio links exhibit independent log-normal shadowing with shadow standard deviation $\sigma_\Omega = 8$ dB. Ignore envelope fading.

- (a) Obtain the probability density function of the total interfering power observed at the mobile station in decibel units using the Fenton–Wilkinson method.
- (b) What is the probability density function of the carrier-to-interference ratio observed at the mobile station in decibel units?
- (c) If the carrier-to-interference ratio must be greater than 6 dB for adequate radio link performance, what is the probability of outage?

3.3. The scenario in Fig. 3.11 depicts the worst case CCI situation for the first tier of co-channel interferers on the forward channel. Assume a reuse cluster size of 7 cells, a cell radius of $R = 3$ km, a path loss exponent of $\beta = 3.5$, and a receiver carrier-to-interference threshold $A_{\text{th}} \text{ (dB)} = 10$ dB. Ignore the effect of handoffs and assume that the mobile station must stay connected to the base station in the center cell.

Fig. 3.10 CCI situation for Problem 3.2

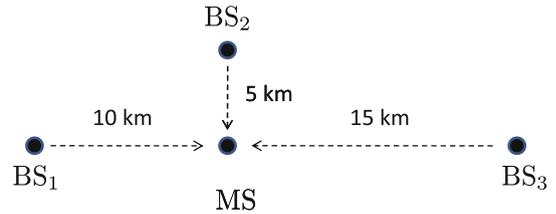
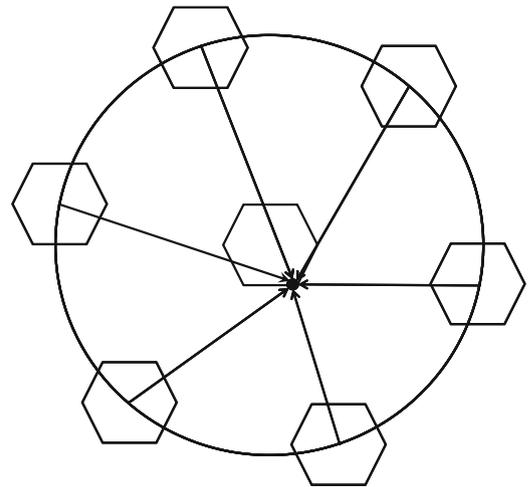


Fig. 3.11 Worst case CCI situation on forward channel in Problem 3.3



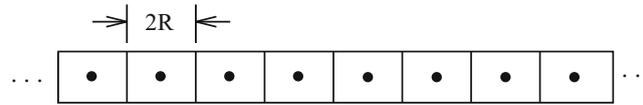


Fig. 3.12 Highway microcell deployment for Problem 3.6

- Using the simple path loss model in (1.6) with $\mu_{\Omega_p \text{ (dBm)}}(d_o) = -10$ dBm at $d_o = 1$ km and a shadow standard deviation $\sigma_{\Omega} = 8$ dB, calculate the probability of outage $O_I(\mathbf{d})$ in (3.46) by using the Fenton–Wilkinson method.
- For $\sigma_{\Omega} = 4$ dB, what is required threshold Λ_{th} such that the probability of outage is less than 1%?
- Repeat b) for $\sigma_{\Omega} = 12$ dB.

3.4. Consider the Fenton–Wilkinson method for approximating the sum of N log-normal random variables

$$I = \sum_{k=0}^N e^{\hat{\Omega}_k},$$

where the $\hat{\Omega}_k$ are independent zero-mean Gaussian random variables with $\sigma_{\hat{\Omega}} = 8$ 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 $N = 2, 3, 4, \dots, 10$.

3.5. This problem uses Monte Carlo simulation techniques to verify the usefulness of the Schwartz & 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 corresponding Gaussian random variables $\Omega_1 \text{ (dB)}$ and $\hat{\Omega}_2 \text{ (dB)}$ are independent and identically distributed with zero mean and variance $\sigma_{\hat{\Omega}}^2$. By using the Schwartz & Yeh method, plot the values of $\mu_Z \text{ (dB)}$ and σ_Z^2 as a function of the variance $\sigma_{\hat{\Omega}}^2$. Repeat for the Fenton–Wilkinson method. Now obtain the same results by using computer simulation and compare the analytical results. What are your conclusions?

3.6. You are asked to design a highway microcell system as shown in Fig. 3.12. Each cell has length $2R$.

- A base station 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.
- Now suppose that directional antennas are used to divide each cell into two sectors with boundaries perpendicular to the highway. Repeat part a).
- 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 CCI outage on a cell boundary? Assume soft handoffs between adjacent cells.

3.7. Derive Eq. (3.50).

3.8. Derive Eq. (3.51).

3.9. Derive Eq. (3.53).

3.10. Derive Eq. (3.74).

3.11. Consider a microcellular environment where a Ricean faded desired 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

Modulation and Power Spectrum

Modulation is the process whereby message information is embedded into a radio frequency carrier. Such information can be transmitted in either the amplitude, frequency, or phase of the carrier, or a combination thereof, in either analog or digital format. Analog modulation schemes include amplitude modulation (AM) and frequency modulation (FM). Analog modulation schemes are still used today for legacy broadcast AM/FM radio and citizens band (CB) radio, but all other communication and broadcast systems now use digital modulation. Digital modulation schemes transmit information using a finite set of waveforms and have a number of advantages over their analog counterparts. Digital modulation is a natural choice for digital sources, e.g., computer communications. Source encoding or data compression techniques can reduce the required transmission bandwidth with a controlled amount of signal distortion. Digitally modulated waveforms are also more robust to channel impairments such as delay and Doppler spread, and co-channel and adjacent channel interference. Finally, encryption and multiplexing is easier with digital modulation schemes.

To achieve high spectral efficiency in wireless systems, signaling schemes are sought that provide power and bandwidth efficient communication. In an information theoretic sense, it is desirable to operate close to the Shannon capacity limit of a channel. This generally requires the use of error control coding along with a jointly designed encoder and modulator. However, this chapter only considers modulation schemes, while the subject of coding and coded modulation is considered in Chap. 8. The bandwidth efficiency of a modulation scheme indicates how much information is transmitted per channel use and is measured in units of bits per second per Hertz of bandwidth (bits/s/Hz). The power efficiency can be measured by the received signal-to-interference-plus-noise ratio (SINR) that is required to achieve reliable communication with a specified bandwidth efficiency in the presence of channel impairments such as delay spread and Doppler spread. In general, modulation techniques for spectrally efficient wireless systems should have the following properties:

- *Compact Power Density Spectrum:* To minimize the effect of adjacent channel interference, the power radiated into the adjacent band is often limited to be 60–80 dB below that in the desired band. This requires modulation techniques having a power spectrum characterized by a narrow main lobe and fast roll-off of side lobes.
- *Robust Communication:* Reliable communication must be achieved in the presence of delay and Doppler spread, adjacent and co-channel interference, and thermal noise. Modulation schemes that promote good power efficiency in the presence of channel impairments are desirable.
- *Envelope Properties:* Portable and mobile devices often employ power efficient nonlinear (Class-C) power amplifiers to minimize battery drain. However, amplifier nonlinearities will degrade the performance of modulation schemes that transmit information in the amplitude of the carrier and/or have a non-constant envelope. To obtain suitable performance, such modulation schemes require a less power efficient linear or a linearized power amplifier. Also, spectral shaping is usually performed prior to up-conversion and nonlinear amplification. To prevent the regrowth of spectral side lobes during nonlinear amplification, modulation schemes having a relatively constant envelope are desirable.

This chapter considers digital modulation techniques that are commonly found in wireless communication systems. Section 4.1 begins the chapter with a mathematical framework for bandpass modulated signals. Section 4.2 discusses Nyquist pulse shaping for ISI-free transmission. Sections 4.3, 4.4, 4.5, 4.6, 4.7, and 4.8 provide a detailed treatment of the various linear and nonlinear digital modulations techniques that are found in wireless systems, including QAM, PSK, $\pi/4$ -DQPSK, orthogonal modulation, OFDM, CPM, GMSK, and others. Finally, Sect. 4.9 considers the power spectrum of digitally modulated signals.

4.1 Representation of Bandpass Modulated Signals

Bandpass modulation schemes refer to modulation schemes that transmit information by using carrier modulation, such that the signal bandwidth is much less than the carrier frequency. A bandpass waveform $s(t)$ can be expressed in terms of its complex envelope as

$$s(t) = \text{Re} \{ \tilde{s}(t) e^{j2\pi f_c t} \}, \quad (4.1)$$

where

$$\tilde{s}(t) = \tilde{s}_I(t) + j\tilde{s}_Q(t) \quad (4.2)$$

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

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \quad (4.3)$$

$$\mathbf{x}_n = (x_n, x_{n-1}, \dots, x_{n-K}), \quad (4.4)$$

where A is the amplitude and $\{x_n\}$ is the sequence of complex data symbols that are chosen from a finite alphabet, and K is the modulator memory order which may be finite or infinite. 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 a *generalized shaping function* whose exact form depends on the type of modulation that is employed. For example, binary phase shift keying (BPSK) with rectangular amplitude pulse shaping has

$$b(t, \mathbf{x}_n) = x_n u_T(t) \quad (4.5)$$

where

$$x_n \in \{-1, +1\} \text{ is the data symbol transmitted at epoch } n$$

$$u_T(t) = u(t) - u(t - T) \text{ is a unit amplitude rectangular pulse of length } T$$

and $u(t)$ is the unit step function. Many types of modulation are considered 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). This is done to streamline the task of finding their power spectra.

By expanding (4.1), the bandpass waveform can also be expressed in the quadrature form

$$s(t) = \tilde{s}_I(t) \cos(2\pi f_c t) - \tilde{s}_Q(t) \sin(2\pi f_c t). \quad (4.6)$$

The waveforms $\tilde{s}_I(t)$ and $\tilde{s}_Q(t)$ are known as the quadrature components $s(t)$, because they modulate the quadrature components of the carrier, $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$, respectively.

Finally $s(t)$ can be expressed in the amplitude-phase form

$$s(t) = a(t) \cos(2\pi f_c t + \phi(t)), \quad (4.7)$$

where

$$a(t) = |\tilde{s}(t)| = \sqrt{\tilde{s}_I^2(t) + \tilde{s}_Q^2(t)} \quad (4.8)$$

$$\phi(t) = \text{Tan}^{-1} \left[\frac{\tilde{s}_Q(t)}{\tilde{s}_I(t)} \right], \quad (4.9)$$

and where $a(t)$ is the amplitude and $\phi(t)$ is the excess phase. The three representations in (4.1), (4.6), and (4.7) are equivalent, but sometimes one representation is more handy than the other two depending on the particular task at hand.

4.1.1 Vector Space Representations

For digital modulation schemes, the bandpass signal that is transmitted at each baud epoch will belong to a finite set of finite energy waveforms with a few exceptions. Let $\{s_1(t), s_2(t), \dots, s_M(t)\}$ be the set of bandpass waveforms, where M denotes the size of the signal set. The corresponding complex envelopes are denoted by $\{\tilde{s}_1(t), \tilde{s}_2(t), \dots, \tilde{s}_M(t)\}$. For now, the development will proceed using the complex envelopes, and the bandpass waveforms will be treated later.

An N -dimensional complex vector space can be defined by a set of N complex orthonormal basis functions $\{\varphi_1(t), \varphi_2(t), \dots, \varphi_N(t)\}$, where

$$\int_{-\infty}^{\infty} \varphi_i(t) \varphi_j^*(t) dt = \delta_{ij} \quad (4.10)$$

and

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}. \quad (4.11)$$

Each waveform $\tilde{s}_m(t)$ in the signal set can be projected onto the set of basis functions to yield a signal vector

$$\tilde{\mathbf{s}}_m = (\tilde{s}_{m1}, \tilde{s}_{m2}, \dots, \tilde{s}_{mN}), \quad m = 1, \dots, M, \quad (4.12)$$

where

$$\tilde{s}_{mi} = \int_{-\infty}^{\infty} \tilde{s}_m(t) \varphi_i^*(t) dt, \quad i = 1, \dots, N. \quad (4.13)$$

The collection of N basis functions is said to constitute a *complete set*, if each waveform in the set $\{\tilde{s}_1(t), \tilde{s}_2(t), \dots, \tilde{s}_M(t)\}$ can be expressed *exactly* as a linear combination of the basis functions. That is,

$$\tilde{s}_m(t) = \sum_{i=1}^N \tilde{s}_{mi} \varphi_i(t), \quad m = 1, \dots, M. \quad (4.14)$$

A systematic procedure for constructing a complete set of basis functions from the set of signal waveforms $\{\tilde{s}_1(t), \tilde{s}_2(t), \dots, \tilde{s}_M(t)\}$ is now described.

4.1.2 Gram–Schmidt Orthonormalization Procedure

Define the inner product between two complex-valued waveforms $u(t)$ and $v(t)$ as

$$(u, v) \triangleq \int_{-\infty}^{\infty} u(t) v^*(t) dt \quad (4.15)$$

and define the norm of the waveform $u(t)$ as

$$\|u\| \triangleq \sqrt{(u, u)}. \quad (4.16)$$

Note that the squared-norm

$$\|u\|^2 \triangleq (u, u) = \int_{-\infty}^{\infty} |u(t)|^2 dt \quad (4.17)$$

is the energy contained in the complex-valued waveform $u(t)$.

Given the finite set of finite energy signals $\{\tilde{s}_1(t), \tilde{s}_2(t), \dots, \tilde{s}_M(t)\}$, a complete set of orthonormal basis functions $\{\varphi_1(t), \varphi_2(t), \dots, \varphi_N(t)\}$ can be constructed by using the following systematic procedure, known as the Gram–Schmidt orthonormalization procedure:

1: Using $\tilde{s}_1(t)$, let

$$g_1(t) = \tilde{s}_1(t)$$

and define

$$\varphi_1(t) = \frac{g_1(t)}{\|g_1\|}. \quad (4.18)$$

2: Using $\tilde{s}_2(t)$, let

$$g_2(t) = \tilde{s}_2(t) - \tilde{s}_{21}\varphi_1(t)$$

where

$$\tilde{s}_{21} = (\tilde{s}_2, \varphi_1)$$

and define

$$\varphi_2(t) = \frac{g_2(t)}{\|g_2\|}. \quad (4.19)$$

3: Using $\tilde{s}_i(t)$, let

$$g_i(t) = \tilde{s}_i(t) - \sum_{j=0}^{i-1} \tilde{s}_{ij}\varphi_j(t)$$

where

$$\tilde{s}_{ij} = (\tilde{s}_i, \varphi_j)$$

and define

$$\varphi_i(t) = \frac{g_i(t)}{\|g_i\|}. \quad (4.20)$$

4: Repeat Step 3 in a recursive fashion until all elements of the waveform set $\{\tilde{s}_1(t), \tilde{s}_2(t), \dots, \tilde{s}_M(t)\}$ have been used.

If one or more steps in the above recursion yields $g_i(t) = 0$, then the corresponding waveform $\tilde{s}_i(t)$ can already be expressed exactly in terms of the basis functions already generated. Consequently, the waveform $\tilde{s}_i(t)$ will not yield an additional basis function and the procedure continues to the next waveform in the set, $\tilde{s}_{i+1}(t)$. In the end, a complete set of N , $1 \leq N \leq M$ complex orthonormal basis functions $\{\varphi_1(t), \varphi_2(t), \dots, \varphi_N(t)\}$ corresponding to the non-zero $g_i(t)$ will be obtained. The dimensionality of the complex vector space N is equal to M if and only if the original set of waveforms $\{\tilde{s}_1(t), \tilde{s}_2(t), \dots, \tilde{s}_M(t)\}$ are linearly independent, i.e., none of the waveforms in the set is a linear combination of the other waveforms in the set.

Example 4.1. Construct an orthonormal basis set for the set of waveforms shown in Fig. 4.1.

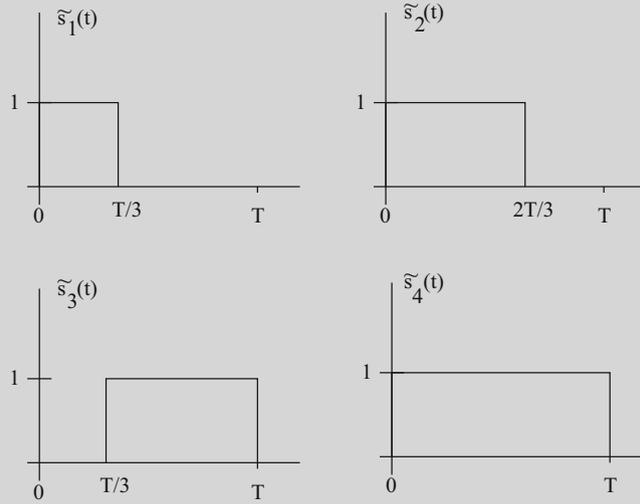


Fig. 4.1 Signal set $\{\tilde{s}_i(t)\}_{i=1}^4$ for Example 4.1

1: Let $g_1(t) = \tilde{s}_1(t)$. Then

$$\varphi_1(t) = \frac{g_1(t)}{\|g_1\|} = \begin{cases} \sqrt{3/T}, & 0 \leq t \leq T/3 \\ 0, & \text{else} \end{cases}.$$

2: Let $g_2(t) = \tilde{s}_2(t) - \tilde{s}_{21}\varphi_1(t)$, where

$$\tilde{s}_{21} = (\tilde{s}_2, \varphi_1) = \int_0^T \tilde{s}_2(t)\varphi_1^*(t)dt = \sqrt{T/3}.$$

Then

$$\varphi_2(t) = \frac{g_2(t)}{\|g_2\|} = \begin{cases} \sqrt{3/T}, & T/3 \leq t \leq 2T/3 \\ 0, & \text{else} \end{cases}$$

3: Let $g_3(t) = \tilde{s}_3(t) - \tilde{s}_{31}\varphi_1(t) - \tilde{s}_{32}\varphi_2(t)$, where

$$\tilde{s}_{31} = (\tilde{s}_3, \varphi_1) = \int_0^T \tilde{s}_3(t)\varphi_1^*(t)dt = 0$$

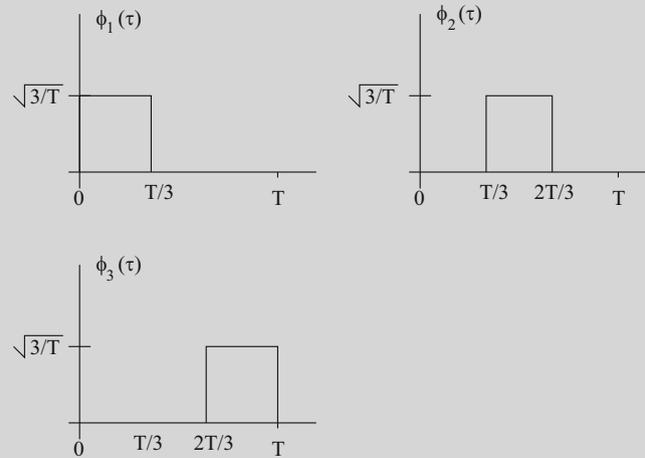
$$\tilde{s}_{32} = (\tilde{s}_3, \varphi_2) = \int_0^T \tilde{s}_3(t)\varphi_2^*(t)dt = \sqrt{T/3}.$$

Then

$$\varphi_3(t) = \frac{g_3(t)}{\|g_3\|} = \begin{cases} \sqrt{3/T}, & 2T/3 \leq t \leq T \\ 0, & \text{else} \end{cases}.$$

4: Let $g_4(t) = \tilde{s}_4(t) - \tilde{s}_{41}\varphi_1(t) - \tilde{s}_{42}\varphi_2(t) - \tilde{s}_{43}\varphi_3(t)$, where

(continued)

Example 4.1 (continued)**Fig. 4.2** Orthonormal basis functions $\{\varphi_i(t)\}_{i=1}^3$ for Example 4.1

$$\tilde{s}_{41} = (\tilde{s}_4, \varphi_1) = \int_0^T \tilde{s}_4(t) \varphi_1^*(t) dt = \sqrt{T/3}$$

$$\tilde{s}_{42} = (\tilde{s}_4, \varphi_2) = \int_0^T \tilde{s}_4(t) \varphi_2^*(t) dt = \sqrt{T/3}$$

$$\tilde{s}_{43} = (\tilde{s}_4, \varphi_3) = \int_0^T \tilde{s}_4(t) \varphi_3^*(t) dt = \sqrt{T/3}.$$

But $g_4(t) = 0$, meaning that $\tilde{s}_4(t)$ is linearly dependent on $\{\tilde{s}_1(t), \tilde{s}_2(t), \tilde{s}_3(t)\}$, and $\tilde{s}_4(t)$ does not yield an additional basis function.

The set of orthonormal basis functions obtained from the above procedure is shown in Fig. 4.2, and they define a 3-dimensional vector space.

Each $\tilde{s}_i(t)$ in the signal set can be expressed as a linear combination of the basis functions, according to (4.14), and the corresponding signal vectors in (4.12) can be constructed. For this example, the signal vectors are

$$\tilde{\mathbf{s}}_1 = (\sqrt{T/3}, 0, 0)$$

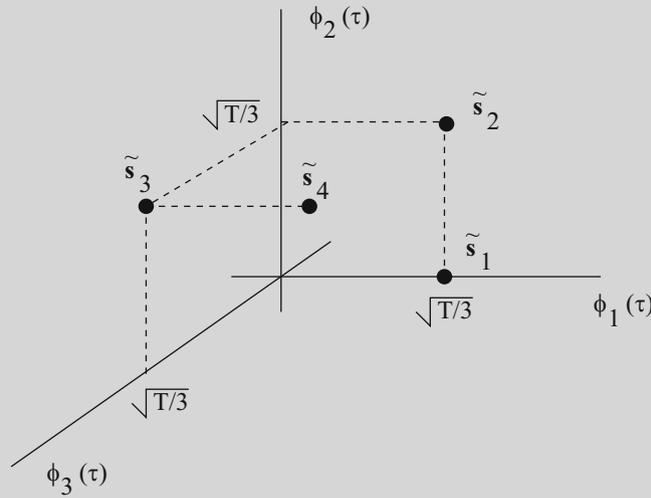
$$\tilde{\mathbf{s}}_2 = (\sqrt{T/3}, \sqrt{T/3}, 0)$$

$$\tilde{\mathbf{s}}_3 = (0, \sqrt{T/3}, \sqrt{T/3})$$

$$\tilde{\mathbf{s}}_4 = (\sqrt{T/3}, \sqrt{T/3}, \sqrt{T/3}).$$

The set of signal vectors $\{\tilde{\mathbf{s}}_i\}$ can be plotted in the 3-D vector space defined by the set of orthonormal basis functions $\{\varphi_i(t)\}$ as shown in Fig. 4.3. The plotted set of signal vectors is sometimes called a signal constellation.

(continued)

Example 4.1 (continued)**Fig. 4.3** Signal constellation for Example 4.1

In the above development, the Gram–Schmidt orthonormalization procedure was applied to the set of complex envelopes $\{\tilde{s}_1(t), \tilde{s}_2(t), \dots, \tilde{s}_M(t)\}$ to produce a complete set of $N \leq M$ complex basis functions $\{\varphi_1(t), \varphi_2(t), \dots, \varphi_N(t)\}$, where N is the dimension of the *complex* vector space. By using the exact same Gram–Schmidt orthonormalization procedure, a complete set of N real-valued orthonormal basis functions $\{\varphi_1(t), \varphi_2(t), \dots, \varphi_N(t)\}$ can be obtained from the real-valued bandpass waveforms $\{s_1(t), s_2(t), \dots, s_M(t)\}$, expressed in quadrature representation (4.6), and where N is the dimension of the *real* vector space. In this case, the complex conjugates in (4.10) and (4.13) may be omitted since all waveforms are real-valued. By using the real-valued basis functions, each bandpass waveform $s_m(t)$ can be projected onto the set of real-valued basis functions to yield the set of signal vectors

$$\mathbf{s}_m = (s_{m1}, s_{m2}, \dots, s_{mN}), \quad m = 1, \dots, M, \quad (4.21)$$

where

$$s_{mi} = \int_{-\infty}^{\infty} s_m(t) \varphi_i(t) dt, \quad i = 1, \dots, N, \quad (4.22)$$

and

$$s_m(t) = \sum_{i=1}^N s_{mi} \varphi_i(t), \quad m = 1, \dots, M. \quad (4.23)$$

Note that the set of orthonormal basis functions and the dimensionality of the vector space needed to represent the bandpass waveforms and their corresponding complex envelopes are related. The complex-valued basis functions each define a 2-dimensional complex plane. So for one-dimensional complex-valued or two-dimensional real-valued signal constellations each complex basis function will yield two orthogonal real-valued basis functions. If the complex envelopes happen to be real-valued, then just a single real-valued basis function, different in each case, is required to represent both the real-valued bandpass waveforms and their complex envelopes.

4.1.3 Signal Energy and Correlations

Define the inner (dot) product between two length- N complex vectors \mathbf{u} and \mathbf{v} as

$$\mathbf{u} \cdot \mathbf{v}^* \triangleq \sum_{i=1}^N u_i v_i^* \quad (4.24)$$

and the norm (length) of the vector \mathbf{u} as

$$\|\mathbf{u}\| \triangleq \sqrt{\mathbf{u} \cdot \mathbf{u}^*} = \sqrt{\sum_{i=1}^N |u_i|^2}. \quad (4.25)$$

If the vectors happen to be real, the complex conjugates and absolute values can be neglected.

Consider the set of bandpass waveforms

$$s_m(t) = \text{Re} \{ \tilde{s}_m(t) e^{j2\pi f_c t} \}, \quad m = 1, \dots, M. \quad (4.26)$$

The energy in the bandpass waveform $s_m(t)$ is

$$E_m = (s_m, s_m) = \int_{-\infty}^{\infty} s_m^2(t) dt. \quad (4.27)$$

Using the amplitude-phase representation of a bandpass waveform in (4.7), and the trig identity $\cos^2(x) = \frac{1}{2}(1 + \cos(2x))$, gives

$$\begin{aligned} E_m &= \int_{-\infty}^{\infty} (|\tilde{s}_m(t)| \cos(2\pi f_c t + \phi_m(t)))^2 dt \\ &= \frac{1}{2} \int_{-\infty}^{\infty} |\tilde{s}_m(t)|^2 dt + \frac{1}{2} \int_{-\infty}^{\infty} |\tilde{s}_m(t)|^2 \cos(4\pi f_c t + 2\phi_m(t)) dt \\ &\approx \frac{1}{2} \int_{-\infty}^{\infty} |\tilde{s}_m(t)|^2 dt \\ &= \frac{1}{2} (\tilde{s}_m, \tilde{s}_m). \end{aligned} \quad (4.28)$$

where $\phi_m(t) = \text{Tan}^{-1} [\tilde{s}_{m,Q}(t)/\tilde{s}_{m,I}(t)]$. The above approximation is valid 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 bandpass modulated signals, this condition is equivalent to $f_c T \gg 1$ so that there are many cycles of the carrier in the baud period T . This condition is satisfied in most wireless systems.

By using the vector representation of the bandpass waveforms in (4.21)–(4.23), it follows that the energy in the bandpass waveform $s_m(t)$ is

$$E_m = \int_{-\infty}^{\infty} \left(\sum_{i=1}^N s_{m_i} \varphi_i(t) \right)^2 dt = \sum_{i=1}^N s_{m_i}^2 = \|\mathbf{s}_m\|^2, \quad (4.29)$$

where the second equality follows from the orthonormal property of the basis functions in (4.10). Notice that the energy in $s_m(t)$ is equal to the squared-norm (length) of the corresponding signal vector \mathbf{s}_m . Likewise, by using the vector representation of the corresponding complex envelope, the energy in the bandpass waveform $s_m(t)$ is also equal to

$$E_m = \frac{1}{2} \int_{-\infty}^{\infty} \left| \sum_{i=1}^N \tilde{s}_{m_i} \varphi_i(t) \right|^2 dt = \frac{1}{2} \sum_{i=1}^N |\tilde{s}_{m_i}|^2 = \frac{1}{2} \|\tilde{\mathbf{s}}_m\|^2. \quad (4.30)$$

Note that the energy in the bandpass waveform is one-half the energy in its complex envelope. This is due to the carrier modulation.

The correlation between the bandpass waveforms $s_m(t)$ and $s_k(t)$ is defined as

$$\begin{aligned}\rho_{km} &= \frac{1}{\sqrt{E_k E_m}} \int_{-\infty}^{\infty} s_m(t) s_k(t) dt \\ &= \frac{\mathbf{s}_m \cdot \mathbf{s}_k}{\|\mathbf{s}_m\| \|\mathbf{s}_k\|} \\ &= \operatorname{Re} \left\{ \frac{\tilde{\mathbf{s}}_m \cdot \tilde{\mathbf{s}}_k^*}{\|\tilde{\mathbf{s}}_m\| \|\tilde{\mathbf{s}}_k\|} \right\}.\end{aligned}\quad (4.31)$$

Finally, the squared Euclidean distance between the bandpass waveforms $s_k(t)$ and $s_m(t)$ is

$$\begin{aligned}d_{km}^2 &= \int_{-\infty}^{\infty} (s_m(t) - s_k(t))^2 dt \\ &= \|\mathbf{s}_m - \mathbf{s}_k\|^2 \\ &= \frac{1}{2} \|\tilde{\mathbf{s}}_m - \tilde{\mathbf{s}}_k\|^2.\end{aligned}\quad (4.32)$$

The results in (4.31) and (4.32) may be obtained using (4.14) and (4.23) along with the orthonormal property of the basis functions.

4.2 Nyquist Pulse Shaping

Consider a modulation scheme where the transmitted complex envelope has the

$$\tilde{s}(t) = A \sum_n x_n h_a(t - nT) \quad (4.33)$$

where $h_a(t)$ is a real-valued amplitude shaping pulse, $\{x_n\}$ is a complex data symbol sequence, and T is the baud period. As will be discussed in Chap. 5, the receiver usually employs a filter that is *matched* to the transmitted pulse, having the form $h_r(t) = h_a(T_o - t)$, where T_o is the duration of the amplitude shaping pulse $h_a(t)$. An overall pulse can be defined that is the cascade of the transmitted pulse $h_a(t)$ and the receiver matched filter $h_r(t)$ as $p(t) = h_a(t) * h_a(T_o - t)$, where $*$ denotes the operation of convolution.

For the time being, consider an ideal channel having impulse response $g(t, \tau) = \delta(\tau)$. In the absence of thermal noise in the receiver, the waveform at the output of the receiver matched filter is

$$\tilde{y}(t) = A \sum_n x_n p(t - nT). \quad (4.34)$$

Now suppose the received complex envelope $\tilde{y}(t)$ is sampled once every T seconds to yield the sample sequence $\{\tilde{y}_k\}$, where

$$\tilde{y}_k = \tilde{y}(kT + t_o) = A \sum_n x_n p(kT + t_o - nT) \quad (4.35)$$

and 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 a non-zero timing offset will be treated later. When $t_o = 0$

$$\tilde{y}_k = A \sum_n x_n p_{k-n} = A x_k p_0 + A \sum_{n \neq k} x_n p_{k-n}, \quad (4.36)$$

where $p_m = p(mT)$ is the sampled overall pulse. The first term in (4.36) is equal to the data symbol transmitted at the k th baud epoch, scaled by the factor $A p_0$. The second term is the contribution of all other data symbols on the sample \tilde{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

$$p_k = \delta_{k0} p_0, \quad (4.37)$$

where δ_{jk} is the Dirac delta function defined in (4.11). This requirement is known as the (first) Nyquist criterion. Under this condition,

$$\tilde{y}_k = Ax_k p_0. \quad (4.38)$$

An equivalent frequency domain requirement is now derived by showing that the continuous-time pulse $p(t)$ satisfies the condition $p_k = \delta_{k0} p_0$ if and only if

$$P_\Sigma(f) \triangleq \frac{1}{T} \sum_{n=-\infty}^{\infty} P\left(f + \frac{n}{T}\right) = p_0. \quad (4.39)$$

The function $P_\Sigma(f)$ is called the folded spectrum, and ISI is avoided if and only if the folded spectrum is *flat*, i.e., it assumes a constant value. To prove the above property, the inverse Fourier transform can be used to write

$$\begin{aligned} p_k &= \int_{-\infty}^{\infty} P(f) e^{j2\pi f k T} df \\ &= \sum_{n=-\infty}^{\infty} \int_{(2n-1)/2T}^{(2n+1)/2T} P(f) e^{j2\pi k f T} df \\ &= \sum_{n=-\infty}^{\infty} \int_{-1/2T}^{1/2T} P\left(f' + \frac{n}{T}\right) e^{j2\pi k \left(f' + \frac{n}{T}\right) T} df' \\ &= \int_{-1/2T}^{1/2T} \left[\sum_{n=-\infty}^{\infty} P\left(f + \frac{n}{T}\right) \right] e^{j2\pi f k T} df \\ &= T \int_{-1/2T}^{1/2T} P_\Sigma(f) e^{j2\pi f k T} df. \end{aligned} \quad (4.40)$$

Since $P_\Sigma(f)$ is periodic with period $1/T$, it follows that the last line in (4.40) represents a Fourier series analysis equation except for the sign of the exponential term. Therefore, $\{p_{-k}\}$ and $P_\Sigma(f)$ are a Fourier series pair, and $P_\Sigma(f)$ can be constructed from $\{p_{-k}\}$ by using the Fourier series synthesis equation, viz.,

$$P_\Sigma(f) = \sum_{k=-\infty}^{\infty} p_{-k} e^{j2\pi k f T} = \sum_{k=-\infty}^{\infty} p_k e^{-j2\pi k f T}. \quad (4.41)$$

To prove that (4.39) is a sufficient condition for ISI-free transmission, suppose that (4.39) holds true. Then $P_\Sigma(f) = p_0 T$ and from the last line of (4.40)

$$p_k = \int_{-1/2T}^{1/2T} e^{j2\pi f k T} p_0 T df = \frac{\sin \pi k}{\pi k} p_0 = \delta_{k0} p_0. \quad (4.42)$$

To prove that (4.39) is a necessary condition for ISI-free transmission, suppose that $p_k = p_0 \delta_{k0}$ holds true. Then from (4.41) $P_\Sigma(f) = p_0$, and the folded spectrum must be flat.

The requirement on the folded spectrum in (4.39) allows us to design pulses in the frequency domain that will admit ISI-free transmission. First consider a pulse having the Fourier transform

$$P_N(f) = T \text{rect}(fT), \quad (4.43)$$

where

$$\text{rect}(fT) = \begin{cases} 1, & |f| \leq \frac{1}{2T} \\ 0, & \text{elsewhere} \end{cases}. \quad (4.44)$$

This pulse has a flat folded spectrum. The corresponding time domain pulse

$$p_N(t) = \text{sinc}(t/T) \quad (4.45)$$

satisfies 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. The edge frequency $f = 1/2T$ is called the Nyquist frequency.

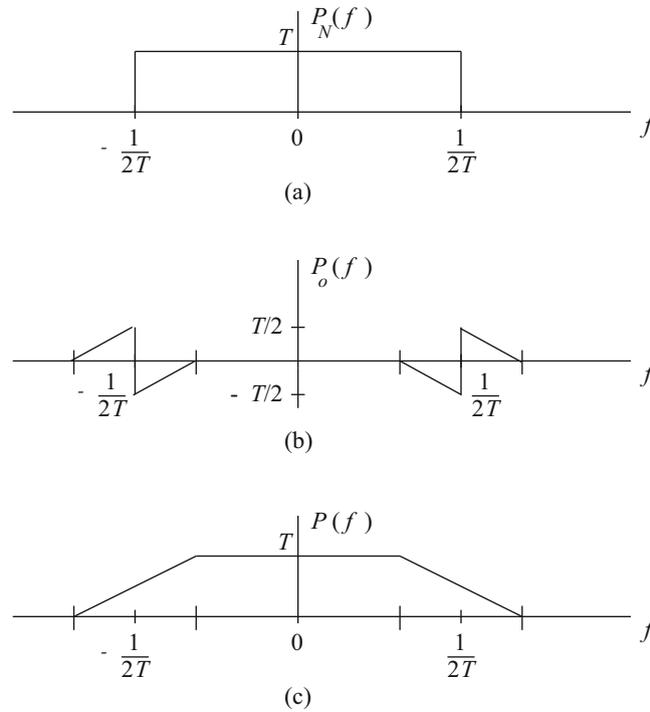


Fig. 4.4 Construction of pulses satisfying the (first) Nyquist criterion

The effect of the sampling or timing offset t_o is now examined with the aid of the ideal Nyquist pulse. With a timing offset

$$\begin{aligned}\tilde{y}_k &= A \sum_n x_n \text{sinc}((kT + t_o - nT)/T) \\ &= Ax_k \text{sinc}(t_o/T) + A \sum_{n \neq k} x_n \text{sinc}((kT + t_o - nT)/T)\end{aligned}\quad (4.46)$$

Observe that the ISI term is non-zero when the timing offset is non-zero. In fact, with an ideal Nyquist pulse, the ISI term is not absolutely summable as shown in Problem 4.1. This is because the tails of the ideal Nyquist pulse in (4.45) decay in time as $1/t$. To reduce this sensitivity to symbol timing errors, pulses need to be designed that satisfy the first Nyquist criterion while having tails that decay faster than $1/t$.

The construction of other Nyquist pulses starts with the ideal Nyquist pulse, $P_N(f)$, shown in Fig. 4.4a. To the pulse $P_N(f)$, a “transmittance” function $P_o(f)$ is added as shown in Fig. 4.4b. The transmittance function must have skew symmetry about the Nyquist frequency $1/2T$, and any skew symmetric function will do. The resulting Nyquist pulse $P(f)$ is shown in Fig. 4.4c. Clearly, the folded spectrum $P_\Sigma(f)$ is flat if the transmittance function is skew symmetric about the Nyquist frequency $1/2T$. The corresponding time domain pulse $p(t)$ can be obtained from the inverse Fourier transform of resulting $P(f)$. Notice that the pulse $P(f)$ takes up additional bandwidth, but the bandwidth expansion will result in a time domain pulse $p(t)$ having tails that decay faster in time than the ideal Nyquist pulse.

4.2.1 Raised Cosine and Root Raised Cosine Pulse

The raised cosine pulse is defined in the frequency domain by

$$P(f) = \begin{cases} T & , 0 \leq |f| \leq (1 - \beta)/2T \\ \frac{T}{2} \left[1 - \sin\left(\frac{\pi|f|T}{\beta} - \frac{\pi}{2\beta}\right) \right] & , (1 - \beta)/2T \leq |f| \leq (1 + \beta)/2T \\ 0 & , |f| \geq (1 + \beta)/2T \end{cases} \quad (4.47)$$

The bandwidth of the raised cosine pulse is $(1 + \beta)/2T$, where the parameter β , $0 \leq \beta \leq 1$ is called the roll-off factor and controls the bandwidth expansion. The term “raised cosine” comes from the fact that pulse spectrum $P(f)$ with $\beta = 1$ has a “raised cosine” shape, i.e., with $\beta = 1$

$$P(f) = \frac{T}{2} [1 + \cos(\pi f T)], \quad 0 \leq |f| \leq 1/T. \quad (4.48)$$

The inverse Fourier transform of $P(f)$ in (4.47) gives the corresponding time domain pulse

$$p(t) = \frac{\sin(\pi t/T)}{\pi t/T} \frac{\cos(\beta \pi t/T)}{1 - (2\beta t/T)^2}. \quad (4.49)$$

For $\beta = 0$, $p(t)$ reduces to the ideal Nyquist pulse in (4.45). Notice that the tails of the raised cosine pulse decay as $1/t^3$.

As mentioned before, the overall pulse produced by the cascade of the transmitter and receiver matched filters is $p(t) = h_a(t) * h_a(T_o - t)$. It follows that the Fourier transform of $p(t)$ is $P(f) = H_a(f)H_a^*(f)e^{-j2\pi f T_o} = |H_a(f)|^2 e^{-j2\pi f T_o}$. Hence, both the transmitted pulse and receiver matched filter have the same magnitude response $|H_a(f)| = |P(f)|^{1/2}$, where $P(f)$ is defined in (4.47). If the overall pulse $p(t)$ is a raised cosine pulse with $P(f)$ defined in (4.47), then the pulse $h_a(t)$ is said to be a root raised cosine pulse. Taking the inverse Fourier transform of $|H_a(f)| = \sqrt{T}|P(f)|^{1/2}$ gives the root raised cosine pulse

$$h_a(t) = \begin{cases} 1 - \beta + 4\beta/\pi & , t = 0 \\ (\beta/\sqrt{2}) ((1 + 2/\pi) \sin(\pi/4\beta) + (1 - 2/\pi) \cos(\pi/4\beta)) & , t = \pm T/4\beta \\ \frac{4\beta(t/T) \cos((1+\beta)\pi t/T) + \sin((1-\beta)\pi t/T)}{\pi(t/T)(1-(4\beta t/T)^2)} & , \text{elsewhere} \end{cases} \quad (4.50)$$

For $\beta = 0$, the root raised cosine pulse reduces to the sinc pulse

$$h_a(t) = \text{sinc}(t/T). \quad (4.51)$$

The raised cosine and root raised cosine pulses corresponding to $\beta = 0.5$ are shown in Fig. 4.5. Strictly speaking, the root-raised cosine pulse in (4.50) is non-causal. Therefore, in practice, a truncated and time shifted approximation of the pulse must be used. For example, in Fig. 4.5 the pulse is truncated to length $6T$ and right time-shifted by $3T$ seconds to yield a causal pulse. The time-shifting makes the pulse have a linear phase response, while the pulse truncation will result in a pulse that is no longer strictly bandlimited. Finally, the raised cosine pulse is a Nyquist pulse having equally spaced zero crossings at the baud period T , while the root-raised cosine pulse by itself is not a Nyquist pulse and does not have equally spaced zero crossings.

4.3 Quadrature Amplitude Modulation

Quadrature amplitude modulation (QAM) is a bandwidth efficient modulation scheme that is used in numerous wireless standards. With QAM, the complex envelope of the transmitted waveform is

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \quad (4.52)$$

where

$$b(t, \mathbf{x}_n) = x_n h_a(t) \quad (4.53)$$

$h_a(t)$ is the amplitude shaping pulse (very often chosen as a root-raised cosine pulse), and $x_n = x_{I,n} + jx_{Q,n}$ is the complex-valued data symbol that is transmitted at baud epoch n . It is apparent that both the amplitude and the excess phase of a QAM waveform depend on the complex data symbols. QAM has the advantage of high bandwidth efficiency, but amplifier nonlinearities will degrade its performance due to the non-constant envelope.

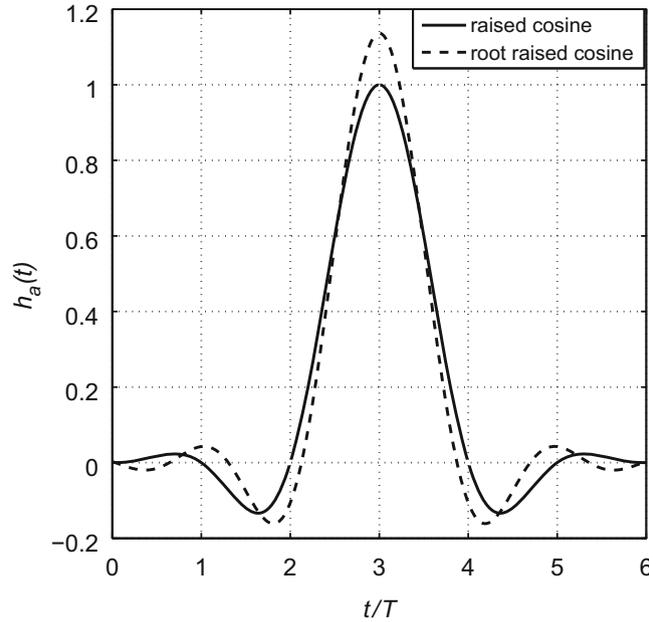


Fig. 4.5 Raised cosine and root raised cosine pulses with roll-off factor $\beta = 0.5$. The pulses are truncated to length $6T$ and right time shifted by $3T$ seconds to yield causal pulses

The set of possible QAM waveforms that are transmitted at each baud epoch have the complex envelopes

$$\tilde{s}_m(t) = Ax_m h_a(t) \quad m = 1, 2, \dots, M. \quad (4.54)$$

To obtain the vector representation of the complex envelopes $\tilde{s}_m(t)$, $m = 1, 2, \dots, M$, the basis function

$$\varphi_1(t) = \sqrt{\frac{A^2}{2E_h}} h_a(t) \quad (4.55)$$

can be used, where

$$E_h = \frac{A^2}{2} \int_{-\infty}^{\infty} h_a^2(t) dt \quad (4.56)$$

is the energy in the bandpass pulse $Ah_a(t) \cos 2\pi f_c t$ under the condition $f_c T \gg 1$. Using this basis function

$$\tilde{s}_m(t) = \sqrt{2E_h} x_m \varphi_1(t), \quad (4.57)$$

and the QAM complex signal vectors are the complex-valued scalars¹

$$\tilde{s}_m = \sqrt{2E_h} x_m, \quad m = 1, 2, \dots, M. \quad (4.58)$$

4.3.1 QAM Signal Constellations

A variety of QAM signal constellations may be constructed. Square QAM constellations can be constructed when M is an even power of 2 by choosing $x_{I,m}, x_{Q,m} \in \{\pm 1, \pm 3, \dots, \pm(\hat{M} - 1)\}$, where $\hat{M} = \sqrt{M}$. 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 an odd power of 2, the QAM signal constellation is not square. In this case,

¹Note that the dimensionality of the complex signal-space is $N = 1$.

Fig. 4.6 Complex signal-space diagram for square QAM constellations

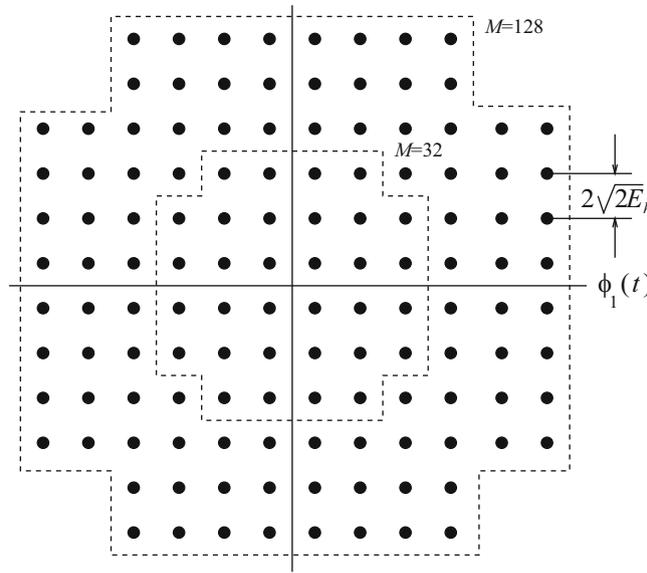
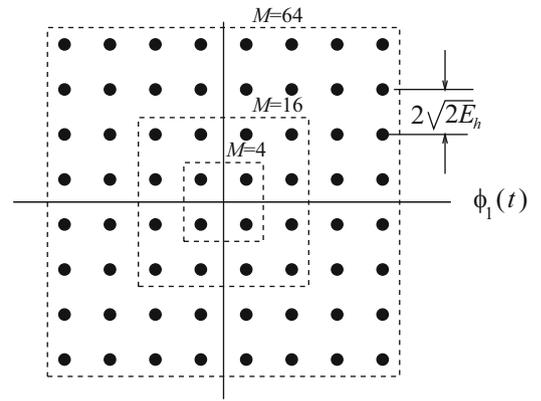


Fig. 4.7 Complex signal-space diagram for cross QAM constellations

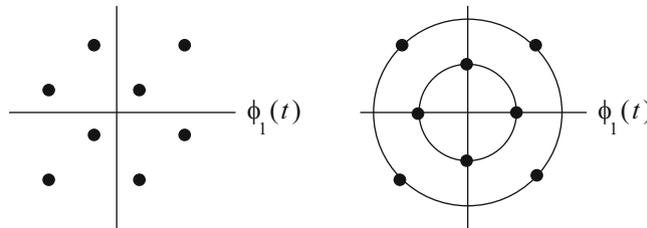


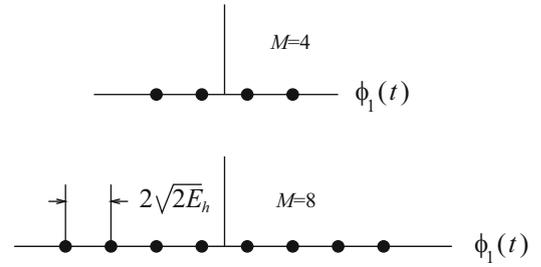
Fig. 4.8 Complex signal-space diagrams for 8-QAM constellations

the constellation assumes the shape of a cross in an attempt to minimize the average energy in the constellation for a given minimum Euclidean distance between signal vectors. Examples of QAM “cross constellations” are shown in Fig. 4.7. Other types of QAM constellations are possible as well. Figure 4.8 shows two different 8-QAM constellations.

4.3.2 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 RF carrier. This can be accomplished by using real data symbols $x_m = x_{I,m}$, where

Fig. 4.9 Complex signal-space diagram for the 4- and 8-PAM constellations



$x_{l,m} \in \{\pm 1, \pm 3, \dots, \pm(M-1)\}$. Equivalently, the set of data symbols can be generated according to $x_{l,m} \in \{2m-1-M : m = 1, 2, \dots, M\}$. The PAM complex signal vectors are the real-valued scalars

$$\tilde{s}_m = \sqrt{2E_h}(2m-1-M), \quad m = 1, 2, \dots, M. \quad (4.59)$$

Typical 4- and 8-PAM signal constellations are shown in Fig. 4.9.

4.4 Phase Shift Keying

The complex envelope of a phase shift keying (PSK) signal has the form

$$\tilde{s}(t) = A \sum_n b(t-nT, \mathbf{x}_n), \quad (4.60)$$

where

$$b(t, \mathbf{x}_n) = x_n h_a(t), \quad (4.61)$$

$h_a(t)$ is the amplitude shaping pulse, $x_n = e^{j\theta_n}$ and the phases θ_n take on values from the set

$$\theta_n \in \left\{ \frac{2\pi}{M}n, \quad n = 0, 1, \dots, M-1 \right\}, \quad (4.62)$$

The set of possible PSK waveforms that are transmitted at each baud epoch have the complex envelopes

$$\tilde{s}_m(t) = A h_a(t) e^{j\theta_m}, \quad m = 1, 2, \dots, M. \quad (4.63)$$

Using the basis function in (4.55)

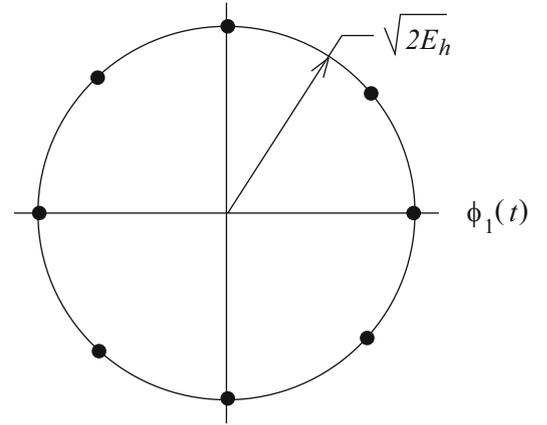
$$\tilde{s}_m(t) = \sqrt{2E_h} e^{j\theta_m} \varphi_0(t), \quad m = 1, \dots, M, \quad (4.64)$$

and the PSK complex signal vectors are the complex-valued scalars

$$\tilde{s}_m = \sqrt{2E_h} e^{j\theta_m}, \quad m = 1, 2, \dots, M. \quad (4.65)$$

The complex signal-space diagram for 8-PSK is shown in Fig. 4.10. Recall the energy in a PSK bandpass waveform is equal to one-half the squared length of its complex signal vector. It follows that the PSK bandpass waveforms all have equal energy E_h .

Fig. 4.10 Complex signal-space diagram for the 8-PSK constellation



4.4.1 Offset QPSK (OQPSK)

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 $\pm 90^\circ$ or 180° shifts of the excess phase from one baud interval to the next. With offset (or staggered) QPSK (OQPSK), the complex envelope is

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \quad (4.66)$$

where

$$b(t, \mathbf{x}_n) = x_{I,n}h_a(t) + jx_{Q,n}h_a(t - T_b) \quad (4.67)$$

and $T_b = T/2$ is the bit interval. With OQPSK signals, the possibility of 180° shifts of the excess phase is eliminated. In fact, the excess phase can only change by $\pm 90^\circ$ every T_b seconds. With OQPSK, the amplitude shaping pulse $h_a(t)$ is often chosen to be the root-raised cosine pulse in (4.50) to yield a compact power spectrum.

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 excess phase transitions. The exact excess phase trajectories depend on the amplitude shaping function. Note that the excess phase trajectories with OQPSK do not pass through the origin, while those with QPSK do. This property reduces the peak-to-average power ratio (PAPR) of the OQPSK waveform as compared to the QPSK waveform, where the PAPR is defined as

$$\text{PAPR} = \lim_{T \rightarrow \infty} \frac{\max_{0 \leq t \leq T} |\tilde{s}(t)|^2}{T^{-1} \int_0^T |\tilde{s}(t)|^2 dt}.$$

A lower PAPR makes the OQPSK waveform less sensitive to power amplifier nonlinearities than the QPSK waveform. For this reason OQPSK waveforms have been used for satellite communication links, where the satellite transponders use power efficient nonlinear amplifiers.

4.4.2 $\pi/4$ -DQPSK

The $\pi/4$ phase-shifted differential quadrature phase shift keyed ($\pi/4$ -DQPSK) modulation scheme was used in some types of now extinct second generation cellular telephone systems. Similar to QPSK and OQPSK, $\pi/4$ -DQPSK transmits 2 bits per modulated symbol. However, unlike QPSK and OQPSK where information is transmitted in the absolute excess phase, $\pi/4$ -DQPSK transmits information in the differential excess phase, and one of 8 absolute excess phases are transmitted at each baud epoch.

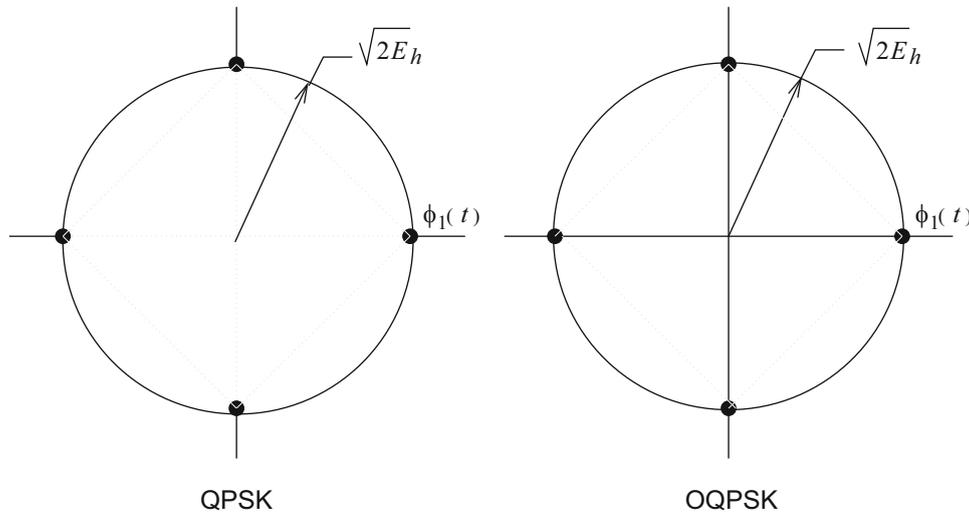


Fig. 4.11 Complex signal-space diagram QPSK and OQPSK signals

Let θ_n be the absolute excess phase for the n th data symbol, and let $\Delta\theta_n = \theta_n - \theta_{n-1}$ be the differential excess phase. With $\pi/4$ -DQPSK, the differential excess phase is related to the quaternary data sequence $\{x_n\}$, $x_n \in \{\pm 1, \pm 3\}$ through the mapping

$$\Delta\theta_n = x_n \frac{\pi}{4}. \quad (4.68)$$

Notice that the excess phase differences are $\pm\pi/4$ and $\pm 3\pi/4$. The complex envelope of the $\pi/4$ -DQPSK signal is

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n), \quad (4.69)$$

where

$$\begin{aligned} b(t, \mathbf{x}_n) &= h_a(t) \exp \left\{ j \left(\theta_{n-1} + x_n \frac{\pi}{4} \right) \right\} \\ &= h_a(t) \exp \left\{ j \frac{\pi}{4} \left(\sum_{k=-\infty}^{n-1} x_k + x_n \right) \right\}. \end{aligned} \quad (4.70)$$

The summation in the exponent of (4.70) represents the accumulated excess phase, while the last term is the excess phase increment due to the n th data symbol. The absolute excess 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 (4.50).

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 PAPR of the complex envelope, making the $\pi/4$ -DQPSK waveform less sensitive to power amplifier nonlinearities. Moreover, since the root-raised cosine pulse is not a Nyquist pulse, the phase trajectories do not go through the signal constellation points either. Finally, observe that the excess phase of $\pi/4$ -DQPSK changes by $\pm\pi/4$ or $\pm 3\pi/4$ radians during every baud interval. This property makes symbol synchronization easier with $\pi/4$ -DQPSK as compared to QPSK.

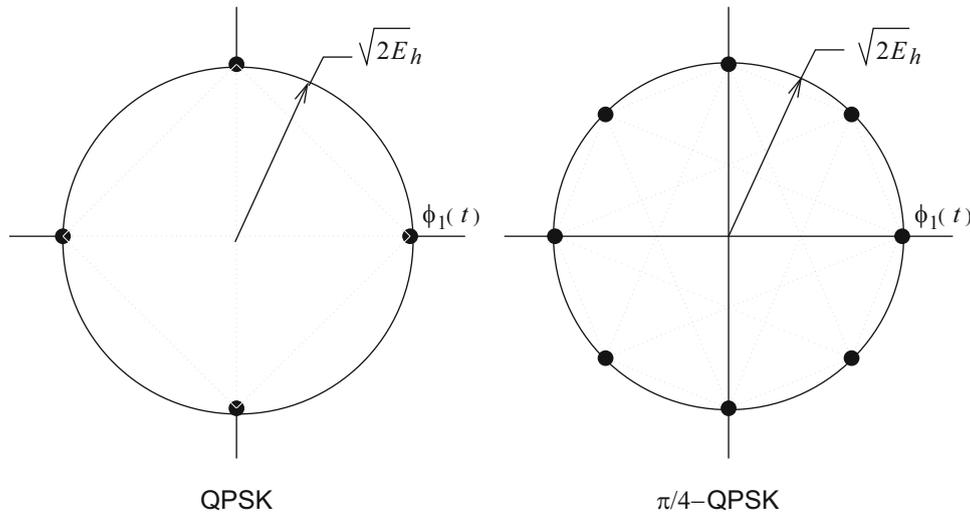


Fig. 4.12 Complex signal-space diagram QPSK and $\pi/4$ -DQPSK signals

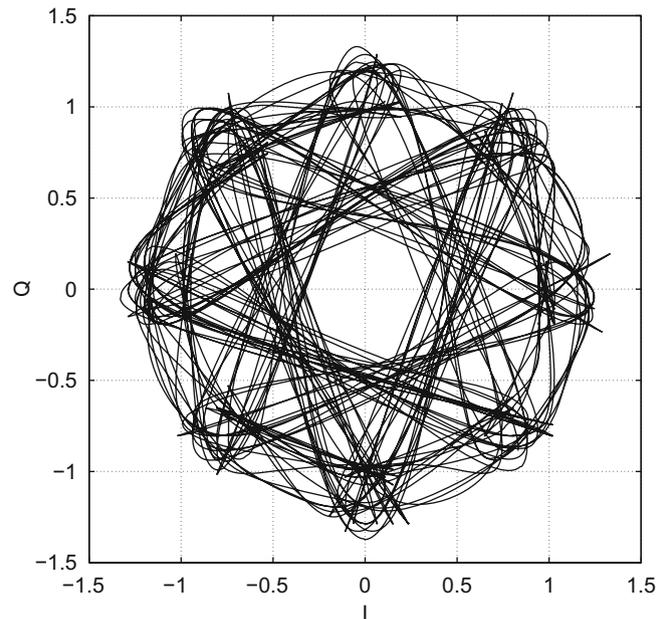


Fig. 4.13 Phaser diagram for $\pi/4$ -DQPSK with root-raised cosine amplitude pulse shaping; $\beta = 0.5$

4.5 Orthogonal Modulation and Variants

Orthogonal modulation schemes transmit information by using a set of waveforms, that may overlap in frequency but are orthogonal in time. Many different types of orthogonal waveforms are possible and here a few methods are considered that are commonly used in wireless systems.

4.5.1 Orthogonal FSK Modulation

Orthogonal M -ary frequency shift keying (MFSK) uses a set of M waveforms that all have different frequencies. The MFSK complex envelope is

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n), \quad (4.71)$$

where

$$b(t, \mathbf{x}_n) = \exp \left\{ j \frac{x_n \pi \Delta_f}{2} t \right\} u_T(t) \quad (4.72)$$

and $x_n \in \{\pm 1, \pm 3, \dots, \pm M - 1\}$. The set of MFSK waveforms that are transmitted at each baud epoch have the complex envelopes

$$\tilde{s}_m(t) = A \exp \left\{ j \frac{x_m \pi \Delta_f}{2} t \right\} u_T(t), \quad m = 1, \dots, M. \quad (4.73)$$

By choosing the frequency separation $\Delta_f = 1/2T$, all the $\tilde{s}_m(t), m = 1, \dots, M$ are mutually orthogonal (see Problem 4.7). Since the $\tilde{s}_m(t)$ are mutually orthogonal, the MFSK signal vectors have dimension $N = M$. The appropriate set of basis functions is

$$\varphi_i(t) = \sqrt{\frac{1}{T}} \exp \left\{ j \frac{x_m \pi \Delta_f}{2} t \right\} u_T(t), \quad i = 1, \dots, M = N. \quad (4.74)$$

The MFSK complex signal vectors are

$$\tilde{\mathbf{s}}_m = \sqrt{2E_b} \mathbf{e}_m, \quad m = 1, \dots, M, \quad (4.75)$$

where $\mathbf{e}_m = (e_1, e_2, \dots, e_M)$, $e_j = \delta_{jm}$, is a length- M unit basis vector with a “1” in the m th coordinate.

4.5.2 Binary Orthogonal Codes

Another set of mutually orthogonal waveforms can be obtained from the rows of a Hadamard matrix. A Hadamard matrix, \mathbf{H}_M , 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, the 8×8 Hadamard matrix obtained from the above recursive procedure is

$$\mathbf{H}_8 = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \end{bmatrix}. \quad (4.76)$$

The rows of the Hadamard matrix are mutually orthogonal. A set of M equal energy orthogonal waveforms can be constructed according to

$$\tilde{s}_m(t) = A \sum_{k=1}^M h_{m_k} h_c(t - kT_c), \quad m = 1, \dots, M, \quad (4.77)$$

where h_{mk} is the k th co-ordinate in the m th row of the Hadamard matrix, $T = MT_c$ is the symbol duration, and $h_c(t)$ is a root Nyquist shaping pulse with a Nyquist frequency of $1/(2T_c)$. Sometimes the above waveforms are called Walsh codes, and find application in the forward link of some cellular code division multiple access (CDMA) systems, such as IS-95A/B and cdma2000.

The bandpass waveforms, $s_m(t)$, all have energy

$$E_h = \frac{MA^2}{2} \int_{-\infty}^{\infty} h_c^2(t) dt. \quad (4.78)$$

To construct signal vectors, the appropriate choice of basis function is

$$\varphi_i(t) = \frac{A}{\sqrt{2E_h}} \sum_{k=1}^M h_{ik} h_c(t - kT_c), \quad i = 1, \dots, M \quad (4.79)$$

and once again the signal vectors are

$$\tilde{\mathbf{s}}_m = \sqrt{2E_h} \mathbf{e}_m, \quad m = 1, \dots, M. \quad (4.80)$$

4.5.3 Biorthogonal Signals

A set of M biorthogonal waveforms can be constructed from a set of $M/2$ orthogonal waveforms. The M -ary biorthogonal waveforms have complex signal vectors

$$\tilde{\mathbf{s}}_i = \begin{cases} \sqrt{2E_h} \mathbf{e}_i, & i = 1, \dots, M/2 \\ -\tilde{\mathbf{s}}_{i-M/2}, & i = M/2 + 1, \dots, M \end{cases}, \quad (4.81)$$

where the unit basis vectors \mathbf{e}_i have length $N = M/2$. By using an appropriate set of basis functions, for example in (4.74) or (4.79), the complex envelopes of the biorthogonal waveforms can be synthesized.

4.5.4 Orthogonal Multipulse Modulation

With binary orthogonal codes only $k = \log_2 M$ bits are transmitted at each baud epoch. A much 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

$$h_i(t) = \sum_{k=0}^{N-1} h_{ik} h_c(t - kT_c), \quad i = 1, \dots, N \quad (4.82)$$

each having duration $T = NT_c$. With orthogonal multipulse modulation, a block of $M = N$ data symbols are transmitted in parallel every T seconds by using the N orthogonal amplitude shaping pulses in (4.82). The transmitted complex envelope is

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n), \quad (4.83)$$

where

$$b(t, \mathbf{x}_n) = \sum_{k=0}^{N-1} x_{nk} h_k(t) \quad (4.84)$$

$T = NT_c$, and $\mathbf{x}_n = (x_{n1}, x_{n2}, \dots, x_{nN})$ is the block of $M = N$ data symbols transmitted at epoch n .

4.6 Orthogonal Frequency Division Multiplexing

All of the modulation techniques discussed so far are *single-carrier* modulation techniques that employ a single RF carrier. Another possibility is to use *multi-carrier* modulation techniques where information is transmitted in parallel by using multiple subcarriers. Orthogonal frequency division multiplexing (OFDM) is perhaps the most popular multi-carrier modulation technique. OFDM was first introduced in the 1960s [55], but it was perhaps the efficient DFT implementation of OFDM developed by Weinstein and Ebert [348] that has led to its popularity and widespread use. OFDM was first suggested for use cellular land mobile radio by Cimini [66] and later implemented in the Motorola IDEN (Integrated Digital Enhanced Network) standard [44]. OFDM is now used in a large number of standards for broadcasting (DVB-T, DVB-H, MediaFLO, and others), wireless LAN or WiFi (IEEE 802.11a/g/n/p), wireless MAN or WiMAX (IEEE 802.16), mobile broadband wireless access (MBWA) (IEEE 802.16e mobile WiMAX), wireless regional area networks (WRAN) (IEEE 802.22), and the cellular land mobile radio [3GPP Long Term Evolution (LTE) air interface named High Speed OFDM Packet Access (HSOPA)], among others.

OFDM is a block modulation scheme where data symbols are transmitted in parallel on orthogonal subcarriers. A block of N data symbols, with symbol period T_s , is converted into a block of N parallel data symbols, each of duration $T = NT_s$. The N parallel data symbols modulate N subcarriers that are spaced in frequency $1/T$ Hz apart. The OFDM complex envelope is given by

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n), \quad (4.85)$$

where

$$b(t, \mathbf{x}_n) = u_T(t) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi kt}{T}} \quad (4.86)$$

n is the block index, k is the subcarrier index, N is the number of subcarriers, and $\mathbf{x}_n = \{x_{n,0}, x_{n,1}, \dots, x_{n,N-1}\}$ is the data symbol block at epoch n .

The data symbols $x_{n,k}$ are usually chosen from a QAM or PSK signal constellation, although any 2-D signal constellation can be used. The $1/T$ Hz frequency separation of the subcarriers ensures that the corresponding waveforms transmitted on the N subcarriers are mutually orthogonal regardless of the random phases that are imparted by the random data symbols $x_{n,k}$ (see Problem 4.7).

A cyclic extension (or guard interval) is usually added to the OFDM waveform in (4.85) and (4.86) to combat intersymbol interference (ISI) as explained in Sect. 10.1 of Chap. 10. The cyclic extension can be in the form of either a cyclic prefix or a cyclic suffix. With a cyclic suffix, the OFDM complex envelope becomes

$$\tilde{s}_g(t) = \begin{cases} \tilde{s}(t) & , 0 \leq t < T \\ \tilde{s}(t - T) & , T \leq t < (1 + \alpha_g)T \end{cases} \quad (4.87)$$

where $\alpha_g T$ is the length of the guard interval and $\tilde{s}(t)$ is defined in (4.85) and (4.86). The OFDM waveform with cyclic suffix can be rewritten in the standard form

$$\tilde{s}_g(t) = A \sum_n b(t - nT_g, \mathbf{x}_n), \quad (4.88)$$

where

$$b(t, \mathbf{x}_n) = u_T(t) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi kt}{T}} + u_{\alpha_g T}(t - T) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi k(t-T)}{T}}, \quad (4.89)$$

and $T_g = (1 + \alpha_g)T$ is the OFDM symbol period with addition of the guard interval. Likewise, with a cyclic prefix, the OFDM complex envelope becomes

$$\tilde{s}_g(t) = \begin{cases} \tilde{s}(t+T) & , -\alpha_g T \leq t < 0 \\ \tilde{s}(t) & , 0 \leq t < T \end{cases}, \quad (4.90)$$

and

$$b(t, \mathbf{x}_n) = u_{\alpha_g T}(t + \alpha_g T) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi k(t+T)}{T}} + u_T(t) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi kt}{T}}. \quad (4.91)$$

4.6.1 DFT-Based OFDM Baseband Modulator

A key advantage of using OFDM is that the baseband modulator can be implemented by using an inverse discrete-time Fourier transform (IDFT). In practice, the computationally efficient inverse fast Fourier transform (IFFT) algorithm is used to implement the IDFT. Consider the OFDM complex envelope defined by (4.85) and (4.86). During the interval $nT \leq t < (n+1)T$, the complex envelope has the form

$$\begin{aligned} \tilde{s}(t) &= Au_T(t - nT) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi k(t-nT)}{T}} \\ &= Au_T(t - nT) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi kt}{NT_s}}, \quad nT \leq t < (n+1)T. \end{aligned} \quad (4.92)$$

Now suppose that the complex envelope in (4.92) is sampled at synchronized T_s second intervals beginning at time nT to yield the sample sequence

$$X_{n,m} = \tilde{s}(mT_s) = A \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi km}{N}}, \quad m = 0, 1, \dots, N-1. \quad (4.93)$$

Observe that the vector $\mathbf{X}_n = \{X_{n,m}\}_{m=0}^{N-1}$ is the IDFT of the vector $A\mathbf{x}_n = A\{x_{n,k}\}_{k=0}^{N-1}$. Contrary to conventional notation, with OFDM it is customary that the lowercase vector $A\mathbf{x}_n$ represents the frequency domain coefficients, as the vector \mathbf{x}_n is the vector of data symbols, while the uppercase vector \mathbf{X}_n represents the IDFT time domain coefficients.

As mentioned earlier, a cyclic extension (or guard interval) is usually added to the OFDM waveform as described in (4.88) and (4.89) to combat ISI. When a cyclic suffix is used, the corresponding sample sequence is

$$X_{n,m}^g = X_{n,(m)_N} \quad (4.94)$$

$$= A \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi km}{N}}, \quad m = 0, 1, \dots, N+G-1, \quad (4.95)$$

where G is the length of the guard interval in samples, and $(m)_N$ is the residue of m modulo N . This gives the vector $\mathbf{X}_n^g = \{X_{n,m}^g\}_{m=0}^{N+G-1}$, where the values in the first and last G coordinates of the vector \mathbf{X}_n^g are the same. Likewise, when a cyclic prefix is used, the corresponding sample sequence is

$$X_{n,m}^g = X_{n,(m)_N} \quad (4.96)$$

$$= A \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi km}{N}}, \quad m = -G, \dots, -1, 0, 1, \dots, N-1. \quad (4.97)$$

This yields the vector $\mathbf{X}_n^g = \{X_{n,m}^g\}_{m=-G}^{N-1}$, where again the first and last G coordinates of the vector \mathbf{X}_n^g are the same. The sample interval after insertion of the guard interval, T_s^g , is compressed in time such that $(N+G)T_s^g = NT_s$.

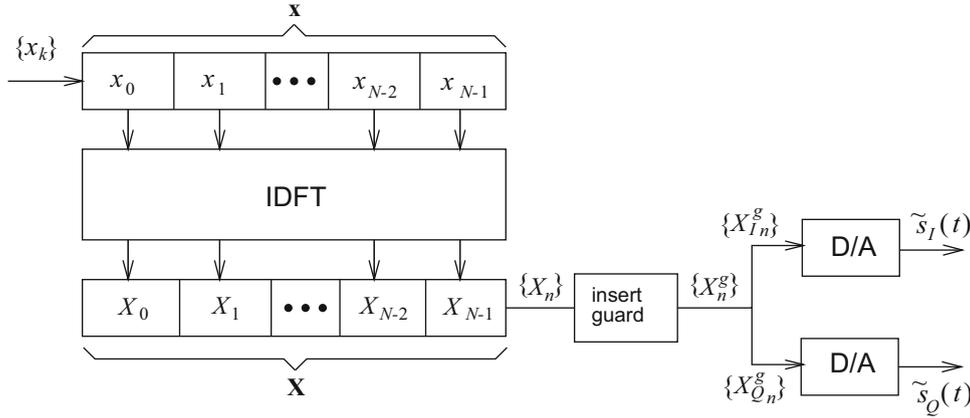


Fig. 4.14 Block diagram of IDFT-based baseband OFDM modulator with guard interval insertion and digital-to-analog conversion

The OFDM complex envelope can be generated by splitting the complex-valued output vector \mathbf{X}_m into its real and imaginary parts, $\text{Re}(\mathbf{X}_m)$ and $\text{Im}(\mathbf{X}_m)$, respectively. The sequences $\{\text{Re}(X_{n,m})\}$ and $\{\text{Im}(X_{n,m})\}$ are then input to a pair of balanced digital-to-analog converters (DACs) to generate the real and imaginary components $\tilde{s}_I(t)$ and $\tilde{s}_Q(t)$, respectively, of the complex envelope $\tilde{s}(t)$, during the time interval $nT \leq t < (n+1)T$. As shown in Fig. 4.14, the OFDM baseband modulator consists of an IDFT operation, followed by guard interval insertion and digital-to-analog conversion.

It is instructive at this stage to note that the waveform generated by using the IDFT OFDM baseband modulator is *not exactly* the same as the waveform generated from the analog waveform definition of OFDM. Consider for example, the OFDM waveform without a cyclic guard in (4.85) and (4.86). The analog waveform definition uses the rectangular amplitude shaping pulse $u_T(t)$ that is strictly time-limited to T seconds. As shown in Sect. 4.9.6, the corresponding power spectrum has infinite bandwidth. Consequently, sampling the complex envelope with any finite sampling rate will lead to aliasing and imperfect reconstruction.

With the IDFT OFDM baseband modulator, the IDFT outputs are applied to a pair of balanced DACs as explained earlier. However, the ideal DAC is an ideal low pass filter with cutoff frequency $1/(2T_s)$, with a corresponding non-causal impulse response $h(t) = \text{sinc}(t/T_s)$. Since the ideal DAC is non-realizable, a causal, finite-length reconstruction filter must be used instead. However, such a filter will necessarily generate a waveform that is not strictly bandlimited. In conclusion, the side lobe structure of the analog waveform definition of OFDM is inherent in the waveform due to rectangular amplitude pulse shaping, whereas the side lobe structure with the IDFT implementation is introduced by the non-ideal (practical) DAC.

Finally, non-rectangular amplitude pulse shaping can be used with OFDM and may yield a more compact power spectrum while still maintaining sub-channel orthogonality. However, such pulse shaping will require an extension of the OFDM symbol beyond T , or with guard interval T_g , seconds in the time domain. This will be discussed in more detail in Chap. 10.

4.6.2 Adaptive Bit Loading and Discrete Multitone Modulation

A wireless OFDM system generally operates over a frequency-selective fading channel with transfer function $T(t, f)$, such that the amplitude response $|T(t, f)|$ varies across the channel bandwidth W . The power spectral density of the additive noise impairment $S_{nn}(f)$ may vary with frequency as well due to the presence of interference. Consider a quasi-static fading channel, such that the channel remains constant over an OFDM block of duration T seconds. For convenience, the time variable t is suppressed with the understanding that $T(t, f) \equiv T(f)$ over an OFDM block, but the channel may change from block to block. Furthermore, knowledge of the channel is assumed to be available at the transmitter. Shannon [300] proved that the capacity of a frequency-selective channel with additive Gaussian noise is achieved when the transmitted power $\Omega_i(f)$ is adjusted across the bandwidth W_s according to

$$\Omega_i(f) = \begin{cases} K - S_{nn}(f)/|T(f)|^2, & f \in W_s \\ 0, & f \notin W_s \end{cases}, \quad (4.98)$$

where K is a constant chosen to satisfy the constraint

$$\int_{W_s} \Omega_t(f) df \leq \Omega_{av}, \quad (4.99)$$

and Ω_{av} is the average available power to the transmitter. One method to achieve capacity is to divide the bandwidth W_s into N sub-bands of width W_s/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.

It is clear from (4.86) that the data symbols $x_{n,k}$ for fixed n modulate the n th subcarrier. From (4.98), 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, a higher transmit power admits the use of a larger size signal constellation in sub-bands where $|T(f)|^2/S_{nn}(f)$ is large, and vice versa. The technique whereby different sized signal constellations are used on the different OFDM subcarriers is sometimes called adaptive bit loading or discrete multitone modulation (DMT).

4.6.3 Multiresolution Modulation

In broadcasting applications, it is sometimes desirable to transmit video or audio information in frames that will simultaneously provide different resolutions, depending on the received signal-to-noise ratio. Low resolution information is typically of high priority (HP) and must be received with high reliability. High resolution information, on the other hand, is typically of low priority (LP) and may be received with a lower reliability. The solution is multi-resolution modulation (MRM), a class of modulation techniques that transmit multiple resolutions in a simultaneous or concurrent fashion, that differ in their bit rates and/or error probabilities. MRM can be implemented in OFDM schemes by using multiplexed, interleaved, embedded signal constellations, and others.

Multiplexed MRM divides the OFDM band into subsets of contiguous subcarriers, for example the upper half subcarriers may be used to transmit HP data symbols and the lower half subcarriers used to transmit an equal number of LP data symbols. The HP low resolution information can be transmitted by using a smaller signal constellation and/or higher transmit power for further robustness and reliability. Likewise, the LP high resolution information can be transmitted using a larger signal constellation and/or lower transmit power. Broadcast service contours can be established for either high definition (both the HP and LP data streams are decodable) or standard definition (only the HP data stream is decodable) reception.

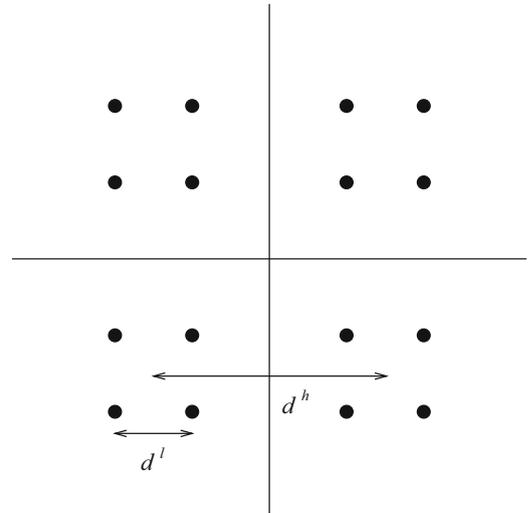
Interleaved MRM interleaves the different resolutions onto the subcarriers in a cyclic fashion. If there are K different resolutions, then subcarriers $\ell, \ell + K, \ell + 2K, \dots$, are assigned to the ℓ th resolution. Each resolution is then transmitted by using a different sized signal constellation and/or transmit power level.

Embedded MRM is more subtle and relies upon the use of an asymmetric signal constellation and finds application in some broadcast video systems. Figure 4.15 shows an example of a 16-QAM embedded MRM signal constellation, that can be used to simultaneously transmit two different resolutions. In Fig. 4.15, two HP low resolution bits are used to select the quadrant of the transmitted signal point, while two low LP high resolution bits are used to select the signal point within the selected quadrant. The relative error probability or reliability between the two priorities is controlled by the parameter $\lambda = d^l/d^h$, $\lambda \leq 0.5$, where d^l is the distance between LP symbols and d^h is the distance between the centroids of the HP symbols. The upper limit on λ is due to the fact that the MRM constellation becomes a symmetric 16-QAM constellation when $\lambda = 0.5$. As λ becomes smaller than 0.5, more power is allocated to the HP low resolution bits than the LP high resolution bits. For broadcasting applications, this can be used to provide high definition reception over some adjustable fraction of the service area where standard definition service can be received. At $\lambda = 0.5$, both resolution classes are treated equally and the coverage areas for standard and high definition service are the same.

4.7 Continuous Phase Modulation

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 by Anderson et al. [17]. 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. CPM waveforms find application in satellite communication systems, and cellular telephone systems notably GSM.

Fig. 4.15 16-QAM embedded MRM signal constellation with two resolutions



The complex envelope of a CPM waveform has the general form

$$\tilde{s}(t) = A e^{j(\phi(t) + \theta_o)}, \tag{4.100}$$

where A is the amplitude, θ_o is the initial carrier phase at $t = 0$, and

$$\phi(t) = 2\pi h \int_0^t \sum_{k=0}^{\infty} x_k h_f(\tau - kT) d\tau \tag{4.101}$$

is the excess phase, where h is the modulation index, $\{x_k\}$ is the data symbol sequence, $h_f(t)$ is the frequency shaping pulse, and T is the baud period. The CPM waveform can be written in the standard form

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \tag{4.102}$$

where

$$b(t, \mathbf{x}_n) = e^{j2\pi h \int_0^t \sum_{k=0}^{\infty} x_k h_f(\tau - kT) d\tau} u_T(t) \tag{4.103}$$

where $\mathbf{x}_n = (x_n, x_{n-1}, \dots, x_0)$, and an initial phase $\theta_o = 0$ at $t = 0$ is assumed. CPM waveforms have the following properties:

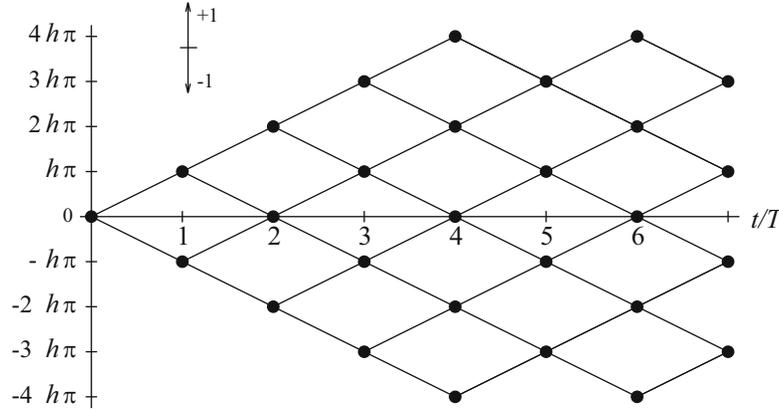
- The data symbols are chosen from the alphabet $\{\pm 1, \pm 3, \dots, \pm(M - 1)\}$, where M is the modulation alphabet size.
- h is the modulation index and is directly proportional to the peak and/or average frequency deviation from the carrier. The instantaneous frequency deviation from the carrier is

$$f_{\text{dev}}(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt} = h \sum_{k=0}^{\infty} x_k h_f(t - kT). \tag{4.104}$$

- $h_f(t)$ is the frequency shaping function, that is zero for $t < 0$ and $t > LT$, and normalized to integrate to 1/2. 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 usually obtained by using frequency shaping functions having continuous higher-order derivatives, such as the raised cosine pulse in Table 4.1. The excess phase is continuous provided that the frequency shaping function $h_f(t)$ does not contain impulses, which is true for all CPM waveforms. When describing CPM waveforms, it is useful to define the phase shaping function,

Table 4.1 CPM frequency shaping functions

Pulse type	$h_f(t)$
L -rectangular (LREC)	$\frac{1}{2LT} u_{LT}(t)$
L -raised cosine (LRC)	$\frac{1}{2LT} [1 - \cos(\frac{2\pi t}{LT})] u_{LT}(t)$
L -half sinusoid (LHS)	$\frac{\pi}{4LT} \sin(\pi t/LT) u_{LT}(t)$
L -triangular (LTR)	$\frac{1}{LT} (1 - \frac{ t-LT/2 }{LT/2}) u_{LT}(t)$

**Fig. 4.16** Phase tree of binary CPFSK with an arbitrary modulation index. CPFSK is characterized by linear excess phase trajectories

$$\beta(t) = \begin{cases} 0 & , t < 0 \\ \int_0^t h_f(\tau) d\tau & , 0 \leq t \leq LT \\ 1/2 & , t \geq LT \end{cases} \quad (4.105)$$

which is the integral of the frequency shaping pulse. An infinite variety of CPM waveforms can be generated by choosing different frequency shaping pulses, modulation indices, and modulation alphabet sizes.

4.7.1 Full Response CPM

For a full response CPM waveform with $L = 1$, the shaping function in (4.103) has the form

$$b(t, \mathbf{x}_n) = e^{j(\pi h \sum_{k=0}^{n-1} x_k + 2\pi h x_n \beta(t))} u_T(t). \quad (4.106)$$

The first term in the exponent of (4.106) represents the accumulated excess phase up to time nT , while the second term represents the excess phase increment during the time interval $nT \leq t < (n+1)T$.

Continuous phase frequency shift keying (CPFSK) is a special type of full response CPM characterized by the rectangular frequency shaping function LREC with $L = 1$. For CPFSK

$$\beta(t) = \begin{cases} 0 & , t < 0 \\ t/2T & , 0 \leq t < T \\ 1/2 & , t \geq T \end{cases} \quad (4.107)$$

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. 4.16 for binary CPFSK. Since the CPFSK frequency shaping function is rectangular, the excess phase trajectories are linear as suggested by (4.107). In each baud interval, T , the excess phase increases by πh if the data symbol is $+1$ and decreases by πh if the data symbol is -1 .

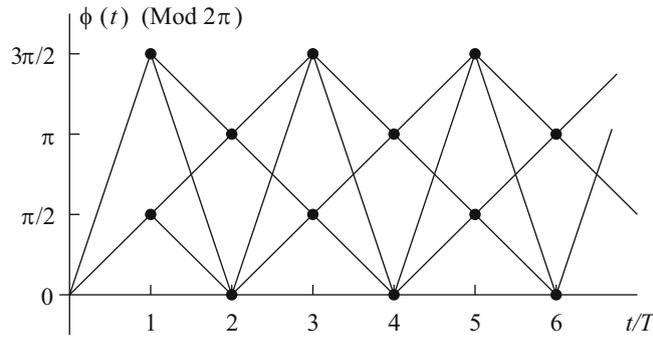


Fig. 4.17 Phase-trellis for MSK

4.7.1.1 Minimum Shift Keying

Minimum shift keying (MSK) is binary CPFSK with modulation index $h = 1/2$. In this case,

$$b(t, \mathbf{x}_n) = e^{j\left(\frac{\pi}{2} \sum_{k=0}^{n-1} x_k + \frac{\pi}{2} x_n \frac{t}{T}\right)} u_T(t). \quad (4.108)$$

The MSK waveform can be described in terms of the phase tree as shown in Fig. 4.16 with $h = 1/2$. At the end of each symbol interval the excess phase $\phi(t)$ 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 belong to the finite set $\{0, \pi/2, \pi, 3\pi/2\}$. The MSK phase tree reduced modulo 2π yields the MSK phase trellis shown in Fig. 4.17.

An interesting property of MSK can be observed from the MSK bandpass waveform. The bandpass waveform on the interval $nT \leq t < (n+1)T$ can be obtained from (4.108) as

$$\begin{aligned} s(t) &= A \cos \left(2\pi f_c t + \frac{\pi}{2} \sum_{k=0}^{n-1} x_k + \frac{\pi}{2} x_n \frac{t - nT}{T} \right) \\ &= A \cos \left(2\pi \left(f_c + \frac{x_n}{4T} \right) t + \frac{\pi}{2} \sum_{k=0}^{n-1} x_k - \frac{\pi n}{2} x_n \right). \end{aligned} \quad (4.109)$$

Observe that the MSK bandpass waveform has one of two possible frequencies in each baud interval, given by

$$f_L = f_c - \frac{1}{4T} \quad \text{and} \quad f_U = f_c + \frac{1}{4T} \quad (4.110)$$

depending on the data symbol x_n . The difference between these two frequencies is $f_U - f_L = 1/(2T)$. This is the minimum frequency separation to ensure orthogonality between two co-phased sinusoids of duration T (Problem 4.7) and, hence, the name *minimum* shift keying.

Another interesting representation for MSK waveforms can be obtained by using Laurent's decomposition [193] (detailed in Sect. 4.8.3) to express the MSK complex envelope in the quadrature form

$$\tilde{s}(t) = A \sum_n b(t - 2nT, \mathbf{x}_n), \quad (4.111)$$

where

$$b(t, \mathbf{x}_n) = \hat{x}_{2n+1} h_a(t - T) + j \hat{x}_{2n} h_a(t) \quad (4.112)$$

and where $\mathbf{x}_n = (\hat{x}_{2n+1}, \hat{x}_{2n})$,

$$\hat{x}_{2n} = \hat{x}_{2n-1} x_{2n} \quad (4.113)$$

$$\hat{x}_{2n+1} = -\hat{x}_{2n}x_{2n+1} \quad (4.114)$$

$$\hat{x}_{-1} = 1 \quad (4.115)$$

and

$$h_a(t) = \sin\left(\frac{\pi t}{2T}\right) u_{2T}(t). \quad (4.116)$$

The sequences, $\{\hat{x}_{2n}\}$ and $\{\hat{x}_{2n+1}\}$, are independent binary symbol sequences taking on elements from the set $\{-1, +1\}$. The symbols \hat{x}_{2n} and \hat{x}_{2n+1} are transmitted on the quadrature branches with a half-sinusoid (HS) amplitude shaping pulse of duration $2T$ seconds and an offset of T seconds. Hence, MSK is equivalent to offset quadrature amplitude shift keying (OQASK) with HS amplitude pulse shaping. This linear representation of MSK is useful in practice for simplified linear detection of MSK waveforms as discussed in Sect. 5.10.1 of Chap. 5.

4.8 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 typically 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

$$\begin{aligned} h_f(t) &= \sum_{k=0}^{L-1} h_f(t) u_T(t - kT) \\ &= \sum_{k=0}^{L-1} h_{f,k}(t - kT), \end{aligned} \quad (4.117)$$

where

$$h_{f,k}(t) = h_f(t + kT) u_T(t). \quad (4.118)$$

Likewise, the partial response phase shaping function can be written as

$$\beta(t) = \sum_{k=0}^{L-1} \beta_k(t - kT), \quad (4.119)$$

where

$$\beta_k(t) = \beta(t + kT) u_T(t). \quad (4.120)$$

Note that

$$\beta_k(t) = \begin{cases} 0 & , t < 0 \\ \int_0^t h_{f,k}(\tau) d\tau & , 0 \leq t < LT \\ \beta_k(T) & , t \geq T \end{cases} \quad (4.121)$$

and

$$\sum_{k=0}^{L-1} \beta_k(T) = \frac{1}{2}. \quad (4.122)$$

An equivalent frequency shaping function of duration T can be derived by noting that the CPM baseband modulating signal has the form

$$\begin{aligned}
 x(t) &= \sum_n x_n h_f(t - nT) \\
 &= \sum_n \sum_{k=0}^{L-1} x_n h_{f,k}(t - (n+k)T) \\
 &= \sum_m \sum_{k=0}^{L-1} x_{m-k} h_{f,k}(t - mT).
 \end{aligned} \tag{4.123}$$

It follows that

$$x(t) = \sum_m h_f(t - mT, \mathbf{x}_m), \tag{4.124}$$

where

$$h_f(t, \mathbf{x}_m) = \sum_{k=0}^{L-1} x_{m-k} h_{f,k}(t) \tag{4.125}$$

and

$$x(t) = \sum_m \beta(t - mT, \mathbf{x}_m), \tag{4.126}$$

where

$$\beta(t, \mathbf{x}_m) = \sum_{k=0}^{L-1} x_{m-k} \beta_k(t) \tag{4.127}$$

and

$$\mathbf{x}_m = (x_m, x_{m-1}, \dots, x_{m-L+1}). \tag{4.128}$$

Finally, the complex envelope of partial response CPM signal can be written in the standard form

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \tag{4.129}$$

where

$$b(t, \mathbf{x}_n) = e^{j2\pi h(\sum_{i=0}^{n-1} \beta(T, \mathbf{x}_i) + \beta(t, \mathbf{x}_n))} u_T(t) \tag{4.130}$$

and an initial excess phase of zero is assumed.

Example 4.2. Consider a partial response CPM waveform with an LREC frequency shaping function. In this case

$$h_f(t) = \frac{1}{2LT} u_{LT}(t).$$

(continued)

Example 4.2 (continued)

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. Consider a partial response CPM waveform with an LRC frequency shaping function. In this case

$$h_f(t) = \frac{1}{2LT} \left(1 - \cos \left(\frac{2\pi t}{LT} \right) \right) u_{LT}(t).$$

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

$$h_{f,k}(t) = \left(1 - \cos \left(\frac{2\pi(t + kT)}{LT} \right) \right) u_T(t).$$

4.8.1 Phase States

The excess phase of a partial response CPM waveform on the interval $nT \leq t < (n+1)T$ is

$$\phi(t) = 2\pi h \int_0^t \sum_{k=0}^n x_k h_f(\tau - kT) d\tau \quad (4.131)$$

$$= \pi h \sum_{k=0}^{n-L} x_k + 2\pi h \sum_{k=n-L+1}^n x_k \beta(t - kT) \quad (4.132)$$

$$= \theta_n + 2\pi h \sum_{k=n-L+1}^n x_k \beta(t - kT) \quad (4.133)$$

where

$$\theta_n = \pi h \sum_{k=0}^{n-L} x_k \quad \text{modulo } 2\pi \quad (4.134)$$

is the accumulated phase. During the interval $nT \leq t < (n+1)T$, the excess phase depends on the input data symbol x_n , the vector of $L-1$ previous data symbols, $\{x_{n-1}, x_{n-2}, \dots, x_{n-L+1}\}$, and the accumulated phase θ_n . The state of the CPM signal at time $t = nT$, is defined by the L -tuple

$$S_n = (\theta_n, x_{n-1}, x_{n-2}, \dots, x_{n-L+1}). \quad (4.135)$$

Since the vector $(x_{n-1}, x_{n-2}, \dots, x_{n-L+1})$ can take on M^{L-1} values, the number of states equals M^{L-1} times the number of values that θ_n can assume.

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 is a useful property for the implementation CPM receivers. If m is even, then

$$\theta_n \in \left\{ 0, \frac{\pi m}{p}, \frac{2\pi m}{p}, \dots, \frac{(p-1)\pi m}{p} \right\} \quad (4.136)$$

while if m is odd

$$\theta_n \in \left\{ 0, \frac{\pi m}{p}, \frac{2\pi m}{p}, \dots, \frac{(2p-1)\pi m}{p} \right\}. \quad (4.137)$$

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

$$|S_n| = \begin{cases} pM^{L-1}, & m \text{ even} \\ 2pM^{L-1}, & m \text{ odd} \end{cases}. \quad (4.138)$$

For example, if $h = 1/4$, $M = 4$, and $L = 2$, then

$$\theta_n \in \left\{ 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4} \right\} \quad (4.139)$$

and the number of CPM states is 32.

CPM waveforms cannot be described in terms of a signal-space diagram, like QAM and PSK waveforms. However, the CPM waveform can be described in terms of the trajectories from one phase state to another. Figures 4.18 and 4.19 show the phase state diagrams for MSK and binary CPM with $h = 1/4$, respectively. Since binary modulation is used, trajectories are only allowed between adjacent phase states as shown by the dotted lines in the figures. Since the CPM waveforms have constant envelope, the complex phaser trajectories will follow along the circle in Figs. 4.18 and 4.19.

4.8.2 Gaussian Minimum Shift Keying

Due to their non-linearity, CPM waveforms have a relatively complicated power spectrum as detailed in Sect. 4.9.7. However, the bandwidth of a CPM waveform can be approximated using Carson's rule developed for analog frequency modulation:

$$BW = (W + f_{\text{peak}}), \quad (4.140)$$

where W is the bandwidth of the frequency shaping pulse $h_f(t)$ and f_{peak} is the peak frequency deviation from the carrier. MSK waveforms have relatively poor spectral characteristics due to the large bandwidth W of the rectangular frequency pulse shaping $h_f(t) = \frac{1}{2T} \text{rect}(t)$. A more compact power spectrum can be achieved by low-pass filtering the MSK modulating signal

Fig. 4.18 Phase state diagram for MSK

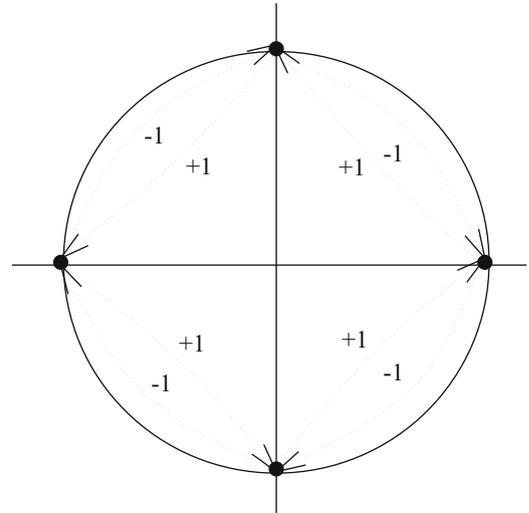


Fig. 4.19 Phase state diagram for binary CPM with $h = 1/4$

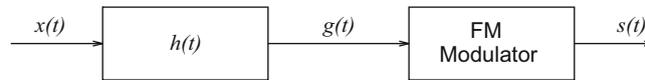
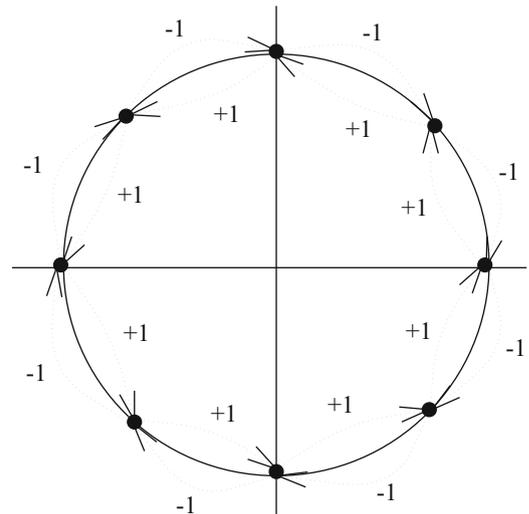


Fig. 4.20 Pre-modulation filtered MSK. The MSK modulating signal is low-pass filtered to remove the high frequency components prior to frequency modulation

$$x(t) = \sum_{n=-\infty}^{\infty} x_n h_f(t - nT) = \frac{1}{2T} \sum_{n=-\infty}^{\infty} x_n u_T(t - nT) \tag{4.141}$$

prior to frequency modulation as shown in Fig. 4.20. Such filtering suppresses the higher frequency components in $x(t)$ thus yielding a more compact power spectrum. Gaussian minimum shift keying (GMSK) is a special type of partial response CPM that uses a low-pass premodulation filter having the transfer function [234]

$$H(f) = \exp \left\{ - \left(\frac{f}{B} \right)^2 \frac{\ln 2}{2} \right\}, \tag{4.142}$$

where B is the 3 dB bandwidth of the filter. It is apparent that $H(f)$ is shaped like a Gaussian probability density function with mean $f = 0$ and, hence, the name ‘‘Gaussian’’ MSK. Convolution of the rectangular pulse

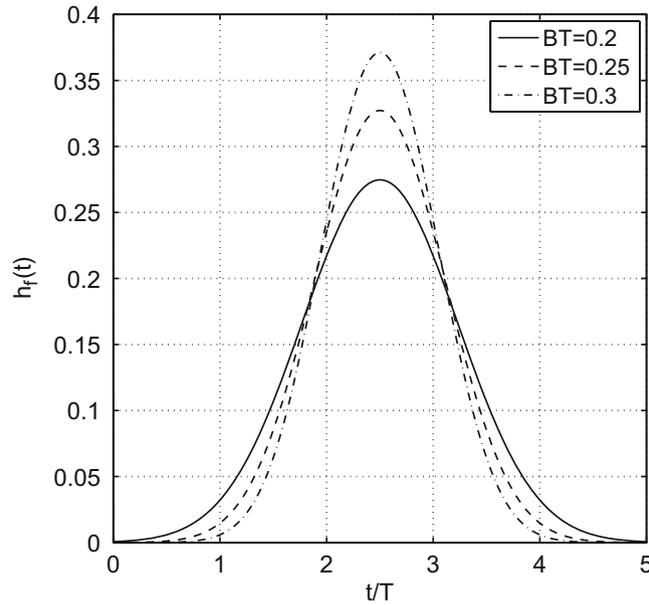


Fig. 4.21 GMSK frequency shaping pulse for various normalized premodulation filter bandwidths BT

$$\frac{1}{2T}\text{rect}(t/T) = \frac{1}{2T}u_T(t + T/2)$$

with the corresponding filter impulse response $h(t)$ yields the frequency shaping pulse

$$\begin{aligned} h_f(t) &= \frac{1}{2T} \sqrt{\frac{2\pi}{\ln 2}} (BT) \int_{t/T-1/2}^{t/T+1/2} \exp\left\{-\frac{2\pi^2(BT)^2 x^2}{\ln 2}\right\} dx \\ &= \frac{1}{2T} \left(Q\left(\frac{t/T - 1/2}{\sigma}\right) - Q\left(\frac{t/T + 1/2}{\sigma}\right) \right), \end{aligned} \quad (4.143)$$

where

$$Q(\alpha) = \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \quad (4.144)$$

$$\sigma^2 = \frac{\ln 2}{4\pi^2(BT)^2}. \quad (4.145)$$

Figure 4.21 plots the GMSK frequency shaping pulse (truncated to $5T$ and time shifted by $2.5T$ to yield a causal pulse) for various normalized premodulation filter bandwidths BT . The GSM cellular standard uses GMSK with $BT = 0.3$.

The phase shaping function is the integral of the frequency shaping function as defined in (4.105). Using $h_f(t)$ in (4.143), and integrating by parts, yields

$$\beta(t) = \int_{-\infty}^t h_f(t) dt = \frac{1}{2} \left(G\left(\frac{t}{T} + \frac{1}{2}\right) - G\left(\frac{t}{T} - \frac{1}{2}\right) \right), \quad (4.146)$$

where

$$G(x) = x \Phi\left(\frac{x}{\sigma}\right) + \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \quad (4.147)$$

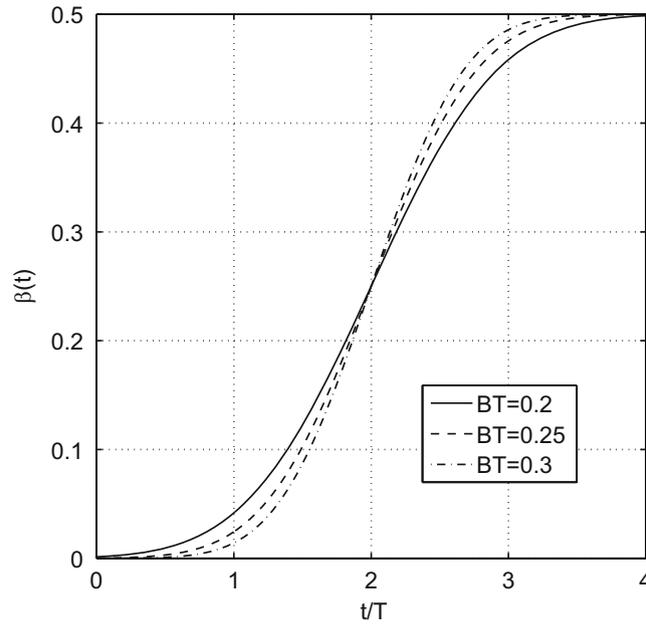


Fig. 4.22 GMSK phase shaping pulse for various normalized premodulation filter bandwidths BT

and

$$\Phi(\alpha) = \int_{-\infty}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \quad (4.148)$$

Figure 4.22 plots the GMSK phase shaping pulse (truncated to $4T$ and time shifted by $2T$ to yield a causal pulse) for $BT = 0.3$. Observe that $\beta(\infty) = 1/2$ and, therefore, the total contribution to the excess phase for each data symbol remains at $\pm\pi/2$.

The change in excess phase over the length- T baud interval from $-T/2$ to $T/2$ is

$$\phi(T/2) - \phi(-T/2) = \pi x_0 \beta_0(T) + \pi \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} x_n \beta_n(T), \quad (4.149)$$

where

$$\beta_n(T) = \int_{-T/2-nT}^{T/2-nT} h_f(\tau) d\tau. \quad (4.150)$$

The first term in (4.149) is the desired term, and the second term is the intersymbol interference (ISI) introduced by the Gaussian premodulation filter. While the premodulation filter will yield a more compact power spectrum, the induced ISI will degrade the bit error rate performance and may necessitate an equalizer in the receiver. However, the induced ISI is not severe and in many cases, including GSM cellular receivers, an equalizer is required anyway to combat the ISI due to channel delay spread.

4.8.3 Linearized GMSK (LGMSK)

Like all other CPM waveforms, GMSK is a nonlinear waveform. Similar to the linearized representation of MSK in Sect. 4.7.1.1, it is desirable to find a linearized representation for GMSK in order to simplify receiver processing. Several linear approximations have been suggested in the literature for GMSK. Here, an approximation based on Laurent's decomposition [193] is considered. Laurent showed that any binary partial response CPM signal can be represented exactly as a linear combination of 2^{L-1} partial-response pulse amplitude modulated (PAM) signals, viz.,

$$\tilde{s}(t) = \sum_{n=0}^{\infty} \sum_{p=0}^{2^{L-1}-1} e^{j\pi h \alpha_{n,p}} c_p(t - nT), \quad (4.151)$$

where

$$c_p(t) = c(t) \prod_{n=1}^{L-1} c(t + (n + L\varepsilon_{n,p})T) \quad (4.152)$$

$$\alpha_{n,p} = \sum_{m=0}^n x_m - \sum_{m=1}^{L-1} x_{n-m} \varepsilon_{m,p} \quad (4.153)$$

and $\varepsilon_{n,p} \in \{0, 1\}$ are the coefficients of the binary representation of the index p , i.e.,

$$p = \varepsilon_{0,p} + 2\varepsilon_{1,p} + \cdots + 2^{L-2}\varepsilon_{L-2,p}. \quad (4.154)$$

The basic signal pulse $c(t)$ in (4.152) is

$$c(t) = \begin{cases} \frac{\sin(2\pi h \beta(t))}{\sin \pi h} & , 0 \leq t < LT \\ \frac{\sin(\pi h - 2\pi h \beta(t-LT))}{\sin \pi h} & , LT \leq t < 2LT \\ 0 & , \text{otherwise} \end{cases} \quad (4.155)$$

where $\beta(t)$ is the CPM phase shaping function.

The above linear decomposition will yield an exact representation of the GMSK waveform. However, the fact that 2^{L-1} pulses are needed to represent the waveform means that the optimum coherent receiver will need 2^{L-1} filters that are matched to the set of 2^{L-1} pulses $\{c_p(t)\}$. Usually, the number of matched filters can be reduced to $K < 2^{L-1}$ when a good approximation to the CPM signal can be obtained with K of the set of 2^{L-1} pulses $\{c_p(t)\}$. Often the pulse $c_0(t)$ contains most of the signal energy, so the $p = 0$ term in (4.151) can provide a good approximation to the CPM signal. From Fig. 4.21, observe that the GMSK frequency shaping pulse spans approximately $L = 4$ symbol periods for practical values of BT . This means that the GMSK waveform can be constructed from the superposition of eight pulses, $c_p(t), p = 0, \dots, 7$. Numerical analysis shows that, with $BT = 0.3$, the pulse $c_0(t)$ contains 99.83% of the energy and, therefore, a linearized GMSK waveform can be derived by using only $c_0(t)$ and neglecting the other pulses. This yields the waveform

$$\tilde{s}(t) = \sum_{n=0}^{\infty} e^{j\pi h \alpha_{n,0}} c_0(t - nT), \quad (4.156)$$

where, with $L = 4$,

$$c_0(t) = \prod_{n=0}^3 c(t + nT) \quad (4.157)$$

$$\alpha_{n,0} = \sum_{m=0}^n x_m. \quad (4.158)$$

Since the GMSK phase shaping pulse is non-causal, when evaluating $c(t)$ in (4.155) the truncated and time shifted GMSK phase shaping pulse

$$\hat{\beta}(t) = \beta(t - 2T) \quad (4.159)$$

is used with $L = 4$ as shown in Fig. 4.22, where $\beta(t)$ is defined in (4.146). Figure 4.23 plots the resulting LGMSK amplitude shaping pulse $c_0(t)$ obtained from (4.157).

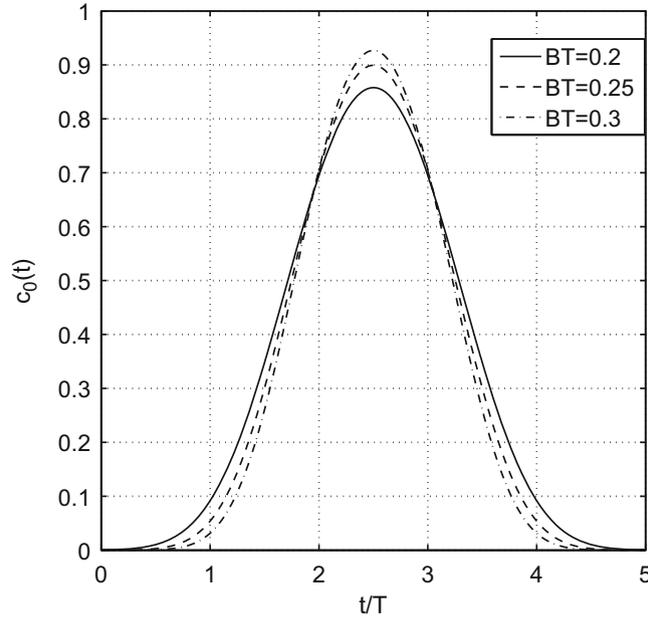


Fig. 4.23 LGMSK amplitude shaping pulse for various normalized premodulation filter bandwidths BT

For the modulation index $h = 1/2$ used in GMSK,

$$a_{n,0} = e^{j\frac{\pi}{2}a_{n,0}} \in \{\pm 1, \pm j\}, \quad (4.160)$$

and it follows that

$$\tilde{s}(t) = A \sum_n \left(\hat{x}_{2n+1} c_0(t - 2nT - T) + j\hat{x}_{2n} c_0(t - 2nT) \right), \quad (4.161)$$

where

$$\hat{x}_{2n} = \hat{x}_{2n-1} x_{2n} \quad (4.162)$$

$$\hat{x}_{2n+1} = -\hat{x}_{2n} x_{2n+1} \quad (4.163)$$

$$\hat{x}_{-1} = 1. \quad (4.164)$$

This is the same as the OQPSK representation for MSK in Sect. 4.7.1.1, except that the half-sinusoid amplitude pulse shaping function in (4.116) is replaced with the LGMSK amplitude pulse shaping function defined in (4.157). Note that the LGMSK pulse has length of approximately $4T$, while the pulses on the quadrature branches are transmitted every $2T$ seconds. Therefore, the LGMSK pulse will introduce ISI that must be corrected by an equalizer to avoid a performance degradation as mentioned earlier. However, as shown in Sect. 4.9.8, GMSK has excellent spectral properties.

4.8.4 Tamed Frequency Modulation

Tamed frequency modulation (TFM) is a special type of partial response binary CPM that was introduced by de Jager and Dekker [91]. TFM also has excellent spectral properties, similar to GMSK. To define TFM waveforms, recall that the MSK excess phase obeys the difference equation

$$\phi(nT + T) - \phi(nT) = \frac{\pi}{2} x_n. \quad (4.165)$$

The TFM excess phase trajectory is “smoothed” by imposing the constraint

$$\phi(nT + T) - \phi(nT) = \frac{\pi}{2} \left(\frac{x_{n-1}}{4} + \frac{x_n}{2} + \frac{x_{n+1}}{4} \right), \quad (4.166)$$

such that the maximum change in excess phase over any bit interval is $\pi/2$. To complete the definition of the TFM signal, an appropriate frequency shaping pulse $h_f(t)$ must be defined. The TFM excess phase can be written as

$$\phi(t) = \pi \sum_{k=0}^{\infty} x_k \beta(t - kT), \quad (4.167)$$

where

$$\beta(t) = \int_0^t h_f(t) dt \quad (4.168)$$

and where a modulation index $h = 1/2$ is assumed. The excess phase change over the time interval $[nT, (n+1)T]$ is

$$\begin{aligned} \phi((n+1)T) - \phi(nT) &= \pi \sum_{k=0}^{\infty} x_k (\beta(nT + T - kT) - \beta(nT - kT)) \\ &= \pi \sum_{\ell=n}^{\infty} x_{n-\ell} (\beta(\ell T + T) - \beta(\ell T)). \end{aligned} \quad (4.169)$$

Expanding (4.166) in more detail gives

$$\phi(nT + T) - \phi(nT) = \frac{\pi}{2} \left(\cdots + x_{n-2} \cdot 0 + \frac{x_{n-1}}{4} + \frac{x_n}{2} + \frac{x_{n+1}}{4} + x_{n+2} \cdot 0 + \cdots \right). \quad (4.170)$$

Comparing (4.169) and (4.170) gives the condition

$$\beta((\ell+1)T) - \beta(\ell T) = \begin{cases} 1/8, & |\ell| = 1 \\ 1/4, & \ell = 0 \\ 0, & \text{otherwise} \end{cases}. \quad (4.171)$$

From the definition of $\beta(t)$ in (4.168) the above equation leads to

$$\int_{\ell T}^{(\ell+1)T} h_f(t) dt = \begin{cases} 1/8, & |\ell| = 1 \\ 1/4, & \ell = 0 \\ 0, & \text{otherwise} \end{cases}. \quad (4.172)$$

One way of obtaining $h_f(t)$ is to use a pulse $h_N(t)$ that satisfies Nyquist's third criterion [244, 258]

$$\int_{(2\ell-1)T/2}^{(2\ell+1)T/2} h_N(t) dt = \begin{cases} 1, & \ell = 0 \\ 0, & \ell \neq 0 \end{cases} \quad (4.173)$$

and generate $h_f(t)$ by using scaling and delay operations through the filter shown in Fig. 4.24. The transfer function of this filter is

$$\begin{aligned} H(f) &= \frac{1}{4} + \frac{1}{8} e^{-j2\pi f T} + \frac{1}{8} e^{j2\pi f T} \\ &= \frac{1}{2} \cos^2(\pi f T). \end{aligned} \quad (4.174)$$

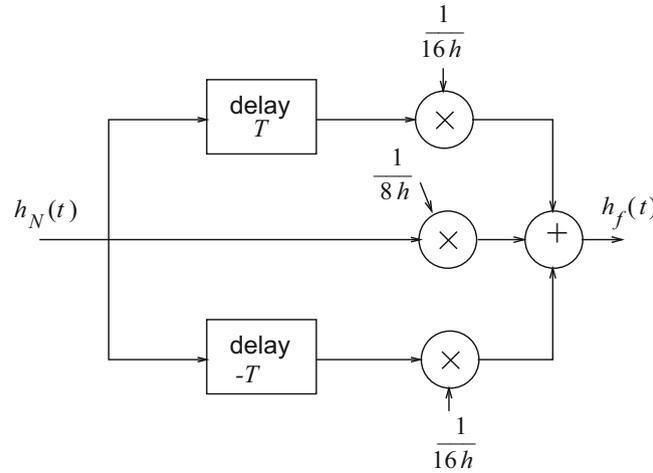


Fig. 4.24 Filter to generate a TFM frequency shaping pulse

The overall pulse $h_f(t)$ has the form

$$\begin{aligned} H_f(f) &= H_N(f)H(f) \\ &= H_N(f)\frac{1}{2}\cos^2(\pi fT). \end{aligned} \quad (4.175)$$

The filter $H(f)$ ensures that the phase constraint in (4.166) is satisfied. However, $H_N(f)$ determines the shape of the phase trajectories and can, therefore, influence the TFM power density spectrum. In general, $H_N(f)$ has the form

$$H_N(f) = \frac{\pi fT}{\sin(\pi fT)}N_1(f), \quad (4.176)$$

where $N_1(f)$ is the Fourier transform of a pulse that satisfies Nyquist's first criterion [244, 258]. One example of such a pulse is the raised cosine pulse $P(f)$ defined in (4.47). Consider, for example, the ideal Nyquist pulse (raised cosine pulse with $\beta = 0$)

$$N_1(f) = \begin{cases} 1, & 0 \leq |f| \leq 1/2T \\ 0, & \text{otherwise} \end{cases}. \quad (4.177)$$

Using (4.175)–(4.177) gives

$$H_f(f) = \begin{cases} \frac{1}{2}\frac{\pi fT}{\sin(\pi fT)}\cos^2(\pi fT), & 0 \leq |f| \leq 1/2T \\ 0, & \text{otherwise} \end{cases}. \quad (4.178)$$

The corresponding frequency shaping pulse $h_f(t)$ is plotted in Fig. 4.25. Note the close similarity to the GMSK frequency shaping pulse in Fig. 4.21.

Generalized TFM (GTFM) is an extension of TFM where the phase difference has the form

$$\phi(nT + T) - \phi(nT) = \frac{\pi}{2}(ax_{n-1} + bx_n + ax_{n+1}). \quad (4.179)$$

The constants a and b satisfy the condition $2a + b = 1$ so that the maximum change in excess phase during one bit period is equal to $\pm\pi/2$. A large variety of waveforms can be constructed by varying the value of b and the pulse response $N_1(f)$ in (4.176). TFM is a special case of GTFM where $b = 0.5$.

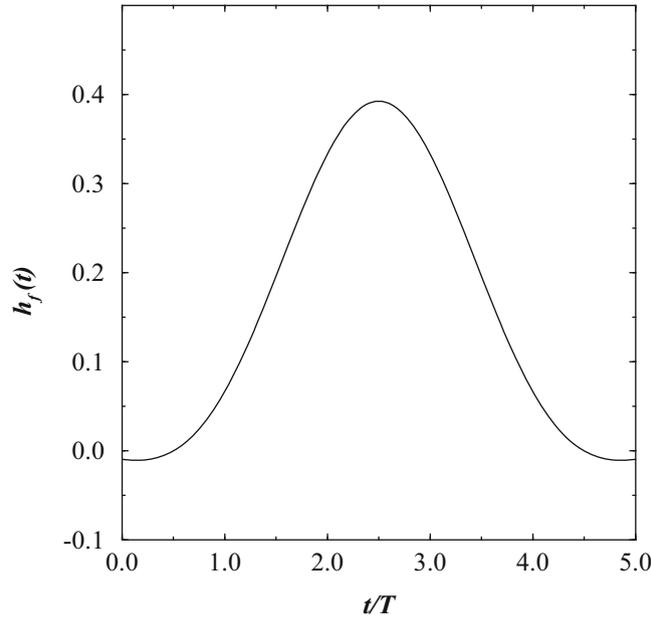


Fig. 4.25 TFM frequency shaping pulse

4.9 Power Spectrum

A digitally modulated bandpass waveform can be written in the generic form

$$\begin{aligned} s(t) &= \text{Re} \{ \tilde{s}(t) e^{j2\pi f_c t} \} \\ &= \frac{1}{2} \{ \tilde{s}(t) e^{j2\pi f_c t} + \tilde{s}^*(t) e^{-j2\pi f_c t} \}. \end{aligned} \quad (4.180)$$

Modulated waveforms belong to the class of cyclostationary or periodic wide-sense stationary random processes. The autocorrelation function of $s(t)$ is

$$\begin{aligned} \phi_{ss}(\tau) &= E[s(t)s(t+\tau)] \\ &= \frac{1}{4} E \left[\left(\tilde{s}(t) e^{j2\pi f_c t} + \tilde{s}^*(t) e^{-j2\pi f_c t} \right) \left(\tilde{s}(t+\tau) e^{j(2\pi f_c t + 2\pi f_c \tau)} + \tilde{s}^*(t+\tau) e^{-j(2\pi f_c t + 2\pi f_c \tau)} \right) \right] \\ &= \frac{1}{4} \left[E[\tilde{s}(t)\tilde{s}(t+\tau)] e^{j(4\pi f_c t + 2\pi f_c \tau)} + E[\tilde{s}(t)\tilde{s}^*(t+\tau)] e^{-j2\pi f_c \tau} \right. \\ &\quad \left. + E[\tilde{s}^*(t)\tilde{s}(t+\tau)] e^{j2\pi f_c \tau} + E[\tilde{s}^*(t)\tilde{s}^*(t+\tau)] e^{-j(4\pi f_c t + 2\pi f_c \tau)} \right]. \end{aligned} \quad (4.181)$$

If $s(t)$ is a wide-sense stationary random process, then the exponential terms that involve the time variable t must vanish, i.e., $E[\tilde{s}(t)\tilde{s}(t+\tau)] = 0$ and $E[\tilde{s}^*(t)\tilde{s}^*(t+\tau)] = 0$. Substituting $\tilde{s}(t) = \tilde{s}_I(t) + j\tilde{s}_Q(t)$ into these two expectations gives the requirement

$$\phi_{\tilde{s}_I\tilde{s}_I}(\tau) = E[\tilde{s}_I(t)\tilde{s}_I(t+\tau)] = E[\tilde{s}_Q(t)\tilde{s}_Q(t+\tau)] = \phi_{\tilde{s}_Q\tilde{s}_Q}(\tau) \quad (4.182)$$

$$\phi_{\tilde{s}_I\tilde{s}_Q}(\tau) = E[\tilde{s}_I(t)\tilde{s}_Q(t+\tau)] = -E[\tilde{s}_Q(t)\tilde{s}_I(t+\tau)] = -\phi_{\tilde{s}_Q\tilde{s}_I}(\tau). \quad (4.183)$$

Using these results,

$$\phi_{ss}(\tau) = \frac{1}{2} \phi_{\tilde{s}\tilde{s}}(\tau) e^{j2\pi f_c \tau} + \frac{1}{2} \phi_{\tilde{s}\tilde{s}}^*(\tau) e^{-j2\pi f_c \tau} \quad (4.184)$$

$$= \text{Re} \{ \phi_{\tilde{s}\tilde{s}}(\tau) e^{j2\pi f_c \tau} \}, \quad (4.185)$$

where

$$\begin{aligned}\phi_{\tilde{s}\tilde{s}}(\tau) &= \frac{1}{2}\mathbb{E}[\tilde{s}^*(t)\tilde{s}(t+\tau)] \\ &= \phi_{\tilde{s}_I\tilde{s}_I}(\tau) + j\phi_{\tilde{s}_I\tilde{s}_Q}(\tau)\end{aligned}\quad (4.186)$$

Finally, the power density spectrum is the Fourier transform of $\phi_{\tilde{s}\tilde{s}}(\tau)$ in (4.184), i.e.,

$$S_{\tilde{s}\tilde{s}}(f) = \frac{1}{2}\left(S_{\tilde{s}\tilde{s}}(f-f_c) + S_{\tilde{s}\tilde{s}}^*(-f-f_c)\right)\quad (4.187)$$

where $S_{\tilde{s}\tilde{s}}(f)$ is the power density spectrum of the complex envelope $\tilde{s}(t)$. Note that $S_{\tilde{s}\tilde{s}}(f)$ is real though not necessarily even, although $\tilde{s}(t)$ and $\phi_{\tilde{s}\tilde{s}}(\tau)$ are complex-valued; this property follows from the fact that $\phi_{\tilde{s}\tilde{s}}(\tau) = \phi_{\tilde{s}\tilde{s}}^*(-\tau)$ as shown in Appendix A. It follows that

$$S_{\tilde{s}\tilde{s}}(f) = \frac{1}{2}\left(S_{\tilde{s}\tilde{s}}(f-f_c) + S_{\tilde{s}\tilde{s}}(-f-f_c)\right).\quad (4.188)$$

From the above expression, observe that the psd of the bandpass waveform $s(t)$ is real and even, and is completely determined by the psd of its complex envelope $\tilde{s}(t)$ and the carrier frequency f_c .

4.9.1 Psd of the Complex Envelope

The complex envelope of any digitally modulated signal can be expressed in the standard form

$$\tilde{s}(t) = A \sum_n b(t-nT, \mathbf{x}_n).\quad (4.189)$$

The autocorrelation of $\tilde{s}(t)$ is

$$\begin{aligned}\phi_{\tilde{s}\tilde{s}}(t, t+\tau) &= \frac{1}{2}\mathbb{E}[\tilde{s}^*(t)\tilde{s}(t+\tau)] \\ &= \frac{A^2}{2} \sum_i \sum_k \mathbb{E}[b^*(t-iT, \mathbf{x}_i)b(t+\tau-kT, \mathbf{x}_k)].\end{aligned}\quad (4.190)$$

Observe that $\tilde{s}(t)$ is a cyclostationary random process, meaning that the autocorrelation function $\phi_{\tilde{s}\tilde{s}}(t, t+\tau)$ is periodic in t with period T . To see this property, first note that

$$\begin{aligned}\phi_{\tilde{s}\tilde{s}}(t+T, t+T+\tau) &= \frac{A^2}{2} \sum_i \sum_k \mathbb{E}[b^*(t+T-iT, \mathbf{x}_i)b(t+T+\tau-kT, \mathbf{x}_k)] \\ &= \frac{A^2}{2} \sum_{i'} \sum_{k'} \mathbb{E}[b^*(t-i'T, \mathbf{x}_{i'+1})b(t+\tau-k'T, \mathbf{x}_{k'+1})].\end{aligned}\quad (4.191)$$

Under the assumption that the information sequence is a stationary random process it follows that

$$\begin{aligned}\phi_{\tilde{s}\tilde{s}}(t+T, t+T+\tau) &= \frac{A^2}{2} \sum_{i'} \sum_{k'} \mathbb{E}[b^*(t-i'T, \mathbf{x}_{i'})b(t+\tau-k'T, \mathbf{x}_{k'})] \\ &= \phi_{\tilde{s}\tilde{s}}(t, t+\tau).\end{aligned}\quad (4.192)$$

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

$$\begin{aligned}
\phi_{\tilde{s}\tilde{s}}(\tau) &= \langle \phi_{\tilde{s}\tilde{s}}(t, t + \tau) \rangle \\
&= \frac{A^2}{2} \sum_i \sum_k \frac{1}{T} \int_0^T \mathbb{E} [b^*(t - iT, \mathbf{x}_i) b(t + \tau - kT, \mathbf{x}_k)] dt \\
&= \frac{A^2}{2T} \sum_i \sum_k \int_{-iT}^{-iT+T} \mathbb{E} [b^*(z, \mathbf{x}_i) b(z + \tau - (k - i)T, \mathbf{x}_k)] dz \\
&= \frac{A^2}{2T} \sum_i \sum_m \int_{-iT}^{-iT+T} \mathbb{E} [b^*(z, \mathbf{x}_i) b(z + \tau - mT, \mathbf{x}_{m+i})] dz \\
&= \frac{A^2}{2T} \sum_i \sum_m \int_{-iT}^{-iT+T} \mathbb{E} [b^*(z, \mathbf{x}_0) b(z + \tau - mT, \mathbf{x}_m)] dz \\
&= \frac{A^2}{2T} \sum_m \int_{-\infty}^{\infty} \mathbb{E} [b^*(z, \mathbf{x}_0) b(z + \tau - mT, \mathbf{x}_m)] dz,
\end{aligned} \tag{4.193}$$

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 $\phi_{\tilde{s}\tilde{s}}(\tau)$,²

$$\begin{aligned}
S_{\tilde{s}\tilde{s}}(f) &= \mathbb{E} \left[\frac{A^2}{2T} \sum_m \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b^*(z, \mathbf{x}_0) b(z + \tau - mT, \mathbf{x}_m) dz e^{-j2\pi f\tau} d\tau \right] \\
&= \mathbb{E} \left[\frac{A^2}{2T} \sum_m \int_{-\infty}^{\infty} b^*(z, \mathbf{x}_0) e^{j2\pi f z} dz \int_{-\infty}^{\infty} b(z + \tau - mT, \mathbf{x}_m) e^{-j2\pi f(z + \tau - mT)} d\tau dz e^{-j2\pi f m T} \right] \\
&= \mathbb{E} \left[\frac{A^2}{2T} \sum_m \int_{-\infty}^{\infty} b^*(z, \mathbf{x}_0) e^{j2\pi f z} dz \int_{-\infty}^{\infty} b(\tau', \mathbf{x}_m) e^{-j2\pi f \tau'} d\tau' e^{-j2\pi f m T} \right] \\
&= \frac{A^2}{2T} \sum_m \mathbb{E} [B^*(f, \mathbf{x}_0) B(f, \mathbf{x}_m)] e^{-j2\pi f m T},
\end{aligned} \tag{4.194}$$

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

$$S_{b,m}(f) = \frac{1}{2} \mathbb{E} [B^*(f, \mathbf{x}_0) B(f, \mathbf{x}_m)]. \tag{4.195}$$

Then

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T} \sum_m S_{b,m}(f) e^{-j2\pi f m T}. \tag{4.196}$$

Note that the psd in (4.196) depends on the correlation properties of the information sequence \mathbf{x}_m and the form of the generalized 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

$$S_{b,m}(f) = S_{b,K}(f), \quad |m| \geq K, \tag{4.197}$$

²Note that expectation and integration are linear operations and their order can be exchanged.

where

$$\begin{aligned}
 S_{b,K}(f) &= \frac{1}{2} \mathbb{E} [B^*(f, \mathbf{x}_0)] \mathbb{E} [B(f, \mathbf{x}_m)] \\
 &= \frac{1}{2} \mathbb{E} [B^*(f, \mathbf{x}_0)] \mathbb{E} [B(f, \mathbf{x}_0)] \\
 &= \frac{1}{2} |\mathbb{E} [B(f, \mathbf{x}_0)]|^2, \quad |m| \geq K.
 \end{aligned} \tag{4.198}$$

It follows that

$$S_{\overline{ss}}(f) = S_{\overline{ss}}^c(f) + S_{\overline{ss}}^d(f), \tag{4.199}$$

where

$$\begin{aligned}
 S_{\overline{ss}}^c(f) &= \frac{A^2}{T} \sum_{|m| < K} (S_{b,m}(f) - S_{b,K}(f)) e^{-j2\pi f m T} \\
 S_{\overline{ss}}^d(f) &= \frac{A^2}{T} S_{b,K}(f) \sum_m e^{-j2\pi f m T}.
 \end{aligned} \tag{4.200}$$

The terms $S_{\overline{ss}}^c(f)$ and $S_{\overline{ss}}^d(f)$ represent the continuous and discrete (line) portions of the psd. The fact that $S_{\overline{ss}}^d(f)$ represents the discrete portion can be seen more clearly by using the identity

$$T \sum_m e^{-j2\pi f m T} = \sum_n \delta \left(f - \frac{n}{T} \right) \tag{4.201}$$

to write

$$S_{\overline{ss}}^d(f) = \left(\frac{A}{T} \right)^2 S_{b,K}(f) \sum_n \delta \left(f - \frac{n}{T} \right). \tag{4.202}$$

Finally, by using the property $S_{b,-m}(f) = S_{b,m}^*(f)$, the continuous portion of the psd can be written as

$$\begin{aligned}
 S_{\overline{ss}}^c(f) &= \frac{A^2}{T} (S_{b,0}(f) - S_{b,K}(f)) + \frac{A^2}{T} \sum_{m=1}^K \left((S_{b,m}(f) - S_{b,K}(f)) e^{-j2\pi f m T} + (S_{b,m}^*(f) - S_{b,K}(f)) e^{j2\pi f m T} \right) \\
 &= \frac{A^2}{T} (S_{b,0}(f) - S_{b,K}(f)) + \frac{A^2}{T} 2\text{Re} \left\{ \sum_{m=1}^K (S_{b,m}(f) - S_{b,K}(f)) e^{-j2\pi f m T} \right\}.
 \end{aligned} \tag{4.203}$$

Note that the ensemble average and Fourier transform are interchangeable linear operators. Therefore, if the complex envelope $\tilde{s}(t)$ has zero mean, i.e., $\mathbb{E}[b(t, \mathbf{x}_0)] = 0$, then $\mathbb{E}[B(f, \mathbf{x}_0)] = 0$. Under this condition

$$S_{b,K}(f) = \frac{1}{2} |\mathbb{E}[B(f, \mathbf{x}_0)]|^2 = 0. \tag{4.204}$$

Hence, if $b(t, \mathbf{x}_0)$ has zero mean, then $S_{\overline{ss}}(f)$ contains no discrete components and $S_{\overline{ss}}(f) = S_{\overline{ss}}^c(f)$. Conversely, if $b(t, \mathbf{x}_0)$ has non-zero mean, then $S_{\overline{ss}}(f)$ will contain discrete (line) components. Another important case arises with uncorrelated zero-mean data, where $S_{b,K}(f) = 0$, $K = 1$. In this case, only the term $S_{b,0}(f)$ remains and

$$S_{\overline{ss}}(f) = \frac{A^2}{T} S_{b,0}(f) \tag{4.205}$$

where

$$S_{b,0}(f) = \frac{1}{2} \mathbb{E} [|B(f, \mathbf{x}_0)|^2]. \tag{4.206}$$

4.9.1.1 Alternative Method

An alternative method of computing the psd is as follows. From the first line in (4.194)

$$\begin{aligned} S_{\overline{ss}}(f) &= \mathbb{E} \left[\frac{A^2}{2T} \sum_m \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b(z, \mathbf{x}_0) b^*(z + \tau - mT, \mathbf{x}_m) dz e^{-j2\pi f \tau} d\tau \right] \\ &= \frac{A^2}{2T} \sum_m \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{E} [b(z, \mathbf{x}_0) b^*(\tau', \mathbf{x}_m)] e^{-j2\pi f(\tau' - z)} dz d\tau' e^{-j2\pi f m T}. \end{aligned} \quad (4.207)$$

Therefore, $S_{b,m}(f)$ is given by the double Fourier transform

$$S_{b,m}(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{b,m}(z, \tau') e^{-j2\pi f(\tau' - z)} dz d\tau'. \quad (4.208)$$

where

$$\phi_{b,m}(z, \tau') = \frac{1}{2} \mathbb{E} [b(z, \mathbf{x}_0) b^*(\tau', \mathbf{x}_m)]. \quad (4.209)$$

4.9.1.2 Linear Full Response Modulation

Consider linear full response modulation schemes where $b(t, \mathbf{x}_n) = x_n h_a(t)$ and $B(f, \mathbf{x}_n) = x_n H_a(f)$. In general, the data sequence $\{x_n\}$ will be correlated with autocorrelation function $\phi_{xx}(m)$. From (4.195),

$$S_{b,m}(f) = \phi_{xx}(m) |H_a(f)|^2, \quad (4.210)$$

where

$$\phi_{xx}(m) = \frac{1}{2} \mathbb{E} [x_k^* x_{k+m}]. \quad (4.211)$$

Hence, from (4.196) the psd of the complex envelope is

$$S_{\overline{ss}}(f) = \frac{A^2}{T} |H_a(f)|^2 S_{xx}(f), \quad (4.212)$$

where

$$S_{xx}(f) = \sum_m \phi_{xx}(m) e^{-j2\pi f m T}. \quad (4.213)$$

Note that the psd is the product of two components; one depends on the squared magnitude of the amplitude shaping function and the other depends on the correlation of the data sequence. With uncorrelated data symbols

$$S_{b,0}(f) = \sigma_x^2 |H_a(f)|^2 \quad (4.214)$$

$$S_{b,m}(f) = \frac{1}{2} |\mu_x|^2 |H_a(f)|^2, \quad |m| \geq 1. \quad (4.215)$$

where $\mu_x = \mathbb{E}[x_m]$ and $\sigma_x^2 = \frac{1}{2} \mathbb{E}[|x_k|^2]$ are the mean and variance of the data symbols, respectively. The psd $S_{\overline{ss}}(f)$ is then given by (4.199), where

$$S^d(f) = \frac{A^2}{T^2} S_{b,1}(f) \sum_n \delta\left(f - \frac{n}{T}\right) \quad (4.216)$$

$$S^c(f) = \frac{A^2}{T} (S_{b,0}(f) - S_{b,1}(f)). \quad (4.217)$$

If $\mu_x = 0$, then $S_{b,1}(f) = 0$ and the psd has the simple form

$$S_{ss}(f) = \frac{A^2}{T} \sigma_x^2 |H_a(f)|^2. \quad (4.218)$$

In this case, the psd only depends on the amplitude shaping pulse $h_a(t)$.

4.9.1.3 Linear Partial Response Modulation

Consider linear partial response modulation schemes where $h_a(t)$ has duration LT . Following the development in Sect. 4.8 the generalized shaping function has the form

$$\begin{aligned} b(t, \mathbf{x}_m) &= h_a(t, \mathbf{x}_m) \\ &= \sum_{k=0}^{L-1} x_{m-k} h_{a,k}(t), \end{aligned} \quad (4.219)$$

where

$$h_{a,k}(t) = h_a(t + kT)u_T(t). \quad (4.220)$$

Taking the Fourier transform of (4.219) gives

$$B(f, \mathbf{x}_m) = \sum_{k=0}^{L-1} x_{m-k} H_{a,k}(f). \quad (4.221)$$

From (4.195),

$$\begin{aligned} S_{b,m}(f) &= \frac{1}{2} \mathbb{E} \left[\sum_{\ell=0}^{L-1} x_{-\ell}^* H_{a,\ell}^*(f) \sum_{k=0}^{L-1} x_{m-k} H_{a,k}(f) \right] \\ &= \sum_{k=0}^{L-1} \sum_{\ell=0}^{L-1} \phi_{xx}(m-k+\ell) H_{a,\ell}^*(f) H_{a,k}(f). \end{aligned} \quad (4.222)$$

For the special case of uncorrelated zero-mean data symbols, $\phi_{xx}(m-k+\ell) = \sigma_x^2 \delta(m-k+\ell)$. Hence,

$$S_{b,m}(f) = \sigma_x^2 \sum_{\ell=0}^{L-1} H_{a,\ell}^*(f) H_{a,m+\ell}(f) \quad (4.223)$$

where

$$\sigma_x^2 = \frac{1}{2} \mathbb{E}[|x_0|^2]$$

is the variance of the data symbols.

Example 4.4 (Duobinary Signaling). For duobinary signaling, $L = 2$ and $h_{a,0}(t) = h_{a,1}(t) = \text{sinc}(t/T)$ and $H_{a,0}(f) = H_{a,1}(f) = T\text{rect}(fT)$, where

$$\text{rect}(fT) = \begin{cases} 1, & |f| \leq \frac{1}{2T} \\ 0, & \text{elsewhere} \end{cases}.$$

With uncorrelated zero-mean data symbols

$$\begin{aligned} S_{b,m}(f) &= \frac{1}{2} \text{E} [(x_0^* H_{a,0}^*(f) + x_{-1}^* H_{a,1}^*(f)) (x_m H_{a,0}(f) + x_{m-1} H_{a,1}(f))] \\ &= \begin{cases} 2\sigma_x^2 T^2 \text{rect}(fT) & , m = 0 \\ \sigma_x^2 T^2 \text{rect}(fT) & , |m| = 1 \\ 0 & , \text{otherwise} \end{cases} \end{aligned}$$

and from (4.196)

$$S_{\overline{ss}}(f) = 2A^2 T \sigma_x^2 \cos^2(\pi f T) \text{rect}(fT). \quad (4.224)$$

Example 4.5 (Modified Duobinary Signaling). For modified duobinary signaling, $L = 3$ and $h_{a,0}(t) = h_{a,2}(t) = \text{sinc}(t/T)$ and $h_{a,1}(t) = 0$. With uncorrelated zero-mean data symbols,

$$S_{b,m}(f) = \begin{cases} 2\sigma_x^2 T^2 \text{rect}(fT) & , m = 0 \\ -\sigma_x^2 T^2 \text{rect}(fT) & , |m| = 2 \\ 0 & , \text{otherwise} \end{cases}$$

and from (4.196)

$$S_{\overline{ss}}(f) = 2A^2 T \sigma_x^2 \sin^2(2\pi f T) \text{rect}(fT).$$

4.9.2 Psd of QAM

The psd of QAM with uncorrelated zero-mean data symbols is given by (4.218). If $h_a(t) = u_T(t)$, then

$$S_{\overline{ss}}(f) = A^2 T \sigma_x^2 \left(\frac{\sin(\pi f T)}{\pi f T} \right)^2. \quad (4.225)$$

With root-raised cosine pulse shaping, $|H_a(f)|^2 = P(f)$ has the form defined in (4.47) with $h_a(t)$ in (4.50). 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 $\tau = LT$. This truncation produces the new pulse $\tilde{h}_a(t) = h_a(t)\text{rect}(t/LT)$. The Fourier transform of the truncated pulse $\tilde{h}_a(t)$ is $\tilde{H}_a(f) = H_a(f) * LT\text{sinc}(\pi f LT)$, 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 (4.218) by simply replacing $H_a(f)$ with $\tilde{H}_a(f)$. As shown in Fig. 4.26, pulse truncation can lead to significant side lobe regeneration.

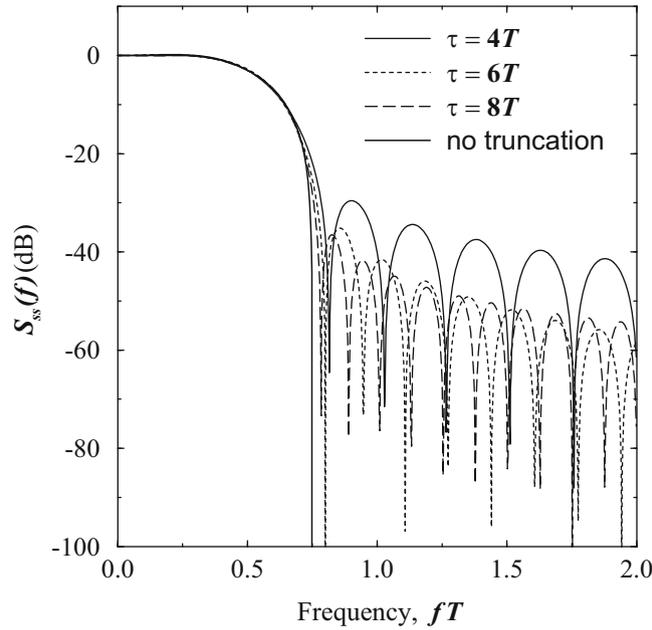


Fig. 4.26 Psd of QAM with a truncated square root-raised cosine amplitude shaping pulse with various truncation lengths; $\beta = 0.5$. Truncation of the amplitude shaping pulse leads to side lobe regeneration

4.9.3 Psd of PSK

For PSK signals with the uncorrelated data symbols and the generalized shaping function in (4.61), the psd is given by (4.218). Hence, PSK signals have the same psd as QAM signals. As such, the psd with truncated root-raised cosine amplitude pulse shaping is shown in Fig. 4.26.

4.9.4 Psd of OQPSK

For OQPSK, the generalized shaping function is

$$b(t, \mathbf{x}_n) = b(t, x_n) = x_{I,n}h_a(t) + jx_{Q,n}h_a(t - T/2) \quad (4.226)$$

where $x_{I,n}, x_{Q,n} \in \{-1/\sqrt{2}, +1/\sqrt{2}\}$. It follows that

$$B(f, \mathbf{x}_n) = (x_{I,n} + jx_{Q,n}e^{-j2\pi fT/2})H_a(f) \quad (4.227)$$

With uncorrelated data symbols,

$$\begin{aligned} S_{b,0}(f) &= \frac{1}{2} \mathbb{E}[|B(f, x_0)|^2] \\ &= \frac{1}{2} |H_a(f)|^2 \end{aligned} \quad (4.228)$$

Therefore,

$$S_{ss}(f) = \frac{A^2}{2T} |H_a(f)|^2 \quad (4.229)$$

Hence, OQPSK has the same psd as QPSK. However, OQASK has a lower peak-to-average power ratio than QPSK.

4.9.5 Psd of $\pi/4$ -DQPSK

To find the psd of $\pi/4$ -DQPSK, the autocorrelation

$$\phi_{b,m}(z, \tau') = \frac{1}{2} \mathbb{E} [b^*(z, \mathbf{x}_0) b(\tau', \mathbf{x}_m)], \quad (4.230)$$

is first computed, where $b(t, \mathbf{x}_n)$ is defined in (4.70). For $m > 0$,

$$\begin{aligned} \phi_{b,m}(z, \tau') &= \frac{1}{2} \mathbb{E} \left[h_a(z) \exp \left\{ j \frac{\pi}{4} \sum_{k=1}^m x_k \right\} h_a(\tau') \right] \\ &= \frac{1}{2} \mathbb{E} \left[\exp \left\{ j \frac{\pi}{4} \sum_{k=1}^m x_k \right\} \right] h_a(z) h_a(\tau') \\ &= 0. \end{aligned} \quad (4.231)$$

For $m = 0$,

$$\phi_{b,0}(z, \tau') = \frac{1}{2} \mathbb{E} [h_a(z) h_a(\tau')] = \frac{1}{2} h_a(z) h_a(\tau'). \quad (4.232)$$

Taking the double Fourier transform gives

$$\begin{aligned} S_{b,0}(f) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{b,0}(z, \tau') e^{-j2\pi f(\tau' - z)} dz d\tau' \\ &= \frac{1}{2} |H_a(f)|^2. \end{aligned} \quad (4.233)$$

Finally, the psd is

$$S_{\overline{ss}}(f) = \frac{A^2}{2T} |H_a(f)|^2. \quad (4.234)$$

Just like OQPSK, $\pi/4$ -DQPSK has the same psd as QPSK. However, as discussed earlier, $\pi/4$ -DQPSK has a lower peak-to-average power ratio than QPSK.

4.9.6 Psd of OFDM

Recall that the OFDM waveform with guard interval is given by (4.88) and (4.89). The data symbols $x_{n,k}$, $k = 0, \dots, N-1$ that modulate the N subcarriers are assumed to have zero mean, variance $\sigma_x^2 = \frac{1}{2} \mathbb{E}[|x_{n,k}|^2]$, and they are mutually uncorrelated. In this case, the psd of the OFDM waveform is

$$S_{\overline{ss}}(f) = \frac{A^2}{T_g} S_{b,0}(f), \quad (4.235)$$

where

$$S_{b,0}(f) = \frac{1}{2} \mathbb{E} [|B(f, \mathbf{x}_0)|^2], \quad (4.236)$$

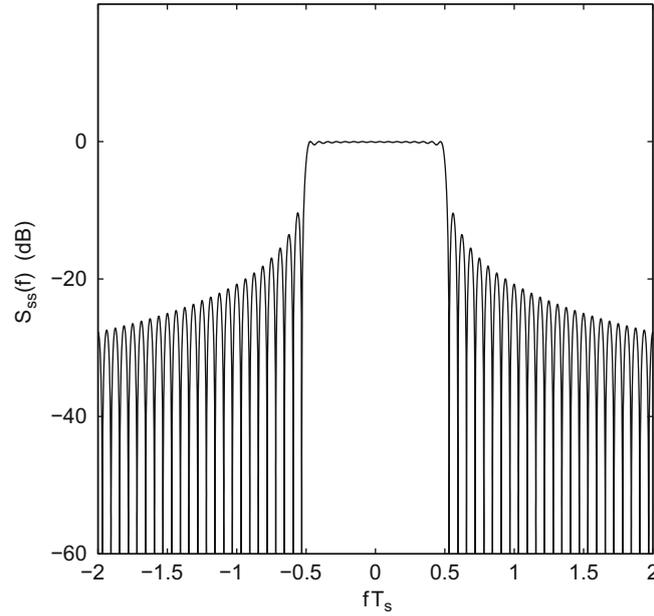


Fig. 4.27 Psd of OFDM with $N = 16, \alpha_g = 0$

and

$$B(f, \mathbf{x}_0) = \sum_{k=0}^{N-1} x_{0,k} T \text{sinc}(fT - k) + \sum_{k=0}^{N-1} x_{0,k} \alpha_g T \text{sinc}(\alpha_g(fT - k)) e^{j2\pi fT}. \quad (4.237)$$

Substituting (4.237) into (4.236) along with $T = NT_s$ and $T_g = (1 + \alpha_g)T$ yields the result

$$\begin{aligned} S_{\overline{ss}}(f) = & \sigma_x^2 A^2 T \left(\frac{1}{1 + \alpha_g} \sum_{k=0}^{N-1} \text{sinc}^2(NfT_s - k) \right. \\ & + \frac{\alpha_g^2}{1 + \alpha_g} \sum_{k=0}^{N-1} \text{sinc}^2(\alpha_g(NfT_s - k)) \\ & \left. + \frac{2\alpha_g}{1 + \alpha_g} \cos(2\pi NfT_s) \sum_{k=0}^{N-1} \text{sinc}(NfT_s - k) \text{sinc}(\alpha_g(NfT_s - k)) \right). \end{aligned} \quad (4.238)$$

The OFDM psd is plotted in Figs. 4.27 and 4.28 for $N = 16, \alpha_g = 0$ and $N = 16, \alpha_g = 0.25$, respectively. Observe the effect of the OFDM guard interval on the psd. Likewise, Figs. 4.29 and 4.30 plot the psd for $N = 1024, \alpha_g = 0$ and $N = 1024, \alpha_g = 0.25$, respectively, where the effect of increasing the block size N can be observed. When plotting the above figures, the index k was replaced with $k - (N - 1)/2$ in the argument of the sinc functions in (4.239) to center the spectrum around 0 Hz. Note that the psd is plotted against the normalized frequency fT_s . To avoid a reduction in data rate, the modulated symbol period with a cyclic extension is $T_s^g = T_s/(1 + \alpha_g)$. Hence, the Nyquist frequency in this case is $1/2T_s^g = (1 + \alpha_g)/2T_s$, which shows a bandwidth expansion due to the guard interval.

4.9.6.1 Psd of OFDM with IDFT Baseband Modulator

It is interesting to examine the OFDM power spectrum when the OFDM complex envelope is generated by using an IDFT baseband modulator followed by a balanced pair DACs as shown in Fig. 4.14. The output of the IDFT baseband modulator is given by $\{\mathbf{X}^g\} = \{X_{n,m}^g\}$, where m is the block index and

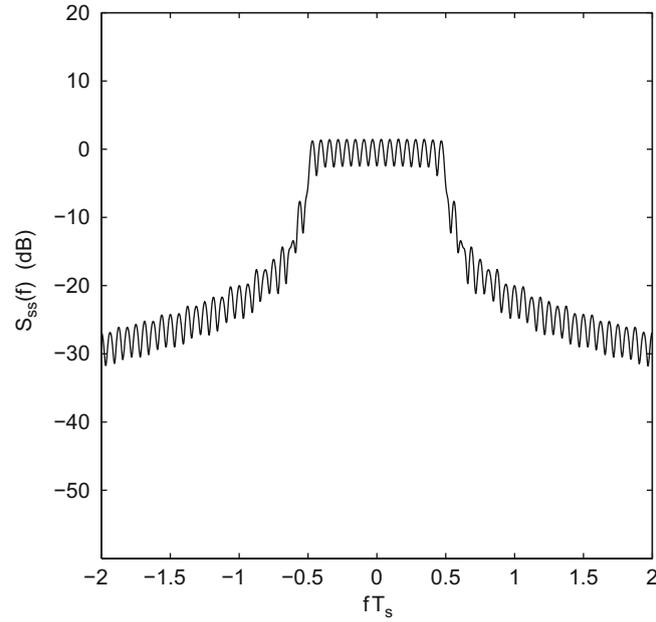


Fig. 4.28 Psd of OFDM with $N = 16$, $\alpha_g = 0.25$

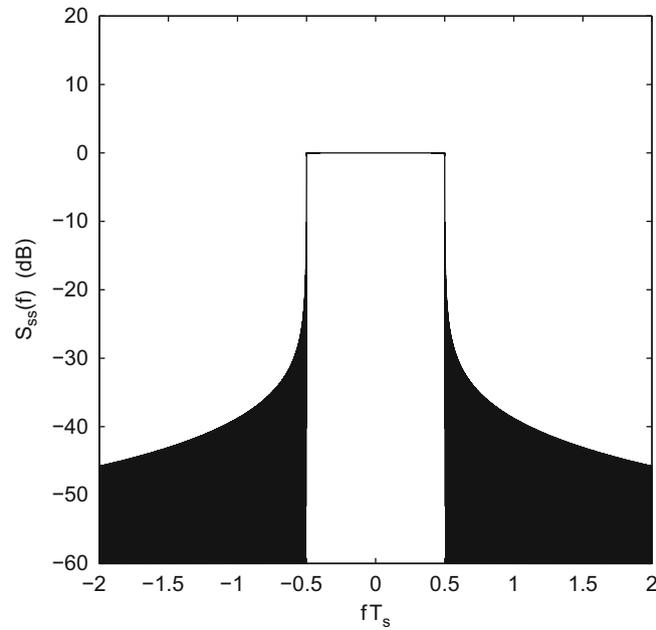


Fig. 4.29 Psd of OFDM with $N = 1024$, $\alpha_g = 0$

$$X_{n,m}^g = X_{n,(m)_N} \quad (4.239)$$

$$= A \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi km}{N}}, \quad m = 0, 1, \dots, N + G - 1. \quad (4.240)$$

The power spectrum of the sequence $\{\mathbf{X}^g\}$ can be calculated by first determining the discrete-time autocorrelation function of the time domain sequence $\{\mathbf{X}^g\}$ and then taking a discrete-time Fourier transform of the discrete-time autocorrelation function. The psd of the OFDM complex envelope with ideal DACs can be obtained by applying the resulting power spectrum to an ideal low-pass filter with a cutoff frequency of $1/(2T_s^g)$ Hz.

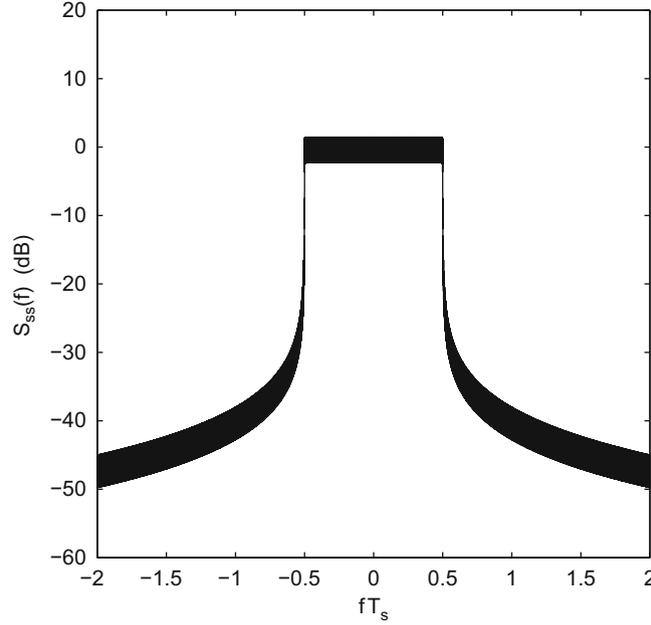


Fig. 4.30 Psd of OFDM with $N = 1024$, $\alpha_g = 0.25$

The time domain sequence $\{\mathbf{X}^g\}$ is a periodic wide-sense stationary sequence having the discrete-time autocorrelation function

$$\phi_{X^g X^g}(m, \ell) = \frac{1}{2} \mathbb{E}[(X_{n,m}^g)^* X_{n,m+\ell}^g] \quad (4.241)$$

$$= A^2 \sum_{k=0}^{N-1} \sum_{i=0}^{N-1} \frac{1}{2} \mathbb{E}[x_{n,k}^* x_{n,i}] e^{j\frac{2\pi}{N}(-km+im+i\ell)}, \quad (4.242)$$

$$\text{for } m = 0, \dots, N + G - 1. \quad (4.243)$$

The data symbols, $x_{n,k}$, are assumed to be mutually uncorrelated with zero mean and variance $\sigma_x^2 = \frac{1}{2} \mathbb{E}[|x_{n,k}|^2]$. Using the fact, that $X_{n,m}^g = X_{n,(m)_N}$, gives

$$\phi_{X^g X^g}(m, \ell) = \begin{cases} A\sigma_x^2 & m = 0, \dots, G-1, \ell = 0, N \\ A\sigma_x^2 & m = G, \dots, N-1, \ell = 0 \\ 0 & m = N, \dots, N+G-1, \ell = 0, -N \\ 0 & \text{otherwise} \end{cases} \quad (4.244)$$

Averaging over all time indices m gives the time-averaged discrete-time autocorrelation function

$$\phi_{X^g X^g}(\ell) = \begin{cases} A\sigma_x^2 & \ell = 0 \\ \frac{G}{N+G} A\sigma_x^2 & \ell = -N, N \\ 0 & \text{otherwise} \end{cases} \quad (4.245)$$

Taking the discrete-time Fourier transform of the discrete-time autocorrelation function in (4.245) gives

$$\begin{aligned} S_{X^g X^g}(f) &= \sum_{\ell=-\infty}^{\infty} \phi_{X^g X^g}(\ell) e^{-j2\pi f \ell T_s^g} \\ &= A\sigma_x^2 \left(1 + \frac{G}{N+G} e^{j2\pi f N T_s^g} + \frac{G}{N+G} e^{-j2\pi f N T_s^g} \right) \\ &= A\sigma_x^2 \left(1 + \frac{2G}{N+G} \cos(2\pi f N T_s^g) \right). \end{aligned} \quad (4.246)$$

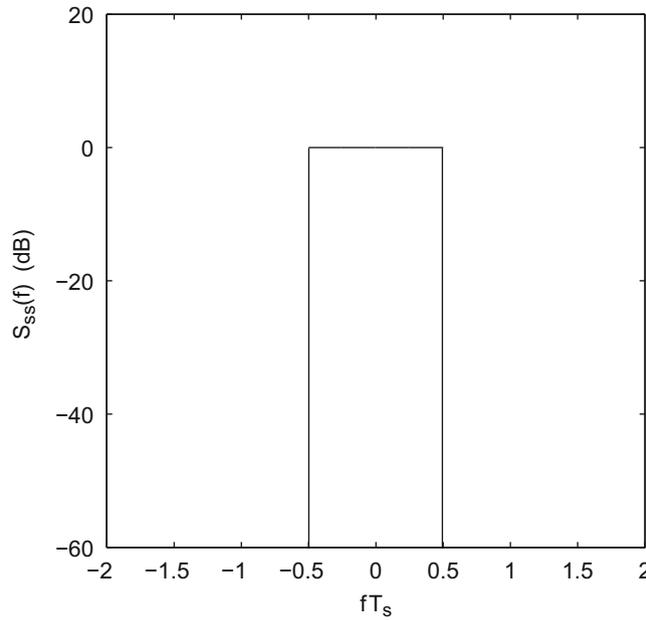


Fig. 4.31 Psd of IDFT-based OFDM with $N = 16$, $G = 0$. Note in this case that $T_s^g = T_s$

Finally, it is assumed here that the sequence $\{\mathbf{X}^g\} = \{X_{n,m}^g\}$ is passed through a pair of ideal DACs. The ideal DAC is a low-pass filter with cutoff frequency $1/(2T_s^g)$. Therefore, the OFDM complex envelope with an ideal DAC has the psd

$$S_{\bar{s}s}(f) = A\sigma_x^2 \left(1 + \frac{2G}{N+G} \cos(2\pi fNT_s^g) \right) \text{rect}(fT_s^g). \quad (4.247)$$

The OFDM psd is plotted in Fig. 4.31 for $G = 0$, where it has the ideal rectangular form $\text{rect}(fT_s)$ for any block size N . Figures 4.32 and 4.33 plot the psd for $N = 16$, $G = 4$, and $N = 1024$, $G = 256$, respectively, where the effect of the cyclic guard interval and an increase in the block size N can be seen.

Finally, the psd plotted in Figs. 4.31, 4.32, and 4.33 assumes an ideal DAC. A practical DAC with a finite-length reconstruction filter will introduce side lobes into the spectrum. Side lobes are inherently present in the continuous-time OFDM waveform in (4.88) and (4.89) due to the use of rectangular amplitude pulse shaping on the subcarriers. However, they are introduced into the IDFT implementation by the non-ideal (practical) DAC.

4.9.7 Psd of Full Response CPM

Recall that the generalized shaping function for a CPM waveform is given by (4.103). To compute the psd, first define the auxiliary function

$$r(t, x_n) \triangleq e^{j2\pi h x_n \beta(t)} u_T(t), \quad (4.248)$$

such that

$$b(t, \mathbf{x}_n) = e^{j\pi h \sum_{k=0}^{n-1} x_k} r(t, x_n) \quad (4.249)$$

and calculate the mean and autocorrelation function of $r(t, x_n)$. If M -ary signaling is used with the values of x_k defined by

$$x_k \in \{2m - 1 - M : m = 1, 2, \dots, M\} \quad (4.250)$$

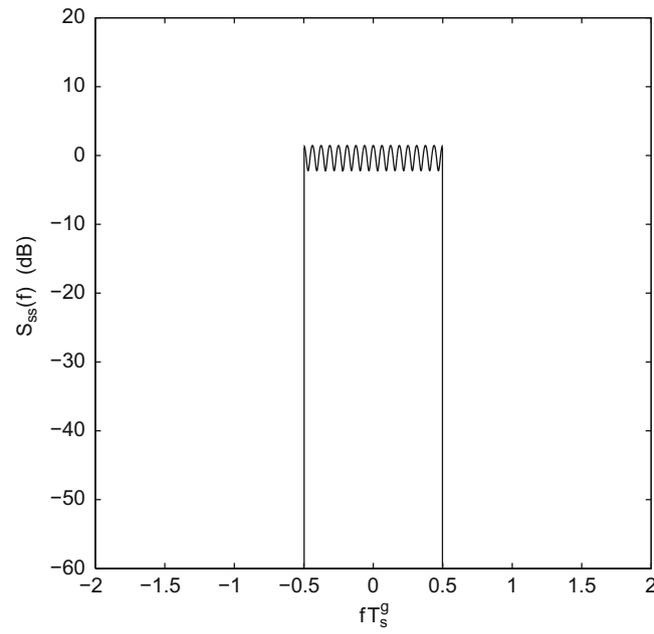


Fig. 4.32 Psd of IDFT-based OFDM with $N = 16, G = 4$

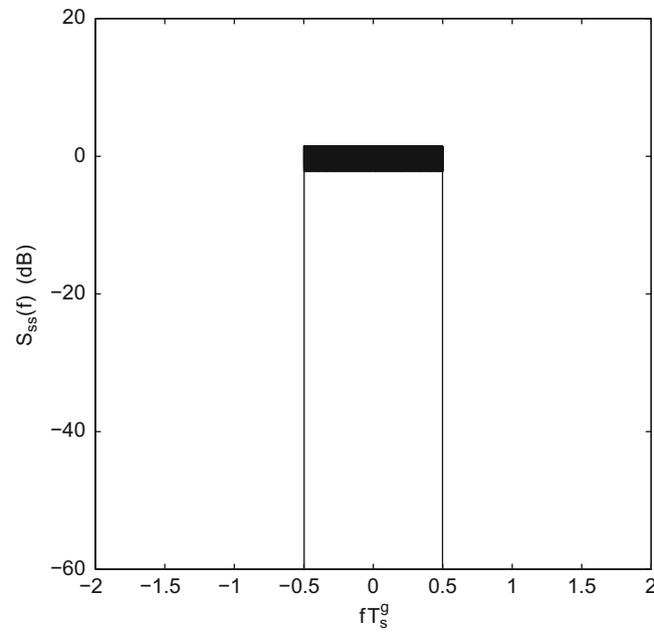


Fig. 4.33 Psd of IDFT-based OFDM with $N = 1024, G = 256$

then

$$\begin{aligned}
 m_r(t) &\triangleq \mathbb{E}[r(t, x_n)] \\
 &= \frac{1}{M} \sum_{i=1}^M e^{j2\pi h(2i-1-M)\beta(t)} u_T(t) \\
 &= D_M(2\pi h\beta(t)) u_T(t).
 \end{aligned} \tag{4.251}$$

where

$$D_M(x) \triangleq \frac{\sin(Mx)}{M \sin x} \quad (4.252)$$

is the Dirichlet function. Also

$$\phi_{r,m}(t, t') = \frac{1}{2} \mathbb{E} [r^*(t, x_0) r(t', x_m)]. \quad (4.253)$$

Evaluating the above expression for $m = 0$ gives the following result which will be used later

$$\begin{aligned} \phi_{r,0}(t, t') &= \frac{1}{2} \mathbb{E} [r^*(t, x_0) r(t', x_0)] \\ &= \frac{1}{2} \mathbb{E} \left[e^{-j2\pi h x_0 \beta(t)} e^{j2\pi h x_0 \beta(t')} u_T(t) u_T(t') \right] \\ &= \frac{1}{2} \mathbb{E} \left[e^{-j2\pi h x_0 (\beta(t) - \beta(t'))} u_T(t) u_T(t') \right] \\ &= \frac{1}{2} D_M \left(2\pi h (\beta(t) - \beta(t')) \right) u_T(t) u_T(t'). \end{aligned} \quad (4.254)$$

To evaluate the psd, it is also necessary to compute the autocorrelation of $b(t, \mathbf{x}_m)$. This can be done as follows:

$$\begin{aligned} \phi_{b,m}(t, t') &= \frac{1}{2} \mathbb{E} [b^*(t, \mathbf{x}_0) b(t', \mathbf{x}_m)] \\ &= \frac{1}{2} \mathbb{E} \left[e^{-j\pi h \sum_{k=0}^{m-1} x_k} r^*(t, x_0) r(t', x_m) \right] \\ &= \frac{1}{2} \mathbb{E} \left[\left(\prod_{k=0}^{m-1} r^*(T, x_k) \right) r^*(t, x_0) r(t', x_m) \right] \\ &= \frac{1}{2} \mathbb{E} \left[\left(\prod_{k=1}^{m-1} r^*(T, x_k) \right) r^*(t, x_0) r^*(T, x_0) r(t', x_m) \right]. \end{aligned} \quad (4.255)$$

Now suppose that the data sequence is uncorrelated. Then for $m > 0$

$$\begin{aligned} \phi_{b,m}(t, t') &= \frac{1}{2} [m_r(T)]^{m-1} m_r(t) \phi_{r,0}(T, t') \\ &= \frac{1}{2} [D_M(\pi h)]^{m-1} D_M(2\pi h \beta(t)) D_M \left(2\pi h (\beta(T) - \beta(t')) \right) u_T(t) u_T(t'), \end{aligned} \quad (4.256)$$

where (4.254) has been used. Likewise, for $m = 0$

$$\begin{aligned} \phi_{b,0}(t, t') &= \frac{1}{2} \mathbb{E} [b^*(t, \mathbf{x}_0) b(t', \mathbf{x}_0)] \\ &= \frac{1}{2} \mathbb{E} \left[e^{-j2\pi h x_0 (\beta(t) - \beta(t'))} u_T(t) u_T(t') \right] \\ &= \frac{1}{2} D_M \left(2\pi h (\beta(t) - \beta(t')) \right) u_T(t) u_T(t') \\ &= \phi_{r,0}(t, t'). \end{aligned} \quad (4.257)$$

Finally, the psd is obtained by using (4.254) and (4.257) along with (4.196) and (4.208).

4.9.7.1 Alternative Method

There is an alternate method for obtaining the full response CPM psd that provides more insight. Using (4.196) along with the property $S_{b,-m}(f) = S_{b,m}^*(f)$ yields

$$S_{\overline{ss}}(f) = \frac{A^2}{T} \left(S_{b,0}(f) + 2\text{Re} \left\{ \sum_{m=1}^{\infty} S_{b,m}(f) e^{-j2\pi f m T} \right\} \right). \quad (4.258)$$

Taking the double Fourier transform of (4.257) and (4.256) gives

$$S_{b,m}(f) = \begin{cases} S_{r,0}(f) & m = 0 \\ m_r^{m-1}(T) M_r(f) \hat{M}_r^*(f) & m > 0 \end{cases}, \quad (4.259)$$

where

$$\begin{aligned} m_r^{m-1}(T) &\triangleq [D_M(\pi h)]^{m-1} \\ M_r(f) &\triangleq \mathcal{F}[m_r(t)] = \mathcal{F}[D_M(2\pi h \beta(t) u_T(t))] \\ \hat{M}_r^*(f) &\triangleq \frac{1}{2} \text{E}[r^*(T, x_0) R(f, x_0)] = \frac{1}{2} \text{E}[e^{-j\pi h x_0} R(f, x_0)] \end{aligned}$$

$\mathcal{F}[\cdot]$ denotes the Fourier transform and

$$R(f, x_0) = \mathcal{F}[r(t, x_0)] = \mathcal{F}[e^{j2\pi h x_0 \beta(t)} u_T(t)]. \quad (4.260)$$

Then,

$$\begin{aligned} S_{\overline{ss}}(f) &= \frac{A^2}{T} \left(S_{r,0}(f) + 2\text{Re} \left\{ M_r(f) \hat{M}_r^*(f) \sum_{m=1}^{\infty} m_r^{m-1}(T) e^{-j2\pi f m T} \right\} \right) \\ &= \frac{A^2}{T} \left(S_{r,0}(f) + 2\text{Re} \left\{ M_r(f) \hat{M}_r^*(f) \sum_{n=0}^{\infty} [m_r(T) e^{-j2\pi f T}]^n e^{-j2\pi f T} \right\} \right). \end{aligned} \quad (4.261)$$

Observe that

$$|m_r(T) e^{-j2\pi f T}| = |m_r(T)| = |D_M(\pi h)| \leq 1. \quad (4.262)$$

The implication of equation (4.262) is that two separate cases must be considered when evaluating the psd.

Case 1: $|m_r(T)| < 1$

In this case the sum in (4.261) converges such that

$$S_{\overline{ss}}(f) = \frac{A^2}{T} \left(S_{r,0}(f) + 2\text{Re} \left\{ \frac{M_r(f) \hat{M}_r^*(f)}{e^{j2\pi f T} - m_r(T)} \right\} \right). \quad (4.263)$$

and the psd has no discrete components.

Case 2: $|m_r(T)| = 1$

This case is possible only if

$$|m_r(T)| = |E[e^{j\pi hx_k}]| = 1. \quad (4.264)$$

For this condition to be true,

$$e^{j\pi hx_k} = e^{jc}, \quad \forall k, \quad (4.265)$$

where c is the same constant for all k . Since this must be true for $x_k = 1$, it follows that $c = \pi h$ and

$$x_k \pi h = \pi h \pmod{2\pi} \quad \forall k. \quad (4.266)$$

This means that h must be an integer, and when h is an integer

$$m_r(T) = E[r(T, x_0)] = e^{j\pi h} \quad (4.267)$$

and

$$\hat{M}_r^*(f) = M_r^*(f)e^{j\pi h}. \quad (4.268)$$

Hence, the psd is

$$\begin{aligned} S_{\overline{sr}}(f) &= \frac{A^2}{T} \left(S_{r,0}(f) + |M_r(f)|^2 2\text{Re} \left\{ \sum_{m=1}^{\infty} e^{j2\pi(f - \frac{h}{2T})mT} \right\} \right) \\ &= \frac{A^2}{T} \left(S_{r,0}(f) - |M_r(f)|^2 + |M_r(f)|^2 \sum_{m=-\infty}^{\infty} e^{-j2\pi(f - \frac{h}{2T})mT} \right) \\ &= \frac{A^2}{T} \left(S_{r,0}(f) - |M_r(f)|^2 + \frac{1}{T} |M_r(f)|^2 \sum_{n=-\infty}^{\infty} \delta \left(f - \frac{h}{2T} - \frac{n}{T} \right) \right) \\ &= \frac{A^2}{T} \left(S_{r,0}(f) - |M_r(f)|^2 \right) + \left(\frac{A}{T} \right)^2 \sum_{n=-\infty}^{\infty} \left| M_r \left(\frac{h}{2T} + \frac{n}{T} \right) \right|^2 \delta \left(f - \frac{h}{2T} - \frac{n}{T} \right). \end{aligned} \quad (4.269)$$

Clearly, the second term in the above expression is a discrete spectral component. Hence, integer values of h lead to discrete spectral components. Since discrete spectral components are generally undesirable, integer values of h are typically not used.

4.9.7.2 Psd of CPFSK

With CPFSK, the phase shaping pulse is given by (4.107). Hence,

$$\begin{aligned} R(f, x_0) &= \int_0^T e^{j\pi \frac{hx_0 t}{T}} \cdot e^{-j2\pi f t} dt \\ &= T e^{-j\pi(fT - hx_0/2)} \text{sinc}((fT - hx_0/2)), \end{aligned} \quad (4.270)$$

where $x_0 \in \{\pm 1, \pm 3, \dots, \pm(M-1)\}$. It follows that

$$\begin{aligned} M_r(f) &= \text{E}[R(f, x_0)] \\ &= \frac{T}{M} \sum_{m=1}^M e^{-j\pi(fT - hx_m/2)} \text{sinc}((fT - hx_m/2)) \end{aligned} \quad (4.271)$$

$$\begin{aligned} S_{r,0}(f) &= \frac{1}{2} \text{E}[|R(f, x_0)|^2] \\ &= \frac{T^2}{2M} \sum_{m=1}^M \text{sinc}^2((fT - hx_m/2)) \end{aligned} \quad (4.272)$$

$$\hat{M}_r^*(f) = \frac{T}{2M} \sum_{m=1}^M e^{j\pi(fT + x_m h/2)} \text{sinc}(fT - x_m h/2). \quad (4.273)$$

These expressions are used in (4.263) to obtain the psd.

For binary $M = 2$ CPFSK,

$$S_{r,0}(f) = \frac{T^2}{4} (\text{sinc}^2(fT - h/2) + \text{sinc}^2(fT + h/2)) \quad (4.274)$$

$$M_r(f) = \frac{T}{2} (e^{-j\pi(fT + h/2)} \text{sinc}(fT + h/2) + e^{-j\pi(fT - h/2)} \text{sinc}(fT - h/2)) \quad (4.275)$$

$$\hat{M}_r^*(f) = \frac{T}{4} (e^{j\pi(fT - h/2)} \text{sinc}(fT + h/2) + e^{j\pi(fT + h/2)} \text{sinc}(fT - h/2)) \quad (4.276)$$

$$m_r(T) = D_2(h\pi). \quad (4.277)$$

When h is an integer, the psd has both continuous and discrete components

$$S_{\tilde{s}\tilde{s}}(f) = S_{\tilde{s}\tilde{s}}^c(f) + S_{\tilde{s}\tilde{s}}^d(f), \quad (4.278)$$

where

$$\begin{aligned} S_{\tilde{s}\tilde{s}}^c(f) &= \frac{A^2 T}{2} \text{sinc}(fT + h/2) \text{sinc}(fT - h/2) \\ S_{\tilde{s}\tilde{s}}^d(f) &= \frac{A^2}{T} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{h}{2T} - \frac{n}{T}\right) \left(\text{sinc}^2(n + h) + \text{sinc}^2(n) - 2\text{sinc}(n + h) \text{sinc}(n) \right) \end{aligned} \quad (4.279)$$

which clearly exhibits line components at frequencies $(\frac{h}{2T} + \frac{n}{T})$. Further simplification may be possible for special cases, but otherwise the psd has an intractable form. Figures 4.34 and 4.35 plot the psd against the normalized frequency fT . MSK corresponds to the case $h = 0.5$. Observe that the CPFSK power spectrum becomes more compact for smaller h , while the converse is true for larger h . Figure 4.35 illustrates the appearance of discrete components at frequencies $(\frac{1}{2} + n) \frac{1}{T}$, n an integer, as $h \rightarrow 1$.

4.9.7.3 Psd of MSK

The psd of CPFSK is complicated for all but a few cases. By using Laurent's decomposition [193], MSK was shown equivalent to OQASK with half-sinusoid amplitude pulse shaping. From (4.112), the MSK baseband signal has the quadrature form

$$\tilde{s}(t) = A \sum_n b(t - 2nT, \mathbf{x}_n), \quad (4.280)$$

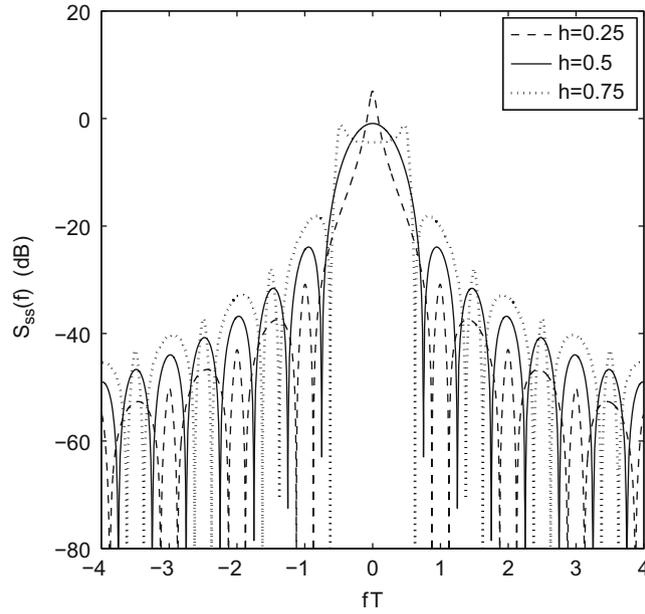


Fig. 4.34 Psd of binary CPFSK with various modulation indices. MSK corresponds to $h = 1/2$

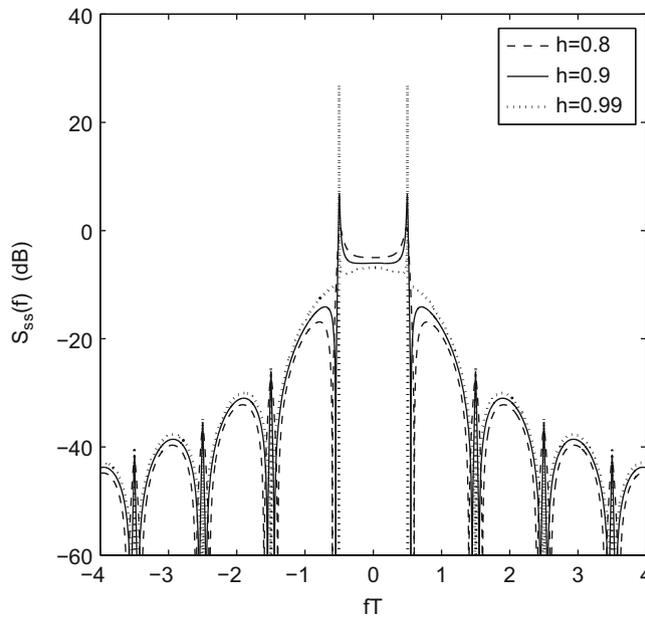


Fig. 4.35 Psd of binary CPFSK as the modulation index $h \rightarrow 1$

where

$$b(t, \mathbf{x}_n) = \hat{x}_{2n+1}h_a(t - T) + j\hat{x}_{2n}h_a(t) \tag{4.281}$$

$$h_a(t) = \sin\left(\frac{\pi t}{2T}\right)u_{2T}(t), \tag{4.282}$$

$\mathbf{x}_n = (\hat{x}_{2n+1}, \hat{x}_{2n})$ is a sequence of odd–even pairs assuming values from the set $\{\pm 1, \pm 1\}$, and T is the bit period. The Fourier transform of (4.281) is

$$B(f, \mathbf{x}_n) = (\hat{x}_{2n+1}e^{-j2\pi fT} + j\hat{x}_{2n})H_a(f). \tag{4.283}$$

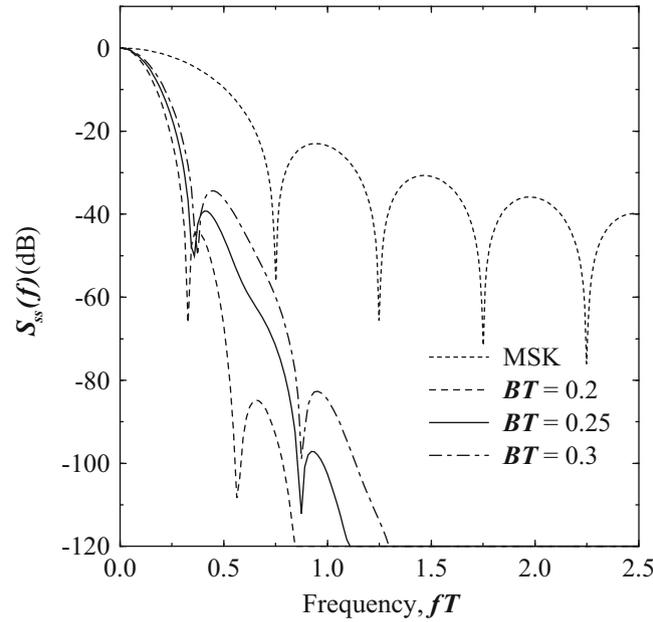


Fig. 4.36 Psd of GMSK with various normalized filter bandwidths BT

Since the data sequence is zero mean and uncorrelated, the MSK psd is

$$\begin{aligned}
 S_{b,0}(f) &= \frac{1}{2} \text{E} [|B(f, \mathbf{x}_0)|^2] \\
 &= \frac{1}{2} \text{E} [\hat{x}_1^2 + \hat{x}_0^2] |H_a(f)|^2 \\
 &= |H_a(f)|^2.
 \end{aligned} \tag{4.284}$$

The Fourier transform of the half-sinusoid pulse in (4.282) is

$$H_a(f) = \frac{2T}{\pi(1 - (4fT)^2)} (1 + e^{-j4\pi fT}). \tag{4.285}$$

Hence, the power spectrum becomes

$$S_{\overline{ss}}(f) = \frac{A^2}{T} |H_a(f)|^2 = \frac{16A^2T}{\pi^2} \left(\frac{\cos^2(2\pi fT)}{1 - (4fT)^2} \right)^2. \tag{4.286}$$

The psd of MSK is plotted in Fig. 4.34.

4.9.8 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 [135], where the modulating pulses are approximated by using a large number of rectangular sub-pulses with properly chosen amplitudes.

Figure 4.36 plots the psd of GMSK with various normalized filter bandwidths BT . Note that a smaller BT results in a more compact psd. Likewise, Fig. 4.37 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 as seen from Figs. 4.21 and 4.25.

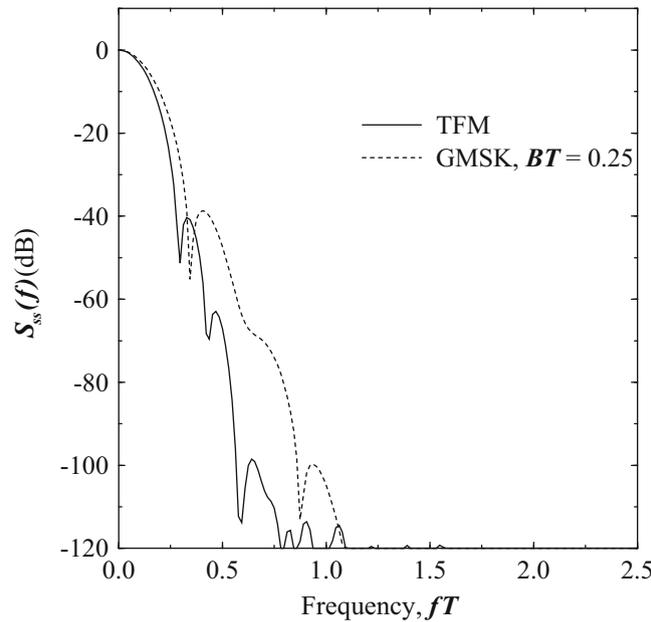


Fig. 4.37 Psd of TFM and GMSK with $BT = 0.25$

Finally, it is interesting to compare the spectral characteristics of GMSK and $\pi/4$ -DQPSK. To make a fair comparison, it must be remembered 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 obtained by dividing the elements on the horizontal axis of Fig. 4.26 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. 4.36, with $f = 1/(2T)$, almost the same side lobe roll-off is obtained. However, for larger values of f , the GMSK pulse side lobes are seen to decay faster in frequency than those of $\pi/4$ -DQPSK.

Problems

4.1. Assume that a received signal is given by

$$\tilde{y}(t) = A \sum_{n=-\infty}^{\infty} x_n p(t - nT),$$

where $x_k = \pm 1$, and $p(t)$ is the ideal Nyquist pulse

$$p(t) = \text{sinc}(t/T)$$

$$P(f) = T \text{rect}(fT).$$

Due to a slight timing error, the received signal is sampled with a timing offset t_o , resulting in the sample sequence $\{\tilde{y}_k\}$ shown in (4.46). Show that

$$\tilde{y}_k = A a_k \text{sinc}(t_o/T) + A \frac{\sin(\pi t_o/T)}{\pi} \sum_{n \neq k} \frac{a_n (-1)^n}{t_o/T - n}.$$

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.,

$$s(t) = G \left(A_n \cos(2\pi f_c t) + B_n \sin(2\pi f_c t) \right) + \left(C_n \cos(2\pi f_c t) + D_n \sin(2\pi f_c t) \right),$$

where G is a gain constant and $\{A_n\}$, $\{B_n\}$, $\{C_n\}$, 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. Consider the two 8-QAM signals constellations shown in Fig. 4.8. Suppose that the distance between nearest-neighbor signal points in each constellation is equal to A .

- For the constellation on the left, determine the cartesian coordinates of the constellation points.
- For the constellation on the right, determine the radii a and b of the inner and outer circles.
- Find the average energy per symbol for each of the two signal constellations in terms of A assuming that each signal point is used with equal probability. Which constellation is more power efficient?

4.4. Two data streams, $\{x_{n,1}\}$ and $\{x_{n,2}\}$, are to be transmitted using *unbalanced* QPSK with rectangular amplitude pulse shaping, such that the data rate for $\{x_{n,1}\}$ is 10 kbps and that for $\{x_{n,2}\}$ is 1 Mbps.

- Relate the amplitudes of the waveforms, A_1 and A_2 , such that both bit sequences have equal energies per bit.
- With A_1 and A_2 so related, find the possible phase shifts for the carrier, where the $x_{n,1}$ and $x_{n,2}$ take on all possible combinations of $+1$ and -1 .

4.5. An important parameter for digital modulation schemes is the peak-to-average power ratio (PAPR), defined by

$$\text{PAPR} = \lim_{T \rightarrow \infty} \frac{\max_{0 \leq t \leq T} |\tilde{s}(t)|^2}{T^{-1} \int_0^T |\tilde{s}(t)|^2 dt}.$$

When nonlinear power amplifiers are used it is desirable to keep the PAPR as small as possible.

- Plot the PAPR for $\pi/4$ -DQPSK with root-raised cosine pulse shaping, as a function of the roll-off factor β .
- Repeat part (a) for QPSK. What conclusions can you draw?

4.6. Two new modulation schemes have been proposed 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

(a_{2k}, a_{2k+1})	b_k
0,0	+3
0,1	+1
1,0	-3
1,1	-1

$$x_k = b_k e^{jk\frac{\pi}{2}}.$$

For B-O-QAM the mapping of data bits a_k to symbols b_k is as follows: The symbols a_k are also used to generate the symbols

a_k	b_k
0	+3
1	-3

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 nonlinear 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 .

4.7. Consider two sinusoids waveforms

$$s_1(t) = A \cos(2\pi f_c t)$$

$$s_2(t) = A \cos(2\pi(f_c + \Delta_f)t).$$

- (a) Determine the minimum value of Δ_f such that the inner product $(s_1, s_2) = 0$ over the interval $0 \leq t \leq T$. Assume that $f_c T \gg 1$.
- (b) Repeat part (a) for the two sinusoids

$$s_1(t) = A \cos(2\pi f_c t + \phi_1)$$

$$s_2(t) = A \cos(2\pi(f_c + \Delta_f)t + \phi_2),$$

where ϕ_1 and ϕ_2 are arbitrary phases.

4.8. A guard interval consisting of a cyclic prefix or cyclic suffix is used in OFDM systems to mitigate the effects of channel time dispersion.

- (a) Assess the cost of the cyclic prefix in terms of
- bandwidth and/or data rate.
 - transmitter power.
- (b) Suppose a guard interval of 500 ns is used. The data rate with 64-QAM modulation is 54 Mb/s. The power penalty due to the guard interval is to be kept less than 1 dB. What is the required value of G (constrained to an integer) and minimum possible OFDM block size (constrained to 2^k for some k)?

4.9. Consider the time-domain sample sequence for the n th OFDM block

$$X_{n,m} = \sum_{k=0}^{N-1} x_{n,k} e^{j \frac{2\pi km}{N}}.$$

The data symbols $x_{n,k}$, $k = 0, \dots, N-1$, are independent and each is chosen with equal probability from a BPSK symbol alphabet, such that $x_{n,k} \in \{-1, +1\}$. The PAPR of the sample sequence for block n can be defined as follows:

$$\text{PAPR} = \frac{\max_m |X_{n,m}|^2}{N^{-1} \sum_{m=0}^{N-1} |X_{n,m}|^2}.$$

By using the triangle inequality, show that $\text{PAPR} \leq N$.

4.10. Consider an OFDM time-domain sequence (without cyclic guard interval)

$$X_{n,m} = \sum_{k=0}^{N-1} x_{n,k} e^{j \frac{2\pi km}{N}}$$

$$= \sum_{k=0}^{N-1} x_{n,k} \cos\left(\frac{2\pi km}{N}\right) + j \sum_{k=0}^{N-1} x_{n,k} \sin\left(\frac{2\pi km}{N}\right), \quad m = 0, 1, \dots, N-1,$$

where the $x_{n,k}$ are symbols are i.i.d. symbols chosen from the binary alphabet $\{-1, +1\}$.

- (a) Invoke the central limit theorem for large N and treat the $X_{n,m} = X_{n,m}^I + jX_{n,m}^Q$ as independent complex Gaussian random variables. What are the means, variances, and cross-correlation of the quadrature components $X_{n,m}^I$ and $X_{n,m}^Q$?
- (b) Suppose that the $X_{n,m}$ can be treated as complex Gaussian random variables with the parameters in part (a). What is the probability density function of the peak power

$$P_{\max} = |X_{\max}|^2 = \max_{0 \leq m \leq N-1} |X_{n,m}|^2?$$

- (c) What is the probability density function of the peak-to-average power ratio

$$\text{PAPR} = \frac{P_{\max}}{P_{\text{av}}}$$

in terms of the block size N ?

4.11. Let $\{X_m\}_{m=0}^{N-1}$ be a finite duration time domain sequence of length N and let $\{x_k\}_{k=0}^{N-1}$ be its N -point DFT. Suppose that $\{X_m\}_{m=0}^{N-1}$ is padded with L zeroes and the $(N+L)$ -point DFT is computed, denoted by $\{\hat{x}_k\}_{k=0}^{N+L-1}$.

- (a) What is the relationship between x_0 and \hat{x}_0 ?
- (b) If $|x_k|$, $k = 0, \dots, N-1$ and $|\hat{x}_k|$, $k = 0, \dots, N+L-1$ are plotted on the same graph, explain the relationships between the two graphs.

4.12. (Computer Exercise) Consider the time-domain sample sequence for the n th OFDM block $\{X_{n,m}\}_{m=0}^{N-1}$. The peak-to-average power ratio (PAPR) for the n th data block can be defined as follows:

$$\text{PAPR}_n = \frac{\max_m |X_{n,m}|^2}{N^{-1} \sum_{m=0}^{N-1} |X_{n,m}|^2},$$

Note that the PAPR for the n th data block, PAPR_n , depends on the random data vector $\{x_{n,0}, x_{n,1}, \dots, x_{n,N-1}\}$.

By averaging over many randomly generated data vectors, determine the mean of the PAPR and the variance of the PAPR. Do this for 16-QAM modulation with block sizes $N = 256, 512, \text{ and } 1024$. Assume in all cases that no guard interval is used, i.e., $G = 0$.

4.13. (Computer Exercise) Consider a selective mapping scheme to reduce the PAPR of an OFDM waveform. The technique begins by generating L different random phase vectors of length N , i.e., first generate

$$\boldsymbol{\phi}_\ell = (\phi_{\ell,0}, \phi_{\ell,1}, \dots, \phi_{\ell,N-1}), \quad \ell = 1, 2, \dots, L,$$

where the $\phi_{\ell,i}$ are independent uniformly distributed random variables on the interval $(-\pi, \pi]$. Then for each $\boldsymbol{\phi}_\ell$, $\ell = 1, \dots, L$, compute the PAPR of the OFDM sample sequence

$$X_{n,m}^\ell = \sum_{k=0}^{N-1} x_{n,k} e^{j\phi_{\ell,k}} e^{j\frac{2\pi km}{N}}, \quad m = 0, 1, \dots, N-1,$$

and select the waveform having the smallest PAPR for transmission.

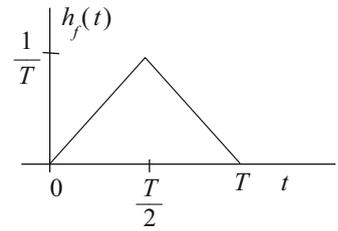
Consider $N = 256$ and 16-QAM symbols, and assume that no guard interval is used, i.e., $G = 0$. Compute the mean PAPR and the variance of the PAPR of the transmitted OFDM waveform for $L = 1, 2, 4$.

4.14. An OFDM signal with a large number of subcarriers N and no guard interval ($G = 0$) has a complex envelope that can be approximated as a zero-mean complex Gaussian random process. Assume an “ideal” OFDM signal spectrum, where the modulated power spectrum is

$$S_{ss}(f) = \begin{cases} S_0 & , |f| \leq 1/2T_s \\ 0 & , \text{elsewhere} \end{cases}$$

where $T = NT_s$.

Fig. 4.38 Frequency shaping pulse for Problem 4.16



- (a) Using the above Gaussian approximation, what is the distribution of the magnitude of the complex envelope, $|\tilde{s}(t)|$, at any time t .
- (b) Suppose that the RF power amplifier will clip the OFDM waveform if the magnitude of the complex envelope $|\tilde{s}(t)|$ exceeds the level ΘR_{rms} , where R_{rms} is the rms envelope level $\sqrt{E[|\tilde{s}(t)|^2]}$. What is the probability that the OFDM waveform will be clipped at any time t ?
- (c) Suppose that a continuous stream of OFDM symbols is transmitted. How many times per second on average will the OFDM waveform be clipped?

4.15. The following problem requires you to design a length $N = 256$ phase vector

$$\boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_{N-1})$$

such that the corresponding OFDM sample sequence

$$X_m = \sum_{k=0}^{N-1} e^{j\phi_k} e^{j\frac{2\pi km}{N}}, \quad m = 0, 1, \dots, N-1,$$

has a PAPR that is no bigger than 3 dB and preferably as small as possible. Using any and all techniques at your disposal, such as analysis and/or computer search, find such a phase vector $\boldsymbol{\phi}$.

4.16. Consider a CPM signal that is generated by using a triangular frequency shaping pulse shown in Fig. 4.38.

- (a) If $h = 1/2$, find the peak frequency deviation from the carrier, where frequency deviation is

$$f_{\text{dev}} = \frac{1}{2\pi} \frac{d\phi(t)}{dt}.$$

- (b) Sketch the phase tree and phase trellis for the binary source symbol sequence

$$\mathbf{x} = (+1, +1, +1, -1, -1, +1, -1, -1)$$

4.17. A CPM signal is generated from a baseband signal with a half-sinusoid frequency shaping function $h_f(t)$.

- (a) If $h = 1/2$, find the peak frequency deviation from the carrier frequency, where frequency deviation is

$$f_{\text{dev}} = \frac{1}{2\pi} \frac{d\phi(t)}{dt}.$$

- (b) Sketch the phase tree and phase trellis if the data symbol sequence is

$$\mathbf{x} = \{+3, -1, +1, +3, -3, +1, -1\}.$$

4.18. Sketch the phase tree, the phase trellis, and phase state diagram for partial response CPM with $h = 1/2$ and

$$h_f(t) = \frac{1}{4T} u_{2T}(t).$$

4.19. Consider a partial response CPM signal

- (a) Generate a frequency shaping function of duration $3T$ by convolving two rectangular shaping functions of duration T and $2T$.
 (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

$$\mathbf{x} = \{+1, -1, +1, -1, -1\}.$$

4.20. 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 $h = 3/4$.

4.21. Equation 4.142 defines the transfer function $H(f)$ of the Gaussian low pass filter that is used to generate the GMSK waveform.

- (a) Obtain the impulse response $h(t)$ and show that it satisfies the properties of a probability density function (pdf).
 (b) Expanding on the interpretation of $h(t)$ as a pdf, determine the variance of the distribution. What is the significance of this interpretation?

4.22. 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 $BT = 0.75$.

4.23. 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 $\text{rect}(t/T)$. Find and sketch the impulse response of the filter $\tilde{h}(t)$.

4.24. Prove the identity

$$T \sum_m e^{-j2\pi f m T} = \sum_n \delta\left(f - \frac{n}{T}\right).$$

4.25. Consider the case of uncorrelated data symbols.

- (a) Show that if the symbols are equiprobable, then

$$\begin{aligned} & \mathbb{E} \left[|B(f, x_0)|^2 \right] - \left| \mathbb{E} [B(f, x_0)] \right|^2 \\ &= \frac{1}{2M^2} \sum_{i=1}^M \sum_{k=1}^M \left| B(f, x_i) - B(f, x_k) \right|^2. \end{aligned}$$

- (b) Compute the value of part (a) for $M = 2$.

4.26. 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

$$\phi_{xx}(n) = \frac{1}{2} \mathbb{E} [x_k^* x_{k+n}] = \rho^{|n|}.$$

Compute the power density spectrum of $\tilde{s}(t)$.

4.27. Suppose that a binary data sequence $\{x_n\}$, $x_i \in \{-1, +1\}$ 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 $S_{xx}(f)$.

4.28. Suppose that an uncorrelated binary data sequence is transmitted by using binary PAM with a root-Gaussian amplitude shaping pulse

$$H_a(f) = \left(\tau e^{-\pi(f\tau)^2} \right)^{1/2}$$

- (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 $h_a(t)$?

4.29. Consider the M -ary orthogonal FSK waveform defined by (4.71) and (4.72). Assuming equally likely messages, determine the psd of the transmitted complex envelope $S_{\tilde{s}\tilde{s}}(f)$.

4.30. 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

$$\tilde{s}_i(t) = \begin{cases} A \sum_{k=0}^7 h_{ik} h_c(t - kT_c), & k = 1, \dots, 8 \\ -\tilde{s}_i(t), & k = 9, \dots, 16 \end{cases}, \quad (4.287)$$

where $T = 8T_c$ 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.

Chapter 5

Detection of Known Signals in Noise

The performance of a communication link is degraded by many transmission impairments including fading, delay spread, Doppler spread, co-channel and adjacent channel interference, noise, and receiver implementation losses. 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 variations and may necessitate a receiver with a fast convergent algorithm. Receiver implementation losses include carrier frequency offset, sample clock frequency offset, symbol timing errors, and channel estimation errors. Co-channel interference, adjacent channel interference, and noise are additive impairments that degrade the bit error rate performance by reducing the SNR or CNR.

This chapter considers the bit error rate performance of digital signaling on frequency non-selective (flat) fading channels with additive white Gaussian noise AWGN. Flat fading channels are characteristic of 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 waveform that is transmitted over the communication link. Frequency selective channels, on the other hand, distort the transmitted signal and will be the subject of Chap. 7. Flat fading channels will be shown to significantly degrade the bit error rate performance unless appropriate countermeasures are taken. Diversity and coding are two 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. This can be accomplished by using multiple receiver antennas with sufficient spatial separation, for example. Multi-antenna techniques will be discussed in Chap. 6. Coding techniques introduce a form of time diversity into the transmitted signal which, along with interleaving, can be exploited to mitigate the effects of fading. Coding techniques are the topic of Chap. 8.

The remainder of this chapter is organized as follows. Section 5.1 introduces a vector representation for digital signaling on flat fading channels with additive white Gaussian noise (AWGN). Section 5.3 provides a generalized analysis of the error rate performance of digital signaling on flat fading channels. Section 5.2 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 Sect. 5.4, QAM in Sect. 5.5, orthogonal signals in Sect. 5.6, and OFDM in Sect. 5.7. Section 5.8 considers differential detection of DPSK and $\pi/4$ -DQPSK waveforms. Section 5.9 considers non-coherent detection and, finally, Sect. 5.10 considers coherent and non-coherent detection of CPM waveforms.

5.1 Vector Space Representation of Received Signals

Consider a general digital modulation scheme having the complex envelope

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n), \quad (5.1)$$

where the generalized shaping function $b(t - nT, \mathbf{x}_n)$ depends on the particular modulation scheme being employed. Suppose the waveform $\tilde{s}(t)$ is transmitted over a flat fading channel having the time-variant channel impulse response

$$g(t, \tau) = g(t)\delta(\tau - \hat{\tau}). \quad (5.2)$$

The received complex envelope is

$$\tilde{r}(t) = g(t)\tilde{s}(t - \hat{\tau}) + \tilde{n}(t), \quad (5.3)$$

where $g(t) = \alpha(t)e^{j\phi(t)}$ is the time-variant complex fading gain introduced by the channel, and $\tilde{n}(t)$ is zero-mean complex additive white Gaussian noise (AWGN) with a power spectral density (psd) of N_o watts/Hz. Note that the fading channel introduces a multiplicative distortion, while the receiver front end introduces AWGN. The flat fading channel will also introduce a random time delay $\hat{\tau}$.

Consider a linear full response modulation scheme, such as QAM or PSK, where one of M message waveforms having a complex envelope chosen from the set $\{\tilde{s}_i(t)\}_{i=1}^M$ is transmitted over a flat fading channel every T seconds. By observing received waveform $\tilde{r}(t)$, the receiver must determine the time sequence of message waveforms that was transmitted over the channel. To do so, the receiver must determine the time delay $\hat{\tau}$, such that the location of the symbol boundaries in the received waveform are known. The process of estimating $\hat{\tau}$ is commonly called symbol or baud timing recovery. For our present purpose, it is assumed that the receiver knows $\hat{\tau}$ exactly and, therefore, it can be assumed that $\hat{\tau} = 0$. Under the above assumptions, the received complex envelope can be written as

$$\tilde{r}(t) = g(t)\tilde{s}(t) + \tilde{n}(t). \quad (5.4)$$

To derive the structure of the optimum coherent receiver, suppose that a single isolated message waveform $\tilde{s}_n(t)$ is chosen from the set $\{\tilde{s}_i(t)\}_{i=1}^M$ and transmitted over the channel. If the channel changes very slowly with respect to the symbol period T , i.e., $f_m T \ll 1$, then $g(t)$ will remain essentially constant over the duration of the amplitude shaping pulse $h_a(t)$. Under this condition, the explicit time dependency of $g(t)$ can be removed so that the received complex envelope is

$$\tilde{r}(t) = g\tilde{s}_n(t) + \tilde{n}(t), \quad (5.5)$$

where $g = \alpha e^{j\phi}$ is the random fading gain.

Chapter 4 showed that the set of waveforms $\{\tilde{s}_i(t)\}_{i=1}^M$ can be represented as a set vectors $\{\tilde{\mathbf{s}}_i\}_{i=1}^M$ in an N -dimensional signal space. The signal vectors and the associated basis functions, $\{\varphi_i(t)\}_{i=1}^N$ may be obtained by using a Gram–Schmidt orthonormalization procedure. To derive the structure of the optimal coherent receiver, it is useful to obtain a vector representation of the received waveform in (5.5). This can be accomplished by projecting the received complex envelope $\tilde{r}(t)$ onto the set of basis functions $\{\varphi_i(t)\}_{i=1}^N$ giving the representation

$$\tilde{r}(t) = \sum_{i=1}^N \tilde{r}_i \varphi_i(t) + \tilde{z}(t), \quad (5.6)$$

where

$$\tilde{r}_i = \int_{-\infty}^{\infty} \tilde{r}(t) \varphi_i^*(t) dt \quad (5.7)$$

$$= g \int_{-\infty}^{\infty} \tilde{s}_n(t) \varphi_i^*(t) dt + \int_{-\infty}^{\infty} \tilde{n}(t) \varphi_i^*(t) dt \quad (5.8)$$

$$= g\tilde{s}_{n_i} + \tilde{n}_i \quad (5.9)$$

and

$$\tilde{z}(t) = \tilde{n}(t) - \sum_{i=1}^N \tilde{n}_i \varphi_i(t) \quad (5.10)$$

is a “remainder” process, which is the portion of the noise process $\tilde{n}(t)$ that lies outside of the signal space. The above process yields the received vector

$$\tilde{\mathbf{r}} = g\tilde{\mathbf{s}}_n + \tilde{\mathbf{n}}, \quad (5.11)$$

where

$$\begin{aligned}\tilde{\mathbf{r}} &= (\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_N) \\ \tilde{\mathbf{s}}_n &= (\tilde{s}_{n1}, \tilde{s}_{n2}, \dots, \tilde{s}_{nN}) \\ \tilde{\mathbf{n}} &= (\tilde{n}_1, \tilde{n}_2, \dots, \tilde{n}_N).\end{aligned}$$

For an AWGN channel, the $\tilde{n}_k, k = 1, \dots, N$ are complex Gaussian random variables that can be completely described by their means and covariances. The means are

$$\mathbb{E}[\tilde{n}_k] = \int_{-\infty}^{\infty} \mathbb{E}[\tilde{n}(t)]\varphi_k(t)dt = 0 \quad (5.12)$$

and covariances are¹

$$\begin{aligned}\phi_{\tilde{n}_j\tilde{n}_k} &= \frac{1}{2}\mathbb{E}[\tilde{n}_j\tilde{n}_k^*] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2}\mathbb{E}[\tilde{n}(t)\tilde{n}^*(s)]\varphi_j(t)\varphi_k^*(s)dt ds \\ &= N_o \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(t-s)\varphi_j(t)\varphi_k^*(s)dt ds \\ &= N_o \int_{-\infty}^{\infty} \varphi_j(t)\varphi_k^*(t)dt \\ &= N_o\delta_{jk}.\end{aligned}$$

It follows that the \tilde{n}_k are all independent identically distributed zero-mean complex Gaussian random variables with variance N_o . Hence, the vector $\tilde{\mathbf{n}}$ has the multivariate complex Gaussian probability density function (pdf) (A.51)

$$\begin{aligned}p(\tilde{\mathbf{n}}) &= \prod_{i=1}^N \frac{1}{2\pi N_o} \exp\left\{-\frac{1}{2N_o}|\tilde{n}_i|^2\right\} \\ &= \frac{1}{(2\pi N_o)^N} \exp\left\{-\frac{1}{2N_o}\|\tilde{\mathbf{n}}\|^2\right\}.\end{aligned} \quad (5.13)$$

The joint pdf $p(\tilde{\mathbf{n}})$ is said to be circularly symmetric, because it appears as a hyperspherical cloud that is centered at the origin in the N -D vector space.

The waveform $\tilde{z}(t)$ is a remainder process due to the fact that $\tilde{z}(t)$ lies outside the vector space that is spanned by the basis functions $\{\varphi_n(t)\}_{n=1}^N$. It is shown below that the remainder process is uncorrelated with received vector $\tilde{\mathbf{r}}$, viz.,

$$\begin{aligned}\frac{1}{2}\mathbb{E}[\tilde{z}(t)r_j^*] &= \frac{1}{2}\mathbb{E}[\tilde{z}(t)]g\tilde{s}_{mj}^* + \mathbb{E}[\tilde{z}(t)\tilde{n}_j^*] \\ &= \frac{1}{2}\mathbb{E}[\tilde{z}(t)\tilde{n}_j^*] \\ &= \frac{1}{2}\mathbb{E}\left[\left(\tilde{n}(t) - \sum_{n=1}^N \tilde{n}_n\varphi_n(t)\right)\tilde{n}_j^*\right] \\ &= \int_{-\infty}^{\infty} \frac{1}{2}\mathbb{E}[\tilde{n}(t)\tilde{n}^*(\tau)]\varphi_j(\tau)d\tau - \sum_{n=1}^N \frac{1}{2}\mathbb{E}[\tilde{n}_n\tilde{n}_j^*]\varphi_n(t) \\ &= N_o\varphi_j(t) - N_o\varphi_j(t) = 0.\end{aligned}$$

¹Since the \tilde{n}_k have zero mean, their covariances $\lambda_{\tilde{n}_j\tilde{n}_k}$ are equal to their autocorrelations $\phi_{\tilde{n}_j\tilde{n}_k}$. The factor of 1/2 in the definition of the covariance function maintains the conventional usage of N_o as representing the power density spectrum of the low-pass noise process $\tilde{n}(t)$.

Since $\frac{1}{2}E[\tilde{z}(t)\tilde{r}_j^*] = 0$, $j = 1, \dots, N$, 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 [357]. In other words, the received vector $\tilde{\mathbf{r}}$ provides "sufficient statistics" for determining which message waveform was transmitted, meaning that $\tilde{\mathbf{r}}$ provides all the necessary information for detection and no other information is required.

5.2 Detection of Known Signals in AWGN

Based on the observation of the noisy received vector $\tilde{\mathbf{r}}$ in (5.11), the receiver should determine which message vector was transmitted such that the probability of decision error is minimized. It is assumed that the receiver has perfect knowledge of the channel gain g . With this in mind, consider the set of a posteriori probabilities

$$P[\mathbf{s}_i \text{ was sent} | g, \tilde{\mathbf{r}}], \quad i = 1, \dots, M, \quad (5.14)$$

which is abbreviated as $P[\mathbf{s}_i | g, \tilde{\mathbf{r}}]$. The maximum a posteriori probability (MAP) receiver decides in favor of the message vector $\tilde{\mathbf{s}}_m$ having the maximum a posteriori probability $P[\tilde{\mathbf{s}}_m | g, \tilde{\mathbf{r}}]$. That is, the MAP decision rule is

$$\text{choose } \tilde{\mathbf{s}}_m \text{ if } P[\tilde{\mathbf{s}}_m | g, \tilde{\mathbf{r}}] \geq P[\tilde{\mathbf{s}}_i | g, \tilde{\mathbf{r}}] \quad \forall i \neq m. \quad (5.15)$$

The probability of error in this decision, denoted by $P_e[\tilde{\mathbf{s}}_m | g, \tilde{\mathbf{r}}]$, is

$$\begin{aligned} P_e[\tilde{\mathbf{s}}_m | g, \tilde{\mathbf{r}}] &= P[\mathbf{s}_m \text{ was not sent} | g, \tilde{\mathbf{r}}] \\ &= 1 - P[\mathbf{s}_m \text{ was sent} | g, \tilde{\mathbf{r}}] \\ &= 1 - P[\mathbf{s}_m | g, \tilde{\mathbf{r}}]. \end{aligned} \quad (5.16)$$

Since the MAP receiver always decides in favor of the message vector \mathbf{s}_m having the maximum a posteriori probability $P[\mathbf{s}_m | g, \tilde{\mathbf{r}}]$ for any received vector $\tilde{\mathbf{r}}$, the probability of error is minimized.

By using Bayes' theorem, the a posteriori probability $P[\tilde{\mathbf{s}}_m | g, \tilde{\mathbf{r}}]$ can be expressed in the form

$$P[\tilde{\mathbf{s}}_m | g, \tilde{\mathbf{r}}] = \frac{p(\tilde{\mathbf{r}} | g, \tilde{\mathbf{s}}_m)P_m}{p(\tilde{\mathbf{r}})}, \quad m = 1, \dots, M, \quad (5.17)$$

where $p(\tilde{\mathbf{r}} | g, \tilde{\mathbf{s}}_m)$ is the joint conditional pdf of the received vector $\tilde{\mathbf{r}}$ given the transmitted message vector $\tilde{\mathbf{s}}_m$ and channel gain g , 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 equivalent to

$$\text{choose } \tilde{\mathbf{s}}_m \text{ if } p(\tilde{\mathbf{r}} | g, \tilde{\mathbf{s}}_m)P_m \geq p(\tilde{\mathbf{r}} | g, \tilde{\mathbf{s}}_i)P_i \quad \forall i \neq m. \quad (5.18)$$

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 message probabilities is called a maximum likelihood (ML) receiver. The ML decision rule is

$$\text{choose } \tilde{\mathbf{s}}_m \text{ if } p(\tilde{\mathbf{r}} | g, \tilde{\mathbf{s}}_m) \geq p(\tilde{\mathbf{r}} | g, \tilde{\mathbf{s}}_i) \quad \forall i \neq m. \quad (5.19)$$

If the prior message probabilities are all equal, i.e., $P_m = 1/M$, $m = 1, \dots, M$, then the signal vector that maximizes $p(\tilde{\mathbf{r}} | g, \tilde{\mathbf{s}}_m)$ also maximizes $p(\tilde{\mathbf{s}}_m | g, \tilde{\mathbf{r}})$. Under this condition the ML receiver minimizes the probability of decision error. In practice, an ML receiver is sometimes implemented regardless of the prior message probabilities, because they may be unknown. Also, the prior message probabilities will be all equal for a well-designed system.

To proceed further, the joint conditional pdf $p(\tilde{\mathbf{r}} | g, \tilde{\mathbf{s}}_m)$ is needed. Since $\tilde{\mathbf{r}} = g\tilde{\mathbf{s}}_m + \tilde{\mathbf{n}}$ and $\tilde{\mathbf{n}}$ has the joint pdf in (5.13), it follows that

$$p(\tilde{\mathbf{r}} | g, \tilde{\mathbf{s}}_m) = \frac{1}{(2\pi N_o)^N} \exp \left\{ -\frac{1}{2N_o} \|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_m\|^2 \right\}. \quad (5.20)$$

By using (5.20) in (5.19), it is apparent that the signal vector $\tilde{\mathbf{s}}_m$ that maximizes $p(\tilde{\mathbf{r}}|g, \tilde{\mathbf{s}}_m)$ also minimizes the exponent in (5.20). Hence, the ML receiver decides in favor of that message $\tilde{\mathbf{s}}_m$ which minimizes the decision metric

$$\mu_1(\tilde{\mathbf{s}}_m) = \|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_m\|^2, \quad m = 1, \dots, M. \quad (5.21)$$

From (5.21), the ML receiver decides in favor of the scaled message vector $g\tilde{\mathbf{s}}_m$ that is closest in squared Euclidean distance (or 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 first expanding (5.21) as

$$\mu_1(\tilde{\mathbf{s}}_m) = \|\tilde{\mathbf{r}}\|^2 - 2\text{Re}\{\tilde{\mathbf{r}} \cdot g^* \tilde{\mathbf{s}}_m^*\} + |g|^2 \|\tilde{\mathbf{s}}_m\|^2. \quad (5.22)$$

Then notice that $\|\tilde{\mathbf{r}}\|^2$ is independent of the choice of $\tilde{\mathbf{s}}_m$, and $\|\tilde{\mathbf{s}}_m\|^2 = 2E_m$, where E_m is the energy in the bandpass waveform corresponding to the signal vector $\tilde{\mathbf{s}}_m$. Hence, the ML receiver decides in favor of that message $\tilde{\mathbf{s}}_m$ which maximizes the decision metric

$$\mu_2(\tilde{\mathbf{s}}_m) = \text{Re}\{\tilde{\mathbf{r}} \cdot g^* \tilde{\mathbf{s}}_m^*\} - |g|^2 E_m, \quad m = 1, \dots, M. \quad (5.23)$$

Using the definition of the inner product, the above decision metric can be rewritten in the alternate form

$$\begin{aligned} \mu_2(\tilde{\mathbf{s}}_m) &= \text{Re} \left\{ \int_{-\infty}^{\infty} \tilde{r}(t) g^* \tilde{s}_m^*(t) dt \right\} - |g|^2 E_m \\ &\equiv \text{Re} \left\{ \int_{-\infty}^{\infty} \tilde{r}(t) e^{-j\phi} \tilde{s}_m^*(t) dt \right\} - \alpha E_m, \quad m = 1, \dots, M. \end{aligned} \quad (5.24)$$

The last line in (5.24) follows because the $\mu_2(\tilde{\mathbf{s}}_m)$ can be divided by α without altering the decision process. In this form of the ML decision metric, the received complex envelope $\tilde{r}(t)$ is correlated directly with the scaled and conjugated signal vector $e^{-j\phi} \tilde{s}_m^*(t)$.

From the above development, the ML receiver can now be constructed. 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 low-pass filter in each branch is used to reject the double frequency term after demodulation. The received bandpass waveform is

$$r(t) = \text{Re}\{\tilde{r}(t)e^{j2\pi f_c t}\} = \tilde{r}_I(t) \cos(2\pi f_c t) - \tilde{r}_Q(t) \sin(2\pi f_c t). \quad (5.25)$$

It follows that

$$[r(t) \cdot 2 \cos(2\pi f_c t)]_{\text{LP}} = \tilde{r}_I(t) \quad (5.26)$$

$$[-r(t) \cdot 2 \sin(2\pi f_c t)]_{\text{LP}} = \tilde{r}_Q(t), \quad (5.27)$$

where $[\cdot]_{\text{LP}}$ indicates low-pass filtering. 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 each of the N basis functions used to define the signal space. This receiver structure is called a correlation detector.

Fig. 5.1 Quadrature demodulator

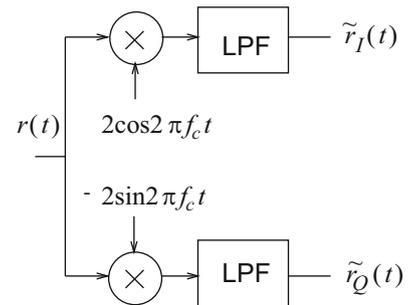


Fig. 5.2 Correlator detector

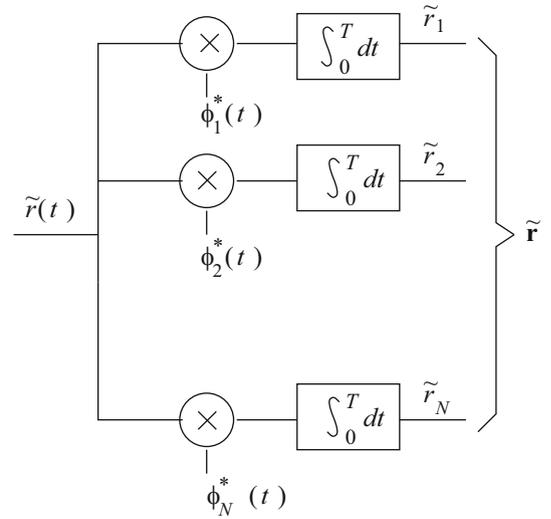
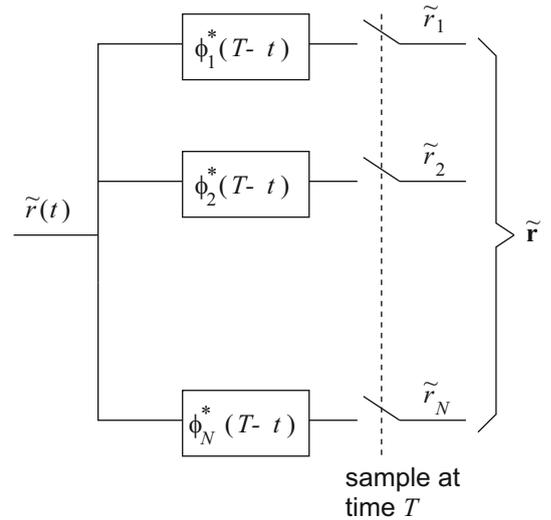


Fig. 5.3 Matched filter detector



A functionally equivalent structure to the correlation detector is shown in Fig. 5.3, where the complex envelope is filtered with a bank of N filters having impulse responses $\phi_i^*(T_o - t)$ and sampling the outputs at time T_o , where T_o is the duration of the $\phi_i(t)$ (or $\tilde{s}_i(t)$). The filter $\phi_i^*(T_o - t)$ is the *matched filter* to $\phi_i(t)$ and, therefore, 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 (see Problem 5.2). Finally, the metric computer in Fig. 5.4 processes the observation vector $\tilde{\mathbf{r}}$ to produce M decision metrics $\mu_2(\tilde{\mathbf{s}}_m)$, $m = 1, \dots, M$. The decision is made in favor of the message $\tilde{\mathbf{s}}_m$ having the largest decision metric.

To show equivalence of the correlation and matched filter detectors in Figs. 5.2 and 5.3, respectively, let $h_i(t) = \phi_i^*(T_o - t)$ denote the filter that is matched to $\phi_i(t)$. Then the output of the matched filter is the convolution

$$\begin{aligned} y(t) &= \int_0^t \tilde{r}(\tau) h_i(t - \tau) d\tau \\ &= \int_0^t \tilde{r}(\tau) \phi_i^*(T_o - t + \tau) d\tau. \end{aligned} \tag{5.28}$$

Sampling the filter output at time T_o gives

$$y(T_o) = \int_0^{T_o} \tilde{r}(\tau) \phi_i^*(\tau) d\tau. \tag{5.29}$$

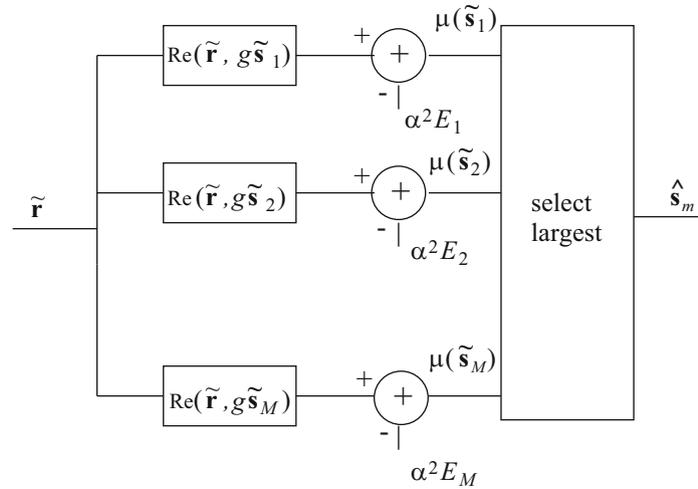


Fig. 5.4 Metric computer

This is exactly the same as the correlation in (5.7). Note that other variations of the ML receiver can be constructed in a similar fashion by direct implementation of (5.24). This will require either a bank of M correlators or a bank of M matched filters, where M is the number of waveforms in the signal set. Since $N \leq M$, the number of correlators or matched filters is usually larger with this latter implementation. However, the outputs of the correlators or matched filters generate the required decision metrics directly, and a subsequent metric computer is not required.

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 (5.23) can be neglected, and the ML receiver decides in favor of that message $\tilde{\mathbf{s}}_m$ which maximizes the decision metric

$$\mu_2(\tilde{\mathbf{s}}_m) = \text{Re} \{ \tilde{\mathbf{r}} \cdot e^{-j\phi} \tilde{\mathbf{s}}_m^* \} \quad (5.30)$$

$$= \text{Re} \left\{ \int_{-\infty}^{\infty} \tilde{r}(t) e^{-j\phi} \tilde{s}_m^*(t) dt \right\}, \quad m = 1, \dots, M. \quad (5.31)$$

In this case, the receiver does not need to know the complete complex channel gain $g = \alpha e^{j\phi}$, but only the phase ϕ .

5.3 Probability of Error

Consider a signal constellation having the set M signal vectors $\{\tilde{\mathbf{s}}_m\}_{m=1}^M$. Assume that the messages are equally likely so that $P_m = 1/M$. By observing the vector $\tilde{\mathbf{r}}$, the ML receiver chooses that 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, first define convex decision regions R_m around each of the scaled signal vectors $g\tilde{\mathbf{s}}_m$ in the N -D signal space. Figure 5.5 shows an example of the decision regions. Formally, the decision regions are defined by

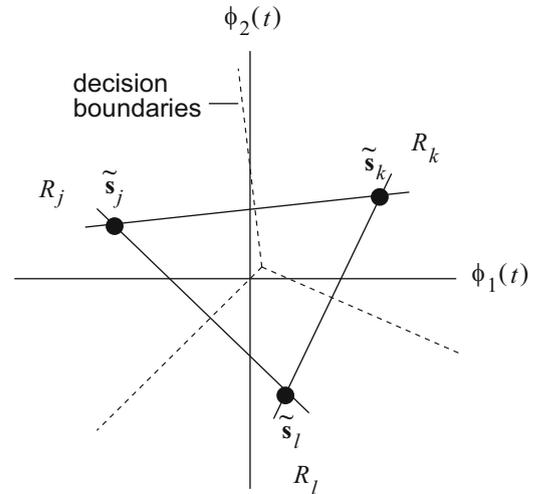
$$R_m = \{ \tilde{\mathbf{r}} : \|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_m\|^2 \leq \|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_i\|^2, \quad \forall i \neq m \}. \quad (5.32)$$

Observe that every $\tilde{\mathbf{r}} \in R_m$ is closer to $g\tilde{\mathbf{s}}_m$ than to any other scaled signal vector $g\tilde{\mathbf{s}}_i, i \neq m$. The ML decision rule becomes

$$\text{choose } \tilde{\mathbf{s}}_m \text{ if } \tilde{\mathbf{r}} \in R_m. \quad (5.33)$$

The decision boundaries are hyperplanes in the N -dimensional signal space that are defined by the locus of signal points that are equidistant from two neighboring scaled signal vectors.

Fig. 5.5 Decision regions in a 2-D signal space



The conditional error probability associated with $\tilde{\mathbf{s}}_m$ is

$$\begin{aligned} P[e|\tilde{\mathbf{s}}_m] &= P[\tilde{\mathbf{r}} \notin R_m] \\ &= 1 - P[\tilde{\mathbf{r}} \in R_m] \\ &= 1 - P[c|\tilde{\mathbf{s}}_m], \end{aligned} \quad (5.34)$$

where $P[c|\tilde{\mathbf{s}}_m]$ is the conditional probability of correct reception. By using the joint conditional pdf in (5.20), it follows that

$$P[e|\tilde{\mathbf{s}}_m] = 1 - \int_{R_m} p(\tilde{\mathbf{r}}|g, \tilde{\mathbf{s}}_m) d\tilde{\mathbf{r}}. \quad (5.35)$$

Finally, the average probability of decision error is

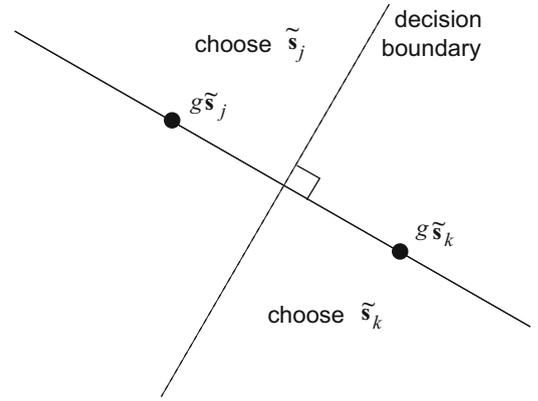
$$P[e] = \frac{1}{M} \sum_{m=1}^M P[e|\tilde{\mathbf{s}}_m]. \quad (5.36)$$

It is often difficult if not impossible to compute the exact probability of decision error, due to the difficulty in defining the decision regions R_m and performing the N -fold integration in (5.35) with the proper limits of integration. In this case, various upper and lower bounds, and approximations on the probability of error are useful. First, the concept of pairwise error probability is introduced.

5.3.1 Pairwise Error Probability

Consider two equally likely signal vectors $\tilde{\mathbf{s}}_j$ and $\tilde{\mathbf{s}}_k$ in a signal constellation of size M , as if these two signal vectors are the only ones that exist. The resulting probability of decision error at the receiver is called the pairwise error probability because it can be defined for each distinct 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 $\|g\tilde{\mathbf{s}}_j - g\tilde{\mathbf{s}}_k\|^2 = \alpha^2 \|\tilde{\mathbf{s}}_j - \tilde{\mathbf{s}}_k\|^2$. 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. This 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 $\tilde{\mathbf{r}} = g\tilde{\mathbf{s}}_j + \tilde{\mathbf{n}}$ to cross the decision boundary. 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 signal space. Hence, the noise component along the line that passes through the two signal vectors will have zero mean and variance N_o . It follows that the error probability is equal to

Fig. 5.6 Two received signal points in an N -D signal space



$$P[e|\tilde{s}_j] = Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{jk}^2}{4N_o}}\right), \quad (5.37)$$

where $\tilde{d}_{jk}^2 = \|\tilde{s}_j - \tilde{s}_k\|^2$ is the squared Euclidean distance between \tilde{s}_j and \tilde{s}_k . Finally, $P[e|\tilde{s}_j] = P[e|\tilde{s}_k]$. Hence, the pairwise error probability between the message vectors \tilde{s}_j and \tilde{s}_k , denoted by $P[\tilde{s}_j, \tilde{s}_k]$, is

$$P[\tilde{s}_j, \tilde{s}_k] = Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{jk}^2}{4N_o}}\right). \quad (5.38)$$

5.3.2 Upper Bounds on Error Probability

Suppose that \tilde{s}_k is transmitted and let \mathcal{E}_j denote the event that the receiver chooses \tilde{s}_j instead, thereby making a symbol error. The probability of the event \mathcal{E}_j is the pairwise error probability $P[\tilde{s}_j, \tilde{s}_k]$. The probability of decision error is the probability of the union of all error events

$$P(e|\tilde{s}_k) = P\left[\bigcup_{j \neq k} \mathcal{E}_j\right]. \quad (5.39)$$

Quite often the error events will overlap and this greatly complicates the calculation of the error probability. However, an upper bound on the error probability can be obtained by employing the union bound

$$P\left[\bigcup_{j \neq k} \mathcal{E}_j\right] \leq \sum_{j \neq k} P[\mathcal{E}_j]. \quad (5.40)$$

This gives the upper bound

$$P[e|\tilde{s}_k] \leq \sum_{j \neq k} P[\tilde{s}_j, \tilde{s}_k]. \quad (5.41)$$

Combining the above result with (5.38) gives

$$P[e|\tilde{s}_k] \leq \sum_{j \neq k} Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{jk}^2}{4N_o}}\right) \quad (5.42)$$

and using (5.36) to average error probability over all messages gives

$$P[e] \leq \frac{1}{M} \sum_{k=1}^M \sum_{j \neq k} Q \left(\sqrt{\frac{\alpha^2 \tilde{d}_{jk}^2}{4N_o}} \right). \quad (5.43)$$

Calculation of the union bound in (5.43) requires the set of squared Euclidean distances $\{\tilde{d}_{jk}^2\}$ between the signal vectors. A simpler upper bound on error probability can be obtained by finding the minimum squared Euclidean distance between any two signal vectors

$$\tilde{d}_{\min}^2 = \min_{n,m} \|\tilde{\mathbf{s}}_n - \tilde{\mathbf{s}}_m\|^2. \quad (5.44)$$

Then the pairwise error probability between $\tilde{\mathbf{s}}_j$ and $\tilde{\mathbf{s}}_k$ is bounded by

$$P[\tilde{\mathbf{s}}_j, \tilde{\mathbf{s}}_k] \leq Q \left(\sqrt{\frac{\alpha^2 \tilde{d}_{\min}^2}{4N_o}} \right), \quad (5.45)$$

since $\tilde{d}_{\min}^2 \leq \tilde{d}_{jk}^2$ and the function $Q(x)$ monotonically decreases with x . Hence,

$$P[e] \leq (M-1)Q \left(\sqrt{\frac{\alpha^2 \tilde{d}_{\min}^2}{4N_o}} \right). \quad (5.46)$$

Finally, some further upper bounds can be obtained by upper bounding the Gaussian Q-function. One such upper bound is (Problem 5.1)

$$Q(x) \leq \frac{1}{2} e^{-x^2/2} \quad x \geq 0. \quad (5.47)$$

Combining with the union bound in (5.43) gives

$$P[e] \leq \frac{1}{2M} \sum_{k=1}^M \sum_{j \neq k} \exp \left\{ -\frac{\alpha^2 \tilde{d}_{jk}^2}{8N_o} \right\}, \quad (5.48)$$

and combining with the upper bound in (5.46) will give the simplest but loosest upper bound of all

$$P[e] \leq \frac{(M-1)}{2} \exp \left\{ -\frac{\alpha^2 \tilde{d}_{\min}^2}{8N_o} \right\}. \quad (5.49)$$

5.3.3 Lower Bound on Error Probability

A useful lower bound on the probability of decision error can be obtained by bounding the error probability

$$P[e|\tilde{\mathbf{s}}_k] \geq \begin{cases} Q \left(\sqrt{\frac{\alpha^2 \tilde{d}_{\min}^2}{4N_o}} \right), & \text{if } \tilde{\mathbf{s}}_k \text{ has at least one neighbor at distance } \tilde{d}_{\min} \\ 0, & \text{otherwise} \end{cases} \quad (5.50)$$

Then

$$P[e] = \frac{1}{M} \sum_{m=1}^M P[e|\tilde{s}_m] \quad (5.51)$$

$$\geq \frac{w_{\min}}{M} Q \left(\sqrt{\frac{\alpha^2 \tilde{d}_{\min}^2}{4N_o}} \right), \quad (5.52)$$

where w_{\min} is the number of signal vectors having at least one minimum distance neighbor. Certainly $w_{\min} \geq 2$, so that

$$P[e] \geq \frac{2}{M} Q \left(\sqrt{\frac{\alpha^2 \tilde{d}_{\min}^2}{4N_o}} \right). \quad (5.53)$$

5.3.4 Bit Versus Symbol Error Probabilities

So far, the probability of decision error $P[e]$ otherwise known as the *symbol* error probability, P_M , has been considered. However, the *bit* error probability, P_b , is often of interest. In general, the bit error probability will depend on the particular mapping between the data bits and the modulated symbols. Since each data symbol corresponds to $\log_2 M$ data bits, the bit error probability can be bounded as follows:

$$\frac{P_M}{\log_2 M} \leq P_b \leq P_M. \quad (5.54)$$

The lower bound results from the fact that each symbol error corresponds to at least one bit error, while the upper bound results from the fact that each symbol error corresponds to at most $\log_2 M$ bit errors.

5.3.4.1 Gray Mapping

For signal constellations such as PSK and QAM, it is possible to map the binary data bits onto the data 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 mapping. When the signal-to-noise ratio is high, symbol errors tend to be made onto the nearest neighboring symbols with high probability. In these cases, symbol errors correspond to single bit errors. Hence,

$$P_b \approx \frac{P_M}{\log_2 M}. \quad (5.55)$$

It turns out that Gray mapping is the optimum mapping for uncoded systems. However, if error control coding is used, Gray mapping is usually not the optimum mapping strategy and other types of mapping are used. This issue will be discussed in more detail in Chap. 8.

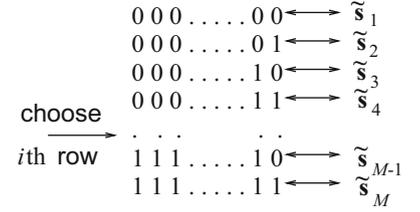
5.3.4.2 Equally Likely Symbol Errors

Suppose that when symbol errors occur, each of the $M-1$ incorrect symbols is chosen with equal probability. To compute the probability of bit error, first note that the set of $M = 2^k$ symbols has a one-to-one mapping onto the set of 2^k binary k -tuples as shown in Fig. 5.7. Now suppose that the all zeroes k -tuple, or first row, corresponds to the correct symbol. Moreover, the receiver makes an error by choosing i th row (symbol), $i \neq 1$, instead. Since there are 2^{k-1} zeros and 2^{k-1} ones in each column, and a zero corresponds to a correct bit, the probability of a particular bit position being in error is

$$P_b = \frac{2^{k-1}}{2^k - 1} P_M = \frac{M}{2(M-1)} P_M. \quad (5.56)$$

It will shown later that this result applies to M -ary orthogonal signals.

Fig. 5.7 Mapping of binary k -tuples onto M -ary symbols



5.3.5 Rotation and Translations

The probability of symbol error in (5.36) is invariant to any rotation of the signal constellation $\{\tilde{\mathbf{s}}_i\}_{i=1}^M$ about the origin of the signal space. This is a consequence of two properties. First, the probability of symbol error depends solely on the set of Euclidean distances $\{\tilde{d}_{jk}\}, j \neq k$ between the signal vectors in the signal constellation. Second, the AWGN is circularly symmetric in all directions of the signal space. A signal constellation can be rotated about the origin of the signal space, by multiplying each N -dimensional signal vector by an $N \times N$ unitary matrix \mathbf{Q} . A unitary matrix has the property $\mathbf{Q}\mathbf{Q}^H = \mathbf{Q}^H\mathbf{Q} = \mathbf{I}$, where \mathbf{Q}^H is the complex conjugate transpose of \mathbf{Q} , and \mathbf{I} is the $N \times N$ identity matrix. The rotated signal vectors are equal to

$$\hat{\mathbf{s}}_i = \tilde{\mathbf{s}}_i\mathbf{Q}, \quad i = 1, \dots, M. \quad (5.57)$$

Correspondingly, the noise vector $\tilde{\mathbf{n}}$ is replaced with its rotated version

$$\hat{\mathbf{n}} = \tilde{\mathbf{n}}\mathbf{Q}. \quad (5.58)$$

The rotated noise vector $\hat{\mathbf{n}}$ is a vector of complex Gaussian random variables that is completely described by its mean and covariance matrix. The mean is

$$\mathbb{E}[\hat{\mathbf{n}}] = \mathbb{E}[\tilde{\mathbf{n}}]\mathbf{Q} = \mathbf{0}. \quad (5.59)$$

The covariance matrix is²

$$\begin{aligned} \Phi_{\hat{\mathbf{n}}\hat{\mathbf{n}}} &= \frac{1}{2}\mathbb{E}[\hat{\mathbf{n}}^H\hat{\mathbf{n}}] \\ &= \frac{1}{2}\mathbb{E}[(\tilde{\mathbf{n}}\mathbf{Q})^H\tilde{\mathbf{n}}\mathbf{Q}] \\ &= \frac{1}{2}\mathbb{E}[\mathbf{Q}^H\tilde{\mathbf{n}}^H\tilde{\mathbf{n}}\mathbf{Q}] \\ &= \mathbf{Q}^H\frac{1}{2}\mathbb{E}[\tilde{\mathbf{n}}^H\tilde{\mathbf{n}}]\mathbf{Q} \\ &= N_o\mathbf{Q}^H\mathbf{Q} = N_o\mathbf{I}. \end{aligned} \quad (5.60)$$

Since, the statistical properties of the noise vector are invariant to rotation, the probability of symbol error is invariant to rotation of the signal constellation about the origin of the signal space.

Next consider a translation of the signal set such that

$$\hat{\mathbf{s}}_i = \tilde{\mathbf{s}}_i - \mathbf{a}, \quad i = 1, \dots, M, \quad (5.61)$$

where \mathbf{a} is a constant vector. In this case, the error probability remains the same since $\hat{d}_{jk} = \tilde{d}_{jk}, j \neq k$. However, the average energy in the signal constellation is altered by the translation and becomes

²Since the vector $\hat{\mathbf{n}}$ has zero mean, its covariance matrix $\Lambda_{\hat{\mathbf{n}}\hat{\mathbf{n}}}$ is equal to its autocorrelation matrix $\Phi_{\hat{\mathbf{n}}\hat{\mathbf{n}}}$.

$$\begin{aligned}
\hat{E}_{\text{av}} &= \frac{1}{2} \sum_{i=1}^M \|\hat{\mathbf{s}}_i\|^2 P_i \\
&= \frac{1}{2} \sum_{i=1}^M \|\tilde{\mathbf{s}}_i - \mathbf{a}\|^2 P_i \\
&= \frac{1}{2} \sum_{i=1}^M \{ \|\tilde{\mathbf{s}}_i\|^2 - 2\text{Re} \{ \tilde{\mathbf{s}}_i \cdot \mathbf{a}^* \} + \|\mathbf{a}\|^2 \} P_i \\
&= \frac{1}{2} \sum_{i=1}^M \|\tilde{\mathbf{s}}_i\|^2 P_i - \text{Re} \left\{ \sum_{i=1}^M \tilde{\mathbf{s}}_i P_i \cdot \mathbf{a}^* \right\} + \frac{1}{2} \|\mathbf{a}\|^2 \sum_{i=1}^M P_i \\
&= E_{\text{av}} - \text{Re} \{ \mathbf{E}[\tilde{\mathbf{s}}] \cdot \mathbf{a}^* \} + \frac{1}{2} \|\mathbf{a}\|^2
\end{aligned} \tag{5.62}$$

where E_{av} is the average energy of the original signal constellation and $\mathbf{E}[\tilde{\mathbf{s}}] = \sum_{i=1}^M \tilde{\mathbf{s}}_i P_i$ is its centroid (or center of mass).

Differentiating (5.62) with respect to \mathbf{a} and setting the result equal to zero will yield the translation that minimizes the average energy in the signal constellation. This gives

$$\mathbf{a}_{\text{opt}} = \mathbf{E}[\tilde{\mathbf{s}}]. \tag{5.63}$$

Note that the center of mass of the translated signal constellation is at the origin, and the minimum average energy in the translated signal constellation is

$$\hat{E}_{\text{min}} = E_{\text{av}} - \frac{1}{2} \|\mathbf{a}_{\text{opt}}\|^2. \tag{5.64}$$

5.4 Error Probability of PSK

This section considers the error probability of various forms of PSK signals. The treatment starts with binary PSK signals, followed by the more complicated forms of PSK signals.

5.4.1 Error Probability of BPSK

The BPSK signal vectors are³

$$\tilde{\mathbf{s}}_1 = -\tilde{\mathbf{s}}_2 = \sqrt{2E_h}. \tag{5.65}$$

Since there are only two signal vectors, the bit error probability is given by the pairwise error probability in (5.38). For BPSK signals, $\tilde{d}_{12} = 2\sqrt{2E_h}$. 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

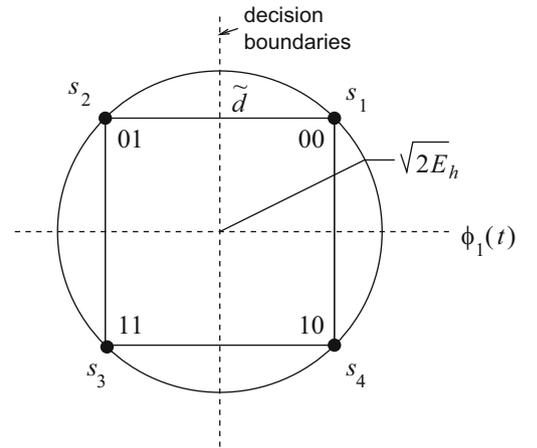
$$P_b(\gamma_b) = Q\left(\sqrt{2\gamma_b}\right) \tag{5.66}$$

where γ_b is defined as the received *bit energy-to-noise ratio*

$$\gamma_b \triangleq \frac{\alpha^2 E_b}{N_o}. \tag{5.67}$$

³When the signal vectors lie in a 1-D complex vector space, the notation is simplified by using the scalars \tilde{s}_i , \tilde{n} , \tilde{r} rather than the vectors $\tilde{\mathbf{s}}_i$, $\tilde{\mathbf{n}}$, and $\tilde{\mathbf{r}}$.

Fig. 5.8 Complex signal-space diagram for QPSK



5.4.2 Error Probability of QPSK and OQPSK

The QPSK (or 4-PSK) signal vectors are

$$\tilde{s}_1 = -\tilde{s}_3 = \sqrt{2E_h} \quad (5.68)$$

$$\tilde{s}_2 = -\tilde{s}_4 = j\sqrt{2E_h}, \quad (5.69)$$

where the signal points are seen to lie on the real and imaginary axes. The QPSK signal constellation can be rotated by 45° as shown in Fig. 5.8 without changing the error probability due to the rotational invariance property. In this case, the decision boundaries correspond to the real and imaginary axes of the complex signal space. The noise vector is $\tilde{n} = \tilde{n}_I + j\tilde{n}_Q$, where \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

$$\begin{aligned} P_M &= P[e|\tilde{s}_1] \\ &= 1 - P[c|\tilde{s}_1] \\ &= 1 - P[\tilde{n}_I > -\alpha\tilde{d}/2, \tilde{n}_Q > -\alpha\tilde{d}/2] \\ &= 1 - P[\tilde{n}_I > -\alpha\tilde{d}/2] P[\tilde{n}_Q > -\alpha\tilde{d}/2] \\ &= 1 - \left(1 - Q\left(\sqrt{\frac{\alpha^2\tilde{d}^2}{4N_o}}\right)\right)^2 \end{aligned}$$

where, again, α is the channel attenuation. Since $\tilde{d}^2 = 4E_h$,

$$P_M = 1 - (1 - Q(\sqrt{\gamma_s}))^2 \quad (5.70)$$

where γ_s is defined as the received *symbol energy-to-noise ratio*

$$\gamma_s \triangleq \frac{\alpha^2 E_h}{N_o}. \quad (5.71)$$

Suppose the data bits are mapped onto the data symbols with the Gray code shown in Fig. 5.8. Letting P_b denote the probability of bit error, it follows that

$$P[c] = (1 - P_b)^2 \quad (5.72)$$

and

$$P_M = 1 - (1 - P_b)^2. \quad (5.73)$$

Comparing (5.73) with (5.70), observe that

$$P_b = Q(\sqrt{\gamma_s}). \quad (5.74)$$

QPSK transmits 2 bits/symbol so the symbol energy is $E_h = 2E_b$, where E_b is the bit energy. Since $\gamma_s = 2\gamma_b$, the probability of bit error is

$$P_b(\gamma_b) = Q(\sqrt{2\gamma_b}). \quad (5.75)$$

Notice that the bit error rate performances of QPSK and BPSK are identical. Finally, since OQPSK is identical to QPSK with the exception that the in-phase and quadrature branches are offset by $T_b = T/2$ seconds, the bit error rate performance of OQPSK is identical to that of QPSK and BPSK as well.

5.4.3 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. Suppose that the message vector $\tilde{s}_1 = \sqrt{2E_h}$ is transmitted. The received signal vector is

$$\tilde{r} = \alpha e^{j\phi} \tilde{s}_1 + \tilde{n}. \quad (5.76)$$

Since the error probability is invariant to the angle rotation ϕ , it is possible to arbitrarily set $\phi = 0$ so that

$$\begin{aligned} \tilde{r} &= \alpha \tilde{s}_1 + \tilde{n} \\ &= \alpha \sqrt{2E_h} + \tilde{n}. \end{aligned} \quad (5.77)$$

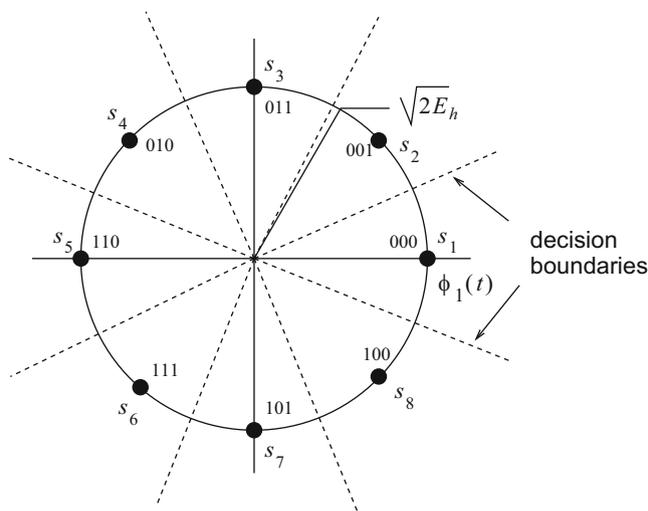


Fig. 5.9 Complex signal-space diagram for 8-PSK along with the associated decision regions

It follows that $\tilde{r} = \tilde{r}_I + j\tilde{r}_Q$ is a complex Gaussian random variable with pdf

$$p_{\tilde{r}}(\tilde{r}) = \frac{1}{\pi N_o} \exp \left\{ -\frac{1}{N_o} \left| \tilde{r} - \alpha \sqrt{2E_h} \right|^2 \right\}. \quad (5.78)$$

Since \tilde{s}_1 was transmitted, the probability of correct symbol reception with minimum distance decisions is the probability that the received vector \tilde{r} falls in the “pie-shaped” region containing \tilde{s}_1 . This is equivalent to the received angle $\Theta = \text{Tan}^{-1}[\tilde{r}_Q/\tilde{r}_I]$ falling in the interval $[-\pi/8, \pi/8]$.

To find the pdf of the angle Θ , first define the random variables

$$R = \sqrt{\tilde{r}_I^2 + \tilde{r}_Q^2}, \quad \Theta = \text{Tan}^{-1}[\tilde{r}_Q/\tilde{r}_I] \quad (5.79)$$

such that

$$\tilde{r}_I = R \cos \Theta, \quad \tilde{r}_Q = R \sin \Theta. \quad (5.80)$$

Then by using a bivariate transformation of random variables as shown in Appendix A, the joint pdf of R and Θ can be obtained as

$$p_{R,\Theta}(r, \theta) = \frac{r}{\pi N_o} e^{-\frac{1}{N_o}(r^2 - 2\alpha\sqrt{2E_h}r\cos\theta + 2\alpha^2E_h^2)}, \quad r \geq 0, -\pi \leq \theta \leq \pi. \quad (5.81)$$

Since only the phase Θ is of interest, the marginal pdf of Θ is obtained as

$$p_{\Theta}(\theta) = \int_0^{\infty} p_{R,\Theta}(r, \theta) dr \quad (5.82)$$

$$= \frac{1}{2\pi} e^{-\gamma_s \sin^2 \theta} \int_0^{\infty} x e^{(x - \sqrt{2\gamma_s} \cos \theta)^2 / 2} dx, \quad (5.83)$$

where $\gamma_s = \alpha^2 E_h / N_o$ is the received symbol energy-to-noise ratio. The probability of symbol error, P_M , is just the probability that Θ falls outside the region $[-\pi/M, \pi/M]$. Thus

$$P_M(\gamma_s) = 1 - \int_{-\pi/M}^{\pi/M} p(\theta) d\theta. \quad (5.84)$$

Unfortunately, a closed form expression for this integral does not exist, except for the cases $M = 2, 4$ which were considered earlier.

5.4.4 Error Probability with Rayleigh Fading

When the channel experiences fading, the error probability must be averaged over the fading distribution. For example, if the channel is Rayleigh faded, then α is a Rayleigh random variable and the squared-envelope α^2 is an exponential random variable at any given time, as discussed in Sect. 2.1.3.1. It follows that the received bit and symbol-energy-to-noise ratios γ_b and γ_s in (5.67) and (5.71), respectively, have the exponential pdfs

$$p_{\gamma_b}(x) = \frac{1}{\bar{\gamma}_b} e^{-x/\bar{\gamma}_b}, \quad x \geq 0 \quad (5.85)$$

and

$$p_{\gamma_s}(x) = \frac{1}{\bar{\gamma}_s} e^{-x/\bar{\gamma}_s}, \quad x \geq 0, \quad (5.86)$$

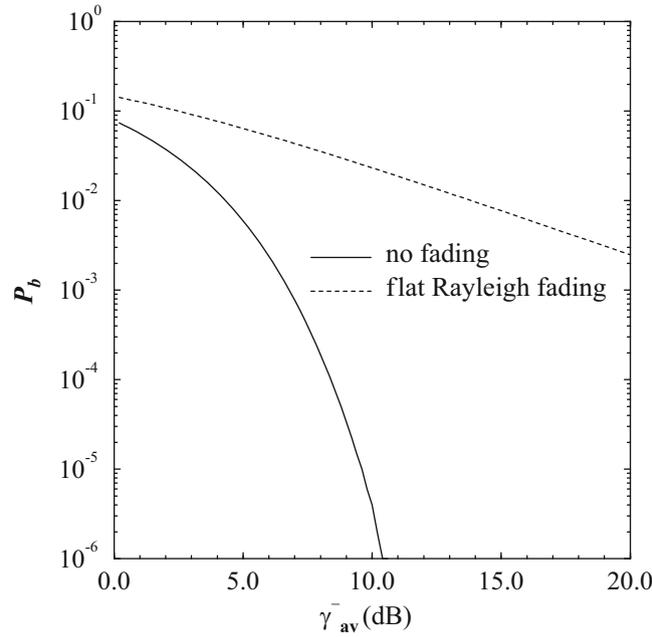


Fig. 5.10 Bit error probability for BPSK and QPSK for a slow flat Rayleigh fading channel with AWGN

where $\bar{\gamma}_b$ and $\bar{\gamma}_s$ are the average received bit and symbol energy-to-noise ratios, respectively. Since there are $\log_2 M$ bits per modulated symbol, it also follows that $\gamma_s = \gamma_b \log_2 M$ and $\bar{\gamma}_s = \bar{\gamma}_b \log_2 M$. Corresponding expressions for the distribution of γ_b and γ_s can be obtained in a similar fashion for other types of fading, such as Ricean and Nakagami fading.

For BPSK and QPSK, the probability of bit error averaged over the distribution of the received bit energy-to-noise ratio in (5.85) is

$$\begin{aligned}
 P_b &= \int_0^{\infty} Q(\sqrt{2x}) p_{\gamma_b}(x) dx \\
 &= \frac{1}{2} \left(1 - \sqrt{\frac{\bar{\gamma}_b}{1 + \bar{\gamma}_b}} \right) \\
 &\approx \frac{1}{4\bar{\gamma}_b} \quad \text{for } \bar{\gamma}_b \gg 1.
 \end{aligned} \tag{5.87}$$

The BPSK and QPSK bit error probability is plotted in Fig. 5.10 for an AWGN channel and a Rayleigh fading channel. Observe that Rayleigh fading converts an exponential dependency of the bit error probability on the average received bit energy-to-noise ratio into an inverse linear one. This behavior with flat Rayleigh fading will be observed for all types of modulation, and it results in a huge loss in performance unless appropriate countermeasures such as diversity and coding are used. For M -PSK, the average symbol error probability is

$$P_M = \int_0^{\infty} P_M(x) p_{\gamma_s}(x) dx \tag{5.88}$$

where $P_M(x)$ is given by (5.84) and γ_s is given by (5.86). Although no closed form expression exists, numerical results will show that the bit error probability depends inversely on the average received bit energy-to-noise ratio $\bar{\gamma}_b$. Recall that with Gray coding the bit error probability is approximately $P_b \approx P_M / \log_2 M$.

5.4.5 Differential PSK

The received carrier phase for PSK signals is

$$\theta_k = \frac{2\pi}{M}x_k + \phi, \quad (5.89)$$

where 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 (5.31). However, in practice this operation is not quite that simple, because the symmetries in the signal constellation create *phase ambiguity*. In particular, any channel induced phase of the form $\phi + 2k\pi/M$, where 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

$$d_k = x_k \oplus x_{k-1}, \quad (5.90)$$

where \oplus denotes modulo- M addition. Then the sequence $\{d_k\}$ is transmitted in the absolute carrier phase according to

$$\theta_k = \frac{2\pi}{M}d_k. \quad (5.91)$$

After carrier recovery, the received carrier phase is

$$\tilde{\theta}_k = \frac{2\pi}{M}d_k + \frac{2\pi\ell}{M}, \quad (5.92)$$

where the additional term $2\pi\ell/M$, ℓ an integer, represents the phase ambiguity. The receiver computes the differential phase

$$\begin{aligned} \tilde{\theta}_k - \tilde{\theta}_{k-1} \text{ modulo } 2\pi &= \frac{2\pi}{M}(d_k - d_{k-1}) \text{ modulo } 2\pi \\ &= \frac{2\pi}{M}(d_k \ominus d_{k-1}) \\ &= \frac{2\pi}{M}x_k \end{aligned} \quad (5.93)$$

where \ominus denotes modulo- M subtraction. Hence, the data sequence $\{x_k\}$ can be 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 $\hat{\theta}_k \neq \tilde{\theta}_k$. Note that an incorrect phase estimate $\hat{\theta}_k$ causes the decisions for 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 Differential PSK (DPSK) is roughly two times that of PSK.

5.5 Error Probability of PAM and QAM

5.5.1 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 of the signal constellation, the probability of symbol error is

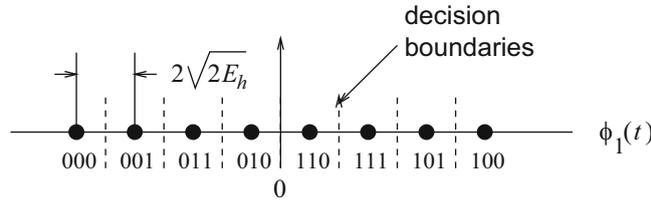


Fig. 5.11 Complex signal-space diagram for 8-PAM

$$P_i = 2Q\left(\frac{2\alpha^2 E_h}{N_o}\right). \quad (5.94)$$

Likewise, for the 2 outer points of the signal constellation, the probability of symbol error is

$$P_o = Q\left(\sqrt{\frac{2\alpha^2 E_h}{N_o}}\right). \quad (5.95)$$

Assuming all points in the signal constellation are used with equal probability, the overall probability of symbol error is

$$\begin{aligned} P_M &= \frac{M-2}{M}P_i + \frac{2}{M}P_o \\ &= 2\left(1 - \frac{1}{M}\right)Q\left(\sqrt{\frac{2\alpha^2 E_h}{N_o}}\right). \end{aligned} \quad (5.96)$$

To proceed further, E_h must be related to the average symbol energy. Since

$$\tilde{s}_m = \sqrt{2E_h}(2m-1-M), \quad m = 1, \dots, M \quad (5.97)$$

the energy in \tilde{s}_m is

$$E_m = \frac{1}{2}\tilde{s}_m^2 = E_h(2m-1-M)^2. \quad (5.98)$$

The average energy is

$$\begin{aligned} E_{av} &= E_h \frac{1}{M} \sum_{m=1}^M (2m-1-M)^2 \\ &= E_h \frac{1}{M} \left(4 \sum_{m=1}^M m^2 - 4(M+1) \sum_{m=1}^M m + M(M+1)^2 \right). \end{aligned} \quad (5.99)$$

Using the identities

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6} \quad (5.100)$$

and simplifying gives the result

$$E_{av} = E_h \frac{M^2 - 1}{3}. \quad (5.101)$$

Hence, from (5.96)

$$P_M(\gamma_s) = 2 \left(1 - \frac{1}{M}\right) Q \left(\sqrt{\frac{6}{M^2 - 1}} \gamma_s \right), \quad (5.102)$$

where

$$\gamma_s = \frac{\alpha^2 E_{av}}{N_o} \quad (5.103)$$

is the average symbol energy-to-noise ratio. Note that in this case, the “average” is over the points in the signal constellation. Since $\gamma_s = (\log_2 M) \gamma_b$,

$$P_M(\gamma_b) = 2 \left(1 - \frac{1}{M}\right) Q \left(\sqrt{\frac{6(\log_2 M)}{M^2 - 1}} \gamma_b \right). \quad (5.104)$$

5.5.2 Error Probability of M -QAM

Consider an M -QAM system having a square constellation of 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 (5.102), the symbol error probability for each \sqrt{M} -PAM system is

$$P_{\sqrt{M}} = 2 \left(1 - \frac{1}{\sqrt{M}}\right) Q \left(\sqrt{\frac{6}{M-1}} \frac{\gamma_s}{2} \right), \quad (5.105)$$

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

$$P[c] = (1 - P_{\sqrt{M}})^2 \quad (5.106)$$

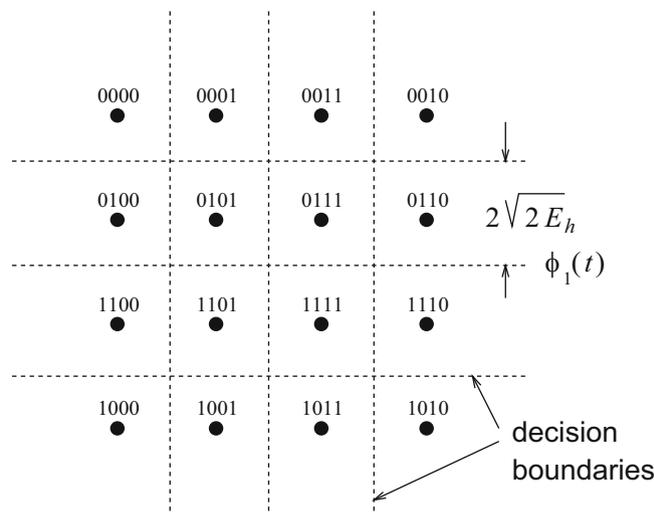


Fig. 5.12 Complex signal-space diagram for 16-QAM constellation

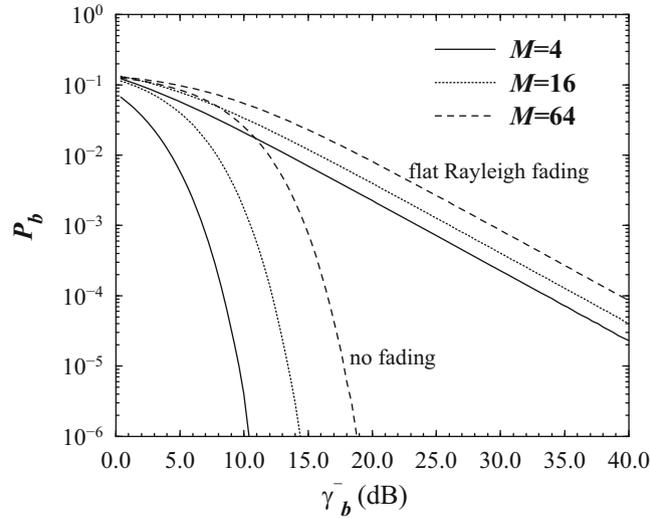


Fig. 5.13 Bit error probability for M -QAM on an AWGN channel and a Rayleigh fading channel with AWGN

and the probability of symbol error is

$$P_M(\gamma_s) = 1 - (1 - P_{\sqrt{M}})^2. \quad (5.107)$$

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 Sect. 5.3.

5.5.2.1 Error Probability with Rayleigh Fading

If the channel is Rayleigh faded, then γ_s has the exponential pdf in (5.86). It follows that the average symbol error probability is

$$P_M = \int_0^\infty P_M(x) p_{\gamma_s}(x) dx. \quad (5.108)$$

Figure 5.13 plots the (approximate) bit error probability $P_b \approx P_M / \log_2 M$ against the average received *bit* energy-to-noise ratio, $\bar{\gamma}_b = \bar{\gamma}_s / \log_2 M$, for several values of M . Once again, Rayleigh fading converts an exponential dependency of the bit error probability on the average received bit energy-to-noise ratio into an inverse linear one. Finally, 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 per modulated symbol.

5.6 Error Probability of Orthogonal Signals

5.6.1 Orthogonal Signals

Consider the M -ary orthogonal signal set

$$\tilde{\mathbf{s}}_i = \sqrt{2E_h} \mathbf{e}_m, \quad m = 1, \dots, M,$$

where \mathbf{e}_m is a length- M unit basis vector with a “1” in the m th coordinate. If the signal $\tilde{\mathbf{s}}_1$ is transmitted, then the received vector is

$$\tilde{\mathbf{r}} = (g\sqrt{2E_h} + \tilde{n}_1, \tilde{n}_2, \dots, \tilde{n}_M), \quad (5.109)$$

where the \tilde{n}_i are independent zero-mean complex Gaussian random variables with variance N_o . The ML receiver computes the M decision variables

$$\mu(\tilde{\mathbf{s}}_m) = \text{Re} \{ \tilde{\mathbf{r}} \cdot \mathbf{g}^* \tilde{\mathbf{s}}_m^* \}, \quad m = 1, \dots, M, \quad (5.110)$$

and decides in favor of the signal having the largest $\mu(\tilde{\mathbf{s}}_m)$. It follows that

$$\begin{aligned} \mu(\tilde{\mathbf{s}}_1) &= 2\alpha^2 E_h + \tilde{n}_{1,1} \alpha \sqrt{2E_h} \\ \mu(\tilde{\mathbf{s}}_m) &= \tilde{n}_{1,m} \alpha \sqrt{2E_h}, \quad m = 2, \dots, M, \end{aligned} \quad (5.111)$$

where the phase rotation on the noise samples has been ignored due to their circular symmetry. The $\mu(\tilde{\mathbf{s}}_i)$, $i = 1, \dots, M$, are independent Gaussian random variables with variance $2\alpha^2 E_h N_o$; the mean of $\mu(\tilde{\mathbf{s}}_1)$ is $2\alpha^2 E_h$ while the $\mu(\tilde{\mathbf{s}}_m)$, $m \neq 1$, have zero mean. The probability of correct symbol decision conditioned on $\mu(\tilde{\mathbf{s}}_1) = x$ is the probability that all the $\mu(\tilde{\mathbf{s}}_m)$, $m \neq 1$ are less than x . This is just

$$P[c | \mu(\tilde{\mathbf{s}}_1) = x] = \left(\Phi \left(\frac{x}{\sqrt{2\alpha^2 E_h N_o}} \right) \right)^{M-1}. \quad (5.112)$$

Hence,

$$P[c] = \int_{-\infty}^{\infty} \left(\Phi \left(\frac{x}{\sqrt{2\alpha^2 E_h N_o}} \right) \right)^{M-1} \frac{1}{\sqrt{4\pi\alpha^2 E_h N_o}} \exp \left\{ -\frac{(x - 2\alpha^2 E_h)^2}{4\alpha^2 E_h N_o} \right\} dx. \quad (5.113)$$

Now let $y = (x - 2\alpha^2 E_h) / \sqrt{2\alpha^2 E_h N_o}$. Then

$$P[c] = \int_{-\infty}^{\infty} \left(\Phi(y + \sqrt{2\gamma_s}) \right)^{M-1} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \quad (5.114)$$

where

$$\gamma_s = \frac{\alpha^2 E_h}{N_o}. \quad (5.115)$$

Finally, the probability of symbol error is

$$P_M = 1 - P[c]. \quad (5.116)$$

An alternate expression for the error probability can be derived by first conditioning on the event that one of the $M - 1$ decision variables $\mu(\tilde{\mathbf{s}}_m)$, $m \neq 1$ is the largest. This gives

$$P_M = (M - 1) \int_{-\infty}^{\infty} \Phi \left(\frac{x - 2\alpha^2 E_h}{\sqrt{2\alpha^2 E_h N_o}} \right) \left(\Phi \left(\frac{x}{\sqrt{2\alpha^2 E_h N_o}} \right) \right)^{M-2} \frac{1}{\sqrt{4\pi\alpha^2 E_h N_o}} \exp \left\{ -\frac{x^2}{4\alpha^2 E_h N_o} \right\} dx. \quad (5.117)$$

Now let $y = x / \sqrt{2\alpha^2 E_h N_o}$. Then

$$P_M = (M - 1) \int_{-\infty}^{\infty} \Phi(y - \sqrt{2\gamma_s}) (\Phi(y))^{M-2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \quad (5.118)$$

For orthogonal signals $\gamma_s = \gamma_b \log_2 M$ and the bit error probability is given by (5.56). Hence,

$$P_b = \frac{M}{2} \int_{-\infty}^{\infty} \Phi\left(y - \sqrt{2\gamma_b \log_2 M}\right) (\Phi(y))^{M-2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \quad (5.119)$$

If the channel is Rayleigh faded, then γ_b has the exponential pdf in (5.85), and the average bit error probability can be calculated as

$$P_b = \int_0^{\infty} P_b(x) p_{\gamma_b}(x) dx. \quad (5.120)$$

5.6.2 Biorthogonal Signals

Consider the biorthogonal signal set

$$\tilde{\mathbf{s}}_i = \begin{cases} \sqrt{2E_h} \mathbf{e}_i, & i = 1, \dots, M/2 \\ -\tilde{\mathbf{s}}_{i-M/2}, & i = M/2 + 1, \dots, M-1 \end{cases}. \quad (5.121)$$

Now suppose that \mathbf{s}_1 is transmitted. The receiver computes the $M/2$ decision variables

$$\mu(\tilde{\mathbf{s}}_m) = \text{Re} \{ \tilde{\mathbf{r}} \cdot \mathbf{g}^* \tilde{\mathbf{s}}_m^* \}, \quad m = 1, \dots, M/2, \quad (5.122)$$

and chooses the one having the largest magnitude. The sign of $\mu(\tilde{\mathbf{s}}_m)$ is used to decide whether $\tilde{\mathbf{s}}_m$ or $\tilde{\mathbf{s}}_{M/2+m} = -\tilde{\mathbf{s}}_m$ was sent. As before, the $\mu(\tilde{\mathbf{s}}_i)$, $i = 1, \dots, M/2$, are independent Gaussian random variables with variance $2\alpha^2 E_h N_o$; the mean of $\mu(\tilde{\mathbf{s}}_1)$ is $2\alpha^2 E_h$ while the $\mu(\tilde{\mathbf{s}}_m)$, $m = 2, \dots, M/2$, have zero mean. The probability of correct decision is the probability that $\mu(\tilde{\mathbf{s}}_1) > 0$ and $|\mu(\tilde{\mathbf{s}}_m)| < \mu(\tilde{\mathbf{s}}_1)$, $m = 2, \dots, M/2$. Conditioned on $\mu(\tilde{\mathbf{s}}_1) = x$, $x > 0$,

$$P[|\mu(\tilde{\mathbf{s}}_m)| < x] = \Phi\left(\frac{x}{\sqrt{2\alpha^2 E_h N_o}}\right) - \Phi\left(-\frac{x}{\sqrt{2\alpha^2 E_h N_o}}\right). \quad (5.123)$$

Hence,

$$P[c] = \int_0^{\infty} \left(\Phi\left(\frac{x}{\sqrt{2\alpha^2 E_h N_o}}\right) - \Phi\left(-\frac{x}{\sqrt{2\alpha^2 E_h N_o}}\right) \right)^{M/2-1} \frac{1}{\sqrt{4\pi\alpha^2 E_h N_o}} \exp\left\{-\frac{(x-2\alpha^2 E_h)^2}{4\alpha^2 E_h N_o}\right\} dx. \quad (5.124)$$

Now let $y = (x - 2\alpha^2 E_h) / \sqrt{2\alpha^2 E_h N_o}$. Then

$$P[c] = \int_{-\sqrt{2\gamma_s}}^{\infty} \left(\Phi(y + \sqrt{2\gamma_s}) - \Phi(-y - \sqrt{2\gamma_s}) \right)^{M/2-1} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \quad (5.125)$$

Finally, $P_M = 1 - P[c]$. For biorthogonal signals $\gamma_s = \gamma_b \log_2 M$. However, the bit error probability is not given by (5.56), because when symbol errors occur the incorrect symbols do not occur with equal probability.

5.7 Error Probability of OFDM

The OFDM baseband demodulator is usually implemented by using a fast Fourier transform (FFT), as discussed in Sect. 4.6. Following the development in Sect. 4.6, suppose that the discrete-time sequence $\mathbf{X}_n^g = \{X_{n,m}^g\}_{m=0}^{N+G-1}$ is passed through a balanced pair of digital-to-analog converters (DACs), as shown in Fig. 4.14, and the resulting complex envelope is transmitted over a quasi-static flat fading channel with complex gain g . The quasi-static assumption means that the channel remains static over an OFDM symbol, but can vary from one OFDM symbol to the next. For flat fading channels, the cyclic guard interval

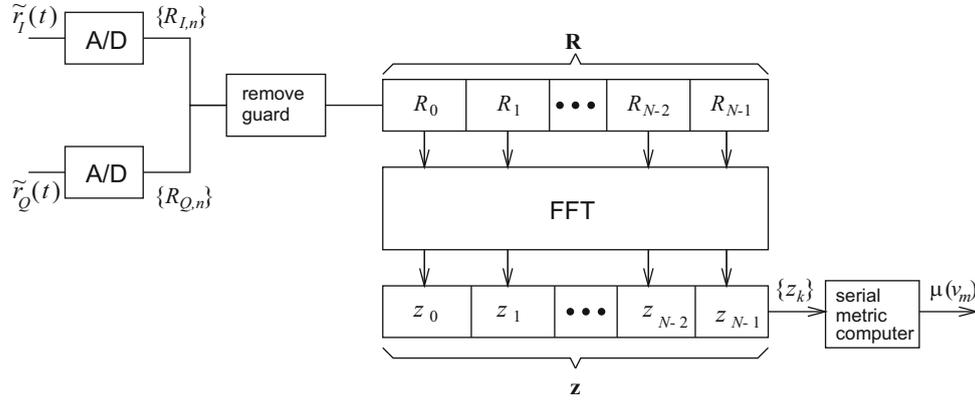


Fig. 5.14 Block diagram of OFDM receiver

is not really necessary, but is included here for simplicity. The receiver uses a quadrature demodulator to extract the received complex envelope $\tilde{r}(t) = \tilde{r}_I(t) + j\tilde{r}_Q(t)$. Suppose that the quadrature components $\tilde{r}_I(t)$ and $\tilde{r}_Q(t)$ are each passed through an ideal anti-aliasing filter (ideal low-pass filter) having a cutoff frequency $1/(2T_s^g)$ followed by an analog-to-digital converter (ADC) as shown in Fig. 5.14. This produces the received complex-valued sample sequence $\mathbf{R}_n^g = \{R_{n,m}^g\}_{m=0}^{N+G-1}$, where

$$R_{n,m}^g = gX_{n,m}^g + \tilde{n}_{n,m}, \quad (5.126)$$

$g = \alpha e^{j\phi}$ is the complex channel gain, and the $\tilde{n}_{n,m}$ are the complex-valued Gaussian noise samples. For an ideal anti-aliasing filter having a cutoff frequency $1/(2T_s^g)$, the $\tilde{n}_{n,m}$ are independent zero-mean complex Gaussian random variables with variance $\sigma^2 = \frac{1}{2}E[|\tilde{n}_{n,m}|^2] = N_o/T_s^g$, where $T_s^g = NT_s/(N+G)$.

Assuming a cyclic suffix as discussed in Sect. 4.6.1, the receiver first removes the guard interval according to

$$R_{n,m} = R_{n,G+(m-G)_N}^g, \quad 0 \leq m \leq N-1, \quad (5.127)$$

where $(m)_N$ is the residue of m modulo N . Demodulation is then performed by computing the FFT on the block $\mathbf{R}_n = \{R_{n,m}\}_{m=0}^{N-1}$ to yield the vector $\mathbf{z}_n = \{z_{n,k}\}_{k=0}^{N-1}$ of N decision variables

$$\begin{aligned} z_{n,k} &= \frac{1}{N} \sum_{m=0}^{N-1} R_{n,m} e^{-\frac{j2\pi km}{N}} \\ &= gAx_{n,k} + v_{n,k}, \quad k = 0, \dots, N-1, \end{aligned} \quad (5.128)$$

where $A = \sqrt{2E_h/T}$, $T = (N+G)T_s^g$, and the noise terms are given by

$$v_{n,k} = \frac{1}{N} \sum_{m=0}^{N-1} \tilde{n}_{n,m} e^{-\frac{j2\pi km}{N}}, \quad k = 0, \dots, N-1. \quad (5.129)$$

It can be shown that the $v_{n,k}$ are zero-mean complex Gaussian random variables with covariance

$$\phi_{j,k} = \frac{1}{2}E[v_{n,j}v_{n,k}^*] = \frac{N_o}{NT_s^g} \delta_{jk}. \quad (5.130)$$

Hence, the $z_{n,k}$ are independent Gaussian random variables with mean $g\sqrt{2E_h/T}x_{n,k}$ and variance N_o/NT_s^g . To be consistent with our earlier results for PSK and QAM signals, the $z_{n,k}$ can be multiplied for convenience by the scalar $\sqrt{NT_s^g}$. Such scaling gives

$$\tilde{z}_{n,k} = g\sqrt{2E_hN/(N+G)}x_{n,k} + \tilde{v}_{n,k}, \quad (5.131)$$

where the $\tilde{v}_{n,k}$ are i.i.d. zero-mean Gaussian random variables with variance N_o . Notice that $\sqrt{2E_h N/(N+G)}x_{n,k} = \tilde{s}_{n,k}$ is equal to the complex signal vector that is transmitted on the i th sub-carrier, where the term $N/(N+G)$ represents the loss in effective symbol energy due to the insertion of the cyclic guard interval. For *each* of the $\tilde{z}_{n,k}$, the receiver decides in favor of the *signal vector* $\tilde{s}_{n,k}$ that minimizes the squared Euclidean distance

$$\mu(\tilde{s}_{n,k}) = \|\tilde{z}_{n,k} - g\tilde{s}_{n,k}\|^2, \quad k = 0, \dots, N-1. \quad (5.132)$$

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. It is apparent from (5.131) 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 mutually orthogonal in time.

5.7.1 Interchannel Interference

The above analysis assumes that the complex channel gain g remains constant over the OFDM symbol duration $T = NT_s = (N+G)T_s^g$. However, as the block size N increases and/or the maximum Doppler frequency increases for a fixed data rate $R_s = 1/T_s$, this assumption becomes invalid. The effect channel time variations on the OFDM link performance are now investigated. Although our analysis will be undertaken for flat fading channels, a similar analysis will apply to frequency selective channels provided that $G \geq L$. It will be shown that variations in the complex channel gain $\{g_k\}_{k=0}^{N-1}$ over the duration of an OFDM symbol cause interchannel interference (ICI) due to a loss of sub-channel orthogonality. The ICI will be shown to behave like additional AWGN that results in an error floor at high signal-to-noise ratios.

To isolate the Doppler effects, AWGN is ignored. The received discrete-time sequence after removal of the guard interval is

$$R_{n,m} = g_{G+(m-G)_N} X_{n,m}. \quad (5.133)$$

The vector $\mathbf{z}_n = \{z_{n,i}\}_{i=0}^{N-1}$ at the output of the FFT baseband demodulator is

$$z_{n,i} = \sqrt{2E_h/T} \sum_{m=0}^{N-1} x_{n,m} H(m-i), \quad (5.134)$$

where

$$H(m-i) = \frac{1}{N} \sum_{k=0}^{N-1} g_{G+(k-G)_N} e^{j\frac{2\pi}{N}(m-i)k}, \quad 0 \leq i \leq N-1. \quad (5.135)$$

To highlight the effect of channel time variations, (5.134) can be rewritten as

$$z_{n,i} = \sqrt{2E_h/T} H(0)x_{n,i} + c_{n,i}, \quad (5.136)$$

where

$$c_{n,i} = \sqrt{2E_h/T} \sum_{\substack{m=0 \\ m \neq i}}^{N-1} x_{n,m} H(m-i). \quad (5.137)$$

Note that $H(0)$ is the *effective* complex channel gain, while $c_{n,i}$ is an additive noise term due to the ICI. Note that if the channel is time-invariant, then $g_k = g$ and $z_{n,i} = g\sqrt{2E_h/T}x_{n,i}$ as before.

If N is sufficiently large in (5.137), the central limit theorem can be invoked and the $c_{n,i}, i = 0, \dots, N-1$ can be treated as complex Gaussian random variables that are characterized by their means, variances, and correlations. Since the $x_{n,m}$ and $H(m-i)$ are independent random variables and $E[x_{n,m}] = 0$, it follows that $E[c_i] = 0$. Since $2E_h \cdot \frac{1}{2}E[x_{n,k}x_{n,m}^*] = E_{av}\delta_{km}$, where E_{av} is the average symbol energy, the autocorrelation of the $c_{n,i}$ is

$$\phi_{cc}(r) = \frac{1}{2} \mathbb{E}[c_{n,i} c_{n,i+r}^*] = \frac{E_{av}}{T} \sum_{m \neq i, i+r} \mathbb{E}[H(m-i)H^*(m-i-r)]. \quad (5.138)$$

Proceeding further requires a model for the time correlation of the channel. If the normalization $\mathbb{E}[|g_k|^2] = 1$ is assumed and Clarke's 2-D isotropic scattering model is assumed with an isotropic receiver antenna (see Chap. 2), then the autocorrelation becomes

$$\phi_{cc}(r) = \frac{E_{av}}{T} \delta_r - \frac{E_{av}}{TN^2} \sum_{k=0}^{N-1} \sum_{k'=0}^{N-1} J_0(2\pi f_m T_s^g (k-k')) \left(e^{j\frac{2\pi k'r}{N}} + (1-\delta_r) e^{j\frac{2\pi kr}{N}} \right), \quad (5.139)$$

where f_m is the maximum Doppler frequency.

For symbol-by-symbol detection, it is sufficient to examine the variance of the ICI term

$$\phi_{cc}(0) = \frac{E_{av}}{T} - \frac{E_{av}}{TN^2} \left(N + 2 \sum_{i=1}^{N-1} (N-i) J_0(2\pi f_m T_s^g i) \right), \quad (5.140)$$

where the fact that $J_0(\cdot)$ is an even function has been used. Note that variance of the $c_{n,i}$ are only a function of E_{av} , N , T_s , and f_m , but is otherwise independent of the signal constellation. Figure 5.15 plots the signal-to-interference ratio, defined as

$$\text{SIR} \triangleq \frac{E_{av}/T}{\phi_{cc}(0)}, \quad (5.141)$$

as a function of $f_m T_s^g$ for several values of N . Observe that the SIR decreases as both the normalized Doppler maximum frequency $f_m T_s^g$ and the block size N increase.

Suppose that the data symbols $x_{n,k}$ are chosen from a 16-QAM alphabet. From Sect. 5.5, the symbol error probability for 16-QAM is

$$P_M = 3Q \left(\sqrt{\frac{1}{5}} \gamma_s \right) \left(1 - \frac{3}{4} Q \left(\sqrt{\frac{1}{5}} \gamma_s \right) \right), \quad (5.142)$$

where γ_s is the average received symbol energy-to-noise ratio. With Rayleigh fading, the symbol error probability is obtained by averaging (5.142) over the pdf in (5.86). Assuming validity of the Gaussian approximation for the ICI, the error floor due to ICI can be obtained by substituting the SIR in (5.141) for $\tilde{\gamma}_s$. The results are shown in Fig. 5.16. Simulation results are also

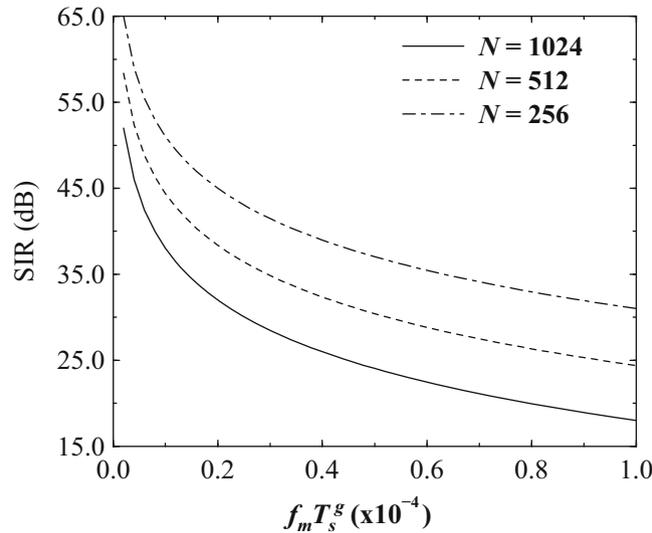


Fig. 5.15 Signal-to-interference ratio of OFDM due to ICI

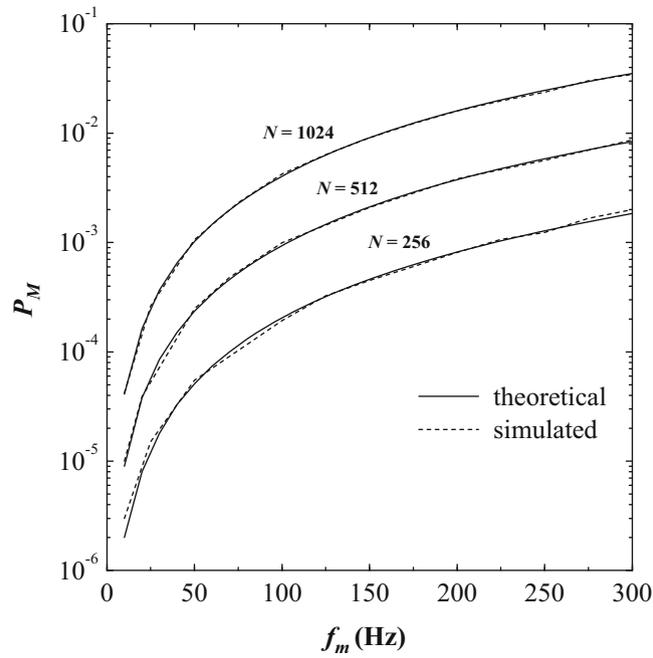


Fig. 5.16 Error floor due to ICI with 16-QAM

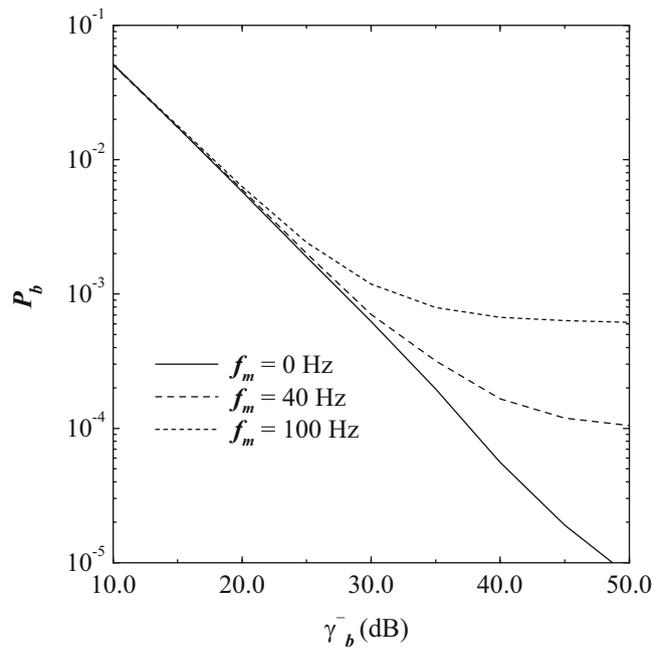


Fig. 5.17 Bit error probability for 16-QAM OFDM on a Rayleigh fading channel with various Doppler frequencies

shown in Fig. 5.16 corroborating the Gaussian approximation for the ICI. Figure 5.17 shows the bit error rate performance of OFDM with $N = 512$ sub-carriers, 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.

5.8 Differential Detection

Differentially encoded PSK (DPSK) can also be detected by using differentially coherent detection, where the receiver estimates the change in the excess phase of the received carrier between two successive baud intervals. Since DPSK transmits data in the differential excess carrier phase from one baud interval to the next, the basic mechanism for differential detection is obvious. For slow fading channels, the phase difference between waveforms received in two successive baud intervals will be independent of the absolute carrier phase. However for fast fading channels, the excess carrier phase will change over two successive baud intervals due to the channel. This leads to an error floor that increases with the fading rate.

5.8.1 Binary DPSK

Consider binary DPSK. Let θ_n denote the absolute transmitted excess carrier phase during the baud interval $nT \leq t \leq (n+1)T$, and let $\Delta\theta_n = \theta_n - \theta_{n-1}$ denote the differential excess carrier phase, where

$$\Delta\theta_n = \begin{cases} 0, & x_n = +1 \\ \pi, & x_n = -1 \end{cases}. \quad (5.143)$$

The DPSK complex envelope is

$$\tilde{s}(t) = A \sum_n h_a(t - nT) e^{j\theta_n} \quad (5.144)$$

and the received complex envelope is

$$\tilde{r}(t) = \alpha e^{j\phi} A \sum_n h_a(t - nT) e^{j\theta_n} + \tilde{n}(t), \quad (5.145)$$

where $g = \alpha e^{j\phi}$ is the complex channel gain. It is assumed that g changes slowly enough to remain essentially constant over two successive baud intervals.

A block diagram of a differentially coherent baseband demodulator for binary DPSK is shown in Fig. 5.18. During the time interval $nT \leq t \leq (n+1)T$, the values of X_n , X_{nd} , Y_n , and Y_{nd} in Fig. 5.18 are

$$\begin{aligned} X_n &= 2\alpha E_h \cos(\theta_n + \phi) + \tilde{n}_I \\ X_{nd} &= 2\alpha E_h \cos(\theta_{n-1} + \phi) + \tilde{n}_{I,d} \\ Y_n &= 2\alpha E_h \sin(\theta_n + \phi) + \tilde{n}_Q \\ Y_{nd} &= 2\alpha E_h \sin(\theta_{n-1} + \phi) + \tilde{n}_{Q,d} \end{aligned} \quad (5.146)$$

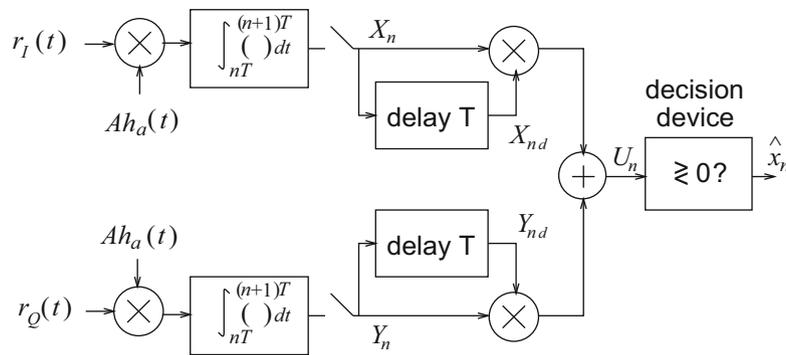


Fig. 5.18 Differentially coherent receiver for binary DPSK

where

$$E_h = \frac{A^2}{2} \int_0^T h_a^2(t) dt \quad (5.147)$$

and the noise terms are

$$\begin{aligned} \tilde{n}_I &= A \int_{nT}^{(n+1)T} \tilde{n}_I(t) h_a(t) dt \\ \tilde{n}_{I,d} &= A \int_{(n-1)T}^{nT} \tilde{n}_I(t) h_a(t) dt \\ \tilde{n}_Q &= A \int_{nT}^{(n+1)T} \tilde{n}_Q(t) h_a(t) dt \\ \tilde{n}_{Q,d} &= A \int_{(n-1)T}^{nT} \tilde{n}_Q(t) h_a(t) dt. \end{aligned} \quad (5.148)$$

One can show that \tilde{n}_I , $\tilde{n}_{I,d}$, \tilde{n}_Q , and $\tilde{n}_{Q,d}$ are independent identically distributed zero-mean Gaussian random variables with variance $2E_h N_o$.

In the absence of noise, it is easy to verify that the input to the decision device is $U_n = 4\alpha^2 E_h^2 x_n$. Hence, the sign of U_n is equal to the sign of x_n and correct decisions are made. When noise is present, U_n is a random variable, and to determine the probability of error the pdf of U_n is required. To determine the pdf of the U_n , it is convenient to express U_n as

$$U_n = \text{Re} \{Z_n Z_{nd}^*\} = \frac{1}{2} (Z_n Z_{nd}^* + Z_n^* Z_{nd}) \quad (5.149)$$

where

$$Z_n = X_n + jY_n \quad (5.150)$$

$$Z_{nd} = X_{nd} + jY_{nd}. \quad (5.151)$$

It can be shown by using characteristic functions that U_n has the differential form $U_n = W_n - Y_n$, where W_n and Y_n are non-central and central chi-square random variables with respective densities [308]

$$f_{W_n}(w) = \frac{1}{2E_h N_o} \exp \left\{ -\frac{(wx_n + 4\alpha^2 E_h^2)}{2E_h N_o} \right\} I_0 \left(\frac{2\sqrt{wx_n \alpha^2 E_h^2}}{E_h N_o} \right), \quad wx_n \geq 0 \quad (5.152)$$

$$f_{Y_n}(y) = \frac{1}{2E_h N_o} \exp \left\{ -\frac{yx_n}{2E_h N_o} \right\}, \quad yx_n \geq 0, \quad (5.153)$$

where $I_0(x)$ is the zero-order modified Bessel function of the first kind defined by

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{-x \cos \theta} d\theta. \quad (5.154)$$

By defining the auxiliary random variable $V_n = W_n$ and using a bivariate transformation of random variables, the pdf of U_n is

$$\begin{aligned} f_{U_n}(u) &= \int_{R_{U_n V_n}} f_{W_n}(v) f_{Y_n}(v-u) dv \\ &= \begin{cases} \frac{1}{4E_h N_o} \exp \left\{ \frac{x_n u - 2\alpha^2 E_h^2}{2E_h N_o} \right\}, & -\infty < x_n u < 0 \\ \frac{1}{4E_h N_o} \exp \left\{ \frac{x_n u - 2\alpha^2 E_h^2}{2E_h N_o} \right\} Q \left(\sqrt{\frac{2\alpha^2 E_h}{N_o}}, \sqrt{\frac{2x_n u}{E_h N_o}} \right), & 0 < x_n u < \infty \end{cases}, \end{aligned} \quad (5.155)$$

where $Q(a, b)$ is the Marcum Q function, defined by

$$Q(a, b) = 1 - \int_0^b ze^{-\frac{z^2+a^2}{2}} I_0(za) dz. \quad (5.156)$$

From (5.155), the bit error probability of DPSK with differential detection is

$$P_b(\gamma_b) = \int_0^\infty \frac{1}{4E_b N_o} \exp\left\{-\frac{u + 2\alpha^2 E_b^2}{2E_b N_o}\right\} du = \frac{1}{2} e^{-\gamma_b}, \quad (5.157)$$

where $\gamma_b = \alpha^2 E_b / N_o$ 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, γ_b , has the exponential pdf in (5.85). It follows that the average bit error probability with slow Rayleigh fading is

$$P_b = \int_0^\infty P_b(x) p_{\gamma_b}(x) dx = \frac{1}{2(1 + \bar{\gamma}_b)} \approx \frac{1}{2\bar{\gamma}_b}. \quad (5.158)$$

Note that the error probability has an inverse linear dependency on $\bar{\gamma}_b$.

5.8.2 Differential Detection of $\pi/4$ -DQPSK

Differential detection can be used with $\pi/4$ -DQPSK as well. Once again the complex envelopes of the transmitted and received signals are given by (5.144) and (5.145), respectively. However, with $\pi/4$ -DQPSK, $\Delta\theta_n = \pi x_n / 4$ where $x_n \in \{\pm 1, \pm 3\}$, so that one of the four possible differential phases must be detected. A block diagram of a differentially coherent baseband demodulator for $\pi/4$ -DQPSK is shown in Fig. 5.19. The values of $X_n, X_{nd}, Y_n,$ and Y_{nd} are again given by (5.146). The detector outputs are

$$U_n = \text{Re}\{Z_n Z_{nd}^*\} = \frac{1}{2} (Z_n Z_{nd}^* + Z_n^* Z_{nd}) \quad (5.159)$$

$$V_n = \text{Im}\{Z_n Z_{nd}^*\} = \frac{1}{j2} (Z_n Z_{nd}^* - Z_n^* Z_{nd}), \quad (5.160)$$

where Z_n and Z_{nd} are defined in (5.150) and (5.151), respectively. In the absence of noise, it can be verified that the detector outputs are

$$\begin{aligned} U_n &= -a, \quad V_n = -a, \quad \text{for } x_n = -3 \\ U_n &= a, \quad V_n = -a, \quad \text{for } x_n = -1 \\ U_n &= a, \quad V_n = a, \quad \text{for } x_n = +1 \\ U_n &= -a, \quad V_n = a, \quad \text{for } x_n = +3 \end{aligned} \quad (5.161)$$

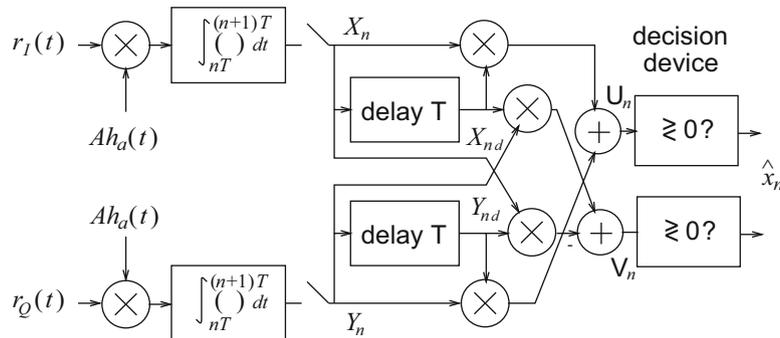


Fig. 5.19 Differentially coherent receiver for $\pi/4$ -DQPSK

where $a = 2\sqrt{2}\alpha^2 E_h^2$. The bit error probability for $\pi/4$ -DQPSK with Gray coding is somewhat complicated to derive and omitted here, but can be expressed in terms of well-known functions [272]

$$P_b(\gamma_b) = Q(a, b) - \frac{1}{2}I_0(ab)e^{-\frac{1}{2}(a^2+b^2)}, \quad (5.162)$$

where

$$a = \sqrt{2\gamma_b \left(1 - \frac{1}{\sqrt{2}}\right)} \quad (5.163)$$

$$b = \sqrt{2\gamma_b \left(1 + \frac{1}{\sqrt{2}}\right)} \quad (5.164)$$

and γ_b is the bit energy-to-noise ratio. Once again, if the channel is flat faded, then the bit error probability can be obtained by averaging (numerically) over the distribution of γ_b in (5.85).

5.9 Non-coherent Detection

If information is transmitted in the amplitude and/or frequency of a waveform, but not the phase, then a non-coherent receiver can be used. Non-coherent receivers make no attempt to determine the carrier phase and are, therefore, easier to implement than coherent receivers. Non-coherent receivers generally trade implementation complexity and robustness to channel impairments, such as high Doppler conditions, for transmitter power and/or bandwidth.

Suppose that one of the M complex low-pass waveforms, $\tilde{s}_m(t)$, $m = 1, \dots, M$, say $\tilde{s}_i(t)$, is transmitted on a flat fading channel with AWGN. The received complex envelope is

$$\tilde{r}(t) = g\tilde{s}_i(t) + \tilde{n}(t), \quad (5.165)$$

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 set of basis functions $\{\varphi_n(t)\}_{n=1}^N$ obtained through the Gram–Schmidt orthonormalization procedure, the received vector is obtained as

$$\tilde{\mathbf{r}} = g\tilde{\mathbf{s}}_i + \tilde{\mathbf{n}}, \quad (5.166)$$

where the joint pdf of $\tilde{\mathbf{n}}$ is given in (5.13).

The maximum likelihood (ML) non-coherent detector does not require knowledge of the random excess received carrier phase ϕ in the decision process, and chooses the message vector $\tilde{\mathbf{s}}_m$ that maximizes the joint conditional pdf $p(\tilde{\mathbf{r}}|\alpha, \tilde{\mathbf{s}}_m)$:

$$\text{choose } \tilde{\mathbf{s}}_m \text{ if } p(\tilde{\mathbf{r}}|\alpha, \tilde{\mathbf{s}}_m) \geq p(\tilde{\mathbf{r}}|\alpha, \tilde{\mathbf{s}}_{\hat{m}}) \quad \forall \hat{m} \neq m. \quad (5.167)$$

Letting $p(\phi)$ denote the pdf of ϕ ,

$$p(\tilde{\mathbf{r}}|\alpha, \tilde{\mathbf{s}}_m) = E_\phi[p(\tilde{\mathbf{r}}|g, \tilde{\mathbf{s}}_m)] = \int_0^{2\pi} p(\tilde{\mathbf{r}}|g, \tilde{\mathbf{s}}_m)p(\phi)d\phi. \quad (5.168)$$

Using the joint conditional pdf of $p(\tilde{\mathbf{r}}|g, \tilde{\mathbf{s}}_m)$ in (5.20) yields

$$\begin{aligned} p(\tilde{\mathbf{r}}|g, \tilde{\mathbf{s}}_m) &= \frac{1}{(2\pi N_o)^N} \exp\left\{-\frac{1}{2N_o} \|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_m\|^2\right\} \\ &= \frac{1}{(2\pi N_o)^N} \exp\left\{-\frac{\|\tilde{\mathbf{r}}\|^2 + 2\alpha^2 E_m}{2N_o}\right\} \exp\left\{\frac{1}{N_o} \text{Re}\{\tilde{\mathbf{r}} \cdot g^* \tilde{\mathbf{s}}_m^*\}\right\}, \end{aligned} \quad (5.169)$$

where, again, E_m is the energy in the bandpass waveform $s_m(t)$ corresponding to the signal vector $\tilde{\mathbf{s}}_m$. Next let $\tilde{\mathbf{r}} \cdot \tilde{\mathbf{s}}_m^* = X_m e^{j\theta_m}$ so that

$$\tilde{\mathbf{r}} \cdot g^* \tilde{\mathbf{s}}_m^* = g^* \tilde{\mathbf{r}} \cdot \tilde{\mathbf{s}}_m^* = g^* X_m e^{j\theta_m} = \alpha X_m e^{j(\theta_m - \phi)}. \quad (5.170)$$

Hence,

$$p(\tilde{\mathbf{r}}|g, \tilde{\mathbf{s}}_m) = \frac{1}{(2\pi N_o)^N} \exp \left\{ -\frac{\|\tilde{\mathbf{r}} + 2\alpha^2 E_m\|^2}{2N_o} \right\} \exp \left\{ \frac{\alpha X_m}{N_o} \cos(\theta_m - \phi) \right\}. \quad (5.171)$$

In the absence of any prior information, the random phase ϕ is assumed to be uniformly distributed on $[-\pi, \pi)$, resulting in

$$\begin{aligned} p(\tilde{\mathbf{r}}|\alpha, \tilde{\mathbf{s}}_m) &= \frac{1}{(2\pi N_o)^N} \exp \left\{ -\frac{2\alpha^2 E_m + \|\tilde{\mathbf{r}}\|^2}{2N_o} \right\} \frac{1}{2\pi} \int_0^{2\pi} \exp \left\{ \frac{\alpha X_m}{N_o} \cos(\theta_m - \phi) \right\} d\phi \\ &= \frac{1}{(2\pi N_o)^N} \exp \left\{ -\frac{2\alpha^2 E_m + \|\tilde{\mathbf{r}}\|^2}{2N_o} \right\} I_0 \left(\frac{\alpha X_m}{N_o} \right). \end{aligned} \quad (5.172)$$

Since the terms $\|\tilde{\mathbf{r}}\|^2$ and $(2\pi N_o)^N$ are independent of the choice of $\tilde{\mathbf{s}}_m$, the signal vector that maximizes $p(\tilde{\mathbf{r}}|\alpha, \tilde{\mathbf{s}}_m)$ also maximizes the decision metric

$$\mu_1(\mathbf{s}_m) = \exp \left\{ -\frac{\alpha^2 E_m}{N_o} \right\} I_0 \left(\frac{\alpha X_m}{N_o} \right). \quad (5.173)$$

If all message waveforms have equal energy, then considerable simplification will result. In this case, the ML receiver can choose $\tilde{\mathbf{s}}_m$ to maximize

$$\mu_2(\tilde{\mathbf{s}}_m) = I_0 \left(\frac{\alpha X_m}{N_o} \right). \quad (5.174)$$

However, $I_0(x)$ increases monotonically with x . Therefore, the ML receiver can simply choose $\tilde{\mathbf{s}}_m$ to maximize

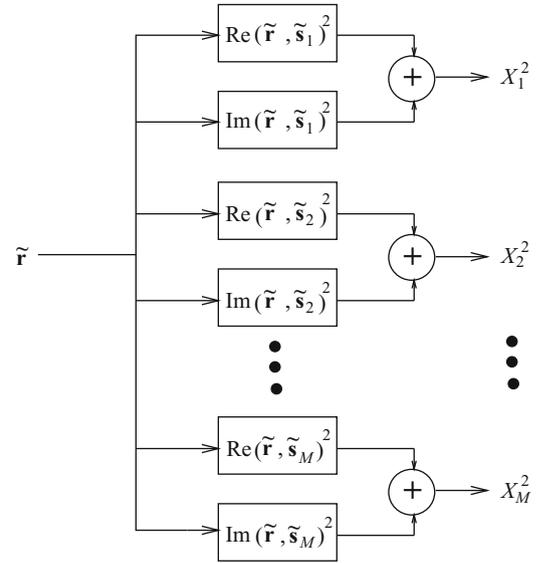
$$\mu_3(\tilde{\mathbf{s}}_m) = X_m. \quad (5.175)$$

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 $\tilde{r}_I(t)$ and $\tilde{r}_Q(t)$. Then it computes the received signal vector \mathbf{r} using the correlator detector in Fig. 5.2 or matched filter detector in Fig. 5.3. For equal energy messages, the decision variables $X_m = |\tilde{\mathbf{r}} \cdot \tilde{\mathbf{s}}_m^*|$, $m = 1, \dots, M$, are computed and the signal vector $\tilde{\mathbf{s}}_m$ is chosen that has the largest X_m . If the messages do not have equal energy, then the metric in (5.173) must be used instead. This will add considerable complexity to the ML receiver, because the channel gain α must be determined and the Bessel function $I_0(x)$ must be calculated. Finally,

$$X_m = [\text{Re}^2(\tilde{\mathbf{r}} \cdot \tilde{\mathbf{s}}_m^*) + \text{Im}^2(\tilde{\mathbf{r}} \cdot \tilde{\mathbf{s}}_m^*)]^{1/2}. \quad (5.176)$$

This leads to the detector structure shown in Fig. 5.20, commonly known as a square-law detector. Note that the square-law detector generates X_m^2 , $m = 1, \dots, M$ rather than X_m , $m = 1, \dots, M$. However, the choice of $\tilde{\mathbf{s}}_m$ that maximizes X_m^2 also maximizes X_m .

Fig. 5.20 Non-coherent square-law detector



5.9.1 Error Probability of M -ary Orthogonal Signals

Consider the case of M -ary orthogonal signals as discussed in Sect. 4.5. Assume without loss of generality that \tilde{s}_1 is sent. Then the received vector $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_N)$ has components

$$\begin{aligned}\tilde{r}_1 &= g\sqrt{2E} + \tilde{n}_1 \\ \tilde{r}_i &= \tilde{n}_i, \quad i = 2, \dots, M.\end{aligned}\quad (5.177)$$

Since the M -ary orthogonal signals have equal energy, the metric in (5.175) can be used. Then

$$\begin{aligned}X_1 &= |\tilde{\mathbf{r}} \cdot \tilde{\mathbf{s}}_1^*| \\ &= |2Eg + \sqrt{2E}\tilde{n}_1| \\ &= \left| 2E\alpha \cos(\phi) + \sqrt{2E}\tilde{n}_{I,1} + j(2E\alpha \sin(\phi) + \sqrt{2E}\tilde{n}_{Q,1}) \right|\end{aligned}\quad (5.178)$$

and

$$\begin{aligned}X_m &= |\tilde{\mathbf{r}} \cdot \tilde{\mathbf{s}}_m^*| \\ &= |\sqrt{2E}\tilde{n}_m| \\ &= \left| \sqrt{2E}\tilde{n}_{I,m} + j\sqrt{2E}\tilde{n}_{Q,m} \right|, \quad m = 2, \dots, M.\end{aligned}\quad (5.179)$$

The receiver will make a correct decision if

$$X_1 > X_i, \quad \forall i \neq 1. \quad (5.180)$$

From Appendix A, X_1 has the Rice distribution

$$p_{X_1}(x) = \frac{x}{2EN_o} \exp\left\{-\frac{x^2 + 4\alpha^2 E^2}{4EN_o}\right\} I_0\left(\frac{\alpha x}{N_o}\right), \quad (5.181)$$

while the $X_i, i \neq 0$, are independent Rayleigh random variables with pdf

$$p_{X_i}(x) = \frac{x}{2EN_o} \exp\left\{-\frac{x^2}{4EN_o}\right\}, \quad i = 2, \dots, M. \quad (5.182)$$

The probability of correct symbol reception is

$$\begin{aligned} P[c] &= P[X_2 < X_1, X_3 < X_1, \dots, X_M < X_1] \\ &= \int_0^\infty \left(\int_0^y \frac{x}{2EN_o} \exp\left\{-\frac{x^2}{4EN_o}\right\} dx \right)^{M-1} p_{X_1}(y) dy \\ &= \int_0^\infty \left(1 - \exp\left\{-\frac{y^2}{4EN_o}\right\} \right)^{M-1} p_{X_1}(y) dy. \end{aligned} \quad (5.183)$$

Using the binomial expansion

$$(1-x)^n = \sum_{k=0}^n \binom{n}{k} (-1)^k x^k$$

gives

$$P[c] = \sum_{k=0}^{M-1} (-1)^k \binom{M-1}{k} \int_0^\infty \exp\left\{-\frac{ky^2}{4EN_o}\right\} p_{X_1}(y) dy. \quad (5.184)$$

The integral in the above expression is

$$\begin{aligned} \mathcal{I} &= \int_0^\infty \exp\left\{-\frac{ky^2}{4EN_o}\right\} p_{X_1}(y) dy \\ &= \int_0^\infty \exp\left\{-\frac{ky^2}{4EN_o}\right\} \frac{y}{2EN_o} \exp\left\{-\frac{y^2 + 4\alpha^2 E^2}{4EN_o}\right\} I_0\left(\frac{\alpha y}{N_o}\right) dy \\ &= \int_0^\infty \frac{y}{2EN_o} \exp\left\{-\frac{(k+1)y^2 + 4\alpha^2 E^2}{4EN_o}\right\} I_0\left(\frac{\alpha y}{N_o}\right) dy. \end{aligned} \quad (5.185)$$

The trick is to manipulate the integrand of \mathcal{I} into the product of a Ricean pdf and a term that does not depend on the variable of integration y . This is accomplished by making the substitutions

$$N'_o = \frac{2EN_o}{k+1} \quad E' = \frac{E}{(k+1)} \quad (5.186)$$

and solving the integral. This gives

$$\mathcal{I} = \frac{1}{k+1} \exp\left\{-\frac{2k\alpha^2 E'^2}{N'_o}\right\} = \frac{1}{k+1} \exp\left\{-\frac{k\gamma_s}{(k+1)}\right\}, \quad (5.187)$$

where $\gamma_s = \alpha^2 E/N_o$ is the symbol energy-to-noise ratio. Hence, the probability of correct symbol reception is

$$P[c] = \sum_{k=0}^{M-1} \frac{(-1)^k \binom{M-1}{k}}{k+1} \exp\left\{-\frac{k\gamma_s}{(k+1)}\right\} \quad (5.188)$$

and the probability of symbol error is

$$P_M = 1 - P[c] = \sum_{k=1}^{M-1} \frac{(-1)^{k+1} \binom{M-1}{k}}{k+1} \exp \left\{ -\frac{k\gamma_s}{(k+1)} \right\}. \quad (5.189)$$

For orthogonal signals $\gamma_s = \gamma_b \log_2 M$ and the bit error probability is given by (5.56). Hence,

$$P_b(\gamma_b) = \frac{M}{2(M-1)} \sum_{k=1}^{M-1} \frac{(-1)^{k+1} \binom{M-1}{k}}{k+1} \exp \left\{ -\frac{k\gamma_b \log_2 M}{(k+1)} \right\}. \quad (5.190)$$

For Rayleigh fading channels, the error probability can be averaged over the distribution of γ_b in (5.85). This gives the following simple closed form for the average bit error probability

$$P_b = \frac{M}{2(M-1)} \sum_{k=1}^{M-1} \frac{(-1)^{k+1} \binom{M-1}{k}}{1+k+k\bar{\gamma}_b \log_2 M}. \quad (5.191)$$

Once again, the error probability has an inverse linear dependency on $\bar{\gamma}_b$.

5.10 Detection of CPM Signals

CPM receivers can be categorized into three different types of detection schemes: coherent detection, differential detection, and non-coherent detection. Furthermore, in each category there are two approaches: symbol-by-symbol detectors and sequence estimators. Sequence estimators will be treated in the context of channel coding in Chap. 8. This section only considers 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 receiver structures 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.

Recall that the partial response CPM complex envelope during the time interval $nT \leq t \leq (n+1)T$ is, from (4.133),

$$\tilde{s}(t) = A e^{j(\theta_n + 2\pi h \sum_{k=n-L+1}^n x_n \beta(t-kT))}, \quad (5.192)$$

and the CPM state at time $t = nT$ is defined by the L -tuple

$$S_n = (\theta_n, x_{n-1}, x_{n-2}, \dots, x_{n-L+1}). \quad (5.193)$$

In the sequel, the CPM complex envelope during the time interval $nT \leq t \leq (n+1)T$ will also be denoted by $\tilde{s}(S_n, x_n, t)$ to emphasize the finite state nature of the signal. For a slow flat fading channel, the received signal is

$$\tilde{r}(t) = g\tilde{s}(t) + \tilde{n}(t), \quad (5.194)$$

where $\tilde{n}(t)$ is a zero-mean complex AWGN with psd N_o watts/Hz.

5.10.1 Coherent Detection and Error Probability of MSK and GMSK

MSK signals can be recovered using a variety of detection techniques. One method uses the linear representation of MSK, where MSK is equivalent to OQASK with a half-sinusoid amplitude shaping function as described in (4.111)–(4.116). The received complex envelope is

$$\tilde{r}(t) = g\tilde{s}(t) + \tilde{n}(t), \quad (5.195)$$

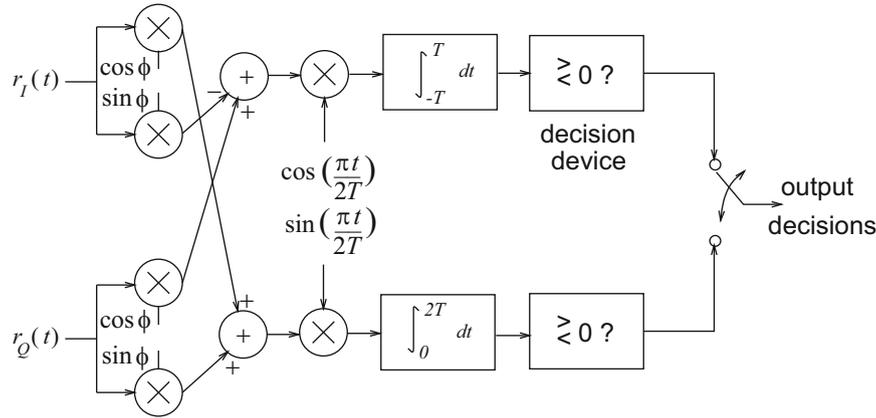


Fig. 5.21 Coherent detector for MSK signals

where $g = \alpha e^{j\phi}$. A coherent MSK receiver first removes the effect of the channel phase rotation according to

$$\begin{aligned} e^{-j\phi} \tilde{r}(t) &= \tilde{r}_I(t) \cos(\phi) + \tilde{r}_Q(t) \sin(\phi) + j(\tilde{r}_Q(t) \cos(\phi) - \tilde{r}_I(t) \sin(\phi)) \\ &= \alpha \tilde{s}_I(t) + \tilde{n}_I(t) + j(\alpha \tilde{s}_Q(t) + \tilde{n}_Q(t)), \end{aligned} \quad (5.196)$$

where the effect of the phase rotation on the noise $\tilde{n}(t)$ has been ignored due to its circular symmetry. Detection then proceeds by processing the real and imaginary parts of $e^{-j\phi} \tilde{r}(t)$ as orthogonal binary PAM streams. The resulting MSK detector is shown in Fig. 5.21. 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 half-sinusoid amplitude shaping pulse $h_a(t)$, and bit decisions are made every T seconds. It follows that coherently detected MSK has the same bit error rate performance as QPSK, OQPSK, and BPSK.

By using the linearized representation of GMSK, LGMSK, in Sect. 4.8.3, replaces the half-sinusoid amplitude shaping pulse in (4.8.3) with the LGMSK pulse in (4.157). This will create ISI that will degrade performance somewhat if a matched filter with symbol-by-symbol decisions are used. For example, with $BT = 0.25$, the E_b/N_o degradation is roughly 0.75 dB in AWGN [122]. The induced ISI can be mitigated by using a variety of channel equalization techniques for linear modulation schemes as detailed in Chap. 7, such that the performance loss due to the Gaussian pulse shaping is negligible.

5.10.2 Coherent CPM Demodulator

A coherent CPM demodulator was proposed by Osborn and Luntz [252], and Schonhoff [295]. The decision metrics for symbol x_n are obtained by observing $\tilde{r}(t)$ over $N_p + 1$ successive symbol intervals and generating decision metrics for all M^{N_p+1} possible symbol vectors $\mathbf{x}_n = \{x_n, \mathbf{b}_n\}$, where $\mathbf{b}_n = \{x_{n+1}, \dots, x_{n+N_p}\}$. 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

$$\mu(S_n, x_n, \mathbf{b}_n) = - \sum_{i=n}^{n+N_p} \int_{iT}^{(i+1)T} |\tilde{r}(t) - g\tilde{s}(S_i, x_i, t)|^2 dt. \quad (5.197)$$

The metrics for x_n can be obtained by averaging (5.197) over the M^{N_p} possible values of \mathbf{b}_n and averaging over all possible initial states S_n . This leads to the decision metric

$$\mu(x_n) = \sum_{S_n} \sum_{\mathbf{b}_n} \mu(S_n, x_n, \mathbf{b}_n) P[\mathbf{b}_n] P[S_n] = \sum_{S_n} \sum_{\mathbf{b}_n} \mu(S_n, x_n, \mathbf{b}_n), \quad (5.198)$$

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 equally likely, and all the S_n are equally likely, for equally likely data symbols. By using (5.198) a set of M metrics is calculated for the M possible x_n . The receiver makes the final decision by choosing the symbol having the largest decision metric.

A simplified receiver that will yield almost the same performance uses the suboptimum decision metric [252, 295]

$$\mu(x_n) = \max_{S_n} \max_{\mathbf{b}_n} \left\{ - \sum_{i=n}^{n+N_p} \int_{iT}^{(i+1)T} |\tilde{r}(t) - g\tilde{s}(S_i, x_i, t)|^2 dt \right\}, \quad (5.199)$$

which is also exactly the same as the decision metric proposed by Kerr and McLane for full response CPFSK [180]. Once again, by using (5.199) a set of M decision metrics is calculated for the M possible x_n and the receiver chooses the symbol having the largest decision metric.

5.10.3 Non-coherent CPM Demodulator

A non-coherent receiver can be constructed by using a multiple-symbol observation interval similar to that suggested for the coherent receiver described in the previous section [368]. After observing $\tilde{r}(t)$ over the N -symbol interval $(n - n_1)T \leq t \leq (n + n_2)T$, where $N = n_1 + n_2 + 1$, the non-coherent CPM demodulator in [368] generates the following set of M^{N+L-2} conditional symbol metrics for each x_n :

$$\mu(x_n, \mathbf{b}_n) = \left| \sum_{i=n-n_1}^{n+n_2} \int_{iT}^{(i+1)T} \tilde{r}(t)\tilde{s}^*(S_i, x_i, t) dt \right|^2, \quad (5.200)$$

where $\mathbf{b}_n = \{x_{n-n_1-L+1}, \dots, x_{n-1}, x_{n+1}, \dots, x_{n+n_2}\}$ is the “adjacent” symbol vector that excludes x_n . Note that the phase term θ_{n-n_1} in S_{n-n_1} does not affect the value of (5.200) and can, therefore, be assumed zero. A simple symbol metric can be formed by choosing the largest among all possible $\mu(x_n, \mathbf{b}_n)$, viz.,

$$\mu(x_n) = \max_{\mathbf{b}_n} \left| \sum_{i=n-n_1}^{n+n_2} \int_{iT}^{(i+1)T} \tilde{r}(t)\tilde{s}^*(S_i, x_i, t) dt \right|^2. \quad (5.201)$$

The set of M symbol metrics so obtained is then used to make decisions on the transmitted symbols by selecting the symbol with the largest symbol metric.

For $N = 1$ ($n_1 = n_2 = 0$), the symbol metric in (5.201) is the same one used by the single-symbol receiver in [5] 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 in an efficient recursive fashion, an approach similar to [304] can be followed to rewrite $\mu(x_n, \mathbf{b}_n)$ as

$$\mu(x_n, \mathbf{b}_n) = \left| \sum_{i=n-n_1}^{n+n_2} \Gamma_i F_i \right|^2, \quad (5.202)$$

where

$$\begin{aligned} \Gamma_i &= \int_{iT}^{(i+1)T} \tilde{r}(t)\tilde{s}^*(x_{i-L+1}, \dots, x_i, t) dt \\ F_i &= e^{-j\pi h x_{i-L}} F_{i-1}; \quad F_{n-n_1} = 1. \end{aligned} \quad (5.203)$$

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} conditional symbol metrics $\mu(x_n, \mathbf{b}_n)$. However, unlike the coherent receiver, the complexity is independent of the modulation index h . Actually, since the term θ_n is not explicitly exploited in S_n , h is not even required to be a rational number, i.e., the CPM waveform is not required to have a finite number of states. Finally, it is observed that the complex channel gain g is not required and, therefore, the receiver complexity is greatly reduced.

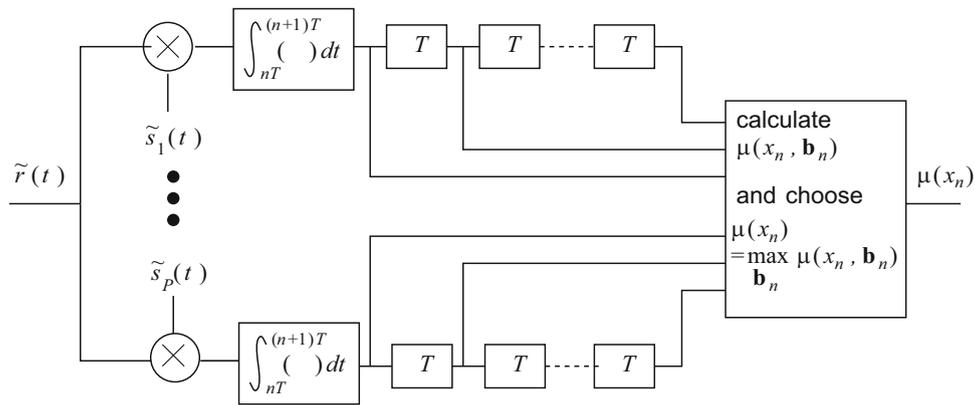


Fig. 5.22 The symbol metric calculator. Note that the signal $\tilde{s}^*(t)$ is labeled to account for $P = M^L$ possible matched filters

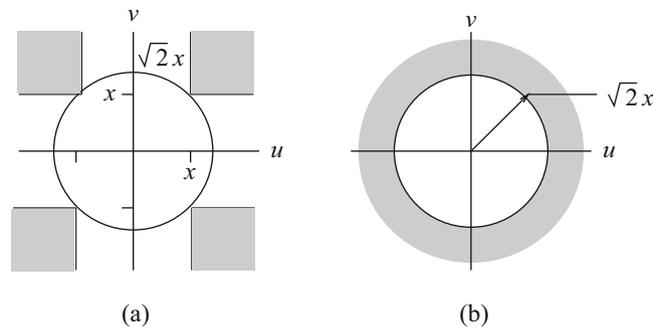


Fig. 5.23 Figure for Problem 5.1

Problems

5.1. Derive the upper bound

$$Q(x) \leq \frac{1}{2} e^{-x^2/2}, \quad x \geq 0.$$

Hint: Note that $4Q^2(x)$ is the probability that a pair of independent zero mean, unit variance, Gaussian random variables u, v lie within the shaded region of Fig. 5.23a. This probability is exceeded by the probability that u, v lies within the shaded region of Fig. 5.23b.

5.2. Consider the receiver model shown in Fig. 5.24, consisting of a linear time-invariant filter $h_r(t)$ followed by a sampler. The input to the filter consists of a pulse $h_a(t)$ of duration T corrupted by AWGN

$$\tilde{r}(t) = h_a(t) + \tilde{n}(t), \quad 0 \leq t \leq T.$$

The output of the filter is

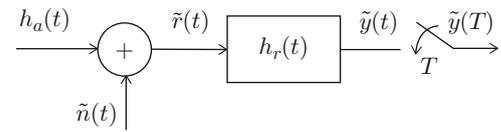
$$\tilde{y}(t) = p(t) + \tilde{z}(t)$$

where $p(t) = h_a(t) * h_r(t)$ and $\tilde{z}(t) = \tilde{n}(t) * h_r(t)$, the filter output is sampled at time T to produce the sample $y(T) = p(T) + \tilde{z}(T)$. The signal-to-noise ratio at the output of the sampler is defined as

$$\text{SNR} = \frac{|p(T)|^2}{\text{E}[|\tilde{z}(T)|^2]}.$$

Find the filter $h_r(t)$, and corresponding transfer function $H_r(f)$, that will maximize the SNR.

Fig. 5.24 Figure for Problem 5.2



5.3. Consider the pulse

$$h_a(t) = \begin{cases} \sin\left(\frac{4\pi t}{T}\right), & 0 \leq t \leq T \\ 0, & \text{otherwise} \end{cases}.$$

- Determine the impulse response of the matched filter for this signal.
- Sketch the waveform $y(t)$ at the output of the matched filter, and determine the output value at time $t = T$.
- Sketch the waveform $y(t)$ at the output of a correlator that correlates $h_a(t)$ with itself, and determine the output value at time $t = T$.

5.4. Derive the expression for the symbol error probability of \sqrt{M} -PAM in (5.102).

5.5. Show that the symbol error probability for coherent M -ary PSK is bounded by $p \leq P_M \leq 2p$, where

$$p = Q\left(\sqrt{2\gamma_s} \sin\frac{\pi}{M}\right)$$

and γ_s is the symbol energy-to-noise ratio.

5.6. 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. The composite squared-envelope has the probability density in (2.371).

- Obtain an expression for the probability density function of the composite envelope.
- Find an expression for the probability of bit error as a function of the average received bit energy-to-noise ratio.
- Plot your results in part (b) for different values of σ_Ω .

5.7 (Computer Exercise). This problem requires that you first complete the computer exercise in Problem 2.46, 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 one of the two possible signal vectors $\tilde{s}_0 = -\tilde{s}_1 = \sqrt{2E_b}$ are sent 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 2.46. 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 $f_m T = 0.1$.

- (b) Plot the simulated bit error probability, P_b , against the average received bit energy-to-noise ratio $\bar{\gamma}_b = E[\alpha^2]E_b/N_o$. 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 fading envelope generator.*

5.8. Consider a non-orthogonal coherent binary FSK system with the bandpass waveforms

$$\begin{aligned} s_1(t) &= A \cos(2\pi f_c t), \quad 0 \leq t \leq T \\ s_2(t) &= A \cos(2\pi(f_c + \Delta_f)t), \quad 0 \leq t \leq T \end{aligned}$$

and assume that $f_c T \gg 1$.

- (a) Show that the correlation between the bandpass waveforms is given by

$$\rho = \frac{1}{E} \int_0^T s_1(t)s_2(t)dt = \text{sinc}(2\Delta_f T)$$

where E is the energy in the waveforms.

- (b) What is the value of Δ_f that minimizes the probability of symbol error?
 (c) For the value of Δ_f obtained in part (b), determine the increase in the received bit energy-to-noise ratio, γ_b , required so that this coherent FSK system has the same bit error probability as a coherent binary PSK system.

5.9. OFDM systems are known to be resilient to timing errors. Consider the following OFDM waveform with a cyclic prefix

$$\tilde{s}_g(t) = A \sum_n b(t - nT_g, \mathbf{x}_n),$$

where

$$b(t, \mathbf{x}_n) = u_{\alpha_g T}(t + \alpha_g T) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi k(t+T)}{NT_s}} + u_T(t) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi k t}{NT_s}},$$

and $T = NT_s$ and $T_g = (1 + \alpha_g)T$. Suppose the waveform $\tilde{s}_g(t)$ is sampled every T_s seconds. For the n th OFDM symbol, this yields the sample sequence $\{X_{n,m}\}_{m=0}^{N-1}$, where

$$X_{n,m} = \tilde{s}(-\alpha_g T + nT_g + mT_s + \Delta_t),$$

and Δ_t is a timing offset. For the n th OFDM symbol an FFT is taken on the sample sequence $\{X_{n,m}\}_{m=0}^{N-1}$.

- (a) Suppose that the timing offset Δ_t lies in the interval $(0, \alpha_g T)$ such that the samples $\{X_{n,m}\}_{m=0}^{N-1}$ all belong to the n th OFDM symbol. Determine the FFT coefficients.
 (b) Now suppose that the timing offset Δ_t lies outside the interval $(0, \alpha_g T)$, such that the samples $\{X_{n,m}\}_{m=0}^{N-1}$ do not all belong to the n th OFDM symbol. Determine the FFT coefficients.

5.10. Consider the following OFDM waveform with a cyclic suffix and a carrier frequency offset Δ_f :

$$\tilde{s}_g(t) = A \sum_n b(t - nT_g, \mathbf{x}_n),$$

where

$$\begin{aligned} b(t, \mathbf{x}_n) &= u_T(t) \sum_{k=0}^{N-1} x_{n,k} \exp \left\{ j2\pi \left(\frac{k}{NT_s} + \Delta_f \right) t \right\} \\ &+ u_{\alpha_g T}(t - T) \sum_{k=0}^{N-1} x_{n,k} \exp \left\{ j2\pi \left(\frac{k}{NT_s} + \Delta_f \right) t \right\}, \end{aligned}$$

$T = NT_s$ and $T_g = (1 + \alpha_g)T$. The waveform $\tilde{s}_g(t)$ is sampled every T_s seconds. For the n th OFDM symbol, this yields the sample sequence $\{X_{n,m}\}_{m=0}^{N-1}$, where

$$X_{n,m} = \tilde{s}(nT_g + mT_s).$$

An FFT is taken on the sample sequence $\{X_{n,m}\}_{m=0}^{N-1}$.

Show that the FFT coefficients (in absence of noise) can be written as

$$Z_{n,i} = \text{FFT}\{X_{n,m}\} = \eta x_{n,i} + c_i,$$

where

$$\eta = A \left\{ \frac{\sin(\pi N \Delta_f T_s)}{N \sin(\pi \Delta_f T_s)} \right\} e^{j\pi(N-1)\Delta_f T_s}$$

and

$$c_i = A \sum_{\substack{m=0 \\ m \neq i}}^{N-1} x_{n,m} H(m, i)$$

is the random ICI term, where

$$H(m, i) = \left\{ \frac{\sin(\pi(m-i+N\Delta_f T_s))}{N \sin(\pi(m-i+N\Delta_f T_s)/N)} \right\} e^{j\pi(\frac{N-1}{N})(m-i+N\Delta_f T_s)}.$$

5.11. 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.

5.12. Consider the differentially coherent receiver shown in Fig. 5.18. Show that the pdf of U_n is given by (5.155).

5.13. Consider a system that uses M -ary orthogonal modulation with non-coherent detection. The error probability on an AWGN channel is known to be

$$P_b = \frac{M}{2(M-1)} \sum_{k=1}^{M-1} \frac{(-1)^{k+1} \binom{M-1}{k}}{k+1} \exp\left\{-\frac{k\gamma_s}{(k+1)}\right\},$$

where $\gamma_s = \alpha^2 E_s / N_o$ is the received *symbol*-energy-to-noise ratio.

(a) Derive the corresponding expression for the probability of bit error on a slow flat Rayleigh fading channel. Express your result in terms of the average received *bit*-energy-to-noise ratio, $\bar{\gamma}_b$, and simplify to closed form.

(b) Repeat part (a) for a slow flat Ricean fading channel. Simplify as much as possible.

5.14. Consider binary CPFSK modulation with modulation index $h \leq 0.5$. Compute the minimum squared Euclidean distance between any pair of bandpass waveforms as given by

$$D_{\min}^2 = \lim_{n \rightarrow \infty} \min_{i,j} \int_0^{nT} (s(t; \mathbf{x}^{(i)}) - s(t; \mathbf{x}^{(j)}))^2 dt,$$

where $s(t; \mathbf{x}^{(i)})$ and $s(t; \mathbf{x}^{(j)})$ are the two bandpass signals whose phase trajectories diverge at time $t = 0$ and merge sometime later. What is the pairwise error probability between two such signals?

5.15. The squared Euclidean distance between a pair of CPM bandpass waveforms, $s(t; \mathbf{x}^{(i)})$ and $s(t; \mathbf{x}^{(j)})$, is

$$D^2 = \int_0^{\infty} (s(t; \mathbf{x}^{(i)}) - s(t; \mathbf{x}^{(j)}))^2 dt.$$

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.

5.16. Construct a differential detector for MSK signaling. Obtain an expression for the probability of bit error for differentially detected MSK on an AWGN channel.

5.17. Suppose that GMSK signaling is used. Unfortunately, the GMSK pulse is non-causal and, therefore, a truncated version of the pulse is employed, i.e., the time domain pulse is

$$h_f(t) = \frac{1}{2T} \left(Q\left(\frac{t/T - 1/2}{\sigma}\right) - Q\left(\frac{t/T + 1/2}{\sigma}\right) \right) \text{rect}\left(\frac{t - L_T T}{2L_T T}\right),$$

where

$$Q(\alpha) = \int_\alpha^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2} dx \quad (5.204)$$

$$\sigma^2 = \frac{\ln 2}{4\pi^2(BT)^2}. \quad (5.205)$$

Compute the maximum value of the ISI term in (4.149) as a function of the normalized filter bandwidth BT when $L_T = 3$.

Chapter 6

Multi-Antenna Techniques

Rayleigh fading converts an exponential dependency of the bit error probability on the average received bit energy-to-noise ratio into an inverse linear one, yielding a very large performance loss. Diversity is a very effective remedy that exploits the principle of providing the receiver with multiple independently faded replicas of the same information bearing signal. Sometimes these replicas are called diversity branches. To illustrate the diversity mechanism, let p denote the probability that the instantaneous bit energy-to-noise ratio for any one diversity branch falls below a critical threshold γ_{th} . Then with independently faded diversity branches, the probability that the instantaneous bit energy-to-noise ratio simultaneously falls below the same critical threshold γ_{th} for all L diversity branches is p^L , and $p^L \ll p$ for small values of p .

The methods by which diversity can be achieved generally fall into six categories: (1) space, (2) angle, (3) polarization, (4) frequency, (5) multipath, and (6) time. Space diversity is achieved by using multiple transmit or receiver antennas. The spatial separation between the antenna elements at the transmitter and/or receiver is chosen so that the diversity branches experience uncorrelated fading. Chapter 2 showed that a spatial separation of about a half-wavelength will suffice for 2-D isotropic scattering and isotropic antenna elements. Angle (or direction) diversity requires a number of directional antennas. Each antenna selects plane waves arriving from a different spatial direction, 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. Frequency diversity uses multiple channels that are separated by at least the coherence bandwidth of the channel. In general, frequency diversity is not a bandwidth efficient solution. However, frequency hopping can be used along with coding where the elements of the codewords are transmitted 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 Chap. 9. 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 a bandwidth efficient method for implementing 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 an unacceptably large interleaving delay. Finally, the above techniques can be combined together. For example, spatial and temporal diversity can be combined together by using space-time coding techniques.

Once the diversity branches are generated, they must be combined together. A large variety of diversity combining techniques have been described in the literature. The type of diversity combining that is most effective will depend, among other things, on the type of additive impairment that is present, i.e., additive white Gaussian noise (AWGN) or co-channel interference (CCI). For AWGN dominant channels, maximal ratio combining is optimum in a maximum likelihood sense. For channels that are dominated by CCI, optimum combining is more effective, where interference correlation across multiple antenna branches is used to reject the CCI.

This chapter concentrates on antenna diversity techniques, although the mathematical concepts will readily apply to other types of diversity as well. Section 6.1 formulates the diversity combining problem when there is a single transmit antenna and multiple receiver antennas. Sections 6.2–6.5 consider selective combining, maximal ratio combining, equal gain combining, and switched combining, respectively. Section 6.6 considers differential detection with equal gain combining. Optimum combining for the purpose of combating fading and co-channel interference is considered in Sect. 6.8. Section 6.9 considers classical antenna beam-forming. Multiple-input multiple-output (MIMO) channels are introduced in Sect. 6.10.

The Alamouti transmit diversity scheme is presented in Sect. 6.11, while spatial multiplexing and detection is considered in Sect. 6.12. Spatial modulation is considered next in Sect. 6.13, where only one antenna in an array is activated. Finally, the chapter wraps up with massive MIMO in Sect. 6.14.

6.1 Diversity Combining

There are many methods for combining the signals that are received on the different diversity branches, and several ways of categorizing them. Diversity combining that takes place before matched filtering or correlation detection is sometimes called predetection combining, while diversity combining that takes place after matched filtering or correlation detection is sometimes called postdetection combining. In many cases there is no difference in theoretical performance between the two approaches, while in other cases there is a performance difference.

Consider the receive diversity system shown in Fig. 6.1. The signal that is received by each receiver antenna is demodulated to baseband with a quadrature demodulator as shown in Fig. 5.1 and then processed with correlator or matched filter detector shown in Figs. 5.2 and 5.3, respectively. Afterwards, the correlator or matched filter outputs are applied to a diversity combiner as shown in Fig. 6.1.

Consider a communication system that uses an M -ary modulation alphabet along with a receiver having L diversity branches. During each baud interval, one of M message waveforms is transmitted having the complex envelope $\tilde{s}_m(t)$, $m = 1, \dots, M$. The received complex envelopes for the L diversity branches are

$$\tilde{r}_k(t) = g_k \tilde{s}_m(t) + \tilde{n}_k(t), \quad k = 1, \dots, L, \quad (6.1)$$

where $g_k = \alpha_k e^{j\phi_k}$ is the complex fading gain associated with the k th branch. The AWGN processes $\tilde{n}_k(t)$, $k = 1, \dots, L$ are independent from branch to branch, since they correspond to the thermal noise that is introduced by the high gain amplifiers that are used to amplify the signals received on each of the receiver antenna elements. After correlating or matched filtering the received waveforms $\tilde{r}_k(t)$ with the basis functions that are obtained from a Gram–Schmidt orthonormalization procedure, the corresponding received signal vectors are

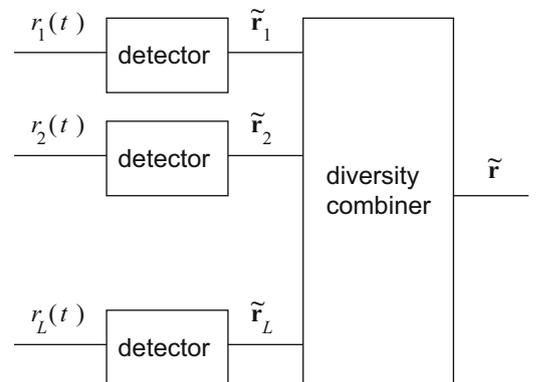
$$\tilde{\mathbf{r}}_k = g_k \tilde{\mathbf{s}}_m + \tilde{\mathbf{n}}_k, \quad k = 1, \dots, L, \quad (6.2)$$

where

$$\tilde{r}_{k_i} = g_k \tilde{s}_{m_i} + \tilde{n}_{k_i}, \quad i = 1, \dots, N. \quad (6.3)$$

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 analytical derivations, the diversity branches are often assumed to be uncorrelated in the literature. However, branch correlation will reduce the achievable diversity gain and, therefore, the uncorrelated branch assumption gives optimistic results. Nevertheless, the performance of the various diversity combining techniques will be evaluated under the assumption of uncorrelated branches.

Fig. 6.1 Postdetection diversity receiver



The fade distribution will affect the diversity gain as well. 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 bit energy-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 bit energy-to-noise ratio and diversity order. In this chapter, the performance with slow flat Rayleigh fading will be considered, although the methodology will apply to slow flat Ricean fading as well.

6.2 Selective Combining

With selective combining (SC), the diversity branch yielding the highest bit energy-to-noise ratio is always selected. In this case, the diversity combiner in Fig. 6.1 performs the operation

$$\tilde{\mathbf{r}} = \max_{|g_k|} \tilde{\mathbf{r}}_k. \quad (6.4)$$

For communication links that use continuous transmission, SC is impractical because it requires continuous monitoring of all diversity branches to estimate and track the time-varying complex gains g_k . If such channel estimation is performed, then it is 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 optimal in AWGN. In systems that transmit information in bursts, a form of SC can sometimes be implemented where the diversity branch is selected on a burst-by-burst basis, by using the synchronization word or training sequence that is inserted into each 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 the burst duration. In this section, however, selection diversity is evaluated under the idealized assumption of continuous branch selection.

With Rayleigh fading, the instantaneous received symbol energy-to-noise ratio on the k th diversity branch has the exponential pdf

$$p_{\gamma_k}(x) = \frac{1}{\bar{\gamma}_c} e^{-x/\bar{\gamma}_c}, \quad (6.5)$$

where $\bar{\gamma}_c$ is the average received branch symbol energy-to-noise ratio, assumed to be the same for all diversity branches. With ideal SC, the branch with the largest symbol energy-to-noise ratio is always selected so that the instantaneous symbol energy-to-noise ratio at the output of the selective combiner is

$$\gamma_s^s = \max \{\gamma_1, \gamma_2, \dots, \gamma_L\}, \quad (6.6)$$

where L is the number of branches. If the branches are independently faded, then order statistics gives the cumulative distribution function (cdf)

$$F_{\gamma_s^s}(x) = \text{P}[\gamma_1 \leq x, \gamma_2 \leq x, \dots, \gamma_L \leq x] = (1 - e^{-x/\bar{\gamma}_c})^L. \quad (6.7)$$

Differentiating the above expression gives the pdf of the instantaneous output symbol energy-to-noise ratio as

$$p_{\gamma_s^s}(x) = \frac{L}{\bar{\gamma}_c} (1 - e^{-x/\bar{\gamma}_c})^{L-1} e^{-x/\bar{\gamma}_c}. \quad (6.8)$$

The average output symbol energy-to-noise ratio with SC is

$$\begin{aligned} \bar{\gamma}_s^s &= \int_0^\infty x p_{\gamma_s^s}(x) dx \\ &= \int_0^\infty \frac{Lx}{\bar{\gamma}_c} (1 - e^{-x/\bar{\gamma}_c})^{L-1} e^{-x/\bar{\gamma}_c} dx \\ &= \bar{\gamma}_c \sum_{k=1}^L \frac{1}{k}. \end{aligned} \quad (6.9)$$

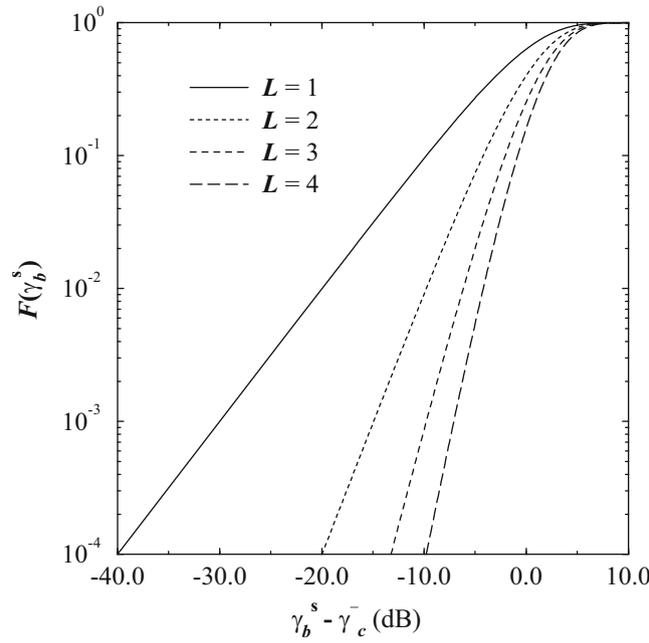


Fig. 6.2 Cdf of γ_b^s for selective combining; $\bar{\gamma}_c$ is the average branch symbol energy-to-noise ratio

Figure 6.2 plots the cdf $F_{\gamma_b^s}(x)$ against the normalized symbol energy-to-noise ratio $x/\bar{\gamma}_c$ in dB units. 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 behavior is typical for all diversity techniques.

The bit error probability with slow flat fading can be obtained by averaging the bit error probability, as a function of the symbol energy-to-noise ratio, over the pdf of γ_s^s . For example, the bit error probability for binary DPSK with differential detection on an AWGN channel is

$$P_b(\gamma_s^s) = \frac{1}{2}e^{-\gamma_s^s}, \quad (6.10)$$

where γ_s^s can be interpreted as the instantaneous bit energy-to-noise ratio since binary modulation is being used. Hence, with SC

$$\begin{aligned} P_b &= \int_0^\infty P_b(x)p_{\gamma_s^s}(x)dx \\ &= \int_0^\infty \frac{L}{2\bar{\gamma}_c} e^{-(1+1/\bar{\gamma}_c)x} (1 - e^{-x/\bar{\gamma}_c})^{L-1} dx \\ &= \frac{L}{2\bar{\gamma}_c} \sum_{n=0}^{L-1} \binom{L-1}{n} (-1)^n \int_0^\infty e^{-(1+(n+1)/\bar{\gamma}_c)x} dx \\ &= \frac{L}{2} \sum_{n=0}^{L-1} \frac{\binom{L-1}{n} (-1)^n}{1+n+\bar{\gamma}_c}, \end{aligned} \quad (6.11)$$

where the binomial expansion

$$(1-x)^{L-1} = \sum_{n=0}^{L-1} \binom{L-1}{n} (-1)^n x^n \quad (6.12)$$

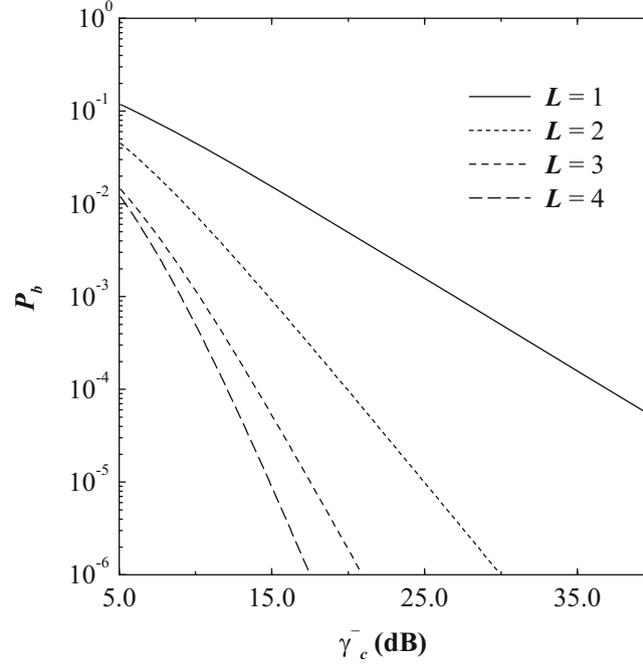


Fig. 6.3 Bit error probability for binary DPSK with differential detection and L -branch selective diversity combining

was used. The bit error probability is plotted in Fig. 6.3, where $\bar{\gamma}_c$ is equal to the branch bit energy-to-noise ratio. SC is seen to give a very large improvement in bit error rate performance. When $\bar{\gamma}_c \gg 1$, (6.11) shows that the bit error probability is proportional to $1/\bar{\gamma}_c^L$. Again, the greatest benefit of using diversity is achieved in going from $L = 1$ to $L = 2$ diversity branches and diminishing returns are obtained with increasing L .

6.3 Maximal Ratio Combining

With maximal ratio combining (MRC), the diversity branches are weighted by their respective complex fading gains and combined. MRC realizes a maximum likelihood receiver as is now shown. Referring to (6.2), the vector

$$\tilde{\mathbf{r}} = \text{vec}(\tilde{\mathbf{r}}_1, \tilde{\mathbf{r}}_2, \dots, \tilde{\mathbf{r}}_L) \quad (6.13)$$

has the multivariate complex Gaussian distribution

$$\begin{aligned} p(\tilde{\mathbf{r}}|\mathbf{g}, \tilde{\mathbf{s}}_m) &= \prod_{k=1}^L \prod_{i=1}^N \frac{1}{2\pi N_o} \exp\left\{-\frac{1}{2N_o} |\tilde{r}_{k,i} - g_k \tilde{s}_{m,i}|^2\right\} \\ &= \frac{1}{(2\pi N_o)^{LN}} \exp\left\{-\frac{1}{2N_o} \sum_{k=1}^L \|\tilde{\mathbf{r}}_k - g_k \tilde{\mathbf{s}}_m\|^2\right\} \end{aligned} \quad (6.14)$$

where $\mathbf{g} = (g_1, g_2, \dots, g_L)$ is the channel vector. The maximum likelihood receiver chooses the message vector $\tilde{\mathbf{s}}_m$ that maximizes the likelihood function $p(\tilde{\mathbf{r}}|\mathbf{g}, \tilde{\mathbf{s}}_m)$. This is equivalent to choosing the message vector $\tilde{\mathbf{s}}_m$ that minimizes the metric

$$\begin{aligned} \mu(\tilde{\mathbf{s}}_m) &= \sum_{k=1}^L \|\tilde{\mathbf{r}}_k - g_k \tilde{\mathbf{s}}_m\|^2 \\ &= \sum_{k=1}^L (\|\tilde{\mathbf{r}}_k\|^2 - 2\text{Re}\{\tilde{\mathbf{r}}_k \cdot g_k^* \tilde{\mathbf{s}}_m^*\} + |g_k|^2 \|\tilde{\mathbf{s}}_m\|^2). \end{aligned} \quad (6.15)$$

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

$$\begin{aligned}\mu_2(\tilde{\mathbf{s}}_m) &= \sum_{k=1}^L \operatorname{Re}\{\tilde{\mathbf{r}}_k \cdot g_k^* \tilde{\mathbf{s}}_m^*\} - E_m \sum_{k=1}^L |g_k|^2 \\ &= \sum_{k=1}^L \operatorname{Re} \left\{ g_k^* \int_{-\infty}^{\infty} \tilde{r}_k(t) \tilde{s}_m^*(t) dt \right\} - E_m \sum_{k=1}^L |g_k|^2.\end{aligned}\quad (6.16)$$

If signals have equal energy, then the last term can be neglected, since it is the same for all message vectors. This results in

$$\begin{aligned}\mu_3(\tilde{\mathbf{s}}_m) &= \sum_{k=1}^L \operatorname{Re}\{\tilde{\mathbf{r}}_k \cdot g_k^* \tilde{\mathbf{s}}_m^*\} \\ &= \sum_{k=1}^L \operatorname{Re} \left\{ g_k^* \int_{-\infty}^{\infty} \tilde{r}_k(t) \tilde{s}_m^*(t) dt \right\}.\end{aligned}\quad (6.17)$$

An alternative form of the ML receiver can also be obtained by rewriting the metric in (6.16) as

$$\begin{aligned}\mu_4(\tilde{\mathbf{s}}_m) &= \operatorname{Re} \left\{ \sum_{k=1}^L g_k^* \tilde{\mathbf{r}}_k \cdot \tilde{\mathbf{s}}_m^* \right\} - E_m \sum_{k=1}^L |g_k|^2 \\ &= \int_{-\infty}^{\infty} \operatorname{Re} \left\{ \left(\sum_{k=1}^L g_k^* \tilde{r}_k(t) \right) \tilde{s}_m^*(t) \right\} dt - E_m \sum_{k=1}^L |g_k|^2.\end{aligned}\quad (6.18)$$

From (6.18), the ML receiver can be constructed. The diversity combiner in Fig. 6.1 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.

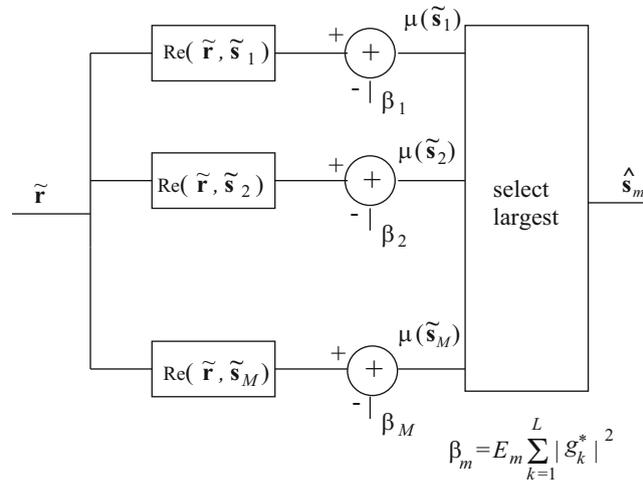


Fig. 6.4 Metric computer for maximal ratio combining

To evaluate the performance gain with MRC, the received branch vectors in (6.2) are substituted into (6.19). This yields

$$\begin{aligned}\tilde{\mathbf{r}} &= \sum_{k=1}^L g_k^* (g_k \tilde{\mathbf{s}}_m + \tilde{\mathbf{n}}_k) \\ &= \left(\sum_{k=1}^L \alpha_k^2 \right) \tilde{\mathbf{s}}_m + \sum_{k=1}^L g_k^* \tilde{\mathbf{n}}_k \\ &\equiv \alpha_M^2 \tilde{\mathbf{s}}_m + \tilde{\mathbf{n}}_M,\end{aligned}\quad (6.20)$$

where $\alpha_M^2 = \sum_{k=1}^L \alpha_k^2$, $\tilde{\mathbf{n}}_M = \sum_{k=1}^L g_k^* \tilde{\mathbf{n}}_k$ and $\alpha_k^2 = |g_k|^2$. The first term in (6.20) is the signal component with average energy $\frac{1}{2} \mathbb{E}[\alpha_M^4 \|\tilde{\mathbf{s}}_m\|^2] = \alpha_M^4 E_{\text{av}}$, where E_{av} is the average symbol energy in the signal constellation. The second term is the noise component with variance per dimension equal to

$$\sigma_{\tilde{\mathbf{n}}_M}^2 = \frac{1}{2N} \mathbb{E}[\|\tilde{\mathbf{n}}_M\|^2] = N_o \sum_{k=1}^L \alpha_k^2 = N_o \alpha_M^2. \quad (6.21)$$

The ratio of the two gives the combiner output symbol energy-to-noise ratio

$$\gamma_s^{\text{mr}} = \frac{\frac{1}{2} \mathbb{E}[\alpha_M^4 \|\tilde{\mathbf{s}}_m\|^2]}{\sigma_{\tilde{\mathbf{n}}_M}^2} = \frac{\alpha_M^2 E_{\text{av}}}{N_o} = \sum_{k=1}^L \frac{\alpha_k^2 E_{\text{av}}}{N_o} = \sum_{k=1}^L \gamma_k, \quad (6.22)$$

where $\gamma_k = \alpha_k^2 E_{\text{av}} / N_o$. Hence, γ_s^{mr} is the sum of the individual symbol energy-to-noise ratios of the L 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 as shown in Appendix A. That is,

$$p_{\gamma_s^{\text{mr}}}(x) = \frac{1}{(L-1)! (\bar{\gamma}_c)^L} x^{L-1} e^{-x/\bar{\gamma}_c}, \quad (6.23)$$

where

$$\bar{\gamma}_c = \mathbb{E}[\gamma_k] \quad k = 1, \dots, L. \quad (6.24)$$

The cdf of γ_s^{mr} is

$$F_{\gamma_s^{\text{mr}}}(x) = 1 - e^{-x/\bar{\gamma}_c} \sum_{k=0}^{L-1} \frac{1}{k!} \left(\frac{x}{\bar{\gamma}_c} \right)^k. \quad (6.25)$$

It follows from (6.22) that the average symbol energy-to-noise ratio with MRC and balanced branches is

$$\bar{\gamma}_s^{\text{mr}} = \sum_{k=1}^L \bar{\gamma}_k = \sum_{k=1}^L \bar{\gamma}_c = L \bar{\gamma}_c. \quad (6.26)$$

Figure 6.5 plots the cdf $F_{\gamma_b^{\text{mr}}}(x)$ against the normalized symbol energy-to-noise ratio $x/\bar{\gamma}_c$ in dB units. Plots of the cumulative distribution function can be used to compare the various diversity combining schemes, independent of the modulation scheme being employed. For example, with $L = 2$ and SC, the cdf in Fig. 6.2 shows that $F_{\gamma_s^{\text{mr}}}(x) = 10^{-4}$ at $x/\bar{\gamma}_c = 0.01$ ($\gamma_s^{\text{mr}} - \bar{\gamma}_c = -20$ dB). However, with $L = 2$ and MRC, Fig. 6.5 shows that $F_{\gamma_s^{\text{mr}}}(x) = 10^{-4}$ at $x/\bar{\gamma}_c = 0.01585$ ($\gamma_s^{\text{mr}} - \bar{\gamma}_c = -18$ dB). The implication is that MRC is 2 dB more power efficient than SC, with all other things being equal.

When computing the probability of bit error, our attention must be limited to coherent signaling techniques since MRC is a coherent detection technique. For example, the bit error probability with BPSK in AWGN is

$$P_b(\gamma_s^{\text{mr}}) = Q(\sqrt{2\gamma_s^{\text{mr}}}), \quad (6.27)$$

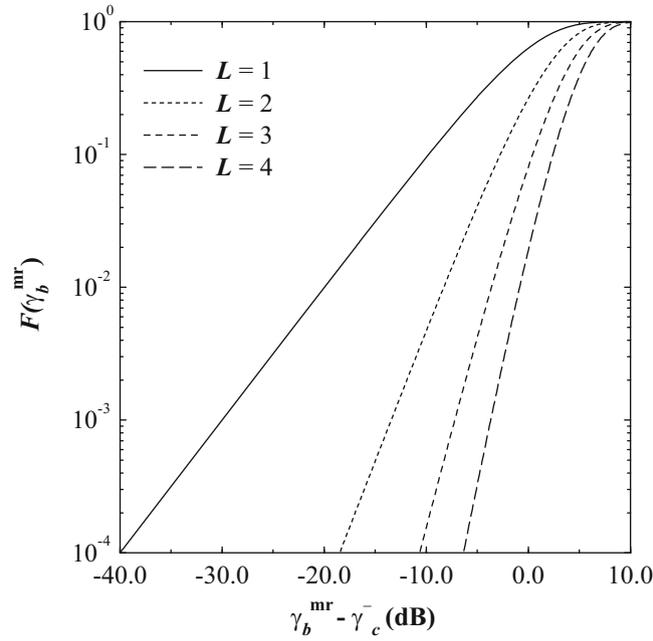


Fig. 6.5 Cdf of γ_s^{mr} for maximal ratio combining; $\bar{\gamma}_c$ is the average branch symbol energy-to-noise ratio

where γ_s^{mr} is the instantaneous *bit* energy-to-noise ratio. Hence, the bit error probability with Rayleigh fading and MRC is

$$\begin{aligned}
 P_b &= \int_0^\infty P_b(x) p_{\gamma_s^{\text{mr}}}(x) dx \\
 &= \int_0^\infty Q(\sqrt{2x}) \frac{1}{(L-1)! (\bar{\gamma}_c)^L} x^{L-1} e^{-x/\bar{\gamma}_c} dx \\
 &= \left(\frac{1-\mu}{2}\right)^L \sum_{k=0}^{L-1} \binom{L-1+k}{k} \left(\frac{1+\mu}{2}\right)^k
 \end{aligned} \tag{6.28}$$

where

$$\mu = \sqrt{\frac{\bar{\gamma}_c}{1 + \bar{\gamma}_c}}. \tag{6.29}$$

The last step follows after some algebra. The bit error probability in (6.28) is plotted in Fig. 6.6. Once again, diversity significantly improves the performance. The largest gain in performance is obtained in going from $L = 1$ to $L = 2$ branches, and diminishing returns are obtained as L increases further.

6.3.1 Granlund Combiner

The Granlund combiner is an ingenious practical method for implementing predetection maximal ratio combining, achieved by cophasing the various diversity branches, maximal ratio weighting them, and coherently combining them using analog circuits. The Granlund combiner was first proposed and studied by Granlund in 1956 [148]. The Granlund combiner, shown in Fig. 6.7, uses a regenerative process to cophase the signals received on the various diversity branches and coherently combine them.

The L input signals to the Granlund combiner are bandpass signals with a carrier frequency f_c and have the form

$$r_k(t) = \text{Re} \{ (\alpha_k \tilde{s}_m(t) e^{j\phi_k} + \tilde{n}_k(t)) e^{j2\pi f_c t} \}, \tag{6.30}$$

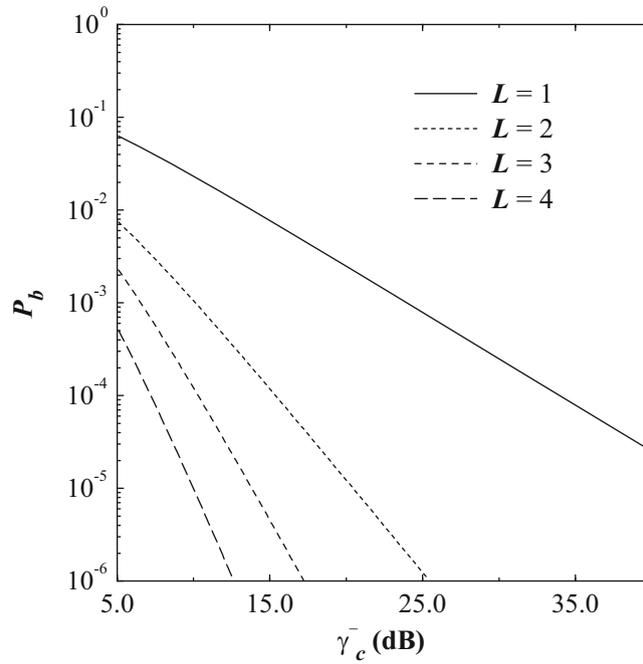


Fig. 6.6 Bit error probability for maximal ratio combining and coherent BPSK signaling

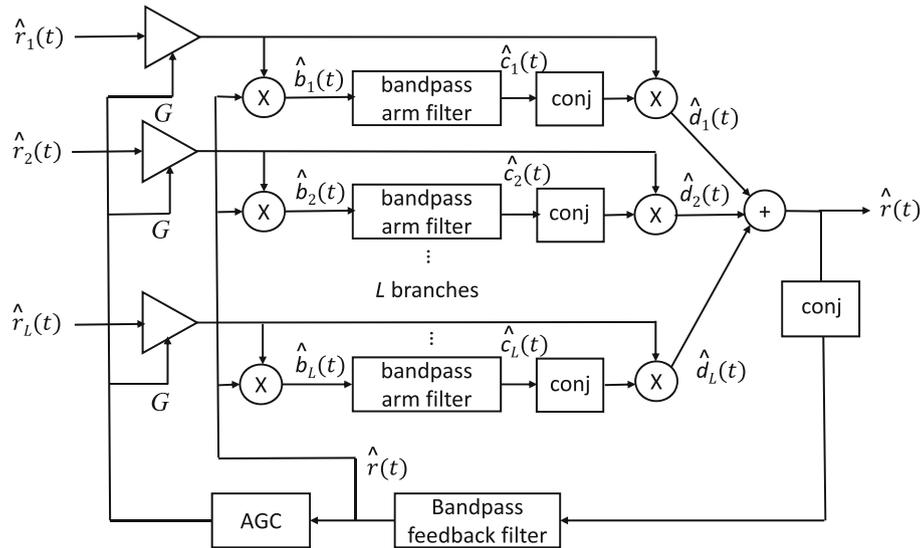


Fig. 6.7 Granlund combiner

In addition to the received complex envelope

$$\tilde{r}_k(t) = \alpha_k \tilde{s}_m(t) e^{j\phi_k} + \tilde{n}_k(t) \tag{6.31}$$

the analysis below requires the received complex exponential signals

$$\hat{r}_k(t) = (\alpha_k \tilde{s}_m(t) e^{j\phi_k} + \tilde{n}_k(t)) e^{j2\pi f_c t} \tag{6.32}$$

Moreover, for convenience of analysis, it is assumed that the signal constellation is normalized to have unit average energy $E_{av} = \frac{1}{2} E[|\tilde{s}_m(t)|^2] = 1$ and the low-pass noise in each branch $\tilde{n}(t)$ has power $\frac{1}{2} E[|\tilde{n}(t)|^2] = N_o$. Thus, the k th diversity branch has symbol energy-to-noise ratio $\gamma_k = \alpha_k^2 / N_o$. Note that the Granlund combiner requires that the noise power in each diversity branch be equal; otherwise, a significant performance degradation will occur [47].

The bandpass feedback filter is assumed to have a center frequency f_o . The bandpass arm filters are centered at $\Delta f = f_c - f_o$, where Δf is greater than f_o . It is assumed that the magnitude of the envelope of the combiner output is kept to a constant value by adjusting a gain G that is applied to all input diversity branches. The combined output signal is assumed to have the form

$$r(t) = \text{Re} \{ (K\tilde{s}_m(t)e^{j\phi_o} + \tilde{n}_o(t)) e^{j2\pi f_o t} \}. \quad (6.33)$$

For this assumption to hold, the arm filters must be narrow enough such that their outputs can be assumed to be an unmodulated tone and the feedback filter must be wide enough to pass the signal without distortion and with negligible delay.

The complex exponential signals that are input to the arm filters are

$$\begin{aligned} \hat{b}_k(t) &= \frac{1}{2}G (\hat{r}_k(t)\hat{r}_k^*(t)) \\ &= \frac{1}{2}G \left(\alpha_k K \tilde{s}_m(t) \tilde{s}_m^*(t) e^{j(\phi_k - \phi_o)} + \tilde{n}_k(t) \tilde{n}_o^*(t) \right. \\ &\quad \left. + \alpha_k \tilde{s}_m(t) \tilde{n}_o^*(t) e^{j\phi_k} + K \tilde{n}_k(t) \tilde{s}_m^*(t) e^{j\phi_o} \right) e^{j2\pi \Delta f t} \end{aligned} \quad (6.34)$$

Since the arm filters are assumed asymptotically narrow, their outputs may be approximated as a carrier with an amplitude equal to the time average of the complex exponential input signal $\hat{b}_k(t)$. If the complex exponential signals at the arm filter outputs are denoted as $\hat{c}_k(t)$, they may be written as

$$\begin{aligned} \hat{c}_k(t) &= \frac{1}{2}G \left(\alpha_k K \langle \tilde{s}_m(t) \tilde{s}_m^*(t) \rangle e^{j(\phi_k - \phi_o)} + \langle \tilde{n}_k(t) \tilde{n}_o^*(t) \rangle \right. \\ &\quad \left. + \alpha_k \langle \tilde{s}_m(t) \tilde{n}_o^*(t) \rangle e^{j\phi_k} + K \langle \tilde{n}_k(t) \tilde{s}_m^*(t) \rangle e^{j\phi_o} \right) e^{j2\pi \Delta f t} \end{aligned} \quad (6.35)$$

where $\langle \cdot \rangle$ is the time average operation. The signal and noise terms are assumed to be uncorrelated, such that

$$\langle \tilde{s}_m(t) \tilde{n}_o^*(t) \rangle = \langle \tilde{n}_k(t) \tilde{s}_m^*(t) \rangle = 0.$$

If the noise cross noise term is neglected (although this is not necessary for proper operation), then the complex exponential signals at the arm filter outputs may be written as

$$\hat{c}_k(t) = \frac{1}{2}G (\alpha_k K e^{j(\phi_k - \phi_o)}) e^{j2\pi \Delta f t}. \quad (6.36)$$

The complex exponential signals at the outputs of the second mixing process, denoted as $\hat{d}_k(t)$, are

$$\begin{aligned} \hat{d}_k(t) &= \frac{1}{2}G (\hat{r}_k(t)\hat{c}_k^*(t)) \\ &= \frac{1}{4}G \alpha_k K (\alpha_k \tilde{s}_m(t) e^{j(\phi_o)} + \alpha_k \tilde{n}_k(t) e^{j(\phi_k - \phi_o)}) e^{j2\pi f_o t} \end{aligned} \quad (6.37)$$

which is a frequency translated and phase shifted version of the input $\hat{r}_k(t)$ scaled by the weighting factor $\frac{1}{4}G \alpha_k K$. Observe that the signal term $\tilde{s}_m(t)$ goes into the k th branch of the combiner with an arbitrary phase ϕ_k and comes out with phase ϕ_o . Thus, the combined complex exponential output signal becomes

$$\hat{r}(t) = \frac{1}{4}G K e^{j2\pi f_o t} \sum_{k=1}^L (\alpha_k^2 \tilde{s}_m(t) e^{j(\phi_o)} + \alpha_k \tilde{n}_k(t) e^{j(\phi_k - \phi_o)}). \quad (6.38)$$

The corresponding complex envelope of the combiner output signal is

$$\tilde{r}(t) = \frac{1}{4}G K \sum_{k=1}^L (\alpha_k^2 \tilde{s}_m(t) e^{j(\phi_o)} + \alpha_k \tilde{n}_k(t) e^{j(\phi_k - \phi_o)}). \quad (6.39)$$

As discussed earlier, the phase rotation $e^{j(\phi_k - \phi_o)}$ has no statistical effect on the additive noise term $\tilde{n}_k(t)$. Other than the phase rotation ϕ_o and the scalar factor $\frac{1}{4}GK$, the complex envelope of the combiner output signal in (6.39) is statistically identical to the sum term

$$\sum_{k=1}^L g_k^* \tilde{r}_k(t) = \sum_{k=1}^L (\alpha_k^2 \tilde{s}_m(t) + \tilde{n}_k(t) e^{-j\phi_k})$$

in (6.18). Since the scalar factor $\frac{1}{4}GK$ affects the signal and noise term similarly, it has no effect on the combiner output symbol energy-to-noise ratio, but in the case of unequal energy symbols it must be applied similarly to the bias term

$$E_m \sum_{k=1}^L |g_k|^2$$

in (6.18) when computing the decision metrics. It also follows that the combiner output symbol energy-to-noise ratio remains the same as (6.22), i.e., $\gamma_s^{\text{mr}} = \sum_{k=1}^L \gamma_k$, where γ_k is the k th branch symbol energy-to-noise ratio.

6.4 Equal Gain Combining

Equal gain combining (EGC) is similar to MRC in that the branches are coherently combined, but is different from MRC in that the 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 $\mathbf{g} = (g_1, g_2, \dots, g_L)$ is required anyway and optimum maximum likelihood MRC might as well be used. With EGC, the receiver maximizes the metric

$$\begin{aligned} \mu(\tilde{\mathbf{s}}_m) &= \sum_{k=1}^L \text{Re}\{e^{-j\phi_k} \tilde{\mathbf{r}}_k \cdot \tilde{\mathbf{s}}_m^*\} \\ &= \sum_{k=1}^L \text{Re} \left\{ e^{-j\phi_k} \int_{-\infty}^{\infty} \tilde{r}_k(t) \tilde{s}_m^*(t) dt \right\}. \end{aligned} \quad (6.40)$$

This metric can be rewritten in the alternate form

$$\begin{aligned} \mu(\tilde{\mathbf{s}}_m) &= \text{Re} \left\{ \sum_{k=1}^L e^{-j\phi_k} \tilde{\mathbf{r}}_k \cdot \tilde{\mathbf{s}}_m^* \right\} \\ &= \int_0^T \text{Re} \left\{ \left(\sum_{k=1}^L e^{-j\phi_k} \tilde{r}_k(t) \right) \tilde{s}_m^*(t) \right\} dt. \end{aligned} \quad (6.41)$$

It is apparent that the combiner in Fig. 6.1 generates the sum

$$\tilde{\mathbf{r}} = \sum_{k=1}^L e^{-j\phi_k} \tilde{\mathbf{r}}_k \quad (6.42)$$

with EGC. The vector $\tilde{\mathbf{r}}$ is then applied to the metric computer shown in Fig. 6.4 with $\beta_m = 0$, $m = 1, \dots, L$. The reason for setting $\beta_m = 0$ comes from the assumption of equal energy signals.

To evaluate the performance gain with EGC, the received branch vectors in (6.2) are substituted into (6.42). This yields

$$\begin{aligned} \tilde{\mathbf{r}} &= \sum_{k=1}^L e^{-j\phi_k} (g_k \tilde{\mathbf{s}}_m + \tilde{\mathbf{n}}_k) \\ &= \left(\sum_{k=1}^L \alpha_k \right) \tilde{\mathbf{s}}_m + \sum_{k=1}^L e^{-j\phi_k} \tilde{\mathbf{n}}_k \\ &\equiv \alpha_E \tilde{\mathbf{s}}_m + \tilde{\mathbf{n}}_E, \end{aligned} \quad (6.43)$$

where $\alpha_E = \sum_{k=1}^L \alpha_k$, $\tilde{\mathbf{n}}_E = \sum_{k=1}^L e^{-j\phi_k} \tilde{\mathbf{n}}_k$ and $\alpha_k = |g_k|$. The first term in (6.43) is the signal component with average energy $\frac{1}{2}E[\alpha_E^2 \|\tilde{\mathbf{s}}_m\|^2] = \alpha_E^2 E_{av}$, where E_{av} is the average symbol energy in the signal constellation. The second term is the noise component with variance per dimension equal to

$$\sigma_{\tilde{\mathbf{n}}_E}^2 = \frac{1}{2N} E[\|\tilde{\mathbf{n}}_E\|^2] = LN_o. \quad (6.44)$$

The ratio of the two gives the symbol energy-to-noise ratio

$$\gamma_s^{\text{eg}} = \frac{\alpha_E^2 E_{av}}{LN_o}. \quad (6.45)$$

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

$$F_{\gamma_s^{\text{eg}}}(x) = 1 - e^{-2x/\bar{\gamma}_c} - \sqrt{\pi \frac{x}{\bar{\gamma}_c}} e^{-x/\bar{\gamma}_c} \left(1 - 2Q \left(\sqrt{2 \frac{x}{\bar{\gamma}_c}} \right) \right). \quad (6.46)$$

Differentiating the above expression yields the pdf

$$p_{\gamma_s^{\text{eg}}}(x) = \frac{1}{\bar{\gamma}_c} e^{-2x/\bar{\gamma}_c} - \sqrt{\pi} e^{-x/\bar{\gamma}_c} \left(\frac{1}{2\sqrt{x\bar{\gamma}_c}} - \frac{1}{\bar{\gamma}_c} \sqrt{\frac{x}{\bar{\gamma}_c}} \right) \left(1 - 2Q \left(\sqrt{2 \frac{x}{\bar{\gamma}_c}} \right) \right). \quad (6.47)$$

The average symbol energy-to-noise ratio with EGC is

$$\begin{aligned} \bar{\gamma}_s^{\text{eg}} &= \frac{E_{av}}{LN_o} E \left[\left(\sum_{k=1}^L \alpha_k \right)^2 \right] \\ &= \frac{E_{av}}{LN_o} \sum_{j=1}^L \sum_{\ell=1}^L E[\alpha_k \alpha_\ell]. \end{aligned} \quad (6.48)$$

With Rayleigh fading, $E[\alpha_k^2] = 2b_0$ and $E[\alpha_k] = \sqrt{\pi b_0/2}$. Furthermore, if the branches experience uncorrelated fading, then $E[\alpha_k \alpha_\ell] = E[\alpha_k]E[\alpha_\ell]$ for $k \neq \ell$. Hence,

$$\begin{aligned} \bar{\gamma}_s^{\text{eg}} &= \frac{E_{av}}{LN_o} \left(2Lb_0 + L(L-1) \frac{\pi b_0}{2} \right) \\ &= \frac{2b_0 E_{av}}{N_o} \left(1 + (L-1) \frac{\pi}{4} \right) \\ &= \bar{\gamma}_c \left(1 + (L-1) \frac{\pi}{4} \right). \end{aligned} \quad (6.49)$$

The bit error probability with 2-branch EGC can be obtained by using the pdf in (6.47). Once again, EGC is a coherent detection technique so our attention must be restricted to coherent signaling techniques. For example, with BPSK the bit error probability is (Problem 6.12)

$$\begin{aligned} P_b &= \int_0^\infty P_b(x) p_{\gamma_s^{\text{eg}}}(x) dx \\ &= \frac{1}{2} \left(1 - \sqrt{1 - \mu^2} \right), \end{aligned} \quad (6.50)$$

where

$$\mu = \frac{1}{1 + \bar{\gamma}_c}. \quad (6.51)$$

6.5 Switched Combining

A switched combiner scans through the diversity branches until it finds one that has a bit energy-to-noise ratio exceeding a specified threshold. This diversity branch is selected and used until the bit energy-to-noise ratio again drops below the threshold. When this happens another diversity branch is chosen which has a bit energy-to-noise ratio exceeding the threshold. The advantage of switched combining is that only one RF chain is needed. There are several variations of switched diversity. Here, two-branch switch and stay combining (SSC) is analyzed. With SSC, the receiver switches to, and stays with, the alternate branch when the bit energy-to-noise ratio drops below a specified threshold. It does this regardless of whether or not the bit energy-to-noise ratio with the alternate branch is above or below the threshold.

Let the bit 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

$$\begin{aligned} q &= \text{P}[\gamma_i < T] \\ &= 1 - e^{-T/\bar{\gamma}_c}, \quad i = 1, 2. \end{aligned} \quad (6.52)$$

Likewise, the probability that γ_i is less than S is

$$p = 1 - e^{-S/\bar{\gamma}_c}, \quad i = 1, 2. \quad (6.53)$$

Let γ_s^{sw} denote the symbol energy-to-noise ratio at the output of the switched combiner. Then

$$\text{P}[\gamma_s^{\text{sw}} \leq S] = \text{P}[\gamma_s^{\text{sw}} \leq S | \gamma_s^{\text{sw}} = \gamma_1 \cup \gamma_s^{\text{sw}} \leq S | \gamma_s^{\text{sw}} = \gamma_2]. \quad (6.54)$$

Since γ_1 is statistically identical to γ_2 , it may be assumed without loss of generality that branch 1 is currently in use. It follows that

$$\text{P}[\gamma_s^{\text{sw}} \leq S] = \begin{cases} \text{P}[\gamma_1 \leq T \cap \gamma_2 \leq S], & S < T \\ \text{P}[T \leq \gamma_1 \leq S \cup \gamma_1 \leq T \cap \gamma_2 \leq S], & S \geq T \end{cases}. \quad (6.55)$$

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_s^{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

$$\text{P}[\gamma_1 \leq T \cap \gamma_2 \leq S] = qp \quad (6.56)$$

and

$$\text{P}[T \leq \gamma_1 \leq S \cup (\gamma_1 \leq T \cap \gamma_2 \leq S)] = p - q + qp. \quad (6.57)$$

Therefore,

$$\text{P}[\gamma_s^{\text{sw}} \leq S] = \begin{cases} qp & S < T \\ p - q + qp & S \geq T \end{cases}. \quad (6.58)$$

Figure 6.8 plots the cdf $F_{\gamma_s^{\text{sw}}}(x)$ against the normalized symbol energy-to-noise ratio $x/\bar{\gamma}_c$, 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 bit energy-to-noise ratio that the communication link can tolerate and still provide an acceptable quality of service. Finally, the optimum threshold, $T = R\bar{\gamma}_c$, depends on $\bar{\gamma}_c$.

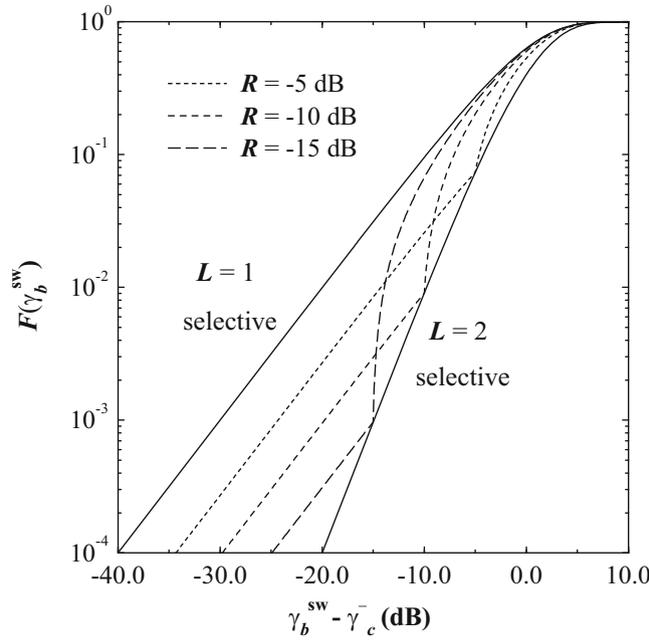


Fig. 6.8 Cdf of γ_{sw} for 2-branch switched diversity for several values of the normalized threshold $R = 10 \log_{10}(T/\bar{\gamma}_c)$; $\bar{\gamma}_c$ is the average branch bit energy-to-noise ratio

The probability of bit error for SSC can be computed from the pdf of the bit energy-to-noise ratio at the output of the switched combiner. The pdf of γ_s^{sw} is

$$p_{\gamma_s^{sw}}(x) = \begin{cases} q \frac{1}{\bar{\gamma}_c} e^{-x/\bar{\gamma}_c}, & x < T \\ (1+q) \frac{1}{\bar{\gamma}_c} e^{-x/\bar{\gamma}_c}, & x \geq T \end{cases} \quad (6.59)$$

If DPSK is used with differential detection, for example, then $P_b(\gamma_s^{sw}) = \frac{1}{2} e^{-\gamma_s^{sw}}$, and the probability of bit error is

$$\begin{aligned} P_b &= \int_0^{\infty} P_b(x) p_{\gamma_s^{sw}}(x) dx \\ &= \frac{1}{2(1+\bar{\gamma}_c)} (q + (1-q)e^{-T}), \end{aligned} \quad (6.60)$$

where $\bar{\gamma}_c$ is the average branch bit energy-to-noise ratio. The above expression is plotted in Fig. 6.9 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. 6.10. For some systems, it may be desirable to keep q as small as possible to minimize the number of switches.

6.6 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. 6.11. The structure of the individual differential detectors depends on the type of modulation that is being used. For DPSK, the detector structure is as shown in Fig. 5.18, while for $\pi/4$ -QPSK the detector is as shown in Fig. 5.19. In the latter case, the U and V branches in Fig. 5.19 are combined separately. For DPSK the decision variable at the output of the combiner at epoch n is, from (5.149) and Fig. 6.11,

$$U_n = \sum_{k=1}^L U_{n,k} = \frac{1}{2} \sum_{k=1}^L (Z_{n,k} Z_{d,n,k}^* + Z_{n,k}^* Z_{d,n,k}). \quad (6.61)$$

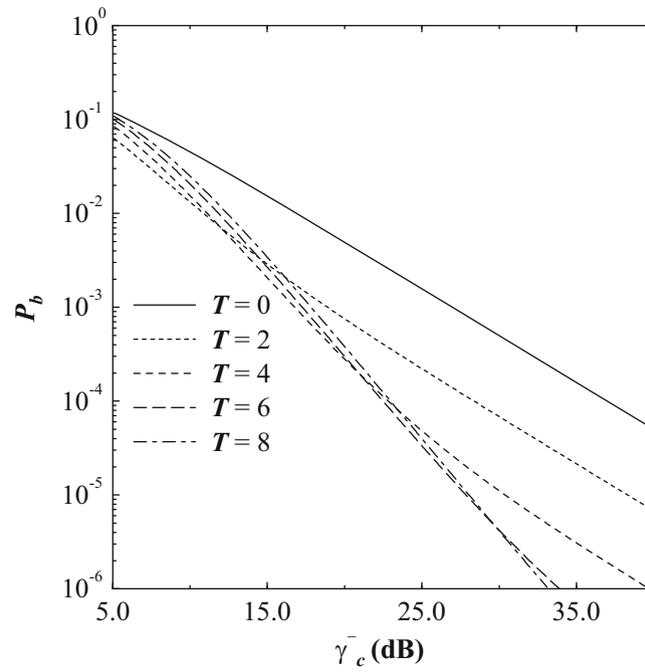


Fig. 6.9 Bit error probability for 2-branch switched combining and differentially detected DPSK

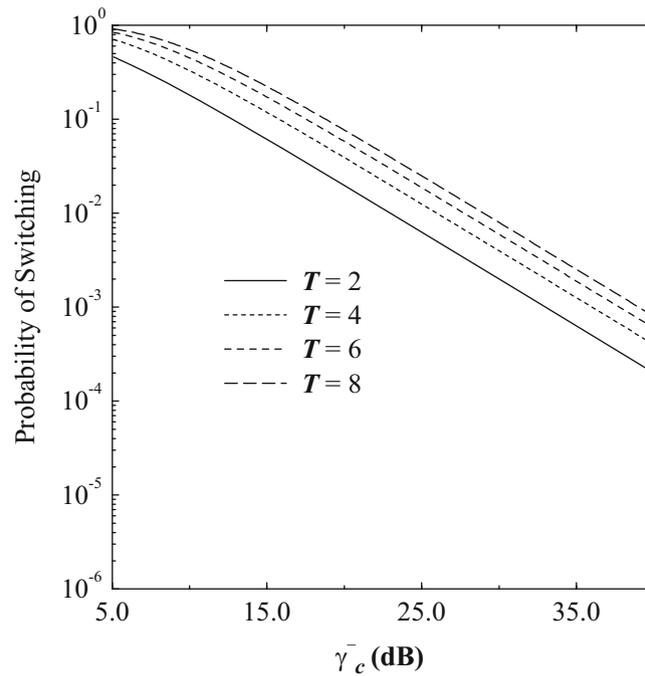


Fig. 6.10 Probability of switching for two-branch switched combining

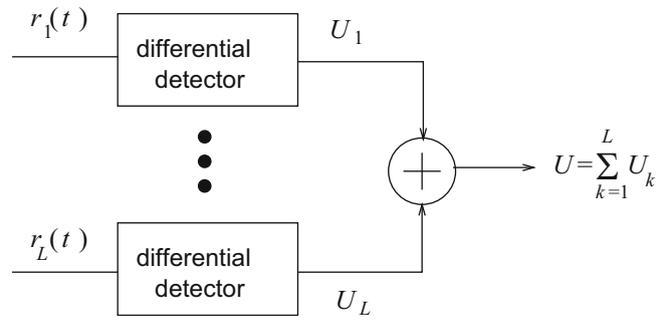


Fig. 6.11 Differential detection with postdetection equal gain combining

By using characteristic functions, it can be shown that the decision variable U_n has the difference form $U_n = W_n - Y_n$, where W_n and Y_n are independent non-central and central chi-square random variables, respectively, each with $2L$ degrees of freedom, i.e.,

$$f_{W_n}(w) = \frac{1}{2E_h N_o} \left(\frac{w}{s^2}\right)^{\frac{L-1}{2}} \exp\left\{-\frac{(s^2+w)}{2E_h N_o}\right\} I_{L-1}\left(\sqrt{w}\frac{s}{E_h N_o}\right) \quad (6.62)$$

$$f_{Y_n}(y) = \left(\frac{1}{2E_h N_o}\right)^L \frac{1}{(L-1)!} y^{L-1} \exp\left\{-\frac{y}{2E_h N_o}\right\}, \quad (6.63)$$

where

$$s^2 = 4E_h \sum_{k=1}^L \alpha_k^2 \quad (6.64)$$

is the non-centrality parameter, and $I_n(x)$ is the n th-order modified Bessel function of the first kind, defined by

$$I_n(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos(\theta)} \cos(n\theta) d\theta. \quad (6.65)$$

After some algebraic detail, the probability of error can be expressed in the closed form [272]

$$P_b(\gamma_t) = \frac{1}{2^{2L-1}} e^{-\gamma_t} \sum_{k=0}^{L-1} b_k \gamma_t^k, \quad (6.66)$$

where

$$b_k = \frac{1}{k!} \sum_{n=0}^{L-1-k} \binom{2L-1}{n} \quad (6.67)$$

and

$$\gamma_t = \sum_{k=1}^L \gamma_k. \quad (6.68)$$

The parameter γ_t has the central chi-square distribution in (6.23), and averaging $P_b(\gamma_t)$ over this distribution gives the result

$$P_b = \frac{1}{2^{2L-1} (L-1)! (1 + \bar{\gamma}_c)^L} \sum_{k=1}^{L-1} b_k (L-1+k)! \left(\frac{\bar{\gamma}_c}{1 + \bar{\gamma}_c}\right)^k. \quad (6.69)$$

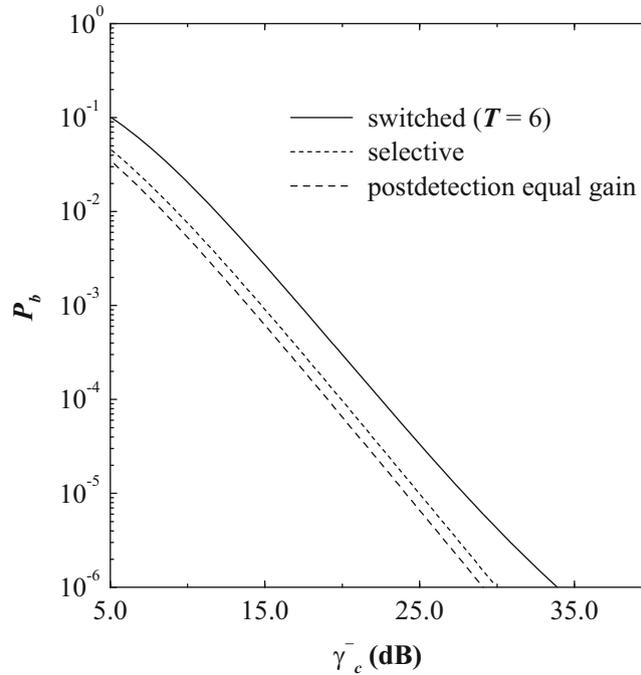


Fig. 6.12 Comparison of 2-branch diversity combining techniques for differentially detected binary DPSK

This can be manipulated in the same form as (6.28) with

$$\mu = \frac{\bar{\gamma}_c}{1 + \bar{\gamma}_c}. \quad (6.70)$$

Switched and postdetection equal gain diversity combining are compared in Fig. 6.12 for differentially detected binary DPSK signals. It is apparent that differentially detected DPSK with postdetection equal gain combining has the best performance, followed by SC and SSC combining. Once again, it does not make sense to use maximal ratio combining or equal gain combining with differential detection, since maximal ratio combining and equal gain combining are coherent combining techniques.

6.7 Non-coherent Square-Law Combining

Square-law combining is a diversity combining technique that is used with non-coherent detection. As such, square-law combining is only useful for orthogonal modulation schemes, including M -ary FSK modulation and binary orthogonal codes as discussed in Sect. 4.5.

Section 5.9 showed that the non-coherent detector calculates the M decision variables X_m^2 , $m = 1, \dots, M$ and chooses the message waveform corresponding to the largest decision variable (assuming equal energy messages). When diversity is employed a separate square-law detector, as shown in Fig. 5.20, is used on each diversity branch. This will yield the set of outputs $X_{m,k}$, $m = 1, \dots, M$, $k = 1, \dots, L$. The square-law combiner then computes the following set of decision variables

$$U_m = \sum_{k=1}^L X_{m,k}^2, \quad m = 1, \dots, M, \quad (6.71)$$

and a decision is made in favor of the message waveform corresponding to the largest U_m .

If waveform $\tilde{s}_1(t)$ was transmitted, then

$$\begin{aligned} U_1 &= \sum_{k=1}^L |\tilde{\mathbf{r}}_k \cdot \tilde{\mathbf{s}}_1^*|^2 \\ &= \sum_{k=1}^L |2Eg_k + \sqrt{2E}\tilde{n}_{k,1}|^2 \\ &= \sum_{k=1}^L \left| 2E\alpha_k \cos(\phi_k) + \sqrt{2E}\tilde{n}_{I,k,1} + j \left(2E\alpha_k \sin(\phi_k) + \sqrt{2E}\tilde{n}_{Q,k,1} \right) \right|^2 \end{aligned}$$

and

$$\begin{aligned} U_m &= \sum_{k=1}^L |\tilde{\mathbf{r}}_k \cdot \tilde{\mathbf{s}}_m^*|^2 \\ &= \sum_{k=1}^L |\sqrt{2E}\tilde{n}_{k,m}|^2 \\ &= \sum_{k=1}^L \left| \sqrt{2E}\tilde{n}_{I,k,m} + j\sqrt{2E}\tilde{n}_{Q,k,m} \right|^2, \quad m = 2, \dots, M. \end{aligned}$$

From Appendix A, the random variables U_i , $i = 1, \dots, M$ have a central chi-square distribution with $2L$ degrees of freedom. It follows that

$$p_{U_1}(u_1) = \frac{1}{(2\sigma_1^2)^L(L-1)!} u_1^{L-1} \exp\left\{-\frac{u_1}{2\sigma_1^2}\right\}, \quad u_1 \geq 0. \quad (6.72)$$

where

$$\begin{aligned} \sigma_1^2 &= \frac{1}{2} \mathbb{E} \left[\left| 2Eg_k + \sqrt{2E}\tilde{n}_{k,1} \right|^2 \right] \\ &= 2EN_o \left(1 + \frac{\mathbb{E}[\alpha_k^2]E}{N_o} \right) \\ &= 2EN_o (1 + \bar{\gamma}_c). \end{aligned} \quad (6.73)$$

Likewise,

$$p_{U_m}(u_m) = \frac{1}{(2\sigma_2^2)^L(L-1)!} u_m^{L-1} \exp\left\{-\frac{u_m}{2\sigma_2^2}\right\}, \quad u_m \geq 0. \quad (6.74)$$

where

$$\sigma_2^2 = 2EN_o, \quad m = 2, \dots, M. \quad (6.75)$$

To evaluate the probability of error with square-law combining, proceed along the lines of Sect. 5.9.1. First assume without loss of generality that waveform $\tilde{s}_1(t)$ is sent. The receiver will make a correct decision if

$$U_i < U_1, \quad \forall i \neq 1. \quad (6.76)$$

Hence, the probability of correct symbol decision is

$$\begin{aligned} \mathbb{P}[c] &= \mathbb{P}[U_2 < U_1, U_3 < U_1, \dots, U_M < U_1] \\ &= \int_0^\infty (\mathbb{P}[U_2 < u_1])^{M-1} p_{U_1}(u_1) du_1. \end{aligned} \quad (6.77)$$

where the second line follows from the fact that the U_m , $m = 2, \dots, M$ are independent and identically distributed. From the cdf of the central chi-square distribution in (A.65),

$$P[U_2 < u_1] = 1 - \exp\left\{-\frac{u_1}{2\sigma_2^2}\right\} \sum_{k=0}^{L-1} \frac{1}{k!} \left(\frac{u_1}{2\sigma_2^2}\right)^k, \quad u_1 \geq 0. \quad (6.78)$$

The $(M-1)$ th power of this probability is then used in (6.77) to obtain the probability of correct decision, and subtracting this result from unity gives the probability of symbol error. This gives

$$\begin{aligned} P_M &= 1 - \int_0^\infty \left(1 - \exp\left\{-\frac{u_1}{2\sigma_2^2}\right\} \sum_{k=0}^{L-1} \frac{1}{k!} \left(\frac{u_1}{2\sigma_2^2}\right)^k\right)^{M-1} \frac{1}{(2\sigma_1^2)^L (L-1)!} u_1^{L-1} \exp\left\{-\frac{u_1}{2\sigma_1^2}\right\} du_1 \\ &= 1 - \int_0^\infty \left(1 - e^{-y} \sum_{k=0}^{L-1} \frac{y^k}{k!}\right)^{M-1} \frac{1}{(1 + \bar{\gamma}_c)^L (L-1)!} y^{L-1} \exp\left\{-\frac{y}{1 + \bar{\gamma}_c}\right\} dy. \end{aligned} \quad (6.79)$$

The above expression can be expressed in closed form by using a multinomial expansion. However, this is mathematically cumbersome for all but small values of M and L , and it may just be easier to evaluate (6.79) numerically. When $L = 1$ (no diversity), (6.79) reduces to the simple form

$$P_M = \sum_{m=1}^{M-1} \frac{(-1)^{m+1} \binom{M-1}{m}}{1 + m + m\bar{\gamma}_c}. \quad (6.80)$$

Finally, from Sect. 5.3.4, the probability of bit error is

$$P_b = \frac{2^{k-1}}{2^k - 1} P_M = \frac{M}{2(M-1)} P_M, \quad (6.81)$$

and the average received bit energy-to-noise ratio is

$$\bar{\gamma}_b = L\bar{\gamma}_c / \log_2 M. \quad (6.82)$$

Using these, (6.80) agrees with the result in (5.191).

For the case of binary non-coherent orthogonal FSK ($M = 2$), it can be shown that (6.79) reduces to the form in (6.28) where the parameter μ is defined as

$$\mu = \frac{\bar{\gamma}_c}{2 + \bar{\gamma}_c}. \quad (6.83)$$

When $\bar{\gamma}_c \gg 1$, the probability of bit error for binary orthogonal FSK with L th order non-coherent square-law combining can be approximated by

$$P_b \approx \left(\frac{1}{\bar{\gamma}_c}\right)^L \binom{2L-1}{L}. \quad (6.84)$$

Figure 6.13 plots the probability of bit error against the bit energy-to-noise ratio for various values of M and L . Note that the performance improves with increasing M and increasing L . A significant improvement is realized by increasing L , and the performance gain realized by increasing M is relatively small when L is small. Since an increase in M implies a decrease in bandwidth efficiency for M -ary orthogonal signals, it is more efficient in terms of bandwidth efficiency to increase L .

6.8 Optimum Combining

Maximal ratio combining (MRC) is the optimal combining method in a maximum likelihood sense when the additive impairment is AWGN. However, when the additive channel impairment is dominated by co-channel interference, it is better to use optimum combining (OC) which is designed to maximize the signal-to-interference-plus-noise ratio at the output of

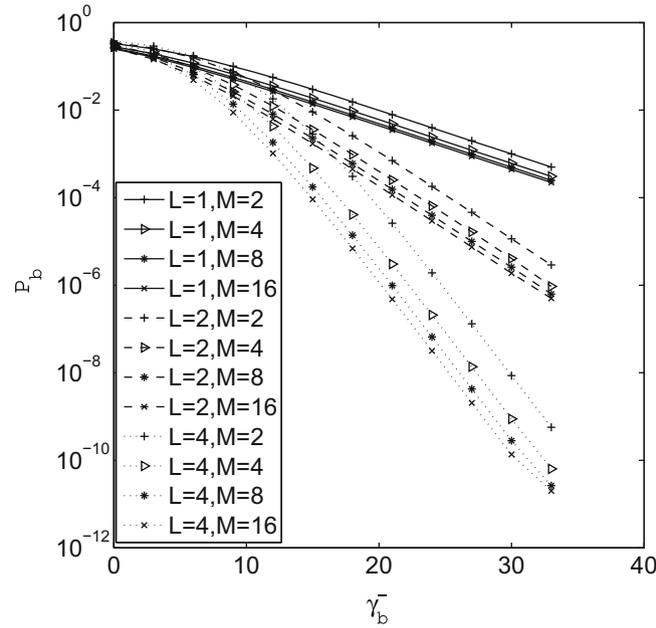


Fig. 6.13 Performance of M -ary orthogonal signals with square-law combining

the combiner. OC uses spatial diversity not only to combat fading of the desired signal, as is the case with MRC, but also to reduce the relative power of the interfering signals at the receiver. This is achieved by exploiting the correlation of the interference across the multiple receiver antenna elements. By combining the signals that are received by multiple antennas, OC can suppress the interference and improve the output signal-to-interference-plus-noise ratio by several decibels. OC was first introduced by Baird and Zahm [32] for the case of a single interferer and later extended to the case of multiple interferers and applied to cellular frequency reuse systems by Winters [352, 353]. In this section, the optimal combiner is described, the distribution of the output signal-to-interference-plus-noise ratio is derived, and the bit error rate of PSK signals when the desired and interfering signals are subject to slow flat Rayleigh fading is evaluated.

Consider a situation where a desired signal is received in the presence of K co-channel interferers. In the following treatment, the signal space dimensionality is assumed to be unity, i.e., $N = 1$, so that the signal vectors are complex-valued scalars chosen from an appropriate constellation such as M -QAM.

The received signal scalars at the L receiver antennas are equal to

$$\tilde{r}_k = g_{k,0}\tilde{s}_0 + \sum_{i=1}^K g_{k,i}\tilde{s}_i + \tilde{n}_k, \quad k = 1, \dots, L, \quad (6.85)$$

where \tilde{s}_0 , \tilde{s}_i , and \tilde{n}_k are the desired signal vector, i th interfering signal vector, and noise vector, respectively, and K is the number of interferers.

The L received signal scalars can be stacked in a column to yield the $L \times 1$ received vector

$$\tilde{\mathbf{r}}_t = \mathbf{g}_0\tilde{s}_0 + \sum_{i=1}^K \mathbf{g}_i\tilde{s}_i + \tilde{\mathbf{n}}, \quad (6.86)$$

where

$$\begin{aligned} \tilde{\mathbf{r}}_t &= (\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_L)^T \\ \mathbf{g}_i &= (g_{i,1}, g_{i,2}, \dots, g_{i,L})^T \\ \tilde{\mathbf{n}} &= (\tilde{n}_1, \tilde{n}_2, \dots, \tilde{n}_L)^T. \end{aligned}$$

The $L \times L$ received desired-signal-plus-interference-plus noise correlation matrix is given by

$$\Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t} = \frac{1}{2} E_{\tilde{s}_0, \tilde{s}_i, \tilde{\mathbf{n}}} \left[\left(\mathbf{g}_0 \tilde{s}_0 + \sum_{i=1}^K \mathbf{g}_i \tilde{s}_i + \tilde{\mathbf{n}} \right) \left(\mathbf{g}_0 \tilde{s}_0 + \sum_{i=1}^K \mathbf{g}_i \tilde{s}_i + \tilde{\mathbf{n}} \right)^H \right], \quad (6.87)$$

where $(\cdot)^H$ denotes complex conjugate transpose. Likewise, the received interference-plus-noise correlation matrix is given by

$$\Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t} = \frac{1}{2} E_{\tilde{s}_i, \tilde{\mathbf{n}}} \left[\left(\sum_{i=1}^K \mathbf{g}_i \tilde{s}_i + \tilde{\mathbf{n}} \right) \left(\sum_{i=1}^K \mathbf{g}_i \tilde{s}_i + \tilde{\mathbf{n}} \right)^H \right]. \quad (6.88)$$

Note that the expectations in (6.87) and (6.88) are taken over a period that is much less than the channel coherence time, i.e., several modulated symbol durations. If the desired signal, interfering signal, and noise vectors are mutually uncorrelated, then (6.87) and (6.88) reduce to

$$\Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t} = \mathbf{g}_0 \mathbf{g}_0^H E_{\text{av}} + \sum_{i=1}^K \mathbf{g}_i \mathbf{g}_i^H E_{\text{av}}^i + N_o \mathbf{I}, \quad (6.89)$$

and

$$\Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t} = \sum_{i=1}^K \mathbf{g}_i \mathbf{g}_i^H E_{\text{av}}^i + N_o \mathbf{I}, \quad (6.90)$$

respectively, where \mathbf{I} is the $L \times L$ identity matrix and E_{av}^i is the average energy in the i th interfering signal. It is important to note that the matrices $\Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t}$ and $\Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t}$ will vary at the channel fading rate.

To mitigate the interference, the received signals scalars \tilde{r}_k , $k = 1, 2, \dots, L$ are multiplied by controllable weights w_k and summed together, i.e., the combiner output is

$$\tilde{r} = \sum_{k=1}^L w_k \tilde{r}_k = \mathbf{w}^T \tilde{\mathbf{r}}_t, \quad (6.91)$$

where $\mathbf{w} = (w_1, w_2, \dots, w_L)^T$ is the weight vector. Several approaches can be taken to find the optimal weight vector \mathbf{w} . One approach is to minimize the mean square error

$$\begin{aligned} J &= E [\|\tilde{r} - \tilde{s}_0\|^2] \\ &= E [\|\mathbf{w}^T \tilde{\mathbf{r}}_t - \tilde{s}_0\|^2] \\ &= 2\mathbf{w}^T \Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t} \mathbf{w}^* - 2\text{Re} \{ \Phi_{\tilde{s}_0 \tilde{\mathbf{r}}_t} \mathbf{w}^* \} - 2E_{\text{av}}, \end{aligned}$$

where $\Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t}$ is defined in (6.89) and

$$\Phi_{\tilde{s}_0 \tilde{\mathbf{r}}_t} = E [\tilde{s}_0 \tilde{\mathbf{r}}_t^H] = 2E_{\text{av}} \mathbf{g}_0^H. \quad (6.92)$$

The weight vector that minimizes the mean square error can be obtained by setting the gradient $\nabla_{\mathbf{w}} J$ to zero. This gives the minimum mean square error (MMSE) solution

$$\nabla_{\mathbf{w}} J = \left(\frac{\partial J}{\partial w_1}, \dots, \frac{\partial J}{\partial w_L} \right) = 2\mathbf{w}^T \Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t} - 2\Phi_{\tilde{s}_0 \tilde{\mathbf{r}}_t} = 0. \quad (6.93)$$

The solution is

$$\mathbf{w}_{\text{opt}} = \Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t}^{-1} \Phi_{\tilde{s}_0 \tilde{\mathbf{r}}_t}^T = 2E_{\text{av}} \Phi_{\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t}^{-1} \mathbf{g}_0^*, \quad (6.94)$$

where the fact that $\Phi_{\tilde{\mathbf{r}}_i}^T = 2\mathbf{g}_0^* E_{\text{av}}$ was used. Since $\Phi_{\tilde{\mathbf{r}}_i} = \mathbf{g}_0 \mathbf{g}_0^H E_{\text{av}} + \Phi_{\tilde{\mathbf{r}}_i}$, it follows that

$$\begin{aligned} \mathbf{w}_{\text{opt}} &= 2E_{\text{av}} (\Phi_{\tilde{\mathbf{r}}_i} + \mathbf{g}_0 \mathbf{g}_0^H E_{\text{av}})^{-1} \mathbf{g}_0^* \\ &= 2E_{\text{av}} (\Phi_{\tilde{\mathbf{r}}_i} + \mathbf{g}_0^* \mathbf{g}_0^T E_{\text{av}})^{-1} \mathbf{g}_0^*. \end{aligned} \quad (6.95)$$

Next, a variation of the matrix inversion lemma is applied

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^H)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^H\mathbf{A}^{-1}}{1 + \mathbf{v}^H\mathbf{A}^{-1}\mathbf{u}} \quad (6.96)$$

to (6.95) resulting in

$$\begin{aligned} \mathbf{w}_{\text{opt}} &= 2E_{\text{av}} \left(\Phi_{\tilde{\mathbf{r}}_i}^{-1} - \frac{E_{\text{av}} \Phi_{\tilde{\mathbf{r}}_i}^{-1} \mathbf{g}_0^* \mathbf{g}_0^T \Phi_{\tilde{\mathbf{r}}_i}^{-1}}{1 + E_{\text{av}} \mathbf{g}_0^T \Phi_{\tilde{\mathbf{r}}_i}^{-1} \mathbf{g}_0^*} \right) \mathbf{g}_0^* \\ &= 2E_{\text{av}} \left(\frac{1}{1 + E_{\text{av}} \mathbf{g}_0^T \Phi_{\tilde{\mathbf{r}}_i}^{-1} \mathbf{g}_0^*} \right) \cdot \Phi_{\tilde{\mathbf{r}}_i}^{-1} \mathbf{g}_0^* \\ &= C \cdot \Phi_{\tilde{\mathbf{r}}_i}^{-1} \mathbf{g}_0^*, \end{aligned} \quad (6.97)$$

where $C = 2E_{\text{av}} / (1 + E_{\text{av}} \mathbf{g}_0^T \Phi_{\tilde{\mathbf{r}}_i}^{-1} \mathbf{g}_0^*)$ is a scalar.

Another criterion for optimizing the weight vector is to maximize the instantaneous signal-to-interference-plus-noise ratio (SINR) at the output of the combiner

$$\omega = \frac{\mathbf{w}^T \mathbf{g}_0 \mathbf{g}_0^H E_{\text{av}} \mathbf{w}^*}{\mathbf{w}^T \Phi_{\tilde{\mathbf{r}}_i}^{-1} \mathbf{w}^*}. \quad (6.98)$$

Solving for the optimum weight vector gives

$$\mathbf{w}_{\text{opt}} = B \cdot \Phi_{\tilde{\mathbf{r}}_i}^{-1} \mathbf{g}_0^*, \quad (6.99)$$

where B is an arbitrary constant. Hence, the *maximum* instantaneous output SINR is

$$\omega = E_{\text{av}} \mathbf{g}_0^H \Phi_{\tilde{\mathbf{r}}_i}^{-1} \mathbf{g}_0. \quad (6.100)$$

Note that the maximum instantaneous output SINR does not depend on the choice of the scalar B . Therefore, the MMSE weight vector in (6.97) also maximizes the instantaneous output SINR. Finally, when no interference is present, $\Phi_{\tilde{\mathbf{r}}_i} = N_o \mathbf{I}$ and the optimal weight vector becomes

$$\mathbf{w}_{\text{opt}} = \frac{\mathbf{g}_0^*}{N_o}, \quad (6.101)$$

so that the combiner output is

$$\tilde{\mathbf{r}} = \sum_{k=1}^L \frac{\mathcal{S}_{0,k}^*}{N_o} \tilde{\mathbf{r}}_k. \quad (6.102)$$

From (6.101) and (6.102), OC reduces to MRC when no co-channel interference is present.

6.8.1 Optimum Combining Performance

In typical land mobile radio environments, there can be several interfering signals whose power level is close to that of the desired signal and numerous lower power interfering signals. The number of interfering signals can be much greater than the number of receiver antenna elements. In this case, the array output SINR may not be changed significantly. However, even a small (few dB) increase in output SINR can result in a large capacity gain. Thus, the array only needs to suppress the dominant interferers so that their power level is below the sum power level of the other interferers.

In a non-fading environment, the array cannot resolve two closely spaced transmitters because the phase differences of the desired and interfering signals across the antenna elements are almost the same. However, for land mobile radio applications, the receiver antenna elements can be separated enough so that the phases at each antenna element are independent. For 2-D isotropic scattering around the receiver antennas, a half-wavelength separation between the receiver antenna elements is sufficient. Likewise, if there is 2-D isotropic scattering around the transmit antennas, then a half-wavelength spatial separation is sufficient to ensure independent phases at the receiver antenna elements, e.g., at a base station. Thus the resolution of the signals from two different transmitters in a land mobile radio environment does not depend on the spatial separation of the transmitters. Instead for all locations there is a small probability that the receiver array cannot resolve the two signals. This occurs when the phase differences across all the receiver antenna elements is nearly the same for both the desired and interfering signals. However, since the phase differences between antenna elements is independent, the probability that the phase difference is nearly the same across all antenna elements decreases significantly as the number of receiver antenna elements increases, and becomes negligible for only a few antennas.

The performance of OC is now considered in terms of the output SINR distribution and the bit error rate for coherent BPSK modulation. Comparisons are made with MRC. It is assumed that all signals are subject to slow flat Rayleigh fading. The performance with optimum combining is quite complicated, since each interferer impacts the performance of the optimum combiner. In this section, the performance with a single dominant interferer is evaluated under the assumption that the remaining co-channel interferers can be combined together and treated as additional lumped interference that is uncorrelated between the receiver antenna elements. Such lumped interference can be treated as additional AWGN. Under the assumption that the combiner cannot suppress the lumped interference (since it is uncorrelated across the antenna elements), a worst case analysis is obtained since the actual combiner performance will always be better.

To evaluate the performance of OC, several definitions are required as follows:

$$\begin{aligned}\Gamma &= \frac{\text{average received desired signal power per antenna}}{\text{average received noise plus interference power per antenna}} \\ \bar{\gamma}_c &= \frac{\text{average received desired signal power per antenna}}{\text{average received noise power per antenna}} = \frac{E[|g_{0,k}|^2]E_{av}}{N_o} \\ \bar{\gamma}_i &= \frac{\text{average received } i\text{th interferer power per antenna}}{\text{average received noise power per antenna}} = \frac{E[|g_{i,k}|^2]E_{av}^i}{N_o} \\ \omega_R &= \frac{\text{instantaneous desired signal power at the array output}}{\text{average noise plus interference power at the array output}} \\ \omega &= \frac{\text{instantaneous desired signal power at the array output}}{\text{instantaneous noise plus interference power at the array output}}.\end{aligned}$$

In the above definitions, “average” refers to the average over the Rayleigh fading, while “instantaneous” refers to an average over a period that is much less than the channel coherence time, i.e., several modulated symbol durations. Finally, note that

$$\Gamma = \frac{C}{I + N} = \frac{\bar{\gamma}_c}{1 + \sum_{k=1}^K \bar{\gamma}_i}. \quad (6.103)$$

In general, two approaches have been taken in the literature to evaluate the performance of OC. The first approach assumes fading of the desired signal only (non-faded interferer), while the second approach assumes fading of both the desired and interfering signals. The second approach is actually more realistic, since the interfering signals are usually subject to the same fading rate as the desired signals. Closed form expressions have been derived for either approach in the case of a single interferer, while bounds and approximations have been used for multiple interferers. For our purpose, attention is restricted to a single interferer.

6.8.1.1 Fading of the Desired Signal Only

From (6.100), ω_R is equal to

$$\omega_R = E_{\text{av}} \mathbf{g}_0^H \Phi_{\tilde{\mathbf{R}}_i, \tilde{\mathbf{R}}_i}^{-1} \mathbf{g}_0, \quad (6.104)$$

where, with a single interferer,

$$\Phi_{\tilde{\mathbf{R}}_i, \tilde{\mathbf{R}}_i} = E_{\text{av}}^1 E[\mathbf{g}_1 \mathbf{g}_1^H] + N_o \mathbf{I}. \quad (6.105)$$

Note that the expectation in (6.105) is over the Rayleigh fading. The pdf of ω_R is [45]

$$p_{\omega_R}(x) = \frac{e^{-x/\bar{\gamma}_c} (x/\bar{\gamma}_c)^{L-1} (1 + L\bar{\gamma}_1)}{\bar{\gamma}_c (L-2)!} \int_0^1 e^{-(x/\bar{\gamma}_c)L\bar{\gamma}_1 t} (1-t)^{L-2} dt \quad (6.106)$$

and the cdf of ω_R is

$$F_{\omega_R}(x) = \int_0^{x/\bar{\gamma}_c} \frac{e^{-y} y^{L-1} (1 + L\bar{\gamma}_1)}{(L-2)!} \int_0^1 e^{-(yL\bar{\gamma}_1)t} (1-t)^{L-2} dt dy. \quad (6.107)$$

which are valid for $L \geq 2$. Note that ω_R in (6.106) and (6.107) is normalized by $\bar{\gamma}_c$. Since $\bar{\gamma}_c = (1 + \bar{\gamma}_1)\Omega$ for the case of a single interferer, it is apparent that ω_R can be normalized by Ω as well. The normalization by Ω allows for a straightforward comparison of OC and MRC. The cdf $F_{\omega_R}(x)$ is plotted against x/Ω in Fig. 6.14 for various values of $\bar{\gamma}_1$. The curve corresponding to $\bar{\gamma}_1 = 0$ corresponds to the performance with MRC. Note that as $\bar{\gamma}_1$ increases the cdf for a fixed value of Ω decreases. This means that OC performs better when the interference becomes a larger fraction of the interference-plus-noise power. Also, the performance improves as the number of antenna elements increases.

The probability of bit error for coherently detected BPSK is given by

$$P_b = \int_0^\infty Q(\sqrt{2x}) p_{\omega_R}(x) dx. \quad (6.108)$$

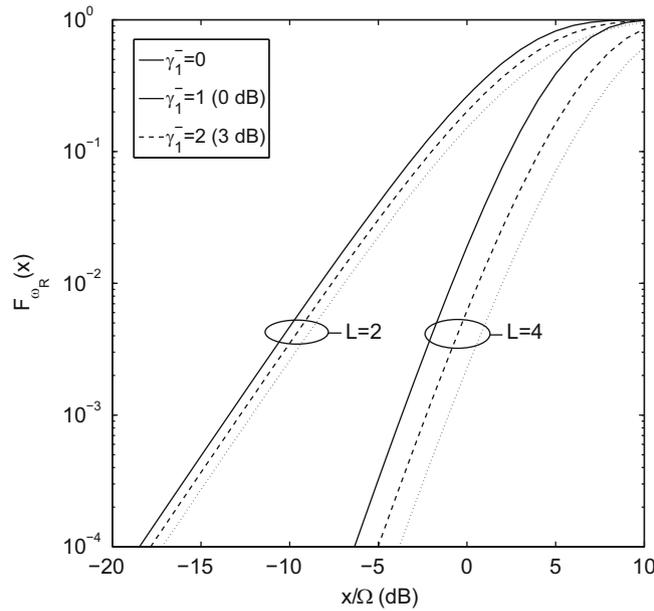


Fig. 6.14 Cdf of ω_R with optimal combining for various values of $\bar{\gamma}_1$ and various number of receiver antenna elements, L

Several approaches have been taken to evaluate the above integral. By using (6.108) and (6.106), and the results in [45], Winters derived the bit error probability as [352]

$$P_b = \frac{(-1)^{L-1}(1 + L\bar{\gamma}_1)}{2(L\bar{\gamma}_1)^{L-1}} \left(-\frac{L\bar{\gamma}_1}{1 + L\bar{\gamma}_1} + \sqrt{\frac{\bar{\gamma}_c}{1 + \bar{\gamma}_c}} - \frac{1}{1 + L\bar{\gamma}_1} \sqrt{\frac{\bar{\gamma}_c}{1 + L\bar{\gamma}_1 + \bar{\gamma}_c}} \right. \\ \left. - \sum_{k=1}^{L-2} (-L\bar{\gamma}_1)^k \left(1 - \sqrt{\frac{\bar{\gamma}_c}{1 + \bar{\gamma}_c}} \left(1 + \sum_{i=1}^k \frac{(2i-1)!!}{i!(2 + 2\bar{\gamma}_c)^i} \right) \right) \right), \quad (6.109)$$

where

$$(2i-1)!! = 1 \cdot 3 \cdot 5 \cdots (2i-1).$$

As observed by Simon and Alouini [303], this expression is only valid for $L \geq 2$. The lack of validity for $L = 1$ can be observed by noting that $P_b = 0$ for $\bar{\gamma}_1 = 0$ (no interference) which is clearly incorrect. By using an alternative form for the Gaussian Q-function, Simon and Alouini have derived the following alternate expression which is valid for $L \geq 1$:

$$P_b = \frac{1}{2} \left(1 - \sqrt{\frac{\bar{\gamma}_c}{1 + \bar{\gamma}_c}} \sum_{k=0}^{L-2} \binom{2k}{k} \frac{1}{(4(1 + \bar{\gamma}_c))^k} \left(1 - \left(-\frac{1}{L\bar{\gamma}_1} \right)^{L-1-k} \right) - \sqrt{\frac{\bar{\gamma}_c}{1 + L\bar{\gamma}_1 + \bar{\gamma}_c}} \left(-\frac{1}{L\bar{\gamma}_1} \right)^{L-1} \right). \quad (6.110)$$

Figure 6.15 shows the probability of bit error with OC and one interferer against Ω for several values of L and $\bar{\gamma}_1$. The performance with $\bar{\gamma}_1 = 0$ is identical to that of MRC, and it was already shown that the bit error probability of coherently detected BPSK with MRC is given by (6.28), where

$$\mu = \sqrt{\frac{\Omega}{1 + \Omega}}$$

and $\Omega = \bar{\gamma}_c$.

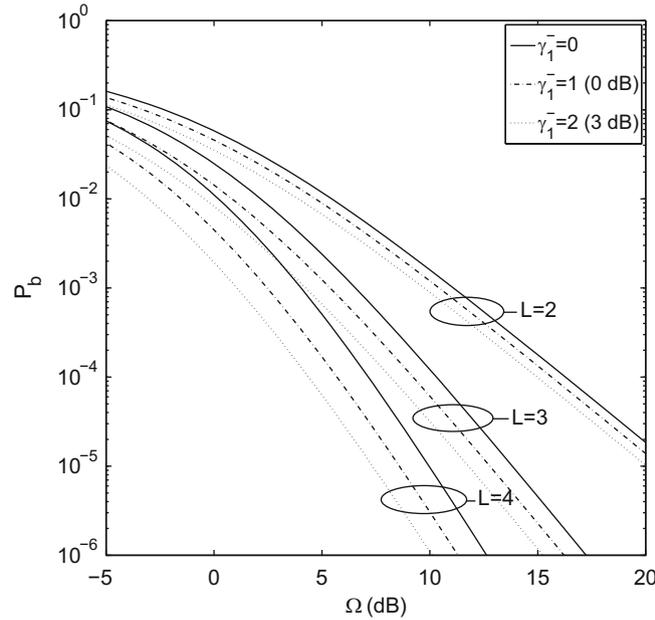


Fig. 6.15 Bit error probability for coherent BPSK and optimal combining for various values of $\bar{\gamma}_1$ and various number of receiver antenna elements, L

6.8.1.2 Fading of the Desired and Interfering Signals

From (6.100), the maximum instantaneous output SINR is equal to

$$\omega = E_{\text{av}} \mathbf{g}_0^H \Phi_{\tilde{\mathbf{r}}_i \tilde{\mathbf{r}}_i}^{-1} \mathbf{g}_0, \quad (6.111)$$

where, with a single interferer,

$$\Phi_{\tilde{\mathbf{r}}_i \tilde{\mathbf{r}}_i} = E_{\text{av}}^1 \mathbf{g}_1 \mathbf{g}_1^H + N_o \mathbf{I}. \quad (6.112)$$

In this case, the matrix $\Phi_{\tilde{\mathbf{r}}_i \tilde{\mathbf{r}}_i}$ varies at the fading rate. The matrix $\Phi_{\tilde{\mathbf{r}}_i \tilde{\mathbf{r}}_i}$ is Hermitian, meaning that $\Phi_{\tilde{\mathbf{r}}_i \tilde{\mathbf{r}}_i}^H = \Phi_{\tilde{\mathbf{r}}_i \tilde{\mathbf{r}}_i}$. Consequently, there exists a diagonalization $\Phi_{\tilde{\mathbf{r}}_i \tilde{\mathbf{r}}_i} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$ such that \mathbf{U} is a unitary matrix and $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_L\}$ is a diagonal matrix consisting of the eigenvalues of $\Phi_{\tilde{\mathbf{r}}_i \tilde{\mathbf{r}}_i}$. It follows that $\Phi_{\tilde{\mathbf{r}}_i \tilde{\mathbf{r}}_i}^{-1} = \mathbf{U}^H \mathbf{\Lambda}^{-1} \mathbf{U}$. Hence,

$$\omega = E_{\text{av}} \mathbf{g}_0^H \mathbf{U}^H \mathbf{\Lambda}^{-1} \mathbf{U} \mathbf{g}_0. \quad (6.113)$$

Since the matrix \mathbf{U} is unitary, the vector $\hat{\mathbf{g}}_0 = \mathbf{U} \mathbf{g}_0 = (\hat{g}_{0,1}, \hat{g}_{0,2}, \dots, \hat{g}_{0,L})^T$ retains the statistical properties of \mathbf{g}_0 . Therefore,

$$\begin{aligned} \omega &= E_{\text{av}} \hat{\mathbf{g}}_0^H \mathbf{\Lambda}^{-1} \hat{\mathbf{g}}_0 \\ &= E_{\text{av}} \sum_{i=1}^L \frac{|\hat{g}_{0,i}|^2}{\lambda_i}. \end{aligned} \quad (6.114)$$

Conditioned on the set of eigenvalues $\lambda_i, i = 1, 2, \dots, L$ (which are now assumed to be random variables), ω is a sum of independent exponentially distributed random variables with mean values $E_{\text{av}} E[|\hat{g}_{0,i}|^2]/\lambda_i, i = 1, 2, \dots, L$. It follows that the characteristic function of ω , conditioned on the set of eigenvalues $\lambda_i, i = 1, 2, \dots, L$, is

$$\phi_{\omega|\mathbf{\Lambda}}(jv) = \prod_{i=1}^L \left(\frac{\lambda_i}{\lambda_i - jv E_{\text{av}} E[|\hat{g}_{0,i}|^2]} \right). \quad (6.115)$$

The main difficulty in proceeding further is that, except for some special cases, it is difficult to find the eigenvalues $\lambda_i, i = 1, 2, \dots, L$, and their associated pdfs. For the case of a single interferer, the eigenvalues are given by [87]

$$\lambda_1 = E_{\text{av}}^1 \mathbf{g}_1^H \mathbf{g}_1 + N_o \quad (6.116)$$

$$\lambda_i = N_o, \quad i = 2, 3, \dots, L. \quad (6.117)$$

The interference-to-noise ratio at the combiner output, $\gamma_1 = E_{\text{av}}^1 \mathbf{g}_1^H \mathbf{g}_1 / N_o$, is a central chi-square random variable with $2L$ degrees of freedom, and has the pdf

$$p_{\gamma_1}(x) = \frac{1}{(L-1)! (\bar{\gamma}_1)^L} x^{L-1} e^{-x/\bar{\gamma}_1}, \quad x \geq 0, \quad (6.118)$$

where $\bar{\gamma}_1 = E_{\text{av}}^1 E[|\mathbf{g}_1|^2] / N_o$. Also, the desired signal-to-noise ratio at the output of the combiner, $\gamma_s = E_{\text{av}} \mathbf{g}_0^H \mathbf{g}_0 / N_o$, is a central chi-square random variable with $2L$ degrees of freedom, and has the pdf

$$p_{\gamma_s}(x) = \frac{1}{(L-1)! (\bar{\gamma}_c)^L} x^{L-1} e^{-x/\bar{\gamma}_c}, \quad x \geq 0, \quad (6.119)$$

where, once again, $\bar{\gamma}_c = E_{\text{av}} E[|\mathbf{g}_0|^2] / N_o$. Using (6.115), the characteristic function of ω , conditioned on γ_1 , is [2]

$$\psi_{\omega|\gamma_1}(jv) = \left(\frac{\frac{\gamma_1+1}{\bar{\gamma}_c}}{\frac{\gamma_1+1}{\bar{\gamma}_c} - jv} \right) \left(\frac{\frac{1}{\bar{\gamma}_c}}{\frac{1}{\bar{\gamma}_c} - jv} \right)^{L-1}. \quad (6.120)$$

For coherent BPSK, it can be shown that the probability of bit error for a given ω is [147, 3.363]

$$\begin{aligned} P_b(\omega) &= Q(\sqrt{2\omega}) \\ &= \frac{1}{2\pi} \int_1^\infty \frac{1}{z\sqrt{z-1}} e^{-\omega z} dz. \end{aligned} \quad (6.121)$$

The bit error probability is then obtained by averaging over the pdf of ω . This gives [2]

$$\begin{aligned} P_b &= \int_0^\infty \frac{1}{2\pi} \int_1^\infty \frac{1}{z\sqrt{z-1}} e^{-xz} p_\omega(x) dz dx \\ &= \frac{1}{2\pi} \int_1^\infty \frac{1}{z\sqrt{z-1}} \int_0^\infty e^{-xz} p_\omega(x) dx dz \\ &= \frac{1}{2\pi} \int_1^\infty \frac{1}{z\sqrt{z-1}} \psi_\omega(-z) dz \end{aligned} \quad (6.122)$$

It follows that the bit error probability conditioned on γ_1 can be obtained by substituting (6.120) into (6.122). The result has the closed form [2]

$$\begin{aligned} P_{b|\gamma_1} &= \frac{1}{2} \left(1 - \sqrt{\frac{\bar{\gamma}_c}{\bar{\gamma}_c + 1}} \sum_{k=0}^{L-2} \binom{2k}{k} \left(\frac{1}{4(\bar{\gamma}_c + 1)} \right)^k \right. \\ &\quad \left. - \left(\sqrt{\frac{\bar{\gamma}_c}{\bar{\gamma}_c + \gamma_1 + 1}} - \sqrt{\frac{\bar{\gamma}_c}{\bar{\gamma}_c + 1}} \sum_{k=0}^{L-2} \binom{2k}{k} \left(\frac{-\gamma_1}{4(\bar{\gamma}_c + 1)} \right)^k \right) (-\gamma_1)^{-(L-1)} \right). \end{aligned} \quad (6.123)$$

The performance with a non-faded interferer can be obtained by replacing γ_1 with its mean value $L\bar{\gamma}_1$. The result agrees with (6.109) and (6.110), where it is assumed that a single interferer with constant power is present. Moreover, as mentioned earlier, OC reduces to MRC when no interfering signals are present. It can be verified that (6.123) with $\gamma_1 = 0$ reduces to

$$P_b = \frac{1}{2} \left(1 - \sqrt{\frac{\bar{\gamma}_c}{\bar{\gamma}_c + 1}} \sum_{k=0}^{L-1} \binom{2k}{k} \left(\frac{1}{4(\bar{\gamma}_c + 1)} \right)^k \right). \quad (6.124)$$

which is also equivalent to (6.28) as expected. In the present case, however, the interference-to-noise ratio, γ_1 is a central chi-square random variable with $2L$ degrees of freedom having the pdf in (6.118). Averaging over the pdf of γ_1 gives the probability of bit error

$$\begin{aligned} P_b &= \int_0^\infty P_{b|\gamma_1}(x) p_{\gamma_1}(x) dx \\ &= \frac{1}{2} \left(1 - \sqrt{\frac{\bar{\gamma}_c}{\bar{\gamma}_c + 1}} \sum_{k=0}^{L-2} \binom{2k}{k} \left(\frac{1}{4(\bar{\gamma}_c + 1)} \right)^k \right) \\ &\quad - \frac{1}{2\Gamma(L)(-\bar{\gamma}_1)^{L-1}} \left(\sqrt{\frac{\pi\bar{\gamma}_c}{\bar{\gamma}_1}} \exp\left(\frac{\bar{\gamma}_c + 1}{\bar{\gamma}_1}\right) \operatorname{erfc}\left(\sqrt{\frac{\bar{\gamma}_c + 1}{\bar{\gamma}_1}}\right) \right. \\ &\quad \left. - \sqrt{\frac{\bar{\gamma}_c}{\bar{\gamma}_c + 1}} \sum_{k=0}^{L-2} \frac{(2k)!}{k!} \left(\frac{-\bar{\gamma}_1}{4(\bar{\gamma}_c + 1)} \right)^k \right). \end{aligned} \quad (6.125)$$

Figure 6.16 plots the bit error probability against

$$\Omega = \frac{\bar{\gamma}_c}{1 + \bar{\gamma}_1}. \quad (6.126)$$

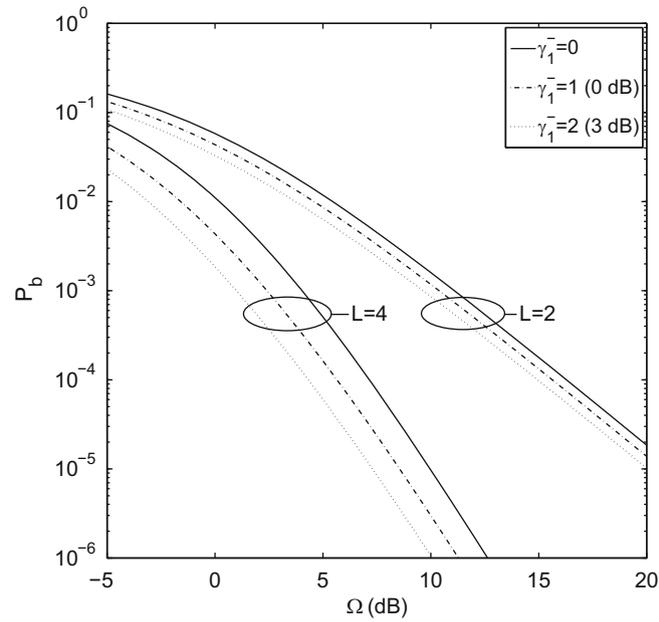


Fig. 6.16 Bit error probability for coherent BPSK and optimal combining for various values of $\bar{\gamma}_1$ and various number of receiver antenna elements, L

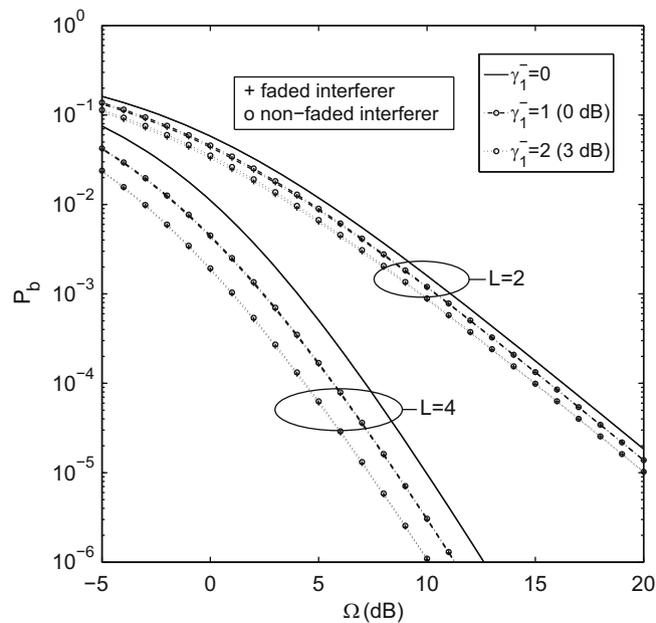


Fig. 6.17 Comparison of the bit error probability for coherent BPSK and optimal combining for a non-faded interferer and a faded interferer; the performance is almost identical

for $\bar{\gamma}_1 = 0, 1, 2$. Figure 6.17 compares the bit error probability for a non-faded interferer in (6.109) or (6.110) with that of a faded interferer in (6.125), where it is seen that the performance is almost identical. The assumption of a non-faded interferer gives a very slightly pessimistic prediction of system performance as compared to the case of a faded interferer. The performance is so close because the array with $L = 2$ or 4 has enough degrees of freedom to reject the co-channel interferer regardless of whether or not it is affected by fading.

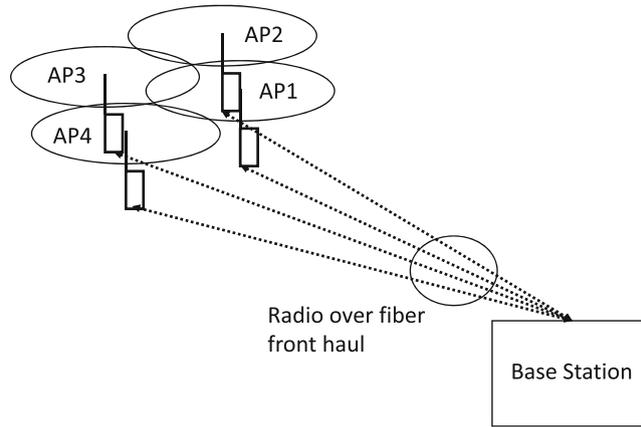


Fig. 6.18 Distributed antenna system using a radio over fiber front haul

6.8.2 Distributed Antenna Systems

The development of MRC in Sect. 6.3 assumed that each diversity branch had equal noise spectral density N_o . It was shown in (6.101) and (6.102) that MRC weighs each signal image $\tilde{\mathbf{r}}_k$ in a manner that is directly proportional to the magnitude of its signal component regardless of phase, the term g_k^* , and inversely proportional to the noise power in the diversity branch, the term $\frac{1}{2}E[|\tilde{n}_k|^2] = N_o$. Optimum performance of the Granlund combiner in Sect. 6.3.1 depends on matched receiver gain and noise figure, meaning that the noise power in each diversity branch is assumed identical and, therefore, it is not necessary to consider the noise explicitly in the combiner weighting process. Sometimes in the literature this is referred to as “ratio squarer diversity.” While such matched receiver gain and noise figure conditions may be possible when the antennas are closely spaced and the preamplifiers are integrated into a single chip as in a cellular handset, these matched conditions generally do not apply to distributed antenna systems where the antennas corresponding to the various diversity branches are widely spaced. Such distributed antenna systems are common for ships, aircraft, and other applications to combat blockage by the ship or aircraft superstructure. Distributed antenna systems are also used for indoor and other types of microcellular systems. For distributed antenna systems, there are typically variations in the noise power in each of the diversity branches contributed by the antennas, cables, and preamplifiers. As such, the noise variances differ from one diversity branch to the other. The low-pass noise in each branch $\tilde{n}(t)$ has power spectral density $\frac{1}{2}E[|\tilde{n}(t)|^2] = N_{o,k}$. Thus, the k th diversity branch has symbol energy-to-noise ratio $\gamma_k = \alpha_k^2/N_{o,k}$. The distributed antenna model is shown in Fig. 6.18.

When the $N_{o,k}$ are distinct, it is important that the MRC combiner account for the different noise power and, according to (6.101) and (6.102) the MRC combining weights become

$$\mathbf{w}_{\text{opt}} = \frac{\mathbf{g}_0^*}{N_{o,k}}, \quad (6.127)$$

and the combiner output is

$$\tilde{\mathbf{r}} = \sum_{k=1}^L \frac{g_{0,k}^*}{N_{o,k}} \tilde{\mathbf{r}}_k. \quad (6.128)$$

The symbol energy-to-noise ratio at the output of the MRC is

$$\gamma_s^{\text{mrc}} = \sum_{k=1}^L \frac{\alpha_k^2 E_{\text{av}}}{N_{o,k}} = \sum_{k=1}^L \gamma_k, \quad (6.129)$$

where $\gamma_k = \alpha_k^2 E_{\text{av}}/N_{o,k}$. Hence, γ_s^{mrc} remains as the sum of the individual symbol energy-to-noise ratios of the L diversity branches, where the γ_k are independent and nonidentically distributed.

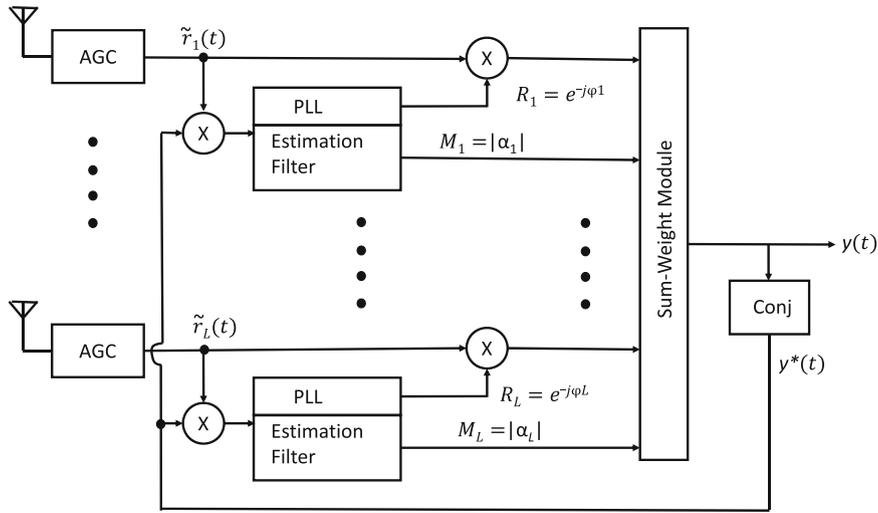


Fig. 6.19 SNR combiner

With only typical variations in the gains and noise figures of distributed antenna systems, a severe degradation in performance will occur if the Granlund combiner described in Sect. 6.3.1 is used [47]. To overcome this problem, Bortnyk and Franke proposed the JTT/SNR combiner [47], where SNR combining yields a combiner performance that is independent of the matching receiver gain or cable loss in each diversity branch. A combining algorithm based on signal-to-noise ratio weighting, due to Bortnyk [46] is elaborated upon below.

Figure 6.19 depicts the SNR combiner, and comparison with Fig. 6.7 highlights the differences between the SNR and Granlund combiners. While the Granlund combiner employs a single automatic gain control (AGC) circuit, the SNR combiner employs independent AGC circuits in each diversity branch. The independent AGC loops maintain the rms signal plus noise envelope level at an input level threshold R_{level} , i.e.,

$$\sqrt{C + N} = R_{\text{level}}, \quad (6.130)$$

where C represents a desired signal power N represents a noise power and R_{level} is the AGC leveling constant. Thus, the rms signal plus noise envelope level of each diversity branch input, $\tilde{r}_k(t)$, to the combiner is adjusted to be equal to R_{level} . The combined output signal is given by

$$y(t) = \frac{1}{\sum_i W_i M_i} \sum_k \{W_k |\alpha_k| \tilde{s}_m(t) + W_k R_k \tilde{n}_k(t)\}. \quad (6.131)$$

where W_i is the weighting coefficient for the k th branch and $R_i = e^{-j\hat{\phi}_k}$ represents an estimate of $e^{-j\phi_k}$ obtained from a phase locked loop.

The combined output $y(t)$ is correlated with each of the received signal images

$$\tilde{r}_k(t) = \alpha_k \tilde{s}_m(t) e^{j\phi_k} + \tilde{n}_k(t) \quad (6.132)$$

to yield the set of phasors $\tilde{v}_k(t)$, $k = 1, \dots, L$, where

$$\begin{aligned} \tilde{v}_k(t) &= (\alpha_k \tilde{s}_m(t) e^{j\phi_k} + \tilde{n}_k(t)) y^*(t) \\ &= (\alpha_k \tilde{s}_m(t) e^{j\phi_k} + \tilde{n}_k(t)) \times \frac{1}{\sum_i W_i M_i} \sum_k \{W_k |\alpha_k| \tilde{s}_m^*(t) + W_k R_k^* \tilde{n}_k^*(t)\} \end{aligned} \quad (6.133)$$

The above expression expands into four terms

$$\begin{aligned} \tilde{v}_k(t) &= \alpha_k \tilde{s}_m(t) \tilde{s}_m^*(t) e^{j\phi_k} \frac{1}{\sum_i W_i M_i} \sum_i \{W_i |\alpha_i|\} + \alpha_k \tilde{s}_m(t) e^{j\phi_k} \frac{1}{\sum_i W_i M_i} \sum_i \{W_i R_i^* \tilde{n}_i^*(t)\} \\ &+ \tilde{n}_k(t) \frac{1}{\sum_i W_i M_i} \sum_i \{W_i |\alpha_i| \tilde{s}_m^*(t)\} + \tilde{n}_k(t) \frac{1}{\sum_i W_i M_i} \sum_i \{W_i R_i^* \tilde{n}_i^*(t)\} \end{aligned} \quad (6.134)$$

The uncorrelated terms due to random noise, data, and channel gains average to zero. It follows that

$$M_k = E[|\tilde{v}_k(t)|] \approx |\alpha_k| E[|\tilde{s}_m(t)|^2] + \frac{W_k}{\sum_i W_i M_i} \|\tilde{n}_k(t)\|^2. \quad (6.135)$$

In practice the expectation is replaced with time averaging. In the event that the signal-to-noise ratio is high, the noise component $n_k(t)$ is small. If the signal-to-noise ratio is small, the weight W_k is small. Hence, the second term is typically small. For simplicity, it can be assumed that $E[|\tilde{s}_m(t)|^2] = 1$ so that $M_k \approx |\alpha_k|$, and M_k is an estimate of the signal magnitude $|\alpha_k|$. It follows that M_k^2 is a measure of the signal power C . The ACG circuit on each branch keeps $E[|r_k(t)|] = R_{\text{level}}$, where R_{level} is the AGC leveling constant. For maximal ratio combining, the weight W_k should be in proportion to the ratio $\sqrt{C/N^2} = M_k/N^2$. Accordingly, $N^2 = E[|n_k(t)|^2] = R_{\text{level}}^2 - M_k^2$, and

$$W_k = \frac{M_k}{R_{\text{level}}^2 - M_k^2}. \quad (6.136)$$

The sum-weight module in Fig. 6.19 sums the L signal images to produce

$$\Sigma_r(t) = \sum_{k=1}^L W_k R_k r_k(t) \quad (6.137)$$

The magnitude of the signal component of $\Sigma_r(t)$ is

$$\left| \sum_{k=1}^L W_k R_k \alpha_k \tilde{s}_m(t) \right| \approx \sum_{k=1}^L W_k M_k \tilde{s}_m(t) \quad (6.138)$$

The output of the combiner $\Sigma_r(t)$ may be normalized to unity by dividing by $\sum_{k=1}^L W_k M_k$. Hence, this gives the combiner output

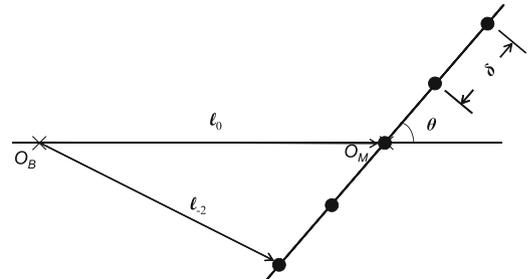
$$\begin{aligned} y(t) &= \frac{1}{\sum_i W_i M_i} \sum_k W_k R_k r_k(t) \\ &= \frac{1}{\sum_i W_i M_i} \sum_k \{W_k |\alpha_k| \tilde{s}_m(t) + W_k R_k \tilde{n}_k(t)\} \end{aligned} \quad (6.139)$$

6.9 Classical Beam-Forming

This section considers line-of-sight reception of a signal by a uniform linear array (ULA) of antenna elements as shown in Fig. 6.20. The distance between the transmitter and receiver is assumed to be large enough so that plane wave propagation can be assumed. The ULA is positioned with an angle θ with respect to the x -axis and the elements are uniformly spaced δ m apart as shown in Fig. 6.20. The transmitted bandpass waveform has the form

$$s(t) = \text{Re} \{ \tilde{s}(t) e^{j2\pi f_c t} \}. \quad (6.140)$$

Fig. 6.20 Plane wave incident on a uniform linear array (ULA)



Then the received bandpass waveform at a distance ℓ from the transmitter is

$$r(t) = \text{Re} \{ \alpha(\ell) \tilde{s}(t - \ell/c) e^{j2\pi f_c(t - \ell/c)} \}, \quad (6.141)$$

where $\alpha(\ell)$ is the attenuation at distance ℓ and c is the speed of light. It follows that the received complex envelope is

$$\tilde{r}(t) = \alpha(\ell) \tilde{s}(t - \ell/c) e^{j\phi(t)}, \quad (6.142)$$

where

$$\phi(t) = -j2\pi f_c \ell / c \quad (6.143)$$

is the excess carrier phase. It is assumed that (1) the antenna elements are spaced close enough together so that $\alpha(\ell) = \alpha(\ell_0) \equiv \alpha_0$ for all antenna elements, (2) the antenna elements are far enough apart so that there is no mutual coupling between antenna elements, and (3) the transmitted signal is a bandpass waveform such that $\tilde{s}(t - \ell/c) = \tilde{s}(t - \ell_0/c) \equiv \tilde{s}_0(t)$.

Assume that the ULA has L elements, where L is odd, that are spaced δ m apart. Antenna k is located at distance ℓ_k from the transmit antenna and the corresponding excess carrier phase is

$$\begin{aligned} \phi_k(t) &= -2\pi f_c \ell_k / c \\ &= -2\pi f_c \ell_0 / c - 2\pi f_c (\ell_k - \ell_0) / c \\ &\equiv \phi_0(t) - 2\pi f_c \Delta \ell_k / c \\ &\equiv \phi_0(t) - \Delta \phi_k. \end{aligned} \quad (6.144)$$

From the geometry in Fig. 6.20, the relative distance $\Delta \ell_k$ is

$$\Delta \ell_k = k\delta \cos(\theta), \quad (6.145)$$

under the assumption that the distance $O_B - O_M$ is large compared to the antenna spacing δ , and where the antenna index k is assumed to run from $-L/2$ to $L/2$. The phase offset $\Delta \phi_k$ is

$$\begin{aligned} \Delta \phi_k &= (2\pi f_c / c) k\delta \cos(\theta) \\ &= 2\pi \left(\frac{k\delta}{\lambda_c} \right) \cos(\theta), \end{aligned} \quad (6.146)$$

where λ_c is the carrier wavelength. Hence, the received complex envelope at antenna element k is

$$\begin{aligned} \tilde{r}_k(t) &= \alpha_0 \tilde{s}_0(t) e^{j\phi_0(t)} e^{j2\pi \left(\frac{k\delta}{\lambda_c} \right) \cos(\theta)} \\ &= \alpha_0 \tilde{s}_0(t) e^{j\phi_0(t)} a_k(\theta), \end{aligned} \quad (6.147)$$

where

$$a_k(\theta) = e^{j2\pi \left(\frac{k\delta}{\lambda_c} \right) \cos(\theta)}. \quad (6.148)$$

A phased array computes the weighted sum

$$\begin{aligned} \tilde{r}_c(t) &= \sum_{k=-L/2}^{L/2} w_k^* a_k(\theta) \alpha_0 \tilde{s}_0(t) e^{j\phi_0(t)} \\ &= \mathbf{w}^H \mathbf{a}(\theta) \alpha_0 \tilde{s}_0(t) e^{j\phi_0(t)}, \end{aligned} \quad (6.149)$$

where

$$\mathbf{a}(\theta) = (a_{-L/2}(\theta), \dots, a_0, \dots, a_{L/2}(\theta))^T, \quad (6.150)$$

and

$$\mathbf{w} = (w_{-L/2}, \dots, w_0, \dots, w_{L/2})^T \quad (6.151)$$

is the *weighting vector*. The weighting vector can be chosen to optimize a variety of criteria. One possibility is to maximize the gain of the antenna array $G(\theta) = \mathbf{w}^H \mathbf{a}(\theta)$ when the desired signal arrives such that the antenna array orientation is equal to θ_o . By using the Cauchy–Schwartz inequality, it can be shown that this is accomplished with the weighting vector

$$\mathbf{w}_{\text{opt}} = \mathbf{a}(\theta_o), \quad (6.152)$$

which yields the antenna gain

$$G(\theta) = \mathbf{a}^H(\theta_o) \mathbf{a}(\theta). \quad (6.153)$$

Example 6.1. Suppose that $\theta_o = 90^\circ$ so that the plane waves arrive broadside to the antenna array. Then

$$\mathbf{w}_{\text{opt}} = \mathbf{a}(\theta_o) = (1, 1, \dots, 1)^T. \quad (6.154)$$

In this case, the antenna gain can be written as

$$\begin{aligned} G(\theta) &= \mathbf{w}_{\text{opt}}^H \mathbf{a}(\theta) \\ &= \sum_{k=-L/2}^{L/2} e^{j2\pi \left(\frac{k\delta}{\lambda_c}\right) \cos(\theta)}. \end{aligned}$$

By letting

$$z = e^{j2\pi \left(\frac{\delta}{\lambda_c}\right) \cos(\theta)} \quad (6.155)$$

it follows that

$$\begin{aligned} G(\theta) &= \sum_{k=-L/2}^{L/2} z^k \\ &= \frac{z^{-(L+1)/2} - z^{(L+1)/2}}{z^{-1/2} - z^{1/2}}. \end{aligned} \quad (6.156)$$

Substituting (6.155) into (6.156) and simplifying by using the inverse Euler identity yields

$$G(\theta) = \frac{\sin(\pi(L+1)(\delta/\lambda_c)\cos(\theta))}{\sin(\pi(\delta/\lambda_c)\cos(\theta))}. \quad (6.157)$$

(continued)

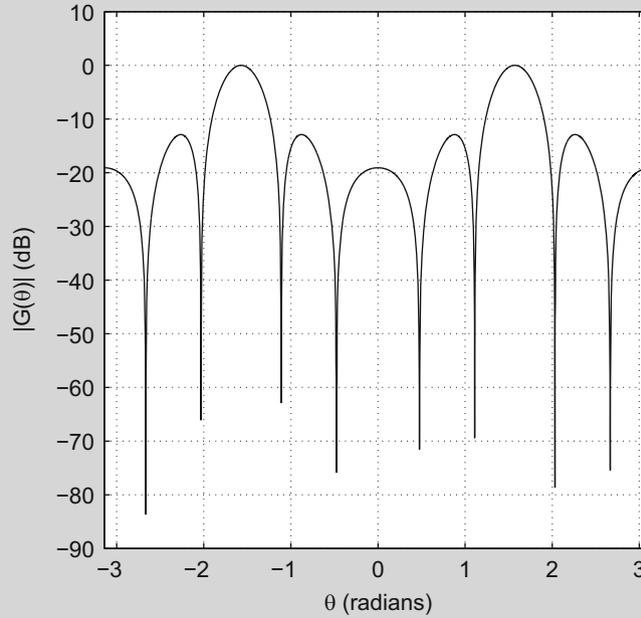
Example 6.1 (continued)

Fig. 6.21 Antenna gain $|G(\theta)|$ dB for a uniform linear array (ULA) optimized for $\theta_o = 90^\circ$; $L = 8$ (9 element array), $\delta/\lambda_c = 0.25$

Figure 6.21 plots the magnitude of the antenna gain

$$G(\theta)_{(\text{dB})} = 20 \log_{10} \{ |G(\theta)| / |G(0)| \} \quad (6.158)$$

against the angle-of-arrival θ for $\delta/\lambda_c = 0.25$ and $L = 8$ (9 element array). Clearly, the antenna exhibits significant gain in the direction $\theta_o = 90^\circ$ ($\pi/2$ radians).

For a ULA the quality of the beam-forming depends on the angle-of-arrival. The best result is obtained for the broadside case when $\theta_o = 90^\circ$ and the worst case is obtained for the in-line case when $\theta_o = 0^\circ$. However, other types of antenna arrays, such as uniform circular arrays, can be used to provide a more uniform performance in all azimuthal directions.

6.10 MIMO Channels

A MIMO system is one that consists, for example, of multiple transmit and receive antennas as shown in Fig. 6.22. For a system consisting of L_T transmit and L_R receive antennas, the channel can be described by the $L_R \times L_T$ matrix

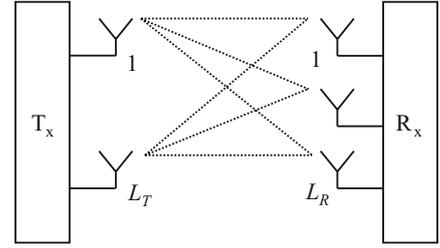
$$\mathbf{G}(t, \tau) = \begin{bmatrix} g_{1,1}(t, \tau) & g_{1,2}(t, \tau) & \cdots & g_{1,L_T}(t, \tau) \\ g_{2,1}(t, \tau) & g_{2,2}(t, \tau) & \cdots & g_{2,L_T}(t, \tau) \\ \vdots & \vdots & \ddots & \vdots \\ g_{L_R,1}(t, \tau) & g_{L_R,2}(t, \tau) & \cdots & g_{L_R,L_T}(t, \tau) \end{bmatrix}, \quad (6.159)$$

where $g_{qp}(t, \tau)$ denotes the time-varying sub-channel impulse response between the p th transmit antenna and q th receive antenna.

Suppose that the complex envelopes of the signals transmitted from the L_T transmit antennas are described by the vector

$$\tilde{\mathbf{s}}(t) = (\tilde{s}_1(t), \tilde{s}_2(t), \dots, \tilde{s}_{L_T}(t))^T, \quad (6.160)$$

Fig. 6.22 MIMO system with multiple transmit and multiple receive antennas



where $\tilde{s}_p(t)$ is the signal transmitted from the p th transmit antenna. Likewise, let

$$\tilde{\mathbf{r}}(t) = (\tilde{r}_1(t), \tilde{r}_2(t), \dots, \tilde{r}_{L_R}(t))^T, \quad (6.161)$$

denote the vector of received complex envelopes, where $\tilde{r}_q(t)$ is the signal received at the q th receiver antenna. Then

$$\tilde{\mathbf{r}}(t) = \int_0^t \mathbf{G}(t, \tau) \tilde{\mathbf{s}}(t - \tau) d\tau \quad (6.162)$$

Under conditions of flat fading

$$\mathbf{G}(t, \tau) = \mathbf{G}(t) \delta(\tau - \hat{\tau}), \quad (6.163)$$

where $\hat{\tau}$ is the delay through the channel and

$$\tilde{\mathbf{r}}(t) = \mathbf{G}(t) \tilde{\mathbf{s}}(t - \hat{\tau}). \quad (6.164)$$

If the MIMO channel is characterized by slow fading, then

$$\tilde{\mathbf{r}}(t) = \int_0^t \mathbf{G}(\tau) \tilde{\mathbf{s}}(t - \tau) d\tau. \quad (6.165)$$

In this case, the channel matrix $\mathbf{G}(\tau)$ remains constant over the duration of the transmitted waveform $\tilde{\mathbf{s}}(t)$, but can vary randomly from one channel use to the next, where a channel use may be defined as the transmission of either a single modulated symbol or a vector of modulated symbols from each antenna. In the case of a vector of modulated symbols, this type of channel is sometimes called a quasi-static fading channel or a block fading channel. Finally, if the MIMO channel is characterized by slow flat fading, then

$$\tilde{\mathbf{r}}(t) = \mathbf{G} \tilde{\mathbf{s}}(t), \quad (6.166)$$

where once again the matrix \mathbf{G} is assumed to vary randomly from one channel use to the next. MIMO channel models can be classified as either *physical* or *analytical* models. Physical MIMO models characterize the channel on the basis of electromagnetic wave propagation between the transmitter and receiver antennas. Such physical channel models can be further classified as deterministic models, geometry-based stochastic models, and non-geometric stochastic models. Deterministic models construct the MIMO channel in a completely deterministic manner, such as ray tracing and stored measurement data. With geometry-based stochastic models, the time-variant MIMO channel impulse response is generated by applying the laws of wave propagation to specific transmitter, receiver, and scattering geometries, which are generated in a stochastic or random fashion. The non-geometric stochastic models determine physical parameters such as the angles of departure and angles of arrival in a completely stochastic fashion by prescribing underlying probability density functions, but not using an underlying geometry. These models include the extended Saleh–Valenzuela models [63, 340].

The analytical MIMO models characterize the MIMO sub-channel impulse responses in a mathematical manner without explicitly considering the underlying electromagnetic wave propagation. Analytic models can be further classified as propagation motivated models or correlation-based models. Propagation motivated models include the finite scattering models [49], maximum entropy models [92], and virtual channel representation [293]. Correlation-based models generate

random realizations of the channel matrix with specified correlations between the matrix elements [179, 350]. These models are easy to implement, which has made them very popular for MIMO channel simulations. Moreover, the analytical models are relatively easy to construct and use, so they are treated here.

6.10.1 Analytical MIMO Channel Models

Analytical MIMO channel models are most often used under quasi-static flat fading conditions. It was shown earlier that the time-variant channel impulses for flat fading channels can be treated as complex Gaussian random processes under conditions of Rayleigh and Ricean fading. The various analytical models generate the MIMO matrices as realizations of complex Gaussian random variables having specified means and correlations. To model Ricean fading, the channel matrix can be divided into a deterministic part and a random part, i.e.,

$$\mathbf{G} = \sqrt{\frac{K}{K+1}} \bar{\mathbf{G}} + \sqrt{\frac{1}{K+1}} \mathbf{G}_s \quad (6.167)$$

where $E[\mathbf{G}] = \sqrt{\frac{K}{K+1}} \bar{\mathbf{G}}$ is the LoS or specular component and $\sqrt{\frac{1}{K+1}} \mathbf{G}_s$ is the scatter component assumed to have zero mean. The elements of the matrices $\bar{\mathbf{G}}$ and \mathbf{G}_s are normalized to have power Ω_p , so that the elements of matrix \mathbf{G} have power Ω_p , i.e., $E[|g_{pq}|^2] = \Omega_p$. In this case, K represents the Rice factor, defined as the ratio of the power in the LoS or specular component to the power in the scatter component.

To simplify our further treatment of the MIMO channel, assume for the time being that $K = 0$, so that $\mathbf{G} = \mathbf{G}_s$. The simplest MIMO model assumes that the entries of the matrix \mathbf{G} are independent and identically distributed (i.i.d) complex Gaussian random variables. This model corresponds to the so-called rich scattering or spatially white environment. Such an independence assumption simplifies the performance analysis of various digital signaling schemes operating on MIMO channels. However, in reality the sub-channels will be correlated and, therefore, the i.i.d. model will lead to optimistic results.

A variety of more sophisticated models have been introduced to account for spatial correlation of the sub-channels. Consider the vector $\mathbf{g} = \text{vec}\{\mathbf{G}\}$, where $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_{L_T}]$ and $\text{vec}\{\mathbf{G}\} = [\mathbf{g}_1^T, \mathbf{g}_2^T, \dots, \mathbf{g}_{L_T}^T]^T$. The vector \mathbf{g} is a column vector of length $n = L_T L_R$. The vector \mathbf{g} is zero-mean complex Gaussian random vector and its statistics are fully specified by the $n \times n$ channel covariance matrix $\mathbf{R}_G = E[\mathbf{g}\mathbf{g}^H]$. Hence, $\mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_G)$ and, if \mathbf{R}_G is invertible, the probability density function of \mathbf{g} is

$$p(\mathbf{g}) = \frac{1}{(2\pi)^n \det(\mathbf{R}_G)} \exp \left\{ -\frac{1}{2} \mathbf{g}^H \mathbf{R}_G^{-1} \mathbf{g} \right\}, \quad \mathbf{g} \in \mathcal{C}^n. \quad (6.168)$$

Realizations of the MIMO channel with the distribution in (6.168) can be generated by

$$\mathbf{G} = \text{unvec}\{\mathbf{g}\} \quad \text{with} \quad \mathbf{g} = \mathbf{R}_G^{1/2} \mathbf{w}. \quad (6.169)$$

Here, $\mathbf{R}_G^{1/2}$ is any matrix square root of \mathbf{R}_G , i.e., $\mathbf{R}_G = \mathbf{R}_G^{1/2} (\mathbf{R}_G^{1/2})^H$, and \mathbf{w} is a length n vector where $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

6.10.1.1 Kronecker Model

The Kronecker model [179] constructs the MIMO channel matrix \mathbf{G} under the assumption that the spatial correlation at the transmitter and receiver is separable. This is equivalent to restricting the correlation matrix \mathbf{R}_G to have the Kronecker product form

$$\mathbf{R}_G = \mathbf{R}_T^T \otimes \mathbf{R}_R, \quad (6.170)$$

where

$$\mathbf{R}_T = E[\mathbf{G}^H \mathbf{G}] \quad \mathbf{R}_R = E[\mathbf{G} \mathbf{G}^H]. \quad (6.171)$$

are the $L_T \times L_T$ and $L_R \times L_R$ transmit and receive correlation matrices respectively, and \otimes is the Kronecker product. For example, the Kronecker product of an $n \times n$ matrix \mathbf{A} and an $m \times m$ matrix \mathbf{B} is

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{n1}\mathbf{B} & \cdots & a_{nn}\mathbf{B} \end{bmatrix}. \quad (6.172)$$

Under the above Kronecker assumption, (6.169) simplifies to the Kronecker model

$$\mathbf{g} = (\mathbf{R}_T^T \otimes \mathbf{R}_R)^{1/2} \mathbf{w} \quad (6.173)$$

or

$$\mathbf{G} = \mathbf{R}_R^{1/2} \mathbf{W} \mathbf{R}_T^{1/2}, \quad (6.174)$$

where \mathbf{W} is an $L_R \times L_T$ matrix having elements that are i.i.d. zero-mean complex Gaussian random variables having unit variance.

The elements of the matrix \mathbf{G} represent correlations between the faded envelopes of the MIMO sub-channels. If the elements of \mathbf{G} could be arbitrarily selected, then the correlation functions would be a function of four sub-channel index parameters, i.e.,

$$\mathbb{E}[g_{qp}^* g_{\tilde{q}\tilde{p}}] = \phi(q, p, \tilde{q}, \tilde{p}), \quad (6.175)$$

where g_{qp} is the channel between the p th transmit and q th receive antenna. However, due to the Kronecker property in (6.170), the elements of \mathbf{G} are structured. One implication of the Kronecker property is spatial stationarity

$$\mathbb{E}[g_{qp}^* g_{\tilde{q}\tilde{p}}] = \phi(q - \tilde{q}, p - \tilde{p}). \quad (6.176)$$

This implies that the sub-channel correlations are determined not by their position in the matrix \mathbf{G} , but by their positional difference. In addition, to the stationary property, manipulation of the Kronecker product form in (6.170) implies that

$$\mathbb{E}[g_{qp}^* g_{\tilde{q}\tilde{p}}] = \phi_T(p - \tilde{p}) \cdot \phi_R(q - \tilde{q}). \quad (6.177)$$

This means that the correlation can be separated into two parts: a transmitter part and a receiver part, and both parts are stationary. Finally, the Kronecker property in (6.170) holds if and only if the separable property in (6.177) holds. The separable property is satisfied by double-bounced channels such as the double ring mobile-to-mobile channel model in Fig. 2.53, where the angles-of-arrival for each ray at the receiver are independent of the angles-of-departure of each ray at the transmitter.

6.10.1.2 Weichselberger Model

The Weichselberger model [350] overcomes the separable requirement of the channel correlation functions in (6.177) of the Kronecker model so as to include a broader range of MIMO channels. Its definition is based on an eigenvalue decomposition of the transmitter and receiver correlation matrices,

$$\mathbf{R}_T = \mathbf{U}_T \mathbf{\Lambda}_T \mathbf{U}_T^H \quad (6.178)$$

$$\mathbf{R}_R = \mathbf{U}_R \mathbf{\Lambda}_R \mathbf{U}_R^H \quad (6.179)$$

Here the matrices $\mathbf{\Lambda}_T$ and $\mathbf{\Lambda}_R$ are diagonal matrices containing the eigenvalues of \mathbf{R}_T and \mathbf{R}_R , respectively, while \mathbf{U}_T and \mathbf{U}_R are unity matrices whose columns are the corresponding eigenvectors of \mathbf{R}_T and \mathbf{R}_R , respectively. Sometimes \mathbf{U}_T and \mathbf{U}_R are called the transmit and receive eigenmodes, respectively. The Weichselberger model describes the power coupling of the transmit and receive eigenmodes by the $L_R \times L_T$ coupling matrix

$$\mathbf{\Omega} = \mathbb{E}_G [(\mathbf{U}_R^H \mathbf{G} \mathbf{U}_T^*) \odot (\mathbf{U}_R^T \mathbf{G}^* \mathbf{U}_T)], \quad (6.180)$$

where \odot denotes the Schur–Hadamard product (element-wise matrix multiplication). The nonnegative real values of the coupling matrix $\mathbf{\Omega}$ determine the average power coupling between the transmitter and receiver eigenvectors. The corresponding channel correlation matrix is

$$\mathbf{R}_G = \sum_{i=1}^{L_T} \sum_{j=1}^{L_R} \omega_{ji} (\mathbf{u}_{T,i} \otimes \mathbf{u}_{R,j}) (\mathbf{u}_{T,i} \otimes \mathbf{u}_{R,j})^H \quad (6.181)$$

where $\omega_{ji} = [\mathbf{\Omega}]_{ji}$, and $\mathbf{u}_{T,i}$ and $\mathbf{u}_{R,j}$ are the i th and j th columns of \mathbf{U}_T and \mathbf{U}_R , respectively.

Realizations of the channel matrix \mathbf{G} are generated as

$$\mathbf{G} = \mathbf{U}_R (\tilde{\mathbf{\Omega}} \odot \mathbf{W}) \mathbf{U}_T^T, \quad (6.182)$$

where \mathbf{W} is an $L_R \times L_T$ matrix consisting of i.i.d. zero-mean complex Gaussian random variables having unit variance, and the matrix $\tilde{\mathbf{\Omega}}$ is the element-wise square root of $\mathbf{\Omega}$.

Finally, the Kronecker model is a special case of the Weichselberger model obtained with the coupling matrix $\mathbf{\Omega} = \boldsymbol{\lambda}_R \boldsymbol{\lambda}_T^T$, where $\boldsymbol{\lambda}_R$ and $\boldsymbol{\lambda}_T$ are column vectors containing the eigenvalues of $\mathbf{\Lambda}_R$ and $\mathbf{\Lambda}_T$, respectively.

6.11 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 antennas 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 user terminals can use just a single antenna and still benefit from a diversity gain.

Transmitter diversity can take on many forms, distinguished by the method of using the multiple transmit antennas. Transmit diversity is straightforward for systems that use time division duplexing (TDD), where different time slots on the same carrier are used for the forward and reverse link transmissions. For TDD systems, the channel impulse response satisfies the reciprocity principle. At the base station the signals received on all antennas can be processed during every received burst, and used to estimate the corresponding channel impulse responses. The antenna that provides the largest received bit or symbol energy-to-noise ratio on the reverse link is selected and used for the next forward burst transmission. This is a form of selective transmit diversity (STD). Obviously, this scheme requires that the channel coherence time be larger than the burst duration.

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). Delay transmit diversity is another method, where copies of the same symbol are transmitted through multiple antennas at different times. This has the effect of creating artificial delay spread so that 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, space-frequency, or space-time-frequency encoding of the transmitted information. In general, these schemes require three functions: (1) a method for encoding and transmitting the information sequence at the transmitter, (2) a combining scheme at the receiver, (3) a rule for making decisions. Alamouti [15] 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 that separate pilot sequences be inserted into the waveforms that are transmitted from each of the transmit antennas.

6.11.1 Alamouti's Transmit Diversity Scheme

Alamouti's transmit diversity scheme [15] uses two transmit antennas and one receiver antenna, referred to as 2×1 diversity. With the Alamouti scheme, two complex data symbols are transmitted over two successive baud intervals by using two transmit antennas. During the first baud interval, the complex symbol vectors transmitted from Antennas 1 and 2 are denoted by $\tilde{\mathbf{s}}_{(1)}$ and $\tilde{\mathbf{s}}_{(2)}$, respectively. During the next baud interval, the complex symbol vectors transmitted from Antennas 1 and

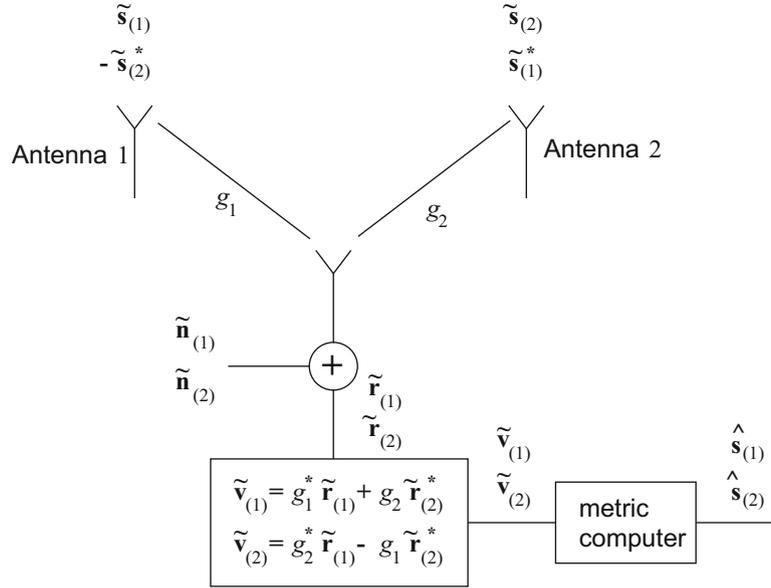


Fig. 6.23 Space-time diversity receiver for 2×1 diversity

2 are $-\tilde{\mathbf{s}}^{(2)*}$ and $\tilde{\mathbf{s}}^{(1)*}$, respectively. Assuming a slow flat fading channel, the complex channel gains associated with transmit Antennas 1 and 2 are g_1 and g_2 , respectively. The complex received vectors are

$$\begin{aligned}\tilde{\mathbf{r}}^{(1)} &= g_1 \tilde{\mathbf{s}}^{(1)} + g_2 \tilde{\mathbf{s}}^{(2)} + \tilde{\mathbf{n}}^{(1)} \\ \tilde{\mathbf{r}}^{(2)} &= -g_1 \tilde{\mathbf{s}}^{(2)*} + g_2 \tilde{\mathbf{s}}^{(1)*} + \tilde{\mathbf{n}}^{(2)},\end{aligned}\quad (6.183)$$

where $\tilde{\mathbf{r}}^{(1)}$ and $\tilde{\mathbf{r}}^{(2)}$ represent the received vectors during the first and second baud intervals, respectively, and $\tilde{\mathbf{n}}^{(1)}$ and $\tilde{\mathbf{n}}^{(2)}$ are the corresponding complex Gaussian noise vectors.

The diversity combiner for this scheme is shown in Fig. 6.23. The combiner constructs the following two signal vectors:

$$\begin{aligned}\tilde{\mathbf{v}}^{(1)} &= g_1^* \tilde{\mathbf{r}}^{(1)} + g_2 \tilde{\mathbf{r}}^{(2)*} \\ \tilde{\mathbf{v}}^{(2)} &= g_2^* \tilde{\mathbf{r}}^{(1)} - g_1 \tilde{\mathbf{r}}^{(2)*}.\end{aligned}\quad (6.184)$$

Afterwards, the receiver applies the vectors $\tilde{\mathbf{v}}^{(1)}$ and $\tilde{\mathbf{v}}^{(2)}$ in a sequential or parallel fashion to the metric computer in Fig. 6.4, to make decisions by maximizing the decision metric

$$\begin{aligned}\mu(\tilde{\mathbf{s}}^{(1),m}) &= \text{Re}(\tilde{\mathbf{v}}^{(1)}, \tilde{\mathbf{s}}^{(1),m}) - E_m(|g_1|^2 + |g_2|^2) \\ \mu(\tilde{\mathbf{s}}^{(2),m}) &= \text{Re}(\tilde{\mathbf{v}}^{(2)}, \tilde{\mathbf{s}}^{(2),m}) - E_m(|g_1|^2 + |g_2|^2).\end{aligned}\quad (6.185)$$

Using (6.183) in (6.184) gives

$$\begin{aligned}\tilde{\mathbf{v}}^{(1)} &= (\alpha_1^2 + \alpha_2^2) \tilde{\mathbf{s}}^{(1)} + g_1^* \tilde{\mathbf{n}}^{(1)} + g_2 \tilde{\mathbf{n}}^{(2)*} \\ \tilde{\mathbf{v}}^{(2)} &= (\alpha_1^2 + \alpha_2^2) \tilde{\mathbf{s}}^{(2)} - g_1 \tilde{\mathbf{n}}^{(2)} + g_2^* \tilde{\mathbf{n}}^{(1)}.\end{aligned}\quad (6.186)$$

This is to be compared with the output of the MRC metric computer in Fig. 6.4. With $L = 2$,

$$\begin{aligned}\tilde{\mathbf{r}} &= g_1^* \tilde{\mathbf{r}}_1 + g_2^* \tilde{\mathbf{r}}_2 \\ &= (\alpha_1^2 + \alpha_2^2) \tilde{\mathbf{s}}_m + g_1^* \tilde{\mathbf{n}}_1 + g_2^* \tilde{\mathbf{n}}_2.\end{aligned}\quad (6.187)$$

Comparison of (6.186) and (6.187) shows that the combined signals in each case are the same. The only difference is the phase rotations of the Gaussian noise vectors which will not matter due to their circular symmetry.

6.11.1.1 $2 \times L$ Diversity

The case of $2 \times L$ diversity is now considered, and it is shown that the performance is equivalent to $1 \times 2L$ diversity with MRC. The approach is illustrated for the case of 2×2 diversity, and the extension to $2 \times L$ diversity will be obvious. To describe the scheme, the following notation is introduced:

$g_{i,j}$ = channel gain between transmit antenna i and receiver antenna j .

$\tilde{\mathbf{r}}_{(1),j}$ = received signal at antenna j during the first baud interval.

$\tilde{\mathbf{r}}_{(2),j}$ = received signal at antenna j during the second baud interval.

The encoding scheme remains the same as before: the complex symbol vectors $\tilde{\mathbf{s}}_{(1)}$ and $\tilde{\mathbf{s}}_{(2)}$ are transmitted from Antennas 1 and 2 during the first baud interval, and complex symbol vectors $-\tilde{\mathbf{s}}_{(2)}^*$ and $\tilde{\mathbf{s}}_{(1)}^*$ are transmitted from Antennas 1 and 2 during the second baud interval. The complex received signal vectors are

$$\tilde{\mathbf{r}}_{(1),1} = g_{1,1}\tilde{\mathbf{s}}_{(1)} + g_{2,1}\tilde{\mathbf{s}}_{(2)} + \tilde{\mathbf{n}}_{(1),1}$$

$$\tilde{\mathbf{r}}_{(2),1} = -g_{1,1}\tilde{\mathbf{s}}_{(2)}^* + g_{2,1}\tilde{\mathbf{s}}_{(1)}^* + \tilde{\mathbf{n}}_{(2),1}$$

$$\tilde{\mathbf{r}}_{(1),2} = g_{1,2}\tilde{\mathbf{s}}_{(1)} + g_{2,2}\tilde{\mathbf{s}}_{(2)} + \tilde{\mathbf{n}}_{(1),2}$$

$$\tilde{\mathbf{r}}_{(2),2} = -g_{1,2}\tilde{\mathbf{s}}_{(2)}^* + g_{2,2}\tilde{\mathbf{s}}_{(1)}^* + \tilde{\mathbf{n}}_{(2),2}.$$

The combiner shown in Fig. 6.24 constructs the following two signal vectors

$$\tilde{\mathbf{v}}_{(1)} = g_{1,1}^*\tilde{\mathbf{r}}_{(1),1} + g_{2,1}\tilde{\mathbf{r}}_{(2),1}^* + g_{1,2}^*\tilde{\mathbf{r}}_{(1),2} + g_{2,2}\tilde{\mathbf{r}}_{(2),2}^* \quad (6.188)$$

$$\tilde{\mathbf{v}}_{(2)} = g_{2,1}^*\tilde{\mathbf{r}}_{(1),1} - g_{1,1}\tilde{\mathbf{r}}_{(2),1}^* + g_{2,2}^*\tilde{\mathbf{r}}_{(1),2} - g_{1,2}\tilde{\mathbf{r}}_{(2),2}^*. \quad (6.189)$$

Again, the receiver applies the vectors $\tilde{\mathbf{v}}_{(1)}$ and $\tilde{\mathbf{v}}_{(2)}$ in a sequential or parallel fashion to the metric computer in Fig. 6.4 and decisions are made by maximizing the metric in (6.185).

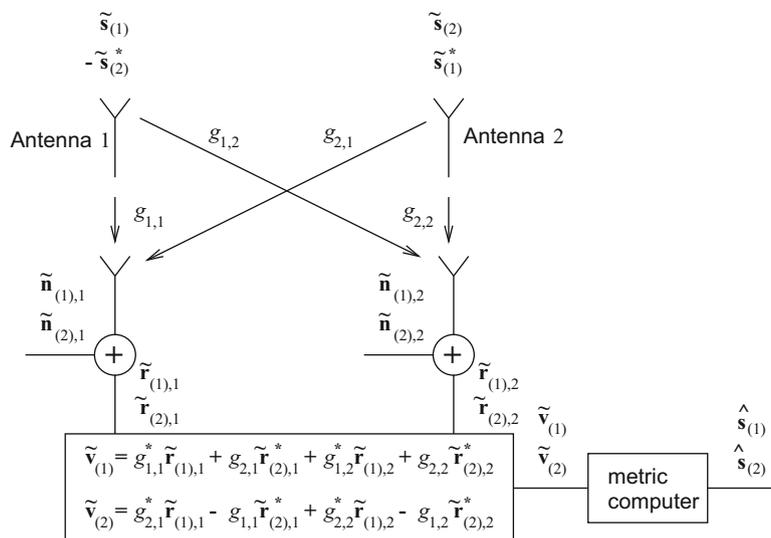


Fig. 6.24 Space-time diversity receiver for 2×2 diversity

To compare the 2×2 transmit diversity scheme to 1×4 receive diversity with MRC, appropriate equations are substituted to obtain

$$\tilde{\mathbf{v}}_{(1)} = (\alpha_{1,1}^2 + \alpha_{1,2}^2 + \alpha_{2,1}^2 + \alpha_{2,2}^2)\tilde{\mathbf{s}}_{(1)} + g_{1,1}^*\tilde{\mathbf{n}}_{(1),1} + g_{2,1}\tilde{\mathbf{n}}_{(2),1}^* + g_{1,2}^*\tilde{\mathbf{n}}_{(1),2} + g_{2,2}\tilde{\mathbf{n}}_{(2),2}^* \quad (6.190)$$

$$\tilde{\mathbf{v}}_{(2)} = (\alpha_{1,1}^2 + \alpha_{1,2}^2 + \alpha_{2,1}^2 + \alpha_{2,2}^2)\tilde{\mathbf{s}}_{(2)} + g_{2,1}^*\tilde{\mathbf{n}}_{(1),1} - g_{1,1}\tilde{\mathbf{n}}_{(2),1}^* + g_{2,2}^*\tilde{\mathbf{n}}_{(1),2} - g_{1,2}\tilde{\mathbf{n}}_{(2),2}^*. \quad (6.191)$$

This is to be compared with the output of the MRC in Fig. 6.4. With $L = 4$,

$$\begin{aligned} \tilde{\mathbf{r}} &= g_1^*\tilde{\mathbf{r}}_1 + g_2^*\tilde{\mathbf{r}}_2 + g_3^*\tilde{\mathbf{r}}_3 + g_4^*\tilde{\mathbf{r}}_4 \\ &= (\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2)\tilde{\mathbf{s}}_m + g_1^*\tilde{\mathbf{n}}_1 + g_2^*\tilde{\mathbf{n}}_2 + g_3^*\tilde{\mathbf{n}}_3 + g_4^*\tilde{\mathbf{n}}_4. \end{aligned} \quad (6.192)$$

Again, Alamouti's 2×2 transmit diversity scheme is equivalent to a 1×4 receive diversity scheme with MRC. The extension to show that Alamouti's $2 \times L$ transmit diversity scheme is equivalent to $1 \times 2L$ diversity with MRC is left as an exercise.

6.11.1.2 Implementation Issues

There are several key implementation issues with Alamouti's transmit diversity scheme, including the following:

- Since there are two transmit antennas, the power per antenna must be halved to maintain a constant transmit power. Hence, the $2 \times L$ transmit diversity scheme has a 3 dB performance loss when compared to a $1 \times 2L$ diversity scheme with MRC.
- With two 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 transmitted simultaneously from the two transmit antennas.
- The transmit antennas must be spaced sufficiently far apart in order to achieve sufficient spatial decorrelation between the sub-channels $g_{1,i}$ and $g_{2,i}$, $i = 1, \dots, L$. Chapter 2 showed that the required antenna separation can be on the order of a half-wavelength with 2-D isotropic scattering, but as much as several tens of wavelengths at a cellular base station.

6.12 MIMO Detection

Consider the MIMO arrangement shown in Fig. 6.22 consisting of L_T transmitter antennas and L_R receiver antennas. As described in Sect. 6.10, consider the case where the MIMO channel exhibits slow flat fading. The complex envelopes of the signals transmitted from the L_T transmit antennas are described by the vector

$$\tilde{\mathbf{s}}(t) = (\tilde{s}_1(t), \tilde{s}_2(t), \dots, \tilde{s}_{L_T}(t))^T, \quad (6.193)$$

where $\tilde{s}_p(t)$ is the signal transmitted from the p th transmit antenna. Likewise, the complex envelopes of the signals received at the L_R receiver antennas are described by the vector

$$\tilde{\mathbf{r}}(t) = (\tilde{r}_1(t), \tilde{r}_2(t), \dots, \tilde{r}_{L_R}(t))^T, \quad (6.194)$$

where $\tilde{r}_q(t)$ is the complex envelope received at the q th receiver antenna. It follows that

$$\tilde{\mathbf{r}}(t) = \mathbf{G}\tilde{\mathbf{s}}(t), \quad (6.195)$$

where \mathbf{G} is the $L_R \times L_T$ channel matrix.

Similar to the development in Sect. 5.1, a correlator matched filter detector may be assumed on each receiver antenna branch. Once again, a complex signal space representation is assumed with dimension $N = 1$, e.g., M -QAM or M -PSK. Let

$$\begin{aligned} \tilde{\mathbf{s}} &= (\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_{L_T})^T \\ \tilde{\mathbf{r}} &= (\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_{L_R})^T \\ \tilde{\mathbf{n}} &= (\tilde{n}_1, \tilde{n}_2, \dots, \tilde{n}_{L_R})^T \end{aligned} \quad (6.196)$$

denote the length- L_T transmitted signal vector, the length- L_R received signal vector, and the length- L_R received noise vector, respectively. It follows that

$$\tilde{\mathbf{r}} = \mathbf{G}\tilde{\mathbf{s}} + \tilde{\mathbf{n}}. \quad (6.197)$$

By processing the received signal vector $\tilde{\mathbf{r}}$, several different detectors are possible.

6.12.1 Maximum Likelihood Detection

The noise vector $\tilde{\mathbf{n}}$ consists of independent identically distributed complex Gaussian random variables, each with variance N_o . From Sect. 5.2, the maximum likelihood receiver decides on the signal vector $\tilde{\mathbf{s}}$ that maximizes the joint conditional density function $p(\tilde{\mathbf{r}}|\tilde{\mathbf{s}})$ which has a multivariate complex Gaussian distribution. It follows that the maximum likelihood receiver maximizes the decision metric

$$\begin{aligned} \mu(\tilde{\mathbf{s}}, \mathbf{G}) &= \|\tilde{\mathbf{r}} - \mathbf{G}\tilde{\mathbf{s}}\|^2 \\ &= \sum_{j=1}^{L_R} \left| \tilde{r}_j - \sum_{i=1}^{L_T} g_{j,i}\tilde{s}_i \right|^2 \end{aligned} \quad (6.198)$$

The maximum likelihood metric in (6.198) has high complexity even when L_T and L_R are relatively small, so many suboptimal detectors have been suggested in the literature. The simplest include the MMSE detector, and inverse channel detector (ICD).

6.12.2 Minimum Mean-Square-Error Detector

The MMSE detector performs linear combining on the received signal vector $\tilde{\mathbf{r}}$ to determine the transmitted signal vector $\tilde{\mathbf{s}}$. The linear combiner forms an estimate of $\tilde{\mathbf{s}}$, denoted as $\hat{\mathbf{s}}$, and represented in matrix form as

$$\hat{\mathbf{s}} = \mathbf{W}^H \tilde{\mathbf{r}} \quad (6.199)$$

where \mathbf{W} is an $L_R \times L_T$ weighting matrix. Each element of the vector $\hat{\mathbf{s}}$ is quantized to the nearest signal vector.

The weighting matrix \mathbf{W} is chosen to minimize the mean square error

$$J = \frac{1}{2} \mathbb{E}[\|\mathbf{s} - \mathbf{W}^H \tilde{\mathbf{r}}\|^2] \quad (6.200)$$

As with the optimal combiner in Sect. 6.8, the optimal weighting matrix, \mathbf{W}_{opt} , can be obtained by setting the gradient $\nabla_{\mathbf{W}} J$ to zero. This gives the MMSE solution

$$\nabla_{\mathbf{W}} J = \begin{bmatrix} \frac{\partial J}{\partial w_{1,1}}, \dots, \frac{\partial J}{\partial w_{L_R,1}} \\ \vdots, \dots, \vdots \\ \frac{\partial J}{\partial w_{1,L_T}}, \dots, \frac{\partial J}{\partial w_{L_R,L_T}} \end{bmatrix} = \mathbf{W} \Phi_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}} - \Phi_{\tilde{\mathbf{r}}\mathbf{s}} = 0. \quad (6.201)$$

where

$$\begin{aligned} \Phi_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}} &= \frac{1}{2} \mathbb{E}[\tilde{\mathbf{r}}\tilde{\mathbf{r}}^H] \\ &= \mathbf{G} \Phi_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}} \mathbf{G}^H + N_o \mathbf{I} \end{aligned} \quad (6.202)$$

is the $L_R \times L_R$ autocorrelation matrix of the received signal vector $\tilde{\mathbf{r}}$, and

$$\Phi_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}} = \frac{1}{2} \mathbf{E}[\tilde{\mathbf{s}}\tilde{\mathbf{s}}^H] \quad (6.203)$$

is the $L_T \times L_T$ signal correlation matrix, and

$$\Phi_{\tilde{\mathbf{s}}\tilde{\mathbf{r}}} = \frac{1}{2} \mathbf{E}[\tilde{\mathbf{s}}^H \tilde{\mathbf{r}}] \quad (6.204)$$

is the $L_T \times L_R$ cross-correlation matrix between $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{r}}$.

In the second line of (6.202), the signal and noise vectors are assumed zero mean and uncorrelated, and the noise vector $\tilde{\mathbf{n}}$ has covariance matrix

$$\frac{1}{2} \mathbf{E}[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H] = N_o \mathbf{I}_{L_R \times L_R} \quad (6.205)$$

The solution to (6.201) is

$$\mathbf{W}_{\text{opt}} = \Phi_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}}^{-1} \Phi_{\tilde{\mathbf{r}}\tilde{\mathbf{s}}} \quad (6.206)$$

Usually the signal vector $\tilde{\mathbf{s}}$ has uncorrelated, zero-mean, elements, so that $\Phi_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}$ in (6.202) reduces to a diagonal matrix.

6.12.3 Inverse Channel Detector

The inverse channel detector (ICD) also forms an estimate of $\tilde{\mathbf{s}}$ by linear combining. If $L_T = L_R$, then the weighting matrix is selected as $\mathbf{W} = \mathbf{G}^{-1}$ so that

$$\hat{\mathbf{s}} = \mathbf{W}\tilde{\mathbf{r}} = \mathbf{G}^{-1}\tilde{\mathbf{r}} = \tilde{\mathbf{s}} + \mathbf{G}^{-1}\tilde{\mathbf{n}} \quad (6.207)$$

As with the MMSE detector, each row of the matrix $\hat{\mathbf{s}}$ is quantized to the nearest signal vector. When $L_R \neq L_T$, the matrix \mathbf{G} is not square so its inverse does not exist. However, the pseudo-inverse can be used instead. If $L_R > L_T$ and the L_T columns of \mathbf{G} are linearly independent, then

$$\mathbf{W} = (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H. \quad (6.208)$$

Likewise, if $L_R < L_T$ and the L_R rows of \mathbf{G} are linearly independent, then

$$\mathbf{W} = \mathbf{G}^H (\mathbf{G} \mathbf{G}^H)^{-1}. \quad (6.209)$$

Note that the ICD detector removes all interchannel interference. However, this also implies that the ICD detector does not exploit any inherent diversity present in the received signal across the diversity branches. Moreover, there may be noise enhancement due to the multiplication of the noise vector $\tilde{\mathbf{n}}$ by the matrix inverse \mathbf{G}^{-1} in cases where the matrix \mathbf{G} is ill-conditioned or, likewise, by multiplication of the noise vector by the pseudo-inverses in (6.208) and (6.209).

6.12.4 Known Channel at the Transmitter and Receiver

In certain types of wireless systems, particularly those using time division duplexing (TDD), knowledge of the channel is available at both the transmitter and receiver. In this case, a singular value decomposition (SVD) of the channel matrix \mathbf{G} may be performed. Suppose that the channel matrix \mathbf{G} has rank r which is at most $\min\{L_T, L_R\}$. Then

$$\mathbf{G} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^H \quad (6.210)$$

where \mathbf{U} is an $L_R \times r$ matrix, \mathbf{V} is an $L_T \times r$ matrix, and $\mathbf{\Lambda}$ is an $r \times r$ diagonal matrix, such that the diagonal elements $\lambda_1, \lambda_2, \dots, \lambda_r$ are the singular values of the channel matrix \mathbf{G} . The matrices \mathbf{U} and \mathbf{V} are unitary matrices, meaning that $\mathbf{U}\mathbf{U}^H = \mathbf{I}_{r \times r}$ and $\mathbf{V}\mathbf{V}^H = \mathbf{I}_{r \times r}$ where $\mathbf{I}_{r \times r}$ is the $r \times r$ identity matrix.

Given knowledge that the channel matrix \mathbf{G} has rank r at the transmitter, r symbols are sent over the channel. The $r \times 1$ transmitted signal vector $\tilde{\mathbf{s}}$ is precoded at the transmitter by using the linear transformation

$$\tilde{\mathbf{s}}_p = \mathbf{V}\tilde{\mathbf{s}}. \quad (6.211)$$

The corresponding received signal vector is

$$\tilde{\mathbf{r}} = \mathbf{G}\tilde{\mathbf{s}}_p + \tilde{\mathbf{n}} = \mathbf{G}\mathbf{V}\tilde{\mathbf{s}} + \tilde{\mathbf{n}}. \quad (6.212)$$

At the receiver, the received signal vector $\tilde{\mathbf{r}}$ is processed by the linear transformation \mathbf{U}^H as follows:

$$\begin{aligned} \hat{\mathbf{s}} &= \mathbf{U}^H \tilde{\mathbf{r}} \\ &= \mathbf{U}^H \mathbf{G}\mathbf{V}\tilde{\mathbf{s}} + \mathbf{U}^H \tilde{\mathbf{n}} \\ &= \mathbf{U}^H \mathbf{U} \mathbf{\Lambda} \mathbf{V}^H \mathbf{V}\tilde{\mathbf{s}} + \mathbf{U}^H \tilde{\mathbf{n}} \\ &= \mathbf{\Lambda} \tilde{\mathbf{s}} + \mathbf{U}^H \tilde{\mathbf{n}}. \end{aligned} \quad (6.213)$$

From (6.213), it follows that there is no interchannel interference and the r data streams may be detected individually. Similar to the ICD detector, the SVD method does not exploit the inherent diversity provided by the channel. Rather the SVD method creates parallel streams of data, which increases the overall data rate. Multiplication of the noise vector $\tilde{\mathbf{n}}$ by the unitary matrix \mathbf{U}^H does not alter the statistics of the noise vector. Due to the multiplication of each transmitted symbol \tilde{s}_k by the corresponding singular value λ_k , the r data streams will have different received bit energy-to-noise ratios.

6.13 Spatial Modulation

Conventional MIMO communication systems transmit multiple data streams simultaneously from all transmit antennas. Both multiplexing and transmit-diversity gains can be obtained using MIMO approaches that lead to high spectral efficiency. Unfortunately, the simultaneous activation all transmit antennas and their associated RF circuitry and power amplifiers will not lead to good energy efficiency if power consumption at the transmitter is accounted for. Under realistic BS power consumption models, MIMO systems equipped with more than two active transmit antennas are unlikely provide any gains in energy efficiency (EE) [163]. In this context, spatial modulation (SM) has recently appeared to improve overall EE, by reducing the radiated power at the transmitter and required signal processing at the receiver. SM is a multi-antenna MIMO concept where the transmitter possesses a larger set of antenna elements than the number of RF chains. The main distinguishing feature of SM is that additional information bits are mapped onto an ‘‘SM constellation,’’ such that each constellation point is comprised of either one or a small subset of the antenna elements.

Figure 6.25 illustrates the difference between spatial multiplexing, transmit diversity, and spatial modulation for the case of $L_T = 2$ transmit antennas and a modulation alphabet size $M = 2$. With spatial multiplexing, the two modulation symbols \mathbf{s}_1 and \mathbf{s}_2 are transmitted simultaneously from the two different transmit antennas. For an arbitrary number of transmit antennas L_T and modulation alphabet size M , the modulation efficiency is $L_T \log_2 M$ bits/channel use. With transmit diversity, the two modulation symbols are transmitted in two channel uses. In Fig. 6.25, the symbols are transmitted using an Alamouti space-time block code. In general, for L_T transmit antennas and modulation alphabet size M , the space-time block code transmits N_M modulation symbols in N_c channel uses. Hence, the overall modulation efficiency is $(N_M/N_c) \log_2 M$ bits/channel use. Finally, with spatial modulation, one of the two modulation symbols \mathbf{s}_1 is explicitly transmitted, while the other modulation symbol \mathbf{s}_2 is implicitly transmitted by the transmit antenna selection. For L_T transmit antennas and a modulation alphabet size M , the modulation efficiency is $\log_2 L_T + \log_2 M$ bits/channel use. For the special case when $M = 1$, information is only transmitted in the activated antenna index. This scheme is called space-shift keying (SSK) and was the first spatial modulation scheme to appear in 2001 [58].

The modulation efficiency of spatial modulation increases as $\log_2 L_T$, whereas the modulation efficiency of spatial multiplexing increases linearly with L_T . This means that more antennas are needed with spatial modulation schemes to achieve the same modulation efficiency. As such, spatial modulation schemes may be more suitable for mm-wave wireless applications, where large arrays of antennas can be implemented in a small volume due to the short wavelength. However,

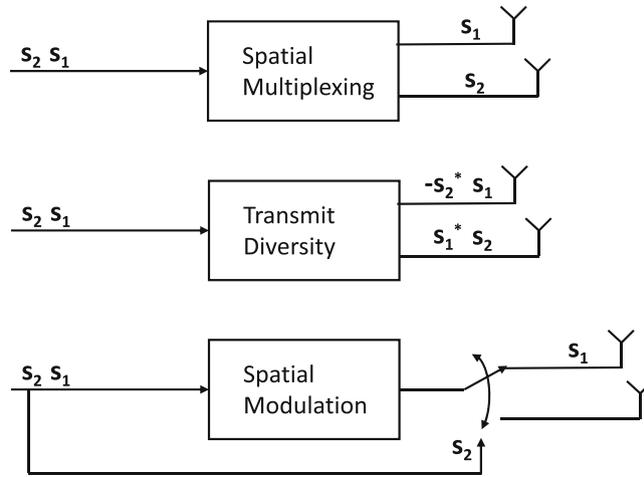


Fig. 6.25 Spatial multiplexing, spatial diversity, and spatial modulation

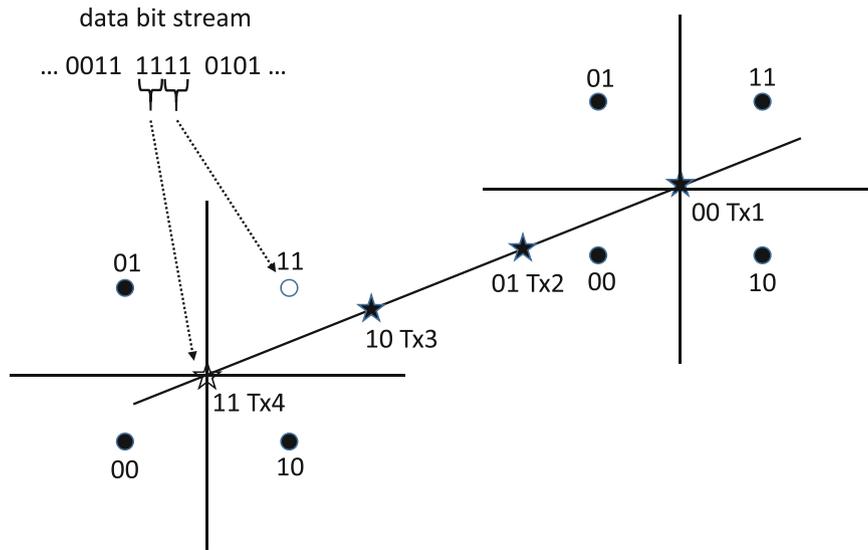


Fig. 6.26 Spatial modulation using 4 transmit antennas with QPSK or 4-QAM modulation

mm-wave cellular links require beam-forming and, therefore, the antenna array would have to provide both a beam-forming gain and a spatial multiplexing gain. As a result, a subset of the antenna elements would be activated for each channel use.

Figure 6.26 shows an example of a spatial modulation system with $L_T = 4$ transmit antennas (Tx1, Tx2, Tx3, Tx4) and an $M = 4$ point signal constellation for each possible transmit antenna. The scheme transmits 4 bits/channel use. The “spatial-constellation diagram” in Fig. 6.26 depicts that the first two bits of each binary data 4-tuple are used to select the antenna for transmission, while the last two bits are used to select the transmitted 4-QAM or QPSK symbol. Note that the activated transmit antenna may change with every channel use.

Spatial modulation works on the principle that the sub-channel impulse responses between each transmit antenna and the array of receiver antennas (possibly a single antenna) are uncorrelated to a certain degree, and can be distinguished. Suppose that the $L_T \times N$ signal matrix $\mathbf{e}_i^T \tilde{\mathbf{s}}_m$ is transmitted, where \mathbf{e}_i is a length L_T row vector of all-zeroes except for a one in the i th coordinate, indicating that antenna i is selected, and $\tilde{\mathbf{s}}_m$ is the length- N signal vector. Note that the signal matrix $\mathbf{e}_i^T \tilde{\mathbf{s}}_m$ contains all-zeroes except for the i th row. It is possible to implement spatial multiplexing schemes where more than one element in the antenna array is activated during each channel use. The number of RF chains required in the transmitter is equal to the number of activated antenna elements during each channel use.

For flat faded sub-channels, the $L_R \times N$ received signal matrix is

$$\tilde{\mathbf{R}} = \mathbf{G} \mathbf{e}_i^T \tilde{\mathbf{s}}_m \tag{6.214}$$

where \mathbf{G} is the $L_R \times L_T$ MIMO channel matrix. By using knowledge of the set of possible transmitted waveforms and knowledge of the sub-channel impulse/frequency responses between each transmit antenna and each receiver antenna, the maximum likelihood or minimum distance receiver can search over the L_T transmit antennas and M modulation symbols to find the solution closest in squared Euclidean distance to the received vector of L_R waveforms. That is, using knowledge of \mathbf{G} , the receiver chooses $\mathbf{e}_i^T \tilde{\mathbf{s}}_m$ to minimize the squared Euclidean distance metric

$$\mu(\mathbf{e}_i^T \tilde{\mathbf{s}}_m) = \|\tilde{\mathbf{R}} - \mathbf{G}\mathbf{e}_i^T \tilde{\mathbf{s}}_m\|^2 \quad (6.215)$$

$$= \sum_{j=1}^{L_R} \sum_{n=1}^N |\tilde{r}_{jn} - g_{j,i} \tilde{s}_{m_n}|^2 \quad (6.216)$$

Spatial modulation has a number of advantages and disadvantages compared to spatial multiplexed MIMO systems. The receiver is simpler with spatial modulation, since only a single transmitter antenna is active at any time. As a result, there is no interchannel interference to mitigate. The transmitter is simpler as well, since only a small number of RF chains are needed along with an RF switching mechanism. Fortunately, such RF switching mechanisms exist that can operate at nanosecond or even at subnanosecond speeds. However, since the RF switching occurs with each modulated symbol, there are limitations on the allowable length of the amplitude shaping pulse $h_a(t)$. Section 4.9.2 showed that the truncation of a root-raised cosine pulse to a short length, for example, will result in significant side lobe generation in the transmitted power spectrum. Spatial modulation, like other MIMO schemes, relies on favorable channel conditions such that the columns in the $L_R \times L_T$ MIMO channel matrix are distinguishable. As with other modulation schemes, the bit error rate performance will depend on the minimum Euclidean distance between points in the spatial-constellation. Finally, the reduced number of RF chains used in the spatial modulation transmitter poses problems for channel estimation, since only a subset of the MIMO sub-channels are active and can be observed at the same time.

6.14 Massive MIMO

Very large MIMO or massive MIMO systems use base station antenna arrays with an order of magnitude more elements than conventional MIMO systems, perhaps a hundred antenna elements or more, while the mobile stations use single antennas. The concept of massive MIMO was eloquently described in the seminal work of Marzetta [218]. A large number of antennas allow the antenna beam pattern at the base station to be adjusted so as to sharply focus the radiated energy in a small region of space at the intended receiver antenna by adding together the signals transmitted from the multiple antennas constructively at the desired receiver, but destructively elsewhere at the other unintended receivers.

Massive MIMO works on the principle of *favorable propagation*, which occurs when the channel vectors between the BS and each MS are nearly orthogonal as the number of BS antennas becomes very large. Favorable propagation allows for simple linear precoding processing schemes, such as matched filtering, to be used that give almost the same performance as more complex precoding schemes such as zero-forcing precoding or dirty paper coding. With a large number of BS antennas, L , the power that is transmitted by each BS antenna element is reduced by the factor $1/L$. Hence, as L becomes large the expensive high-power RF components used in traditional MIMO systems can be replaced by a larger number of low-power inexpensive RF components. Therefore, the constraints on the accuracy and linearity of each RF chain are reduced and, moreover, the system is robust to the failure of any one antenna element. Finally, the large number of antennas tends to reduce the depths of envelope fades through spatial averaging and makes the overall channel less time and frequency selective. This is important for multi-carrier schemes such as OFDM, since the effects of frequency-selectivity are removed and the computationally complex algorithms for power and bit allocation on the OFDM subcarriers to maximize the overall bit rate are no longer needed.

Massive MIMO has its challenges. To estimate the channels with massive MIMO, time-division duplex (TDD) operation is used with reverse link pilot sequences to enable the BS to estimate the reciprocal forward and reverse link channels. For the purpose of channel estimation a distinct pilot sequence is assigned to each MS in a cell, such that the pilots transmitted by the MSs in the same cell are orthogonal in time and frequency. The conjugate-transpose of the channel estimates are then used for linear precoding and combining, respectively, on the forward and reverse links. Neither the BS nor the MSs have any prior knowledge of the channels. Therefore, all channel state information (CSI) is acquired from reverse-link pilots which must be scheduled, along with forward- and reverse-link data transmissions, in a time interval (called a slot) over which the channel can be assumed to be constant. The slot duration can be set equal to the channel coherence time, which is inversely proportional to the mobility or speed of the MSs. Thus, the number of users that can be served in a particular cell

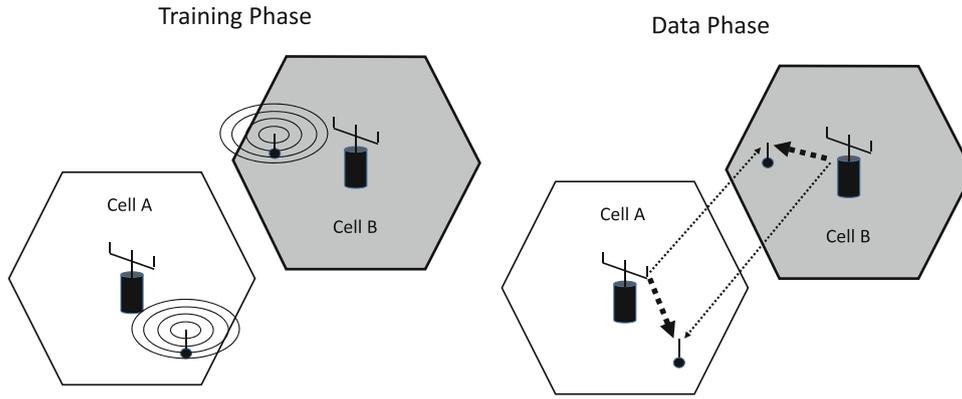


Fig. 6.27 Pilot contamination in massive MIMO. During the pilot phase (*left*) the BS overhears the pilot transmissions from MSs in other cells. During the data phase (*right*) the BS partially beamforms to the MSs it has overhead in other cells [285]

is determined by the number of time orthogonal pilot sequences that can be assigned. Since the number of MSs that can be served is proportional to the time spent sending pilots, and the sum-rate is proportional to the number of MSs served, the net sum-rate is maximized by spending one half of the slot sending pilots and the other half of the slot sending data [217]. When the channel exhibits delay spread, the number of MSs that can be served is equal to one-half of the channel coherence time divided by the channel delay spread.

With massive MIMO, there is no sharing of channel information between BSs and no power control is used. Since the number of orthogonal pilot sequences is limited, the pilot sequences are reused in the different cells of a multi-cell system according to some pilot re-use factor. When estimating the channels to the MS it serves in the pilot period, a BS inadvertently learns the channel to MSs in other cells that reuse the same pilot sequence. This phenomenon is called *pilot contamination*. Thus, on the forward link channels, part of the energy sent to a particular MS will also be directed to an unintended MS in another co-channel cell during the data period. This concept is illustrated in Fig. 6.27. Likewise, when the BS combines its reverse-link signals to receive the individual data transmissions of the MSs it serves, it is also coherently combining the signals from MSs in other co-channel cells. This out-of-cell interference on the forward and reverse links persists even when the number of BS antennas becomes very large and presents a fundamental performance limitation in massive MIMO systems. Millimeter wave frequencies admit themselves well to massive MIMO in regards to antenna technology since the antenna spacing is small. However, the interference models for UHF frequencies and mm-wave frequencies are different. With classical massive MIMO, the MS typically have a single antenna. However, at mm-wave frequencies both the BS and MS employ MIMO arrays to achieve sufficient transmit and receive antenna gain in the link budget. The means the antenna aperture is small such that the probability of interference is small and the mm-wave link budget is typically noise limited rather than interference limited.

6.14.1 Massive MIMO System Model

Consider a massive MIMO cellular multi-access system consisting of a collection of cells, where the BSs at the center of each cell have L omnidirectional antennas and each BS serves K_{MS} MSs per cell, and where the MSs have a single omnidirectional antenna. The $L \times K_{\text{MS}}$ matrix of channel gains between the j th BS and the K_{MS} MSs in cell ℓ is denoted as $\mathbf{H}_{j,\ell}$ which is assumed to have the form

$$\mathbf{H}_{j,\ell} = \mathbf{G}_{j,\ell} \mathbf{\Omega}_{j,\ell}^{1/2}. \quad (6.217)$$

The elements of the matrix $\mathbf{G}_{j,\ell} = [g_{j,\ell}]_{m,k} = g_{j,\ell,m,k}$ are assumed to be zero-mean complex Gaussian random variables with an envelope power of unity, i.e., $E[|g_{j,\ell,m,k}|^2] = 1$, which is characteristic of Rayleigh fading. The matrix $\mathbf{\Omega}_{j,\ell}$ is a $K_{\text{MS}} \times K_{\text{MS}}$ diagonal matrix accounting for the effects of shadowing and path loss, where the k diagonal element of the matrix, denoted as $\Omega_{j,\ell,k}$ represents the received local mean envelope power, $\Omega_{j,\ell,k} = E[|h_{j,\ell,m,k}|^2]$, between the j th BS and MS k in cell ℓ , assumed to be the same for all L antenna elements for the ℓ th BS. The signal model is shown in Fig. 6.28, showing the propagation coefficient $h_{j,\ell,m,k}$ between the k th MS in Cell ℓ and the m th antenna element in Cell j .

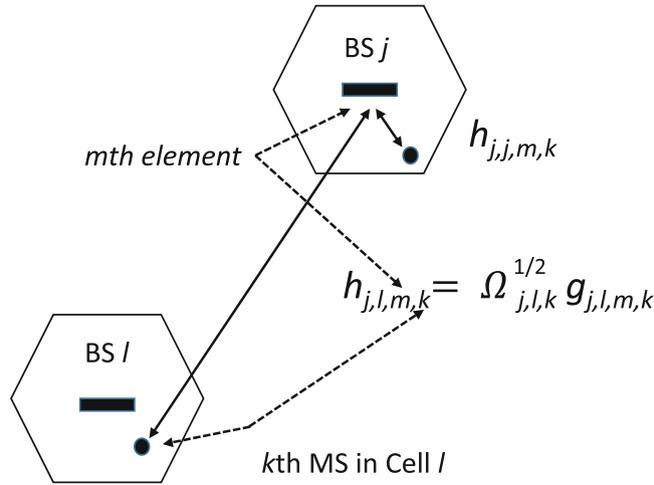


Fig. 6.28 Propagation coefficient between k th MS in Cell ℓ and the m th antenna element in Cell j

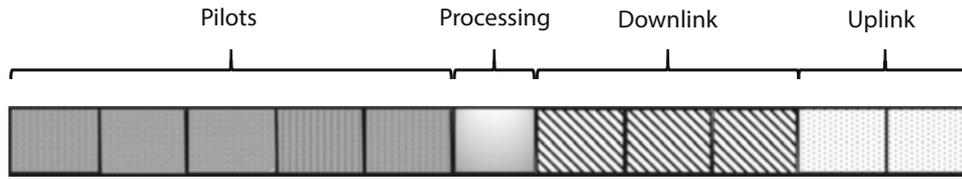


Fig. 6.29 Time domain aligned pilots for massive MIMO. In this example, the slot consists of 5 sub-slots for pilots, one sub-slot for pilot processing, 3 sub-slots for downlink transmissions, and 2 slots for uplink transmissions from [125]

Using the simple path loss model in (1.4), the $\Omega_{j,\ell,k}$ are log-normally distributed and distance dependent according to

$$\Omega_{j,\ell,k}(d) = \frac{\mu_{\Omega_p}(d_o)\epsilon}{(d/d_o)^\beta}, \quad (6.218)$$

where $\mu_{\Omega_p}(d_o) = E[\Omega_p(d_o)]$ is the average received signal power at the reference distance d_o , β is the propagation path loss exponent, and ϵ is a log-normal random variable, such that $\epsilon_{(\text{dB})}$ is a zero-mean Gaussian random variable with standard deviation equal to σ_Ω .

6.14.2 Reverse Link Pilots

Reverse-link pilots are required to estimate the channel for both forward and reverse data transmission. The same band of frequencies is reused in a multiplicity of cells according to a frequency reuse plan. Consequently, the reverse link pilots that are received at a particular base station will be contaminated by the reverse link pilots that are transmitted by MSs in other cells sharing the same time-frequency resource. It is assumed that a total of N base stations in the deployment share the same band of frequencies, and the same set of K_{MS} pilot signals, one for each MS in their cells. Moreover, the forward and reverse link transmissions are assumed to be perfectly synchronized, which constitutes the worst case condition for pilot contamination. The cells use the time-domain aligned pilot (TDAP) structure shown in Fig. 6.29. First, pilots are transmitted by the MSs. The BSs then process the received pilots to extract channel estimates. The channel estimates are used by the BSs to precode the data transmitted on the downlink and to perform matched filtering on the uplink transmissions from the MSs.

Let $\hat{\mathbf{H}}_{j,j}$ denote the estimate of the $L \times K_{\text{MS}}$ channel matrix between the K_{MS} MSs in the j th cell and the L antennas at the j th BS. Accordingly,

$$\hat{\mathbf{H}}_{j,j} = \sqrt{\gamma_p} \sum_{\ell=1}^N \mathbf{H}_{j,\ell} + \mathbf{v}_j, \quad (6.219)$$

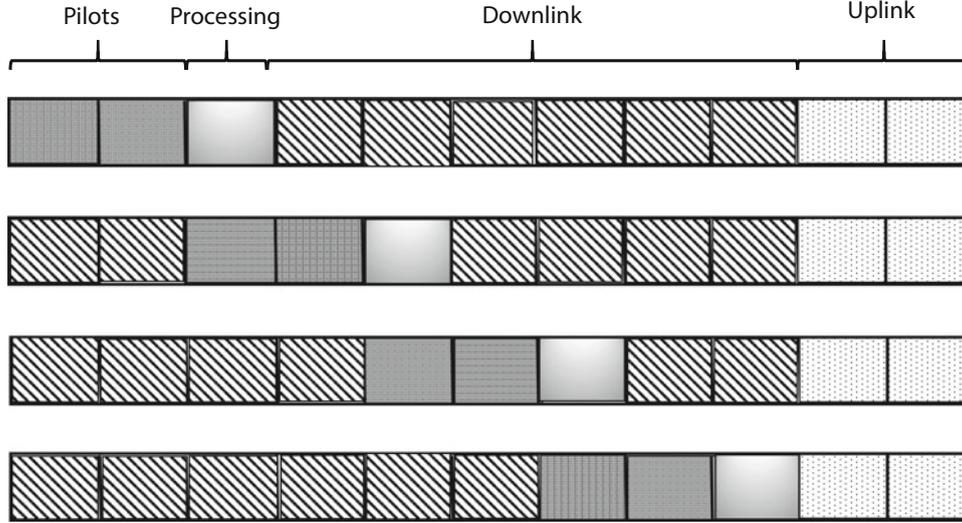


Fig. 6.30 Time domain shifted pilots for massive MIMO. In this example, the slot consists of 11 sub-slots, such that 2 sub-slots are used for pilots in each cell group, 1 sub-slot is used for processing, 6 sub-slots are used for downlink data, and 2 sub-slots are used for uplink data, from [125]

where, again, $\mathbf{H}_{j,\ell}$ is the $L \times K_{\text{MS}}$ propagation matrix between the K_{MS} MSs in the ℓ th cell and the L antennas at the base station in the j th cell, and \mathbf{v}_j is an $L \times K_{\text{MS}}$ receiver noise matrix, whose components are zero-mean unit-variance complex Gaussian random variables that are mutually uncorrelated under the assumption of orthogonal pilots and uncorrelated with the propagation matrices $\mathbf{H}_{j,\ell}$. The quantity γ_p is the pilot signal-to-noise ratio (SNR), which will be shown later to be irrelevant as the effects of noise vanish when the number of uncorrelated BS antennas, L , becomes unbounded.

Pilot contamination is the key bottleneck in massive MIMO systems. Fortunately, pilot contamination can be mitigated by using a number of approaches that will work better than the simple TDAP design discussed above. One possibility is the time domain shifted pilot (TDSP) design that mitigates pilot contamination through asynchronous transmission among adjacent cells as shown in Fig. 6.30 [125]. The TDAP scheme partitions the cellular deployment into groups of cell groups labelled A_1, A_2, \dots, A_{N_p} . In Fig. 6.30, there are $N_p = 4$ cell groups. Communication during each coherence interval is divided into stages. During the first stage, the MSs from cells belonging to cell group A_1 transmit their pilot sequences simultaneously, while MSs from all other groups receive their downlink data. Once MSs in A_1 finish their pilot sequence, they start receiving downlink data while MSs belonging to a different cell group start sending their pilots. This procedure is repeated until MSs in all cell groups have transmitted their pilots. Afterwards, all MSs switch to transmitting uplink data to their BTSs. With unbounded BS antennas, TDSP can remove the pilot contamination from cells adopting non-overlapping pilot phase in uplink transmission, but it still has mutual interference between data and pilot in downlink transmission [125].

6.14.3 Reverse Link Data

The K_{MS} MSs in each cell transmit their data over the reverse channel to their respective base stations. The BSs use estimates of the channel matrices, $\hat{\mathbf{H}}_{j,j}$ to perform maximal ratio combining. Again, assuming perfect synchronization of the signals received from the N MSs sharing the same time-frequency resource, the $L \times 1$ received signal vector at the L antennas of the j th BS is

$$\tilde{\mathbf{r}}_j = \sqrt{\gamma_r} \sum_{\ell=1}^N \mathbf{H}_{j,\ell} \tilde{\mathbf{s}}_\ell + \tilde{\mathbf{n}}_j, \quad (6.220)$$

where $\tilde{\mathbf{s}}_\ell$ is the $K_{\text{MS}} \times 1$ vector of K_{MS} modulation symbols transmitted from the K_{MS} MSs in the ℓ th cell, $\tilde{\mathbf{n}}_j$ is the $L \times 1$ received noise vector whose components are zero-mean, mutually uncorrelated, and uncorrelated with the propagation matrices. Note that in (6.220) it is assumed that the complex signal constellation has a dimensionality of unity, such that all signal vectors are complex-valued scalars. The parameter γ_r in (6.220) is the symbol-energy-to-noise ratio which, like the pilot SNR, will be shown later to be irrelevant as the number of uncorrelated BS antennas, L , becomes unbounded.

The j th BS processes the received vector $\tilde{\mathbf{r}}_j$ by multiplying it by the conjugate-transpose of the channel estimate to perform maximal ratio combining. From (6.219) and (6.220),

$$\begin{aligned}\tilde{\mathbf{r}}_j &= \hat{\mathbf{H}}_{j,j}^H \tilde{\mathbf{r}}_j \\ &= \left(\sqrt{\gamma_p} \sum_{\ell_1=1}^N \mathbf{H}_{j,\ell_1} + \mathbf{v}_j \right)^H \left(\sqrt{\gamma_r} \sum_{\ell_2=1}^N \mathbf{H}_{j,\ell_2} \tilde{\mathbf{s}}_{\ell_2} + \tilde{\mathbf{n}}_j \right).\end{aligned}\quad (6.221)$$

The components of the vector $\tilde{\mathbf{r}}_j$ consist of the sum of inner products between random vectors of length L . As the number of BS antennas becomes very large, the squared-length or L_2 norm of these vectors grows with L , while the inner products of uncorrelated vectors, by assumption of favorable propagation conditions, grow at a lesser rate. For very large L , only the products of identical terms in the bracketed expressions in (6.221) are significant. From (6.217) and (6.218),

$$\frac{1}{L} \mathbf{H}_{j,\ell_1}^H \mathbf{H}_{j,\ell_2} = \boldsymbol{\Omega}_{j,\ell_1}^{1/2} \left(\frac{\mathbf{G}_{j,\ell_1}^H \mathbf{G}_{j,\ell_2}}{L} \right) \boldsymbol{\Omega}_{j,\ell_2}^{1/2}, \quad (6.222)$$

where, once again, $\boldsymbol{\Omega}_{j,\ell_1}$ and $\boldsymbol{\Omega}_{j,\ell_2}$ are $K_{\text{MS}} \times K_{\text{MS}}$ diagonal matrices. As the number of BS antennas L grows without bound,

$$\frac{\mathbf{G}_{j,\ell_1}^H \mathbf{G}_{j,\ell_2}}{L} \longrightarrow \mathbf{I}_{K_{\text{MS}} \times K_{\text{MS}}} \delta_{\ell_1, \ell_2} \quad (6.223)$$

due to the assumption of favorable propagation conditions. Substituting (6.223) and (6.222) into (6.221) gives

$$\frac{1}{L \sqrt{\gamma_p \gamma_r}} \tilde{\mathbf{r}}_j \longrightarrow \sum_{\ell=1}^N \boldsymbol{\Omega}_{j,\ell} \tilde{\mathbf{s}}_{\ell}, \quad j = 1, \dots, N. \quad (6.224)$$

The k th element of the vector $\tilde{\mathbf{r}}_j$ normalized by $1/L \sqrt{\gamma_p \gamma_r}$ is

$$\frac{1}{L \sqrt{\gamma_p \gamma_r}} \tilde{r}_{j,k} \longrightarrow \Omega_{j,k,j} \tilde{s}_{j,k} + \sum_{\substack{\ell=1 \\ \ell \neq j}}^N \Omega_{j,k,\ell} \tilde{s}_{\ell,k}, \quad k = 1, \dots, K_{\text{MS}} \quad (6.225)$$

where $\Omega_{j,k,\ell}$ is the k th diagonal element of the $K_{\text{MS}} \times K_{\text{MS}}$ diagonal matrix $\boldsymbol{\Omega}_{j,\ell}$. The noise terms due to \mathbf{v}_j and $\tilde{\mathbf{n}}_j$ in (6.224) vanish as L goes to infinity, due to the multiplicative $1/L$ factor. Moreover, the effects of fast envelope fading are completely eliminated, due to spatial (antenna) averaging, and the transmissions from the K_{MS} MSs in each cell do not interfere with each other. The second term in the R.H.S. of (6.225) represents co-channel interference received from MSs in other cells sharing the same pilot. Thus, with idealized (synchronized) massive MIMO, the interference-to-noise ratio is infinite and the signal-to-interference-plus noise ratio (SINR) becomes equal to the signal-to-interference ratio (SIR). In the R.H.S. of (6.225), both the desired symbol \tilde{s}_k and interfering symbols in the summation are multiplied by their respective local mean powers $\Omega_{j,k}$ that are received at the j th BS. Consequently, due to the vanishing effects of noise, the SIR at the j th BS involves the squares of the $\Omega_{j,k,j}$ and $\Omega_{j,k,\ell}$, and is given by

$$\text{SIR}_{r,k} = \frac{\Omega_{j,k,j}^2}{\sum_{\substack{\ell=1 \\ \ell \neq j}}^N \Omega_{j,k,\ell}^2}, \quad k = 1, \dots, K_{\text{MS}}, \quad (6.226)$$

where the assumption is made that the co-channel interferers sharing the same time-frequency resource block are uncorrelated, which is easily satisfied if their associated information symbol sequences are uncorrelated.

6.14.4 Forward Link Data

The forward channel uses precoding such that each BS transmits a length- L vector from its L transmit antennas, which is proportional to the complex conjugate transpose of its propagation matrix as estimated from reverse link pilots. The transmission from the j th BS to its k th MS suffers from co-channel interference from the forward link BS transmissions in co-channels cells to their respective MSs that are reusing the same time-frequency resource.

The ℓ th BS transmits the $L \times 1$ vector, $\hat{\mathbf{H}}_{\ell,\ell}^H \tilde{\mathbf{s}}_\ell$, where $\tilde{\mathbf{s}}_\ell$ is a length- K_{MS} column vector of information symbols transmitted to the K_{MS} MS served by BS ℓ . In practice, there is a normalizing factor corresponding to the transmit power constraints of the BSs. Under the assumption that the normalizing factor is the same for all BSs, the value of the normalizing factor becomes irrelevant as the number of BS antennas, L , becomes unbounded.

The K_{MS} MSs served by the j th BS each receive their respective component of a length K_{MS} vector that is comprised of the signals transmitted from each of the BS in the N surrounding co-channel cells. That is,

$$\begin{aligned} \bar{\mathbf{r}}_j &= \sqrt{\gamma_f} \sum_{\ell=1}^N \hat{\mathbf{H}}_{\ell,j}^T \mathbf{H}_{\ell,\ell}^H \tilde{\mathbf{s}}_\ell + \mathbf{w}_j \\ &= \sqrt{\gamma_f} \sum_{\ell_1=1}^N \hat{\mathbf{H}}_{\ell_1,j}^T \left(\sqrt{\gamma_p} \sum_{\ell_2=1}^N \mathbf{H}_{\ell_1,\ell_2} + \mathbf{v}_{\ell_1} \right)^H \tilde{\mathbf{s}}_{\ell_1} + \mathbf{w}_j, \end{aligned} \quad (6.227)$$

where \mathbf{w}_j is a $K_{\text{MS}} \times 1$ vector of uncorrelated noise components, and γ_f is a measure of the forward channel SNR which will be shown to be irrelevant as the number of BS antennas L grows without bound. As with the reverse channel, as the number of BS antennas L goes to infinity, (6.222) and (6.223) are once again used to yield

$$\frac{1}{L\sqrt{\gamma_p\gamma_f}} \bar{\mathbf{r}}_j \rightarrow \sum_{\ell=1}^N \Omega_{\ell,j} \tilde{\mathbf{s}}_\ell, \quad j = 1, \dots, N. \quad (6.228)$$

The k th MS in the j th cell receives

$$\frac{1}{L\sqrt{\gamma_p\gamma_f}} \bar{r}_{kj} \rightarrow \Omega_{j,k,j} \tilde{s}_{k,j} + \sum_{\substack{\ell=1 \\ \ell \neq j}}^N \Omega_{\ell,k,j} \tilde{s}_{k,\ell}, \quad j = 1, \dots, N. \quad (6.229)$$

As with the reverse link, the effective signal-to-interference ratio for the k th MS in the j th cell is

$$\text{SIR}_{f,k} = \frac{\Omega_{j,k,j}^2}{\sum_{\substack{\ell=1 \\ \ell \neq j}}^N \Omega_{\ell,k,j}^2}, \quad k = 1, \dots, K_{\text{MS}} \quad (6.230)$$

Although the forward and reverse SIRs in (6.230) and (6.226), respectively, look similar, they are in fact different meaning that the forward and reverse channels exhibit link imbalance. The numerator term of (6.230) and (6.226) in each case is the same. However, the denominators are different. In the case of the reverse channel, the denominator is the sum of squares of the local means $\Omega_{\ell,k,j}$ from the k th MS in the $N-1$ other cells than the j th cell. These local means are statistically uncorrelated. In the case of the forward channel, the denominator is the sum of squares of the local means $\Omega_{\ell,k,j}$ from the $N-1$ BSs other than the j th BS that are transmitting to their respective k th MSs and received at the k th MS in the j th cell. These local means are statistically uncorrelated as well. However, the set of propagation distances to the co-channel interferers are not the same for the forward and reverse links, meaning their respective area mean powers are different. Therefore, in the absence of power control, forward and reverse link imbalance exists as in Chap. 1, Sect. 1.7.

6.14.5 Favorable Propagation Conditions

Massive MIMO systems rely upon favorable propagation conditions in (6.223). This section explores the concept further to define the circumstances under which favorable propagation conditions occur. To simplify the treatment, consider a single isolated cell. Let \mathbf{h}_k be the length- L row vector representing the propagation vector between the k th MS and the L BS antennas. Favorable propagation implies that as L becomes very large

$$\frac{\mathbf{h}_k^H \mathbf{h}_j}{L} \longrightarrow \Omega_k \delta_{kj}. \quad (6.231)$$

This means that the propagation vectors become mutually orthogonal. Another way of writing (6.231) is to observe that

$$\frac{\mathbf{H}^H \mathbf{H}}{L} \longrightarrow \mathbf{\Omega} \quad (6.232)$$

where $\mathbf{\Omega}$ is a $K_{\text{MS}} \times K_{\text{MS}}$ diagonal matrix whose k th diagonal element is Ω_k . One method for ascertaining favorable propagation is the condition number of the matrix \mathbf{H} , defined as the ratio of the maximum and minimum singular values of the associated $K_{\text{MS}} \times K_{\text{MS}}$ Gram matrix $\mathbf{H}^H \mathbf{H}$. The singular values of \mathbf{H} are equal to the square root of the eigenvalues of the Gram matrix $\mathbf{H}^H \mathbf{H}$. If the condition number is unity, then favorable propagation is achieved. From the R.H.S. of (6.232), it is apparent that if the local mean Ω_k happens to be the same for all K_{MS} MSs, then the condition number is equal to 1. However, the Ω_k are typically different, so that condition number is not always meaningful. Moreover, the condition number as a measure of favorable propagation only considers the largest and smallest eigenvalues, and disregards all the other eigenvalues.

The sum capacity of the massive MIMO channel in bits per channel use or bits per symbol is

$$C = \log_2 \left| \mathbf{I} + \frac{\gamma}{L} \mathbf{H}^H \mathbf{H} \right|, \quad (6.233)$$

where γ is the symbol energy-to-noise ratio. Using the Hadamard inequality, it follows that

$$\begin{aligned} C &= \log_2 \left| \mathbf{I} + \frac{\gamma}{L} \mathbf{H}^H \mathbf{H} \right| \\ &\leq \log_2 \left(\prod_{k=1}^{K_{\text{MS}}} \left[\mathbf{I} + \frac{\gamma}{L} \mathbf{H}^H \mathbf{H} \right]_{k,k} \right) \\ &= \sum_{k=1}^{K_{\text{MS}}} \log_2 \left(\left[\mathbf{I} + \frac{\gamma}{L} \mathbf{H}^H \mathbf{H} \right]_{k,k} \right) \\ &= \sum_{k=1}^{K_{\text{MS}}} \log_2 \left(1 + \frac{\gamma}{L} \|\mathbf{h}_k\|^2 \right). \end{aligned} \quad (6.234)$$

This motivates the *distance from favorable propagation* as a metric for favorable propagation, defined as [243]

$$\Delta_C = \frac{\sum_{k=1}^{K_{\text{MS}}} \log_2 \left(1 + \frac{\gamma}{L} \|\mathbf{h}_k\|^2 \right) - \log_2 \left| \mathbf{I} + \frac{\gamma}{L} \mathbf{H}^H \mathbf{H} \right|}{\log_2 \left| \mathbf{I} + \frac{\gamma}{L} \mathbf{H}^H \mathbf{H} \right|}. \quad (6.235)$$

Observe that $\Delta_C \geq 0$ and ideal propagation occurs when $\Delta_C = 0$.

Yet another possibility is the *factor of favorable propagation*, defined as the ratio of the capacity to the Hadamard upper bound on capacity, viz.,

$$R_C = \frac{\log_2 \left| \mathbf{I} + \frac{\gamma}{L} \mathbf{H}^H \mathbf{H} \right|}{\sum_{k=1}^{K_{\text{MS}}} \log_2 \left(1 + \frac{\gamma}{L} \|\mathbf{h}_k\|^2 \right)}, \quad 0 \leq R_C \leq 1. \quad (6.236)$$

where $R_C = 1$ means that favorable propagation occurs.

For the remainder of this section, shadowing is ignored and only envelope fading is considered. $\Omega_k = 1$ for all k MSs, so that $h_k = g_k$ and $\mathbf{H} = \mathbf{G}$ in what follows.

6.14.5.1 Independent Rayleigh Fading

For practical applications L is finite, so that the L.H.S. of (6.231) is a random variable. The g_k are statistically independent. Therefore, $\mathbf{g}_k^H \mathbf{g}_j / L$ has a mean of zero and variance equal to

$$\text{Var} \left(\frac{\mathbf{g}_k^H \mathbf{g}_j}{L} \right) = \frac{1}{L}$$

If matched filtering is used, then the desired signal power at the massive MIMO combiner output is proportional to $\|\mathbf{g}_k\|^4$, while the power of each co-channel interferer is proportional to $\|\mathbf{g}_k^H \mathbf{g}_j\|^2$. The interfering power due to the j th interferer $\|\mathbf{g}_k^H \mathbf{g}_j\|^2 / L^2$ has a mean of zero and a variance of

$$\text{Var} \left(\frac{\|\mathbf{g}_k^H \mathbf{g}_j\|^2}{L^2} \right) = \frac{L+2}{L^3} \approx \frac{1}{L^2}$$

The cdf function of the singular values of \mathbf{G} can be used to predict whether or not a given channel realization \mathbf{G} exhibits favorable propagation. The cdf of the sorted singular values resulting from 10,000 random realizations of the $L \times K_{\text{MS}}$ matrix \mathbf{G} with $L = 100$ and $K_{\text{MS}} = 10$ is plotted in Fig. 6.31. Observe that with i.i.d. Rayleigh fading, the singular values tend to be uniformly spread between the minimum and maximum singular values.

6.14.5.2 Uniform Random Line-of-Sight

Uniform Random Line-of-Sight (UR-LoS) conditions represent the other extreme of scattering environments where free space line-of-sight propagation exists between the BS and MSs. Consider a uniform linear array of antennas at the BS with antenna element spacing δ m as described in Sect. 6.9. From (6.148), the propagation vector \mathbf{g}_k can be written as

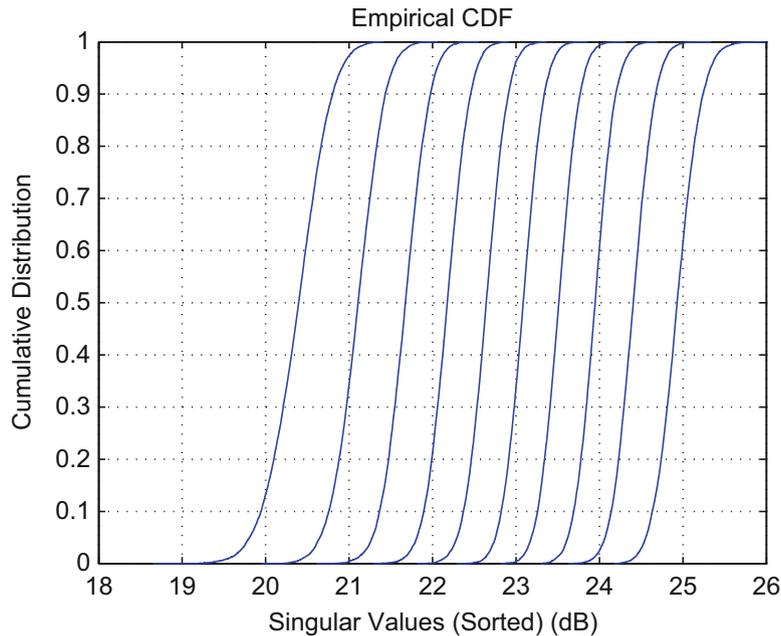


Fig. 6.31 Cdfs of sorted singular values of \mathbf{G} for i.i.d. Rayleigh fading with $L = 100$, $K_{\text{MS}} = 10$. For $K_{\text{MS}} = 10$ MSs, there are 10 sorted singular values and 10 corresponding cdf curves

$$\mathbf{g}_k = \left(1, e^{j2\pi\left(\frac{\delta}{\lambda c}\right)\cos(\theta_k)}, \dots, e^{j2\pi\left(\frac{k\delta}{\lambda c}\right)\cos(\theta_k)}, \dots, e^{j2\pi\left(\frac{(L-1)\delta}{\lambda c}\right)\cos(\theta_k)} \right) \quad (6.237)$$

where θ_k is the array orientation angle with respect to the k th MS, assumed to be independently and uniformly distributed on the interval $[-\pi, \pi)$ for each MS. If $\delta = \lambda/2$, then $\mathbf{g}_k^H \mathbf{g}_j / L$ has a mean of zero and variance

$$\text{Var} \left(\frac{\mathbf{g}_k^H \mathbf{g}_j}{L} \right) = \frac{1}{L} - \frac{1}{L^2}.$$

Compared to the case of independent Rayleigh fading, the variance is almost the same for large values of L . The squared magnitude of the inner product, $\|\mathbf{g}_k^H \mathbf{g}_j\|^2 / L^2$, has a mean of zero and variance

$$\text{Var} \left(\frac{\|\mathbf{g}_k^H \mathbf{g}_j\|^2}{L^2} \right) = \frac{(L-1)L(2L-1)}{3L^4} \approx \frac{2}{3L}$$

Hence, the variance of the interfering power decays more slowly with UR-LoS conditions than independent Rayleigh fading. The cdf of the sorted singular values resulting from 10,000 random realizations of the $L \times K_{\text{MS}}$ matrix \mathbf{G} with $L = 100$ and $K_{\text{MS}} = 10$ is plotted in Fig. 6.32. Notice from Fig. 6.32 that a few singular values are small with high probability, while the remainder are concentrated around their median value. The small singular values occur when two MSs have array orientation angles θ_k and θ_j that are very close, such that $|\theta_k - \theta_j| \approx 1/L$. Under this condition the normalized inner product $\mathbf{g}_k^H \mathbf{g}_j / L$ does not converge to 0 as L becomes large as shown by

$$\frac{\mathbf{g}_k^H \mathbf{g}_j}{L} = \frac{1}{L} \frac{1 - e^{j\pi(\sin(\theta_k) - \sin(\theta_j))L}}{1 - e^{j\pi(\sin(\theta_k) - \sin(\theta_j))}} = \frac{1}{L} \frac{1 - e^{j\pi}}{1 - e^{j\pi/L}} \rightarrow \frac{j2}{\pi} \neq 0. \quad (6.238)$$

In order to recover favorable propagation conditions it is necessary to drop a typically small number, often one or two, MSs that cause unfavorable propagation conditions. Essentially, to guarantee that the propagation conditions are favorable, each BS antenna beam must contain at most one MS. For UR-LoS conditions, Ngo, Larsson, and Marzetta derived the probability of dropping n MSs as [243]

$$P(N_{\text{drop}} = n) = \binom{L}{N + L - K_{\text{MS}}} \sum_{k=1}^{K_{\text{MS}}-n} (-1)^k \binom{K_{\text{MS}}-n}{k} \left(1 - \frac{n + L - K_{\text{MS}} + k}{L} \right)_{\text{MS}}^K \quad (6.239)$$

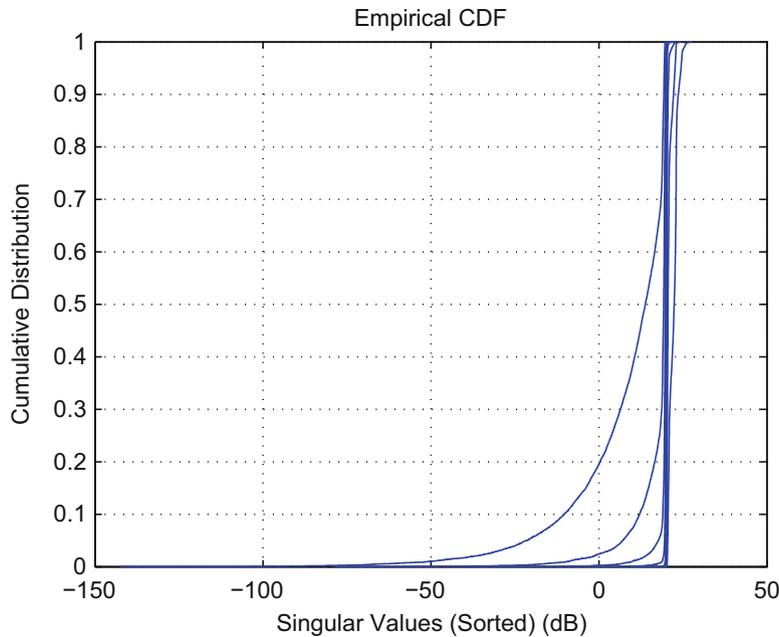


Fig. 6.32 Cdfs of singular sorted values of \mathbf{G} for uniform random line-of-sight conditions with $L = 100$, $K_{\text{MS}} = 10$. For $K_{\text{MS}} = 10$ MSs, there are 10 sorted singular values and 10 corresponding cdf curves

The mean number of MSs dropped is

$$\bar{N}_{\text{drop}} = \sum_{n=1}^{K_{\text{MS}}-1} P(N_{\text{drop}} = n). \quad (6.240)$$

6.14.5.3 Rician Fading Conditions

To model Rician fading, the channel matrix \mathbf{G} can be divided into a deterministic part and a random part, i.e.,

$$\mathbf{G} = \sqrt{\frac{K}{K+1}} \bar{\mathbf{G}} + \sqrt{\frac{1}{K+1}} \mathbf{G}_s \quad (6.241)$$

where $E[\mathbf{G}] = \sqrt{\frac{K}{K+1}} \bar{\mathbf{G}}$ is the LoS or specular component and $\sqrt{\frac{1}{K+1}} \mathbf{G}_s$ is the scatter component assumed to have zero mean. With Rician fading conditions, the inner product of the propagation vectors has variance

$$\text{Var}\left(\frac{\mathbf{g}_k^H \mathbf{g}_j}{L}\right) = \frac{1}{L} - \frac{1}{L^2} \sqrt{\frac{K}{K+1}}$$

The variance of the square of the inner product is

$$\text{Var}\left(\frac{\|\mathbf{g}_k^H \mathbf{g}_j\|^2}{L^2}\right) = \frac{1}{L} \sqrt{\frac{1}{K+1}} \left(\frac{1}{L} + \frac{2}{3} \sqrt{K}\right)$$

Figure 6.33 shows the cdf of the sorted singular values for a Rician channel with a Rice factor $K = 10$ dB. The results are obtained by combining the UR-LoS propagation vector in (6.237) with independent Rayleigh fading. As with UR-LoS conditions, a few singular values are small with high probability, while the remainder are concentrated around their median values.

6.14.5.4 Sufficiently Favorable Channels

Suppose that the propagation channel \mathbf{G} is considered to be sufficiently favorable when the condition number is below 8 dB 90% of the time. Figure 6.34 depicts the 90% maximum condition number as a function of the Rice factor K when

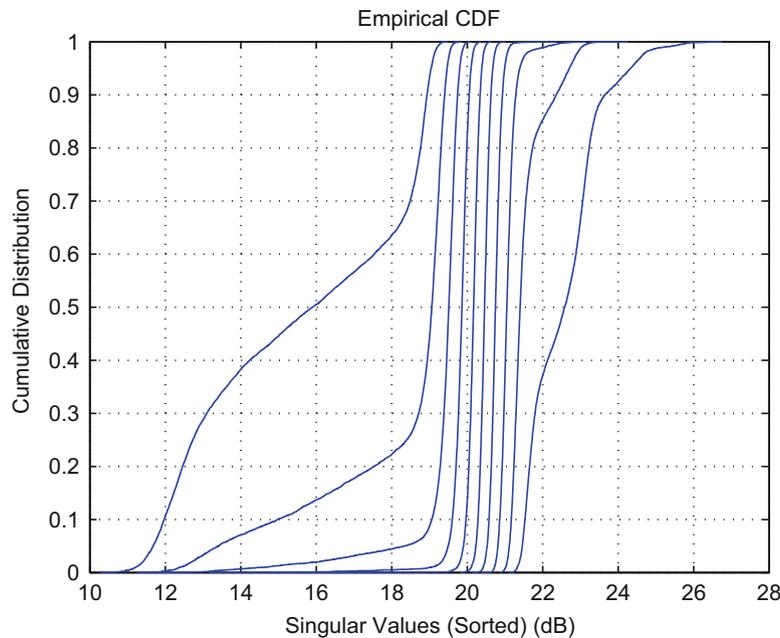


Fig. 6.33 Cdfs of singular sorted values of \mathbf{G} for Rician fading conditions with $L = 100$, $K = 10$. For $K_{\text{MS}} = 10$ MSs, there are 10 sorted singular values and 10 corresponding cdf curves

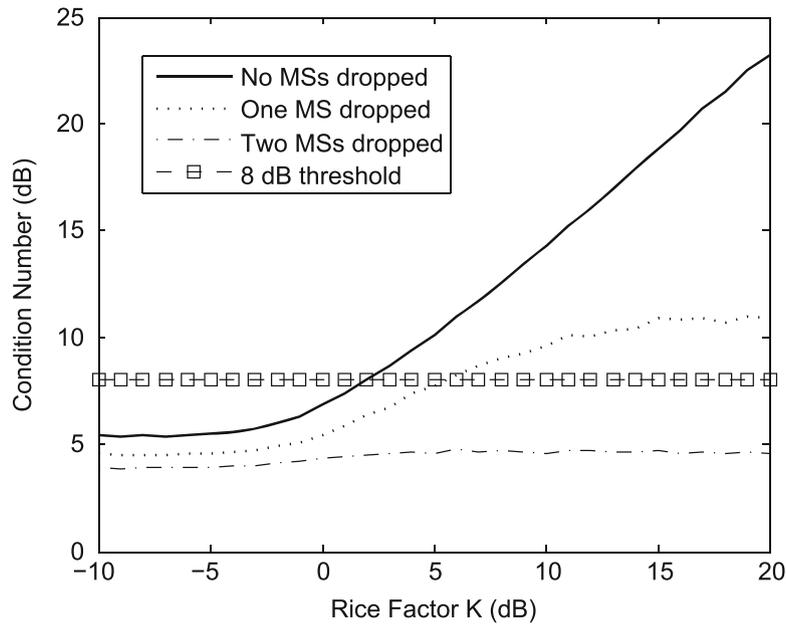


Fig. 6.34 Plot of condition number against Rice factor K when $N_{\text{drop}} = 0, 1,$ and 2 with $L = 100$ and $K_{\text{MS}} = 10$. With 10,000 realizations of the channel matrix \mathbf{G} , the condition number plotted is the 1000th highest, such that the value plotted is only exceeded 10% of the time

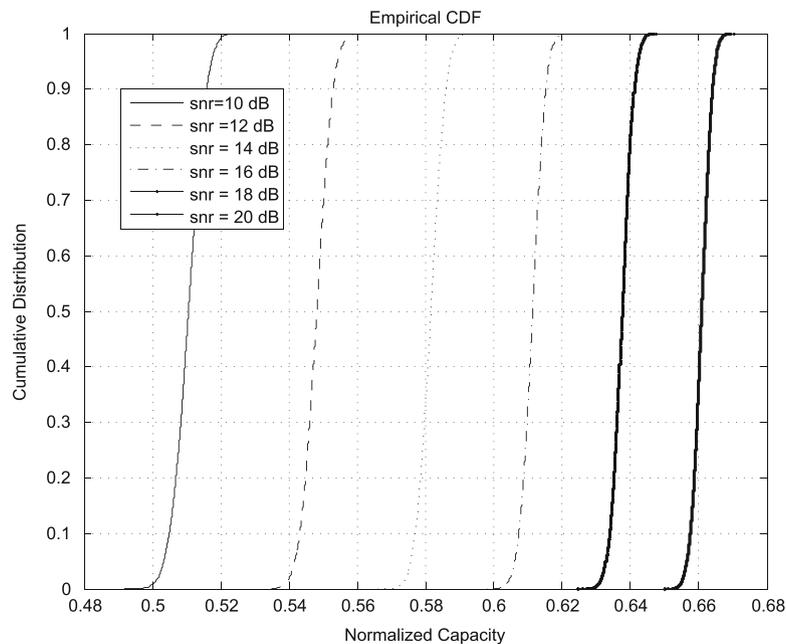


Fig. 6.35 Plot of factor of favorable propagation, defined in (6.236) as the capacity divided by the Hadamard upper bound on capacity for each random channel realization. The curves plot different symbol SNR, γ_s , progressing from 10 to 20 dB

$N_{\text{drop}} = 0, 1,$ and 2 . In Fig. 6.34, the condition number is below the various traces that are plotted 90% of the time. The outage is sufficiently small with no MSs dropped for $K < 2$ dB. For $2 \leq K \leq 5$, it is recommended that one MS be dropped. For $K \geq 5$, it is recommended that 2 MSs be dropped. There is no Rice factor K where more than 2 out of 10 MSs would need to be dropped to keep the condition number less than 8 dB for 90% of the time.

Another possibility is to consider the *factor of favorable propagation*, R_C , in (6.236). Figure 6.35 plots the cumulative distribution function of R_C against the symbol SNR, γ_s for $L = 100$ antennas. Observe that the median of R_C shifts to the right as γ_s increases, and the cdf of R_C is steep around a median value that ranges from 0.5 at $\gamma_s = 10$ dB to 0.65 at $\gamma_s = 20$ dB.

Problems

6.1. Let $\{X_1, X_2, \dots, X_N\}$ be a set of independent Rayleigh random variables each with an rms value of $1/\sqrt{N}$.

(a) Derive the pdf of

$$Y = \max(X_1^2, X_2^2, \dots, X_N^2).$$

This result is useful for the study of *selective combining* diversity systems.

(b) Again, using the set $\{X_1, X_2, \dots, X_N\}$, derive the pdf of

$$Z = X_1^2 + X_2^2 + \dots + X_N^2.$$

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 $10 \log_{10}\{\gamma_s^s/\bar{\gamma}_i\}$, where $\bar{\gamma}_i = (\bar{\gamma}_1 + \bar{\gamma}_2)/2$. Show several curves while varying the ratio $\xi = \bar{\gamma}_1/\bar{\gamma}_2$.

6.3. Consider using selective combining with coherent BPSK. For coherent 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

$$P_b = \int_0^\infty P_b(x) p_{\gamma_s^s}(x) dx.$$

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 W/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 $\bar{\gamma}_c = 20$ dB, where $\bar{\gamma}_c$ is the average received branch bit energy-to-noise ratio.

(b) Evaluate the probability of bit error for $\bar{\gamma}_c = 20$ 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 (6.28) for BPSK and maximal ratio combining.

6.6. Derive (6.66) for DPSK with differential detection followed by equal gain combining.

6.7. Consider an AWGN channel where the channel gain, α , has the following probability density function

$$p_\alpha(x) = 0.2\delta(x) + 0.5\delta(x-1) + 0.3\delta(x-2).$$

(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.

6.8. Consider a BPSK modulated system with simple repetition code and time interleaving, such that each data bit is transmitted L times and each transmission experiences independent identically distributed (i.i.d.) Rayleigh fading. If symbol \tilde{s} is transmitted, the corresponding L correlator or matched filter outputs at the receiver are

$$\tilde{r}_k = \alpha_k \tilde{s} + \tilde{n}_k, \quad k = 1, \dots, L$$

where \tilde{s} is the transmitted BPSK symbol chosen from the alphabet $\{\pm\sqrt{2E}\}$, the α_k are i.i.d. Rayleigh random variables, and the \tilde{n}_k are i.i.d. zero-mean complex Gaussian random variables with variance $\frac{1}{2}E[|\tilde{n}_k|^2] = N_o$.

- (a) One decoding strategy is to combine the \tilde{r}_k , $k = 1, \dots, L$ using maximal ratio combining and then make a bit decision. What is the probability of decision error in terms of the average received bit energy-to-noise ratio, $\bar{\gamma}_b$?
- (b) Another decoding strategy is to make a hard decision as to which symbol was transmitted for each of the \tilde{r}_k , $k = 1, \dots, L$, and then make a majority logic decision (assuming L is odd) as to which data bit was transmitted, i.e., if more of the L symbols comprising each bit are decided to be $+\sqrt{2E}$ than $-\sqrt{2E}$, then choose the data bit corresponding to symbol $+\sqrt{2E}$. What is the probability of decision error in terms of the average received bit energy-to-noise ratio, $\bar{\gamma}_b$?
- (c) Evaluate the probability of bit error in parts (a) and (b) when $L = 3$ and $\bar{\gamma}_b = 20$ dB.

6.9. Consider digital transmission using BPSK modulation and $L = 2$ receiver diversity. The channel gain for Antenna i , $i = 1, 2$, and symbol epoch n , $\alpha_{i,n}$, has the following probability density function

$$p_{\alpha_{i,n}}(x) = 0.9\delta(x - 1.0) + 0.1\delta(x - 0.05),$$

and the $\alpha_{i,n}$ are independent for $i = 1, 2$ and all n . Each receiver branch is affected by independent complex AWGN with noise power spectral density N_o W/Hz. Derive an expression for the bit error rate probability with maximal ratio combining.

6.10. The bit error probability of 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 bound (see Appendix A) on the probability of bit error 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 $\bar{\gamma}_1 = \bar{\gamma}_2 = \dots = \bar{\gamma}_L = \bar{\gamma}_c$.

6.11. The bit error probability of binary orthogonal FSK signaling with non-coherent square-law combining on a Rayleigh fading channel with additive white Gaussian noise is given by (6.28) where the parameter μ is defined in (6.83).

- (a) Derive a Chernoff bound (see Appendix A) on the probability of bit error and compare the Chernoff bound with the exact probability of error.
- (b) Derive a union-Chernoff bound on the probability of bit error with M -ary orthogonal FSK signaling and non-coherent square-law combining.
- (c) By using the union-Chernoff bound obtained in part (b) determine the diversity order L that will minimize the error probability.

6.12. 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 (6.50).

6.13. Consider a system that employs 2-branch selection diversity, where each diversity branch consists of L antennas with maximal ratio combining as shown in Fig. 6.36. Assume that all input MRC diversity branches are equal, i.e., $\bar{\gamma}_{ij} = \bar{\gamma}_c$, $i = 1, 2, j = 1, \dots, L$.

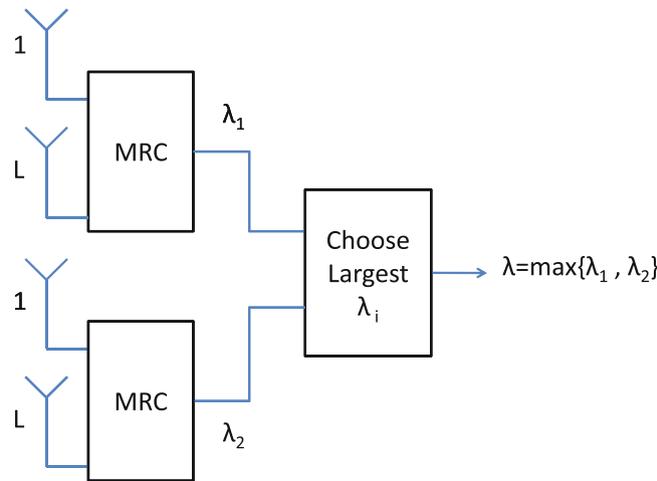


Fig. 6.36 Combiner for Problem 6.13

- (a) Derive an expression for the cumulative distribution function of the symbol energy-to-noise ratio, λ , at the output of the selective combiner.
- (b) Derive an expression for the probability density function of the symbol energy-to-noise ratio, λ , at the output of the selective combiner.
- (c) Write down an integral expression for the probability of bit error with BPSK modulation. You do not have to solve the integral!

6.14. Consider a system that uses L -branch selection diversity. The instantaneous received signal power on each diversity branch, $s_{0,i}$, $i = 1, \dots, L$, has the non-central chi-square (Ricean fading) distribution in (2.63). The instantaneous received signal power from each interferer on each diversity branch, $s_{k,i}$, $i = 1, \dots, L$ has the exponential (Rayleigh fading) distribution in (2.56). Note that all the $s_{0,i}$ and $s_{k,i}$ are all independent. Let $\lambda_i = s_{0,i} / \sum_{k=1}^{N_I} s_{k,i}$, $i = 1, \dots, L$ 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

$$O_I = P[\lambda_s < \lambda_{th}].$$

Plot O_I against λ_{th} for various L .

6.15. Consider the reception of a desired signal in the presence of a single co-channel interferer and neglect the effect of AWGN. The received signal power, s_0 , and interference power, s_1 , due to Rayleigh fading have the exponential distributions

$$p_{s_0}(x) = \frac{1}{\Omega_0} e^{-x/\Omega_0}$$

$$p_{s_1}(y) = \frac{1}{\Omega_1} e^{-y/\Omega_1},$$

where Ω_0 and Ω_1 are the average received signal power and interference power, respectively.

- (a) Assuming that s_0 and s_1 are independent random variables, find the probability density function for the carrier-to-interference ratio

$$\lambda = \frac{s_0}{s_1}.$$

Hint: If X and Y are independent random variables, then the probability density function of $U = X/Y$ is

$$p_U(u) = \int p_{XY}(v, v/u) |v/u^2| dv.$$

- (b) What is the mean value of λ ?
- (c) Now suppose that the system uses L -branch selection diversity. The branches are independent and balanced (i.e., the distribution of λ_i , $i = 1 \dots, L$, is the same for each branch. What is the probability density function of

$$\lambda_s = \max(\lambda_1, \lambda_2, \dots, \lambda_L),$$

the carrier-to-interference ratio at the output of the selective combiner?

6.16. 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 $\bar{\gamma}_c = E[\gamma_i]$. It is known that joint pdf of γ_1 and γ_2 is

$$p_{\gamma_1, \gamma_2}(x_1, x_2) = \frac{1}{\bar{\gamma}_c^2(1 - |\rho|^2)} I_0 \left(\frac{2|\rho|\sqrt{x_1 x_2}}{\bar{\gamma}_c(1 - |\rho|^2)} \right) \exp \left\{ -\frac{x_1 + x_2}{\bar{\gamma}_c(1 - |\rho|^2)} \right\},$$

where $|\rho|$ is the magnitude of the correlation coefficient of the two complex Gaussian random variables that are associated with the two diversity branches. Derive an expression for the cdf of the bit energy-to-noise ratio at the output of the selective combiner

$$\gamma_s = \max\{\gamma_1, \gamma_2\}.$$

Plot the cdf for various ρ . What conclusions can you make?

6.17. Show that with optimal combining $\nabla_{\mathbf{w}} J$ in (6.93) is given by

$$\nabla_{\mathbf{w}} J = \left(\frac{\partial J}{\partial w_1}, \dots, \frac{\partial J}{\partial w_L} \right) = 2\mathbf{w}^T \Phi_{\tilde{\mathbf{R}}, \tilde{\mathbf{R}}} - 2\Phi_{\tilde{\mathbf{s}}_0, \tilde{\mathbf{R}}} = 0.$$

6.18. Suppose that the received signal consists of two strong multipath components arriving from different directions. Explain how you would design a phased array to capture the energy in both rays.

6.19. Consider Alamouti's transmit diversity scheme. Section 6.11.1 shows how to construct the combiner for the case of 2×1 and 2×2 diversity. Construct the combiner for the case of $2 \times L$ diversity.

6.20. Describe how you would combine Alamouti's transmit diversity scheme with OFDM. Show a block diagram of the transmitter and receiver.

Chapter 7

Time Domain Equalization and Interference Cancellation

Chapters 5 and 6 have considered digital signaling on frequency non-selective or flat fading channels. Such channels are typical for low data rate systems that occupy a bandwidth that is smaller than the channel coherence bandwidth. However, as the data rate increases, the bandwidth of the transmitted waveform will typically be larger than the channel coherence bandwidth. Under this condition the channel is non-ideal and will exhibit frequency selectivity or time delay spread. Such time delay spread causes interference between modulated symbols, a phenomenon known as inter-symbol interference (ISI). This chapter concentrates on the modeling of ISI channels and the various signal processing methods for recovering digital information transmitted over such channels.

This chapter begins with a treatment of ISI channel modeling in Sect. 7.1 that includes a vector representation of digital signaling on ISI channels. Section 7.2 then develops the maximum likelihood receiver for ISI channels, leading to an equivalent model of the ISI channel known as the discrete-time white noise channel model. The effects of using fractional sampling or over-sampling at the receiver are also considered, where the sampling rate is an integer multiple of the modulated symbol rate. Section 7.3 provides a treatment of symbol-by-symbol equalizers, including the linear zero-forcing and minimum mean-square-error (MMSE) equalizers, and the nonlinear decision feedback equalizer (DFE). Section 7.4 provides a treatment of sequence estimators beginning with maximum likelihood sequence estimation (MLSE) and the Viterbi algorithm. Since the MLSE receiver can have high complexity for channels that have a long impulse response, some reduced complexity sequence estimation techniques are considered such as reduced state sequence estimation (RSSE) and delayed decision feedback sequence estimation (DDFSE). Section 7.5 provides an analysis of the bit error rate performance of MLSE on static ISI channels and multipath-fading ISI channels. Finally, Sect. 7.6 considers fractionally spaced MLSE receivers for ISI channels.

Finally, Sect. 7.7 concludes the chapter with a discussion of co-channel demodulation for digital signals on ISI channels. The basic idea is to simultaneously recover the data from multiple users that transmit in the same bandwidth. The problem is formulated as a multiple-input multiple-output (MIMO) channel, where the inputs are the waveforms transmitted by multiple users in the same bandwidth and the outputs are the signals received at multiple antenna elements. A vector representation of the received signals is first developed, along with the optimum receiver that uses joint maximum likelihood sequence estimation (J-MLSE). Similar to the single-user case, the effects of using fractional sampling will be considered as well. Finally, the chapter ends with a receiver structure that incorporates a combination of optimal combining as discussed in Chap. 6 and sequence estimation as implemented with the Viterbi algorithm.

7.1 Modeling of ISI Channels

Chapter 4 showed that the complex envelope of any modulated signal can be expressed in the general form

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n). \quad (7.1)$$

This chapter restricts attention to linear full-response modulation schemes where

$$b(t, \mathbf{x}_n) = x_n h_a(t), \quad (7.2)$$

$h_a(t)$ is the amplitude shaping pulse, and $\{x_n\}$ is the complex data symbol sequence.

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

$$\tilde{r}(t) = \sum_n x_n h(t - nT) + \tilde{n}(t), \quad (7.3)$$

where

$$h(t) = \int_{-\infty}^{\infty} h_a(\tau) g(t - \tau) d\tau \quad (7.4)$$

is the received pulse, given by the convolution of the transmitted pulse $h_a(t)$ and the channel impulse response $g(t)$, and $\tilde{n}(t)$ is complex-valued additive white Gaussian noise (AWGN) with a power spectral density of N_o watts/Hz. Since the transmitted pulse $h_a(t)$ is causal ($h_a(t) = 0, t < 0$) the lower limit of integration can be replaced by zero, and since the physical channel is causal ($g(t) = 0, t < 0$) the upper limit of integration in (7.4) can be replaced by t , so that

$$h(t) = \int_0^t h_a(\tau) g(t - \tau) d\tau, \quad t \geq 0. \quad (7.5)$$

Finally, the received pulse $h(t)$ is assumed to have a finite duration so that $h(t) = 0$ for $t < 0$ and $t \geq (L + 1)T$, where L is some positive integer.

7.1.1 Vector Representation of Received Signals

Using a Gram–Schmidt orthonormalization procedure, the received signal $\tilde{r}(t)$ in (7.3) can be expressed in the form

$$\tilde{r}(t) = \lim_{N \rightarrow \infty} \sum_{k=1}^N \tilde{r}_k \varphi_k(t), \quad (7.6)$$

where the $\{\varphi_n(t)\}$ are a complete set of complex orthonormal basis functions. Note that the basis functions span over the entire length of the waveform $\tilde{r}(t)$, and for the present purpose it is not necessary to actually generate the basis functions. Also, this set of basis functions should not be confused with the set of basis functions that is used to represent the signal set as in Sect. 5.1. It can be readily shown that

$$\tilde{r}_k = \sum_n x_n h_{nk} + \tilde{n}_k, \quad (7.7)$$

where

$$\begin{aligned} h_{nk} &= \int_{-\infty}^{\infty} h(t - nT) \varphi_k^*(t) dt \\ \tilde{n}_k &= \int_{-\infty}^{\infty} \tilde{n}(t) \varphi_k^*(t) dt. \end{aligned} \quad (7.8)$$

The noise samples \tilde{n}_k are zero-mean complex Gaussian random variables with covariance $\phi_{\tilde{n}_k \tilde{n}_m} = \frac{1}{2} E[\tilde{n}_k^* \tilde{n}_m] = N_o \delta_{km}$. Hence, it follows that the observation vector $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_N)$ has the conditional multivariate complex Gaussian distribution

$$p(\tilde{\mathbf{r}} | \mathbf{x}, \mathbf{h}) = \prod_{k=1}^N \frac{1}{2\pi N_o} \exp \left\{ -\frac{1}{2N_o} \left| \tilde{r}_k - \sum_n x_n h_{nk} \right|^2 \right\}, \quad (7.9)$$

where $\mathbf{h} = \{h_{nk}\}$.

7.2 Maximum Likelihood Receiver for ISI Channels with AWGN

The maximum likelihood receiver decides in favor of the symbol sequence \mathbf{x} that maximizes the likelihood function $p(\tilde{\mathbf{r}}|\mathbf{x}, \mathbf{h})$ or the log-likelihood function $\log\{p(\tilde{\mathbf{r}}|\mathbf{x}, \mathbf{h})\}$, i.e.,

$$\text{choose } \mathbf{x} \text{ if } \log\{p(\tilde{\mathbf{r}}|\mathbf{x}, \mathbf{h})\} > \log\{p(\tilde{\mathbf{r}}|\hat{\mathbf{x}}, \mathbf{h})\} \quad \forall \hat{\mathbf{x}} \neq \mathbf{x}. \quad (7.10)$$

For an AWGN channel, $p(\tilde{\mathbf{r}}|\mathbf{x}, \mathbf{h})$ has the form in (7.9) and the decision rule in (7.10) is equivalent choosing \mathbf{x} to maximize the quantity

$$\begin{aligned} \mu(\mathbf{x}) &= -\sum_{k=1}^N \left| \tilde{r}_k - \sum_n x_n h_{nk} \right|^2 \\ &= -\sum_{k=1}^N |\tilde{r}_k|^2 + \sum_{k=1}^N \left(\tilde{r}_k^* \sum_n x_n h_{nk} + \tilde{r}_k \sum_n x_n^* h_{nk}^* \right) \\ &\quad - \sum_{k=1}^N \left(\sum_n x_n h_{nk} \right) \left(\sum_m x_m^* h_{mk}^* \right). \end{aligned} \quad (7.11)$$

Since the term $\sum_{k=1}^N |\tilde{r}_k|^2$ is independent of \mathbf{x} , it may be omitted so that the maximum likelihood receiver chooses \mathbf{x} to maximize

$$\mu(\mathbf{x}) = 2\text{Re} \left\{ \sum_n x_n^* \sum_{k=1}^N \tilde{r}_k h_{nk}^* \right\} - \sum_n \sum_m x_n x_m^* \sum_{k=1}^N h_{nk} h_{mk}^*, \quad (7.12)$$

where $\text{Re}\{z\}$ denotes the real part of z . In the limit as the number of observable random variables N approaches infinity, define the following

$$y_n \triangleq \lim_{N \rightarrow \infty} \sum_{k=1}^N \tilde{r}_k h_{nk}^* = \int_{-\infty}^{\infty} \tilde{r}(t) h^*(t - nT) dt \quad (7.13)$$

$$f_{m-n} \triangleq \lim_{N \rightarrow \infty} \sum_{k=1}^N h_{nk} h_{mk}^* = \int_{-\infty}^{\infty} h(t - nT) h^*(t - mT) dt. \quad (7.14)$$

Using (7.13) and (7.14) in (7.12) gives the final form

$$\mu(\mathbf{x}) = 2\text{Re} \left\{ \sum_n x_n^* y_n \right\} - \sum_n \sum_m x_n x_m^* f_{m-n}. \quad (7.15)$$

The variables $\{y_n\}$ are obtained by passing the received complex low-pass waveform $\tilde{r}(t)$ 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. Hence, the optimum front-end processing is as shown in Fig. 7.1. Finally by changing the variable of integration, the $\{f_{m-n}\}$ in (7.14) can be rewritten in the form

$$f_\ell = \int_{-\infty}^{\infty} h(t + \ell T) h^*(t) dt, \quad (7.16)$$

where $\ell = m - n$. From (7.16), it is seen that the $\{f_\ell\}$ represent the sampled autocorrelation function of the received pulse $h(t)$ with sample spacing T , and have the property that $f_n^* = f_{-n}$. Sometimes the $\{f_\ell\}$ are called the ISI coefficients.

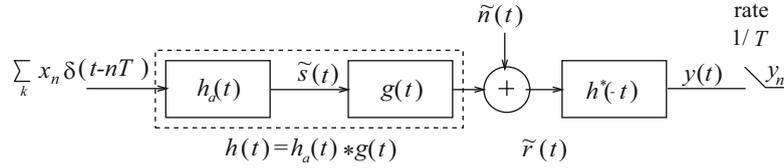


Fig. 7.1 Digital signaling on an ISI channel. The optimum front-end processor implements a filter that is matched to the received pulse $h(t)$ followed by a symbol rate sampler

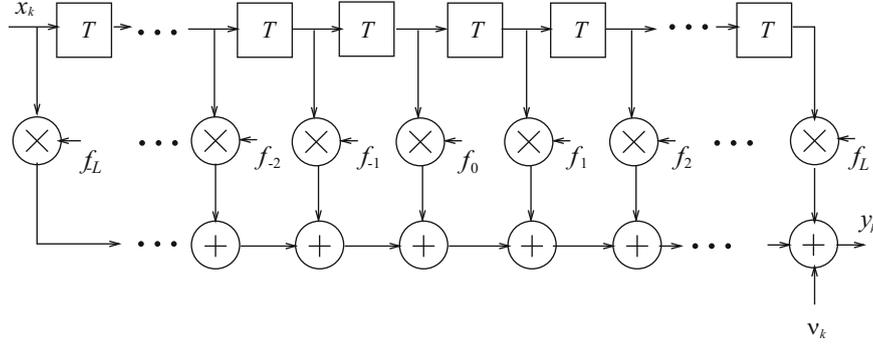


Fig. 7.2 Discrete-time model for digital signaling on an ISI channel

7.2.1 Discrete-Time White Noise Channel Model

Sampling the output of the matched filter $h^*(-t)$ in Fig. 7.1 every T seconds yields the sample sequence $\{y_k\}$, where

$$\begin{aligned} y_k &= \sum_n x_n f_{k-n} + v_k \\ &= \sum_{\ell=-L}^L x_{k-\ell} f_{\ell} + v_k \end{aligned} \quad (7.17)$$

and

$$v_k = \int_{-\infty}^{\infty} \tilde{n}(\tau) h^*(\tau - kT) d\tau \quad (7.18)$$

is the noise sample at the output of 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

$$\mathbf{f} = (f_{-L}, f_{-L+1}, \dots, f_{-1}, f_0, f_1, \dots, f_{L-1}, f_L). \quad (7.19)$$

This representation is depicted in Fig. 7.2.

As shown in (7.15), the maximum likelihood receiver uses the observation vector $\mathbf{y} = \{y_k\}$ and knowledge of the ISI coefficients $\{f_{\ell}\}$ to determine the most likely transmitted sequence \mathbf{x} . Equivalently, the maximum likelihood receiver decides in favor of the symbol sequence \mathbf{x} that maximizes the likelihood function $p(\mathbf{y}|\mathbf{x}, \mathbf{f})$ or the log-likelihood function $\log\{p(\mathbf{y}|\mathbf{x}, \mathbf{f})\}$, i.e.,

$$\text{choose } \mathbf{x} \text{ if } \log\{p(\mathbf{y}|\mathbf{x}, \mathbf{f})\} > \log\{p(\mathbf{y}|\hat{\mathbf{x}}, \mathbf{f})\} \quad \forall \hat{\mathbf{x}} \neq \mathbf{x}. \quad (7.20)$$

The noise samples at the matched filter output in (7.18) have the discrete autocorrelation function

$$\phi_{vv}(n) = \frac{1}{2} E[v_{k+n} v_k^*] = N_a f_n. \quad (7.21)$$

Hence, the noise sequence $\{v_k\}$ will be correlated unless $f_n = 0, n \neq 0$, meaning that the overall pulse

$$f(t) = \int_{-\infty}^{\infty} h(t + \tau)h^*(\tau)d\tau \quad (7.22)$$

satisfies the first Nyquist criterion. Such a condition will not be true for ISI channels due to the non-ideal channel $g(t)$, and the resulting correlation between the noise samples $\{v_k\}$ results a log-likelihood function $\log\{p(\mathbf{y}|\mathbf{x}, \mathbf{f})\}$ that has a complicated form. This difficulty can be overcome by passing the sample sequence at the output of the matched filter, $\{y_k\}$, through a noise whitening filter as described below, to whiten the noise samples.

The z -transform of the vector \mathbf{f} is

$$F(z) = \sum_{n=-L}^L f_n z^{-n}. \quad (7.23)$$

Using the property $f_n^* = f_{-n}$ yields

$$F^*(1/z^*) = F(z). \quad (7.24)$$

This means that if z is a root of $F(z)$ then $1/z^*$ is a root of $F(z)$, i.e., the roots of $F(z)$ occur in conjugate reciprocal pairs. It follows that $F(z)$ has $2L$ roots with the factorization

$$F(z) = G(z)G^*(1/z^*), \quad (7.25)$$

where $G(z)$ and $G^*(1/z^*)$ are polynomials each of degree L . The roots of $G(z)$ are z_1, z_2, \dots, z_L , while the roots of $G^*(1/z^*)$ are $1/z_1^*, 1/z_2^*, \dots, 1/z_L^*$. Hence, there are $2L$ possible choices for the roots of $G^*(1/z^*)$ and any one will suffice for a noise whitening filter $1/G^*(1/z^*)$. However, some reduced state equalization techniques such as RSSE and DDFSE require that the polynomial of the overall system $G(z) = F(z) \cdot 1/G^*(1/z^*)$ be minimum-phase, meaning all the poles and zeroes of $G(z)$ lie inside the unit circle. For such cases, the noise whitening filter $1/G^*(1/z^*)$ will be a stable filter, but it is noncausal since all its poles are outside the unit circle. In practice, such a noncausal noise whitening filter can be approximated to sufficient accuracy by using a long enough filter delay. If the overall response $G(z)$ need not have minimum phase, then $G^*(1/z^*)$ can be chosen to have minimum phase, i.e., all the poles and zeros of the noise whitening filter $1/G^*(1/z^*)$ are inside the unit circle. This choice will ensure that the noise whitening filter is both causal and stable.

If the noise whitening filter is chosen such that $G(z)$ has minimum phase, then the resulting discrete-time white noise channel satisfies the *minimum energy-delay property*. To explain this further, let $G_{\min}(e^{j2\pi f})$ be the frequency response function corresponding to the $G(z)$ having minimum phase, and let $g_{k \min}$ be the corresponding time-domain impulse response. All 2^L choices for $G(z)$ will have the same magnitude response, i.e., $|G(e^{j2\pi f})| = |G_{\min}(e^{j2\pi f})|$. Consequently, all impulse responses g_k whose magnitude response $|G(e^{j2\pi f})|$ is equal to $|G_{\min}(e^{j2\pi f})|$ will have the same total energy by Parseval's theorem, i.e.,

$$\sum_{n=0}^{\infty} |g_k|^2 = \int_{-1/2}^{1/2} |G(e^{j2\pi f})|^2 df = \int_{-1/2}^{1/2} |G_{\min}(e^{j2\pi f})|^2 df = \sum_{n=0}^{\infty} |g_{k \min}|^2. \quad (7.26)$$

If the partial energy of the impulse response is defined as

$$E(k) \triangleq \sum_{n=0}^k |g_k|^2, \quad (7.27)$$

then it can be shown that [251]

$$E(k) = \sum_{n=0}^k |g_k|^2 \leq \sum_{n=0}^k |g_{k \min}|^2 = E_{\min}(k), \quad (7.28)$$

for all impulse responses g_k that have the same magnitude response. Accordingly, the energy of the system having minimum phase is most concentrated around $k = 0$. This means that the energy of the minimum phase system has the least delay among all systems that have the same magnitude response function. For this reason the minimum phase system is said to satisfy the minimum energy-delay property.

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, i.e., $\frac{1}{2} \int_{-\infty}^{\infty} h_a^2(t) dt = 1$. The corresponding ISI coefficients are

$$\begin{aligned} f_\ell &= \int_{-\infty}^{\infty} h^*(t)h(t + \ell T)dt \\ &= \begin{cases} 1 + |a|^2 & \ell = 0 \\ a & \ell = 1 \\ a^* & \ell = -1 \end{cases} \end{aligned}$$

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.

Case 1: Under the assumption that $|a| < 1$, suppose that the zero of $G^*(1/z^*)$ is chosen to be outside the unit circle. That is,

$$\begin{aligned} G(z) &= 1 + az^{-1} \\ G^*(1/z^*) &= 1 + a^*z. \end{aligned}$$

In this case, the noise whitening filter $1/G^*(1/z^*)$ is noncausal yet stable, and the overall system is characterized by the minimum phase polynomial

$$G(z) = 1 + az^{-1}.$$

Note that the zero of $G(z)$ is inside the unit circle, with a pole at the origin.

Case 2: Under the assumption that $|a| < 1$, suppose that the zero of $G^*(1/z^*)$ is instead chosen to be inside 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 $1/G^*(1/z^*)$ is a minimum phase filter that is both stable and causal. However, the overall system $G(z)$ is characterized by the non-minimum phase polynomial

$$G(z) = 1 + a^*z.$$

Note that the zero of $G(z)$ is outside the unit circle with a pole at infinity.

The sequence of noise samples at the output of the noise whitening filter is now shown to be uncorrelated. From (7.17), the z-transform of the sample sequence at the output of the noise whitening filter is

$$\begin{aligned} V(z) &= (X(z)F(z) + v(z)) \frac{1}{G^*(1/z^*)} \\ &= X(z)G(z) + v(z) \frac{1}{G^*(1/z^*)} \\ &= X(z)G(z) + W(z). \end{aligned} \tag{7.29}$$

From (7.21), the power spectral density of the noise samples $\{v_k\}$ at the input to the noise whitening filter is

$$S_{vv}(e^{j2\pi fT}) = N_o F(e^{j2\pi fT}), \quad |f| \leq \frac{1}{2T}. \tag{7.30}$$

Therefore, the power spectral density of the noise samples $\{w_k\}$ at the output of noise whitening filter $1/G^*(1/z^*)$ is

$$\begin{aligned} S_{ww}(e^{j2\pi fT}) &= N_o \frac{F(e^{j2\pi fT})}{|G^*(e^{j2\pi fT})|^2} \\ &= N_o \frac{G(e^{j2\pi fT})G^*(e^{j2\pi fT})}{G(e^{j2\pi fT})G^*(e^{j2\pi fT})} \\ &= N_o, \quad |f| \leq \frac{1}{2T} \end{aligned} \tag{7.31}$$

which is clearly white.

The above development leads to the system shown in Fig. 7.3, and the discrete-time white noise channel model shown in Fig. 7.4. Sometimes the concatenation of the matched filter and noise whitening filter in Fig. 7.3 is called a whitened matched filter. The overall system function $G(z)$ can be viewed as a finite impulse response (FIR) filter with tap coefficients $\{g_n\}$. The discrete-time samples at the output of the noise whitening filter are

$$v_k = \sum_{n=0}^L g_n x_{k-n} + w_k. \tag{7.32}$$

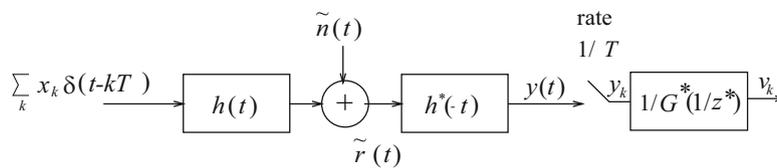


Fig. 7.3 Block diagram of system that implements a filter matched to $h(t)$ followed by a discrete-time noise whitening filter

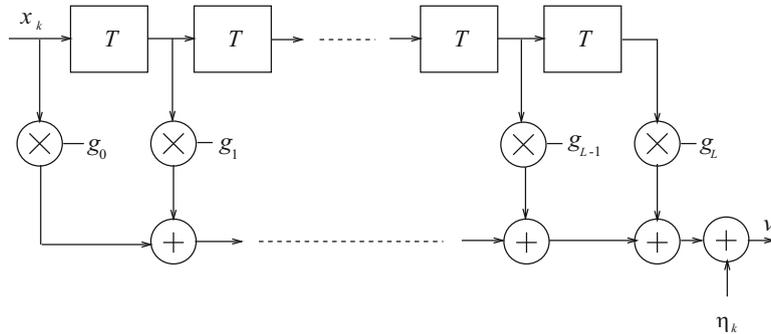


Fig. 7.4 Discrete-time white noise channel model

The maximum likelihood receiver uses the observation vector $\mathbf{v} = \{v_k\}_{k=0}^L$ to decide in favor of the symbol sequence \mathbf{x} that maximizes the likelihood function $p(\mathbf{v}|\mathbf{x}, \mathbf{g})$ or the log-likelihood function $\log\{p(\mathbf{v}|\mathbf{x}, \mathbf{g})\}$, i.e.,

$$\text{choose } \mathbf{x} \text{ if } \log\{p(\mathbf{v}|\mathbf{x}, \mathbf{g})\} > \log\{p(\mathbf{v}|\hat{\mathbf{x}}, \mathbf{g})\} \quad \forall \hat{\mathbf{x}} \neq \mathbf{x}, \quad (7.33)$$

where

$$\mathbf{g} = (g_0, g_1, \dots, g_L)^T \quad (7.34)$$

is the overall channel impulse response. Since the noise samples $\{w_k\}$ are white, the likelihood function has the simple product form

$$p(\mathbf{v}|\mathbf{x}, \mathbf{g}) = \prod_k \frac{1}{2\pi N_o} \exp \left\{ -\frac{1}{2N_o} \left| v_k - \sum_{n=0}^L g_n x_{k-n} \right|^2 \right\}. \quad (7.35)$$

The log-likelihood function $\log\{p(\mathbf{v}|\mathbf{x}, \mathbf{g})\}$ results in the decision rule

$$\text{choose } \mathbf{x} \text{ if } \mu(\mathbf{x}) > \mu(\hat{\mathbf{x}}) \quad \forall \hat{\mathbf{x}} \neq \mathbf{x}, \quad (7.36)$$

where

$$\mu(\mathbf{x}) = - \sum_k \left| v_k - \sum_{n=0}^L g_n x_{k-n} \right|^2. \quad (7.37)$$

An efficient method for finding the sequence \mathbf{x} is the Viterbi algorithm as discussed in Sect. 7.4.1.

Finally, for an ISI channel, the received symbol energy-to-noise ratio is defined as

$$\gamma_s = \frac{\mathbb{E}[|x_k|^2] \sum_{i=0}^L |g_i|^2}{2N_o} = \frac{\mathbb{E}[|x_k|^2] f_0}{2N_o} = \frac{2\sigma_x^2 E_{h_r}}{N_o} = \frac{E}{N_o} \quad (7.38)$$

where

$$E_{h_r} = \frac{1}{2} \int_{-\infty}^{\infty} |h(t)|^2 dt \quad (7.39)$$

is the energy in the received pulse $h(t)$. The bit energy-to-noise ratio is $\gamma_b = \gamma_s / \log_2 M$ where M is the modulation alphabet size.

7.2.1.1 Slowly Fading ISI Channels with Diversity

Consider a fading channel with D -branch receiver diversity. The received pulse on each diversity branch is equal to the convolution

$$h_d(t) = \int_{-\infty}^{\infty} g_d(t, \tau) h_a(t - \tau) d\tau, \quad d = 1, \dots, D, \quad (7.40)$$

where $g_d(t, \tau)$ is the time-variant channel impulse response for branch d . For slow fading, the channel impulse responses $g_d(t, \tau)$ can be assumed to change slowly with respect to the duration of the received pulses. When data is transmitted in short frames, e.g., 10–20 ms long, the channel may remain constant over the duration of the frame. This is sometimes called a block fading channel or quasi-static fading channel. In any case, at the k th epoch the received pulses can be accurately approximated as

$$h_{d,k}(t) = \int_{-\infty}^{\infty} g_d(kT, \tau) h_a(t - \tau) d\tau, \quad d = 1, \dots, D. \quad (7.41)$$

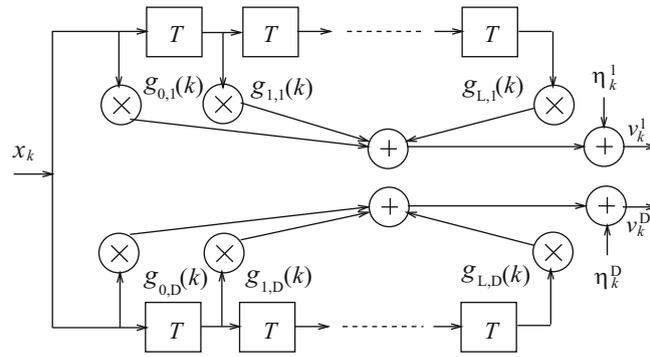


Fig. 7.5 Discrete-time white noise channel model with D -branch diversity

The receiver then implements a matched filter on each diversity branch having the impulse response $h_{d,k}^*(-t)$, and samples are taken at the output of the matched filter every T seconds. The samples at the output of each matched filter are passed through a corresponding noise whitening filter $1/G_{d,k}^*(1/z^*)$. This results in the discrete-time white noise channel model shown in Fig. 7.5. At epoch k , the tap gains associated with diversity branch d are described by the vector

$$\mathbf{g}_d(k) = (g_{0,d}(k), g_{1,d}(k), \dots, g_{L,d}(k))^T. \quad (7.42)$$

The $\{g_{i,d}(k)\}$ are discrete-time complex Gaussian random processes that are generally correlated with the correlation matrix

$$\Phi_{\mathbf{g}_d}(m) = \frac{1}{2} \mathbb{E}[\mathbf{g}_d(k) \mathbf{g}_d^H(k+m)] \quad (7.43)$$

where $(\mathbf{x})^H$ is the complex conjugate transpose of the vector \mathbf{x} . The received sample on branch d at epoch k is

$$v_{k,d} = \sum_{i=0}^{L-1} g_{i,d}(k) x_{k-i} + w_{k,d}, \quad (7.44)$$

where the $w_{k,d}$ are independent complex zero-mean white Gaussian noise samples with variance $\frac{1}{2} \mathbb{E}[|w_{k,d}|^2] = N_o$.

For a fading ISI channel, the average received symbol energy-to-noise ratio for branch d is

$$\bar{\gamma}_s^d = \frac{\mathbb{E}[|x_k|^2] \sum_{i=0}^{L-1} \mathbb{E}[|g_{i,d}|^2]}{2N_o} = \frac{\mathbb{E}[|x_k|^2] \mathbb{E}[f_{0,d}]}{N_o} = \frac{2\sigma_x^2 \mathbb{E}[E_{h,d}]}{N_o} = \frac{\bar{E}}{N_o}. \quad (7.45)$$

In many cases, the branches are balanced so that $\bar{\gamma}_s^d = \bar{\gamma}_s$, $d = 1, \dots, D$. The averaged received branch bit energy-to-noise ratio is $\bar{\gamma}_s / \log_2 M$.

Note that the matched filter and noise whitening filter impulse responses change slowly with time due to variations in the underlying channels. This presents a practical difficulty because implementation and adjustment of the matched filter and noise whitening filter requires knowledge of the underlying channel. It will be shown later that this difficulty can be overcome by implementing a filter that is matched to the transmitted pulse $h_a(t)$, over-sampling the output, and processing the output samples with a fractionally spaced noise whitening filter. First, however, the effect of over-sampling the matched filter output is considered.

7.2.1.2 $T/2$ -Spaced Receiver

In practice the matched filter outputs are often over-sampled for the purpose of symbol timing synchronization 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 matched filter is sampled with rate $2/T$. In this case, the overall channel impulse response and sampler can be represented by a discrete-time transversal filter with coefficients

$$\mathbf{f}^{(2)} = (f_{-2L}^{(2)}, f_{-2L+1}^{(2)}, \dots, f_{-1}^{(2)}, f_0^{(2)}, f_1^{(2)}, \dots, f_{2L-1}^{(2)}, f_{2L}^{(2)}). \quad (7.46)$$

where $(\cdot)^{(2)}$ indicates rate $2/T$ sampling. If the samples in (7.46) are obtained with the correct timing phase, i.e., $f_n^{(2)} = f(nT/2)$, then

$$\begin{aligned} \mathbf{f} &= (f_{-L}, f_{-L+1}, \dots, f_{-1}, f_0, f_1, \dots, f_{L-1}, f_L) \\ &= (f_{-2L}^{(2)}, f_{-2L+2}^{(2)}, \dots, f_{-2}^{(2)}, f_0^{(2)}, f_2^{(2)}, \dots, f_{2L-2}^{(2)}, f_{2L}^{(2)}), \end{aligned} \quad (7.47)$$

where $f_n^{(2)} = (f_{-n}^{(2)})^*$ and $f_n = f_{2n}^{(2)}$. More details on timing phase sensitivity will be provided in Sect. 7.6.3.

The $T/2$ -spaced noise samples at the matched filter output have the discrete-time autocorrelation function

$$\phi_{vv}(n) = N_a f_n^{(2)}. \quad (7.48)$$

The z -transform of $\mathbf{f}^{(2)}$, denoted as $F^{(2)}(z)$, has $4L$ roots with the factorization

$$F^{(2)}(z) = G^{(2)}(z)G^{(2)*}(1/z^*) \quad (7.49)$$

where $G^{(2)}(z)$ and $G^{(2)*}(1/z^*)$ are polynomials of degree $2L$ having conjugate reciprocal roots. The 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 as a noncausal stable filter such that the overall system function $G^{(2)}(z)$ has minimum phase with all its roots inside the unit circle. The output of the noise whitening filter is

$$v_n^{(2)} = \sum_{k=0}^{2L} g_k^{(2)} x_{n-k}^{(2)} + w_n^{(2)}, \quad (7.50)$$

where $\{w_n^{(2)}\}$ is a white Gaussian noise sequence with variance $\frac{1}{2}E[|w_n^{(2)}|^2] = N_o$ and the $\{g_n^{(2)}\}$ are the coefficients of a 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

$$x_n^{(2)} = \begin{cases} x_{n/2}, & n = 0, 2, 4, \dots \\ 0, & n = 1, 3, 5, \dots \end{cases} \quad (7.51)$$

Notice that each transmitted symbol is padded with a zero. In general, if rate K/T sampling is used, then each input symbol is padded with $K - 1$ zeros. The overall system and equivalent discrete-time white noise channel models are shown in Figs. 7.6 and 7.7, respectively.

Comparing (7.25) and (7.49) gives

$$\sum_{k=0}^{2L} |g_k^{(2)}|^2 = \sum_{k=0}^L |g_k|^2 = f_0^{(2)} = f_0. \quad (7.52)$$

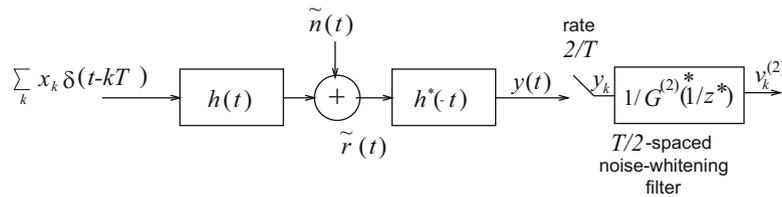


Fig. 7.6 Block diagram of system that implements a filter matched to $h(t)$ followed by a $T/2$ -spaced sampler and a discrete-time noise whitening filter

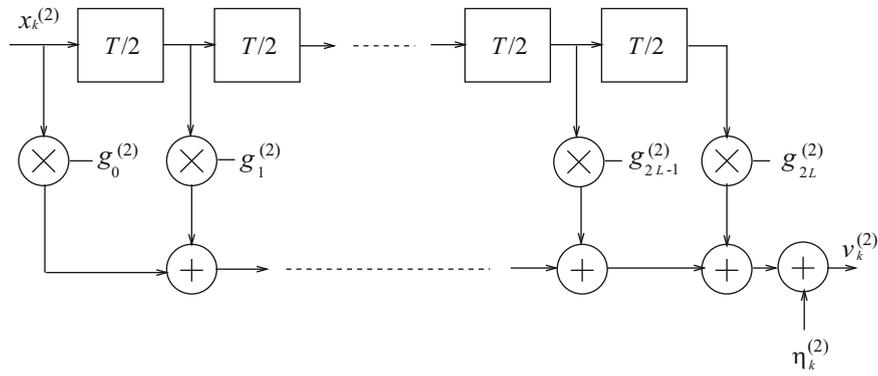


Fig. 7.7 Discrete-time white noise channel model with rate- $2/T$ sampling

Note that the samples $v_{2n}^{(2)}$ and $v_{2n+1}^{(2)}$ correspond to the n th received baud, where

$$v_{2n}^{(2)} = \sum_{k=0}^L g_{2k}^{(2)} x_{n-k} + w_{2n}^{(2)} \quad (7.53)$$

$$v_{2n+1}^{(2)} = \sum_{i=0}^{L-1} g_{2i+1}^{(2)} x_{n-k} + w_{2n+1}^{(2)}. \quad (7.54)$$

Finally, by comparing (7.32) and (7.54), $v_{2n}^{(2)}$ is not necessarily equal to v_n because a different noise whitening filter is used to whiten the $T/2$ -spaced noise samples than is used to whiten T -spaced noise samples.

7.3 Symbol-by-Symbol Equalizers

As shown in Fig. 7.8, a linear forward equalizer consists of a transversal filter with adjustable tap coefficients. The tap coefficients of the equalizer are denoted by the column vector

$$\mathbf{c} = (c_0, c_1, \dots, c_{N-1})^T, \quad (7.55)$$

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

$$\tilde{x}_n = \sum_{j=0}^{N-1} c_j v_{n-j}, \quad (7.56)$$

where the v_n are given by (7.32). The equalizer output \tilde{x}_k is quantized to the nearest (in Euclidean distance) information symbol to form the decision \hat{x}_k .

Observe that the overall discrete-time white noise channel and equalizer can be represented by a single filter having the sampled impulse response

$$\mathbf{q} = (q_0, q_1, \dots, q_{N+L-1})^T, \quad (7.57)$$

where

$$\begin{aligned} q_n &= \sum_{j=0}^{N-1} c_j g_{n-j} \\ &= \mathbf{c}^T \mathbf{g}(n) \end{aligned} \quad (7.58)$$

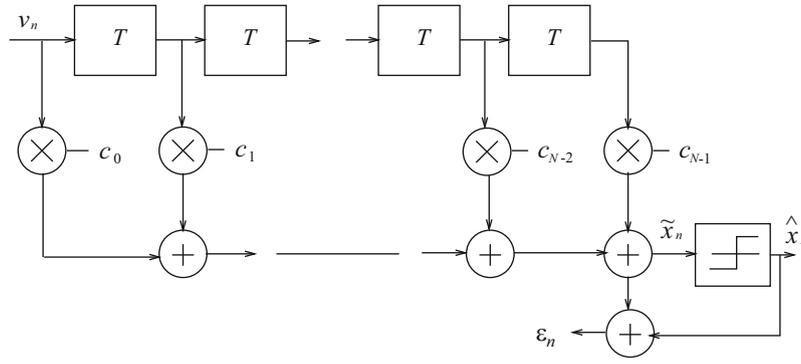


Fig. 7.8 Linear transversal equalizer with adjustable T -spaced taps

with

$$\mathbf{g}(n) = (g_n, g_{n-1}, g_{n-2}, \dots, g_{n-N+1})^T \quad (7.59)$$

and $g_i = 0, i < 0, i > L$. That is, \mathbf{q} is the discrete convolution of \mathbf{g} and \mathbf{c} .

If the equalizer is preceded by a noise whitening filter, then the discrete-time white noise channel has a system function $G(z)$ with minimum phase. Although the minimum phase system function $G(z)$ satisfies the minimum energy-delay property as discussed previously, it is not necessarily true that $|g_0|^2 \geq |g_k|^2, \forall k \geq 1$, i.e., the first tap g_0 does not necessarily have the largest magnitude. Let the component of \mathbf{g} of greatest magnitude be denoted by g_{d_1} . Also, let the number of equalizer taps be equal to $N = 2d_2 + 1$ where d_2 is an integer. Perfect equalization means that

$$\mathbf{q} = \mathbf{e}_d = (\underbrace{0, 0, \dots, 0}_{d-1 \text{ zeroes}}, 1, 0, \dots, 0)^T \quad (7.60)$$

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.

7.3.1 Linear Equalizer

7.3.1.1 Zero-Forcing (ZF) Solution

Lucky [211, 212] 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. With a ZF equalizer, the tap coefficients \mathbf{c} are chosen to minimize the peak distortion of the equalized channel, defined as

$$D_p = \frac{1}{|q_d|} \sum_{\substack{n=0 \\ n \neq d}}^{N+L-1} |q_n - \hat{q}_n|, \quad (7.61)$$

where $\hat{\mathbf{q}} = (\hat{q}_0, \dots, \hat{q}_{N+L-1})^T$ is the desired equalized channel and the delay d is a positive integer optimized to have the value $d = d_1 + d_2$ [71]. Lucky showed that if the initial distortion before equalization is less than unity, i.e.,

$$D = \frac{1}{|g_{d_1}|} \sum_{\substack{n=0 \\ n \neq d_1}}^L |g_n| < 1, \quad (7.62)$$

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 $\hat{\mathbf{q}} = \mathbf{e}_d$ the equalized channel is given by

$$\mathbf{q} = (q_0, \dots, q_{d_1-1}, 0, \dots, 0, 1, 0, \dots, 0, q_{d_1+N}, \dots, q_{N+L-1})^T. \quad (7.63)$$

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 [71]. To do so, the matrix

$$\mathbf{G} = [\mathbf{g}(d_1), \dots, \mathbf{g}(d), \dots, \mathbf{g}(N + d_1 - 1)] \quad (7.64)$$

is formed along with the vector

$$\tilde{\mathbf{q}} = (\hat{q}_{d_1}, \dots, \hat{q}_d, \dots, \hat{q}_{N+d_1-1})^T. \quad (7.65)$$

Then the vector of optimal tap gains, \mathbf{c}_{op} , is

$$\mathbf{c}_{\text{op}}^T \mathbf{G} = \tilde{\mathbf{q}}^T \longrightarrow \mathbf{c}_{\text{op}} = (\mathbf{G}^{-1})^T \tilde{\mathbf{q}}. \quad (7.66)$$

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 a 3-tap ZF equalizer is to be designed. Since g_0 is the component of \mathbf{g} having the largest magnitude, $d_1 = 0$ and the equalizer delay is chosen as $d = d_1 + d_2 = 1$. Suppose that the desired response is $\hat{\mathbf{q}}_0 = \mathbf{e}_1^T$ so that $\tilde{\mathbf{q}} = (0, 1, 0)$. The matrix

$$\begin{aligned} \mathbf{G} &= [\mathbf{g}(0), \mathbf{g}(1), \mathbf{g}(2)] \\ &= \begin{bmatrix} 0.90 & -0.15 & 0.20 \\ 0.00 & 0.90 & -0.15 \\ 0.00 & 0.00 & 0.90 \end{bmatrix} \end{aligned}$$

is then constructed to obtain the optimal tap solution

$$\mathbf{c}_{\text{op}} = (\mathbf{G}^{-1})^T \tilde{\mathbf{q}} = (0, 1.1111, -0.185185)^T.$$

The overall response of the channel and equalizer is

$$\mathbf{q} = (0.0, 1.0, 0, 0.194, 0.148, -0.037, -0.009, 0, \dots)^T.$$

Finally, the distortion after equalization is

$$D_p = \frac{1}{|q_1|} \sum_{\substack{n=0 \\ n \neq 1}}^6 |q_n - \hat{q}_n| = 0.388.$$

Adaptive Solution

In practice, the channel impulse response is unknown to the receiver and a known finite length sequence \mathbf{x} is used to train the equalizer. During this training mode, the equalizer taps can be obtained by using the following steepest-descent recursive algorithm:

$$c_j^{n+1} = c_j^n + \alpha \epsilon_n x_{n-j-d_1}^*, \quad j = 0, \dots, N-1, \quad (7.67)$$

where

$$\begin{aligned} \epsilon_n &= x_{n-d} - \tilde{x}_n \\ &= x_{n-d} - \sum_{i=0}^{N-1} c_i v_{n-i} \end{aligned} \quad (7.68)$$

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 trade off convergence rate and steady state bit error rate performance. Notice that adaptation rule in (7.67) attempts to force the cross-correlations $\epsilon_n x_{n-j-d_1}^*, j = 0, \dots, N-1$, to zero. To see that (7.67) leads to the desired solution, (7.32) and (7.68) are used along with the fact that the symbol sequence $\{x_n\}$ is uncorrelated with the noise sequence $\{w_n\}$ to obtain

$$\begin{aligned} \frac{1}{2} \mathbb{E}[\epsilon_n x_{n-j-d_1}^*] &= \frac{1}{2} \mathbb{E}[x_{n-d} x_{n-j-d_1}^*] - \frac{1}{2} \sum_{i=0}^{N-1} \sum_{\ell=0}^L c_i g_\ell \mathbb{E}[x_{n-i-\ell} x_{n-j-d_1}^*] \\ &= \sigma_x^2 \left(\delta_{d_2-j} - \sum_{i=0}^{N-1} c_i g_{j+d_1-i} \right) \\ &= \sigma_x^2 (\delta_{d_2-j} - q_{j+d_1}), \quad j = 0, 1, \dots, N-1, \end{aligned} \quad (7.69)$$

where $\sigma_x^2 = \frac{1}{2} \mathbb{E}[|x_k|^2]$. Note that the conditions $\frac{1}{2} \mathbb{E}[\epsilon_n x_{n-j-d_1}^*] = 0$ are satisfied when $q_d = 1$ and $q_i = 0$ for $d - d_2 \leq i < d$ and $d < i \leq d + d_2$, which is the zero-forcing solution.

After training the equalizer, a decision-feedback mechanism is typically employed where the sequence of symbol decisions $\hat{\mathbf{x}}$ is used to update the tap coefficients. This mode is called the data mode and allows the equalizer to track variations in the channel vector \mathbf{g} . In the data mode,

$$c_j^{n+1} = c_j^n + \alpha \epsilon_n \hat{x}_{n-j-d_1}^*, \quad j = 0, \dots, N-1, \quad (7.70)$$

where the error term ϵ_n in (7.68) becomes

$$\epsilon_n = \hat{x}_{n-d} - \sum_{i=0}^{N-1} c_i v_{n-i} \quad (7.71)$$

and, again, \hat{x}_{n-d} is the decision on the equalizer output \tilde{x}_n delayed by d samples.

Performance of the ZF Equalizer

If the ZF equalizer has an infinite number of taps, it is possible to select the tap weights so that $D_p = 0$, i.e., $\mathbf{q} = \hat{\mathbf{q}}$. Assuming that $\hat{q}_n = \delta_{n0}$ this condition means that

$$Q(z) = 1 = C(z)G(z). \quad (7.72)$$

Therefore,

$$C(z) = \frac{1}{G(z)} \quad (7.73)$$

and the ideal ZF equalizer has a discrete transfer function that is simply the inverse of overall channel $G(z)$. The cascade of the noise whitening filter with transfer function

$$W(z) = \sqrt{2E_{h_r}}/G^*(1/z^*) \quad (7.74)$$

and the ZF equalizer with transfer function $1/G(z)$ results in an equivalent equalizer with transfer function¹

$$C'(z) = \frac{\sqrt{2E_{h_r}}}{G^*(1/z^*)G(z)} = \frac{\sqrt{2E_{h_r}}}{F(z)}. \quad (7.75)$$

Recall from (7.30) that the noise sequence at the input to the equivalent equalizer $C'(z)$ has the discrete autocorrelation function $\phi_{vv}(n) = \frac{N_o}{2E_{h_r}}f_n$ and power spectral density (psd)

$$S_{\bar{v}\bar{v}}(f) = \frac{N_o}{2E_{h_r}}F(e^{j2\pi fT}), \quad |f| \leq \frac{1}{2T}. \quad (7.76)$$

Therefore, the psd of the noise sequence $\{\zeta_n\}$ at the output of the equalizer is

$$S_{\zeta\zeta}(f) = \frac{N_o}{F(e^{j2\pi fT})}, \quad |f| \leq \frac{1}{2T}, \quad (7.77)$$

and the noise samples have variance

$$\begin{aligned} \sigma_\zeta^2 &= T \int_{-1/2T}^{1/2T} S_{\zeta\zeta}(f) df \\ &= T \int_{-1/2T}^{1/2T} \frac{N_o}{F(e^{j2\pi fT})} df. \end{aligned} \quad (7.78)$$

If $\sigma_x^2 = \frac{1}{2}E[|x_k|^2]$ and $\hat{q}_n = \delta_{n0}$, then the signal-to-noise ratio at the output of the infinite-tap equalizer is

$$\gamma_\infty = \frac{\sigma_x^2}{\sigma_\zeta^2}. \quad (7.79)$$

Finally, it can be shown that (Problem 7.1)

$$F(e^{j2\pi fT}) = F_\Sigma(f), \quad |f| \leq \frac{1}{2T}, \quad (7.80)$$

where $F_\Sigma(f)$ is the folded spectrum of $F(f)$ defined by

$$F_\Sigma(f) \triangleq \frac{1}{T} \sum_{n=-\infty}^{\infty} F\left(f + \frac{n}{T}\right). \quad (7.81)$$

¹The scaling of the noise whitening filter gain by $\sqrt{2E_{h_r}}$ is not necessary in practice and is done here for mathematical convenience.

and $F(f)$ is the Fourier transform of the pulse $f(t) = h(t) * h^*(-t)$. Hence, the signal-to-noise ratio at the equalizer output can be written in the final form

$$\gamma_\infty = \sigma_x^2 \left(TN_o \int_{-1/2T}^{1/2T} \left(\frac{1}{T} \sum_{n=-\infty}^{\infty} F\left(f + \frac{n}{T}\right) \right)^{-1} df \right)^{-1}. \quad (7.82)$$

It is clear from (7.82) 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 the nulls in the folded spectrum by introducing infinite gain at their frequencies. Unfortunately, this results in severe noise enhancement at the output of the equalizer at these same frequencies. Land mobile radio channels often exhibit spectral nulls and, therefore, linear ZF equalizers are not used for land mobile radio applications.

On the other hand, when the overall channel $f(t)$ satisfies the conditions for ISI-free transmissions as discussed in Sect. 4.2, then the ISI coefficients satisfy the property $f_n = f_0 \delta_{n0}$ and the matched filter output is

$$\tilde{r}_k = \sqrt{2E_{h_r}} x_k + \tilde{n}_k. \quad (7.83)$$

The noise samples $\{\tilde{n}_k\}$ in the case are white due to the fact that the overall pulse $f(t)$ satisfies the first Nyquist criterion. From Sect. 4.2, an equivalent condition in the frequency domain is that the folded spectrum $F_\Sigma(f)$ is flat, i.e.,

$$F_\Sigma(f) = \frac{1}{T} \sum_{n=-\infty}^{\infty} F\left(f + \frac{n}{T}\right) = f_0 = 2E_{h_r}. \quad (7.84)$$

Under this condition the signal-to-noise ratio in (7.82) reaches its maximum value $\gamma_\infty = 2\sigma_x^2 E_{h_r} / N_o$.

7.3.1.2 MMSE Solution

Soon after Lucky introduced the ZF equalizer, Proakis and Miller [271], Lucky et. al. [213], and Gersho [138] developed the linear MMSE equalizer based on the least mean square (LMS) criterion. The MMSE equalizer is more robust and superior to the ZF equalizer in its performance and convergence properties [271, 272, 275]. By defining the vector

$$\mathbf{v}_n = (v_n, v_{n-1}, \dots, v_{n-N+1}), \quad (7.85)$$

where v_k is the output of the whitened matched filter in (7.32), the output of the equalizer in (7.56) can be expressed in the form

$$\tilde{x}_n = \mathbf{c}^T \mathbf{v}_n = \mathbf{v}_n^T \mathbf{c}. \quad (7.86)$$

An MMSE equalizer adjusts the tap coefficients to minimize the mean square error (MSE)

$$\begin{aligned} J &\triangleq \frac{1}{2} \mathbb{E}[|x_{n-d} - \tilde{x}_n|^2] \\ &= \frac{1}{2} \mathbb{E}[\mathbf{c}^T \mathbf{v}_n \mathbf{v}_n^H \mathbf{c}^* - 2\text{Re}\{\mathbf{v}_n^H \mathbf{c}^* x_{n-d}\} + |x_{n-d}|^2]. \end{aligned} \quad (7.87)$$

where, again, d is the equalizer delay assumed here to be chosen as $d = d_1 + d_2$.

Equalizer Tap Solution

If the channel impulse response is known, the optimum equalizer taps can be obtained by direct solution. Define

$$\begin{aligned}\mathbf{M}_v &\triangleq \frac{1}{2} \mathbf{E}[\mathbf{v}_n \mathbf{v}_n^H] \\ \mathbf{v}_x^H &\triangleq \frac{1}{2} \mathbf{E}[\mathbf{v}_n^H x_{n-d}]\end{aligned}\quad (7.88)$$

where \mathbf{M}_v is an $N \times N$ Hermitian matrix (meaning that $\mathbf{M}_v = \mathbf{M}_v^H$) and \mathbf{v}_x is a length N column vector. Using these definitions and assuming that $\frac{1}{2} \mathbf{E}[|x_{n-d}|^2] = \sigma_x^2$, the mean-square-error is

$$J = \mathbf{c}^T \mathbf{M}_v \mathbf{c}^* - 2 \operatorname{Re}\{\mathbf{v}_x^H \mathbf{c}^*\} + \sigma_x^2. \quad (7.89)$$

The tap vector \mathbf{c} that minimizes the mean square error can be obtained by equating the gradient $\nabla_{\mathbf{c}} J$ to zero. It can be shown that (Problem 7.15)

$$\nabla_{\mathbf{c}} J = \left(\frac{\partial J}{\partial c_0}, \dots, \frac{\partial J}{\partial c_{N-1}} \right) = 2\mathbf{c}^T \mathbf{M}_v - 2\mathbf{v}_x^H. \quad (7.90)$$

Setting $\nabla_{\mathbf{c}} J = 0$ gives the MMSE tap solution

$$\mathbf{c}_{\text{op}} = (\mathbf{M}_v^T)^{-1} \mathbf{v}_x^*. \quad (7.91)$$

By substituting (7.91) into (7.89), using the identity $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$ and the fact that \mathbf{M}_v is Hermitian, the MMSE can be expressed as

$$\begin{aligned}J_{\min} &= \mathbf{c}_{\text{op}}^T \mathbf{M}_v \mathbf{c}_{\text{op}}^* - 2 \operatorname{Re}\{\mathbf{v}_x^H \mathbf{c}_{\text{op}}^*\} + \sigma_x^2 \\ &= \sigma_x^2 - \mathbf{v}_x^H \mathbf{M}_v^{-1} \mathbf{v}_x.\end{aligned}\quad (7.92)$$

To proceed further, the i th component of the vector \mathbf{v}_x^H is

$$\frac{1}{2} \mathbf{E}[x_{n-d} v_{n-i}^*] = \sigma_x^2 \sum_{\ell=0}^L g_{\ell}^* \delta_{d-i-\ell} = \sigma_x^2 g_{d-i}^*, \quad i = 0, \dots, N-1, \quad (7.93)$$

so that

$$\mathbf{v}_x^H = \sigma_x^2 (g_d^*, g_{d-1}^*, \dots, g_0^*, 0, \dots, 0)^T. \quad (7.94)$$

Also,

$$\frac{1}{2} \mathbf{E}[v_{k-i} v_{k-j}^*] = \begin{cases} \sigma_x^2 f_{j-i} + N_o \delta_{ij}, & |i-j| \leq L \\ 0, & \text{otherwise} \end{cases}, \quad (7.95)$$

where the property

$$f_n = \sum_{k=0}^{L-n} g_k^* g_{k+n}, \quad n = 0, \dots, L \quad (7.96)$$

have been used. Hence, the $N \times N$ matrix \mathbf{M}_v has the form

$$\mathbf{M}_v = \sigma_x^2 \begin{bmatrix} f_0 + N_o/\sigma_x^2 & f_1 & f_2 & \cdots & f_{N-1} \\ f_1^* & f_0 + N_o/\sigma_x^2 & f_1 & \cdots & f_{N-2} \\ f_2^* & f_1^* & f_0 + N_o/\sigma_x^2 & \cdots & f_{N-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ f_{N-1}^* & \cdots & f_2^* & f_1^* & f_0 + N_o/\sigma_x^2 \end{bmatrix}. \quad (7.97)$$

Example 7.3. Consider a system having the same channel vector \mathbf{g} as in Example 7.2. Suppose that a 3-tap MMSE equalizer is to be designed. In this case, $g_{d_1} = 0$ and $N = 2d_2 + 1 = 3$, so that $d = d_1 + d_2 = 1$. Hence,

$$\mathbf{v}_x^H = \sigma_x^2(g_1^*, g_0^*, 0) = \sigma_x^2(-0.15, 0.90, 0.00)$$

and

$$\mathbf{M}_v = \sigma_x^2 \begin{bmatrix} \beta & -0.1500 & 0.1550 \\ -0.1500 & \beta & -0.1500 \\ 0.1550 & -0.1500 & \beta \end{bmatrix}$$

where $\beta = 0.8850 + N_o/\sigma_x^2$. The inverse of \mathbf{M}_v is

$$\mathbf{M}_v^{-1} = \frac{\text{adj}(\mathbf{M}_v)}{\det(\mathbf{M}_v)}$$

where $\det(\mathbf{M}_v) = (\sigma_x^2)^3(\beta(\beta^2 - 0.069025) + 0.006975)$ and

$$\text{adj}(\mathbf{M}_v) = (\sigma_x^2)^2 \begin{bmatrix} \beta^2 - 0.0225 & 0.15\beta - 0.02325 & 0.0225 - 0.155\beta \\ 0.15\beta - 0.02325 & \beta^2 - 0.024025 & 0.15\beta - 0.02325 \\ 0.0225 - 0.155\beta & 0.15\beta - 0.02325 & \beta^2 - 0.0225 \end{bmatrix}$$

Hence,

$$\mathbf{c}_{\text{op}} = \frac{(\sigma_x^2)^3}{\det(\mathbf{M}_v)} \begin{pmatrix} -0.15\beta^2 + 0.135\beta - 0.1755 \\ 0.90\beta^2 - 0.0225\beta - 0.018135 \\ 0.15825\beta - 0.0243 \end{pmatrix}.$$

With this tap solution,

$$J_{\min} = \sigma_x^2 \left(1 - \frac{0.8325\beta^2 - 0.013689}{\beta(\beta^2 - 0.069025) + 0.006975} \right)$$

and as $N_o \rightarrow 0$, $J_{\min} = 0.001089424\sigma_x^2$.

Adaptive Solution

In practice, the channel impulse response is unknown beforehand so that the MMSE solution cannot be obtained by the matrix inversion in (7.91). However, the equalizer taps can be obtained by using the stochastic gradient algorithm

$$c_j^{n+1} = c_j^n + \alpha \epsilon_n v_{n-j}^*, \quad j = 0, \dots, N-1, \quad (7.98)$$

where ϵ_n is given by (7.68). To show that (7.98) leads to the desired solution, from (7.90)

$$\begin{aligned} \nabla_{\mathbf{c}} J &= \mathbf{E}[\mathbf{c}^T \mathbf{v}_n \mathbf{v}_n^H - x_{n-d} \mathbf{v}_n^H] \\ &= \mathbf{E}[(\mathbf{c}^T \mathbf{v}_n - x_{n-d}) \mathbf{v}_n^H] \\ &= \mathbf{E}[\epsilon_n \mathbf{v}_n^H] = \mathbf{0}. \end{aligned} \quad (7.99)$$

It follows that

$$\frac{1}{2}E[\epsilon_n v_{n-j}^*] = 0, \quad j = 0, \dots, N-1, \quad (7.100)$$

and, therefore, the adaptive solution tends to force the cross-correlations $\epsilon_n v_{n-j}^*$, $j = 0, \dots, N-1$ to zero.

Performance of the MMSE Equalizer

The performance of an MMSE equalizer having an infinite number of taps provides some useful insight. In this case

$$\begin{aligned} \mathbf{c} &= (c_{-\infty}, \dots, c_0, \dots, c_{\infty}) \\ \mathbf{v}_n &= (v_{n+\infty}, \dots, v_n, \dots, v_{n-\infty}). \end{aligned}$$

Since the decision delay d with an infinite-tap equalizer is irrelevant, $d = 0$ can be chosen so that

$$\frac{1}{2}E[x_n v_{n-j}^*] = \begin{cases} \sigma_x^2 g_{-j}^*, & -L \leq j \leq 0 \\ 0, & \text{otherwise} \end{cases}. \quad (7.101)$$

The equation for the optimal tap gain vector $\mathbf{c}^T \mathbf{M}_v = \mathbf{v}_x^H$ can be written in the form

$$\sum_{i=-\infty}^{\infty} c_i (f_{j-i} + N_o \delta_{ij}) = g_{-j}^*, \quad -\infty < j < \infty. \quad (7.102)$$

Taking the z -transform of both sides of (7.102) gives

$$C(z) \left(G(z)G^*(1/z^*) + N_o \right) = G^*(1/z^*) \quad (7.103)$$

and, therefore,

$$C(z) = \frac{G^*(1/z^*)}{G(z)G^*(1/z^*) + N_o}. \quad (7.104)$$

The equivalent MMSE equalizer $C'(z) = C(z)W(z)$ that includes the noise whitening filter in (7.74) is

$$C'(z) = \frac{\sqrt{2E_{h_r}}}{G(z)G^*(1/z^*) + N_o} = \frac{\sqrt{2E_{h_r}}}{F(z) + N_o}. \quad (7.105)$$

Notice that $C'(z)$ has the same form as the ZF equalizer in (7.75), except for the noise term N_o in the denominator. Clearly, the ZF and MMSE 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 coefficients. Another measure of performance is the mean square error. The MMSE of an infinite-length MMSE equalizer is given by [272]

$$J_{\min} = \sigma_x^2 T \int_{-1/2T}^{1/2T} \frac{N_o}{F_{\Sigma}(f) + N_o} df, \quad (7.106)$$

where $\sigma_x^2 = \frac{1}{2}E[|x_k|^2]$. Note that $0 \leq J_{\min} \leq \sigma_x^2$, and that $J_{\min} = 0$ when there is no noise and $J_{\min} = \sigma_x^2$ when the folded spectrum $F_{\Sigma}(f)$ exhibits a spectral null. Furthermore, the relationship between the signal-to-noise ratio at the equalizer output and J_{\min} is

$$\gamma_{\infty} = \sigma_x^2 \cdot \frac{\sigma_x^2 - J_{\min}}{J_{\min}}. \quad (7.107)$$

When there is no ISI, $F_{\Sigma}(f) = f_0 = 2E_{h_r}$, giving

$$J_{\min} = \frac{\sigma_x^2 N_o}{2E_{h_r} + N_o}, \quad (7.108)$$

and the equalizer reaches its maximum output signal-to-noise ratio $\gamma_{\infty} = 2\sigma_x^2 E_{h_r} / N_o$. Finally, another useful measure for the effectiveness of linear equalization techniques is the *signal-to-interference-plus-noise ratio* (SINR) defined as

$$\text{SINR} = \frac{2\sigma_x^2 |q_d|^2}{2\sigma_x^2 \sum_{\substack{j=0 \\ j \neq d}}^{N+L-1} |q_j|^2 + N_o \sum_{j=0}^{N-1} |c_j|^2}. \quad (7.109)$$

Although the MMSE 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 [118, 272]. Another problem with a linear equalizer is the adaptation of the equalizer during data mode. This problem is especially acute when bandwidth efficient trellis-coded modulation schemes are used with non-iterative receivers. In this case, equalizer-based decisions are unreliable and inferior to those in uncoded systems due to the reduced separation between the points in the enlarged signal constellation. This problem can be partially alleviated by using periodic training, where the equalizer taps are allowed to converge in periodic training modes [99].

7.3.2 Decision Feedback Equalizer

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 [23] proposed the nonlinear DFE to mitigate noise enhancement. The DFE consists of two sections; a feedforward section and a feedback section as 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. It has been shown that the optimum tap setting of a zero-forcing DFE having infinite length feedforward and feedback filters is such that the feedforward filter is identical to a noise whitening filter with system function $1/G^*(1/z^*)$, such that the system function $G(z)$ has minimum phase [275]. Such a filter suppresses the postcursor of the channel response and whitens the noise. The combination of the matched filter, sampler, and feedforward filter yields an equivalent discrete-time white noise channel having the system function $G(z)$.

To eliminate the postcursor ISI, decisions made on the equalizer outputs are propagated through the feedback filter. The optimal coefficients of the feedback filter are the sampled impulse response of the tail of the overall system impulse response that includes the forward part of the DFE. This feedback mechanism introduces error propagation which can degrade the performance of the DFE and complicate its performance analysis.

The output of the DFE shown in Fig. 7.9 is

$$\tilde{x}_n = \sum_{i=0}^{N-1} c_i \tilde{r}_{n-i} - \sum_{i=1}^M F_i \hat{x}_{n-i}, \quad (7.110)$$

where $\{c_n\}$ and $\{F_n\}$ are the tap coefficients of the feedforward and feedback filters, respectively, $\{y_n\}$ is the sample sequence at the output of the matched filter, and $\{\hat{x}_n\}$ is the sequence of previously detected data symbols. Recall that the overall channel and feedforward portion of the equalizer can be represented by the sampled impulse response \mathbf{q} in (7.57). Hence, the DFE output can be written as

$$\tilde{x}_n = \sum_{i=0}^{N+L-1} q_i x_{n-i} - \sum_{i=1}^M F_i \hat{x}_{n-i} + \tilde{\eta}_n, \quad (7.111)$$

where

$$\tilde{\eta}_n = \sum_{i=0}^{N-1} c_i \tilde{r}_{n-i} \quad (7.112)$$

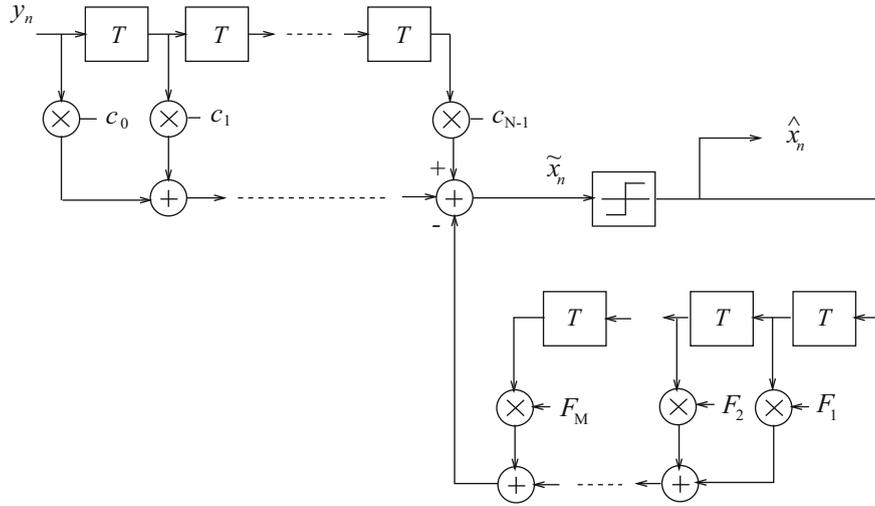


Fig. 7.9 Decision feedback equalizer

is the n th noise sample at the output of the feedforward filter. By adding and subtracting terms, the output of the DFE can be rewritten as

$$\begin{aligned} \tilde{x}_n &= x_n q_0 + \sum_{i=1}^M q_i (x_{n-i} - \tilde{x}_{n-i}) + \sum_{i=1}^M (q_i - F_i) \hat{x}_{n-i} \\ &+ \sum_{i=M+1}^{N+L-1} q_i x_{n-i} + \tilde{\eta}_n. \end{aligned} \quad (7.113)$$

Choosing

$$F_i = q_i = \mathbf{c}^T \mathbf{g}(i), \quad i = 1, 2, \dots, M \quad (7.114)$$

so that the second summation is zero, and if correct decisions are made so that the first summation is zero, then

$$\tilde{x}_n = x_n q_0 + \sum_{i=M+1}^{N+L-1} q_i x_{n-i} + \tilde{\eta}_n. \quad (7.115)$$

The summation in (7.115) represents the residual ISI that remains from the feedforward filter, which is zero if $M = N + L - 1$.

Equalizer Tap Solution

The coefficients $\{c_i\}$ and $\{F_i\}$ can be adjusted simultaneously to minimize the mean square error, resulting in an equalizer that is sometimes called an MMSE-DFE. Define

$$\mathbf{c} = (c_0, c_1, \dots, c_{N-1})^T \quad (7.116)$$

$$\tilde{\mathbf{r}}_n = (\tilde{r}_n, \tilde{r}_{n-1}, \dots, \tilde{r}_{n-N})^T \quad (7.117)$$

$$\hat{\mathbf{x}}_n = (\hat{x}_{n-1}, \hat{x}_{n-2}, \dots, \hat{x}_{n-M})^T \quad (7.118)$$

$$\mathbf{F} = (F_1, F_2, \dots, F_M)^T, \quad (7.119)$$

and define the error at the n th epoch, ϵ_n , as

$$\begin{aligned}\epsilon_n &= \tilde{x}_n - x_n \\ &= \mathbf{c}^T \tilde{\mathbf{r}}_n - \mathbf{F}^T \hat{\mathbf{x}}_n - x_n.\end{aligned}\quad (7.120)$$

Now define

$$\tilde{\mathbf{t}} \triangleq \text{vec}(\mathbf{c}^T, \mathbf{F}^T)^T \quad (7.121)$$

$$\tilde{\mathbf{y}}_n \triangleq \text{vec}(\tilde{\mathbf{r}}_n^T, -\hat{\mathbf{x}}_n^T)^T \quad (7.122)$$

so that $\epsilon_n = \tilde{\mathbf{t}}^T \tilde{\mathbf{y}}_n - x_n$. Then the MSE can be expressed as

$$\begin{aligned}J &= \frac{1}{2} \text{E}[|\epsilon_n|^2] \\ &= \frac{1}{2} \text{E}[\tilde{\mathbf{t}}^T \tilde{\mathbf{y}}_n \tilde{\mathbf{y}}_n^H \tilde{\mathbf{t}}^* - 2\text{Re}\{\tilde{\mathbf{y}}_n^H \tilde{\mathbf{t}}^* x_n\} + |x_n|^2].\end{aligned}\quad (7.123)$$

Notice that (7.123) and (7.87) have the exact same form. Therefore, the MMSE tap solution can be obtained by defining

$$\tilde{\mathbf{M}}_y \triangleq \frac{1}{2} \text{E}[\tilde{\mathbf{y}}_n \tilde{\mathbf{y}}_n^H] \quad (7.124)$$

$$\tilde{\mathbf{y}}_x^H \triangleq \frac{1}{2} \text{E}[\tilde{\mathbf{y}}_n^H x_n]. \quad (7.125)$$

Using the same argument that leads to (7.91) gives the MMSE-DFE tap solution

$$\tilde{\mathbf{t}}_{\text{op}} = (\tilde{\mathbf{M}}_y^T)^{-1} \tilde{\mathbf{y}}_x^*. \quad (7.126)$$

Adaptive Solution

The feedforward taps of the DFE can be adjusted by using

$$c_j^{n+1} = c_j^n + \alpha \epsilon_n v_{n+j}^*, \quad j = 0, \dots, N-1, \quad (7.127)$$

while the feedback coefficients can be adjusted according to

$$F_j^{n+1} = F_j^n + \alpha \epsilon_n \hat{x}_{n-j}^*, \quad j = 1, \dots, M, \quad (7.128)$$

where α is the step size. To see that this leads to the desired solution, the same argument that leads to (7.99) can be used. Then observe that $\nabla_{\mathbf{c}} J = \text{E}[\epsilon_n \tilde{\mathbf{y}}_n^H] = \mathbf{0}$ implies that

$$\frac{1}{2} \text{E}[\epsilon_n v_{n+j}^*] = 0, \quad j = 0, \dots, N-1 \quad (7.129)$$

$$\frac{1}{2} \text{E}[\epsilon_n \hat{x}_{n-j}^*] = 0, \quad j = 1, \dots, M, \quad (7.130)$$

where the second expectation is zero under the assumption that the DFE makes correct decision so that $\hat{x}_{n-j}^* = x_{n-j}^*$.

Performance of the DFE

The performance of a DFE is complicated by the fact that incorrect decisions in the feedback portion of the equalizer result in error propagation. 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 an infinite-length forward filter is identical to a stable, non-causal, noise whitening filter that results in an overall system response $G(z)$ having minimum phase [275]. The MMSE for the infinite length DFE is [287]

$$J_{\min} = \sigma_x^2 \exp \left\{ T \int_{-1/2T}^{1/2T} \ln \left(\frac{N_o}{F_{\Sigma}(f) + N_o} \right) df \right\}, \quad (7.131)$$

where $0 \leq J_{\min} \leq \sigma_x^2$. The corresponding signal-to-noise ratio at the output of the DFE is

$$\gamma_{\infty} = \sigma_x^2 \cdot \frac{\sigma_x^2 - J_{\min}}{J_{\min}}. \quad (7.132)$$

Once again, when there is no ISI $F_{\Sigma}(f) = f_0 = 2E_{h_r}$ and

$$J_{\min} = \frac{\sigma_x^2 N_o}{2E_{h_r} + N_o}, \quad (7.133)$$

and the equalizer reaches its maximum output signal-to-noise ratio $\gamma_{\infty} = 2\sigma_x^2 E_{h_r} / N_o$.

7.4 Sequence Estimation

7.4.1 Maximum Likelihood Sequence Estimation

The Viterbi algorithm was originally devised by Viterbi for maximum likelihood decoding of convolutional codes [334, 335]. 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 [131]. Because of the efficiency of the Viterbi algorithm, the implementation of optimum maximum likelihood sequence estimation (MLSE) for detecting ISI-corrupted signals is feasible.

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 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 defined by contents of the L memory elements in the tapped delay lines. If the size of the signal constellation is 2^n , there are total of $N_S = (2^n)^L$ states. The state at epoch k is

$$Q_k = (x_{k-1}, x_{k-2}, \dots, x_{k-L}). \quad (7.134)$$

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 ($N_S = 4$) and the system can be described the state diagram shown in Fig. 7.10. Note that there are two branches entering and leaving each state since binary modulation is used. In general there are $M = 2^n$ such branches for an M -ary modulation alphabet. The dashed lines correspond to an input symbol equal to “−1” while the solid lines correspond to an input symbol equal to “1.”

The system state diagram can be used to construct the trellis diagram shown in Fig. 7.11, where the initial zero state is assumed to be $Q_0^{(0)} = (-1, -1)$. Again, state transitions with a solid line correspond to an input symbol +1, while those with a dashed line correspond to an input symbol −1.

Suppose that the data sequence $\mathbf{x} = (-1, 1, 1, -1, 1, 1, -1, -1, \dots)$ is transmitted over the channel \mathbf{g} . Then the state sequence follows the shaded path in Fig. 7.11. The noiseless received sequence is $\mathbf{v} = (v_0, v_1, v_2, v_3, v_4, \dots)$, where

(continued)

Example 7.4 (continued)
Fig. 7.10 State diagram for binary signaling on a three-tap ISI channel

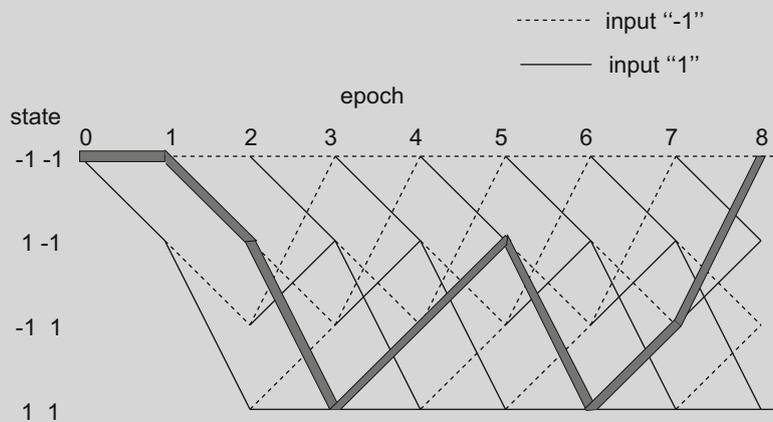
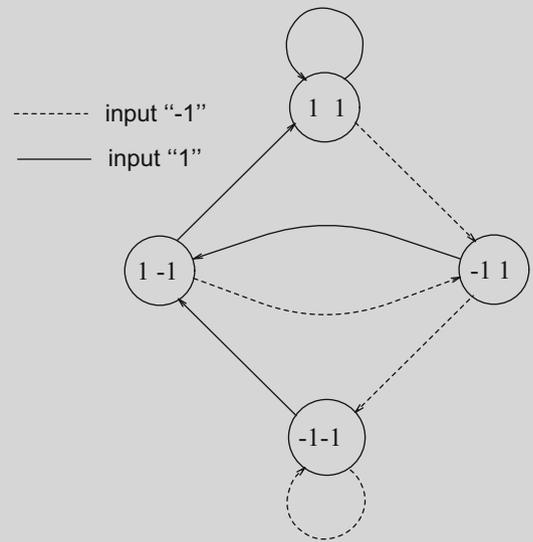


Fig. 7.11 Trellis diagram for binary signaling on a three-tap ISI channel

$$v_n = g_0x_n + g_1x_{n-1} + g_2x_{n-2}$$

$$= x_n + x_{n-1} + x_{n-2}.$$

Hence, for the data sequence $\mathbf{x} = (-1, 1, 1, -1, 1, 1, -1, -1, \dots)$ the noiseless received sequence is $\mathbf{v} = (-3, -1, 1, 1, 1, 1, 1, -1, \dots)$.

Assume that k symbols have been transmitted over the channel. Let $\mathbf{V}_n = (v_{n,1}, v_{n,2}, \dots, v_{n,D})$ denote the vector of signals received on all D 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

$$p(\mathbf{V}_k, \dots, \mathbf{V}_1 | x_k, \dots, x_1) \tag{7.135}$$

or, equivalently, the log-likelihood function

$$\log\{p(\mathbf{V}_k, \dots, \mathbf{V}_1 | x_k, \dots, x_1)\}. \tag{7.136}$$

Since the noise samples $\{w_{n,d}\}$ in (7.32) are mutually independent with respect to the indices n and d , and \mathbf{V}_n depends only on the L most recent transmitted symbols, the log-likelihood function (7.136) can be rewritten as

$$\begin{aligned} & \log\{p(\mathbf{V}_k, \dots, \mathbf{V}_1 | x_k, \dots, x_1)\} \\ &= \log\{p(\mathbf{V}_k | x_k, \dots, x_{k-L})\} + \log\{p(\mathbf{V}_{k-1}, \dots, \mathbf{V}_1 | x_{k-1}, \dots, x_1)\}, \end{aligned} \quad (7.137)$$

where $x_{k-L} = 0$ for $k - L \leq 0$. If the second term on the right side of (7.137) has been calculated previously at epoch $k - 1$ and stored in memory, 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

$$p(\mathbf{V}_k | x_k, \dots, x_{k-L}) = \frac{1}{(2\pi N_o)^D} \exp \left\{ -\frac{1}{2N_o} \sum_{d=1}^D \left| v_{k,d} - \sum_{i=0}^L g_{i,d} x_{k-i} \right|^2 \right\} \quad (7.138)$$

so that $\log p(\mathbf{V}_k | x_k, \dots, x_{k-L})$ yields the branch metric

$$\mu_k = -\sum_{d=1}^D \left| v_{k,d} - \sum_{i=0}^L g_{i,d} x_{k-i} \right|^2, \quad (7.139)$$

using the same argument that was used to arrive at (7.37). Note that the receiver requires knowledge of the set of channel vectors $\{\mathbf{g}_d, d = 1, \dots, D\}$ to compute the branch metric. As discussed later, this can be obtained by using a separate channel estimator.

Based on the recursion in (7.137) and the branch metric in (7.139), the well-known Viterbi algorithm [335] can be used to implement the ML receiver by searching through the N_S -state trellis for the most likely transmitted sequence $\mathbf{x} = \{x_k\}$ given the sequence of observation vectors $\mathbf{V} = \{\mathbf{V}_k\}$. This search process is called maximum likelihood sequence estimation (MLSE). At epoch k , the Viterbi algorithm stores N_S surviving sequences known as *survivors* $\check{\mathbf{x}}(\varrho_k^{(i)})$ (paths through the trellis) along with their associated *path metrics* $\Gamma(\varrho_k^{(i)})$ (distances from the received sequence) that terminate at state $\varrho_k^{(i)}$, $i = 0, \dots, N_S - 1$. The path metric is defined as

$$\Gamma(\varrho_k^{(i)}) = \sum_{n=1}^k \mu_n^{(i)}, \quad i = 0, \dots, N_S - 1, \quad (7.140)$$

where $\{\mu_n^{(i)}\}$ is the sequence of branch metrics along the surviving path $\check{\mathbf{x}}(\varrho_k^{(i)})$. MLSE as implemented by the Viterbi algorithm is now outlined followed by an example.

The Viterbi algorithm is initialized at time index $k = 0$, by setting all path metrics to zero, i.e., $\Gamma(\varrho_0^{(i)}), i = 1, \dots, N_S - 1$.

1. After the vector \mathbf{V}_{k+1} has been received, compute the set of path metrics $\Gamma(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)}) = \Gamma(\varrho_k^{(i)}) + \mu(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)})$ for all possible paths through the trellis that terminate in each state $\varrho_{k+1}^{(j)}, j = 0, \dots, N_S - 1$, where $\mu(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)})$ is the branch metric defined below. For a modulation alphabet of size M , there will be M such paths that terminate in each state $\varrho_{k+1}^{(j)}$.
2. Find $\Gamma(\varrho_{k+1}^{(j)}) = \max_i \Gamma(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)}), j = 0, \dots, N_S - 1$ where, again, the maximization is over all M 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 that terminate in state $\varrho_{k+1}^{(j)}$.
4. Increment the time index k , go to Step 1, and repeat the entire algorithm.

In Step 1 above, $\mu(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)})$ is the branch metric associated with the state transition $\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)}$ and is computed according to the following variation of (7.139)

$$\mu(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)}) = -\sum_{d=1}^D \left| v_{k,d} - g_{0,d} x_k(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)}) - \sum_{m=1}^L g_{m,d} x_{k-m}(\varrho_k^{(i)}) \right|^2, \quad (7.141)$$

where $x_k(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)})$ is a symbol that is uniquely determined by the state transition $\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)}$, and the L most recent symbols $\{x_{k-m}(\varrho_k^{(i)})\}_{m=1}^L$ are uniquely specified by the state $\varrho_k^{(i)}$.

Example 7.5. Consider again binary 4-state system in Example 7.4. In the presence of noise, the noise received sequence is

$$\begin{aligned} v_n &= g_0x_n + g_1x_{n-1} + g_2x_{n-2} + w_k \\ &= x_n + x_{n-1} + x_{n-2} + w_k, \end{aligned}$$

where the w_k are i.i.d. zero-mean Gaussian random variables with variance N_o . Suppose that due to additive white Gaussian noise, the noisy received sequence is

$$\begin{aligned} \mathbf{v} &= (v_0, v_1, v_2, v_3, v_4, \dots) \\ &= (-3.2, -1.1, 0.9, 0.1, 1.2, 1.5, 0.7, -1.3, \dots). \end{aligned}$$

The Viterbi algorithm is initialized with $\Gamma(\varrho_0^{(i)}) = 0$ for $i = 0, \dots, 3$. The initial state is assumed to be $\varrho_0^{(0)} = (-1, -1)$. Executing the Viterbi algorithm yields the result shown in Fig. 7.12, where the X's on the branches in the trellis denote dropped paths (the other path at each state is the survivor), and the numbers in the trellis are the path metrics corresponding to the surviving sequences at each state. The path metrics at each state $\varrho_k^{(j)}, j = 0, \dots, 3$, are equal to the squared Euclidean distance between the corresponding surviving sequence $\check{\mathbf{x}}(\varrho_k^{(j)})$ and the received sequence \mathbf{v} at the output of the noise whitening filter.

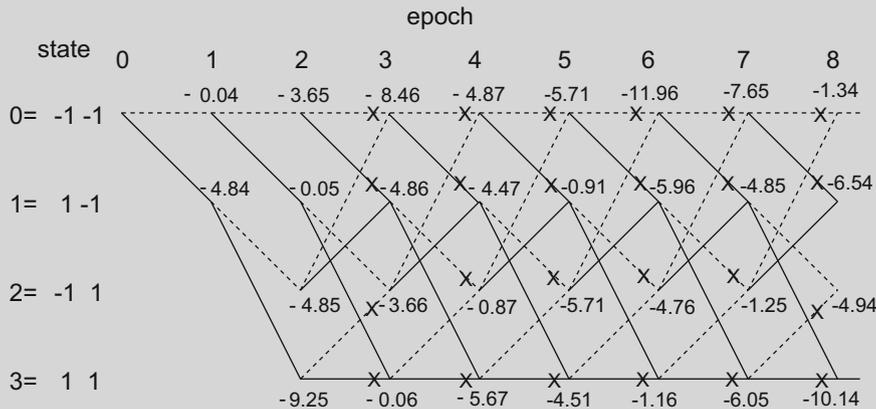


Fig. 7.12 Cumulative path metrics and surviving sequences with the Viterbi algorithm

As implied in (7.137), the ML receiver theoretically 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. One method for solving this problem is to modify the Viterbi algorithm to introduce a fixed decoding delay. Typically, the N_S surviving sequences $\check{\mathbf{x}}(\varrho_k^{(i)}), i = 1, \dots, N_S - 1$ at time index k will be identical for bit (or symbol) indices $k - Q$ or less where Q is some sufficiently large number. That is, all the surviving sequences will share a common parent sequence for bit (or symbol) indices $k - Q$ or less. With this in mind, one possibility is to modify the Viterbi algorithm to implement a finite fixed decoding delay, by storing only the Q most recent bits or symbols for each surviving sequence. When the channel vector \mathbf{V}_k is received and the path metrics $\Gamma(\varrho_{k+1}^{(j)}), j = 1, \dots, N_S - 1$ are being calculated, the final decision is made on the bit (or symbol) Q branches back in the trellis by deciding in favor of the bit (or symbol) at index $k - Q$ in the surviving sequence $\check{\mathbf{x}}(\varrho_k^{(i)})$ having the largest path metric $\Gamma(\varrho_k^{(i)})$. It is well known that if $Q > 5L$, the performance degradation caused by the resulting path metric truncation is negligible [335].

Another possibility is to transmit the data in blocks $\mathbf{x} = (x_1, x_2, \dots, x_N)$ of length N , and use tail symbols to terminate the trellis in a known state. Recall that the state at epoch k is $Q_k = (x_{k-1}, x_{k-2}, \dots, x_{k-L})$. Therefore, if each block of N data symbols is appended with a known L -symbol tail sequence, the trellis will terminate in a known state that is uniquely determined by the L -symbol tail sequence. After the received vector \mathbf{V}_{N+L} has been processed there will be only one surviving sequence left, and this surviving sequence is then used to make a decision on the entire packet of bits (or symbols). This approach suffers a loss in power and data rate by the factor $N/(N+L)$ due to the insertion of tail symbols.

7.4.1.1 Adaptive MLSE Receiver

The Viterbi algorithm requires knowledge of the channel vectors \mathbf{g}_d , $d = 1, \dots, D$ to compute the branch metrics in (7.139) so that an adaptive channel estimator is needed. Various channel estimators have been proposed in the literature [72, 106, 215]. Often the LMS algorithm is used for this purpose, because of its good performance, numerical stability, and simplicity of implementation [162, 215]. Another possible adaptation algorithm is the Recursive Least Squares (RLS) or the Kalman algorithm [162]. 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 [272]. 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 [106, 205, 302]. 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. As mentioned previously a decision on the data symbol x_{k-Q} , denoted by \hat{x}_{k-Q} , is made when the vector \mathbf{V}_k is received and processed. With the LMS algorithm, the tap coefficients are updated according to

$$\begin{aligned} \hat{g}_{i,d}(k+1) &= \hat{g}_{i,d}(k) + \alpha \epsilon_{k-Q,d} \hat{x}_{k-i-Q}^*, \quad i = 0, \dots, L \\ & \quad d = 1, \dots, D, \end{aligned} \quad (7.142)$$

where α is the adaptation step size, and

$$\epsilon_{k-Q,d} = v_{k-Q,d} - \sum_{i=0}^L \hat{g}_{i,d}(k) \hat{x}_{k-i-Q} \quad (7.143)$$

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, write

$$v_{k-Q,d} = \sum_{i=0}^L g_{i,d}(k-Q) x_{k-i-Q} + \eta_{k-Q,d} \quad (7.144)$$

so that

$$\epsilon_{k-Q,d} = \sum_{i=0}^L \left(g_{i,d}(k-Q) - \hat{g}_{i,d}(k) \right) x_{k-i-Q} + \eta_{k-Q,d}. \quad (7.145)$$

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 unfortunately reduce the reliability of the data decisions \hat{x}_{k-i-Q} . Since decision errors will also degrade the performance of the channel estimator, the overall performance improvement obtained by reducing Q is minimal at best.

One effective solution to this problem is to use per-survivor processing [204, 277, 297, 298], where *each state* maintains a separate channel estimator to track the channel. After storing the path metrics $\Gamma(Q_{k+1}^{(j)})$ and the associated surviving sequences $\check{\mathbf{x}}(Q_{k+1}^{(j)})$ for each state $j = 0, \dots, N_S - 1$, the channel tap estimates are updated according to

$$\begin{aligned}
\hat{g}_{i,d}^{(j)}(k+1) &= \hat{g}_{i,d}^{(j)}(k) + \alpha \epsilon_{k,d}^{(j)} \check{\mathbf{x}}_{k+1-i}^*(\varrho_{k+1}^{(j)}), \quad i = 0, \dots, L \\
&\quad d = 1, \dots, D \\
&\quad j = 0, \dots, N_S - 1,
\end{aligned} \tag{7.146}$$

where

$$\epsilon_{k,d}^{(j)} = v_{k,d} - \sum_{i=0}^L \hat{g}_{i,d}^{(j)}(k) \check{\mathbf{x}}_{k+1-i}^*(\varrho_{k+1}^{(j)}) \tag{7.147}$$

and $\check{\mathbf{x}}_{k+1-i}(\varrho_{k+1}^{(j)})$ is element $k+1-i$ of the surviving sequence $\check{\mathbf{x}}(\varrho_{k+1}^{(j)})$ associated with state $\varrho_{k+1}^{(j)}$. 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. The zero-delay symbols that are used to update the channel tap estimates associated with state $\varrho_{k+1}^{(j)}$ are uniquely defined by surviving sequence $\check{\mathbf{x}}(\varrho_{k+1}^{(j)})$ that is associated with $\varrho_{k+1}^{(j)}$.

7.4.1.2 Fractionally Spaced MLSE Receiver

As mentioned previously, the matched filter outputs are often over-sampled to aid symbol timing synchronization and to mitigate the effects of sample timing errors. 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 (7.134). However, the Viterbi decoder searches for the most likely path in the trellis based on 2 samples per branch. For each state transition $\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)}$ at epoch k , the samples $v_{2k}^{(2)}$ and $v_{2k+1}^{(2)}$ are used by the Viterbi algorithm to evaluate the branch metric²

$$\begin{aligned}
\mu(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)}) &= \left| v_{2k}^{(2)} - g_0^{(2)} x_k(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)}) - \sum_{m=1}^L g_{2m}^{(2)} x_{k-m}(\varrho_k^{(i)}) \right|^2 \\
&\quad + \left| v_{2k+1}^{(2)} - g_1^{(2)} x_k(\varrho_k^{(i)} \rightarrow \varrho_{k+1}^{(j)}) - \sum_{m=1}^{L-1} g_{2m+1}^{(2)} x_{k-m}(\varrho_k^{(i)}) \right|^2.
\end{aligned} \tag{7.148}$$

Other than the change in the branch metric, the Viterbi algorithm proceeds as before. Note also that the adaptive channel estimator must estimate and track two different channel vectors; $\mathbf{g}_e^{(2)} = \{g_0^{(2)}, g_2^{(2)}, \dots, g_{2L}^{(2)}\}$ and $\mathbf{g}_o^{(2)} = \{g_1^{(2)}, g_3^{(2)}, \dots, g_{2L-1}^{(2)}\}$.

7.4.2 Delayed Decision-Feedback Sequence Estimation

The complexity of the MLSE receiver grows exponentially with the channel memory length. When the channel memory length becomes large, the MLSE receiver quickly becomes impractical. Considerable research has been undertaken to reduce the complexity of MLSE while retaining most of its performance. Duel-Hallen and Heegard [100, 101] proposed delayed decision-feedback sequence estimation DDFSE, a technique that reduces 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 receiver is obtained with a complexity that is controlled by the parameter μ .

The system function $G(z)$ of the overall discrete-time white noise channel can be written as

$$G(z) = G_\mu(z) + z^{-(\mu+1)} G^+(z), \tag{7.149}$$

²For notational simplicity, it is assumed that $D = 1$.

where

$$G_\mu(z) = \sum_{i=0}^{\mu} g_i z^{-i} \quad (7.150)$$

$$G^+(z) = \sum_{i=0}^{L-\mu-1} g_{i+\mu+1} z^{-i}. \quad (7.151)$$

Let $U(z) = G^+(z)X(z)$, where $X(z)$ is the z -transform of the input sequence. Then

$$u_k = \sum_{i=0}^{L-\mu-1} g_{i+\mu+1} x_{k-i} \quad (7.152)$$

and

$$v_k = \sum_{i=0}^{\mu} g_i x_{k-i} + u_{k-\mu-1} + w_k. \quad (7.153)$$

From (7.152) and (7.153), the system state at epoch k can be decomposed into the state

$$\varrho_k^\mu = (x_{k-1}, \dots, x_{k-\mu}) \quad (7.154)$$

and a partial state

$$\kappa_k = (x_{k-\mu-1}, \dots, x_{k-L}). \quad (7.155)$$

There are $N_\mu = 2^{\mu}$ states in (7.154).

The DDFSE receiver can be viewed as a Viterbi algorithm with a decision feedback mechanism. For each state transition $\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)}$, the DDFSE receiver uses the branch metric

$$\mu_k(\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)}) = - \left| v_k - g_0 x_k(\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)}) - \sum_{\ell=1}^{\mu} g_\ell x_{k-\ell}(\varrho_k^{\mu(i)}) - \sum_{\ell=\mu+1}^L g_\ell \check{x}_{k-\ell}(\varrho_k^{\mu(i)}) \right|^2, \quad (7.156)$$

where $\check{x}_{k-\ell}(\varrho_k^{\mu(i)})$ is element $k-\ell$ 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 as severe as with a DFE receiver. When $\mu = 0$ the DDFSE receiver is equivalent to Driscoll's decoder [99] 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 (7.154), 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. If $G(z)$ does not have minimum phase, DDFSE does not work as well.

Example 7.6. Consider again the system in Example 7.4, where the received sequence is

$$\begin{aligned} \mathbf{v} &= (v_0, v_1, v_2, v_3, v_4, \dots) \\ &= (-3.2, -1.1, 0.9, 0.1, 1.2, 1.5, 0.7, -1.3, \dots) \end{aligned}$$

Recall that $\varrho_k^{(i)} = (x_{k-1}, x_{k-2})$ so there are 4 system states. However, DDFSE is to be used with the state $\varrho_k^{\mu(i)} = x_{k-1}$, $i = 0, 1$. The initial state is assumed to be $\varrho_0^{\mu(0)} = -1$. Since the channel has finite length, (7.156) gives the

(continued)

Example 7.6 (continued)
branch metric

$$\mu_k(\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)}) = \left| y_k - x_k(\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)}) - x_{k-1}(\varrho_k^{\mu(i)}) - \check{x}_{k-2}(\varrho_k^{\mu(i)}) \right|^2.$$

Applying DDFSE with the Viterbi algorithm gives the result shown in Fig. 7.13. Once again, the X 's on the branches in the trellis denote the dropped paths and the numbers on the trellis nodes denote the path metrics.

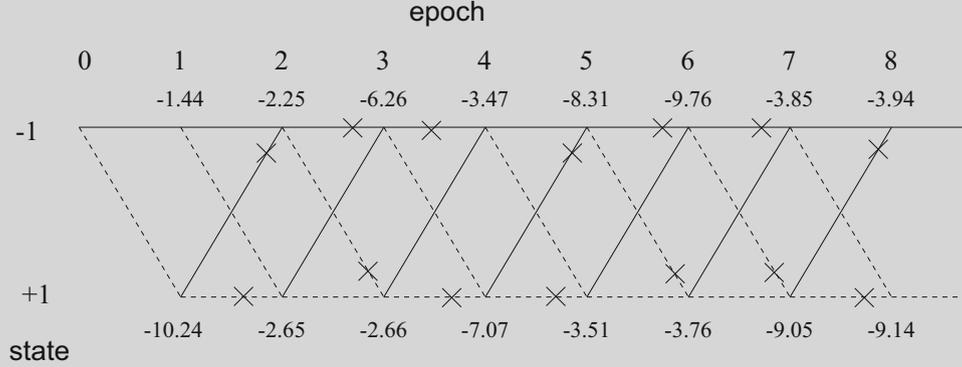


Fig. 7.13 Cumulative path metrics and surviving sequences with DDFSE

7.4.3 Reduced-State Sequence Estimation

For large signal constellations the number of states with DDFSE, $2^{n\mu}$, is substantial even for small μ . Eyuboğlu and Qureshi [119] proposed a reduced complexity method called reduced-state sequence estimation (RSSE), a technique that is especially useful for systems with large signal constellations. RSSE reduces the number of states by using Ungerboeck-like set partitioning principles that were developed for trellis-coded modulation [325]. As described in [119], for each element x_{k-n} in $\varrho_k^\mu = (x_{k-1}, \dots, x_{k-\mu})$, 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 $J_1 \geq J_2 \dots \geq J_\mu$. In this case this following subset-state can be defined

$$t_k^\mu = (c_1(x_{k-1}), c_2(x_{k-2}), \dots, c_\mu(x_{k-\mu})). \quad (7.157)$$

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 $c_0(x_k)$ 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 $2^{n\mu}$. Note that if $J_1 < 2^n$, there are parallel transitions associated with each subset-transition. The number of the parallel transitions is equal to the number of symbols in the corresponding subset.

³If $J_1 = J_2 = \dots = J_\mu = M$ and $\mu < L$, then RSSE becomes DDFSE.

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 first⁵ and then the steps are executed as defined in Sect. 7.4.1.

With RSSE, the branch metric in (7.139) 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 [101, 119]. The RSSE branch metric for a particular parallel transition associated with the subset-transition $t_k^{\mu(i)} \rightarrow t_{k+1}^{\mu(j)}$ is

$$\mu(t_k^{\mu(i)} \rightarrow t_{k+1}^{\mu(j)}) = - \left| v_k - g_0 x_k(t_k^{\mu(i)} \rightarrow t_{k+1}^{\mu(j)}) - \sum_{\ell=1}^L g_\ell \check{x}_{k-\ell}(t_k^{\mu(i)}) \right|^2 \quad (7.158)$$

where $x_k(t_k^{\mu(i)} \rightarrow t_{k+1}^{\mu(j)})$ is the data symbol corresponding to the particular parallel transition, and $\check{x}_{k-\ell}(t_k^{\mu(i)})$ is element $k - \ell$ of the surviving sequence $\check{\mathbf{x}}(t_k^{\mu(i)})$ associated with the subset-state $t_k^{\mu(i)}$.

7.5 Error Probability for MLSE on ISI Channels

The error probability performance of MLSE on ISI channels is now considered. It is impossible to derive the exact error probabilities so that bounding techniques, such as the union bound, must be used. Let \mathbf{x} and $\hat{\mathbf{x}}$ be the transmitted and estimated symbol sequences, respectively. For every pair \mathbf{x} and $\hat{\mathbf{x}}$, an error sequence $\boldsymbol{\epsilon} = \{\epsilon_i\}$ can be formed by defining $\epsilon_i = x_i - \hat{x}_i$. It is arbitrarily assumed 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:

$\mathcal{E}'(\boldsymbol{\epsilon})$: 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

$$\mathcal{E}'_G = \bigcup_{\boldsymbol{\epsilon} \in G} \mathcal{E}'(\boldsymbol{\epsilon}) \quad (7.159)$$

and

$$\mathcal{E}_F = \bigcup_{\boldsymbol{\epsilon} \in F} \mathcal{E}(\boldsymbol{\epsilon}), \quad (7.160)$$

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 non-zero 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

$$\varrho_{k_1} = \hat{\varrho}_{k_1}, \varrho_{k_2} = \hat{\varrho}_{k_2}, \text{ and } \varrho_j \neq \hat{\varrho}_j \text{ for } k_1 < j < k_2, \quad (7.161)$$

where $k_1 \leq j_1 \leq k_2$. The symbol error probability at epoch j_1 is

$$\begin{aligned} P_s(j_1) &= \text{P}[x_{j_1} \neq \hat{x}_{j_1}] \\ &= \text{P}[\mathcal{E}'_G] \\ &= \sum_{\boldsymbol{\epsilon} \in G} \sum_{\mathbf{x} \in \mathcal{X}'(\boldsymbol{\epsilon})} \text{P}[\mathcal{E}'(\boldsymbol{\epsilon})|\mathbf{x}] \text{P}[\mathbf{x}], \end{aligned} \quad (7.162)$$

⁴With DDFSE there are no parallel transitions.

⁵If the signal constellation has some symmetries, this step can be done by using a slicing operation [119].

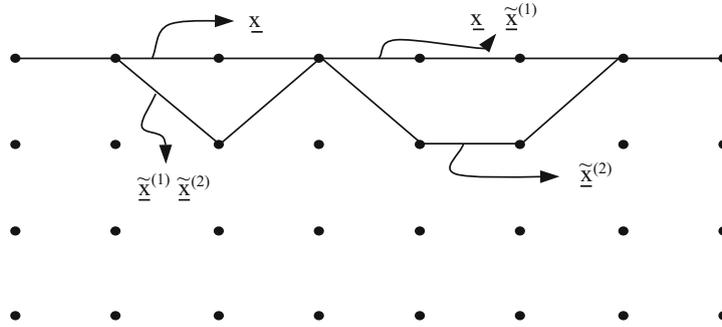


Fig. 7.14 A typical error state trellis diagram

where $\mathcal{X}(\epsilon)$ is the set of symbol sequences that can have ϵ as the error sequence. For different ϵ , the set $\mathcal{X}(\epsilon)$ might be different. The third equation in (7.162) is obtained by using the property that the events $\mathcal{E}^l(\epsilon)$ are disjoint for $\epsilon \in G$. Unfortunately, (7.162) 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, it is now proven that the symbol error probability at epoch j_1 is

$$P_s(j_1) = P[\mathcal{E}_F]. \quad (7.163)$$

Consider the typical trellis diagram as shown in Fig. 7.14, 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 $\epsilon^{(1)}$ associated with $\tilde{\mathbf{x}}^{(1)}$ and the error sequence $\epsilon^{(2)}$ associated with $\tilde{\mathbf{x}}^{(2)}$ belong to sets F and $G \setminus F$, respectively. For every $\epsilon^{(2)} \in G \setminus F$ there always exists an $\epsilon^{(1)} \in F$. If the sequence $\mathbf{x} - \epsilon^{(2)}$ is the ML sequence, i.e., the event \mathcal{E}_G^l has occurred, then the sequence $\mathbf{x} - \epsilon^{(1)}$ 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^l implies \mathcal{E}_F . On the other hand, if $\epsilon^{(1)} \in F$ and the sequence $\mathbf{x} - \epsilon^{(1)}$ has a larger path metric than sequence \mathbf{x} , then there exists a sequence $\epsilon \in G$ such that the sequence $\mathbf{x} - \epsilon$ is the ML sequence. Therefore, \mathcal{E}_F implies \mathcal{E}_G^l , and (7.163) is proven.

The union bound on (7.163) yields

$$\begin{aligned} P_s(j_1) &\leq \sum_{\epsilon \in F} P[\mathcal{E}^l(\epsilon)] \\ &= \sum_{\epsilon \in F} \sum_{\mathbf{x} \in \mathcal{X}(\epsilon)} P[\mathcal{E}^l(\epsilon)|\mathbf{x}]P[\mathbf{x}] \end{aligned} \quad (7.164)$$

or, equivalently,

$$P_s(j_1) \leq \sum_{\epsilon \in E} w_s(\epsilon) \sum_{\mathbf{x} \in \mathcal{X}(\epsilon)} P[\mathcal{E}^l(\epsilon)|\mathbf{x}]P[\mathbf{x}], \quad (7.165)$$

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 (7.165), the following observations were used; 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 $P[\mathcal{E}^l(\epsilon)|\mathbf{x}]$ 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 $\{\mathcal{E}^l(\epsilon)\}$ for $\epsilon \in F$ in (7.164) 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 $\mathcal{E}^l(\epsilon)$ and, hence, the union bound (7.164) becomes looser.

From the definition of event $\mathcal{E}(\epsilon)$, the union bound (7.165) becomes

$$P_s \leq \sum_{\epsilon \in E} w_s(\epsilon) \sum_{\mathbf{x} \in \mathcal{X}(\epsilon)} P[\Gamma(\mathbf{x} - \epsilon) \geq \Gamma(\mathbf{x}) | \mathbf{x}] P[\mathbf{x}], \quad (7.166)$$

where $\Gamma(\mathbf{x})$ is the path metric associated with the input sequence \mathbf{x} . To obtain the bit error probability, (7.166) can be easily modified as

$$P_b \leq \frac{1}{n} \sum_{\epsilon \in E} w_b(\epsilon) \sum_{\mathbf{x} \in \mathcal{X}(\epsilon)} P[\Gamma(\mathbf{x} - \epsilon) \geq \Gamma(\mathbf{x}) | \mathbf{x}] P[\mathbf{x}], \quad (7.167)$$

where n is the number bits transmitted per unit time, and $w_b(\epsilon)$ is the number of bit errors associated with the error sequence ϵ . The probability

$$P[\Gamma(\mathbf{x} - \epsilon) \geq \Gamma(\mathbf{x}) | \mathbf{x}] \quad (7.168)$$

is called the pairwise error probability. It will be shown in the following two sections that the pairwise error probability is independent of the transmitted symbol sequence \mathbf{x} . Therefore, the union bounds (7.166) and (7.167) simplify to

$$P_s \leq \sum_{\epsilon \in E} w_s(\epsilon) P[\Gamma(\mathbf{x} - \epsilon) \geq \Gamma(\mathbf{x}) | \mathcal{X}(\epsilon)] P[\mathcal{X}(\epsilon)] \quad (7.169)$$

and

$$P_b \leq \frac{1}{n} \sum_{\epsilon \in E} w_b(\epsilon) P[\Gamma(\mathbf{x} - \epsilon) \geq \Gamma(\mathbf{x}) | \mathcal{X}(\epsilon)] P[\mathcal{X}(\epsilon)], \quad (7.170)$$

respectively. The expressions in (7.169) and (7.170) are easier to calculate than those in (7.166) and (7.167), because not all of the symbol sequences need to be considered in the calculation.

7.5.1 Static ISI Channels

The pairwise error probability associated with the error event of length $\ell = k_2 - k_1$ in (7.161) is (Problem 7.16)

$$P[\Gamma(\mathbf{x} - \epsilon) \geq \Gamma(\mathbf{x}) | \mathbf{x}] = Q\left(\sqrt{\frac{\Delta^2}{4N_o}}\right), \quad (7.171)$$

where

$$\Delta^2 = \sum_{k=k_1}^{k_1+\ell-1} \left| \sum_{i=0}^L g_i \epsilon_{k-i} \right|^2 \quad (7.172)$$

and Δ^2 is the squared Euclidean path distance. At high signal-to-noise ratios the error event probability is approximately

$$P_e \approx N_{\min} Q\left(\sqrt{\frac{d_{\min}^2}{4N_o}}\right) \quad (7.173)$$

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 (7.172) can be rewritten as

$$\Delta^2 = \sum_{k=k_1}^{k_1+\ell-1} \Delta_k^2, \quad (7.174)$$

where

$$\Delta_k^2 = \mathbf{g}^H \mathbf{E}_k \mathbf{g} \quad (7.175)$$

is the squared branch distance and

$$\mathbf{E}_k = [(e_{mn})_k] \quad (7.176)$$

is the $(L+1) \times (L+1)$ branch distance matrix with elements $(e_{mn})_k = \epsilon_{k-m+1}^* \epsilon_{k-n+1}$. Define the error vector

$$\mathbf{e}_k = (\epsilon_k, \epsilon_{k-1}, \dots, \epsilon_{k-L})^H. \quad (7.177)$$

It follows that $\mathbf{E}_k = \mathbf{e}_k \mathbf{e}_k^H$ and, hence, \mathbf{E}_k has rank one. Note that $\mathbf{E}_k \mathbf{e}_k = (\mathbf{e}_k^H \mathbf{e}_k) \mathbf{e}_k$ and, therefore, \mathbf{e}_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 (7.161) is defined as

$$\mathbf{E} \triangleq \sum_{k=k_1}^{k_1+\ell-1} \mathbf{E}_k. \quad (7.178)$$

Using (7.134) and (7.161), the elements of \mathbf{E} are

$$e_{mn} = r_\ell(n-m) \quad (7.179)$$

where

$$r_\ell(i) = \begin{cases} \sum_{k=k_1}^{k_1+\ell-(L+1)-i} \epsilon_k \epsilon_{k+i}^* & i \geq 0 \\ r_\ell^*(-i) & i < 0 \end{cases}. \quad (7.180)$$

It follows that (7.174) 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 noting that $\mathbf{g}^H \mathbf{g} = f_0 = 2E_{h_r}$, the squared Euclidean path distance can be expressed in the form

$$\Delta^2 = 2E_{h_r} \frac{\mathbf{g}^H \mathbf{E} \mathbf{g}}{\mathbf{g}^H \mathbf{g}} = 2E_{h_r} R(\mathbf{g}), \quad (7.181)$$

where 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 $R(\mathbf{g})$ [162]. 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 $R(\mathbf{g})$ corresponds to the smallest possible Δ^2 (or largest pairwise error probability) for a given error sequence ϵ , and occurs when $\mathbf{g} = \mathbf{v}_{\min}$. Likewise, the maximum value of $R(\mathbf{g})$ corresponds to the largest possible Δ^2 (or smallest pairwise error probability) for a given error sequence ϵ , and occurs when $\mathbf{g} = \mathbf{v}_{\max}$. Note that the eigenvalues of \mathbf{E} are only a function of the error sequence ϵ and do not depend on the channel vector \mathbf{g} . While the eigenvalues of \mathbf{E} can be calculated exactly for a given error sequence ϵ , the eigenvalues of \mathbf{E} are bounded by [162]

$$\lambda_{\max} \leq \sum_{i=0}^L |r_\ell(i)| \quad \text{and} \quad \lambda_{\min} \geq r_\ell(0) - \sum_{i=1}^L |r_\ell(i)|. \quad (7.182)$$

The upper and lower bounds on the eigenvalues are easier to calculate than the eigenvalues themselves. Finally, the condition number of \mathbf{E} is defined as $c(\mathbf{E}) \triangleq \lambda_{\max}/\lambda_{\min}$, and $c(\mathbf{E}) \geq 1$. If the condition number of a particular error sequence $\boldsymbol{\epsilon}$ is large (small), then the corresponding pairwise error probability will have a large (small) variation with the channel vector \mathbf{g} .

7.5.2 Fading ISI Channels

Consider the case of a fading ISI channel with D -branch diversity reception and maximal ratio combining. The pairwise error probability is still given by (7.171) but the squared Euclidean path distance associated with an error event of length ℓ is [301]

$$\Delta^2 = \sum_{d=1}^D \Delta_d^2, \quad (7.183)$$

where

$$\Delta_d^2 = \sum_{k=k_1}^{k_1+\ell-1} \left| \sum_{i=0}^L g_{i,d}(k) \epsilon_{k-i} \right|^2. \quad (7.184)$$

The above expression can be written in the form

$$\Delta_d^2 = \sum_{k=k_1}^{k_1+\ell-1} \mathbf{g}_d^H(k) \mathbf{E}_k \mathbf{g}_d(k). \quad (7.185)$$

In general, the correlation matrix $\Phi_{\mathbf{g}_d}(0)$, where $\Phi_{\mathbf{g}_d}(m)$ is defined in (7.43) is not diagonal. Although a non-diagonal $\Phi_{\mathbf{g}_d}(0)$ matrix is more realistic, it leads to considerable analytical difficulty and loss of insight. Therefore, our attention is restricted to the case where $\Phi_{\mathbf{g}_d}(0)$ is a diagonal matrix, meaning that the channel tap gains are all mutually uncorrelated (assuming they have zero mean). In such a case, a normalized channel vector $\hat{\mathbf{g}}_d(k)$ can be defined such that $\Phi_{\hat{\mathbf{g}}_d}(0) = \mathbf{I}_{L+1}$. As a result, (7.185) can be rewritten as

$$\Delta_d^2 = \sum_{k=k_1}^{k_1+\ell-1} \hat{\mathbf{g}}_d^H(k) \mathbf{A}_{k,d} \hat{\mathbf{g}}_d(k), \quad (7.186)$$

where

$$\mathbf{A}_{k,d} = \boldsymbol{\Sigma}_d \mathbf{E}_k \boldsymbol{\Sigma}_d \quad (7.187)$$

and

$$\boldsymbol{\Sigma}_d = \text{diag}[\sigma_{0,d}, \sigma_{1,d}, \dots, \sigma_{L,d}] \quad (7.188)$$

with $\sigma_{i,d}^2 = \frac{1}{2} \mathbb{E}[|g_{i,d}|^2]$. It follows that $\mathbf{A}_{k,d} = \mathbf{u}_{k,d} \mathbf{u}_{k,d}^H$ where $\mathbf{u}_{k,d} = \boldsymbol{\Sigma}_d \mathbf{e}_k$ 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 non-zero eigenvalue of $\mathbf{A}_{k,d}$ is $\lambda_d = \sum_{i=0}^L \sigma_{i,d}^2 |\epsilon_{k-i}|^2$.

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 $J_0(2\pi f_m |k|T) \geq J_0(2\pi f_m 20T) = 0.9755 \approx 1$. By using the above assumption, (7.186) can be written as

$$\Delta_d^2 = \hat{\mathbf{g}}_d^H \mathbf{A}_d \hat{\mathbf{g}}_d, \quad (7.189)$$

where

$$\begin{aligned}\mathbf{A}_d &= \sum_{k=k_1}^{k_1+\ell-1} \mathbf{A}_{k,d} \\ &= \boldsymbol{\Sigma}_d \mathbf{E} \boldsymbol{\Sigma}_d.\end{aligned}\quad (7.190)$$

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 $[(a_{mm})]_d = \sigma_{m-1,d} \sigma_{n-1,d} r_\ell(n-m)$ where $r_\ell(i)$ is given by (7.180). The trace of the matrix \mathbf{A}_d is

$$\text{tr}(\mathbf{A}_d) = \sum_{i=0}^L \lambda_{i,d} = (\bar{E}/\sigma_x^2) r_\ell(0), \quad (7.191)$$

where the $\lambda_{i,d}$, $i = 0, \dots, L$ are the eigenvalues of \mathbf{A}_d . The last equality in (7.191) is obtained by using (7.45). 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 $\boldsymbol{\omega}_d = \mathbf{U}_d^H \hat{\mathbf{g}}_d$ be the corresponding diagonal transformation. Hence,

$$\Delta_d^2 = \boldsymbol{\omega}_d^H \boldsymbol{\Lambda}_d \boldsymbol{\omega}_d = \sum_{i=0}^L \lambda_{i,d} |\omega_{i,d}|^2, \quad (7.192)$$

where $\frac{1}{2} \mathbf{E}[\boldsymbol{\omega}_d \boldsymbol{\omega}_d^H] = \mathbf{I}_{L+1}$ so that the $\{\omega_{i,d}\}$ are independent zero-mean unit-variance Gaussian random variables. Using (7.183) and (7.192) gives

$$\Delta^2 = \sum_{d=1}^D \sum_{i=0}^L \alpha_{i,d}, \quad (7.193)$$

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

$$\psi_{\Delta^2}(z) = \prod_{d=1}^D \prod_{i=0}^L \frac{1}{1 - \bar{\alpha}_{i,d} z}, \quad (7.194)$$

where $\bar{\alpha}_{i,d} = 2\lambda_{i,d}$. Finally, the pairwise error probability is

$$\text{P}[\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x}) | \mathbf{x}] = \int_0^\infty \mathcal{Q}(\sqrt{2x}) f_{\Delta^2}(x) dx, \quad (7.195)$$

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 (7.194). This will be the case for balanced diversity branches, and will also be the case if the channel vectors \mathbf{g}_d have some equal strength taps. Consider the case where balanced D -branch diversity is used, but where the channel taps in each diversity branch are not of equal strength. In this case, $\lambda_{i,d} \equiv \lambda_i$, $d = 1, \dots, D$ and the characteristic function in (7.194) has the form

$$\begin{aligned}\psi_{\Delta^2}(z) &= \prod_{i=0}^L \frac{1}{(1 - z\bar{\alpha}_i)^D} \\ &= \sum_{i=0}^L \sum_{d=1}^D \frac{A_{id}}{(1 - z\bar{\alpha}_i)^d},\end{aligned}\quad (7.196)$$

where

$$A_{id} = \frac{1}{(D-d)!(-\bar{\alpha}_i)^{D-d}} \left\{ \frac{d^{D-d}}{dz^{D-d}} (1 - z\bar{\alpha}_i)^D \psi_{\Delta^2}(z) \right\}_{z=1/\bar{\alpha}_i} \quad (7.197)$$

and $\bar{\alpha}_i = 2\lambda_i$. The pdf of Δ^2 is

$$f_{\Delta^2}(x) = \sum_{i=0}^L \sum_{d=1}^D A_{id} \frac{1}{(d-1)!(\bar{\alpha}_i)^d} x^{d-1} e^{-x/\bar{\alpha}_i}, \quad x \geq 0. \quad (7.198)$$

From (7.195) and (7.198), the exact pairwise error probability is

$$P[\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x}] = \sum_{i=0}^L \sum_{d=1}^D A_{id} \left(\frac{1-\mu_i}{2}\right)^d \sum_{m=0}^{d-1} \binom{d-1+m}{m} \left(\frac{1+\mu_i}{2}\right)^m, \quad (7.199)$$

where

$$\mu_i = \sqrt{\frac{\bar{\alpha}_i}{1+\bar{\alpha}_i}}. \quad (7.200)$$

From (7.191), the $\bar{\alpha}_{i,d}$ have the sum value constraint

$$\sum_{i=0}^L \bar{\alpha}_{i,d} = 2 \sum_{i=0}^L \lambda_{i,d} = (2\bar{E}/\sigma_x^2)r_\ell(0). \quad (7.201)$$

Define $S \subseteq R^{L+1}$ as the set of all $(L+1)$ -component vectors $\{\bar{\boldsymbol{\alpha}}_d : \sum_{i=0}^L \bar{\alpha}_{i,d} = (2\bar{E}/\sigma_x^2)r_\ell(0)\}$. The set S is convex, since for any pair of vectors $\bar{\boldsymbol{\alpha}}_{d,k}$ and $\bar{\boldsymbol{\alpha}}_{d,j}$ the convex combination $\theta\bar{\boldsymbol{\alpha}}_{d,k} + (1-\theta)\bar{\boldsymbol{\alpha}}_{d,j}$ is contained in S for any $0 \leq \theta \leq 1$. If the pairwise error probability is treated as a mapping from S to R , then it is a convex function of $\bar{\boldsymbol{\alpha}}_d$ and, hence, has a unique minimum. For example, Fig. 7.15 shows the pairwise error probability for a three-tap channel ($L = 2, D = 1$) with equal strength taps ($\bar{\alpha}_0 = \bar{\alpha}_1 = \bar{\alpha}_2$). Note that the value of $\bar{\alpha}_2$ is determined uniquely by the values of $\bar{\alpha}_0$ and $\bar{\alpha}_1$, and that is why a three dimensional graph is used. By using variational calculus, it can be shown (Appendix 7A) that the pairwise error probability is minimized when the $\bar{\alpha}_{i,d}$ are all equal, i.e., $\lambda_{i,d} = \lambda = (\bar{E}/\sigma_x^2)r_\ell(0)/(L+1)$, resulting in the minimum pairwise error probability

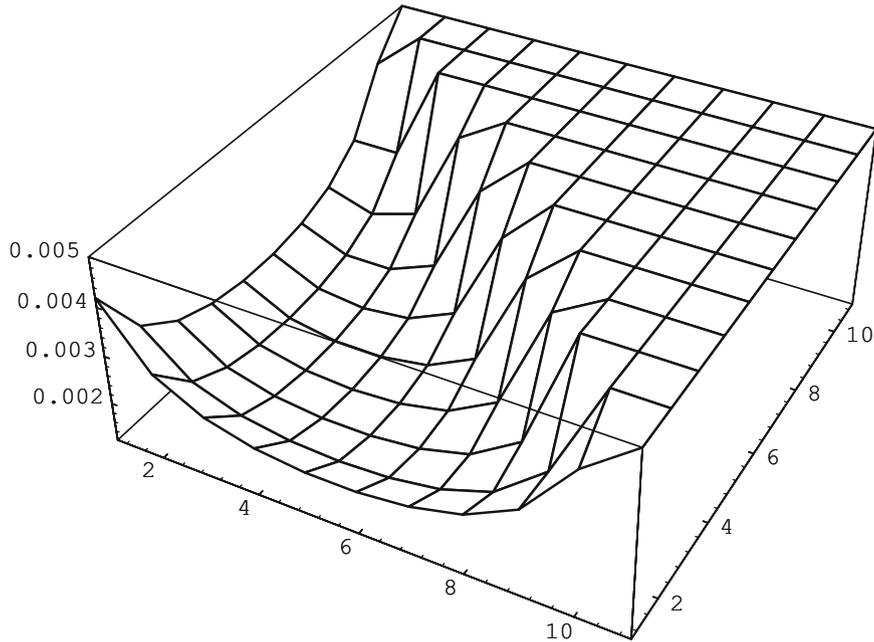


Fig. 7.15 Pairwise error probability for a three-tap fading ISI channel

$$P_{\min} = \left(\frac{1-\mu}{2}\right)^{D(L+1)} \sum_{m=0}^{D(L+1)-1} \binom{D(L+1)-1+m}{m} \left(\frac{1+\mu}{2}\right)^m, \quad (7.202)$$

where

$$\mu = \sqrt{\frac{\lambda/4N_o}{1 + \lambda/4N_o}}. \quad (7.203)$$

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 $c(\mathbf{A}_d) = c(\mathbf{\Sigma}_d \mathbf{E} \mathbf{\Sigma}_d) \leq (c(\mathbf{\Sigma}_d))^2 c(\mathbf{E})$, where $(c(\mathbf{\Sigma}_d))^2$ represents the ratio of the maximum and minimum channel tap variances $(\sigma_d^2)_{\max}/(\sigma_d^2)_{\min}$ for diversity branch d . It has been shown previously that \mathbf{E} depends only on 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 $c(\mathbf{A}_d) \leq c(\mathbf{E})$ 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.

7.6 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 received sequences that are corrupted by noise samples with variance N_o . Therefore, it is only necessary to compare the Euclidean distances between allowed sequences of channel outputs to determine the relative performance of the T - and $T/2$ -spaced receivers.

7.6.1 T -Spaced MLSE Receiver

From the definition of the error event in (7.161), the z -transform of the length- ℓ error sequence $\boldsymbol{\epsilon} = \{\epsilon_{k_1}, \epsilon_{k_1+1}, \dots, \epsilon_{k_1+\ell-1}\}$ is

$$\mathcal{E}(z) = \epsilon_{k_1} + \epsilon_{k_1+1}z^{-1} + \dots + \epsilon_{k_1+\ell-1}z^{-\ell+1} \quad (7.204)$$

where $\epsilon_k = x_k - \hat{x}_k$. The z -transform of the received signal error sequence associated with the length- ℓ error sequence $\boldsymbol{\epsilon}$ is

$$\begin{aligned} \mathcal{E}_v(z) &= (v_{k_1} - \hat{v}_{k_1}) + (v_{k_1+1} - \hat{v}_{k_1+1})z^{-1} + \dots \\ &\quad \dots + (v_{k_1+\ell+L-1} - \hat{v}_{k_1+\ell+L-1})z^{-(\ell+L-1)} \end{aligned} \quad (7.205)$$

and

$$\mathcal{E}_v(z) = \mathcal{E}(z)G(z). \quad (7.206)$$

From (7.172), the squared Euclidean distance Δ^2 of the error event in (7.161) is [131]

$$\begin{aligned} \Delta^2 &= \sum_{k=k_1}^{k_1+\ell-1} \left| \sum_{i=0}^L g_i \epsilon_{k-i} \right|^2 \\ &= [\mathcal{E}_v(z) \mathcal{E}_v^*(1/z^*)]_0 \\ &= [\mathcal{E}(z)F(z) \mathcal{E}^*(1/z^*)]_0 \end{aligned} \quad (7.207)$$

where $[\cdot]_0$ is the coefficient of z^0 .

7.6.2 $T/2$ -Spaced MLSE Receiver

For the same error event described in (7.161), the corresponding $T/2$ -spaced error sequence is $\epsilon^{(2)} = \{\epsilon_{k_1}^{(2)}, \epsilon_{k_1+1}^{(2)}, \dots, \epsilon_{k_1+2\ell}^{(2)}\} - 1^{(2)}$, where

$$\epsilon_{k_1+i}^{(2)} = \begin{cases} \epsilon_{k_1+i/2}, & i \text{ even} \\ 0, & i \text{ odd} \end{cases}. \quad (7.208)$$

The z -transform of $\epsilon^{(2)}$ is

$$\mathcal{E}^{(2)}(z) = \epsilon_{k_1} + \epsilon_{k_1+1}z^{-2} + \dots + \epsilon_{k_1+\ell+L-1}z^{-2(\ell+L-1)}, \quad (7.209)$$

and because $\epsilon^{(2)}$ is zero in the odd coordinates $\mathcal{E}^{(2)}(z) = \mathcal{E}(z^2)$. The corresponding z -transform of the $T/2$ -spaced received signal error sequence associated with the error sequence $\epsilon^{(2)}$ is

$$\mathcal{E}_v^{(2)}(z) = \mathcal{E}^{(2)}(z)G^{(2)}(z). \quad (7.210)$$

From (7.172), the squared Euclidean distance of the error event in (7.161) is

$$\begin{aligned} (\Delta^{(2)})^2 &= \sum_{k=k_1}^{k_1+2\ell-1} \left| \sum_{i=0}^{2L} g_i^{(2)} \epsilon_{k-i}^{(2)} \right|^2 \\ &= [\mathcal{E}_v^{(2)}(z)\mathcal{E}_v^{(2)*}(1/z^*)]_0 \\ &= [\mathcal{E}^{(2)}(z)F^{(2)}(z)\mathcal{E}^{(2)*}(1/z^*)]_0 \\ &= [\mathcal{E}(z^2)F^{(2)}(z)\mathcal{E}^*(1/z^{*2})]_0, \end{aligned} \quad (7.211)$$

where $F^{(2)}(z) = G^{(2)}(z)(G^{(2)}(1/z^*))^*$. Note that the 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 (7.46) and (7.47) 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,

$$(\Delta^{(2)})^2 = [\mathcal{E}(z^2)F^{(2)}(z)\mathcal{E}^*(1/z^{*2})]_0 = [\mathcal{E}(z)F(z)\mathcal{E}^*(1/z^*)]_0 = \Delta^2. \quad (7.212)$$

Consequently, the T - and $T/2$ -spaced MLSE receivers have identical error probability performance.

Example 7.9. 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

$$\begin{aligned} F(z) &= f_{-1} z + f_0 + f_1 z^{-1} = f_{-2}^{(2)} z + f_0^{(2)} + f_2^{(2)} z^{-1} \\ \mathcal{E}^{(2)}(z) &= \epsilon_0 + \epsilon_1 z^{-2} + \epsilon_2 z^{-4}. \end{aligned}$$

(continued)

Example 7.9 (continued)

Therefore,

$$\begin{aligned} [\mathcal{E}(z)F(z)\mathcal{E}^*(1/z^*)]_0 &= (|\epsilon_0|^2 + |\epsilon_1|^2 + |\epsilon_2|^2)^2 f_0 \\ &\quad + \epsilon_0 \epsilon_1^* f_1 + \epsilon_1 \epsilon_0^* f_{-1} \end{aligned}$$

and

$$\begin{aligned} [\mathcal{E}^{(2)}(z)F^{(2)}(z)\mathcal{E}^{(2)*}(1/z^*)]_0 &= (|\epsilon_0|^2 + |\epsilon_1|^2 + |\epsilon_2|^2)^2 f_0^{(2)} \\ &\quad + \epsilon_0 \epsilon_1^* f_2^{(2)} + \epsilon_1 \epsilon_0^* f_{-2}^{(2)}. \end{aligned}$$

Hence, $\Delta^2 = (\Delta^{(2)})^2$.

7.6.3 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 [275]. A $T/2$ -spaced MLSE receiver is insensitive to the sampler timing phase. This can be readily seen by examining the sampled spectrum at the output of the matched filter. Suppose that the received pulse has less than 100% excess bandwidth, i.e., $H(f) = 0$ for $|f| > 1/T$. With T -spaced sampling and a timing offset of t_o , the sampled spectrum is equal to

$$F_{\Sigma}(f) = \frac{1}{T} \sum_n F\left(f + \frac{n}{T}\right) e^{j2\pi(f+n/T)t_o}. \quad (7.213)$$

This spectrum exhibits aliasing for any received pulse having a bandwidth greater than the Nyquist frequency of $1/(2T)$ Hz. Consider, for example, the sampled spectrum at the Nyquist frequency $f = 1/(2T)$ Hz. If there is no timing offset, then

$$F_{\Sigma}(1/(2T)) = \frac{1}{T} (F(1/(2T)) + F(-1/(2T))).$$

In this case, the sampled spectrum at frequency $f = 1/(2T)$ Hz adds constructively. Now suppose there is a timing offset $t_o = T/2$. Then,

$$F_{\Sigma}(1/(2T)) = \frac{1}{T} (F(1/(2T))e^{j\pi/2} + F(-1/(2T))e^{-j\pi/2}) = 0.$$

In this case, the sampled spectrum at frequency $f = 1/(2T)$ Hz adds destructively and there is a null. Hence, the folded spectrum is sensitive to the timing offset with T -spaced sampling. With $T/2$ -spaced sampling, on the other hand,

$$F_{\Sigma}(f) = \frac{2}{T} \sum_n F\left(f + \frac{2n}{T}\right) e^{j2\pi(f+2n/T)t_o}. \quad (7.214)$$

Notice that the sampled spectrum is a function of the timing offset t_o , but does not exhibit aliasing since $H(f)$ has no more than 100% excess bandwidth.

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 $f_{t_0,n}^{(2)} \neq (f_{t_0,-n}^{(2)})^*$ when a timing offset is present. The discrete-time Fourier transform (DTFT) of $\mathbf{f}_{t_0}^{(2)}$ is

$$F_{t_0}^{(2)}(e^{j\omega}) = F^{(2)}(e^{j\omega})e^{\pm j\omega t_0}, \quad (7.215)$$

where $\tau_0 = t_0/T'$. If the sampler phase is known, then a discrete-time filter with a frequency response function $e^{\mp j\omega\tau_0}$ after the sampler will give the sampled response $\mathbf{f}^{(2)}$ at its output. However, there is no need to correct this phase offset prior to performing MLSE equalization when $T/2$ -spaced sampling is used, as is now shown.

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

$$S_{vv}(f) = N_o F^{(2)}(e^{j\omega}). \quad (7.216)$$

Since the DTFT of the noise whitening filter is

$$1/(G^{(2)}(1/z^*))^*|_{z=e^{j\omega}} = 1/(G^{(2)}(e^{j\omega}))^* \quad (7.217)$$

and

$$F^{(2)}(e^{j\omega}) = G^{(2)}(e^{j\omega}) (G^{(2)}(e^{j\omega}))^* = |G^{(2)}(e^{j\omega})|^2 \quad (7.218)$$

it follows that the noise samples are white at the output of the noise whitening filter with variance N_o . The DTFT of the message signal at the output of the noise whitening filter is

$$X^{(2)}(e^{j\omega})G_{t_0}^{(2)}(e^{j\omega}) = X^{(2)}(e^{j\omega})G^{(2)}(e^{j\omega})e^{\pm j\omega\tau_0} \quad (7.219)$$

and

$$\sum_i |g_i^{(2)}|^2 = \sum_i |g_{t_0,i}^{(2)}|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G^{(2)}(e^{j\omega})|^2 d\omega \quad (7.220)$$

according to Parseval's theorem. Also

$$G_{t_0}^{(2)}(z)(G_{t_0}^{(2)}(1/z^*))^* = F^{(2)}(z). \quad (7.221)$$

Hence, the distances between allowed sequences of channel outputs with the $T/2$ -spaced MLSE receiver in (7.211) remain the same and the performance is insensitive to the sampler phase $e^{\pm j\omega\tau_0}$. The MLSE receiver just needs to estimate the set of tap coefficients $\{g_{t_0,i}^{(2)}\}$ instead of $\{g_i^{(2)}\}$ for use in the branch metric calculations.

7.6.4 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 [327] showed that this receiver is optimum; however, it has some drawbacks. First, it is not suitable for bandwidth efficient systems that use frequency division multiplexing, because the cutoff frequency of the low-pass filter will extend significantly into the adjacent band and pass large amounts of adjacent channel interference. Second, the ideal low-pass filter is non-causal and difficult to approximate. Another solution is to use a receiver filter that is matched to the transmitted pulse $h_a(t)$ followed by a $T/2$ -spaced sampler and a noise whitening filter [65, 156]. If the transmitted signals are strictly bandlimited, e.g., $h_a(t)$ is a root-raised cosine pulse, such that the received pulse $h(t)$ has at most 100% excess bandwidth,⁶ then rate- $2/T$ sampling satisfies the Shannon sampling theorem and the $T/2$ -spaced samples will provide sufficient statistics as is now shown.

For simplicity, consider a time-invariant channel with impulse response $g(t)$. Let $H_a^{(2)}(z)$, $G^{(2)}(z)$, and $H^{(2)}(z)$ be the z -transforms of the $T/2$ -spaced discrete-time signals corresponding to $h_a(t)$, $g(t)$, 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 $F_h^{(2)}(z) = H_a^{(2)}(z)(H_a^{(2)}(1/z^*))^*$. Using the factorization

⁶The received pulse will have a larger bandwidth due to Doppler spreading.

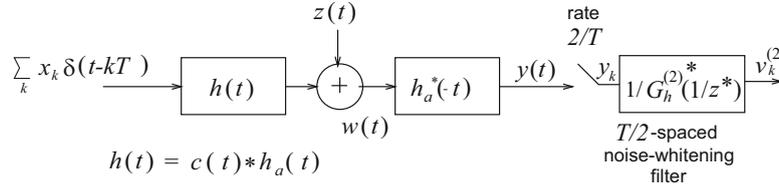


Fig. 7.16 Block diagram of system that implements a filter matched to $h_a(t)$ followed by a $T/2$ -spaced noise whitening filter. The structure of the noise whitening filter depends only on the pulse $h_a(t)$

$$F_h^{(2)}(z) = G_h^{(2)}(z) \left(G_h^{(2)}(1/z^*) \right)^* \quad (7.222)$$

the $T/2$ -spaced noise sequence can be whitened by using a filter with transfer function $1/\left(G_h^{(2)}(1/z^*)\right)^*$ such that $G_h^{(2)}(z)$ has minimum phase. The resulting system is shown in Fig. 7.16. The receivers in Figs. 7.6 and 7.16 are now shown to yield identical performance.

The z -transform of the $T/2$ -spaced discrete-time white noise channel in Fig. 7.16 is

$$\begin{aligned} G_{\text{eq}}^{(2)}(z) &= H_a^{(2)}(z) G^{(2)}(z) \left(H_a^{(2)}(1/z^*) \right)^* / \left(G_h^{(2)}(1/z^*) \right)^* \\ &= G_h^{(2)}(z) G^{(2)}(z). \end{aligned} \quad (7.223)$$

On the other hand, referring to the conventional system shown in Fig. 7.6,

$$H^{(2)}(z) = H_a^{(2)}(z) G^{(2)}(z) \quad (7.224)$$

and

$$F^{(2)}(z) = H_a^{(2)}(z) \left(H_a^{(2)}(1/z^*) \right)^* G^{(2)}(z) \left(G^{(2)}(1/z^*) \right)^*. \quad (7.225)$$

Let

$$G^{(2)}(z) \left(G^{(2)}(1/z^*) \right)^* = G_c^{(2)}(z) \left(G_c^{(2)}(1/z^*) \right)^* \quad (7.226)$$

be a factorization of $G^{(2)}(z) \left(G^{(2)}(1/z^*) \right)^*$ such that $G_c^{(2)}(z)$ has minimum phase. Using (7.222), (7.225), and (7.226) yields

$$F^{(2)}(z) = G_h^{(2)}(z) \left(G_h^{(2)}(1/z^*) \right)^* G_c^{(2)}(z) \left(G_c^{(2)}(1/z^*) \right)^*. \quad (7.227)$$

If the system function of the noise whitening filter is chosen as

$$1 / \left(\left(G_h^{(2)}(1/z^*) \right)^* \left(G_c^{(2)}(1/z^*) \right)^* \right), \quad (7.228)$$

then the system function in Fig. 7.6 is

$$G^{(2)}(z) = G_h^{(2)}(z) G_c^{(2)}(z). \quad (7.229)$$

The system responses $G_{\text{eq}}^{(2)}(e^{j\omega})$ and $G^{(2)}(e^{j\omega})$ have the same amplitude response but a different phase response. Also

$$G_{\text{eq}}^{(2)}(z) \left(G_{\text{eq}}^{(2)}(1/z^*) \right)^* = F^{(2)}(z). \quad (7.230)$$

Therefore, the Euclidean distance between sequences of noise whitening filter outputs for the T -spaced and $T/2$ -spaced systems are the same. It follows that the systems shown in Figs. 7.6 and 7.16 yield identical performance. The main advantage

of the system in Fig. 7.16 is that the noise whitening filter does not depend on the unknown channel and has a fixed structure. Moreover, channel estimation can be performed after the noise whitening filter and the Viterbi algorithm can be implemented using the metric in (7.148). Although the number of computations needed in the $T/2$ -spaced MLSE receiver is twice that of a T -spaced receiver, the latter cannot 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 [249].

7.7 Co-channel Demodulation

By extending Forney's maximum likelihood receiver [131], Van Etten [114] proposed MLSE for joint maximum likelihood sequence estimation of co-channel signals. This section derives the optimum and suboptimum MLSE receivers for co-channel demodulation of digital signals corrupted by ISI. By modeling the overall system as a discrete-time multiple-input multiple-output (MIMO) channel, the optimum MIMO joint MLSE (J-MLSE) receiver is derived. By following a parallel argument used for single-input single-output (SISO) channels, a $T/2$ -spaced MIMO J-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 J-MLSE receiver requires complete knowledge of the complex channel gains associated with all co-channel signals. In many cases, such complete knowledge is impractical or even infeasible to obtain. For such cases, an interference rejection combining MLSE (IRC-MLSE) receiver is proposed that only requires knowledge of the complex channel gain for the desired signal but not the co-channel interferers.

7.7.1 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 $K \times J$ multiple-input multiple-output (MIMO) channel, where the MIMO channel inputs are the symbol sequences from the K co-channel users and the MIMO 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 co-channel detection problem is mathematically very similar to CDMA multiuser detection problem. However, while the users in a CDMA system are distinguished by different spreading sequences as discussed in Chap. 9, the K co-channel users will all use the same transmitter pulse shaping filter $h_a(t)$. The impulse response of the channel between the k th user and the j th antenna element is denoted by $g^{(k,j)}(t, \tau)$, where the channels are modeled as time-variant linear filters. While the channel introduces frequency and time selectivity into the received signals, it also allows the co-channel signals to be distinguished at the receiver, since the received pulses $h^{(k,j)}(t) = g^{(k,j)}(t, \tau) * h_a(t)$ are all distinct due to the different channel impulse responses. The received pulses $h^{(k,j)}(t)$ are all assumed to have a length of at most $L^{(k,j)}T$, i.e., $h^{(k,j)}(t) = 0$ for $t \leq 0$ and $t \geq (L^{(k,j)} + 1)T$. The received complex envelope at the j th antenna element is

$$\tilde{r}^{(j)}(t) = \sum_{k=1}^K \sum_{\ell} x_{\ell}^{(k)} h^{(k,j)}(t - \ell T - \tau_k) + \tilde{n}^{(j)}(t), \quad (7.231)$$

where τ_k , $0 \leq \tau_k \leq T$ is the random transmission delay due to symbol asynchronous users and $\tilde{n}^{(j)}(t)$ is additive white Gaussian noise (AWGN) assumed to be independent on the different antenna branches.

7.7.2 Joint Maximum Likelihood Co-channel Receiver

The joint maximum likelihood sequence estimation (J-MLSE) receiver processes the total received vector

$$\tilde{\mathbf{r}}(t) = (\tilde{r}^{(1)}(t), \tilde{r}^{(2)}(t), \dots, \tilde{r}^{(J)}(t)) \quad (7.232)$$

to generate an ML estimate of the information sequence

$$\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(K)}), \quad (7.233)$$

where $\mathbf{x}^{(k)} = \{x_n^{(k)}\}$. To derive the structure of the joint ML receiver, the same approach used in Sect. 7.1.1 is followed. Let $\{\varphi_n(t)\}$ denote a complete set of complex orthonormal basis functions that span the entire duration of the observation vector $\tilde{\mathbf{r}}(t)$. Then

$$\tilde{r}^{(j)}(t) = \sum_{n=1}^N \tilde{r}_n^{(j)} \varphi_n(t), \quad (7.234)$$

where

$$\tilde{r}_n^{(j)} = \sum_{k=1}^K \sum_{\ell} x_{\ell}^{(k)} h_{\ell_n}^{(k,j)} + \tilde{n}_n^{(j)} \quad (7.235)$$

and

$$h_{\ell_n}^{(k,j)} = \int_{-\infty}^{\infty} h^{(k,j)}(t - \ell T - \tau_k) \varphi_n^*(t) dt \quad (7.236)$$

$$\tilde{n}_n^{(j)} = \int_{-\infty}^{\infty} \tilde{n}^{(j)}(t) \varphi_n^*(t) dt. \quad (7.237)$$

Define the received vector

$$\tilde{\mathbf{r}} = \text{vec}(\tilde{\mathbf{r}}^{(1)}, \tilde{\mathbf{r}}^{(2)}, \dots, \tilde{\mathbf{r}}^{(J)}), \quad (7.238)$$

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

$$p(\tilde{\mathbf{r}}|\mathbf{x}, \mathbf{h}) = \prod_{n=1}^N \frac{1}{2\pi N_o} \exp \left\{ -\frac{1}{2N_o} \sum_{j=1}^J \left| \tilde{r}_n^{(j)} - \sum_{k=1}^K \sum_{\ell} x_{\ell}^{(k)} h_{\ell_n}^{(k,j)} \right|^2 \right\}, \quad (7.239)$$

where $\mathbf{h} = \{h_{\ell_n}^{(k,j)}\}$.

The optimum receiver chooses \mathbf{x} to maximize $p(\tilde{\mathbf{r}}|\mathbf{x}, \mathbf{h})$ or, equivalently, the metric

$$\mu(\mathbf{x}) = - \sum_{n=1}^N \sum_{j=1}^J \left| \tilde{r}_n^{(j)} - \sum_{k=1}^K \sum_{\ell} x_{\ell}^{(k)} h_{\ell_n}^{(k,j)} \right|^2. \quad (7.240)$$

Since $\sum_{n=1}^N \sum_{j=1}^J |\tilde{r}_n^{(j)}|^2$ is independent of \mathbf{x} , maximizing (7.240) is equivalent to maximizing

$$\mu(\mathbf{x}) = \sum_{j=1}^J \left(2\text{Re} \left\{ \sum_{k=1}^K \sum_{\ell} x_{\ell}^{(k)*} \sum_{n=1}^N \tilde{r}_n^{(j)} h_{\ell_n}^{(k,j)*} \right\} - \sum_{k=1}^K \sum_{k'=1}^K \sum_{\ell} \sum_{\ell'} x_{\ell}^{(k)} x_{\ell'}^{(k')*} \sum_{n=1}^N h_{\ell_n}^{(k,j)} h_{\ell'_n}^{(k',j)*} \right). \quad (7.241)$$

In the limit as the number of observable random variables N approaches infinity, define

$$\begin{aligned} y_\ell^{(k,j)} &\triangleq \lim_{N \rightarrow \infty} \sum_{n=1}^N \tilde{r}_n^{(j)} h_{\ell_n}^{(k,j)*} \\ &= \int_{-\infty}^{\infty} \tilde{r}^{(j)}(t) h^{(k,j)*}(t - \ell T - \tau_k) dt \end{aligned} \quad (7.242)$$

$$\begin{aligned} f_{\ell' - \ell}^{(k,k',j)} &\triangleq \lim_{N \rightarrow \infty} \sum_{n=1}^N h_{\ell_n}^{(k,j)} h_{\ell'_n}^{(k',j)*} \\ &= \int_{-\infty}^{\infty} h^{(k,j)}(t - \ell T - \tau_k) h^{(k',j)*}(t - \ell' T - \tau_{k'}) dt. \end{aligned} \quad (7.243)$$

Using (7.242) and (7.243) in (7.241) yields the final form

$$\mu(\mathbf{x}) = \sum_{j=1}^J \left\{ 2\text{Re} \left(\sum_{k=1}^K \sum_{\ell} x_\ell^{(k)*} y_\ell^{(k,j)} \right) - \sum_{k=1}^K \sum_{k'=1}^K \sum_{\ell} \sum_{\ell'} x_\ell^{(k)} x_{\ell'}^{(k')*} f_{\ell' - \ell}^{(k,k',j)} \right\}. \quad (7.244)$$

The variables $\{y_\ell^{(k,j)}\}$ in (7.242) are obtained by passing the received complex envelope $\tilde{r}_n^{(j)}(t)$ through the matched filter $h^{(k,j)*}(-t - \tau_k)$ and sampling the output.⁷ Note that there are K such matched filters on each of the J receiver antenna elements, and the samples must be taken with the timing delay $\tau_k, k = 1, \dots, K$. The above development leads to the overall system model shown in Fig. 7.17. Finally, by changing the variable of integration, the $\{f_{\ell' - \ell}^{(k,k',j)}\}$ in (7.243) can be rewritten in the form

$$f_n^{(k,k',j)} = \int_{-\infty}^{\infty} h^{(k,j)}(t - nT - \tau_k) h^{(k',j)*}(t - \tau_{k'}) dt. \quad (7.245)$$

where $n = \ell' - \ell$. By using (7.245), it can be shown that the ISI coefficients have the symmetric property

$$f_n^{(k,k',j)} = f_{-n}^{(k',k,j)*}. \quad (7.246)$$

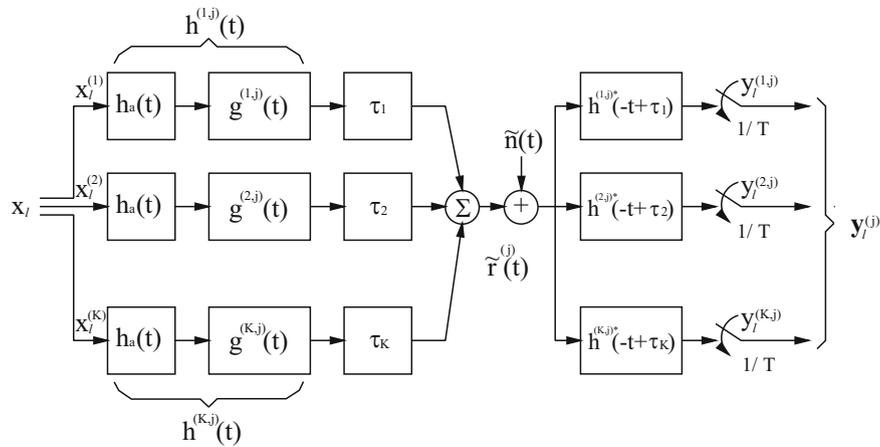


Fig. 7.17 Overall MIMO system model for co-channel demodulation

⁷It is assumed that the differential delay across the multiple receiver antenna elements is negligible.

7.7.3 Discrete-Time MIMO Channel Model

Sampling the outputs of the matched filter $h^{(k,j)*}(-t - \tau_k)$ in Fig. 7.17 every T seconds yields the sample sequence $\{y_\ell^{(k,j)}\}$, where

$$\begin{aligned} y_\ell^{(k,j)} &= \sum_{k'=1}^K \sum_n x_n^{(k')} f_{\ell-n}^{(k',k,j)} + v_\ell^{(k,j)} \\ &= \sum_{k'=1}^K \sum_{m=-L_j-1}^{L_j+1} x_{\ell-m}^{(k')} f_m^{(k',k,j)} + v_\ell^{(k,j)} \end{aligned} \quad (7.247)$$

and

$$v_\ell^{(k,j)} = \int_{-\infty}^{\infty} \tilde{n}^{(j)}(t) h^{(k,j)*}(t - \ell T - \tau_k) dt, \quad (7.248)$$

and where $L_j = \max_k L^{(k,j)}$ is the maximum discrete-time channel length. This is an extension of the result in (7.17) for SISO ISI channels. However, for MIMO ISI channels, the range of summation must be expanded from $(-L_j, L_j)$ to $(-L_j-1, L_j+1)$ to account for the random user delays in the case of symbol asynchronous users. By defining

$$\mathbf{y}_\ell^{(j)} \triangleq (y_\ell^{(1,j)}, \dots, y_\ell^{(K,j)})^T \quad (7.249)$$

$$\mathbf{x}_\ell \triangleq (x_\ell^{(1)}, \dots, x_\ell^{(K)})^T \quad (7.250)$$

$$\mathbf{v}_\ell^{(j)} \triangleq (v_\ell^{(1,j)}, \dots, v_\ell^{(K,j)})^T \quad (7.251)$$

$$\mathbf{F}_m^{(j)} \triangleq [f_m^{(k',k,j)}]_{K \times K} \quad (7.252)$$

(7.247) leads to the convenient matrix representation

$$\mathbf{y}_\ell^{(j)} = \sum_{m=-L_j-1}^{L_j+1} \mathbf{F}_m^{(j)} \mathbf{x}_{\ell-m} + \mathbf{v}_\ell^{(j)}. \quad (7.253)$$

From (7.244) and (7.253), it follows that the joint maximum likelihood receiver uses the entire set of J observation vectors $\mathbf{y} = \{\mathbf{y}_\ell^{(j)}, j = 1, \dots, J\}$ along with knowledge of the entire set of ISI coefficient matrices $\mathbf{F} = \{\mathbf{F}_m^{(j)}, m = -L_j-1, \dots, L_j+1, j = 1, \dots, J\}$ to determine the most likely transmitted sequence $\mathbf{x} = \{\mathbf{x}_\ell\}$. Hence, the joint maximum likelihood receiver decides in favor of the symbol sequence \mathbf{x} that maximizes the joint likelihood function $p(\mathbf{y}|\mathbf{x}, \mathbf{F})$ or the log-likelihood function $\log\{p(\mathbf{y}|\mathbf{x}, \mathbf{F})\}$.

The noise samples in (7.248) are zero-mean complex Gaussian variables that are correlated by the matched filters $h^{(k,j)*}(-t - \tau_k), k = 1, \dots, K$. The vector of zero-mean noise samples $\mathbf{v}_\ell^{(j)}$ has the covariance matrix

$$\Phi_{\mathbf{v}^{(j)}\mathbf{v}^{(j)}}(n) = \frac{1}{2} \mathbb{E} [\mathbf{v}_\ell^{(j)} \mathbf{v}_{\ell+n}^{(j)H}] = N_o \mathbf{F}_n^{(j)}. \quad (7.254)$$

As is the case with SISO ISI channels, the correlation of the noise samples complicates the implementation of various equalization schemes. However, a *matrix* noise whitening filter can be used to whiten the vector of noise samples $\mathbf{v}_\ell^{(j)}$. The z -transform of the noise covariance matrix for the j th antenna element can be defined as $\Phi_{\mathbf{v}^{(j)}\mathbf{v}^{(j)}}(z) = N_o \mathbf{F}^{(j)}(z)$, where [102]

$$\mathbf{F}^{(j)}(z) = \sum_{n=-L_j-1}^{L_j+1} \mathbf{F}_n^{(j)} z^{-n} \quad (7.255)$$

and, once again, $L_j = \max_k L^{(k,j)}$ is the maximum discrete-time channel length. This is a straightforward extension of the SISO ISI channels where $F(z) = \sum_{n=-L}^L f_n z^{-n}$. However, a MIMO channel is described by a covariance matrix and the range of summation in (7.255) must be expanded from $(-L_j, L_j)$ to $(-L_j - 1, L_j + 1)$ to account for the random user delays in the case of symbol asynchronous users.

By using the symmetric property $f_n^{(k,m,j)} = f_{-n}^{(m,k,j)*}$ of the ISI coefficients, it follows that $\mathbf{F}_n^{(j)} = \mathbf{F}_{-n}^{(j)H}$. Therefore, $\mathbf{F}^{(j)}(z)$ has the symmetric form

$$\mathbf{F}^{(j)}(z) = \mathbf{F}^{(j)H}(1/z^*). \quad (7.256)$$

It follows that the matrix $\mathbf{F}^{(j)}(z)$ can be factored as

$$\mathbf{F}^{(j)}(z) = \mathbf{G}^{(j)}(z)\mathbf{G}^{(j)H}(1/z^*) \quad (7.257)$$

and the filter $[\mathbf{G}^{(j)H}(1/z^*)]^{-1}$ will serve as a noise whitening matrix filter.

Example 7.10. Consider a two user system with a single receiver antenna. Since $J = 1$, the index (j) can be omitted. 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

$$\begin{aligned} \mathbf{F}(z) &= \mathbf{G}^H(1/z^*)\mathbf{G}(z) \\ &= [\mathbf{G}_0 + \mathbf{G}_1 z^*]^H [\mathbf{G}_0 + \mathbf{G}_1 z^{-1}]. \end{aligned}$$

\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 non-causal and stable with an infinite length.

$$\begin{aligned} [\mathbf{G}^H(1/z^*)]^{-1} &= \begin{bmatrix} 0.8 & 0.6 \\ 0.6z & 0.8 \end{bmatrix}^{-1} \\ &= \frac{1}{0.64 - 0.36z} \begin{bmatrix} 0.8 & -0.6 \\ -0.6z & 0.8 \end{bmatrix}. \end{aligned}$$

(continued)

Example 7.10 (continued)

In practice the filter $[\mathbf{G}^H(1/z^*)]^{-1}$ can be approximated with a finite length filter by using a sufficient filter delay. Finally, the overall discrete-time white noise matrix channel has transfer function

$$\mathbf{G}(z) = \begin{bmatrix} 0.8 & 0 \\ 0.6 & 0.8 \end{bmatrix} + \begin{bmatrix} 0 & 0.6 \\ 0 & 0 \end{bmatrix} z^{-1}.$$

Let $\mathbf{x}(z) = (x^{(1)}(z), x^{(2)}(z), \dots, x^{(K)}(z))^T$ be the z-transform of the input sequence $\{\mathbf{x}_n\}$ to the channel, where $\mathbf{x}_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(K)})^T$, and let $\mathbf{v}^{(j)}(z) = (v^{(1,j)}(z), v^{(2,j)}(z), \dots, v^{(K,j)}(z))^T$ be the z-transform of the output sequence $\{\mathbf{v}_n^{(j)}\}$ from the noise whitening filter on antenna j , where $\mathbf{v}_n^{(j)} = (v_n^{(1,j)}, v_n^{(2,j)}, \dots, v_n^{(K,j)})^T$. Then

$$\mathbf{v}^{(j)}(z) = \mathbf{G}^{(j)}(z)\mathbf{x}(z) + \mathbf{w}^{(j)}(z), \quad (7.258)$$

where $\mathbf{w}^{(j)}(z) = (w^{(1,j)}(z), w^{(2,j)}(z), \dots, w^{(K,j)}(z))^T$ is the z-transform of the sequence of white Gaussian noise vectors $\mathbf{w}_n^{(j)} = (w_n^{(1,j)}, w_n^{(2,j)}, \dots, w_n^{(K,j)})^T$. In the time domain

$$\mathbf{v}_n^{(j)} = \sum_{k=0}^{L_j} \mathbf{G}_k^{(j)} \mathbf{x}_{n-k} + \mathbf{w}_n^{(j)}, \quad (7.259)$$

where the zero-mean noise vector $\mathbf{w}_n^{(j)}$ has covariance matrix

$$\Phi_{\mathbf{w}_n^{(j)} \mathbf{w}_n^{(j)}}(n) = \frac{1}{2} \mathbf{E} [\tilde{\mathbf{w}}_\ell^{(j)} \tilde{\mathbf{w}}_{\ell+n}^{(j)H}] = N_o \delta_n \mathbf{I}. \quad (7.260)$$

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. 7.18. To determine the number of states in the overall channel model, first define $L_k = \max_j L^{(k,j)}$ as the length of the channel for the k th input. Then there are $2^{n \sum_{k=1}^K L_k}$ states, where 2^n is the size of the signal constellation.

7.7.4 The Viterbi Algorithm

Suppose that m symbols from each of the K transmitters have been transmitted over the channel. Let

$$\mathbf{V}_n = \text{vec}(\mathbf{v}_n^{(1)T}, \mathbf{v}_n^{(2)T}, \dots, \mathbf{v}_n^{(J)T})^T,$$

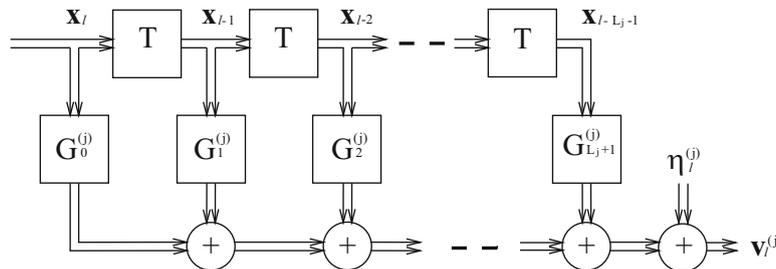


Fig. 7.18 Discrete-time white noise MIMO channel model

where $\mathbf{v}_n^{(j)} = (v_n^{(1,j)}, \dots, v_n^{(K,j)})^T$ denotes the collection of vectors at the outputs of the matrix noise whitening filters for 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

$$\begin{aligned} & \log\{p(\mathbf{V}_m, \dots, \mathbf{V}_1 | \mathbf{x}_m, \dots, \mathbf{x}_1)\} \\ &= \log\{p(\mathbf{V}_m | x_m^{(1)}, \dots, x_{m-L_1}^{(1)}; x_m^{(2)}, \dots, x_{m-L_2}^{(2)}; \dots; x_m^{(K)}, \dots, x_{m-L_K}^{(K)})\} + \log\{p(\mathbf{V}_{m-1}, \dots, \mathbf{V}_1 | \mathbf{x}_{m-1}, \dots, \mathbf{x}_1)\}. \end{aligned} \quad (7.261)$$

The first term on the right-hand side of (7.261) is the *branch metric* used in the Viterbi algorithm. The discrete-time white noise matrix channel model leads to the conditional density function

$$\begin{aligned} & \log\{p(\mathbf{V}_m | x_m^{(1)}, \dots, x_{m-L_1}^{(1)}; x_m^{(2)}, \dots, x_{m-L_2}^{(2)}; \dots; x_m^{(K)}, \dots, x_{m-L_K}^{(K)})\} \\ &= \frac{1}{(2\pi N_o)^{KJ}} \exp \left\{ -\frac{1}{2N_o} \sum_{j=1}^J \|\mathbf{v}_m^{(j)} - \sum_{n=0}^L \mathbf{G}_n^{(j)} \mathbf{x}_{m-n}\|^2 \right\}, \end{aligned} \quad (7.262)$$

where $L = \max_k L_k$. Note that some elements in the matrix $\mathbf{G}_n^{(j)}$ may be zero if $L_k \neq L$, $k = 1, \dots, K$ in which case the branch metric computation can be somewhat simplified. The density in (7.262) leads to the branch metric

$$\mu_m = -\sum_{j=1}^J \|\mathbf{v}_m^{(j)} - \sum_{n=0}^L \mathbf{G}_n^{(j)} \mathbf{x}_{m-n}\|^2. \quad (7.263)$$

7.7.5 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

$$P[\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x}) | \mathbf{x}] = P[\Gamma(\hat{\mathbf{x}}) > \Gamma(\mathbf{x})] \quad (7.264)$$

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 (7.263). From (7.263),

$$\begin{aligned} P[\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x}) | \mathbf{x}] &= P \left[\sum_m \sum_{j=1}^J \left\| \sum_{n=0}^L \mathbf{G}_n^{(j)} \boldsymbol{\epsilon}_{m-n} + \mathbf{w}_m^{(j)} \right\|^2 < \sum_m \sum_{j=1}^J \|\mathbf{w}_m^{(j)}\|^2 \right] \\ &= P \left[\sum_m \sum_{j=1}^J 2\text{Re} \left\{ \sum_{n=0}^L \mathbf{w}_m^{(j)H} \mathbf{G}_n^{(j)} \boldsymbol{\epsilon}_{m-n} \right\} \right. \\ &\quad \left. < -\sum_m \sum_{j=1}^J \left\| \sum_{n=0}^L \mathbf{G}_n^{(j)} \boldsymbol{\epsilon}_{m-n} \right\|^2 \right]. \end{aligned} \quad (7.265)$$

Define

$$\Delta^2 \triangleq \sum_m \sum_{j=1}^J \left\| \sum_{n=0}^L \mathbf{G}_n^{(j)} \boldsymbol{\epsilon}_{m-n} \right\|^2 \quad (7.266)$$

$$\chi \triangleq \sum_m \sum_{j=1}^J 2\text{Re} \left\{ \sum_{n=0}^L \mathbf{w}_m^{(j)H} \mathbf{G}_n^{(j)} \boldsymbol{\epsilon}_{m-n} \right\} \quad (7.267)$$

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

$$P[\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x}] = Q\left(\sqrt{\frac{\Delta^2}{4N_o}}\right). \quad (7.268)$$

7.7.6 $T/2$ -Spaced MIMO J-MLSE Receiver

Suppose that the matched filter outputs $h^{(k,j)*}(-t - \tau_k)$, $k = 1, \dots, K$ are sampled every $T/2$ seconds at the correct timing phase. Again, the zero-mean noise samples at the output of the matched filters are correlated due to the matched filtering and their covariance matrix is

$$\tilde{\Phi}_{\tilde{\mathbf{v}}^{(j)}\tilde{\mathbf{v}}^{(j)}}(n) = \frac{1}{2}\mathbb{E}\left[\tilde{\mathbf{v}}_\ell^{(j)}\tilde{\mathbf{v}}_{\ell+n}^{(j)H}\right] = N_o\tilde{\mathbf{F}}_n^{(j)}, \quad (7.269)$$

where the elements of the matrix $\tilde{\mathbf{F}}_n^{(j)}$ are $\tilde{f}_n^{(k,k',j)} = \tilde{f}^{(k,k',j)}(nT/2)$, the function $f^{(k,k',j)}(t)$ is defined in (7.245), and where the tilde denotes rate $2/T$ sampling. It follows that the z -transform of the noise covariance matrix for the j th antenna element can be defined as $\tilde{\Phi}_{\tilde{\mathbf{v}}^{(j)}\tilde{\mathbf{v}}^{(j)}}(z) = N_o\tilde{\mathbf{F}}^{(j)}(z)$, where

$$\tilde{\mathbf{F}}^{(j)}(z) = \sum_{n=-2L_j-1}^{2L_j+1} \tilde{\mathbf{F}}_n^{(j)}z^{-n}. \quad (7.270)$$

Since $\tilde{f}_\ell^{(k,k',j)} = \tilde{f}_{-\ell}^{(k',k,j)*}$, it follows that $\tilde{\mathbf{F}}^{(j)}(z)$ has the factorization

$$\tilde{\mathbf{F}}^{(j)}(z) = \tilde{\mathbf{G}}^{(j)}(z)\tilde{\mathbf{G}}^{(j)H}(1/z^*). \quad (7.271)$$

As with baud-rate sampling, the $T/2$ -spaced noise samples can be whitened by using a stable noncausal matrix noise whitening filter with the transfer function $[\tilde{\mathbf{G}}^{(j)H}(1/z^*)]^{-1}$. Analogous to (7.258), the z -transform of the output of the matrix noise whitening filter is

$$\tilde{\mathbf{v}}^{(j)}(z) = \tilde{\mathbf{G}}^{(j)}(z)\tilde{\mathbf{x}}(z) + \tilde{\mathbf{w}}^{(j)}(z), \quad (7.272)$$

or in the time domain

$$\tilde{\mathbf{v}}_\ell^{(j)} = \sum_{n=0}^{2L_j} \tilde{\mathbf{G}}_n^{(j)}\tilde{\mathbf{x}}_{\ell-n} + \tilde{\mathbf{w}}_\ell^{(j)}, \quad (7.273)$$

where $\{\tilde{\mathbf{w}}_\ell^{(j)}\}$ is a $T/2$ -spaced white noise sequence with power spectrum $S_{\tilde{\mathbf{w}}\tilde{\mathbf{w}}}(f) = N_o\mathbf{I}$. The sequence $\{\tilde{\mathbf{x}}_n\}$ is the corresponding $T/2$ -spaced input symbol sequence and is given by

$$\tilde{\mathbf{x}}_n = \begin{cases} \mathbf{x}_{n/2} & , n = 0, 2, 4, \dots \\ 0 & , n = 1, 3, 5, \dots \end{cases} \quad (7.274)$$

The overall system and equivalent discrete-time white noise models are shown in Figs. 7.19 and 7.20, respectively.

Note that the vector samples $\tilde{\mathbf{v}}_{2\ell}^{(j)}$ and $\tilde{\mathbf{v}}_{2\ell+1}^{(j)}$ correspond to the ℓ th received baud, such that

$$\tilde{\mathbf{v}}_{2\ell}^{(j)} = \sum_{n=0}^{L_j} \tilde{\mathbf{G}}_{2n}^{(j)}\mathbf{x}_{\ell-n} + \tilde{\mathbf{w}}_{2\ell}^{(j)} \quad (7.275)$$

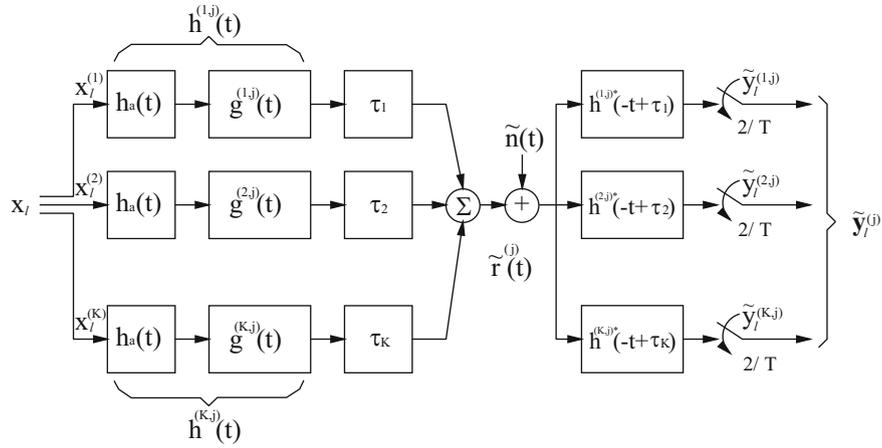


Fig. 7.19 Overall MIMO system model with $T/2$ -spaced sampling

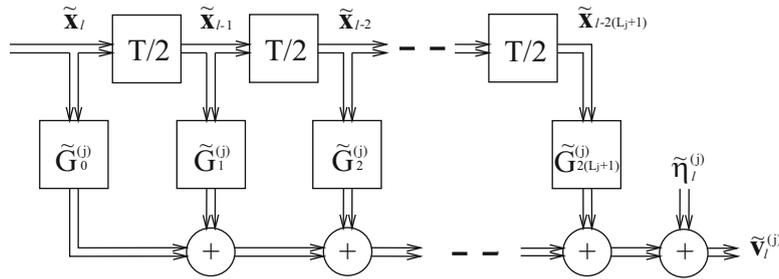


Fig. 7.20 Overall discrete-time white noise MIMO channel model with $T/2$ -spaced sampling

$$\tilde{\mathbf{v}}_{2\ell+1}^{(j)} = \sum_{n=0}^{L_j-1} \tilde{\mathbf{G}}_{2n+1}^{(j)} \mathbf{x}_{\ell-n} + \tilde{\mathbf{w}}_{2\ell+1}^{(j)}. \quad (7.276)$$

With $T/2$ -spaced fractional sampling there are two samples per baud, and the branch metric becomes

$$\mu_m = - \sum_{j=1}^J \left(\|\tilde{\mathbf{v}}_{2m}^{(j)} - \sum_{n=0}^L \tilde{\mathbf{G}}_{2n}^{(j)} \mathbf{x}_{m-n}\|^2 + \|\tilde{\mathbf{v}}_{2m+1}^{(j)} - \sum_{n=0}^{L-1} \tilde{\mathbf{G}}_{2n+1}^{(j)} \mathbf{x}_{m-n}\|^2 \right). \quad (7.277)$$

Once again, if $L_k \neq L, k = 1, \dots, K$, then some of the $\tilde{\mathbf{G}}_{2n}^{(j)}$ and $\tilde{\mathbf{G}}_{2n+1}^{(j)}$ 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.

7.7.6.1 Error Probability

The result for SISO channels is now generalized to show that the T -spaced and $T/2$ -spaced MIMO J-MLSE receivers for co-channel demodulation have identical performance. For T -spaced sampling, define

$$\mathbf{E}(z) \triangleq \sum_n \boldsymbol{\epsilon}_n z^{-n} \quad (7.278)$$

$$\mathbf{E}_v^{(j)}(z) \triangleq \sum_n \mathbf{v}_n^{(j)} z^{-n}. \quad (7.279)$$

Then

$$\mathbf{E}_v^{(j)}(z) = \mathbf{E}(z) \mathbf{G}^{(j)}(z) \quad (7.280)$$

and

$$\Delta^2 = \sum_m \sum_{j=1}^J \left\| \sum_{n=0}^L \mathbf{G}_n^{(j)} \epsilon_{n-m} \right\|^2 \quad (7.281)$$

$$\begin{aligned} &= \sum_{j=1}^J \left[\mathbf{E}_v^{(j)}(z) \mathbf{E}_v^{(j)H}(1/z^*) \right]_0 \\ &= \sum_{j=1}^J \left[\mathbf{E}(z) \mathbf{G}^{(j)}(z) \mathbf{G}^{(j)H}(1/z^*) \mathbf{E}^*(1/z^*) \right]_0 \\ &= \sum_{j=1}^J \left[\mathbf{E}(z) \mathbf{F}^{(j)}(z) \mathbf{E}^*(1/z^*) \right]_0. \end{aligned} \quad (7.282)$$

For $T/2$ -spaced sampling, define

$$\tilde{\mathbf{E}}(z) \triangleq \sum_n \tilde{\epsilon}_n z^{-2n}. \quad (7.283)$$

Since $\tilde{\epsilon}_n = \mathbf{x}_n - \hat{\mathbf{x}}_n$ is zero for even k , it follows that $\tilde{\mathbf{E}}(z) = \mathbf{E}(z^2)$. Also,

$$\tilde{\mathbf{E}}_v^{(j)}(z) = \tilde{\mathbf{E}}(z) \tilde{\mathbf{G}}^{(j)}(z). \quad (7.284)$$

Therefore,

$$\tilde{\Delta}^2 = \sum_m \sum_{j=1}^J \left\| \sum_{n=0}^L \tilde{\mathbf{G}}_n^{(j)} \tilde{\epsilon}_{n-m} \right\|^2 \quad (7.285)$$

$$\begin{aligned} &= \sum_{j=1}^J \left[\tilde{\mathbf{E}}_v^{(j)}(z) \tilde{\mathbf{E}}_v^{(j)H}(1/z^*) \right]_0 \\ &= \sum_{j=1}^J \left[\tilde{\mathbf{E}}(z) \mathbf{G}^{(j)}(z) \mathbf{G}^{(j)H}(1/z^*) \tilde{\mathbf{E}}^*(1/z^*) \right]_0 \\ &= \sum_{j=1}^J \left[\tilde{\mathbf{E}}(z) \mathbf{F}^{(j)}(z) \tilde{\mathbf{E}}^*(1/z^*) \right]_0 \\ &= \sum_{j=1}^J \left[\mathbf{E}(z^2) \mathbf{F}^{(j)}(z) \mathbf{E}^*(1/z^{*2}) \right]_0, \end{aligned} \quad (7.286)$$

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 $\mathbf{F}_\ell = \tilde{\mathbf{F}}_{2\ell}$ it follows that $\tilde{\Delta}^2 = \Delta^2$. Therefore, the T -spaced and $T/2$ -spaced receivers have identical bit error probability performance.

7.7.6.2 Timing Phase Sensitivity

The T -spaced MIMO J-MLSE receiver is sensitive to sampler timing phase due to aliasing, similar to the SISO MLSE receiver. As a result the T -spaced receiver must accurately estimate the delays $\tau_k, k = 1, \dots, K$. A $T/2$ -spaced MIMO J-MLSE receiver will avoid aliasing (assuming that the received pulse exhibits less than 100% excess bandwidth) and is insensitive to sampler timing phase as is now shown. Suppose that the timing phase offset for the k th sampler and the j th antenna branch is $t^{(k,j)}$ seconds. The $T/2$ -spaced sampled pulse $f^{(k,j)}(t)$ at the output of the matched filter $h^{(k,j)*}(-t - \tau_k)$

is denoted as $\tilde{f}_{\ell,t}^{(k,k',j)} = \tilde{f}^{(k,k',j)}(\ell T' \pm t^{(k,j)})$, 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

$$\tilde{\mathbf{F}}_n^{(j)} \triangleq \left[\tilde{f}_n^{(k,k',j)} \right]_{K \times K} \quad (7.287)$$

$$\tilde{\mathbf{F}}_{n,t}^{(j)} \triangleq \left[\tilde{f}_{n,t}^{(k,k',j)} \right]_{K \times K}. \quad (7.288)$$

The discrete-time Fourier transform (DTFT) of $\tilde{\mathbf{F}}_{n,t}^{(j)}$ is

$$\begin{aligned} \tilde{\mathbf{F}}_t^{(j)}(e^{j\omega}) &= \sum_{n=-2L_j-1}^{2L_j+1} \tilde{\mathbf{F}}_{n,t}^{(j)}(e^{j\omega})^{-n} \\ &= e^{\pm j\omega\tau^{(j)}} \tilde{\mathbf{F}}^{(j)}(e^{j\omega}), \end{aligned} \quad (7.289)$$

where $e^{\pm j\omega\tau^{(j)}} = (e^{\pm j\omega\tau^{(1,j)}}, \dots, e^{\pm j\omega\tau^{(K,j)}})$ and $\tau^{(k,j)} = t^{(k,j)}/T'$.

Since the noise is circularly symmetric, the psd of the noise at the output of the j th matched filter is independent of the timing offset vector $\mathbf{t}^{(j)} = (t^{(1,j)}, \dots, t^{(K,j)})$ and is given by

$$S_{\tilde{\mathbf{w}}^{(j)}\tilde{\mathbf{w}}^{(j)}}(f) = N_o \tilde{\mathbf{F}}^{(j)}(e^{j\omega}) \quad (7.290)$$

The DTFT of the noise whitening matrix filter is

$$\left[\tilde{\mathbf{G}}^{(j)H}(1/z^*) \right]_{z=e^{j\omega}}^{-1} = \left[\tilde{\mathbf{G}}^{(j)H}(e^{j\omega}) \right]^{-1} \quad (7.291)$$

and

$$\tilde{\mathbf{F}}^{(j)}(e^{j\omega}) = \tilde{\mathbf{G}}^{(j)}(e^{j\omega})\tilde{\mathbf{G}}^{(j)H}(e^{j\omega}) = |\tilde{\mathbf{G}}^{(j)H}(e^{j\omega})|^2. \quad (7.292)$$

Hence, the noise at the output of the matrix noise whitening filter is white. The DTFT of the message vector at the output of the noise whitening filter is

$$\tilde{\mathbf{x}}(e^{j\omega})\tilde{\mathbf{G}}_t^{(j)}(e^{j\omega}) = \tilde{\mathbf{x}}(e^{j\omega})e^{\pm j\omega\tau^{(j)}}\tilde{\mathbf{G}}^{(j)}(e^{j\omega}) \quad (7.293)$$

and

$$\sum_n \tilde{\mathbf{G}}_n^{(j)}\tilde{\mathbf{G}}_n^{(j)H} = \sum_n \tilde{\mathbf{G}}_{n,t}^{(j)}\tilde{\mathbf{G}}_{n,t}^{(j)H} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\tilde{\mathbf{G}}^{(j)}(e^{j\omega})|^2 d\omega. \quad (7.294)$$

by Parseval's theorem. Also

$$\tilde{\mathbf{G}}_t^{(j)}(z)\tilde{\mathbf{G}}_t^{(j)H}(1/z^*) = \tilde{\mathbf{F}}^{(j)}(z). \quad (7.295)$$

It follows that the distances between allowed sequences of channel outputs in (7.286) remain the same and the $T/2$ -spaced J-MLSE receiver is not sensitive to the sampler phase $e^{\pm j\omega\tau^{(j)}}$. The J-MLSE receiver just needs to estimate the $\{\tilde{\mathbf{G}}_{n,t}^{(j)}\}$ rather than the $\{\tilde{\mathbf{G}}_n^{(j)}\}$ to calculate the branch metrics in the Viterbi algorithm. Finally, (7.294) does not hold for the T -spaced receiver due to aliasing of the signal spectrum.

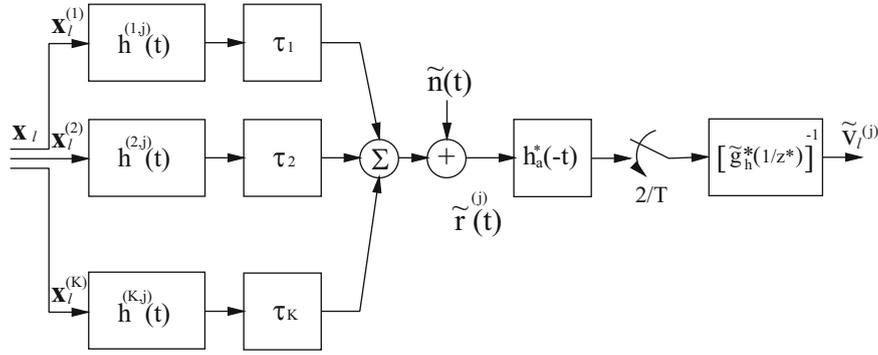


Fig. 7.21 Practical MIMO system with $T/2$ -spaced sampling

7.7.6.3 Practical Receiver

Section 7.6 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 noise whitening filter. Here this concept is generalized to MIMO ISI channels. For a MIMO system where all co-channel waveforms are transmitted with the same pulse shape $h_a(t)$, a significant complexity reduction is realized by using this receiver since a matched filter bank is no longer required at each antenna element. As shown in Fig. 7.21, the receiver simply consists of a single matched filter for each antenna element followed by a rate- $2/T$ sampler and a noise whitening filter. Although the T -spaced samples at the output of the filter $h_a^*(-t)$ are white since the pulse $p(t) = h_a(t) * h_a(-t)$ satisfies the first Nyquist criterion, 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 only depends on the known filter $h_a^*(-t)$.

The systems in Figs. 7.19 and 7.21 are now shown to 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 root-raised cosine pulse shaping with less than 100% excess bandwidth. Once again, assume that the channels are time-invariant and have impulse responses $g^{(k,j)}(t)$, $k = 1, \dots, K$. Define

$$\mathbf{g}^{(j)}(t) \triangleq (g^{(1,j)}(t), \dots, g^{(K,j)}(t))^T \quad (7.296)$$

$$\mathbf{h}^{(j)}(t) \triangleq (h^{(1,j)}(t), \dots, h^{(K,j)}(t))^T \quad (7.297)$$

where $h^{(k,j)}(t) = h_a(t) * g^{(k,j)}(t)$. Let $\tilde{H}_a(z)$, $\tilde{\mathbf{G}}^{(j)}(z)$, and $\tilde{\mathbf{H}}^{(j)}(z)$ be the z -transforms of the $T/2$ -spaced samples of the waveforms $h_a(t)$, and waveform vectors $\mathbf{g}^{(j)}(t)$ and $\mathbf{h}^{(j)}(t)$, respectively, where the tilde indicates $T/2$ -spaced sampling. 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 $\tilde{F}_h(z) = \tilde{H}_a(z) \tilde{H}_a^*(1/z^*)$. By using the factorization

$$\tilde{F}_h(z) = \tilde{G}_h(z) \tilde{G}_h^*(1/z^*) \quad (7.298)$$

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 $1/\tilde{G}_h^*(1/z^*)$ as shown in Fig. 7.21. 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. 7.21 that includes the noise whitening filter is

$$\begin{aligned} \tilde{\mathbf{G}}_{\text{eq}}^{(j)}(z) &= \tilde{H}_a(z) \tilde{\mathbf{G}}^{(j)}(z) \tilde{H}_a^*(1/z^*) / \tilde{G}_h^*(1/z^*) \\ &= \tilde{\mathbf{G}}^{(j)}(z) \tilde{G}_h(z). \end{aligned} \quad (7.299)$$

On the other hand, referring to the conventional system shown in Fig. 7.19,

$$\tilde{\mathbf{H}}^{(j)}(z) = \tilde{H}_a(z) \tilde{\mathbf{G}}^{(j)}(z) \quad (7.300)$$

and

$$\begin{aligned}\tilde{\mathbf{F}}^{(j)}(z) &= \tilde{H}_a(z)\tilde{\mathbf{G}}^{(j)}(z)(\tilde{H}_a(1/z^*)\mathbf{G}^{(j)}(1/z^*))^H \\ &= \tilde{H}_a(z)\tilde{\mathbf{G}}^{(j)}(z)\tilde{\mathbf{G}}^{(j)H}(1/z^*)\tilde{H}_a^*(1/z^*).\end{aligned}\quad (7.301)$$

Let

$$\tilde{\mathbf{G}}_c^{(j)}(z)\mathbf{G}_c^{(j)H}(1/z^*) = \tilde{\mathbf{G}}^{(j)}(z)\tilde{\mathbf{G}}^{(j)H}(1/z^*) \quad (7.302)$$

be a factorization of the matrix $\tilde{\mathbf{G}}^{(j)}(z)\tilde{\mathbf{G}}^{(j)H}(1/z^*)$ such that $\tilde{\mathbf{G}}_c^{(j)}(z)$ has minimum phase. Combining (7.298), (7.301), and (7.302) gives

$$\tilde{\mathbf{F}}^{(j)}(z) = \tilde{G}_h(z)\tilde{\mathbf{G}}_c^{(j)H}(z)\tilde{\mathbf{G}}_c^{(j)}(1/z^*)\tilde{G}_h^*(1/z^*). \quad (7.303)$$

The transfer function of the matrix noise whitening filter is chosen as

$$\left[\tilde{\mathbf{G}}_c^{(j)H}(1/z^*)\tilde{G}_h^*(1/z^*)\right]^{-1}. \quad (7.304)$$

Therefore, the overall transfer function at the output of the matrix noise whitening filter is

$$\tilde{\mathbf{G}}^{(j)}(z) = \tilde{G}_h(z)\tilde{\mathbf{G}}_c^{(j)}(z). \quad (7.305)$$

Finally,

$$\tilde{\mathbf{G}}_{\text{eq}}^{(j)}(z)\tilde{\mathbf{G}}_{\text{eq}}^{(j)H}(1/z^*) = \tilde{\mathbf{F}}^{(j)}(z) = \tilde{\mathbf{G}}^{(j)}(z)\tilde{\mathbf{G}}^{(j)H}(1/z^*). \quad (7.306)$$

Therefore, the Euclidean distance between sequences of noise whitening filter outputs for the MIMO receiver in Fig. 7.21 is the same as those for the $T/2$ -spaced MIMO J-MLSE receiver in Fig. 7.19. Consequently, the system shown in Fig. 7.21 achieves ML performance. The main advantage of the system in Fig. 7.21 is that the noise whitening filter does not depend on the unknown channel and has a fixed structure.

The receiver shown in Fig. 7.21 has a scalar output, while the receiver in Fig. 7.19 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 (7.277)

$$\mu_m = -\sum_{j=1}^J \left(\left| \tilde{v}_{2m}^{(j)} - \sum_{n=0}^{L_j} \tilde{\mathbf{G}}_{\text{eq},2n}^{(j)T} \mathbf{x}_{m-n} \right|^2 + \left| \tilde{v}_{2m+1}^{(j)} - \sum_{n=0}^{L_j-1} \tilde{\mathbf{G}}_{\text{eq},2n+1}^{(j)T} \mathbf{x}_{m-n} \right|^2 \right). \quad (7.307)$$

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 channel vectors $\{\tilde{\mathbf{G}}_{\text{eq},n}^{(j)}\}$. This synchronization and training problem is particularly challenging for an asynchronous co-channel waveforms where the training sequences are not coincident. Second, the receiver must be able to track the channel vectors $\{\tilde{\mathbf{G}}_{\text{eq},n}^{(j)}\}$ during data demodulation if necessary. Perhaps a per-survivor processing approach such as the one suggested in Sect. 7.4.1.1 could be used for such cases.

7.7.7 Interference Rejection Combining MLSE Receiver

In many cases, the structure of the co-channel interferers are often unknown. In other cases, it is only necessary to recover the data from a single desired co-channel user. Bottomley and Jamal [48] developed a scheme called the interference rejection combining MLSE (IRC-MLSE) that combines adaptive antenna arrays and MLSE equalization. Co-channel interference

cancellation is performed in the Viterbi metric and the receiver is equivalent to Winter's optimum linear combiner [352, 354] under flat fading channel conditions.

Once again, the receiver filter on each antenna element is assumed to be matched to the transmitted pulse and followed by a rate $2/T$ sampler. Since the co-channel waveforms are assumed to have an unknown structure, a matched filter is only required for the desired signal. The overall pulse response consisting of the transmit filter, channel, and receiver filter is $f^{(j)}(t) = h_a(t) * g^{(1,j)}(t) * h_a^*(-t)$. The vector of matched filter outputs from the J receiver antenna elements is

$$\tilde{\mathbf{y}}(t) = \sum_{\ell=0}^L x_{\ell} \mathbf{f}(t - \ell T) + \tilde{\mathbf{z}}(t), \quad (7.308)$$

where

$$\begin{aligned} \tilde{\mathbf{y}}(t) &= (y^{(1)}(t), \dots, y^{(J)}(t))^T \\ \mathbf{f}(t) &= (f^{(1)}(t), \dots, f^{(J)}(t))^T \\ \tilde{\mathbf{z}}(t) &= (z^{(1)}(t), \dots, z^{(J)}(t))^T \end{aligned} \quad (7.309)$$

and where LT is the length of the pulse $f^{(j)}(t)$. The impairment vector $\tilde{\mathbf{z}}(t)$ at the output of the matched filter is due to the $K - 1$ interfering co-channel signals plus AWGN, and has the form

$$\tilde{\mathbf{z}}(t) = \sum_{k=1}^{K-1} \tilde{\mathbf{I}}_k(t) + \tilde{\mathbf{v}}(t), \quad (7.310)$$

where

$$\begin{aligned} \tilde{\mathbf{I}}_k(t) &= (I_k^{(1)}(t), \dots, I_k^{(J)}(t))^T \\ \tilde{\mathbf{v}}(t) &= (v^{(1)}(t), \dots, v^{(J)}(t))^T. \end{aligned} \quad (7.311)$$

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 co-channel interference, since the co-channel interference 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 co-channel interference. 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 (7.299). It follows that the overall channel consisting of the transmit filter, channel, and receiver filter, $T/2$ -spaced sampler, and noise whitening filter, can be modeled as a $T/2$ -spaced tapped delay line with tap coefficients

$$\tilde{\mathbf{g}}^{(j)} = (\tilde{g}_0^{(j)}, \dots, \tilde{g}_{2L}^{(j)})^T.$$

Define

$$\tilde{\mathbf{g}}_{\ell} \triangleq (\tilde{g}_{\ell}^{(1)}, \dots, \tilde{g}_{\ell}^{(J)})^T. \quad (7.312)$$

Then the vectors $\tilde{\mathbf{v}}_{2k} = (\tilde{v}_{2k}^{(1)}, \dots, \tilde{v}_{2k}^{(J)})^T$ and $\tilde{\mathbf{v}}_{2k+1} = (\tilde{v}_{2k+1}^{(1)}, \dots, \tilde{v}_{2k+1}^{(J)})^T$ at the output of the noise whitening filter corresponding to the k th received baud are

$$\tilde{\mathbf{v}}_{2k} = \sum_{\ell=0}^{L_h} x(k - \ell) \tilde{\mathbf{g}}_{2\ell} + \tilde{\mathbf{n}}_{2k} \quad (7.313)$$

$$\tilde{\mathbf{v}}_{2k+1} = \sum_{\ell=0}^{L_h-1} x(k - \ell) \tilde{\mathbf{g}}_{2\ell+1} + \tilde{\mathbf{n}}_{2k+1}, \quad (7.314)$$

where $\tilde{\mathbf{n}}_{2k} = (\tilde{n}_{2k}^{(1)}, \dots, \tilde{n}_{2k}^{(J)})^T$ and $\tilde{\mathbf{n}}_{2k+1} = (\tilde{n}_{2k+1}^{(1)}, \dots, \tilde{n}_{2k+1}^{(J)})^T$. Once again, the noise vectors $\tilde{\mathbf{n}}_{2k}$ and $\tilde{\mathbf{n}}_{2k+1}$ are not white due to the presence of the co-channel signals.

To derive a feasible receiver structure, it is now *assumed* that the sampled impairment vector $\tilde{\mathbf{n}}_\ell$ at the output of the noise whitening filter at epoch ℓ is a vector of J correlated zero-mean complex Gaussian random variables having the joint pdf

$$p(\tilde{\mathbf{n}}_\ell) = \frac{1}{(2\pi)^J |\mathbf{R}_\ell|} \exp \left\{ -\frac{1}{2} \tilde{\mathbf{n}}_\ell^H \mathbf{R}_\ell^{-1} \tilde{\mathbf{n}}_\ell \right\}, \quad (7.315)$$

where $|\mathbf{R}_\ell|$ is the determinant of \mathbf{R}_ℓ and

$$\mathbf{R}_\ell = \frac{1}{2} \mathbb{E} [\tilde{\mathbf{n}}_\ell \tilde{\mathbf{n}}_\ell^H] \quad (7.316)$$

is the impairment correlation matrix. Assuming an MLSE-like algorithm, the branch metric should be related to the likelihood of the impairment vector. For each state transition $\varrho_k^{(j)} \rightarrow \varrho_{k+1}^{(i)}$ 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

$$\begin{aligned} \mu_k &= (\tilde{\mathbf{v}}_{2k} - \hat{\mathbf{u}}_{2k}^1)^H \mathbf{R}_{2k}^{-1} (\tilde{\mathbf{v}}_{2k} - \hat{\mathbf{u}}_{2k}^1) \\ &\quad + (\tilde{\mathbf{v}}_{2k+1} - \hat{\mathbf{u}}_{2k+1}^2)^H \mathbf{R}_{2k+1}^{-1} (\tilde{\mathbf{v}}_{2k+1} - \hat{\mathbf{u}}_{2k+1}^2), \end{aligned} \quad (7.317)$$

where

$$\hat{\mathbf{u}}_{2k}^1 = \sum_{\ell=0}^{L_h} \hat{x}(k-\ell) \tilde{\mathbf{g}}_{2\ell} \quad (7.318)$$

$$\hat{\mathbf{u}}_{2k+1}^2 = \sum_{\ell=0}^{L_h-1} \hat{x}(k-\ell) \tilde{\mathbf{g}}_{2\ell+1}. \quad (7.319)$$

Notice that the metric calculation requires the impairment correlation matrix \mathbf{R}_k and its inverse, and the channel impulse vectors

$$\tilde{\mathbf{g}}^1 = (\tilde{\mathbf{g}}_0, \tilde{\mathbf{g}}_2, \dots, \tilde{\mathbf{g}}_{2L_h})^T \quad (7.320)$$

$$\tilde{\mathbf{g}}^2 = (\tilde{\mathbf{g}}_1, \tilde{\mathbf{g}}_3, \dots, \tilde{\mathbf{g}}_{2L_h-1})^T. \quad (7.321)$$

Computing the inverse of the $J \times J$ matrix \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

$$\mathbf{R}_k^{-1} = \frac{\text{adj}(\mathbf{R}_k)}{|\mathbf{R}_k|} = \frac{1}{r_{k11}r_{k22} - r_{k12}r_{k21}} \begin{bmatrix} r_{k22} & -r_{k12} \\ -r_{k21} & r_{k11} \end{bmatrix}. \quad (7.322)$$

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 will just scale all the path metrics. For such cases, the Viterbi algorithm can use the simplified branch metric

$$\begin{aligned} \mu_k &= (\tilde{\mathbf{v}}_{2k} - \hat{\mathbf{u}}_{2k}^1)^H \text{adj}(\mathbf{R}_{2k}) (\tilde{\mathbf{v}}_{2k} - \hat{\mathbf{u}}_{2k}^1) \\ &\quad + (\tilde{\mathbf{v}}_{2k+1} - \hat{\mathbf{u}}_{2k+1}^2)^H \text{adj}(\mathbf{R}_{2k+1}) (\tilde{\mathbf{v}}_{2k+1} - \hat{\mathbf{u}}_{2k+1}^2), \end{aligned} \quad (7.323)$$

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 additive impairment is white Gaussian noise.

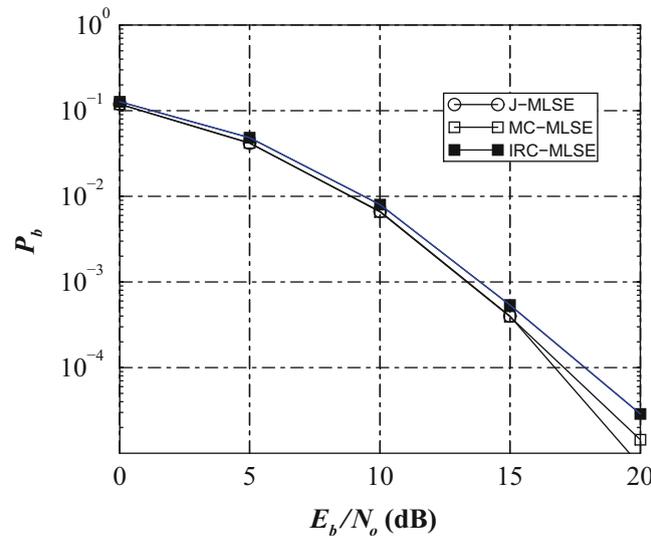


Fig. 7.22 Relative E_b/N_o performance of the J-MLSE, MC-MLSE, and IRC-MLSE receivers; $C/I = 30$ dB

7.7.8 Examples

The performances of the J-MLSE, IRC-MLSE, and MC-MLSE receivers discussed in the previous sections are now compared and contrasted. For this purpose, an EDGE burst format is assumed. The EDGE 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.

Figure 7.22 shows the E_b/N_o performance of the three receivers for a fixed carrier-to-interference ratio $C/I = 30$ dB. Under this condition the interference is very light. Perfect channel knowledge is assumed. 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 gives the worst E_b/N_o performance, although with coding the degradation is a relatively small 1 dB.

Figure 7.23 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.

It is curious that the IRC-MLSE receiver outperforms the J-MLSE receiver. First, the J-MLSE receiver that has been implemented does not have a sufficient number of receiver states due to pulse truncation effects. Hence, there is some residual ISI that is 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 when the signal constellation points overlap. In this case, errors can occur even for large E_b/N_o values.

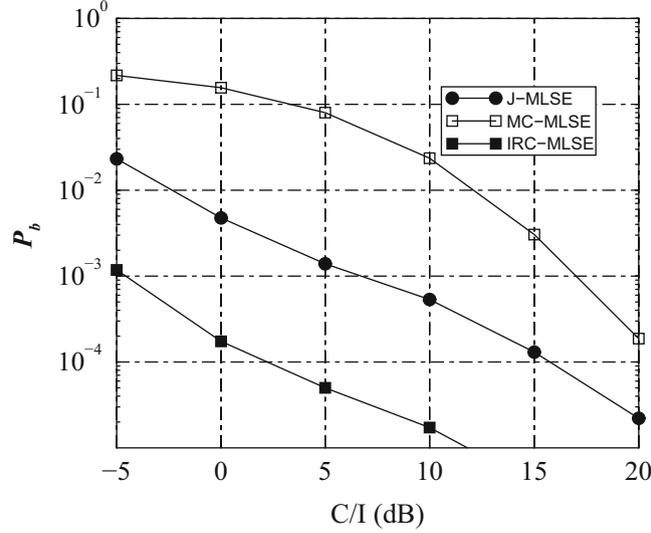


Fig. 7.23 Relative C/I performance of the J-MLSE, MC-MLSE and IRC-MLSE receivers; $E_b/N_o = 30$ dB

Appendix 7A: Derivation of Eq. (7.202)

Consider the case where $D = 1$ and suppose that the characteristic function in (7.194) has $L + 1$ different poles $\bar{\alpha} = (\bar{\alpha}_0, \bar{\alpha}_1, \dots, \bar{\alpha}_L)$. Then the pairwise error probability is equal to

$$P(\bar{\alpha}) = \sum_{i=0}^L \left(\left(\frac{1}{2} - \frac{1}{2} \sqrt{\frac{\bar{\alpha}_i}{1 + \bar{\alpha}_i}} \right) \prod_{j \neq i} \left(1 - \frac{\bar{\alpha}_j}{\bar{\alpha}_i} \right)^{-1} \right). \quad (7A.1)$$

Define the function $\phi(\bar{\alpha}) = \sum_{i=0}^L \bar{\alpha}_i - C = 0$, where C is a constant. The method of Lagrange multipliers suggests that

$$\frac{\partial P(\bar{\alpha})}{\partial \bar{\alpha}_i} + \lambda \frac{\partial \phi}{\partial \bar{\alpha}_i} = 0 \quad i = 1, \dots, L \quad (7A.2)$$

for any real number λ . It can be shown by induction that

$$\begin{aligned} \frac{\partial P(\bar{\alpha})}{\partial \bar{\alpha}_k} = & - \left(\frac{1}{2} - \frac{1}{2} \sqrt{\frac{\bar{\alpha}_k}{1 + \bar{\alpha}_k}} \right) \sum_{i \neq k} \left(\frac{\bar{\alpha}_i}{\bar{\alpha}_k^2} \left(1 - \frac{\bar{\alpha}_i}{\bar{\alpha}_k} \right)^{-2} \prod_{j \neq i, k} \left(1 - \frac{\bar{\alpha}_j}{\bar{\alpha}_k} \right)^{-1} \right) \\ & + \sum_{i \neq k} \left(\frac{1}{\bar{\alpha}_i} \left(1 - \frac{\bar{\alpha}_k}{\bar{\alpha}_i} \right)^{-2} \left(\frac{1}{2} - \frac{1}{2} \sqrt{\frac{\bar{\alpha}_i}{1 + \bar{\alpha}_i}} \right) \prod_{j \neq i, k} \left(1 - \frac{\bar{\alpha}_j}{\bar{\alpha}_i} \right)^{-1} \right) \\ & - \left(\frac{1}{4\bar{\alpha}_k^{1/2} (1 + \bar{\alpha}_k)^{3/2}} \right) \prod_{j \neq k} \left(1 - \frac{\bar{\alpha}_j}{\bar{\alpha}_k} \right)^{-1}. \end{aligned} \quad (7A.3)$$

By solving (7A.2) and observing the symmetry of $P(\bar{\alpha})$ and the derivative (7A.3) with respect to the permutations of $\bar{\alpha}$, it is apparent that the minimum of $P(\bar{\alpha})$ is achieved when $\bar{\alpha}_0 = \bar{\alpha}_1 = \dots = \bar{\alpha}_L$.

Problems

7.1. Starting with

$$f_k = \int_{-\infty}^{\infty} h^*(\tau)h(\tau + kT)d\tau$$

show that

$$F(e^{j2\pi fT}) = F_{\Sigma}(f).$$

7.2. Suppose that the impulse response of an overall channel consisting of the transmit filter, channel, and receive filter is

$$F(f) = \begin{cases} 1 & , |f| \leq f_{\ell} \\ \frac{f_u - |f|}{f_u - f_{\ell}} & , f_{\ell} \leq |f| \leq f_u \end{cases}.$$

- Find the overall impulse response $f(t)$.
- Is it possible to transmit data without ISI?
- How do the magnitudes of the tails of the overall impulse response decay with large values of t ?
- 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?

7.3. Suppose a digital communication system operates over an “ideal channel” and employs an overall pulse $p(t)$ that has the Gaussian-shaped form

$$p(t) = e^{-\pi a^2 t^2}.$$

- Explain why $p(t)$ does not admit intersymbol interference (ISI) free transmission.
- To reduce the level of ISI to a relatively small amount, the condition that $p(T) = 0.01$ is imposed, where T is the symbol interval. The bandwidth W of the pulse $p(t)$ is defined as that value of W for which $P(W)/P(0) = 0.01$, where $P(f)$ is the Fourier transform of $p(t)$. Determine the value W and compare this value to that of a raised cosine spectrum with 100% roll-off.

7.4. Show that the ISI coefficients $\{f_n\}$ may be expressed in terms of the channel vector coefficients $\{g_n\}$ as

$$f_n = \sum_{k=0}^{L-n} g_k^* g_{k+n} \quad n = 0, 1, 2, \dots, L.$$

7.5. Suppose that BPSK is used on a static ISI channel. The complex envelope has the form

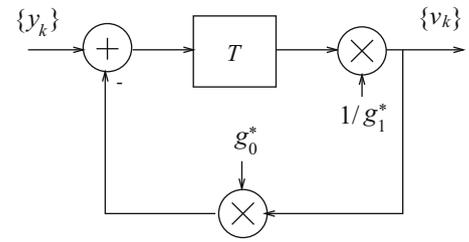
$$\tilde{s}(t) = A \sum_{k=-\infty}^{\infty} x_k h_a(t - kT),$$

where $x_k \in \{-1, +1\}$ and $h_a(t)$ is the amplitude shaping pulse. The full response rectangular 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 $0 < \tau < T$.

Fig. 7.24 Digital filter for Problem 7.8



- (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 ISI. By using mathematical arguments, show that (1) the matched filter tends to accentuate the effect of ISI, and (2) 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. 7.24. The values of g_0 and g_1 are chosen to satisfy

$$\begin{aligned} |g_0|^2 + |g_1|^2 &= f_0 \\ g_0 g_1^* &= f_1. \end{aligned}$$

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 .

7.9. The z -transform of the channel vector \mathbf{g} of a communication system is equal to

$$G(z) = 0.1 + 1.0z^{-1} - 0.1z^{-2}.$$

A binary sequence \mathbf{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

$$\{x_k\} = (-1)^k, \quad k = 0, 1, 2, 3 \text{ ?}$$

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

$$h(t) = \sqrt{2a}e^{-at}, \quad 0 \leq t \leq \infty.$$

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 coefficients 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

$$y_n = x_n f_0 + \sum_{k \neq n} x_k f_{n-k} + v_n,$$

where the source symbols are from the set $\{\pm 1, \pm 3, \dots, \pm(M-1)\}$. Decisions are made on the $\{y_n\}$ without any equalization by using a threshold detector. The ℓ th ISI pattern can be written as

$$D(\ell) = \sum_{k \neq n} x_{\ell,k} f_{n-k}$$

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

$$D_{\max} = \frac{1}{f_0} \sum_{n \neq 0} |f_n|.$$

- (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

$$f_i = 0.0, \quad |i| \geq 3$$

$$f_2 = f_{-2} = 0.1$$

$$f_1 = f_{-1} = -0.2$$

$$f_0 = 1.0.$$

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 \mathbf{c} satisfies

$$\mathbf{c} = \mathbf{c}_{\text{op}} + \mathbf{c}_e$$

where \mathbf{c}_e is the tap gain error vector. Show that the mean square error that is achieved with the tap gain vector \mathbf{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

$$\mathbf{x}_k \mathbf{M}_v = \lambda_k \mathbf{x}_k, \quad k = 1, \dots, N.$$

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 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 ISI in addition to the noise.

7.15. In this question, it will be shown in steps that

$$\nabla_{\mathbf{c}} J = 2\mathbf{c}^T \mathbf{M}_v - 2\mathbf{v}_x^H.$$

Define

$$\begin{aligned} \mathbf{M}_v &\triangleq \mathbf{M}_{v_R} + j\mathbf{M}_{v_I} \\ \mathbf{c} &\triangleq \mathbf{c}_R + j\mathbf{c}_I \\ \mathbf{v}_x &\triangleq \mathbf{v}_{x_R} + j\mathbf{v}_{x_I}. \end{aligned}$$

(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

$$\begin{aligned} \nabla_{\mathbf{c}_R} \text{Re}\{\mathbf{v}_x^H \mathbf{c}^*\} &= \mathbf{v}_{x_R}^T \\ \nabla_{\mathbf{c}_I} \text{Re}\{\mathbf{v}_x^H \mathbf{c}^*\} &= -\mathbf{v}_{x_I}^T \\ \nabla_{\mathbf{c}_R} \mathbf{c}^T \mathbf{M}_v \mathbf{c}^* &= 2\mathbf{c}_R^T \mathbf{M}_{v_R} - 2\mathbf{c}_I^T \mathbf{M}_{v_I} \\ \nabla_{\mathbf{c}_I} \mathbf{c}^T \mathbf{M}_v \mathbf{c}^* &= 2\mathbf{c}_I^T \mathbf{M}_{v_R} + 2\mathbf{c}_R^T \mathbf{M}_{v_I} \end{aligned}$$

where $\nabla_{\mathbf{x}}$ is the gradient with respect to vector \mathbf{x} .

(c) By defining the gradient of a real-valued function with respect to a complex vector \mathbf{c} as

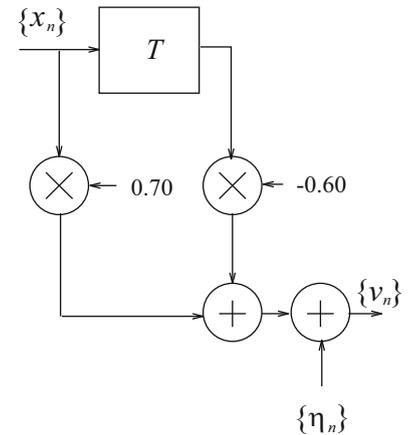
$$\nabla_{\mathbf{c}} = \nabla_{\mathbf{c}_R} + j\nabla_{\mathbf{c}_I}$$

show that

$$\begin{aligned} \nabla_{\mathbf{c}} \text{Re}\{\mathbf{v}_x^H \mathbf{c}^*\} &= \mathbf{v}_x^H \\ \nabla_{\mathbf{c}} \mathbf{c}^T \mathbf{M}_v \mathbf{c}^* &= 2\mathbf{c}^T \mathbf{M}_v. \end{aligned}$$

7.16. Show that the pairwise error probability for digital signaling on an ISI channel is given by (7.171).

Fig. 7.25 Discrete-time white noise channel model for Problem 7.17



7.17. Consider the transmission of the binary sequence \mathbf{x} , $x_n \in \{-1, +1\}$ over the equivalent discrete-time white noise channel model shown in Fig. 7.25. The received sequence is

$$v_0 = 0.70x_0 + \eta_1$$

$$v_1 = 0.70x_1 - 0.60x_0 + \eta_2$$

$$v_2 = 0.70x_2 - 0.60x_1 + \eta_3$$

$$\vdots$$

$$v_k = 0.70x_k - 0.60x_{k-1} + \eta_k.$$

- Draw the state diagram for this system.
- Draw the trellis diagram.
- 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 have been received.

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.

- Find $G(z)$.
- Draw the state diagram and the trellis diagram for the discrete-time white noise channel model.
- 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 $x_i = 0$, $i < 9$ and suppose that the sampled sequence at the output of the noise whitening filter is

$$\begin{aligned} \mathbf{v} &= \{v_0, v_1, v_2, v_3, \dots, v_9\} \\ &= \{1/2, 1/4, -3/4, 3/4, -3/4, -1/4, 3/4, -3/4, -1/4, -1/4\}. \end{aligned}$$

What sequence \mathbf{x} was most likely transmitted?

Chapter 8

Error Control Coding

Channel coding and interleaving techniques have long been used for combating 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 subsequently exploited at the receiver for 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 applications such as voice or real-time video. 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. Both block codes and convolutional codes find potential applications in wireless systems. 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.

In the early application of coding to digital communications, the modulator and coder were treated as 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 approach may be feasible 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 [220] 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) [325] 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 wireless applications where high spectral efficiency is needed due to limited bandwidth resources and good power efficiency is needed to extend battery life in portable devices. 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 [43]. In 1985, a TCM-based modem operating at 19.2 kb/s was introduced by Codex [326].

Ungerboeck's work [325] captured the attention of the coding community and laid the foundation for intensified research. Calderbank and Mazo introduced an analytic description of trellis codes [51]. They showed how to realize the two operations (coding and mapping) in Ungerboeck's codes by using a single-step procedure. Calderbank and Sloane [52] and Wei [346]

proposed multi-dimensional trellis codes. Spaces with larger dimensionality are attractive, because the signals are spaced at larger Euclidean distances [43]. Calderbank and Sloan [52] and Forney [132] 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 [53, 94, 95, 103]. Divsalar and Simon [95, 96] 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 [53] 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 [347] 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 performance of trellis codes on intersymbol interference (ISI) channels [99, 118, 318, 355]. 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 [60]. 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 [42]. When used in conjunction with an iterative decoding scheme, PCCCs achieve near Shannon capacity limit performance on both the AWGN channel and the interleaved flat fading channel. Simulations of a rate-1/2 turbo code in [42] showed 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 capacity 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 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 [39]. In general, SCCC outperform PCCC at high E_b/N_o , whereas the opposite is true for low E_b/N_o .

In the late 1990s, the study of MIMO channels led to the development of space-time coding schemes. These include both space-time block codes and space-time trellis codes. The first space-time trellis codes were introduced by Tarokh et al. [314]. Alamouti introduced the first space-time block code [15], while other orthogonal space-time block codes were developed by Tarokh et al. [315, 316].

The remainder of this chapter is organized as follows. Section 8.1 gives an introduction to block codes and space-time block codes. Section 8.2 introduces convolutional codes, and decoding algorithms for convolutional codes, including the Viterbi algorithm and BCJR algorithm. Section 8.3 introduces trellis-coded modulation. This is followed by the performance analysis of convolutional and trellis codes on AWGN channels in Sect. 8.4. Section 8.5 considers block and convolutional interleavers that are useful for coding on fading channels. This is followed by a consideration of the design and performance analysis of trellis codes on interleaved flat fading channels in Sect. 8.6. The performance of space-time codes and the decoding of space-time codes is considered in Sect. 8.7. Finally, Sect. 8.8 provides a treatment of parallel and serial turbo codes.

8.1 Block Codes

8.1.1 Binary Block Codes

A binary block encoder accepts a length- k input vector $\mathbf{a} = (a_1, a_2, \dots, a_k)$, $a_i \in \{0, 1\}$, and generates a length- n codeword $\mathbf{c} = (c_1, c_2, \dots, c_n)$, $c_i \in \{0, 1\}$, through the linear mapping $\mathbf{c} = \mathbf{a}\mathbf{G}$, where $\mathbf{G} = [g_{ij}]_{k \times n}$ is a $k \times n$ 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 task of designing block codes reduces to one of finding generator matrices that produce 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 $\mathbf{G}\mathbf{H}^T = \mathbf{0}_{k \times (n-k)}$. The matrix \mathbf{H} has full row rank $n-k$ and is orthogonal to all codewords, i.e., $\mathbf{c}\mathbf{H}^T = \mathbf{0}_{n-k}$. The matrix \mathbf{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 generator matrix of the form

$$\mathbf{G} = [\mathbf{I}_{k \times k} | \mathbf{P}], \quad (8.1)$$

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 \mathbf{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

$$\mathbf{H} = [-\mathbf{P}^T | \mathbf{I}_{(n-k) \times (n-k)}]. \quad (8.2)$$

The parity check matrix in (8.2) is a general form that applies to both binary and non-binary systematic block codes. For binary codes the negative sign in front of the \mathbf{P}^T matrix is not necessary. For a binary systematic block code, $\mathbf{G}\mathbf{H}^T = \mathbf{I}_{k \times k} \mathbf{P} \oplus \mathbf{P} \mathbf{I}_{(n-k) \times (n-k)} = \mathbf{P} \oplus \mathbf{P} = \mathbf{0}_{k \times (n-k)}$, where \oplus indicates modulo-2 addition.

Example 8.1. The $(n-k) \times n$ parity check matrix of an (n, k) Hamming code is constructed by listing as columns all non-zero binary $(n-k)$ -tuples. There are $n = 2^{n-k} - 1$ such non-zero $(n-k)$ -tuples. For example, a systematic $(7, 4)$ Hamming code has the parity check matrix

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 1 & \vdots & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & \vdots & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & \vdots & 0 & 0 & 1 \end{bmatrix}, \quad (8.3)$$

where the columns of \mathbf{H} consists of all non-zero binary 3-tuples. Note that the last three columns form a 3×3 identity matrix since the code is in systematic form, while the first four columns that constitute the \mathbf{P}^T matrix may be placed in any random order with no effect on the performance of the code. Such codes are said to be equivalent.

The generator matrix of this particular $(7, 4)$ systematic Hamming code is

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & \vdots & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & \vdots & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & \vdots & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & \vdots & 1 & 0 & 1 \end{bmatrix} \quad (8.4)$$

and it can be verified that $\mathbf{G}\mathbf{H}^T = \mathbf{0}_{k \times (n-k)}$. The 16 codewords of the $(7, 4)$ Hamming code are generated by taking all linear combinations of the rows of \mathbf{G} using modulo-2 arithmetic.

8.1.1.1 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 $d(\mathbf{c}_1, \mathbf{c}_2) = w(\mathbf{c}_1 \oplus \mathbf{c}_2)$, where $w(\mathbf{c}_1 \oplus \mathbf{c}_2)$ is the weight of $\mathbf{c}_1 \oplus \mathbf{c}_2$, equal to the number of non-zero coordinates of $\mathbf{c}_1 \oplus \mathbf{c}_2$. 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 $\mathbf{c}_1 \oplus \mathbf{c}_2$ is another codeword \mathbf{c} and the all zeroes vector is a codeword. Hence, the free Hamming distance is

$$d_{\text{free}} = \min_{\mathbf{c}_1, \mathbf{c}_2} d(\mathbf{c}_1, \mathbf{c}_2) = \min_{\mathbf{c} \neq \mathbf{0}} d(\mathbf{c}, \mathbf{0}) = \min_{\mathbf{c} \neq \mathbf{0}} w(\mathbf{c}). \quad (8.5)$$

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 the generator matrix of any linear block code can be put into systematic form $\mathbf{G} = [\mathbf{I}_{k \times k} | \mathbf{P}]$, where \mathbf{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 the rows of the generator matrix \mathbf{G} are themselves valid codewords, it must be true that

$$d_{\text{free}} \leq n - k + 1 \quad (8.6)$$

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. Intuitively, this means that all codewords are as far apart from each other as possible in Hamming distance, and such codes should perform well.

An example of a simple block code that meets the Singleton bound is the binary repetition code

$$\begin{aligned} 0 &\longrightarrow \mathbf{c}_0 = (0, 0, \dots, 0)_n \\ 1 &\longrightarrow \mathbf{c}_1 = (1, 1, \dots, 1)_n \end{aligned}$$

In this case, $d_{\text{free}} = d(\mathbf{c}_0, \mathbf{c}_1) = n - k + 1$. The repetition code happens to be the only binary MDS code, and no other binary MDS codes exist. The well-known Reed–Solomon codes are good examples for non-binary MDS codes.

8.1.1.2 Syndromes

Suppose that the codeword \mathbf{c} is transmitted and the vector $\mathbf{y} = \mathbf{c} \oplus \mathbf{e}$ is received, where \mathbf{e} is defined as the error vector. The syndrome of the received vector \mathbf{y} is defined as the length $n - k$ vector

$$\mathbf{s} = \mathbf{y}\mathbf{H}^T. \quad (8.7)$$

If $\mathbf{s} = \mathbf{0}$, then \mathbf{y} is a codeword; conversely if $\mathbf{s} \neq \mathbf{0}$, then an error must have occurred. Note that if \mathbf{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 the sum of any two codewords is another codeword for a linear code, it follows that the number of undetectable error vectors is equal to $2^k - 1$, the number of non-zero codewords. The syndrome only depends upon the error vector because

$$\mathbf{s} = \mathbf{y}\mathbf{H}^T = \mathbf{c}\mathbf{H}^T \oplus \mathbf{e}\mathbf{H}^T = \mathbf{0} \oplus \mathbf{e}\mathbf{H}^T = \mathbf{e}\mathbf{H}^T. \quad (8.8)$$

In general, $\mathbf{s} = \mathbf{e}\mathbf{H}^T$ is a system of $n - k$ equations in n variables. Hence, for any given syndrome \mathbf{s} , there are 2^k possible solutions having the same error vector \mathbf{e} . However, the most likely error pattern \mathbf{e} is the one that has minimum Hamming weight.

8.1.1.3 Error Detection

A linear block code can detect all error patterns of $d_{\text{free}} - 1$ or fewer errors. 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 8.8, there are $2^4 - 1 = 15$ undetectable error patterns and $2^7 - 2^4 = 112$ detectable error patterns.

8.1.1.4 Weight Distribution

Consider a block code \mathcal{C} and let A_i be the number of codewords of weight i . The set $\{A_0, A_1, \dots, A_n\}$ is called the weight distribution of \mathcal{C} . The weight distribution can be expressed as a weight enumerator polynomial

$$A(z) = A_0z^0 + A_1z^1 + \dots + A_nz^n. \quad (8.9)$$

For the (7,4) Hamming code in Example 8.8,

$$A_0 = 1, A_2 = 0, A_3 = 7, A_4 = 7, A_5 = 0, A_6 = 0, A_7 = 1.$$

Hence,

$$A(z) = 1 + 7z^3 + 7z^4 + z^7.$$

The weight enumerator polynomial can be used to evaluate the exact performance of a code. Unfortunately, weight enumerator polynomials are generally difficult to find and are known only for a few classes of codes such as the Hamming codes.

8.1.1.5 Probability of Undetected Error

The probability of undetected error is

$$\begin{aligned} P_e[u] &= P[\mathbf{e} \text{ is a non-zero codeword}] \\ &= \sum_{i=1}^n A_i P[w(\mathbf{e}) = i]. \end{aligned} \quad (8.10)$$

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

$$P[y_i \neq c_i] = p = 1 - P[y_i = c_i]. \quad (8.11)$$

For a BSC, $P[w(\mathbf{e}) = i] = p^i(1-p)^{n-i}$ and, hence,

$$P_e[u] = \sum_{i=1}^n A_i p^i (1-p)^{n-i}. \quad (8.12)$$

The (7,4) Hamming code in Example 8.1 has an undetected error probability of

$$P_e[u] = 7p^3(1-p)^4 + 7p^4(1-p)^3 + p^7. \quad (8.13)$$

For a raw channel error rate of $p = 10^{-2}$, it follows $P_e[u] = 7 \times 10^{-6}$. Hence, the undetected error rate can be very small even for a fairly simple block code.

8.1.1.6 Error Correction

A linear block code can correct all error patterns of t or fewer errors, where

$$t \leq \left\lfloor \frac{d_{\text{free}} - 1}{2} \right\rfloor \quad (8.14)$$

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

$$\begin{aligned} P[e] &\leq 1 - P[t \text{ or fewer errors}] \\ &= 1 - \sum_{i=0}^t \binom{n}{i} p^i (1-p)^{n-i}. \end{aligned} \quad (8.15)$$

8.1.1.7 Standard Array Decoding

One conceptually simple method for decoding 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 vector \mathbf{e}_2 of weight 1 and place it under \mathbf{c}_0 . Under each codeword put $\mathbf{c}_i \oplus \mathbf{e}_2, i = 1, \dots, 2^k - 1$.
3. Select a minimum weight error vector \mathbf{e}_3 from the remaining unused n -tuples and place it under $\mathbf{c}_0 = \mathbf{0}$. Under each codeword put $\mathbf{c}_i \oplus \mathbf{e}_3, i = 1, \dots, 2^k - 1$.
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 i th row has the elements

$$F_i = \{\mathbf{e}_i, \mathbf{e}_i \oplus \mathbf{c}_1, \dots, \mathbf{e}_i \oplus \mathbf{c}_{2^k-1}\}.$$

The first element, \mathbf{e}_i , is called the coset leader. The standard array also consists of 2^k disjoint columns. The j th column has the elements

$$D_j = \{\mathbf{c}_j, \mathbf{c}_j \oplus \mathbf{e}_2, \dots, \mathbf{c}_j \oplus \mathbf{e}_{2^{n-k}}\}.$$

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 $\mathbf{y} \oplus \mathbf{e}_i = \mathbf{c}_j$. 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. Obviously, standard array decoding is only useful for simple codes, since 2^n vectors must be stored in memory. A somewhat simpler decoding strategy is syndrome decoding.

8.1.1.8 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}_ℓ where $\mathbf{e}_\ell\mathbf{H}^T = \mathbf{s}$.
3. Decode \mathbf{y} into $\mathbf{y} \oplus \mathbf{e}_\ell = \hat{\mathbf{c}}$.

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.

8.1.2 Space-Time Block Codes

Space-time block coding is a technique for coding across multiple antennas. In Sect. 6.11, a simple transmit diversity scheme by Alamouti [15] was introduced that used two transmit antennas in a $2 \times L_R$ arrangement, where L_R is the number of receiver antennas. The Alamouti scheme achieves a diversity order of $2L_R$ and has a very simple maximum likelihood decoding algorithm. It can be thought of as a very simple 2×2 space-time block code having the code matrix

$$S = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} \\ -\tilde{\mathbf{s}}_{(2)}^* & \tilde{\mathbf{s}}_{(1)}^* \end{bmatrix}, \quad (8.16)$$

where the rows represent time slots and the columns represent the transmissions from the antennas over time. For the Alamouti code two symbols are transmitted from two antennas over two time slots. Note that the transmitted symbols in the code matrix are modulation symbols, rather than binary information bits. The basic idea of the Alamouti code was later generalized by Tarokh, Jafarkhani, and Calderbank to an arbitrary number, L_T , of transmit antennas by applying orthogonal designs [315]. While a variety of space-time block codes exist, orthogonal space-time block codes are considered here. Orthogonal space-time block codes achieve a diversity order $L_T L_R$, while allowing for maximum likelihood based on computationally simple linear processing [316].

Let L_T be the number of transmit antennas and p represent the number of time slots that are used to transmit one space-time codeword. This gives rise to a $p \times L_T$ space-time code matrix. Each group of k information symbols, chosen from an $M = 2^m$ -ary alphabet, is encoded according to the space-time code matrix to generate a $p \times L_T$ space-time codeword. Since the codewords are transmitted simultaneously from L_T transmit antennas in p time slots, the space-time code rate is equal to $R = k/p$. Similar to the Alamouti code, the entries of the $p \times L_T$ code matrix are chosen to be a combination of the block of k modulating symbols $\{\tilde{\mathbf{s}}_{(1)}, \dots, \tilde{\mathbf{s}}_{(k)}\}$ and their complex conjugates $\{\tilde{\mathbf{s}}_{(1)}^*, \dots, \tilde{\mathbf{s}}_{(k)}^*\}$. For orthogonal space-time codes, the $p \times L_T$ code matrix S satisfies the following orthogonal property [315]:

$$S^H S = I_{L_T \times L_T} \sum_{i=1}^k |\tilde{\mathbf{s}}_{(i)}|^2. \quad (8.17)$$

Depending on the type of signal constellation from which the symbols $\tilde{\mathbf{s}}_{(i)}$ are drawn, either real-or complex-valued orthogonal space-time block codes can be constructed.

8.1.2.1 Real Orthogonal Space-Time Block Codes

A real orthogonal design is a $p \times L_T$ orthogonal matrix containing real-valued elements satisfying the orthogonal property in (8.17). First consider the case where S is an $L_T \times L_T$ square matrix. It is known that real orthogonal designs having square code matrices only exist for $L_T = 2, 4$ or 8 transmit antennas. Examples of real orthogonal space-time block codes are the 2×2 design

$$S_2 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} \\ -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} \end{bmatrix}, \quad (8.18)$$

the 4×4 design

$$S_4 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} \\ -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(3)} \\ -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(1)} & -\tilde{\mathbf{s}}_{(2)} \\ -\tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} \end{bmatrix}, \quad (8.19)$$

and the 8×8 design

$$S_8 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(5)} & \tilde{\mathbf{s}}_{(6)} & \tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(8)} \\ -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(7)} \\ -\tilde{\mathbf{s}}_{(3)} & -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(8)} & -\tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(6)} \\ -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(3)} & -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(8)} & -\tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(5)} \\ -\tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(7)} & -\tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} \\ -\tilde{\mathbf{s}}_{(6)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(7)} & -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(3)} \\ -\tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(1)} & -\tilde{\mathbf{s}}_{(2)} \\ -\tilde{\mathbf{s}}_{(8)} & -\tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(6)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} \end{bmatrix}. \quad (8.20)$$

The reader can verify that the columns of these matrices are mutually orthogonal, i.e., their dot product is zero. Hence, the orthogonal property in (8.17) is satisfied. Also, the rate achieved with these code matrices is $R = 1$. For example, with the 8×8 code matrix, $k = 8$ symbols are transmitted in $p = 8$ time slots so that $R = k/p = 1$. Finally, these codes all achieve full transmit diversity of order L_T .

The above orthogonal space-time code designs are based on $L_T \times L_T$ square matrices. Tarokh et al. [315] have developed generalized real orthogonal designs for any number of transmit antennas that achieve rate $R = 1$. They have shown that the minimum number of transmission periods p to achieve full rate is given by [315]

$$p = \min \{2^{4c+d}\}, \quad (8.21)$$

where the minimization is taken over the set

$$\{c, d \mid 0 \leq c, 0 \leq d < 4 \text{ and } 8c + 2^d \geq L_T\}. \quad (8.22)$$

Based on (8.21) and (8.22), real orthogonal designs for $L_T = 3, 5, 6,$ and 7 transmit antennas can be constructed that have full transmit diversity of order L_T and rate $R = 1$ as follows:

$$S_3 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(3)} \\ -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & -\tilde{\mathbf{s}}_{(4)} \\ -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(1)} \\ -\tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(2)} \end{bmatrix}, \quad (8.23)$$

$$S_5 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(5)} \\ -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(6)} \\ -\tilde{\mathbf{s}}_{(3)} & -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(7)} \\ -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(3)} & -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(8)} \\ -\tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(7)} & -\tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(1)} \\ -\tilde{\mathbf{s}}_{(6)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(7)} & -\tilde{\mathbf{s}}_{(2)} \\ -\tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(3)} \\ -\tilde{\mathbf{s}}_{(8)} & -\tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(6)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(4)} \end{bmatrix}. \quad (8.24)$$

$$S_6 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(5)} & \tilde{\mathbf{s}}_{(6)} \\ -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(5)} \\ -\tilde{\mathbf{s}}_{(3)} & -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(8)} \\ -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(3)} & -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(8)} & -\tilde{\mathbf{s}}_{(7)} \\ -\tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(7)} & -\tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} \\ -\tilde{\mathbf{s}}_{(6)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(7)} & -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} \\ -\tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} \\ -\tilde{\mathbf{s}}_{(8)} & -\tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(6)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} \end{bmatrix}. \quad (8.25)$$

and

$$S_7 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(5)} & \tilde{\mathbf{s}}_{(6)} & \tilde{\mathbf{s}}_{(7)} \\ -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(8)} \\ -\tilde{\mathbf{s}}_{(3)} & -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(8)} & -\tilde{\mathbf{s}}_{(5)} \\ -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(3)} & -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(8)} & -\tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(6)} \\ -\tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(7)} & -\tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(3)} \\ -\tilde{\mathbf{s}}_{(6)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(7)} & -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & -\tilde{\mathbf{s}}_{(4)} \\ -\tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(8)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(6)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(1)} \\ -\tilde{\mathbf{s}}_{(8)} & -\tilde{\mathbf{s}}_{(7)} & \tilde{\mathbf{s}}_{(6)} & \tilde{\mathbf{s}}_{(5)} & -\tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(2)} \end{bmatrix}. \quad (8.26)$$

Taking S_7 as an example, observe that $k = 8$ symbols are transmitted from $L_T = 7$ antennas in $p = 8$ time slots. Hence, S_7 has rate $R = k/p = 1$. Since the columns are mutually orthogonal the code achieves full transmit diversity of order $L_T = 7$.

8.1.2.2 Complex Orthogonal Space-Time Block Codes

Complex orthogonal space-time codes have $p \times L_T$ code matrices containing the elements $\{\tilde{\mathbf{s}}_{(1)}, \dots, \tilde{\mathbf{s}}_{(k)}\}$ and their complex conjugates $\{\tilde{\mathbf{s}}_{(1)}^*, \dots, \tilde{\mathbf{s}}_{(k)}^*\}$, and they satisfy the orthogonal property in (8.17). Such codes provide full transmit diversity of order L_T and have rate $R = k/p$. The Alamouti space-time code matrix in (8.16) is one such 2×2 scheme, that achieves a transmit diversity of order 2 with a full code rate $R = 1$. The Alamouti scheme is unique in that it is the only complex orthogonal space-time block code having full transmit diversity and rate $R = 1$. Tarkoh et al. [315] have constructed complex orthogonal space-time block codes with rate $R = 1/2$ for $L_T = 3$ and 4 transmit antennas as follows:

$$C_3 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(3)} \\ -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & -\tilde{\mathbf{s}}_{(4)} \\ -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(1)} \\ -\tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(2)} \\ \tilde{\mathbf{s}}_{(1)}^* & \tilde{\mathbf{s}}_{(2)}^* & \tilde{\mathbf{s}}_{(3)}^* \\ -\tilde{\mathbf{s}}_{(2)}^* & \tilde{\mathbf{s}}_{(1)}^* & -\tilde{\mathbf{s}}_{(4)}^* \\ -\tilde{\mathbf{s}}_{(3)}^* & \tilde{\mathbf{s}}_{(4)}^* & \tilde{\mathbf{s}}_{(1)}^* \\ -\tilde{\mathbf{s}}_{(4)}^* & -\tilde{\mathbf{s}}_{(3)}^* & \tilde{\mathbf{s}}_{(2)}^* \end{bmatrix}, \quad (8.27)$$

and

$$C_4 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} \\ -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & -\tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(3)} \\ -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(4)} & \tilde{\mathbf{s}}_{(1)} & -\tilde{\mathbf{s}}_{(2)} \\ -\tilde{\mathbf{s}}_{(4)} & -\tilde{\mathbf{s}}_{(3)} & \tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} \\ \tilde{\mathbf{s}}_{(1)}^* & \tilde{\mathbf{s}}_{(2)}^* & \tilde{\mathbf{s}}_{(3)}^* & \tilde{\mathbf{s}}_{(4)}^* \\ -\tilde{\mathbf{s}}_{(2)}^* & \tilde{\mathbf{s}}_{(1)}^* & -\tilde{\mathbf{s}}_{(4)}^* & \tilde{\mathbf{s}}_{(3)}^* \\ -\tilde{\mathbf{s}}_{(3)}^* & \tilde{\mathbf{s}}_{(4)}^* & \tilde{\mathbf{s}}_{(1)}^* & -\tilde{\mathbf{s}}_{(2)}^* \\ -\tilde{\mathbf{s}}_{(4)}^* & -\tilde{\mathbf{s}}_{(3)}^* & \tilde{\mathbf{s}}_{(2)}^* & \tilde{\mathbf{s}}_{(1)}^* \end{bmatrix}. \quad (8.28)$$

The reader can verify that the columns of these code matrices are all mutually orthogonal. Thus, these schemes achieve full transmit diversity of order L_T , but lose half of the theoretical bandwidth efficiency since they are only $R = 1/2$ codes.

By allowing linear processing at the transmitter, it is possible to construct higher rate complex space-time block codes. Tarokh et al. [315] have identified the following complex orthogonal designs: the $L_T = 3$ rate $R = 3/4$ code

$$H_3 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \frac{\tilde{\mathbf{s}}_{(3)}}{\sqrt{2}} \\ -\tilde{\mathbf{s}}_{(2)}^* & \tilde{\mathbf{s}}_{(1)}^* & \frac{\tilde{\mathbf{s}}_{(3)}}{\sqrt{2}} \\ \frac{\tilde{\mathbf{s}}_{(3)}^*}{\sqrt{2}} & \frac{\tilde{\mathbf{s}}_{(3)}^*}{\sqrt{2}} & \frac{(-\tilde{\mathbf{s}}_{(1)} - \tilde{\mathbf{s}}_{(1)}^* + \tilde{\mathbf{s}}_{(2)} - \tilde{\mathbf{s}}_{(2)}^*)}{2} \\ \frac{\tilde{\mathbf{s}}_{(3)}^*}{\sqrt{2}} & -\frac{\tilde{\mathbf{s}}_{(3)}^*}{\sqrt{2}} & \frac{(\tilde{\mathbf{s}}_{(2)} - \tilde{\mathbf{s}}_{(2)}^* + \tilde{\mathbf{s}}_{(1)} - \tilde{\mathbf{s}}_{(1)}^*)}{2} \end{bmatrix}. \quad (8.29)$$

and the $L_T = 4$ rate $R = 3/4$ code

$$H_4 = \begin{bmatrix} \tilde{\mathbf{s}}_{(1)} & \tilde{\mathbf{s}}_{(2)} & \frac{\tilde{\mathbf{s}}_{(3)}}{\sqrt{2}} & \frac{\tilde{\mathbf{s}}_{(3)}}{\sqrt{2}} \\ -\tilde{\mathbf{s}}_{(2)} & \tilde{\mathbf{s}}_{(1)} & \frac{\tilde{\mathbf{s}}_{(3)}}{\sqrt{2}} & -\frac{\tilde{\mathbf{s}}_{(3)}}{\sqrt{2}} \\ \frac{\tilde{\mathbf{s}}_{(3)}^*}{\sqrt{2}} & \frac{\tilde{\mathbf{s}}_{(3)}^*}{\sqrt{2}} & \frac{(-\tilde{\mathbf{s}}_{(1)} - \tilde{\mathbf{s}}_{(1)}^* + \tilde{\mathbf{s}}_{(2)} - \tilde{\mathbf{s}}_{(2)}^*)}{2} & \frac{(-\tilde{\mathbf{s}}_{(2)} + \tilde{\mathbf{s}}_{(2)}^* + \tilde{\mathbf{s}}_{(1)} - \tilde{\mathbf{s}}_{(1)}^*)}{2} \\ \frac{\tilde{\mathbf{s}}_{(3)}^*}{\sqrt{2}} & -\frac{\tilde{\mathbf{s}}_{(3)}^*}{\sqrt{2}} & \frac{(\tilde{\mathbf{s}}_{(2)} + \tilde{\mathbf{s}}_{(2)}^* + \tilde{\mathbf{s}}_{(1)} - \tilde{\mathbf{s}}_{(1)}^*)}{2} & \frac{(-\tilde{\mathbf{s}}_{(1)} - \tilde{\mathbf{s}}_{(1)}^* - \tilde{\mathbf{s}}_{(2)} + \tilde{\mathbf{s}}_{(2)}^*)}{2} \end{bmatrix}. \quad (8.30)$$

8.1.2.3 Decoding Orthogonal Space-Time Block Codes

The Alamouti space-time block code was shown to have an efficient and simple maximum likelihood decoding method in Sect. 6.11. We now discuss the decoding of orthogonal $p \times L_T$ space-time block codes. Following the notation in Sect. 6.11, the received signal vector at antenna j and time t is given by

$$\tilde{\mathbf{r}}_{(t),j} = \sum_{i=1}^{L_T} g_{i,j} \tilde{\mathbf{s}}_{(t),i} + \tilde{\mathbf{n}}_{(t),j}, \quad (8.31)$$

where $g_{i,j}$ is the complex channel gain from transmit antenna i to receiver antenna j , $\tilde{\mathbf{s}}_{(t),i}$ is the symbol vector transmitted from antenna i in time slot t , and the $\tilde{\mathbf{n}}_{(t),j}$ are independent zero-mean complex Gaussian random vectors with variance N_o in

each dimension of the signal space. Assuming perfect channel state information, the maximum likelihood receiver computes the decision metric

$$\mu(C) = \sum_{t=1}^p \sum_{j=1}^{L_R} \left\| \tilde{\mathbf{r}}_{(t),j} - \sum_{i=1}^{L_T} g_{i,j} \tilde{\mathbf{s}}_{(t),i} \right\|^2 \quad (8.32)$$

over all codewords $C = [\tilde{\mathbf{s}}_{(t),i}]_{p \times L_T}$ and chooses the codeword with the minimum metric.

First consider the real orthogonal space-time block codes with square matrices S_2 , S_4 , and S_8 . Note that the rows of code matrices S_2 , S_4 , and S_8 are all permutations of the first row, possibly with different signs. Let $\epsilon_1, \dots, \epsilon_{L_t}$ denote the permutations corresponding to these rows, and let $\delta_t^{(i)}$ denote the sign of the entry in row t and column i of the code matrix. Then $\epsilon_t^{(i)} = q$ means that $\tilde{\mathbf{s}}_{(t),i}$ is up to a sign change equal to the element in row t and column q of the code matrix. Since the columns of the space-time orthogonal code matrices S_2 , S_4 , and S_8 are mutually orthogonal, minimizing the metric in (8.32) is equivalent to minimizing [315]

$$\mu(C) = \sum_{i=1}^{L_T} P_i, \quad (8.33)$$

where

$$P_i = \left(\left\| \left[\sum_{t=1}^{L_T} \sum_{j=1}^{L_R} \tilde{\mathbf{r}}_{(t),j} g_{\epsilon_t^{(i)},j}^* \delta_t^{(i)} \right] - \tilde{\mathbf{s}}_i \right\|^2 + \left(-1 + \sum_{i,j} |g_{i,j}|^2 \right) \|\tilde{\mathbf{s}}_i\|^2 \right). \quad (8.34)$$

Note that $P_i, i = 1, \dots, L_T$ only depends on the choice of code symbol $\tilde{\mathbf{s}}_i$, the set of received vectors $\{\tilde{\mathbf{r}}_{(t),j}, j = 1, \dots, L_R\}$, the channel fading coefficients $\{g_{i,j}\}$, and the structure of the code matrix. It follows that minimizing the sum in (8.33) is equivalent to minimizing P_i in (8.34) for each $i, 1 \leq i \leq L_T$. This separable property results in a very simple decoding strategy that provides spatial diversity of order $L_T L_R$. The maximum likelihood receiver simply forms the L_T decision variables

$$R_i = \sum_{t=1}^{L_T} \sum_{j=1}^{L_R} \tilde{\mathbf{r}}_{(t),j} g_{\epsilon_t^{(i)},j}^* \delta_t^{(i)}, \quad i = 1, \dots, L_T \quad (8.35)$$

and decides in favor of symbol $\hat{\mathbf{s}}_i$ if

$$\hat{\mathbf{s}}_i = \arg \min_{\mathbf{s}} \|\mathbf{R}_i - \mathbf{s}\|^2 + \left(-1 + \sum_{i,j} |g_{i,j}|^2 \right) \|\tilde{\mathbf{s}}_i\|^2, \quad i = 1, \dots, L_T. \quad (8.36)$$

Similar low complexity decoding strategies for the other space-time block codes C_3 , C_4 , H_3 , and H_4 in (8.27), (8.28), (8.29) and (8.30), respectively, are not presented here but are available in [316].

8.2 Convolutional Codes

8.2.1 Encoder Description

The encoder for a rate-1/ n binary convolutional code can be viewed as a finite-state machine (FSM) that consists of a ν -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 single ν -stage shift register, the constraint length is equal to $K = \nu + 1$. A very simple rate-1/2, constraint length $K = 3$, binary convolutional encoder is shown in Fig. 8.1.

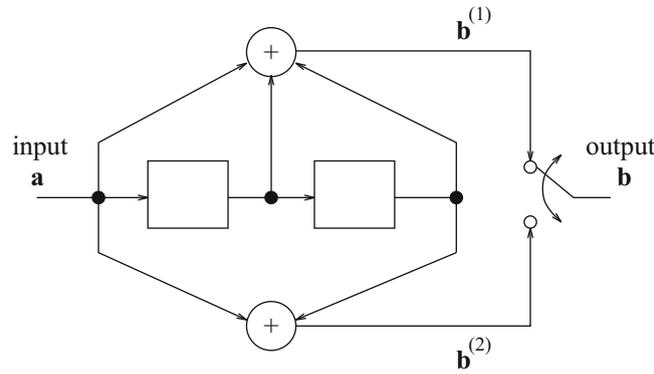


Fig. 8.1 Binary convolutional encoder; $R_c = 1/2, K = 3$

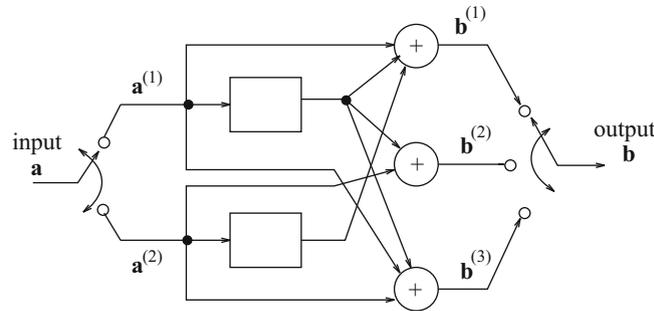


Fig. 8.2 Binary convolutional encoder; $R_c = 2/3, K = 2$

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 $\mathbf{a}_\ell = (a_\ell^{(1)}, \dots, a_\ell^{(k)})$ is input to the encoder at epoch ℓ to generate the n -bit code vector $\mathbf{b}_\ell = (b_\ell^{(1)}, \dots, b_\ell^{(n)})$. If K_i denotes the constraint length of the i th shift register, then the overall constraint length is defined as $K = \max_i K_i$. Figure 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 j th output sequence $\mathbf{b}^{(j)}$ that results from the i th input sequence $\mathbf{a}^{(i)} = (1, 0, 0, 0, \dots)$. The impulse responses can have a duration of at most K and have the form $\mathbf{g}_i^{(j)} = (g_{i,0}^{(j)}, g_{i,1}^{(j)}, \dots, g_{i,K-1}^{(j)})$. Sometimes the $\{\mathbf{g}_i^{(j)}\}$ are called generator sequences. For the encoder in Fig. 8.1

$$\mathbf{g}^{(1)} = (1, 1, 1) \quad \mathbf{g}^{(2)} = (1, 0, 1) \tag{8.37}$$

and for the encoder in Fig. 8.2

$$\begin{aligned} \mathbf{g}_1^{(1)} &= (1, 1), \quad \mathbf{g}_1^{(2)} = (0, 1), \quad \mathbf{g}_1^{(3)} = (1, 1) \\ \mathbf{g}_2^{(1)} &= (0, 1), \quad \mathbf{g}_2^{(2)} = (1, 0), \quad \mathbf{g}_2^{(3)} = (1, 0). \end{aligned} \tag{8.38}$$

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 $\mathbf{b}_i^{(j)} = \mathbf{a}^{(i)} \otimes \mathbf{g}_i^{(j)}$, where \otimes denotes modulo-2 convolution. The time domain convolutions can be conveniently replaced by polynomial multiplications in a D -transform domain according to

$$\mathbf{b}_i^{(j)}(D) = \mathbf{a}^{(i)}(D)\mathbf{g}_i^{(j)}(D), \tag{8.39}$$

where

$$\mathbf{a}^{(i)}(D) = \sum_{k=0}^{\infty} a_{i,k} D^k \quad (8.40)$$

is the i th input data polynomial,

$$\mathbf{b}_i^{(j)}(D) = \sum_{k=0}^{\infty} b_{i,k}^{(j)} D^k \quad (8.41)$$

is the j th output polynomial corresponding to the i th input, and

$$\mathbf{g}_i^{(j)}(D) = \sum_{k=0}^{K-1} g_{i,k}^{(j)} D^k \quad (8.42)$$

is the associated generator polynomial. It follows that the j th output sequence is

$$\mathbf{b}^{(j)}(D) = \sum_{i=1}^k \mathbf{b}_i^{(j)}(D) = \sum_{i=1}^k \mathbf{a}^{(i)}(D) \mathbf{g}_i^{(j)}(D). \quad (8.43)$$

The above expression leads to the matrix form

$$(\mathbf{b}^{(1)}(D), \dots, \mathbf{b}^{(n)}(D)) = (\mathbf{a}^{(1)}(D), \dots, \mathbf{a}^{(k)}(D)) \mathbf{G}(D), \quad (8.44)$$

where

$$\mathbf{G}(D) = \begin{bmatrix} \mathbf{g}_1^{(1)}(D), \dots, \mathbf{g}_1^{(n)}(D) \\ \vdots \\ \mathbf{g}_k^{(1)}(D), \dots, \mathbf{g}_k^{(n)}(D) \end{bmatrix} \quad (8.45)$$

is the generator matrix of the code. For the encoder in Fig. 8.1

$$\mathbf{G}(D) = [1 + D + D^2 \quad 1 + D^2] \quad (8.46)$$

while for the encoder in Fig. 8.2

$$\mathbf{G}(D) = \begin{bmatrix} 1 + D & D & 1 + D \\ D & 1 & 1 \end{bmatrix}. \quad (8.47)$$

After multiplexing the outputs, the final codeword has the polynomial representation

$$\mathbf{b}(D) = \sum_{j=1}^n D^{j-1} \mathbf{b}^{(j)}(D^n). \quad (8.48)$$

Systematic convolutional codes are those where first k of the n encoder output sequences, $\mathbf{b}^{(1)}, \dots, \mathbf{b}^{(k)}$ are equal to the k encoder input sequences $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(k)}$.

8.2.2 State and Trellis Diagrams, and Weight Distribution

Since the convolutional encoder is an FSM, its operation can be described by a state diagram and trellis diagram in a manner very similar to the state- and trellis-diagram descriptions of the discrete-time white noise channel model in Chap. 7. The

Fig. 8.3 State diagram for the binary convolutional encoder in Fig. 8.1

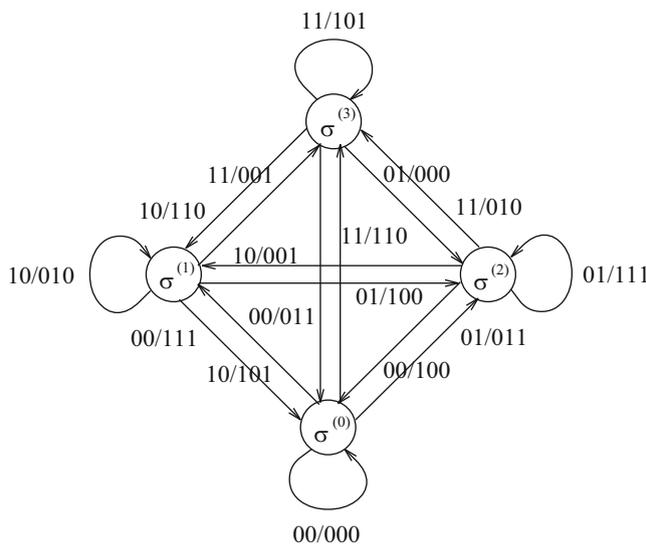
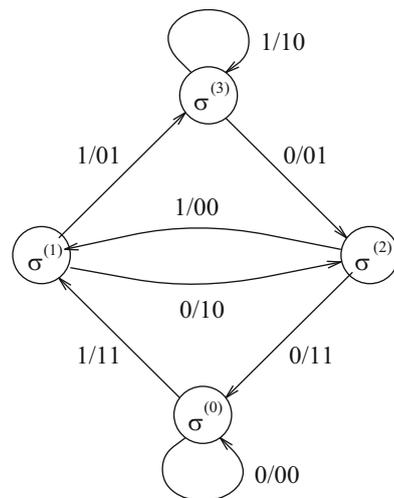


Fig. 8.4 State diagram for the binary convolutional encoder in Fig. 8.2

state of the encoder is defined by the shift register contents. For a rate- k/n code, the i th shift register contains v_i previous information bits. The state of the encoder at epoch ℓ is defined as

$$\sigma_\ell = (a_{\ell-1}^{(1)}, \dots, a_{\ell-v_1}^{(1)}; \dots; a_{\ell-1}^{(k)}, \dots, a_{\ell-v_m}^{(k)}). \tag{8.49}$$

There are a total of $N_S = 2^{v_T}$ encoder states, where $v_T \triangleq \sum_{i=1}^k v_i$ is defined as the total encoder memory. For a rate- $1/n$ code, the encoder state at epoch ℓ is simply $\sigma_\ell = (a_{\ell-1}, \dots, a_{\ell-v})$.

Figures 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 $\sigma^{(i)}$, $i = 0, \dots, v_T - 1$, where $\sigma^{(i)}$ represents the encoder state (c_0, \dots, c_{v_T-1}) corresponding to the integer $i = \sum_{j=0}^{v_T-1} c_j 2^j$. 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 $\mathbf{a}/\mathbf{b} = (a^{(1)}, a^{(2)}, \dots, a^{(k)}) / (b^{(1)}, b^{(2)}, \dots, b^{(n)})$. For example, the state transition $\sigma^{(1)} \rightarrow \sigma^{(3)}$ in Fig. 8.3 has the label 1/01. This means if the encoder in Fig. 8.1 is in state state $\sigma^{(1)} = (01)$ and the input bit is a 1, the encoder will output the code bits 01 and transition 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, 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, the zero state $\sigma^{(0)}$ can be split into two

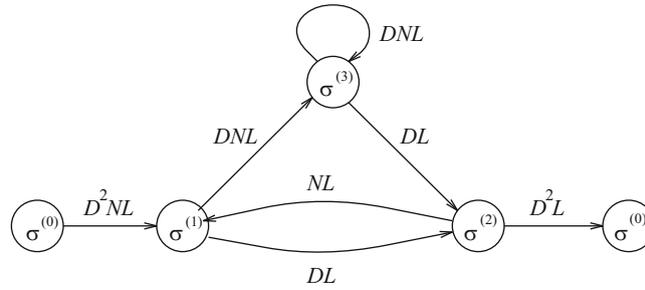


Fig. 8.5 Modified state diagram for the binary convolutional encoder in Fig. 8.1

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^iN^jL , 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 that transition. Every branch is labeled with the letter L , and the exponent of L is unity because each branch has length one. Each possible path through the modified state diagram corresponds to a non-all-zeroes codeword.

The weight distribution 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, including flow-graph reduction techniques and Mason's formula [219]. For the example shown in Fig. 8.5, the transfer function is

$$\begin{aligned} T(D, N, L) &= \frac{D^5NL^3}{1 - DNL(L + 1)} \\ &= D^5L^3N + D^6N^2L^4(L + 1) + D^7N^3L^5(L + 1)^2 \\ &\quad + \dots + D^{k+5}N^{k+1}L^{k+3}(L + 1)^k + \dots \end{aligned} \quad (8.50)$$

where the second line was obtained using polynomial division. The terms in the second line of (8.50) enumerate the weight distribution and distance properties of the code. For example, consider the term $D^{k+5}N^{k+1}L^{k+3}(L + 1)^k$ appearing in the transfer function. Using the binomial expansion, this term can be rewritten as

$$D^{k+5}N^{k+1} \sum_{n=0}^k \binom{k}{n} L^{n+k+3}.$$

Hence, there are 2^k paths through the modified state diagram that are at Hamming distance $k + 5$ from the all-zeroes path, i.e., have $k + 5$ output 1's, that are caused by $k + 1$ input 1's. Of these 2^k paths, $\binom{k}{n}$ have length $k + n + 3$ branches.

Sometimes the transfer function can be simplified if only certain distance properties of the convolutional code are to be extracted. For example, the output 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 (8.50) this leads to

$$\begin{aligned} T(D) &= \frac{D^5}{1 - 2D} \\ &= D^5 + 2D^6 + 4D^7 + \dots + 2^k D^{5+k} + \dots \end{aligned} \quad (8.51)$$

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. This means that the free Hamming distance of the code is $d_{\text{free}} = 5$. The free Hamming distance for this simple example can also be seen by inspecting the trellis diagram in Fig. 8.6, where the branches in the trellis diagram 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 [272], Lin and Costello [203], and Clark and Cain [73].

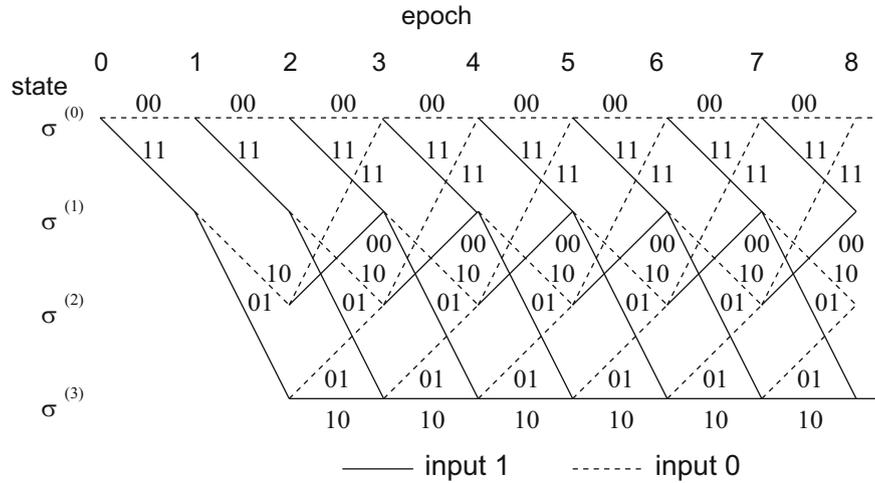


Fig. 8.6 Trellis diagram for the binary convolutional encoder in Fig. 8.1

8.2.3 Recursive Systematic Convolutional Codes

Forney [129] and Costello [82] showed that it is possible to construct a recursive systematic convolutional (RSC) encoder from every rate $R_c = 1/n$ feedforward non-systematic convolutional encoder, such that the output weight distributions of the codes are identical. Consider a rate- $1/n$ code with generator polynomials $\mathbf{g}_1(D), \dots, \mathbf{g}_n(D)$. The output sequences are described by the polynomials

$$\mathbf{b}^{(j)}(D) = \mathbf{a}(D)\mathbf{g}^{(j)}(D), \quad j = 1, \dots, n. \quad (8.52)$$

To obtain a systematic code, it is required that $\mathbf{b}^{(1)}(D) = \mathbf{a}(D)$. To obtain this, suppose that both sides of (8.52) are divided by $\mathbf{g}^{(1)}(D)$, so that

$$\tilde{\mathbf{b}}^{(1)}(D) = \frac{\mathbf{b}^{(1)}(D)}{\mathbf{g}^{(1)}(D)} = \mathbf{a}(D) \quad (8.53)$$

$$\tilde{\mathbf{b}}^{(j)}(D) = \frac{\mathbf{b}^{(j)}(D)}{\mathbf{g}^{(1)}(D)} = \mathbf{a}(D) \frac{\mathbf{g}^{(j)}(D)}{\mathbf{g}^{(1)}(D)}, \quad j = 2, \dots, n. \quad (8.54)$$

Sometimes the $\mathbf{g}^{(j)}(D)$ are called the feedforward polynomials, while $\mathbf{g}^{(1)}(D)$ is called the feed-back polynomial. Define a new input sequence $\tilde{\mathbf{a}}(D)$ as

$$\tilde{\mathbf{a}}(D) \triangleq \frac{\mathbf{a}(D)}{\mathbf{g}^{(1)}(D)} \quad (8.55)$$

so that

$$\tilde{\mathbf{b}}^{(1)}(D) = \tilde{\mathbf{a}}(D)\mathbf{g}^{(1)}(D) \quad (8.56)$$

$$\tilde{\mathbf{b}}^{(j)}(D) = \tilde{\mathbf{a}}(D)\mathbf{g}^{(j)}(D), \quad j = 2, \dots, n. \quad (8.57)$$

Observe that the transformation between $\mathbf{a}(D)$ and $\tilde{\mathbf{a}}(D)$ in (8.55) is that of a recursive digital filter with modulo-2 operations. This transformation simply reorders the input sequence $\mathbf{a}(D)$. 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 output weight distribution functions. However, the input weight distributions for the two codes are completely different as will be shown.

Example 8.3. Consider, for example, the rate-1/2 encoder in Fig. 8.1 with generators

$$\mathbf{g}^{(1)}(D) = 1 + D + D^2 \quad (8.58)$$

$$\mathbf{g}^{(2)}(D) = 1 + D^2. \quad (8.59)$$

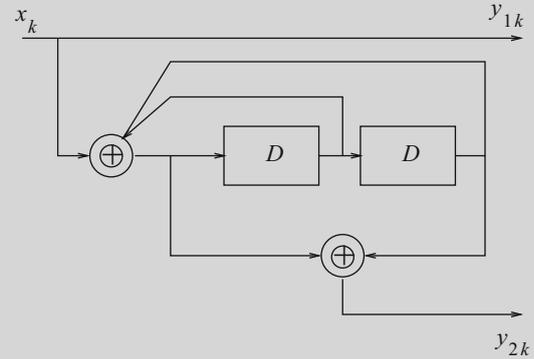
By following the above described procedure a RSC code is obtained with generators

$$\hat{\mathbf{g}}^{(1)}(D) = 1$$

$$\hat{\mathbf{g}}^{(2)}(D) = \frac{\mathbf{g}^{(2)}(D)}{\mathbf{g}^{(1)}(D)} = \frac{1 + D^2}{1 + D + D^2}.$$

The RSC is shown in Fig. 8.7.

Fig. 8.7 Recursive systematic convolutional (RSC) encoder derived from the feedforward non-systematic encoder in Fig. 8.1



Similar to their feedforward 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

$$T(D, N, L) = \frac{D^5 N^3 L^3 - D^6 N^4 L^4 + D^6 N^2 L^4}{1 - DNL - DNL^2 - D^2 L^3 + D^2 N^2 L^3} \quad (8.60)$$

$$= D^5 N^3 L^3 + D^6 N^2 L^4 + D^6 N^4 L^5 + \dots \quad (8.61)$$

Setting $N = 1$ and $L = 1$ in the transfer function $T(D, N, L)$ gives the output weight distribution of the code, $T(D)$, which is identical to the output weight distribution of the corresponding feedforward non-systematic encoder in (8.51). However, by comparing the first few terms in their respective transfer functions in (8.50) and (8.61) 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 feedforward 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 $1 + D + D^2$. It will be shown later that these properties are crucial for turbo codes.

Finally, both the feedforward 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.

8.2.4 Maximum Likelihood Decoding

The Viterbi algorithm was devised by Viterbi for maximum likelihood decoding of convolutional codes [334, 335]. To develop the decoder, first note that the convolutional encoder outputs n bits per branch, that are subsequently mapped onto an $M = 2^k$ -point signal constellation. Here, it is assumed that there are $\ell = n/k$ modulated symbols per branch where ℓ is an integer. For example, a rate-1/2 having 2 code bits per branch may have the 2 code bits mapped onto either 2 binary symbols or just one quaternary symbol. To derive the Viterbi algorithm, suppose that a sequence of modulated symbols $\tilde{\mathbf{s}} = \{\tilde{\mathbf{s}}_n\}_{n=1}^k$, $\tilde{\mathbf{s}}_n = (\tilde{\mathbf{s}}_{n,1}, \dots, \tilde{\mathbf{s}}_{n,\ell})$, corresponding to k branches in the code trellis are transmitted over an interleaved flat fading channel with AWGN.¹ After receiving the sequence $\tilde{\mathbf{r}} = \{\tilde{\mathbf{r}}_n\}_{n=1}^k$, $\tilde{\mathbf{r}}_n = (\tilde{\mathbf{r}}_{n,1}, \dots, \tilde{\mathbf{r}}_{n,\ell})$, the maximum likelihood receiver uses knowledge of the sequence of complex channel gains $\mathbf{g} = \{\mathbf{g}_n\}_{n=1}^k$, $\mathbf{g}_n = (g_{n,1}, \dots, g_{n,\ell})$ (obtained from a separate channel estimator) to decide in favor of the sequence $\tilde{\mathbf{s}}$ that maximizes the likelihood function

$$p(\tilde{\mathbf{r}}_k, \dots, \tilde{\mathbf{r}}_1 | g_{k,\ell}\tilde{\mathbf{s}}_{k,\ell}, \dots, g_{1,1}\tilde{\mathbf{s}}_{1,1}) \quad (8.62)$$

or, equivalently, the log-likelihood function

$$\log\{p(\tilde{\mathbf{r}}_k, \dots, \tilde{\mathbf{r}}_1 | g_{k,\ell}\tilde{\mathbf{s}}_{k,\ell}, \dots, g_{1,1}\tilde{\mathbf{s}}_{1,1})\}. \quad (8.63)$$

The log-likelihood function in (8.63) can be rewritten as

$$\begin{aligned} & \log\{p(\tilde{\mathbf{r}}_k, \dots, \tilde{\mathbf{r}}_1 | g_{k,\ell}\tilde{\mathbf{s}}_{k,\ell}, \dots, g_{1,1}\tilde{\mathbf{s}}_{1,1})\} \\ &= \log\{p(\tilde{\mathbf{r}}_k | g_{k,1}\tilde{\mathbf{s}}_{k,1}, \dots, g_{k,\ell}\tilde{\mathbf{s}}_{k,\ell})\} + \log\{p(\tilde{\mathbf{r}}_{k-1}, \dots, \tilde{\mathbf{r}}_1 | g_{k-1,\ell}\tilde{\mathbf{s}}_{k-1,\ell}, \dots, g_{1,1}\tilde{\mathbf{s}}_{1,1})\}. \end{aligned} \quad (8.64)$$

If the second term on the right side of (8.64) has been calculated previously at epoch $k-1$ and stored in memory, then only the first term, called the branch metric, has to be computed for the incoming signal vector \mathbf{r}_k at epoch k . From our treatment in Chap. 5

$$p(\tilde{\mathbf{r}}_k | g_{k,1}\tilde{\mathbf{s}}_{k,1}, \dots, g_{k,\ell}\tilde{\mathbf{s}}_{k,\ell}) = \frac{1}{(2\pi N_o)^N} \exp\left\{-\frac{1}{2N_o} \sum_{m=1}^{\ell} \|\tilde{\mathbf{r}}_{k,m} - g_{k,m}\tilde{\mathbf{s}}_{k,m}\|^2\right\} \quad (8.65)$$

so that $\log\{p(\tilde{\mathbf{r}}_k | g_{k,1}\tilde{\mathbf{s}}_{k,1}, \dots, g_{k,\ell}\tilde{\mathbf{s}}_{k,\ell})\}$ yields the Euclidean branch metric

$$\mu_k = -\sum_{m=1}^{\ell} \|\tilde{\mathbf{r}}_{k,m} - g_{k,m}\tilde{\mathbf{s}}_{k,m}\|^2. \quad (8.66)$$

Based on the recursion in (8.64) and the branch metric in (8.66), the Viterbi algorithm searches through the N_S -state code trellis for the most likely transmitted sequence $\tilde{\mathbf{s}}$ given the sequence of received vectors $\tilde{\mathbf{r}}$ and knowledge of the sequence of complex channel gains \mathbf{g} . At epoch k , the Viterbi algorithm stores N_S survivors $\check{\mathbf{s}}(\sigma_k^{(i)})$ along with their associated path metrics $\Gamma(\sigma_k^{(i)})$ that terminate at state $\sigma_k^{(i)}$, $i = 0, \dots, N_S - 1$. The path metric is defined as

$$\Gamma(\sigma_k^{(i)}) = \sum_{n=1}^k \mu_n^{(i)}, \quad i = 0, \dots, N_S - 1, \quad (8.67)$$

where $\{\mu_n^{(i)}\}$ is the sequence of branch metrics along the surviving path $\check{\mathbf{s}}(\sigma_k^{(i)})$.

¹Here, complex low-pass vector notation is used.

8.2.4.1 Viterbi Algorithm

The Viterbi algorithm is initialized at time $k = 0$, by setting all path metrics to zero, i.e., $\Gamma(\sigma_0^{(i)}) = 0$, $i = 1, \dots, N_S - 1$.

1. After the vector $\tilde{\mathbf{r}}_{k+1}$ has been received, compute the set of path metrics $\Gamma(\sigma_k^{(i)} \rightarrow \sigma_{k+1}^{(j)}) = \Gamma(\sigma_k^{(i)}) + \mu(\sigma_k^{(i)} \rightarrow \sigma_{k+1}^{(j)})$ for all possible paths through the trellis that terminate in each state $\sigma_{k+1}^{(j)}$, $j = 0, \dots, N_S - 1$, where $\mu(\sigma_k^{(i)} \rightarrow \sigma_{k+1}^{(j)})$ is the branch metric defined below. For a modulation alphabet of size M , there will be M such paths that terminate in each state $\sigma_{k+1}^{(j)}$.
2. Find $\Gamma(\sigma_{k+1}^{(j)}) = \max_i \Gamma(\sigma_k^{(i)} \rightarrow \sigma_{k+1}^{(j)})$, $j = 0, \dots, N_S - 1$ where the maximization is over all M possible paths through the trellis that terminate in state $\sigma_{k+1}^{(j)}$.
3. Store $\Gamma(\sigma_{k+1}^{(j)})$ and its associated surviving sequence $\check{s}(\sigma_{k+1}^{(j)})$, $j = 0, \dots, N_S - 1$. Drop all other paths.
4. Increment the time index k , go to Step 1 and repeat the entire algorithm.

In Step 1 above, $\mu(\sigma_k^{(i)} \rightarrow \sigma_{k+1}^{(j)})$ is the branch metric associated with the state transition $\sigma_k^{(i)} \rightarrow \sigma_{k+1}^{(j)}$ and is computed according to the following variation of (8.66)

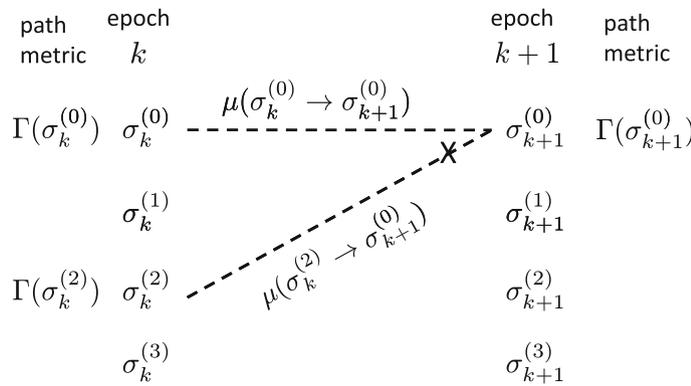
$$\mu(\sigma_k^{(i)} \rightarrow \sigma_{k+1}^{(j)}) = - \sum_{m=1}^{\ell} \|\tilde{\mathbf{r}}_{k,m} - g_{k,m} \tilde{s}_{k,m}(\sigma_k^{(i)} \rightarrow \sigma_{k+1}^{(j)})\|^2, \quad (8.68)$$

where $\tilde{s}_{k,m}(\sigma_k^{(i)} \rightarrow \sigma_{k+1}^{(j)})$ is a symbol that is uniquely determined by the state transition $\sigma_k^{(i)} \rightarrow \sigma_{k+1}^{(j)}$ and the symbol mapping being used.

The calculation of the path metric $\Gamma(\sigma_{k+1}^{(0)})$ for state $\sigma^{(0)}$ at epoch $k + 1$ is illustrated in Fig. 8.8 for the code trellis shown in Fig. 8.6. In this case, there are two paths merging into state $\sigma^{(0)}$ at epoch $k + 1$. The Viterbi algorithm determines the path metric $\Gamma(\sigma_{k+1}^{(0)})$ in this particular example as

$$\begin{aligned} \Gamma(\sigma_{k+1}^{(0)}) &= \max \left\{ \Gamma(\sigma_k^{(0)}) + \mu(\sigma_k^{(0)} \rightarrow \sigma_{k+1}^{(0)}), \Gamma(\sigma_k^{(2)}) + \mu(\sigma_k^{(2)} \rightarrow \sigma_{k+1}^{(0)}) \right\} \\ &= \Gamma(\sigma_k^{(0)}) + \mu(\sigma_k^{(0)} \rightarrow \sigma_{k+1}^{(0)}). \end{aligned}$$

Hence, the path that includes the state transition $\sigma_k^{(0)} \rightarrow \sigma_{k+1}^{(0)}$ is the survivor and the path that includes the state transition $\sigma_k^{(2)} \rightarrow \sigma_{k+1}^{(0)}$ is “dropped” as indicated by the “X” on the path.



$$\begin{aligned} \Gamma(\sigma_{k+1}^{(0)}) &= \max[\Gamma(\sigma_k^{(0)}) + \mu(\sigma_k^{(0)} \rightarrow \sigma_{k+1}^{(0)}), \Gamma(\sigma_k^{(2)}) + \mu(\sigma_k^{(2)} \rightarrow \sigma_{k+1}^{(0)})] \\ &= \Gamma(\sigma_k^{(0)}) + \mu(\sigma_k^{(0)} \rightarrow \sigma_{k+1}^{(0)}) \end{aligned}$$

Fig. 8.8 Path metric update for the code trellis in Fig. 8.6

8.2.5 BCJR Algorithm

The Viterbi algorithm uses maximum likelihood sequence estimation (MLSE) to find the most likely input sequence. The BCJR algorithm, named after Bahl et al. [31], is a symbol-by-symbol maximum a posteriori probability (MAP) algorithm for decoding convolutional codes. Here, the BCJR algorithm is described for rate-1/n convolutional codes with binary modulation.

Having observed the received sequence $\tilde{\mathbf{r}}$, the decoder calculates the a posteriori probabilities (APPs) $P[a_k = +1|\tilde{\mathbf{r}}]$ and $P[a_k = -1|\tilde{\mathbf{r}}]$, and decides $a_k = +1$ if $P[a_k = +1|\tilde{\mathbf{r}}] > P[a_k = -1|\tilde{\mathbf{r}}]$ and $a_k = -1$ otherwise. Alternatively, the decoder can calculate the a posteriori log-likelihood ratio (LLR)

$$L(a_k|\tilde{\mathbf{r}}) \triangleq \log \left(\frac{P[a_k = +1|\tilde{\mathbf{r}}]}{P[a_k = -1|\tilde{\mathbf{r}}]} \right), \quad (8.69)$$

and make the decision $\hat{a}_k = \text{sign}(L(a_k|\tilde{\mathbf{r}}))$. First note that code bit a_k is output for the state transition $\sigma_{k-1} \rightarrow \sigma_k$, and there may be several such state transitions that will output the same code bit. It follows that the APP is

$$\begin{aligned} P[a_k = u|\tilde{\mathbf{r}}] &= \frac{p(\tilde{\mathbf{r}}, a_k = u)}{p(\tilde{\mathbf{r}})} \\ &\propto p(\tilde{\mathbf{r}}, a_k = u) \\ &= \sum_{\sigma_k \rightarrow \sigma_{k+1}: a_k = u} p(\tilde{\mathbf{r}}, \sigma_k, \sigma_{k+1}) \end{aligned} \quad (8.70)$$

Hence, the a posteriori LLR can be written as

$$L(a_k|\tilde{\mathbf{r}}) = \log \left\{ \frac{\sum_{\sigma_k \rightarrow \sigma_{k+1}: a_k = +1} p(\tilde{\mathbf{r}}, \sigma_k, \sigma_{k+1})}{\sum_{\sigma_k \rightarrow \sigma_{k+1}: a_k = -1} p(\tilde{\mathbf{r}}, \sigma_k, \sigma_{k+1})} \right\}. \quad (8.71)$$

One elegant form of the APPs can be obtained by first defining the sequences

$$\tilde{\mathbf{r}}_{j < k} = \{\tilde{\mathbf{r}}_{j,1}, \dots, \tilde{\mathbf{r}}_{j,n}\}_{j=1}^{k-1} \quad (8.72)$$

$$\tilde{\mathbf{r}}_{j > k} = \{\tilde{\mathbf{r}}_{j,1}, \dots, \tilde{\mathbf{r}}_{j,n}\}_{j=k+1}^N \quad (8.73)$$

$$\tilde{\mathbf{r}}_k = \{\tilde{\mathbf{r}}_{k,1}, \dots, \tilde{\mathbf{r}}_{k,n}\} \quad (8.74)$$

such that $\tilde{\mathbf{r}} = (\tilde{\mathbf{r}}_{j < k}, \tilde{\mathbf{r}}_k, \tilde{\mathbf{r}}_{j > k})$, where N is the block length (equal to the number of input bits). Then the joint probability $p(\tilde{\mathbf{r}}, \sigma_k, \sigma_{k+1})$ can be split as follows [154]:

$$\begin{aligned} p(\tilde{\mathbf{r}}, \sigma_k, \sigma_{k+1}) &= p(\sigma_k, \sigma_{k+1}, \tilde{\mathbf{r}}_{j < k}, \tilde{\mathbf{r}}_k, \tilde{\mathbf{r}}_{j > k}) \\ &= p(\sigma_k, \sigma_{k+1}, \tilde{\mathbf{r}}_{j < k}, \tilde{\mathbf{r}}_k) p(\tilde{\mathbf{r}}_{j > k} | \sigma_k, \sigma_{k+1}, \tilde{\mathbf{r}}_{j < k}, \tilde{\mathbf{r}}_k) \\ &= p(\sigma_k, \tilde{\mathbf{r}}_{j < k}) p(\tilde{\mathbf{r}}_k, \sigma_{k+1} | \sigma_k, \tilde{\mathbf{r}}_{j < k}) p(\tilde{\mathbf{r}}_{j > k} | \sigma_k, \sigma_{k+1}, \tilde{\mathbf{r}}_{j < k}, \tilde{\mathbf{r}}_k) \\ &= p(\sigma_k, \tilde{\mathbf{r}}_{j < k}) \cdot p(\tilde{\mathbf{r}}_k, \sigma_{k+1} | \sigma_k) \cdot p(\tilde{\mathbf{r}}_{j > k} | \sigma_{k+1}) \\ &= \alpha_k(\sigma_k) \cdot \gamma_k(\sigma_k \rightarrow \sigma_{k+1}) \cdot \beta_{k+1}(\sigma_{k+1}), \end{aligned} \quad (8.75)$$

leading to three terms; the branch metric $\gamma_k(\sigma_k \rightarrow \sigma_{k+1})$, and the terms $\alpha_k(\sigma_k)$ and $\beta_{k+1}(\sigma_{k+1})$. The second last equality in (8.75) comes from the properties of the code trellis, which causes $\tilde{\mathbf{r}}_{j > k}$ to depend on $\sigma_k, \sigma_{k+1}, \tilde{\mathbf{r}}_{j < k}, \tilde{\mathbf{r}}_k$ only through the state σ_{k+1} , and the pair $\tilde{\mathbf{r}}_k, \sigma_{k+1}$ depends on $\sigma_k, \tilde{\mathbf{r}}_{j < k}$ only through the state σ_k .

The terms $\alpha_k(\sigma_k)$ and $\beta_{k+1}(\sigma_{k+1})$ can be calculated according to a forward recursion and a backward recursion, respectively. The forward recursion is given by

$$\begin{aligned} \alpha_{k+1}(\sigma_{k+1}) &= p(\sigma_{k+1}, \tilde{\mathbf{r}}_{j < k+1}) \\ &= p(\sigma_{k+1}, \tilde{\mathbf{r}}_k, \tilde{\mathbf{r}}_{j < k}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{\sigma_k} p(\sigma_{k+1}, \sigma_k, \tilde{\mathbf{r}}_k, \tilde{\mathbf{r}}_{j < k}) \\
&= \sum_{\sigma_k} p(\tilde{\mathbf{r}}_{j < k}, \sigma_k) p(\tilde{\mathbf{r}}_k, \sigma_{k+1} | \sigma_k) \\
&= \sum_{\sigma_k} \alpha_k(\sigma_k) \cdot \gamma_k(\sigma_k \rightarrow \sigma_{k+1})
\end{aligned} \tag{8.76}$$

with the initial condition $\alpha_0(\sigma_0) = 1$, i.e., the all-zeroes state σ_0 is the initial state in the code trellis.

Similarly, the backward recursion is given by

$$\begin{aligned}
\beta_k(\sigma_k) &= p(\tilde{\mathbf{r}}_{j > k-1} | \sigma_k) \\
&= \sum_{\sigma_{k+1}} p(\sigma_{k+1}, \tilde{\mathbf{r}}_k, \tilde{\mathbf{r}}_{j > k} | \sigma_k) \\
&= \sum_{\sigma_{k+1}} p(\tilde{\mathbf{r}}_k, \sigma_{k+1} | \sigma_k) p(\tilde{\mathbf{r}}_{j > k} | \sigma_{k+1}) \\
&= \sum_{\sigma_{k+1}} \gamma_k(\sigma_k \rightarrow \sigma_{k+1}) \cdot \beta_{k+1}(\sigma_{k+1})
\end{aligned} \tag{8.77}$$

with the initial condition $\beta_N(\sigma_0) = 1$, i.e., the all-zeroes state σ_0 is the ending state of the code trellis. Note that tail bits are required to terminate the trellis in state σ_0 .

The branch metric $\gamma_k(\sigma_k \rightarrow \sigma_{k+1})$ has the form

$$\begin{aligned}
\gamma_k(\sigma_k \rightarrow \sigma_{k+1}) &= p(\sigma_k, \tilde{\mathbf{r}}_i, \sigma_{k+1}) \\
&\quad - p(\sigma_{k+1} | \sigma_k) p(\tilde{\mathbf{r}}_i | \sigma_{k+1}, \sigma_k) \\
&= \mathbb{P}[a_k = u] p(\tilde{\mathbf{r}}_k | a_k = u) \\
&= \mathbb{P}[a_k = u] p(\tilde{\mathbf{r}}_k | g_{k,1} \tilde{\mathbf{s}}_{k,1}, \dots, g_{k,n} \tilde{\mathbf{s}}_{k,n}),
\end{aligned} \tag{8.78}$$

where $\tilde{\mathbf{s}}_k = \{\tilde{s}_{k,1}, \dots, \tilde{s}_{k,n}\}$ is the sequence of binary modulated symbols transmitted at epoch k , and $\mathbf{g}_k = \{g_{k,1}, \dots, g_{k,n}\}$ is the corresponding sequence of complex channel gains at epoch k . The third line in (8.78) used the fact that there is a one-to-one correspondence between the state transition $\sigma_k \rightarrow \sigma_{k+1}$ in the code trellis and the input bit $a_k = u$, and where it has been assumed that the state transition $\sigma_k \rightarrow \sigma_{k+1}$ is possible in the code trellis. Note that the branch metric depends on the prior probability $\mathbb{P}[a_k = u]$ of the information bit at epoch k , and on the conditional probability $p(\tilde{\mathbf{r}}_k | g_{k,1} \tilde{\mathbf{s}}_{k,1}, \dots, g_{k,n} \tilde{\mathbf{s}}_{k,n})$ which is given by (8.65).

Finally, by using (8.75) and (8.71) along with (8.76)–(8.78) the *a posteriori* LLR is obtained as

$$L(a_k | \tilde{\mathbf{r}}) = \log \left\{ \frac{\sum_{\sigma_k \rightarrow \sigma_{k+1}: a_k = +1} \alpha_k(\sigma_k) \cdot \gamma_k(\sigma_k \rightarrow \sigma_{k+1}) \cdot \beta_{k+1}(\sigma_{k+1})}{\sum_{\sigma_k \rightarrow \sigma_{k+1}: a_k = -1} \alpha_k(\sigma_k) \cdot \gamma_k(\sigma_k \rightarrow \sigma_{k+1}) \cdot \beta_{k+1}(\sigma_{k+1})} \right\}, \tag{8.79}$$

and decisions are made according to $\hat{a}_k = \text{sign}(L(a_k | \tilde{\mathbf{r}}))$. Note that the $L(a_k | \tilde{\mathbf{r}})$ provide a level of certainty of the decoder about the value of a_k and are called soft outputs. These soft outputs are essential for the decoding of turbo codes that will be discussed later in this chapter.

8.2.5.1 Log-MAP Algorithm

The BCJR algorithm as described above exhibits numerical instability in the form of underflows and overflows. An alternative to this algorithm is its log-domain version known as the log-APP or log-MAP algorithm. Instead of using $\alpha_k(\sigma_k)$, $\beta_{k+1}(\sigma_{k+1})$ and $\gamma_k(\sigma_k \rightarrow \sigma_{k+1})$, their logarithms are defined as follows:

$$\tilde{\alpha}_k(\sigma_k) = \log\{\alpha_k(\sigma_k)\} \quad (8.80)$$

$$\tilde{\beta}_{k+1}(\sigma_{k+1}) = \log\{\beta_{k+1}(\sigma_{k+1})\} \quad (8.81)$$

$$\tilde{\gamma}_k(\sigma_k \rightarrow \sigma_{k+1}) = \log\{\gamma_k(\sigma_k \rightarrow \sigma_{k+1})\} \quad (8.82)$$

By examining (8.76) and (8.77), it can be seen that

$$\tilde{\alpha}_{k+1}(\sigma_{k+1}) = \log \left\{ \sum_{\sigma_k} e^{\tilde{\alpha}_k(\sigma_k) + \tilde{\gamma}_k(\sigma_k \rightarrow \sigma_{k+1})} \right\} \quad (8.83)$$

$$\tilde{\beta}_k(\sigma_k) = \log \left\{ \sum_{\sigma_{k+1}} e^{\tilde{\gamma}_k(\sigma_k \rightarrow \sigma_{k+1}) + \tilde{\beta}_{k+1}(\sigma_{k+1})} \right\} \quad (8.84)$$

with the initial conditions $\tilde{\alpha}_0(\sigma_0) = 0$ and $\tilde{\beta}_N(\sigma_N) = 0$, assuming that the code trellis begins and ends in state σ_0 . The a posteriori LLR values are calculated according to

$$L(a_k|\tilde{\mathbf{r}}) = \log \left\{ \frac{\sum_{\sigma_k \rightarrow \sigma_{k+1}: a_k=+1} e^{\tilde{\alpha}_k(\sigma_k) + \tilde{\gamma}_k(\sigma_k \rightarrow \sigma_{k+1}) + \tilde{\beta}_{k+1}(\sigma_{k+1})}}{\sum_{\sigma_k \rightarrow \sigma_{k+1}: a_k=-1} e^{\tilde{\alpha}_k(\sigma_k) + \tilde{\gamma}_k(\sigma_k \rightarrow \sigma_{k+1}) + \tilde{\beta}_{k+1}(\sigma_{k+1})}} \right\}. \quad (8.85)$$

To proceed further, define the Jacobian logarithm

$$\max^*\{x, y\} \triangleq \log\{e^x + e^y\} \quad (8.86)$$

$$\max^*\{x, y, z\} \triangleq \log\{e^x + e^y + e^z\}. \quad (8.87)$$

Note that

$$\max^*\{x, y\} = \max\{x, y\} + \log\{1 + e^{-|x-y|}\}. \quad (8.88)$$

The second term $\log\{1 + e^{-|x-y|}\}$ is small when x and y are not close, and its maximum value is equal to $\log\{2\}$ when $x = y$. Hence, when x and y are not close, the approximation

$$\max^*\{x, y\} \approx \max\{x, y\} \quad (8.89)$$

can be used. Using the above approximation in place of the Jacobian logarithm yields a suboptimal (but simpler) implementation the log-MAP algorithm called the max-log-MAP algorithm.

Using the Jacobian logarithm, it follows that

$$\tilde{\alpha}_{k+1}(\sigma_{k+1}) = \max_{\sigma_k}^* (\tilde{\alpha}_k(\sigma_k) + \tilde{\gamma}_k(\sigma_k \rightarrow \sigma_{k+1})) \quad (8.90)$$

$$\tilde{\beta}_k(\sigma_k) = \max_{\sigma_{k+1}}^* (\tilde{\gamma}_k(\sigma_k \rightarrow \sigma_{k+1}) + \tilde{\beta}_{k+1}(\sigma_{k+1})). \quad (8.91)$$

and the a posteriori LLR values are

$$L(a_k|\tilde{\mathbf{r}}) = \max_{\substack{\sigma_k \rightarrow \sigma_{k+1} \\ a_k=+1}}^* (\tilde{\alpha}_k(\sigma_k) + \tilde{\gamma}_k(\sigma_k \rightarrow \sigma_{k+1}) + \tilde{\beta}_{k+1}(\sigma_{k+1})) \quad (8.92)$$

$$- \max_{\substack{\sigma_k \rightarrow \sigma_{k+1} \\ a_k=-1}}^* (\tilde{\alpha}_k(\sigma_k) + \tilde{\gamma}_k(\sigma_k \rightarrow \sigma_{k+1}) + \tilde{\beta}_{k+1}(\sigma_{k+1})).$$

Example 8.4. Consider the case of a rate-1/2 recursive systematic convolutional code, for example, the encoder shown in Fig. 8.7. In this case, the transmitted sequence of code symbols is $\tilde{\mathbf{s}} = \{\tilde{\mathbf{s}}_n\}_{n=1}^k$, where $\tilde{\mathbf{s}}_n = (\tilde{s}_{n,s}, \tilde{s}_{n,p})$ and the terms with subscripts s and p correspond to the systematic (information bit) and parity check bit, respectively. For binary modulation, $\tilde{s}_{n,s}, \tilde{s}_{n,p} \in \{-\sqrt{2E_h}, \sqrt{2E_h}\}$, where $E_h = E_c$ is the energy per code bit. Likewise, the sequence of received vectors is $\tilde{\mathbf{r}} = \{\tilde{\mathbf{r}}_n\}_{n=1}^k$, where $\tilde{\mathbf{r}}_n = (\tilde{r}_{n,s}, \tilde{r}_{n,p})$.

Returning to the branch metric in (8.78),

$$\begin{aligned} \gamma_k(\sigma_k \rightarrow \sigma_{k+1}) &= \text{P}[a_k = u] p(\tilde{r}_{k,s}, \tilde{r}_{k,p} | g_{k,s} \tilde{s}_{k,s}, g_{k,p} \tilde{s}_{k,p}) \\ &= \frac{\text{P}[a_k = u]}{(2\pi N_o)^N} \exp \left\{ -\frac{1}{2N_o} (|\tilde{r}_{k,s} - g_{k,s} \tilde{s}_{k,s}|^2 + |\tilde{r}_{k,p} - g_{k,p} \tilde{s}_{k,p}|^2) \right\} \\ &= \frac{1}{(2\pi N_o)^N} \exp \left\{ -\frac{1}{2N_o} (|\tilde{r}_{k,s}|^2 + |\tilde{r}_{k,p}|^2 + 4E_c) \right\} \\ &\quad \times \text{P}[a_k = u] \exp \left\{ \frac{1}{N_o} (\text{Re}\{g_{k,s}^* \tilde{r}_{k,s} \tilde{s}_{k,s}\} + \text{Re}\{g_{k,p}^* \tilde{r}_{k,p} \tilde{s}_{k,p}\}) \right\} \end{aligned} \quad (8.93)$$

Note that the term

$$\frac{1}{(2\pi N_o)^N} \exp \left\{ -\frac{1}{2N_o} (|\tilde{r}_{k,s}|^2 + |\tilde{r}_{k,p}|^2 + 4E_c) \right\}$$

is independent of a_k and will cancel in the numerator and denominator of the a posteriori LLR in (8.85) and, therefore, can be ignored. Also, the numerator in (8.85) has $\tilde{s}_{k,s} = \sqrt{2E_c}$ corresponding to $a_k = +1$, while the denominator has $\tilde{s}_{k,s} = -\sqrt{2E_c}$ corresponding to $a_k = -1$. It follows that the a posteriori LLR in (8.79) becomes

$$\begin{aligned} L(a_k | \tilde{\mathbf{r}}) &= \frac{2\sqrt{2E_c} g_{k,s}^* \tilde{r}_{k,s}}{N_o} + \log \left\{ \frac{\text{P}[a_k = +1]}{\text{P}[a_k = -1]} \right\} \\ &\quad + \log \left\{ \frac{\sum_{\sigma_k \rightarrow \sigma_{k+1}: a_k = +1} \alpha_k(\sigma_k) \cdot \exp \left\{ \frac{\text{Re}\{g_{k,p}^* \tilde{r}_{k,p} \tilde{s}_{k,p}\}}{N_o} \right\} \cdot \beta_{k+1}(\sigma_{k+1})}{\sum_{\sigma_k \rightarrow \sigma_{k+1}: a_k = -1} \alpha_k(\sigma_k) \cdot \exp \left\{ \frac{\text{Re}\{g_{k,p}^* \tilde{r}_{k,p} \tilde{s}_{k,p}\}}{N_o} \right\} \cdot \beta_{k+1}(\sigma_{k+1})} \right\}. \end{aligned} \quad (8.94)$$

Note that the symbols $\tilde{s}_{k,p}$ are not conjugated in the above expression since they are real-valued in this example. If the Log-MAP algorithm is employed, then the a posteriori LLR becomes

$$L(a_k | \tilde{\mathbf{r}}) = L_{\text{sys}}(a_k | \tilde{\mathbf{r}}_s) + L_a(a_k) + L_e(a_k | \tilde{\mathbf{r}}_p), \quad (8.95)$$

where $L_{\text{sys}}(a_k | \tilde{\mathbf{r}}_s)$, $L_a(a_k)$, and $L_e(a_k | \tilde{\mathbf{r}}_p)$ are, respectively, defined as

$$L_{\text{sys}}(a_k | \tilde{\mathbf{r}}_s) = \frac{2\sqrt{2E_c} g_{k,s}^* \tilde{r}_{k,s}}{N_o} \quad (8.96)$$

$$L_a(a_k) = \log \left\{ \frac{\text{P}[a_k = +1]}{\text{P}[a_k = -1]} \right\} \quad (8.97)$$

$$\begin{aligned} L_e(a_k | \tilde{\mathbf{r}}_p) &= \max_{\substack{\sigma_k \rightarrow \sigma_{k+1} \\ a_k = +1}}^* \left(\tilde{\alpha}_k(\sigma_k) + \frac{\text{Re}\{g_{k,p}^* \tilde{r}_{k,p} \tilde{s}_{k,p}\}}{N_o} + \tilde{\beta}_{k+1}(\sigma_{k+1}) \right) \\ &\quad - \max_{\substack{\sigma_k \rightarrow \sigma_{k+1} \\ a_k = -1}}^* \left(\tilde{\alpha}_k(\sigma_k) + \frac{\text{Re}\{g_{k,p}^* \tilde{r}_{k,p} \tilde{s}_{k,p}\}}{N_o} + \tilde{\beta}_{k+1}(\sigma_{k+1}) \right). \end{aligned} \quad (8.98)$$

(continued)

Example 8.4 (continued)

The a posteriori LLR consists of three terms; the first depends on the channel output due to the systematic component, the second depends on the a priori probabilities of the information bits, and the third depends on the channel outputs due to the parity bits. Usually $P[a_k = +1] = P[a_k = -1] = 1/2$ for convolutional decoders, so that the a priori term $L_a(a_k)$ is zero. However, for the iterative decoders that are used with the turbo codes discussed later in this chapter, the decoder will receive *extrinsic* or *soft* information for each a_k which serves as a priori information. Once the a posteriori LLR has been calculated, the extrinsic information can be calculated as

$$L_e(a_k|\tilde{\mathbf{r}}_p) = L(a_k|\tilde{\mathbf{r}}) - L_{\text{sys}}(a_k|\tilde{\mathbf{r}}_s) - L_a(a_k). \quad (8.99)$$

8.3 Trellis Coded Modulation

8.3.1 Encoder Description

Conventional convolutional codes realize a coding gain at the expense of the coded modulation efficiency or bits/s/Hz. Although convolutional codes may be attractive for power-limited applications, they are less suitable for bandwidth-limited applications. One approach for overcoming the loss of coded modulation efficiency is to map the encoder output bits onto a higher-order signal constellation, such as M -PSK or M -QAM. However, the symbol mapping that is used in this case is very critical. Employing a straightforward mapping, such as Gray mapping or natural mapping, will give disappointing results because the decreased Euclidean distance between the signals points in the higher-order constellation will tend to offset the benefits gained from using the convolutional code. 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 [325]. 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.
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.

Figure 8.9 shows the basic encoder structure for Ungerboeck's trellis codes. The n -bit information vector $\mathbf{a}_k = (a_k^{(1)}, \dots, a_k^{(n)})$ is transmitted at epoch k . At each epoch k , $m \leq n$ data 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

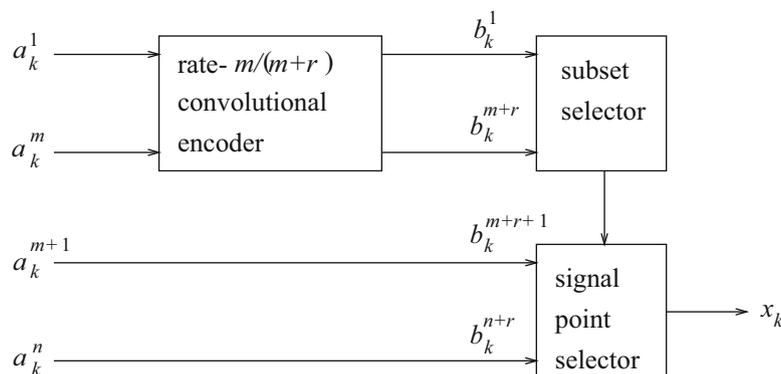


Fig. 8.9 Ungerboeck trellis encoder

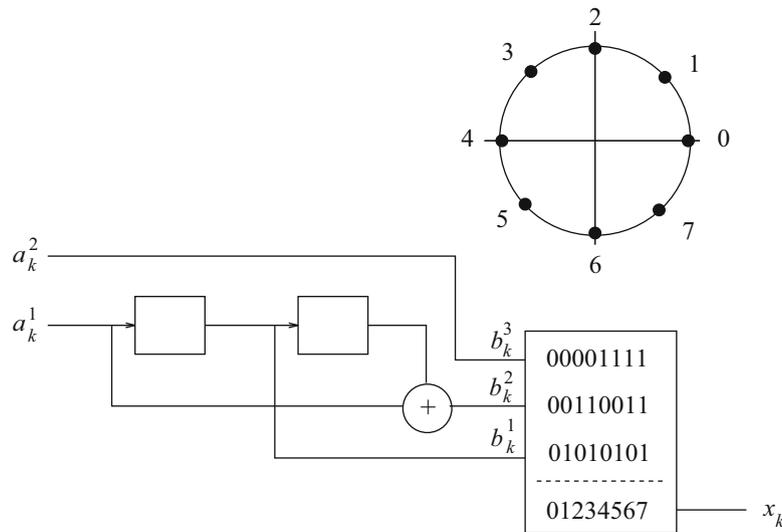


Fig. 8.10 Encoder and signal mapping for the 4-state 8-PSK Ungerboeck trellis code

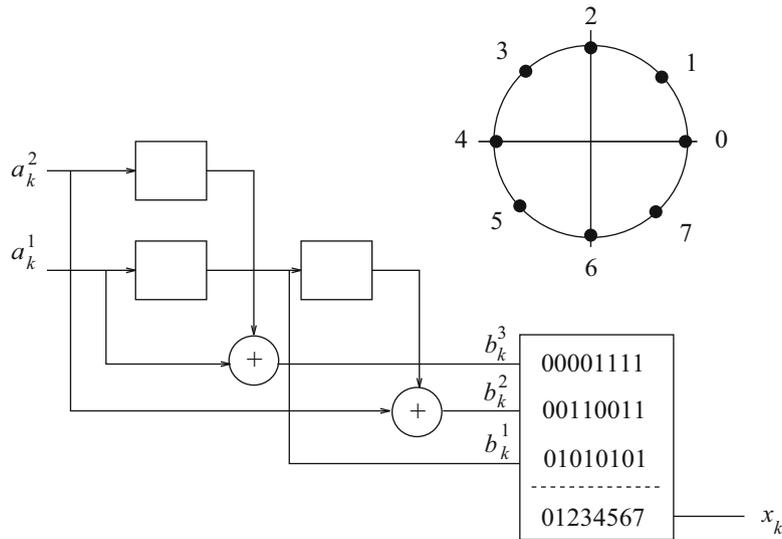


Fig. 8.11 Encoder and signal mapping for the 8-state 8-PSK Ungerboeck trellis code

constellation. The remaining $n - m$ data bits are used to select one of the 2^{n-m} signal points within the selected subset. This principle is best explained by example, and Fig. 8.10 shows a 4-state 8-PSK Ungerboeck trellis code. The equivalent uncoded system is 4-PSK which has a 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. Figure 8.11 shows another example of 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.

8.3.2 Mapping by Set Partitioning

The critical step in the design of Ungerboeck's codes is the method of mapping by set partitioning. Figure 8.12 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. Here a normalized 8-PSK signal constellation is assumed having eight signal

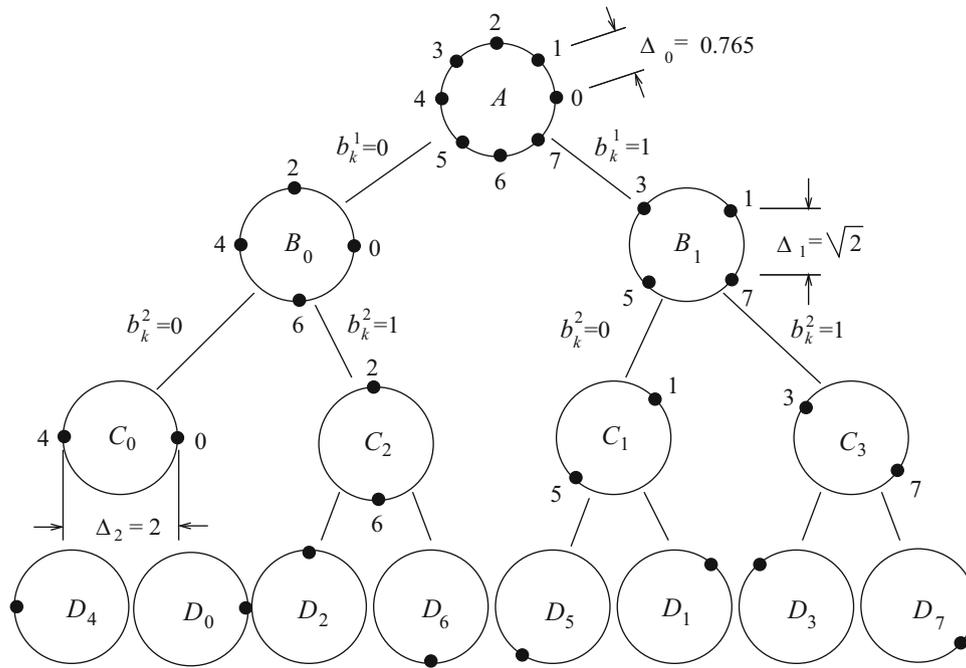


Fig. 8.12 Set partitioning for an 8-PSK signal constellation

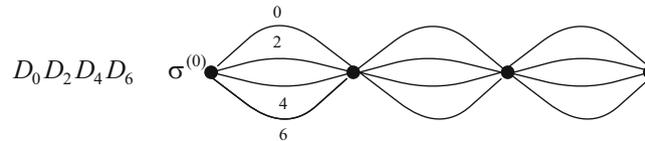


Fig. 8.13 Trellis diagram for uncoded 4-PSK

points uniformly spaced around a circle of unit radius. Notice that the minimum Euclidean distance between signal points in the normalized 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. Notice that 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. 8.13. The trellis only has one state and there are four parallel 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. 8.14. 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. 8.12, 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 with a Euclidean distance of $d = 2$. Also, any two signal paths that diverge from a state and merge with the same state after more than 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 Euclidean 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 $d_{\min} = 2$.

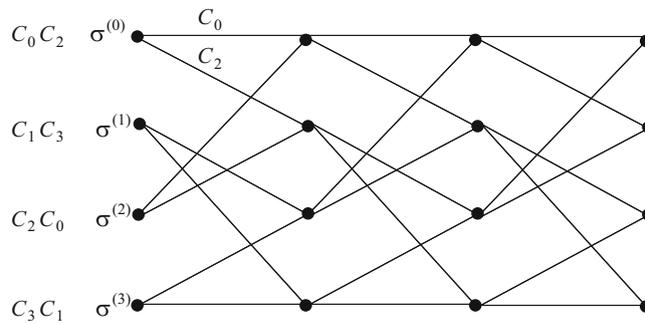


Fig. 8.14 Trellis diagram for 4-state 8-PSK Ungerboeck trellis code

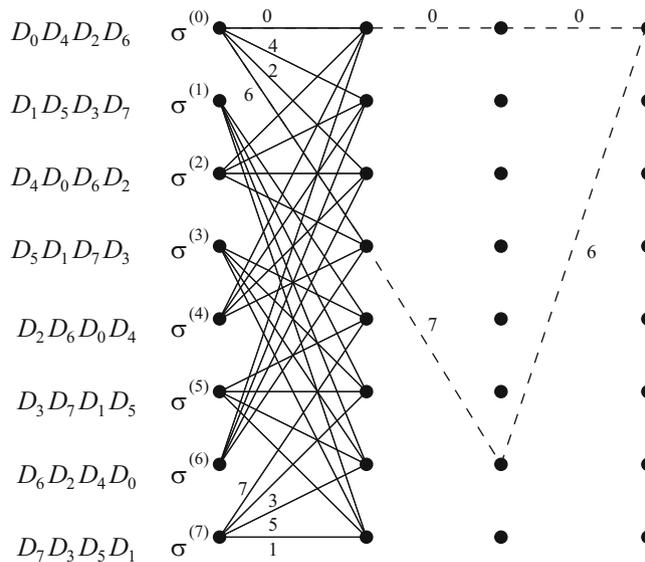


Fig. 8.15 Trellis diagram for 8-state 8-PSK Ungerboeck trellis code. The dashed lines show two minimum distance paths

The concept of mapping by set partitioning was developed by Ungerboeck as a method for maximizing the minimum Euclidean distance of a code and consequently to optimize its performance on an AWGN channel. Ungerboeck’s construction of the optimum 4-state 8-PSK code was based on the following heuristic rules [326];

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.

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. 8.15. In this case, the minimum Euclidean distance is $d_{\min} = \sqrt{\Delta_1^2 + \Delta_0^2 + \Delta_1^2} = 2.141$.

8.4 Code Performance on AWGN Channels

Viterbi originally exploited the trellis structure of convolutional codes and developed the Viterbi algorithm for ML decoding of convolutional codes [334, 335]. 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 as discussed in Sect. 7.4.1 [131]. Given the similarity between the trellis structures of ISI channels, convolutional codes, and

trellis codes (e.g., compare Figs. 7.11, 8.6 and 8.14), it is not surprising that the union bounding techniques used to evaluate the error probability of digital signaling on ISI channels with an MLSE receiver in Sect. 7.5 can be applied, with appropriate modification, to evaluate the error probability of convolutional and trellis codes with an MLSE receiver.

To develop the union bound on decoded bit error probability, let $\mathbf{a} = \{\mathbf{a}_k\}$ 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 that have the first non-zero element starting at time j_1 . Then, the average bit error probability is bounded by

$$P_b \leq \frac{1}{n} \sum_{\mathbf{e} \in E} w_b(\mathbf{e}) \sum_{\mathbf{a}} P[\mathbf{a}] P\left[\Gamma(\mathbf{a} \oplus \mathbf{e}) \geq \Gamma(\mathbf{a}) \mid \mathbf{a}\right], \quad (8.100)$$

where $\Gamma(\mathbf{a})$ is the path metric of \mathbf{a} , and $w_b(\mathbf{e})$ is the number of bit errors associated with \mathbf{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-coded modulation because the signal mapping and, hence, the codes are nonlinear.

Another way of writing the bound on the bit error probability in (8.100) is

$$P_b \leq \sum_{\tilde{\mathbf{s}} \in C} \sum_{\hat{\mathbf{s}} \in C} w_b(\tilde{\mathbf{s}}, \hat{\mathbf{s}}) P[\tilde{\mathbf{s}}] P[\tilde{\mathbf{s}} \rightarrow \hat{\mathbf{s}}], \quad (8.101)$$

where C is the set of all coded symbol sequences, $w_b(\tilde{\mathbf{s}}, \hat{\mathbf{s}})$ is the number of bit errors that occur when the complex symbol sequence $\tilde{\mathbf{s}} = \{\tilde{s}_i\}$ is transmitted and the complex symbol sequence $\hat{\mathbf{s}} \neq \tilde{\mathbf{s}}$ is chosen by the decoder, $P[\tilde{\mathbf{s}}]$ is the a priori probability of transmitting $\tilde{\mathbf{s}}$, and $P[\tilde{\mathbf{s}} \rightarrow \hat{\mathbf{s}}]$ is the pairwise error probability.

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 symbol sequences $\tilde{\mathbf{s}}$ and $\hat{\mathbf{s}}$ separated by Euclidean distance d_{\min} is

$$P[\tilde{\mathbf{s}} \rightarrow \hat{\mathbf{s}}] = Q\left(\sqrt{\frac{d_{\min}^2}{4N_o}}\right). \quad (8.102)$$

The asymptotic coding gain (at high signal-to-noise ratio) is defined by [43]

$$G_a = 10 \log_{10} \frac{(d_{\min, \text{coded}}^2 / E_{\text{av, coded}})}{(d_{\min, \text{uncoded}}^2 / E_{\text{av, uncoded}})} \text{ dB} \quad (8.103)$$

where E_{av} is the average energy per symbol in the signal constellation. For the 4-state 8-PSK code shown in Fig. 8.10, the asymptotic coding gain is $G_a = 3$ dB over uncoded 4-PSK. Likewise, the 8-state 8-PSK code in Fig. 8.11 has an asymptotic coding gain of 3.6 dB over uncoded 4-PSK.

8.4.1 Union Bound for Convolutional Codes

For convolutional codes the upper bound in (8.100) 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 [203]. Because of this property, the all-zeroes sequence can be assumed to be transmitted, i.e., $\mathbf{a} = \mathbf{0}$, so that the union bound becomes

$$P_b \leq \frac{1}{k} \sum_{\mathbf{e} \in E} w_b(\mathbf{e}) P\left[\Gamma(\mathbf{e}) \geq \Gamma(\mathbf{0})\right]. \quad (8.104)$$

Note that division is 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.

For convolutional codes the set E in (8.104) 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 $P[\Gamma(\mathbf{e}) \geq \Gamma(\mathbf{0})]$ for that path and dividing by the number of input bits per branch, k , gives the bit error probability associated with that path. Summing over the set of all possible incorrect sequences E yields a union bound on the bit error probability.

The pairwise error probability in (8.104) 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. Assuming an AWGN channel and a coherent receiver, the received vector (see Sect. 5.1) at epoch n is

$$\tilde{\mathbf{r}}_n = \tilde{\mathbf{s}}_n + \tilde{\mathbf{n}}_n, \quad (8.105)$$

where $\tilde{\mathbf{s}}_n$ is the transmitted symbol vector and $\tilde{\mathbf{n}}_n$ is the Gaussian noise vector at epoch n . For convolutional codes, two types of decoding are commonly used, hard decision decoding and soft decision decoding. Soft decision decoders use the sequence of received signal vectors $\mathbf{r} = \{\mathbf{r}_n\}$ to make sequence decisions. For an AWGN channel, the MLSE receiver searches for the sequence of symbol vectors $\tilde{\mathbf{s}} = \{\tilde{\mathbf{s}}_n\}$ that is closest in Euclidean distance to the received sequence of signal vectors \mathbf{r} . To do so, the MLSE receiver chooses the sequence $\hat{\mathbf{s}}$ that minimizes the metric

$$\mu(\hat{\mathbf{s}}) = \|\mathbf{r} - \hat{\mathbf{s}}\|^2. \quad (8.106)$$

The decided sequence $\hat{\mathbf{s}}$ maps one-to-one onto the data bit sequence $\hat{\mathbf{a}}$ that is the final estimate of the transmitted information sequence.

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

$$P_2(\ell) = Q\left(\sqrt{\frac{\Delta^2}{4N_o}}\right), \quad (8.107)$$

where

$$\Delta^2 = \sum_{k=k_1}^{k_1+\ell+1} \delta_k^2 \quad (8.108)$$

$$\delta_k^2 = \|\tilde{\mathbf{s}}_k - \hat{\mathbf{s}}_k\|^2 \quad (8.109)$$

and $\tilde{\mathbf{s}} = \{\tilde{\mathbf{s}}_k\}$ and $\hat{\mathbf{s}} = \{\hat{\mathbf{s}}_k\}$ are the symbol sequences corresponding to the data bit sequences $\tilde{\mathbf{a}}$ 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 $\tilde{\mathbf{b}}$ and $\hat{\mathbf{b}}$ that differ in d positions is

$$P_2(d) = Q(\sqrt{2R_c d \gamma_b}) \quad (8.110)$$

where γ_b is the received bit energy-to-noise ratio.² Note that the pairwise error probability has been explicitly shown to be a function of the Hamming distance between the codewords in (8.110). However, it is important to realize that this property does not always apply. For example, if the outputs of the rate-2/3 convolutional encoder in Fig. 8.2 are mapped onto symbols from an 8-PSK signal constellation, then 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.

²The received symbol energy-to-noise ratio is $\gamma_s = R_c \gamma_b$.

In general, the transfer function $T(D, N)$ for a convolutional code has the form

$$T(D, N) = \sum_{d=d_{\text{free}}}^{\infty} a_d D^d N^{f(d)}, \quad (8.111)$$

where $f(d)$ is the exponent of N as a function of d . For the example in (8.50), $a_d = 2^{d-5}$ and $f(d) = d - 4$. Differentiating $T(D, N)$ with respect to N and setting $N = 1$ gives

$$\left. \frac{dT(D, N)}{dN} \right|_{N=1} = \sum_{d=d_{\text{free}}}^{\infty} a_d f(d) D^d. \quad (8.112)$$

Once again, for the example in (8.50) this leads to

$$\left. \frac{dT(D, N)}{dN} \right|_{N=1} = \sum_{d=d_{\text{free}}}^{\infty} 2^{d-5} (d-4) D^d. \quad (8.113)$$

Using this notation, the union bound on bit error probability for convolutional coding with BPSK modulation is

$$P_b \leq \frac{1}{k} \sum_{d=d_{\text{free}}}^{\infty} a_d f(d) P_2(d), \quad (8.114)$$

where $P_2(d)$ is given by (8.110).

In contrast to soft decision decoders, hard decision decoders first make symbol-by-symbol decisions on the sequence of received vectors $\mathbf{r} = \{\mathbf{r}_k\}$ to yield the sequence of symbol decisions $\tilde{\mathbf{s}} = \{\tilde{s}_k\}$. The decoder then operates on the sequence $\tilde{\mathbf{s}}$ to estimate the most likely transmitted data sequence. A minimum distance decoder is one that decides in favor of the symbol sequence $\hat{\mathbf{s}}$ that is closest in Hamming distance to the received symbol sequence $\tilde{\mathbf{s}}$. 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 $\tilde{\mathbf{b}}$ and $\hat{\mathbf{b}}$ at Hamming distance d is

$$P_2(d) = \begin{cases} \sum_{k=(d+1)/2}^d \binom{d}{k} p^k (1-p)^{d-k}, & d \text{ odd} \\ \sum_{k=d/2+1}^d \binom{d}{k} p^k (1-p)^{d-k} + \frac{1}{2} \binom{d}{d/2} p^{d/2} (1-p)^{d/2}, & d \text{ even} \end{cases} \quad (8.115)$$

where

$$p = Q(\sqrt{2R_c \gamma_b}) \quad (8.116)$$

is the probability of symbol error. Once again, the pairwise error probability for BPSK is a function of the Hamming distance between the codewords.

8.4.1.1 Union-Chernoff Bound for Convolutional Codes

The union bound in (8.114) 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 $\tilde{\mathbf{s}}$ is transmitted and \mathbf{r} is the received sequence. Then the pairwise error probability between sequences $\tilde{\mathbf{s}}$ and $\hat{\mathbf{s}}$ with a maximum likelihood receiver can be Chernoff bounded by

$$\begin{aligned} P[\tilde{\mathbf{s}} \rightarrow \hat{\mathbf{s}}] &= P[\|\mathbf{r} - \hat{\mathbf{s}}\|^2 < \|\mathbf{r} - \tilde{\mathbf{s}}\|^2] \\ &\leq \mathbf{E} \left[e^{\lambda(\|\mathbf{r} - \tilde{\mathbf{s}}\|^2 - \|\mathbf{r} - \hat{\mathbf{s}}\|^2)} \mid \tilde{\mathbf{s}} \right]. \end{aligned} \quad (8.117)$$

Substituting $\mathbf{r} = \tilde{\mathbf{s}} + \mathbf{n}$, taking the expectation over the complex Gaussian random vector \mathbf{n} , and simplifying gives

$$P[\tilde{\mathbf{s}} \rightarrow \hat{\mathbf{s}}] \leq e^{-\lambda \|\tilde{\mathbf{s}} - \hat{\mathbf{s}}\|^2 (1 - \lambda 2N_o)}. \quad (8.118)$$

The tightest upper bound is obtained with $\lambda^* = 1/(4N_o)$ yielding

$$P[\tilde{\mathbf{s}} \rightarrow \hat{\mathbf{s}}] \leq e^{-\|\tilde{\mathbf{s}} - \hat{\mathbf{s}}\|^2 / 8N_o}. \quad (8.119)$$

For the case of BPSK signaling on an AWGN channel, the Chernoff bound on the pairwise error probability becomes

$$P_2(d) \leq e^{-d\gamma_s} = e^{-R_c d \gamma_b}, \quad (8.120)$$

where d is the number of coordinates in which the sequences $\tilde{\mathbf{s}}$ and $\hat{\mathbf{s}}$ differ, $\gamma_s = E_h/N_o$ is the received symbol energy-to-noise ratio, R_c is the code rate, and γ_b is the received bit energy-to-noise ratio. Likewise, if BPSK signaling is used with hard decision decoding, it can be shown that the pairwise error probability has the Chernoff bound

$$P_2(d) \leq [4p(1-p)]^{d/2}, \quad (8.121)$$

where the code bit error probability p is given by (8.116).

Notice how the Hamming distance d appears in the exponent of the Chernoff bound on pairwise error probability with either hard or soft decision decoding. From (8.112) and (8.114), it is apparent that the decoded bit error probability has the following union-Chernoff bound:

$$P_b \leq \frac{1}{k} \frac{dT(D, N)}{dN} \Big|_{N=1, D=Z}, \quad (8.122)$$

where

$$Z = \begin{cases} \sqrt{4p(1-p)}, & \text{hard decision decoding} \\ e^{-R_c \gamma_b}, & \text{soft decision decoding} \end{cases}. \quad (8.123)$$

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 decoded bit error probability with BPSK at high SNR is approximately

$$P_b \approx \frac{1}{k} a_{d_{\text{free}}} f(d_{\text{free}}) P_2(d_{\text{free}}) \leq \frac{1}{k} a_{d_{\text{free}}} f(d_{\text{free}}) Z^{d_{\text{free}}}. \quad (8.124)$$

The above procedure for upper bounding the decoded bit error probability is sometimes called the transfer function approach, because it uses the transfer function of the augmented state diagram. However, the transfer function approach has its limitations. First, as the number of encoder states becomes large, it becomes difficult to compute the transfer function $T(D, N)$. Second, for non-binary signal constellations, the pairwise error probability will depend on the particular pair of modulated symbol sequences, and not just on the Hamming distance between them. In this case, 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 using a different approach to compute the upper bound, such as a stack algorithm [301].

8.5 Interleaving

Most error control codes are designed for memoryless channels, where successively transmitted code symbols experience independent channel conditions. Although this memoryless property holds for AWGN channels, it is not usually the case for wireless channels due to the temporal correlation of the channel, as discussed in Chap. 2. An effective method for dealing with such channels is to use interleaving. There are two basic approaches to interleaving, namely symbol interleaving and bit interleaving. A symbol interleaver performs symbol-by-symbol interleaving of the code symbols *after* they have been mapped onto a signal constellation. A bit interleaver, on the other hand, performs bit-by-bit interleaving of the encoder outputs *before* they are mapped onto a signal constellation.

The interleaving/deinterleaving operation serves to reduce the correlation between the fades experienced by successive code bits at the output of the encoder. Since, most error control codes are designed for memoryless channels, this will generally improve the decoded bit error rate performance. Indeed if the interleaver/deinterleaver were eliminated and the

channel fades slowly, it is possible that entire codewords could be received with either a low or high signal-to-noise ratio. Under a low instantaneous signal-to-noise ratio condition even a very low-rate code will fail, while under a high instantaneous signal-to-noise ratio condition even very high-rate codes will succeed. Under such conditions, the error control code is ineffective.

8.5.1 Block Interleaving

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. Such an interleaver is called a (J, M) block interleaver, and the length of the interleaver is JM . The block interleaver can be used to perform either symbol interleaving or bit interleaving. In the case of symbol (bit) interleaving, the code symbols (bits) are input to the buffer row-wise from top to bottom and output from the buffer column-wise from left to right. The deinterleaver simply performs the reverse operation. The block interleaver has the following characteristics:

1. Any burst of symbol (bit) errors of length $j \leq J$ results in single symbol (bit) errors at the deinterleaver output that are each separated by at least M symbols (bits).
2. Any burst of symbol (bit) errors of length $j = kJ, k > 1$ results in bursts of no more than $\lceil k \rceil$ symbol (bit) errors that are separated by no less than $M - \lceil k \rceil$ symbols (bits).
3. A periodic sequence of single symbol (bit) errors that are spaced J code symbols (bits) apart results in a single burst of symbol (bit) errors of length M at the deinterleaver output.
4. The end-to-end interleaving delay is equal to $2JM$ code symbols (bits), and the memory requirements for the interleaver and deinterleaver is equal to JM code symbols (bits).

In practice, the interleaver depth J should be chosen so that successive code symbols (bits), which are transmitted J code symbol (bit) durations apart, are independently faded. Chapter 2 showed that the temporal autocorrelation function of a cellular land mobile radio channel observed at either the mobile station or base station with 2-D isotropic scattering around the mobile station is $\phi_{gg}(\tau) = \frac{\sigma_p}{2} J_0(2\pi f_m \tau)$, where $f_m = v/\lambda_c$ with v being the mobile station velocity and λ_c the carrier wavelength. From Fig. 2.5, observe that the temporal autocorrelation is small provided that $f_m \tau > 0.5$. Hence, effective interleaving should have $J > 0.5\lambda_c/vT$, where T is the code symbol (bit) duration. The choice of the parameter M depends on the type of coding used. For block codes M should be larger than the block length. For convolutional and trellis codes, the interleaver should separate any $L_D + 1$ successive code symbols (bits) as far apart as possible, where L_D is the decoding depth. Hence, the interleaver should have $M \geq L_D + 1$. For convolutional and trellis codes, it is well known that the decoding depth $L_D \approx 5K$, where K is the constraint length.

Observe that the required interleaving depth is inversely proportional to the speed of the mobile station and, therefore, slower moving mobile stations require larger interleaving depths. Unfortunately, interleaving introduces delay into the link, since the $J \times M$ interleaver array must be filled before the symbols (bits) can be transmitted, and afterwards the $J \times M$ deinterleaver array must be filled before the symbols (bits) can be decoded. However, delay critical traffic such as real-time voice will impose a limit on the interleaving delay defined as $t_d = JMT$ and, therefore, the interleaving will be insufficient at low speeds. For example, if $f_c = 900$ MHz, $R = 1/T = 24 \times 10^3$, and $v = 30$ km/h, then $J > 478$ is required. For a constraint length $K = 3$ convolutional code with a decoding depth $L_D = 5K = 15$, the minimum required interleaving delay satisfies

$$t_d = JMT > 0.5\lambda_c(L_D + 1)/v = 319 \text{ ms.}$$

Such a delay is quite large, especially for real-time voice applications, and the problem is exasperated by lower mobile station speeds. One possible solution is to use a code with a smaller decoding depth L_D . The other solution is to use improved interleaving techniques to reduce the interleaving delay. Once such interleaver is the convolutional interleaver discussed below. However, at low speeds the required interleaving delay may be excessive with any type of interleaver. Under such conditions, adaptive closed-loop power control techniques are effective for combatting fading.

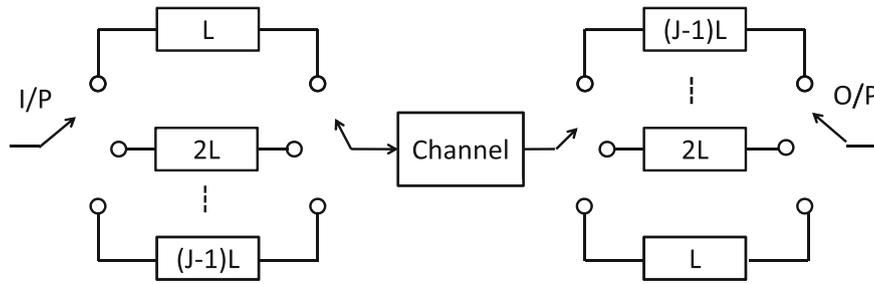


Fig. 8.16 Convolutional interleaver implemented with shift registers

8.5.2 Convolutional Interleaving

Convolutional interleavers have been proposed by Ramsey [279] and Forney [130], and the structure proposed by Forney is discussed below. Defining the parameter $M = LJ$, the interleaver is referred to as a (J, M) interleaver and has properties that are similar to a (J, M) block interleaver.

As shown in Fig. 8.16, the code symbols (bits) are input sequentially into the bank of J registers of increasing lengths. With each new code symbol (bit), the commutator switches to the next register, the code symbol (bit) is input to the register and the oldest code symbol (bit) in that register is shifted out to the channel. The input and output commutators of the interleaver operate in a synchronous fashion. The deinterleaver performs the reverse operation, and the deinterleaving commutators must be synchronized with the interleaving commutators. The most important properties of the convolutional interleaver are as follows:

1. The minimum separation at the interleaver output is J symbols (bits) for any two symbols that are separated by less than M symbols (bits) at the interleaver output.
2. Any burst of symbol (bit) errors of length $j \leq J$ results in single symbol (bit) errors at the deinterleaver output that are each separated by at least M symbols (bits).
3. A periodic sequence of single symbol (bit) errors that are spaced $M + 1$ code symbols (bits) apart results in a single burst of symbol (bit) errors of length J at the deinterleaver output.
4. The end-to-end interleaving delay is equal to $M(J - 1)$ code symbols (bits), and the memory requirements for both the interleaver and deinterleaver are equal to $M(J - 1)/2$ code symbols (bits). This is half the delay and memory requirement of a (J, M) block interleaver.

The parameters J and M are chosen in the same manner as for a block interleaver, and the performance is very similar.

8.6 Code Performance on Interleaved Flat Fading Channels

8.6.1 TCM with Symbol Interleaving

Figure 8.17 is a block diagram of a coded communication system operating on a flat fading channel with symbol interleaving. The information sequence \mathbf{a} is encoded and mapped onto a signal set to generate the modulated symbol sequence \mathbf{s} by using either a convolutional code or trellis code. The modulated symbol sequence is then time interleaved (or scrambled) and the resulting sequence $\check{\mathbf{s}}$ is transmitted over the channel. If a coherent detection is used, the receiver uses a correlator or matched filter detector as discussed in Sect. 5.2 to generate the sequence of received vectors $\mathbf{r} = \{\mathbf{r}_k\}$. With hard decision decoding, the received vector \mathbf{r}_k at each epoch k is applied to a decision device to yield an estimate of the transmitted symbol sequence $\hat{\mathbf{s}}$. The estimated symbol sequence $\hat{\mathbf{s}}$ is then deinterleaved and input to the decoder. With soft decision decoding on the other hand, the received vector \mathbf{r} is deinterleaved and applied directly to the decoder.

For analytical purposes, an infinite interleaving depth is often assumed so that the deinterleaved sequence of complex channel gains $\mathbf{g} = \{g_k\}$ constitutes a sequence of independent random variables. In this case the conditional density of \mathbf{r} has the product from

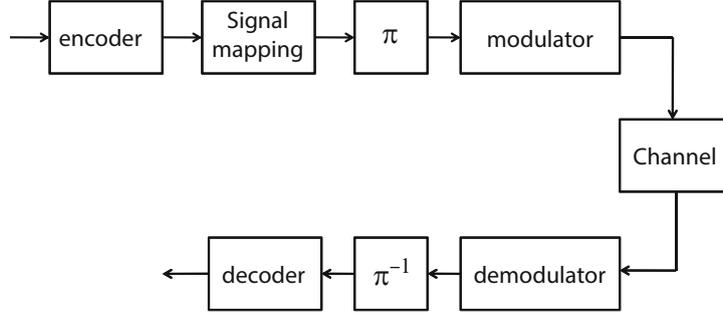


Fig. 8.17 TCM with symbol interleaving on a flat fading channel

$$p(\mathbf{r}|\mathbf{g} \cdot \tilde{\mathbf{s}}) = \prod_k p(\mathbf{r}_k|g_k \tilde{\mathbf{s}}_k), \quad (8.125)$$

where $\mathbf{g} \cdot \tilde{\mathbf{s}}$ is the vector dot product of \mathbf{g} and $\tilde{\mathbf{s}}$. Suppose that sequence $\tilde{\mathbf{s}}$ is transmitted and the sequence $\mathbf{r} = \mathbf{g} \cdot \tilde{\mathbf{s}} + \mathbf{n}$ is received. An ML receiver having perfect knowledge of \mathbf{g} chooses the sequence $\hat{\mathbf{s}}$ that minimizes the squared Euclidean distance metric

$$\mu(\hat{\mathbf{s}}) = \|\mathbf{r} - \mathbf{g} \cdot \hat{\mathbf{s}}\|^2. \quad (8.126)$$

The pairwise error probability between the sequences $\tilde{\mathbf{s}}$ and $\hat{\mathbf{s}}$ has the Chernoff bound (see Appendix A)

$$P[\tilde{\mathbf{s}} \rightarrow \hat{\mathbf{s}}] \leq \exp \left\{ -\frac{\|\mathbf{g} \cdot (\tilde{\mathbf{s}} - \hat{\mathbf{s}})\|^2}{8N_o} \right\}. \quad (8.127)$$

If the normalization $E[|x_k|^2] = 1$ is assumed so that $E_s = \frac{1}{2}E[\|\mathbf{s}_k\|^2] = E_s E[|x_k|^2]$, i.e., $\mathbf{s}_k = \sqrt{2E_s}x_k$, then the Chernoff bound becomes

$$P[\tilde{\mathbf{s}} \rightarrow \hat{\mathbf{s}}] \leq \exp \left\{ -\frac{E_s}{4N_o} \|\mathbf{g} \cdot (\tilde{\mathbf{x}} - \hat{\mathbf{x}})\|^2 \right\}, \quad (8.128)$$

where E_s is the symbol energy.

The channel is assumed to be characterized by flat Ricean fading, so that the pdf of α_k is

$$p_{\alpha}(x) = \frac{2x(1+K)}{\Omega_p} \exp \left\{ -K - \frac{(K+1)x^2}{\Omega_p} \right\} I_0 \left(2x \sqrt{\frac{K(K+1)}{\Omega_p}} \right) \quad (8.129)$$

where $\Omega_p = E[\alpha_k^2]$ is the average envelope power. Averaging (8.128) over the probability density function in (8.129) gives [94]

$$P[\mathbf{s} \rightarrow \hat{\mathbf{s}}] \leq \prod_{i \in A} \frac{1+K}{1+K + \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2} \exp \left\{ -\frac{K \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2}{1+K + \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2} \right\} \quad (8.130)$$

where $\bar{\gamma}_s = E[\alpha^2]E_s/N_o$ is the average received symbol energy-to-noise ratio, and $A = \{i|x_i \neq \hat{x}_i\}$. At sufficiently high $\bar{\gamma}_s$, (8.130) simplifies to

$$P[\mathbf{s} \rightarrow \hat{\mathbf{s}}] \leq \prod_{i \in A} \frac{4(1+K)}{\bar{\gamma}_s|x_i - \hat{x}_i|^2} e^{-K}. \quad (8.131)$$

It follows that the bound in (8.100) will be dominated by the error event path having the smallest number of elements in set A . Divsalar and Simon [94, 95] 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

$$P_b \simeq C \left(\frac{(1+K)e^{-K}}{\bar{\gamma}_s} \right)^{L_{\min}}, \quad \bar{\gamma}_s \gg K \quad (8.132)$$

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 [347] 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 (8.130) can be written in the form

$$P[\mathbf{s} \rightarrow \hat{\mathbf{s}}] \leq e^{-\frac{\bar{\gamma}_s}{4} d^2}, \quad (8.133)$$

where

$$\begin{aligned} d^2 &= \sum_{i \in A} \frac{|x_i - \hat{x}_i|^2 K}{1 + K + \frac{\bar{\gamma}_s}{4} |x_i - \hat{x}_i|^2} + \left(\frac{\bar{\gamma}_s}{4} \right)^{-1} \ln \left(\frac{1 + K + \frac{\bar{\gamma}_s}{4} |x_i - \hat{x}_i|^2}{1 + K} \right) \\ &= \sum_{i \in A} d_{1i}^2 + d_{2i}^2 \end{aligned} \quad (8.134)$$

Two special cases are associated with (8.134), $K = \infty$ and $K = 0$. For $K = \infty$ (no fading),

$$d_{1i}^2 = |x_i - \hat{x}_i|^2, \quad d_{2i}^2 = 0 \quad (8.135)$$

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),

$$d_{1i}^2 = 0, \quad d_{2i}^2 = \left(\frac{\bar{\gamma}_s}{4} \right)^{-1} \ln \left(1 + \frac{\bar{\gamma}_s}{4} |x_i - \hat{x}_i|^2 \right). \quad (8.136)$$

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

$$P[\mathbf{s} \rightarrow \hat{\mathbf{s}}] \leq \left(\prod_{i \in A} \frac{\bar{\gamma}_s}{4} |x_i - \hat{x}_i|^2 \right)^{-1} \quad (8.137)$$

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,

$$\min_{\mathbf{x}, \hat{\mathbf{x}}} \prod_{i \in A} |x_i - \hat{x}_i|^2 \quad (8.138)$$

is another design parameter for Rayleigh fading channels. For values of K between 0 and ∞ , the equivalent squared Euclidean distance in (8.134) 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 the metric in (8.126), the bit error probability at high SNR is approximately,

$$P_b \simeq C_1 E \left[e^{-\gamma_s d_{\min}^2 / 4} \right], \quad (8.139)$$

where C_1 is a constant, $\gamma_s = \alpha^2 E_s / N_o$ is the received symbol energy-to-noise ratio, d_{\min}^2 is the minimum Euclidean distance of the code, and the expectation is with respect to the density in (8.129). Taking this average gives

$$P_b \simeq C_1 \frac{1 + K}{1 + K + d_{\min}^2 \frac{\bar{\gamma}_s}{4} |x_i - \hat{x}_i|^2} \exp \left\{ - \frac{K d_{\min}^2 \frac{\bar{\gamma}_s}{4} |x_i - \hat{x}_i|^2}{1 + K + d_{\min}^2 \frac{\bar{\gamma}_s}{4} |x_i - \hat{x}_i|^2} \right\} \quad (8.140)$$

which can be approximated at large $\bar{\gamma}_s$ by

$$P_b \simeq 4C_1 \frac{1 + K}{d_{\min}^2 \bar{\gamma}_s} e^{-K}. \quad (8.141)$$

Observe that without interleaving, P_b is asymptotically inverse linear with $\bar{\gamma}_s$, independent of the trellis code. It follows that interleaving is required to achieve diversity with TCM on a flat fading channel.

8.6.1.1 Design Rules for Symbol Interleaved TCM on Flat Fading Channels

According to the previous section, when symbol interleaved TCM is used on Ricean fading channel, 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 can be 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.

In general, the following guidelines are followed when designing trellis codes for symbol 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 [325], but now the subsets used may be different. The fifth rule is used to reduce the decoding depth of the code. Wei [347] 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 $\tilde{\mathbf{x}} = \{\tilde{x}_k\}$, which originates from the same state as another valid sequence $\mathbf{x} = \{x_k\}$ and merges into the same last state as \mathbf{x} and whose Hamming distance from \mathbf{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 differs 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 symbol interleaver should separate the output symbols corresponding to each sequence of MDD1+1 consecutive input symbols as far apart as possible.

MDD2 is defined as the length of the longest unmerged valid sequence of signal points, say $\check{\mathbf{x}}$, which originates from the same state as another valid sequence, say \mathbf{x} , and whose Hamming distance from \mathbf{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 symbol interleaver should separate the output symbols corresponding to each sequence of MDD2+1 consecutive input symbols as far apart as possible.

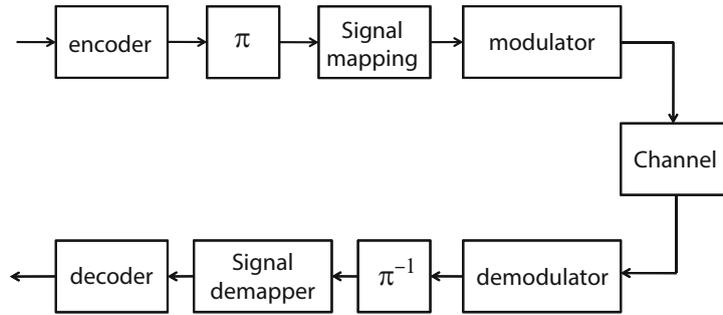


Fig. 8.18 Bit interleaved coded modulation (BICM) on a flat fading channel

8.6.2 Bit Interleaved Coded Modulation

Bit interleaved coded modulation (BICM) interleaves the code bits at the output of a convolutional encoder prior to symbol mapping, along with an appropriately defined soft-decision metric as an input to the Viterbi decoder. The fundamental idea of BICM is to separate the encoder and modulator to make the code diversity equal to the number of distinct *bits* rather than the number of distinct modulated *symbols* along any error event. Such an approach cannot achieve optimum Euclidean distance. However, it will yield better performance on fading channels than TCM with symbol interleaving due to the increased code diversity. The idea is to transform the channel that is generated by a multilevel constellation of size $M = 2^k$ into a parallel collection of k channels that each carries one binary symbol from the signal mapping. To make the k channels independent, they are interleaved prior to the signal mapping. The decoder in a BICM system must reflect the fact that the bits are interleaved prior to signal mapping. The basic structure of a BICM system is shown in Fig. 8.18.

Suppose that the length- n codeword

$$\mathbf{s} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n) \quad (8.142)$$

is transmitted and the sequence

$$\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) \quad (8.143)$$

is received at the output of the bank of receiver matched filters or correlators. With TCM on a memoryless channel, decoding is performed by using the log-likelihood metric

$$\mu(\hat{\mathbf{s}}) = \log\{p(\mathbf{r}|\mathbf{g} \cdot \hat{\mathbf{s}})\} = \log\left\{\prod_{k=1}^n p(\mathbf{r}_k|g_k \hat{\mathbf{s}}_k)\right\} = \sum_{k=1}^n \log\{p(\mathbf{r}_k|g_k \hat{\mathbf{s}}_k)\}. \quad (8.144)$$

When BICM is used, on the other hand, instead of the *symbol* metric $p(\mathbf{r}_k|g_k \hat{\mathbf{s}}_k)$, the *bit* metric $\mu(b)$ is used, where

$$\mu(b) = \log\left\{\sum_{\mathbf{s}_i \in \mathcal{X}(b,j)} p(\mathbf{r}_i|g_i \mathbf{s}_i)\right\}, \quad b = 0, 1, \quad j = 1, 2, \dots, k = \log_2 M, \quad (8.145)$$

and where $\mathcal{X}(b,j)$ denotes the subset of the size $M = 2^k$ point signal constellation \mathcal{X} that has bit b in position j of its signal mapping label. The computation of this metric may be complicated due to calculation of the logarithm. A sub-optimum metric can be defined based on

$$\log\left\{\sum_j z_j\right\} \approx \max_j \log\{z_j\}, \quad (8.146)$$

which yields

$$\mu(b) = \max_{\mathbf{s}_i \in \mathcal{X}(b,j)} \log\{p(\mathbf{r}_i|g_i \mathbf{s}_i)\}. \quad (8.147)$$

Except for the maximization operation, this metric is the same as the TCM symbol metric. Finally, the performance of BICM strongly depends on the signal mapping that is used. Gray mapping is known to perform better than mapping by set partitioning without iterative decoding. Optimized mapping schemes for iteratively decoded BICM are described in [312].

8.7 Performance of Space-Time Codes

The orthogonal space-time block codes described in Sect. 8.1.2 are just one possible construction for the space-time code matrix. We now discuss some useful criteria for designing more general $p \times L_T$ space-time codes for quasi-static fading channels, where the channel remains static over p time slots; the rank and determinant criteria. These criteria are derived under the condition of independent fades, where the complex channel gains g_{ij} are all independent.

To obtain the rank and determinant criteria for space-time codes, the pairwise error probability is needed between any two space-time codewords. Suppose that a space-time codeword $C = [\tilde{\mathbf{s}}_{(t,i)}]_{p \times L_T}$ is transmitted over an $L_T \times L_R$ MIMO channel in p time slots, where the modulated symbol vectors are chosen from the symbol alphabet of a linear modulation scheme, such as QAM and PSK. By observing the noisy received signal vectors in (8.31), the receiver decides which space-time codeword was transmitted using the maximum likelihood metric in (8.32). Suppose that the receiver erroneously decides in favor of another space-time codeword $\hat{C} = [\hat{\mathbf{s}}_{(t,i)}]_{p \times L_T}$. This is the pairwise error probability $P[C \rightarrow \hat{C}]$. Assuming ideal channel state information, the pairwise error probability can be written as

$$\begin{aligned}
P[C \rightarrow \hat{C}] &= \mathbb{P} \left[\sum_{j=1}^{L_R} \sum_{t=1}^p \left\| \tilde{\mathbf{r}}_{(t),j} - \sum_{i=1}^{L_T} g_{i,j} \tilde{\mathbf{s}}_{(t),i} \right\|^2 > \sum_{j=1}^{L_R} \sum_{t=1}^p \left\| \tilde{\mathbf{r}}_{(t),j} - \sum_{i=1}^{L_T} g_{i,j} \hat{\mathbf{s}}_{(t),i} \right\|^2 \right] \\
&= \mathbb{P} \left[\sum_{j=1}^{L_R} \sum_{t=1}^p 2 \operatorname{Re} \left\{ \tilde{\mathbf{n}}_{(t),j}^H \sum_{i=1}^{L_T} g_{i,j} (\hat{\mathbf{s}}_{(t),i} - \tilde{\mathbf{s}}_{(t),i}) \right\} \right. \\
&\quad \left. > \sum_{j=1}^{L_R} \sum_{t=1}^p \left\| \sum_{i=1}^{L_T} g_{i,j} (\hat{\mathbf{s}}_{(t),i} - \tilde{\mathbf{s}}_{(t),i}) \right\|^2 \right] \\
&= Q \left(\sqrt{\frac{\Delta^2 E_s}{4N_o}} \right), \tag{8.148}
\end{aligned}$$

where

$$\Delta^2 = \frac{1}{E_s} \sum_{j=1}^{L_R} \sum_{t=1}^p \left\| \sum_{i=1}^{L_T} g_{i,j} (\tilde{\mathbf{s}}_{(t),i} - \hat{\mathbf{s}}_{(t),i}) \right\|^2. \tag{8.149}$$

Applying a Chernoff bound on the Gaussian Q function (see Appendix A) gives

$$P[C \rightarrow \hat{C}] \leq \frac{1}{2} \exp \left\{ -\frac{\Delta^2 E_s}{4N_o} \right\}. \tag{8.150}$$

We can rewrite (8.149) as

$$\Delta^2 = \frac{1}{E_s} \sum_{j=1}^{L_R} \sum_{i=1}^{L_T} \sum_{i'=1}^{L_T} g_{i,j} g_{i',j}^* \sum_{t=1}^p (\tilde{\mathbf{s}}_{(t),i} - \hat{\mathbf{s}}_{(t),i}) (\tilde{\mathbf{s}}_{(t),i'} - \hat{\mathbf{s}}_{(t),i'})^H. \tag{8.151}$$

Now let $\mathbf{g}_j = (g_{1,j}, \dots, g_{L_T,j})$. After some manipulations, (8.151) can be rewritten as

$$\Delta^2 = \sum_{j=1}^{L_R} \mathbf{g}_j \mathbf{A}(C, \hat{C}) \mathbf{g}_j^H, \tag{8.152}$$

where $\mathbf{A}(C, \hat{C}) = [A_{p,q}]_{L_T \times L_R} = \frac{1}{E_s} \mathbf{s}_p \mathbf{s}_q^T$, and where

$$\mathbf{s}_p = (\tilde{\mathbf{s}}_{(p),1} - \hat{\mathbf{s}}_{(p),1}, \dots, \tilde{\mathbf{s}}_{(p),L_T} - \hat{\mathbf{s}}_{(p),L_T}) \quad (8.153)$$

$$\mathbf{s}_q = (\tilde{\mathbf{s}}_{(q),1} - \hat{\mathbf{s}}_{(q),1}, \dots, \tilde{\mathbf{s}}_{(q),L_T} - \hat{\mathbf{s}}_{(q),L_T}). \quad (8.154)$$

Hence, the Chernoff bound in (8.150) becomes

$$P[C \rightarrow \hat{C}] \leq \frac{1}{2} \prod_{j=1}^{L_R} \exp \left\{ -\frac{\mathbf{g}_j \mathbf{A}(C, \hat{C}) \mathbf{g}_j^H E_s}{4N_o} \right\}. \quad (8.155)$$

Since $\mathbf{A}(C, \hat{C}) = \mathbf{A}^H(C, \hat{C})$, there exists a unitary matrix \mathbf{V} , such that $\mathbf{V}\mathbf{V}^H = \mathbf{I}_{L_T \times L_T}$, and a real diagonal matrix \mathbf{D} such that $\mathbf{V}\mathbf{A}(C, \hat{C})\mathbf{V}^H = \mathbf{D}$. The rows $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{L_T}$ of \mathbf{V} are a complete complex orthonormal basis for an L_T -dimensional complex vector space. Also, the diagonal elements of \mathbf{D} are the eigenvalues, $\lambda_i, i = 1, \dots, L_T$ of $\mathbf{A}(C, \hat{C})$ that may include repeated eigenvalues. By construction, the matrix

$$\mathbf{B}(C, \hat{C}) = \frac{1}{\sqrt{E_s}} \begin{bmatrix} \tilde{\mathbf{s}}_{(1),1} - \hat{\mathbf{s}}_{(1),1} & \tilde{\mathbf{s}}_{(2),1} - \hat{\mathbf{s}}_{(2),1} & \cdots & \tilde{\mathbf{s}}_{(p),1} - \hat{\mathbf{s}}_{(p),1} \\ \tilde{\mathbf{s}}_{(1),2} - \hat{\mathbf{s}}_{(1),2} & \tilde{\mathbf{s}}_{(2),2} - \hat{\mathbf{s}}_{(2),2} & \cdots & \tilde{\mathbf{s}}_{(p),2} - \hat{\mathbf{s}}_{(p),2} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\mathbf{s}}_{(1),L_T} - \hat{\mathbf{s}}_{(1),L_T} & \tilde{\mathbf{s}}_{(2),L_T} - \hat{\mathbf{s}}_{(2),L_T} & \cdots & \tilde{\mathbf{s}}_{(p),L_T} - \hat{\mathbf{s}}_{(p),L_T} \end{bmatrix} \quad (8.156)$$

is the matrix square root of $\mathbf{A}(C, \hat{C})$ i.e., $\mathbf{A}(C, \hat{C}) = \mathbf{B}(C, \hat{C})\mathbf{B}(C, \hat{C})^H$. Therefore, the eigenvalues of $\mathbf{A}(C, \hat{C})$ are all nonnegative real numbers. Next, Δ^2 is expressed as a function of the eigenvalues of the matrix $\mathbf{A}(C, \hat{C})$. Let $(\beta_{1,j}, \beta_{2,j}, \dots, \beta_{L_T,j}) = \mathbf{g}_j \mathbf{V}^H$. Then

$$\mathbf{g}_j \mathbf{A}(C, \hat{C}) \mathbf{g}_j^H = \mathbf{g}_j \mathbf{V}^H \mathbf{D} \mathbf{V} \mathbf{g}_j^H = \sum_{i=1}^{L_T} \lambda_i |\beta_{i,j}|^2. \quad (8.157)$$

Next recall that $g_{i,j}$ are all complex Gaussian random variables. Since \mathbf{V} is unitary, the $\beta_{i,j}$ are independent complex Gaussian random variables with a normalized variance 1/2 per dimension. Assuming that the $g_{i,j}$ have zero mean, the $|\beta_{i,j}|$ are independent Rayleigh random variables with the density function

$$p_{\beta_{i,j}}(x) = 2xe^{x^2}, \quad x \geq 0. \quad (8.158)$$

Thus to obtain Chernoff bound on the pairwise error probability,

$$\prod_{j=1}^{L_R} \exp \left\{ -\frac{E_s}{4N_o} \sum_{i=1}^{L_T} \lambda_i |\beta_{i,j}|^2 \right\} \quad (8.159)$$

is averaged over the distribution of the $|\beta_{i,j}|$ in (8.158). This leads to the bound on the pairwise error probability

$$P[C \rightarrow \hat{C}] \leq \frac{1}{2} \left(\frac{1}{\prod_{i=1}^{L_T} (1 + \lambda_i E_s / 4N_o)} \right)^{L_R}. \quad (8.160)$$

Let r denote the rank of the matrix $\mathbf{A}(C, \hat{C})$, where $1 \leq r \leq L_T$. Then for sufficiently high E_s/N_o such that $1 + \lambda_i E_s / 4N_o \approx \lambda_i E_s / 4N_o$, it follows that the pairwise error probability has the upper bound

$$P[C \rightarrow \hat{C}] \leq \frac{1}{2} \left[\left(\prod_{i=1}^r \lambda_i \right)^{1/r} \right]^{rL_R} \left(\frac{E_s}{4N_o} \right)^{rL_R}. \quad (8.161)$$

Therefore, this space-time codeword pair provides diversity of order rL_R and a coding gain of $(\lambda_1 \cdot \lambda_2 \cdots \lambda_r)^{1/r}$ which is equal to the geometric mean of the non-zero eigenvalues. The coding gain is measured with respect to an uncoded system with maximal ratio combining that is operating with the same diversity order. Finally, the ranks of $\mathbf{A}(C, \hat{C})$ and $\mathbf{B}(C, \hat{C})$ are equal.

The above development leads to the following two simple design criteria for space-time codes on quasi-static Rayleigh fading channels:

Rank Criterion: To achieve the maximum diversity order $L_T L_R$, the matrix $\mathbf{B}(C, \hat{C})$ must have full rank L_T for all pairs of distinct codewords. If the minimum rank of $\mathbf{B}(C, \hat{C})$ is equal to r , then a diversity order of rL_R is achieved.

Determinant Criterion: If a diversity order of $L_T L_R$ is achieved, then the coding gain is optimized by maximizing the minimum determinant of $\mathbf{A}(C, \hat{C})$ over all pairs of distinct codewords. If a diversity order of rL_R is achieved, then the minimum value of $(\lambda_1 \cdot \lambda_2 \cdots \lambda_r)^{1/r}$ should be maximized over all pairs of distinct codewords.

8.7.1 Space-Time Trellis Codes

8.7.1.1 Encoder Description

A space-time trellis code (STTC) maps the information bit stream into L_T streams of modulated symbols that are transmitted simultaneously from L_T transmit antennas. The total transmission power is divided equally among the L_T transmit antennas. STTCs can be designed according to the rank and determinant criteria as discussed in Sect. 8.7, although other criteria exist as well.

STTCs are an extension of trellis-coded modulation to multi-antenna systems. The encoder for a STTC is similar to that for a trellis code, except that the encoder must begin and end each frame in the all-zeroes state. The STTC encoder can be described in terms of its generator sequences or generator polynomials. For a system with L_T transmit antennas, the i th input sequence $\mathbf{a}^{(i)}$, $i = 1, \dots, L_T$ has the polynomial representation

$$\mathbf{a}^{(i)}(D) = \sum_{k=0}^{\infty} a_{i,k} D^k, \quad i = 1, \dots, L_T, \quad (8.162)$$

where the $a_{i,k} \in \{0, 1\}$ are binary input data bits. The generator polynomial corresponding to the i th input sequence, where $i = 1, \dots, L_T$, and the j th transmit antenna, where $j = 1, \dots, L_T$, can be written as

$$\mathbf{g}_i^{(j)}(D) = \sum_{k=0}^{K-1} g_{i,k}^{(j)} D^k, \quad (8.163)$$

where the $g_{i,k}^{(j)}$ are non-binary coefficients chosen from the set $\{0, 1, \dots, M-1\}$, where M is the size of the signal constellation, and K is the overall constraint length of the encoder (similar to convolutional codes). Very often the generator sequences for STTCs are tabulated in the following format:

$$\mathbf{g}_i = [(g_{i,0}^{(1)}, g_{i,0}^{(2)}, \dots, g_{i,0}^{(L_T)}), (g_{i,1}^{(1)}, g_{i,1}^{(2)}, \dots, g_{i,1}^{(L_T)}), \dots, (g_{i,K-1}^{(1)}, g_{i,K-1}^{(2)}, \dots, g_{i,K-1}^{(L_T)})], \quad i = 1, \dots, L_T. \quad (8.164)$$

A simple 2-transmit antenna, 16-state, STTC encoder is shown in Fig. 8.19. The constraint length $K = 3$ encoder shown in Fig. 8.19 may have

$$\mathbf{g}_1 = [(0, 1), (1, 2), (2, 0)], \quad \mathbf{g}_2 = [(0, 2), (2, 0), (0, 2)], \quad (8.165)$$

which just happens to satisfy the rank and determinant criteria in Sect. 8.7.

The multiplier outputs, as illustrated in Fig. 8.19, are summed with modulo- M addition. Hence, the encoded sequence that is transmitted from antenna j is given by

$$\mathbf{b}^{(j)}(D) = \sum_{i=1}^{L_T} \mathbf{a}^{(i)}(D) \mathbf{g}_i^{(j)}(D) \text{ modulo } M, \quad i = 1, \dots, L_T. \quad (8.166)$$

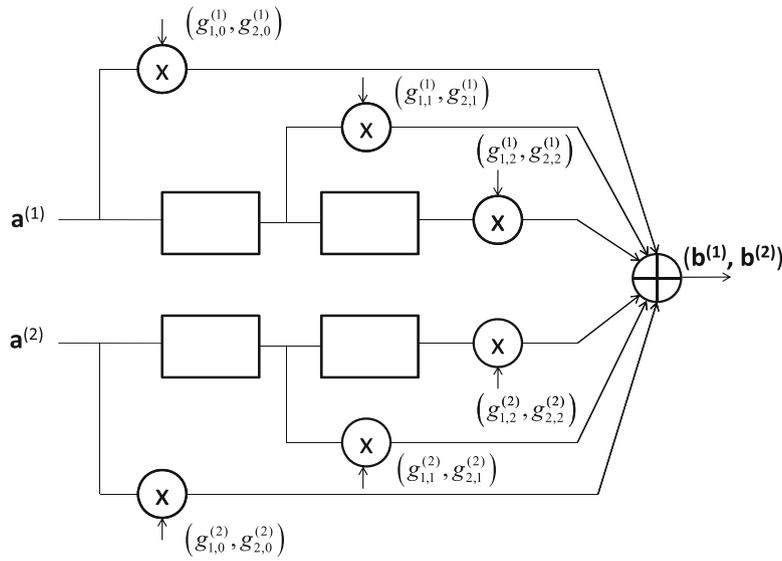


Fig. 8.19 Encoder for a 2-transmit antenna, 16-state, STTC

The above expression can be expressed in the matrix form

$$[\mathbf{b}^{(1)}(D), \dots, \mathbf{b}^{(L_T)}(D)] = [\mathbf{a}^{(1)}(D), \dots, \mathbf{a}^{(L_T)}(D)] \begin{bmatrix} \mathbf{g}_1^{(1)}(D), \dots, \mathbf{g}_1^{(L_T)}(D) \\ \vdots \\ \mathbf{g}_{L_T}^{(1)}(D), \dots, \mathbf{g}_{L_T}^{(L_T)}(D) \end{bmatrix}, \quad (8.167)$$

where

$$\mathbf{G}(D) = \begin{bmatrix} \mathbf{g}_1^{(1)}(D), \dots, \mathbf{g}_1^{(L_T)}(D) \\ \vdots \\ \mathbf{g}_{L_T}^{(1)}(D), \dots, \mathbf{g}_{L_T}^{(L_T)}(D) \end{bmatrix} \quad (8.168)$$

is the generator matrix of the STTC. The elements of the code symbol sequences $\mathbf{b}_i = \{b_{i,k}\}$, $i = 1, \dots, L_T$ at the output of the encoder are mapped onto symbols chosen from an M -ary signal constellation, such as M -PSK or M -QAM according to a one-to-one mapping. This yields the sequence of modulated symbols $\mathbf{s}_i = \{s_{i,k}\}$, $i = 1, \dots, L_T$ that are transmitted from the L_T transmit antennas.

8.7.1.2 STTC Code Trellis and the Viterbi Algorithm

Similar to convolutional and trellis codes, the STTC encoder can be described by a state diagram and trellis diagram. As an example, consider the 2-transmit antenna, 4-state, 4-PSK STTC shown in Fig. 8.20, where the generators are

$$\mathbf{g}_1 = [(0, 1), (1, 0)], \quad \mathbf{g}_2 = [(0, 2), (2, 0)]. \quad (8.169)$$

The corresponding trellis diagram for this code is shown in Fig. 8.21. The branch label b_1b_2 means that symbol b_1 is transmitted from the first antenna, while the symbol b_2 from the second antenna. The symbol pairs in each row label the branch transitions out of a given state, in order, from top to bottom. The symbols $b_1, b_2 \in \{0, 1, 2, 3\}$ are mapped onto a 4-PSK signal constellation using the mapping shown in Fig. 8.21, which can be expressed mathematically as $s_i = \sqrt{2E_b}(j)^{b_i}$. The encoder is required to begin and end each frame in the zero-state. Beginning at State 0, if the two input bits are 11, then the encoder outputs symbol 0 on Antenna 1 and symbol 3 on Antenna 2, and transitions to State 3.

Fig. 8.20 Encoder for a 2-transmit antenna, 4-state, 4-PSK STTC

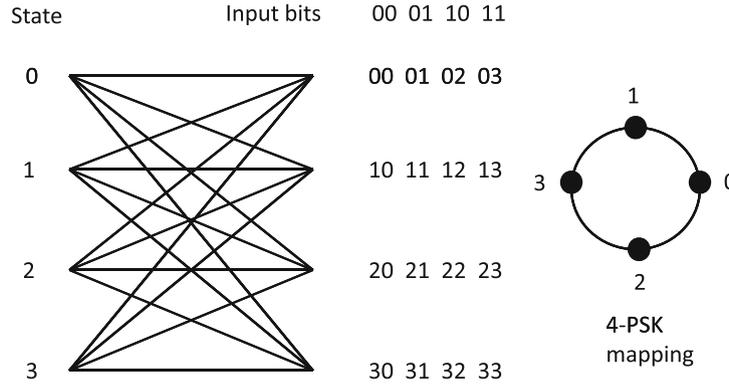
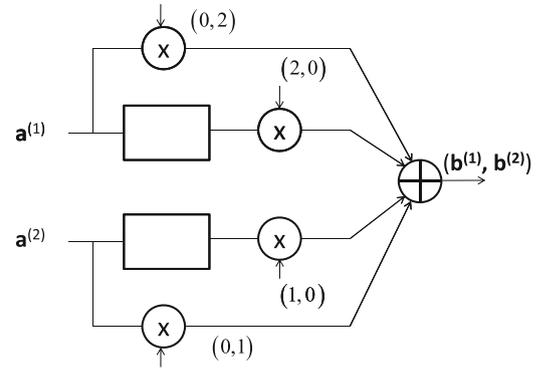


Fig. 8.21 STTC trellis for the 2-transmit antenna, 4-state, 4-PSK STTC in Fig. 8.20

Given their trellis structure, it is apparent that STTCs can be efficiently decoded by using the Viterbi algorithm with an appropriately defined branch metric. For quasi-static flat fading channels, the branch metric is

$$\mu(q_k^{(i)} \rightarrow q_{k+1}^{(j)}) = - \sum_{d=1}^{L_R} \left\| \tilde{\mathbf{r}}_{(i),j} - \sum_{i=1}^{L_T} g_{i,j} \hat{\mathbf{s}}_{(i),i}(q_k^{(i)} \rightarrow q_{k+1}^{(j)}) \right\|^2, \tag{8.170}$$

where $\hat{\mathbf{s}}_{(i),i}(q_k^{(i)} \rightarrow q_{k+1}^{(j)})$ is a symbol that is uniquely determined by the state transition $q_k^{(i)} \rightarrow q_{k+1}^{(j)}$.

8.8 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 [33]. 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 that are realized in systematic form. Here, only convolutional component codes are considered. Parallel concatenated convolutional codes (PCCCs) use recursive systematic convolutional (RSC) component codes. Serial concatenated convolutional codes (SCCCs) use a recursive or non-recursive convolutional outer code along with a recursive convolutional inner code.

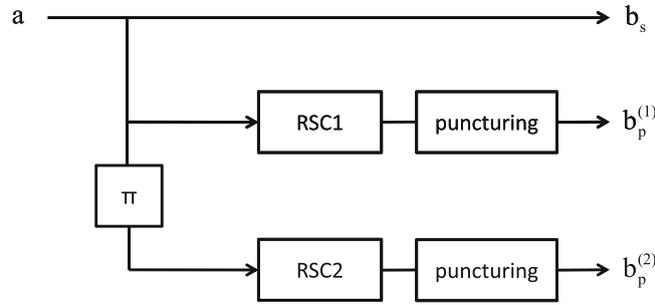


Fig. 8.22 PCCC encoder

8.8.1 PCCC Encoder

Figure 8.22 shows a PCCC encoder structure which is a parallel concatenation of two RSC component codes.³ The component codes must be recursive for reasons that will be apparent later. Notice that both the systematic and parity bits of the first encoder are used, while only the parity bits of the second encoder are used. 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

$$R_T = \frac{R_c^{(1)} R_c^{(2)}}{R_c^{(1)} + R_c^{(2)} - R_c^{(1)} R_c^{(2)}} = \frac{k}{n_1 + n_2 - k}. \quad (8.171)$$

The input data sequence \mathbf{a} is first encoded by RSC1. Suppose, for example, that RSC1 and RSC2 happen to be identical rate-1/2 codes. Then feedforward and feedback generator polynomials of RSC1 and RSC2 are $\mathbf{g}^{(2)}(D) = g_0^{(2)} + g_1^{(2)}D + \dots + g_\nu^{(2)}D^\nu$ and $\mathbf{g}^{(1)}(D) = g_0^{(1)} + g_1^{(1)}D + \dots + g_\nu^{(1)}D^\nu$, 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

$$b_{s_k} = a_k$$

$$b_{p_k}^{(1)} = \sum_{i=0}^{\nu} g_i^{(1)} d_{k-i},$$

where

$$d_k = a_k \oplus \sum_{i=1}^{\nu} g_i^{(2)} d_{k-i}. \quad (8.172)$$

The data sequence \mathbf{a} is interleaved by a turbo 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

$$\{0, 1, 2, 3, \dots, N-1\}_N \rightarrow \{23, 12, 6, 7, \dots, 1\}_N$$

For turbo codes the choice of interleaver is crucial. Interleavers that have a structure, such as block interleavers, are not suitable. Usually, random interleavers are employed, where the interleaving mapping is randomly generated. 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 multiplexing) of the systematic component and the parity sequences. If higher code rates are desired, then the parity outputs of the RSC component encoders

³The parallel concatenation of more than two component codes is possible, but only two component codes are considered for simplicity.

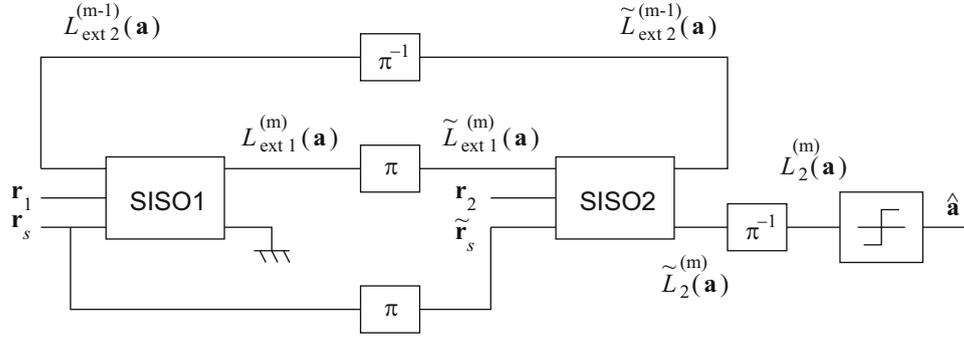


Fig. 8.23 PCCC decoder

can be punctured according to a puncturing pattern. For example, if RSC1 and RSC2 in Fig. 8.22 are rate-1/2 codes and no puncturing is used, then the code rate will be $R_T = 1/3$. However, suppose that puncturing is used with the pattern

$$\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (8.173)$$

where a “1” in the puncturing pattern means that the code bit is transmitted, while a “0” means that the code bit is not transmitted. This particular puncturing pattern means that the parity bits from RSC1 and RSC2 are transmitted in an alternate fashion, and overall code rate is increased to $R_T = 1/2$. Finally, 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.”

8.8.2 PCCC Decoder

The turbo decoder is an iterative structure consisting of several identical stages, each consisting of two soft-input soft-output (SISO) decoding units for the case of two constituent codes. Suppose that the codeword $\mathbf{b} = (\mathbf{b}_s, \mathbf{b}_p^{(1)}, \mathbf{b}_p^{(2)})$ is transmitted and the received vector is $\tilde{\mathbf{r}} = (\tilde{\mathbf{r}}_s, \tilde{\mathbf{r}}_p^{(1)}, \tilde{\mathbf{r}}_p^{(2)})$. The decoder structure for PCCCs is shown in Fig. 8.23. The (SISO) modules generate a posteriori probabilities (APPs)

$$P(a_k | \tilde{\mathbf{r}}_s, \tilde{\mathbf{r}}_p^{(1)}, \tilde{\mathbf{r}}_p^{(2)}) \quad (8.174)$$

or, for binary codes, a posteriori log-likelihood ratios (LLRs)

$$L(a_k | \tilde{\mathbf{r}}_s, \tilde{\mathbf{r}}_p^{(1)}, \tilde{\mathbf{r}}_p^{(2)}) = \log \left\{ \frac{P(a_k = +1 | \tilde{\mathbf{r}}_s, \tilde{\mathbf{r}}_p^{(1)}, \tilde{\mathbf{r}}_p^{(2)})}{P(a_k = -1 | \tilde{\mathbf{r}}_s, \tilde{\mathbf{r}}_p^{(1)}, \tilde{\mathbf{r}}_p^{(2)})} \right\} \quad (8.175)$$

of each information bit a_k based on the received vector $\tilde{\mathbf{r}} = (\tilde{\mathbf{r}}_s, \tilde{\mathbf{r}}_p^{(1)}, \tilde{\mathbf{r}}_p^{(2)})$ 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 m th iteration, $m \geq 1$, the LLRs generated by the SISO decoders for data bit a_k are

$$L_1^{(m)}(a_k | \tilde{\mathbf{r}}_p^{(1)}) = L_{\text{sys}}(a_k | \tilde{\mathbf{r}}_s) + L_{\text{ext2}}^{(m-1)}(a_k | \tilde{\mathbf{r}}_p^{(2)}) + L_{\text{ext1}}^{(m)}(a_k | \tilde{\mathbf{r}}_p^{(1)}) \quad (8.176)$$

$$L_2^{(m)}(a_k | \tilde{\mathbf{r}}_p^{(2)}) = L_{\text{sys}}(a_k | \tilde{\mathbf{r}}_s) + L_{\text{ext1}}^{(m)}(a_k | \tilde{\mathbf{r}}_p^{(1)}) + L_{\text{ext2}}^{(m)}(a_k | \tilde{\mathbf{r}}_p^{(2)}), \quad (8.177)$$

where $L_{\text{sys}}(a_k | \tilde{\mathbf{r}}_s)$ is the a posteriori LLR due to the systematic component, and $L_{\text{ext1}}^{(m)}(a_k | \tilde{\mathbf{r}}_p^{(1)})$ and $L_{\text{ext2}}^{(m)}(a_k | \tilde{\mathbf{r}}_p^{(2)})$ are the extrinsic information for each bit generated at the m th decoding stage by SISO1 and SISO2, respectively, and can be expressed as

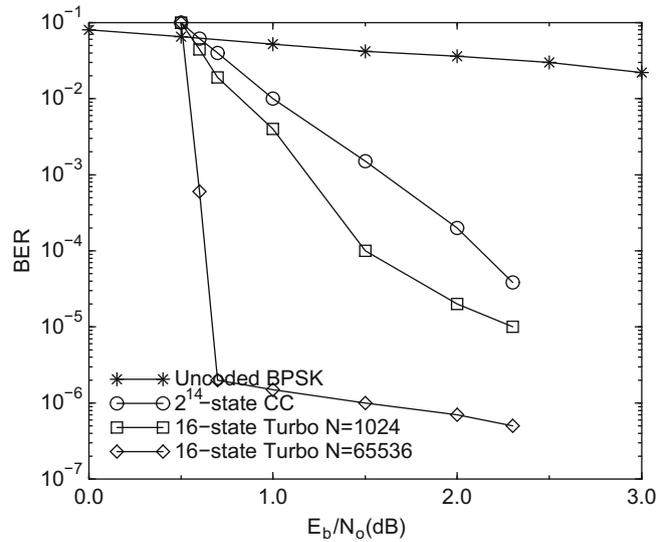


Fig. 8.24 Typical PCCC performance on an AWGN channel

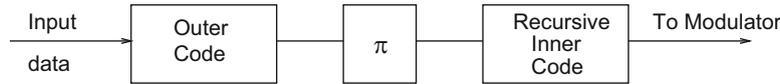


Fig. 8.25 SCCC encoder

$$L_{\text{ext1}}^{(m)}(a_k | \tilde{\mathbf{r}}_p^{(1)}) = f(L_{\text{sys}}(a_k | \tilde{\mathbf{r}}_s), L_{\text{ext2}}^{(m-1)}(a_k | \tilde{\mathbf{r}}_p^{(2)})) \quad (8.178)$$

$$L_{\text{ext2}}^{(m)}(a_k | \tilde{\mathbf{r}}_p^{(2)}) = f(L_{\text{sys}}(a_k | \tilde{\mathbf{r}}_s), L_{\text{ext1}}^{(m)}(a_k | \tilde{\mathbf{r}}_p^{(1)})), \quad (8.179)$$

where $f(\cdot)$ denotes the SISO decoding unit. The extrinsic information can be calculated by using the BCJR algorithm as discussed in Sect. 8.2.5. For binary turbo codes $L_{\text{sys}}(a_k | \tilde{\mathbf{r}}_s)$ is given by (8.96). Note that the extrinsic information that is generated by one SISO decoding unit serves as the a priori information for the other SISO decoding unit. The iterative procedure is started with initial condition $L_{\text{ext2}}^{(0)}(a_k | \tilde{\mathbf{r}}_p^{(2)}) = 0$, since the data symbols are assumed random and equally likely. The final bit decision for a_k is determined as $\hat{a}_k = \text{sign}(L_2^{(m)}(a_k))$.

As mentioned previously, turbo codes can provide near Shannon limit performance. Figure 8.24 shows the typical performance of a rate-1/2, 16-state, PCCC on an AWGN channel for different random interleaver sizes. The capacity limit for a rate-1/2 binary turbo code is -0.817 dB. Also included, is a 2^{16} -state convolutional code for comparison. Observe that a simple 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 bit error rate (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 Hamming distance of turbo codes.

8.8.3 SCCC Encoder and Decoder

Figure 8.25 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

$$R_T = R_c^{(1)} R_c^{(2)} = (k/p)(p/n) = k/n. \quad (8.180)$$

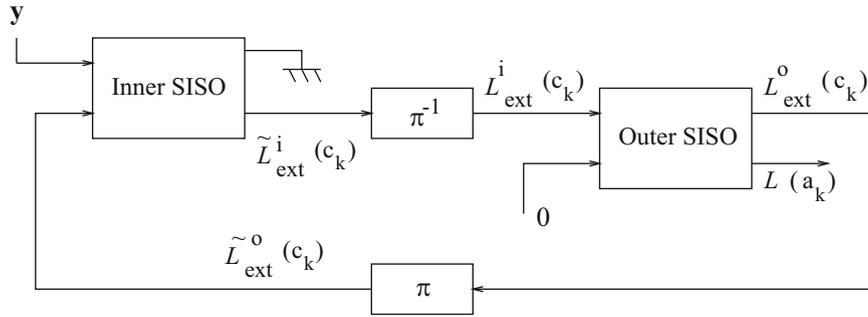


Fig. 8.26 SCCC decoder

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 SCCC, 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. 8.26. 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.

8.8.4 Weight Distribution

It is sometimes useful to view PCCCs and SCCC 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 is defined as [37]

$$A_w(z) \triangleq \sum_d A_{w,d} z^d, \quad (8.181)$$

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

$$P_b(e) \leq \frac{1}{N'} \sum_w \sum_{d=d_{\text{free}}} w A_{w,d} P_2(d), \quad (8.182)$$

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; either decrease $A_{w,d}$ or increase d_{free} . With convolutional codes, A_d increases rapidly with d and, as a result, convolutional codes are said to have a *dense distance spectrum*.⁴ 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 obtained by increasing d_{free} , which is 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

$$w(\mathbf{b}) = w(\mathbf{b}_s) + w(\mathbf{b}_p^{(1)}) + w(\mathbf{b}_p^{(2)}). \quad (8.183)$$

⁴It 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.

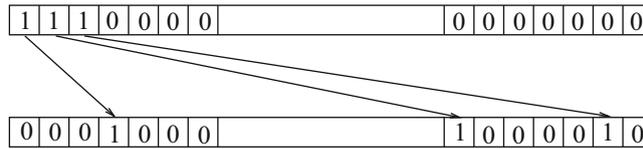


Fig. 8.27 Random turbo interleaver

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. 8.27. Certain input sequences \mathbf{a} will lead to low output weights $w(\mathbf{b}_p^{(1)})$ from the first encoder RSC1. For example, the input sequence $\mathbf{a}(D) = 1 + D^3$ 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 event occurs with high probability [98]. At high E_b/N_o the error events corresponding 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* [37, 83].

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 $A_{w,d,i}^c$ 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 $A_{w,d}^c = \sum_{i=1}^{n_{\max}} A_{w,d,i}^c$, 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 [37] solved this problem by introducing a hypothetical interleaver called a *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.

8.8.4.1 Weight Distribution of PCCCs

With a uniform interleaver the number of weight- d turbo codewords with weight- w input sequences is, for large N , [36]

$$A_{w,d}^T \approx \sum_{l=0}^d \sum_{n_1=1}^{n_{\max}} \sum_{n_2=1}^{n_{\max}} \frac{\binom{N}{n_1} \binom{N}{n_2}}{\binom{N}{w}} A_{w,l,n_1}^c A_{w,d-l,n_2}^c. \quad (8.184)$$

Using the approximation $\binom{N}{n} \approx \frac{N^n}{n!}$ gives

$$A_{w,d}^T \approx \sum_{l=0}^d \sum_{n_1=1}^{n_{\max}} \sum_{n_2=1}^{n_{\max}} \frac{w!}{n_1! \cdot n_2!} N^{n_1+n_2-w} A_{w,l,n_1}^c A_{w,d-l,n_2}^c. \quad (8.185)$$

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, 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 [98]. To see why, consider a rate-1/3, 8-state, PCCC code where the RSC component encoders have

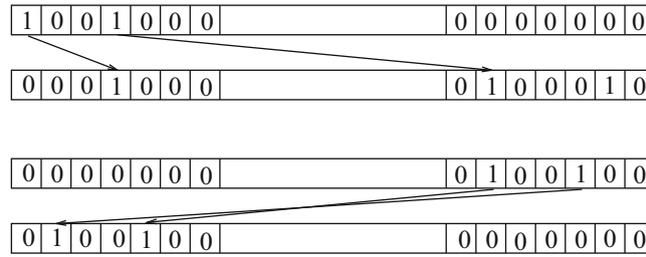


Fig. 8.28 Bad random interleaver mappings

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 [98]. 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 $D^i(1+D^3)$ with high probability [98]. 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. 8.28. 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} [98]. 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} , then the free Hamming distance of the PCCC code is $d_{\text{free}}^T = 2 + 2d_{\text{ceff}}$. For our example PCCC code, the free distance is $d_{\text{free}} = 2 + 4 + 4 = 10$, which is rather small. This small free Hamming distance is typical of PCCCs precisely the reason for the so-called BER and frame error rate (FER) floor of PCCCs. Finally, 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 the dominant error events. Such bad mappings affect only a very few bit positions, but they nevertheless 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 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

$$P[D^i(1+D^3) \rightarrow D^j(1+D^3)] = \frac{2N(N-2)!}{N!} = \frac{2}{N-1}. \quad (8.186)$$

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.,⁵

$$P[\text{total number of bad mappings} = k] = \binom{N}{k} \left(\frac{2}{N-1}\right)^k \left(1 - \frac{2}{N-1}\right)^{N-k}.$$

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.

⁵The 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.

8.8.4.2 Weight Distribution of SCCCs

Consider the serial concatenation system in Fig. 8.25. 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 [38]

$$A_{w,d}^S = \sum_{l=d_f^o}^N \sum_{n^o=1}^{n_M^o} \sum_{n^i=1}^{n_M^i} \frac{\binom{N'/p}{n^o} \binom{N/p}{n^i}}{\binom{N}{l}} A_{w,l,n^o}^{C_o} A_{l,d,n^i}^{C_i}, \quad (8.187)$$

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 $\binom{N}{n} \approx \frac{N^n}{n!}$ [38]

$$A_{w,d}^S \approx \sum_{l=d_f^o}^N \sum_{n^o=1}^{n_M^o} \sum_{n^i=1}^{n_M^i} N^{n^o+n^i-l-1} \frac{l!}{p^{n^o+n^i} n^o! n^i!} \frac{1}{n} A_{w,l,n^o}^{C_o} A_{l,h,n^i}^{C_i}, \quad (8.188)$$

where w_m^o is the minimum-weight of all input sequences that will produce an error event for the outer code.

Observe from (8.188) 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 [38]. 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 $n^o - \lfloor l/2 \rfloor - 1$. 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 simple repetition code that generates codewords by simply repeating each information symbol L times.

- What is the rate of the code?
- How many codewords are there in the code?
- What is the minimum distance of the code?
- How many channel errors would have to occur to confuse one codeword with another?

8.2. The generator matrix for a (6,3) linear binary block code is

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- What is the parity check matrix for this code?
- Generate the standard array for this code.
- Calculate the syndrome vector for all of the correctable error patterns.
- Decode the received sequence $\mathbf{y} = 101101$.

8.3. The parity check matrix \mathbf{H} for a linear block code is given as follows:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

- Construct a standard array decoding table for this code.
- How many error patterns can this code correct.
- If this code is used for error detection on a binary symmetric channel with crossover probability p , what is the probability of undetected error?

8.4. Consider a systematic (15,11) Hamming code.

- Construct a parity check matrix for a systematic (15,11) Hamming code.
- Construct a syndrome table for the code defined by the parity check matrix.
- If the (15,11) Hamming code is used for error detection on a binary symmetric channel with crossover probability p , what is the probability of undetected error?

8.5. Determine the decision variables for the separate maximum likelihood decoding of the symbols in the following rate-3/4 space-time block code

$$C = \begin{bmatrix} \mathbf{s}_{(1)} & \mathbf{s}_{(2)} & \mathbf{s}_{(3)} \\ -\mathbf{s}_{(2)}^* & \mathbf{s}_{(1)}^* & 0 \\ \mathbf{s}_{(3)}^* & 0 & -\mathbf{s}_{(1)}^* \\ 0 & \mathbf{s}_{(3)}^* & -\mathbf{s}_{(2)}^* \end{bmatrix}.$$

8.6. Determine the decision variables for the separate maximum likelihood decoding of the symbols in the following rate-1/2 orthogonal space-time block code in (8.28).

8.7. The generator matrix for a rate-1 space-time block code is given as

$$C = \begin{bmatrix} \mathbf{s}_{(1)} & \mathbf{s}_{(2)} & \mathbf{s}_{(3)} & \mathbf{s}_{(4)} \\ -\mathbf{s}_{(2)}^* & \mathbf{s}_{(1)}^* & -\mathbf{s}_{(4)}^* & \mathbf{s}_{(3)}^* \\ -\mathbf{s}_{(3)}^* & -\mathbf{s}_{(4)}^* & \mathbf{s}_{(1)}^* & \mathbf{s}_{(2)}^* \\ \mathbf{s}_{(4)} & -\mathbf{s}_{(3)} & -\mathbf{s}_{(2)} & \mathbf{s}_{(1)} \end{bmatrix}.$$

- Determine the matrix $C^H C$ and, thus, show that the code is not orthogonal.
- Show that the maximum likelihood decoder can perform pairwise maximum likelihood detection.
- What is the diversity order achieved by the code.

8.8. Consider a rate-1/3 convolutional code with generators $\mathbf{g}^{(1)} = (111)$, $\mathbf{g}^{(2)} = (111)$, and $\mathbf{g}^{(3)} = (101)$.

- Draw a block diagram of the encoder structure.
- Draw the state diagram and trellis diagram.
- Determine the output sequence corresponding to the input sequence 1110101.

8.9. The output of a rate-1/3 convolutional encoder with constraint length 3 to the input $\mathbf{a} = (1, 1, 0, \dots)$ is $\mathbf{b} = (111, 110, 010, \dots)$

- Determine the transfer function $T(D, N, L)$.
- Determine the number of paths through the state diagram or trellis that diverge from the all-zeroes state and merge with the all-zeroes state 7 branches later.
- Determine the number of paths of Hamming distance 20 from the all-zeroes sequence.

8.10. Consider the rate-1/3 code in Problem 8.8.

- Determine the transfer function $T(D, N, L)$ of the code. What is the free Hamming distance d_{free} ?

Fig. 8.29 Signal constellations for Problem 8.11. (a) 8-PAM. (b) 32-CROSS

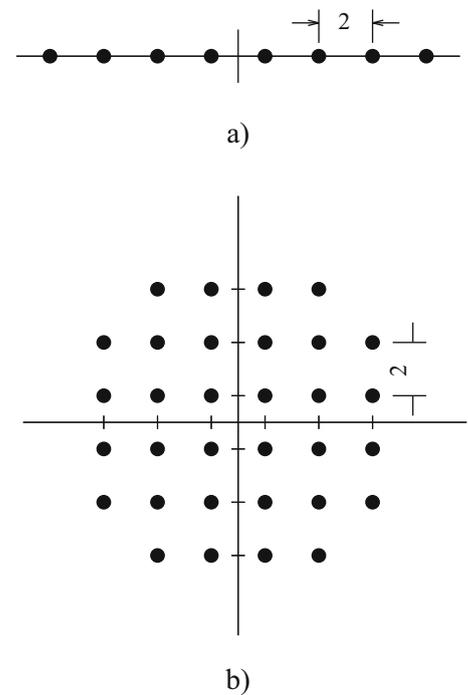
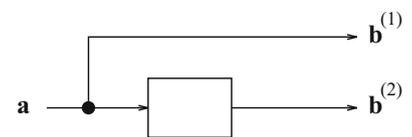


Fig. 8.30 Trellis encoder for Problem 8.12



- (b) Assuming the use of BPSK signaling and an AWGN channel, derive a union-Chernoff bound on the decoded bit error probability with (1) hard decision decoding and (2) soft decision decoding.
- (c) Repeat part (b) assuming an interleaved flat Rayleigh fading channel, where the receiver has perfect knowledge of the channel.

8.11. Consider the 8-PAM and 32-CROSS signal constellations in Fig. 8.29.

- (a) Construct the partition chain as in Fig. 8.12 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.

8.12. Consider the 2-state, rate-1/2, trellis encoder shown in Fig. 8.30. By using this encoder with a 4-PAM and 8-PAM signal constellation TCM systems can be constructed 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.
- (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.

8.13. 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 (8.120).
- (b) Derive the Chernoff bound on the pairwise error probability for an AWGN channel with hard decision decoding, given by (8.121).
- (c) Derive the Chernoff bound on the pairwise error probability for an interleaved flat fading channel with soft decision decoding, given by (8.130).

Chapter 9

Spread Spectrum Techniques

Spread spectrum systems were originally developed for military applications, to provide anti-jam 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 [97, 263, 305]. However, there are important commercial applications as well. Code division multiple access (CDMA) based on spread spectrum technology is used in 3G cellular systems, including WCDMA and cdma2000. It is also used in wireless local area network standards such as IEEE 801.11 (commonly known as WiFi), and wireless personal area networks such as Bluetooth.

Spread spectrum signals have the distinguishing characteristic that the bandwidth used to transmit a message 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. This independence property means that modulation schemes such as continuous phase modulation with a large modulation index do not qualify as spread spectrum techniques. The receiver uses a synchronized replica of the PN sequence to despread the received signal allowing recovery of the message. The bandwidth expansion does not combat additive white Gaussian noise (AWGN), but the wide-band character of spread spectrum signals can be utilized to mitigate the effects of intentional and non-intentional sources of additive interference and to exploit the inherent diversity that is present in frequency-selective fading channels.

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 concepts can be found in the books by Simon et al. [305] and Ziemer and Peterson [378]. Some of the early proposals that applied spread spectrum to cellular radio, such as the system proposed by Cooper and Nettleton [76], were based on FH spread spectrum. However, modern 3G cellular systems use DS spread spectrum. As a result, the focus of this chapter is on DS spread spectrum and 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 (1) universal frequency reuse, (2) narrow band interference rejection, (3) inherent multipath diversity in DS spread spectrum, (4) ability to exploit variable rate speech coding, (5) soft hand-off capability, (6) soft capacity limit, and (7) inherent message privacy. The disadvantages of CDMA for cellular applications include (1) stringent power control requirements with DS CDMA, (2) difficulties in determining the base station (BS) power levels for deployments that have cells of differing sizes.

This chapter begins with an introduction to DS and FH spread spectrum in Sect. 9.1. PN sequences are fundamental to all spread spectrum systems and are the subject of Sect. 9.2. A variety of sequences are considered including m -sequences, Gold sequences, Kasami sequences, Barker sequences, Walsh-Hadamard sequences, variable length orthogonal codes, and complementary code keying (CCK). The remainder of the chapter concentrates on DS spread spectrum. The power spectral density of DS spread spectrum signals is considered in Sect. 9.3. Section 9.4 considers the bit error rate performance of DS spread spectrum signals in the presence of tone interference. Section 9.5 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. Section 9.6 considers the performance of a CDMA correlator detector in the presence of multiple-access interference. The chapter concludes with a discussion of CDMA multiuser detection techniques in Sect. 9.7, including optimum CDMA multiuser detection, decorrelation detection, and minimum mean square error (MMSE) detection.

9.1 Basic Principles of Spread Spectrum

9.1.1 Direct Sequence Spread Spectrum

A simplified direct sequence (DS) spread spectrum system with QPSK modulation, termed DS/QPSK, is shown in Fig. 9.1. The pseudo-random (PN) sequence generator produces a spreading sequence $\mathbf{a} = \{a_k\}_{k=1}^N$, which is actually a deterministic sequence with period N . This spreading sequence is used to generate the spreading waveform

$$a(t) = A \sum_k a_k h_c(t - kT_c), \quad (9.1)$$

where $\mathbf{a} = \{a_k\}_{k=1}^N$ is in general a complex-valued spreading sequence such that $a_k \in \{\pm 1 \pm j\}$, T_c is the PN symbol or chip period, and $h_c(t)$ is a real-valued chip amplitude shaping pulse. The energy per chip is

$$E_c = A^2 \sigma_a^2 \int_{-\infty}^{\infty} h_c^2(t) dt = A^2 \int_{-\infty}^{\infty} h_c^2(t) dt \quad (9.2)$$

since $\sigma_a^2 = \frac{1}{2} E[|a_k|^2] = 1$. Notice that spectral control is achieved in the DS spread spectrum waveform by shaping the PN chips rather than the data symbols.

The data symbol sequence $\{x_n\}$ is used to generate the waveform

$$x(t) = \sum_n x_n u_T(t - nT), \quad (9.3)$$

where A is the amplitude, $\mathbf{x} = \{x_n\}$, $x_n \in \{\pm 1/\sqrt{2} \pm j/\sqrt{2}\}$ is the complex QPSK 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 as the number of PN chips per data symbol. Two categories of spreading codes can be defined according to 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 of the spreading sequence.

The DS/QPSK complex envelope, obtained by multiplying $a(t)$ and $x(t)$, is

$$\tilde{s}(t) = A \sum_n x_n h_n(t - nT), \quad (9.4)$$

where

$$h_n(t) = \sum_{k=0}^{G-1} a_{nG+k} h_c(t - kT_c). \quad (9.5)$$

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 n th data symbol is shaped with the amplitude shaping pulse $h_n(t)$ in (9.5). For short codes $h_n(t)$ is the

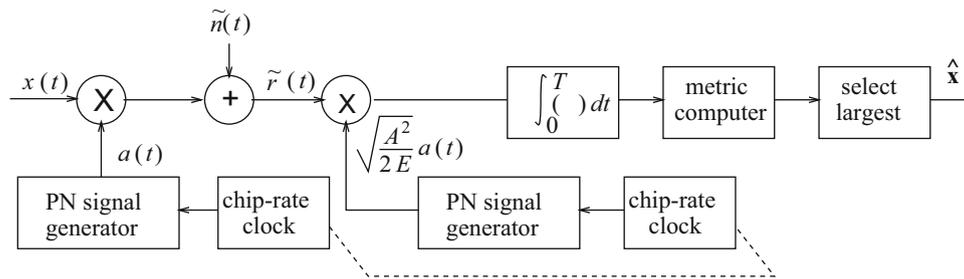
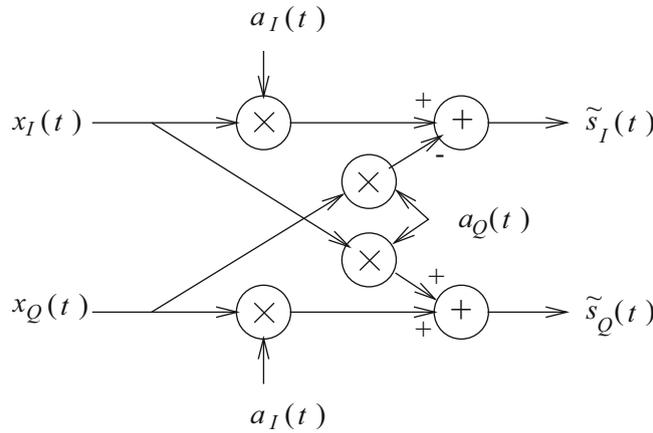


Fig. 9.1 Simplified DS/QPSK system


Fig. 9.2 Complex spreading

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. Offset QPSK (OQPSK) should not be used with complex spreading, since it will increase the peak-to-average ratio. The complex envelope $\tilde{s}(t)$ is applied to a quadrature modulator to produce the bandpass waveform

$$s(t) = A \sum_n \left(\left(x_{I,n} h_{I,n}(t - nT) - x_{Q,n} h_{Q,n}(t - nT) \right) \cos(2\pi f_c t) - \left(x_{Q,n} h_{I,n}(t - nT) + x_{I,n} h_{Q,n}(t - nT) \right) \sin(2\pi f_c t) \right), \quad (9.6)$$

where

$$h_n(t) = h_{I,n}(t) + j h_{Q,n}(t) \quad (9.7)$$

$$x_n = x_{I,n} + j x_{Q,n}. \quad (9.8)$$

During the time interval $[nT, (n+1)T)$, the DS/QPSK complex envelope can assume one of the four possible values

$$\tilde{s}_i(t) = A h_n(t) x_i, \quad i = 1, \dots, 4. \quad (9.9)$$

Using the basis function

$$\phi_n(t) = \sqrt{\frac{A^2}{2E}} h_n(t), \quad (9.10)$$

where $E = GE_c$ is the symbol energy, gives

$$\tilde{s}_i(t) = \sqrt{2E} x_i \phi_n(t), \quad i = 1, \dots, 4 \quad (9.11)$$

and it follows that the complex DS/QPSK signal vectors are

$$\tilde{s}_i = \sqrt{2E} x_i, \quad i = 1, \dots, 4. \quad (9.12)$$

Notice that the basis function $\phi_n(t)$ is indexed with the baud epoch n in the case of long spreading codes.

Besides complex spreading, other types of PN spreading can be used. One possibility is dual-channel quaternary spreading as shown in Fig. 9.3. Usually this scheme is used with OQPSK modulation to reduce the peak-to-average ratio of the magnitude of the complex envelope. If only one data sequence is to be transmitted, then either simple binary spreading or balanced quaternary spreading could be used as shown in Fig. 9.4. Balanced quaternary spreading is known to be less sensitive to interference than simple binary spreading.

Fig. 9.3 Dual-channel quaternary spreading

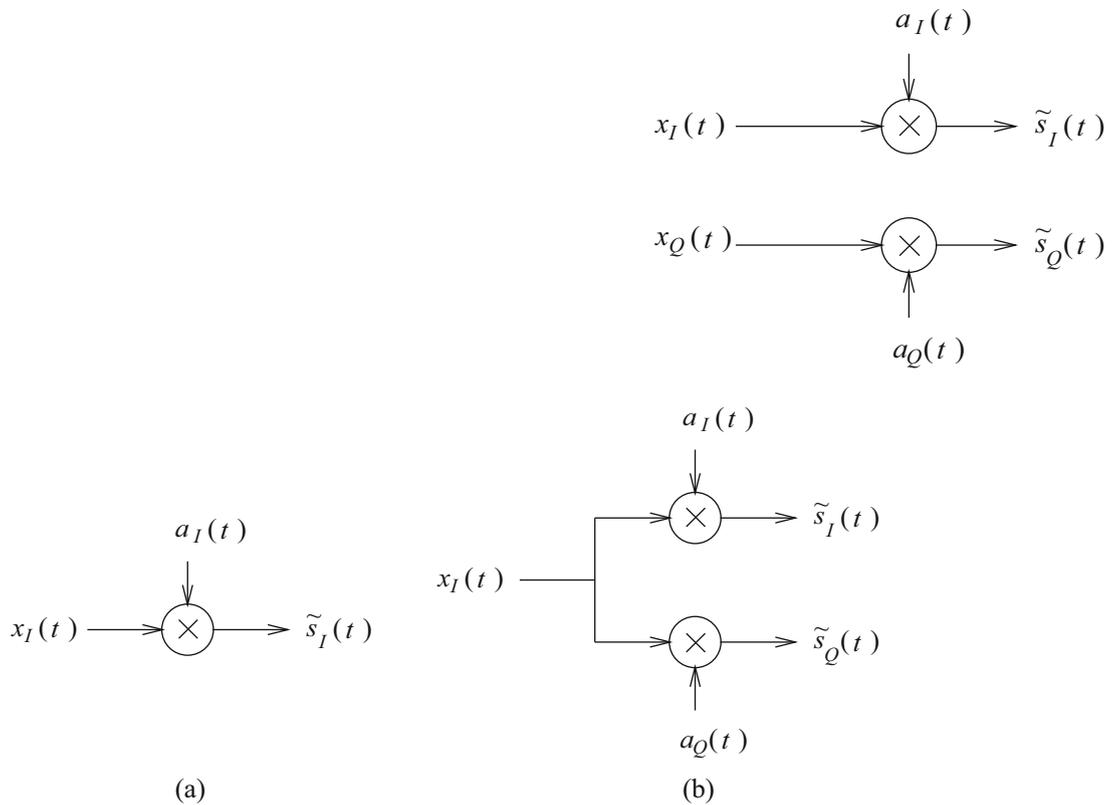


Fig. 9.4 Spreading binary data using (a) simple binary spreading, and (b) balanced quaternary spreading

Figure 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 despreaders/detector is

$$\tilde{r} = g\tilde{s}_i + \tilde{n}, \quad (9.13)$$

where \tilde{n} is a zero-mean Gaussian random variable with variance $\frac{1}{2}E[|\tilde{n}|^2] = N_o$.

The ML receiver observes \tilde{r} and decides in favor of the signal vector \tilde{s}_m that minimizes the squared Euclidean distance

$$\mu(\tilde{s}_m) = \|\tilde{r} - g\tilde{s}_m\|^2. \quad (9.14)$$

It follows that the bit error probability of DS/QPSK with Gray coding is identical to that of QPSK, and is given by

$$P_b = Q(\sqrt{2\gamma_b}), \quad (9.15)$$

where $\gamma_b = \alpha^2 E_b / N_o$ is the received bit energy-to-noise ratio. Note that spread spectrum signaling does nothing to improve the error probability performance on an AWGN channel. However, in the sequel spread spectrum signaling will be shown to offer significant error probability performance gains in the presence of additive intentional and non-intentional interference, multipath-fading, and other channel impairments.

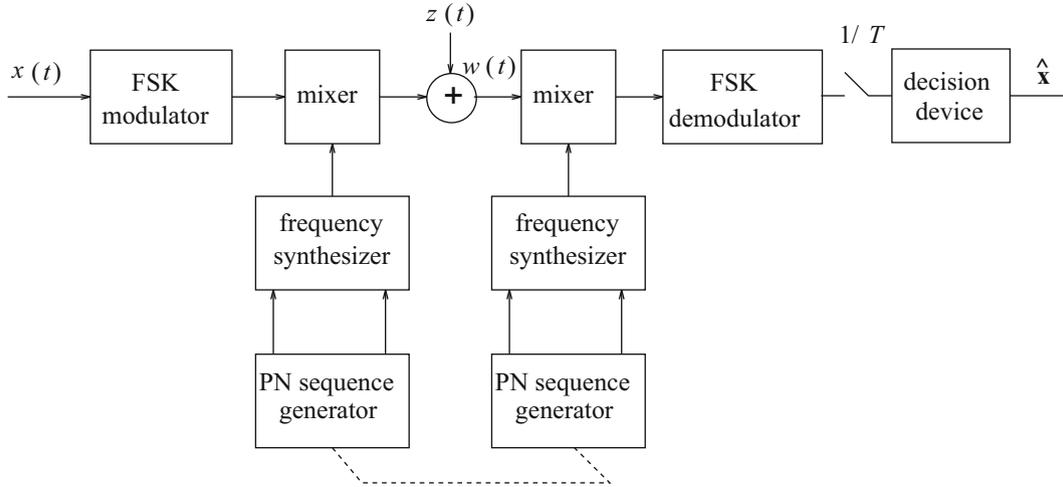


Fig. 9.5 Simplified FH system operating on an AWGN channel

9.1.2 Frequency Hop Spread Spectrum

Frequency hop (FH) 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

$$\tilde{s}(t) = A \sum_n e^{jx_n \pi \Delta_f t} u_T(t - nT), \quad (9.16)$$

where Δ_f is the frequency separation, and $x_n \in \{\pm 1, \pm 3, \dots, \pm M - 1\}$. A FH/MFSK waveform can be generated by using a digital frequency synthesizer whose inputs consist of the complex envelope of the modulating waveform $\tilde{s}(t)$ and the contents of a pseudo-noise sequence generator. A conceptual FH/MFSK spread spectrum system is shown in Fig. 9.5.

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

$$\tilde{s}(t) = A \sum_n \sum_{i=1}^L e^{jx_{nL+i} \pi \Delta_f t + 2\pi f_{nL+i} t} u_T(t - (nL + i)T), \quad (9.17)$$

where the first sum indexes the sequence of hopped carrier frequencies $\{f_n\}$, and the second sum indexes the vector of L data symbols $\mathbf{x}_n = (x_{nL+1}, x_{nL+2}, \dots, x_{(n+1)L})$ that are transmitted at the n th hop.

FFH systems transmit the same data symbol on multiple (in general L) hopped carrier frequencies. If independent channel conditions are experienced on each of the hop frequencies, then a diversity gain can be achieved by using diversity combining. The FFH/MFSK complex envelope is

$$\tilde{s}(t) = A \sum_n \sum_{i=1}^L e^{jx_n \pi \Delta_f t + 2\pi f_{nL+i} t} u_{T/L}(t - (nL + i)T/L), \quad (9.18)$$

where the first sum indexes the sequence of data symbols, $\{x_n\}$, and the second sum indexes the vector of hop frequencies $\mathbf{f}_n = (f_{nL+1}, f_{nL+2}, \dots, f_{(n+1)L})$ that are used when transmitting the n th 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 $\Delta_f = 1/2T$, while non-coherent detection requires $\Delta_f = 1/T$ (Problem 4.7). If coherent detection can be used, then the error probability of SFH/MFSK or FFH/MFSK on an AWGN channel is given by (5.119). 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 (5.191). If FFH/MFSK is used on an AWGN channel, then the error probability assumes a more complicated form due to a square-law combining loss [272].

9.2 Spreading Sequences

CDMA systems achieve their multiple-access capability by using sets spreading sequences that are chosen to have three desirable attributes: (1) the sequences are balanced so that each element of the sequence alphabet occurs with equal frequency, (2) the autocorrelations have small off-peak values, to allow for rapid sequence acquisition at the receiver and to minimize self-interference due to multipath, (3) 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 $\mathbf{a}^{(k)} = \{a_n^{(k)}\}$ denote the k th 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 $\mathbf{a}^{(k)}$ is²

$$\phi_{k,k}(n) = \frac{1}{2N} \sum_{i=0}^{N-1} a_i^{(k)*} a_{i+n}^{(k)} \quad (9.19)$$

and the full period cross-correlation between the sequences $\mathbf{a}^{(k)}$ and $\mathbf{a}^{(m)}$ is

$$\phi_{k,m}(n) = \frac{1}{2N} \sum_{i=0}^{N-1} a_i^{(k)*} a_{i+n}^{(m)}, \quad (9.20)$$

where N is the length or period of the spreading sequence(s).

The aperiodic autocorrelation of $\mathbf{a}^{(k)}$ is defined as

$$\phi_{k,k}^a(n) = \begin{cases} \frac{1}{2N} \sum_{i=1}^{N-n} a_i^{(k)*} a_{i+n}^{(k)}, & 0 \leq n \leq N-1 \\ \frac{1}{2N} \sum_{i=1}^{N+n} a_i^{(k)*} a_{i-n}^{(k)}, & -N+1 \leq n \leq 0 \\ 0, & |n| \geq N \end{cases} \quad (9.21)$$

For spread spectrum systems that employ long codes, each data symbol is spread by only a length- G subsequence of the spreading code. In this case, the partial period correlations are of interest. The partial period auto- and cross-correlations are

$$\phi_{k,k}^p(n) = \frac{1}{2G} \sum_{i=0}^{G-1} a_i^{(k)*} a_{i+n}^{(k)} \quad (9.22)$$

$$\phi_{k,m}^p(n) = \frac{1}{2G} \sum_{i=0}^{G-1} a_i^{(k)*} a_{i+n}^{(m)}. \quad (9.23)$$

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, a statistical treatment is often used under the assumption that the sequences are randomly generated, i.e., the sequence elements are chosen randomly from the set $\{\pm 1, \pm j\}$ independently and with equal probability. For random sequences

$$\frac{1}{2} \mathbb{E}[a_n^{(k)}] = 0 \quad \frac{1}{2} \mathbb{E}[|a_n^{(k)}|^2] = 1 \quad \frac{1}{2} \mathbb{E}[a_n^{(k)*} a_n^{(m)}] = 0. \quad (9.24)$$

Hence, the mean value of the partial period autocorrelation is

$$\mu_{\phi_{k,k}^p(n)} = \mathbb{E}[\phi_{k,k}^p(n)] = \frac{1}{2G} \sum_{i=0}^{G-1} \mathbb{E}[a_i^{(k)*} a_{i+n}^{(k)}] = \delta_{n, \ell N}, \quad (9.25)$$

¹The following development also applies to real spreading sequences.

²Throughout this section complex spreading sequences are assumed. For real spreading sequences, the correlation functions are similar but are normalized by N rather than $2N$.

where

$$\delta_{n,\ell N} = \begin{cases} 1, & n = \ell N \\ 0, & n \neq \ell N \end{cases} \quad (9.26)$$

and ℓ is an integer. The variance of the partial period autocorrelation is

$$\begin{aligned} \sigma_{\phi_{k,k}^p(n)}^2 &= \text{E}[|\phi_{k,k}^p(n)|^2] - \mu_{\phi_{k,k}^p(n)}^2 \\ &= \frac{1}{(2G)^2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \text{E}[a_i^{(k)*} a_{i+n}^{(k)} a_j^{(k)*} a_{j+n}^{(k)}] - \mu_{\phi_{k,k}^p(n)}^2 \\ &= (1 - \delta_{n,\ell N})(1/G). \end{aligned} \quad (9.27)$$

Likewise, the mean and variance of the partial period cross-correlation are

$$\mu_{\phi_{k,m}^p(n)} = \text{E}[\phi_{k,m}^p(n)] = 0, \quad \forall n \quad (9.28)$$

$$\sigma_{\phi_{k,m}^p(n)}^2 = \text{E}[|\phi_{k,m}^p(n)|^2] - \mu_{\phi_{k,m}^p(n)}^2 = 1/G, \quad \forall n. \quad (9.29)$$

9.2.1 Spreading Waveforms

The full period cross-correlation between two spreading waveforms $a^{(k)}(t)$ and $a^{(m)}(t)$, each of period T , is³

$$\begin{aligned} R_{k,m}(\tau) &= \frac{1}{2T} \int_0^T a^{(k)*}(t) a^{(m)}(t + \tau) dt \\ &= \frac{1}{2T} \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_i^{(k)*} a_j^{(m)} \int_0^T h_c(t - iT_c) h_c(t + \tau - jT_c) dt. \end{aligned} \quad (9.30)$$

The integral in (9.30) is non-zero 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 $0 \leq \delta < T_c$. 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

$$\begin{aligned} R_{k,m}(\tau) &= \frac{1}{2N} \sum_{i=0}^{N-1} a_i^{(k)*} a_{\ell+i}^{(m)} \frac{1}{T_c} \int_0^{T_c-\delta} h_c(t') h_c(t' + \delta) dt' \\ &\quad + \frac{1}{2N} \sum_{i=0}^{N-1} a_i^{(k)*} a_{\ell+i+1}^{(m)} \frac{1}{T_c} \int_{T_c-\delta}^{T_c} h_c(t') h_c(t' - T_c + \delta) dt'. \end{aligned} \quad (9.31)$$

The continuous-time partial autocorrelation functions of the chip waveform $h_c(t)$ (of duration T_c) are defined as [273]

$$R_h(\delta) = \frac{1}{T_c} \int_0^{T_c-\delta} h_c(t') h_c(t' + \delta) dt' \quad (9.32)$$

$$\hat{R}_h(\delta) = \frac{1}{T_c} \int_{T_c-\delta}^{T_c} h_c(t') h_c(t' - T_c + \delta) dt' \quad (9.33)$$

allowing us to write

$$R_{k,m}(\tau) = \phi_{k,m}(\ell) R_h(\delta) + \phi_{k,m}(\ell + 1) \hat{R}_h(\delta), \quad (9.34)$$

³For real-value spreading waveforms, the full period cross-correlation function is similar except for the factor of 1/2 in front of the integral.

where $\phi_{k,m}(\ell)$ is the full period cross-correlation defined in (9.20). For example, if $h_c(t) = u_{T_c}(t)$, then

$$R_{k,m}(\tau) = \phi_{k,m}(\ell) \left(1 - \frac{\delta}{T_c}\right) + \phi_{k,m}(\ell + 1) \frac{\delta}{T_c}. \quad (9.35)$$

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

$$\mu_{R_{k,m}(\tau)} = \mu_{\phi_{k,m}(\ell)} R_h(\delta) + \mu_{\phi_{k,m}(\ell+1)} \hat{R}_h(\delta) = 0 \quad (9.36)$$

$$\sigma_{R_{k,m}(\tau)}^2 = \sigma_{\phi_{k,m}(\ell)}^2 R_h^2(\delta) + \sigma_{\phi_{k,m}(\ell+1)}^2 \hat{R}_h^2(\delta) = \frac{1}{G} \left(R_h^2(\delta) + \hat{R}_h^2(\delta) \right). \quad (9.37)$$

Likewise, the autocorrelation is also a random variable that (for random spreading sequences) has mean and variance

$$\begin{aligned} \mu_{R_{k,k}(\tau)} &= \mu_{\phi_{k,k}(\ell)} R_h(\delta) + \mu_{\phi_{k,k}(\ell+1)} \hat{R}_h(\delta) \\ &= \begin{cases} R_h(\delta), & \ell = iG \\ \hat{R}_h(\delta), & \ell + 1 = iG \\ 0, & \text{elsewhere} \end{cases} \end{aligned} \quad (9.38)$$

$$\begin{aligned} \sigma_{R_{k,k}(\tau)}^2 &= \sigma_{\phi_{k,k}(\ell)}^2 R_h^2(\delta) + \sigma_{\phi_{k,k}(\ell+1)}^2 \hat{R}_h^2(\delta) \\ &= \begin{cases} R_h^2(\delta), & \ell = iG \\ \hat{R}_h^2(\delta), & \ell + 1 = iG \\ 1/G, & \text{elsewhere} \end{cases} \end{aligned} \quad (9.39)$$

where i is an integer.

9.2.2 m -sequences

One very well-known and important class of spreading sequences are the maximal-length sequences or m -sequences. As shown in Fig. 9.6, an m -sequence $\tilde{\mathbf{a}} = \{\tilde{a}_k\}$, $\tilde{a}_k \in \{0, 1\}$, 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 in the LFSR is a primitive polynomial of degree m over GF(2), having the form

$$p(x) = 1 + p_1x + p_2x^2 + p_3x^3 + \cdots + p_{m-1}x^{m-1} + x^m, \quad (9.40)$$

where $p_i \in \{0, 1\}$. Tables of primitive polynomials, $p(x)$, are tabulated in many texts, e.g., [203]. Notice that $p_0 = 1$, since this represents the feedback connection tap. Also, $p_m = 1$; otherwise, if $p_m = 0$, the effective length of the shift register is 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 length $N = 2^m - 1$ is generated. As an m -sequence generator cycles through

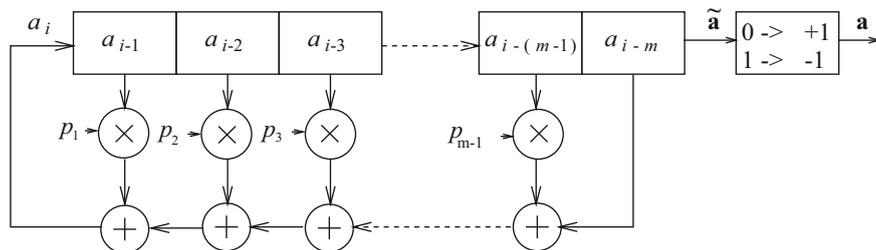


Fig. 9.6 m -sequence generator

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 LFSR 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 sequence is nearly 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

$$n_P = \begin{cases} 2^{m-P-1}, & P = 1, 2, \dots, m-1 \\ 1, & P = m \end{cases} \quad (9.41)$$

The full period autocorrelation of an m -sequence is

$$\phi(n) = \begin{cases} 1 & , n = \ell N \\ -1/N & , n \neq \ell N \end{cases} \quad (9.42)$$

For large values of N , $\phi(n) \approx \delta(n)$ so that m -sequences are almost ideal 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

$$R(\tau) = \phi(\ell) \left(1 - \frac{\delta}{T_c}\right) + \phi(\ell + 1) \frac{\delta}{T_c} \quad (9.43)$$

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 straightforward fashion by replacing the expectations in (9.25) and (9.27) with averages over all possible starting positions. This gives

$$\mu_{\phi(n)} = \begin{cases} 1 & , n = \ell N \\ -1/G & , n \neq \ell N \end{cases} \quad (9.44)$$

$$\sigma_{\phi(n)}^2 = \begin{cases} 0 & , n = \ell N \\ \frac{1}{G} \left(1 + \frac{1}{N}\right) \left(1 - \frac{G}{N}\right) & , n \neq \ell N \end{cases} \quad (9.45)$$

Unfortunately, m -sequences also have a number of undesirable properties. First, the number of m -sequences that can be generated by an LFSR of length m 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

$$\Phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right), \quad (9.46)$$

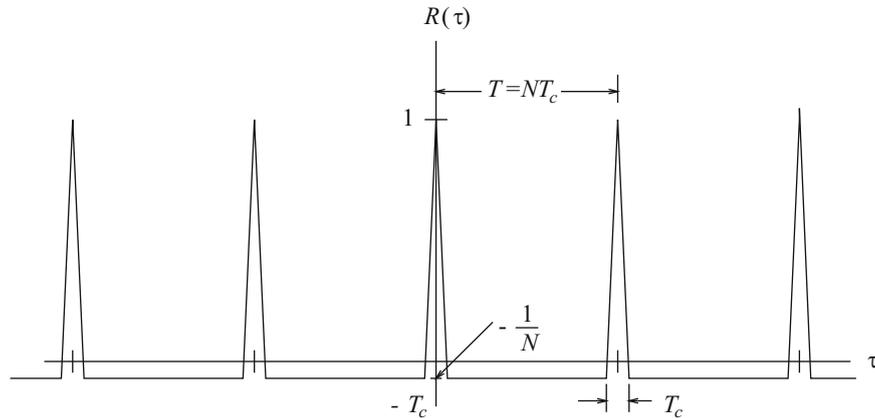


Fig. 9.7 Typical full period autocorrelation function of an m -sequence spreading waveform

Table 9.1 Best and worst case average cross-correlations for m -sequences

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

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

$$\theta = \frac{1}{N} \sum_{n=0}^{N-1} \phi_{k,m}(n). \quad (9.47)$$

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.

9.2.3 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 $\text{GF}(2^m)$ be an extension field of $\text{GF}(2)$. Let α be a primitive N th root of unity in the extension field $\text{GF}(2^m)$, where $N = 2^m - 1$. Let $p_1(x)$ and $p_2(x)$ be a pair of primitive polynomials over $\text{GF}(2)$ each having degree m such that $p_1(\alpha) = 0$ and $p_2(\alpha^d) = 0$ for some integer d . Consider the case when $m \not\equiv 0 \pmod{4}$. If $d = 2^h + 1$ or $d = 2^{2h} - 2^h + 1$ and if $e = \text{GCD}(m, h)$ 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 either $p_2(x^3) = 0$ or $p_2(x^5) = 0$ may be chosen. To find the corresponding polynomials, the reader is referred to Peterson's table of irreducible polynomials [262]. 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

$$t(m) = \begin{cases} 2^{(m+1)/2} + 1 & , m \text{ odd} \\ 2^{(m+2)/2} + 1 & , m \text{ even} \end{cases} \quad (9.48)$$

By using the preferred pair of sequences $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$, a set of Gold sequences can be constructed 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 give a set of $2^m + 1$ Gold sequences.

It is important to note that not 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 so 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 autocorrelation functions 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). The cross-correlation properties of m -sequences and Gold sequences are summarized in Table 9.2. Notice that Gold sequences have much smaller peak cross-correlations than m sequences.

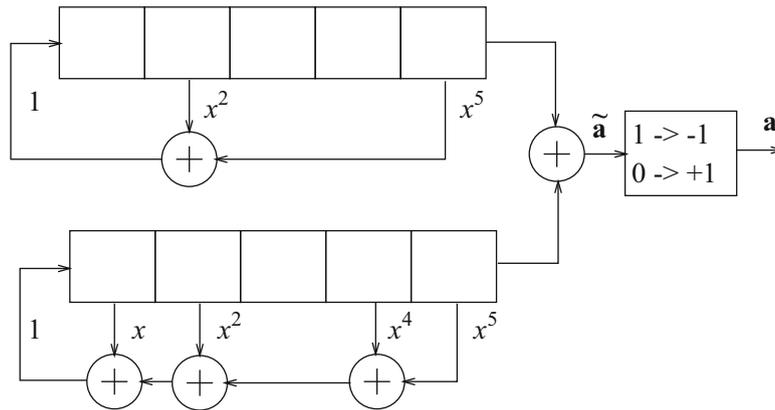


Fig. 9.8 A Gold sequence generator with $p_1(x) = 1 + x^2 + x^5$ and $p_2(x) = 1 + x + x^2 + x^4 + x^5$. This sequence generator can produce 32 Gold sequences of length 31

Table 9.2 Peak cross-correlation of m -sequences and Gold sequences

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

9.2.4 Kasami Sequences

The construction of Kasami sequences proceeds as follows [175, 176]. Let m be even. Let $p_1(x)$ be a primitive polynomial over the binary field $GF(2)$ with degree m and α 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 [262]. 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 $N = 2^m - 1$. 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 $s(m) = 2^{m/2} + 1$.

9.2.5 Barker Sequences

Barker sequences only exist for lengths 2, 3, 4, 5, 7, 11, and 13, given as follows:

$$\begin{aligned} \mathbf{a} &= (+1 - 1) \\ \mathbf{a} &= (+1 + 1 - 1) \\ \mathbf{a} &= (+1 + 1 - 1 + 1) \end{aligned}$$

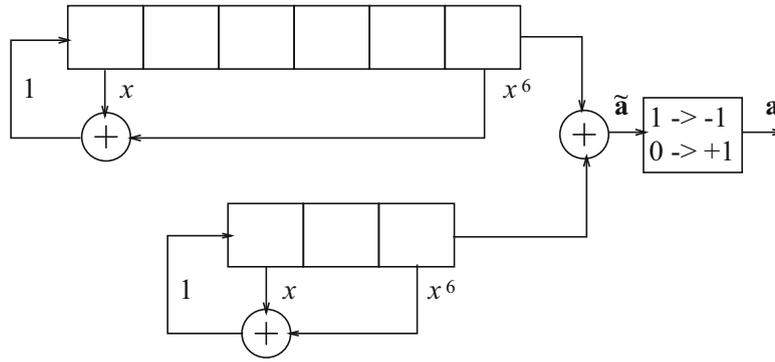


Fig. 9.9 A Kasami sequence generator with $p_1(x) = 1+x+x^6$ and $p_2(x) = 1+x+x^3$. This sequence generator can produce 8 Kasami sequences of length 63

$$\begin{aligned}
 \mathbf{a} &= (+1 + 1 + 1 - 1 + 1) \\
 \mathbf{a} &= (+1 + 1 + 1 - 1 - 1 + 1 - 1) \\
 \mathbf{a} &= (+1 + 1 + 1 - 1 - 1 - 1 + 1 - 1 - 1 + 1 - 1) \\
 \mathbf{a} &= (+1 + 1 + 1 + 1 + 1 - 1 - 1 + 1 + 1 - 1 + 1 - 1 + 1)
 \end{aligned}$$

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, as defined in (9.21). For the Barker sequences

$$\phi_{k,k}^a(n) = \begin{cases} 1 & , n = 0 \\ 0, 1/N, \text{ or } -1/N & , 1 \leq |n| \leq N - 1 \end{cases} \quad (9.49)$$

9.2.6 Walsh-Hadamard Sequences

Walsh-Hadamard sequences are obtained by selecting as sequences the rows of a Hadamard matrix \mathbf{H}_M . For $M = 2$ the Hadamard matrix is

$$\mathbf{H}_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}. \quad (9.50)$$

Larger Hadamard matrices are obtained by using the recursion

$$\mathbf{H}_{2M} = \begin{bmatrix} \mathbf{H}_M & \mathbf{H}_M \\ \mathbf{H}_M & -\mathbf{H}_M \end{bmatrix}. \quad (9.51)$$

For example,

$$\mathbf{H}_8 = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \end{bmatrix}. \quad (9.52)$$

The rows in the Hadamard matrix define the Walsh-Hadamard sequences, and have the property that they are mutually orthogonal.

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. This will lead to multiple-access interference in orthogonal CDMA systems.

9.2.6.1 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 Chap. 5. Another possibility is a variant of biorthogonal modulation, where each row of the Hadamard matrix is used to send one bit of information. In this case M bits are sent at one time. This type of modulation requires coherent detection.

9.2.7 Variable Length Orthogonal Codes

For multimedia applications, it is necessary to support a variety of data streams ranging from low to very high bit rates. Quite often these streams are used concurrently and they all use the same spread bandwidth. Consider a system where each data symbol in the highest bit rate stream $R = R_{\max}$ is spread by an orthogonal sequence of length $N = 2^m$. Then the data symbols in a stream with bit rate $R = R_{\max}/2^k$ are 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:

$$\mathbf{c}_{2n} = \begin{bmatrix} \mathbf{c}_{2n,1} \\ \mathbf{c}_{2n,2} \\ \vdots \\ \mathbf{c}_{2n,2n} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \mathbf{c}_{n,1} & \mathbf{c}_{n,1} \\ \mathbf{c}_{n,1} & -\mathbf{c}_{n,1} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \mathbf{c}_{n,n} & \mathbf{c}_{n,n} \\ \mathbf{c}_{n,n} & -\mathbf{c}_{n,n} \end{bmatrix} \end{bmatrix} \tag{9.53}$$

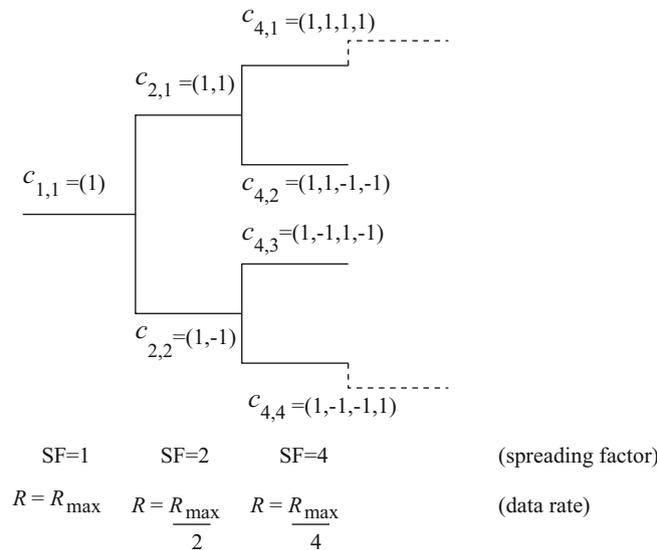


Fig. 9.10 Construction of orthogonal spreading codes with different spreading factors

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.

9.2.8 Complementary Code Keying

Complementary codes have the property that the sum of their aperiodic autocorrelation functions is zero for all delays except zero delay. That is,

$$\frac{1}{M} \sum_{k=1}^M \phi_{k,k}^a(n) = \delta(n). \quad (9.54)$$

A variety of constructions exist for complementary codes and two examples are given here. The IEEE 802.11b standard uses length-8 CCK sequences for 11 Mb/s transmission. The eight complex chip values for CCK code words are

$$\begin{aligned} \mathcal{C} = \{ & e^{j(\phi_1+\phi_2+\phi_3+\phi_4)}, e^{j(\phi_1+\phi_3+\phi_4)}, e^{j(\phi_1+\phi_2+\phi_4)}, \\ & -e^{j(\phi_1+\phi_4)}, e^{j(\phi_1+\phi_2+\phi_3)}, e^{j(\phi_1+\phi_3)}, -e^{j(\phi_1+\phi_2)}, e^{j(\phi_1)} \}, \end{aligned} \quad (9.55)$$

where the phases $\{\phi_1, \phi_2, \phi_3, \phi_4\}$ are QPSK phases. The phases ϕ_2, ϕ_3 , and ϕ_4 each take on four different values, leading to a code alphabet of size 64. The phase ϕ_1 is differentially encoded across successive codewords. Since each of the four phases ϕ_1 to ϕ_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

$$\mathcal{C} = \{ e^{j(\phi_1+\phi_2+\phi_3)}, e^{j(\phi_1+\phi_3)}, e^{j(\phi_1+\phi_2)}, -e^{j(\phi_1)} \}, \quad (9.56)$$

where, again, the phases $\{\phi_1, \phi_2, \phi_3\}$ are QPSK phases.

9.2.9 Zadoff–Chu Sequences

A Zadoff–Chu sequence, sometimes called a Chu sequence [64], is a complex-valued sequence defined in the time domain by the following:

$$a_n^{(m)} = \begin{cases} e^{-j\frac{\pi m n^2}{N}}, & \text{when } N \text{ is even} \\ e^{-j\frac{\pi m n(n+1)}{N}}, & \text{when } N \text{ is odd} \end{cases} \quad 0 \leq n \leq N-1 \quad (9.57)$$

where m is called the sequence index or family, such that m is relatively prime to the sequence length N (meaning that the only integer that divides m and N is 1). The number of Zadoff–Chu sequences for a given length N is the number of integers that are relatively prime to N . Thus, to maximize the number of available Zadoff–Chu sequences for a given sequence length N , N may be chosen to be a prime number.

The DFT of the Zadoff–Chu sequence in (9.57) for N even can be obtained from the table of DFT pairs in [75] as

$$A_k^{(m)} = e^{j\pi\left(\frac{(N-k)^2}{mN} - \frac{1}{4}\right)} \frac{1}{\sqrt{m}} \sum_{\ell=0}^{m-1} \xi^{\ell(n-k) - \ell^2 N/2} \quad (9.58)$$

where $m > 0$, mN is even, and $\xi = e^{-j2\pi/m}$. When $m = 1$ this reduces to

$$A_k^{(1)} = e^{-j\pi/4} e^{j\pi k^2}. \quad (9.59)$$

An important property of Zadoff–Chu sequences is their orthogonality to their cyclic shifts. Lemma 9.1 below provides a general theorem for circular *auto-orthogonality*.

Lemma 9.1. *If and only if a sequence $\{A_k\}$, $k = 0, \dots, N-1$, is circular auto-orthogonal, i.e.,*

$$\frac{1}{N} \sum_{k=0}^{N-1} A_k A_{(k-l)_N}^* = \delta_l = \begin{cases} 1, & l = 0 \\ 0, & l \neq 0 \end{cases}, \quad (9.60)$$

meaning that A_k is orthogonal to any of its cyclic shifts, then the time domain sequence $\{a_n\}$, $k = 0, \dots, N-1$, has constant envelope.

Proof. The proof can be carried out by manipulating properties of the DFT. Since $|a_n|^2 = a_n a_n^*$, $a_n \xrightarrow{\text{DFT}} A_k$, and $a_n^* \xrightarrow{\text{DFT}} A_{(-k)_N}^*$,

$$|a_n|^2 = a_n a_n^* \xrightarrow{\text{DFT}} A_k \otimes A_{(-k)_N}^* = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} A_l A_{(l-k)_N}^*, \quad (9.61)$$

where \otimes is circular convolution. If $\{A_k\}$ is circular auto-orthogonal, then the right side of (9.61) is equal to $\sqrt{N}\delta_k$. This leads to $a_n a_n^* = 1$ and, therefore, $|a_n| = 1$. Conversely, if $|a_n| = 1$, then $a_n a_n^* = 1$, and its DFT is $\sqrt{N}\delta_k$. From (9.61), $A_k \otimes A_{(-k)_N}^* = \sqrt{N}\delta_k$. Therefore, $\{A_k\}$ is circular auto-orthogonal. From the fact $a_n \xrightarrow{\text{DFT}} A_k$ and $|a_n| = |A_k| = 1$ both sequences a_n and A_k are circular auto-orthogonal, based on Lemma 9.1.

Zadoff–Chu sequences have the following useful properties:

- Constant Envelope: The Zadoff–Chu sequences have constant envelope in the time domain.
- Auto-orthogonal: The Zadoff–Chu sequences have a perfect periodic autocorrelation in the time domain

$$\phi_{m,m}(n) = \frac{1}{N_{\text{ZC}}} \sum_{i=0}^{N_{\text{ZC}}-1} a_i^{(m)*} a_{i+n}^{(m)} = \delta(n), \quad n = 0, \dots, N_{\text{ZC}} - 1. \quad (9.62)$$

- Low Cross-correlations: The cross-correlation of two different prime length Zadoff–Chu sequences of family m_1 and m_2 is a constant $1/\sqrt{N_{\text{ZC}}}$ regardless of time shift, provided that $m_1 - m_2$ is relatively prime to N_{ZC} .

Example 9.1. Consider the special DFT pair [75]

$$\begin{aligned} a_n &= e^{-j\frac{\pi}{8}} e^{j\frac{\pi n^2}{N}}, \quad n = 0, \dots, N-1 \\ A_k &= e^{j\frac{\pi}{8}} e^{-j\frac{\pi k^2}{N}}, \quad k = 0, \dots, N-1, \end{aligned} \quad (9.63)$$

This particular DFT pair satisfies all the properties of Zadoff–Chu sequences and, moreover, $A_k = a_k^*$ so the DFT sequence of the a_n is simply its conjugate. In the time domain, the sequence a_n serves as a perfect synchronization sequence, having unity peak-to-average power ratio (PAPR). Meanwhile, the sequence A_k has constant envelope in the frequency domain. This property simplifies the process of channel estimation in the frequency domain since the sub-carriers are excited with known training symbols of equal power, such that the power spectrum of the pilot waveform is flat over the bandwidth of interest.

9.3 Power Spectral Density of DS Spread Spectrum Signals

The DS/QPSK waveform can be thought of as a QPSK waveform where the n th data symbol is transmitted with the amplitude shaping pulse in (9.5). For uncorrelated zero-mean data symbols, the results in Sect. 4.9.1.2 show that the power spectral density (psd) of the DS/QPSK complex envelope is c.f. (4.218)

$$S_{\overline{ss}}(f) = \frac{A^2}{T} \sigma_x^2 |H_a(f)|^2, \quad (9.64)$$

where $h_a(t)$ is the amplitude shaping pulse. In the case of a short code, the amplitude shaping pulse is

$$h_a(t) = \sum_{k=0}^{N-1} a_k h_c(t - kT_c). \quad (9.65)$$

Taking the Fourier transform of $h_a(t)$ gives

$$H_a(f) = H_c(f) \sum_{k=0}^{N-1} a_k e^{-j2\pi f k T_c} \quad (9.66)$$

and

$$|H_a(f)|^2 = |H_c(f)|^2 \sum_{k=0}^{N-1} \sum_{\ell=0}^{N-1} a_k a_\ell^* e^{-j2\pi f(k-\ell)T_c}. \quad (9.67)$$

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

$$|H_a(f)|^2 = |H_c(f)|^2 2N \Phi^a(f), \quad (9.68)$$

where $\Phi^a(f)$ is the discrete-time Fourier transform (DTFT) of the aperiodic autocorrelation function, defined by

$$\Phi^a(f) = \sum_{n=-N+1}^{N-1} \phi^a(n) e^{-j2\pi f n T_c}. \quad (9.69)$$

Using $T = NT_c$ and $\sigma_x^2 = \frac{1}{2} \mathbb{E}[|x_i|^2] = 1/2$ gives

$$S_{\overline{ss}}(f) = \frac{A^2}{T_c} |H_c(f)|^2 \Phi^a(f). \quad (9.70)$$

Observe that the psd depends on both $|H_c(f)|$ and $\Phi^a(f)$. Suppose the spreading sequence has an ideal ‘‘thumbtack’’ aperiodic autocorrelation function

$$\phi^a(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}. \quad (9.71)$$

Then $\Phi^a(f) = 1$ and

$$S_{\overline{ss}}(f) = \frac{A^2}{T_c} |H_c(f)|^2. \quad (9.72)$$

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 $S_{\overline{ss}}(f) = A^2 T_c \text{sinc}^2(fT_c)$. Unfortunately, spreading sequences having the ideal aperiodic autocorrelation function in (9.71) do not exist for any non-trivial length. Consider the two sequences:

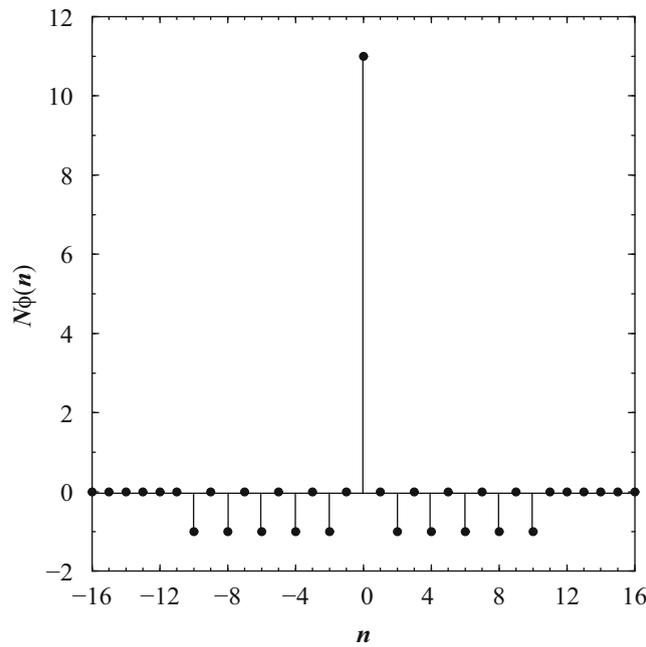


Fig. 9.11 Aperiodic autocorrelation function for the length-11 Barker sequence

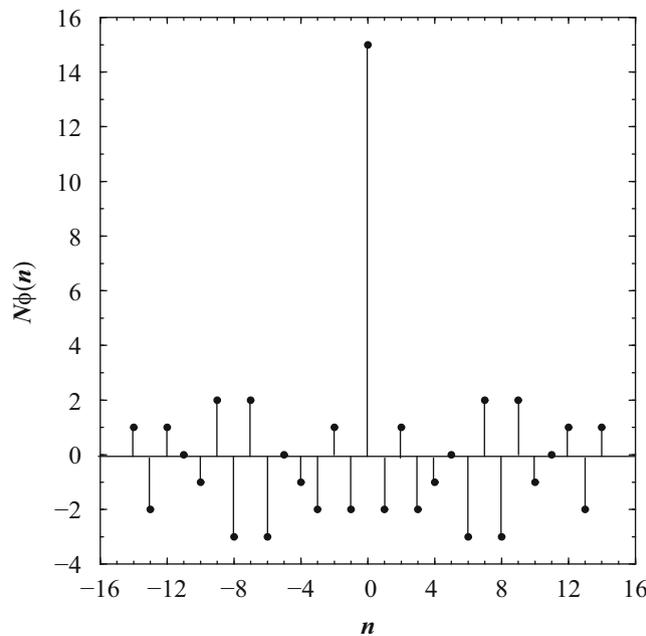


Fig. 9.12 Aperiodic autocorrelation function for the length-15 m -sequence

$$\mathbf{a}^{(1)} = (-1 + 1 - 1 - 1 + 1 - 1 - 1 - 1 + 1 + 1 + 1) \quad (9.73)$$

$$\mathbf{a}^{(2)} = (+1 - 1 - 1 + 1 - 1 - 1 - 1 + 1 + 1 + 1 + 1 - 1 + 1 - 1 + 1).$$

The first is a length-11 Barker sequence and the second is a length-15 m -sequence. The scaled aperiodic autocorrelation functions $N\phi^a(n)$ for these sequences are shown in Figs. 9.11 and 9.12, respectively.

The corresponding power spectral densities with 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 (9.71). This leads to peaks and nulls in the spectrum shown in Fig. 9.14. Such spectral peaks are undesirable.

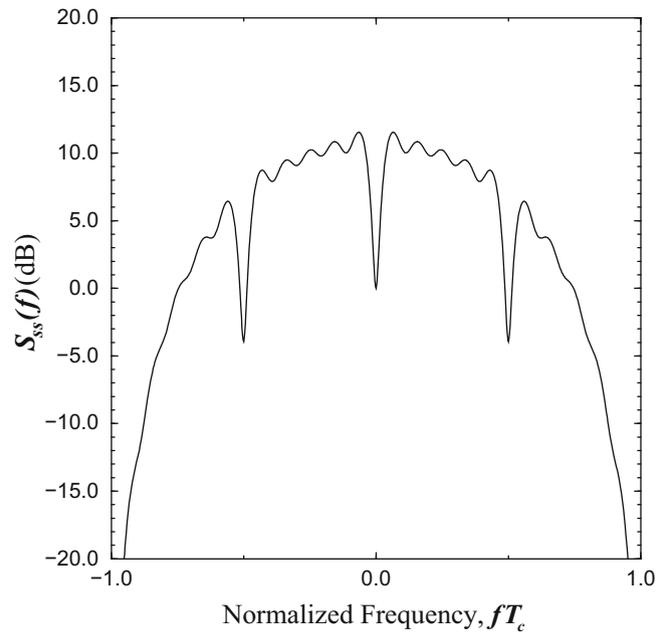


Fig. 9.13 Psd with the length-11 Barker sequence

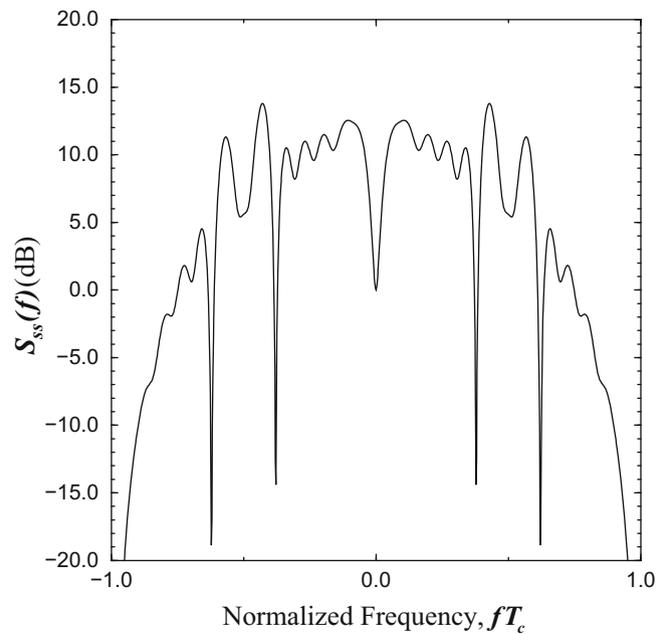


Fig. 9.14 Psd with the length-15 m -sequence

The length-11 Barker sequence is seen to provide a much smoother psd without any large peaks. For this reason, and the fact that a minimum length-10 sequence is required by the IEEE 802.11 WLAN standard, the length-11 Barker sequence was chosen.

It is interesting to note that if CCK is used, then the psd depends on the DTFT of the average aperiodic autocorrelation function in (9.54). In this case,

$$\frac{1}{M} \sum_{k=1}^M \Phi_{k,k}^a(f) = 1, \quad (9.74)$$

and the psd has the ideal form in (9.72). 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.

9.4 Performance of DS/QPSK in CW Tone Interference

Spread spectrum systems must often operate in the presence of narrowband interference. This section examines the effect of continuous wave (CW) tone interference on the performance of DS/QPSK. Consider a DS/QPSK system with dual-channel quaternary spreading as shown in Fig. 9.3. The bandpass DS/QPSK waveform is

$$s(t) = A \sum_n \left(x_{I,n} h_{I,n}(t - nT) \cos(2\pi f_c t) - x_{Q,n} h_{Q,n}(t - nT) \sin(2\pi f_c t) \right), \quad (9.75)$$

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

$$h_{I,n}(t) = \sum_{k=0}^{G-1} a_{I,nG+k} h_c(t - nT_c), \quad h_{Q,n}(t) = \sum_{k=0}^{G-1} a_{Q,nG+k} h_c(t - nT_c). \quad (9.76)$$

With dual-channel quaternary spreading, the energy per modulated symbol is

$$\begin{aligned} E &= \int_0^T s^2(t) dt \\ &= A^2 \int_0^T (x_{I,n}^2 h_{I,n}^2(t) \cos^2(2\pi f_c t) + x_{Q,n}^2 h_{Q,n}^2(t) \sin^2(2\pi f_c t)) \\ &= \frac{A^2}{4} \int_0^T (h_{I,n}^2(t) + h_{Q,n}^2(t)) \\ &= \frac{A^2}{4} \sum_{k=0}^{G-1} (a_{I,nG+k}^2 + a_{Q,nG+k}^2) \int_0^{T_c} h_c^2(t) dt \\ &= G \frac{A^2}{2} \int_0^{T_c} h_c^2(t) dt \\ &= GE_c, \end{aligned} \quad (9.77)$$

where

$$E_c = \frac{A^2}{2} \int_0^{T_c} h_c^2(t) dt \quad (9.78)$$

is the energy per PN chip. Note that (9.78) and (9.2) differ by a factor of 2, because (9.2) assumes complex spreading while (9.78) assumes quadrature spreading. This can be seen by comparing the energy of the bandpass waveforms in (9.6) and (9.75) over the interval $[nT, (n+1)T)$.

The received bandpass signal in the presence of continuous wave (CW) tone interference and additive white Gaussian noise (AWGN) is

$$r(t) = s(t) + n(t) + J(t), \quad (9.79)$$

where $n(t)$ is AWGN with two-sided power spectral density $N_o/2$ and $J(t)$ is the CW tone interference of the form

$$J(t) = A_J \cos(2\pi f_J t + \theta), \quad (9.80)$$

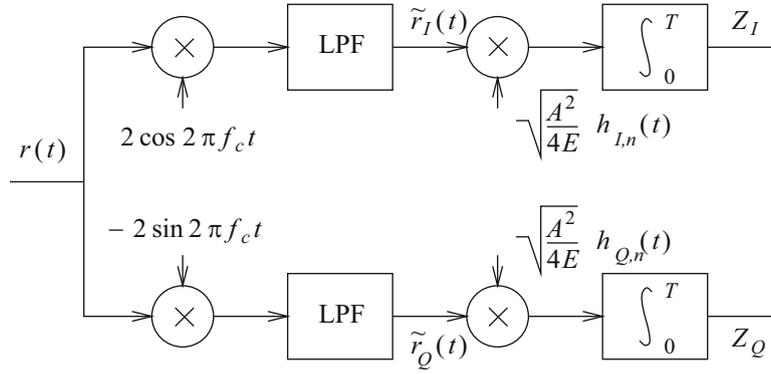


Fig. 9.15 Quadrature demodulator for DS/QPSK

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

$$E_J = \frac{A_J^2 T}{2}. \quad (9.81)$$

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 the signal, noise, and interference terms are considered separately. During the time interval $[nT, (n+1)T)$ the contribution of the signal term to Z_I and Z_Q is

$$\begin{aligned} Z_I(s) &= \int_0^T \tilde{s}_I(t) \sqrt{\frac{A^2}{4E}} h_{I,n}(t) dt \\ &= \int_0^T A x_{I,n} h_{I,n}(t) \sqrt{\frac{A^2}{4E}} h_{I,n}(t) dt \\ &= x_{I,n} A \sqrt{\frac{A^2}{4E}} \int_0^T h_{I,n}^2(t) dt \\ &= x_{I,n} \sqrt{E}, \end{aligned} \quad (9.82)$$

where (9.77) was used. Likewise,

$$Z_Q(s) = x_{Q,n} \sqrt{E}. \quad (9.83)$$

The contribution of the AWGN term to Z_I and Z_Q is

$$Z_I(n) = \int_0^T \tilde{n}_I(t) \sqrt{\frac{A^2}{4E}} h_{I,n}(t) dt \quad (9.84)$$

$$Z_Q(n) = \int_0^T \tilde{n}_Q(t) \sqrt{\frac{A^2}{4E}} h_{Q,n}(t) dt. \quad (9.85)$$

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 CW tone interference term to Z_I and Z_Q can be calculated as follows:

$$\begin{aligned} Z_I(J) &= \int_0^T J(t) 2 \cos(2\pi f_c t) \sqrt{\frac{A^2}{4E}} h_{I,n}(t) dt \\ &= \int_0^T A_J \cos(2\pi f_J t) 2 \cos(2\pi f_c t) \sqrt{\frac{A^2}{4E}} h_{I,n}(t) dt \end{aligned}$$

$$\begin{aligned}
&= A_J \sqrt{\frac{A^2}{4E}} \int_0^T 2h_{I,n}(t) \cos(2\pi f_c t) \cos(2\pi f_J t + \theta) dt \\
&= A_J \sqrt{\frac{A^2}{4E}} \int_0^T h_{I,n}(t) \{ \cos(2\pi \Sigma_f t + \theta) + \cos(2\pi \Delta_f t + \theta) \} dt,
\end{aligned} \tag{9.86}$$

where

$$\Sigma_f = f_c + f_J \tag{9.87}$$

$$\Delta_f = f_J - f_c. \tag{9.88}$$

Using $A_J = \sqrt{2E_J/T}$ gives

$$Z_I(J) = \sqrt{E_J} \sqrt{\frac{A^2 T}{2E}} \frac{1}{T} \int_0^T h_{I,n}(t) \left(\cos(2\pi \Sigma_f t + \theta) + \cos(2\pi \Delta_f t + \theta) \right) dt \tag{9.89}$$

Finally, using

$$E = G \frac{A^2}{2} \int_0^{T_c} h_c^2(t) dt \tag{9.90}$$

gives

$$Z_I(J) = \sqrt{E_J/\bar{h}_c} \frac{1}{T} \int_0^T h_{I,n}(t) \left(\cos(2\pi \Sigma_f t + \theta) + \cos(2\pi \Delta_f t + \theta) \right) dt, \tag{9.91}$$

where

$$\bar{h}_c = \frac{1}{T_c} \int_0^{T_c} h_c(t) dt. \tag{9.92}$$

Using further trigonometric identities gives

$$\begin{aligned}
Z_I(J) &= \sqrt{E_J/\bar{h}_c} \left(\cos(\theta) \frac{1}{T} \int_0^T h_{I,n}(t) \left(\cos(2\pi \Sigma_f t) + \cos(2\pi \Delta_f t) \right) dt \right. \\
&\quad \left. - \sin(\theta) \frac{1}{T} \int_0^T h_{I,n}(t) \left(\sin(2\pi \Sigma_f t) + \sin(2\pi \Delta_f t) \right) dt \right).
\end{aligned} \tag{9.93}$$

In a similar fashion

$$\begin{aligned}
Z_I(J) &= \int_0^T J(t) 2 \sin(2\pi f_c t) \sqrt{\frac{A^2}{4E}} h_{Q,n}(t) dt \\
&= \sqrt{E_J/\bar{h}_c} \left(\cos(\theta) \frac{1}{T} \int_0^T h_{Q,n}(t) \left(\sin(2\pi \Delta_f t) - \sin(2\pi \Sigma_f t) \right) dt \right. \\
&\quad \left. + \sin(\theta) \frac{1}{T} \int_0^T h_{Q,n}(t) \left(\cos(2\pi \Delta_f t) - \cos(2\pi \Sigma_f t) \right) dt \right).
\end{aligned} \tag{9.94}$$

Combining the signal, noise, and CW tone interference terms gives

$$\begin{aligned}
Z_I &= Z_I(s) + Z_I(n) + Z_I(J) \\
Z_Q &= Z_Q(s) + Z_Q(n) + Z_Q(J).
\end{aligned} \tag{9.95}$$

Hence, Z_I and Z_Q are independent Gaussian random variables with means

$$\begin{aligned} E[Z_I] &= x_{I,n}\sqrt{E} + I_I \sqrt{E_J/\bar{h}_c} \\ E[Z_Q] &= x_{Q,n}\sqrt{E} + I_Q \sqrt{E_J/\bar{h}_c}, \end{aligned} \quad (9.96)$$

and variance $N_o/2$, where

$$\begin{aligned} I_I &= \cos(\theta) \frac{1}{T} \int_0^T h_{I,n}(t) \left(\cos(2\pi \Sigma_f t) + \cos(2\pi \Delta_f t) \right) dt \\ &\quad - \sin(\theta) \frac{1}{T} \int_0^T h_{I,n}(t) \left(\sin(2\pi \Sigma_f t) + \sin(2\pi \Delta_f t) \right) dt \end{aligned} \quad (9.97)$$

$$\begin{aligned} I_Q &= \cos(\theta) \frac{1}{T} \int_0^T h_{Q,n}(t) \left(\sin(2\pi \Delta_f t) - \sin(2\pi \Sigma_f t) \right) dt \\ &\quad + \sin(\theta) \frac{1}{T} \int_0^T h_{Q,n}(t) \left(\cos(2\pi \Delta_f t) - \sin(2\pi \Sigma_f t) \right) dt. \end{aligned} \quad (9.98)$$

9.4.1 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 (9.96), 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 in-phase and quadrature components of the modulated carrier such that

$$h(t) = h_{I,n}(t) = h_{Q,n}(t) = \sum_{k=0}^{G-1} a_k u_{T_c}(t - kT_c). \quad (9.99)$$

It follows that

$$\begin{aligned} I_I &= \cos(\theta) \frac{1}{NT_c} \int_0^{NT_c} \sum_{k=0}^{N-1} a_k u_{T_c}(t - kT_c) \left(\cos(2\pi \Sigma_f t) + \cos(2\pi \Delta_f t) \right) dt \\ &\quad - \sin(\theta) \frac{1}{NT_c} \int_0^{NT_c} \sum_{k=0}^{N-1} a_k u_{T_c}(t - kT_c) \left(\sin(2\pi \Sigma_f t) + \sin(2\pi \Delta_f t) \right) dt \\ &= \frac{1}{N} \sum_{k=0}^{N-1} a_k \left(\cos(\theta) \int_k^{k+1} \left(\cos(2\pi \Sigma_f T_c t) + \cos(2\pi \Delta_f T_c t) \right) dt \right. \\ &\quad \left. - \sin(\theta) \int_k^{k+1} \left(\sin(2\pi \Sigma_f T_c t) + \sin(2\pi \Delta_f T_c t) \right) dt \right). \end{aligned} \quad (9.100)$$

Likewise,

$$\begin{aligned} I_Q &= \frac{1}{N} \sum_{k=0}^{N-1} a_k \left(\cos(\theta) \int_k^{k+1} \left(\sin(2\pi \Delta_f T_c t) - \sin(2\pi \Sigma_f T_c t) \right) dt \right. \\ &\quad \left. + \sin(\theta) \int_k^{k+1} \left(\cos(2\pi \Delta_f T_c t) - \sin(2\pi \Sigma_f T_c t) \right) dt \right). \end{aligned} \quad (9.101)$$

Fortunately, the above integrals exist in closed form. Defining

$$\alpha \triangleq 2\pi \Sigma_f T_c \quad (9.102)$$

$$\beta \triangleq 2\pi \Delta_f T_c \quad (9.103)$$

yields

$$I_I = \frac{1}{N} \sum_{k=0}^{N-1} a_k \left(\cos(\theta) \left(\frac{\sin((k+1)\alpha) - \sin(k\alpha)}{\alpha} + \frac{\sin((k+1)\beta) - \sin(k\beta)}{\beta} \right) - \sin(\theta) \left(\frac{\cos(k\alpha) - \cos((k+1)\alpha)}{\alpha} + \frac{\cos(k\beta) - \cos((k+1)\beta)}{\beta} \right) \right) \quad (9.104)$$

and

$$I_Q = \frac{1}{N} \sum_{k=0}^{N-1} a_k \left(\cos(\theta) \left(\frac{\cos(k\beta) - \cos((k+1)\beta)}{\beta} - \frac{\cos(k\alpha) - \cos((k+1)\alpha)}{\alpha} \right) + \sin(\theta) \left(\frac{\sin((k+1)\beta) - \sin(k\beta)}{\beta} - \frac{\cos(k\alpha) - \cos((k+1)\alpha)}{\alpha} \right) \right). \quad (9.105)$$

Due to the random phase of the CW tone interferer, the CW tone interference is circularly symmetric similar to 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 CW tone interference, the probability of correct symbol reception is

$$P[c] = (1 - P_b)^2, \quad (9.106)$$

where

$$P_b = Q\left(\sqrt{2\gamma_b}\right) \quad (9.107)$$

is the bit error probability, and $\gamma_b = E_b/N_o$ is the received bit energy-to-noise ratio.

When CW tone interference is present, the probability of correct reception is

$$P_{C|b_0b_1} = (1 - P_{b1})(1 - P_{b2}). \quad (9.108)$$

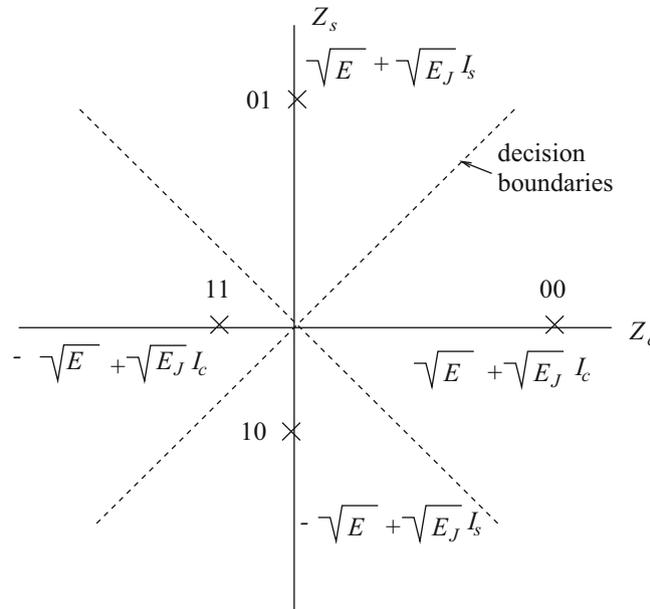


Fig. 9.16 QPSK signal constellation with tone interference

Observe that the error probability depends on the transmitted symbol and the interference impairment I_I and I_Q . Referring to Fig. 9.16

$$P_{C|00} = (1 - P_{b1})(1 - P_{b2}) \quad (9.109)$$

where

$$P_{b1} = P_{b2} = Q \left(\sqrt{\frac{2E_b}{N_o} \left(1 + 2\sqrt{E_J/E}I_I + (E_J/E)^2I_I^2 \right)} \right). \quad (9.110)$$

Hence,

$$P_{b|00} = Q \left(\sqrt{\frac{2E_b}{N_o} \left(1 + 2\sqrt{E_J/E}I_I + (E_J/E)^2I_I^2 \right)} \right). \quad (9.111)$$

In a similar fashion,

$$P_{b|01} = Q \left(\sqrt{\frac{2E_b}{N_o} \left(1 + 2\sqrt{E_J/E}I_Q + (E_J/E)^2I_Q^2 \right)} \right) \quad (9.112)$$

$$P_{b|11} = Q \left(\sqrt{\frac{2E_b}{N_o} \left(1 - 2\sqrt{E_J/E}I_I + (E_J/E)^2I_I^2 \right)} \right) \quad (9.113)$$

$$P_{b|10} = Q \left(\sqrt{\frac{2E_b}{N_o} \left(1 - 2\sqrt{E_J/E}I_Q + (E_J/E)^2I_Q^2 \right)} \right). \quad (9.114)$$

Since all symbols are equally likely, the bit error probability is

$$P_b = \frac{1}{4}(P_{b|00} + P_{b|11} + P_{b|10} + P_{b|01}). \quad (9.115)$$

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 be calculated by averaging (9.115) over random phase of the CW tone interferer.

Figure 9.17 shows the bit error probability when the length-15 m -sequence $\mathbf{a}^{(2)}$ in (9.73) is used as a short code ($G = 15$). Figure 9.17 arbitrarily assumes that $f_c = 280$ MHz, and $T_c = 191 \times 10^{-9}$ s. Observe that the bit error probability varies greatly with the frequency of the CW tone interferer. It is interesting to note that an interfering tone placed at the carrier frequency f_c does not give the worst case performance. Also, the bit error probability is seen to exhibit an error floor due to the AWGN.

Figure 9.18 shows the bit error probability when the length-11 Barker sequence $\mathbf{a}^{(1)}$ in (9.73) 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.

Figure 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 CW tone interferer and AWGN. Likewise, Fig. 9.20 inverts Fig. 9.18 for the length-11 Barker sequence. Observe that the sensitivity to CW 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 in-phase and quadrature channels are spread by using the amplitude shaping pulse

$$h(t) = \sum_{k=0}^{N-1} a_k h_c(t - kT_c), \quad (9.116)$$

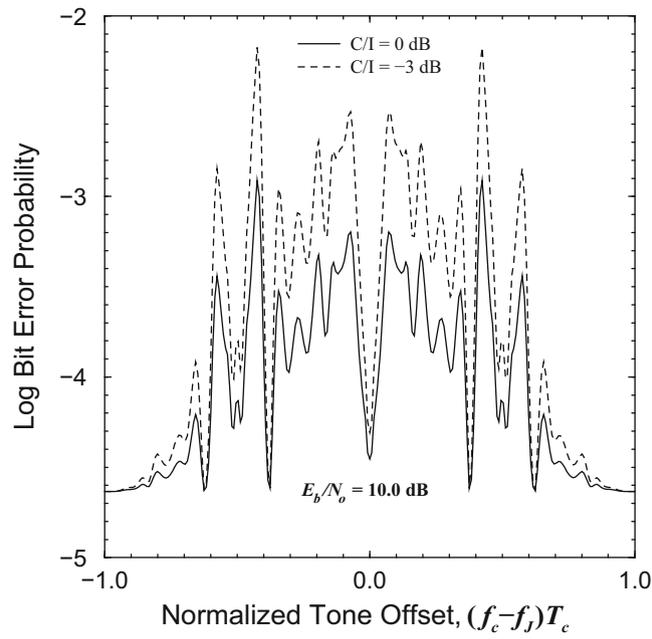


Fig. 9.17 Bit error probability with length-15 m -sequence

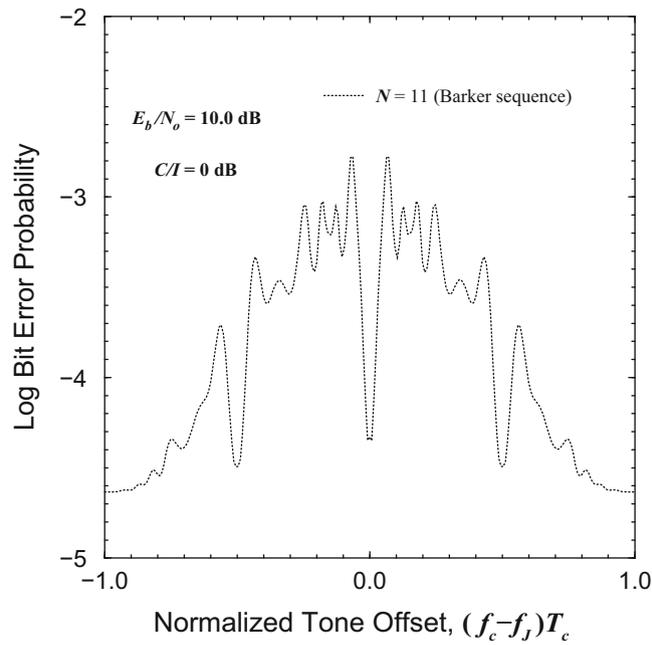


Fig. 9.18 Bit error probability with the length-11 Barker sequence

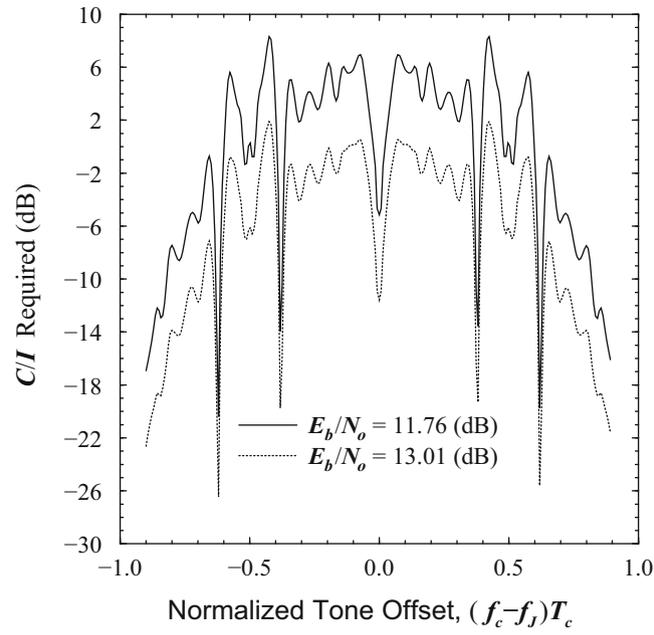


Fig. 9.19 Required C/I to achieve 10^{-6} bit error rate with a length-15 m -sequence

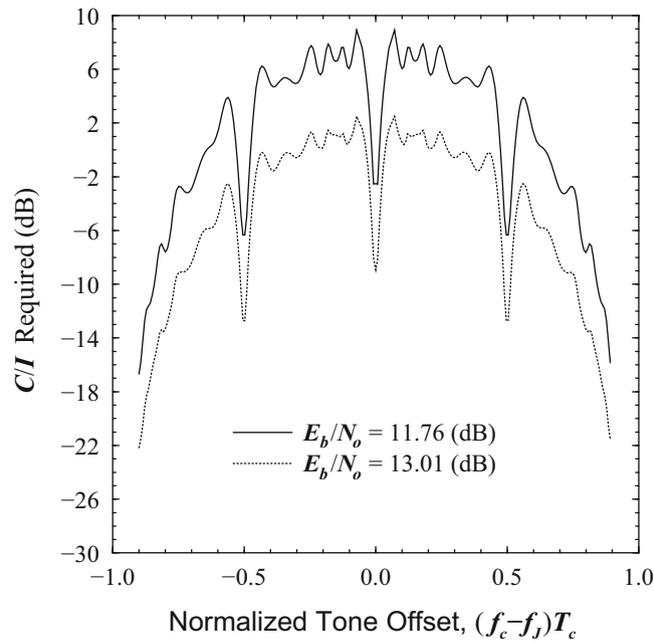


Fig. 9.20 Required C/I to achieve 10^{-6} bit error rate with a length-11 Barker sequence

where $\{a_k\}_{k=0}^{N-1}$ is the periodic spreading sequence of length N . After quadrature demodulation, the receiver employs a correlator or matched filter detector having the impulse response⁴

$$\begin{aligned} h_r(t) &= h^*(NT_c - t) \\ &= \sum_{k=0}^{N-1} a_{N-k} h_c(t - kT_c). \end{aligned} \quad (9.117)$$

The filter $h_r(t)$ has transfer function

$$\begin{aligned} H_r(f) &= \int_{-\infty}^{\infty} h_r(t) e^{-j2\pi ft} dt \\ &= \int_{-\infty}^{\infty} \sum_{k=0}^{N-1} a_{N-k} h_c(t - kT_c) e^{-j2\pi ft} dt \\ &= \sum_{k=0}^{N-1} a_{N-k} \int_{-\infty}^{\infty} h_c(t - kT_c) e^{-j2\pi ft} dt \\ &= H_c(f) \sum_{k=0}^{N-1} a_{N-k} e^{-j2\pi f k T_c} \\ &= H_c(f) A(f), \end{aligned} \quad (9.118)$$

where

$$A(f) = \sum_{k=0}^{N-1} a_{N-k} e^{-j2\pi f k T_c}. \quad (9.119)$$

For a rectangular chip shaping function $h_c(t) = u_{T_c}(t)$,

$$H_c(f) = \text{sinc}(fT_c) e^{-j\pi f T_c}. \quad (9.120)$$

The corresponding amplitude response $|H_r(f)|$ for the length-15 m -sequence and the length-11 Barker sequence is identical in form to the corresponding transmitted psds shown in Figs. 9.14 and 9.13, respectively. The frequencies where $|H_r(f)|$ has the highest relative gain are the exact same frequencies where the CW tone interferer causes a large error probability. If the length of the short code is increased, the sensitivity to CW tone interference will not necessarily diminish. To make the receiver less sensitive to CW tone interference, the aperiodic autocorrelation function in (9.21) should be as close to ideal as possible. In other words, the power spectrum $\Phi_{k,k}(f)$ defined in (9.69) is as flat as possible. Although some types of sequences, such as Gold and Kasami sequences, have excellent cross-correlation properties, their aperiodic autocorrelation functions are usually far from ideal resulting in a power spectrum $\Phi_{k,k}(f)$ is typically full of peaks and nulls. The Barker sequences have the best aperiodic autocorrelation properties for a given sequence length N when they exist, and will result in the least sensitivity to tone interference.

9.4.2 Short Code Design

Sometimes it is desirable to construct short sequences of a given sequence length N that have good aperiodic autocorrelation properties with a corresponding power spectrum that is as flat as possible. One possibility is to design such sequences based on a MMSE criterion based on their aperiodic autocorrelation function in (9.21). The ideal aperiodic autocorrelation function

⁴The usual case is assumed, where $h_c(-t) = h_c(t)$.

is the perfect “thumb-tack” function $\phi_{k,k}^{a(\text{ideal})}(n) = \delta(n)$. In this case, the discrete-time Fourier transform of $\phi_{k,k}^{a(\text{ideal})}(n)$ gives the flat power spectrum $\Phi^a(f) = 1$. However, spreading codes having ideal aperiodic autocorrelation functions do not exist for any length. Here, short spreading sequences are designed to minimize the mean square error

$$\varepsilon = \frac{1}{N} \sum_{n=0}^{N-1} \left(\phi_{k,k}^a(n) - \phi_{k,k}^{a(\text{ideal})}(n) \right)^2 \quad (9.121)$$

$$= \frac{1}{N} \sum_{n=1}^{N-1} \left(\phi_{k,k}^a(n) \right)^2, \quad (9.122)$$

where the summation may be started at $n = 1$ since $\phi_{k,k}^a(0) = 1$ for any sequence. Unfortunately, there is no easy analytical method to find such sequences. However, for relatively small N the sequences can be found by an exhaustive search. For sequence lengths $N = 2, 3, 4, 5, 7, 11$ and 13 , the above process will generate the Barker sequences in Sect. 9.2.5, along with their mirror images (time reversed sequences). So the Barker sequences are optimal in a minimum mean square sense.

Example 9.2. Suppose our objective is to find MMSE binary spreading sequences of length $N = 15$. In this case, the following four sequences will minimize the mean square error in (9.122):

$$\mathbf{x}_1 = \{-1 -1 -1 -1 -1 +1 +1 -1 -1 +1 +1 -1 +1 -1 +1\}$$

$$\mathbf{x}_2 = \{+1 -1 +1 -1 +1 +1 -1 -1 +1 +1 -1 -1 -1 -1 -1\}$$

$$\mathbf{x}_3 = \{-1 -1 -1 +1 +1 +1 -1 +1 +1 +1 -1 +1 +1 -1 +1\}$$

$$\mathbf{x}_4 = \{+1 -1 +1 +1 -1 +1 +1 +1 -1 +1 +1 +1 -1 -1 -1\}$$

Observe that \mathbf{x}_1 and \mathbf{x}_3 are the mirror images (time reversals) of \mathbf{x}_2 and \mathbf{x}_4 , respectively. The aperiodic autocorrelation function for all four sequences is identical and the scaled version $N\phi^a(n)$ is plotted in Fig. 9.21. Each sequence has $\varepsilon = 1$ with a maximum “off-peak” aperiodic autocorrelation value equal to $\max_n N\phi^a(n) = 3$.

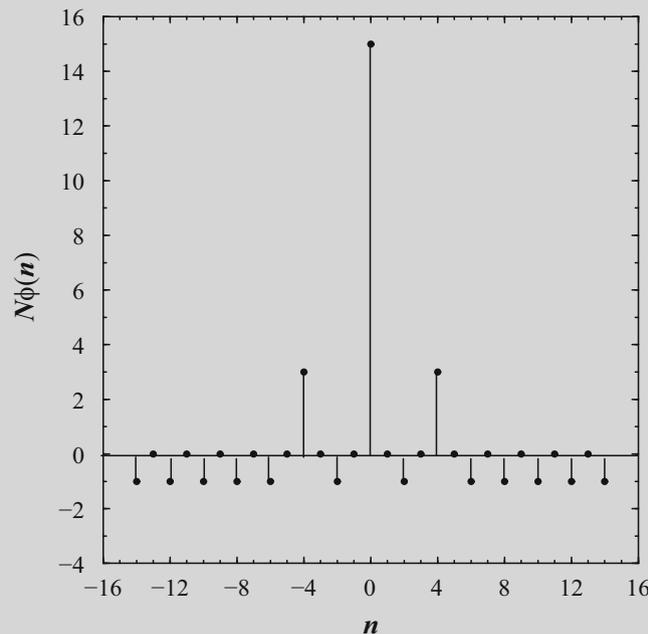


Fig. 9.21 Aperiodic autocorrelation for optimized length-15 sequences

(continued)

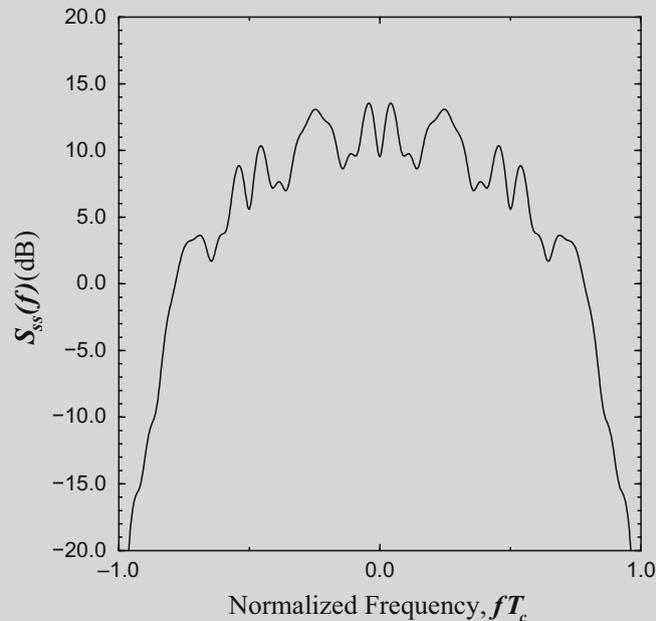
Example 9.2 (continued)

Fig. 9.22 Power spectrum with optimized length-15 sequences

The corresponding power spectral density achieved with these sequences is shown in Fig. 9.22 for the case of rectangular chip shaping, $h_c(t) = u_{T_c}(t)$. Note that the power spectrum exhibits considerably less variation than that corresponding to the length-15 m -sequence in Fig. 9.14, but is not as good as that corresponding to the length-11 Barker sequence in Fig. 9.13.

9.4.3 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.

Figure 9.23 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 $1 + x + x^{11}$. Figure 9.24 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 CW tone frequency as compared to a short code. For sequence lengths of 127 and 2047, the bit error probability is maximized when $f_I = f_c$. 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 with all possible binary N -tuples, thus leading to the irregularities observed in Fig. 9.23. It is interesting to note that the length-2047 m -sequence seems to be more sensitive to an interfering tone at the carrier frequency 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 9.23 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) with a processing gain of 15 only over 4 narrow ranges of interfering tone frequencies.

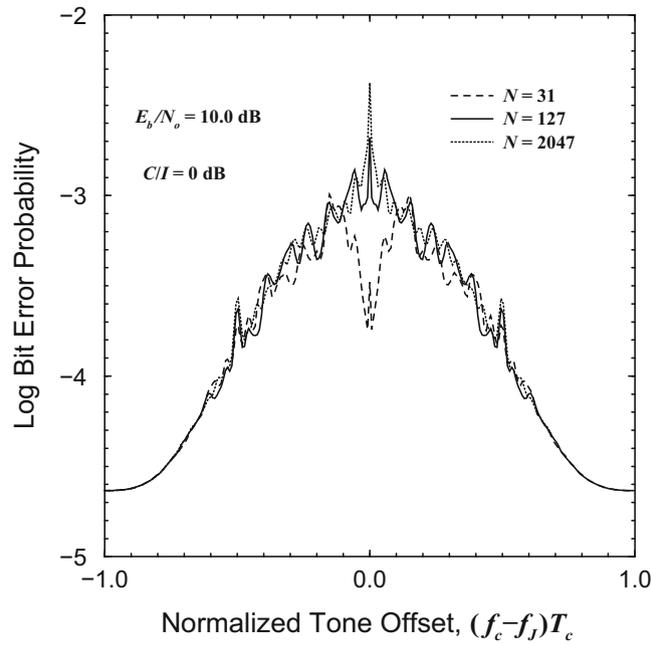


Fig. 9.23 Bit error probability with length-31, 127, and 2047 m -sequences

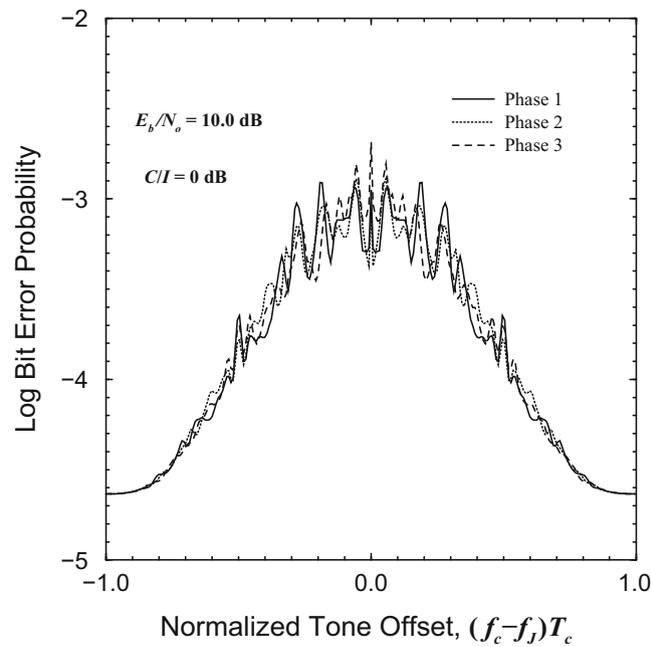


Fig. 9.24 Bit error probability with length-63 m -sequence

9.5 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, root-raised cosine pulse shaping. Since the low-pass signal $\tilde{s}(t)$ is band-limited to $|f| \leq W/2$, 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

$$\tilde{s}_\delta(t) = \sum_{n=-\infty}^{\infty} \tilde{s}\left(\frac{n}{W}\right) \delta\left(t - \frac{n}{W}\right) \quad (9.123)$$

$$= \tilde{s}(t) \sum_{n=-\infty}^{\infty} \delta\left(t - \frac{n}{W}\right). \quad (9.124)$$

Taking the Fourier transform of both sides of (9.124) gives

$$\begin{aligned} \tilde{S}_\delta(f) &= \tilde{S}(f) * W \sum_{n=-\infty}^{\infty} \delta(f - nW) \\ &= W \sum_{n=-\infty}^{\infty} \tilde{S}(f) * \delta(f - nW) \\ &= W \sum_{n=-\infty}^{\infty} \tilde{S}(f - nW). \end{aligned} \quad (9.125)$$

From (9.125),

$$\tilde{S}(f) = \frac{1}{W} \tilde{S}_\delta(f), \quad 0 \leq |f| \leq W/2. \quad (9.126)$$

Another useful expression can be obtained by taking the Fourier transform of both sides of (9.123) giving

$$\tilde{S}_\delta(f) = \sum_{n=-\infty}^{\infty} \tilde{s}\left(\frac{n}{W}\right) e^{-j2\pi n f/W}. \quad (9.127)$$

Combining (9.126) and (9.127) gives

$$\tilde{S}(f) = \frac{1}{W} \sum_{n=-\infty}^{\infty} \tilde{s}\left(\frac{n}{W}\right) e^{-j\pi n f/W}, \quad 0 \leq |f| \leq W/2. \quad (9.128)$$

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

$$\tilde{r}(t) = \int_{-\infty}^{\infty} \tilde{S}(f) T(f, t) e^{j2\pi f t} df. \quad (9.129)$$

Substituting $\tilde{S}(f)$ from (9.128) gives

$$\begin{aligned} \tilde{r}(t) &= \frac{1}{W} \sum_{n=-\infty}^{\infty} \tilde{s}\left(\frac{n}{W}\right) \int_{-\infty}^{\infty} T(f, t) e^{-j2\pi f(t-n/W)} df \\ &= \frac{1}{W} \sum_{n=-\infty}^{\infty} \tilde{s}\left(\frac{n}{W}\right) g\left(t - \frac{n}{W}, t\right) \\ &= \frac{1}{W} \sum_{n=-\infty}^{\infty} \tilde{s}\left(t - \frac{n}{W}\right) g\left(\frac{n}{W}, t\right), \end{aligned} \quad (9.130)$$

where $g(\tau, t)$ is the time-variant impulse response of the channel. By defining

$$g_n(t) = \frac{1}{W} g\left(\frac{n}{W}, t\right) \quad (9.131)$$

the noiseless received complex envelope can be written as

$$\tilde{r}(t) = \sum_{n=-\infty}^{\infty} g_n(t) \tilde{s}\left(t - \frac{n}{W}\right) \quad (9.132)$$

and it follows that the complex low-pass impulse response of the channel is

$$g(t, \tau) = \sum_{n=-\infty}^{\infty} g_n(t) \delta\left(\tau - \frac{n}{W}\right). \quad (9.133)$$

For WWSUS channels, the $\{g_n(t)\}$ in (9.131) are independent complex Gaussian random processes. For all practical purposes, the channel will be causal with an impulse response that is non-zero 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

$$g(t, \tau) = \sum_{n=1}^L g_n(t) \delta\left(\tau - \frac{n}{W}\right). \quad (9.134)$$

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. 9.25. 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 Sect. 2.5.6.

If ideal Nyquist chip amplitude pulse shaping is used such that $h_c(t) = \text{sinc}(t/T_c)$, 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 root-raised cosine pulse, then the tapped delay line channel model in Fig. 9.25 is not T_c -spaced, e.g., a root-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 approximated as consisting of uncorrelated T_c -spaced rays, i.e.,

$$g(t, \tau) = \sum_{n=-\infty}^{\infty} g_n(t) \delta(\tau - nT_c). \quad (9.135)$$

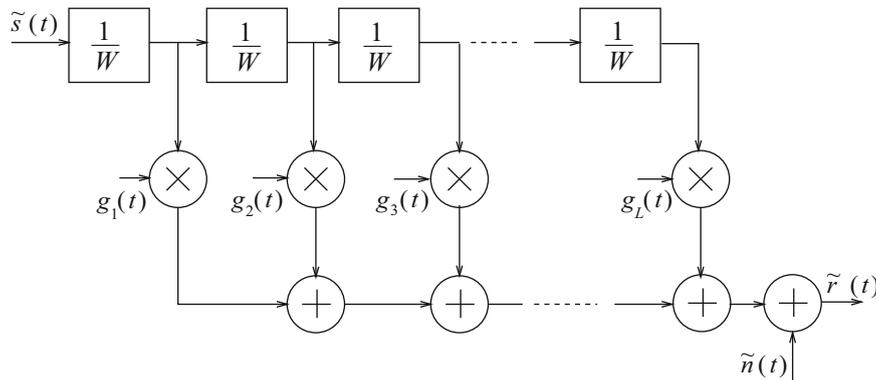


Fig. 9.25 Tapped delay line model of a frequency-selective fading channel, from [272]

9.5.1 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 correlator or matched filter detector. 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 to simultaneously detect all the signals that are present. This section concentrates on conventional detectors for DS spread spectrum signaling on multipath-fading channels.

A simple type of conventional detector uses the autocorrelation properties of the spreading sequences to reject the multipath interference [136, 137]. Sometimes this is called a multipath rejection receiver. Another approach is to exploit 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 $\tilde{s}_m(t)$, $m = 0, \dots, M-1$ are transmitted at each baud epoch. With the frequency-selective fading channel shown in Fig. 9.25, the corresponding received complex envelope is

$$\tilde{r}(t) = \sum_{\ell=1}^L g_{\ell}(t) \tilde{s}_m \left(t - \frac{\ell}{W} \right) + \tilde{n}(t) = \hat{s}_m(t) + \tilde{n}(t) \quad (9.136)$$

where

$$\hat{s}_m(t) = \sum_{\ell=1}^L g_{\ell}(t) \tilde{s}_m \left(t - \frac{\ell}{W} \right). \quad (9.137)$$

As discussed in Sect. 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

$$\begin{aligned} \mu(m) &= \operatorname{Re} \left\{ \int_0^T \tilde{r}(t) \hat{s}_m^*(t) dt \right\} - E_{\hat{m}} \\ &= \operatorname{Re} \left\{ \int_0^T \tilde{r}(t) \sum_{\ell=1}^L g_{\ell}^*(t) \tilde{s}_m^*(t - \ell/W) dt \right\} - E_{\hat{m}}, \end{aligned} \quad (9.138)$$

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 (9.138) 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. 9.26. By changing the variable of integration in (9.138) an alternate form of the RAKE receiver can be obtained as shown in Fig. 9.27. In this case the waveform $\tilde{s}_m(t)$ is correlated with delayed versions of the received complex envelope $\tilde{r}(t)$. The receivers in Figs. 9.26 and 9.27 were first derived by Price and Green [270], and are commonly called RAKE receivers due to their function and resemblance to an ordinary garden rake.

9.5.2 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

$$\tilde{s}_0(t) = -\tilde{s}_1(t) = Ah(t), \quad (9.139)$$

where

$$h(t) = \sum_{k=0}^{N-1} a_k h_c(t - nT_c). \quad (9.140)$$

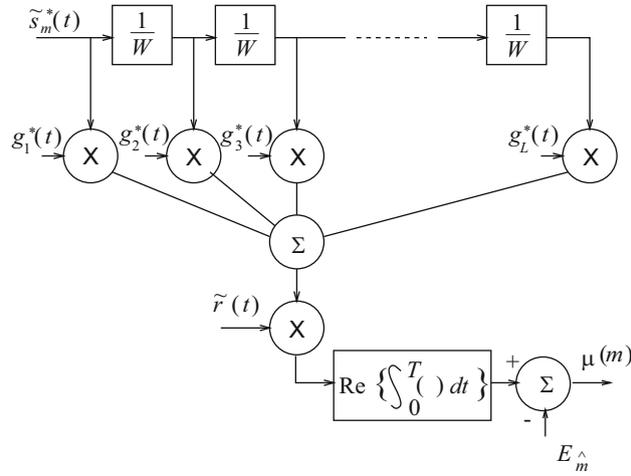


Fig. 9.26 RAKE receiver for DS/QPSK signals

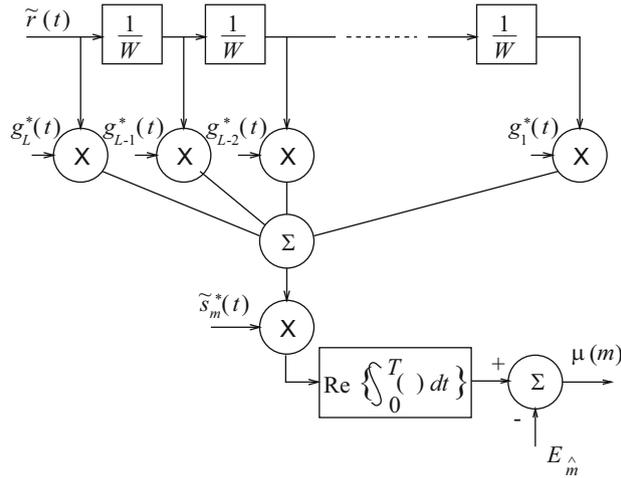


Fig. 9.27 Alternate form of RAKE receiver for DS/QPSK signals

With DS/BPSK the received waveforms $\hat{s}_m(t)$ have equal energy so the bias term $E_{\hat{n}}$ in (9.138) is not needed. Assume that $\tilde{s}_0(t)$ is transmitted. Then using (9.138)

$$\mu = \sum_{m=1}^L \sum_{\ell=1}^L \text{Re} \left\{ g_m g_\ell^* \int_0^T \tilde{s}_0(t - m/W) \tilde{s}_0^*(t - \ell/W) dt \right\} + \tilde{n}, \tag{9.141}$$

where

$$\tilde{n} = \sum_{m=1}^L \text{Re} \left\{ g_m^* \int_0^T \tilde{n}(t) \tilde{s}_0^*(t - m/W) dt \right\} \tag{9.142}$$

and $g_m = \alpha_m e^{j\phi_m}$. The random variable \tilde{n} is Gaussian with zero mean and variance

$$\sigma_{\tilde{n}}^2 = 2EN_o \sum_{m=1}^L \alpha_m^2. \tag{9.143}$$

In general, the integral in (9.141) 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 $h_c(t) = \text{sinc}(t/T_c)$ with bandwidth $W = 1/T_c$ is used. Strictly speaking this pulse is non-causal so the limits of integration in (9.141) must be from $-\infty$ to ∞ . This leads to⁵

$$\begin{aligned}
I &\equiv \int_{-\infty}^{\infty} \tilde{s}_0(t - m/W) \tilde{s}_0^*(t - \ell/W) dt \\
&= A^2 \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} a_k a_j \int_{-\infty}^{\infty} h_c(t - (m+k)T_c) h_c(t - (\ell+j)T_c) dt \\
&= A^2 \sum_{k=0}^{N-1} a_k a_{m+k-\ell} \int_{-\infty}^{\infty} h_c^2(t) dt \\
&= 2E_c N \phi_{aa}(m - \ell) = 2E \phi_{aa}(m - \ell)
\end{aligned} \tag{9.144}$$

where the second last step follows under the assumption of a short code. Therefore, (9.141) becomes

$$\mu = 2E \sum_{m=1}^L \alpha_m^2 + 2E \sum_{m=1}^L \sum_{\substack{\ell=1 \\ \ell \neq m}}^L \text{Re} \{g_m g_\ell^*\} \phi_{aa}(m - \ell) + \tilde{n}. \tag{9.145}$$

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

$$\begin{aligned}
Y_{m,\ell} &= \text{Re} \{g_m g_\ell^*\} \\
&= \alpha_m \cos(\phi_m) \alpha_\ell \cos(\phi_\ell) + \alpha_m \sin(\phi_m) \alpha_\ell \sin(\phi_\ell).
\end{aligned} \tag{9.146}$$

Define the new random variables

$$X_{I,k} = \alpha_k \cos(\phi_k) \quad X_{Q,k} = \alpha_k \sin(\phi_k). \tag{9.147}$$

Then

$$\alpha_k = \sqrt{X_{I,k}^2 + X_{Q,k}^2} \quad \phi_k = \text{Tan}^{-1} \frac{X_{I,k}}{X_{Q,k}}. \tag{9.148}$$

Therefore,

$$Y_{m,\ell} = X_{I,m} X_{I,\ell} + X_{I,m} X_{I,\ell}. \tag{9.149}$$

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

$$p_{Y_{m,\ell}}(y) = \frac{1}{2\sigma_m \sigma_\ell} \exp \left\{ -\frac{|y|}{\sigma_m \sigma_\ell} \right\}. \tag{9.150}$$

Making the substitution for $Y_{m,\ell}$ and rearranging the sum in the second term in (9.145) gives

$$\mu = 2E \sum_{m=1}^L \alpha_m^2 + 4E \sum_{k=1}^{L-1} \sum_{i=1}^{L-1-k} Y_{i,i+k} \phi_{aa}(k) + \tilde{n}. \tag{9.151}$$

⁵Since DS/BPSK signaling is used the spreading sequence \mathbf{a} is real with autocorrelation function $\phi_{aa}(n) = E[a_i a_{i+n}]$.

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 autocorrelation 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 applications, 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 channels having impulse responses up to nT_c seconds long.

If the spreading sequences have an ideal autocorrelation function, i.e., $\phi_{aa}(n-m) = \delta_{nm}$, then there is no self-interference and (9.151) becomes

$$\mu = 2E \sum_{m=1}^L \alpha_m^2 + \tilde{n}. \quad (9.152)$$

Under this condition, the probability of bit error is

$$P_b(\gamma_b) = Q\left(\sqrt{2\gamma_b}\right), \quad (9.153)$$

where γ_b is the received bit energy-to-noise ratio given by

$$\gamma_b = \frac{1}{\sigma_n^2} \left(2E \sum_{m=1}^L \alpha_m^2\right)^2 = \sum_{m=1}^L \gamma_m \quad (9.154)$$

where

$$\gamma_m = \frac{\alpha_m^2 E}{N_o}. \quad (9.155)$$

With Rayleigh fading, each of the γ_m is exponentially distributed with density function

$$p(\gamma_m) = \frac{1}{\bar{\gamma}_m} \exp\left\{-\frac{\gamma_m}{\bar{\gamma}_m}\right\}, \quad (9.156)$$

where $\bar{\gamma}_m$ is the average received bit energy-to-noise ratio for the k th channel tap. To compute the density of γ_b , first note that the characteristic function of γ_m is

$$\psi_{\gamma_m}(jv) = \frac{1}{1 - jv\bar{\gamma}_m} \quad (9.157)$$

so that the characteristic function of γ_b is

$$\psi_{\gamma_b}(jv) = \prod_{m=1}^L \frac{1}{1 - jv\bar{\gamma}_m}. \quad (9.158)$$

By using a partial fraction expansion and taking the inverse characteristic function, the density of γ_b is

$$p_{\gamma_b}(x) = \sum_{m=1}^L \frac{A_m}{\bar{\gamma}_m} \exp\left\{-\frac{x}{\bar{\gamma}_k}\right\}, \quad (9.159)$$

where

$$A_m = \prod_{\substack{i=1 \\ i \neq m}}^L \frac{\bar{\gamma}_m}{\bar{\gamma}_m - \bar{\gamma}_i}. \quad (9.160)$$

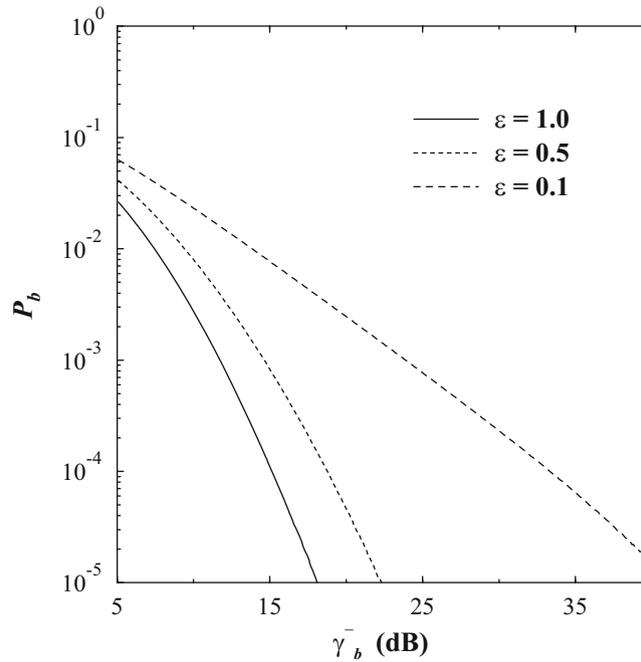


Fig. 9.28 Bit error probability with a RAKE receiver for DS/BPSK signaling on a multipath-fading channel. The channel has $L = 4$ taps and a 4-tap RAKE receiver is used

Therefore, with Rayleigh fading the average probability of bit error is

$$\begin{aligned}
 P_b &= \int_0^{\infty} Q(\sqrt{2x}) p_{\gamma_b}(x) dx \\
 &= \frac{1}{2} \sum_{m=1}^L A_m \left(1 - \sqrt{\frac{\bar{\gamma}_m}{1 + \bar{\gamma}_m}} \right).
 \end{aligned} \tag{9.161}$$

In order to proceed further the $\bar{\gamma}_m$ must be specified. One plausible model assumes an exponentially decaying power-delay profile, e.g.,

$$\bar{\gamma}_m = C e^{-k/\varepsilon}, \tag{9.162}$$

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

$$\bar{\gamma}_m = \frac{(1 - e^{-1/\varepsilon}) e^{-k/\varepsilon}}{e^{-1/\varepsilon} - e^{-(L+1)/\varepsilon}} \bar{\gamma}_b. \tag{9.163}$$

The probability of bit error is plotted in Fig. 9.28 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, 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 will suffer from some loss in performance.

9.6 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 short spreading codes of length N . The transmitted complex envelope for the i th user is

$$\tilde{s}^{(i)}(t) = A \sum_n x_n^{(i)} h^{(i)}(t - nT) \tag{9.164}$$

where

$$h^{(i)}(t) = \sum_{k=0}^{N-1} a_k^{(i)} h_c(t - kT_c) \tag{9.165}$$

and $\mathbf{a}^{(i)} = \{a_k^{(i)}\}$ and $\mathbf{x}^{(i)} = \{x_n^{(i)}\}$ are the i th user's spreading and data sequences, respectively. A real-valued spreading sequence assumed with chips $a_k^{(i)}$ chosen from the set $\{-1, +1\}$, corresponding to simple binary spreading or dual channel quaternary spreading. 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.

With CDMA, the signals from the various MSs in a cell will arrive at the corresponding BS with different power levels. DS CDMA systems must be power controlled such that all signals arrive at the intended receiver with the same power level; otherwise, a large performance loss will result. Power control is needed to combat the near-far effect, where strong signals will capture the receiver and prevent detection of the weaker signals. The power control loop must be fast enough to compensate not only for path loss and shadowing, but envelope fading as well. Under the assumption of perfect power control and a frequency non-selective channel, the received complex envelope will be⁶

$$\tilde{r}(t) = \sum_{i=1}^K e^{j\phi_i} \tilde{s}^{(i)}(t - \tau_i) + \tilde{n}(t) \tag{9.166}$$

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. 9.29.

This section considers the performance of an ideal correlation detector, where the composite received signal at a BS is multiplied by a synchronized and conjugated replica of the spreading sequence corresponding to a particular target MS whose data sequence is to be detected. Multiplication of a spreading sequence by a synchronized version of itself will despread the received signal corresponding to the particular target MS, while further spreading the received signals corresponding

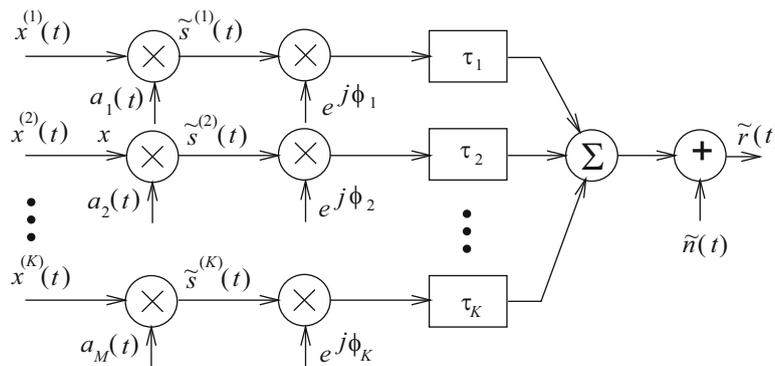


Fig. 9.29 DS CDMA signaling on a perfectly power controlled flat fading channel

⁶Here the normalization $\alpha = 1$ is assumed.

to other MSs in the composite received signal. As a result, the data sequence of the target MS can be obtained from the output of the correlator. Due to symmetry, only the decorrelator that recovers the data sequence from the first MS is considered. Furthermore, since only the relative delays and phases of the K signals comprising the composite received signal are important, it can be assumed that $\tau_1 = 0$ and $\phi_1 = 0$, while the remaining ϕ_i are uniformly distributed on $[0, 2\pi)$ and the remaining τ_i are uniformly distributed on $[0, T)$. Hence, it is assumed that the composite received signal at the BS is despread by multiplying by the spreading sequence $a^{(1)}(t)$.

The decision variable at the output of the correlator in Fig. 9.1 has been derived by Lehnert and Pursley [200] as

$$\mu_n = \sqrt{2E} \left[x_n^{(1)} + \sum_{k=2}^K B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) \cos \phi_k \right] + \tilde{n}_l \quad (9.167)$$

where \tilde{n}_l is a zero-mean complex Gaussian random variable with variance N_o

$$B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) = x_{n-1}^{(k)} R_{k,1}^p(\tau_k) + x_n^{(k)} \hat{R}_{k,1}^p(\tau_k) \quad (9.168)$$

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

$$R_{k,m}^p(\tau) = \frac{1}{T} \int_0^\tau a^{(k)}(t-\tau) a^{(m)}(t) dt \quad (9.169)$$

$$\hat{R}_{k,m}^p(\tau) = \frac{1}{T} \int_\tau^T a^{(k)}(t-\tau) a^{(m)}(t) dt. \quad (9.170)$$

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

$$R_{k,m}^p(\tau) = \phi_{k,m}^a(\ell - N) \hat{R}_h(\delta) + \phi_{k,m}^a(\ell + 1 - N) R_h(\delta) \quad (9.171)$$

$$\hat{R}_{k,m}^p(\tau) = \phi_{k,m}^a(\ell) \hat{R}_h(\delta) + \phi_{k,m}^a(\ell + 1) R_h(\delta) \quad (9.172)$$

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 $\{0, 1, \dots, N-1\}$. Combining (9.168), (9.171), and (9.172) gives

$$\begin{aligned} B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) = & \left[x_{n-1}^{(k)} \phi_{k,1}^a(\ell_k - N) + x_n^{(k)} \phi_{k,1}^a(\ell_k) \right] \hat{R}_h(\delta_k) \\ & + \left[x_{n-1}^{(k)} \phi_{k,1}^a(\ell_k + 1 - N) + x_n^{(k)} \phi_{k,1}^a(\ell_k + 1) \right] R_h(\delta_k) \end{aligned} \quad (9.173)$$

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. One possibility is to assume random spreading sequences, where the length- N spreading sequences $\mathbf{a}^{(i)} = \{a_k^{(i)}\}_{k=1}^N$ are generated randomly such that the $a_k^{(i)}$ are randomly chosen from the set $\{-1, +1\}$ with equal probability. For random spreading sequences and a rectangular chip shaping function $h_c(t) = u_{T_c}(t)$, Morrow and Lehnert [230] have shown that

$$B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) = x_{n-1}^{(k)} \frac{\delta_k}{N} + x_n^{(k)} \frac{(1-\delta_k)}{N} + \frac{X_k}{N} + \frac{Y_k}{N} (1-2\delta_k), \quad (9.174)$$

where δ_k is uniform on $[0, 1)$, $x_{n-1}^{(k)}$ and $x_n^{(k)}$ are uniform on $\{-1, +1\}$, and X_k and Y_k are discrete random variables having the probability distribution functions

$$p_{X_k}(i) = \frac{1}{2^{-A}} \binom{A}{\frac{i+A}{2}}, \quad i \in \{-A, -A+2, \dots, A-2, A\} \quad (9.175)$$

$$p_{Y_k}(i) = \frac{1}{2^{-B}} \binom{B}{\frac{j+B}{2}}, \quad i \in \{-B, -B+2, \dots, B-2, B\}. \quad (9.176)$$

The quantities A and B are related to

$$C \triangleq N\phi_{1,1}^a(1) = \sum_{j=0}^{N-2} a_j^{(1)} a_{j+1}^{(1)} \quad (9.177)$$

by

$$A = \frac{N-1+C}{2} \quad (9.178)$$

$$B = \frac{N-1-C}{2} \quad (9.179)$$

where $\phi_{1,1}^a(1)$ is the aperiodic cross-correlation of the spreading sequence of the first user. 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

$$p_C(i) = \frac{1}{2^{N-1}} \binom{N-1}{\frac{i+N-1}{2}}, \quad i \in \{-N+1, -N+3, \dots, N-3, N-1\}. \quad (9.180)$$

9.6.1 Standard Gaussian Approximation

The standard Gaussian approximation assumes that the multiple-access interference

$$I = \sum_{k=2}^K B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) \cos \phi_k \quad (9.181)$$

for sufficiently large K 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 $\{\delta_k, \phi_k, B \text{ or } C\}$ followed by ensemble averaging. Starting with the result in (9.174), it is straightforward to show that

$$\mathbb{E}[B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) | C] = 0 \quad (9.182)$$

since the conditional density functions for $x_n^{(k)}, x_{n-1}^{(k)}, X_k$, and Y_k are symmetrical about zero. Hence, $\mathbb{E}[W_k] = 0$ and $\mathbb{E}[I] = 0$.

To compute the variance of the multiple-access interference, it is convenient to define the vectors $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_K)$ and $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_K)$. The variance of the multiple-access interference is

$$\begin{aligned} \sigma_I^2 &= \mathbb{E}[I^2 | \boldsymbol{\delta}, \boldsymbol{\phi}, B] \\ &= \frac{1}{N^2} \sum_{k=2}^K \mathbb{E} \left[B_{k,1}^2(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) | \delta_k, B \right] \mathbb{E}[\cos^2 \phi_k | \phi_k] \\ &= \frac{1}{2} \sum_{k=2}^K (1 + \cos(2\phi_k)) \mathbb{E} \left[B_{k,1}^2(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) | \delta_k, B \right] \end{aligned} \quad (9.183)$$

Since all the Z_k are independent it follows that

$$\mathbb{E} \left[B_{k,1}^2(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) | \delta_k, B \right] = \frac{2(2B+1)(\delta_k^2 - \delta_k) + N}{N^2} \quad (9.184)$$

so that

$$\sigma_I^2 = \frac{1}{2} \sum_{k=2}^K (1 + \cos(2\phi_k)) \frac{2(2B + 1)(\delta_k^2 - \delta_k) + N}{N^2}. \quad (9.185)$$

If the intended sequence is known, then B is known. For random sequences $E[B] = (N - 1)/2$, and using this value for B

$$\sigma_I^2 = \frac{1}{N} \sum_{k=2}^K (1 + \cos(2\phi_k)) (\delta_k^2 - \delta_k + 1/2). \quad (9.186)$$

Several possibilities can be examined from here, including the following two important cases.

9.6.1.1 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 (9.167) 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

$$P_b = Q\left(\sqrt{\frac{3N}{K - 1}}\right). \quad (9.187)$$

The carrier to interference ratio C/I can be defined as the carrier power divided by the total noise power

$$\frac{C}{I} = \frac{1}{K - 1}. \quad (9.188)$$

By comparing (9.187) with the probability of bit error for binary signaling on an AWGN channel, i.e., $P_b = Q(\sqrt{2\gamma_b})$ observe that the C/I and the effective bit energy-to-noise ratio, γ_b , are related by

$$\gamma_b = \frac{2N}{3} \frac{C}{I}. \quad (9.189)$$

9.6.1.2 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

$$P_b = Q\left(\sqrt{\frac{N}{K - 1}}\right). \quad (9.190)$$

For chip and phase synchronous signals C/I and the effective γ_b are related by

$$\gamma_b = 2N \frac{C}{I}. \quad (9.191)$$

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.

9.6.2 Improved Gaussian Approximation

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.

A simple and more accurate Gaussian approximation has been derived by Holtzman [166]. To describe this method, let $P(\theta)$ be any function of a random variable θ having mean μ and variance σ^2 . Then using a Taylor series expansion about the mean μ gives

$$P(\theta) = P(\mu) + (\theta - \mu)P'(\mu) + \frac{1}{2}(\theta - \mu)^2P''(\mu) + \dots \quad (9.192)$$

Taking expectations

$$E[P(\theta)] \approx P(\mu) + \frac{1}{2}P''(\mu)\sigma^2. \quad (9.193)$$

Instead of using the Taylor series expansion, start instead with differences (Stirling's formula) and write

$$\begin{aligned} P(\theta) &= P(\mu) + (\theta - \mu) \frac{P(\mu + h) - P(\mu - h)}{2h} \\ &\quad + \frac{1}{2}(\theta - \mu)^2 \frac{P(\mu + h) - 2P(\mu) + P(\mu - h)}{h^2} + \dots \end{aligned} \quad (9.194)$$

Taking expectations

$$E[P(\theta)] \approx P(\mu) + \frac{1}{2} \frac{P(\mu + h) - 2P(\mu) + P(\mu - h)}{h^2} \sigma^2. \quad (9.195)$$

Holtzman [166] has shown that $h = \sqrt{3}\sigma$ yields good results so that

$$E[P(\theta)] \approx \frac{2}{3}P(\mu) + \frac{1}{6}P(\mu + \sqrt{3}\sigma) + \frac{1}{6}P(\mu - \sqrt{3}\sigma). \quad (9.196)$$

To apply the above result, let μ and σ^2 be the mean and variance of σ_l^2 in (9.185). It follows that

$$\mu = \frac{(K-1)}{N^2} E[L_k]. \quad (9.197)$$

where

$$L_k = \frac{1}{2}(1 + \cos(2\phi_k)) (2(2B+1)(\delta_k^2 - \delta_k) + N). \quad (9.198)$$

Then

$$\begin{aligned} \mu &= \frac{K-1}{N^2} \left(\frac{N}{2} - \frac{E[B]}{3} - \frac{1}{6} \right) \\ &= (K-1)/3N \end{aligned} \quad (9.199)$$

where the last line assumes ensemble averaging with random spreading sequences. The variance is

$$\begin{aligned} \sigma^2 &= \frac{K-1}{N^4} (E[L_k^2] - E^2[L_k] + (K-2)\text{cov}(L_j, L_k)) \quad \text{for any } j \neq k. \\ &= \frac{K-1}{N^4} \left(\frac{23}{360}N^2 + \left(\frac{1}{20} + \frac{K-2}{36} \right)N - \frac{1}{20} - \frac{K-2}{36} \right). \end{aligned} \quad (9.200)$$

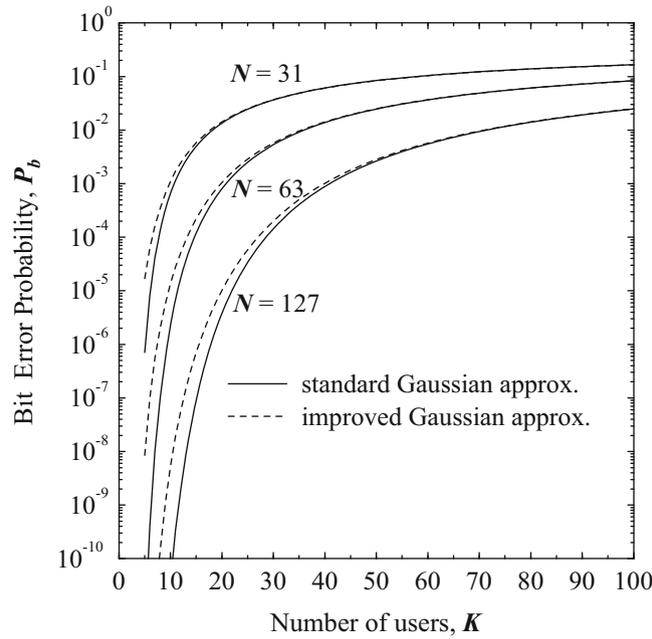


Fig. 9.30 Bit error probability against the number of users and various processing gains. The standard Gaussian approximation is shown to underestimate the error probability for small numbers of users

This yields

$$\begin{aligned}
 P_b \approx & \frac{2}{3}Q\left(\sqrt{\frac{3N}{K-1}}\right) + \frac{1}{6}Q\left(\frac{1}{\sqrt{(K-1)/3N + \sqrt{3}\sigma}}\right) \\
 & + \frac{1}{6}Q\left(\frac{1}{\sqrt{(K-1)/3N - \sqrt{3}\sigma}}\right). \tag{9.201}
 \end{aligned}$$

The above calculations are very simple and lead to quite accurate results for all values of K and N . Figure 9.30 compares the standard and improved Gaussian approximations for various processing gains and number of simultaneous users. Note that the standard Gaussian approximation underestimates 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.

9.7 CDMA Multiuser Detection

The previous sections have considered the performance of DS CDMA with conventional correlation detection where the multiple-access interference was treated as an unwanted impairment that can have a detrimental affect on system performance. This section considers the use of CDMA multi-user detection to combat multiple-access interference. The concept of DS CDMA multi-user detection is very similar to the concept of co-channel demodulation discussed in Sect. 7.7. However, instead of distinguishing users by their different received pulses due to their signals propagation through different frequency-selective fading channels, DS CDMA multi-user detection distinguishes the different users by assigning them different spreading sequences. In any case, the mathematical development is very similar and results from Sect. 7.7 will be used in our treatment of CDMA multi-user detection.

Once again, consider a CDMA system consisting of K users that use short spreading codes. The transmitted complex envelope for the i th user is given by

$$\tilde{s}^{(i)}(t) = A \sum_{n=1}^L x_n^{(i)} a^{(i)}(t - nT), \quad (9.202)$$

where

$$a^{(i)}(t) = \sum_{k=0}^{N-1} a_k^{(i)} h_c(t - kT_c), \quad (9.203)$$

and $\mathbf{a}^{(i)} = \{a_0^{(i)}, \dots, a_{N-1}^{(i)}\}$ and $\mathbf{x}^{(i)} = \{x_1^{(i)}, \dots, x_L^{(i)}\}$ are the i th user's length- N spreading sequence and length- L data sequence, respectively. The data symbols $x_n^{(i)}$ are independently chosen from the set $\{-1, +1\}$ with equal probability. The received complex envelope is given by

$$\tilde{r}(t) = A \sum_{i=1}^K g_i \tilde{s}^{(i)}(t - \tau_i) + \tilde{n}(t), \quad (9.204)$$

where the $\{g_i\}$, $g_i = \alpha_i e^{j\phi_i}$ and $\{\tau_i\}$ are the sets of random complex gains and delays that are introduced by the channel. The delays are assumed to satisfy the condition $0 \leq \tau_i \leq T$ for $1 \leq i \leq K$, where T is the data symbol duration. In the case of synchronous CDMA transmission, $\tau_i = 0$ for $1 \leq i \leq K$. Unlike the case of conventional correlation detection, the CDMA signals are not power controlled and may be received with different power levels due to the different channel gains α_i for $1 \leq i \leq K$. The fact that power control is not needed is a benefit of CDMA multi-user detection.

9.7.1 Optimum CDMA Multiuser Detection

The optimum CDMA multiuser detector determines the most likely sequence of transmitted bits $\mathbf{x}^{(i)} = \{x_1^{(i)}, \dots, x_L^{(i)}\}$ for all users $1 \leq i \leq K$, given the observation of the received waveform $\tilde{r}(t)$. There are generally two cases to consider: synchronous CDMA and asynchronous CDMA. However, for our purpose only synchronous CDMA is considered where the received complex envelope is given by

$$\tilde{r}(t) = A \sum_{i=1}^K g_i \tilde{s}^{(i)}(t) + \tilde{n}(t). \quad (9.205)$$

In this case, it is sufficient to consider the received pulse corresponding to the k th baud interval, given by

$$\tilde{r}(t) = A \sum_{i=1}^K g_i x^{(i)} a^{(i)}(t) + \tilde{n}(t), \quad (9.206)$$

and to detect the corresponding data vector $\mathbf{x} = \{x^{(1)}, \dots, x^{(K)}\}$. Similar to the development in Sect. 7.7, a complete set of complex orthonormal basis functions can be used to represent the received waveform $\tilde{r}(t)$ as a vector $\tilde{\mathbf{r}} = \{\tilde{r}_n\}$, such that

$$\tilde{r}_n = A \sum_{i=1}^K g_i x^{(i)} a_n^{(i)} + \tilde{n}_n, \quad (9.207)$$

where

$$a_n^{(i)} = \int_{-\infty}^{\infty} a^{(i)}(t) \varphi_n^*(t) dt \quad (9.208)$$

$$\tilde{n}_n = \int_{-\infty}^{\infty} \tilde{n}(t) \varphi_n^*(t) dt. \quad (9.209)$$

The \tilde{n}_n are uncorrelated zero-mean complex Gaussian random variables with variance $\frac{1}{2}E[|\tilde{n}_n|^2] = N_o$, so that the received vector $\tilde{\mathbf{r}}$ has the joint conditional multivariate Gaussian density

$$p(\tilde{\mathbf{r}}|\mathbf{x}) = \prod_{n=1}^N \frac{1}{2\pi N_o} \exp \left\{ -\frac{1}{2N_o} \left| \tilde{r}_n - A \sum_{i=1}^K g_i x^{(i)} a_n^{(i)} \right|^2 \right\}. \quad (9.210)$$

The optimum receiver chooses vector of symbols $\mathbf{x} = \{x^{(1)}, \dots, x^{(K)}\}$ to maximize $p(\tilde{\mathbf{r}}|\mathbf{x})$ or, equivalently, to minimize the metric

$$\mu(\mathbf{x}) = \sum_{n=1}^N \left| \tilde{r}_n - A \sum_{i=1}^K g_i x^{(i)} a_n^{(i)} \right|^2. \quad (9.211)$$

Since $\sum_{n=1}^N |\tilde{r}_n|^2$ is independent of \mathbf{x} , maximizing (9.211) is equivalent to maximizing

$$\mu(\mathbf{x}) = 2\text{Re} \left\{ A \sum_{i=1}^K g_i^* x^{(i)*} \sum_{n=1}^N \tilde{r}_n a_n^{(i)*} \right\} - A^2 \sum_{i=1}^K \sum_{i'=1}^K g_i x^{(i)} g_{i'}^* x^{(i')*} \sum_{n=1}^N a_n^{(i)} a_n^{(i')*}. \quad (9.212)$$

In the limit as the number of observable random variables N approaches infinity, define

$$y^{(i)} \triangleq \lim_{N \rightarrow \infty} \sum_{n=1}^N \tilde{r}_n a_n^{(i)*} = \int_{-\infty}^{\infty} \tilde{r}(t) a^{(i)*}(t) dt \quad (9.213)$$

$$R_{i,i'} \triangleq \lim_{N \rightarrow \infty} \sum_{n=1}^N a_n^{(i)} a_n^{(i')*} = \int_{-\infty}^{\infty} a^{(i)}(t) a^{(i')*}(t) dt. \quad (9.214)$$

Using (9.213) and (9.214) in (9.212) yields the final form

$$\mu(\mathbf{x}) = 2\text{Re} \left\{ A \sum_{i=1}^K g_i^* x^{(i)*} y^{(i)} \right\} - A^2 \sum_{i=1}^K \sum_{i'=1}^K g_i x^{(i)} g_{i'}^* x^{(i')*} R_{i,i'}. \quad (9.215)$$

The variables $\{y^{(i)}\}$, $1 \leq i \leq K$ in (9.213) are obtained by passing the received complex envelope $\tilde{r}(t)$ through a bank of K correlators or matched filters and sampling the outputs. This essentially despreads and detects the received signal. The parameter $R_{i,i'}$ in (9.214) is equal to the full period cross-correlation between the spreading waveforms $a^{(i)}(t)$ and $a^{(i')}(t)$.

The decision metric in (9.215) can be put into a convenient vector product form as follows:

$$\mu(\mathbf{x}) = 2\text{Re} \{ \mathbf{x}^H \mathbf{G}^* \mathbf{y} \} - \mathbf{x}^H \mathbf{G}^* \mathbf{R} \mathbf{G} \mathbf{x}, \quad (9.216)$$

where

$$\mathbf{y} = (y^{(1)}, y^{(2)}, \dots, y^{(K)})^T \quad (9.217)$$

$$\mathbf{x} = (x^{(1)}, x^{(2)}, \dots, x^{(K)})^T \quad (9.218)$$

$$\mathbf{G} = \text{diag}[Ag_1, Ag_2, \dots, Ag_K] \quad (9.219)$$

$$\mathbf{R} = [R_{i,i'}]_{K \times K}, \quad (9.220)$$

and \mathbf{x}^H is the complex conjugate transpose of \mathbf{x} . Note that the optimum detector must know or estimate the complex channel gains g_i , $1 \leq i \leq K$ to compute the decision metrics. The matrix \mathbf{R} may be assumed known, since the spreading sequences are known. It is apparent that the complexity of the optimal detector grows exponentially with the number of users K and becomes impractical for large K . For this reason, a number of sub-optimum CDMA multi-user detectors have been suggested in the literature.

9.7.2 Decorrelating Detector

The vector of samples at the output of the K correlators or matched filters is

$$\mathbf{y} = \mathbf{G}\mathbf{R}\mathbf{x} + \mathbf{v}, \quad (9.221)$$

where the zero-mean Gaussian noise vector

$$\mathbf{v} = (v^{(1)}, v^{(2)}, \dots, v^{(K)})^T \quad (9.222)$$

has covariance matrix

$$\Phi_{\mathbf{v}\mathbf{v}} = \frac{1}{2} \mathbf{E}[\mathbf{v}\mathbf{v}^H] = N_o \mathbf{R}. \quad (9.223)$$

It follows that the vector \mathbf{y} has a joint conditional complex Gaussian distribution with mean $\mathbf{G}\mathbf{R}\mathbf{x}$ and covariance matrix $N_o \mathbf{R}$. That is,

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi N_o)^K |\mathbf{R}|^{1/2}} \exp \left\{ -\frac{1}{2N_o} (\mathbf{y} - \mathbf{G}\mathbf{R}\mathbf{x})^H \mathbf{R}^{-1} (\mathbf{y} - \mathbf{G}\mathbf{R}\mathbf{x}) \right\}. \quad (9.224)$$

The best linear estimate of \mathbf{x} is the one that minimizes the metric

$$\mu(\mathbf{x}) = (\mathbf{y} - \mathbf{G}\mathbf{R}\mathbf{x})^H \mathbf{R}^{-1} (\mathbf{y} - \mathbf{G}\mathbf{R}\mathbf{x}) \quad (9.225)$$

which results in the solution

$$\tilde{\mathbf{x}} = \mathbf{R}^{-1} \mathbf{G}^{-1} \mathbf{y}. \quad (9.226)$$

To see how the decorrelator works, substitute (9.221) into (9.226) to obtain

$$\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{R}^{-1} \mathbf{G}^{-1} \mathbf{v} \quad (9.227)$$

Finally, the binary symbol decisions are obtained by

$$\hat{\mathbf{x}} = \text{sign}(\tilde{\mathbf{x}}). \quad (9.228)$$

Note that the decorrelator detector removes the multi-user interference in the signal terms, and this makes the detector near-far resistance so that power control is not necessary. Also, the decorrelator detector inverts the channel \mathbf{G} . This will result in noise enhancement for users that have small channel gains, g_i . Finally, if one or more pairs of spreading sequences are highly correlated, the matrix \mathbf{R} is ill-conditioned and will also cause noise enhancement.

9.7.3 Minimum Mean Square Error Detector

The MMSE detector is one that chooses the linear solution $\tilde{\mathbf{x}} = \mathbf{A}\mathbf{y}$ to minimize the mean square error (MSE)

$$J(\mathbf{x}) = \frac{1}{2} \mathbf{E}_{\mathbf{x}, \mathbf{v}}[\|\mathbf{x} - \tilde{\mathbf{x}}\|^2] = \frac{1}{2} \mathbf{E}[\|\mathbf{x} - \mathbf{A}\mathbf{y}\|^2], \quad (9.229)$$

where the expectation is over both the random data and noise vectors. The MMSE solution occurs when the error vector $\mathbf{x} - \mathbf{A}\mathbf{y}$ is orthogonal to the observation vector \mathbf{y} , i.e.,

$$\begin{aligned} \frac{1}{2} \mathbf{E}_{\mathbf{x}, \mathbf{v}}[(\mathbf{x} - \mathbf{A}\mathbf{y})\mathbf{y}^H] &= \mathbf{0} \\ \frac{1}{2} \mathbf{E}_{\mathbf{x}, \mathbf{v}}[\mathbf{x}\mathbf{y}^H] - \mathbf{A} \frac{1}{2} \mathbf{E}_{\mathbf{x}, \mathbf{v}}[\mathbf{y}\mathbf{y}^H] &= \mathbf{0}. \end{aligned} \quad (9.230)$$

For the case of synchronous CDMA,

$$\frac{1}{2}E_{\mathbf{x},\mathbf{v}}[\mathbf{xy}^H] = \frac{1}{2}E_{\mathbf{x}}[\mathbf{xx}^H\mathbf{R}^H\mathbf{G}^*] + \frac{1}{2}E_{\mathbf{v}}[\mathbf{xv}^H] = \frac{1}{2}\mathbf{R}^H\mathbf{G}^* = \frac{1}{2}\mathbf{R}\mathbf{G}^*, \quad (9.231)$$

since $E_{\mathbf{x}}[\mathbf{xx}^H] = \mathbf{I}$ with uncorrelated data, the noise vector \mathbf{v} has zero mean, and \mathbf{R} is a Hermitian matrix, i.e., $\mathbf{R}^H = \mathbf{R}$. Also,

$$\begin{aligned} \frac{1}{2}\mathbf{A}E_{\mathbf{x},\mathbf{v}}[\mathbf{yy}^H] &= \frac{1}{2}\mathbf{A}E_{\mathbf{x},\mathbf{v}}[(\mathbf{G}\mathbf{R}\mathbf{x} + \mathbf{v})(\mathbf{G}\mathbf{R}\mathbf{x} + \mathbf{v})^H] \\ &= \mathbf{A} \left(\frac{1}{2}E_{\mathbf{x}}[\mathbf{G}\mathbf{R}\mathbf{xx}^H\mathbf{R}^H\mathbf{G}^*] + \frac{1}{2}E_{\mathbf{v}}[\mathbf{vv}^H] \right) \\ &= \mathbf{A} \left(\frac{1}{2}\mathbf{G}\mathbf{R}\mathbf{R}^H\mathbf{G}^* + N_o\mathbf{R} \right) \\ &= \mathbf{A} \left(\frac{1}{2}\mathbf{G}\mathbf{R}\mathbf{R}\mathbf{G}^* + N_o\mathbf{R} \right). \end{aligned} \quad (9.232)$$

Substituting (9.231) and (9.232) into (9.230) and solving for \mathbf{A} gives the equation

$$\frac{1}{2}\mathbf{R}\mathbf{G}^* - \mathbf{A} \left(\frac{1}{2}\mathbf{G}\mathbf{R}\mathbf{R}\mathbf{G}^* + N_o\mathbf{R} \right) = \mathbf{0} \quad (9.233)$$

which leads to the solution

$$\mathbf{A}_{\text{opt}} = [\mathbf{G}\mathbf{R} + 2N_o(\mathbf{G}^*)^{-1}]^{-1}. \quad (9.234)$$

Using \mathbf{A}_{opt} , the MMSE receiver computes the vector

$$\tilde{\mathbf{x}} = \mathbf{A}_{\text{opt}}\mathbf{y} \quad (9.235)$$

and makes the decisions

$$\hat{\mathbf{x}} = \text{sign}(\tilde{\mathbf{x}}). \quad (9.236)$$

Like the decorrelator detector, the MMSE detector is near-far resistant. Also, from (9.234), if the noise is negligible compared to the multi-user interference, i.e., $N_o = 0$, then $\mathbf{A}_{\text{opt}} = \mathbf{R}^{-1}\mathbf{G}^{-1}$ and the MMSE detector reduces to the decorrelator detector. On the other hand, if the noise is dominant compared to the multi-user interference, i.e., N_o is large, then from (9.234), $\mathbf{A}_{\text{opt}} = \mathbf{G}^*/2N_o$. In this case, the MMSE detector reduces to the conventional correlation detector.

Problems

9.1. Suppose that a DS/BPSK spread spectrum signal is corrupted by a single, phase-asynchronous, CW 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. Consider a DS/BPSK waveform that is subjected to broadband pulse jamming. In particular, a jammer with total power J jams the DS/BPSK waveform a fraction ρ of the time with additive white Gaussian noise having a two-sided power spectral density

$$\frac{N_J}{2\rho} = \frac{J}{2W_{ss}\rho},$$

where W_{ss} is the spread spectrum bandwidth, and $N_J = J/W_{ss}$ is the one-sided spectral density achieved by spreading the total jammer power J over the spread spectrum bandwidth W_{ss} . During the remaining fraction of time $1 - \rho$, the jammer is off and the received DS/BPSK waveform is assumed to be noise-free. It is assumed that each modulated and spread BPSK symbol is either completely jammed or not at all. The average received bit energy-to-jammer-noise ratio is defined as

$$\gamma_b = \frac{E_b}{N_J} = \frac{W_{ss}S}{R_b J},$$

where S is the received desired signal power and R_b is the bit rate.

- Obtain an expression for the probability of bit error as a function of the jamming fraction ρ .
- Derive an expression for the jamming fraction ρ that will maximize the probability of bit error. Note that jamming fraction must satisfy the constraint $0 < \rho \leq 1$.
- What is the maximum bit error probability corresponding to the worst case jamming fraction found in part (b)?

9.3. Consider a non-coherently detected FH/BFSK waveform that is subjected to partial band noise jamming. In particular, a jammer with total power J jams a fraction ρ of the spread spectrum bandwidth W_{ss} with additive white Gaussian noise having a two-sided power spectral density

$$\frac{N_J}{2\rho} = \frac{J}{2W_{ss}\rho},$$

where $N_J = J/W_{ss}$ is the one-sided spectral density achieved by spreading the total jammer power J over the spread spectrum bandwidth W_{ss} . The remaining fraction $1 - \rho$ of bandwidth is free of jamming and is also assumed to be noise-free. It is assumed that each modulated and frequency hopped BFSK symbol is either completely jammed or not at all. The average received bit energy-to-jammer-noise ratio is defined as

$$\gamma_b = \frac{E_b}{N_J} = \frac{W_{ss}S}{R_b J},$$

where S is the received desired signal power and R_b is the bit rate.

- Obtain an expression for the probability of bit error as a function of the jamming fraction ρ .
- Derive an expression for the jamming fraction ρ that will maximize the probability of bit error. Note that jamming fraction must satisfy the constraint $0 < \rho \leq 1$.
- What is the maximum bit error probability corresponding to the worst case jamming fraction found in part (b)?

9.4. The generator polynomials for constructing “Gold-like” code sequences of length $N = 7$ are

$$\begin{aligned} p_1(x) &= 1 + x + x^3 \\ p_2(x) &= 1 + x^2 + x^3. \end{aligned}$$

The sequences are said to be “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.

9.5. Consider the length-7 m -sequences that are generated by the polynomials

$$\begin{aligned} p_1(x) &= 1 + x + x^3 \\ p_2(x) &= 1 + x^2 + x^3. \end{aligned}$$

- Find and sketch the full period autocorrelation functions of the sequences.
- Find and sketch the aperiodic autocorrelation function of the sequences.
- Find and sketch the full period cross-correlation function of the two sequences.
- Find the discrete-time Fourier transform of the aperiodic autocorrelation function, $\Phi^a(f)$, for each sequence.

9.6. Consider the length-7 Barker sequence

$$\mathbf{a} = (+1, +1, +1, -1, -1, +1, -1)$$

and its mirror image sequence

$$\mathbf{a} = (-1, +1, -1, -1, +1, +1, +1).$$

- Find and sketch the full period autocorrelation functions of the sequences.
- Find and sketch the aperiodic autocorrelation function of the sequences.
- Find and sketch the full period cross-correlation function of the two sequences.
- Sketch the discrete-time Fourier transform of the aperiodic autocorrelation function, $\Phi^a(f)$, and find the frequencies where it attains its maximum and minimum values.

9.7. Suppose that the following length-7 time-domain sequences are being proposed for the synchronization word in an OFDM system having block size $N = 7$.

$$\mathbf{x}_1 = (+1, +1, +1, -1, -1, +1, -1)$$

$$\mathbf{x}_2 = (+1, +1, +1, -1, +1, -1, -1)$$

- Compute and plot the periodic autocorrelation function and aperiodic autocorrelation function of these two sequences.
- Which sequence has better periodic autocorrelation characteristics for the purpose of synchronization.
- By taking the discrete-time Fourier transform of the aperiodic autocorrelation functions compute the power spectrum $\Phi^a(f)$ in each case. Plot your results.

9.8 (Computer Exercise). Write a computer program to generate a set of Gold sequences of length 127.

- Plot the mean and variance of the partial period autocorrelation as a function of the processing gain $10 \leq G \leq 20$ for this set of Gold codes.
- Repeat part (a) for the partial period cross-correlation.

9.9 (Computer Exercise). Consider a DS/BPSK CDMA system that uses length-31 Gold codes. The k th 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

$$h^{(k)}(t) = \sum_{j=1}^{31} a_j^{(k)} u_{T_c}(t - jT_c),$$

$\{x_n^{(k)}\}$ is uncorrelated zero-mean binary data sequence, T_c is the chip duration, and $T = 31T_c$ is the data bit duration.

- Pick one of the length-31 Gold codes at random for the k th user and write down the sequence. Plot the aperiodic autocorrelation function of sequence.
- Plot the power spectrum of the k th user $S_{ss}^{(k)}(f)$ against the normalized frequency fT_b .
- Repeat parts (a) and (b) for a randomly chosen length-63 Gold code for the k th user.

9.10. 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:

$$h_a(t) = \begin{cases} u_{T_c}(t) & \text{non-return-to-zero} \\ \sin(\pi t/T_c)u_{T_c}(t) & \text{half-sinusoid} \\ 1 - 2|t - T_c/2|/T_c u_{T_c}(t) & \text{triangular} \end{cases}$$

9.11. 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

$$\boldsymbol{\rho} = [\rho_{k,n}]_{k \times n},$$

where $\rho_{k,n} = \phi_{k,k}(n)$.

9.12. A WLAN 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

$$\begin{aligned} \tilde{s}_m(t) &= A \sum_{k=1}^8 h_{mk} h_c(t - kT_c), & m &= 1, \dots, 8 \\ &= -\tilde{s}_m(t), & m &= 9, \dots, 16, \end{aligned}$$

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.

- Assuming that $h_c(t) = u_{T_c}(t)$, find the power density spectrum of the complex envelope $S_{\tilde{s}_m}^-(f)$.
- Plot the power spectrum $S_{\tilde{s}_m}^-(f)$ against the normalized frequency fT_b , where T_b is the *bit* duration.

9.13. 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

$$\mathbf{a} = (-1, +1, +1).$$

- Compute and plot the full period autocorrelation of the sequence \mathbf{a} .
- Compute and plot the aperiodic autocorrelation of the sequence \mathbf{a} .
- 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 $S_{\tilde{s}_m}^-(f)$?

- At which frequencies is the receiver most sensitive to tone interference?

9.14 (Computer exercise). Consider the four length-15 spreading sequences identified in Example 9.1.

- Write a computer program to exhaustively search for and find the 4 sequences.
- The periodic autocorrelation and cross-correlation of these sequences are also important to ensure rapid acquisition and to minimize multiuser interference. Define the 4×4 matrix $\Phi = [\phi_{k,k}]$, where $\phi_{k,k}$ is defined as the *maximum* off-peak periodic autocorrelation of sequence $\mathbf{x}^{(k)}$ and $\phi_{j,k}$ is the *maximum* cross-correlation between $\mathbf{x}^{(j)}$ and $\mathbf{x}^{(k)}$. These quantities are defined by

$$\phi_{j,k} = \max_n \left| \frac{1}{N} \sum_{i=1}^N x_i^{(k)} x_{i+n}^{(j)} \right|$$

Find the 4×4 matrix Φ .

9.15 (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.

- Assuming that $h_c(t) = u_{T_c}(t)$, find the power density spectrum of the complex envelope $S_{ss}^-(f)$. You must average over all possible subsequences of length 10.
- Plot the power spectrum $S_{ss}^-(f)$ 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.

9.16. Suppose that the multipath intensity profile of a channel is given by

$$\phi_g(\tau) = \frac{P}{\mu_\tau} e^{-\tau/\mu_\tau}.$$

- What is the average delay and delay spread of the channel?
- 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.
- If the bit error probability for a non-dispersive 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} ?

9.17. 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.

9.18. Suppose that a 4-user synchronous CDMA system uses the following four length-15 spreading sequences:

$$\mathbf{x}_1 = \{-1 - 1 - 1 - 1 - 1 + 1 + 1 - 1 - 1 + 1 + 1 - 1 + 1 - 1 + 1\}$$

$$\mathbf{x}_2 = \{+1 - 1 + 1 - 1 + 1 + 1 - 1 - 1 + 1 + 1 - 1 - 1 - 1 - 1\}$$

$$\mathbf{x}_3 = \{-1 - 1 - 1 + 1 + 1 + 1 - 1 + 1 + 1 + 1 - 1 + 1 + 1 - 1 + 1\}$$

$$\mathbf{x}_4 = \{+1 - 1 + 1 + 1 - 1 + 1 + 1 + 1 - 1 + 1 + 1 + 1 - 1 - 1 - 1\}$$

Design a decorrelating detector, assuming an AWGN channel.

Chapter 10

Multi-Carrier Techniques

Multi-carrier techniques are widely used in wireless LANs and fourth generation cellular systems. Orthogonal frequency division multiplexing (OFDM) is widely used modulation technique in high-rate wireless systems such as IEEE 802.11a/g wireless local area network (WLAN) standard due to its robustness to frequency-selective fading. Multiple access with OFDM can be achieved by using TDMA or random access approaches. Orthogonal frequency division multiple access (OFDMA) is an extension of OFDM to accommodate multiple simultaneous users. OFDMA was originally proposed by Sari and Karam for cable television networks [289]. Since then OFDMA has been adopted for the forward channel in 3GPP Long Term Evolution (LTE) and in both the forward and reverse channels for IEEE 802.16e Worldwide Interoperability for Microwave Access (WiMAX) standard.

OFDMA achieves multiple access by dividing the available sub-carriers into mutually exclusive sets that are assigned to distinct users for simultaneous transmission. The orthogonality of the sub-carriers ensures protection against multiple-access interference. There are three basic forms of OFDMA depending on how the sub-carriers are allocated to the users. The first is clustered-carrier OFDMA (CC-OFDMA) where each user is allocated a contiguous group of sub-carriers. The second is spaced-carrier OFDMA (SC-OFDMA) where each user is assigned a group of sub-carriers that is regularly spaced across the channel bandwidth. The last is random interleaving OFDMA (RI-OFDMA) where sub-carriers are assigned in a random fashion to each user.

OFDMA has essentially the same advantages and disadvantages as OFDM when compared to single-carrier modulation schemes. It achieves robustness to frequency-selective fading by using closely spaced orthogonal sub-carriers, such that frequency domain equalization (FDE) can be employed. However, it also suffers from a high peak-to-average power ratio (PAPR) that requires the use of either PAPR reduction techniques or a highly linear power amplifier. OFDMA is attractive for use on the forward link of a cellular system, since all forward link transmissions can all use the same RF local oscillator and sample clock reference in their digital-to-analog converters (DACs). However, the use of OFDMA on the cellular reverse link is complicated considerably by the fact that waveform received at the BS from each MS will have a different carrier frequency offset, timing offset, and sampling clock offset

To overcome the difficulties of using OFDMA on the cellular reverse link, a modified form of OFDMA, called single-carrier frequency division multiple access (SC-FDMA) was introduced by Myung et al. [236]. SC-FDMA can be viewed as DFT-spread OFDMA, where a block of time-domain data symbols are first transformed to the frequency domain by using a discrete Fourier transform (DFT) before being applied to an OFDMA modulator. Similar to OFDMA, multiple access is achieved by assigning the users disjoint sets of sub-carriers. This can be done using localized FDMA (L-FDMA) where contiguous sub-carriers are assigned to a particular user, or Interleaved FDMA (I-FDMA) where users are assigned regularly spaced sub-carriers that are distributed over the entire bandwidth. The resulting SC-FDMA waveform is a single-carrier modulated waveform having a characteristically much lower PAPR than the corresponding multi-carrier OFDMA waveform. This lower PAPR benefits the MS in terms of transmit power efficiency, thereby making SC-FDMA very attractive for the cellular reverse link. For this reason SC-FDMA has been adopted as the reverse channel multiple-access scheme for 3GPP Long Term Evolution (LTE).

Section 10.1 begins the chapter with a discussion of OFDM on frequency-selective channels, and describes how a cyclic guard interval can be used to completely remove any intersymbol interference (ISI) in a very efficient fashion provided that the length of the cyclic guard interval is at least as long as the length of the overall discrete-time channel impulse response. The performance of OFDM on static ISI channels and fading ISI channels is then considered, in cases where the length of the guard interval is less than the length of the overall discrete-time channel impulse response. In this case, residual ISI is present which is shown to be devastating to the performance of OFDM. Section 10.1.2 then presents an effective technique to mitigate

residual ISI, called residual ISI cancellation (RISIC) that uses a combination of tail cancellation and cyclic reconstruction. Section 10.2 then considers the combination of single-carrier modulation with frequency domain equalization (FDE), a technique known as single-carrier frequency domain equalization (SC-FDE). FDE is especially attractive on channels having long impulse responses where the complexity of time-domain equalizers can become prohibitive. Afterwards, Sect. 10.3 treats a variety of issues related to OFDMA. The use of OFDMA on both the forward and reverse link is covered, and issues such as sub-carrier allocation and time-domain windowing are considered. Section 10.4 concludes the chapter with single-carrier frequency division multiple access (SC-FDMA), including multiplexing methods and analysis of peak-to-average power ratio.

10.1 Orthogonal Frequency Division Multiplexing

Consider an orthogonal frequency division multiplexing (OFDM) system where the number of sub-carriers N is chosen to be large enough so that the channel transfer function $T(t, f)$ is essentially constant across sub-bands of width $1/T$. Then

$$T(t, k) \doteq T(t, k\Delta_f), \quad k\Delta_f - 1/(2T) \leq f \leq k\Delta_f + 1/(2T), \quad (10.1)$$

and no equalization is necessary because the ISI is negligible. Viewing the problem another way, if the block length N is chosen so that $N \gg L$, where L is the length of the discrete-time channel impulse response, then the ISI will only affect a small fraction of the symbol transmitted on each sub-carrier. Weinstein and Ebert [348] came up with an ingenious solution, whereby they inserted a guard interval in the form of a length- G cyclic prefix or cyclic suffix to each IDFT output vector $\mathbf{X}_n = \{X_{n,m}\}$. In fact, if the discrete-time channel impulse response has duration $L \leq G$, their method can completely remove the ISI in a very efficient fashion as is now described.

Suppose that the IDFT output vector $\mathbf{X}_n = \{X_{n,m}\}_{m=0}^{N-1}$ is appended with a cyclic suffix to yield the vector $\mathbf{X}_n^g = \{X_{n,m}^g\}_{m=0}^{N+G-1}$, where

$$X_{n,m}^g = X_{n,(m)_N} \quad (10.2)$$

$$= A \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi km}{N}}, \quad m = 0, 1, \dots, N+G-1, \quad (10.3)$$

G is the length of the guard interval in samples, and $(m)_N$ is the residue of m modulo N . To maintain the data rate $R_s = 1/T_s$, the DAC in the transmitter is clocked with rate $R_s^g = \frac{N+G}{N}R_s$, due to the insertion of the cyclic guard interval.

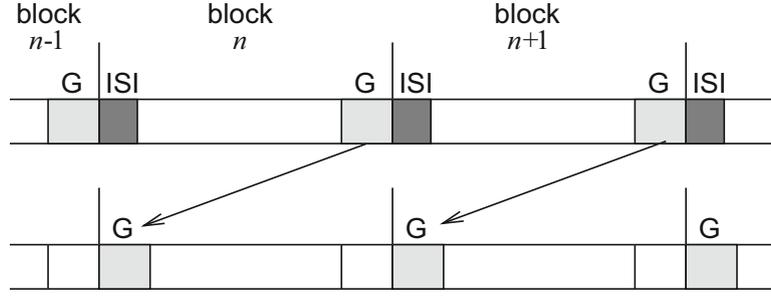
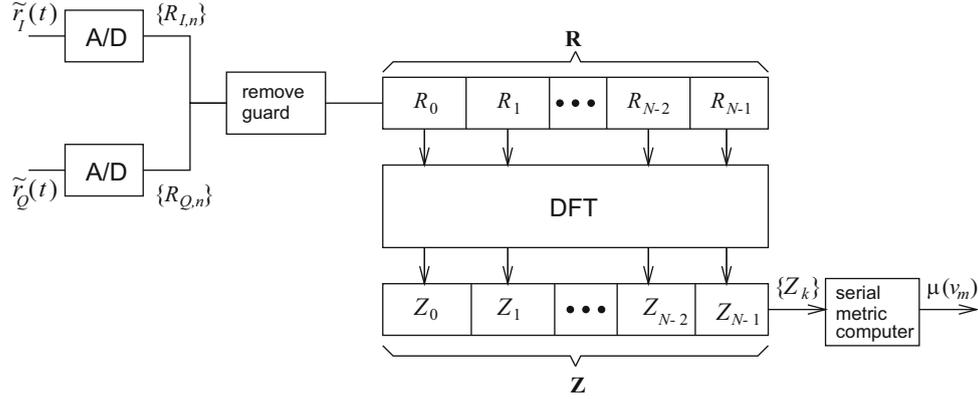
Consider a time-invariant ISI channel with impulse response $g(t)$. The combination of the DAC, waveform channel $g(t)$, anti-aliasing filter, and DAC yields an overall discrete-time channel with sampled impulse response $\mathbf{g} = \{g_m\}_{m=0}^L$, where L is the length of the discrete-time channel impulse response. The discrete-time linear convolution of the transmitted sequence $\{\mathbf{X}_n^g\}$ with the discrete-time channel produces the discrete-time received sequence $\{R_{n,m}^g\}$, where

$$R_{n,m}^g = \begin{cases} \sum_{i=0}^m g_i X_{n,m-i}^g + \sum_{i=m+1}^L g_i X_{n-1,N+G+m-i}^g + \tilde{n}_{n,m}, & 0 \leq m < L \\ \sum_{i=0}^L g_i X_{n,m-i}^g + \tilde{n}_{n,m}, & L \leq m \leq N+G-1 \end{cases}. \quad (10.4)$$

To remove the ISI introduced by the channel, the first G received samples $\{R_{n,m}^g\}_{m=0}^{G-1}$ are discarded and replaced with the last G received samples $\{R_{n,m}^g\}_{m=N}^{N+G-1}$, as shown in Fig. 10.1. If the length of the guard interval satisfies $G \geq L$, this gives the received sequence

$$\begin{aligned} R_{n,m} &= R_{n,G+(m-G)_N}^g \\ &= \sum_{i=0}^L g_i X_{n,(m-i)_N} + \tilde{n}_{n,(m-i)_N}, \quad 0 \leq m \leq N-1. \end{aligned} \quad (10.5)$$

Note that the first term in (10.5) represents a circular convolution of the transmitted sequence $\mathbf{X}_n = \{X_{n,m}\}$ with the discrete-time channel $\mathbf{g} = \{g_m\}_{m=0}^L$.


Fig. 10.1 Removal of ISI by using the cyclic suffix

Fig. 10.2 Block diagram of OFDM receiver

As shown in Fig. 10.2, the OFDM baseband demodulator computes the DFT of the vector \mathbf{R}_n . This yields the output vector

$$\begin{aligned} z_{n,i} &= \frac{1}{N} \sum_{m=0}^{N-1} R_{n,m} e^{-j \frac{2\pi m i}{N}} \\ &= T_i A x_{n,i} + v_{n,i}, \quad 0 \leq i \leq N-1, \end{aligned} \quad (10.6)$$

where

$$T_i = \sum_{m=0}^L g_m e^{-j \frac{2\pi m i}{N}} \quad (10.7)$$

and the noise samples $\{v_{n,i}\}$ are i.i.d with zero mean and variance $N_o/(NT_s^g)$. Note that $\mathbf{T} = \{T_i\}_{i=0}^{N-1}$ is the DFT of the zero padded sequence $\mathbf{g} = \{g_m\}_{m=0}^{N-1}$ and is equal to the sampled frequency response of the channel. To be consistent with our earlier results in Chap. 5, the $z_{n,i}$ can be multiplied by the scalar $\sqrt{NT_s^g}$. Following the same argument used in Sect. 5.7, such scaling gives

$$\tilde{z}_{n,i} = T_i \hat{A} x_{n,i} + \tilde{v}_{n,i} \quad i = 0, \dots, N-1, \quad (10.8)$$

where $\hat{A} = \sqrt{2E_h N / (N + G)}$ and the $\tilde{v}_{n,i}$ are i.i.d. zero-mean Gaussian random variables with variance N_o . Note that in this case, E_h is defined in (4.56). Observe that each $\tilde{z}_{n,i}$ depends only on the corresponding data symbol $x_{n,i}$ and, therefore, the ISI has been completely removed. Once again, for *each* of the $\tilde{z}_{n,i}$, the receiver decides in favor of the data symbol $x_{n,i}$ that minimizes the squared Euclidean distance

$$\mu(x_{n,i}) = \|\tilde{z}_{n,i} - T_i \hat{A} x_{n,i}\|^2. \quad (10.9)$$

Thus, for each OFDM block, N data symbol decisions must be made, one for each of the N sub-carriers.

10.1.1 Performance of OFDM on ISI Channels

For channels having long delay spreads or with deployment of single frequency simulcast networks (SFNs), the possibility exists that duration of the ISI will exceed the length of the guard interval. Such ISI is called *residual* ISI, and is devastating even in small amounts. Increasing the length of guard interval to reduce residual ISI has its limitations, because it introduces a bandwidth and power penalty. An attempt to equalize residual ISI has been suggested in [337], but the technique has high complexity and provides only a limited improvement. This section presents an effective technique to mitigate residual ISI, called *residual ISI cancellation* (RISIC). The RISIC technique is based on a very efficient method for echo cancellation [70]. The RISIC technique can be thought of as an iterative version of the echo cancellation method in [70]. The RISIC technique will be shown to be highly effective for combating residual ISI with reasonable complexity.

10.1.1.1 Static ISI Channels

Once again, in an attempt to remove the ISI introduced by the channel, the first G received samples $\{R_{n,m}^g\}_{m=0}^{G-1}$ in (10.4) are discarded and replaced with the last G received samples $\{R_{n,m}^g\}_{m=N}^{N+G-1}$. If the length of the channel exceeds the length of the guard interval such that $G < L$, then the received samples after removal of the guard interval can be rewritten as the sum of two components, viz.,

$$R_{n,m} = R_{n|n-1,m} + R_{n|n,m}, \quad (10.10)$$

where $R_{n|n-1,m}$ is the received sample component with contributions only from block $n - 1$ and $R_{n|n,m}$ is the received sample component with contributions only from block n . Then, in the absence of noise¹

$$R_{n|n-1,m} = \sum_{i=G+1}^L g_i X_{n-1,(m-i)_N} (1 - u(m - i + G)) \quad (10.11)$$

$$R_{n|n,m} = \sum_{i=0}^L g_i X_{n,(m-i)_N} u(m - i + G), \quad (10.12)$$

where $u(n)$ is the unit step function. The received sample sequence $\{R_{n,m}\}_{m=0}^{N-1}$ is demodulated by taking the N -point DFT

$$\text{DFT}\{R_{n,m}\} = \text{DFT}\{R_{n|n-1,m}\} + \text{DFT}\{R_{n|n,m}\}. \quad (10.13)$$

$\text{DFT}\{R_{n|n-1,m}\}$ can be expressed as

$$\text{DFT}\{R_{n|n-1,m}\} = \frac{1}{N} \sum_{i=G+1}^L g_i \sum_{\ell=0}^{N-1} x_{n-1,\ell} \exp\left\{-j\frac{2\pi\ell i}{N}\right\} \sum_{m=0}^{N-1} u(i - m - G - 1) \exp\left\{j\frac{2\pi(\ell - k)m}{N}\right\} \quad (10.14)$$

and $\text{DFT}\{R_{n|n,m}\}$ by

$$\begin{aligned} \text{DFT}\{R_{n|n,m}\} &= x_{n,k} \left\{ \sum_{i=0}^G g_i \exp\left\{-j\frac{2\pi ki}{N}\right\} + \sum_{i=G+1}^L g_i \exp\left\{-j\frac{2\pi ki}{N}\right\} \left(1 + \frac{G}{N} - \frac{i}{N}\right) \right\} \\ &\quad - \frac{1}{N} \sum_{i=G+1}^L g_i \sum_{\substack{\ell=0 \\ \ell \neq k}}^{N-1} x_{n,\ell} \exp\left\{-j\frac{2\pi\ell i}{N}\right\} \sum_{m=0}^{N-1} u(i - m - G - 1) \exp\left\{j\frac{2\pi(\ell - k)m}{N}\right\}. \end{aligned} \quad (10.15)$$

¹Here we ignore noise to highlight the effect of ISI.

For symbol k of block n , (10.14) is the ISI contribution from block $n-1$, the top half of (10.15) is the useful signal term, and the bottom half of (10.15) is the ICI term. Equation (10.13) can be expressed as

$$z_{n,k} = \text{DFT}\{R_{n,m}\} \quad (10.16)$$

$$= \eta_k x_{n,k} + I_k \quad (10.17)$$

where

$$\eta_k = \sum_{i=0}^G g_i \exp\left\{-j\frac{2\pi ki}{N}\right\} + \sum_{i=G+1}^L g_i \exp\left\{-j\frac{2\pi ki}{N}\right\} \left(1 + \frac{G}{N} - \frac{i}{N}\right) \quad (10.18)$$

$$I_k = I_{n,k} + I_{n-1,k} \quad (10.19)$$

and where $I_{n,k}$ the ICI term and $I_{n-1,k}$ the ISI term.

Next, the signal-to-interference ratio (SIR) for symbol k defined by

$$\text{SIR}(k) = E_u(k)/E_I(k) \quad (10.20)$$

is found where the useful signal energy is

$$E_u(k) = \frac{1}{2} \text{E}|\eta_k x_{n,k}|^2. \quad (10.21)$$

Since the input symbols $\{x_{n,k}\}$ are assumed independent, $x_{n,k}$ and I_k are also independent. Furthermore, $I_{n,k}$ and $I_{n-1,k}$ are independent, too. Then, interference energy is

$$\begin{aligned} E_I(k) &= \frac{1}{2} \text{E}|I_k|^2 \\ &= \frac{1}{2} \text{E}|I_{n-1,k}|^2 + \frac{1}{2} \text{E}|I_{n,k}|^2 \\ &= E_{\text{ISI}}(k) + E_{\text{ICI}}(k). \end{aligned} \quad (10.22)$$

The signal energy in (10.21) can be expressed as

$$\begin{aligned} E_u(k) &= E_s \left| \sum_{i=0}^G g_i \exp\left\{-j\frac{2\pi ki}{N}\right\} + \sum_{i=G+1}^L g_i \exp\left\{-j\frac{2\pi ki}{N}\right\} \left(1 + \frac{G}{N} - \frac{i}{N}\right) \right|^2 \\ &\approx E_s \left| \sum_{i=0}^L g_i \exp\left\{-j\frac{2\pi ki}{N}\right\} \right|^2, \quad \text{if } L - G \ll N. \end{aligned} \quad (10.23)$$

From (10.22),

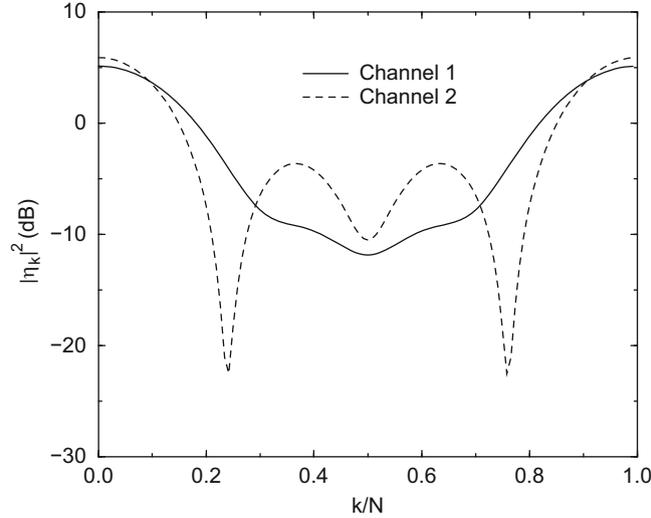
$$E_{\text{ISI}}(k) = \frac{E_s}{N} \left\{ \sum_{i=G+1}^L \sum_{i'=i}^L 2(i-G) \text{Re} \left\{ g_i g_{i'}^* \exp\left\{-j\frac{2\pi k(i-i')}{N}\right\} \right\} - \sum_{i=G+1}^L |g_i|^2 (i-G) \right\} \quad (10.24)$$

$$E_{\text{ICI}}(k) = E_{\text{ISI}}(k) - \frac{E_s}{N^2} \left| \sum_{i=G+1}^L g_i \exp\left\{-j\frac{2\pi ki}{N}(i-G)\right\} \right|^2. \quad (10.25)$$

Note that the second term in (10.25) is relatively small when $L - G \ll N$, in which case $E_I(k) \approx 2E_{\text{ISI}}(k)$.

Table 10.1 4-Tap static ISI channel models

Tap #	Delay (μs)	Fractional power Channel 1	Fractional power Channel 2
0	0.0	0.15	0.39
1	0.2	0.65	0.16
2	0.4	0.15	0.26
3	0.6	0.05	0.19

**Fig. 10.3** Amplitude spectrum of static ISI channel with OFDM for $N = 128$

QAM Performance

The symbol error rate (SER) for 16-QAM is

$$\text{SER} = 3Q\left(\sqrt{\frac{1}{5}}\gamma_s\right)\left(1 - \frac{3}{4}Q\left(\sqrt{\frac{1}{5}}\gamma_s\right)\right), \quad (10.26)$$

where γ_s is the received symbol energy-to-noise ratio. With OFDM, the SER on a static ISI channel is

$$\text{SER} = \frac{1}{N} \sum_{k=0}^{N-1} \text{SER}(k) \quad (10.27)$$

where $\text{SER}(k)$ is obtained from (10.26) with γ_s replaced by $\text{SIR}(k)$ from (10.20).

Two static ISI channels are considered as shown in Table 10.1. Channel 1 has sub-channels with a moderate null, whereas Channel 2 has several sub-channels with a severe null, as shown in Fig. 10.3.

Figure 10.4 shows the SER floors due to ISI without noise for different block sizes when a guard interval is not used ($G = 0$). As expected the error floor due to the ISI decreases with increasing block size. However, when there exists deep null sub-channels as in Channel 2, the improvement from increasing block size is quite small. Therefore, increasing the OFDM block size is not always a very efficient countermeasure against ISI.

10.1.1.2 Fading ISI Channel

On fading ISI channels, channel time variations during a block cause ICI. For large block sizes, the central limit theorem can be invoked and the ICI can be treated like AWGN as shown in Sect. 5.7.1. If 2-D isotropic scattering and Rayleigh fading is assumed, then for large N the SIR is

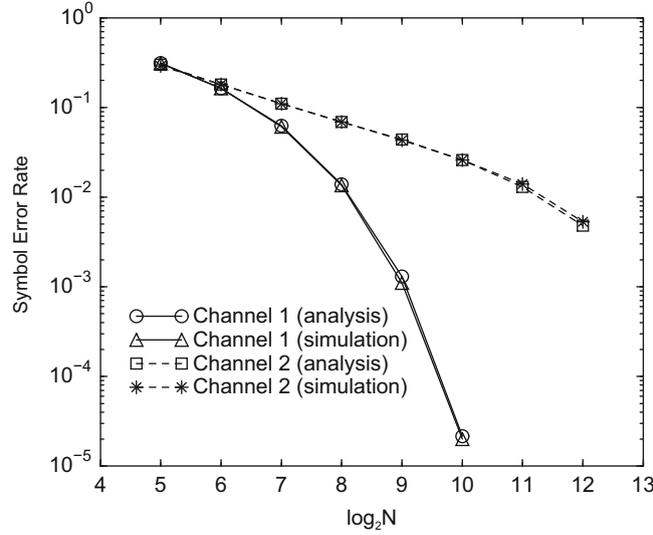


Fig. 10.4 Performance of OFDM signaling on static ISI channels with different block sizes; $G = 0$

$$\text{SIR} = \frac{E_u}{E_s - E_u}, \quad (10.28)$$

where E_u is

$$E_u = \frac{E_s}{N^2} \left\{ \sum_{i=0}^G E|g_i|^2 \left(N + 2 \sum_{m=1}^{N-1} (N-m) J_0(2\pi f_m T_s m) \right) + \sum_{i=G+1}^L E|g_i|^2 \left(N - i + G + 2 \sum_{m=1}^{N-i-1+G} (N-i-m+G) J_0(2\pi f_m T_s m) \right) \right\}. \quad (10.29)$$

The SER for a Rayleigh fading channel is obtained by averaging (10.26) over the probability density function

$$p_{\bar{\gamma}_s}(x) = \frac{1}{\bar{\gamma}_s} e^{-x/\bar{\gamma}_s}, \quad x \geq 0, \quad (10.30)$$

where $\bar{\gamma}_s$ is replaced by the SIR. The SER obtained by using the SIR in (10.28) is actually an upper bound when the interference caused by ISI is a dominant factor. This is due to the correlation between the useful signal and interference term. An intuitive explanation is as follows. If the channel varies slowly, the channel impulse response can be assumed constant over a duration of an OFDM symbol. Hence, from (10.20), the conditional SIR for the sub-channel k given the channel impulse response $\mathbf{g} = \{g_m\}_{m=0}^L$, $\text{SIR}(k|\mathbf{g})$, is

$$\text{SIR}(k|\mathbf{g}) = \frac{E[|\eta_k x_{n,k}|^2 | \mathbf{g}]}{E[|I_k|^2 | \mathbf{g}]}. \quad (10.31)$$

For $L - G \ll N$ $\text{SIR}(k|\mathbf{g})$ is well approximated by

$$\begin{aligned} \text{SIR}(k|\mathbf{g}) = \frac{E_s}{2} & \left\{ \frac{\sum_{i=0}^G \sum_{i'=0}^G g_i g_{i'}^* \exp\{-j\frac{2\pi k}{N}(i-i')\}}{E_{\text{ISI}}(k|\mathbf{g})} \right. \\ & + \frac{\sum_{i=0}^G \sum_{i'=G+1}^L g_i g_{i'}^* \exp\{-j\frac{2\pi k}{N}(i-i')\}}{E_{\text{ISI}}(k|\mathbf{g})} \\ & \left. + \frac{\sum_{i=G+1}^L \sum_{i'=G+1}^L g_i g_{i'}^* \exp\{-j\frac{2\pi k}{N}(i-i')\}}{E_{\text{ISI}}(k|\mathbf{g})} \right\} \quad (10.32) \end{aligned}$$

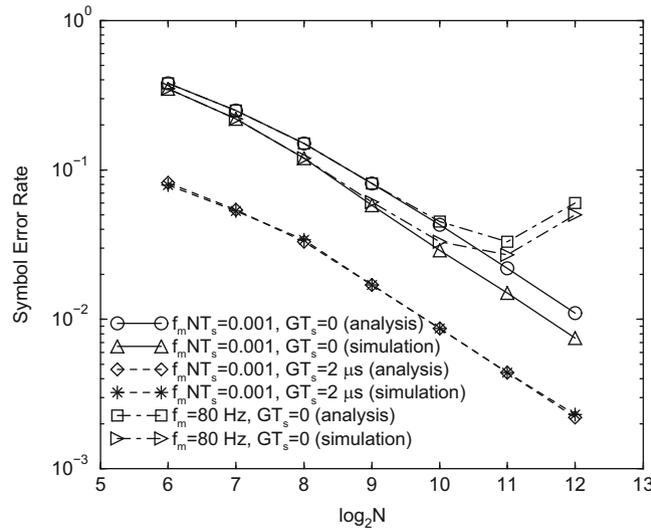


Fig. 10.5 Performance of OFDM signaling on fading ISI channels with different block sizes

where $E_{\text{ISI}}(k|\mathbf{g})$ is from (10.24) but expressed differently as

$$E_{\text{ISI}}(k|\mathbf{g}) = \frac{E_s}{N} \sum_{i=G+1}^L \sum_{i'=G+1}^L g_i g_{i'}^* \min[i - G, i' - G] \exp \left\{ -j \frac{2\pi k(i - i')}{N} \right\} \quad (10.33)$$

In (10.32), the first and second fractions represent the portion of SIR for which the Rayleigh assumption is valid in computing the SER, because the useful signal term and the interference term are uncorrelated. However, the last fraction shows that the useful signal term and the interference term are correlated. Hence, when the useful signal term is faded so is the interference. Consequently, the Rayleigh assumption leads to pessimistic performance estimates, i.e., the use of (10.28) to compute the SER gives an upper bound. On the other hand, if the portion of the energy contained within the guard interval is relatively large, then the last fraction becomes insignificant relative to the first two fractions in (10.32) and, hence, the SER found by using the SIR from (10.28) is accurate.

Figure 10.5 shows the performance on a fading ISI channel that is based on the COST207 6-tap reduced typical urban channel model (see Sect. 2.5.4.1). Similar to the case of static ISI channels, the performance improves as the block size is increased if the normalized Doppler frequency, $f_m NT_s$, is small. Unlike the static ISI channels, however, the block size cannot be made too large since this increases the ICI caused by channel time variations over block (Fig. 10.5). Therefore, the block size must be small enough to keep the ICI small, while it must be large enough to keep the ISI small for channels with a long impulse response. Also, as explained above, the SER from the analysis when $GT_s = 2 \mu\text{s}$ agrees very well with simulation result but gives an upper bound when $G = 0$.

10.1.2 Residual ISI Cancellation

If the channel changes little over an OFDM symbol period and the guard interval is sufficiently large, i.e., $G \geq L$, then the channel output is

$$\tilde{R}_{n,m} = \sum_{i=0}^L g_i X_{n,(m-i)_N}, \quad 0 \leq m \leq N - 1 \quad (10.34)$$

where $\tilde{R}_{n,m}$ represents the *desired* channel output that is free of ISI. However, when $G < L$ there will be residual ISI. To achieve the *desired* channel output $\tilde{R}_{n,m}$ in the presence of residual ISI, two steps must be taken. The first is to remove the residual ISI from the received signal, and the second is to use reconstruction to restore cyclicity and avoid ICI. These two procedures are called *tail cancellation* and *cyclic reconstruction*, respectively [70]. The procedure can be described by

$$\tilde{R}_{n,m} = R_{n,m} - R_{n|n-1,m} + \sum_{i=G+1}^L g_i X_{n,(m-i)_N} (1 - u(m - i + G)). \quad (10.35)$$

The residual ISI is removed from the received signal by subtracting the second term in (10.35). Cyclicity is restored by the last term in (10.35).

The feasibility of implementing the tail cancellation and cyclic reconstruction procedures depends on the availability of the transmitted sample sequence $\{\mathbf{X}_n\}$ at the receiver. Echo cancellers have exact knowledge of the transmitted symbols and, therefore, the above procedures have been successfully implemented [70]. However, the vast majority of communication applications that require mitigation of ISI do not enjoy this luxury. A method is now described for reducing the effect of ISI using the aforementioned procedures when the transmitted symbols are not available to the receiver *a priori*.

10.1.2.1 Residual ISI Cancellation Algorithm

Assume that the channel impulse response is constant over an OFDM symbol period, i.e., $g_{n,k} = g_n$, $0 \leq k \leq N + G - 1$. The RISIC algorithm proceeds as follows:

1. An estimate of the channel impulse response, \hat{g}_n , is obtained from a training sequence and updated in a decision-directed mode. Channel estimation will be treated in more detail in Sect. 10.1.3.3.
2. Decisions on the transmitted data symbols $\{\hat{x}_{n-1,k}\}$ from block $n - 1$ are obtained for use in tail cancellation. Since the decisions are affected by residual ISI, some may be erroneous. These symbols are converted back to time domain using an IDFT giving $\{\hat{X}_{n-1,m}\}$.
3. For the block of index n , *tail cancellation* is performed by calculating the residual ISI and subtracting it from $R_{n,m}$, i.e.,

$$\tilde{R}_{n,m}^{(0)} = R_{n,m} - \sum_{i=G+1}^{\hat{L}} \hat{g}_i \hat{X}_{n-1,(m-i)_N} (1 - u(m - i + G)) \quad 0 \leq m \leq N - 1 \quad (10.36)$$

where \hat{L} is the estimate of the maximum channel impulse response length.

4. The $\{\tilde{R}_{n,m}^{(0)}\}$ obtained in Step 3 are converted to the frequency domain and decisions are made. Afterwards, the decisions are converted back to time domain to give $\{\hat{X}_{n,m}^{(0)}\}$.
5. Next *cyclic reconstruction* is performed by forming

$$\tilde{R}_{n,m}^{(I)} = \tilde{R}_{n,m}^{(0)} + \sum_{i=G+1}^{\hat{L}} \hat{g}_i \hat{X}_{n,(m-i)_N}^{(I-1)} (1 - u(m - i + G)), \quad 0 \leq m \leq N - 1 \quad (10.37)$$

where I represents an iteration number with an initial value of $I = 1$.

6. The $\{\tilde{R}_{n,m}^{(I)}\}$ are converted to the frequency domain and decisions are made yielding $\{\hat{x}_{n,k}^{(I)}\}$. This completes the I th iteration in the RISIC algorithm.
7. To continue iterations, convert the $\{\hat{x}_{n,k}^{(I)}\}$ to $\{\hat{X}_{n,m}^{(I)}\}$ and repeat Steps 5–7 with $I \leftarrow I + 1$.
8. End of the RISIC algorithm for block n .

10.1.3 Performance of the RISIC Algorithm

The performance of the RISIC algorithm is now evaluated on both static and fading ISI channels. The stability of the RISIC algorithm and the effect of using imperfect estimates of the channel impulse response is also investigated. Here, it is assumed that $\hat{L} = L$, and choosing $\hat{L} > L$, e.g., $\hat{L} = 2L$, will cause only a small degradation in performance.

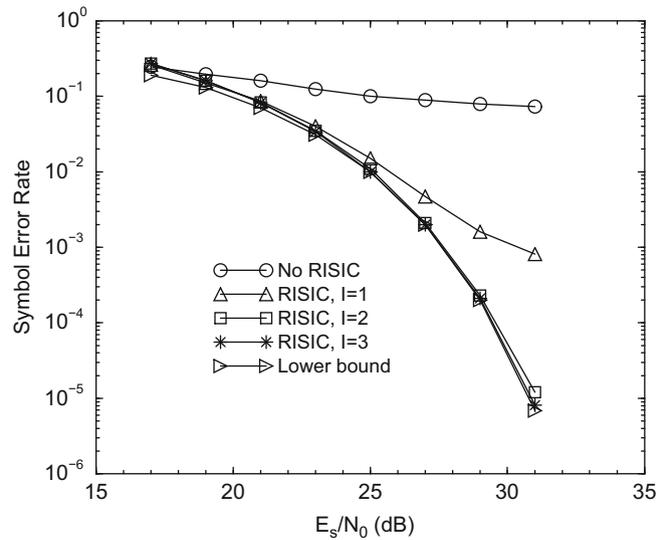


Fig. 10.6 Performance of the RISIC technique on Channel 1; $G = 0$, $N = 128$

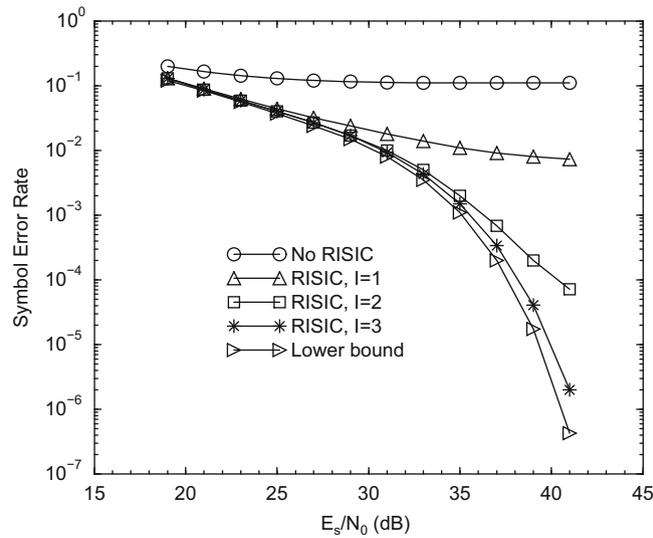


Fig. 10.7 Performance of the RISIC technique on Channel 2; $G = 0$, $N = 128$

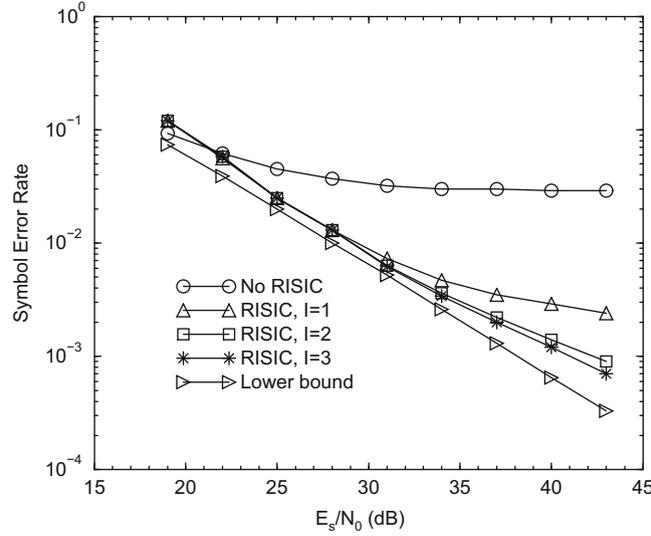
10.1.3.1 Static ISI Channel

Suppose that the receiver has perfect channel information. The two static ISI channels in Table 10.1 are considered. Figures 10.6 and 10.7 illustrate the performance of the RISIC for Channels 1 and 2, respectively. RISIC offers a huge improvement in SER even after the first iteration, especially at high SNR. Several orders of magnitude improvement in SER are achievable for both channels after 2 or 3 iterations. The lower bounds shown in these figures are obtained by computing the SER using (10.27) without ISI and with noise only. On Channel 1, two iterations are required to effectively achieve the lower bound, while three iterations are required on Channel 2.

A periodic pilot sequence is needed even if perfect channel information is available, to prevent instability when feeding back a highly erroneous signal in the tail cancellation procedure. Table 10.2 shows the performance of RISIC on Channel 1 in Table 10.1 with $N = 64$ for $E_s/N_0 = 35$ dB and various training sequence overheads. Even with small overhead (2%), the SER improves dramatically, whereas the algorithm diverges when a pilot sequence is not used. Furthermore, only small degradation is observed with 2% overhead as compared to 5% overhead. If the block size is increased to $N = 128$, the RISIC algorithm converges for Channel 1 even when a pilot sequence is not used. Since many communications systems do utilize periodic training sequences for the purpose of synchronization or channel estimation, the need for additional overhead to maintain stability when applying RISIC may be unnecessary.

Table 10.2 Effect of using pilot sequence with RISIC technique

Overhead	Without RISIC	SER ($I = 1$)	SER ($I = 2$)	SER ($I = 3$)
0 %	0.17	0.84	0.84	0.85
2 %	0.17	7.5×10^{-3}	2.9×10^{-4}	5.8×10^{-5}
5 %	0.17	7.5×10^{-3}	2.6×10^{-4}	3.1×10^{-5}

**Fig. 10.8** Performance of the RISIC technique on a fading ISI channel; $G = 0$, $N = 1024$, $f_m N T_s = 0.001$

10.1.3.2 Fading ISI Channel

For fading ISI channels, the channel impulse response varies with time. However, if the channel changes little over an OFDM symbol duration and $L - G \ll N$, then

$$\{g_m\} \approx \frac{1}{\sqrt{N}} \text{IDFT}\{\eta_k\} \quad (10.38)$$

When the channel variations are rapid, Fig. 10.5 shows that the performance is dominated by ICI rather than residual ISI. Assume η_k is known at the receiver and the channel impulse response is found by (10.38). Figures 10.8 and 10.9 show the performance of RISIC on a typical urban channel with $f_m N T_s = 0.001$ and $f_m N T_s = 0.005$, respectively, and 5% training overhead.

10.1.3.3 Channel Estimation

All the remarkable improvements that RISIC has shown so far are achieved under the premise of perfect channel information. For a strictly static ISI channel, the channel estimation process needs to be carried out only once. In fact, almost perfect channel estimates can be obtained using a training sequence [70]. Time-varying channels present the real challenge for channel estimation.

For fading ISI channels, a channel estimation technique is proposed that can provide accurate estimates even in the presence of residual ISI. An OFDM training block can be used consisting of $N/2$ symbols chosen from a chirp sequence [70]

$$c_k = \exp \left\{ j \frac{2\pi}{N} k^2 \right\}, \quad 0 \leq k \leq \frac{N}{2} - 1. \quad (10.39)$$

The chirp sequence provides an optimal peak-to-average power ratio of 1 and weighs each sub-channel equally in estimating the channel since the DFT of a chirp sequence is also chirp sequence. Unlike [70], however, the training sequence is modified to increase resilience to residual ISI, by inserting zeros in every odd sub-channel. The training block symbols are

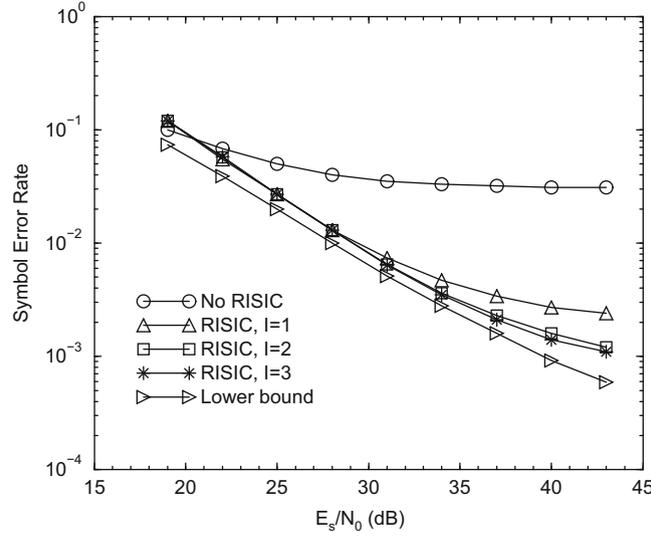


Fig. 10.9 Performance of the RISIC technique on a fading ISI channel; $G = 0$, $N = 1024$, $f_m N T_s = 0.005$

$$d_k = \begin{cases} \sqrt{2}c_{k/2} & , k = 0, 2, \dots, N-2 \\ 0 & , k = 1, 3, \dots, N-1, \end{cases} \quad (10.40)$$

where $\sqrt{2}$ is a normalization factor. The N -point IDFT of $\{d_k\}$ is²

$$D_m = C_{(m)N/2}, \quad 0 \leq m \leq N-1. \quad (10.41)$$

From (10.41), the first half of the time-domain training sequence $\{D_m\}$ is identical to the second half. This is a valuable property for long channel impulse responses, because the first half of $\{D_m\}$ can be used just like a guard interval while $\{D_m\}$ still possesses a peak-to-average power ratio of 1. The channel estimation procedure using the proposed training block is as follows:

1. After removal of the guard interval, the received samples are rearranged as

$$\bar{R}_{ts,m} = R_{ts,m+N/2}, \quad 0 \leq m \leq N/2-1, \quad (10.42)$$

where the subscript ts indicates that the received samples are for a training block.

2. $N/2$ channel estimates, $\tilde{\eta}_k$, are calculated by

$$\tilde{\eta}_k = \frac{z_{ts,k}}{c_k}, \quad 0 \leq k \leq N/2-1, \quad (10.43)$$

where $\{z_{ts,k}\}$ is the $N/2$ -point DFT of $\{\bar{R}_{ts,m}\}$.

3. The estimates $\tilde{\eta}_k$ are converted to \tilde{g}_m by

$$\{\tilde{g}_m\} = \frac{1}{\sqrt{N/2}} \text{IDFT}\{\tilde{\eta}_k\}, \quad (10.44)$$

where an $N/2$ -point IDFT is used. Then, $\{\tilde{g}_m\}$ is passed through a rectangular window to zero \tilde{g}_m , $m > \hat{L}$, and the result is $\{\hat{g}_m\}$ which is used for RISIC.

²The C_n can also be obtained by the $N/2$ -point IDFT of $\{c_k\}$.

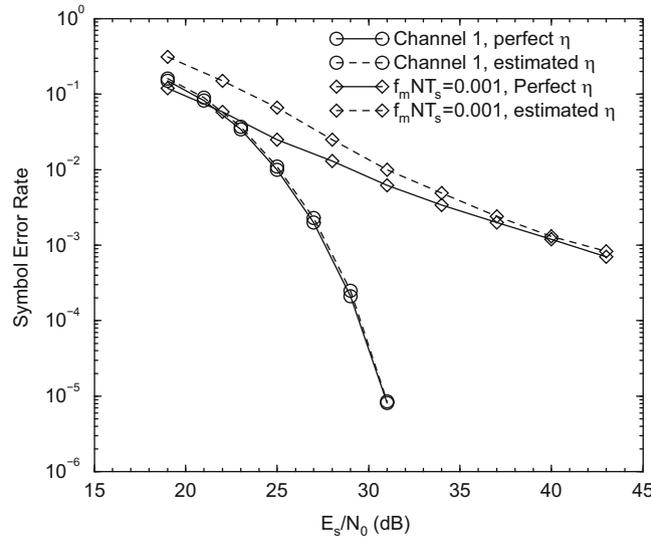


Fig. 10.10 Effect of imperfect channel estimation with the RISIC technique for static and slowly fading ISI channels with $N = 128$ and $N = 1024$, respectively; $G = 0$, $I = 3$

4. $\{\hat{g}_m\}$ is converted to $\{\hat{\eta}_k\}$ by

$$\hat{\eta}_k = \sqrt{N} \text{DFT}\{\hat{g}_m\} \quad (10.45)$$

where this time an N -point DFT is used. The channel estimates for all N sub-channels are now available for data demodulation.

When L is not known, the window size must be large enough to include the entire channel impulse response yet as small as possible to minimize the effect of noise.

On static ISI channels, the channel estimation is performed only once at the beginning of each simulation run and the channel estimates are averaged over four blocks. On fading ISI channels, the channel estimation is performed by periodically sending one training block out of every 20 blocks transmitted, while updating the channel estimate in the remaining blocks in a decision-directed mode [224]. In the decision-directed mode, the decisions made after the final iteration of the RISIC algorithm are used to update the channel estimates. Decision errors do not degrade the accuracy of the estimates excessively, because the rectangular window can smooth out the aberrations caused by the decision errors.

Figure 10.10 shows the performance of RISIC technique with channel estimation on static and very slowly fading ISI channels. For a static ISI channel, the SER obtained is virtually identical to that obtained with perfect channel estimation. For a slowly varying fading ISI channel ($f_m N T_s = 0.001$), the RISIC technique works well with the channel estimator, especially at high SNR. However, for faster fading ($f_m N T_s = 0.005$), the degradation is severe as shown in Fig. 10.11. This suggests that the proposed RISIC technique with channel estimation is well suited for static and slowly time-varying ISI channels.

10.2 Single-Carrier Frequency Domain Equalization

As discussed in Sect. 10.1, OFDM can easily mitigate ISI by using a cyclic guard interval. However, one of the major drawbacks of OFDM is its high peak-to-average power ratio (PAPR) and high sensitivity to oscillator frequency offsets. An attractive alternative to OFDM is the use of single-carrier (SC) modulation combined with frequency domain equalization (FDE), a technique known as single-carrier frequency domain equalization (SC-FDE) [121, 290–292]. Such a system has the benefit of having the PAPR of single-carrier modulation while at the same time the ease of equalization of OFDM. For channels that span 10–20 data symbols, single-carrier systems with conventional time-domain equalization will have a significantly higher processing complexity than comparable OFDM approaches. However, the complexity of single-carrier approaches becomes similar to OFDM if FDE is employed in the single-carrier receiver.

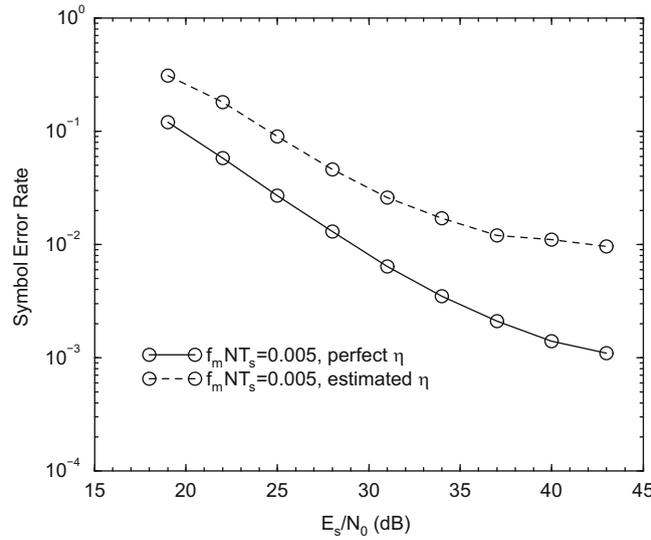


Fig. 10.11 Effect of imperfect channel estimation with the RISIC technique for a fading ISI channel; $N = 1024$, $G = 0$, $I = 3$

Fig. 10.12 SC-FDE transmitter with guard interval insertion

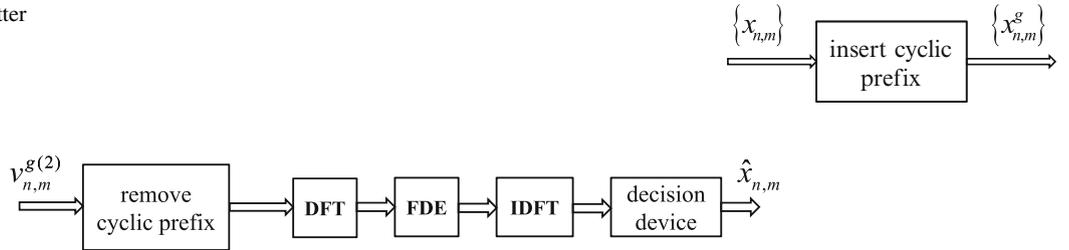
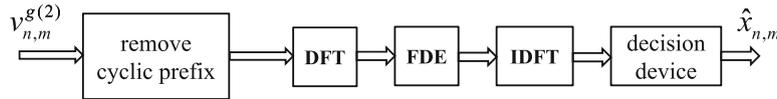


Fig. 10.13 SC-FDE receiver



A typical SC-FDE transmitter is shown in Fig. 10.12, while the receiver is shown in Fig 10.13. In contrast to OFDM that uses an IDFT at the transmitter and an DFT at the receiver, a SC-FDE system has both the DFT and IDFT at the receiver side. SC-FDE can be used advantageously in wireless systems, by using OFDM on the downlink and SC modulation on the uplink. With this arrangement, the mobile terminal transmits SC modulation and, thus, it is simple to implement and can use a more efficient power amplifier. Also, the signal processing complexity is concentrated in the access point. For full duplex operation, the access point has two IDFT and one DFT operations, while the mobile terminal has just one DFT operation.

One important difference between OFDM and SC-FDE is the exploitation of frequency diversity. OFDM can achieve maximum likelihood (ML) performance on a frequency-selective channel through maximal ratio combining of the symbols that are transmitted on the various sub-carriers. SC-FDE approaches are sub-optimal in this regard and do not achieve ML performance. Nevertheless, OFDM only reaps the full benefits of frequency diversity by coding across the sub-carriers or by using receiver antenna diversity. Uncoded OFDM will lose the inherent frequency diversity in the channel, since information data on the sub-carriers affected by the frequency-selective fading cannot be recovered from the other sub-carriers. On the other hand, the energy of each SC symbol is spread over the entire frequency band, providing inherent frequency diversity. Consequently, uncoded SC-FDE outperforms uncoded OFDM. However, for a single-user, multi-antenna system (receive diversity) the advantage of SC-FDE over OFDM decreases as the number of receiver antennas increases.

Similar to the guard interval used in OFDM and SC-FDE for linear modulation schemes, a length- G cyclic guard interval is appended to each block of N data symbols, such that G equals or exceeds the maximum expected discrete-time channel length L . Here a cyclic prefix is assumed. The cyclic prefix is added to the sequence $\mathbf{x}_n = \{x_{n,k}, k = 0, \dots, N-1\}$ by simply copying the last G symbols of $\{x_{n,k}\}$ and appending them to the beginning of $\{x_{n,k}\}$. The symbol sequence with guard interval, denoted as $\{x_{n,k}^g\}$, is

$$\mathbf{x}_n^g = \{x_{n,(k)N}, k = -G, -G+1, \dots, -1, 0, 1, \dots, N-1\}, \quad (10.46)$$

where $(k)_N$ is the residue of k modulo- N . The sequence of $N + G$ symbols is then transmitted using SC modulation. Here, a linear modulation scheme is assumed such as PSK or QAM, although SC-FDE can be applied with appropriate modification to non-linear CPM as well [313]. With this assumption, the transmitted SC waveform has the form

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n^g), \quad (10.47)$$

where

$$b(t, \mathbf{x}_n^g) = \sum_{k=-G}^{N-1} x_{n,k}^g h_a(t - kT_s^g) \quad (10.48)$$

and $T = (N + G)T_s^g$ is the duration of each data block consisting of N data symbols and G guard symbols.

10.2.1 ZF and MMSE SC-FDE

Suppose that the waveform $\tilde{s}(t)$ is transmitted over a time-invariant ISI channel with impulse response $g(t)$. Section 7.6.4 established that the optimal front-end processing may be implemented as a filter that is matched to the transmitted pulse $h_a(t)$ followed by rate- $2/T$ sampling and a $T/2$ -spaced noise whitening filter $1/\left(G_h^{(2)}(1/z^*)\right)^*$. Similar to the development leading to (10.4), the received sample sequence at the output of the $T/2$ -spaced whitened matched filter is

$$v_{n,k}^{g(2)} = \begin{cases} \sum_{i=0}^{k+2G} g_{\text{eq},i}^{(2)} x_{n,k-i}^{g(2)} + \sum_{i=k+2G+1}^{2L} g_{\text{eq},i}^{(2)} x_{n-1,2N+2G+k-i}^{g(2)} + \tilde{w}_{n,k}^{(2)}, & -2G \leq k < 2(-G + L) \\ \sum_{i=0}^{2L} g_{\text{eq},i}^{(2)} x_{n,k-i}^{g(2)} + \tilde{w}_{n,k}^{(2)}, & 2(-G + L) \leq k \leq 2N - 1 \end{cases}, \quad (10.49)$$

where the sequence $\mathbf{x}_n^{g(2)} = \{x_{n,k}^{g(2)}\}$ is the corresponding $T/2$ -spaced input symbol sequence given by

$$x_{n,k}^{g(2)} = \begin{cases} x_{k/2}^g, & k = -2G, \dots, 0, 2, 4, \dots, 2N - 2 \\ 0, & k = -2G + 1, \dots, 1, 3, 5, \dots, 2N - 1 \end{cases} \quad (10.50)$$

and $\{g_{\text{eq},i}^{(2)}\}$ is the $T/2$ -spaced discrete-time channel impulse response with system function $G_{\text{eq}}^{(2)}(z) = G_h^{(2)}(z)G^{(2)}(z)$.

The SC-FDE first removes the guard interval by simply deleting the first $2G$ samples in each block, i.e., samples in the range $-2G \leq k \leq -1$. Under the assumption that $G \geq L$, this yields the length- $2N$ sequence

$$v_{n,k}^{(2)} = \sum_{i=0}^{2L} g_{\text{eq},i}^{(2)} x_{n,k-i}^{g(2)} + \tilde{w}_{n,k}^{(2)} \quad (10.51)$$

$$= \sum_{i=0}^{2L} g_{\text{eq},i}^{(2)} x_{n,(k-i)_N}^{(2)} + \tilde{w}_{n,k}^{(2)}, \quad 0 \leq k \leq 2N - 1, \quad (10.52)$$

where

$$x_{n,k}^{(2)} = \begin{cases} x_{k/2}, & k = 0, 2, 4, \dots, 2N - 2 \\ 0, & k = 1, 3, 5, \dots, 2N - 1 \end{cases}. \quad (10.53)$$

Note that the length- $2N$ vector $\mathbf{v}_n^{(2)} = \{v_{n,k}^{(2)}\}$ in (10.52) is the circular convolution of the length- $2N$ data sequence $\mathbf{x}_n^{(2)} = \{x_{n,k}^{(2)}\}$ with the length- $2L$ channel vector $\mathbf{g}_{\text{eq}}^{(2)} = \{g_{\text{eq},i}^{(2)}\}$.

To perform FDE, the length- $2N$ vector $\mathbf{v}_m^{(2)}$ is applied to a $2N$ -point discrete Fourier transform (DFT) or fast Fourier transform (FFT) to yield the output vector

$$\begin{aligned} V_{n,m}^{(2)} &= \frac{1}{2N} \sum_{k=0}^{2N-1} v_{n,k}^{(2)} e^{-j\frac{2\pi km}{2N}} \\ &= T_{\text{eq},m}^{(2)} X_{n,m}^{(2)} + W_{n,m}^{(2)}, \quad 0 \leq m \leq 2N-1, \end{aligned} \quad (10.54)$$

where

$$X_{n,m}^{(2)} = \frac{1}{2N} \sum_{k=0}^{2N-1} x_{n,k}^{(2)} e^{-j\frac{2\pi km}{2N}}, \quad m = 0, \dots, 2N-1 \quad (10.55)$$

is the $2N$ -point DFT of the data symbol vector $\mathbf{x}_n^{(2)}$,

$$T_{\text{eq},m}^{(2)} = \sum_{k=0}^L g_{\text{eq},m}^{(2)} e^{-j\frac{2\pi km}{N}}, \quad m = 0, \dots, 2N-1 \quad (10.56)$$

is the channel frequency response, and the noise samples $\{W_{n,m}^{(2)}\}$ are i.i.d with zero mean and variance $N_o/(2NT_s^g)$. Note that $\mathbf{T}_{\text{eq}}^{(2)} = \{T_{\text{eq},m}^{(2)}\}_{m=0}^{2N-1}$ is the DFT of the zero padded sequence $\mathbf{g}_{\text{eq}}^{(2)} = \{g_{\text{eq},k}^{(2)}\}_{k=0}^{N-1}$ and is equal to the sampled frequency response of the channel. Also, every other coordinate of the vector $\mathbf{x}_n^{(2)}$ being zero leads to the property $X_{n,m}^{(2)} = X_{n,m+N}^{(2)}$ for $m = 0, \dots, N-1$, i.e., the first half of the length- $2N$ vector $\mathbf{X}_n^{(2)} = \{X_{n,m}^{(2)}\}_{m=0}^{2N-1}$ is identical to the last half.

FDE may now be carried out by processing the vector $\mathbf{V}_n^{(2)} = \{V_{n,m}^{(2)}\}$. One option is to estimate $T_{\text{eq},m}^{(2)}$ and calculate

$$\begin{aligned} \hat{V}_{n,m}^{(2)} &= \frac{V_{n,m}^{(2)}}{T_{\text{eq},m}^{(2)}} \\ &= X_{n,m}^{(2)} + \frac{W_{n,m}^{(2)}}{T_{\text{eq},m}^{(2)}}, \quad m = 0, \dots, 2N-1. \end{aligned} \quad (10.57)$$

This is sometimes called zero-forcing frequency domain equalization (ZF-FDE), because the equalizer tries to invert the channel in the frequency domain. Afterwards, the $2N$ -point IDFT of the vector $\hat{\mathbf{V}}_n^{(2)} = \{\hat{V}_{n,m}^{(2)}\}_{m=0}^{2N-1}$ is calculated to yield the length- $2N$ vector $\tilde{\mathbf{x}}_n^{(2)} = \{\tilde{x}_{n,k}^{(2)}\}$, where

$$\tilde{x}_{n,k}^{(2)} = x_{n,k}^{(2)} + \tilde{w}_{n,k}^{(2)}, \quad k = 0, \dots, 2N-1 \quad (10.58)$$

and

$$\tilde{w}_{n,k}^{(2)} = \sum_{m=0}^{N-1} \left(\frac{W_{n,m}^{(2)}}{T_{\text{eq},m}^{(2)}} \right) e^{j\frac{2\pi mk}{2N}}, \quad k = 0, 1, \dots, 2N-1. \quad (10.59)$$

The length- $2N$ vector $\tilde{\mathbf{x}}_n^{(2)}$ is then decimated by taking every other sample, i.e., $\tilde{x}_{n,k} = \tilde{x}_{n,2k}^{(2)}$, $k = 0, \dots, N-1$, and decisions are made from the resulting vector $\tilde{\mathbf{x}}_n = \{\tilde{x}_{n,k}\}$. For *each* of the $\tilde{x}_{n,k}$, the receiver decides in favor of the data symbol $x_{n,k}$ that minimizes the squared Euclidean distance

$$\mu(x_{n,k}) = \|\tilde{x}_{n,k} - x_{n,k}\|^2, \quad k = 0, \dots, N-1. \quad (10.60)$$

Observe that the noise samples $W_{n,m}^{(2)}/T_{\text{eq},m}^{(2)}$ in (10.57) are independent zero-mean Gaussian random variables with variance $N_o/(2NT_s^g |T_{\text{eq},m}^{(2)}|^2)$. Hence, the noise is amplified in sub-bands where $|T_{\text{eq},m}^{(2)}|$ is small, and if $|T_{\text{eq},m}^{(2)}|$ is very small $\hat{V}_{n,m}^{(2)}$ will be corrupted by heavy noise. This effect is similar to noise enhancement in time-domain ZF linear equalizers and can severely degrade the bit error probability.

Another possibility is to account for the effects of noise enhancement by performing minimum mean square error frequency domain equalization (MMSE-FDE). Assume that the MMSE-FDE has tap gains $C_m, m = 0, \dots, N - 1$. Then the output of the MMSE-FDE is

$$\begin{aligned}\hat{V}_{n,m}^{(2)} &= V_{n,m}^{(2)} C_m \\ &= T_{\text{eq},m}^{(2)} X_{n,m}^{(2)} C_m + W_{n,m}^{(2)} C_m, \quad 0 \leq m \leq 2N - 1.\end{aligned}\quad (10.61)$$

The mean square error is defined as

$$\begin{aligned}J &= \frac{1}{2} \text{E} \left[|\hat{V}_{n,m}^{(2)} - X_{n,m}^{(2)}|^2 \right] \\ &= |T_{\text{eq},m}^{(2)} C_m - 1|^2 \sigma_x^2 / (2N) + |C_m|^2 N_o / (2N) \\ &= \left(T_{\text{eq},m}^{(2)} \sigma_x^2 + N_o \right) |C_m|^2 / (2N) - 2\hat{A} \sigma_x^2 \text{Re} \left\{ T_{\text{eq},m}^{(2)} C_m \right\} / (2N) + \sigma_x^2 / (2N),\end{aligned}\quad (10.62)$$

where the second line uses $\frac{1}{2} \text{E} [|X_{n,m}^{(2)}|^2] = \sigma_x^2 / (2N)$ as seen from Parseval's theorem

$$\frac{1}{2N} \sum_{k=0}^{2N-1} |\tilde{x}_{n,k}^{(2)}|^2 = \sum_{m=0}^{2N-1} |X_{n,m}^{(2)}|^2. \quad (10.63)$$

The MMSE solution can be obtained by solving

$$\frac{dJ}{dC_m} = \frac{dJ}{dC_{R,m}} + j \frac{dJ}{dC_{I,m}} = 0, \quad m = 0, \dots, N - 1, \quad (10.64)$$

where $C_m = C_{R,m} + jC_{I,m}$. This gives the solution

$$C_m = \frac{\sigma_x^2 T_{\text{eq},m}^{(2)*}}{\sigma_x^2 |T_{\text{eq},m}^{(2)}|^2 + N_o}, \quad m = 0, \dots, 2N - 1. \quad (10.65)$$

Note that under high signal-to-noise ratio conditions where $N_o \rightarrow 0$, the tap solution for the MMSE-FDE in (10.65) reduces to the ZF-FDE tap solution $C_m = 1/T_{\text{eq},m}^{(2)}$ used in (10.57). The output of the MMSE-FDE is

$$\begin{aligned}\hat{V}_{n,m}^{(2)} &= V_{n,m}^{(2)} C_m \\ &= \frac{\sigma_x^2 |T_{\text{eq},m}^{(2)}|}{\sigma_x^2 |T_{\text{eq},m}^{(2)}|^2 + N_o} X_{n,m}^{(2)} + \frac{\sigma_x^2 T_{\text{eq},m}^{(2)*}}{\sigma_x^2 |T_{\text{eq},m}^{(2)}|^2 + N_o} W_{n,m}^{(2)}, \quad m = 0, \dots, 2N - 1.\end{aligned}\quad (10.66)$$

Once again, the IDFT of the vector $\hat{\mathbf{V}}_n^{(2)} = \{\hat{V}_{n,m}^{(2)}\}_{m=0}^{2N-1}$ is obtained to yield the vector $\tilde{\mathbf{x}}_n^{(2)}$, which is decimated to yield the vector $\tilde{\mathbf{x}}_n = \{\tilde{x}_{n,k} = \tilde{x}_{n,2k}^{(2)}, k = 0, \dots, N - 1\}$, and the vector $\tilde{\mathbf{x}}_n$ is applied to a minimum distance decision device as in (10.60).

Finally, while decisions variables $\tilde{Z}_{n,m}$ in (10.8) can be used to implement maximum likelihood decisions in an OFDM receiver, the corresponding decision variables $\tilde{x}_{n,k}$ in a SC-FDE receiver will *not* yield maximum likelihood decisions. In this sense, the SC-FDE approach is sub-optimal.

10.3 Orthogonal Frequency Division Multiple Access

Orthogonal frequency division multiple access (OFDMA) achieves multiple access by assigning different users disjoint sets of sub-carriers. Assume that there are a total of M sub-carriers that are evenly distributed among Q users, such that each user is allocated $N = M/Q$ sub-carriers. The overall sub-carriers are labeled with indices from 0 to $M - 1$, while the N

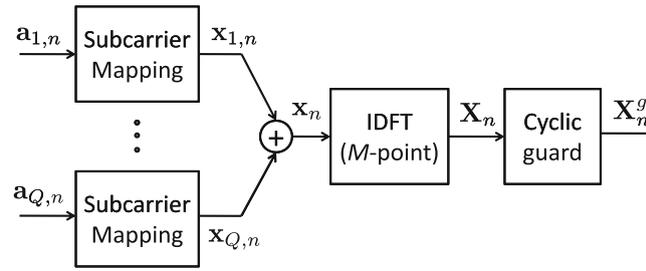


Fig. 10.14 Baseband OFDMA forward link BS transmitter. There are M sub-carriers of which N are occupied by the input data for each of Q users

sub-carriers allocated to the j th MS have indices that belong to the set \mathcal{T}_j . Clearly, the sets \mathcal{T}_j must be disjoint such that each sub-carrier is assigned to at most one MS. The sub-carrier allocation can be performed by extending the n th data vector for the j th MS, denoted by $\mathbf{a}_{j,n}$ with the insertion of $M - N$ zeros in the sub-carriers belonging to the set $\tilde{\mathcal{T}}_j$ which is the complement of \mathcal{T}_j , i.e.,

$$x_{j,n,i} = \begin{cases} a_{j,n,i}, & \text{if } i \in \mathcal{T}_j \\ 0, & \text{otherwise} \end{cases}, \quad (10.67)$$

where $a_{j,n,i}$ is the data symbol transmitted to the j th MS in block n on the i th sub-carrier.

10.3.1 OFDMA Forward Link

10.3.1.1 Transmitter

A block diagram of an OFDMA baseband forward link transmitter is shown in Fig. 10.14. On the forward link, the vectors $\mathbf{x}_{j,n} = \{x_{j,n,i}\}_{i=0}^{M-1}$ are summed up to produce the n th data block

$$\mathbf{x}_n = \sum_{j=1}^Q \mathbf{x}_{j,n} \quad (10.68)$$

that is subsequently applied to an M -point IDFT to produce the length- M time-domain sequence \mathbf{X}_n . After the IDFT, a length- G guard interval is appended to each block in the form of a cyclic prefix or cyclic suffix, in the same manner as conventional OFDM as described in Sect. 4.6.1, to yield the transmitted time-domain sequence \mathbf{X}_n^g . In the case of a cyclic prefix, the last G symbols of the sequence $\mathbf{X}_n = \{X_{n,m}, m = 0, \dots, M-1\}$ are copied and appended to the beginning of \mathbf{X}_n . The transmitted time-domain sequence for the n th block with guard interval, denoted as \mathbf{X}_n^g , is

$$\mathbf{X}_n^g = \{X_{n,(m)_M}, m = -G, -G+1, \dots, -1, 0, 1, \dots, M-1\}, \quad (10.69)$$

where $(m)_M$ is the residue of m modulo- M .

10.3.1.2 Time-Domain Windowing

The elements of the vector \mathbf{X}_n^g are converted into the continuous-time complex-valued waveform by using pair of balanced DACs, as discussed in Sect. 4.6.1. The resulting continuous-time complex envelope corresponding to the n th block is

$$\tilde{s}_n(t) = \sum_{k=-G}^{M-1} \tilde{x}_{n,k} h_{\text{DAC}}(t - k\tilde{T}_s^g), \quad (10.70)$$

where $h_{\text{DAC}}(t)$ is the impulse response of the DAC (reconstruction filter) and \tilde{T}_s^g is the sample period. For OFDMA, the sample period is equal to $\tilde{T}_s^g = (M/(M+G))\tilde{T}_s = (N/(M+G))T_s$ seconds.

For greater spectral control and to reduce out-of-band emissions, the vector of samples \mathbf{X}_n^g can be windowed prior to digital-to-analog conversion, such that

$$\tilde{s}_n(t) = \sum_{k=-G}^{M-1} \tilde{x}_{n,k} w_k h_{\text{DAC}}(t - k\tilde{T}_s^g), \quad (10.71)$$

where the vector $\mathbf{w} = (w_{-G}, \dots, w_{-1}, w_0, w_1, \dots, w_{M-1})$ is the window function. The corresponding continuous-time window function $w(t)$ can be obtained by passing the weighted impulse train

$$\delta_{\tilde{T}_s^g}(t) = \sum_{k=-G}^{M-1} w_k \delta(t - k\tilde{T}_s^g) \quad (10.72)$$

through the reconstruction filter $h_{\text{DAC}}(t)$. Here, an ideal DAC is assumed with impulse response

$$h_{\text{DAC}}(t) = \text{sinc}(t/\tilde{T}_s^g). \quad (10.73)$$

For a rectangular window, $w_n = 1$, $-G \leq n \leq M-1$. This yields the continuous-time window function

$$w(t) = \sum_{k=-G}^{M-1} \text{sinc}(t/\tilde{T}_s^g - k), \quad (10.74)$$

which is plotted in Fig. 10.15 for the case $M = 32$, $G = 4$. Notice that the continuous-time window function $w(t)$ is non-causal, since the ideal DAC is itself a non-causal ideal low-pass filter.

With an ideal DAC, the power spectrum of the baseband OFDMA waveform is confined to $-1/2\tilde{T}_s^g$ to $+1/2\tilde{T}_s^g$ Hz. However, a practical DAC will produce side lobes outside of this range of frequencies and, in some cases, the rectangular window may result in large side lobes. Such out-of-band emissions can be reduced by using a time-domain window that

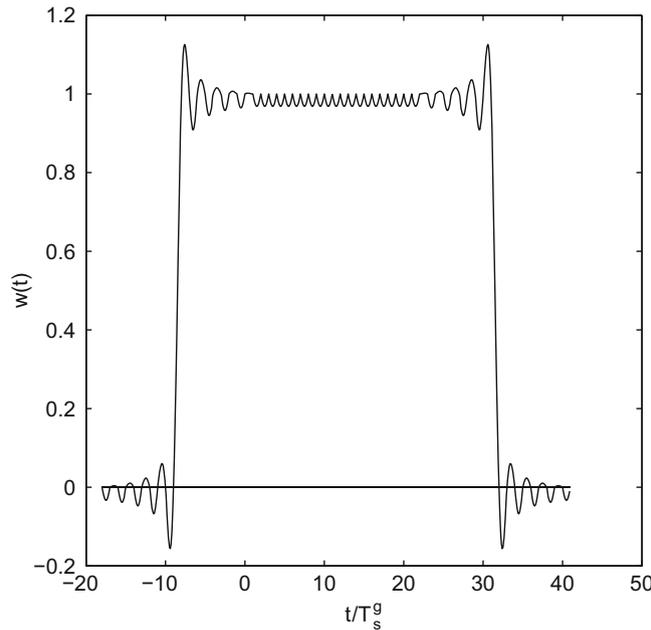


Fig. 10.15 Continuous-time rectangular window function $w(t)$, $M = 32$, $G = 8$

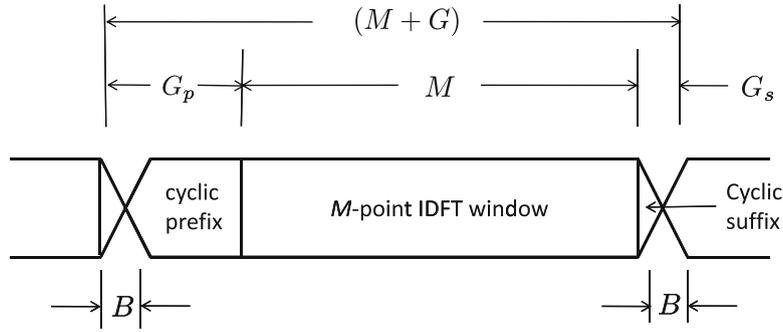


Fig. 10.16 OFDMA symbols with time-domain windowing. Note that $B = G_s + 1$, where G_s is the length of the cyclic suffix and B is the length of the transition region

smooths the ends of the OFDMA symbols. Such windowing can be implemented by extending the OFDMA symbol with both a cyclic prefix and cyclic suffix as shown in Fig. 10.16. That is,

$$\mathbf{X}_n^g = \{X_{n,(m)M}, m = -G_p, \dots, -1, 0, 1, \dots, M-1, \dots, M + G_s - 1\}, \quad (10.75)$$

where G_p and G_s are the length of the cyclic prefix and cyclic suffix, respectively, and $G = G_p + G_s$ is the total length of the guard interval. In this way, the required extension to implement the time-domain window can be absorbed into the guard interval. This will save bandwidth and power efficiency but will also reduce the delay spread tolerance by $2G_s$ samples.

One commonly used time-domain window function with OFDM/OFDMA is the raised cosine window

$$w_k^{\text{rc}} = \begin{cases} \frac{1}{2} + \frac{1}{2} \sin \left[\frac{\pi}{B} \left(k - \frac{B-1}{2} \right) \right], & 0 \leq k < B \\ 1, & B \leq k < M + G - B \\ \frac{1}{2} - \frac{1}{2} \sin \left[\frac{\pi}{B} \left(k - M - G + \frac{B+1}{2} \right) \right], & M + G - B \leq k \leq M + G - 1 \\ 0, & \text{elsewhere} \end{cases} \quad (10.76)$$

defined here on the interval $0 \leq k \leq M + G - 1$. For the sequence \mathbf{X}_n^g in (10.75), the required window is the raised cosine window in (10.76) left shifted by G_p samples, such that

$$w_k = w_{k+G_p}^{\text{rc}}. \quad (10.77)$$

After passing through an ideal D/A converter, this will yield the continuous-time pulse

$$w(t) = \sum_{k=-G}^{M-1} w_k \text{sinc}(t/\tilde{T}_s^g - k) \quad (10.78)$$

which is plotted in Fig. 10.17 for the case $M = 32$, $G_p = 4$, $G_s = 4$. Notice that the raised cosine windowing has also significantly reduced the peaks on the amplitude shaping pulse. Hence, the time-domain windowing not only produces a more compact power spectral density, but it also reduces the PAPR of the transmitted waveform.

10.3.1.3 Sub-carrier Allocation

Several different methods for sub-carrier allocation are used with OFDMA. The choice of sub-carrier allocation will affect the sensitivity of the OFDMA waveform to frequency-selective fading.

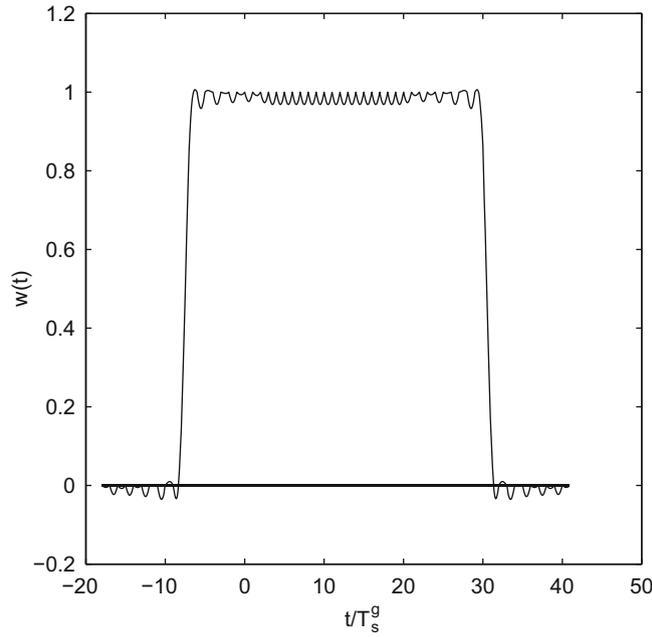


Fig. 10.17 Continuous-time raised cosine window function $w(t)$, $M = 32$, $G_p = 4$, $G_s = 4$

Clustered Carrier (CC-OFDMA)

With CC-OFDMA, the M sub-carriers are divided into Q groups where each group consists of N contiguous sub-carriers called clusters. The set of sub-carrier indices allocated to the k th user is $\{kN, kN + 1, \dots, kN + N - 1\}$, where $0 \leq k < Q$. CC-OFDMA is sensitive to frequency-selective fading, because all sub-carriers assigned to a particular user may fade simultaneously.

Spaced Carrier (SC-OFDMA)

With SC-OFDMA, the M sub-carriers are partitioned into N groups, where each group has Q contiguous sub-carriers. Then the k th sub-carrier of each group is assigned to the k th user. That is, the k th user is assigned the set of sub-carrier indices $\{k, Q + k, \dots, (N - 1)Q + k\}$, where $0 \leq k < Q$. SC-OFDMA is less sensitive to frequency-selective fading, since the sub-carriers assigned to each user span the entire bandwidth.

Random Interleaving (RI-OFDMA)

RI-OFDMA has been adopted by IEEE802.16a. While the sub-carriers are partitioned into N groups as in SC-OFDMA, the sub-carrier index in each of the N groups that is assigned to a particular user is a random variable. The sub-carrier indices allocated to the k th user are $\{\epsilon_{k,1}, Q + \epsilon_{k,2}, \dots, (M - 1)Q + \epsilon_{k,M-1}\}$, where the $\epsilon_{k,i}$ are independent and identically distributed uniform random variables on the set $\{0, 1, \dots, Q - 1\}$.

10.3.1.4 Receiver

Similar to the case of OFDM in Sect. 10.1, consider the transmission of the OFDMA waveform over a frequency-selective quasi-static fading channel. Assuming a length- G cyclic prefix, the discrete-time linear convolution of the transmitted sequence $\{\mathbf{X}_n^g\} = \{X_{n,m}\}_{m=-G}^{M-1}$ with the discrete-time channel impulse response $\mathbf{g} = \{g_m\}_{m=0}^L$ produces the discrete-time received sequence $\mathbf{R}_n^g = \{R_{n,m}^g\}$, where

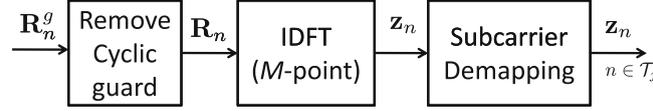


Fig. 10.18 Baseband OFDMA forward link receiver

$$R_{n,m}^g = \begin{cases} \sum_{i=0}^{m+G} g_i X_{n,m-i}^g + \sum_{i=m+G+1}^L g_i X_{n-1,M+G+m-i}^g + \tilde{n}_{n,m}, & -G \leq m < -G + L \\ \sum_{i=0}^L g_i X_{n,m-i}^g + \tilde{n}_{n,m}, & -G + L \leq m \leq M - 1 \end{cases} \quad (10.79)$$

and the $\tilde{n}_{n,m}$ are complex-valued Gaussian noise samples having zero mean and variance $\sigma^2 = \frac{1}{2}E[|\tilde{n}_{n,m}|^2] = N_o/T_s^g$, where $T_s^g = NT_s/(M + G)$, and T_s is the modulation symbol rate for each user.

The OFDMA baseband receiver is shown in Fig. 10.18. To remove the ISI introduced by the channel, the first G received samples $\{R_{n,m}^g\}_{m=-G}^{-1}$ of each block are simply discarded. If the length of the cyclic prefix is at least as long as the discrete-time channel length, i.e., $G \geq L$, then the received sequence is

$$\begin{aligned} R_{n,m} &= R_{n,m}^g \\ &= \sum_{i=0}^L g_i X_{n,(m-i)_M} + \tilde{n}_{n,m}, \quad 0 \leq m \leq M - 1, \end{aligned} \quad (10.80)$$

which is just the circular convolution of the transmitted sequence $\mathbf{X}_n = \{X_{n,m}\}$ with the channel $\mathbf{g} = \{g_m\}_{m=0}^L$. Afterwards, an M -point IDFT is taken to transform to the frequency domain. This yields the output vector

$$\begin{aligned} z_{n,i} &= \frac{1}{M} \sum_{m=0}^{M-1} R_{n,m} e^{-j\frac{2\pi mi}{M}} \\ &= T_i A x_{n,i} + v_{n,i}, \quad 0 \leq i \leq M - 1, \end{aligned} \quad (10.81)$$

where

$$T_i = \sum_{m=0}^L g_m e^{-j\frac{2\pi mi}{M}} \quad (10.82)$$

and the noise samples $\{v_{n,i}\}$ are i.i.d with zero mean and variance $N_o/(MT_s^g)$. On the forward link each MS will only be interested in the N data symbols that are transmitted by the BS on its allocated sub-carriers. Hence, only the DFT outputs with indices in the set \mathcal{T}_j are used by the j th MS for data detection. The resulting time-domain sequence is further processed, using, for example, maximum likelihood sequence estimation (MLSE) along with the estimates of the channel gains $\{T_i\}$, $i \in \mathcal{T}_j$.

10.3.2 OFDMA Reverse Link

On the OFDMA reverse link, Q users transmit their signals to a central BS. The reverse link MS transmitter is shown in Fig. 10.19, and is similar to the forward link BS transmitter in Fig. 10.14. The main difference is that the MS transmitter only transmits its own data stream, whereas the BS transmitter sends data streams simultaneously for all the MSs. Similar to the OFDMA forward link, the j th MS performs sub-carrier allocation according to (10.67), and the resulting vector $\mathbf{x}_{j,n}$ is applied to an M -point IDFT, and appended with a length- G cyclic guard interval.

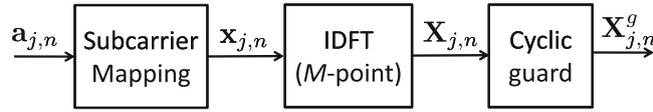


Fig. 10.19 Baseband OFDMA reverse link MS transmitter. There are M sub-carriers of which N are occupied by the input data

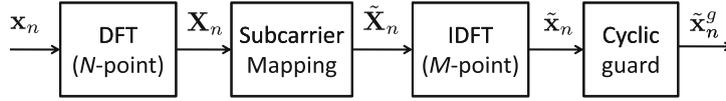


Fig. 10.20 Baseband SC-FDMA transmitter. There are M sub-carriers of which N are occupied by the input data

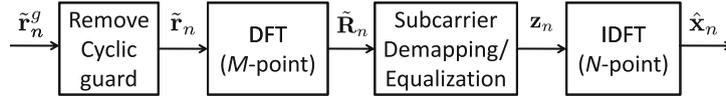


Fig. 10.21 Baseband SC-FDMA receiver with SC-FDE

10.3.3 Peak-to-Average Power Ratio

One of the biggest drawbacks of OFDMA is its high PAPR. A high PAPR may be acceptable on the forward link, since the BS can use a less power efficient linear amplifier. However, for the reverse link, a high PAPR is undesirable since the MS is often battery powered. For this reason, OFDMA is not used on the reverse link of LTE and LTE-A.

10.4 Single-Carrier Frequency Division Multiple Access

A block diagram of a single-carrier frequency division multiple access (SC-FDMA) transmitter is shown in Fig. 10.20. The SC-FDMA transmitter groups the modulation symbols into blocks of N symbols. Let

$$\mathbf{x}_n = (x_{n,1}, x_{n,2}, \dots, x_{n,N}) \quad (10.83)$$

denote the n th block of modulation symbols. An N -point DFT (N -DFT) is taken on each block \mathbf{x}_n , to yield length- N vectors

$$\mathbf{X}_n = (X_{n,1}, X_{n,2}, \dots, X_{n,N}) \quad (10.84)$$

that are the frequency domain representation of the blocks of input symbols. The sub-carrier mapper then maps the N components of the vector \mathbf{X}_n onto a larger set of M sub-carriers such that $M = NQ$, where Q is an integer. There are several different types of sub-carrier mappings, including the interleaved (I-FDMA) and localized (L-FDMA) mappings that are considered below. The sub-carrier mapping generates the sequence $\tilde{\mathbf{X}}_n$. An M -point IDFT is then taken of the sequence $\tilde{\mathbf{X}}_n$ to produce the output sequence $\tilde{\mathbf{x}}_n$. The time domain input symbols $x_{n,k}$ have duration T_s seconds. However, after going through the SC-FDMA modulator the time-domain output symbols $\tilde{x}_{n,k}$ are compressed and have duration $\tilde{T}_s = (N/M)T_s$ seconds.

The SC-FDMA baseband receiver is shown in Fig. 10.21. First the cyclic guard interval is removed. Afterwards, an M -point DFT is taken to transform to the frequency domain. Sub-carrier demapping and equalization is then performed in the frequency domain. Finally, an N -point IDFT is used to convert the samples back to the time domain for detection and further processing.

Interleaved (I-FDMA)

The I-FDMA sub-carrier mapping is illustrated in Fig. 10.22 and can be described as follows:

$$\tilde{X}_{n,\ell} = \begin{cases} X_{n,\ell/Q}, & \ell = kQ, \quad 0 \leq k \leq N-1 \\ 0, & \text{otherwise} \end{cases} \quad (10.85)$$

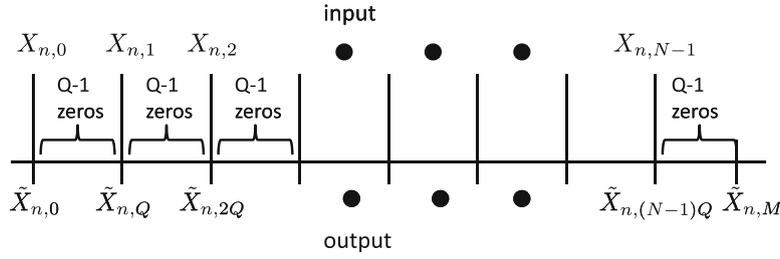


Fig. 10.22 Interleaved FDMA (I-FDMA) sub-carrier mapping

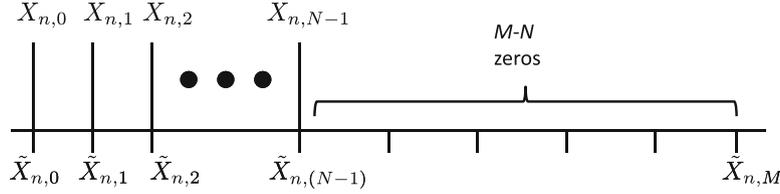


Fig. 10.23 Localized FDMA (L-FDMA) sub-carrier mapping

The time-domain vector $\tilde{\mathbf{x}}_n$ is obtained by taking the M -point inverse DFT of the vector $\tilde{\mathbf{X}}_n$. Let $k = Nq + m$, where $0 \leq q \leq Q - 1$ and $0 \leq m \leq N - 1$. Then

$$\begin{aligned}
 \tilde{x}_{n,k} &\equiv \tilde{x}_{n,Nq+m} = \frac{1}{M} \sum_{\ell=0}^{M-1} \tilde{X}_{n,\ell} e^{j\frac{2\pi\ell k}{M}} \\
 &= \frac{1}{QN} \sum_{\ell=0}^{N-1} X_{n,\ell} e^{j\frac{2\pi\ell k}{N}} \\
 &= \frac{1}{QN} \sum_{\ell=0}^{N-1} X_{n,\ell} e^{j\frac{2\pi\ell(Nq+m)}{N}} \\
 &= \frac{1}{Q} \left(\frac{1}{N} \sum_{\ell=0}^{N-1} X_{n,\ell} e^{j\frac{2\pi\ell m}{N}} \right) \\
 &= \frac{1}{Q} x_{n,m}.
 \end{aligned} \tag{10.86}$$

It follows that the time-domain vector $\tilde{\mathbf{x}}_n$ is equal to

$$\tilde{\mathbf{x}}_n = \frac{1}{Q} \text{vec}(\mathbf{x}_n, \mathbf{x}_n, \dots, \mathbf{x}_n) \tag{10.87}$$

which is just the Q -fold repetition of the time-domain vector \mathbf{x}_n . It also follows that the PAPR of the time-domain sequence $\tilde{\mathbf{x}}_n$ is exactly the same as that of the time-domain sequence \mathbf{x}_n .

Localized (L-FDMA)

The L-FDMA, the sub-carrier mapping is illustrated in Fig. 10.23 and can be described as follows:

$$\tilde{X}_{n,\ell} = \begin{cases} X_{n,\ell}, & 0 \leq \ell \leq N - 1 \\ 0, & N \leq \ell \leq M - 1 \end{cases}. \tag{10.88}$$

The time-domain vector $\tilde{\mathbf{x}}_n$ is obtained by taking the M -point inverse DFT of the vector $\tilde{\mathbf{X}}_n$. Let $k = Qm + q$, where $0 \leq m \leq N - 1$ and $0 \leq q \leq Q - 1$. Then

$$\begin{aligned}
\tilde{x}_{n,k} &\equiv \tilde{x}_{n,Qm+q} = \frac{1}{M} \sum_{\ell=0}^{M-1} \tilde{X}_{n,\ell} e^{j\frac{2\pi\ell k}{M}} \\
&= \frac{1}{QN} \sum_{\ell=0}^{N-1} X_{n,\ell} e^{j\frac{2\pi\ell(Qm+q)}{QN}}.
\end{aligned} \tag{10.89}$$

If $q = 0$, then

$$\begin{aligned}
\tilde{x}_{n,k} &\equiv \tilde{x}_{n,Qm} = \frac{1}{Q} \left(\frac{1}{N} \sum_{\ell=0}^{N-1} X_{n,\ell} e^{j\frac{2\pi\ell Qm}{QN}} \right) \\
&= \frac{1}{Q} \left(\frac{1}{N} \sum_{\ell=0}^{N-1} X_{n,\ell} e^{j\frac{2\pi\ell m}{N}} \right) \\
&= \frac{1}{Q} x_{n,m}.
\end{aligned} \tag{10.90}$$

If $q \neq 0$, then using

$$X_{n,\ell} = \sum_{i=0}^{N-1} x_{n,i} e^{j\frac{2\pi i\ell}{N}} \tag{10.91}$$

in (10.89) gives [235]

$$\tilde{x}_{n,k} \equiv \tilde{x}_{n,Qm+q} = \frac{1}{Q} \left(1 - e^{j2\pi\frac{q}{Q}} \right) \frac{1}{N} \sum_{i=0}^{N-1} \frac{x_{n,i}}{1 - e^{j2\pi\left(\frac{m-i}{N} + \frac{q}{QN}\right)}}. \tag{10.92}$$

It follows that the time-domain vector $\tilde{\mathbf{x}}_n$ has an exact copy of the time-domain vector \mathbf{x}_n in N coordinates defined by $k = Qm$, $0 \leq m \leq N - 1$. In between these positions, the vector $\tilde{\mathbf{x}}_n$ has values that depend on a complex weighted sum of all the time-domain symbols in the time-domain vector \mathbf{x}_n . This will increase the PAPR significantly.

L-FDMA is the scheme adopted in 3GPP Long Term Evolution (LTE). While it may seem that L-FDMA is less desirable due to the high PAPR, the waveform is less sensitive to frequency offsets and phase noise, as compared to I-FDMA. While it may seem that L-FDMA also loses the diversity advantages of I-FDMA, it may be possible that all sub-carriers fade together and only one channel estimator is needed.

Cyclic Guard Interval and SC-FDE

After performing sub-carrier mapping, the transmitter inserts a cyclic guard interval of length- G , such that G equals or exceeds the maximum expected discrete-time channel length L . The purpose of the cyclic guard interval is to isolate the transmitted blocks and to permit single-carrier frequency domain equalization (SC-FDE) at the receiver as discussed in Sect. 10.2. Here a length- G cyclic prefix is assumed. The cyclic prefix is added to the sequence $\tilde{\mathbf{x}}_n = \{\tilde{x}_{n,m}, m = 0, \dots, M-1\}$ by simply copying the last G symbols of $\tilde{\mathbf{x}}_n$ and appending them to the beginning of $\tilde{\mathbf{x}}_n$. The symbol sequence with guard interval, denoted as $\tilde{\mathbf{x}}_n^g = \{\tilde{x}_{n,m}^g\}$, is

$$\tilde{\mathbf{x}}_n^g = \{\tilde{x}_{n,(m)_M}, m = -G, -G+1, \dots, -1, 0, 1, \dots, M-1\}, \tag{10.93}$$

where $(m)_M$ is the residue of m modulo- M . When a cyclic guard is appended, the time-domain output symbols $\tilde{x}_{n,m}^g$ have duration $\tilde{T}_s^g = (M/(M+G))\tilde{T}_s = (N/(M+G))T_s$ seconds.

Once the cyclic guard is removed and a DFT taken on the received block, SC-FDE can be applied. Typically, either a zero-forcing (ZF) or minimum mean square error (MMSE) SC-FDE is employed. The required algorithms are described in detail in Sect. 10.2 and are not repeated here. ZF-FDE and MMSE-FDE can be applied to SC-FDMA as well, but there are better solutions. These include maximal ratio combining (MRC) for flat fading AWGN channels, and optimum combining (OC) with receiver antenna diversity. With OC, the signal-to-interference-plus-noise ratio is maximized for a flat faded desired signal in the presence of one or more flat faded and dominant co-channel interferers.

10.4.1 Peak-to-Average Power Ratio

One of the biggest advantages of SC-FDMA is its low PAPR. The PAPR for block n can be defined in terms of the discrete-time SC-FDMA samples as

$$\text{PAPR}_n = \frac{\max_m \{|\tilde{x}_{n,m}^g|^2\}}{(M+G)^{-1} \sum_{n=-G}^{M-1} |\tilde{x}_{n,m}^g|^2}. \quad (10.94)$$

The PAPR in this case depends only on the modulation signal constellation and sub-carrier mapping. With SC-FDMA pulse shaping can be used to reduce the PAPR and to reduce out-of-band emissions. Since each user in the SC-FDMA uplink uses only a subset of the sub-carriers, different approaches for implementing the pulse shaping have been suggested. With the conventional approach, the sample sequence \tilde{x}_n^g is transmitted by using an amplitude shaping pulse $h_a(t)$ such that the continuous-time complex envelope corresponding to the n th block is

$$\tilde{s}_n(t) = \sum_{m=0}^{M+G-1} \tilde{x}_{n,m}^g h_a(t - m\tilde{T}_s^g). \quad (10.95)$$

For the continuous-time complex envelope $\tilde{s}_n(t)$, the PAPR for block n is defined as

$$\text{PAPR}_n = \frac{\max_{0 \leq t \leq (M+G)\tilde{T}_s^g} |\tilde{s}_n(t)|^2}{\frac{1}{(M+G)\tilde{T}_s^g} \int_0^{(M+G)\tilde{T}_s^g} |\tilde{s}_n(t)|^2 dt}. \quad (10.96)$$

In this case, the PAPR depends not only on the modulation signal constellation and sub-carrier mapping, but also on the pulse shaping waveform $h_a(t)$. For SC-FDMA, the shaping pulse $h_a(t)$ can be chosen to be a frequency domain root-raised cosine pulse, since receiver matched filtering will yield an overall raised cosine pulse that satisfies the first Nyquist criterion for intersymbol interference (ISI) free transmission. The required magnitude response is given by $|H_a(f)| = \sqrt{\tilde{T}_s^g} |P(f)|^{1/2}$, where $P(f)$ is the spectral raised cosine pulse defined in (4.47) with $T = \tilde{T}_s^g$. The root-raised cosine filtering can be implemented either in the time domain or the frequency domain. With time-domain filtering, the sequence $\tilde{x}_{n,m}$ is passed through a digital filter $h_{a,n}$ having the corresponding analog filter impulse response $h_a(t)$ in (4.50) with $T = \tilde{T}_s^g$. Implementation of the root-raised cosine time-domain filter requires that the sequence $\tilde{\mathbf{x}}_n^g = \{\tilde{x}_{n,m}\}_{m=-G}^{M-1}$ be upsampled prior to filtering, since the bandwidth of the root-raised cosine pulse exceeds the Nyquist frequency $1/2\tilde{T}_s^g$. Assuming an upsampling factor of F , the upsampled sequence is

$$\tilde{x}_{n,\ell}^u = \begin{cases} \tilde{x}_{n,\ell/F}, & \ell = k \cdot F, \quad k \text{ an integer} \\ 0, & \text{otherwise} \end{cases}. \quad (10.97)$$

If the roll-off factor β lies in the range $0 \leq \beta \leq 1$, then an upsampling factor $F = 2$ will suffice. The upsampled digital filter impulse response in this case is given by $h_{a,n}^u = h_a(n\tilde{T}_s^g/F)$, where $h_a(t)$ is defined in (4.50) with $T = \tilde{T}_s^g$.

An entirely different approach is to implement the root-raised cosine filtering in the frequency domain. This approach is feasible with SC-FDMA because of the DFT/IDFT processing that is used in the transmitter. Figure 10.24 shows a transmitter block diagram for SC-FDMA with a spectrum shaping filter in the frequency domain. As before, a block of data symbols $\mathbf{x}_n = (x_{n,1}, x_{n,2}, \dots, x_{n,N})$ is input to an N -point DFT to yield the length- N vector $\mathbf{X}_n = (X_{n,1}, X_{n,2}, \dots, X_{n,N})$. Once again, implementation of the root-raised cosine filtering requires that the input sequence \mathbf{x}_n be upsampled by a factor of F to yield the sequence $\mathbf{x}_n^u = \{x_{n,\ell}^u\}$, where

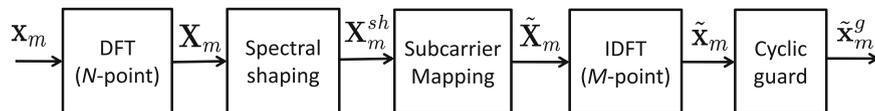


Fig. 10.24 SC-FDMA with frequency-domain root-raised cosine filtering

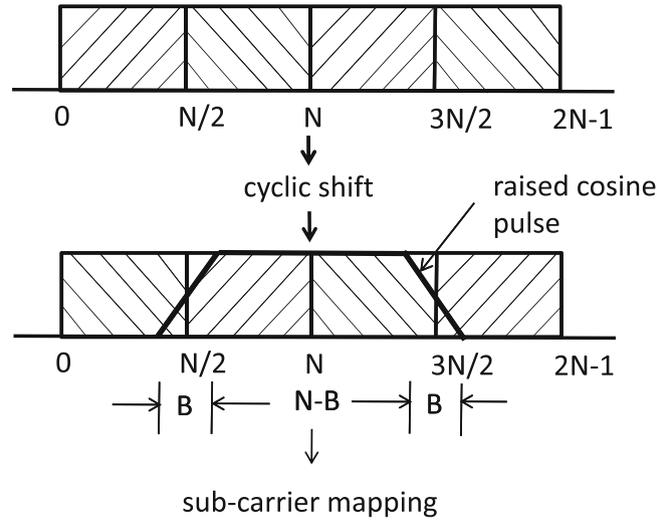


Fig. 10.25 Frequency-domain root-raised cosine filtering

$$x_{n,\ell}^u = \begin{cases} x_{n,\ell/F}, & \ell = k \cdot F, \quad k \text{ an integer} \\ 0, & \text{otherwise} \end{cases} \quad (10.98)$$

Rather than implementing the upsampling process in the time domain, the upsampling process can be emulated in the discrete frequency-domain by generating periodic replicas of the vector \mathbf{X}_n , i.e.,

$$X_{n,\ell N+m}^u = X_{n,m}, \quad \ell = 0, \dots, F-1, \quad 0 \leq m \leq N-1. \quad (10.99)$$

Once again, if the roll-off factor β lies in the range $0 \leq \beta \leq 1$, then an upsampling factor $F = 2$ will suffice, and this value is assumed here. A cyclic shift of $N/2$ positions is then applied to the vector \mathbf{X}_n^u and, afterwards, apply frequency-domain raised cosine filtering to generate the vector \mathbf{X}_n^{sh} as shown in Fig. 10.25. After filtering, the sub-carrier mapping is applied.

The required frequency-domain pulse description can be obtained by modifying the pulse description in (10.76), where the subscript k now refers to the sub-carrier frequency. The pulse defined in (10.76) begins at index $k = 0$ and ends at $k = M + G - 1$. However, the block size should be set to N rather than M , and the guard interval in this case serves only to contain the required bandwidth expansion, which is equal to B . Hence, setting $M = N$ and $G = B$ in (10.76) gives

$$W_k^{rc} = \begin{cases} \sqrt{\frac{1}{2} + \frac{1}{2} \sin\left(\frac{\pi}{B}\left(k - \frac{B-1}{2}\right)\right)}, & 0 \leq k < B \\ 1, & B \leq k < N \\ \sqrt{\frac{1}{2} - \frac{1}{2} \sin\left(\frac{\pi}{B}\left(k - N + \frac{B+1}{2}\right)\right)}, & N \leq k \leq N + B - 1 \\ 0, & \text{elsewhere} \end{cases} \quad (10.100)$$

The pulse in (10.100) has a flat top of length $N - B$, and two roll-off sections of length B . A relationship is needed between the parameters B and N , and the roll-off factor β used in the continuous-time root-raised cosine pulse description. The required values can be obtained by noting that $B = \beta/\tilde{T}_s^g$ and that the flat top portion has width $N - B = (1 - \beta)/\tilde{T}_s^g$ and, hence, B is chosen according to $\beta = B/N$. The required frequency domain description can be obtained by right-shifting the pulse in (10.100) by $N - (N + B)/2$ positions.

Note that the overall length of the frequency-domain root-raised cosine pulse is $N + B$. Hence, with root-raised cosine filtering there are $N + B$ samples that are applied to the sub-carrier mapping, as opposed to just N samples without frequency domain filtering. The extra B frequency-domain samples per user is the bandwidth cost that is required to implement the root-raised cosine filtering.

Problems

10.1. Consider an OFDM system having $N = 1024$ sub-carriers that are spaced 4 kHz apart. Suppose that the waveform channel is a linear time-invariant channel consisting of two non-faded equal rays spaced τ_d seconds apart, i.e.,

$$g(t, \tau) = g\delta(\tau) + g\delta(\tau_d),$$

where $g = \alpha e^{j\phi}$ is the gain associated with each of the two channel taps.

Find the value of the delay τ_d in the above channel model that will yield the worst possible bit error rate performance for this OFDM system.

10.2. Suppose that an OFDM system is implemented with a guard interval that is a cyclic extension of the IDFT coefficients as shown in (4.95).

(a) Show that the output of the OFDM demodulator is given by (10.6).

(b) Now suppose that the guard interval simply consists of a blank interval where nothing at all is transmitted. Assuming that $G \geq L > 0$ can the data block \mathbf{x}_n be recovered by taking a DFT of the received block $\mathbf{R}_n = \{R_{n,m}\}_{m=0}^{N-1}$?

10.3. Assume the transmission of a known OFDM symbol, $\mathbf{c}_n = (c_{n,0}, \dots, c_{n,N-1})$, given by the chirp sequence

$$c_{n,k} = \exp \left\{ j \frac{2\pi}{N} k^2 \right\}, \quad 0 \leq k \leq N-1.$$

Coarse channel estimates can be obtained by computing

$$\begin{aligned} \tilde{T}_{n,i} &= \frac{z_{n,i}}{c_{n,i}} \\ &= T_{n,i} + \tilde{n}_{n,i}, \quad i = 0, \dots, N-1, \end{aligned}$$

where $\tilde{\mathbf{n}}_n = \{\tilde{n}_{n,i}\}_{i=0}^{N-1}$ is a vector of independent and identically distributed complex Gaussian random variables with zero mean and variance σ_n^2 . This yields the channel estimate $\tilde{\mathbf{T}}_n = \{\tilde{T}_{n,i}\}_{i=0}^{N-1}$. The receiver then computes the vector $\mathbf{v}_n = (v_{n,1}, \dots, v_{n,N})$, where $\mathbf{v}_n = \text{IDFT}[\tilde{\mathbf{T}}_n]$. Afterwards, the $\{v_{n,m}\}_{m=0}^{N-1}$ are passed through a rectangular window to zero $v_{n,m}$, $m \geq G+1$, and a DFT is taken of the resulting sequence to yield the refined channel estimate $\bar{\mathbf{T}}_n = \{\bar{T}_{n,i}\}_{i=0}^{N-1}$, such that

$$\bar{T}_{n,i} = T_{n,i} + \bar{n}_{n,i}, \quad i = 0, \dots, N-1.$$

Show that the noise components $\bar{n}_{n,i}$ are zero mean and have a variance equal to $\sigma_n^2 \frac{G+1}{N}$.

10.4. Derive the MMSE-FDE solution in (10.65).

10.5. Derive (10.92) for localized FDMA.

Chapter 11

Frequency Planning Techniques

This chapter considers cellular frequency planning techniques for TDMA and OFDMA/SC-FDMA based cellular systems. System concepts for CDMA cellular systems will be covered in Chap. 12. However, some schemes for high capacity, such as cell sectoring, are flexible enough to be applied to any standard. Regardless of the chosen access method, the ultimate goal is to achieve high capacity while satisfying quality of Service (QoS) requirements. An architecture must also be implemented that will easily accommodate system growth.

Microcells are a straightforward solution for achieving high capacity. However, as microcells are introduced, a mixed cell architectures naturally evolves, consisting of overlaid macrocells and underlaid microcells. Such an arrangement is called a *hierarchical architecture*. Hierarchical architectures can be implemented for TDMA, CDMA, and OFDMA/SC-FDMA cellular 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 bandwidth among the hierarchical layers. If the entire bandwidth is used in each hierarchical layer, then both high capacity and high flexibility will be achieved. CDMA systems employ universal frequency reuse, but require sophisticated power control algorithms if the bandwidth is not partitioned between the hierarchical layers.

Macrodiversity architectures are another method for achieving high capacity, where the signal transmitted by a mobile station (MS) is received by multiple base stations (BSs). Likewise, the signal that is received by a MS may be transmitted by multiple BSs. Macrodiversity is an effective method for combatting shadow and envelope fading. In fact, cellular handoff algorithms implement a form of macrodiversity. The soft handoff techniques used in CDMA systems are a well-known method for realizing macrodiversity. TDMA systems that use hard handoff algorithms will not yield as much macrodiversity gain due to latencies in the hard handoff algorithms. The requirement for hard handoff in TDMA systems arises a result of frequency planning with fixed channel assignment (FCA). However, if dynamic channel assignment (DCA) techniques are used, then TDMA systems can realize benefits from soft handoff similar to those obtained in CDMA systems. DCA techniques do not permanently assign channels to cells.

The remainder of this chapter begins with a discussion of basic cellular frequency planning techniques, including cell sectoring in Sect. 11.1 and cell splitting and reuse partitioning in Sect. 11.2. Afterwards, Sect. 11.3 considers issues related to frequency planning in OFDMA/SC-FDMA cellular networks. Section 11.4 considers a novel TDMA hierarchical cellular architecture based on the concept of cluster planning, where macrocells and microcells can share the same frequencies. Finally, Sect. 11.5 considers macrodiversity TDMA architectures, where a mobile station is simultaneously served by multiple base stations.

11.1 Cell Sectoring

11.1.1 Cell Sectoring with Wide-Beam Directional Antennas

One of the simplest and most practical methods for controlling co-channel interference (CCI) is to use wide-beam directional antennas at the BSs. On the forward channel, wide-beam directional antennas reduce the generation of CCI by transmitting the signals to the MSs with a narrower angle-of-departure (AoD) spread than omnidirectional antennas. On the reverse channel, wide-beam directional antennas reduce the effect of the CCI because they respond to CCI that is received within a narrower angle-of-arrival (AoA) spread about the target MS than would be the case with omnidirectional antennas.

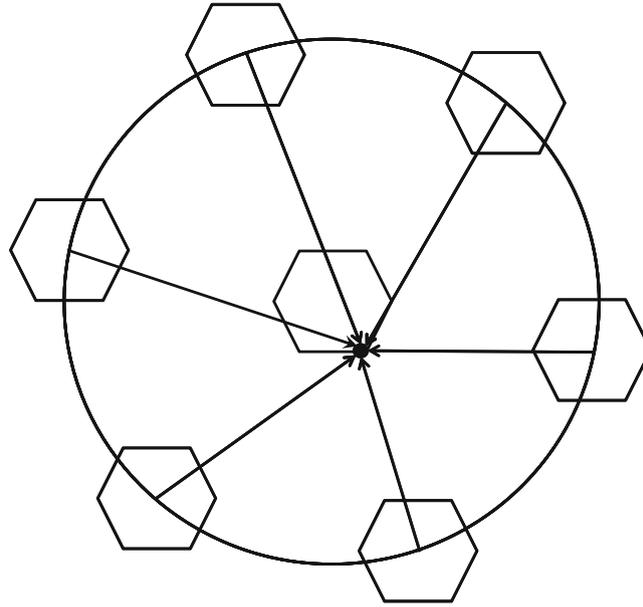


Fig. 11.1 Worst case CCI situation on the forward channel

Consider a uniform deployment of hexagonal cells, where the BSs employ omnidirectional antennas. Suppose the effects of shadowing and multipath-fading are ignored, and assume the following simple path loss model adapted from (2.386), such that the received area mean power at distance d is

$$\mu_{\Omega_p} = \Omega_i G_T G_R \frac{(h_b h_m)^2}{d^\beta}, \quad (11.1)$$

where Ω_i 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. 11.1, the worst case forward channel CCI situation occurs when the MS is located at the corner of a cell, furthest from its serving BS. For $N = 7$, there are six first-tier co-channel BSs, located at distances $\{\sqrt{13}R, 4R, \sqrt{19}R, 5R, \sqrt{28}R, \sqrt{31}R\}$ from the MS. If the values of Ω_i , G_T , and h_b are assumed to be the same for all BS antennas, then it follows that the worst case carrier-to-interference ratio, Λ , is

$$\begin{aligned} \Lambda &= \frac{R^{-\beta}}{(\sqrt{13}R)^{-\beta} + (4R)^{-\beta} + (\sqrt{19}R)^{-\beta} + (5R)^{-\beta} + (\sqrt{28}R)^{-\beta} + (\sqrt{31}R)^{-\beta}} \\ &= \frac{1}{(\sqrt{13})^{-\beta} + (4)^{-\beta} + (\sqrt{19})^{-\beta} + (5)^{-\beta} + (\sqrt{28})^{-\beta} + (\sqrt{31})^{-\beta}}. \end{aligned} \quad (11.2)$$

A similar expression can be derived for the case of $N = 3$ with co-channel distances $\{2R, \sqrt{7}R, \sqrt{7}R, \sqrt{13}R, \sqrt{13}R, 4R\}$, and $N = 4$ with co-channel distances $\{\sqrt{7}R, \sqrt{7}R, \sqrt{13}R, \sqrt{13}R, \sqrt{19}R, \sqrt{19}R\}$. With a path loss exponent $\beta = 3.5$, the worst case Λ is

$$\Lambda_{(\text{dB})} = \begin{cases} 14.56 \text{ dB} & \text{for } N = 7 \\ 9.98 \text{ dB} & \text{for } N = 4 \\ 7.33 \text{ dB} & \text{for } N = 3 \end{cases}.$$

The minimum allowable cluster size is determined by the threshold carrier-to-interference ratio requirement, Λ_{th} , of the radio receiver. Unfortunately, the above worst case Λ values may be too small to yield acceptable performance, especially when shadowing and multipath-fading are accounted for.

Sectoring is a very common method that is employed in cellular systems to improve the worst case Λ , 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

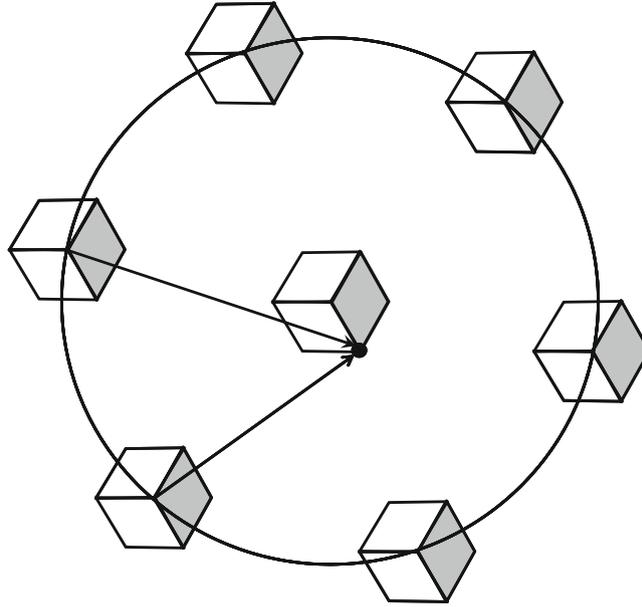


Fig. 11.2 Worst case CCI situation on the forward channel with 120° cell sectoring

$3N$ sectors). As shown in Fig. 11.2, 120° cell sectoring reduces the number of first-tier co-channel interferers from six to two. For $N = 7$, the two first tier interferers are located at distances $\sqrt{19}R$, $\sqrt{28}R$ from the MS. The resulting worst case carrier-to-interference ratio, Λ , is

$$\begin{aligned}\Lambda &= \frac{R^{-\beta}}{(\sqrt{19}R)^{-\beta} + (\sqrt{28}R)^{-\beta}} \\ &= \frac{1}{(\sqrt{19})^{-\beta} + (\sqrt{28})^{-\beta}}.\end{aligned}\quad (11.3)$$

A similar expression can be derived for the case of $N = 3$ with co-channel distances $\{\sqrt{7}R, \sqrt{13}R\}$, and $N = 4$ with co-channel distances $\{\sqrt{13}R, \sqrt{19}R\}$. Hence,

$$\Lambda_{(\text{dB})} = \begin{cases} 20.60 \text{ dB} & \text{for } N = 7 \\ 17.69 \text{ dB} & \text{for } N = 4 \\ 13.52 \text{ dB} & \text{for } N = 3 \end{cases}.$$

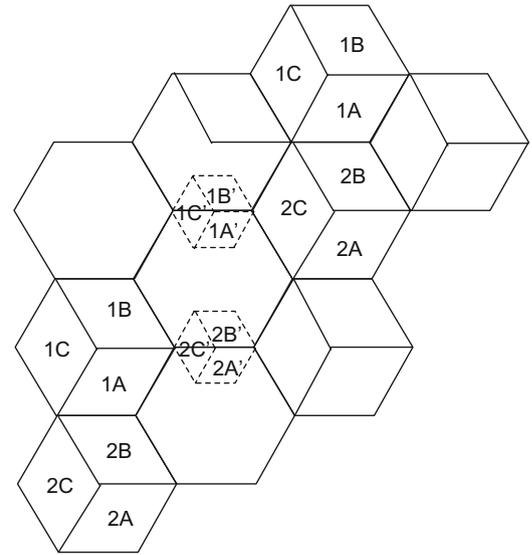
For $N = 7$, 120° sectoring yields a 6.04 dB gain in the worst case Λ over the case when omnidirectional antennas are used.

To derive a benefit from sectoring, the channel set that is assigned to each cell must be further partitioned into disjoint sets, such that each sector uses a disjoint set of channels. This finer partitioning of the channel sets results in a loss in trunking efficiency, as discussed in Sect. 1.7.3. Hence, cell sectoring improves the worst case carrier-to-interference ratio performance at the cost of reduced trunking efficiency.

11.2 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. 11.3. 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

Fig. 11.3 Conventional cell splitting is used to accommodate an increased traffic load by introducing smaller cells



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 labeled **2'** can be located at the midpoint between the two cells labeled **2**.

Because the split cells are smaller, the transmit power can be reduced. To estimate the transmit power requirements in the split cells, assume the simple path loss model in (1.6). Then the received power for a MS located at the corner of a parent cell is

$$\Omega(R_o) = k\Omega_o R_o^{-\beta}, \quad (11.4)$$

while the received power at the boundary of a split cell is

$$\Omega(R_s) = k\Omega_s R_s^{-\beta}. \quad (11.5)$$

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, and k is a constant. To keep the received signal power that is associated with a MS located on the cell boundary at a constant value, the transmit power requirements of the split cell and parent cell are related as follows:

$$\Omega_s = \Omega_o \left(\frac{R_s}{R_o} \right)^{-\beta}. \quad (11.6)$$

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 straightforward 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 cell splitting occurs throughout the system area, requiring the installation of a large number of additional cell sites. Therefore, channel segmenting may not always be a good option.

Another solution is shown in Fig. 11.4, 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.

Fig. 11.4 Overlaid inner cells can be used to maintain the frequency reuse constraint when cell splitting is used

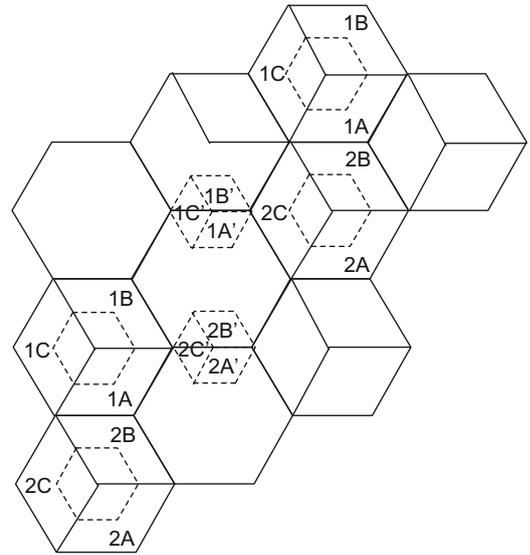
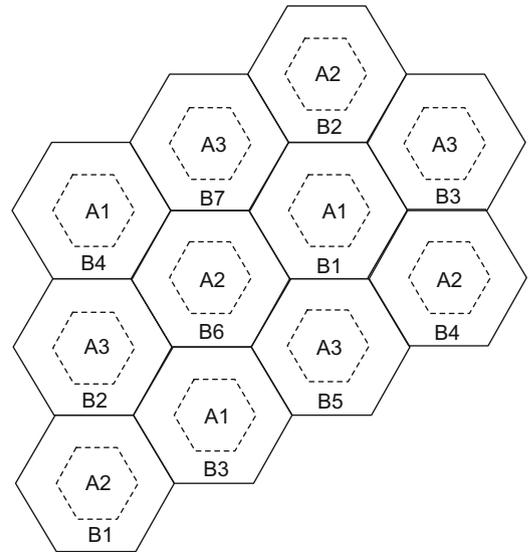


Fig. 11.5 Reuse partitioning can be used to increase the channel reuse efficiency, from [155]



11.2.1 Reuse Partitioning

Halpern [155] suggested an overlay/underlay scheme based on the concept of reuse partitioning, where multiple co-channel reuse factors are used in the same deployment. Sometimes this is called fractional reuse. An inner cell is created within each of the existing cells as shown in Fig. 11.5. For the example in Fig. 11.5, 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 an inner and outer cell. The reduced radii of the inner cells lead to an increase in cell capacity. To quantify this increase let

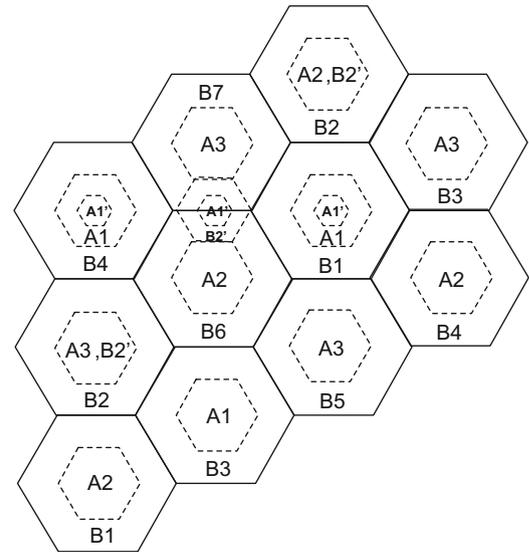
R_i = radius of the inner cells.

R_o = radius of the outer cells.

D_i = reuse distance for the inner cells.

D_o = reuse distance for the outer cells.

Fig. 11.6 Cell splitting can be used in combination with reuse partitioning



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

$$\frac{D_i/R_i}{D_i/R_o} = \frac{4.6}{3}. \quad (11.7)$$

Hence, the inner and outer cell radii are related by $R_i = 0.65R_o$ and, therefore, the inner and outer cell areas are related by $A_i = (0.65)^2 A_o = 0.43A_o$. If a total of N_T channels are available, then $0.43N_T$ channels should be assigned to the inner cells and $0.57N_T$ channels assigned to the outer cell area (assuming a homogenous traffic distribution throughout the system area). The resulting cell capacity is

$$N_\mu = 0.57N_T/7 + 0.43N_T/3 = 0.225N_T \text{ channels/cell}. \quad (11.8)$$

On the other hand, with a conventional 7-cell reuse plan

$$N_\mu = N_T/7 \text{ channels/cell}. \quad (11.9)$$

Hence, an improvement of 1.575 in cell capacity is realized.

11.2.1.1 Cell Splitting with Reuse Partitioning

Cell splitting can also be used with reuse partitioning. An example is shown in Fig. 11.6 where a split cell is added between the parent cells labeled **B2**. 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, there is a drawback when cell splitting is used with the reuse partitioning, in that 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 such rings.

11.3 OFDMA/SC-FDMA Radio Planning

Radio planning for OFDMA/SC-FDMA networks is more akin to radio planning in TDMA cellular networks, such as GSM, than radio planning in CDMA networks, such as WCDMA and cdma2000. The reason is that the intra-cell interference is essentially eliminated due to the orthogonal property of the OFDMA sub-carriers. However, to achieve high capacity with

Fig. 11.7 Deployment with a sector reuse factor of 1

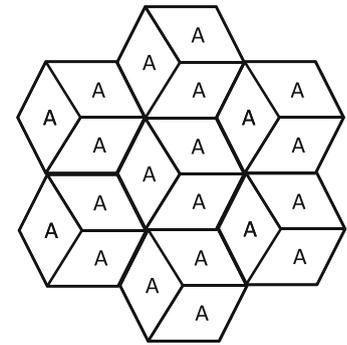
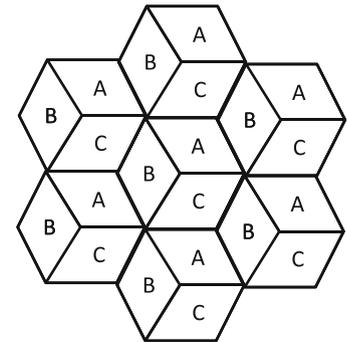


Fig. 11.8 Deployment with a sector reuse factor of 3



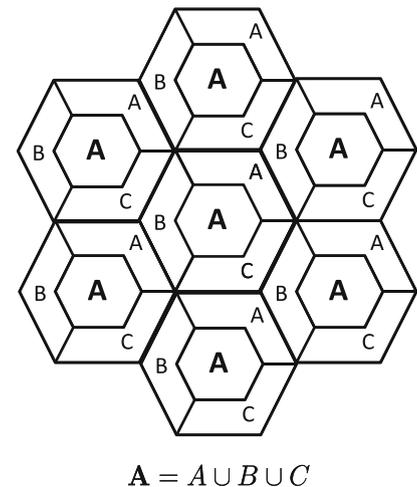
OFDMA/SC-FDMA networks, aggressive frequency planning is necessary, and this will tend to increase the level of inter-cell interference. For OFDMA/SC-FDMA networks adjacent channel interference is controlled by the orthogonal nature of the sub-carriers. However, co-channel interference will be present. The reuse partitioning schemes discussed in Sect. 11.2.1 can be used for deployment, along with the cell spitting techniques in Sect. 11.2.1.1. The frequency reuse configurations that are of most interest for OFDMA/SC-FDMA networks are BSs with 120° sectoring, and a sector reuse of 1 or 3.

With a sector reuse of 1, also known as universal frequency reuse, all sectors can use the same sub-carrier frequencies, as shown in Fig. 11.7. With universal frequency reuse, two different schemes can be used to mitigate the co-channel interference between the cells and along the sector boundaries, namely interference avoidance and interference randomization. Interference avoidance assigns sets of sub-carriers dynamically so as to avoid interference. Randomization can be achieved by scrambling the set of sub-carriers that are used within the same band, or by frequency hopping over a larger band. In either case, the interference experienced by the users is randomized and a strong interferer is shared by all the users.

With a sector reuse of 3, the three reuse sectors in each cell are assigned a disjoint set of sub-carrier frequencies, as shown in Fig. 11.8. Reuse 3 will eliminate the co-channel interference that occurs along the sector boundaries in the same cell. Furthermore, it also increases the co-channel reuse distance. This will permit greater usage of the sub-carrier frequencies in each sector which will improve spectral efficiency. However, a sector reuse of 3 does require that the set of frequencies be partitioned into three groups which will tend to decrease spectral efficiency.

A third alternative is to employ the concept of reuse partitioning as detailed in Sect. 11.2.1. With reuse partitioning a cell is divided into multiple regions, for example, an inner cell and three outer cell sectors as shown in Fig. 11.9. Users that are located in the inner cells can reuse all the available sub-carrier frequencies. Users that are located in the outer cell areas use a fraction of the available sub-carriers according to a 3 sector reuse. This arrangement will yield high spectral efficiency by using all the available sub-carrier frequencies in the inner cells, while at the same time provide robustness against interference by using a 3 sector reuse of sub-carrier frequencies in the outer cell areas. The spectral efficiency will be somewhere between that of universal frequency reuse and 3 sector reuse. The scheme also requires that the locations of the users be tracked and the set of sub-carriers allocated to the users be dynamically updated according to a policy that promotes uniform sub-carrier-to-interference-plus-noise ratio (sub-carrier CIR) while maximizing a metric such as the sum capacity. For cellular downlink applications, power and sub-carrier allocation schemes can be used under the assumption that the BS has complete channel state information for all forward channels it serves, and the total BS transmit power is constrained. The uplink is similar, but more complicated in the sense that each MS may have a different transmit power constraint.

Fig. 11.9 Deployment with reuse partitioning



A variety of techniques described previously for single-carrier systems can be applied with simple modification to multi-user multi-carrier systems. On the forward channel beam-forming can be used to promote good spatial efficiency by minimizing BS emissions. For line-of-sight (LoS) channels, classical beam-forming is described in Sect. 6.9. In this case, it is possible that the same antenna weighting vector could be used for all sub-carriers. On the reverse channel, multi-antenna interference cancelation techniques can be performed, such as optimum combining as described in Sect. 6.8. Since the channel varies with each uplink user and each sub-carrier, optimum combining would have to be employed on a per user per sub-carrier basis.

11.4 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 necessarily correspond to the locations that are experiencing the highest traffic growth. A TDMA hierarchical architecture is now described based on the concept of 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.

11.4.1 System Architecture

A traditional 7/21 frequency reuse system is shown in Fig. 11.10. 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:

1. Assign the same channels to each cell site as in the traditional 7/21 frequency reuse plan shown in Fig. 11.10.
2. Divide the macrocell reuse clusters into three groups as shown in Fig. 11.11.
3. Let the first group be the reference group.
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.

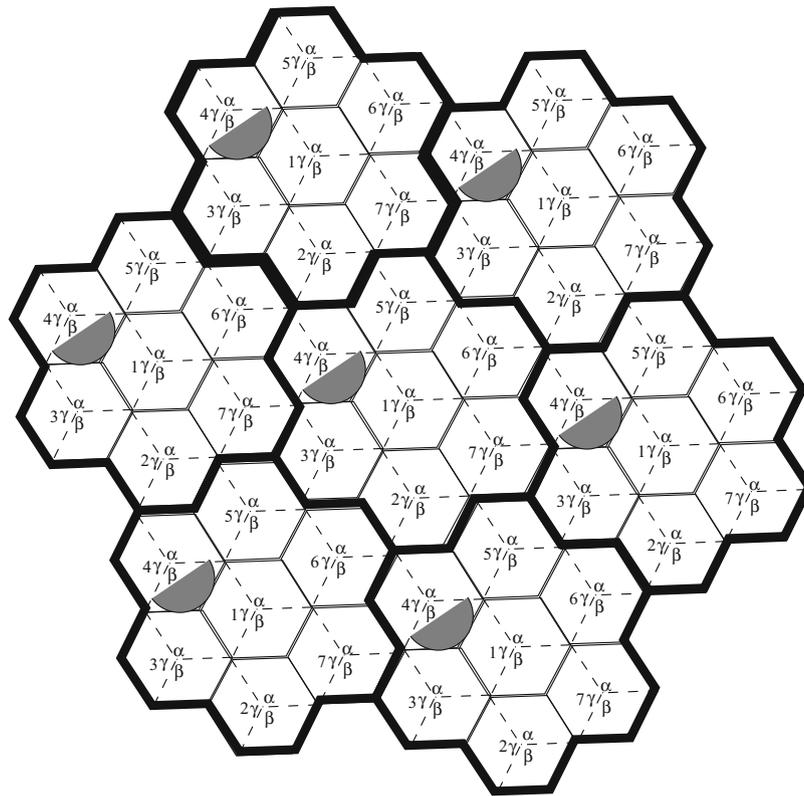


Fig. 11.10 Traditional 7/21 frequency reuse plan

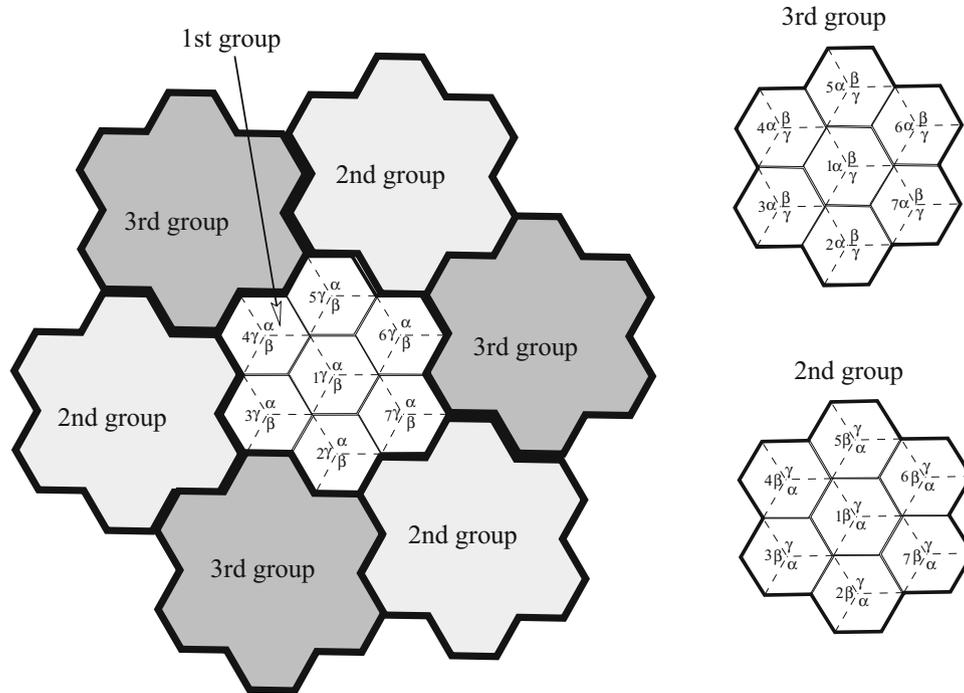


Fig. 11.11 Proposed 7/21 frequency reuse plan with cluster planning

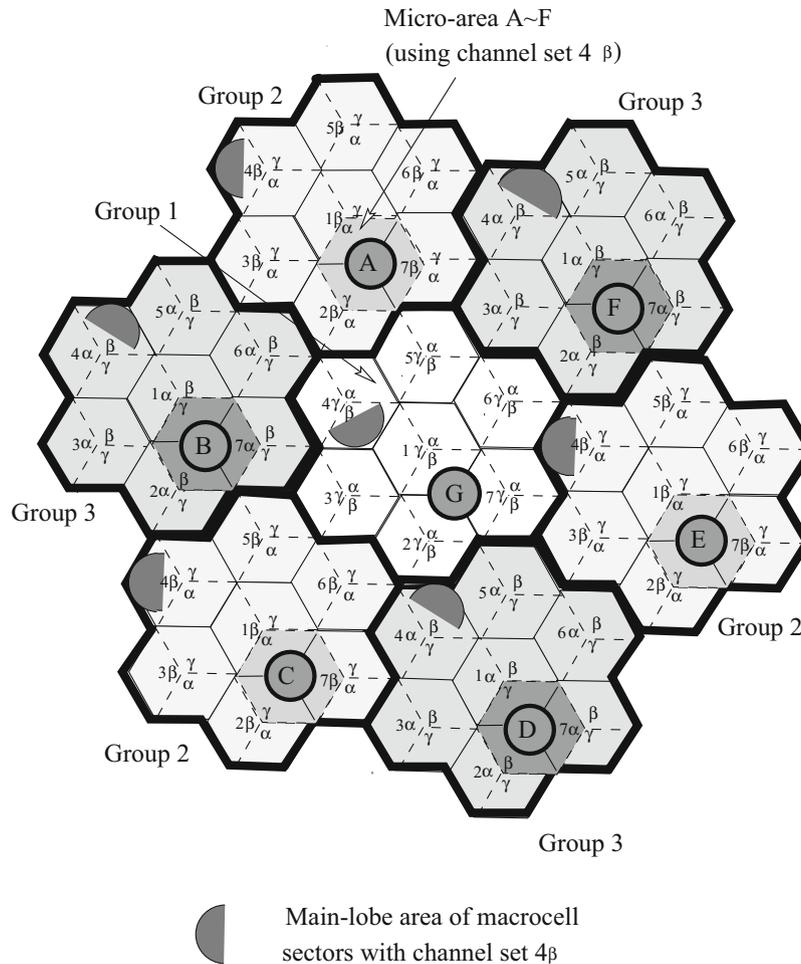


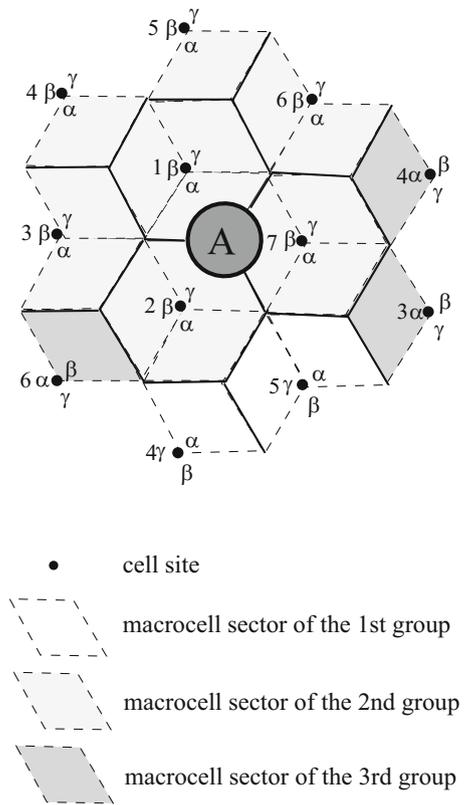
Fig. 11.12 Microcells can reuse low-interference macrocell channels in the proposed hierarchical architecture. The macrocell channel set 4β can be reused in the micro-area A ~ F

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. Figure 11.12 shows the result of rotating the sectors. Zones A ~ F show a very low-interference for channel set 4β , since they are located in the back-lobe areas of the macrocell sectors using channel set 4β . Thus microcells can be introduced in these areas by using channel set 4β .

11.4.2 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 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. 11.10), the back-lobe area of each channel set will still encounter some first-tier interferers. To protect the back-lobe areas from the first tier interferers, the sectors are rotated through the cluster planning procedure. Cluster planning creates low-interference micro-areas as shown in Fig. 11.12, 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. Figure 11.13 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.

Fig. 11.13 Interference neighborhood for micro-area A in Fig. 11.12



The following algorithm systematically determines the channels that can be used in each micro-area. To describe the algorithm, let c_i^j represent the channel set in sector $i = \alpha, \beta, \gamma$, of the cell site c , where $c = 1, \dots, 7$. 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

$$\Theta = \{c_i^j \in M\}$$

denote the union of channel sets c_i^j in the interference neighborhood M .

- From Θ , construct a 3×3 “indicator matrix” $\mathbf{B}_c = [b_{ij}]$ for BSs $c = 1, \dots, 7$, 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 11.1: According to Fig. 11.13, the interference neighborhood for micro-area A is

$$\Theta = \{1_\alpha^2, 1_\beta^2, 1_\gamma^2, 2_\alpha^2, 2_\beta^2, 2_\gamma^2, 3_\alpha^2, 3_\alpha^3, 3_\gamma^2, 4_\alpha^1, 4_\alpha^2, 4_\alpha^3, 5_\alpha^1, 5_\alpha^2, 5_\gamma^1, 6_\alpha^2, 6_\beta^2, 6_\beta^3, 7_\alpha^2, 7_\beta^2, 7_\gamma^2\}.$$

Then the indicating matrices are

(continued)

Example 11.1 (continued)

$$\mathbf{B}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}; \mathbf{B}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}; \mathbf{B}_3 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\mathbf{B}_4 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \mathbf{B}_5 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}; \mathbf{B}_6 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{B}_7 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Examining the indicator matrices $\mathbf{B}_c, c = 1, \dots, 7$, observe 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. 11.14 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 11.1. Note that the micro-areas are capable of reusing two macrocell channel sets and microcells can be deployed throughout the whole service area.

Since a micro-area consists of three macrocell sectors, each macrocell area has five available channel sets; three are assigned to macrocells and two are 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 sectorized 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, it will be shown that $C_\mu = 6$ is possible and, hence, giving a capacity increase of seven times.

11.4.3 Performance Analysis of Cluster-Planned Architecture

11.4.3.1 Propagation Model and System Assumptions

Our analysis uses the following path loss model from (2.386):

$$\mu_{\Omega_r} = \frac{\Omega_t (h_b h_m)^2}{d^\beta}, \quad (11.10)$$

where μ_{Ω_r} 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 (11.10) is more suitable for a macrocell 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 omnidirectional antennas. It is possible to improve the C/I performance by sectoring the microcells as well, but this is not considered here. The MSs use omnidirectional antennas.

Uplink power control: An uplink power control scheme is adopted such that the transmitted power is adjusted in six levels from -22 to -2 dBW in steps of 4 dB. Downlink power control is not required in the proposed architecture.

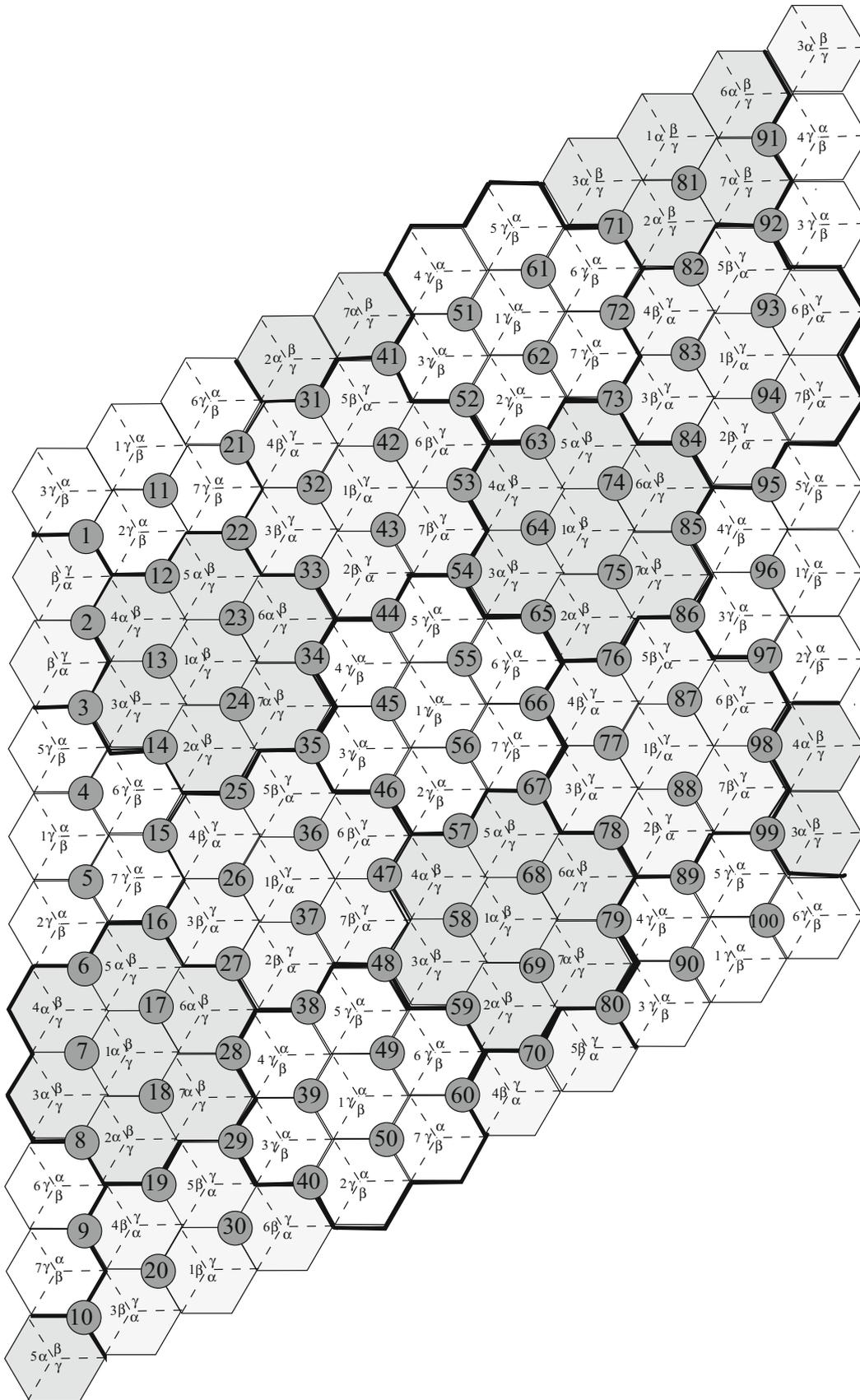


Fig. 11.14 Channel reuse in the proposed 3-sector cellular system, where each micro-area consists of three sectors that belong to three different macrocells

Table 11.1 The macrocell channel sets that can be used in the underlaid microcells

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	$2_\beta, 2_\gamma$

Before proceeding, the notation is first clarified. 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.

11.4.3.2 Macrocell Performance

Section 11.4.2 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. 11.10 and the proposed hierarchical cellular system in Fig. 11.12 without the underlaid microcells. Figure 11.15 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. The uplink case is considered 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, the effect of underlaid microcells will be further included when analyzing the performance of the proposed hierarchical cellular system. For ease of analysis, the worst case scenario is hereafter adopted, where a MS is situated on a cell boundary.

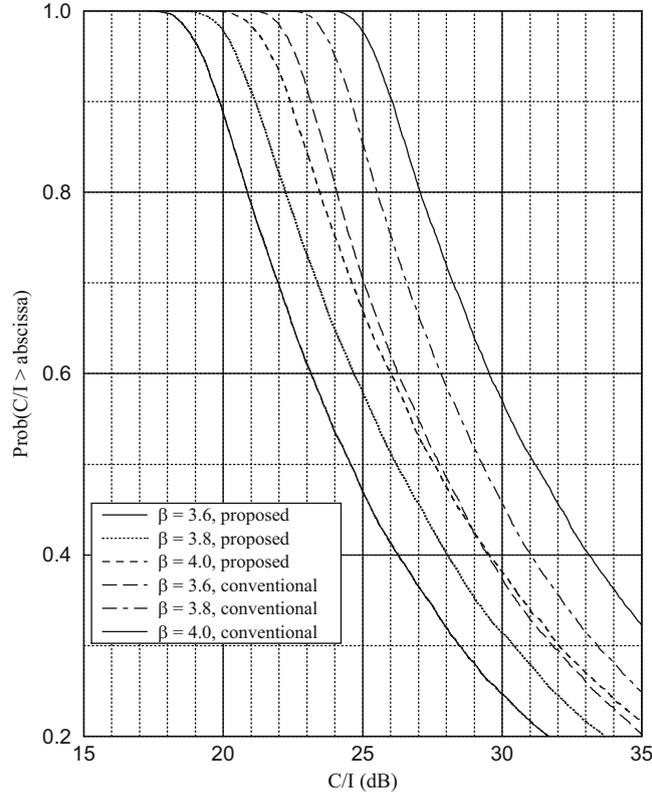


Fig. 11.15 Comparison of the uplink C/I performance of conventional macrocells and the proposed hierarchical cellular system without the underlaid microcells for different path loss exponent β

Downlink C/I Analysis

By applying (11.10) with $\beta = 4$, the C/I received by the MS at the macrocell boundary can be expressed as

$$\frac{C_M^d}{I_M^d + J_{\mu M}^d} = \frac{\Omega_{t,b}^M \frac{(h_b^M h_m)^2}{R_M^4}}{\sum_{i=1}^{N_M} \frac{\Omega_{t,b}^M (h_b^M h_m)^2}{D_i^4} + \sum_{j=1}^{Z_\mu} \sum_{k=1}^{C_\mu} \frac{\Omega_{t,b}^\mu (h_b^\mu h_m)^2}{d_{jk}^4}} \quad (11.11)$$

where

C_M^d = MS received power from the desired macrocell BS

I_M^d = downlink macrocell-to-macrocell CCI

$J_{\mu M}^d$ = downlink microcell-to-macrocell CCI

$\Omega_{t,b}^M$ = macrocell BS transmitted power

$\Omega_{t,b}^\mu$ = microcell BS transmitted power

N_M = the number of macrocell co-channel interferers

Z_μ = the number of interfering micro-areas

C_μ = the number of microcell clusters in a micro-area

- D_i = MS distance to the i th interfering macrocell BS
 $d_{j,k}$ = MS distance to the k th interfering microcell BS in the j th micro-area
 h_b^M = macrocell BS antenna height
 h_b^μ = microcell BS antenna height
 h_m = MS antenna height
 R_M = macrocell radius

Referring to Fig. 11.14 and Table 11.1, the downlink interference is examined 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, the three second-tier interfering BSs are also considered, 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 $[\bar{d}_1, \bar{d}_2, \bar{d}_3, \bar{d}_4, \bar{d}_5, \bar{d}_6] = [3, 4.58, 3.46, 6, 5.2, 6.25]R_M$, respectively. The second-tier interfering micro-areas 3, 29, 41, and 92 have distances $[\bar{d}_7, \bar{d}_8, \bar{d}_9, \bar{d}_{10}] = [7.55, 9, 7.94, 12]R_M$, respectively. Each micro-area is assumed to have C_μ microcell reuse clusters, and each of these clusters has K_μ microcells. Through the channel selection algorithm in Sect. 11.4.2, each micro-area is assigned two macrocell channel sets. These two sets of channels are further partitioned into K_μ groups and each group is then assigned 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, it is assumed that the distance \bar{d}_j approximates $d_{j,k}$, where \bar{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 (11.11). 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 above assumptions in (11.11)

$$\frac{C_M^d}{I_M^d + J_{\mu M}^d} = \frac{1}{1.02875 \times 10^{-2} + C_\mu \left(\frac{\Omega_{t,b}^\mu}{\Omega_{t,b}^M} \right) \times 2.79449 \times 10^{-3}}. \quad (11.12)$$

The downlink C/I performance is shown in terms of C_μ and $\Omega_{t,b}^\mu/\Omega_{t,b}^M$ in Fig. 11.16 with consideration of only first-tier interfering BSs and in Table 11.2 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 11.2 with Fig. 11.16, 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.

Uplink CCI Analysis

By modifying (11.11) slightly, the uplink C/I can be formulated as

$$\frac{C_M^u}{I_M^u + J_{\mu M}^u} = \frac{\Omega_{t,m}^M (h_b^M h_m)^2}{R_M^4} \frac{1}{\sum_{i=1}^{N_M} \frac{\Omega_{t,m}^M (h_b^M h_m)^2}{D_i^4} + \sum_{j=1}^{Z_\mu} \sum_{k=1}^{C_\mu} \frac{\Omega_{t,m}^\mu (h_b^M h_m)^2}{d_{jk}^4}} \quad (11.13)$$

where

- C_M^u = macrocell BS received power from the desired MS
 I_M^u = uplink macrocell-to-macrocell interference

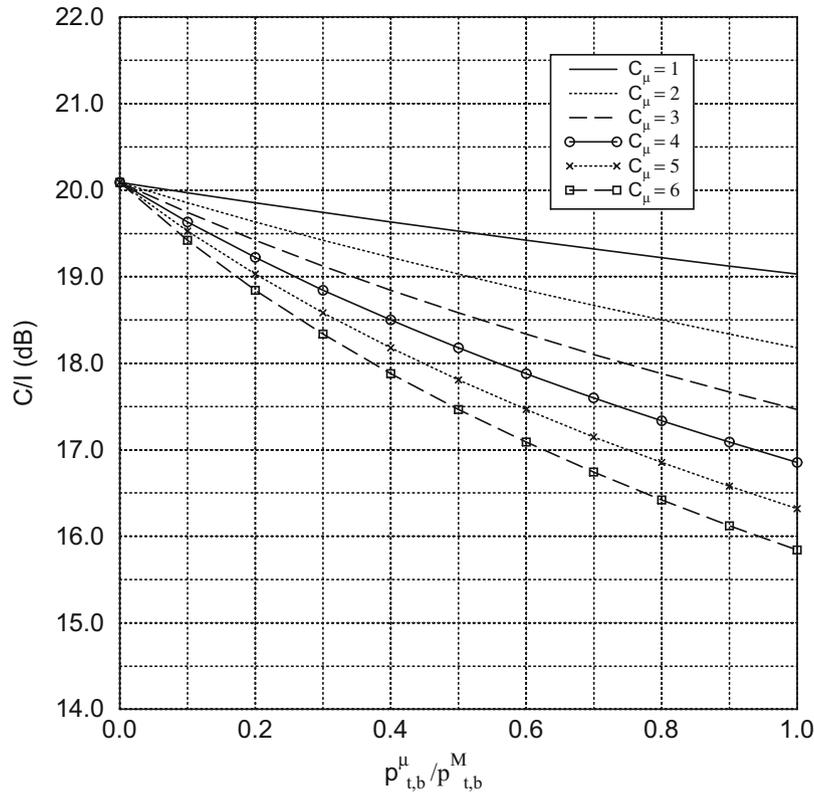


Fig. 11.16 Macrocell downlink C/I performance against $\Omega_{t,b}^{\mu}/\Omega_{t,b}^M$ for different values of C_{μ} , where $\Omega_{t,b}^{\mu}/\Omega_{t,b}^M$ is the microcell to macrocell BS transmit power, and C_{μ} is the number of microcell clusters in a micro-area

Table 11.2 Downlink C/I performance for overlaying macrocells, where $h_b^{\mu}/h_b^M = 1/3$

$\Omega_{t,b}^{\mu}/\Omega_{t,b}^M$	$C_M^d/(J_M^d + J_{\mu M}^d)$ (dB)			
	$C_{\mu} = 1$	$C_{\mu} = 2$	$C_{\mu} = 4$	$C_{\mu} = 6$
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

$$J_{\mu M}^{\mu} = \text{uplink microcell-to-macrocell interference}$$

$$\Omega_{t,m}^M = \text{macrocell MS transmitted power}$$

$$\Omega_{t,m}^{\mu} = \text{microcell MS transmitted power}$$

and where the remaining parameters have already been defined following (11.11). 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

$$[d_1, d_2, d_3, d_4, d_5, d_6] = [7.0, 7.0, 14.7, 5.3, 11.5, 9.53]R_m.$$

The effect of the three other interfering micro-areas 4, 17, 49 can be ignored because they are located in the back-lobe area of the sector using channel set 2γ . By substituting the above values into (11.13), the uplink C/I performance for this example becomes

$$\frac{C_M^u}{I_M^u + J_{\mu M}^u} = \frac{1}{1.2677 \times 10^{-2} + C_\mu \left(\frac{\Omega_{t,m}^\mu}{\Omega_{t,m}^M} \right) \times 2.11 \times 10^{-3}}. \quad (11.14)$$

Figure 11.17 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

$$\frac{\Omega_{t,m}^\mu}{\Omega_{t,m}^M} \leq 0.2. \quad (11.15)$$

Note that (11.15) was obtained under the assumption that the interfering macrocell MSs are on the cell boundary and are transmitting with the maximum power. Thus (11.15) can be used to determine the maximum microcell MS's transmitted power. For example, consider a MS that can adjust its transmitted power in six levels from -22 to -2 dBW. Then (11.15) implies that the maximum microcell MS transmitted power is -9 dBW, which is still in the operational range of the MS.

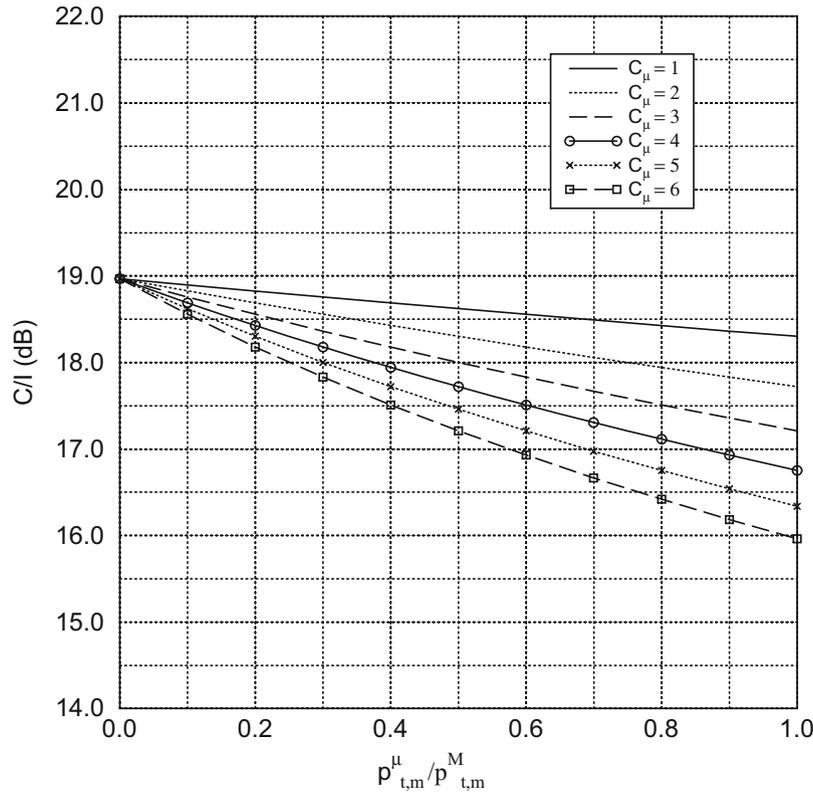


Fig. 11.17 Macrocell uplink C/I performance against $\Omega_{t,m}^\mu/\Omega_{t,m}^M$ for different values of C_μ , where $\Omega_{t,m}^\mu/\Omega_{t,m}^M$ is the ratio of the transmitted power of the microcell MS to that of the macrocell MS, and C_μ is the number of microcell clusters in a micro-area

11.4.3.3 Microcell Performance

It is now shown 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: (1) C -criterion: a MS will receive stronger power, C , at the microcell boundary than at the macrocell boundary; (2) C/I -criterion: the C/I at the microcell boundary equals or exceeds that at the macrocell boundary.

C -criterion: From the path loss model in (11.10), the microcell radius R_μ can be calculated as

$$R_\mu \leq \left(\left(\frac{\Omega_{t,b}^\mu}{\Omega_{t,b}^M} \right) \left(\frac{h_b^\mu}{h_b^M} \right)^2 \right)^{1/4} R_M, \quad (11.16)$$

where R_M , h_b^μ , h_b^M , $\Omega_{t,b}^\mu$, and $\Omega_{t,b}^M$ are defined in (11.11).

C/I -criterion: The C/I received by the MS at the microcell boundary can be written as

$$\frac{C_\mu^d}{I_\mu^d + J_{M\mu}^d} = \frac{\Omega_{t,b}^\mu (h_{\mu,b} h_m)^2}{R_\mu^4} \left/ \left(\sum_{i=1}^{C_\mu-1} \frac{\Omega_{t,b}^\mu (h_{\mu,b} h_m)^2}{D_{\mu i}^4} + \sum_{i=1}^{N_{Mf}} \frac{\Omega_{t,b}^M (h_{M,b} h_m)^2}{D_{Mf,i}^4} + \frac{1}{\eta} \left(\sum_{i=1}^{N_{Mb}} \frac{\Omega_{t,b}^M (h_{M,b} h_m)^2}{D_{Mb,i}^4} \right) \right) \right., \quad (11.17)$$

where the parameters $\Omega_{t,b}^\mu$, $\Omega_{t,b}^M$, h_b^μ , h_b^M , C_μ , and h_m have already been defined in (11.11) and

- C_μ^d = MS received power from its desired microcell BS
- I_μ^d = downlink microcell-to-microcell interference
- $J_{M\mu}^d$ = downlink macrocell-to-microcell interference
- N_{Mf} = the number of main-lobe macrocell interferers
- N_{Mb} = the number of back-lobe macrocell interferers
- $D_{Mf,i}$ = MS distance to the i th main-lobe interfering BS
- $D_{Mb,j}$ = MS distance to the j th back-lobe interfering BS
- $D_{\mu,i}$ = MS distance to the i th microcell interfering BS
- R_M = macrocell radius
- R_μ = microcell radius
- η = the front-to-back ratio of the directional antenna in macrocells

Let $(C/I)_{\text{req}}$ denote the required C/I . Then (11.17) becomes

$$\frac{R_\mu}{R_M} \leq \left(\frac{(C/I)_{\text{req}}^{-1}}{\sum_{i=1}^{C_\mu-1} \left(\frac{1}{D_{\mu i}} \right)^4 + \left(\sum_{i=1}^{N_{Mf}} \left(\frac{1}{D_{Mf,i}} \right)^4 + \frac{1}{\eta} \sum_{j=1}^{N_{Mb}} \left(\frac{1}{D_{Mb,j}} \right)^4 \right) \left(\frac{\Omega_{t,b}^M}{\Omega_{t,b}^\mu} \right) \left(\frac{h_{M,b}}{h_{\mu,b}} \right)^2} \right)^{1/4}, \quad (11.18)$$

where $\widehat{D}_{M_{f,i}} = D_{M_{f,i}}/R_M$, $\widehat{D}_{M_{b,j}} = D_{M_{b,j}}/R_M$, and $\widehat{D}_{\mu_i} = D_{\mu_i}/R_M$ are the normalized distances of interferers with respect to macrocell radius R_M . Here the microcells and macrocells are assumed to 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 $i = 1, \dots, C_\mu$). Then

$$\frac{D_\mu}{R_\mu} = \sqrt{3K_\mu}, \quad (11.19)$$

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

$$\frac{R_M}{R_\mu} \geq \sqrt{C_\mu K_\mu}. \quad (11.20)$$

Substituting (11.19), (11.20) into (11.18) gives

$$\frac{R_\mu}{R_M} \leq \left(\frac{(C/I)_{\text{req}}^{-1}}{\frac{(C_\mu - 1) C_\mu^2}{9} + \left(\sum_{i=1}^{N_{M_f}} \left(\frac{1}{\widehat{D}_{M_{f,i}}} \right)^4 + \frac{1}{\eta} \sum_{j=1}^{N_{M_b}} \left(\frac{1}{\widehat{D}_{M_{b,j}}} \right)^4 \right) \left(\frac{\Omega_{t,b}^M}{\Omega_{t,b}^\mu} \right) \left(\frac{h_b^M}{h_b^\mu} \right)^2} \right)^{\frac{1}{4}}, \quad (11.21)$$

Notice that N_{M_b} back-lobe macrocell interferers are considered in (11.21). 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 11.2: Referring to Fig. 11.14 and Table 11.1, 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

$$[\widehat{D}_{M_{b,1}}, \widehat{D}_{M_{b,2}}, \widehat{D}_{M_{b,3}}] = [2.65, 2.65, 2.65] \quad (11.22)$$

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

$$[\widehat{D}_{M_{f,1}}, \widehat{D}_{M_{f,2}}, \widehat{D}_{M_{f,3}}] = [5.29, 5.29, 5.29]. \quad (11.23)$$

Furthermore, three main-lobe interfering macrocell BSs in the third tier are located near micro-areas 13, 70, and 85 with distances of

$$[\widehat{D}_{M_{f,4}}, \widehat{D}_{M_{f,5}}, \widehat{D}_{M_{f,6}}] = [7.0, 7.0, 7.0]. \quad (11.24)$$

It is also important to determine if there exist interfering microcell BSs from neighboring micro-areas. Figure 11.14 and Table 11.1 show one feature of the proposed system architecture; the adjacent micro-areas are assigned different macrocell channel sets. For instance, micro-area 56 in Fig. 11.14 is assigned channel sets $[4_\alpha, 4_\gamma]$, while the

(continued)

Example 11.2 (continued)
 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 (11.22), (11.23), and (11.24) into (11.21), one can obtain

$$\frac{R_\mu}{R_M} \leq \left(\frac{(C/I)_{\text{req}}^{-1}}{\frac{(C_\mu - 1)C_\mu^2}{9} + \left(5.0803 \times 10^{-3} + 0.0608 \times \frac{1}{\eta} \right) \left(\frac{\Omega_{t,b}^M}{\Omega_{t,b}^\mu} \right) \left(\frac{h_b^M}{h_b^\mu} \right)^2} \right)^{\frac{1}{4}} \quad (11.25)$$

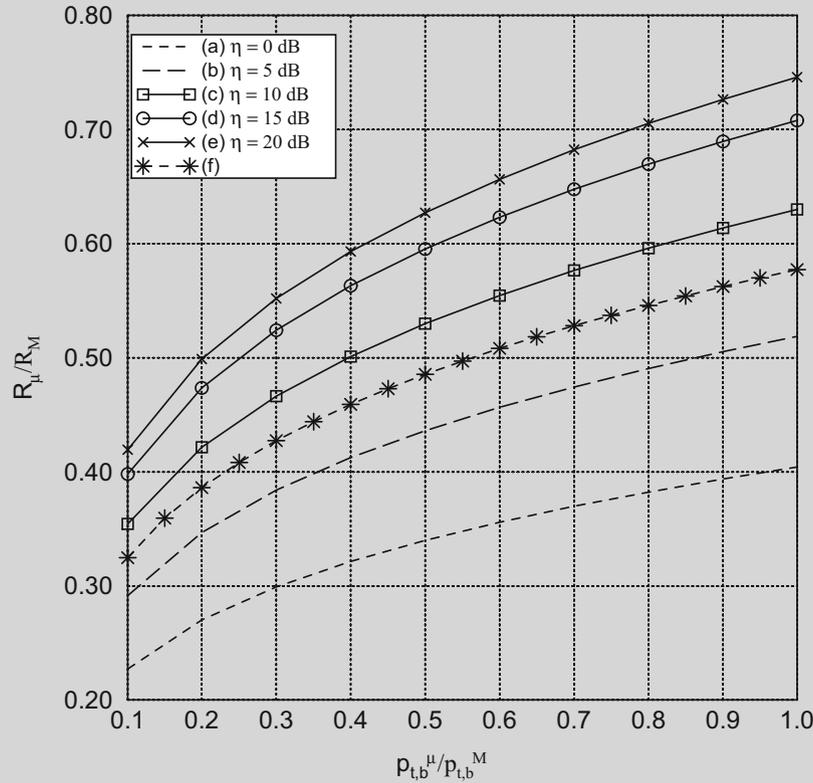


Fig. 11.18 Effect of front-to-back ratio η on the microcell radius based on downlink microcell C/I performance analysis, where R_μ/R_M and $\Omega_{t,b}^\mu/\Omega_{t,b}^M$ are the cell radius ratio and transmitted power ratio of microcells over macrocells, respectively. With $(C/I)_{\text{req}} = 18$ dB and $h_b^\mu/h_b^M = 1/3$, 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

(a) $C_\mu = 1$: First consider the special case where only one microcell is installed in each micro-area. This situation may occur with initial microcell deployment. Figure 11.18 shows the effect of the front-to-back ratio η on the microcell radius, in the case $(C/I)_{\text{req}} = 18$ 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. 11.18, one can observe that if front-to-back ratio $\eta \geq 10$ dB, the

(continued)

Example 11.2 (continued)

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 $R_\mu \leq 0.5R_M$ by the C/I-criterion and $R_\mu \leq 0.46R_M$ by the C-criterion, respectively. The more stringent requirement must be satisfied and, therefore, the microcell radius is $0.46R_M$. 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, consider the case where many microcells are deployed in each microarea. Figure 11.19 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 $R_\mu \leq 0.165R_M$ from Fig. 11.19.

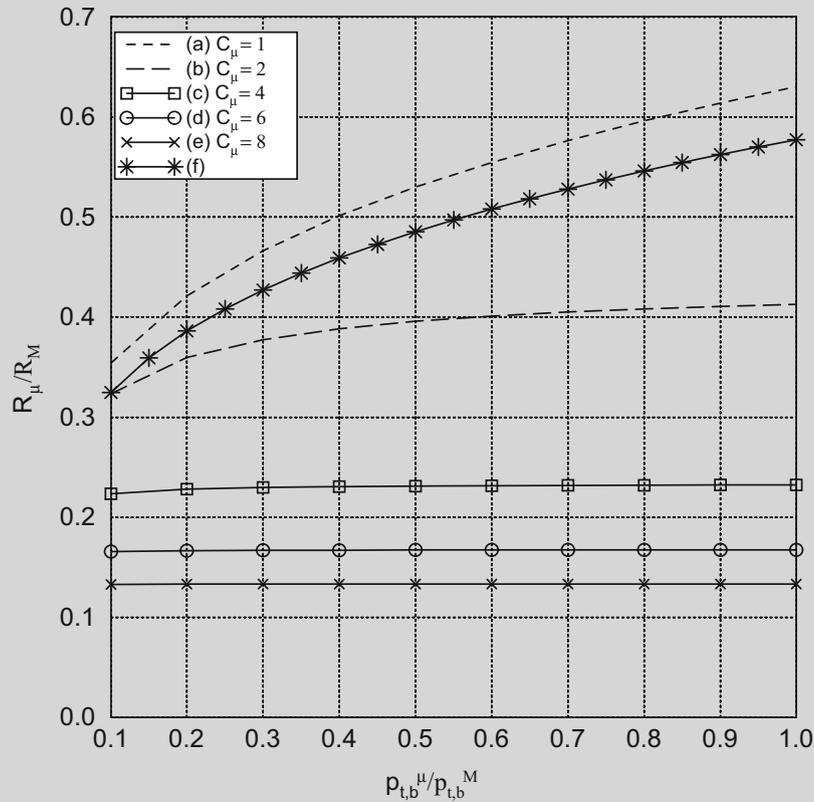


Fig. 11.19 Downlink microcell radius R_μ against $\Omega_{t,b}^\mu/\Omega_{t,b}^M$ for different values of C_μ in the case $\eta = 10$ dB, $(C/I)_{\text{req}} = 18$ dB, and $h_b^\mu/h_b^M = 1/3$, whereby the microcell radius is normalized with respect to the macrocell radius R_M ; $\Omega_{t,b}^\mu/\Omega_{t,b}^M$ represents the ratio of the transmitted power of microcell BS to that of macrocell BS; C_μ is the number of clusters in a micro-area; η is the front-to-back ratio of the directional antenna; h_b^μ/h_b^M is the ratio of the microcell BS antenna to the macrocell BS antenna. Curves (a) ~ (e) are obtained by C/I-criterion for $C_\mu = 1, 2, 4, 6, 8$, while curve (f) is obtained by C-criterion

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 specifically,

$$\frac{C_\mu^u}{I_\mu^u + J_{M\mu}^u} = \frac{\Omega_{t,m}^\mu \frac{(h_b^\mu h_m)^\mu}{R_{\mu,\text{up}}^4}}{\sum_{i=1}^{C_\mu-1} \frac{\Omega_{t,m}^\mu (h_b^\mu h_m)^\mu}{D_{\mu,i}^4} + \sum_{i=1}^{N_{Mf,i}} \frac{\Omega_{t,m}^M (h_b^M h_m)^\mu}{D_{M_i}^4}}, \quad (11.26)$$

where the parameters $\Omega_{t,b}^\mu$, $\Omega_{t,b}^M$, C_μ , h_b^μ , $D_{\mu,i}$, R_M , h_b^μ , and h_m have been defined in (11.11) and (11.17) and

- C_μ^u = microcell BS received power from the desired microcell MS
- $I_{\mu M}^u$ = uplink microcell-to-macrocell interference
- $J_{M\mu}^u$ = uplink macrocell-to-microcell interference
- N_{Mf} = the number of macrocell interfering MSs
- $D_{M,i}$ = BS distance to the i th interfering macrocell MS
- $R_{\mu,\text{up}}$ = uplink microcell radius

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 (11.18), one can simplify (11.26) as

$$\frac{R_{\mu,\text{up}}}{R_M} \leq \left(\frac{(C/I)_{\text{req}}^{-1}}{\frac{(C_\mu - 1) C_\mu^2}{9} + \left(\sum_{i=1}^{N_{Mf}} \left(\frac{1}{\widehat{D}_{Mf,i}} \right)^4 \right) \left(\frac{\Omega_{t,m}^M}{\Omega_{t,m}^\mu} \right)} \right)^{\frac{1}{4}}. \quad (11.27)$$

It has been 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 performance, (11.27) can be approximated as

$$\frac{R_{\mu,\text{up}}}{R_M} \leq \left(\frac{1}{(C/I)_{\text{req}} \frac{(C_\mu - 1) C_\mu^2}{9}} \right)^{\frac{1}{4}}. \quad (11.28)$$

By combining (11.19), (11.20) and (11.28), upper and lower bounds are obtained on K_μ as

$$\frac{1}{3} \sqrt{(C/I)_{\text{req}} (C_\mu - 1)} \leq K_\mu \leq \frac{1}{C_\mu} \left(\frac{R_M}{R_\mu} \right)^2. \quad (11.29)$$

The relation between K_μ and C_μ with R_μ/R_M as a parameter is shown in Fig. 11.20.

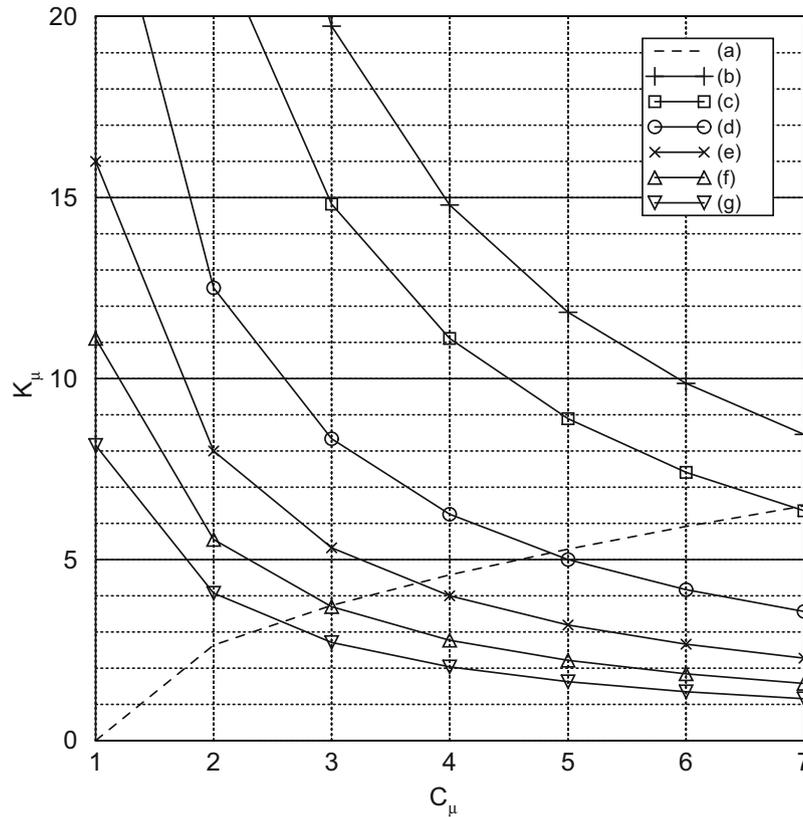


Fig. 11.20 K_μ against C_μ with R_μ/R_M as a parameter, whereby K_μ is the microcell cluster size, C_μ is the number of clusters in a micro-area, and R_μ/R_M is the ratio of the microcell radius to the macrocell radius. Curve (a) represents the lower bound of K_μ , while curves (b) ~ (g) represent the upper bound of K_μ for $R_\mu/R_M = 0.13, 0.15, 0.20, 0.25, 0.30, 0.35$, respectively

Example 11.3: Consider again micro-area 56 in Fig. 11.14. Referring to Table 11.1, micro-area 56 can be assigned channel sets $[4_\alpha, 4_\gamma]$. Take channel set 4_α for example. The worst case interference 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 (11.27) and then letting $(C/I)_{\text{req}} = 18$ dB, Fig. 11.21 shows 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. The feasible cluster size K_μ and the microcell radius are first determined. From Fig. 11.20, $K_\mu = 7$ and $R_\mu = 0.15 \times R_M$. Then from Fig. 11.21, 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 transmitting at -2 dBW. The microcell MS transmitted power should be larger than -20 dBW in this case. If the MS transmitted power ranges from -22 to -2 dBW as was assumed earlier, then the MS can be used in both the macrocells and microcells of the cluster-planned architecture without changing the MS transmit power specification.

(continued)

Example 11.3 (continued)

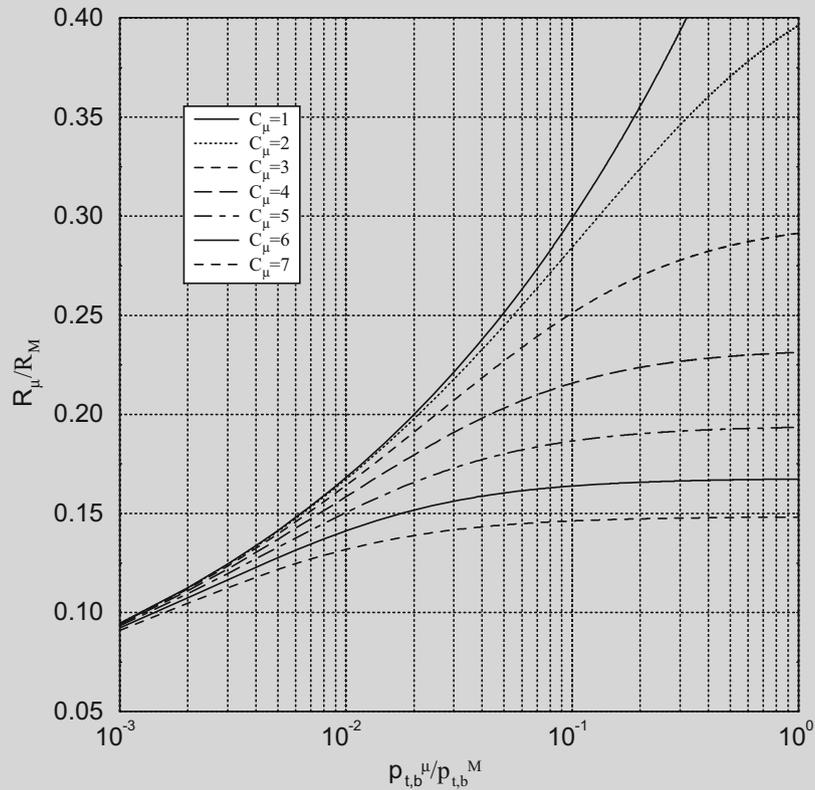


Fig. 11.21 Uplink microcell radius R_μ against $\Omega_{t,m}^\mu/\Omega_{t,m}^M$ for different values of C_μ , where the microcell radius is normalized by the macrocell radius R_M , $\Omega_{t,m}^\mu/\Omega_{t,m}^M$ is the ratio of the transmitted power of the microcell MS to that of the macrocell MS, C_μ is the number of microcell clusters in a micro-area, and $(C/I)_{\text{req}} = 18$ dB

11.4.3.4 Adjacent Channel Interference Analysis

To avoid excessive adjacent channel interference, it is desirable not to use the same channel sets in adjacent sectors. A frequency plan designed to avoid adjacent channel interference in the conventional macrocellular system is first reviewed. Then it will be shown that the same plan works for the cluster-planned hierarchical architecture. As shown in Fig. 11.10, 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 11.3 [196]. 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 11.3 is applied to the cluster-planned architecture in Fig. 11.14, there is no adjacent channel interference between macrocell 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 within a micro-area. For example, referring to Fig. 11.14 and Table 11.1, 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 11.1.

Table 11.3 Frequency management plan for avoiding adjacent channel interference

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	

11.5 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 [171, 196], although it is effective against envelope fading as well. A TDMA macrodiversity system serves a mobile station (MS) simultaneously by L different 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 system. Usually, the branch selection is based on the local mean power rather than the instantaneous power [7, 8, 171, 322, 343, 370], because the branch selection algorithm cannot react to the rapidly varying instantaneous signal power. Here, the focus is on *local-mean-based* branch selection schemes.

Previous studies on macrodiversity systems have evaluated the co-channel interference performance with shadowing only [41, 341, 366] and shadowed Rayleigh fading channels [342]. The co-channel interference performance was also discussed in [207], 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 [7, 8, 322, 323, 370]. However, these studies did not consider co-channel interference. The analysis in [344] 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.

11.5.1 Co-channel Interference Outage

An analytical model is now considered for calculating the probability of co-channel interference outage, O_I , for an L -branch local-mean-based macrodiversity system with log-normal shadowing. The model assumes that the local mean envelope power of the desired signal, $\Omega_{d,k}$, is available for each branch k , where $k = 1, \dots, L$. 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

$$\Omega_d = \max(\Omega_{d,1}, \Omega_{d,2}, \dots, \Omega_{d,L}). \quad (11.30)$$

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 treated as independent random variables with the pdf in (2.355), then Ω_d has the pdf $p_{\Omega_d}(x) = L[F_k(x)]^{L-1}p_k(x)$. The probability of co-channel interference outage is

$$\begin{aligned} O_I &= \text{P}[\Omega_d/\Omega_I < \Lambda_{\text{th}}] \\ &= 1 - \int_0^\infty \left(\int_{-\infty}^{x/\Lambda_{\text{th}}} p_{\Omega_I}(x)(y)dy \right) p_{\Omega_d}(x)dx, \end{aligned} \quad (11.31)$$

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 non-coherently so that the total interference power on the k th 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 i th interferer on the k th 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 Sect. 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 & Yeh methods.

If the $\{\Omega_{I,k}\}_{k=1}^n$ are independent and identically distributed (i.i.d.), and the $\{\Omega_{d,k}\}_{k=1}^L$ are also i.i.d. and independent of the $\{\Omega_{I,k}\}_{k=1}^n$, then [341, 366]

$$\begin{aligned} O_I &= 1 - L \int_0^\infty \left(\int_0^{x/\Lambda_{\text{th}}} \frac{1}{\sqrt{2\pi}\sigma_{\Omega_I}\xi y} \exp\left\{ -\frac{(10\log_{10}\{y\} - \mu_{\Omega_I}(\text{dB}))^2}{2\sigma_{\Omega_I}^2} \right\} dy \right) \\ &\quad \times \left(1 - Q\left(\frac{10\log_{10}\{x\} - \mu_{\Omega_d}(\text{dB})}{\sigma_{\Omega_d}} \right) \right)^{L-1} \frac{1}{\sqrt{2\pi}\sigma_{\Omega_d}\xi x} \exp\left\{ -\frac{(10\log_{10}\{x\} - \mu_{\Omega_d}(\text{dB}))^2}{2\sigma_{\Omega_d}^2} \right\} dx, \end{aligned} \quad (11.32)$$

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 k th diversity branch, respectively.

For ease of evaluation, let $w = (10\log_{10}\{x\} - \ln \mu_{\Omega_d}(\text{dB}))/\sqrt{2}\sigma_{\Omega_d}$ and transform (11.32) into a Hermite integration form. That is,

$$O_I = 1 - \int_{-\infty}^\infty g(w)e^{-w^2}dw \simeq 1 - \sum_{i=1}^n g(w_i)h_i, \quad (11.33)$$

where

$$g(w) = \frac{L}{\sqrt{\pi}} \left(1 - Q\left(\frac{\sqrt{2}\sigma_{\Omega_d}w + \xi(\mu_{\Omega_d}(\text{dB}) - \mu_{\Omega_I}(\text{dB}) - \Lambda_{\text{th}}(\text{dB}))}{\sigma_{\Omega_I}} \right) \right) \left(1 - Q(\sqrt{2}w) \right)^{L-1}, \quad (11.34)$$

and w_i and h_i are the roots and weight factors of the n th-order Hermite polynomial, respectively [4].

11.5.2 Shadow Correlation

Until now, independent shadowing has been assumed on the macrodiversity branches. However, in many cases the macrodiversity branches will be correlated. Define

$$\mathbf{\Omega}_d = (\Omega_{d,1}, \Omega_{d,2}, \dots, \Omega_{d,L}). \quad (11.35)$$

For a correlated L -branch macrodiversity system, the joint pdf of $\mathbf{\Omega}_d$ is [89]

$$p_{\mathbf{\Omega}_d}(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^L \det(\mathbf{M})} \xi^L z_1 \cdots z_L} \exp\left\{ -\frac{1}{2} \mathbf{Y}^T \mathbf{M}^{-1} \mathbf{Y} \right\}, \quad (11.36)$$

where $\mathbf{z} = (z_1, \dots, z_L)$, $\mathbf{Y}^T = [y_1, \dots, y_L]$ denotes the transpose of column vector

$$\mathbf{Y} = \begin{bmatrix} 10\log_{10}(z_1) - \mu_{\Omega_{d,1}} \text{ (dB)} \\ \vdots \\ 10\log_{10}(z_L) - \mu_{\Omega_{d,L}} \text{ (dB)} \end{bmatrix} \quad (11.37)$$

and $\mu_{\Omega_{d,1}} \text{ (dB)}, \dots, \mu_{\Omega_{d,L}} \text{ (dB)}$ are the area means of each diversity branch. The covariance matrix \mathbf{M} is expressed as

$$\mathbf{M} = \begin{bmatrix} \sigma_{\Omega_1}^2 & \cdots & v_{1L} \\ \vdots & \ddots & \vdots \\ v_{L1} & \cdots & \sigma_{\Omega_L}^2 \end{bmatrix}, \quad (11.38)$$

where σ_{Ω} is the shadowing standard deviation and v_{ij} is the covariance of $\Omega_{d,i} \text{ (dB)}$ and $\Omega_{d,j} \text{ (dB)}$

$$v_{ij} = E[(\Omega_{d,i} \text{ (dB)} - \mu_{\Omega_{d,i}} \text{ (dB)}) (\Omega_{d,j} \text{ (dB)} - \mu_{\Omega_{d,j}} \text{ (dB)})]. \quad (11.39)$$

It is convenient to define $\mathbf{N} = \mathbf{M}^{-1}$ and express the matrix multiplication in (11.36) in the form

$$\mathbf{Y}^T \mathbf{N} \mathbf{Y} = \sum_{i=1}^L N_{ii} y_i^2 + 2 \sum_{i=1}^{L-1} \sum_{j=i+1}^L N_{ij} y_i y_j \quad (11.40)$$

where N_{ij} is the element in the i th row and j th column.

According to (11.30), (11.36), and (11.40), the probability that $\Omega_d < y$ is

$$P[\Omega_d < y] = \int_0^y \cdots \int_0^y \frac{1}{\sqrt{(2\pi)^L \det(\mathbf{M})} \xi^L z_1 \cdots z_L} \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^L N_{ii} y_i^2 + 2 \sum_{i=1}^{L-1} \sum_{j=i+1}^L N_{ij} y_i y_j \right) \right\} d\mathbf{z}, \quad (11.41)$$

where N_{ij} and y_i are defined in (11.40) and (11.37), 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 cannot 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 $\mu_{\Omega_{d,j}} \text{ (dB)} = \mu_{\Omega_d} \text{ (dB)}$

$$p_{\Omega_d}(z) = \frac{1}{\sqrt{2\pi \det(\mathbf{M})} \xi^2} \left\{ \frac{1}{\sqrt{N_{22}}} \exp \left\{ -\frac{z^2}{2} \left(N_{11} - \frac{N_{12}}{N_{22}} \right) \right\} \left(1 - Q \left(\left(\sqrt{N_{22}} + \frac{N_{12}}{\sqrt{N_{22}}} \right) z \right) \right) \right. \\ \left. + \frac{1}{\sqrt{N_{11}}} \exp \left\{ -\frac{z^2}{2} \left(N_{22} - \frac{N_{12}}{N_{11}} \right) \right\} \left(1 - Q \left(\left(\sqrt{N_{11}} + \frac{N_{12}}{\sqrt{N_{11}}} \right) z \right) \right) \right\}, \quad (11.42)$$

where $z = (10\log_{10}y - \mu_{\Omega_d} \text{ (dB)})$. Consider the following covariance matrix \mathbf{M}

$$\mathbf{M} = \begin{bmatrix} \sigma^2 & v \\ v & \sigma^2 \end{bmatrix}, \quad (11.43)$$

and

$$\mathbf{N} = \mathbf{M}^{-1} = \frac{1}{\sigma^4 - v^2} \begin{bmatrix} \sigma^2 & -v \\ -v & \sigma^2 \end{bmatrix}. \quad (11.44)$$

By substituting (11.44) into (11.42), the pdf of the output local-mean power of the dual macrodiversity system can be expressed as

$$p_{\Omega_d}(y) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma\xi^2y} \left\{ 1 - Q \left(\left(\frac{1-r}{\sqrt{1-r^2}} \right) \left(\frac{10\log_{10}y - \mu_{\Omega_d} \text{ (dB)}}{\sigma} \right) \right) \exp \left\{ -\frac{(10\log_{10}y - \mu_{\Omega_d} \text{ (dB)})^2}{2\sigma^2} \right\} \right\}, \quad (11.45)$$

where the correlation coefficient r is defined as $r = v/\sigma^2$. Combining (11.31) and (11.45) gives

$$O_I = 1 - \int_{-\infty}^{\infty} g(w)e^{-w^2} dw \simeq 1 - \sum_{i=1}^n g(w_i)h_i, \quad (11.46)$$

where

$$g(w) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(1 - Q \left(\frac{\sqrt{2}\sigma_d w + [\mu_{\Omega_d} \text{ (dB)} - \mu_{\Omega_I} \text{ (dB)} - \Lambda_{th} \text{ (dB)}]}{\sigma_{\Omega_I}} \right) \right) \left(1 - Q \left(\frac{1-r}{\sqrt{1-r^2}} \sqrt{2}w \right) \right) dw. \quad (11.47)$$

11.5.3 Numerical 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 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 $g = 0.15R$ in (2.248).

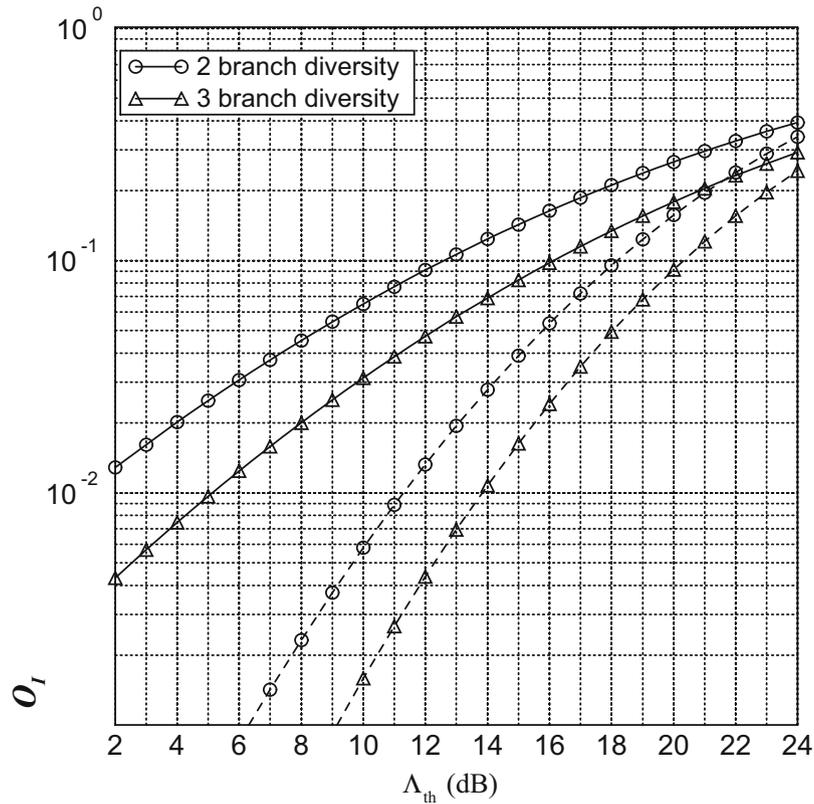


Fig. 11.22 Probability of co-channel interference outage, O_I , against the required threshold, Λ_{th} , at the receiver for the local-mean-based macrodiversity system, where the *solid lines* denote the case for shadowing standard deviation $\sigma = 10$ dB and the *dashed lines* for $\sigma = 6$ dB; $a = b = 2$, $g = 0.15R$; two interferers are located at a distance of $5.2R$

Table 11.4 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

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 11.5 Effects of branch correlation on a 2-branch macrodiversity

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	—

Figure 11.22 shows the probability of co-channel interference outage performance, while Table 11.4 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.

The effects of correlation coefficient r on a 2-branch macrodiversity system are evaluated with various σ ; $\sigma = 6$ dB in Fig. 11.23 and $\sigma = 10$ dB in Fig. 11.24. With respect to a 10% outage, Table 11.5 lists Λ_{th} with different r . Observe 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 $r = 0$.

Problems

11.1. Consider a regular hexagonal cell deployment, where the MSs and BSs use omnidirectional antennas. Consider the forward channel performance and consider only the first tier of co-channel interferers as shown in Fig. 1.12. 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).

- 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.
- What is the minimum cluster size that is needed if the radio receivers have $\Lambda_{\text{th}} = 9$ dB?
- Referring to Fig. 1.13, repeat parts (a) and (b) for the reverse channel.

Note: In this problem you must use exact radio path distances.

11.2. Consider a cellular system that uses a 7-cell hexagonal reuse cluster. The base stations employ 120° wide-beam directional antennas and they all have the same antenna height and transmit with the same power level. Consider the forward channel (base-to-mobile). Ignore shadowing and envelope fading and consider only the path loss. A mobile station will experience the lowest co-channel interference ratio, Λ , when it is located in the corner of a cell.

- Considering only the first tier of co-channel base stations, what is the worst case Λ with a path loss exponent of 4?
- Considering the first two tiers of co-channel base stations, what is the worst case Λ with path loss exponent of 4?
- From parts a) and b) what conclusions can you make about the effect of the second-tier co-channel base stations?
- What happens if the path loss exponent is equal to 3?

Note: In this problem you must use exact radio path distances.

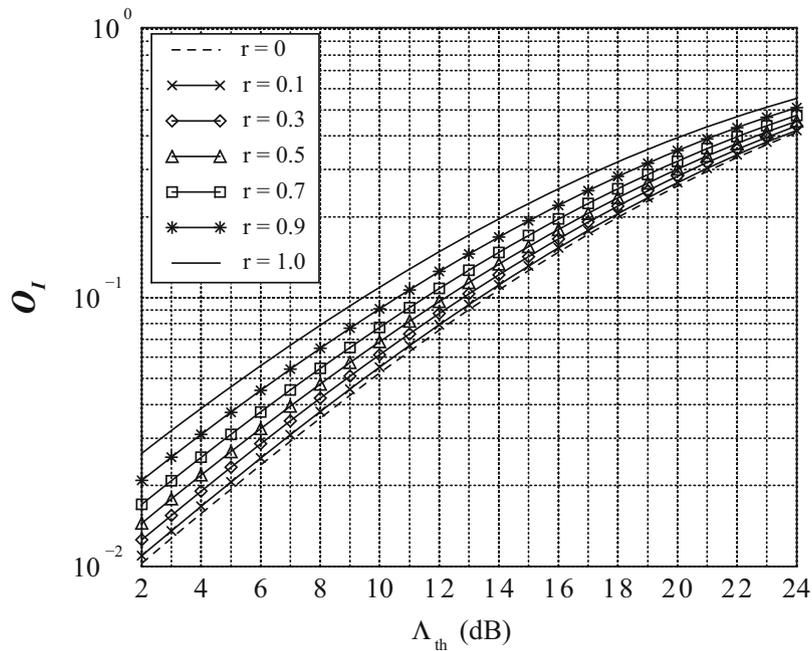


Fig. 11.23 Effect of branch correlation coefficient r on the local-mean-based macrodiversity system with $\sigma = 6$ dB; $a = b = 2$, $g = 0.15R$; two interferers are located at a distance of $5.2R$

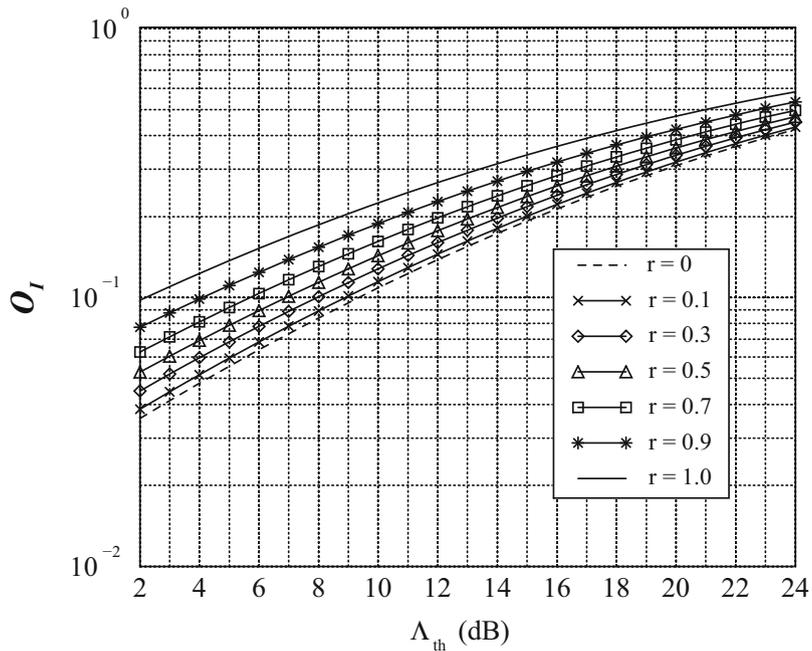


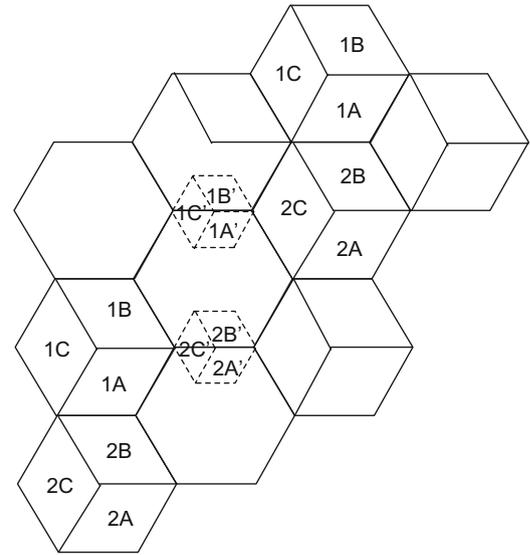
Fig. 11.24 Effect of branch correlation coefficient r on the local-mean-based macrodiversity system with $\sigma = 10$ dB; $a = b = 2$, $g = 0.15R$; two interferers are located at a distance of $5.2R$

11.3. Consider the reverse link of a cellular system that uses a 7-cell reuse cluster with omnidirectional base station antennas. Ignore envelope fading and shadowing, and assume the simple path loss model

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

where $\beta = 3.5$. Also assume that all mobile stations are transmitting at the same power level.

Fig. 11.25 Split cells for Problem 11.4



- Show graphically the worst case co-channel interference geometry for the reverse channel.
- Calculate the worst case carrier-to-interference ratio, Λ in terms of the co-channel reuse factor D/R .
- Repeat parts (a) and (b) if 120° cell sectoring is used.

Note: In this problem you must use exact radio path distances.

11.4. Cell splitting is a process whereby smaller cells are introduced into an existing cellular deployment. If heavy traffic loading is experienced at the midpoint between the two cells labeled **1** in Fig. 11.25 above, then a split cell labeled **1'** can be introduced at that location. Likewise, the split cell labeled **2'** can be introduced at the midpoint between the two cells labeled **2**.

Assume propagation path loss according to the following model:

$$\mu_{\Omega_p}(d) = \mu_{\Omega_p}(d_o, h_{b_o}, h_{m_o}) \frac{[(h_b/h_{b_o})(h_m/h_{m_o})]^2}{(d/d_o)^\beta}$$

where $\mu_{\Omega_p}(\text{dBm}) = -70$ dBm at $d_o = 1$ km, $\beta = 3.5$, $h_{b_o} = 70$ m, $h_{m_o} = 1.5$ m. Assume that $h_b = h_{b_o} = 70$ m and $h_m = h_{m_o} = 1.5$ m unless otherwise specified.

- Convert the path loss model to dB units, with power levels referenced to 1 mW.
- If the receiver sensitivity is -115 dBm, and the transmit and receive antenna gains are unity (0 dB), what is the radius of the original cells?
- What is the radius of the split cells?
- If the MS is to receive a signal level of -115 dBm on the corner of the split cell, by how much should $\mu_{\Omega_p}(d_o, h_{b_o}, h_{m_o})$ or, equivalently, the BS transmit power be reduced or increased?
- Repeat part c) if the split cell uses an antenna height of only 30 m.

11.5. A cellular network provider uses a TDMA scheme that can tolerate a C/I of 9 dB in the worst case. The propagation environment is characterized by a path loss exponent $\beta = 4$.

- Find the best value of N for (1) omnidirectional antennas, (2) 120° sectoring, and (3) 60° sectoring.
- Suppose that the system bandwidth supports a total of N_{tot} voice channels. The objective is to maximize the “cell capacity” defined as the number of voice channels per cell. Should sectoring be used?
- If sectoring is used, should you use 120° or 60° sectoring? Explain.

11.6. One method for improving the capacity of a cellular system is to employ a *two-channel bandwidth* scheme, where a hexagonal cell is divided into two concentric hexagons as shown in Fig. 11.26 below. The inner hexagon is serviced by half-rate channels, while the outer hexagon is serviced by full-rate channels. When a mobile station crosses the boundary between the inner and outer portions of a cell a handoff occurs.

Fig. 11.26 Cell division with two channel bandwidth scheme

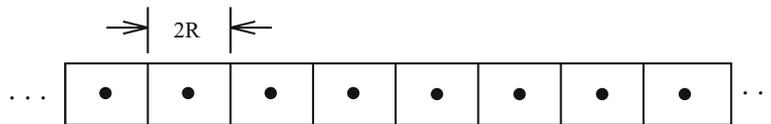
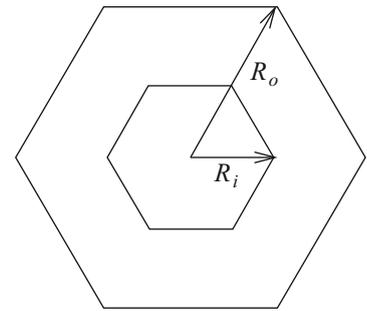


Fig. 11.27 Highway cell reuse pattern for Problem 11.7

Suppose that the full-rate channels require $C/I = 7$ dB to maintain an acceptable radio link quality, while the half-rate channels require $C/I = 10$ dB.

Assume a fourth-law path loss model and ignore envelope fading and shadowing. Consider the reverse link and assume that there are six co-channel interferers at distance D from the serving base station. It follows that the worst case C/I when a mobile station is located at distance d from the base station is $C/I = (D/d)^4/6$.

Hence, a $C/I = 7$ dB requires a 7-cell reuse cluster with $D/R_o = 4.6$, where R_o is the radius of the outer cell.

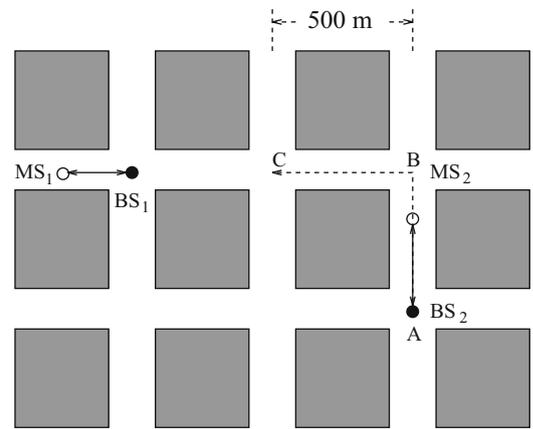
- Find the required value of D/R_o so that $C/I = 7$ dB in the outer cell, where R_i is the radii of the outer cell.
- Find the required value of D/R_i so that $C/I = 10$ dB in the inner cell, where R_i is the radii of the inner cell.
- Use the values of D/R_i and D/R_o to determine the ratio of the inner and outer cell areas, A_i/A_o . Use the exact area of a hexagon in terms of its radius.
- Let N_i and N_o be the number of channels that are allocated to the inner and outer areas of each cell, and assume that the channels are assigned such that $N_i/N_o = A_i/A_o$. Determine the increase in cell capacity (as measured in channels per cell) over a conventional *one-channel bandwidth* system that uses only full-rate channels.

11.7. You are asked to design a highway cell reuse pattern as shown in Fig. 11.27. Each cell has length $2R$.

- A base station with an omnidirectional antenna is placed at the center of each cell. Shadowing is ignored under assumed line-of-sight conditions. Derive an expression for C/I in terms of the linear reuse cluster size N . Afterwards, determine the minimum reuse factor needed to yield a worst case carrier-to-interference ratio (C/I) of 9 dB or more. Consider only the nearest neighboring co-channel interferers and assume a propagation path loss exponent $\beta = 4$.
- Now suppose that directional antennas are used to divide each cell into two semicircular sectors with boundaries perpendicular to the highway. Repeat part (a).
- Consider again the sectored cell arrangement in part (b). However, log-normal shadowing (due to heavy vehicle traffic) is now present with shadow standard deviation of σ_Ω dB. Derive an expression for the C/I outage on a cell boundary, again considering only the nearest neighbouring co-channel interferers.

11.8. 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. 11.28 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 non-line-of-sight corner path loss model in (2.412). Suppose that $a = 2$, $b = 4$, and $g = 150$ m. Plot \mathcal{L} 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 1 m.

Fig. 11.28 Microcellular propagation environment for Problem 11.8



Chapter 12

CDMA Cellular Systems

CDMA has been adopted in third generation (3G) cellular systems, notably in the cdma2000 and WCDMA standards, due to its many benefits such as universal frequency reuse, soft handoff capability, and the ability to exploit multipath-fading with RAKE receivers. CDMA waveforms are broadband and noise-like, so that the multiple-access interference behaves in a manner that is approximately equivalent to AWGN. CDMA cellular systems are interference limited and, therefore, their capacities are closely related to the amount of multiple-access interference that is generated by the transmitters on one hand and tolerated by the receivers on the other. Because of the universal frequency reuse, multiple-access interference in CDMA cellular systems is due to both intra-cell interference from transmissions in the same cell and inter-cell interference from transmissions in neighboring cells. To alleviate the level of multiple-access interference, and thus increase the capacity and quality, CDMA systems must employ power control in conjunction with soft handoff. The fundamental idea behind power control is to limit the generation of multiple-access interference by restraining mobile stations (MSs) and base stations (BSs) from transmitting any more power than is necessary. With power control, each MS and BS transmits just enough power to meet the required carrier-to-interference ratio (CIR) at the intended receiver.

This chapter considers capacity and performance of CDMA cellular systems. Section 12.1 begins the chapter with a discussion of the power control mechanism in the CDMA reverse and forward links. Afterwards, the reverse and forward link capacity of CDMA cellular systems is considered, and the impact of imperfect power control is demonstrated in Sect. 12.2. Section 12.3 considers hierarchical CDMA cellular architectures consisting of macrocells and underlaid macrocells, where both hierarchical layers use the entire system bandwidth. On the reverse link, this is accomplished by using macrodiversity maximal ratio combining where the signals received at multiple base stations (BSs) are coherently combined. On the forward link, only one BS can transmit to a given mobile station (MS) at any given time. The forward link transmit power is determined according to a neighboring cell pilot power scheme, where the forward transmit power to each MS is determined according to link conditions between the MS and surrounding BSs. Additional improvement can be gained by using selective transmit diversity at the BSs on the forward link.

12.1 CDMA Power Control

CDMA cellular systems must use reverse channel power control to combat the near-far effect, a phenomenon where the signals received from distant MSs will experience excessive multiple-access interference from close-in MSs due to differences in the propagation path loss. Shadowing and envelope fading will also contribute significantly to the near-far effect, although these propagation factors are not distance dependent. To combat variations in the received CIR due to path loss, shadowing, and envelope fading, the reverse link in CDMA cellular systems employs a fast closed-loop power control algorithm. For cdma2000 cellular systems, the transmit power from each MS is adjusted every 1.25 ms to equalize the corresponding received CIR at the serving BS. Such adjustments are based on measurements of the received E_b/N_t (bit energy to interference-plus-noise spectral density ratio). The reverse closed-loop power control algorithm consists of outer and inner power control loops. The outer power control loop occurs between the base station controller and the BSs, and adjusts a “set point” or target received E_b/N_t according to measurements of the frame error rate that is required to maintain a desired quality of service. The outer loop adjustments are performed at the frame interval, for example, every 20 ms in cdma2000. The inner power control loop occurs between the BSs and MSs. The serving BS compares measurements of the

received E_b/N_t from each MS to the corresponding target E_b/N_t . Based on the results of these comparisons the BS instructs each MS to either increase or decrease its transmit power by a power increment at short periodic intervals, for example, ± 1 dB every 1.25 ms in cdma2000.

CDMA cellular systems also use forward link power control to combat the corner effect, a condition where a MS experiences a decrease in received CIR near the corner of a cell. While the IS-95A/B cellular system uses open loop power control on the forward link, 3G CDMA cellular systems such as cdma2000 and WCDMA also use fast closed-loop power control on the forward link based on measurements of the received E_c/I_o (chip energy to interference-plus-noise spectral density ratio) that are obtained from the forward pilot channel. Once again, the forward closed-loop power control algorithm consists of outer and inner power control loops. The outer power control loop is carried out entirely within the MS, where the MS adjusts a “set point” or target received E_c/I_o according to measurements of the frame error rate that is required to maintain a desired quality of service. The outer loop adjustments can be performed at the frame interval, for example, every 20 ms in cdma2000. The inner power control loop occurs between the MSs and BSs. The MS compares measurements of the received E_c/I_o from the BS to the corresponding target E_c/I_o . Based on the results of these comparisons, the MS instructs the BS to either increase or decrease its transmit power by a predetermined power increment at short periodic intervals, for example, ± 1 dB every 1.25 ms in cdma2000.

The basic idea of CDMA power control can be described as follows [197, 321]. Consider the CDMA reverse link where a BS serves K MSs (although the same argument can be applied to forward link power control). The reverse link CIR experienced by MS i , $(C/I)_i$, is

$$(C/I)_i = \frac{G_i P_i}{\sum_{j \neq i}^K G_j P_j + I_o + N_o}, \quad (12.1)$$

where G_i is the link gain (due to path loss, shadowing, and fading) between MS $_i$ and the BS, and P_i is the reverse transmit power MS $_i$. The term I_o represents the out-of-cell interference generated by the surrounding cells and N_o is the additive white Gaussian noise (AWGN). The goal is to develop a recursive algorithm to adjust the transmit power of MS $_i$ to drive $(C/I)_i$ to meet or exceed a desired target value, $(C/I)_{\text{target},i}$, i.e.,

$$(C/I)_i \geq (C/I)_{\text{target},i}, \quad i \leq i \leq K. \quad (12.2)$$

Rearranging (12.1) gives

$$P_i = \sum_{\substack{j=1 \\ j \neq i}}^K (C/I)_i \frac{G_j}{G_i} P_j + (C/I)_i \frac{I_o + N_o}{G_i}. \quad (12.3)$$

Based on the above derivation, define the $K \times K$ matrix $\mathbf{W} = [W_{ij}]$ and vector $\mathbf{U} = (U_i)$ having elements

$$W_{ij} = \begin{cases} 1 & , \quad i = j \\ -(C/I)_{\text{target},i} \frac{G_j}{G_i} & , \quad i \neq j \end{cases} \quad (12.4)$$

$$U_i = (C/I)_{\text{target},i} \frac{I_o + N_o}{G_i}, \quad 1 \leq i \leq K. \quad (12.5)$$

The existence of a power vector, $\mathbf{P} = (P_1, P_2, \dots, P_K)^T$, such that $\mathbf{W}\mathbf{P} = \mathbf{U}$ implies that $(C/I)_i \geq (C/I)_{\text{target},i}$, $1 \leq i \leq K$. When such a matrix \mathbf{P} exists, \mathbf{W} has only strictly positive eigenvalues and, therefore, \mathbf{P} is given by

$$\mathbf{P} = \mathbf{W}^{-1}\mathbf{U}. \quad (12.6)$$

It is proven in [321] that whenever a unique power vector \mathbf{P} that satisfies $(C/I)_{\text{target},i}$, $1 \leq i \leq K$, exists, $\mathbf{P}(n)$ converges to $\mathbf{P} = \mathbf{W}^{-1}\mathbf{U}$ for any initial power vector $\mathbf{P}(0)$ according to the following recursion:

$$\mathbf{P}(n+1) = (\mathbf{I} - \mathbf{W})\mathbf{P}(n) + \mathbf{U}. \quad (12.7)$$

The above recursion can be rewritten as

$$P_i(n+1) = P_i(n) \frac{(C/I)_{\text{target},i}}{(C/I)_i(n)}, \quad (12.8)$$

where $P_i(n)$ and $(C/I)_i(n)$ are the transmit power of MS_i and $(C/I)_i$ at step n , respectively.

12.2 Capacity of Cellular CDMA

Numerous studies have investigated the capacity and performance of CDMA cellular systems for radio propagation environments characterized by path loss and shadowing, including Gilhousen et al. [139], Kudoh and Matsumoto [186], and Newson and Heath [242]. Mokhtar and Gupta [229] considered reverse channel capacity on shadowed Nakagami fading channels, where the desired and interfering signals have the same statistical fading characteristics.

CDMA cellular systems employ universal frequency reuse, where the bandwidth is shared by all the cells and users are distinguished through the assignment of unique spreading sequences. For such systems, multiple-access interference from both in-cell and out-of-cell sources must be carefully accounted for. The propagation path loss associated with out-of-cell interferers is relatively small due to the short reuse distance and, hence, the associated interference is significant. With cellular CDMA systems, any technique that reduces or mitigates multiple-access interference translates directly into a capacity gain. For this reason CDMA systems use voice activity detection along with variable rate speech coding, where the rate of the speech coder is reduced when silent periods are detected in the speech waveform.

Our analysis of cellular CDMA capacity assumes 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 cell sector. For hexagonal cells of radius R , this yields a subscriber density of

$$\rho = \frac{2K}{3\sqrt{3}R^2} \text{ per unit area.} \quad (12.9)$$

The effects of variable rate speech coding can be modeled by assuming that each transmitter is independently active at any time 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 .

12.2.1 Reverse Link Capacity

Our analysis assumes a simple closed-loop reverse channel power control scheme that equalizes the received power C at a BS from all MSs being served by that BS. Let $\text{MS}_{i,j}$ denote the j th MS located in cell i . The power transmitted by $\text{MS}_{i,j}$, located at distance $d_{i,j}$ from its serving BS, BS_i , is $P_{i,j}$. The received power at BS_i is

$$C = P_{i,j} 10^{\frac{\zeta_{i,j}}{10}}, \quad (12.10)$$

where $\zeta_{i,j}$ is a random gain due to path loss, shadowing, and fading. $\text{MS}_{i,j}$ is also located at distance $d_{i,0}$ to the reference BS, BS_0 , and will produce an out-of-cell interference equal to

$$\begin{aligned} \frac{I_o(i,j)}{C} &= 10^{\zeta_{i,0}/10} \cdot \left(\frac{1}{10^{\zeta_{i,j}/10}} \right) \\ &= 10^{(\zeta_{i,0} - \zeta_{i,j})/10} \leq 1. \end{aligned} \quad (12.11)$$

The first term is due to path loss, shadowing, and fading to BS_0 , while the second term is the effect of the power control to compensate for the random attenuation between $\text{MS}_{i,j}$ and BS_j . Note that $I_o(i,j)/C$ is always less than unity; otherwise, the MS will execute a handoff to another BS that will make it less than unity.

For our purpose, log-normal shadowing with Nakagami fading is assumed, where the received squared-envelope has the composite *Gamma*-log-normal pdf in (2.373). The composite *Gamma*-log-normal pdf can be approximated by a purely

log-normal pdf with mean and standard deviation given by (2.374). Hence, the random variables $\varsigma_{i,j}$ and $\varsigma_{i,0}$ are treated as Gaussian random variables with means and variances, respectively,

$$\begin{aligned}\mu_{i,j} \text{ (dBm)} &= \xi^{-1}(\psi(m_{i,j}) - \ln(m_{i,j})) + \mu_{\Omega_p \text{ (dBm)}}(i,j) \\ \mu_{i,0} \text{ (dBm)} &= \xi^{-1}(\psi(m_{i,0}) - \ln(m_{i,0})) + \mu_{\Omega_p \text{ (dBm)}}(i,0) \\ \sigma_{i,j}^2 &= \xi^{-2}\zeta(2, m_{i,j}) + \sigma_{\Omega}^2 \\ \sigma_{i,0}^2 &= \xi^{-2}\zeta(2, m_{i,0}) + \sigma_{\Omega}^2,\end{aligned}\tag{12.12}$$

where $\xi = (\ln 10)/10$, $m_{i,j}$ and $m_{i,0}$ are the Nakagami shape factors, and σ_{Ω} is the shadow standard deviation. The parameters $\mu_{\Omega_p \text{ (dBm)}}(i,j)$ and $\mu_{\Omega_p \text{ (dBm)}}(i,0)$ are determined by the path loss. Using the simple path loss model in (1.6), their difference is

$$\mu_{\Omega_p \text{ (dBm)}}(i,j) - \mu_{\Omega_p \text{ (dBm)}}(i,0) = 10\beta \log_{10} \{d_{i,j}/d_{i,0}\}.\tag{12.13}$$

The total out-of-cell interference-to-signal ratio is equal to

$$\frac{I_o}{C} = \int \int \chi I_o(i,j) \Phi_{0,j} \rho dA,\tag{12.14}$$

where

$$\Phi_{0,j} = \begin{cases} 1, & \text{if } 10^{(\varsigma_{i,0} - \varsigma_{i,j})/10} < 1 \\ 0, & \text{otherwise} \end{cases},\tag{12.15}$$

ρ is the user density over the area A , and χ is the voice activity variable

$$\chi = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1 - p \end{cases}.\tag{12.16}$$

The total out-of-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

$$\begin{aligned}E[I_o/C] &= \int \int E[\chi 10^{(\varsigma_{i,0} - \varsigma_{i,j})/10} \Phi_{0,j}] \rho dA \\ &= \int \int p E[10^{(\varsigma_{i,0} - \varsigma_{i,j})/10} \Phi_{0,j}] \rho dA.\end{aligned}\tag{12.17}$$

Let $x = \varsigma_{i,0} - \varsigma_{i,j}$ and define

$$\begin{aligned}\mu_x \text{ (dB)} &= \mu_{i,0} \text{ (dBm)} - \mu_{i,j} \text{ (dBm)} \\ \sigma_x^2 &= \sigma_{i,0}^2 + \sigma_{i,j}^2.\end{aligned}\tag{12.18}$$

The expectation in the integrand of (12.17) is

$$\begin{aligned}E[10^{(\varsigma_{i,0} - \varsigma_{i,j})/10} \Phi_{0,j}] &= \int_{-\infty}^0 e^{\xi x} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{(x - \mu_x \text{ (dB)})^2}{2\sigma_x^2}\right\} dx \\ &= \exp\left\{\frac{\xi^2\sigma_x^2}{2} + \xi\mu_x \text{ (dB)}\right\} \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^0 \exp\left\{-\left(\frac{x}{\sqrt{2}\sigma_x} - \frac{\xi\sigma_x}{\sqrt{2}} - \frac{\mu_x \text{ (dB)}}{\sqrt{2}\sigma_x}\right)^2\right\} dx \\ &= \exp\left\{\frac{\xi^2\sigma_x^2}{2} + \xi\mu_x \text{ (dB)}\right\} \left(1 - Q\left(-\xi\sigma_x - \frac{\mu_x \text{ (dB)}}{\sigma_x}\right)\right).\end{aligned}\tag{12.19}$$

Therefore,

$$E[I_o/C] = p \int \int \exp \left\{ \frac{\xi^2 \sigma_x^2}{2} + \xi \mu_x \text{ (dB)} \right\} \left(1 - Q \left(-\xi \sigma_x - \frac{\mu_x \text{ (dB)}}{\sigma_x} \right) \right) \rho dA. \quad (12.20)$$

In a similar fashion,

$$\begin{aligned} E[(I_o/C)^2] &= \int \int E[\chi^2 10^{(\zeta_{i0} - \zeta_{ij})/5} \Phi_{0j}^2] \rho dA \\ &= p \int \int \int_{-\infty}^0 e^{2\xi x} \frac{1}{\sqrt{2\pi}\sigma_x} \exp \left\{ -\frac{(x - \mu_x \text{ (dB)})^2}{2\sigma_x^2} \right\} dx \rho dA \\ &= p \int \int \exp \{ 2\xi^2 \sigma_x^2 + 2\xi \mu_x \text{ (dB)} \} \left(1 - Q \left(-2\xi \sigma_x - \frac{\mu_x \text{ (dB)}}{\sigma_x} \right) \right) \rho dA. \end{aligned} \quad (12.21)$$

Finally, the variance of I_o/C is

$$\begin{aligned} \text{Var}[I_o/C] &= p \int \int \exp \{ 2\xi^2 \sigma_x^2 + 2\xi \mu_x \text{ (dB)} \} \left(1 - Q \left(-2\xi \sigma_x - \frac{\mu_x \text{ (dB)}}{\sigma_x} \right) \right) \rho dA \\ &\quad - E^2[I_o/C]. \end{aligned} \quad (12.22)$$

The integrals in (12.20) and (12.22) must be evaluated numerically over the random location of mobile locations in the area A , as defined by the set of shaded sectors shown in Fig. 12.1 (minus the sector where the reference MS is located) where only the adjacent cells are shown.

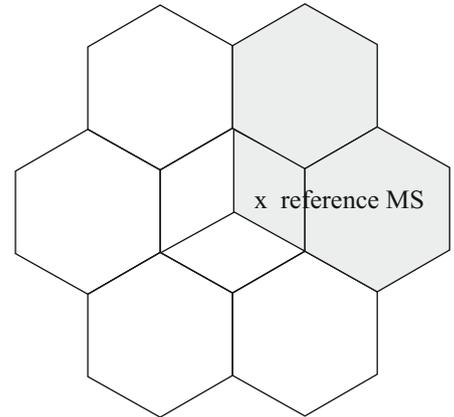
With perfect power control, the in-cell interference I_{in} is

$$I_{in} = C \sum_{i=1}^{K-1} \chi_i, \quad (12.23)$$

where χ_i is a Bernoulli random variable equal to 1 with probability p and 0 with probability $1 - p$. Let $I = I_o + I_{in}$ be the total interference. The carrier-to-interference ratio C/I and baseband E_b/N_o are related by

$$C/I = \frac{E_b/N_o}{W_{ss}/W}, \quad (12.24)$$

Fig. 12.1 Reverse channel transmissions from MSs located in the shaded area will cause out-of-cell interference with the reverse channel transmission from the reference MS



where E_b is bit energy, N_o is the one-sided noise power spectral density, W_{ss} is the spread spectrum bandwidth, and W is the data signal bandwidth. Many factors must be considered to establish the required $\gamma_b = E_b/N_o$, $\gamma_{b \text{ req}}$, such as the channel characteristics, type of modulation, diversity, receiver structure, and coding/interleaving techniques. In the cdma2000 cellular system with radio configuration (RC1), $\gamma_{b \text{ req}} = 7$ dB on the reverse link and $\gamma_{b \text{ req}} = 5$ dB on the forward link [139]. The difference in $\gamma_{b \text{ req}}$ for the forward and reverse links arises, because the reverse link of cdma2000 RC1 uses non-coherent detection while the forward link uses coherent detection.

From (12.24),

$$(I/C)_{\gamma_b} = \frac{W_{ss}/W}{\gamma_b}. \quad (12.25)$$

It follows that the probability that the received γ_b at a BS is below a required value, $\gamma_{b \text{ req}}$, is

$$\begin{aligned} P_{\text{out}} &= \text{P}[\gamma_b < \gamma_{b \text{ req}}] \\ &= \text{P}[(I/C)_{\gamma_b} > (I/C)_{\gamma_{b \text{ req}}}] \\ &= \text{P}\left[\sum_{i=1}^{K-1} \chi_i + I_o/C > (I/C)_{\gamma_{b \text{ req}}}\right] \\ &= \text{P}\left[I_o/C > (I/C)_{\gamma_{b \text{ req}}} - \sum_{i=1}^{K-1} \chi_i\right] \\ &= \sum_{k=0}^{K-1} \text{P}\left[I_o/C > (I/C)_{\gamma_{b \text{ req}}} - k \mid \sum_{i=1}^{K-1} \chi_i = k\right] \text{P}\left[\sum_{i=1}^{K-1} \chi_i = k\right] \\ &= \sum_{k=0}^{K-1} \binom{K-1}{k} p^k (1-p)^{K-1-k} Q\left(\frac{(I/C)_{\gamma_{b \text{ req}}} - k - \text{E}[I_o/C]}{\sqrt{\text{Var}[I_o/C]}}\right). \end{aligned} \quad (12.26)$$

Assume a chip rate of $W_{ss} = 1.25$ MHz, $W = 8$ kHz, $p = 3/8$, and a required outage probability of $P_{\text{out}} = 10^{-2}$. Figure 12.2 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. Figure 12.3 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. Figure 12.4 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 mostly affect the out-of-cell interference.

The ratio of the mean out-of-cell interference to the mean in-cell interference is

$$\theta = \frac{\text{E}[I_o]}{\text{E}[I_{\text{in}}]} = \frac{\text{E}[I_o/C]}{\text{E}[I_{\text{in}}/C]} = \frac{\text{E}[I_o/C]}{pK}. \quad (12.27)$$

With a 4th-order path loss exponent, Newson and Heath [242] 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 12.1 tabulates the corresponding values of θ and $f = 1/(1 + \theta)$ 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.

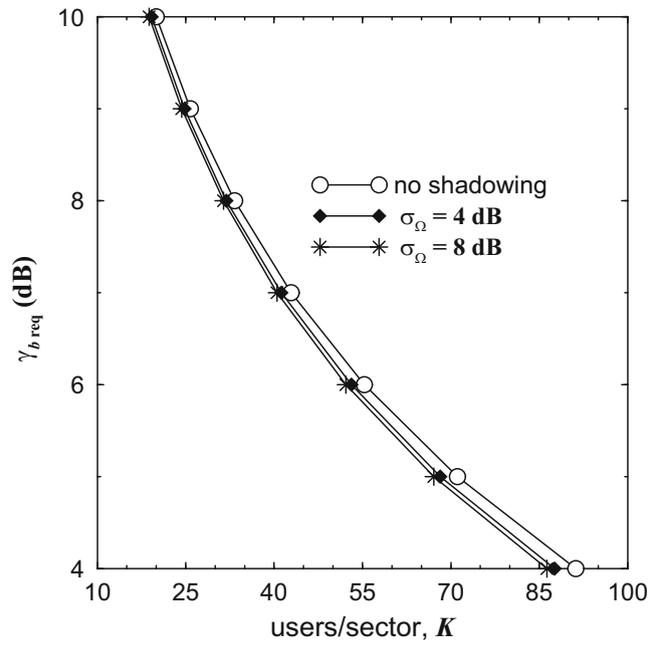


Fig. 12.2 Reverse channel capacity with $P_{\text{out}} = 10^{-2}$ for different $\gamma_{b, \text{req}}$; the Nakagami shape factors are $m_d = 1$ and $m_l = 1$ (Rayleigh fading), and σ_{Ω} is the shadow standard deviation

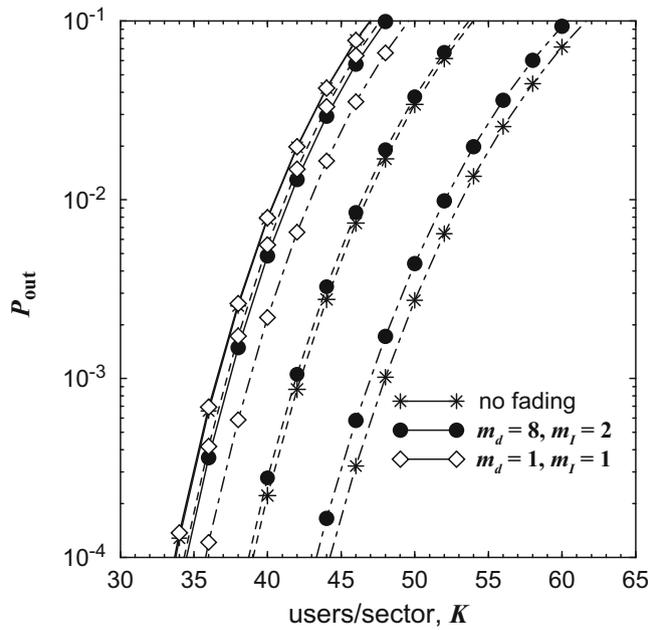


Fig. 12.3 Reverse channel capacity for different propagation environments with $\gamma_{b, \text{req}} = 7$ dB. Solid lines denote $\sigma_{\Omega} = 0$ dB, dotted lines denote $\sigma_{\Omega} = 4$ dB, and dot-dashed lines denote $\sigma_{\Omega} = 8$ dB

To show that the values of θ and f in Table 12.1 do not depend on the number of users per cell, K , the cdf of the out-of-cell interference to the in-cell interference I_o/I_{in} for the reverse channel is derived as

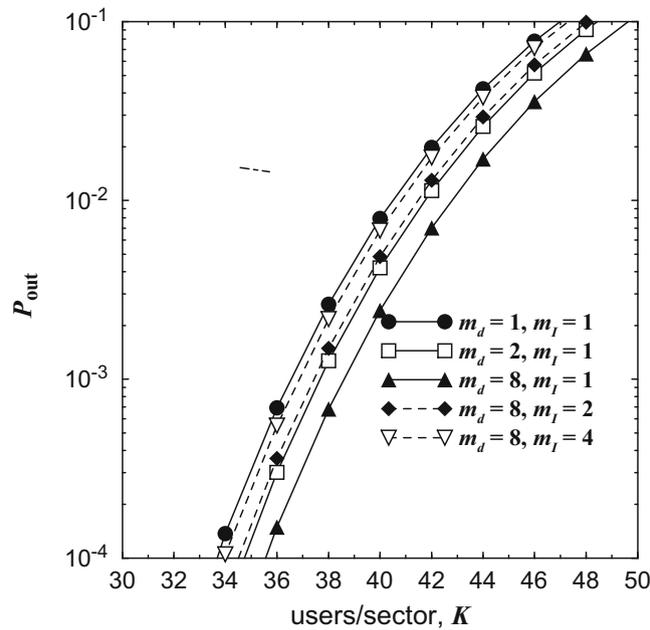


Fig. 12.4 Reverse channel capacity for different Nakagami shape factors with $\gamma_{b \text{ req}} = 7$ dB, $\sigma_{\Omega} = 8$ dB

Table 12.1 Ratio of the mean of out-of-cell interference to the mean in-cell interference, θ , and frequency reuse efficiency, f , under different propagation conditions

m_d	m_l	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	66.18
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

$$\begin{aligned}
 P[I_o/I_{in} < z] &= P\left[\frac{I_o/C}{I_{in}/C} < z\right] \\
 &= 1.0 - P[I_o/C > z I_{in}/C] \\
 &= 1.0 - \sum_{k=0}^{K-1} \binom{K-1}{k} \eta^k (1-\eta)^{K-1-k} Q\left(\frac{kz - E[I_o/C]}{\sqrt{\text{Var}[I_o/C]}}\right).
 \end{aligned} \tag{12.28}$$

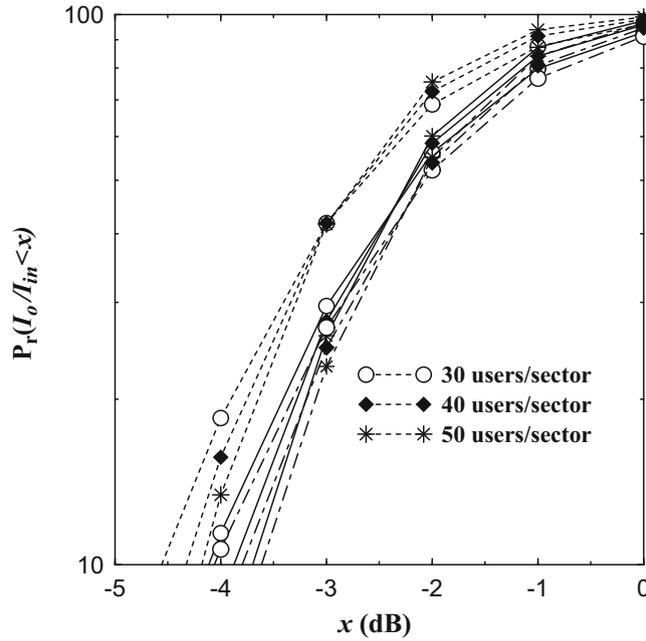


Fig. 12.5 Distribution of out-of-cell interference to in-cell interference, I_o/I_{in} , for the reverse channel. *Solid lines* denote $\sigma_\Omega = 0$ dB, *dotted lines* denote $\sigma_\Omega = 4$ dB, and *dot-dashed lines* denote $\sigma_\Omega = 8$ dB; $m_d = 1, m_l = 1$ (Rayleigh fading)

Figure 12.5 plots the distribution of I_o/I_{in} (in dB) with different shadow standard deviations. Although the distribution varies with K , the mean value $E[I_o/I_{in}]$ 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 12.1 do not depend on K .

12.2.2 Forward Link Capacity

For the forward channel, a pilot signal is transmitted from each BS. The pilot signal causes interference in every cell, thereby reducing the capacity. However, this is offset by a decrease in $\gamma_{b \text{ req}}$ due to coherent demodulation. With forward channel balancing power control, the mobile measures the received signal and periodically transmits the measurement to its serving BS [139]. 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.

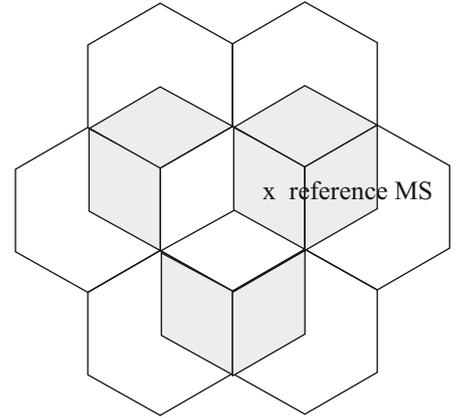
In the worst case situation, each BS always transmits with the maximum allowable power P_{\max} . From (12.24), the γ_b at the i th mobile under this condition is

$$\gamma_{b,i} = \frac{W_{ss}/W}{(I/C)_i} = \frac{W_{ss}/W}{(\sum_{j=0}^M C_{T_j} - \delta\phi_i C_{T_0})/\delta\phi_i C_{T_0}} \quad (12.29)$$

$$C_{T_j} = P_{\max} 10^{\frac{\zeta_j}{10}}, \quad (12.30)$$

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 i th mobile. Forward channel transmissions to MSs located in the shaded sectors of Fig. 12.6 will cause multiple-access interference with the forward channel transmission to the reference MS. As in [139] our results assume that 20% of the total BS transmit power is allocated to the pilot signal. Once again, the ζ_i are Gaussian random variables due to shadow and fading variations, with means and variances obtained from (2.374).

Fig. 12.6 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 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 (12.29). To account for the random voice activity, a modified weighting factor is then calculated as [186]

$$\bar{\phi}_i = \frac{(\phi_i)_{\text{req}}}{\sum_{\substack{j=1 \\ j \neq i}}^K \chi_j (\phi_j)_{\text{req}} + (\phi_i)_{\text{req}}} \quad (12.31)$$

The power balancing scheme in [56] does the same thing, except that the voice activity factors, χ_j , are not considered. The outage probability then becomes

$$\begin{aligned} P_{\text{out}} &= \text{P}[\gamma_b < \gamma_{b \text{ req}}] \\ &= \text{P}[\bar{\phi}_i < (\phi_i)_{\text{req}}]. \end{aligned} \quad (12.32)$$

Numerical results can be obtained from the last equation in (12.32) 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, the required fraction of power, $(\phi_i)_{\text{req}}$, needed to meet the $\gamma_{b \text{ req}}$ requirement is first determined. Afterwards, it is determined if the actual power allocation for each MS, $\bar{\phi}_i$, is sufficient.

Figure 12.7 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. Figure 12.8 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.

12.2.3 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 [186, 242]. For the reverse channel, (12.10) has the modified form

$$C10^{\zeta_{E_j}/10} = P_{i,j}10^{\zeta_{d}/10}, \quad (12.33)$$

where ζ_{E_j} is the power control error. The mean and variance of $x = \zeta_{i,0} - \zeta_{i,j}$ with imperfect power control are similar to (12.18), but have the form

$$\mu_x \text{ (dB)} = \mu_{i,0} \text{ (dBm)} - \mu_{i,j} \text{ (dBm)} \quad (12.34)$$

$$\sigma_x^2 = \sigma_{i,0}^2 + \sigma_{i,j}^2 + \sigma_E^2. \quad (12.35)$$

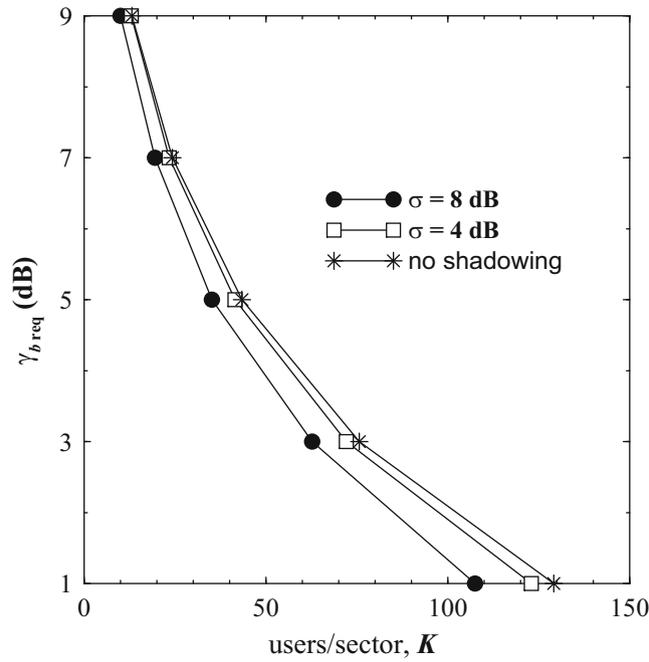


Fig. 12.7 Forward channel capacity with $P_{out} = 10^{-2}$ for different $\gamma_{b, req}$; the Nakagami shape factors are $m_d = 1$ and $m_l = 1$ (Rayleigh fading), and σ_{Ω} is the shadow standard deviation

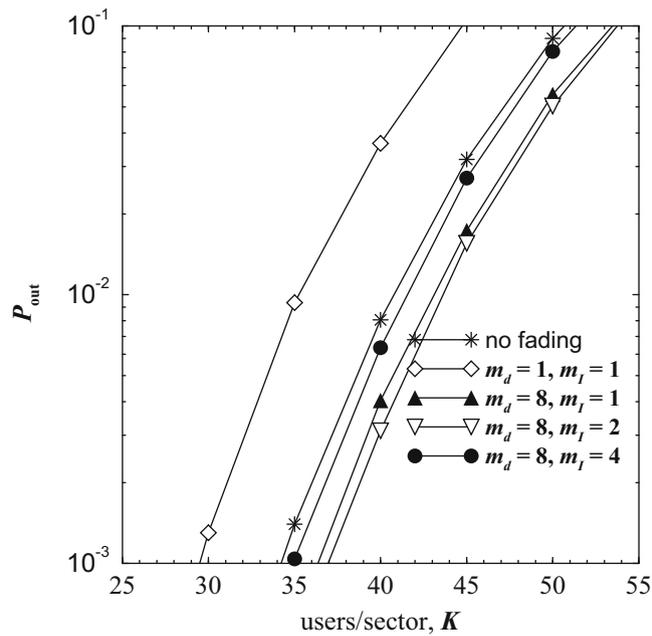


Fig. 12.8 Forward channel capacity for different Nakagami shape factors with $\gamma_{b, req} = 5$ dB, $\sigma_{\Omega} = 8$ dB

With imperfect power control, the in-cell interference experienced by $MS_{0,i}$ at the reference BS, BS_0 , is $I_{in} = C\kappa$, where

$$\kappa = \sum_{\substack{j=1 \\ j \neq i}}^K 10^{SE_j/10} \chi_j. \tag{12.36}$$

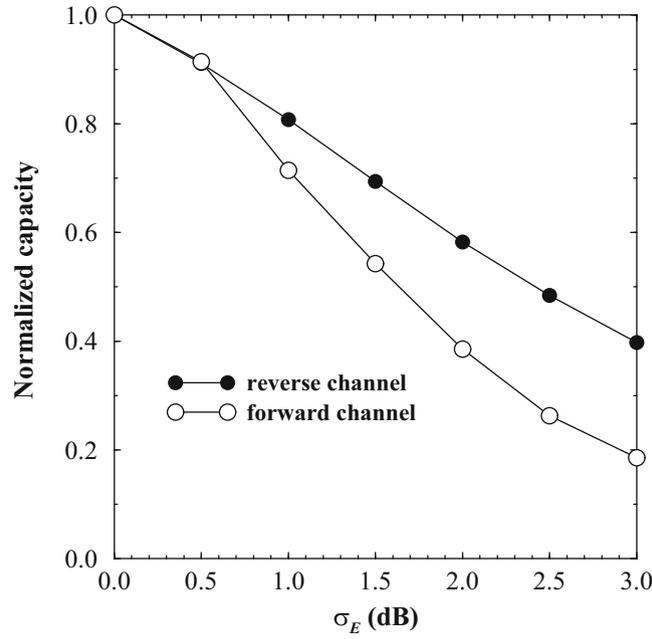


Fig. 12.9 Forward and reverse channel capacity with imperfect power control. The capacity is normalized with respect to the capacity with perfect power control; σ_E is the standard deviation of the power control error, $m_d = 1$, $m_f = 1$, and $\sigma_\Omega = 8$ dB

Then

$$\begin{aligned}
 P_{\text{out}} &= \text{P}[\gamma_b < \gamma_{b \text{ req}}] \\
 &= \text{P}[I_o/C + I_{\text{in}}/C > (I/C)\gamma_{b \text{ req}}] \\
 &= \text{P}[I_o/C > (I/C)\gamma_{b \text{ req}} - \kappa] \\
 &= \sum_{k=0}^{K-1} \binom{K-1}{k} p^k (1-p)^{K-1-k} \int_0^\infty \text{P}[I_o/C > (I_o/C)\gamma_{b \text{ req}} - \kappa | \kappa] p(\kappa|k) d\kappa.
 \end{aligned} \tag{12.37}$$

Note that the conditional pdf of κ given k , $p(\kappa|k)$, is approximately log-normal. The log-normal approximation can be calculated using the Fenton-Wilkinson or Schwartz & Yeh methods discussed in Sect. 3.1. Observe from Fig. 12.9 that the reverse channel capacity is dramatically decreased as the power control error increases. For $P_{\text{out}} = 0.01$ and power control errors of $\sigma_E = 1$ dB, 2 dB, 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, (12.31) becomes

$$\bar{\phi}_i = \frac{\phi_i 10^{\sigma E_i/10}}{\sum_{\substack{j=1 \\ j \neq i}}^K \chi_j \phi_j 10^{\sigma E_j/10} + \phi_i 10^{\sigma E_i/10}}. \tag{12.38}$$

Figure 12.9 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.

12.3 Hierarchical Macrodiversity CDMA Cellular Architectures

CDMA cellular systems often employ hierarchical architectures consisting of macrocells with underlaid microcells. In such architectures, macrocells cover large areas with sparse traffic densities, whereas microcells serve small areas with high traffic densities. However, due to their effective frequency reuse factor of one, hierarchical CDMA systems must deal with cross

interference between the hierarchical layers. This cross-layer interference can be subdued by assigning distinct spectrum to each layer, but this will make inefficient use of the available spectrum. A better approach allows the hierarchical layers to share the same spectrum.

For the reverse link, one possibility is hierarchical maximum ratio combining (HMRC), where the signal from each mobile station (MS) is received by several base stations (BSs) in both hierarchical layers and coherently combined. If independent interference is assumed at the different BS locations, the combined CIR is the algebraic sum of the CIRs at each BS,

$$(C/I)_{\text{HMRC}} = \sum_{i=1}^N (C/I)_{\mu,i} + \sum_{j=1}^M (C/I)_{\text{M},j}, \quad (12.39)$$

where N and M are the number of BSs involved in the combining, and the subscripts μ and M designate “microcell” and “macrocell,” respectively.

Macrodiversity maximal ratio combining (MMRC) has been proven to be an effective way of improving the capacity in cellular CDMA systems [146, 159]. In [159], Hanly proved the existence of a power control solution using MMRC, and showed that the capacity is unaffected by outside interference. In [146], by assuming equal reverse interference level at each BS in non-hierarchical settings, Gorricho and Paradells constructed a simple proof showing that MMRC reverse link capacity is close to an isolated cell capacity. Here, the results in [146] are generalized and applied to hierarchical CDMA systems. An analytical solution will be derived for HMRC reverse performance without assuming an equal level of reverse link interference among cells, and show that both microcell and macrocell performances are nearly unaffected by each other’s presence.

The reverse link is commonly considered to limit the CDMA system capacity. However, with the emergence of asymmetric wireless data services, the forward link performance has become increasingly important. For the forward link, HMRC-like combining schemes are not suitable because such schemes will increase the forward link interference [168]. One possible solution is Selective Transmit Diversity (STD) where each BS provides multiple transmit paths by means of spatially separated antennas, and the system allows each MS to connect to the most robust path among the multiple paths [278, 284].

12.3.1 System Model

Consider a hierarchical CDMA deployment consisting of a group of omnidirectional macrocells and a cluster of omnidirectional microcells embedded within the macrocells. The MSs are assumed to be uniformly distributed in both macrocells and microcells, yet the load condition of each cell might differ. Although no particular load conditions are assumed, microcells are generally more densely populated by MSs than macrocells. With macrodiversity, there are no longer distinct boundaries between cells and hierarchical layers. Therefore, MSs are referenced to their respective locations. For example, a microcell MS means that the MS is physically located in the designated microcell area, but not necessarily served by it.

The radio links are assumed to be affected by Rayleigh fading and log-normal shadowing. The composite distribution of the link gain, G , is given by the Gamma-log-normal density in (2.373) with $m = 1$, i.e.,

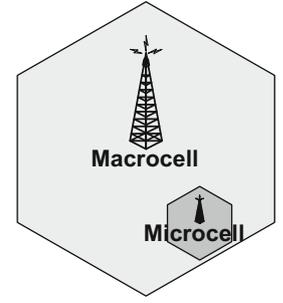
$$f_G(x) = \int_0^\infty \frac{1}{\Omega_p} e^{-x/\Omega_p} \frac{1}{\sqrt{2\pi\xi\sigma_\Omega\Omega_p}} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p(\text{dBm})})^2}{2\sigma_\Omega^2}\right\} d\Omega_p, \quad (12.40)$$

where σ_Ω is the shadow standard deviation, $\xi = (\ln 10)/10$ and $\mu_{\Omega_p(\text{dBm})}$ is the area mean that depends on the path loss model. Here, the simple path loss model

$$\mu_{\Omega_p(\text{dBm})}(d) = \mu_{\Omega_p(\text{dBm})}(d_o) - 10\beta \log_{10}\{d/d_o\} \quad (\text{dBm}) \quad (12.41)$$

is used, where $\mu_{\Omega_p(\text{dBm})}(d_o)$ is the area mean at a reference distance d_o , d is the distance between the MS and BS, β is the path loss exponent. As shown in (2.374), the composite Gamma-log-normal distribution in (12.40) can be approximated by a log-normal distribution with parameters

$$\begin{aligned} \mu_{(\text{dBm})} &= \mu_{\Omega_p(\text{dBm})} - 2.50675 \\ \sigma^2 &= \sigma_\Omega^2 + 31.0215. \end{aligned} \quad (12.42)$$

Fig. 12.10 Single cell model

In other words, the link gain G in decibels, denoted by $\zeta = 10\log_{10}\{G\}$ is normally distributed with the mean $\mu_{(\text{dBm})}$ and variance σ^2 in (12.42).

12.3.2 Reverse Link Analysis

First consider a simple single macrocell and microcell system to introduce the method of HMRC analysis. Afterwards, the approach will be extended to a multi-cell system.

12.3.2.1 Single Cell Model

Consider a single microcell embedded within a macrocell as shown in Fig. 12.10. Using HMRC, the reverse link CIR of MS_i is

$$\begin{aligned} (C/I)_i &= (C/I)_{\mu,i} + (C/I)_{\text{M},i} \\ &= \frac{C_{\mu,i}}{I_{\mu,i}} + \frac{C_{\text{M},i}}{I_{\text{M},i}}, \end{aligned} \quad (12.43)$$

where $C_{\mu,i}$ and $C_{\text{M},i}$ are the received signal power at the Microcell BS and Macrocell BS, respectively. Let $\lambda_i = I_{\mu,i}/I_{\text{M},i}$ be the ratio of the microcell interference to macrocell interference. Then $(C/I)_i$ becomes

$$\begin{aligned} (C/I)_i &= \frac{C_{\mu,i} + \lambda_i C_{\text{M},i}}{I_{\mu,i}} \\ &= \frac{(G_{\mu,i} + \lambda_i G_{\text{M},i})P_i}{I_{\mu,i}} \\ &= \frac{C_{\text{reverse},i}}{I_{\mu,i}}, \end{aligned} \quad (12.44)$$

where $G_{\mu,i}$ and $G_{\text{M},i}$ are the reverse link gains associated with the Microcell BS and Macrocell BS.

The HMRC reverse power control algorithm controls each MS transmit power, P_i , so that all MS experience a uniform CIR level. The convergence of such power control policy has been proven in [364]. Let us assume that the Microcell BS and Macrocell BS both serve a large number of MSs such that the microcell and macrocell interference levels experienced by each MS are nearly the same, i.e.,

$$I_{\mu,i} \approx I_{\mu}, \quad I_{\text{M},i} \approx I_{\text{M}}, \quad \forall i. \quad (12.45)$$

Since the interference power is the difference between the total received power and the desired signal power, the differences in the desired signal components have minimal effect on interference values when the system load is relatively large. This also suggests that the variation in λ_i is minimal. The above assumption is justified in the numerical results section that follows

below. Based on the assumption in (12.45), HMRC power control now results in all MSs having the same uniform combined signal power, C_{reverse} .

Let N and M be the numbers of MSs located in the Microcell and Macrocell, respectively, and express $I_{\mu,i}$ and $I_{M,i}$ as follows:

$$\begin{aligned} I_{\mu,i} &= \sum_{j \neq i}^N C_{\mu,j|\mu} + \sum_{k=1}^M C_{\mu,k|M} \\ I_{M,i} &= \sum_{j \neq i}^N C_{M,j|\mu} + \sum_{k=1}^M C_{M,k|M}, \quad i \in \text{Microcell} \end{aligned} \quad (12.46)$$

and

$$\begin{aligned} I_{\mu,i} &= \sum_{j=1}^N C_{\mu,j|\mu} + \sum_{k \neq i}^M C_{\mu,k|M} \\ I_{M,i} &= \sum_{j=1}^N C_{M,j|\mu} + \sum_{k \neq i}^M C_{M,k|M}, \quad i \in \text{Macrocell}, \end{aligned} \quad (12.47)$$

where $C_{BSj,i|BSk}$ is the received signal power at BS j from MS i given that MS i is located in the cell served by BS k , and BS j , $BSk \in \{\mu, M\}$. First consider the case for the Microcell MS. From (12.44), it can be deduced that $C_{M,i} = (C_{\text{reverse}} - C_{\mu,i})/\lambda_i$. Therefore,

$$\begin{aligned} \lambda_i &= \frac{I_{\mu,i}}{I_{M,i}} \\ &= \frac{I_{\mu,i}}{\sum_{j \neq i}^N C_{M,j|\mu} + \sum_{k=1}^M C_{M,k|M}} \\ &= \frac{I_{\mu,i}}{\sum_{j \neq i}^N (C_{\text{reverse}} - C_{\mu,j|\mu})/\lambda_j + \sum_{k=1}^M (C_{\text{reverse}} - C_{\mu,k|M})/\lambda_k} \\ &\approx \frac{\lambda_i I_{\mu,i}}{\sum_{j \neq i}^N (C_{\text{reverse}} - C_{\mu,j|\mu}) + \sum_{k=1}^M (C_{\text{reverse}} - C_{\mu,k|M})}, \\ &= \frac{\lambda_i I_{\mu,i}}{(N + M - 1)C_{\text{reverse}} - I_{\mu,i}}. \end{aligned} \quad (12.48)$$

where the approximation in the second last line uses the fact that $\lambda_j \approx \lambda_k \forall j, k$. Solving the above equation for $I_{\mu,i}$ gives

$$I_{\mu,i} \approx \frac{(N + M - 1)C_{\text{reverse}}}{2}. \quad (12.49)$$

By using the similar approach, one can see that the macrocell MS yields the same result. Then, the reverse link CIR can be approximated as follows:

$$(C/I)_i = \frac{C_{\text{reverse}}}{I_{\mu,i}} \approx \frac{2}{N + M - 1}. \quad (12.50)$$

Some important observations about HMRC can be made from (12.50). First, the CIR performance is independent of the microcell location. Without HMRC, the overall performance suffers from increased level of inter-layer cross interference in cases where the microcell is closer to the macrocell BS. For HMRC, however, the combining effect is directly related to the proximity of the two BSs. Therefore, the increase in combining effect compensates for the increase in interference due to the microcell. Another noteworthy observation is that the HMRC performance is only limited by the overall system load, $N + M$, and not by individual cell loads. An overloaded microcell does not affect the system performance as long as the overall system load is kept under check, whereas it can dictate the system performance for non-HMRC systems. This suggests that HMRC is an effective way to share available resources between hierarchical layers.

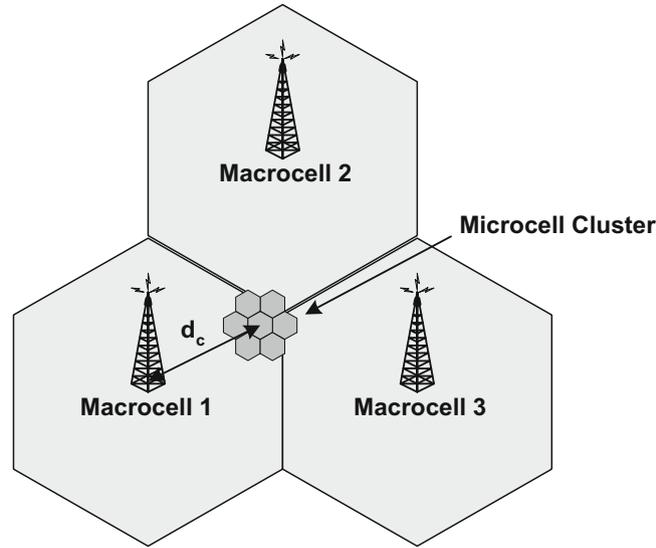


Fig. 12.11 Multiple cell model

12.3.2.2 Multiple Cell Model

The above analysis is now extended to multiple cell environments. Our multi-cell model consists of three macrocells and a cluster of K microcells embedded within the macrocells as shown in Fig. 12.11. Then,

$$\begin{aligned}
 (C/I)_i &= (C/I)_{\mu 1,i} + \cdots + (C/I)_{\mu K,i} + (C/I)_{M1,i} + \cdots + (C/I)_{M3,i} \\
 &= \frac{C_{\mu 1,i} + \cdots + \lambda_{\mu K,i} C_{\mu K,i} + \lambda_{M1,i} C_{M1,i} + \cdots + \lambda_{M3,i} C_{M3,i}}{I_{\mu 1,i}} \\
 &= \frac{C_{\text{reverse},i}}{I_{\mu 1,i}}, \tag{12.51}
 \end{aligned}$$

where $\lambda_{BSj,i} = I_{\mu 1,i}/I_{BSj,i}$. Let N_i and M_j be the numbers of MSs in Microcell i and Macrocell j , respectively. Then, for the Microcell 1 MSs,

$$\begin{aligned}
 I_{M1,i} &= \sum_{j \neq i}^{N_1} C_{M1,j} |_{\mu 1} + \cdots + \sum_{l=1}^{N_K} C_{M1,l} |_{\mu K} + \sum_{p=1}^{M_1} C_{M1,p} |_{M1} + \cdots + \sum_{r=1}^{M_3} C_{M1,r} |_{M3} \\
 &\approx (N_1 + \cdots + N_K + M_1 + M_2 + M_3 - 1) \frac{C_{\text{reverse}}}{\lambda_{M1,i}} - (2 + K) \frac{I_{\mu 1,i}}{\lambda_{M1,i}}. \tag{12.52}
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 I_{\mu 1,i} &\approx \frac{(N_1 + \cdots + N_K + M_1 + M_2 + M_3 - 1) C_{\text{reverse}}}{3 + K} \\
 (C/I)_i &\approx \frac{3 + K}{N_1 + \cdots + N_K + M_1 + M_2 + M_3 - 1}. \tag{12.53}
 \end{aligned}$$

Using MSs in other cells with their corresponding λ s yields the same result. The same important observations can be made as in the single cell case for multi-cell case. Note also that if both microcells and macrocells are loaded with an equal number of MSs, the result in (12.53) is the same as the result obtained in [146]. This tells us that the macrocell capacity is nearly unaffected by the introduction of microcells when inter-layer HMRC is allowed.

Table 12.2 Single cell model result

N	M	$CIR_{anal.}$ (dB)	$CIR_{sim.}$ (dB)
10	10	-9.60	-9.78
15	10	-10.66	-10.79
20	15	-12.21	-12.30
20	20	-12.82	-12.90

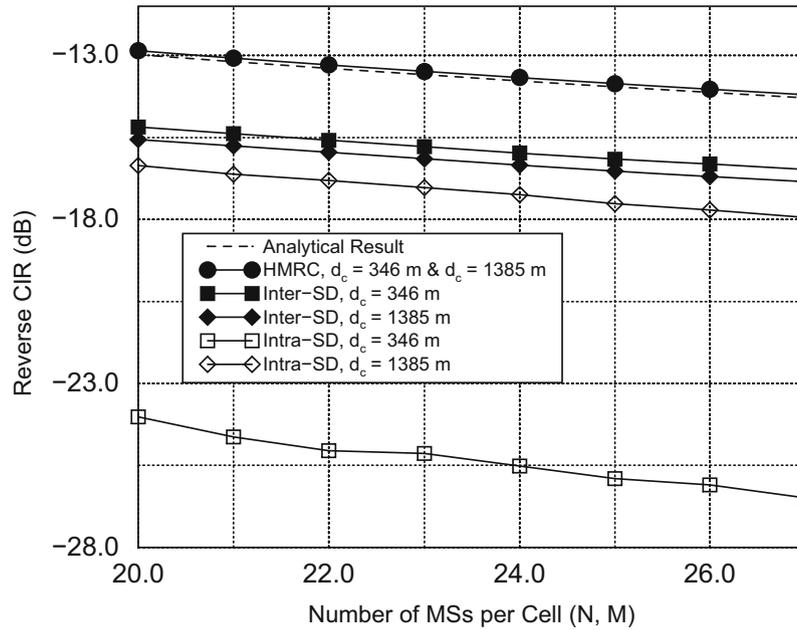


Fig. 12.12 Reverse link CIR performance comparison

12.3.2.3 Numerical Results

Consider the deployment in Fig. 12.10 where the radii of the macrocell and microcell regions are 1500 m and 200 m, respectively. The propagation environment is characterized by a path loss exponent of 4 and a shadow standard deviation of 8 dB. Table 12.2 shows the average performance comparison between analytical and simulation results obtained by using the single cell model. Small deviations between the analytical and simulation results are most likely caused by the equal interference assumption in (12.45) while deriving the analytical solution. The fact is that MSs located close to a BS will experience less interference, while MSs further away from the BS will face higher interference. However, the difference is marginal and it becomes smaller as the system load increases.

Figure 12.12 shows the multi-cell reverse link CIR performance comparison between HMRC and non-HMRC diversity power control schemes at various microcell cluster locations (d_c). A three-macrocell and three-microcell model is used to obtain the simulation results. Both macrocells and microcells are loaded with the same number of MSs. Two non-HMRC diversity schemes are compared; intra-layer selection diversity (Intra-SD) and inter-layer selection diversity (Inter-SD). With Intra-SD the most robust link within each layer is selected, while Inter-SD allows each MS to connect the best BS in any layer. Clearly, HMRC performance is superior to that of non-HMRC schemes. It is observed that both macrocell and microcell capacities are nearly unaffected by each other’s presence (i.e., both macrocells and microcells retain a near isolated cell capacity). The performances of both SD schemes are dependent on the microcell cluster location, d_c . For intra-SD, the cross interference between the layers increases as the microcell cluster gets closer to a macrocell BS, and causes overall system performance degradation. For inter-SD, the diversity gain decreases as d_c increases, which also causes the performance to suffer. However, employing HMRC eliminates the effect of the cluster location on the CIR performance. From (12.51) one can see that the reverse CIR performance of HMRC is proportional to the λ_i and macrocell gains. These two factors balance the performance as the microcell cluster location changes. When d_c is small, the cross interference causes the λ_i to decrease, but the loss is compensated for by the increase in macrocell gains. Conversely, when d_c is large, the macrocell gains decrease while the λ_i increase.

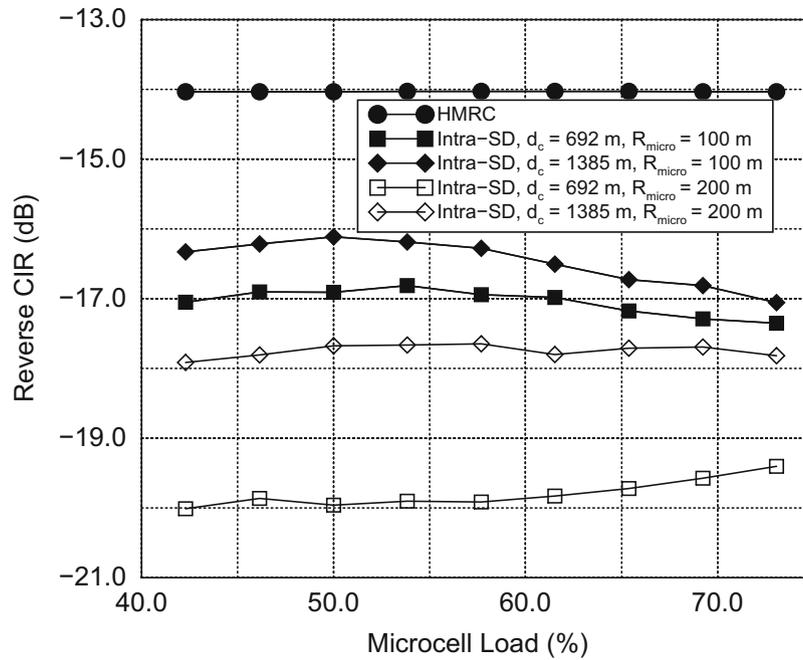


Fig. 12.13 Intra-SD CIR performance against microcell load

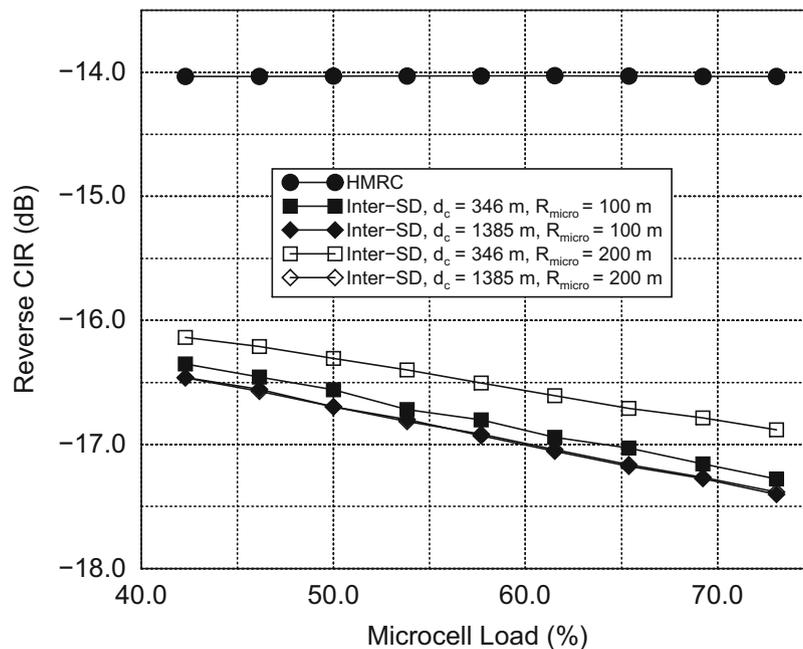


Fig. 12.14 Inter-SD CIR performance against microcell load

The analytical results indicate that the HMRC performance is a function of the overall system load, and does not depend on either the load distribution between the layers or the cell sizes. Figures 12.13 and 12.14 show the effect of microcell load and size on the reverse link CIR performance. The same three-three multi-cell model is used in this simulation also. The overall system load is kept at 156 MSs while the microcell load percentage to the overall load is varied. It is observed that in nearly all instances both non-HMRC schemes suffer performance losses with an increase in the microcell load percentage and microcell radius, R_{micro} . It can be understood that the increase in the overall interference due to an increased microcell load cannot be relieved entirely by a decreased macrocell load. Larger microcell sizes also increase interference, since MSs that belong to microcells need to transmit at higher power levels. The performance of HMRC is not affected by the microcell

load and size changes, as predicted. The increase in microcell interference also increases the λ_i , which offsets the negative effect of the microcell interference. Therefore, HMRC allows flexible resource sharing between hierarchical layers.

Our observations suggest that implementing HMRC gives system planners and administrators almost unlimited freedom and flexibility when contemplating microcell placements. With HMRC, microcell(s) can be placed anywhere within the existing macrocell layer, significant performance and capacity gains can be obtained while guaranteeing robust resource sharing between the layers.

12.3.3 Forward Link Analysis

The forward power control consider here is based on the neighboring-cell pilot power (NPP) scheme proposed in [168], where the forward transmit power to each MS is determined according to link conditions between the MS and surrounding BSs. To see how the NPP scheme can be applied to a hierarchical CDMA architecture, consider a deployment of K microcells surrounded by three macrocells as in Fig. 12.11. Let P_{Mj}^t and $P_{\mu k}^t$ be the total powers transmitted by the BSs serving Macrocell j and Microcell k , respectively. Then, the interference experienced by MS_i can be estimated as

$$I_i \approx G_{M1,i}P_{M1}^t + \cdots + G_{M3,i}P_{M3}^t + G_{\mu 1,i}P_{\mu 1}^t + \cdots + G_{\mu K,i}P_{\mu K}^t. \quad (12.54)$$

Define the power ratio $\gamma_{BSj} = P_{\mu 1}^t/P_{BSj}^t$. Only one BS can transmit to MS_i at any given time and the transmit power is determined by

$$P_{BSj,i} = \frac{\frac{1}{\gamma_{M1}}G_{M1,i} + \cdots + \frac{1}{\gamma_{M3}}G_{M3,i} + G_{\mu 1,i} + \cdots + \frac{1}{\gamma_{\mu K}}G_{\mu K,i}}{G_{BSj,i}}P_T \quad (12.55)$$

when MS_i is connected to BSj and P_T is the predetermined forward transmit power constant. The resulting forward CIR for MS_i is then

$$\begin{aligned} (C/I)_i &= \frac{C_i}{I_i} \\ &\approx \frac{\left(\frac{1}{\gamma_{M1}}G_{M1,i} + \cdots + \frac{1}{\gamma_{M3}G_{M3,i} + G_{\mu 1,i} + \cdots + \frac{1}{\gamma_{\mu K}}G_{\mu K,i}} \right) P_T}{\left(\frac{1}{\gamma_{M1}}G_{M1,i} + \cdots + \frac{1}{\gamma_{M3}}G_{M3,i} + G_{\mu 1,i} + \cdots + \frac{1}{\gamma_{\mu K}}G_{\mu K,i} \right) P_{\mu 1}^t} \\ &\approx \frac{P_T}{P_{\mu 1}^t}. \end{aligned} \quad (12.56)$$

Notice that with NPP, every MS experiences the same forward CIR regardless of its location. While NPP does not offer any significant performance gain over conventional power control schemes, it does guarantee that every MS experiences the same forward link CIR.

The forward analysis consists of two parts; non-STD and STD cases. For non-STD case, there is no transmit diversity; each BS has only one antenna and provides single forward transmit path. Each MS connects to the BS which provides the most robust path. A single-cell hierarchical model as in Fig. 12.10 is used for the forward analysis, since the concept can be readily extended to multi-cell environments. For the single macrocell/microcell model,

$$P_{BSj,i} = \frac{G_{\mu,i} + (1/\gamma)G_{M,i}}{G_{BSj,i}}P_T, \quad (12.57)$$

where $\gamma = P_{\mu 1}^t/P_M^t$ is the ratio between the microcell and macrocell total forward transmit powers, and $BSj \in \{\mu, M\}$.

12.3.3.1 Non-selective Transmit Diversity

From the model in Sect. 12.3.1, the gain $G_{BSj,i}$ has a composite Gamma-log-normal distribution that can be approximated by a purely log-normal distributed with the parameters given in (12.42). The mean $\mu_{(\text{dBm})}$ in (12.42) depends on the distance between base station BSj and MS_i according to the path loss model in (12.41). This distance can be expressed as a function of the polar coordinates (r, θ) of the MS location referenced to BSj, i.e., $d_{BSj}(r, \theta)$. Hence, $G_{BSj,i} = 10^{\varsigma_{BSj}(r, \theta)/10}$, where $\varsigma_{BSj}(r, \theta)$ is a normal random variable with mean $\mu_{(\text{dBm})}$ and σ^2 in (12.42).

Suppose that the location of MS_i, (r, θ) , is referenced with respect to the BS it is connected to. Then, the forward transmit power according to NPP is

$$\begin{aligned}
 P_{\mu,i}(r, \theta) &= \frac{(1/\gamma)G_{M,i} + G_{\mu,i}}{G_{\mu,i}} P_T \\
 &= \left(\frac{G_{M,i}}{\gamma G_{\mu,i}} + 1 \right) P_T \\
 &= \left(\frac{1}{\gamma} 10^{(\varsigma_M(r, \theta) - \varsigma_{\mu}(r, \theta))/10} + 1 \right) P_T \\
 &= \left(\frac{1}{\gamma} 10^{x(r, \theta)/10} + 1 \right) P_T, \text{ if } G_{\mu,i} > G_{M,i} \\
 P_{M,j}(r, \theta) &= \frac{(1/\gamma)G_{M,j} + G_{\mu,j}}{G_{M,j}} P_T \\
 &= \left(\frac{1}{\gamma} + \frac{G_{\mu,j}}{G_{M,j}} \right) P_T \\
 &= \left(\frac{1}{\gamma} + 10^{(\varsigma_{\mu}(r, \theta) - \varsigma_M(r, \theta))/10} \right) P_T \\
 &= \left(\frac{1}{\gamma} + 10^{y(r, \theta)/10} \right) P_T, \text{ if } G_{\mu,j} < G_{M,j},
 \end{aligned} \tag{12.58}$$

where $x(r, \theta) = \varsigma_M(r, \theta) - \varsigma_{\mu}(r, \theta)$, $y(r, \theta) = \varsigma_{\mu}(r, \theta) - \varsigma_M(r, \theta)$, and P_T is the forward transmit power constant. Let us now compute the conditional cumulative distribution functions (cdfs) of $Z = 10^{x(r, \theta)/10}$ and $W = 10^{y(r, \theta)/10}$ for the deployment shown in Fig. 12.10. First note that a Microcell MS is physically located in the Microcell area but may be served by either the Microcell or Macrocell BS. Also, a Macrocell MS can be physically located anywhere in the Macrocell area including the Microcell area, and may be served by either the Microcell or Macrocell BS. It follows that

$$\begin{aligned}
 F_Z(z|r, \theta) &= \frac{\text{P}[10^{x(r, \theta)/10} < z|r, \theta]}{\text{P}[10^{x(r, \theta)/10} < 1|r, \theta]} = \frac{\text{P}[x(r, \theta) < 10 \log_{10}(z)|r, \theta]}{\text{P}[x(r, \theta) < 0|r, \theta]} \\
 &= \Phi \left(\frac{10 \log_{10}\{z\} - \mu_x(r, \theta)}{\sigma_x} \right) \left(\Phi \left(\frac{-\mu_x(r, \theta)}{\sigma_x} \right) \right)^{-1}, \quad 0 < z < 1
 \end{aligned} \tag{12.59}$$

$$F_W(w|r, \theta) = \Phi \left(\frac{10 \log_{10}\{w\} - \mu_y(r, \theta)}{\sigma_y} \right) \left(\Phi \left(\frac{-\mu_y(r, \theta)}{\sigma_y} \right) \right)^{-1}, \quad 0 < w < 1, \tag{12.60}$$

where

$$\begin{aligned}
 \mu_x(r, \theta) &= \mu_{\varsigma_M}(r, \theta) - \mu_{\varsigma_{\mu}}(r, \theta) \\
 &= -\beta 10 \log_{10}\{d_M(r, \theta)\} + \beta 10 \log_{10}\{d_{\mu}(r, \theta)\} \\
 \mu_y(r, \theta) &= \mu_{\varsigma_{\mu}}(r, \theta) - \mu_{\varsigma_M}(r, \theta) \\
 &= -\beta 10 \log_{10}\{d_{\mu}(r, \theta)\} + \beta 10 \log_{10}\{d_M(r, \theta)\}
 \end{aligned}$$

$$\sigma_x^2 = \sigma_y^2 = 2(\sigma_\Omega^2 + 31.0254). \quad (12.61)$$

Then the cdfs of Z and W are

$$F_Z(z) = \begin{cases} \int_0^{2\pi} \int_0^{R_\mu} \frac{r}{\pi R_\mu^2} F_Z(z|r, \theta) dr d\theta, & \text{if MS} \in \text{Microcell} \\ \int_0^{2\pi} \int_0^{R_M} \frac{r}{\pi R_M^2} F_Z(z|r, \theta) dr d\theta, & \text{if MS} \in \text{Macrocell} \end{cases} \quad (12.62)$$

$$F_W(w) = \begin{cases} \int_0^{2\pi} \int_0^{R_M} \frac{r}{\pi R_M^2} F_W(w|r, \theta) dr d\theta, & \text{if MS} \in \text{Macrocell} \\ \int_0^{2\pi} \int_0^{R_\mu} \frac{r}{\pi R_\mu^2} F_W(w|r, \theta) dr d\theta, & \text{if MS} \in \text{Microcell} \end{cases}. \quad (12.63)$$

Since both Z and W are nonnegative random variables, their expected values are as follows:

$$\begin{aligned} E[Z] &= \int_0^\infty (1 - F_Z(z)) dz = \int_0^1 (1 - F_Z(z)) dz \\ E[W] &= \int_0^1 (1 - F_W(w)) dw. \end{aligned} \quad (12.64)$$

Let N and M be the number of MSs physically located in Microcell and Macrocell, respectively. Assuming there are N' Microcell MS connected to Microcell and M' Macrocell MSs connected to Macrocell, the expected value of the total forward transmit power by each BS is

$$E[P_\mu|N', M'] = N' P_T \left(\frac{E[Z|\text{MS} \in \text{Microcell}]}{\gamma} + 1 \right) + (M - M') P_T \left(\frac{E[Z|\text{MS} \in \text{Macrocell}]}{\gamma} + 1 \right) \quad (12.65)$$

$$E[P_M|N', M'] = (N - N') P_T \left(\frac{1}{\gamma} + E[W|\text{MS} \in \text{Microcell}] \right) + M' P_T \left(\frac{1}{\gamma} + E[W|\text{MS} \in \text{Macrocell}] \right). \quad (12.66)$$

Note that N' and M' are (N, Pr_μ) and (M, Pr_M) binomial random variables, respectively, where Pr_μ and Pr_M are the probabilities that a MS located in Microcell and Macrocell will connect to Microcell and Macrocell, respectively. Then,

$$E[P_\mu] = \sum_{N'=0}^N \binom{N}{N'} \text{Pr}_\mu^{N'} (1 - \text{Pr}_\mu)^{N-N'} \sum_{M'=0}^M \binom{M}{M'} \text{Pr}_M^{M'} (1 - \text{Pr}_M)^{M-M'} E[P_\mu|N', M'] \quad (12.67)$$

$$E[P_M] = \sum_{N'=0}^N \binom{N}{N'} \text{Pr}_\mu^{N'} (1 - \text{Pr}_\mu)^{N-N'} \sum_{M'=0}^M \binom{M}{M'} \text{Pr}_M^{M'} (1 - \text{Pr}_M)^{M-M'} E[P_M|N', M'], \quad (12.68)$$

where

$$\begin{aligned} \text{Pr}_\mu &= P[G_{-,i} > G_{M,i} | i \in \text{Microcell}] \\ &= 1 - \int_0^{2\pi} \int_0^{R_\mu} \frac{r}{\pi R_\mu^2} \Phi \left(\frac{\mu_M(r, \theta) - \mu_\mu(r, \theta)}{\sqrt{2(\sigma_\Omega^2 + 31.0254)}} \right) dr d\theta \end{aligned} \quad (12.69)$$

$$\begin{aligned} \text{Pr}_M &= P[G_{M,j} > G_{-,j} | j \in \text{Macrocell}] \\ &= \int_0^{2\pi} \int_0^{R_M} \frac{r}{\pi R_M^2} \Phi \left(\frac{\mu_M(r, \theta) - \mu_\mu(r, \theta)}{\sqrt{2(\sigma_\Omega^2 + 31.0254)}} \right) dr d\theta. \end{aligned} \quad (12.70)$$

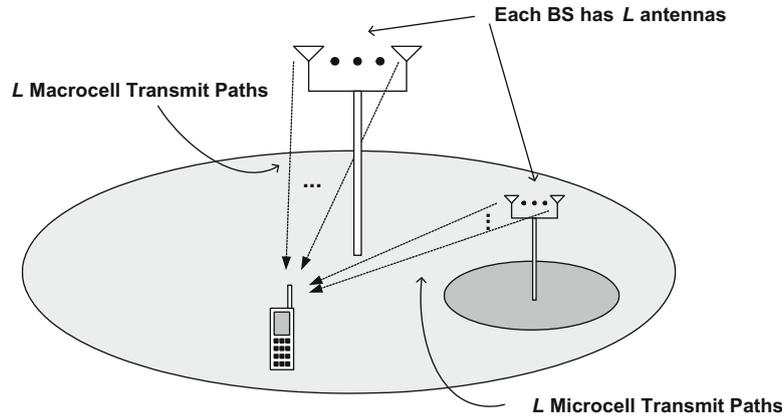


Fig. 12.15 Selective transmit diversity

Let $\gamma = E[P_\mu]/E[P_M]$ and iterate until γ converges. Then the forward CIR is

$$\text{CIR}_{\text{non-STD}} = \frac{P_T}{P_\mu} \approx \frac{P_T}{E[P_\mu]}. \quad (12.71)$$

12.3.3.2 Selective Transmit Diversity

In STD, each BS has a number of spatially separated antennas and an orthogonal pilot signal is transmitted from each antenna as shown in Fig. 12.15. In order for the fading conditions associated with different antennas to be sufficiently decorrelated (less than 0.7 correlation), the BS antenna separation needs to be on the order of ten wavelengths as shown in Sect. 2.1.6.2. By way of monitoring the pilot signals, a MS can select (mobile-assisted) the BS antenna that provides the most robust forward link until a better BS antenna is found. Therefore, only one BS antenna is selected to transmit at a time with STD, but the selected BS antenna will provide the best link among the multiple BS antennas. The main difference between STD and non-STD schemes is that STD provides multiple potential forward links per BS each with uncorrelated fading, while non-STD provides one forward link per BS. However, both STD and non-STD use only one BS antenna to transmit at a time. Diversity gain through antenna separation is viable at BS sites where the space and system complexity are less of limiting factors. The antennas are assumed separated sufficiently far apart that all potential forward links from the same BS have uncorrelated fading, but correlated shadowing.

Let us now assume that the Microcell BS and Macrocell BS each have L antennas, which means that the MS selects the best antenna out of a total of $2L$ antennas. Let $P_{\text{BS}j}^{(l)}$ be the total forward transmit power by the l th transmit branch of $\text{BS}j$, where $\text{BS}j \in \{\mu, M\}$. Since the MSs are uniformly distributed, each transmit branch within a cell has an equal probability of being selected by the MSs and, therefore, it may be assumed that

$$\begin{aligned} P_\mu^{(i)} &\approx P_\mu^{(j)} \approx P_\mu^{\text{STD}} \\ P_M^{(i)} &\approx P_M^{(j)} \approx P_M^{\text{STD}}, \quad i \neq j. \end{aligned} \quad (12.72)$$

Then, the forward transmit power for MS_i located at (r, θ) is

$$P_{\mu,i}^{\text{STD}}(r, \theta) = \frac{(1/\gamma_{\text{STD}})(G_{M,i}^{(1)} + \dots + G_{M,i}^{(L)}) + G_{\mu,i}^{(1)} + \dots + G_{\mu,i}^{(L)}}{\max[G_{\mu,i}^{(1)}, \dots, G_{\mu,i}^{(L)}]} P_T, \quad \text{if } \max[G_{\mu,i}^{(1)}, \dots, G_{\mu,i}^{(L)}] > \max[G_{M,i}^{(1)}, \dots, G_{M,i}^{(L)}] \quad (12.73)$$

$$P_{M,i}^{\text{STD}}(r, \theta) = \frac{(1/\gamma_{\text{STD}})(G_{M,i}^{(1)} + \dots + G_{M,i}^{(L)}) + G_{\mu,i}^{(1)} + \dots + G_{\mu,i}^{(L)}}{\max[G_{M,i}^{(1)}, \dots, G_{M,i}^{(L)}]} P_T, \quad \text{if } \max[G_{\mu,i}^{(1)}, \dots, G_{\mu,i}^{(L)}] < \max[G_{M,i}^{(1)}, \dots, G_{M,i}^{(L)}], \quad (12.74)$$

where $\gamma_{\text{STD}} = P_{\mu}^{\text{STD}}/P_{\text{M}}^{\text{STD}}$ and $G_{\text{BS},i}^{(l)}$ is the gain associated with l th transmit branch of BS j to MS i . Let $G_{\text{BS},i} = \max[G_{\text{BS},i}^{(1)}, \dots, G_{\text{BS},i}^{(L)}]$, then the cdf of $G_{\text{BS},i}^{(l)}/G_{\text{BS},i}$ is

$$\text{P}[G_{\text{BS},i}^{(l)}/G_{\text{BS},i} < x] = \frac{\text{P}[G_{\text{BS},i}^{(l)}/G_{\text{BS},i} < x]}{\text{P}[G_{\text{BS},i}^{(l)}/G_{\text{BS},i} < 1]} \approx \frac{2x}{x+1}, \quad 0 < x < 1, \quad (12.75)$$

where $\text{P}[G_{\text{BS},i}^{(l)}/G_{\text{BS},i} < x] = x/(x+1)$. The above result is accurate for $L = 2$ and adequate for $L = 3$ since the greatest diversity gain occurs between $L = 1$ and $L = 2$ as shown in Chap. 6. For two different cell locations, it can be assumed that $G_{\text{BS},i}^{(l)}$ and $G_{\text{BS},i}$ are independent. Because the transmit paths from the same BS experience independent Rayleigh fading and the same shadowing, the conditional pdf and pdf of $G_{\text{BS},i}$ are, respectively,

$$f_{G_{\text{BS},i}}(x|\Omega_p) = \frac{L}{\Omega_p} e^{-x/\Omega_p} (1 - e^{-x/\Omega_p})^{L-1} \quad (12.76)$$

and

$$f_{G_{\text{BS},i}}(x) = \int_0^\infty f_{G_{\text{BS},i}}(x|\Omega_p) \frac{1}{\sqrt{2\pi\xi\sigma_\Omega\Omega_p}} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p}(\text{dBm}))^2}{2\sigma_\Omega^2}\right\} d\Omega_p. \quad (12.77)$$

It is shown in Appendix 12A in this chapter that $f_{G_{\text{BS},i}}(x)$ can be approximated by a purely log-normal distribution for $L = 2$ with mean and variance given by

$$\begin{aligned} \mu_G &= \xi^{-1}(\ln 2 - C) + \mu_\Omega = 0.503552 - \beta \log_{10}\{d\} \\ \sigma_G^2 &= \xi^{-2}(\zeta(2, 1) - 2(\ln 2)^2) + \sigma_\Omega^2 = 12.9016 + \sigma_\Omega^2, \end{aligned} \quad (12.78)$$

where $C \simeq 0.5772$ is Euler's constant and $\zeta(2, 1) = \sum_{k=0}^\infty 1/(1+k)^2$ is Reimann's zeta function. Appendix 12B in this chapter derives a similar approximation for $L = 3$, yielding the following mean and variance:

$$\begin{aligned} \mu_G &= \xi^{-1}(3\ln 2 - \ln 3 - C) + \mu_\Omega = 1.75294 - \beta \log_{10}\{d\} \\ \sigma_G^2 &= \xi^{-2}(\zeta(2, 1) - 12(\ln 2)^2 + 6(\ln 2)(\ln 3)) + \sigma_\Omega^2 = 8.4592 + \sigma_\Omega^2. \end{aligned} \quad (12.79)$$

Then the conditional cdf and cdf of $G_{\text{BS},i}^{(l)}/G_{\text{BS},i}$ are

$$\begin{aligned} \text{P}[G_{\text{BS},i}^{(l)}/G_{\text{BS},i} < x|r, \theta] &= \frac{\text{P}[G_{\text{BS},i}^{(l)}/G_{\text{BS},i} < x|r, \theta]}{\text{P}[G_{\text{BS},i}^{(l)}/G_{\text{BS},i} < 1|r, \theta]} \\ &= \Phi\left(\frac{10\log_{10}(x) - \mu_{kj}(r, \theta)}{\sigma_{kj}}\right) \left(\Phi\left(\frac{-\mu_{kj}(r, \theta)}{\sigma_{kj}}\right)\right)^{-1} \end{aligned} \quad (12.80)$$

$$\text{P}[G_{\text{BS},i}^{(l)}/G_{\text{BS},i} < x] = \int_0^{2\pi} \int_0^{R_{\text{BS},j}} \frac{r}{\pi R_{\text{BS},j}^2} \text{P}[G_{\text{BS},i}^{(l)}/G_{\text{BS},i} < x|r, \theta] dr d\theta, \quad 0 < x < 1, \quad (12.81)$$

where

$$\begin{aligned} \mu_{kj}(r, \theta) &= \mu_{\text{BS},i}(r, \theta) - \mu_{G_{\text{BS},i}}(r, \theta) \\ &= -\beta 10 \log_{10}\{d_{\text{BS},i}(r, \theta)\} - 2.50675 + \mu_{G_{\text{BS},i}}(r, \theta) \\ \sigma_{kj}^2 &= \sigma_{\text{BS},i}^2 + \sigma_G^2 \\ &= \sigma_\Omega^2 + 31.0254 + \sigma_G^2. \end{aligned} \quad (12.82)$$

The expected values of $P_{\mu,i}^{\text{STD}}$ and $P_{M,i}^{\text{STD}}$ can be derived using the same approach used in the previous section to get

$$\begin{aligned} E[P_{\mu,i}^{\text{STD}}] &= \left(\frac{L}{\gamma_{\text{STD}}} E[Z] + (L-1)E[X] + 1 \right) P_T \\ E[P_{M,i}^{\text{STD}}] &= \left(\frac{L-1}{\gamma_{\text{STD}}} E[Y] + LE[W] + \frac{1}{\gamma_{\text{STD}}} \right) P_T, \end{aligned} \quad (12.83)$$

where

$$\begin{aligned} Z &= G_{M,i}^{(l)} / G_{\mu,i} \\ X &= G_{\mu,i}^{(l)} / G_{\mu,i} \\ Y &= G_{M,i}^{(l)} / G_{M,i} \\ W &= G_{\mu,i}^{(l)} / G_{M,i}, \quad 0 < Z, X, Y, W < 1. \end{aligned} \quad (12.84)$$

Since each transmit branch has equal chance of being selected, the expected value of the total forward transmit power by each BS branch, given N' and M' , is

$$E[P_{\mu}^{\text{STD}} | N', M'] = \frac{N'}{L} P_T E[P_{\mu,i}^{\text{STD}} | \text{MS}_i \in \text{Microcell}] + \frac{M-M'}{L} P_T E[P_{\mu,i}^{\text{STD}} | \text{MS}_i \in \text{Macrocell}] \quad (12.85)$$

$$E[P_M^{\text{STD}} | N', M'] = \frac{N-N'}{L} P_T E[P_{M,i}^{\text{STD}} | \text{MS}_i \in \text{Microcell}] + \frac{M'}{L} P_T E[P_{M,i}^{\text{STD}} | \text{MS}_i \in \text{Macrocell}]. \quad (12.86)$$

The expectations $E[P_{\mu}^{\text{STD}}]$ and $E[P_M^{\text{STD}}]$ can now be computed as in (12.67) with the following Pr_{μ} and Pr_M :

$$\begin{aligned} \text{Pr}_{\mu} &= P[\max[G_{\mu,i}^{(1)}, \dots, G_{\mu,i}^{(L)}] > \max[G_{M,i}^{(1)}, \dots, G_{M,i}^{(L)}] | \text{MS}_i \in \text{Microcell}] \\ &= P[G_{\mu,i} > G_{M,i} | \text{MS}_i \in \text{Microcell}] \end{aligned} \quad (12.87)$$

$$\text{Pr}_M = P[G_{M,i} > G_{\mu,i} | \text{MS}_i \in \text{Macrocell}]. \quad (12.88)$$

12.3.3.3 Numerical Results

Table 12.3 compares the analytical and simulation results of the single cell model. These results are obtained with $d_c = 692$ m. The analytical results require less than 15 iteration loops for the γ 's to converge. The analytical and simulation results are in good agreement, although the analytical results exhibit lower values than the simulation results. This is due to the analytical interference assumption in (12.54) which results in a pessimistic interference level. However, the differences become marginal as the number of MSs increases. Table 12.3 also shows the results obtained with STD. Again, the analytical results closely follow the simulation results. One can immediately see the benefit of employing STD on forward link performance. With 2-branch transmit diversity the forward link CIR performance improves by 1.5 dB, and it improves by 2.5 dB with 3-branch transmit diversity.

Table 12.3 Single cell model forward performance results

N	M	Non-STD		STD, $L = 2$		STD, $L = 3$	
		CIR _{anal.} (dB)	CIR _{sim.} (dB)	CIR _{anal.} (dB)	CIR _{sim.} (dB)	CIR _{anal.} (dB)	CIR _{sim.} (dB)
10	10	-11.54	-11.21	-9.99	-9.65	-9.31	-8.86
15	10	-12.80	-12.56	-11.25	-11.02	-10.57	-10.20
20	15	-14.19	-14.02	-12.64	-12.46	-11.96	-11.67
20	20	-14.56	-14.40	-13.01	-12.84	-12.34	-12.04

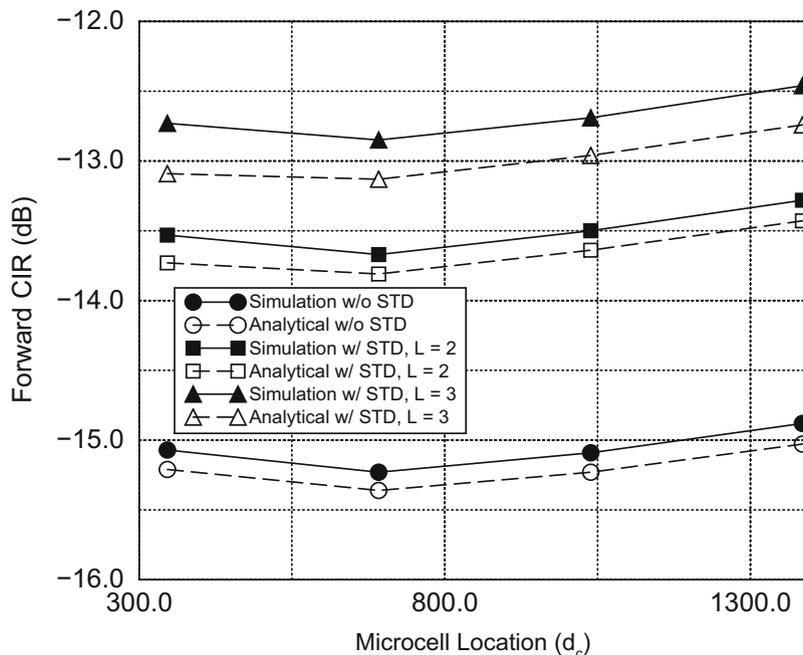


Fig. 12.16 Forward link performance for different microcell locations

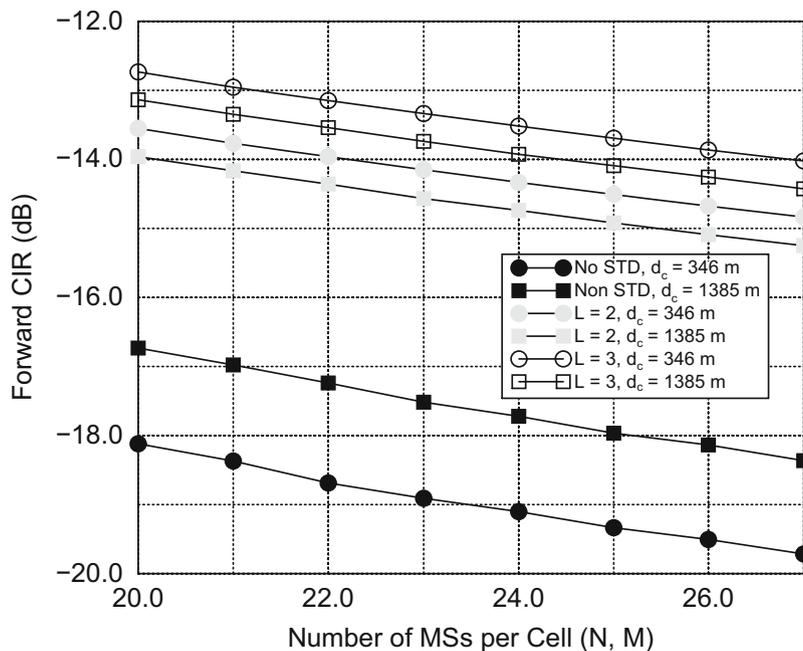


Fig. 12.17 Forward performance for the multi-cell model in Fig. 12.11

Figure 12.16 compares the analytical and simulation results as a function of the microcell location, d_c , while the system load is fixed at 24 MSs per cell. Again, they are in good agreement. The differences between the analytical and simulation results in the figure are larger for $L = 3$ than $L = 2$ due to the assumption in (12.75). But, the differences are still small considering they are within 0.3 dB. Figure 12.17 shows the average forward link performance results obtained from the multi-cell model in Fig. 12.11 consisting of three macrocells and a cluster of three microcells. The benefit of STD is more evident from these results where the system benefits from the added diversity effect due to multiple cell locations. Although the performance varies slightly, STD effectively neutralizes the effect of the microcell cluster location on forward performance. Without STD, the performance suffers if the cluster is located too close to a macrocell BS. Figure 12.17 shows about 1.5 dB

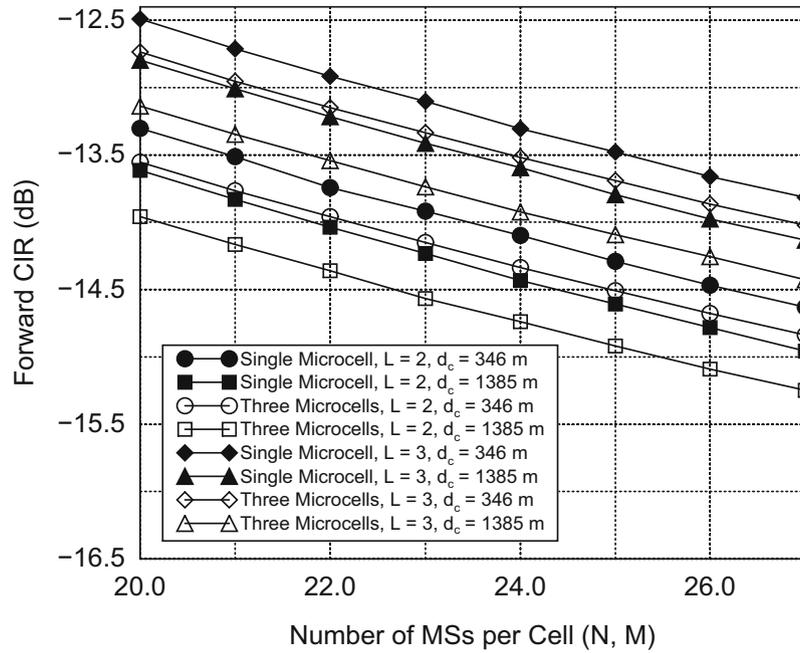


Fig. 12.18 Forward performance for different microcell cluster sizes

performance difference between the cluster locations, $d_c = 346$ m and 1385 m without STD. One interesting observation is that performance improves slightly as d_c decreases with STD, which is opposite of that observed in non-STD case. Although the cross interference between the hierarchical layers increases with smaller d_c , STD takes advantage of increased inter-layer diversity effect which ultimately results in improved performance.

Figure 12.18 shows the effect of the microcell cluster size on the forward link performance. The plot shows the performance comparison between a single-cell cluster and a three-cell cluster. The previous section has shown that HMRC allows microcell(s) to be added without impacting the existing reverse link capacity. With STD, however, the forward link performance does depend on the microcell cluster size, and an increase in the number of microcells in the cluster causes the performance to suffer. The added diversity effect is apparently not enough to fully compensate for increase in interference resulting from the microcell increase. However, the performance degradation is negligible considering the capacity gain obtained by adding microcell(s). For example, assuming the target CIR is set to -14 dB, a three macrocell system with a single-cell cluster has capacity of 26 MSs per cell at $d_c = 1385$ m with $L = 3$. With the same exact setting, the same system with a three-cell cluster achieves 24 MSs per cell, yet its overall system capacity is far greater due to the added microcells.

From Figs. 12.12 and 12.17, observe that a forward link performance comparable to the reverse link HMRC performance can be achieved by implementing STD with $L = 3$. Not only does STD improve the forward performance, but it can also benefit the reverse link performance by providing additional BS antenna elements for stronger combining.

Appendix 12A: Derivation of Eq. (12.78)

The conditional pdf of the link gain, G , is

$$p_{G|\Omega_p}(x) = \frac{2}{\Omega_p} e^{-x/\Omega_p} (1 - e^{-x/\Omega_p}). \quad (12A.1)$$

Averaging over distribution of log-normal shadowing yields the composite pdf for squared-envelope

$$p_G(g) = \int_0^\infty \frac{2}{\Omega_p} e^{-g/\Omega_p} (1 - e^{-g/\Omega_p}) \frac{1}{\sqrt{2\pi\xi\sigma_\Omega\Omega_p}} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p}(\text{dBm}))^2}{2\sigma_\Omega^2}\right\} d\Omega_p.$$

The mean of the approximate log-normal distribution is

$$\begin{aligned}\mu_{G \text{ (dBm)}} &= E[10\log_{10}\{g\}] \\ &= \int_0^\infty \int_0^\infty (10\log_{10}\{g\}) \frac{2}{\Omega_p} e^{-g/\Omega_p} (1 - e^{-g/\Omega_p}) \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega\Omega_p} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} dw d\Omega_p \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_\Omega\xi^2\Omega_p^2} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} \int_0^\infty 2\ln(g)e^{-g/\Omega_p} (1 - e^{-g/\Omega_p}) dg d\Omega_p.\end{aligned}\quad (12A.2)$$

From [147, 4.352.1], the inner integral becomes

$$\int_0^\infty 2\ln(g)e^{-g/\Omega_p}(1 - e^{-g/\Omega_p})dg = \Omega_p(\ln(2) + \ln(\Omega_p) - C),\quad (12A.3)$$

where $C \simeq 0.5772$ is Euler's constant. Hence,

$$\mu_{G \text{ (dBm)}} = \xi^{-1}(\ln(2) - C) + \mu_{\Omega_p \text{ (dBm)}}.\quad (12A.4)$$

In a similar fashion, the mean square value is

$$\begin{aligned}E[(10\log_{10}\{g\})^2] &= \int_0^\infty \int_0^\infty (10\log_{10}\{g\})^2 \frac{2}{\Omega_p} e^{-g/\Omega_p} (1 - e^{-g/\Omega_p}) \\ &\quad \times \frac{1}{\sqrt{2\pi}\sigma_\Omega\xi\Omega_p} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} dg d\Omega_p \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_\Omega\xi^3\Omega_p^2} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} \\ &\quad \times \int_0^\infty 2(\ln g)^2 e^{-g/\Omega_p} (1 - e^{-g/\Omega_p}) dg d\Omega_p.\end{aligned}\quad (12A.5)$$

From [147, 4.358.2], the inner integrals become

$$\begin{aligned}\int_0^\infty 2(\ln g)^2 e^{-g/\Omega_p} dg &= 2\Omega_p((\ln \Omega_p - C) + \zeta(2, 1)) \\ \int_0^\infty (\ln g)^2 e^{-2g/\Omega_p} dg &= \Omega_p((\ln \Omega_p - \ln(2) - C)^2 + \zeta(2, 1)),\end{aligned}\quad (12A.6)$$

where

$$\zeta(2, m) = \sum_{k=0}^{\infty} \frac{1}{(m+k)^2}\quad (12A.7)$$

is Reimann's zeta function. Finally, the variance is

$$\begin{aligned}\sigma_{G \text{ (dBm)}}^2 &= E[(10\log_{10}\{g\})^2] - E^2[10\log_{10}\{g\}] \\ &= \xi^{-2}(\zeta(2, 1) - 2(\ln 2)^2) + \sigma_\Omega^2.\end{aligned}\quad (12A.8)$$

Appendix 12B: Derivation of Eq. (12.79)

For $L = 3$, the mean of log-normal approximation is

$$\begin{aligned} E[10\log_{10}\{g\}] &= \int_0^\infty \int_0^\infty 10\log_{10}\{g\} \frac{3}{\Omega_p} e^{-g/\Omega_p} (1 - e^{-g/\Omega_p})^2 \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega\Omega_p} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} dg d\Omega_p \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}\xi^2\sigma_\Omega\Omega_p^2} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} \\ &\quad \times \int_0^\infty 3\ln(g)(e^{-g/\Omega_p} - 2e^{-2g/\Omega_p} + e^{-3g/\Omega_p}) dg d\Omega_p. \end{aligned} \quad (12B.1)$$

Using [147, 4.352.1]

$$\int_0^\infty 3\ln(g)(e^{-g/\Omega_p} - 2e^{-2g/\Omega_p} + e^{-3g/\Omega_p}) dg = \Omega_p(3\ln(2) - \ln(3) - C + \ln(\Omega_p)) \quad (12B.2)$$

and, therefore,

$$\mu_G \text{ (dBm)} = E[10\log_{10}\{g\}] = \xi^{-1}(3\ln(2) - \ln(3) - C) + \mu_{\Omega_p \text{ (dBm)}}. \quad (12B.3)$$

Similarly, the second moment of the approximation is

$$\begin{aligned} E[(10\log_{10}\{g\})^2] &= \int_0^\infty \int_0^\infty (10\log_{10}\{g\})^2 \frac{3}{\Omega_p} e^{-g/\Omega_p} (1 - e^{-g/\Omega_p})^2 \\ &\quad \times \frac{1}{\sqrt{2\pi}\sigma_\Omega\xi\Omega_p} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} dg d\Omega_p \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_\Omega\xi^3\Omega_p^2} \exp\left\{-\frac{(10\log_{10}\{\Omega_p\} - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} \\ &\quad \times \int_0^\infty 3(\ln(g))^2(e^{-g/\Omega_p} - 2e^{-2g/\Omega_p} + e^{-3g/\Omega_p}) dg d\Omega_p. \end{aligned} \quad (12B.4)$$

Using [147, 4.358.2]

$$\begin{aligned} \int_0^\infty 3(\ln(g))^2(e^{-g/\Omega_p} - 2e^{-2g/\Omega_p} + e^{-3g/\Omega_p}) dg &= \Omega_p \left(-6C\ln(2) + 6(2)\ln(\Omega_p) - 3(\ln(2))^2 + C^2 - 2C\ln(\Omega_p) \right. \\ &\quad \left. + (\ln(\Omega_p))^2 + 2C\ln(3) - 2\ln(3)\ln(\Omega_p) + (\ln(3))^2 + \zeta(2, 1) \right) \end{aligned} \quad (12B.5)$$

and, therefore,

$$\begin{aligned} E[(10\log_{10}\{g\})^2] &= \xi^{-2} \left(-6C\ln(2) - 3(\ln(2))^2 + C^2 + 2C\ln(3) + (\ln(3))^2 + \zeta(2, 1) \right) \\ &\quad + 6\ln(2)\mu_{\Omega_p \text{ (dBm)}} - 2C\mu_{\Omega_p \text{ (dBm)}} - 2\ln(3)\mu_{\Omega_p \text{ (dBm)}} + \sigma_\Omega^2 + \mu_{\Omega_p \text{ (dBm)}}^2. \end{aligned} \quad (12B.6)$$

Therefore, the variance of the log-normal approximation is

$$\begin{aligned} \sigma_G^2 \text{ (dBm)} &= E[(10\log_{10}\{g\})^2] - \mu_G^2 \text{ (dBm)} \\ &= \xi^{-2} \left(\zeta(2, 1) - 12(\ln(2))^2 + 6\ln(2)\ln(3) \right) + \sigma_\Omega^2. \end{aligned} \quad (12B.7)$$

Problems

12.1. 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 ζ_{E_j} . The power control errors, ζ_{E_j} (in dB) are independent zero-mean Gaussian random variables with variance σ_E^2 . Hence, the in-cell interference is

$$I_{\text{in}} = C \sum_{j=1}^5 10^{\zeta_{E_j}/10} \chi_j,$$

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 $(I_{\text{in}})_{\text{dB}}$ is Gaussian when conditioned on the number of active interferers, write down and expression for the pdf of $(I_{\text{in}})_{\text{dB}}$.

12.2 (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, 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

$$\begin{aligned} \mu_x &= \mu_{i0} - \mu_{ij} \\ \sigma_x^2 &= 2\sigma_\Omega^2. \end{aligned}$$

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?

12.3. 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

$$\gamma_b^s = \max \{ \gamma_1, \gamma_2, \dots, \gamma_L \}.$$

An outage occurs if $\gamma_b^s < \gamma_{\text{th}}$. What is the probability of outage?

- (b) The forward link uses maximal ratio combining such that

$$\gamma_b^{\text{mr}} = \gamma_1 + \gamma_2 + \dots + \gamma_L.$$

Again, an outage occurs if $\gamma_b^{\text{mr}} < \gamma_{\text{th}}$. What is the probability of outage if $\bar{\gamma}_1 = \bar{\gamma}_2 = \dots = \bar{\gamma}_L$?

- (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 13

Radio Resource Management

When a new call arrives, mobile station (MS) must be connected to a suitable base station (BS) or perhaps a set of BSs. Also, as a MS traverses from one cell to the next handoffs are required so that an acceptable link quality can be maintained; otherwise the required increase in transmit power to compensate for path loss will result in excessive co-channel and adjacent channel interference. The failure to handoff a MS as it moves from one cell to the next will also tend to increase the call blocking probability, because some cells may carry more traffic than planned. For TDMA cellular systems that do not employ frequency hopping, an intra-cell handoff is sometimes 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 mechanism consists of two processes: (1) link quality evaluation and handoff initiation, (2) 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 [190] has shown that the number of cell boundary crossings is inversely proportional to the cell size. Furthermore, Nanda [239] has shown that the handoff rate increases with only the square root of the call density in macrocells, but it increases 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 increased co-channel interference and/or excessive control traffic. A variety of parameters such as bit error rate (BER) [77], carrier-to-interference ratio, C/I [126], distance [112, 223], traffic load, signal strength [62, 149, 150, 223, 233, 331], and various combinations of these fundamental schemes have been suggested for evaluating the link quality and deciding when a handoff should be performed. One possibility is to measure the received carrier plus interference power, $C + I$. However, a large $C + I$ does not necessarily imply a large C/I . Since the radio link quality depends more on the C/I than the $C + I$, it is apparent that C/I based handoff algorithms are necessary for high capacity cellular systems. 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 was used in first generation cellular systems. With a NCHO algorithm, the reverse link quality is monitored by the serving BS and the surrounding BSs, typically using measurements of the received $C + I$. The reverse link $C + I$ measurements are forwarded to a mobile switching center (MSC), which ultimately makes the handoff decision. Typically, these centralized NCHO algorithms only supported inter-cell handoffs, had handoff delays on the order of several seconds, and relied on relatively infrequent measurements and reports of the reverse link $C + I$ values.

The second type of handoff algorithm is the mobile-assisted handoff (MAHO) algorithm which is widely used in modern cellular systems. MAHO algorithms use both the serving BS and the MS to measure the reverse and forward link quality, respectively, usually based on the received C/I . However, link quality measurements of the alternate BSs are only obtained by the MS. The MS periodically relays the link quality measurements of the serving and alternate BSs back to the serving BS, and the handoff decision is still made by the serving BS along with the MSC. MAHO algorithms typically support both intra-cell and inter-cell handoffs. MAHO algorithms typically use frequent updates of the link quality measurements, and have much lower handoff delays than a NCHO algorithm.

The third type of handoff algorithm is the decentralized mobile-controlled handoff (MCHO) algorithm. 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 intra-cell and inter-cell 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 and are the most reliable. However, since the handoff decisions are made by the MS, network-wide radio resource optimization is more difficult. MCHO algorithms are typically used with cordless telephone systems.

Handoff algorithms can be categorized into forward and backward types depending on how the handoff process is initiated. 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 widely used in TDMA cellular systems such as GSM. Forward handoff algorithms initiate the handoff process via a channel with 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. Finally, handoffs can also be categorized as hard handoffs and soft handoffs.

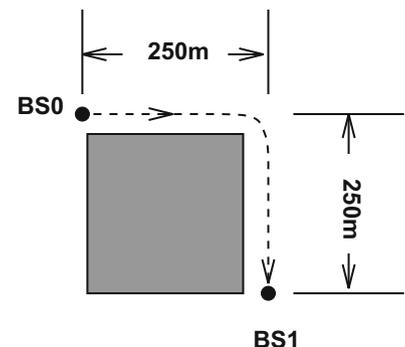
With hard handoffs, a MS can connect to only one BS at a time. An absolute (binary) decision is made to initiate and execute a handoff without making a number of simultaneous connections among candidate BSs. The handoff is initiated based on a hysteresis imposed on the current link. The target BS is 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 the MS can only connect to one BS at a time. Hard handoffs do not take advantage of the 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 TDMA cellular systems such as GSM.

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 control signaling and resource consumption on the network. Several authors [150, 223, 233, 331] 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 chosen to correspond to a spatial distance of 20–40 carrier wavelengths, and the hysteresis H is chosen on the order of the shadow standard deviation.

Murase [233] 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. 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 sometimes called the “corner effect” [62, 233] since it occurs while turning corners in urban microcellular settings like the one shown in Fig. 13.1, where the MS traverses along a route from BS₀ to BS₁. 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 [233].

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

Fig. 13.1 Typical NLoS handoff scenario. The MS rounds the corner, losing the LoS from BS₀, and gaining the LoS from BS₁



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 [233]. 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 this problem, and are known to be robust to the severe propagation environments that are typical of urban microcellular networks [28].

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 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 signals from surrounding BSs. Soft handoff is a type of MAHO. However, all these advantages do not imply that soft handoff is without its weaknesses. Soft handoff is complex and expensive to implement. Also, forward link interference actually increases with soft handoff since several BSs, instead of one, can connect to the same MS. This increase in forward link 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 with power control. CDMA systems are interference-limited meaning their capacities are closely related to the amount of interference the systems can tolerate. Due to universal frequency reuse, a CDMA system cell is affected by not only interference within its own cell, but also interference from its neighboring cells. To alleviate the 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 mobile stations (MSs) and base stations (BSs) from transmitting more power than is necessary in order to limit excess interference. With power control, each MS (or BS) is disciplined to transmit just enough power to meet the target C/I level. However, in order for power control to work properly, the system must ensure that each MS is connected to the BS having the least path attenuation at all times; otherwise, a positive feedback problem can destabilize the entire system. Soft handoff ensures that each MS is served by the best BS a majority of the time, by allowing connections to multiple BSs with macroscopic selection diversity.

To illustrate the necessity for soft handoff in CDMA cellular systems, consider a simple system consisting of two BSs and two MSs as shown in Fig. 13.2. 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/I 's by increasing the MS transmit powers. But for the given setting, increasing the MS transmit powers also increases

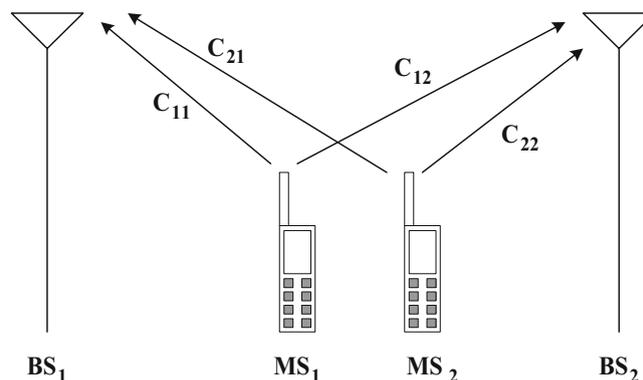


Fig. 13.2 Soft handoff example, from [274]

the respective interference levels and the C/I 's 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 CDMA cellular systems.

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. Moreover, it may be desirable 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 13.1 presents several different types of signal strength based handoff algorithms and Sect. 13.2 introduces pilot-to-interference ratio based soft handoff algorithms. This is followed by a detailed treatment of spatial signal strength averaging in Sect. 13.3. 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 13.4 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 Sect. 13.5, the velocity estimators are incorporated into a velocity adaptive handoff algorithm. Section 13.6 provides an analytical treatment of conventional signal strength based hard handoff algorithms while Sect. 13.7 does the same for soft handoff algorithms. Finally, Sect. 13.8 discusses methods for C/I measurements in TDMA cellular systems.

13.1 Signal Strength Based Hard Handoff Algorithms

Traditional MAHO algorithms use signal strength estimates that are obtained by calculating time averages of the received squared-envelope, $\langle |\tilde{r}_i(t)|^2 \rangle$, from N neighboring BSs, BS_i , $i = 0, \dots, N - 1$. 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 a factor of H . For example, a handoff is performed between two BSs, BS_0 and BS_1 , when

$$\begin{aligned} \bar{\Omega}_{p,1}(n) &> \bar{\Omega}_{p,0}(n)H \quad \text{if the serving BS is } BS_0 \\ \bar{\Omega}_{p,0}(n) &> \bar{\Omega}_{p,1}(n)H \quad \text{if the serving BS is } BS_1, \end{aligned} \quad (13.1)$$

where H denotes the hysteresis, and $\bar{\Omega}_{p,0}(n)$ and $\bar{\Omega}_{p,1}(n)$ are the estimated local mean signal strengths of BS_0 and BS_1 , given by

$$\bar{\Omega}_{p,0}(n) = \frac{1}{N} \sum_{k=n-N+1}^n |r_0(k)|^2 \quad (13.2)$$

$$\bar{\Omega}_{p,1}(n) = \frac{1}{N} \sum_{k=n-N+1}^n |r_1(k)|^2, \quad (13.3)$$

respectively, where $|\tilde{r}_i(kT_s)|^2$ is the k th sample of the received squared-envelope, 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

$$\begin{aligned} \bar{\Omega}_{p,1}(n) &> \bar{\Omega}_{p,0}(n)H \text{ and } \bar{\Omega}_{p,0}(n) > \Omega_L, \quad \text{if the serving BS is } BS_0 \\ \bar{\Omega}_{p,1}(n) &> \bar{\Omega}_{p,0}(n) \text{ and } \bar{\Omega}_{p,0}(n) < \Omega_L, \quad \text{if the serving BS is } BS_0 \end{aligned}$$

$$\begin{aligned} \bar{\Omega}_{p,0}(n) &> \bar{\Omega}_{p,1}(n)H \text{ and } \bar{\Omega}_{p,1}(n) > \Omega_L, \text{ if the serving BS is BS}_1 \\ \bar{\Omega}_{p,0}(n) &> \bar{\Omega}_{p,1}(n) \text{ and } \bar{\Omega}_{p,1}(n) < \Omega_L, \text{ if the serving BS is BS}_1. \end{aligned} \quad (13.4)$$

This scheme encourages a handoff whenever the received signal strength from the serving BS drops below the threshold Ω_L , thereby reducing the probability of a 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 BS₀ and BS₁ when

$$\begin{aligned} \bar{\Omega}_{p,1}(n) &> \bar{\Omega}_{p,0}(n)H \text{ and } \Omega_L < \bar{\Omega}_{p,0}(n) < \Omega_U, \text{ if the serving BS is BS}_0 \\ \bar{\Omega}_{p,1}(n) &> \bar{\Omega}_{p,0}(n) \text{ and } \bar{\Omega}_{p,0}(n) < \Omega_L, \text{ if the serving BS is BS}_0 \\ \bar{\Omega}_{p,0}(n) &> \bar{\Omega}_{p,1}(n)H \text{ and } \Omega_L < \bar{\Omega}_{p,1}(n) < \Omega_U, \text{ if the serving BS is BS}_1 \\ \bar{\Omega}_{p,0}(n) &> \bar{\Omega}_{p,1}(n) \text{ and } \bar{\Omega}_{p,1}(n) < \Omega_L, \text{ if the serving BS is BS}_1. \end{aligned} \quad (13.5)$$

This scheme avoids unnecessary handoffs, thereby reducing the network signaling load and network delay.

Direction biased handoff algorithms have also been suggested for improving the handoff performance in urban microcells [26]. These algorithms 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 the serving base station BS_s, into two sets based on their direction information. Define the following sets:

$$\mathcal{A} := \text{the set of BSs the MS is approaching} \quad (13.6)$$

$$\mathcal{R} := \text{the set of BSs the MS is moving away from.} \quad (13.7)$$

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

$$\begin{aligned} \bar{\Omega}_{p,j}(n) &> \bar{\Omega}_{p,s}(n)H, \text{ if BS}_s \in \mathcal{R} \\ \bar{\Omega}_{p,j}(n) &> \bar{\Omega}_{p,s}(n)H_d, \text{ if BS}_s \in \mathcal{A} \end{aligned} \quad (13.8)$$

or if BS_j $\in \mathcal{A}$ and

$$\begin{aligned} \bar{\Omega}_{p,j}(n) &> \bar{\Omega}_{p,s}(n)H_e, \text{ if BS}_s \in \mathcal{R} \\ \bar{\Omega}_{p,j}(n) &> \bar{\Omega}_{p,s}(n)H, \text{ if BS}_s \in \mathcal{A}. \end{aligned} \quad (13.9)$$

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 (13.1). 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.

13.2 Pilot-to-Interference Ratio Based Soft Handoff Algorithms

In CDMA based systems each BS transmits a pilot signal to assist soft handoff [105]. In synchronous CDMA systems such as cdma2000, all BSs use the same pilot sequence and the BSs are distinguished by using different phase shifts of the same sequence. In asynchronous CDMA systems such as WCDMA, each cell is allocated a distinct pilot sequence. In any case, the MSs use the pilot signals to initiate and complete handoffs. Each pilot signal is used to measure the pilot-to-interference ratio (PIR), which is the ratio of received pilot energy per chip to total interference spectral density:

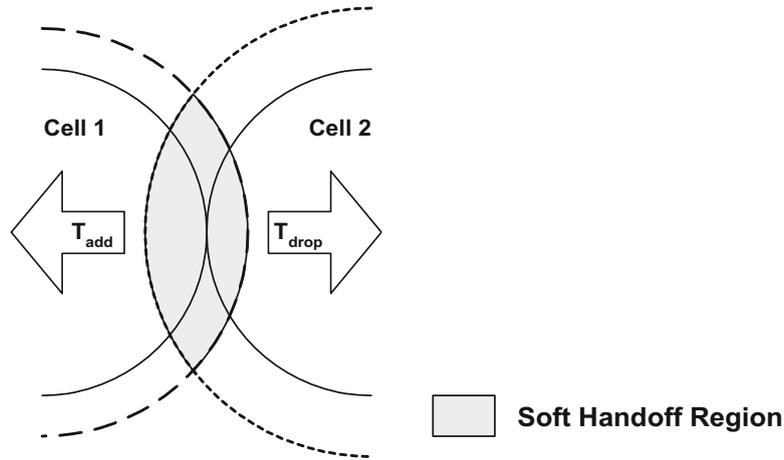


Fig. 13.3 Soft handoff boundaries and region

$$\text{PIR} = \frac{E_c}{I_o}. \quad (13.10)$$

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 active 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. Figure 13.3 shows how changes in the handoff parameters affect 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 decrease the 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 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.

13.3 Signal Strength Averaging

The received squared-envelopes $|\tilde{r}(t)|^2$ are affected by Ricean fading, log-normal shadowing, and path loss attenuation. Here, the two Ricean fading models discussed in Sect. 2.1.3.2 are considered. The first model assumes that the quadrature components of the complex fading envelope $g(t) = g_I(t) + jg_Q(t)$ are independent Gaussian random processes with variance b_0 and means $m_I(t) = m_I$ and $m_Q(t) = m_Q$, respectively. In this case, the envelope $\alpha(t) = |g(t)|$ is Ricean distributed with Rice factor $K = s^2/(2b_0)$, where $s^2 = m_I^2 + m_Q^2$. The second model is more realistic and uses the time-varying means $m_I(t)$ and $m_Q(t)$ in (2.59) and (2.60), 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 Sect. 2.6.1, shadow correlation can be described by the model in (2.6.1). Finally, for LoS propagation, the two-slope path loss model in (2.410) is assumed. For NLoS propagation, the model in (2.412) is assumed yielding, for example, the received signal strength profile in Fig. 2.67.

13.3.1 Choosing the Proper Window Length

When a time average of the received complex squared-envelope ($|\tilde{r}_i(t)|^2$) is calculated, an averaging interval or temporal window length must be chosen. One method for determining the proper window length is to use analog averaging. The following development extends the original work of Lee [199] by incorporating Aulin's Ricean fading model. With Lee's multiplicative model, the squared-envelope of the composite signal at position y is

$$\hat{\alpha}_c^2(y)^2 = \alpha^2(y) \cdot \Omega_p(y) \quad (13.11)$$

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). Since the local mean remains constant over short spatial distances, it is assumed that $\Omega_p(y) = \Omega_p$ over the spatial averaging interval. However, if the spatial averaging interval is too large, this assumption will no longer be valid since the time-average ($|\tilde{r}_i(t)|^2$) will average over the shadowing process as well. Assuming that the envelope fading is an ergodic random process, an integral spatial average of $\hat{\alpha}_c^2(y)$ can be used to estimate the local mean Ω_p , i.e.,

$$\bar{\Omega}_p = \frac{1}{2L} \int_{x-L}^{x+L} \hat{\alpha}_c^2(y) dy = \frac{\Omega_p}{2L} \int_{x-L}^{x+L} \alpha^2(y) dy, \quad (13.12)$$

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 (13.12), calculated as [195]

$$\sigma_{\bar{\Omega}_p}^2 = \frac{1}{L} \int_0^{2L} \left(1 - \frac{\ell}{2L}\right) \lambda_{\alpha^2\alpha^2}(\ell) d\ell, \quad (13.13)$$

where $\lambda_{\alpha^2\alpha^2}(\ell) = E[\alpha^2(y)\alpha^2(y+\ell)] - E[\alpha^2(y)]E[\alpha^2(y+\ell)]$ is the *spatial* autocovariance of the squared-envelope, and $E[x]$ denotes the ensemble average of x . The spatial autocovariance of the received squared-envelope at a MS with 2-D isotropic scattering can be obtained from (2.94) along with the time-distance transformation $f_m\tau = \ell/\lambda_c$. Substituting (2.94) into (13.13) and using $f_m\tau = \ell/\lambda_c$ yields

$$\sigma_{\bar{\Omega}_p}^2 = \left(\frac{\Omega_p}{K+1}\right)^2 \frac{1}{L} \int_0^{2L} \left(1 - \frac{\ell}{2L}\right) J_0(2\pi\ell/\lambda_c) \left(J_0(2\pi\ell/\lambda_c) + 2K \cos(2\pi\ell \cos(\theta_0)/\lambda_c)\right) d\ell. \quad (13.14)$$

As desired, $\sigma_{\bar{\Omega}_p}^2 \rightarrow 0$ as $L \rightarrow \infty$. When L is large, $\bar{\Omega}_p$ 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) [143, 144].

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

$$1\sigma \text{ spread} = 10 \log_{10} \left\{ \frac{\Omega_p + \sigma_{\bar{\Omega}_p}}{\Omega_p - \sigma_{\bar{\Omega}_p}} \right\} \quad (13.15)$$

with the interpretation that $P[|\bar{\Omega}_p(\text{dB}) - \Omega_p(\text{dB})| \leq 1\sigma \text{ spread}] = 0.68$.¹ Observe from (13.14) and (13.15) that the accuracy of the local mean estimate depends on K , L , and θ_0 . Figure 13.4 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. 13.5. When $\theta_0 = 90^\circ$ the 1σ spread is minimized, resulting in the best estimate of the local mean. Conversely, the worst estimates

¹The probability of lying within one standard deviation of the mean of a Gaussian random variable is 0.68.

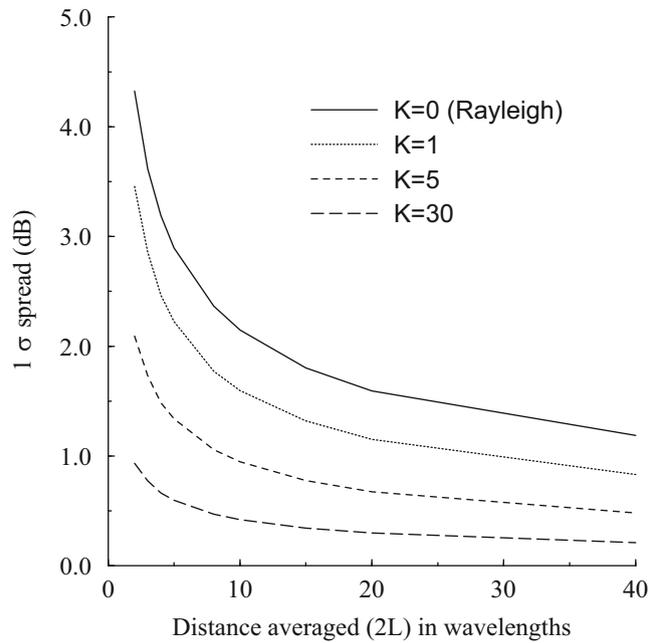


Fig. 13.4 Dependency of the 1σ spread for squared-envelope samples on the averaging distance ($2L$) and Rice factor K when $\theta_0 = 60^\circ$, from [28]

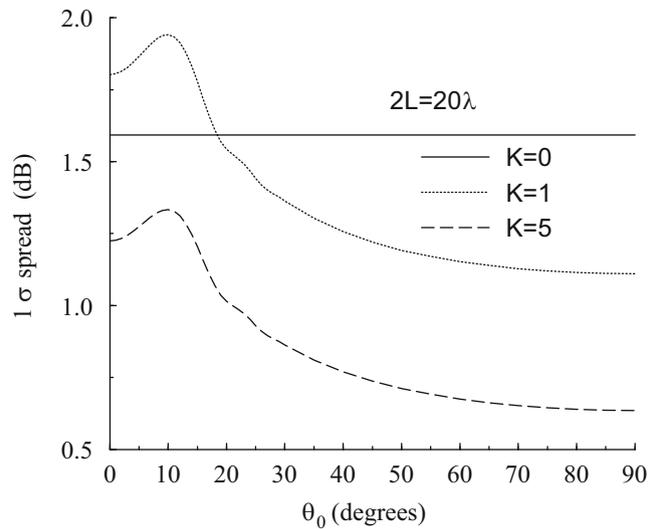


Fig. 13.5 Dependency of the 1σ spread on the specular angle $\theta_0 \bmod 90^\circ$, from [28]

occur for small θ_0 (in the neighborhood of 10° in Fig. 13.5). The actual angle that the maximum occurs is a function of L , and it can easily be shown that the 1σ 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 depends on K and θ_0 .

13.3.2 Choosing the Proper Number of Samples to Average

Most practical signal strength estimators use samples of the signal strength rather than analog averaging. The number and spacing of samples to provide a good estimate of the local mean must be determined. Consider the sampled composite squared-envelope

$$\alpha_c^2[i] \triangleq \alpha_c^2(iS) \quad (13.16)$$

where S is the spatial sampling period, and i is an integer. Then the spatial average

$$\bar{\Omega}_p = \frac{1}{N} \sum_{i=0}^{N-1} \alpha_c^2[i] \quad (13.17)$$

can be used to determine an unbiased estimate of the local mean Ω_p . As with analog averaging, the variance of this estimate can be used to measure its accuracy, where

$$\sigma_{\bar{\Omega}_p}^2 = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E[\alpha_c^2[i]\alpha_c^2[j]] - (E[\bar{\Omega}_p])^2. \quad (13.18)$$

By using (2.94), the transformation $f_m\tau = \ell/\lambda_c$, and the symmetric properties of the autocovariance function, (13.18) becomes

$$\begin{aligned} \sigma_{\bar{\Omega}_p}^2 &= \frac{\lambda_{\alpha^2\alpha^2}(0)}{N} + 2 \sum_{j=1}^{N-1} \left(\frac{N-j}{N^2} \right) \lambda_{\alpha^2\alpha^2}(Sj) \\ &= \left(\frac{\Omega_p}{K+1} \right)^2 \left(\frac{1+2K}{N} + 2 \sum_{j=1}^{N-1} \left(\frac{N-j}{N^2} \right) J_0(2\pi Sj/\lambda_c) \left(J_0(2\pi Sj/\lambda_c) + 2K \cos(2\pi Sj \cos(\theta_0)/\lambda_c) \right) \right), \end{aligned} \quad (13.19)$$

where S is measured in wavelengths (λ_c). Note that $\sigma_{\bar{\Omega}_p}$ depends on N , K , S , and θ_0 . Fortunately, the effect of each parameter is nearly independent of the others. Figure 13.6 illustrates the relationship between S and K for $\theta_0 = 0^\circ$, 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 1σ spread in a manner similar to analog averaging in Fig. 13.4. The discontinuities in Fig. 13.6 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 ($\theta_0 = 0^\circ$ in Fig. 13.5) over the same spatial distance. Similar to Fig. 13.5, it is also observe that small Rice factors, e.g., $K = 0.1$ and $K = 1$, at $\theta_0 = 0^\circ$ increase the 1σ 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 (2.94) and plotted in Fig. 2.16.

Although it is often assumed that $\theta_0 = 0^\circ$ in our treatment, Fig. 13.7 shows the relationship between the 1σ 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 1σ spread except for some small angles as shown in Fig. 13.5; it also shifts the spike at $0.5\lambda_c$ to the right, because the first side lobe of (2.94) shifts as θ_0 increases.

To summarize, the spatial averaging distance that is needed to obtain the local mean 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 $20\lambda_c$ should be sufficient for most applications.

13.4 Velocity Estimation

Temporal based hard 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. 13.1, where a MS traveling from BS_0 has a log-normal shadowed Ricean faded LoS signal from BS_0 and a log-normal shadowed Rayleigh faded 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 hard 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 [233]. 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. 13.1

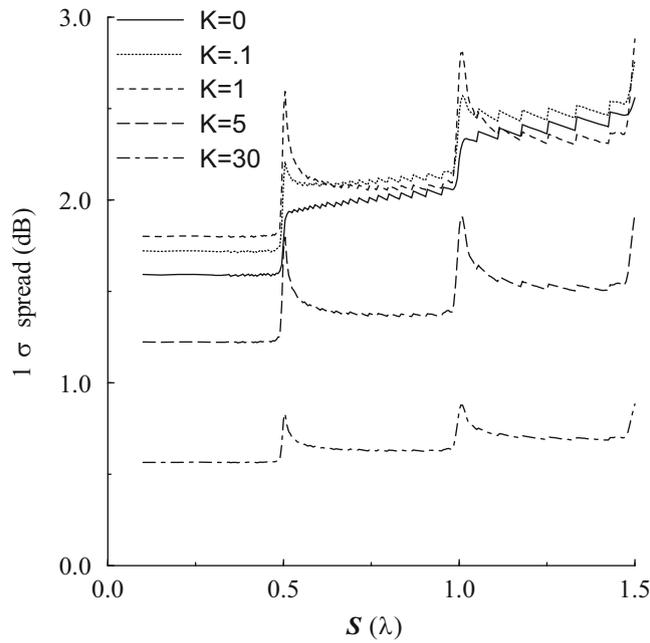


Fig. 13.6 1 σ spread versus S for various K , with $\theta_0 = 0^\circ$, distance averaged = $20\lambda_c$, from [28]

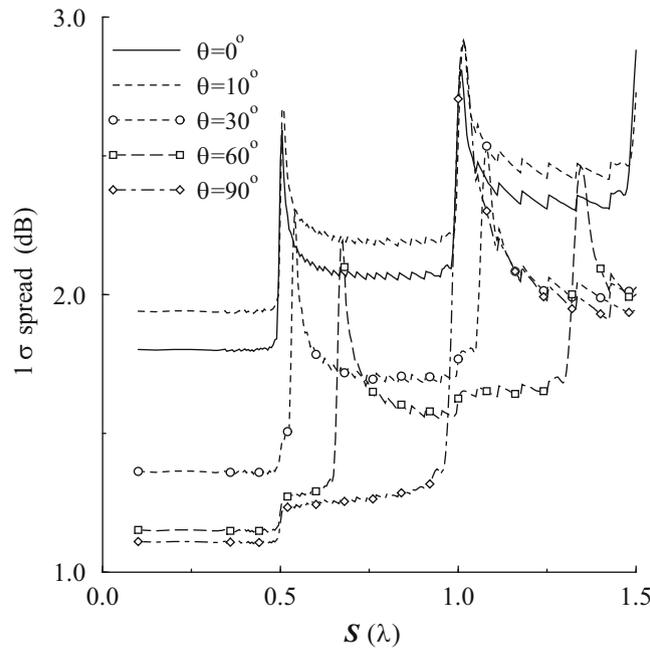


Fig. 13.7 1 σ spread versus S (in wavelengths) for various θ_0 , $K = 1$, distance averaged = $20\lambda_c$, from [28]

along with the received signal strength profile in Fig. 2.67. 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 (2.6.1). The simulation of a temporal power averaging hard handoff algorithm with a hysteresis $H = 8$ dB, averaging window duration of 2.27 s, and a window overlap of $2.27/2 = 1.135$ s gives the handoff performance shown by the lines in Fig. 13.13.² The hard handoff performance is evaluated by the mean number of

²A 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 13.5 further details the simulation.

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 $\Pr(BS_1) = 0.9$ at the abscissa. This distance gives a measure of the handoff delay, assuming that handoffs will occur between BS_0 and BS_1 only.

Figure 13.13 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 hard handoff algorithm should have at least 90% of the MSs handed off before a distance d_{cutoff} , where d_{cutoff} is chosen as the distance where the mean signal strength is $2\sigma_\Omega$ above -90 dBm. If $\sigma_\Omega = 6$ dB and the data from Fig. 2.67 is used, then a signal strength of $-90 + 12 = -78$ dBm occurs at 283 m for BS_0 . Hence, if the velocity adaptive hard handoff algorithm can adapt to the point at 5 km/h in Fig. 13.13, corresponding to handoff requests at a distance 262 m, and the maximum speed of a MS turning the corner is 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 macrocellular systems this network delay is acceptable, implying the usefulness of velocity adaptive hard handoff algorithms discussed here. In the above example the 5 km/h point on the curve in Fig. 13.13 was chosen as the desired operating point, because the best hard 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 hard handoff algorithms.

Hierarchical cellular architectures use microcells to serve high density low mobility MSs and macrocells to serve low density high mobility MSs. Velocity estimation is necessary so that MSs can be assigned to the proper hierarchical layer depending on their mobility. If high mobility MSs are connected to microcells, then excessive handoffs may occur. On the other hand, if low mobility MSs are connected to macrocells, then the overall system capacity will suffer. Finally, for hierarchical TDMA cellular architectures the micro-to-macrocell and micro-to-microcell handoffs must be performed quickly to prevent dropped calls, especially for MSs that are connected to microcells and subsequently transition from low to high mobility. For these cases, velocity adaptive handoff algorithms are necessary to achieve the required low network handoff delay.

13.4.1 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 of the envelope $\alpha(t) = |g(t)| = |g_I(t) + jg_Q(t)|$ are functions of the MS velocity as discussed in Sect. 2.1.5. The envelope level crossing rate (LCR) is defined as the average number of positive going crossings per second, that the fading envelope $\alpha(t)$ makes at a predetermined envelope level R . Likewise, the zero crossing rate (ZCR) is defined as the average number of positive going zero crossings of the quadrature components $\hat{g}_I(t) = g_I(t) - m_I(t)$ and $\hat{g}_Q(t) = g_Q(t) - m_Q(t)$. Rice has derived the ZCR of $\hat{g}_I(t)$ or $\hat{g}_Q(t)$ as (2.112)

$$L_{\text{ZCR}_1} = \frac{1}{\pi} \sqrt{\frac{b_2}{b_0}} \quad (13.20)$$

and the envelope LCR with respect to the level R as, (2.100) and (2.101),

$$L_R = \int_0^\infty \dot{\alpha} p(R, \dot{\alpha}) d\dot{\alpha} = \frac{R(2\pi)^{-3/2}}{\sqrt{Bb_0}} \int_0^\infty \int_{-\pi}^\pi \dot{\alpha} \times \exp \left\{ -\frac{1}{2Bb_0} \left(B(R^2 - 2Rs \cos(\theta) + s^2) + (b_0\dot{\alpha} + b_1s \sin(\theta))^2 \right) \right\} d\theta d\dot{\alpha}, \quad (13.21)$$

where $B = b_0 b_2 - b_1^2$. The b_n can be obtained from (2.103) with appropriate modification to account for the presence of additive white Gaussian noise in the received complex envelope. That is, the LCR is computed from the quadrature components of the received complex envelope $\tilde{r}_I(t) = g_I(t) + \tilde{n}_I(t)$ and $\tilde{r}_Q(t) = g_Q(t) + \tilde{n}_Q(t)$ instead of the quadrature components of the received fading envelope $g_I(t)$ and $g_Q(t)$. If an isotropic antenna is assumed such that $G(\theta) = 1$, then

$$b_n = (2\pi)^n b_0 \int_0^{2\pi} \hat{p}(\theta) (f_m \cos(\alpha) - f_q)^n d\theta + (2\pi)^n \int_{-B_w/2}^{B_w/2} \frac{N_o}{2} f^n df, \quad (13.22)$$

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 [171]. The second term in (13.22) is due to additive white Gaussian noise (AWGN), 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. Note that B_w is determined by the receiver bandwidth and must be at least equal to the maximum Doppler frequency f_m in order to accurately estimate the LCR. However, if B_w is larger than f_m , then the LCR estimates will be corrupted by unnecessary noise. For the special case when $\theta_0 = 0^\circ$ and there is 2-D isotropic scattering, then $\hat{p}(\theta) = 1/(2\pi)$, $-\pi \leq \theta \leq \pi$ and (13.22) can be written as

$$b_n = (2\pi)^n \frac{b_0}{\pi} \int_{-f_m}^{f_m} \frac{f^n}{\sqrt{f_m^2 - f^2}} df + (2\pi)^n \int_{-B_w/2}^{B_w/2} \frac{N_o}{2} f^n df. \quad (13.23)$$

With Aulin's Ricean fading model with the means of $g_I(t)$ and $g_Q(t)$ defined in (2.59) and (2.60), respectively, the ZCR of $g_I(t)$ or $g_Q(t)$ is [283]

$$L_{\text{ZCR}_2} = L_{\text{ZCR}_1} \left(e^{-\gamma} I_0(\beta) + \frac{b^2}{2\gamma} I_e \left(\frac{\beta}{\gamma}, \gamma \right) \right), \quad (13.24)$$

where $I_0(x)$ is the zero-order modified Bessel function of the first kind, and

$$\gamma = \frac{a^2 + b^2}{4}, \quad \beta = \frac{a^2 - b^2}{4}, \quad a = \sqrt{2K}$$

$$I_e(k, x) = \int_0^x e^{-u} I_0(ku) du, \quad b = 2\pi f_m \cos(\theta_0) \sqrt{\frac{2Kb_0}{b_2}}, \quad (13.25)$$

and L_{ZCR_1} was defined in (13.20).

For macrocells, 2-D isotropic scattering is a reasonable assumption. However, for microcells the scattering is often non-isotropic. Nevertheless, one approach is to derive velocity estimators under the assumption of 2-D isotropic scattering with no AWGN, and afterwards study the effects of the mismatch caused by non-isotropic scattering and noise. Using (13.20) along with the definition for the b_n in (13.23) and $N_o = 0$ gives

$$L_{\text{ZCR}_1} = \sqrt{2}v/\lambda_c \quad (13.26)$$

and (2.109)

$$L_R = (v/\lambda_c) \sqrt{2\pi(K+1)} \rho e^{-K-(K+1)\rho^2} I_0 \left(2\rho \sqrt{K(K+1)} \right), \quad (13.27)$$

where $\rho = R/R_{\text{rms}}$, where $R_{\text{rms}} = \sqrt{\Omega_p}$ is the rms envelope level. Likewise, for Aulin's Ricean fading model with the means in (2.59) and (2.60), and $\theta_0 = 0^\circ$,

$$\gamma = \frac{3K}{2}, \quad \beta = -\frac{K}{2}, \quad a = \sqrt{2K}, \quad b = 2\sqrt{K}, \quad (13.28)$$

and L_{ZCR_2} reduces to

$$L_{ZCR_2} = (v/\lambda_c)\sqrt{2} \left(e^{-3K/2} I_0 \left(-\frac{K}{2} \right) + \frac{4}{3} I_e \left(-\frac{1}{3}, \frac{3K}{2} \right) \right). \quad (13.29)$$

Clearly, the LCR and ZCR are proportional to the velocity v and, hence, can be used to derive a velocity estimate. However, it remains to be seen if they are robust to the Rice factor K , non-isotropic scattering, additive noise, and other factors. The robustness with respect to K is first considered, and the other factors are treated afterwards.

L_{ZCR_1} is not affected by K . Figure 2.17 in Sect. 2.1.5 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 LCR can be used to provide a velocity estimate that is robust to K . Consequently, the steps for using the LCR (or ZCR) of $\alpha(t)$ (or $g_I(t)$ or $g_Q(t)$) for velocity estimation are; determine R_{rms} (or $m_I(t)$ or $m_Q(t)$), estimate the number of crossings per second $\hat{L}_{R_{\text{rms}}}$ (or \hat{L}_{ZCR_1}), and use (13.27) to solve for v , with $\rho = 1$ and $K = 0$ (or (13.26) for ZCR_1). Thus, the following velocity estimators are robust with respect to Rice factor K :

$$\hat{v}_{ZCR_1} \approx \frac{\lambda_c \hat{L}_{ZCR}}{\sqrt{2}}, \quad \hat{v}_{LCR} \approx \frac{\lambda_c \hat{L}_{R_{\text{rms}}}}{\sqrt{2\pi} e^{-1}}. \quad (13.30)$$

Figure 13.8 shows the effect of K and θ_0 on L_{ZCR_2} . Notice that if the angle of the specular component is $\theta_0 = 0^\circ$ or 180° , then L_{ZCR_2} can have up to 40% relative error. Consequently, a non-zero value of K should be chosen as a default value to minimize the effect of K . Choosing $K \approx 0.61$ yields a maximum error of at most 20%. In this case, the velocity estimate from (13.29) becomes

$$\hat{v}_{ZCR_2} \approx \frac{\lambda_c \hat{L}_{ZCR_2}}{1.2\sqrt{2}}. \quad (13.31)$$

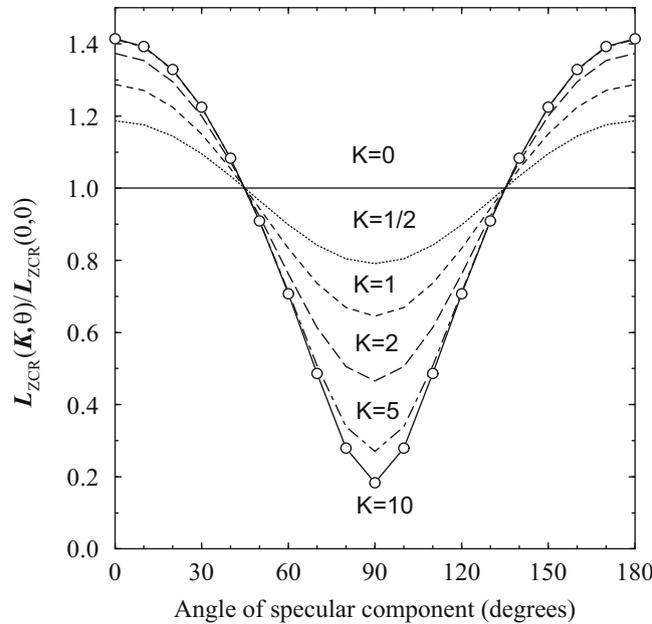


Fig. 13.8 Normalized zero crossing rate versus K and θ_0 , from [28]

13.4.2 Covariance Approximation Methods

A velocity estimator has been proposed by Holtzman and Sampath [167, 288] that relies upon an estimate of the autocovariance between faded samples $r[i]$, where the $r[i]$ can be samples of the envelope, squared-envelope, or log-envelope. With this method, referred to here as the covariance (COV) method, the statistic

$$V = \frac{1}{N} \sum_{k=1}^N (r[k + \tau] - r[k])^2 \quad (13.32)$$

is calculated. If N is large and ergodicity applies, then V can be replaced by the ensemble average

$$E[V] = 2\lambda_{rr}(0) - 2\lambda_{rr}(\tau), \quad (13.33)$$

where $\lambda_{rr}(\tau)$ denotes the autocovariance of $r[k]$. The general form for $\lambda_{rr}(\tau)$, assuming squared-envelope samples, can be derived from [22] and [288] as

$$\begin{aligned} \lambda_{rr}(\tau) = & 4a(\tau) \left(a(\tau) + s^2 \cdot \cos(2\pi f_m \tau \cos(\theta_0)) \right) + 4c^2(\tau) + \frac{2N_o a(\tau) \sin(B_w \pi \tau)}{\pi \tau} \\ & + \frac{4KN_o b_0 \cos(2\pi f_m \tau \cos(\theta_0)) \sin(B_w \pi \tau)}{\pi \tau} + \frac{N_o^2 \sin(B_w \pi \tau)^2}{\pi^2 \tau^2}, \end{aligned} \quad (13.34)$$

where [22]

$$a(\tau) = b_0 \int_0^{2\pi} \hat{p}(\theta) \cos(2\pi f_m \tau \cos(\theta)) d\theta \quad (13.35)$$

$$c(\tau) = b_0 \int_0^{2\pi} \hat{p}(\theta) \sin(2\pi f_m \tau \cos(\theta)) d\theta. \quad (13.36)$$

This estimator depends on $\hat{p}(\theta)$ and, hence, is also a function of the scattering environment. Like the LCR estimator, isotropic scattering is first assumed without additive noise to derive a velocity estimator and afterwards the effect of non-isotropic scattering and noise is considered.³ Consider the scattering environment shown in Fig. 2.7, characterized by 2-D isotropic scattering plus a LoS or specular component arriving at angle $\theta_0 = \pi/2$. If squared-envelope samples are used, then using (2.94) gives

$$\bar{V} = E[V] = 2 \left(\frac{\Omega_p}{K+1} \right)^2 \left((1+2K) - J_0(2\pi f_m \tau) \left(J_0(2\pi f_m \tau) + 2K \cos(2\pi f_m \tau \cos(\theta_0)) \right) \right), \quad (13.37)$$

which is dependent on K and θ_0 . If $\lambda_{rr}(0)$ is known exactly, then the bias with respect to K can be eliminated for small τ by the normalization [288]

$$\frac{\bar{V}}{\lambda_{rr}(0)} \approx (2\pi v \tau_i / \lambda_c)^2 \frac{1 + 2K + K \cos(2\theta_0)}{(1 + 2K)} \quad (13.38)$$

so that [288]

$$\hat{v}_{\text{COV}} \approx \frac{\lambda_c}{2\pi \tau_i} \sqrt{\frac{\bar{V}}{\lambda_{rr}(0)}}, \quad (13.39)$$

where τ_i is the sample spacing in seconds/sample.

³Only isotropic scattering was considered in [167, 288].

In large co-channel interference situations it may be preferable to modify the above scheme since the empirical average in (13.32), and in particular $\lambda_{rr}(0)$, is sensitive to co-channel interference as shown in [183]. Consequently, defining

$$U(\tau) = \frac{1}{N} \sum_{k=1}^N r[k + \tau]r[k] - \left(\frac{1}{N} \sum_{k=1}^N r[k] \right)^2 \quad (13.40)$$

and $V_2 = 2U(\tau_1) - 2U(\tau_2)$, yields

$$E[V_2] = 2\lambda_{rr}(\tau_1) - 2\lambda_{rr}(\tau_2) \quad (13.41)$$

so that $E[V_2]/\lambda_{rr}(0)$ is equal to (13.38) with $\tau_t = \tau_2^2 - \tau_1^2$, and a result similar to (13.39) follows.

Whether V or V_2 is used, $\lambda_{rr}(0)$ is never known exactly and must be estimated by the MS in the same way that $m_I(t)$, $m_Q(t)$, and R_{rms} must be estimated in the ZCR and LCR methods, respectively. Consequently, to actually use (13.39) it must be shown or verified that

$$v \propto E \left[\sqrt{\frac{\bar{V}}{\lambda_{rr}(0)}} \right]. \quad (13.42)$$

This is analytically difficult, but simulation results in Sect. 13.5 suggest that (13.39) is a useful approximation to (13.42).

It is also shown in Appendix 13A that

$$\lim_{\tau \rightarrow 0} \hat{v}_{\text{COV}} = \lim_{\tau \rightarrow 0} \frac{\lambda_c}{2\pi\tau} \sqrt{\frac{\bar{V}}{\lambda_{rr}(0)}} = v \sqrt{\frac{1 + 2K + K \cos(2\theta_0)}{1 + 2K}}. \quad (13.43)$$

It follows from (13.39) and (13.43) that K and θ_0 cause at most 20% error in v [288], thus providing a velocity estimator that is reasonably robust with respect to K .

13.4.3 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,

$$\frac{\tilde{v}}{v} = \frac{\tilde{L}_{R_{\text{rms}}}(\tilde{b}_0, \tilde{b}_1, \tilde{b}_2)}{L_{R_{\text{rms}}}(b_0, b_1, b_2)} \quad (13.44)$$

and

$$\frac{\tilde{v}}{v} = \frac{\tilde{L}_{ZCR_1}}{L_{ZCR_1}} = \sqrt{\frac{\tilde{b}_2}{\tilde{b}_0} \cdot \frac{b_0}{b_2}}, \quad (13.45)$$

respectively, where \tilde{v} denotes the corrupted velocity estimate, and the level crossing rates $\tilde{L}_{R_{\text{rms}}}(\tilde{b}_0, \tilde{b}_1, \tilde{b}_2)$ and $L_{R_{\text{rms}}}(b_0, b_1, b_2)$ are given by (13.21) with the appropriate values of \tilde{b}_n and b_n , respectively. Little simplification results for the LCR method in general. However, when $K = 0$ (13.45) simplifies to [171]

$$\frac{\tilde{v}}{v} = \sqrt{\frac{\tilde{b}_2/\tilde{b}_0 - \tilde{b}_1^2/\tilde{b}_0^2}{b_2/b_0 - b_1^2/b_0^2}}. \quad (13.46)$$

For Aulin's fading model with the means in (2.59) and (2.60), the sensitivity of the ZCR is

$$\frac{\tilde{v}}{v} = \frac{\tilde{L}_{ZCR_2}}{L_{ZCR_2}} = \frac{\tilde{L}_{ZCR_1}}{L_{ZCR_1}} \cdot \frac{e^{-\tilde{\gamma}} I_0(\tilde{\beta}) + \frac{\tilde{b}_2^2}{2\tilde{\gamma}} I_e\left(\frac{\tilde{\beta}}{\tilde{\gamma}}, \tilde{\gamma}\right)}{e^{-\gamma} I_0(\beta) + \frac{b_2^2}{2\gamma} I_e\left(\frac{\beta}{\gamma}, \gamma\right)}, \quad (13.47)$$

where $\tilde{\beta}$, $\tilde{\gamma}$, and \tilde{b} are given by (13.25) using (13.22) where appropriate. Likewise, for the covariance method using squared-envelope samples,

$$\frac{\tilde{v}}{v} = \frac{\sqrt{\frac{2\tilde{\lambda}_{rr}(0) - 2\tilde{\lambda}_{rr}(\tau)}{\tilde{\lambda}_{rr}(0)}}}{\sqrt{\frac{2\lambda_{rr}(0) - 2\lambda_{rr}(\tau)}{\lambda_{rr}(0)}}}. \quad (13.48)$$

13.4.3.1 Effect of the Scattering Distribution

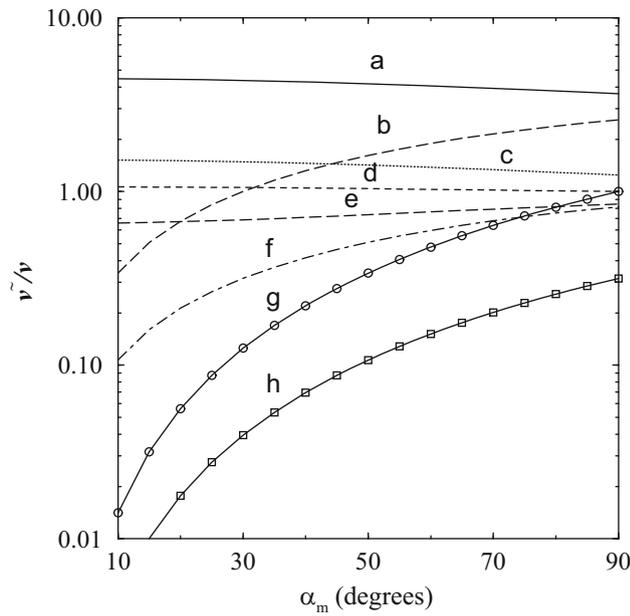
Here the sensitivity of the velocity estimators to the scattering distribution is studied by using four different non-isotropic scattering models. With the first model S1, plane waves arrive from one direction only with a varying degree of directivity as might happen when signals are channelized along 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 (2.52), where the vehicle motion is in the direction of $\theta = 0^\circ$, and θ_m determines the directivity of the incoming plane waves. Figure 2.9 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 $\hat{p}(\theta)/2$ and $\hat{p}(\theta - \pi)/2$ are combined to form the overall angle-of-arrival distribution. The resulting density is similar to Fig. 2.9 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 (1) isotropic scattering and, (2) 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 (13.44)–(13.47), 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 (13.48) can be found from the results in Appendix 13A with $N_o = 0$, or by using small values of τ in (13.48). Here, the latter is chosen with $\tau = 1/50$.

Figure 13.9 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. 13.9 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. 13.9 correspond to those that are designed for scattering model S1 with $\theta_m = 90^\circ$. By using Fig. 13.9 the relative robustness of the various velocity estimators to the scattering distribution has been summarized by the ranking in Table 13.1.

For urban microcell deployments, 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 $\theta_m = 90^\circ$ 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 (13.46) depends on \tilde{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 (13.21) for small θ_m . For large $\theta_m > 80^\circ$ 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 $\text{ZCR}_2, \text{ZCR}_{2,d}$ (ZCR_1 and $\text{ZCR}_{1,d}$ are independent of K). Also, velocity estimators that have been designed for scattering model S1 with $\theta_m = 90^\circ$ 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 , $\tilde{v}/v = 0.316$ 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 the signals that arrive at a MS from a distant cell or an umbrella cell, since they will tend to experience 2-D isotropic scattering.



Method	Scattering model							
	1		2		3		4	
	$K = 0$	$K = 1$	$K = 0$	$K = 1$	$K = 0$	$K = 1$	$K = 0$	$K = 1$
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. 13.9 Non-isotropic scattering effects.—corresponds to a curve that had insufficient precision to be reported. Superscript x^* denotes that the curve is approximately equal the curve labeled x , from [28]

Table 13.1 Robustness to the scattering distribution

Curve	Rank	\tilde{v}/v at $\theta_m = 10^\circ$	\tilde{v}/v at $\theta_m = 90^\circ$
d	Excellent	1.06	1.0
e, c	Very good	0.66, 1.52	0.85, 1.24
b, f	Good	0.34, 0.107	2.6, 0.82
a, g, h	Poor	4.5, 0.014, 0.004	3.6, 1.0, 0.32

13.4.3.2 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 Eqs. (13.44) to (13.48) under the assumption of 2-D isotropic scattering. With AWGN the rms value of the received complex envelope is $\tilde{R} = \sqrt{s^2 + 2b_0 + N_o B_w}$, and the values of \tilde{b}_n and b_n in (13.44) are

$$\tilde{b}_0 = b_0 + \frac{N_o B_w}{2} \tag{13.49}$$

$$\tilde{b}_2 = 2b_0 (\pi f_m)^2 + \frac{N_o B_w^3 \pi^2}{6} \tag{13.50}$$

$$b_2 = 2b_0 (\pi f_m)^2. \tag{13.51}$$

For the LCR velocity estimator, (13.44) and (13.21) yield, after considerable algebra,

$$\frac{\tilde{v}}{v} = \left(1 + \frac{K+1}{6\gamma_S} \left(\frac{B_w}{f_m}\right)^2\right)^{\frac{1}{2}} \frac{\sqrt{\gamma_S(\gamma_S+1)} I_0\left(\frac{2\sqrt{\gamma_S(\gamma_S+1)K(K+1)}}{\gamma_S+K+1}\right)}{\gamma_S+K+1} \frac{I_0\left(\frac{2\sqrt{K(K+1)}}{\gamma_S+K+1}\right)}{I_0\left(\frac{2\sqrt{K(K+1)}}{\gamma_S+K+1}\right)} \exp\left\{2K+1 - \frac{\gamma_S(2K+1)+K+1}{\gamma_S+K+1}\right\}, \quad (13.52)$$

where

$$\gamma_S \triangleq \frac{s^2 + 2b_0}{N_o B_w} = \frac{\Omega_p}{N_o B_w} \quad (13.53)$$

is defined as the signal-to-noise ratio. Likewise, for the ZCR velocity estimator ZCR₁, (13.45) becomes

$$\frac{\tilde{v}}{v} = \sqrt{\frac{\gamma_S + \left(\frac{B_w}{f_m}\right)^2 \frac{K+1}{6}}{\gamma_S + K + 1}}. \quad (13.54)$$

For Aulin's fading model with the means in (2.59) and (2.60), the effect of AWGN on ZCR₂ can be obtained from (13.47) with $\tilde{L}_{ZCR_1}/L_{ZCR_1}$ in (13.54),

$$\tilde{\beta} = \frac{K}{2} \left(1 - 2\cos^2(\theta_0) \left(\frac{\gamma_S + K + 1}{\gamma_S + \frac{K+1}{6} \left(\frac{B_w}{f_m}\right)^2}\right)\right) \quad (13.55)$$

$$\tilde{\gamma} = \frac{K}{2} \left(1 + 2\cos^2(\theta_0) \left(\frac{\gamma_S + K + 1}{\gamma_S + \frac{K+1}{6} \left(\frac{B_w}{f_m}\right)^2}\right)\right) \quad (13.56)$$

$$\tilde{b} = 2\sqrt{K} \cos(\theta) \sqrt{\frac{\gamma_S + K + 1}{\gamma_S + \frac{K+1}{6} \left(\frac{B_w}{f_m}\right)^2}} \quad (13.57)$$

and a , b , γ , and β given by (13.28).

In [288], the effect of AWGN on the COV velocity estimator has been derived as a function of $\tau > 0$. Here a closed form analytic result is provided for the effect of AWGN on the COV velocity estimate for the limiting case when $\tau \rightarrow 0$. The limiting case is important when analyzing the effects of AWGN, since (13.38) is only valid for small τ . Consequently, the $\lim_{\tau \rightarrow 0} \tilde{v}/v$ in (13.48) is found, and afterwards, the effect of $\tau > 0$ in (13.48) is compared. It is shown in Appendix 13A that

$$\lim_{\tau \rightarrow 0} \tilde{v}/v = \frac{\sqrt{\zeta}}{\sqrt{\frac{(1+2K+K\cos(2\theta_0))}{(1+2K)}}}, \quad (13.58)$$

where ζ is given by (13A.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 2-D isotropic scattering.

It is apparent from (13.52), and (13.54)–(13.58) 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., $B_w \gtrsim \hat{v}/\lambda_c$. Figure 13.10 shows the effect of AWGN on each of the velocity estimators with respect to K , γ_S , and v , assuming $\theta_0 = 0^\circ$ (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 . It must also be remembered that Fig. 13.10 shows the worst case performance of the COV method as $\tau \rightarrow 0$. 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 (13.48), then

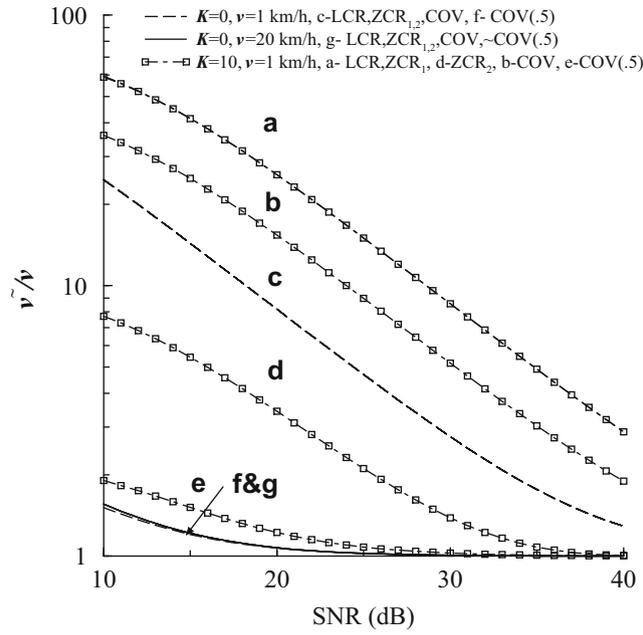


Fig. 13.10 The effect of AWGN on the velocity estimates. $\text{COV}(0.5) \Rightarrow 2\pi v\tau/\lambda_c = 0.5$, from [28]

a large reduction in the effect of AWGN is realized, as shown by the curves labeled $\text{COV}(0.5)$ in Fig. 13.10. 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.

13.5 Velocity Adaptive Hard Handoff Algorithms

To study velocity adaptive handoff algorithms Rayleigh fading ($K = 0$) is assumed, 2-D 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, the latter is assumed. To reduce the variance in the velocity estimate, a sum of weighted past velocity estimates is performed using an exponential window average of past estimates, i.e.,

$$\check{v}(n) = a\check{v}(n-1) + (1-a)\hat{v}(n), \quad (13.59)$$

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 Fig. 13.1 along with the corresponding signal strength profile in Fig. 2.67 was performed. The path loss was assumed to follow the model in (2.412) with $a = 2$, $b = 2$, $g = 150$ m, and $d_c = 255$ m. Spatially correlated log-normal shadows were simulated having a standard deviation of $\sigma_\Omega = 6$ dB and $\zeta_{30} = 0.1$ using the approach in Sect. 2.6.1. Envelope fading was simulated using Jakes' method in Sect. 2.5.2.2. Two-branch receiver antenna diversity was assumed, so that the $\hat{v}(n)$ in (13.59) represent the average estimate out of the diversity branches at position n , and $a = 0.1$ in (13.59) unless otherwise specified.

As mentioned previously, Fig. 13.13 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 approximately $20\lambda_c$ with a window overlap of approximately $10\lambda_c$.

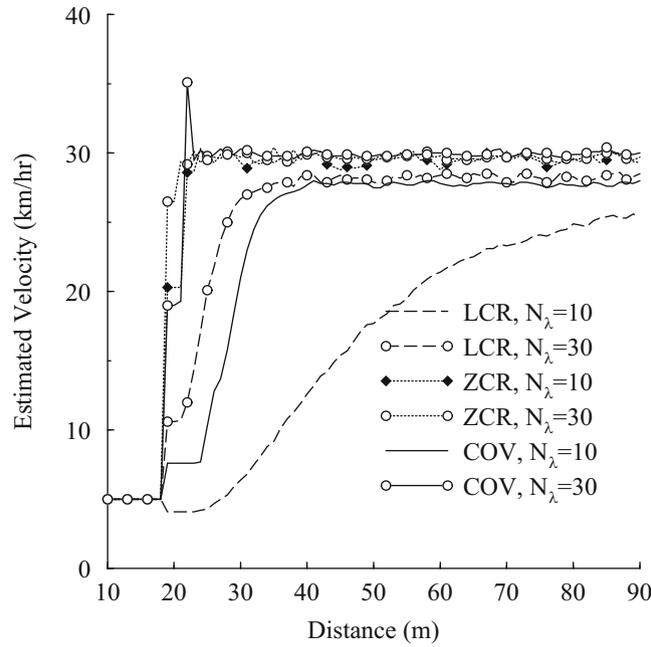


Fig. 13.11 The effect of N_λ on the mean response time to a change in velocity. $a = 0.1$, $W_l = 10 \lambda_c$, from [28]

A total of 1000 runs were made from BS_0 to BS_1 , and the 95% confidence intervals were calculated for (1) the velocity at 100 m, (2) the corner at 255 m, and (3) the probability of being assigned to BS_0 at 255 m. This resulted in a 95% confidence interval spread of $\hat{v} \pm 0.5$ km/h and $\Pr(BS_0) \pm 0.025$. Likewise, the mean number of handoff values had a 95% confidence interval spread of approximately 0.05 (mean number of handoffs ± 0.025).

13.5.1 Effect of N_λ

To examine the effect of N_λ , assume that $a = 0.1$ 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. 13.1 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. Figure 13.11 shows the effect of N_λ on the velocity estimate, in the first 90 m of the trajectory 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 Sect. 13.4.3.2 where $2\pi v\tau/\lambda_c = 0.5$ confirm that the effects of AWGN can be mitigated by using a larger sample spacing without drastically affecting the velocity estimate. The simulations used an estimate of the rms value R_{rms} in the LCR method and an estimate of the variance $\lambda_{rr}(0)$ 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 has little effect, confirming the claimed robustness of the estimators. Over the 1000 runs, the ZCR had the smallest velocity estimate error variance followed by the COV and LCR methods, respectively.

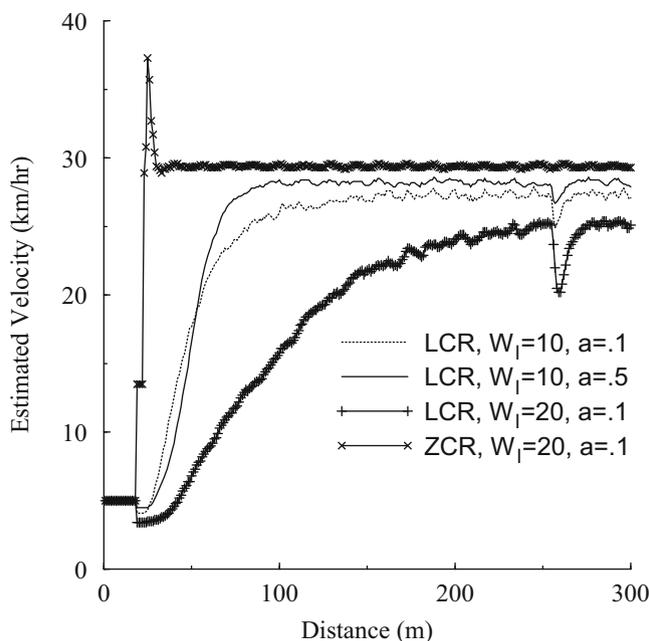


Fig. 13.12 The effect of a and W_l on the mean response and corner effects, from [28]

13.5.2 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. 13.12 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 [24].

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, 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 overlapped windows are used 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.

13.5.3 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. 13.13 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 $a = 0.1$, $W_l = 10 \lambda_c$ with an initial startup velocity of 5 km/h. The mean number of handoffs was 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.

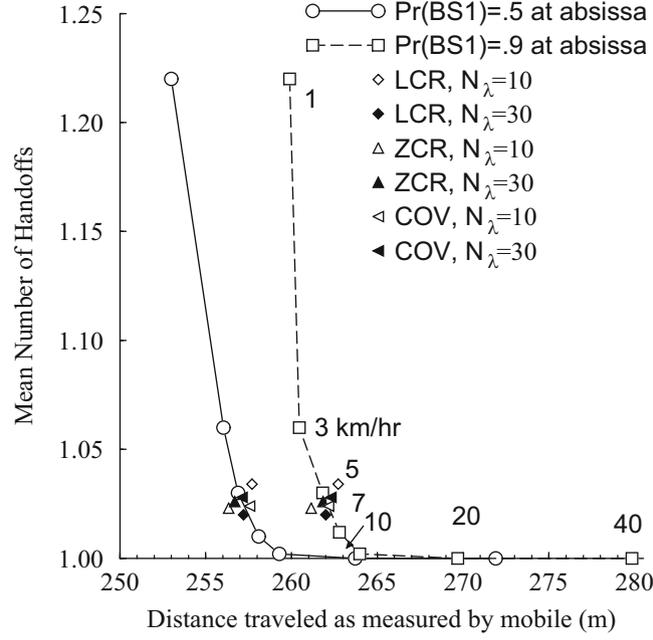


Fig. 13.13 Handoff performance of a 2.27 s averaging handoff algorithm in comparison with a velocity adaptive handoff algorithm using the LCR, ZCR, or covariance method for velocity control. $H = 8$ dB, $\sigma_{\Omega} = 6$ dB, from [28]

13.6 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 handoff if the average signal strength of the target BS is at least H dB larger than that of the serving BS [28, 150, 332, 333]. 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 [332, 333] 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 [373], similar to the one in (13.4).

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. Envelope fading is neglected under the assumption that the received signal strength estimates are averaged by using a window with an appropriate spatial length as explained in Sect. 13.3 and using velocity adaptive averaging as discussed in Sect. 13.5. 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 $\Omega_{0 \text{ (dB)}}(d)$ and $\Omega_{1 \text{ (dB)}}(d)$ that are received from BS_0 and BS_1 , respectively, are (1.4)

$$\Omega_{0 \text{ (dB)}}(d) = \Omega_{\text{(dB)}}(d_0) - 10\beta \log_{10}\{d/d_0\} + \epsilon_0 \text{ (dB)} \quad (13.60)$$

$$\Omega_{1 \text{ (dB)}}(d) = \Omega_{\text{(dB)}}(d_0) - 10\beta \log_{10}\{(D-d)/d_0\} + \epsilon_1 \text{ (dB)}, \quad (13.61)$$

where d is the distance between BS_0 and the MS. The parameters $\epsilon_0 \text{ (dB)}$ and $\epsilon_1 \text{ (dB)}$ 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,

$$\bar{\Omega}_{0 \text{ (dB)}}(d) = \frac{1}{d_{av}} \int_0^d \Omega_{0 \text{ (dB)}}(d-x) e^{-x/d_{av}} dx \quad (13.62)$$

$$\bar{\Omega}_{1 \text{ (dB)}}(d) = \frac{1}{d_{av}} \int_0^d \Omega_{1 \text{ (dB)}}(d-x) e^{-x/d_{av}} dx. \quad (13.63)$$

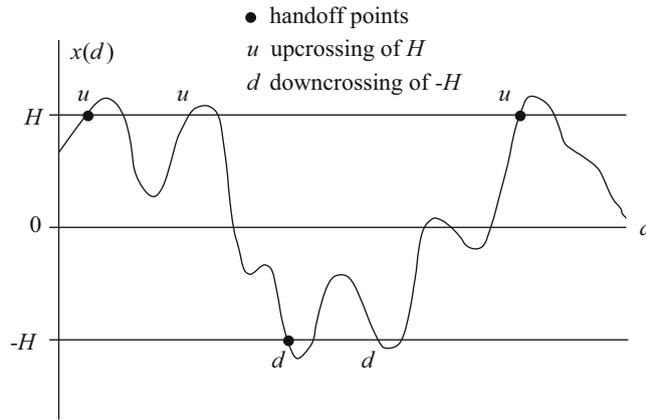


Fig. 13.14 Handoff initiation points with their associated hysteresis level crossings. The MS is moving from BS₀ to BS₁ and assumed to be communicating with BS₁ at the beginning of the interval shown, from [333]

To describe the signal strength based handoff algorithm, let

$$x(d) = \overline{\Omega}_0 \text{ (dB)}(d) - \overline{\Omega}_1 \text{ (dB)}(d) \quad (13.64)$$

denote the difference between the averaged signal strength estimates for BS₀ and BS₁. Consider the crossings of $x(d)$ with respect to the hysteresis levels $\pm H$ dB as illustrated in Fig. 13.14. 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 [333]. 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} [333].

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 [333]

$$p_{ho}(n) = p_d(n)p_{lu}(n) + p_u(n)(1 - p_{lu}(n)), \quad (13.65)$$

where $p_u(n)$ and $p_d(n)$ are the probability of an up-crossing or down-crossing in the n th interval, and $p_{lu}(n)$ is the probability that the last event was an up-crossing. In other words, the MS was assigned to BS₀ at the beginning of the n th interval. This can happen in one of the two mutually exclusive ways: (1) there is an up-crossing but no down-crossing in the $(n - 1)$ th interval, and (2) there are no crossings in 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 $p_{lu}(n - 1)$ [333]:

$$p_{lu}(n) = p_u(n - 1)(1 - p_d(n - 1)) + (1 - p_u(n - 1))(1 - p_d(n - 1))p_{lu}(n - 1). \quad (13.66)$$

As detailed in [333], 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 $\sigma_x'^2(d)$ of $x(d)$. These in turn are functions of the statistics of $\Omega_0 \text{ (dB)}(d)$ and $\Omega_1 \text{ (dB)}(d)$, which depend on the path loss and shadowing. The statistics of $\Omega_0 \text{ (dB)}(d)$ and $\Omega_1 \text{ (dB)}(d)$ are first evaluated, and afterwards the appropriate expressions for $p_d(n)$ and $p_u(n)$ are derived.

As discussed in Chap. 3, co-channel interference can be assumed to add on a power basis [265, 296]. Hence, in the presence of N_I co-channel interferers the signals received from BS₀ and BS₁ are, respectively,

$$\Omega_0 \text{ (dB)}(d) = 10 \log_{10} \left\{ \sum_{k=0}^{N_I} \Omega_{0,k \text{ (dB)}}(d) \right\} \quad (13.67)$$

$$\Omega_{1 \text{ (dB)}}(d) = 10 \log_{10} \left\{ \sum_{k=0}^{N_I} \Omega_{1,k \text{ (dB)}}(d) \right\}, \quad (13.68)$$

where $\Omega_{0,0 \text{ (dB)}}(d)$ and $\Omega_{1,0 \text{ (dB)}}(d)$ are the power of the desired signals from BS₀ and BS₁, respectively, and $\Omega_{0,k \text{ (dB)}}(d)$ and $\Omega_{1,k \text{ (dB)}}(d)$, $k = 1, \dots, N_I$ are the powers of the interfering co-channel signals received at the same BSs. Once again, the $\Omega_{0,k \text{ (dB)}}(d)$ and $\Omega_{1,k \text{ (dB)}}(d)$ are log-normally distributed. As discussed in Sect. 3.1, the sum of log-normal random variables can be approximated by another log-normal random variable and, hence, $\Omega_{0 \text{ (dB)}}(d)$ and $\Omega_{1 \text{ (dB)}}(d)$ remain Gaussian. Here, the approximations suggested by Fenton–Wilkinson [265, 296], and Schwartz & Yeh [296] are considered. As a further benefit, this will allow us to compare the usefulness of these two log-normal approximations for this application.

Following the notation established in Sect. 3.1, define $\hat{\Omega} = \xi \Omega_{\text{(dB)}}$, where $\xi = \ln(10)/10 = 0.23026$. If the N_I interferers for BS₀ have means $\mu_{\hat{\Omega}_{0,k}}(d)$ and variance $\sigma_{\hat{\Omega}_{0,k}}^2$, then the mean and variance of $\hat{\Omega}_0(d)$ using the Fenton–Wilkinson approach are

$$\mu_{\hat{\Omega}_0}(d) = \frac{\sigma_{\hat{\Omega}}^2 - \sigma_{\hat{\Omega}_0}^2(d)}{2} + \ln \left(\sum_{k=0}^{N_I} e^{\mu_{\hat{\Omega}_{0,k}}(d)} \right) \quad (13.69)$$

$$\sigma_{\hat{\Omega}_0}^2(d) = \ln \left((e^{\sigma_{\hat{\Omega}}^2} - 1) \frac{\sum_{k=0}^{N_I} e^{2\mu_{\hat{\Omega}_{0,k}}(d)}}{\left(\sum_{k=0}^{N_I} e^{\mu_{\hat{\Omega}_{0,k}}(d)} \right)^2} + 1 \right), \quad (13.70)$$

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 & Yeh's approach is a 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

$$\mu_{\hat{\Omega}_0}(d) = \mu_{\hat{\Omega}_{0,0}}(d) + G_1 \quad (13.71)$$

$$\sigma_{\hat{\Omega}_0}^2(d) = \sigma_{\hat{\Omega}_{0,0}}^2 - G_1^2 - 2\sigma_{\hat{\Omega}_{0,0}}^2 G_3 + G_2, \quad (13.72)$$

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 of the two log-normal approximations, 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 Sect. 2.1.5. In particular,

$$\begin{aligned} p_u(n) &= d_s \int_0^\infty \dot{x} p(H, \dot{x}) d\dot{x} \\ p_d(n) &= d_s \int_{-\infty}^0 |\dot{x}| p(H, \dot{x}) d\dot{x}, \end{aligned} \quad (13.73)$$

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,

$$p_u(kd_s) = \frac{d_s}{\sqrt{2\pi b_0}} \exp \left\{ -\frac{(H - \mu_x(kd_s))^2}{2b_0} \right\} \left(\mu_{\dot{x}}(kd_s) \mathcal{Q} \left(-\frac{\mu_{\dot{x}}(kd_s)}{\sqrt{b_2}} \right) + \sqrt{\frac{b_2}{2\pi}} \exp \left\{ -\frac{\mu_{\dot{x}}^2(kd_s)}{2b_2} \right\} \right), \quad (13.74)$$

where, from (2.102) with $f_q = 0$,

$$b_0 = \sigma_x^2(d) = 2 \int_0^\infty \hat{S}_{xx}(f) df \quad (13.75)$$

$$b_2 = \sigma_{\dot{x}}^2(d) = 2(2\pi)^2 \int_0^\infty f^2 \hat{S}_{xx}(f) df, \quad (13.76)$$

and $\hat{S}_{xx}(f)$ in this case represents the power spectrum of $x(d)$ that includes the effect of co-channel interference. Likewise,

$$p_d(kd_s) = \frac{d_s}{\sqrt{2\pi}b_0} \exp\left\{-\frac{(H + \mu_x(kd_s))^2}{2b_0}\right\} \left(-\mu_x(kd_s)Q\left(\frac{\mu_x(kd_s)}{\sqrt{b_2}}\right) + \sqrt{\frac{b_2}{2\pi}} \exp\left\{-\frac{\mu_x^2(kd_s)}{2b_2}\right\} \right). \quad (13.77)$$

The autocorrelation of $\Omega_{0 \text{ (dB)}}(d)$ or $\Omega_{1 \text{ (dB)}}(d)$ (equal to the shadow autocorrelation) without co-channel interference is modeled by

$$\phi_{\Omega_{\text{(dB)}}\Omega_{\text{(dB)}}}(d) = \sigma_{\Omega}^2 \exp(-|d_1 - d_2|/d_0), \quad (13.78)$$

where $d = d_1 - d_2$, and d_0 controls the decorrelation with distance. Let $\tilde{\phi}_{\Omega_{\text{(dB)}}\Omega_{\text{(dB)}}}(d)$ denote the same autocorrelation function when co-channel interference is present. The value $\tilde{\phi}_{\Omega_{\text{(dB)}}\Omega_{\text{(dB)}}}(0)$ can be accurately approximated by using either (13.70) or (13.72). An approximation of $\tilde{\phi}_{\Omega_{\text{(dB)}}\Omega_{\text{(dB)}}}(d)$ for $d > 0$ can be obtained by substituting σ_{Ω}^2 in (13.78) with the value obtained in (13.70) or (13.72). The accuracy of this approximation was tested through the simulation of mutually uncorrelated log-normal interferers, each having the shadow autocorrelation in (13.78) with $\sigma_{\Omega} = 6$ dB and $d_0 = 20$ m. Figure 13.15 shows the results and verifies that the proposed approximation of $\tilde{\phi}_{\Omega_{\text{(dB)}}\Omega_{\text{(dB)}}}(d)$ is fairly accurate. Also, very accurate modeling of $\tilde{\phi}_{\Omega_{\text{(dB)}}\Omega_{\text{(dB)}}}(d)$ is not essential in handoff analysis [332].

Using the above approximation gives

$$\hat{S}_{xx}(f) = \frac{2(\sigma_{\Omega_0}^2(d) + \sigma_{\Omega_1}^2(d))d_0}{(1 + d_0^2(2\pi f)^2)(1 + d_{\text{av}}^2(2\pi f)^2)}, \quad (13.79)$$

where $\sigma_{\Omega_0}^2(d)$ and $\sigma_{\Omega_1}^2(d)$ are obtained using either (13.70) or (13.72). This leads to

$$\sigma_x^2(d) = \frac{(\sigma_{\Omega_0}^2(d) + \sigma_{\Omega_1}^2(d))d_0}{d_0 + d_{\text{av}}}, \quad (13.80)$$

$$\sigma_x^2(d) = \frac{\sigma_x^2(d)}{d_{\text{av}} d_0}. \quad (13.81)$$

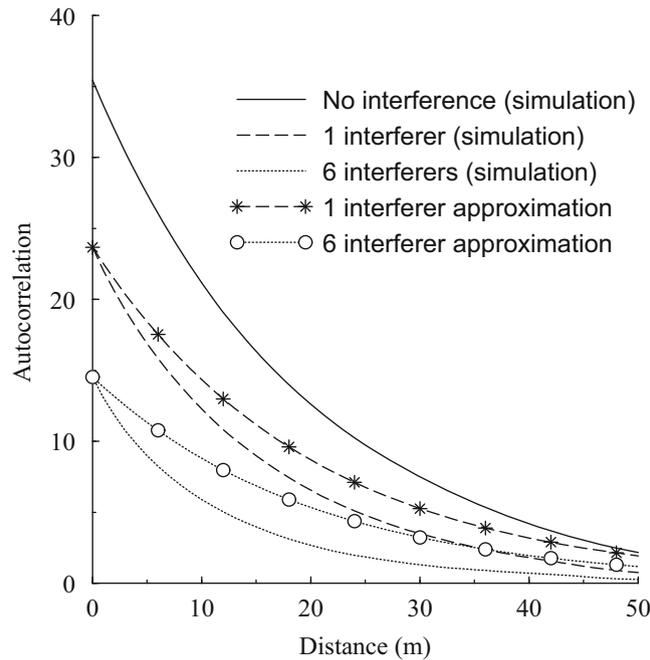


Fig. 13.15 Shadow autocorrelation with and without co-channel interference, from [25]

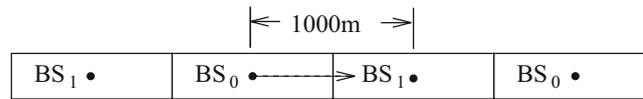


Fig. 13.16 Base station layout, MS route (dotted line), and location of co-channel interferers

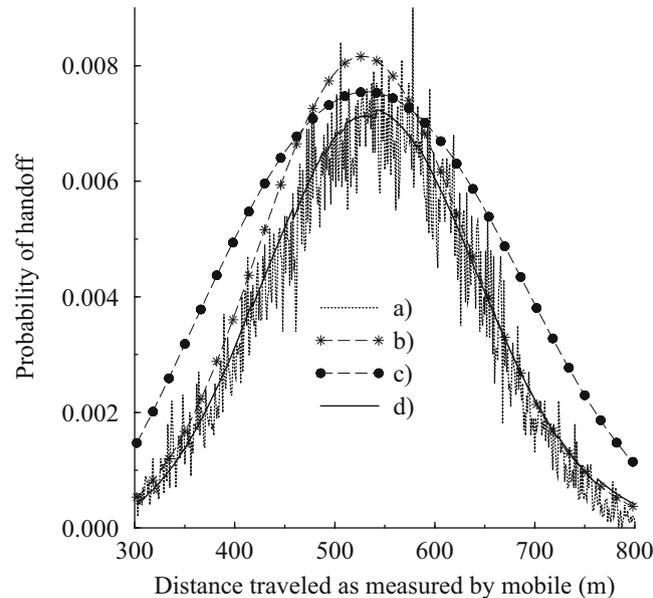


Fig. 13.17 Simulation vs. analytical model performance. a) Simulation of LoS handoff with co-channel interference, b) handoff analysis model in the absence of co-channel interference, c) handoff analysis model including co-channel interference and using the Fenton–Wilkinson log-normal approximation, d) handoff analysis model including co-channel interference and using the Schwartz & Yeh log-normal approximation, from [25]

13.6.1 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. 13.16. Assume a square-law path loss with distance (used here to accentuate the co-channel interference effects), $d_{av} = 10$ m, $d_0 = 20$ m, and choose $\sigma_\Omega = 4$ dB so that both the Fenton–Wilkinson and Schwartz & Yeh log-normal approximations are accurate. Figure 13.17 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 & Yeh’s method leads to an accurate prediction of the handoff probabilities while the Fenton–Wilkinson method leads to less 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.

13.7 CDMA Soft Handoff Analysis

In order to successfully deploy CDMA cellular systems it is essential to understand soft handoff behavior. As mentioned earlier, soft handoff has great impact on CDMA cellular system performance/capacity and studying its performance can provide crucial information on how the CDMA system performance can be optimized. Here, soft handoff in hierarchical CDMA architectures is considered. The treatment is divided into two main parts: interference analysis and handoff analysis. The first part of this section introduces an interference analysis, where the emphasis is on reverse link C/I performance and interference imbalance in hierarchical CDMA systems. When microcells are introduced to an existing macrocell layer the resulting interference imbalance between the layers can greatly impact the overall system performance. Therefore, it is important to characterize the interference in hierarchical CDMA architectures, and the interference analysis presented here provides a tool for studying the performance under soft handoff.

The second part of this section considers a soft handoff analysis method similar to that proposed in [10, 374], where a moving MS is tracked to determine its soft handoff active set membership. Such analysis is useful for determining cell boundaries and overall handoff efficiencies for a given set of handoff parameters. The studies in [10, 374] are limited to single MS and are not accurate when interference is taken into account. Here, soft handoff performance is considered in the presence of interference. This is accomplished by augmenting a user tracking soft handoff model with the results obtained from the interference analysis. The resulting model is useful for studying the impact of soft handoff parameters on soft handoff performance measures, such as handoff error probability and average active set membership. Finally, the effect of dynamic soft handoff parameter assignment is considered, where the soft handoff parameters are dynamically adjusted based on the given interference conditions. Dynamic parameter assignment offers a more efficient soft handoff mechanism than fixed assignment by reducing unnecessary soft handoff overhead.

13.7.1 System Model and Analysis

The channel model used here accounts for log-normal shadowing and path loss attenuation with distance.⁴ The link gain between a MS located at (r, θ) and BS_{*i*} is

$$\begin{aligned} G_i(r, \theta) &= d_i(r, \theta)^{-\beta} 10^{\epsilon_i/10} \\ G_i(r, \theta)_{(\text{dB})} &= -\beta 10 \log_{10}\{d_i(r, \theta)\} + \epsilon_i, \end{aligned} \quad (13.82)$$

where $d_i(r, \theta)$ is the distance between the MS and BS_{*i*}, β is the path loss exponent, and $10^{\epsilon_i/10}$ is the shadowing component such that ϵ_i has the normal distribution

$$\epsilon_i \sim \mathcal{N}(0, \sigma_\Omega^2), \quad (13.83)$$

where σ_Ω is the shadow standard deviation. Therefore, $G_i(r, \theta)$ also has log-normal distribution

$$G_i(r, \theta)_{(\text{dB})} \sim \mathcal{N}(\mu_i(r, \theta), \sigma_\Omega^2), \quad (13.84)$$

where

$$\mu_i(r, \theta) = -\beta 10 \log_{10}\{d_i(r, \theta)\}. \quad (13.85)$$

Since our analysis involves a multi-cell system, the propagation model also accommodates shadow correlation between the multiple BS links:

$$E[\epsilon_i \epsilon_j] = \rho \sigma_\Omega^2, \quad i \neq j. \quad (13.86)$$

13.7.1.1 Interference Analysis

The system model used here for illustration consists of three macrocells and single microcell embedded within the macrocell layer as shown in Fig. 13.18, with the understanding that the methodology can readily be extended to larger deployments. The macrocells and microcell both use omnidirectional BS antennas. The microcell location is specified by the distance, d_μ , and angle, θ_μ , with respect to BS₁. Each macrocell area contains N MSs and the microcell area contains M MSs. The MSs are assumed to be uniformly distributed within each cell area. It is important to realize that the MSs located within a macro- or microcell area are not necessarily served by the BS located at the center of that macro- or micro-cell. Moreover, the model is not restricted to uniform macrocells either. Different MS densities within the macrocells can be realized by assigning different values of N to the macrocells and, likewise, by assigning different values of M to the microcells should there be more than one microcell. The introduction of the microcell in Fig. 13.18 will introduce interference imbalance into

⁴One can incorporate Rayleigh/Nakagami fading into our analysis by using a log-normal approximation for the composite log-normal Rayleigh/Nakagami distribution.

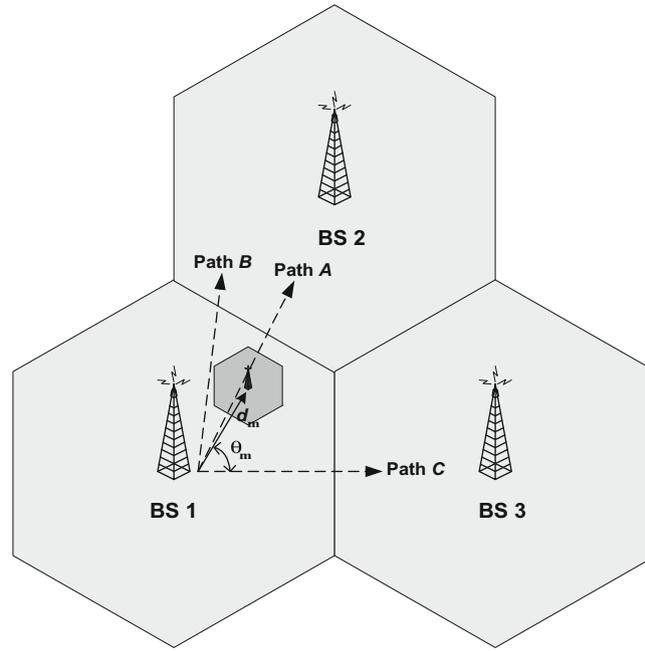


Fig. 13.18 Hierarchical CDMA system model

the overall system. Additional interference imbalance can be introduced by assigning different values of N to the macrocells as well. However, for exemplary purposes it is assumed that each macrocell area contains N uniformly distributed MSs.

Suppose that each MS connects to the BS that provides the least attenuation link. Given the location of a reference MS, r and θ in Cell 1, the probability that the MS is connected to BS _{i} is

$$\begin{aligned} P_{\text{H}1i}(r, \theta) &= P[G_i(r, \theta) \geq G_j(r, \theta), j = 1, 2, 3, \mu] \\ &= \int_{-\infty}^{\infty} \prod_{j \neq i} \Phi \left(\frac{\epsilon_i - \mu_j(r, \theta) - \rho[\epsilon_i - \mu_i(r, \theta)]}{\sqrt{(1 - \rho^2)\sigma_\Omega}} \right) \frac{1}{\sqrt{2\pi}\sigma_\Omega} \exp \left\{ -\frac{(\epsilon_i - \mu_i(r, \theta))^2}{2\sigma_\Omega^2} \right\} d\epsilon_i. \end{aligned} \quad (13.87)$$

Therefore, the probability of the reference MS in Cell 1 being connected to BS _{i} is

$$P_{\text{H}1i} = \int_0^{2\pi} \int_0^{R_M} \frac{r}{\pi R_M^2} P_{\text{H}1i}(r, \theta) dr d\theta. \quad (13.88)$$

where R_M is the macrocell radius. Similarly, $P_{\text{H}2}$, $P_{\text{H}3}$, and $P_{\text{H}micro}$ can be calculated for the MSs located in different cells.

It may be argued that (13.87) is not representative of CDMA systems that employ power control and soft handoff. For these systems, a MS will connect to the BS that minimizes the transmit power required to achieve a target carrier-to-interference ratio (CIR). Looking at this another way, if a MS were to transmit with a given power P_T , it would connect to the BS that provides the largest carrier-to-interference ratio (CIR). Hence, under the assumption of ideal soft handoff, (13.87) becomes:

$$\begin{aligned} P_{\text{H}1i}(r, \theta) &= P[\text{CIR}_i(r, \theta) \geq \text{CIR}_j(r, \theta), j = 1, 2, 3, \mu] \\ &= P \left[\frac{G_i(r, \theta)P_T}{I_i} \geq \frac{G_j(r, \theta)P_T}{I_j}, j = 1, 2, 3, \mu \right]. \end{aligned} \quad (13.89)$$

Later Sect. 13.7.2 shows that there is barely any difference between the two approaches in terms of handoff analysis. Moreover, handoff errors will be shown to sometimes occur where the MSs fail to connect to their ideal BSs. So an analysis based on ideal soft handoff is really an approximation as well. For these reasons, the analysis proceeds with the interference analysis based on (13.87). Later, this approach by extending the analysis to ideal soft handoff will be justified, using (13.89) instead of (13.87).

The total reverse link desired plus interfering signal power that is received by BS_{*i*} is equal to the sum of received power contributions from MSs located in different cells, i.e.,

$$S_i = S_i(1) + S_i(2) + S_i(3) + S_i(\mu), \quad (13.90)$$

where $S_i(j)$ is the signal contribution to BS_{*i*} from all MSs located in Cell j . With the introduction of a microcell, the level of interference that each BS experiences, I_i , will be uneven. Define γ_i as the interference power ratio between BS_{*i*} and the microcell:

$$\gamma_i \triangleq \frac{I_i}{I_\mu}. \quad (13.91)$$

Assuming a uniform CIR requirement and perfect power control, the signal power that is received from a MS connected to BS_{*i*} must satisfy

$$C_i = \gamma_i C_\mu, \quad (13.92)$$

where C_μ is the power controlled received power level of a MS connected to the microcell, which is used as a reference. Therefore,

$$I_i = S_i - C_i. \quad (13.93)$$

The power that is received at BS_{*i*} from MSs that are located in the Cell i cell, but are connected to BSs other than BS_{*i*} is now investigated. Let N_{ij} ($M_{\mu j}$) be the number of MSs in Cell i (Microcell) connected to BS_{*j*}:

$$\begin{aligned} N &= N_{i1} + N_{i2} + N_{i3} + N_{i\mu} \\ M &= M_{\mu 1} + M_{\mu 2} + M_{\mu 3} + M_{\mu \mu}. \end{aligned} \quad (13.94)$$

Define \mathbf{N}_i as a vector containing the N_{ij} :

$$\mathbf{N}_i = (N_{i1}, N_{i2}, N_{i3}, N_{i\mu}). \quad (13.95)$$

Let us consider $S_1(1)$ as an example. Given \mathbf{N}_1 ,

$$S_1(1) = N_{11}C_1 + \sum_{k=1}^{N_{12}} S_{12}(1, k) + \sum_{k=1}^{N_{13}} S_{13}(1, k) + \sum_{k=1}^{N_{1\mu}} S_{1\mu}(1, k), \quad (13.96)$$

where $S_{ij}(q, k)$ is the interference contribution to BS_{*i*} from the k th MS located in Cell q and connected to BS_{*j*}. Under the assumption of perfect power control,

$$S_{ij}(q, k) = \frac{G_i(r, \theta)}{G_j(r, \theta)} C_j, \quad 0 < S_{ij}(q, k) < C_j. \quad (13.97)$$

The cumulative distribution function of $S_{ij}(q, k)$ for any MS, $S_{ij}(q)$, is then,

$$F_{S_{ij}(q)}(z) = \frac{1}{P_{Hqj}} \int_0^{2\pi} \int_0^{R_M} \frac{r}{\pi R_M^2} P \left[\frac{G_i(r, \theta)}{G_j(r, \theta)} C_j < z \mid \text{MS connected to BS}_j \right] dr d\theta \quad (13.98)$$

Using the Gaussian distribution in (13.84) gives

$$F_{S_{ij}(q)}(z) = \frac{1}{P_{Hqj}} \int_0^{2\pi} \int_0^{R_M} \int_{-\infty}^{\infty} \frac{r}{\pi R_M^2} \frac{1}{\sqrt{2\pi} \sigma_\Omega} \exp \left\{ -\frac{(\varepsilon_j - \mu_j(r, \theta))^2}{2\sigma_\Omega^2} \right\}$$

$$\begin{aligned}
& \times \Phi \left(\frac{\varepsilon_j - 10 \log_{10}[C_j/z] - \mu_i(r, \theta) - \rho[\varepsilon_j - \mu_j(r, \theta)]}{\sqrt{(1 - \rho^2)\sigma_s}} \right) \\
& \times \prod_{l \neq i, j} \Phi \left(\frac{\varepsilon_j - \mu_l(r, \theta) - \rho[\varepsilon_j - \mu_j(r, \theta)]}{\sqrt{(1 - \rho^2)\sigma_s}} \right) dr d\theta d\varepsilon_j.
\end{aligned} \tag{13.99}$$

Since $S_{ij}(q)$ is a nonnegative random variable, its expected value and the second moment are given as follows:

$$\begin{aligned}
E[S_{ij}(q)] &= \int_0^\infty (1 - F_{S_{ij}(q)}(z)) dz \\
&= \int_0^{C_j} (1 - F_{S_{ij}(q)}(z)) dz \\
E[S_{ij}^2(q)] &= \int_0^{C_j} 2z(1 - F_{S_{ij}(q)}(z)) dz,
\end{aligned} \tag{13.100}$$

and

$$\text{Var}[S_{ij}(q)] = E[S_{ij}^2(q)] - E[S_{ij}(q)]^2. \tag{13.101}$$

Then, given \mathbf{N}_1 , the mean and second moment of $S_1(1)$ are

$$\begin{aligned}
E[S_1(1)|\mathbf{N}_1] &= N_{11}C_1 + N_{12}E[S_{12}(1)] + N_{13}E[S_{13}(1)] + N_{1\mu}E[S_{1\mu}(1)] \\
E[S_1^2(1)|\mathbf{N}_1] &= N_{12}\text{Var}[S_{12}(1)] + N_{13}\text{Var}[S_{13}(1)] + N_{1\mu}\text{Var}[S_{1\mu}(1)] \\
&\quad + N_{11}^2 C_1^2 + N_{12}^2 E[S_{12}(1)]^2 + N_{13}^2 E[S_{13}(1)]^2 + N_{1\mu}^2 E[S_{1\mu}(1)]^2 \\
&\quad + \sum_{i=1} N_{1i} E[S_{1i}(1)] \left(\sum_{j \neq i} N_{1j} E[S_{1j}(1)] \right).
\end{aligned} \tag{13.102}$$

The N_{1j} are binomial random variables with parameters P_{H1j} . Applying the chain rule of probability gives

$$\begin{aligned}
E[S_1(1)] &= \sum_{N_{11}=0}^N \binom{N}{N_{11}} P_{H11}^{N_{11}} (1 - P_{H11})^{N-N_{11}} \\
&\quad \times \sum_{N_{12}=0}^{N-N_{11}} \binom{N-N_{11}}{N_{12}} \bar{P}_{H12}^{N_{12}} (1 - \bar{P}_{H12})^{N-N_{11}-N_{12}} \\
&\quad \times \sum_{N_{13}=0}^{N-N_{11}-N_{12}} \binom{N-N_{11}-N_{12}}{N_{13}} \bar{P}_{H13}^{N_{13}} (1 - \bar{P}_{H13})^{N-N_{11}-N_{12}-N_{13}} E[S_1(1)|\mathbf{N}_1] \\
&= \sum_{N_{11}=0}^N \sum_{N_{12}=0}^{N-N_{11}} \sum_{N_{13}=0}^{N-N_{11}-N_{12}} \frac{N!}{N_{11}! N_{12}! N_{13}! N_{1\mu}!} P_{H11}^{N_{11}} (1 - P_{H11})^{N-N_{11}} \\
&\quad \times \bar{P}_{H12}^{N_{12}} (1 - \bar{P}_{H12})^{N-N_{11}-N_{12}} \bar{P}_{H13}^{N_{13}} (1 - \bar{P}_{H13})^{N_{1\mu}} E[S_1(1)|\mathbf{N}_1]
\end{aligned} \tag{13.103}$$

$$\begin{aligned}
E[S_1^2(1)] &= \sum_{N_{11}=0}^N \sum_{N_{12}=0}^{N-N_{11}} \sum_{N_{13}=0}^{N-N_{11}-N_{12}} \frac{N!}{N_{11}! N_{12}! N_{13}! N_{1\mu}!} P_{H11}^{N_{11}} (1 - P_{H11})^{N-N_{11}} \\
&\quad \times \bar{P}_{H12}^{N_{12}} (1 - \bar{P}_{H12})^{N-N_{11}-N_{12}} \bar{P}_{H13}^{N_{13}} (1 - \bar{P}_{H13})^{N_{1\mu}} E[S_1^2(1)|\mathbf{N}_1]
\end{aligned} \tag{13.104}$$

where

$$\bar{P}_{H12} = \frac{P_{H12}}{P_{H12} + P_{H13} + P_{H1\mu}} \quad (13.105)$$

$$\bar{P}_{H13} = \frac{P_{H13}}{P_{H13} + P_{H1\mu}}. \quad (13.106)$$

Finally, the variance of $S_1(1)$ is

$$\text{Var}[S_1(1)] = E[S_1^2(1)] - E[S_1(1)]^2. \quad (13.107)$$

Similarly, the means and variances of $S_1(2)$, $S_1(3)$, and $S_1(\mu)$ can be obtained. Finally,

$$\begin{aligned} E[S_1] &= E[S_1(1)] + E[S_1(2)] + E[S_1(3)] + E[S_1(\mu)] \\ \text{Var}[S_1] &= \text{Var}[S_1(1)] + \text{Var}[S_1(2)] + \text{Var}[S_1(3)] + \text{Var}[S_1(\mu)] \\ E[I_1] &= E[S_1] - C_1 \\ \text{Var}[I_1] &= \text{Var}[S_1]. \end{aligned} \quad (13.108)$$

Since I_1 represents a power sum, I_1 can be modeled as a log-normal random variable with the probability density function

$$p_{I_1}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right\}, \quad (13.109)$$

where

$$E[I_1] = e^{\mu + \sigma^2/2} \quad (13.110)$$

$$\text{Var}[I_1] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}. \quad (13.111)$$

In a likewise fashion, the means and variances of I_2 , I_3 , and I_μ can be computed. The interference analysis can be run using the following iterative steps:

1. Set $C_1 = C_2 = C_3 = C_\mu$.
2. Compute means and variances of I_1 , I_2 , I_3 , and I_μ .
3. Compute $E[\gamma_i] = E[I_i/I_\mu]$.
4. Set $C_i = E[\gamma_i]C_\mu$.
5. Goto Step 2.

Usually, 15 iteration loops are sufficient for the γ_i to converge. Then the reverse link CIR becomes

$$(C/I)_i = C_i/I_i = C_\mu/I_\mu. \quad (13.112)$$

13.7.1.2 Soft Handoff Analysis

In CDMA cellular systems such as cdma2000 and WCDMA, each BS transmits a pilot signal, consisting of an unmodulated spreading sequence, to assist soft handoff. The MSs use the pilot signals to initiate and complete handoffs among other things. An *active set* refers to a set of BSs to which a MS is connected to at any given time. The active set contains multiple BSs when the MS is in soft handoff mode.

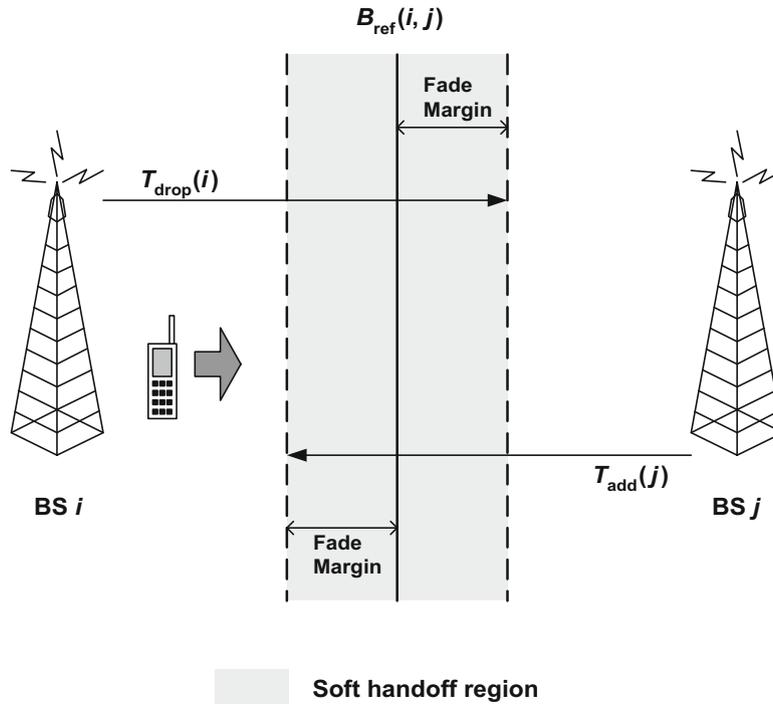


Fig. 13.19 Soft handoff parameters and corresponding handoff region

Suppose that the active set membership is based on the received pilot signal power.⁵ The upper threshold, T_{add} , is the pilot signal level where qualifying BSs are added to the active set, whereas the lower threshold, T_{drop} , determines when the BSs are removed from the active set. The difference between T_{add} and T_{drop} is an indicator of how long a soft handoff will take on average. This is illustrated graphically in Fig. 13.19. Considering a MS that is traveling from BS_{*i*} to BS_{*j*}, the soft handoff region is determined by T_{drop} imposed on BS_{*i*} and T_{add} imposed on BS_{*j*}. The values of T_{add} and T_{drop} are determined by defining the reference boundary, B_{ref} , and adding a fade margin to combat the effect of shadow fading [336].

This section introduces a hierarchical soft handoff analysis similar to the analysis presented in [10, 374], which tracks a moving MS to observe its active set membership while incorporating the spatial correlation property of shadow fading. As mentioned previously, the introduction of microcell(s) into a macrocell layer results in interference imbalance which can impact the soft handoff decisions and performance. A handoff analysis based on received pilot signal strength and a single MS will not accurately depict the actual system behavior. On the other hand, a comprehensive analysis that includes multiple MSs while incorporating interference effects is prohibitively complicated and computationally exhaustive. Here, a soft handoff analysis model is presented that allows the study of soft handoff performance in conjunction with interference performance, by integrating the results obtained in interference analysis in Sect. 13.7.1.1. The approach accurately depicts the handoff performance of hierarchical systems, while being computationally efficient. Some detailed derivations are omitted in the following section, and refer the reader to [10, 374].

A MS traveling along a specific path is now considered to study its active set membership. Let $A(n)$ be the active set membership at epoch n for the MS under consideration. Let $P_i(n)$ be the probability that BS_{*i*} is in the active set at epoch n :

$$P_i(n) = P[\text{BS}_i \in A(n)]. \quad (13.113)$$

When $A(n)$ contains more than two BSs, the MS connects to the BS in the set that will minimize its transmit power, thereby limiting interference. This means that the BS selection within the set depends not only on the forward link received pilot strengths, but also the reverse link interference conditions. Let $B(n)$ be the BS in the active set that minimizes the MS transmit power. Since $A(n)$ is constantly being updated, the selection of $B(n)$ is based on the active set membership at epoch $n - 1$:

⁵CDMA cellular systems actually use the forward link E_c/I_o , the ratio of the received pilot chip energy to total interference spectral density, to determine active set memberships. For the present purpose, the received pilot signal power is used instead and later in Sect. 13.7.2 the difference between these two methods for determining active set membership in terms of their soft handoff performance will be illustrated.

$$\begin{aligned}
B(n) &= \max \left\{ \frac{G_i(n)}{I_i(n)} \mid \text{BS}_i \in A(n-1) \right\} \\
&= \max \left\{ \frac{G_i(n)}{\gamma_i(n)} \mid \text{BS}_i \in A(n-1) \right\}.
\end{aligned} \tag{13.114}$$

As mentioned before CDMA systems measure the forward link E_c/I_o to determine the active set memberships. However, for now just use the received pilot signal strength is used. It is also assumed that the BSs transmit their pilot signals with equal power. A BS is added to a MS's active set when its path gain exceeds its add threshold, $T_{\text{add}}(i)$. Therefore, the probability that it will be added to the active set at epoch n is:

$$P_{N \rightarrow i}(n) = P[G_i(n) > T_{\text{add}}(i) \mid \text{BS}_i \notin A(n-1)]. \tag{13.115}$$

A BS is dropped from the active set by using both absolute and relative thresholds. First, the associated path gain must fall below the absolute drop threshold, $T_{\text{drop}}(i)$. When it does, its gain is compared to the largest path gain in the active set, $B(n)$. When the difference between the two exceeds the relative drop threshold, $T_{\text{rel}}(i)$, the BS is dropped from the active set. The relative threshold causes a BS to be dropped from the active set only when its link has deteriorated far below the best link. This also ensures that the active set contains at least one candidate BS at all times. The probability that BS_i is dropped from the active set at epoch n is:

$$P_{i \rightarrow N}(n) = P[B(n) - G_i(n) > T_{\text{rel}}(i), G_i(n) < T_{\text{drop}}(i) \mid \text{BS}_i \in A(n-1)]. \tag{13.116}$$

Finally,

$$P_i(n) = P_i(n-1)[1 - P_{i \rightarrow N}(n)] + [1 - P_i(n-1)]P_{N \rightarrow i}(n). \tag{13.117}$$

The main purpose of soft handoff is to ensure that the MS is connected to the BS which minimizes its transmit power. Therefore, a handoff error occurs when $B(n)$ is not the best available choice:

$$\begin{aligned}
H_{\text{error}}(n) &= P \left[B(n) \neq \max \left\{ \frac{G_1(n)}{I_1(n)}, \frac{G_2(n)}{I_2(n)}, \frac{G_3(n)}{I_3(n)}, \frac{G_\mu(n)}{I_\mu(n)} \right\} \right] \\
&= P \left[B(n) \neq \max \left\{ \frac{G_1(n)}{\gamma_1(n)}, \frac{G_2(n)}{\gamma_2(n)}, \frac{G_3(n)}{\gamma_3(n)}, G_\mu(n) \right\} \right].
\end{aligned} \tag{13.118}$$

Another measure of soft handoff efficiency is the average number of BSs in active set at epoch n , $\bar{A}(n)$:

$$\bar{A}(n) = P_1(n) + P_2(n) + P_3(n) + P_\mu(n). \tag{13.119}$$

A smaller value of $\bar{A}(n)$ implies a lower infrastructure overhead to support soft handoff.

13.7.2 Numerical Results

A path loss exponent $\beta = 4$ and shadow standard deviation $\sigma_\Omega = 8$ dB are used in the simulation. Log-normal shadows are generated using the approach suggested in Sect. 2.6.1, where the parameter ζ_D is the spatial shadow correlation between two points separated by distance D m. The radii of the macrocell and microcell regions are set to 1500 and 100 m, respectively. Other important simulation parameters include:

- $\zeta_{100} = 0.82$
- $\rho = 0$ (uncorrelated shadowing)
- MS velocity = 60 km/h
- sampling period = 1 s
- $T_{\text{rel}} = 3$ dB.

Table 13.2 Comparison of analytical and numerical results. Microcell load (M) is fixed at 12 while macrocell load (N) is varied

N	M	Simulation				Analysis			
		E[CIR] (dB)	$E[\gamma_1]$	$E[\gamma_2]$	$E[\gamma_3]$	E[CIR] (dB)	$E[\gamma_1]$	$E[\gamma_2]$	$E[\gamma_3]$
12	12	-12.70	0.135	0.096	0.093	-12.70	0.136	0.096	0.090
13	12	-12.86	0.148	0.110	0.107	-12.86	0.148	0.108	0.102
14	12	-13.03	0.162	0.128	0.125	-13.01	0.160	0.120	0.115
15	12	-13.19	0.176	0.145	0.142	-13.16	0.171	0.134	0.129
16	12	-13.34	0.188	0.163	0.162	-13.30	0.183	0.149	0.143
17	12	-13.50	0.203	0.181	0.182	-13.44	0.195	0.164	0.159
18	12	-13.64	0.219	0.204	0.208	-13.58	0.207	0.180	0.177
19	12	-13.79	0.232	0.225	0.229	-13.72	0.219	0.198	0.195

Table 13.3 Comparison of analytical and numerical results. Macrocell load (N) is fixed at 12 while microcell load (M) is varied

N	M	Simulation				Analysis			
		E[CIR] (dB)	$E[\gamma_1]$	$E[\gamma_2]$	$E[\gamma_3]$	E[CIR] (dB)	$E[\gamma_1]$	$E[\gamma_2]$	$E[\gamma_3]$
12	10	-12.25 dB	0.164	0.141	0.138	-12.23 dB	0.164	0.126	0.120
12	11	-12.48 dB	0.146	0.113	0.111	-12.47 dB	0.146	0.109	0.103
12	12	-12.70 dB	0.135	0.096	0.093	-12.70 dB	0.136	0.096	0.090
12	13	-12.92 dB	0.122	0.084	0.080	-12.92 dB	0.126	0.085	0.079
12	14	-13.13 dB	0.117	0.074	0.070	-13.13 dB	0.117	0.077	0.071
12	15	-13.34 dB	0.105	0.067	0.063	-13.33 dB	0.109	0.070	0.065
12	16	-13.53 dB	0.100	0.062	0.058	-13.52 dB	0.102	0.064	0.059
12	17	-13.72 dB	0.093	0.057	0.053	-13.71 dB	0.096	0.059	0.054

13.7.2.1 Interference Results

Tables 13.2 and 13.3 show the average CIR and interference performance comparisons between the analytical and simulation results. The microcell is placed at $d_\mu = 600$ m and $\theta_\mu = \pi/3$. Table 13.2 contains the results for varying macrocell load, N , while Table 13.3 shows the results when the microcell load, M , is varied. It is observed that the analytical and simulation results are in very close agreement, for both $E[\text{CIR}]$ and $E[\gamma_i]$. It is also seen that the accuracy of the analytical results improves as the interference discrepancy between the layers increases (smaller γ_i). As expected, increasing the system load (N and M) results in a decrease in system CIR performance, since it causes the overall interference to increase. Since the density of MSs in the microcell is higher than the density of MSs in the macrocell by nature, the microcell experiences a higher level of interference than the macrocells. The γ_i indicate the degree of interference imbalance between the hierarchical layers, and the obtained results agree with our basic intuition; a larger microcell load increases the interference imbalance (smaller γ_i) while a smaller macrocell load decreases the interference imbalance (larger γ_i).

Figures 13.20 and 13.21 show the effect of microcell location on the average CIR and interference performance. The results are obtained by varying d_μ while θ_μ is fixed at $\pi/3$. Again, the analytical results are in close agreement with the simulation results. Figure 13.20 shows that the average CIR performance varies insignificantly with changes in microcell location, although it seems to benefit somewhat from diversity gain when the microcell is located very close to a macrocell BS. Figure 13.21 shows how the γ_i are affected by different microcell locations. It is observed that the corresponding γ_i increases as the microcell moves closer to a macrocell BS. This is expected since the level of inter-layer interference between the microcell and macrocell increases as the microcell gets closer to a macrocell BS, which in turn causes the macrocell interference to increase. Observe from Fig. 13.21 that as d_μ increases γ_1 decreases while γ_2 and γ_3 increase.

13.7.2.2 Soft Handoff Results

The previous section has shown how various system loads and microcell locations can affect the interference levels in hierarchical CDMA cellular systems. The interference imbalance factors (γ_i) are important parameters for determining the soft handoff performance since, along with the received pilot signal strengths, they can be used to portray the system behavior during soft handoff and provide information on how to improve the handoff performance. A fixed parameter handoff

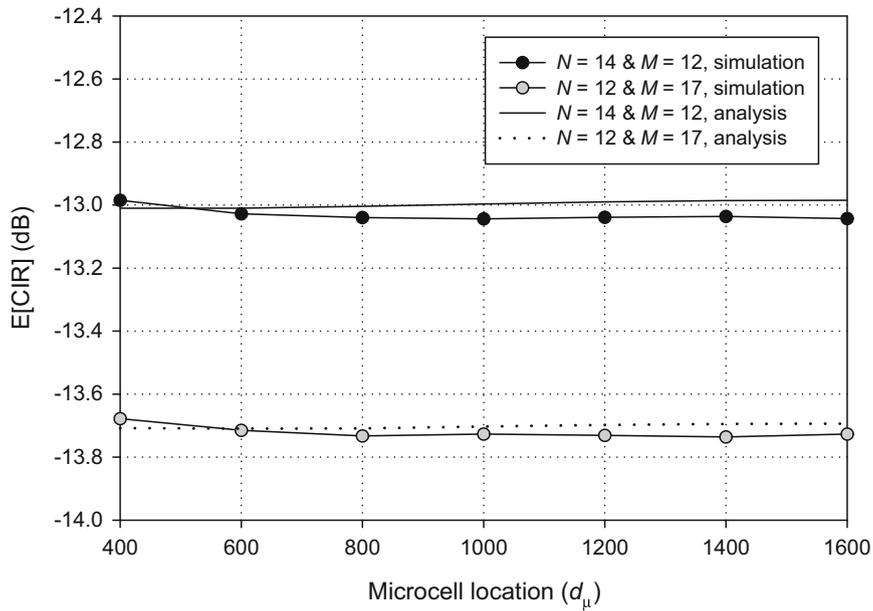


Fig. 13.20 Average CIR performance as a function of microcell location

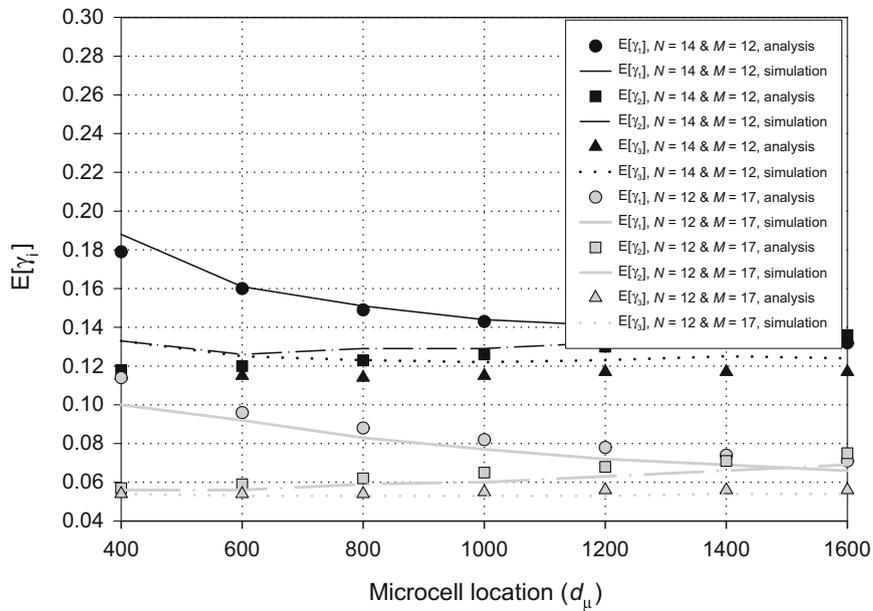


Fig. 13.21 $E[\gamma_i]$ as a function of microcell location

algorithm is first considered, where the values of T_{add} and T_{drop} are fixed regardless of the changing interference conditions, and determined by defining B_{ref} at an equal distance location and assigning a fade margin of 8 dB,⁶ viz.:

$$\begin{aligned}
 T_{\text{add}}(i) &= B_{\text{ref}}(i, j) + 8 \text{ dB} \\
 T_{\text{drop}}(j) &= B_{\text{ref}}(i, j) - 8 \text{ dB}.
 \end{aligned}
 \tag{13.120}$$

⁶Other fade margins can be chosen.

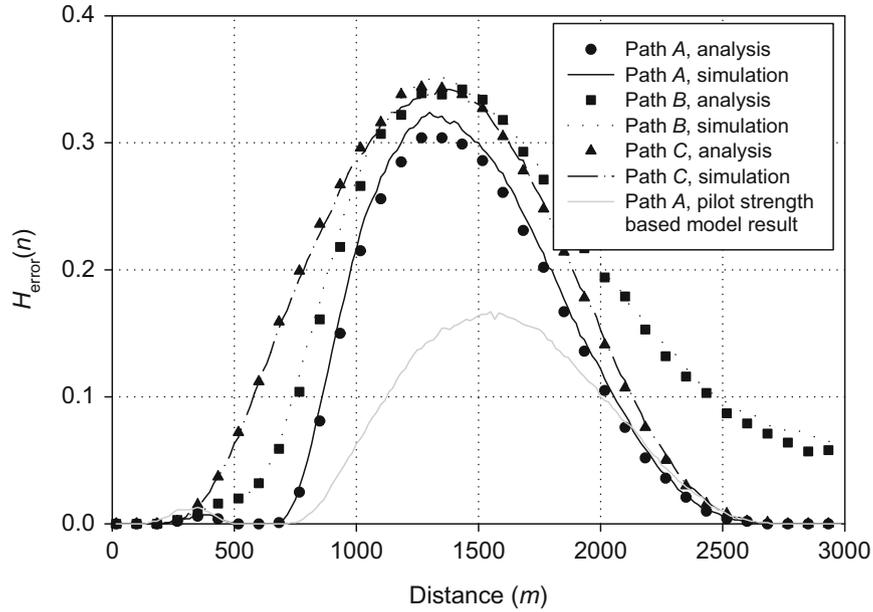


Fig. 13.22 Error performance of fixed handoff parameter assignment; $N = 13$ and $M = 12$

Figure 13.22 shows the handoff error probability for fixed handoff parameter assignment. The microcell is located at $d_\mu = 600$ m and $\theta_\mu = \pi/3$. The analytical results are obtained using $E[\gamma_i]_{\text{analytical}}$ while the simulation results are obtained using actual γ_i . The figure shows the handoff error probability for three traveling paths, all starting from BS_1 as shown in Fig. 13.18. It is observed that the analytical and simulation results are in good agreement. Figure 13.22 also shows the handoff error probability for pilot-strength based handoff model [10, 374] and shows how it grossly underestimates the actual handoff error probability when interference levels are not uniform. By incorporating the interference analysis results, a far more accurate handoff performance analysis is obtained. The handoff error probability is observed to increase around the vicinity of physical cell boundaries. It is also observed that the handoff error probability is significantly higher between 1000 and 2000 m. This phenomenon is largely due to the value of σ_Ω and the selection of T_{drop} for BS_1 ($T_{\text{drop}}(1)$). $T_{\text{drop}}(1)$ has been set so that BS_1 is dropped from the active set once the MS enters the microcell. However, with $\sigma_\Omega = 8$ dB and with the effect of γ_1 , BS_1 provides the best connection in the 1000–2000 m region a significant number of times, and that is why high values of $H_{\text{error}}(n)$ are observed. The handoff error probability can be improved by relaxing $T_{\text{drop}}(1)$ to cover the region, but this will definitely increase $\bar{A}(n)$, thereby leading to additional system resource requirements. However, the handoff error probability depends on the microcell location as shown in Fig. 13.23. The figure contains the handoff error probability plots for path A at three different microcell locations. It is seen that the handoff error probability decreases if d_μ is increased without changing $T_{\text{drop}}(1)$.

Next, the performance of dynamic handoff parameter assignment is examined. In dynamic parameter assignment B_{ref} is not fixed, but is dynamically updated as a function of the γ_i to improve the handoff performance. The concept is similar to the phenomenon of “cell breathing” [54, 319] where a heavily loaded cell shrinks in size and forces handoffs that will reduce interference. Our objective is to control the microcell handoff region according to given interference imbalance condition (as defined by the γ_i) to limit unnecessary overhead. This is accomplished by defining $B_{\text{ref}}(i, \mu)$ at the equilibrium point, d_e , where

$$\frac{d_e^{-\beta}(\mu)}{\gamma_i} = d_e^{-\beta}(i), \quad (13.121)$$

where $d_e(i)$ is the distance between d_e and BS_i . It is easily observed that $B_{\text{ref}}(i, \mu)$ moves toward the microcell BS as γ_i decreases which reduces the microcell soft handoff region accordingly. Figure 13.24 compares the performance between fixed and dynamic parameter assignment for path A with the microcell location at $d_\mu = 600$ m and $\theta_\mu = \pi/3$. While dynamic handoff parameter assignment does not offer any significant gain in handoff error probability, it provides a more efficient handoff mechanism over fixed handoff parameter assignment by reducing $\bar{A}(n)$. Fixed handoff parameter assignment requires a larger system overhead since it does not incorporate the system interference information into its handoff decisions.

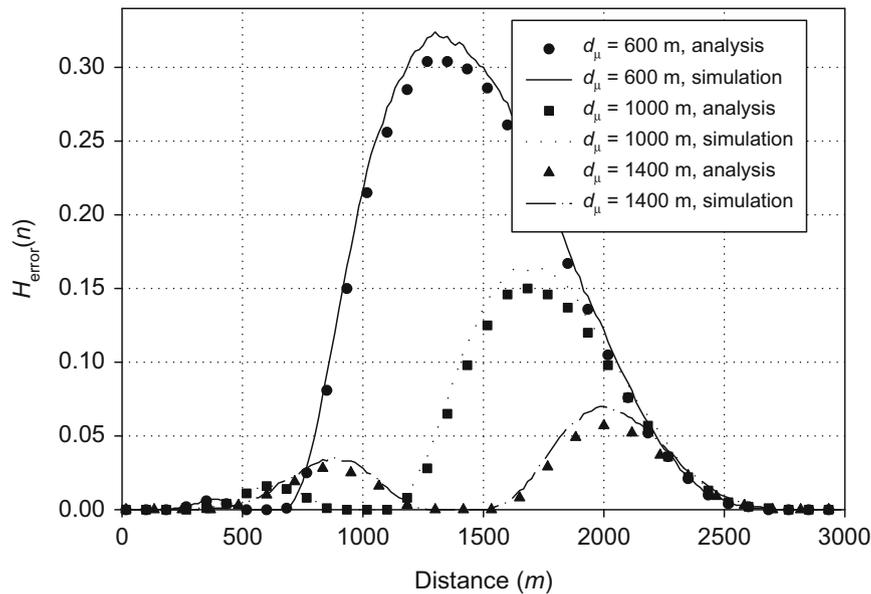


Fig. 13.23 Effect of microcell location on soft handoff performance; $N = 13$ and $M = 12$

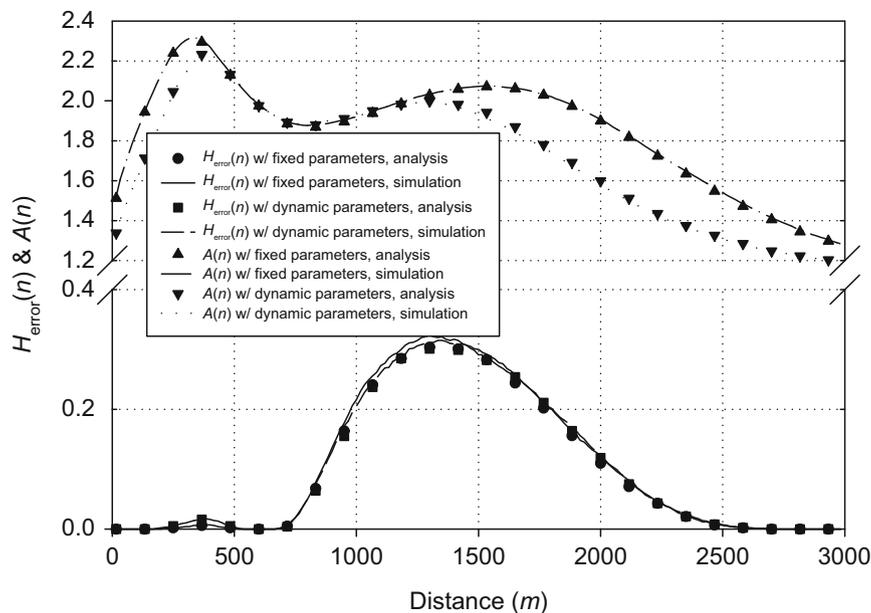


Fig. 13.24 Performance comparison between fixed and dynamic handoff parameter assignments; $N = 13$ and $M = 12$

Table 13.4 Comparison of fixed and dynamic soft handoff parameter assignment performances; $N = 13$ and $M = 12$

Path	Fixed		Dynamic	
	$E[H_{\text{error}}(n)]$	$E[\bar{A}(n)]$	$E[H_{\text{error}}(n)]$	$E[\bar{A}(n)]$
A	0.1043	1.8618	0.1033	1.7135
B	0.1598	1.7573	0.1584	1.6013
C	0.1406	1.8366	0.1314	1.6510

Dynamic handoff parameter assignment dynamically adjusts the microcell handoff region so that the system can prevent MSs from being prematurely subjected to soft handoff. Table 13.4 shows the average error probability and active set membership for three specified MS paths. For all three paths, dynamic handoff parameter assignment provides superior performance in $E[\bar{A}(n)]$ while slightly improving $E[H_{\text{error}}(n)]$.

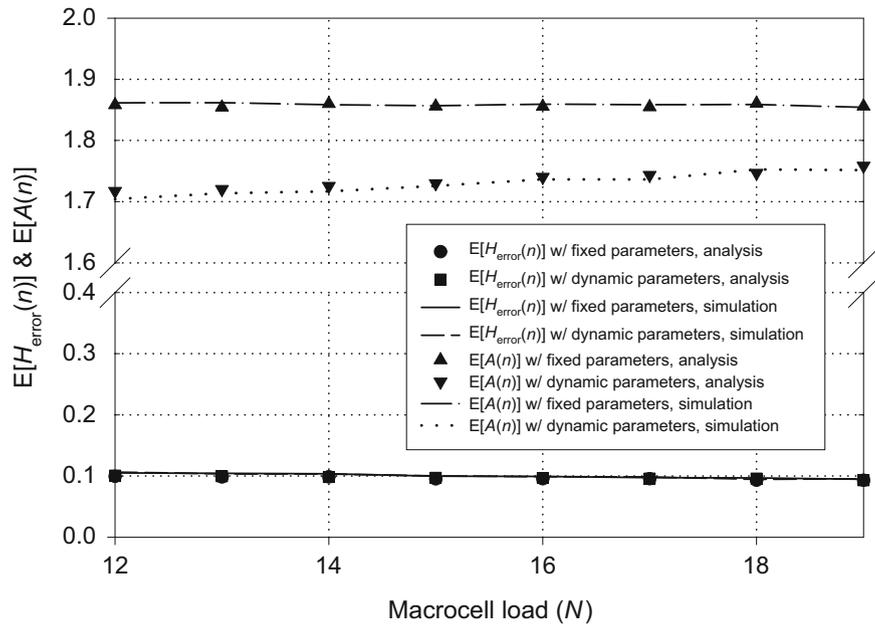


Fig. 13.25 Effect of interference imbalance on soft handoff performance. Microcell load (M) is fixed at 12 while macrocell load (N) is varied

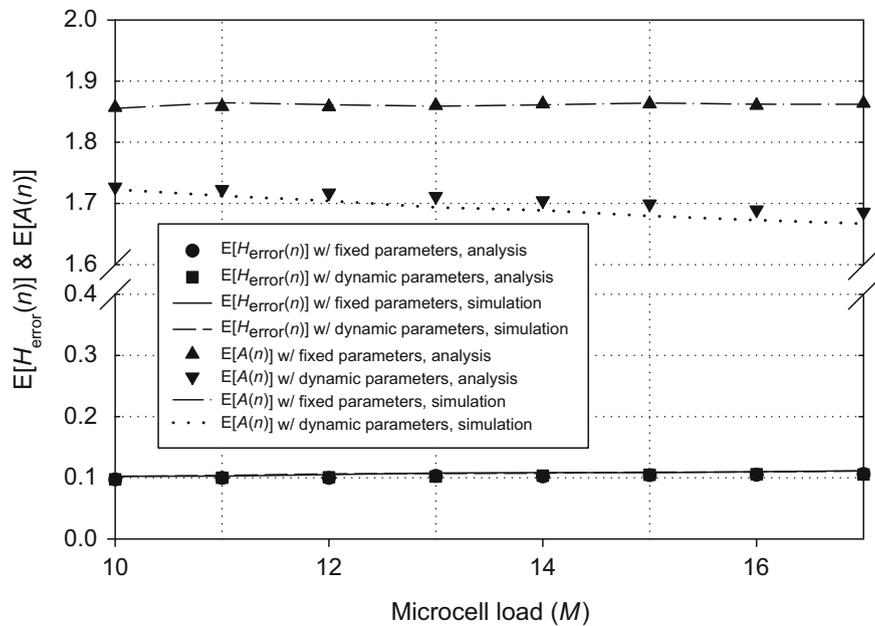


Fig. 13.26 Effect of interference imbalance on soft handoff performance. Macrocell load (N) is fixed at 12 while microcell load (M) is varied

Figures 13.25 and 13.26 compare the performance of fixed and dynamic handoff parameter assignment as the system load is varied. As was observed in Figs. 13.20 and 13.21, increasing the macrocell load, N , increases the γ_i while increasing the microcell load, M , reduces the γ_i . It is seen that $E[A(n)]$ stays nearly uniform with various system loads for fixed handoff parameter assignment while it changes according to the γ_i for dynamic handoff parameter assignment. As expected, a larger interference imbalance (lower γ_i) causes the microcell handoff region to shrink and thereby reducing $E[A(n)]$ for dynamic handoff parameter assignment. The average handoff error probabilities for both fixed and dynamic handoff parameter assignments do not change significantly with varying system load.

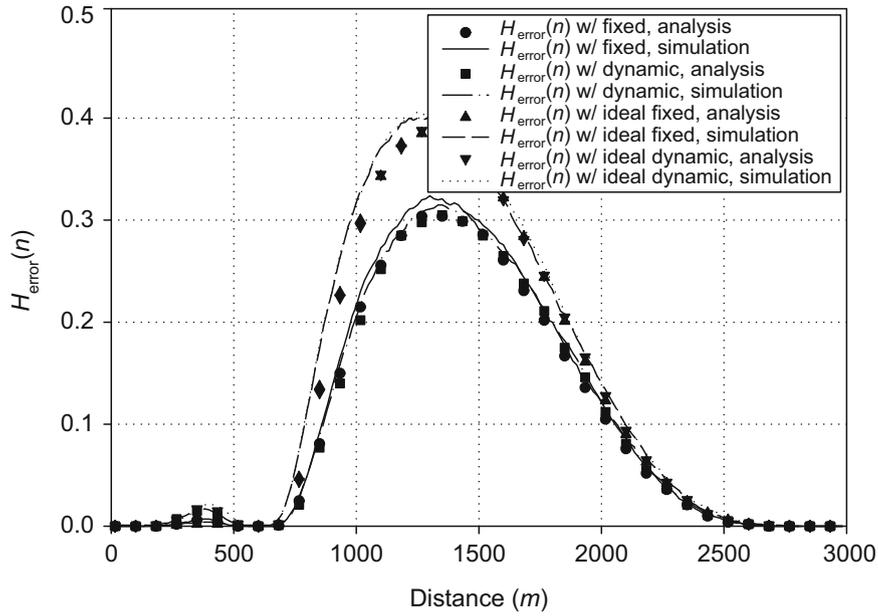


Fig. 13.27 Error performance comparison between our soft handoff assumption and ideal soft handoff; $N = 13$ and $M = 12$

13.7.2.3 Ideal Handoff and E_c/I_o -Based Active Set Membership

The soft analysis to this point has used some simplifying assumptions regarding soft handoff and its active set membership. The validity of these assumptions is now examined. In Sect. 13.7.1.1, it was assumed that a MS connects to the BS that provides the most robust path gain according to (13.87). However, during ideal soft handoff, a MS connects to the BS that minimizes its transmit power according to (13.89). Figures 13.27 and 13.28 compare the handoff error and the active set membership performance between the soft handoff analysis based on (13.87) and ideal soft handoff analysis based on (13.89). There is no significant performance difference between the two approaches. Also, dynamic handoff parameter assignment yields a more efficient handoff mechanism than fixed handoff parameter assignment in either case.

The analysis in Sect. 13.7.1.2 used the forward link received pilot signal power to determine active set membership, while practical CDMA systems use forward link E_c/I_o measurements instead. The difference between these two approaches is now examined. Let $P_T(i)$ be the total forward transmit power from BS i , including its pilot power. Then, for a MS located at (r, θ) ,

$$\frac{E_c(i)}{I_o(i)} \approx \frac{G_i(r, \theta) P_{\text{pilot}}}{\sum_{j=1}^{2,3,\mu} G_j(r, \theta) P_T(j)} G_{\text{spread}}, \quad (13.122)$$

where G_{spread} is the processing gain. There are two main difficulties when incorporating E_c/I_o into the soft handoff analysis. First, it is difficult to model the E_c/I_o behavior mathematically. A power controlled forward link is harder to model than its reverse link counterpart, especially with open loop power control. Second, the total forward transmit power from each BS, $P_T(i)$, depends on the number of MSs served by that BS including the MSs in soft handoff.

Once again the deployment in Fig. 13.18 introduces interference imbalance on the forward link due to the presence of the microcell. This interference imbalance will impact the received E_c/I_o from each BS. Figures 13.29 and 13.30 compare the pilot signal power and E_c/I_o methods for determining active set membership, in terms of the handoff error probability and average number of BSs in active set. These E_c/I_o based results are obtained by assuming that $P_T(i)$ is same for all BSs in the system, although this may not be true. There are some significant differences in performance between received pilot power and E_c/I_o based active set memberships. In particular, the E_c/I_o method requires much less overhead for a comparable handoff error performance. The observation may be attributed to the fact that E_c/I_o follows a slope up to $d^{-2\beta}$ and it has angular dependency. In either case, however, dynamic handoff parameter assignment yields a more efficient handoff mechanism than fixed handoff parameter assignment.

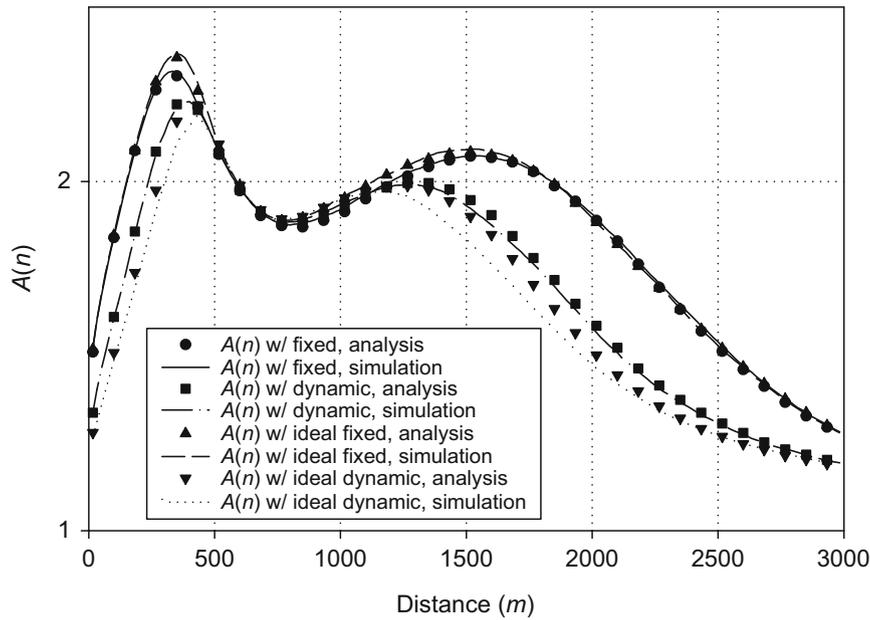


Fig. 13.28 Handoff overhead comparison between our soft handoff assumption and ideal soft handoff; $N = 13$ and $M = 12$

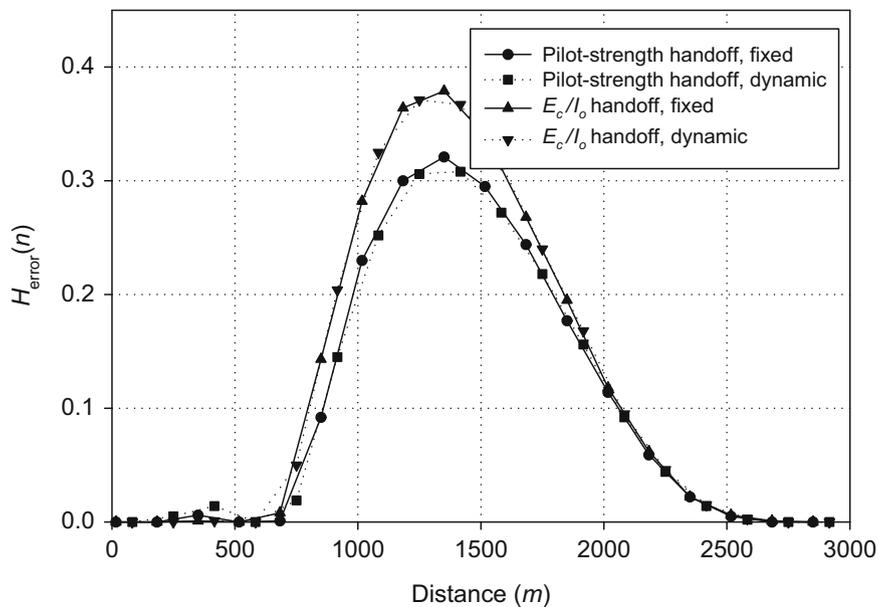


Fig. 13.29 Error performance comparison between pilot-strength and E_c/I_o based active set membership; $N = 13$ and $M = 12$

13.8 CINR Based Link Quality Measurements

Cellular radio resource allocation algorithms have been developed for handoffs [126], dynamic channel assignment [145, 241], and power control [20, 21], under the assumption that the MSs and/or BSs have access to real-time measurements of the received carrier-to-interference plus noise ratio $C/(I + N)$ or CINR. This section presents a technique for estimating $C + I + N$ and $C/(I + N)$ that could be used in resource management algorithms [29, 30].

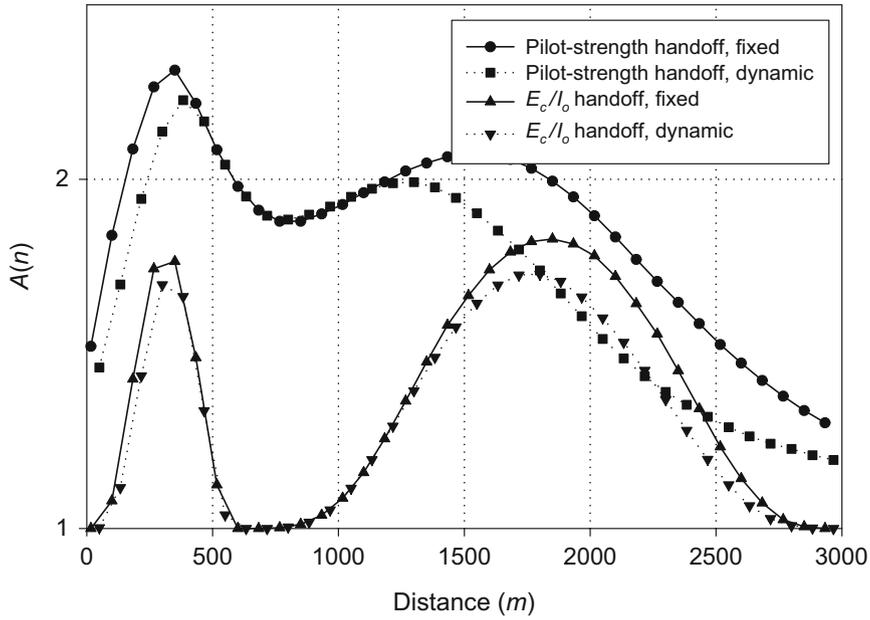


Fig. 13.30 Handoff overhead comparison between pilot-strength and E_c/I_o based active set membership; $N = 13$ and $M = 12$

13.8.1 Discrete-Time Model for Signal Quality Estimation

As shown in Sect. 7.2.1, the overall channel consisting of the transmit filter, waveform channel, matched filter, sampler, and noise whitening filter can be modeled as a T -spaced, $L + 1$ -tap, transversal filter.⁷ The overall discrete-time channel is described by the channel vector $\mathbf{g} = (g_0, g_1, \dots, g_L)^T$, where $(\)^T$ denotes transpose. Let $\mathbf{v} = (v_0, \dots, v_M)^T$ denote the received signal vector consisting of M samples, where $v_k = \sum_{i=0}^L g_i x_{k-i} + w_k$. 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

$$\mathbf{v} = \mathbf{X}\mathbf{g} + \mathbf{w}, \quad (13.123)$$

where \mathbf{X} is an $(M + 1) \times (L + 1)$ Toeplitz matrix consisting of the transmitted symbols of the form

$$\mathbf{X} = [x_{i,j}] = \begin{bmatrix} x_0 & x_{-1} & \cdots & x_{-L} \\ x_1 & x_0 & \cdots & x_{1-L} \\ \vdots & \vdots & \cdots & \vdots \\ x_M & x_{M-1} & \cdots & x_{M-L} \end{bmatrix} \quad (13.124)$$

and $\mathbf{w} = (w_0, \dots, w_M)^T$ is an impairment vector consisting of the samples of the received interference plus AWGN.

13.8.1.1 Estimation of $I + N$

An $I + N$ or $C/(I + N)$ estimator requires a method for separating \mathbf{g} and \mathbf{w} from the observation of \mathbf{v} . Consider the situation where $M > L$, so that \mathbf{X} has more rows than columns. Then, there exists a vector $\mathbf{c} = (c_0, \dots, c_M)^T$ 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

$$\mathbf{c}^T \mathbf{v} = \mathbf{0} + \mathbf{c}^T \mathbf{w}, \quad (13.125)$$

⁷If rate $2/T$ sampling is used, then the overall channel is a $T/2$ -spaced, $2L + 1$ -tap, transversal filter.

and, therefore, \mathbf{g} and \mathbf{w} are completely separated from the observation \mathbf{v} . However, with the exception of training and pilot sequences, \mathbf{X} is not known exactly because the data symbols comprising \mathbf{X} must be obtained from symbol decisions. In such cases, a matrix of symbol decisions $\hat{\mathbf{X}}$ must be used instead, where $\hat{\mathbf{X}} = \mathbf{X} + \mathbf{\Delta}$ and $\mathbf{\Delta} = [\delta_{ij}]$ is the symbol error matrix. Nevertheless, a vector $\hat{\mathbf{c}}$ can still be found in the null space of $\hat{\mathbf{X}}$ so that

$$\hat{\mathbf{c}}^T \mathbf{v} = \hat{\mathbf{c}}^T \mathbf{X} \mathbf{g} + \hat{\mathbf{c}}^T \mathbf{w}. \quad (13.126)$$

Hence, an (I+N) estimate can be obtained from

$$\begin{aligned} \hat{\sigma}_{I+N}^2 &= \frac{1}{2} \mathbb{E} \left[\frac{\mathbf{v}^H \hat{\mathbf{c}}^* \hat{\mathbf{c}}^T \mathbf{v}}{\|\hat{\mathbf{c}}\|^2} \right] \\ &= \frac{1}{2} \mathbb{E} \left[\frac{\mathbf{g}^H \mathbf{\Delta}^H \hat{\mathbf{c}}^* \hat{\mathbf{c}}^T \mathbf{\Delta} \mathbf{g}}{\|\hat{\mathbf{c}}\|^2} \right] + \frac{1}{2} \mathbb{E} \left[\frac{\mathbf{w}^H \hat{\mathbf{c}}^* \hat{\mathbf{c}}^T \mathbf{w}}{\|\hat{\mathbf{c}}\|^2} \right] \\ &= \frac{1}{2} \left(\sum_{i=0}^L \sum_{j=0}^M \sum_{k=0}^L \sum_{\ell=0}^M \mathbb{E} \left[\frac{\delta_{j,i}^* g_i^* \hat{c}_j^* \delta_{\ell,k} g_k \hat{c}_\ell}{\|\hat{\mathbf{c}}\|^2} \right] + \sum_{i=0}^M \sum_{j=0}^M \mathbb{E} \left[\frac{\hat{c}_i \hat{c}_j^*}{\|\hat{\mathbf{c}}\|^2} \right] \mathbb{E}[w_i w_j^*] \right), \end{aligned} \quad (13.127)$$

where $()^H$ denotes complex conjugate transpose, and where the second equality is obtained by using $\hat{\mathbf{X}} = \mathbf{X} + \mathbf{\Delta}$ along with the reasonable assumption that the impairment vector \mathbf{w} has zero mean and is uncorrelated with \mathbf{X} and \mathbf{g} . It is also reasonable to assume that the symbol errors are independent with a constant variance, i.e.,

$$\frac{1}{2} \mathbb{E}[|\delta_{j,i}|^2] = \sigma_{\Delta}^2. \quad (13.128)$$

It follows that

$$\hat{\sigma}_{I+N}^2 = \Omega_p \sigma_{\Delta}^2 + \frac{1}{2} \sum_{i=0}^M \sum_{j=0}^M \mathbb{E} \left[\frac{\hat{c}_i \hat{c}_j^*}{\|\hat{\mathbf{c}}\|^2} \right] \mathbb{E}[w_i w_j^*], \quad (13.129)$$

where $\Omega_p = \sum_{i=0}^L \Omega_i$ is the fading envelope power and $\Omega_i = \mathbb{E}[|g_i|^2]$ is the mean square value of the i th channel tap. To determine $\mathbb{E}[w_i w_j^*]$, rewrite the impairment vector \mathbf{w} as

$$\mathbf{w} = \sum_{k=1}^{N_I} \mathbf{B}_k \mathbf{g}_k + \boldsymbol{\eta}, \quad (13.130)$$

where \mathbf{B}_k is an $(M+1) \times (L+1)$ matrix consisting of the symbols from the k th interferer with associated channel tap vector \mathbf{g}_k , N_I is the number of interferers, and $\boldsymbol{\eta}$ is the vector of additive white Gaussian noise samples. The elements of \mathbf{w} are

$$w_i = \sum_{k=1}^{N_I} \sum_{\ell=0}^L b_{k,i,\ell} g_{k,\ell} + \eta_i, \quad i = 0, \dots, M, \quad (13.131)$$

where $\mathbf{B}_k = [b_{k,i,\ell}]$ and $\mathbf{g}_k = (g_{k,0}, \dots, g_{k,L})$. It is now assumed 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 η_i are independent zero-mean Gaussian random variables with variance σ_{η}^2 . Then $\frac{1}{2} \mathbb{E}[w_i w_j^*] = 0$ for $i \neq j$ and

$$\begin{aligned} \sigma_w^2 &= \frac{1}{2} \mathbb{E}[|w_i|^2] = \frac{1}{2} \mathbb{E} \left[\sum_{k=1}^{N_I} \sum_{\ell=0}^L |b_{k,i,\ell}|^2 |g_{k,\ell}|^2 + |\eta_i|^2 \right] \\ &= \sum_{k=1}^{N_I} \sigma_b^2 \sum_{\ell=0}^L \Omega_{k,\ell} + \sigma_{\eta}^2 \end{aligned}$$

$$\begin{aligned}
&= \sigma_I^2 + \sigma_\eta^2 \\
&= \sigma_{I+N}^2,
\end{aligned} \tag{13.132}$$

where $\sigma_b^2 = \frac{1}{2}\mathbb{E}[|b_{k,i,\ell}|^2]$ is the interferer symbol variance, $\Omega_{k,\ell} = \mathbb{E}[|g_{k,\ell}|^2]$ denotes the mean square value of the ℓ th channel tap gain associated with the k th interferer, and σ_I^2 denotes the total interference power. Using this result, (13.129) becomes

$$\hat{\sigma}_{I+N}^2 = \Omega_p \sigma_\Delta^2 + \sigma_w^2 = \Omega_p \sigma_\Delta^2 + \sigma_{I+N}^2. \tag{13.133}$$

In practice, the ensemble averaging in the first line of (13.127) is replaced by an empirical average over P observation vectors \mathbf{v}_i to provide the unbiased estimate

$$\hat{\sigma}_{I+N}^2 = \frac{1}{2P} \sum_{i=1}^P \frac{\mathbf{v}_i^H \hat{\mathbf{c}}_i^* \hat{\mathbf{c}}_i^T \mathbf{v}_i}{\|\hat{\mathbf{c}}_i\|^2}. \tag{13.134}$$

13.8.1.2 Estimation of $C/(I+N)$

A $C/(I+N)$ estimator can be formed by using $\hat{\sigma}_{I+N}^2$ as follows. The total received signal power from the desired signal, interfering signals, and noise is

$$\begin{aligned}
\sigma_{C+I+N}^2 &= \frac{1}{M+1} \frac{1}{2} \mathbb{E}[\mathbf{v}^H \mathbf{v}] \\
&= \frac{1}{M+1} \frac{1}{2} \mathbb{E}[\mathbf{g}^H \mathbf{X}^H \mathbf{X} \mathbf{g} + \mathbf{w}^H \mathbf{w}] \\
&= \frac{1}{M+1} \left(\sum_{j=0}^L \Omega_j \sum_{i=0}^M \frac{1}{2} \mathbb{E}[|x_{ij}|^2] + (M+1) \sigma_w^2 \right),
\end{aligned} \tag{13.135}$$

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{X} matrix or the channel tap gains be uncorrelated. Now let $\sigma_X^2 = \frac{1}{2}\mathbb{E}[|x_{ij}|^2]$. Then

$$\sigma_{C+I+N}^2 = \sigma_X^2 \Omega_p + \sigma_w^2 = \sigma_C^2 + \sigma_{I+N}^2. \tag{13.136}$$

Using (13.133) and assuming that σ_Δ^2 is sufficiently small such that $\Omega_p \sigma_\Delta^2 \approx 0$, the following $C/(I+N)$ estimate is obtained

$$\widehat{\text{CIR}} = \left(\frac{\sigma_{C+I+N}^2}{\hat{\sigma}_{I+N}^2} - 1 \right) \approx \frac{\sigma_C^2}{\sigma_{I+N}^2}. \tag{13.137}$$

The above approximation becomes exact when \mathbf{X} is known exactly. This would be the case if \mathbf{X} is formed by using elements from a known training sequence. Finally, by replacing ensemble averages with empirical averages, the following empirical $C/(I+N)$ estimate is obtained

$$\widehat{\text{CIR}} = \frac{\frac{1}{M+1} \sum_{i=1}^P \mathbf{v}_i^H \mathbf{v}_i}{\sum_{i=1}^P \frac{\mathbf{v}_i^H \hat{\mathbf{c}}_i^* \hat{\mathbf{c}}_i^T \mathbf{v}_i}{\|\hat{\mathbf{c}}_i\|^2}} - 1. \tag{13.138}$$

13.8.2 Training Sequence Based $C/(I + N)$ Estimation

The bursts in TDMA cellular systems contain known training and/or color code sequences. The color code sequences are used for BS and sector identification, while the training sequences are used for synchronization and channel estimation. As can be seen from the previous section, the $I + N$ and $C/(I + N)$ estimators will only work well when σ_{Δ}^2 is small. Fortunately, if the $I + N$ and $C/(I + N)$ estimators are constructed from the training and color code sequences,⁸ then $\sigma_{\Delta}^2 = 0$.

The $I + N$ and $C/(I + N)$ estimators of the previous section were evaluated through the software simulation for an IS-54/136 cellular system [104]. Although the IS-54/136 standard is now extinct, it still serves as a useful example on how the $I + N$ and $C/(I + N)$ estimators can be applied. The IS-54 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 six symbol color code sequence in the middle of the burst as shown in Fig. 1.2. For simulation purposes, a two-equal-ray, T -spaced, Rayleigh fading channel was chosen with uncorrelated taps. If tap correlation is present, no changes are required to the proposed $I + N$ and $C/(I + N)$ estimators because the estimates depend only on the fading envelope power Ω_p . Shadowing is assumed to remain constant over the estimates and is neglected. Finally, it is assumed that the receiver has correctly synchronize onto each of the received bursts, i.e., perfect timing recovery is assumed.

Four consecutive symbols were used to form a 3×2 Toeplitz non-symmetric matrix \mathbf{X} . Let $\{v_1(i), \dots, v_{14}(i)\}$ denote the 14 received samples corresponding to the training sequence and $\{v_{15}(i), \dots, v_{21}(i)\}$ denote the six received samples corresponding to the color code of the i th frame. From the training sequence 4 estimates of $I + N$ and $C/(I + N)$ were formed by using the following four sets:

$$\{\{v_1(i), \dots, v_4(i)\}, \{v_5(i), \dots, v_8(i)\}, \{v_9(i), \dots, v_{12}(i)\}, \{v_{11}(i), \dots, v_{14}(i)\}\},$$

where the fourth set shares two samples with the third set. Likewise, 2 estimates of $I + N$ and $C/(I + N)$ were formed from the six symbol color code sequence by using the two sets

$$\{\{v_{15}(i), \dots, v_{18}(i)\}, \{v_{17}(i), \dots, v_{21}(i)\}\} \quad (13.139)$$

which share two common samples. Although the $I + N$ and $C/(I + N)$ estimators in (13.134) and (13.138) assume independent received sample vectors the additional estimates of $I + N$ and $C/(I + N)$ which use overlapped symbols at the ends of the training and color code sequences were 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 symbol sequences. AWGN at 20 dB below the interference power was also included.

To evaluate the performance of the $I + N$ estimator, the average absolute percentage error between the $I + N$ estimate and the true interference plus noise power is defined as

$$\frac{|\hat{\sigma}_{I+N}^2 - \sigma_{I+N}^2|}{\sigma_{I+N}^2} \times 100. \quad (13.140)$$

Figure 13.31 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. 13.31 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.

Figure 13.32 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. 13.32.

⁸In the IS-54 and PDC cellular systems, the color code sequence is known provided that the MS has correctly determined its serving BS.

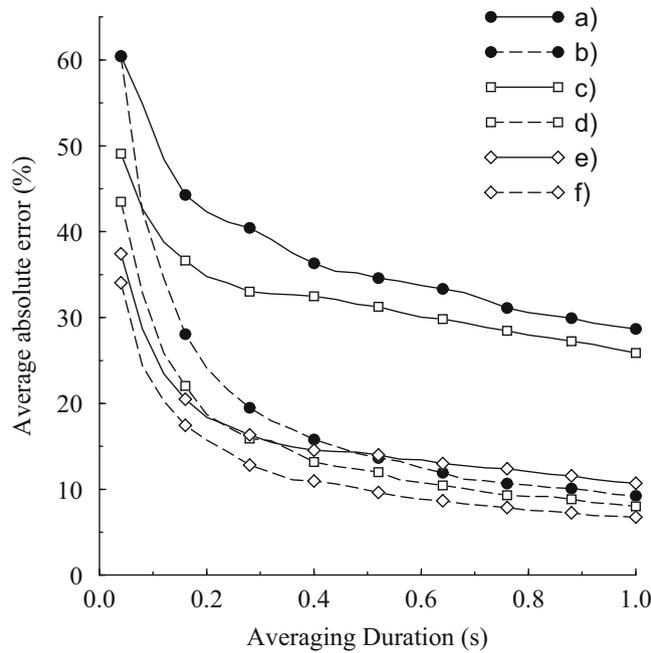


Fig. 13.31 Average absolute percent error of the $I + N$ estimator against the averaging time, from [30], [29]. The frame duration is 40 ms. Legend: a) $N_l = 1, v = 5$ km/h, b) $N_l = 1, v = 100$ km/h, c) $N_l = 2, v = 5$ km/h, d) $N_l = 2, v = 100$ km/h, e) $N_l = 6, v = 5$ km/h, f) $N_l = 6, v = 100$ km/h

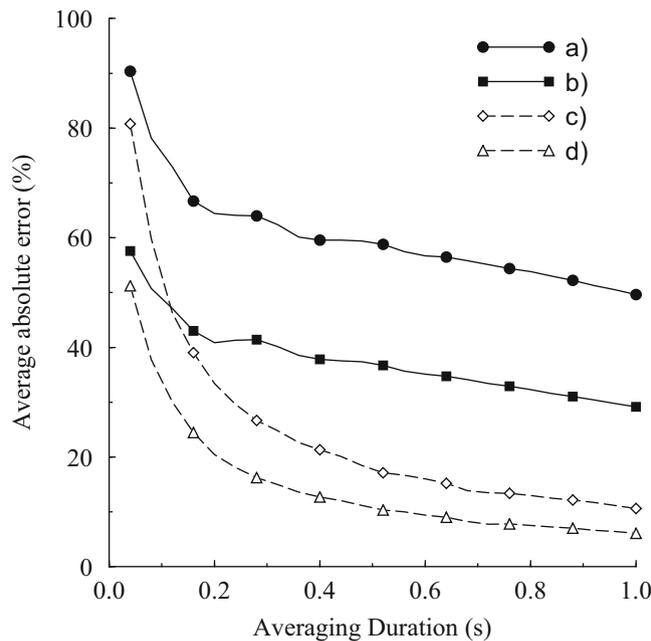


Fig. 13.32 Average absolute percent error of the $C + I + N$ estimator against the averaging time, from [29], [30]. The frame duration is 40 ms. Legend: a) $N_l = 1, v = 5$ km/h, $C/I = 5$ dB, b) $N_l = 1, v = 5$ km/h, $C/I = 20$ dB, c) $N_l = 1, v = 100$ km/h, $C/I = 5$ dB, d) $N_l = 1, v = 100$ km/h, $C/I = 20$ dB

Finally, Fig. 13.33 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 a high speed MS, the $C/(I + N)$ can be estimated to within 2 dB in less 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. 13.34.

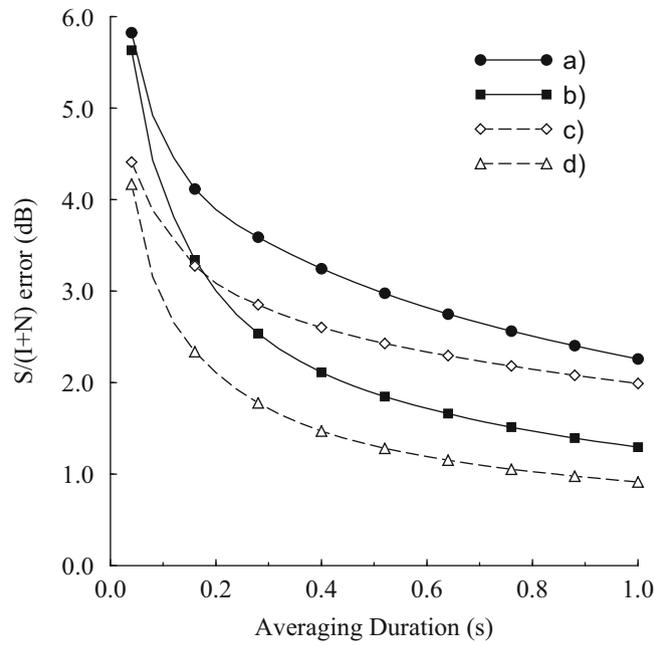


Fig. 13.33 Average error of the $C/(I+N)$ estimator against the averaging time, from [29], [30]. The frame duration is 40 ms. Legend: a) $N_I = 1$, $v = 5$ km/h, $C/I = 5$ dB, b) $N_I = 1$, $v = 100$ km/h, $C/I = 5$ dB, c) $N_I = 6$, $v = 5$ km/h, $C/I = 5$ dB, d) $N_I = 6$, $v = 100$ km/h, $C/I = 5$ dB

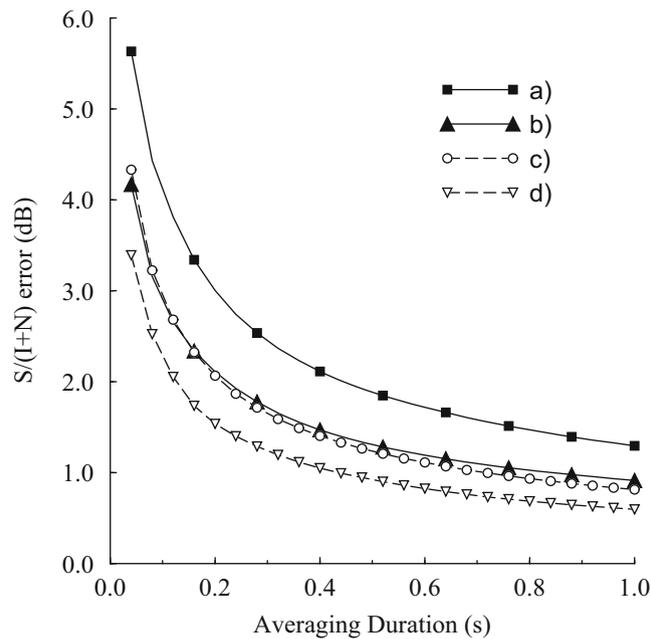


Fig. 13.34 Average error of the $C/(I+N)$ estimator for half rate and full rate channels against the averaging time, from [29], [30]. The frame duration is 40 ms. Legend: a) $N_I = 1$, $v = 100$ km/h, $C/I = 5$ dB, half rate channel, b) $N_I = 6$, $v = 100$ km/h, $C/I = 5$ dB, half rate channel, c) $N_I = 1$, $v = 100$ km/h, $C/I = 5$ dB, full rate channel, d) $N_I = 6$, $v = 100$ km/h, $C/I = 5$ dB, full rate channel

Appendix 13A: Derivation of Equations (13.43) and (13.58)

The limit in (13.58) can be written as

$$\lim_{\tau \rightarrow 0} \frac{\frac{\lambda_c}{2\pi\tau} \sqrt{\frac{2\tilde{\lambda}_{rr}(0) - 2\tilde{\lambda}_{rr}(\tau)}{\tilde{\lambda}_{rr}(0)}}}{\frac{\lambda_c}{2\pi\tau} \sqrt{\frac{2\lambda_{rr}(0) - 2\lambda_{rr}(\tau)}{\lambda_{rr}(0)}}} = \frac{\lim_{\tau \rightarrow 0} \frac{\lambda_c}{2\pi\tau} \sqrt{\frac{2\tilde{\lambda}_{rr}(0) - 2\tilde{\lambda}_{rr}(\tau)}{\tilde{\lambda}_{rr}(0)}}}{\lim_{\tau \rightarrow 0} \frac{\lambda_c}{2\pi\tau} \sqrt{\frac{2\lambda_{rr}(0) - 2\lambda_{rr}(\tau)}{\lambda_{rr}(0)}}}. \quad (13A.1)$$

Note that the limit of the denominator gives (13.43) and is a special case of the numerator limit with $N_o = 0$. To find the numerator limit, the following property can be used [309]:

If a function $f(\tau)$ has a limit as τ approaches a , then

$$\lim_{\tau \rightarrow a} \sqrt[n]{f(\tau)} = \sqrt[n]{\lim_{\tau \rightarrow a} f(\tau)} \quad (13A.2)$$

provided that either τ is an odd positive integer or n is an even positive integer and $\lim_{\tau \rightarrow a} f(\tau) > 0$.

Therefore, if the limit

$$\zeta = \lim_{\tau \rightarrow 0} f^2(\tau) = \lim_{\tau \rightarrow 0} \frac{\lambda_c^2}{(2\pi\tau)^2} \frac{2\tilde{\lambda}_{rr}(0) - 2\tilde{\lambda}_{rr}(\tau)}{\tilde{\lambda}_{rr}(0)} \quad (13A.3)$$

exists and is positive, the solution to (13A.1) can be determined. It is apparent that L'Hôpital's Rule should be applied to determine the limit in (13A.3). After substituting $\tilde{\lambda}_{rr}(\tau)$ from (13.34) and applying L'Hôpital's Rule four times, the limit is

$$\begin{aligned} \zeta = & \frac{\lambda_c^2 (B_w^4 N_o^2 \pi^2 + 3 B_w K N_o (2\pi f_m)^2 b_0 + 2 B_w^3 K N_o \pi^2 b_0 + 2 B_w^3 N_o \pi^2 a(0))}{6 \pi^2 (B_w N_o + 2 a(0)) (B_w N_o + 4 K b_0 + 2 a(0))} \\ & + \frac{\lambda_c^2 (6 K (2\pi f_m)^2 b_0 a(0) + 3 B_w K N_o (2\pi f_m)^2 b_0 \cos(2 \theta_0))}{6 \pi^2 (B_w N_o + 2 a(0)) (B_w N_o + 4 K b_0 + 2 a(0))} \\ & + \frac{\lambda_c^2 (6 K (2\pi f_m)^2 b_0 a(0) \cos(2 \theta_0))}{6 \pi^2 (B_w N_o + 2 a(0)) (B_w N_o + 4 K b_0 + 2 a(0))} \\ & + \frac{\lambda_c^2 (-12 a'(0)^2 - 12 c'(0)^2 - 6 B_w N_o a''(0) - 12 K b_0 a''(0))}{6 \pi^2 (B_w N_o + 2 a(0)) (B_w N_o + 4 K b_0 + 2 a(0))} \\ & + \frac{\lambda_c^2 (-12 a(0) a''(0) - 12 c(0) c''(0))}{6 \pi^2 (B_w N_o + 2 a(0)) (B_w N_o + 4 K b_0 + 2 a(0))}, \end{aligned} \quad (13A.4)$$

where $a(\tau)$ and $c(\tau)$ are given by (13.35) and (13.36), respectively, and $x'(0)$ denotes the derivative of $x(t)$ evaluated at $t = 0$. Consequently, $a(0) = b_0$, $a'(0) = c(0) = 0$, and

$$\begin{aligned} a''(0) &= b_0 (2\pi f_m)^2 \int_0^{2\pi} \hat{p}(\theta) \cos^2(\theta) d\theta \\ c'(0) &= b_0 2\pi f_m \int_0^{2\pi} \hat{p}(\theta) \cos(\theta) d\theta. \end{aligned} \quad (13A.5)$$

Using these, and the identity $\cos^2(\theta) = (1 + \cos(2\theta))/2$, it can be shown that (13A.4) is positive for all θ for the scattering distributions considered in Sect. 13.4.3.1. Consequently, applying the theorem in (13A.2), the limit of the numerator of (13A.1) is the square root of (13A.4), which if desired can be expressed in terms of the signal-to-noise ratio γ_S using

$$b_0 = \frac{\gamma_S N_o B_w}{2(K + 1)} \quad (13A.6)$$

Fig. 13.35 MS traversing from BS₀ to BS₁ along a handoff route



from (13.53) with $K = s^2/2b_0$. The denominator of (13A.1), which is also (13.43), is obtained by assuming isotropic scattering and no noise, so that $a(0) = b_0$, $a'(0) = c(0) = c'(0) = c''(0) = 0$, $a''(0) = 2b_0(\pi f_m)^2$, and $N_o = 0$ in (13A.4). After taking the square root, the result is

$$\lim_{\tau \rightarrow 0} \frac{\lambda_c}{2\pi\tau} \sqrt{\frac{2\lambda_{rr}(0) - 2\lambda_{rr}(\tau)}{\lambda_{rr}(0)}} = v \sqrt{\frac{(1 + 2K + K \cos(2\theta_0))}{(1 + 2K)}}. \quad (13A.7)$$

Problems

13.1. Suppose that a MS is traveling along a straight line from BS₁ to BS₂, as shown in Fig. 13.35. The BSs are separated by distance D , and the MS is at distance r from BS₁ and distance $D - r$ from BS₂. 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 is given by (1.4).

- A handoff from BS₁ to BS₂, or vice versa, can never occur if $|\Omega_{1 \text{ (dB)}} - \Omega_{2 \text{ (dB)}}| < H$ but may or may not occur otherwise.
 - A handoff from BS₁ to BS₂ will occur if the MS is currently assigned to BS₁ and $\Omega_{2 \text{ (dB)}} \geq \Omega_{1 \text{ (dB)}} + H$.
- (a) Find an expression for the probability that a handoff can never occur from BS₁ to BS₂, or vice versa.
 (b) Given that the MS is currently assigned to BS₁ what is the probability that a handoff will occur from BS₁ to BS₂.

13.2. 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.

13.3. A MS is moving with speed 100 km/h along a straight line between two base stations, BS₁ and BS₂. For simplicity ignore envelope fading and shadowing, and consider only the path loss. The received power (in dBm) follows the characteristic

$$\mu_{\Omega_p \text{ (dBm)}}(d_i) = \mu_{\Omega_p \text{ (dBm)}}(d_o) - 10\beta \log_{10}(d_i/d_o) \text{ (dBm)},$$

where d_i is distance from BS _{i} in meters. Assume $\mu_o = 0$ dBm at $d_o = 1$ m, and let the path loss exponent be $\beta = 3.0$.

Assume that the minimum usable signal quality at the receiver (at either link end) is $\mu_{\min} = -88$ dBm. The mobile station is connected to BS₁ and signal level at/from BS₁ is measured. The measured signal level is compared to a threshold μ_{HO} ; if it drops below the threshold a handoff is initiated. Once the handoff is initiated it takes 0.5 s to complete.

- (a) Determine the minimum margin $\Delta = \mu_{\text{HO}} - \mu_{\min}$ so that calls are not lost due to weak signal strength during a handoff.
 (b) What is the maximum allowable distance between BS₁ and BS₂?
 (c) Describe the effects of the margin on the *link quality performance* and *capacity* of a cellular system.

13.4. Derive Eq. (13.19).

13.5. Derive Eq. (13.21).

13.6. Derive Eq. (13.34).

13.7. Derive Eq. (13.52).

13.8. Consider a communication link operating over a channel with propagation path loss exponent $\beta = 3.5$ and a shadow standard deviation $\sigma_{\Omega} = 8$ dB.

- (a) Consider the case of two adjacent cells. A mobile station is transmitting at its maximum power and is located exactly on the cell boundary between the two base stations. In the absence of shadowing, the received power level would be equal to the receiver sensitivity, S_{RX} , i.e., the mobile station is located at distance d_{max} from each base station.

Now assume that shadowing is present and a soft handoff algorithm is used, such that the least attenuation link is always selected. The shadows experienced on the two possible links are independent. If an outage probability of 10% is desired for the given mobile station location (averaged over a large ensemble of realizations), what is the required margin ($M_{shad} - G_{HO}$), where M_{shad} is the required shadow margin for a single isolated cell and G_{HO} is the soft handoff gain?

- (b) What is the value of G_{HO} ?
- (c) Repeat parts (a) and (b) assuming that the mobile is located on the boundaries of and equidistant from three base stations, i.e., the mobile station is located at distance d_{max} from three base stations.

Chapter 14

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 service quality for the MS at any instant, and relieves undesired network congestion. As shown in Fig. 14.1, there are three basic types of channel assignment algorithms, fixed, flexible, and dynamic [317].

Fixed channel assignment (FCA) was used by first 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.

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 non-stationary 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 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 non-stationary 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.

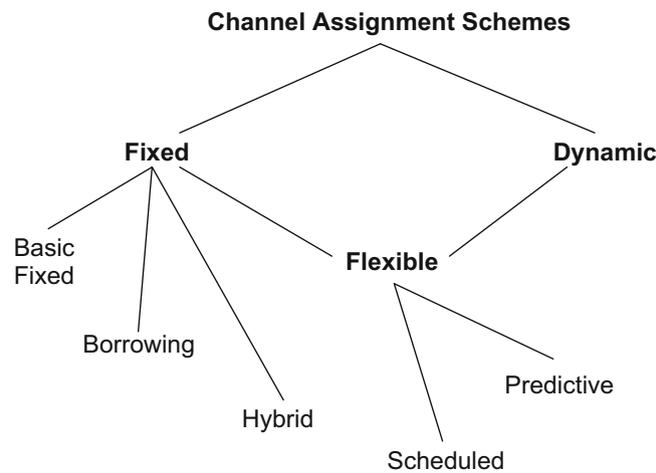


Fig. 14.1 Basic classifications of channel assignment schemes, from [317]

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) [117], 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 [281].

Fully decentralized DCA schemes are the other extreme and require no network planning or communication among BSs [11, 120, 253]. 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) [240]. 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 [240]. Another distributed DCA scheme is simple dynamic channel assignment (SDCA) [349]. SDCA performs slightly worse than DRA, but requires communication among a smaller set of cells called the interference neighborhood [349].

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 [310]. 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 than 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 [169]. Another method is to queue the handoff requests (but not the new call arrivals) [134, 169]. 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, some of the many different DCA schemes will be briefly outlined, 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 Sect. 14.1. Decentralized DCA strategies such as First Available (FA), Nearest Neighbor (NN), and dynamic resource acquisition (DRA) are discussed in Sect. 14.2. Fully decentralized DCA schemes are the topic of Sect. 14.3, including Channel Segregation (CS) and Minimum Interference (MI), along with aggressive and timid strategies. Hybrid FCA/DCA schemes are the subject of Sect. 14.4. The important class of borrowing schemes are the topic of Sect. 14.5, 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 and reuse partitioning, and handoff priority schemes in Sects. 14.6, 14.7, 14.8 and 14.9 respectively.

Following our results in [349], Sect. 14.10 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.

14.1 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.

14.1.1 Maximum Packing

The maximum packing (MP) algorithm was originally presented by Everitt and Macfadyen in 1983 [116]. 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 [178] 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 rearrangements; otherwise, the call is blocked. For example, consider the simple system consisting of five cells shown in Fig. 14.2. 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:

\mathcal{R} = set of cells in the system.

$K = |\mathcal{R}|$ = number of cells in the system.

N_T = total number of channels available.

n_i = number of calls in progress in cell i .

$\mathbf{n} = (n_i, i \in \mathcal{R})$ = state vector.

\mathcal{S} = set of admissible states.

ρ_i = traffic load in cell i .

The set of admissible states depends on the particular cell layout. Let J be the number of complete or partial reuse clusters $CL_i, i = 1, \dots, J$ that can be defined such that (1) each reuse cluster differs by at least one cell, i.e., they are not totally overlapping, and (2) all cells are contained in at least one such cluster. For the example in Fig. 14.2, $J = 3$. 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. 14.2

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}. \quad (14.1)$$

Fig. 14.2 Five cell deployment with maximum packing (MP)

1	2	3	4	5
---	---	---	---	---

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 $a_{3,2} = 0$. Finally, let $N_i, i = 1, \dots, J$ be the number of channels that are available in CL_i , $N_i \leq N_T$, and $\mathbf{N} = (N_1, \dots, N_J)$. Then the set of admissible states is then given by

$$\mathcal{S} = \{\mathbf{n} : \mathbf{A}\mathbf{n}^T \leq \mathbf{N}\}. \quad (14.2)$$

It is well known (e.g., [177]) that \mathbf{n} has the steady-state distribution

$$\pi(\mathbf{n}) = G(\mathbf{N}) \prod_{i \in \mathcal{R}} \frac{\rho_i^{n_i}}{n_i!}, \quad \mathbf{n} \in \mathcal{S} \quad (14.3)$$

where $G(\mathbf{N})$ is the normalizing constant

$$G(\mathbf{N}) = \left(\sum_{\mathbf{n} \in \mathcal{S}} \prod_{i \in \mathcal{R}} \frac{\rho_i^{n_i}}{n_i!} \right)^{-1}. \quad (14.4)$$

Then the steady-state probability that a call arrival in cell i is blocked is

$$B_i = 1 - \frac{G(\mathbf{N})}{G(\mathbf{N} - \mathbf{A}\mathbf{e}_i^T)} \quad (14.5)$$

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

$$B_i \approx 1 - \prod_{j \in \text{CL}_i} (1 - E_j) \quad (14.6)$$

where the $E_j, j = 1, \dots, J$ solve the nonlinear equations

$$E_j = E \left(\sum_{r \in \mathcal{R}} a_{jr} \rho_r \prod_{i \in r - \{j\}} (1 - E_i), N_j \right), \quad j = 1, \dots, J \quad (14.7)$$

with

$$E(\rho, N) = \frac{\rho^N}{N!} \left(\sum_{n=0}^N \frac{\rho^n}{n!} \right)^{-1} \quad (14.8)$$

being the Erlang-B formula. Kelly [178] 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 $\text{CL}_i, i \in r - \{j\}$ before being offered to CL_j .

14.1.2 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

$$\Lambda(\mathbf{d})_{(\text{dB})} = \Omega(d_i)_{(\text{dB})} - 10 \log_{10} \sum_{k \in I} 10^{\Omega(d_k)_{(\text{dB})}/10}. \quad (14.9)$$

where the $\Omega(d_j)_{(\text{dB})}$ 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

$$\max_{j \in C} \min_{i \in S} \{\Lambda_i\} \quad (14.10)$$

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. Methods for C/I monitoring have been presented in Sect. 13.8.

14.2 Decentralized DCA

14.2.1 First Available and Nearest Neighbor

In 1972, Cox and Reudnik [85] 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.

14.2.2 Dynamic Resource Acquisition

Nanda and Goodman [240] 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 [240], 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. Figure 14.3 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.

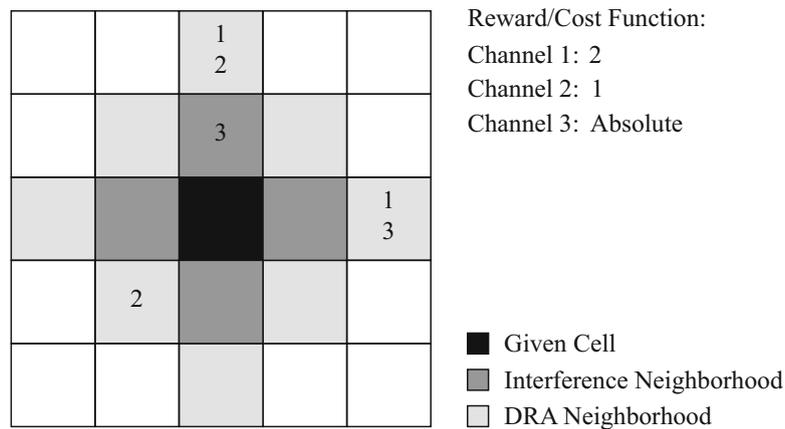


Fig. 14.3 The DRA (light shaded) and interference (dark shaded) neighborhoods of a cell under consideration (shaded black). The numbers within each cell indicate active channels

14.3 Fully Decentralized DCA

14.3.1 Channel Segregation

Akaiwa and Andoh [12] 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 inaccessible 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 simultaneously 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. 14.4. 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 [12] $P(i) = N_s/N_t$, 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 inaccessible, 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 priority 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 [12] demonstrated by simulation that the CS policy outperforms FCA and the FA DCA algorithm.

The steps within the dashed box in Fig. 14.4 are a modification so that the original CS algorithm developed by Akaiwa [11] 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.

14.3.2 Channel Segregation with Variable Threshold

Another Channel Segregation scheme has been proposed by Hanabe et al. [157] 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 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. proposed is as follows:

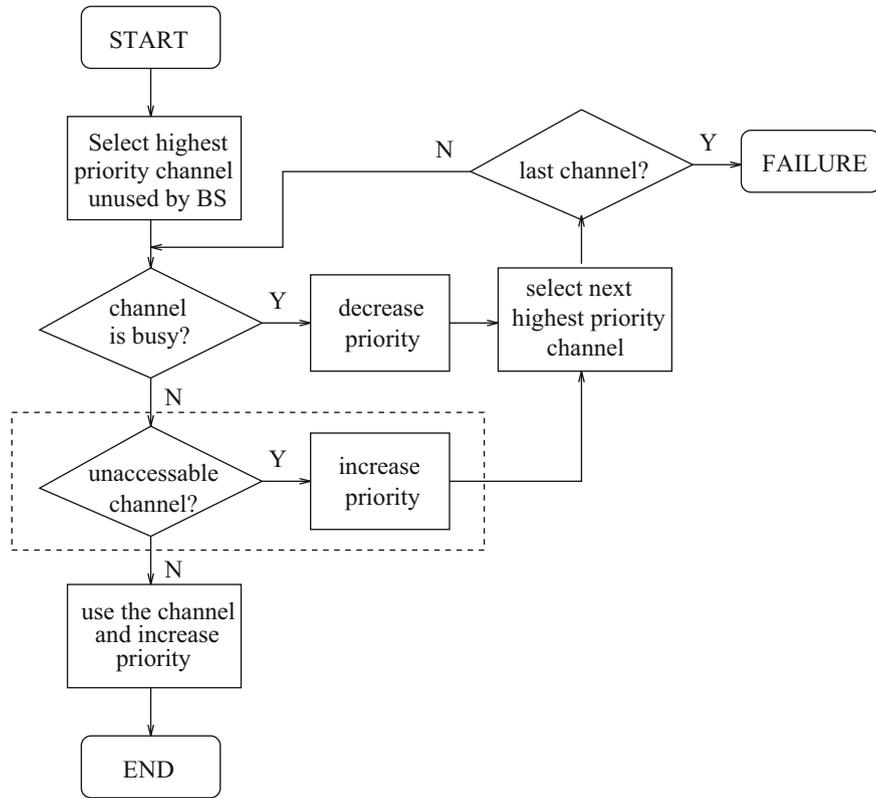


Fig. 14.4 Channel segregation (CS) algorithm

$$T(k) = \begin{cases} T_o + 15 + S, & k = 1, 2, \dots, N_1 \\ T_o + 10 + S, & k = 9, 10, \dots, N_2 \\ T_o + 5 + S, & k = 17, 18, \dots, N_3 \end{cases} \quad (14.11)$$

For the example in [157], $N_1 = 8$, $N_2 = 16$, 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 (14.11) 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, deadlocks, 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.

14.3.3 Minimum Interference 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.

14.3.4 Aggressive and Timid DCA Strategies

Distributed self-organizing DCA algorithms that use aggressive and timid strategies were first introduced by Cimini and Foshini [67]. 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 [67] 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 array. 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 [69]

$$C_{\text{ran}} = \int_0^1 \exp \left\{ 2 \sum_{i=0}^{R-1} \frac{(v^{i+1} - 1)}{i + 1} \right\} dv. \quad (14.12)$$

The saturation density can also be obtained as a function of the traffic load ρ as [69]

$$C(\rho) (1 - C(\rho)R)^R = \rho (1 - (R + 1)C(\rho))^{R+1} \quad (14.13)$$

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

$$C_{\min} = \frac{1}{1 + 3R(R + 1)} \quad (14.14)$$

$$C_{\max} = \begin{cases} \frac{4}{3(R+1)^2} & , \quad R \text{ even} \\ \frac{4}{1+3(R+1)^2} & , \quad R \text{ odd} \end{cases} \quad (14.15)$$

However, expressions for C_{ran} and $C(\rho)$ for the hexagonal planar array are unknown.

¹The 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 $j = R/2 + 1$.

Table 14.1 Normalized channel utilizations $\delta = C_{\text{ran}}/C_{\text{max}}$ for the timid DCA algorithm, from [68]

R	Linear	Planar hexagonal
1	0.864	0.693
2	0.825	0.658
3	0.804	0.627

For the case of a single channel the blocking probability has the exact form [69]

$$P_b = 1 - C(\rho)/\rho. \quad (14.16)$$

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 [68]. 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 14.1. 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 (14.8), 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, ρ is replaced 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, ρ is replaced 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, 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.

14.4 Hybrid FCA/DCA Schemes

DCA schemes perform very well for light non-stationary nonhomogeneous 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 [86] 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, and NN+1. 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.

14.5 Borrowing Schemes

Engel and Peritsky [109] introduced a 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 [16]. 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.

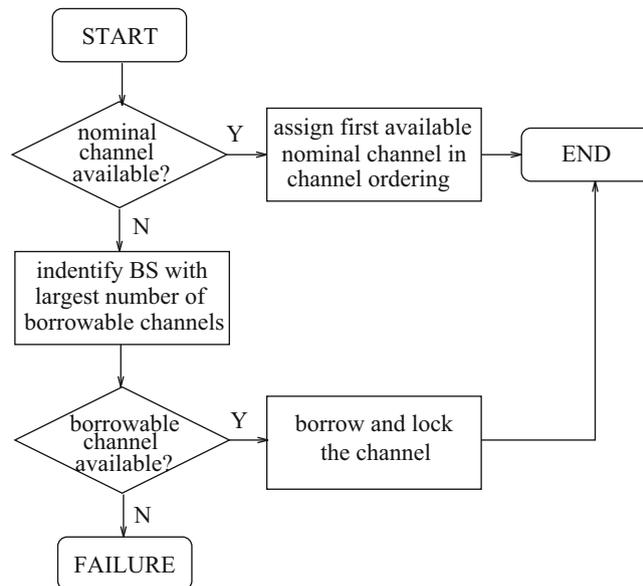


Fig. 14.5 Call arrival policy for the BCO algorithm

14.5.1 Borrowing with Channel Ordering

Elnoubi et al. [108] proposed a channel borrowing strategy that makes use of 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 flowchart in Fig. 14.5. 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. 14.7. 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 will 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. 14.6. 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 [187] 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 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 the 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.

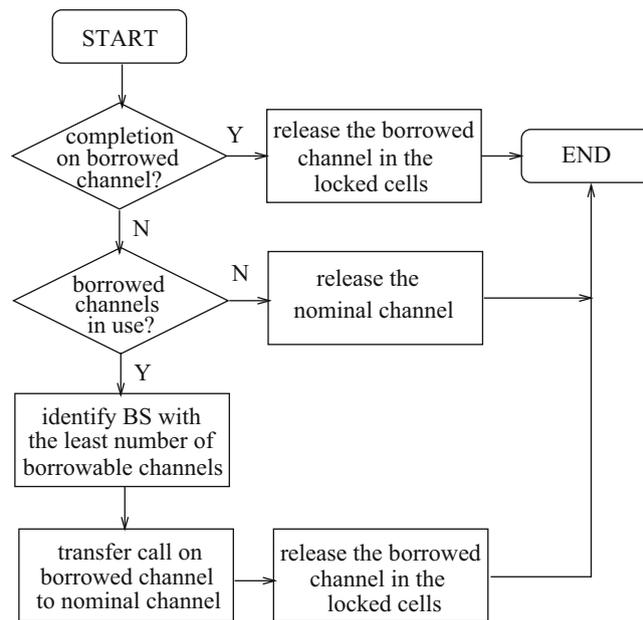


Fig. 14.6 Call departure policy for the BCO algorithm

14.5.2 Borrowing with Directional Locking

Zhang and Yum [375] 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. 14.7, 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 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. [108], except that the directional locking mechanism is accounted for. Zhang and Yum [375] 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.

14.5.3 Borrowing Without Locking

A borrowing scheme, Channel Borrowing Without Locking (CBWL), has been proposed by Jiang and Rappaport [172] that does not require channel locking by using borrowed channels with a reduced power level to limit interference 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 i th group can only be lent to the i th adjacent cell. This principle of directional lending is illustrated in Fig. 14.8, 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 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.

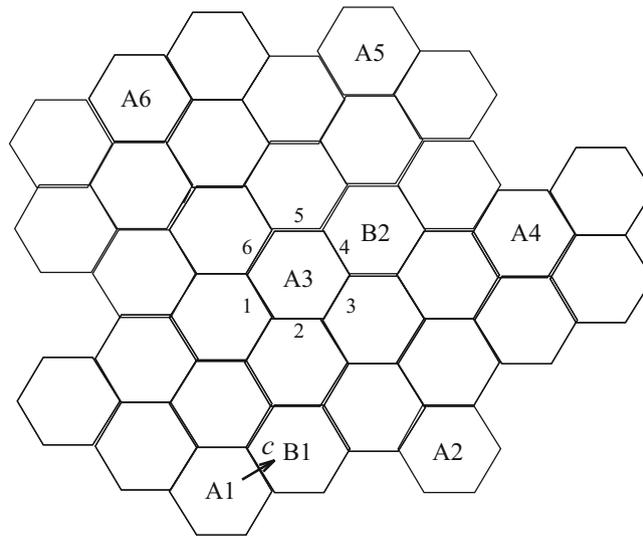


Fig. 14.7 Principle of Borrowing with Directional Channel Locking (BDCL)

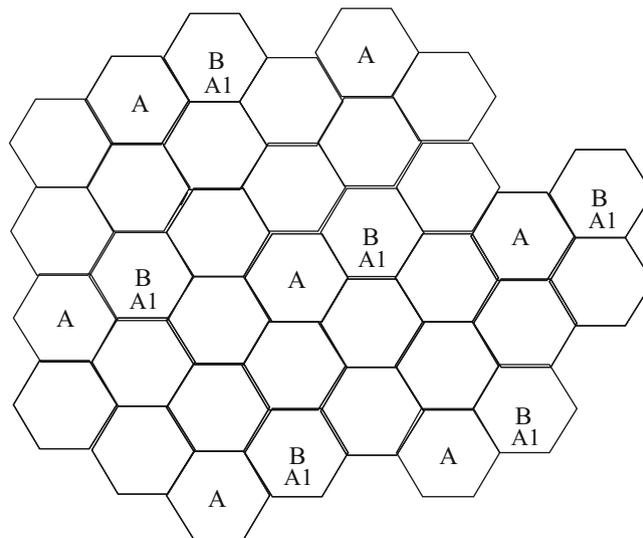


Fig. 14.8 Principle of Channel Borrowing Without Locking (CBWL)

Various forms of channel rearrangements can be used to 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.

14.5.4 Compact Pattern Based DCA

Yeung and Yum [367] 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: (1) channel acquisitions where an optimal idle channel is assigned to the MS, and (2) 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 noncompact 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.

14.6 Directed Retry and Directed Handoff

Everitt [115] 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.

14.7 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 [247, 248]. 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 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 [248]. 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) [246]. MS are divided into two classes: high speed MS (HSMS) who are traveling at 50–60 km/h and low speed MS (LSMS) who are traveling at 0–4 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, a variety of velocity estimation techniques are available as discussed in Sect. 13.4. Moving direction information can be obtained by using the past signal strength history in LOS environments [26] or the sign of the Doppler.

14.8 Reduced Transceiver Coverage

Takeo et al. [311] 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 [311] 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. [311] 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.

14.8.1 Reuse Partitioning

Reuse partitioning employs a two-level cell plan where clusters of size M are overlaid on clusters of size N , $N > M$. Figure 9.20 shows a FCA scheme using reuse partitioning with $M = 3$ and $N = 9$. As discussed in Sect. 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 [174]. 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 [133]. 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 .

14.9 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 [169], 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 [134, 169]. 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 [317]. 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.

14.10 Example DCA Schemes for TDMA Systems

A TDMA system is assumed 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

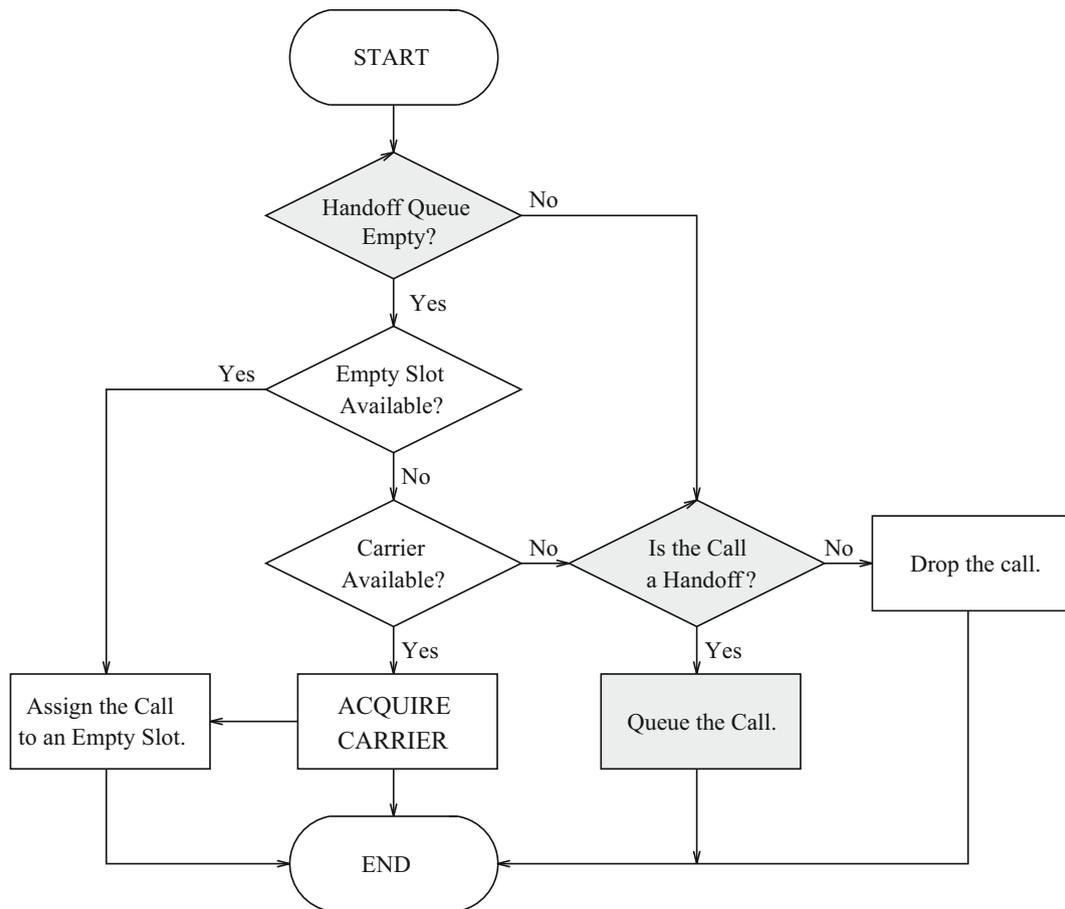


Fig. 14.9 ACQUIRE CARRIER policy that is executed upon a new call or handoff arrival

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 flowcharts in Figs. 14.9 and 14.10 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. 14.9 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. 14.10 shows, the following policy is executed upon a call completion or a handoff departure:

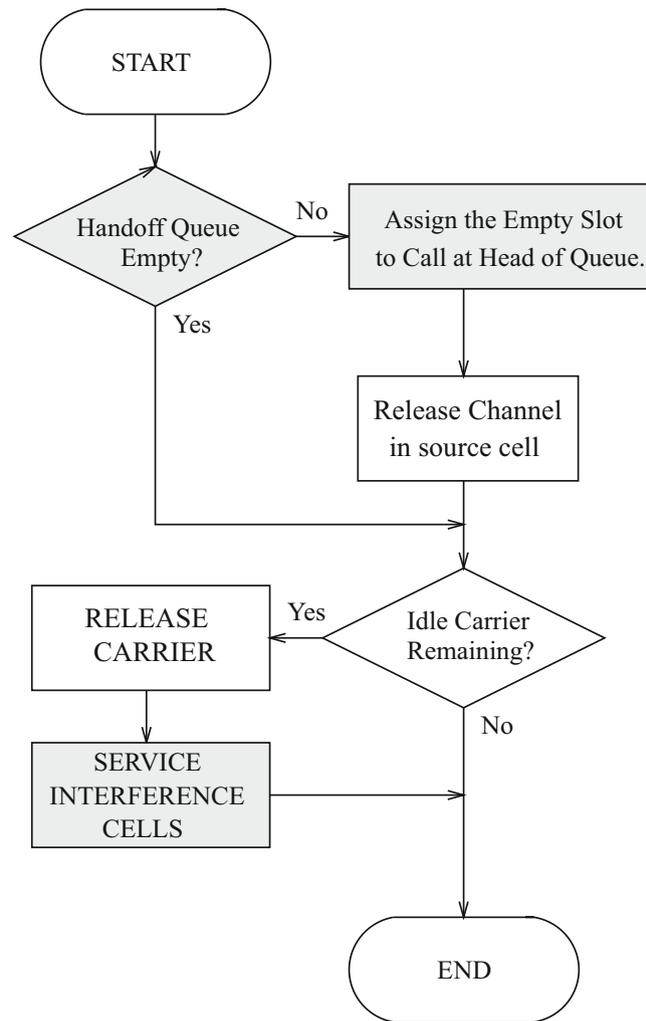


Fig. 14.10 RELEASE CARRIER policy that is executed upon a call completion or handoff departure

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.

14.10.1 The Simple DCA (SDCA) Strategy

Elnoubi et al. [108] proposed the BCO strategy that makes use of different channel orderings in each cell. Here, a channel assignment strategy is considered that uses carrier orderings rather than channel orderings [349]. 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.

$$\begin{aligned}
 A &: \{1, 4, 7, 8, 9, 5, 6, 2, 3\} \\
 B &: \{2, 5, 8, 9, 7, 6, 4, 3, 1\} \\
 C &: \{3, 6, 9, 7, 8, 4, 5, 1, 2\}
 \end{aligned}
 \tag{14.17}$$

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 third column, columns 6 and 7 are permutations of the second column, and columns 8 and 9 are permutations of the first 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.

14.10.2 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 [134, 169], 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 handoff queueing is combined 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, it may be assumed 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 (14.17) where 9 carriers were distributed among three carrier orderings, the owned and borrowed carrier orderings are

$$\begin{array}{cc}
 \text{Owned} & \text{Borrowed} \\
 A &: \{1, 4, 7\} \{8, 9, 5, 6, 2, 3\} \\
 B &: \{2, 5, 8\} \{9, 7, 6, 4, 3, 1\} \\
 C &: \{3, 6, 9\} \{7, 8, 4, 5, 1, 2\}
 \end{array}
 \tag{14.18}$$

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 combining DCA and handoff queueing is described below along with the flowcharts in Figs. 14.9, 14.10, 14.11, and 14.12.

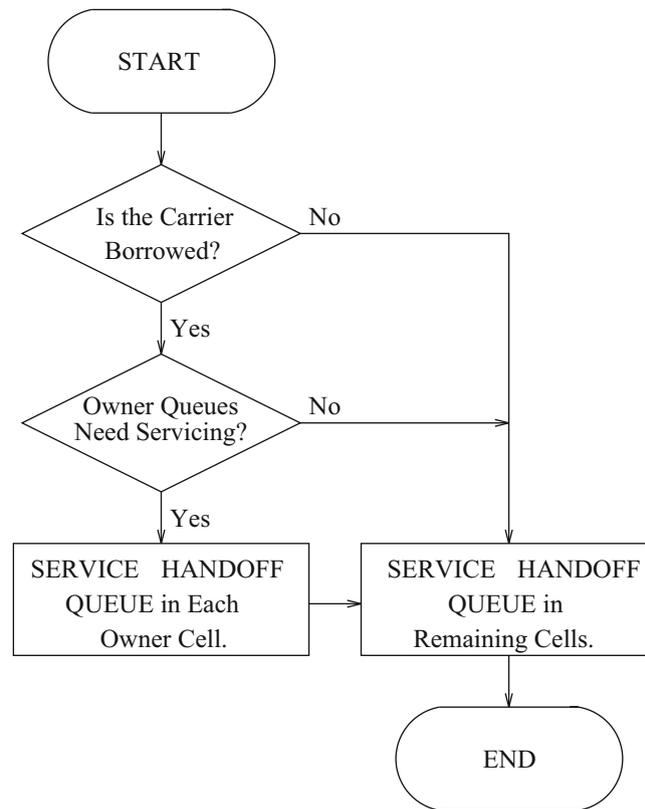


Fig. 14.11 SERVICE INTERFERENCE CELLS Policy

ACQUIRE CARRIER Policy (Queueing Case)

Referring to Fig. 14.9, 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 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. 14.10, 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. 14.11, the following policy is executed whenever a carrier is released:

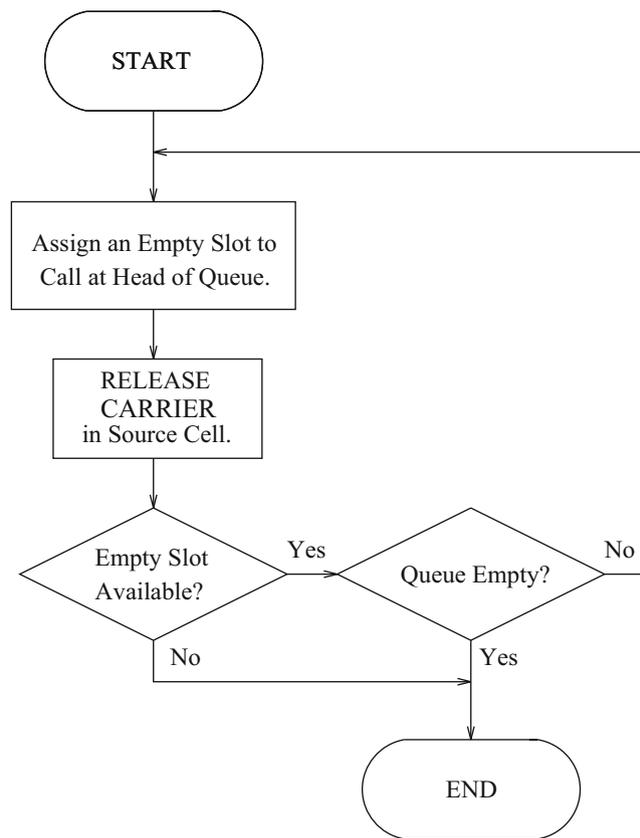


Fig. 14.12 SERVICE HANDOFF QUEUE Policy

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. 14.12, 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.
 - (a) Step 1. is executed until either all of the available slots are filled or the handoff queue is empty.

14.10.3 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 [171] 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 assigned, rather than where they should be assigned, and capacity suffers [117]. Under such conditions DCA strategies usually perform *worse* than FCA strategies.

As discussed in Sect. 14.3.4 the performance at high traffic loads can be improved by using an aggressive policy where, under certain conditions, a cell 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. 14.13, 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

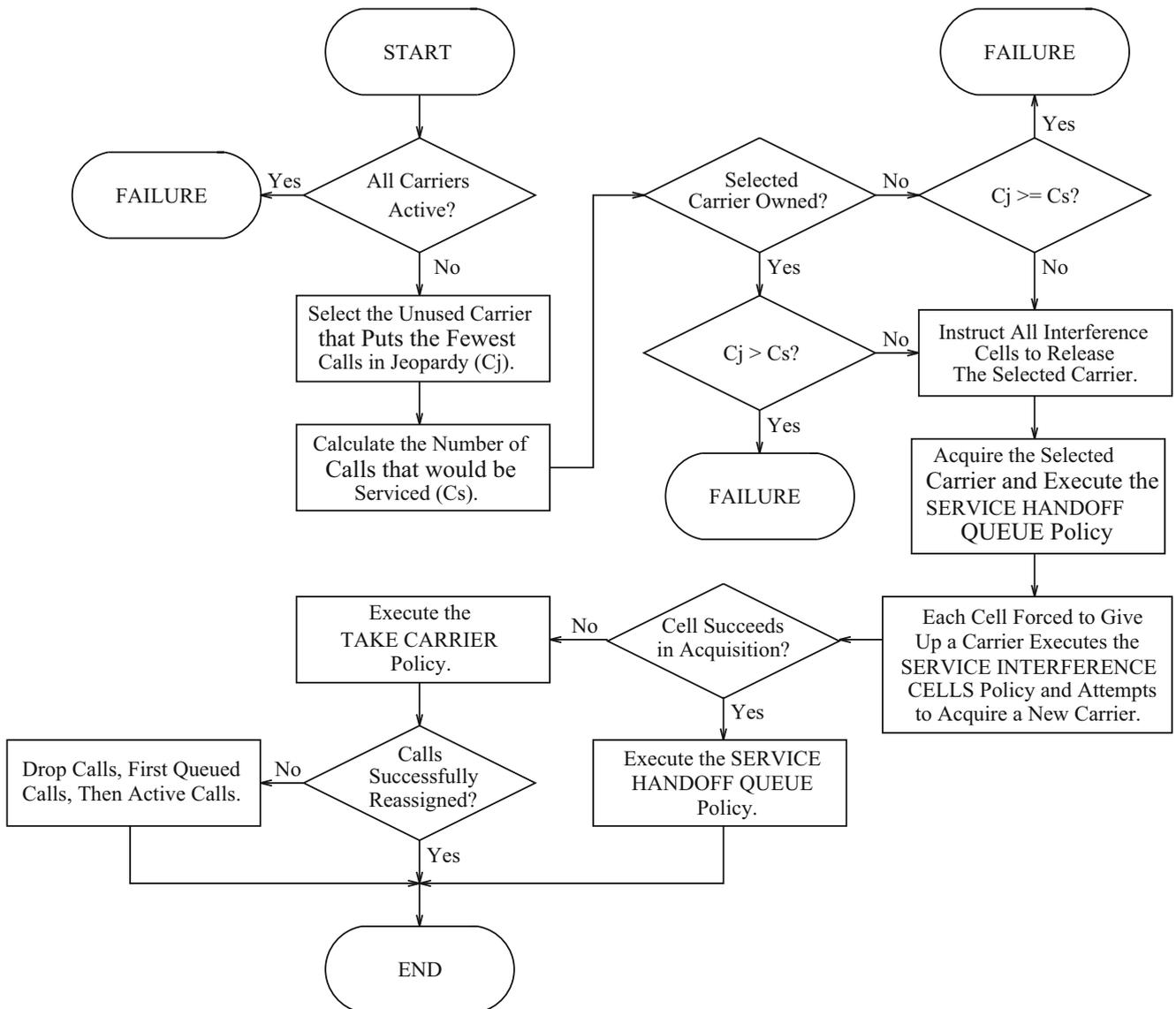


Fig. 14.13 TAKE CARRIER Policy

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 SERVICE 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 SERVICE 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 CARRIER policy acquires carriers that place the fewest number of calls in jeopardy. From a practical standpoint, the aggressive SCDA strategy requires communication among BSs in the interference neighborhood to execute the TAKE CARRIER policy.

14.10.4 Simulation Model, Results, and Discussion

Consider a microcellular environment consisting of a rectangular grid of intersecting streets, as shown in Fig. 14.14. 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.

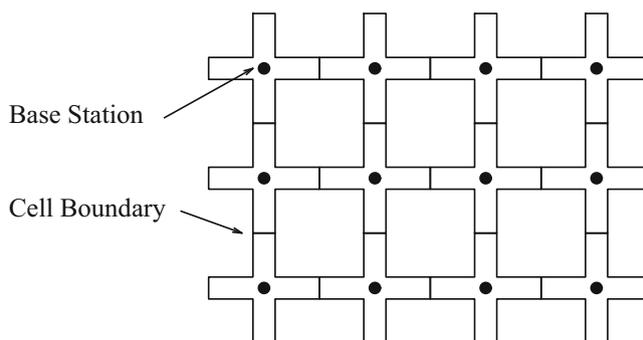


Fig. 14.14 Cells and BSs in an urban microcellular environment

²When 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.

Fig. 14.15 Interference neighborhood and cell reuse pattern

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

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 are shown in Fig. 14.15.

To account for the uneven distribution of teletraffic in the microcellular environment the identical active-dormant Markov model from [240] 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 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 $\lambda \in \{\lambda_{\text{ACT}}, \lambda_{\text{DOR}}\}$. 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

$$\Lambda = \sum_i \lambda_i. \tag{14.19}$$

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

$$r_{c,i} = \frac{N_{\text{active},i} + N_{\text{queued},i}}{\mu}. \tag{14.20}$$

The completion of calls in different cells are assumed to be independent. Therefore, the global call completion rate is:

$$r_c = \sum_i r_{c,i}. \tag{14.21}$$

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

$$r_{h,i} = \frac{hN_{\text{active},i}}{\mu}. \quad (14.22)$$

Call handoffs in different cells are assumed to be independent, so the global handoff attempt rate is

$$r_h = \sum_i r_{h,i}. \quad (14.23)$$

Mode Transitions

Each cell remains in its current mode for duration D , where D is exponentially distributed with mean $1/\bar{D}$. If the cell is in active mode, then $\bar{D} = \bar{D}_{\text{ACT}}$, and if the cell is in dormant mode, then $\bar{D} = \bar{D}_{\text{DOR}}$. 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,

$$r_{\text{ACT} \rightarrow \text{DOR}} = \frac{N_{\text{ACT}}}{\bar{D}_{\text{ACT}}} \quad (14.24)$$

$$r_{\text{DOR} \rightarrow \text{ACT}} = \frac{N_{\text{DOR}}}{\bar{D}_{\text{DOR}}} \quad (14.25)$$

The probability of a cell being in the active mode is

$$P_{\text{ACT}} = \frac{\bar{D}_{\text{ACT}}}{\bar{D}_{\text{ACT}} + \bar{D}_{\text{DOR}}}. \quad (14.26)$$

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:	λ_i/Λ
Call completion:	$r_{c,i}/r_c$
Handoff Attempt:	$r_{h,i}/r_h$
ACT \rightarrow DOR transition:	$\begin{cases} 1/N_{\text{ACT}}, & \text{if cell } i \text{ is active} \\ 0, & \text{if cell } i \text{ is dormant} \end{cases}$
DOR \rightarrow ACT transition:	$\begin{cases} 1/N_{\text{DOR}}, & \text{if cell } i \text{ is dormant} \\ 0, & \text{if cell } i \text{ is active} \end{cases}$

The active to dormant traffic ratio $R_{\text{ACT/DOR}} = \lambda_{\text{ACT}}/\lambda_{\text{DOR}}$ specifies the ratio of the new call arrival rates in the active and dormant cells. To complete the model, the offered traffic per cell, ρ , is specified. Then the active and dormant call arrival rates are:

$$\lambda_{\text{ACT}} = \frac{\rho R_{\text{ACT/DOR}}}{\mu(1 + P_{\text{ACT}}(R_{\text{ACT/DOR}} - 1))} \quad (14.27)$$

$$\lambda_{DOR} = \frac{\rho}{\mu(1 + P_{ACT}(R_{ACT/DOR} - 1))} \tag{14.28}$$

where P_{ACT} is as defined in (14.26). The parameters used in the simulations are as follows:

Number of cells: 144 (12 × 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:	120 s
Average number of handoffs per call:	3
Average duration of the ACTIVE mode:	60 s
Average duration of the DORMANT mode:	600 s
ACTIVE to DORMANT traffic ratio:	5
Offered traffic:	0–50 Erlangs
Queue size:	10

Figures 14.16 through 14.17 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. 14.16 that a substantial reduction in the probability of new call blocking is achieved by using SDCA as compared to FCA. Handoff queueing causes a slight 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. Figure 14.17 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.

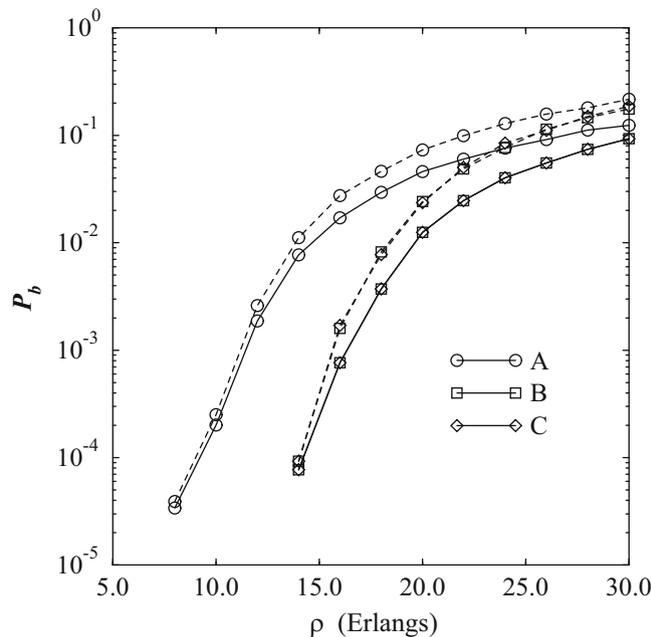


Fig. 14.16 Probability of new call blocking versus offered traffic without queueing (*solid*) and 5-second handoff queue (*dashed*). Legend: A = FCA, B = SDCA, C = Aggressive SDCA, from [349]

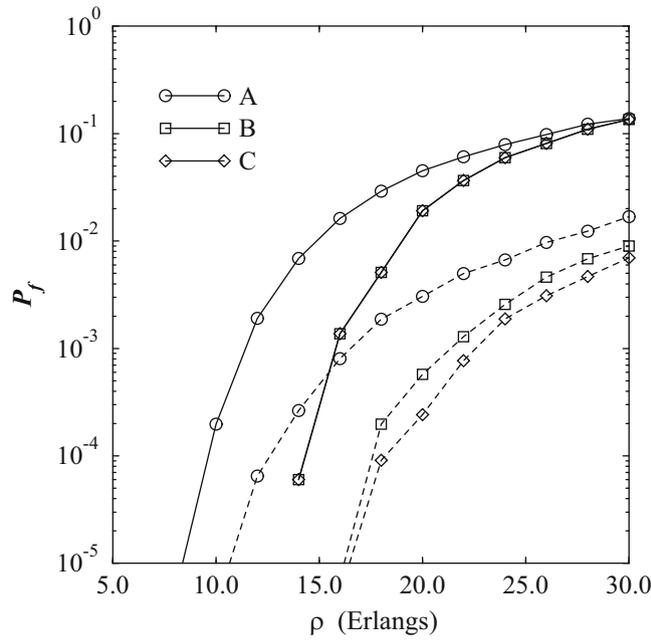


Fig. 14.17 Probability of forced termination versus offered traffic without queueing (*solid*) and a 5-second handoff queue (*dashed*). Legend: A = FCA, B = SDCA, C = Aggressive SDCA, from [349]

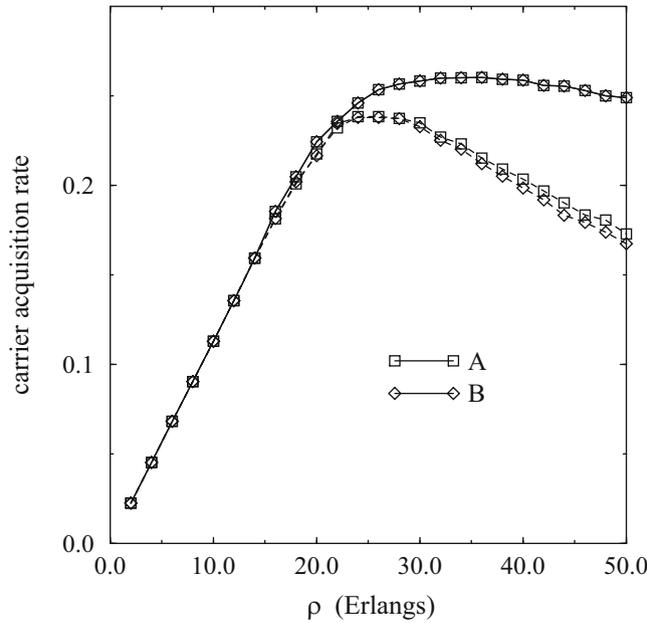


Fig. 14.18 Carrier acquisitions per second (per cell) versus offered traffic without queueing (*solid*) and a 5-second handoff queue (*dashed*). Legend: A = SDCA, B = Aggressive SDCA, from [349]

Carrier Acquisitions

It is useful to determine the increase in the rate of carrier acquisitions that results from using aggressive SDCA. Figure 14.18 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 handoff queueing has the largest effect on the carrier 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 [117] 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 present in microcells. Finally, the use of handoff queuing 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.

14.11 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. Moreover, channel assignment schemes have to support a variety of multimedia applications that have different quality of service requirements and require different types and amounts of network resources including channel resources, delay, etc.

Problems

14.1. Suppose that the maximum packing (MP) policy is used with the system shown in Fig. 14.2. Suppose that ten channels are available for use within each of the three reuse clusters $CL_i, i = 1, 2, 3$.

- a) Compute the number of admissible states $|\mathcal{S}|$.
- b) By using the approximation in (14.6) 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.

14.2. Show that the maximum and minimum saturation densities for a linear array of cells are $C_{\max} = (R + 1)^{-1}$ and $C_{\min} = (2R + 1)^{-1}$.

14.3. Show that the maximum and minimum saturation densities for a planar array of cells are

$$C_{\min} = \frac{1}{1 + 3R(R + 1)}$$

$$C_{\max} = \begin{cases} \frac{4}{3(R+1)^2} & , R \text{ even} \\ \frac{4}{1+3(R+1)^2} & , R \text{ odd} \end{cases} .$$

14.4. 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?

14.5. Derive Eq. (14.12).

Appendix A

Probability and Random Processes

The theory of probability and random processes is essential in the design and performance analysis of wireless communication systems. This appendix presents a brief review of the basic concepts of probability theory and random processes, with emphasis on the concept needed to understand this book. 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 [201, 254].

This appendix begins in Sect. A.1 with the basic axioms of probability, conditional probability, and Bayes' theorem. It then goes onto means, moments, and generating functions in Sect. A.2. Afterwards, Sect. A.3 presents a variety of discrete probability distributions and continuous probability density functions. Particular emphasis is placed on Gaussian, complex Gaussian, multi-variate Gaussian, multivariate complex Gaussian density functions, and functions of Gaussian random variables. After a brief treatment of upper bounds on probability in Sect. A.4, the appendix then goes onto a treatment of random processes, including means and correlation functions in Sect. A.5.1, cross-correlation and crosscovariance for joint random processes in Sect. A.5.2, complex random processes in Sect. A.5.3, power spectral density in Sect. A.5.4, and filtering of random processes in Sect. A.5.5. The important class of cyclostationary random processes is considered in Sect. A.5.6 and the appendix wraps up with a brief treatment of discrete-time random processes in Sect. A.5.7.

A.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

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \quad (\text{A.1})$$

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 $A \cap B = \emptyset$ and $P[A \cap B] = 0 \neq P[A]P[B]$. Therefore, mutually exclusive events with non-zero probability cannot be independent. Thus, the disjointness of events is a property of the events themselves, while independence is a property of their probabilities.

In general, the events $A_i, i = 1, \dots, n$, are independent if and only if for all collections of k distinct integers (i_1, i_2, \dots, i_k) chosen from the set $(1, 2, \dots, n)$,

$$P[A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}] = P[A_{i_1}]P[A_{i_2}] \cdots P[A_{i_k}]$$

for $2 \leq k \leq n$.

In summary

- If $A_i, i = 1, \dots, n$ is a sequence of mutually exclusive events, then

$$P\left[\bigcup_{i=1}^n A_i\right] = \sum_{i=1}^n P[A_i]. \quad (\text{A.2})$$

- If $A_i, i = 1, \dots, n$ is a sequence of independent events, then

$$P\left[\bigcap_{i=1}^n A_i\right] = \prod_{i=1}^n P[A_i]. \quad (\text{A.3})$$

A.1.1 Total Probability

The collection of sets $\{B_i\}, i = 1, \dots, n$ forms a *partition* of the sample space S if $B_i \cap B_j = \emptyset, i \neq j$ and $\bigcup_{i=1}^n B_i = S$. For any event $A \subset S$

$$A = \bigcup_{i=1}^n (A \cap B_i). \quad (\text{A.4})$$

That is, every element of A is contained in one and only one B_i . Since $(A \cap B_i) \cap (A \cap B_j) = \emptyset, i \neq j$, the sets $A \cap B_i$ are mutually exclusive. Therefore,

$$\begin{aligned} P[A] &= \sum_{i=1}^n P[A \cap B_i] \\ &= \sum_{i=1}^n P[A|B_i]P[B_i]. \end{aligned} \quad (\text{A.5})$$

This last equation is often referred to as the theorem of total probability.

A.1.2 Bayes' Theorem

Let the events $B_i, i = 1, \dots, n$ be mutually exclusive such that $\bigcup_{i=1}^n B_i = S$, where S is the sample space. Let A be an event with non-zero probability. Then as a result of conditional probability and total probability:

$$\begin{aligned} P[B_i|A] &= \frac{P[B_i \cap A]}{P[A]} \\ &= \frac{P[A|B_i]P[B_i]}{\sum_{i=1}^n P[A|B_i]P[B_i]}, \end{aligned}$$

a result known as Bayes' theorem.

A.2 Means, Moments, and Moment Generating Functions

The k th moment of a random variable, $E[X^k]$, is defined as

$$E[X^k] \triangleq \begin{cases} \sum_{x_i \in R_X} x_i^k p_X(x_i) & \text{if } X \text{ is discrete} \\ \int_{R_X} x^k p_X(x) dx & \text{if } X \text{ is continuous} \end{cases}, \quad (\text{A.6})$$

where $p_X(x_i) \triangleq P[X = x_i]$ is the probability distribution function of X , and $p_X(x)$ is the probability density function (pdf) of X . The k th central moment of the random variable X is $E[(X - E[X])^k]$. The mean is the first moment

$$\mu_X = E[X] \quad (\text{A.7})$$

and the variance is the second central moment

$$\sigma_X^2 = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2. \quad (\text{A.8})$$

The moment generating function or characteristic function of a random variable X is

$$\psi_X(jv) \triangleq E[e^{jvX}] = \begin{cases} \sum_{x_i \in R_X} e^{jvx_i} p_X(x_i) & \text{if } X \text{ is discrete} \\ \int_{R_X} e^{jvx} p_X(x) dx & \text{if } X \text{ is continuous} \end{cases}, \quad (\text{A.9})$$

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.,

$$p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_X(jv) e^{-jvx} dv \quad (\text{A.10})$$

and

$$p_X(x_k) = \frac{1}{2\pi} \oint_C \psi_X(jv) e^{-jvx_k} dv. \quad (\text{A.11})$$

The cumulative distribution function (cdf) of a random variable X is defined as

$$F_X(x) \triangleq P[X \leq x] = \begin{cases} \sum_{x_i \leq x} p_X(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^x p_X(x) dx & \text{if } X \text{ is continuous} \end{cases}, \quad (\text{A.12})$$

and $0 \leq F_X(x) \leq 1$. The complementary distribution function (cdfc) is defined as

$$F_X^c(x) \triangleq 1 - F_X(x). \quad (\text{A.13})$$

The probability density function of a continuous random variable X is related to the cdf by

$$p_X(x) = \frac{dF_X(x)}{dx}. \quad (\text{A.14})$$

A.2.1 Bivariate Random Variables

Consider a pair of random variables X and Y . The joint cdf of X and Y is

$$F_{XY}(x, y) = P[X \leq x, Y \leq y], \quad 0 \leq F_{XY}(x, y) \leq 1, \quad (\text{A.15})$$

and the joint (cdfc) of X and Y is

$$F_{XY}^c(x, y) = P[X > x, Y > y] = 1 - F_{XY}(x, y), \quad 0 \leq F_{XY}^c(x, y) \leq 1. \quad (\text{A.16})$$

The joint pdf of X and Y is

$$p_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}, \quad F_{XY}(x) = \int_{-\infty}^x \int_{-\infty}^y p_{XY}(x, y) dx dy. \quad (\text{A.17})$$

The marginal pdfs of X and Y are

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x, y) dy \quad p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x, y) dx. \quad (\text{A.18})$$

If X and Y are independent random variables, then the joint pdf has the product form

$$p_{XY}(x, y) = p_X(x)p_Y(y). \quad (\text{A.19})$$

The conditional pdfs of X and Y are

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)} \quad p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}. \quad (\text{A.20})$$

The joint moments of X and Y are

$$E[X^i Y^j] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^i y^j p_{XY}(x, y) dx dy. \quad (\text{A.21})$$

The covariance of X and Y is

$$\begin{aligned} \lambda_{XY} &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ &= E[XY] - \mu_X\mu_Y \end{aligned} \quad (\text{A.22})$$

The correlation coefficient of X and Y is

$$\rho_{XY} = \frac{\lambda_{XY}}{\sigma_X \sigma_Y}. \quad (\text{A.23})$$

Two random variables X and Y are *uncorrelated* if and only if $\lambda_{X,Y} = 0$. Two random variables X and Y are *orthogonal* if and only if $E[XY] = 0$.

The joint characteristic function is

$$\Phi_{XY}(v_1, v_2) = E[e^{jv_1 X + jv_2 Y}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{XY}(x, y) e^{jv_1 x + jv_2 y} dx dy. \quad (\text{A.24})$$

If X and Y are independent, then

$$\begin{aligned}\Phi_{XY}(v_1, v_2) &= E[e^{jv_1X+jv_2Y}] \\ &= \int_{-\infty}^{\infty} p_X(x)e^{jv_1x} dx \int_{-\infty}^{\infty} p_Y(y)e^{jv_2y} dy \\ &= \Phi_X(v_1)\Phi_Y(v_2).\end{aligned}\tag{A.25}$$

Moments can be generated according to

$$E[XY] = -\frac{\partial^2 \Phi_{XY}(v_1, v_2)}{\partial v_1 \partial v_2} \Big|_{v_1=v_2=0}.\tag{A.26}$$

with higher order moments generated in a straightforward extension.

A.3 Some Useful Probability Distributions

A.3.1 Discrete Distributions

A.3.1.1 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 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

$$p_X(x) = (1 - p)\delta(x) + p\delta(x - 1).\tag{A.27}$$

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 $\{0, 1, \dots, n\}$ and the probability distribution of Y is the binomial distribution

$$p_Y(k) \equiv P[Y = k] = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.\tag{A.28}$$

The random variable Y also has the pdf

$$p_Y(y) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta(y - k).\tag{A.29}$$

A.3.1.2 Poisson Distribution

The random variable X has a Poisson distribution if

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \dots, \infty\tag{A.30}$$

A.3.1.3 Geometric Distribution

The random variable X has a geometric distribution if

$$p_X(k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots, \infty.\tag{A.31}$$

A.3.2 Continuous Distributions

Many communication systems are affected by Gaussian random processes. Therefore, Gaussian random variables and various functions of Gaussian random variables play a central role in the characterization and analysis of communication systems.

A.3.2.1 Gaussian Distribution

A real-valued Gaussian or normal random variable X has the pdf

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad (\text{A.32})$$

where $\mu = E[X]$ is the mean of X and $\sigma^2 = E[(X-\mu)^2]$ is the variance of X . Sometimes the shorthand notation $X \sim \mathcal{N}(\mu, \sigma^2)$ is used 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 $X \sim \mathcal{N}(0, 1)$.

The cumulative distribution function (cdf) of a Gaussian random variable X is

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} dy. \quad (\text{A.33})$$

The cdf of a standard normal distribution defines the Gaussian Q function

$$Q(x) \triangleq \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \quad (\text{A.34})$$

and the cdfc defines the Gaussian Φ function

$$\Phi(x) \triangleq 1 - Q(x). \quad (\text{A.35})$$

If X is a non-standard normal random variable, $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad (\text{A.36})$$

$$F_X^c(x) = Q\left(\frac{x-\mu}{\sigma}\right). \quad (\text{A.37})$$

Sometimes the cumulative distribution function of a Gaussian random variable is described in terms of the complementary error function $\text{erfc}(x)$, defined as

$$\text{erfc}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-y^2} dy. \quad (\text{A.38})$$

The complementary error function and the Gaussian Q function are related as follows:

$$\text{erfc}(x) = 2Q(\sqrt{2}x) \quad (\text{A.39})$$

$$Q(x) = \frac{1}{2} \text{erfc}\left(\frac{x}{\sqrt{2}}\right). \quad (\text{A.40})$$

These identities can be established by using the Gaussian Q function in (A.34). The error function of a Gaussian random variable is defined as

$$\text{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy. \quad (\text{A.41})$$

Note that $\operatorname{erfc}(x) + \operatorname{erf}(x) \neq 1$. Also,

$$Q(x) = \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right), \quad x \geq 0. \quad (\text{A.42})$$

A.3.2.2 Multivariate Gaussian Distribution

Let $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$, be a collection of n real-valued Gaussian random variables having means $\mu_i = \mathbb{E}[X_i]$ and covariances

$$\begin{aligned} \lambda_{X_i X_j} &= \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \mathbb{E}[X_i X_j] - \mu_i \mu_j, \quad 1 \leq i, j \leq n. \end{aligned}$$

Let

$$\begin{aligned} \mathbf{X} &= (X_1, X_2, \dots, X_n)^T \\ \mathbf{x} &= (x_1, x_2, \dots, x_n)^T \\ \boldsymbol{\mu}_X &= (\mu_1, \mu_2, \dots, \mu_n)^T \\ \boldsymbol{\Lambda} &= \begin{bmatrix} \lambda_{X_1 X_1} & \cdots & \lambda_{X_1 X_n} \\ \vdots & & \vdots \\ \lambda_{X_n X_1} & \cdots & \lambda_{X_n X_n} \end{bmatrix}, \end{aligned}$$

where \mathbf{X}^T is the transpose of \mathbf{X} . The random vector \mathbf{X} has the multivariate Gaussian distribution

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Lambda}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)^T \boldsymbol{\Lambda}^{-1} (\mathbf{x} - \boldsymbol{\mu}_X)\right\}, \quad (\text{A.43})$$

where $|\boldsymbol{\Lambda}|$ is the determinant of $\boldsymbol{\Lambda}$.

A.3.2.3 Multivariate Complex Gaussian Distribution

Complex Gaussian distributions often arise in the treatment of fading channels and narrow-band Gaussian noise. Let

$$\begin{aligned} \mathbf{X} &= (X_1, X_2, \dots, X_n)^T \\ \mathbf{Y} &= (Y_1, Y_2, \dots, Y_n)^T \end{aligned}$$

be length- n vectors of real-valued Gaussian random variables, such that $X_i \sim \mathcal{N}(\mu_{X_i}, \sigma_{X_i}^2)$, and $Y_i \sim \mathcal{N}(\mu_{Y_i}, \sigma_{Y_i}^2)$, $i = 1, \dots, n$. The complex random vector $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$ has a complex Gaussian distribution that can be described with the following three parameters:

$$\begin{aligned} \boldsymbol{\mu}_Z &= \mathbb{E}[\mathbf{Z}] = \boldsymbol{\mu}_X + j\boldsymbol{\mu}_Y \\ \boldsymbol{\Gamma} &= \frac{1}{2} \mathbb{E}[(\mathbf{Z} - \boldsymbol{\mu}_Z)(\mathbf{Z} - \boldsymbol{\mu}_Z)^H] \\ \mathbf{C} &= \frac{1}{2} \mathbb{E}[(\mathbf{Z} - \boldsymbol{\mu}_Z)(\mathbf{Z} - \boldsymbol{\mu}_Z)^T], \end{aligned}$$

where \mathbf{X}^T and \mathbf{X}^H are the transpose and complex conjugate transpose of \mathbf{X} , respectively. The covariance matrix $\boldsymbol{\Gamma}$ must be Hermitian ($\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^H$) and the relation matrix \mathbf{C} should be symmetric ($\mathbf{C} = \mathbf{C}^T$). Matrices $\boldsymbol{\Gamma}$ and \mathbf{C} can be related to the covariance matrices of \mathbf{X} and \mathbf{Y} as follows:

$$\mathbf{\Lambda}_{\mathbf{X}\mathbf{X}} = E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^T] = \text{Re}\{\boldsymbol{\Gamma} + \mathbf{C}\} \quad (\text{A.44})$$

$$\mathbf{\Lambda}_{\mathbf{X}\mathbf{Y}} = E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{Y} - \boldsymbol{\mu}_Y)^T] = \text{Im}\{-\boldsymbol{\Gamma} + \mathbf{C}\} \quad (\text{A.45})$$

$$\mathbf{\Lambda}_{\mathbf{Y}\mathbf{X}} = E[(\mathbf{Y} - \boldsymbol{\mu}_Y)(\mathbf{X} - \boldsymbol{\mu}_X)^T] = \text{Im}\{\boldsymbol{\Gamma} + \mathbf{C}\} \quad (\text{A.46})$$

$$\mathbf{\Lambda}_{\mathbf{Y}\mathbf{Y}} = E[(\mathbf{Y} - \boldsymbol{\mu}_Y)(\mathbf{Y} - \boldsymbol{\mu}_Y)^T] = \text{Re}\{\boldsymbol{\Gamma} - \mathbf{C}\} \quad (\text{A.47})$$

and, conversely,

$$\begin{aligned} \boldsymbol{\Gamma} &= \frac{1}{2}(\mathbf{\Lambda}_{\mathbf{X}\mathbf{X}} + \mathbf{\Lambda}_{\mathbf{Y}\mathbf{Y}} + j(\mathbf{\Lambda}_{\mathbf{Y}\mathbf{X}} - \mathbf{\Lambda}_{\mathbf{X}\mathbf{Y}})) \\ \mathbf{C} &= \frac{1}{2}(\mathbf{\Lambda}_{\mathbf{X}\mathbf{X}} - \mathbf{\Lambda}_{\mathbf{Y}\mathbf{Y}} + j(\mathbf{\Lambda}_{\mathbf{Y}\mathbf{X}} + \mathbf{\Lambda}_{\mathbf{X}\mathbf{Y}})). \end{aligned} \quad (\text{A.48})$$

The complex random vector \mathbf{Z} has the complex multivariate Gaussian distribution

$$p_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^n \sqrt{\det(\boldsymbol{\Gamma})\det(\mathbf{P})}} \exp \left\{ -\frac{1}{4} ((\mathbf{z} - \boldsymbol{\mu}_Z)^H, (\mathbf{z} - \boldsymbol{\mu}_Z)^T) \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{C} \\ \mathbf{C}^H & \boldsymbol{\Gamma}^* \end{pmatrix}^{-1} \begin{pmatrix} (\mathbf{z} - \boldsymbol{\mu}_Z) \\ (\mathbf{z}^* - \boldsymbol{\mu}_Z^*) \end{pmatrix} \right\}, \quad (\text{A.49})$$

where

$$\mathbf{P} = \boldsymbol{\Gamma}^* - \mathbf{C}^H \boldsymbol{\Gamma}^{-1} \mathbf{C}. \quad (\text{A.50})$$

For a circular-symmetric complex Gaussian distribution $\mathbf{C} = \mathbf{0}$, and the complex multivariate Gaussian distribution simplifies considerably as

$$p_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^n \det(\boldsymbol{\Gamma})} \exp \left\{ -\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu}_Z)^H \boldsymbol{\Gamma}^{-1} (\mathbf{z} - \boldsymbol{\mu}_Z) \right\}. \quad (\text{A.51})$$

The circular-symmetric scalar complex Gaussian random variable $Z = X + jY$ has the density

$$p_Z(z) = \frac{1}{2\pi\sigma_Z^2} \exp \left\{ -\frac{|z - \mu_Z|^2}{2\sigma_Z^2} \right\}, \quad (\text{A.52})$$

where $\mu_Z = E[Z]$ and $\sigma_Z^2 = \frac{1}{2}E[|z - \mu_Z|^2]$. Sometimes this is denoted with the shorthand notation $Z_i \sim \mathcal{CN}(\mu_Z, \sigma_Z^2)$. The standard complex Gaussian distribution $Z_i \sim \mathcal{CN}(0, 1)$ has the density

$$p_Z(z) = \frac{1}{2\pi} \exp \left\{ -\frac{|z|^2}{2} \right\}. \quad (\text{A.53})$$

A.3.2.4 Rayleigh Distribution

Let $X \sim \mathcal{N}(0, \sigma^2)$ and $Y \sim \mathcal{N}(0, \sigma^2)$ be independent real-valued normal random variables. The random variable $R = \sqrt{X^2 + Y^2}$ is said to be Rayleigh distributed. To find the pdf and cdf of R first define the auxiliary variable

$$V = \text{Tan}^{-1}(Y/X).$$

Then

$$X = R \cos V$$

$$Y = R \sin V.$$

By using a bivariate transformation of random variables

$$p_{RV}(r, v) = p_{XY}(r \cos v, r \sin v) |J(r, v)|$$

where

$$J(r, v) = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \cos v & r \sin v \\ \sin v & r \cos v \end{vmatrix} = r(\cos^2 v + \sin^2 v) = r$$

Since

$$p_{XY}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2 + y^2}{2\sigma^2}\right\}$$

it follows that

$$p_{RV}(r, v) = \frac{r}{2\pi\sigma^2} \exp\left\{-\frac{r^2}{2\sigma^2}\right\}. \quad (\text{A.54})$$

The marginal pdf of R has the Rayleigh distribution

$$\begin{aligned} p_R(r) &= \int_0^{2\pi} p_{RV}(r, v) dv \\ &= \frac{r}{\sigma^2} \exp\left\{-\frac{r^2}{2\sigma^2}\right\}, \quad r \geq 0. \end{aligned} \quad (\text{A.55})$$

The cdf of R is

$$F_R(r) = 1 - \exp\left\{-\frac{r^2}{2\sigma^2}\right\}, \quad r \geq 0. \quad (\text{A.56})$$

The marginal pdf of V is

$$\begin{aligned} p_V(v) &= \int_0^\infty p_{RV}(r, v) dr \\ &= \frac{1}{2\pi}, \quad \pi \leq v \leq \pi. \end{aligned} \quad (\text{A.57})$$

which is a uniform distribution on the interval $[-\pi, \pi)$.

A.3.2.5 Rice Distribution

Let $X \sim \mathcal{N}(\mu_1, \sigma^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma^2)$ be independent normal random variables with non-zero means. The random variable $R = \sqrt{X^2 + Y^2}$ has a Rice distribution or is said to be Ricean distributed. To find the pdf and cdf of R again define the auxiliary variable $V = \tan^{-1}(Y/X)$. Then by using a bivariate transformation $J(r, v) = r$ and

$$p_{RV}(r, v) = r \cdot p_{XY}(r \cos v, r \sin v). \quad (\text{A.58})$$

However,

$$\begin{aligned} p_{XY}(x, y) &= \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{(x - \mu_1)^2 + (y - \mu_2)^2}{2\sigma^2}\right\} \\ &= \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2 + y^2 + \mu_1^2 + \mu_2^2 - 2(x\mu_1 + y\mu_2)}{2\sigma^2}\right\}. \end{aligned}$$

Hence,

$$p_{RV}(r, v) = \frac{r}{2\pi\sigma^2} \exp \left\{ -\frac{r^2 + \mu_1^2 + \mu_2^2 - 2r(\mu_1 \cos v + \mu_2 \sin v)}{2\sigma^2} \right\}.$$

Now define $s \triangleq \sqrt{\mu_1^2 + \mu_2^2}$ and $t \triangleq \tan^{-1} \mu_2/\mu_1$, $-\pi \leq t \leq \pi$, so that $\mu_1 = s \cos t$ and $\mu_2 = s \sin t$. Then

$$\begin{aligned} p_{RV}(r, v) &= \frac{r}{2\pi\sigma^2} \exp \left\{ -\frac{r^2 + s^2 - 2rs(\cos t \cos v + \sin t \sin v)}{2\sigma^2} \right\} \\ &= \frac{r}{2\pi\sigma^2} \exp \left\{ -\frac{r^2 + s^2 - 2rs \cos(v-t)}{2\sigma^2} \right\}. \end{aligned}$$

The marginal pdf of R is

$$P_R(r) = \frac{r}{\sigma^2} \exp \left\{ -\frac{r^2 + s^2}{2\sigma^2} \right\} \frac{1}{2\pi} \int_0^{2\pi} \exp \left\{ \frac{rs}{\sigma^2} \cos(v-t) \right\} dv. \quad (\text{A.59})$$

The zero order modified Bessel function of the first kind is defined as

$$I_0(x) \triangleq \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos \theta} d\theta. \quad (\text{A.60})$$

This gives the Rice distribution

$$P_R(r) = \frac{r}{\sigma^2} \exp \left\{ -\frac{r^2 + s^2}{2\sigma^2} \right\} I_0 \left(\frac{rs}{\sigma^2} \right), \quad r \geq 0. \quad (\text{A.61})$$

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.

A.3.2.6 Central Chi-Square Distribution

Let $X \sim \mathcal{N}(0, \sigma^2)$ and $Y = X^2$. Then it can be shown that

$$\begin{aligned} p_Y(y) &= \frac{p_X(\sqrt{y}) + p_X(-\sqrt{y})}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi y \sigma^2}} \exp \left\{ -\frac{y}{2\sigma^2} \right\}, \quad y \geq 0. \end{aligned}$$

The characteristic function of Y is

$$\begin{aligned} \psi_Y(jv) &= \int_{-\infty}^{\infty} e^{jvy} p_Y(y) dy \\ &= \frac{1}{\sqrt{1 - j2v\sigma^2}}. \end{aligned} \quad (\text{A.62})$$

Now define the random variable $Y = \sum_{i=1}^n X_i^2$, where the X_i are independent and $X_i \sim \mathcal{N}(0, \sigma^2)$. Then

$$\psi_Y(jv) = \frac{1}{(1 - j2v\sigma^2)^{n/2}}. \quad (\text{A.63})$$

Taking the inverse transform gives

$$\begin{aligned} p_Y(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_Y(jv) e^{-jvy} dv \\ &= \frac{1}{(2\sigma^2)^{n/2} \Gamma(n/2)} y^{n/2-1} \exp\left\{-\frac{y}{2\sigma^2}\right\}, \quad y \geq 0. \end{aligned}$$

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 $m = n/2$, then the pdf of Y defines the central chi-square distribution with $2m$ degrees of freedom

$$p_Y(y) = \frac{1}{(2\sigma^2)^m (m-1)!} y^{m-1} \exp\left\{-\frac{y}{2\sigma^2}\right\}, \quad y \geq 0. \quad (\text{A.64})$$

The cdf of Y is

$$F_Y(y) = 1 - \exp\left\{-\frac{y}{2\sigma^2}\right\} \sum_{k=0}^{m-1} \frac{1}{k!} \left(\frac{y}{2\sigma^2}\right)^k, \quad y \geq 0. \quad (\text{A.65})$$

The exponential distribution is a special case of the central chi-square distribution with $m = 1$ (2 degrees of freedom). In this case

$$\begin{aligned} p_Y(y) &= \frac{1}{2\sigma^2} \exp\left\{-\frac{y}{2\sigma^2}\right\}, \quad y \geq 0 \\ F_Y(y) &= 1 - \exp\left\{-\frac{y}{2\sigma^2}\right\}, \quad y \geq 0. \end{aligned} \quad (\text{A.66})$$

A.3.2.7 Non-central Chi-Square Distribution

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = X^2$. Then

$$\begin{aligned} p_Y(y) &= \frac{p_X(\sqrt{y}) + p_X(-\sqrt{y})}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi}y\sigma} \exp\left\{-\frac{(y + \mu^2)}{2\sigma^2}\right\} \cosh\left(\frac{\sqrt{y}\mu}{\sigma^2}\right), \quad y \geq 0. \end{aligned}$$

The characteristic function of Y is

$$\begin{aligned} \psi_Y(jv) &= \int_{-\infty}^{\infty} e^{jvy} p_Y(y) dy \\ &= \frac{1}{\sqrt{1 - j2v\sigma^2}} \exp\left\{\frac{jv\mu^2}{1 - j2v\sigma^2}\right\}. \end{aligned}$$

Now define the random variable $Y = \sum_{i=1}^n X_i^2$, where the X_i are independent normal random variables and $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$. 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

$$p_Y(y) = \frac{1}{2\sigma^2} \left(\frac{y}{s^2} \right)^{\frac{n-2}{4}} \exp \left\{ -\frac{(s^2 + y)}{2\sigma^2} \right\} I_{n/2-1} \left(\sqrt{y} \frac{s}{\sigma^2} \right), \quad y \geq 0,$$

where

$$s^2 = \sum_{i=1}^n \mu_i^2$$

and $I_k(x)$ is the modified Bessel function of the first kind and order k , defined by

$$I_k(x) \triangleq \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos \theta} \cos(k\theta) d\theta.$$

If n is even (which is usually the case in practice) and $m = n/2$, then the pdf of Y defines the non-central chi-square distribution with $2m$ degrees of freedom

$$p_Y(y) = \frac{1}{2\sigma^2} \left(\frac{y}{s^2} \right)^{\frac{m-1}{2}} \exp \left\{ -\frac{(s^2 + y)}{2\sigma^2} \right\} I_{m-1} \left(\sqrt{y} \frac{s}{\sigma^2} \right), \quad y \geq 0 \quad (\text{A.67})$$

and the cdf of Y is

$$F_Y(y) = 1 - Q_m \left(\frac{s}{\sigma}, \frac{\sqrt{y}}{\sigma} \right), \quad y \geq 0, \quad (\text{A.68})$$

where $Q_m(a, b)$ is called the generalized Q -function.

A.4 Upper Bounds on the cdfc

Several different approaches can be used to upper bound the tail area under a probability density function including the Chebyshev and Chernoff bounds.

A.4.1 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

$$\begin{aligned} \sigma_X^2 &= \int_{-\infty}^{\infty} (x - \mu_X)^2 p_X(x) dx \\ &\geq \int_{|x - \mu_X| \geq \delta} (x - \mu_X)^2 p_X(x) dx \\ &\geq \delta^2 \int_{|x - \mu_X| \geq \delta} p_X(x) dx \\ &= \delta^2 \mathbf{P}[|X - \mu_X| \geq \delta]. \end{aligned}$$

Hence,

$$P[|X - \mu_X| \geq \delta] \leq \frac{\sigma_X^2}{\delta^2}. \quad (\text{A.69})$$

The Chebyshev bound is straightforward to apply but it tends to be quite loose.

A.4.2 Chernoff Bound

The Chernoff bound is more difficult to compute but is much tighter than the Chebyshev bound. To derive the Chernoff bound the following inequality is used

$$u(x) \leq e^{\lambda x}, \quad \forall x \text{ and } \forall \lambda \geq 0,$$

where $u(x)$ is the unit step function. Then,

$$\begin{aligned} P[X \geq 0] &= \int_0^{\infty} p_X(x) dx \\ &= \int_{-\infty}^{\infty} u(x) p_X(x) dx \\ &\leq \int_{-\infty}^{\infty} e^{\lambda x} p_X(x) dx \\ &= E[e^{\lambda X}]. \end{aligned}$$

The Chernoff bound parameter, λ , $\lambda > 0$, can be optimized to give the tightest upper bound. This can be accomplished by setting the derivative to zero

$$\frac{d}{d\lambda} E[e^{\lambda X}] = E\left[\frac{d}{d\lambda} e^{\lambda X}\right] = E[X e^{\lambda X}] = 0.$$

Let $\lambda^* = \arg \min_{\lambda \geq 0} E[e^{\lambda X}]$ be the solution to the above equation. Then

$$P[X \geq 0] \leq E[e^{\lambda^* X}]. \quad (\text{A.70})$$

Example A.1: Let X_i , $i = 1, \dots, n$ be independent and identically distributed random variables with density

$$p_X(x) = p\delta(x - 1) + (1 - p)\delta(x + 1).$$

Let

$$Y = \sum_{i=1}^n X_i.$$

Consider the quantity $P[Y \geq 0]$. To compute this probability exactly,

$$\begin{aligned} P[Y \geq 0] &= P[\lceil n/2 \rceil \text{ or more of the } X_i \text{ are ones}] \\ &= \sum_{k=\lceil n/2 \rceil}^n \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

(continued)

Example A.1 (continued)

For $n = 10$ and $p = 0.1$

$$P[Y \geq 0] = 0.0016349. \quad (\text{A.71})$$

Chebyshev Bound

To compute the Chebyshev bound, first determine the mean and variance of Y .

$$\begin{aligned} \mu_Y &= nE[X_i] \\ &= n[p - 1 + p] \\ &= n(2p - 1). \end{aligned}$$

$$\begin{aligned} \sigma_Y^2 &= n\sigma_X^2 \\ &= n(E[X_i^2] - E^2[X_i]) \\ &= n(1 - (2p - 1)^2) \\ &= n(1 - 4p^2 + 4p - 1) \\ &= 4np(1 - p). \end{aligned}$$

Hence,

$$P[|Y - \mu_Y| \geq \mu_Y] \leq \frac{\sigma_Y^2}{\mu_Y^2} = \frac{4np(1 - p)}{n^2(2p - 1)^2}.$$

Then by symmetry

$$\begin{aligned} P[Y \geq 0] &= \frac{1}{2}P[|Y - \mu_Y| \geq \mu_Y] \\ &\leq \frac{2p(1 - p)}{n(2p - 1)^2}. \end{aligned}$$

For $n = 10$ and $p = 0.1$

$$P[Y \geq 0] \leq 0.028125. \quad (\text{A.72})$$

Chernoff Bound

The Chernoff bound is given by

$$\begin{aligned} P[Y \geq 0] &\leq E[e^{\lambda Y}] \\ &= (E[e^{\lambda X_i}])^n. \end{aligned}$$

However,

$$E[e^{\lambda X_i}] = pe^\lambda + (1 - p)e^{-\lambda}.$$

(continued)

Example A.1 (continued)

To find the optimal Chernoff bound parameter, solve

$$\frac{d}{d\lambda} E[e^{\lambda x}] = pe^{\lambda} - (1-p)e^{-\lambda} = 0$$

giving

$$\lambda^* = \ln \left(\sqrt{\frac{1-p}{p}} \right).$$

Hence,

$$\begin{aligned} P[Y \geq 0] &\leq \left(E[e^{\lambda^* x}] \right)^n \\ &= (4p(1-p))^{n/2}. \end{aligned}$$

For $n = 10$ and $p = 0.1$

$$P[Y \geq 0] \leq 0.0060466.$$

Notice that the Chernoff bound is much tighter than the Chebyshev bound in this case.

A.5 Random Processes

A random process, or stochastic process, $X(t)$, is an ensemble of sample functions $\{X_1(t), X_2(t), \dots, X_\xi(t)\}$ together with a probability rule which assigns a probability to any meaningful event associated with the observation of these sample functions. Consider the set of sample functions shown in Fig. A.1. The sample function x_i corresponds to the sample point s_i in the sample space and occurs with probability $P[s_i]$. 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

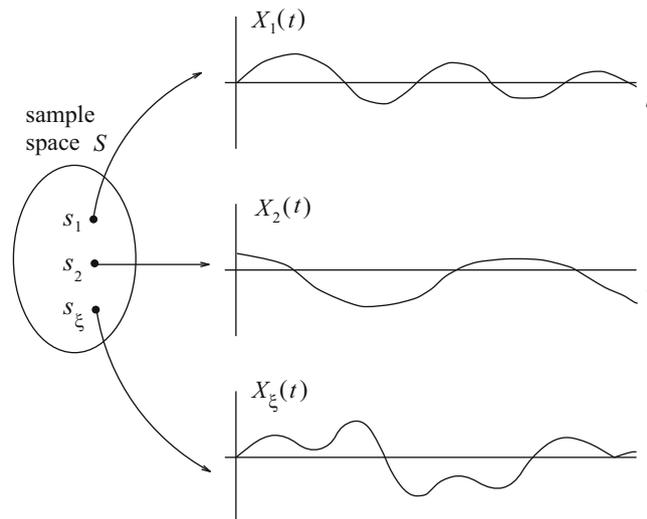


Fig. A.1 Ensemble of sample functions for a random process

continuous instants in time, which define discrete- or continuous-time random processes, respectively. Furthermore, their values (or parameters) at these time instants may be either discrete or continuous valued as well, which defines a discrete- or continuous-parameter random process, respectively. Hence, random processes may be discrete-time discrete-parameter, discrete-time continuous-parameter, continuous-time discrete-parameter, or continuous-time continuous-parameter.

Suppose that all sample functions of a random process are observed at some time instant t_1 . Their values form the set of numbers $\{X_i(t_1)\}$, $i = 1, 2, \dots, \xi$. Since $X_i(t_1)$ occurs with probability $P[s_i]$, the collection of numbers $\{X_i(t_1)\}$, $i = 1, 2, \dots, \xi$, forms a random variable, denoted by $X(t_1)$. By observing the set of waveforms at another time instant t_2 a different random variable $X(t_2)$ is obtained. The collection of n such random variables, $X(t_1), X(t_2), \dots, X(t_n)$, has the joint cdf

$$F_{X(t_1), \dots, X(t_n)}(x_1, \dots, x_n) = P[X(t_1) < x_1, \dots, X(t_n) < x_n].$$

A more compact notation can be obtained by defining the vectors

$$\mathbf{x} \triangleq (x_1, x_2, \dots, x_n)^T$$

$$\mathbf{X}(t) \triangleq (X(t_1), X(t_2), \dots, X(t_n))^T.$$

Then the joint cdf and joint pdf of the random vector $\mathbf{X}(t)$ are, respectively,

$$F_{\mathbf{X}(t)}(\mathbf{x}) = P(\mathbf{X}(t) \leq \mathbf{x}) \quad (\text{A.73})$$

$$p_{\mathbf{X}(t)}(\mathbf{x}) = \frac{\partial^n F_{\mathbf{X}(t)}(\mathbf{x})}{\partial x_1 \partial x_2 \cdots \partial x_n}. \quad (\text{A.74})$$

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

$$p_{\mathbf{X}(t)}(\mathbf{x}) = p_{\mathbf{X}(t+\tau)}(\mathbf{x}) \quad (\text{A.75})$$

holds for all sets of time instants $\{t_1, t_2, \dots, t_n\}$ and all time shifts τ . Some important random processes that are encountered in practice are strictly stationary, while many are not.

A.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:

$$E[\cdot] \triangleq \text{ensemble average}$$

$$\langle \cdot \rangle \triangleq \text{time average.}$$

The ensemble average of a random process at time t is

$$\mu_X(t) = E[X(t)] = \int_{-\infty}^{\infty} xp_{X(t)}(x)dx. \quad (\text{A.76})$$

Note that the ensemble average is generally a function of time. However, if the ensemble average changes with time, then the process is not strictly stationary. The time average of a random process is

$$\langle X(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T X(t)dt. \quad (\text{A.77})$$

In general, the time average $\langle X(t) \rangle$ 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

$$\phi_{XX}(t_1, t_2) = E[X(t_1)X(t_2)]. \quad (\text{A.78})$$

The autocovariance of a random process $X(t)$ is defined as

$$\begin{aligned} \lambda_{XX}(t_1, t_2) &= E[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))] \\ &= \phi_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2). \end{aligned} \quad (\text{A.79})$$

A random process that is strictly stationary must have

$$E[X^n(t)] = E[X^n] \quad \forall t, n.$$

Hence, strictly stationary random process must have

$$\begin{aligned} \mu_X(t) &= \mu \\ \sigma_X^2(t) &= \sigma_X^2 \\ \phi_{XX}(t_1, t_2) &= \phi_{XX}(t_2 - t_1) \equiv \phi_{XX}(\tau) \\ \lambda_{XX}(t_1, t_2) &= \lambda_{XX}(t_2 - t_1) \equiv \lambda_{XX}(\tau), \end{aligned}$$

where $\tau = t_2 - t_1$.

If a random process satisfies the following two conditions

$$\begin{aligned} \mu_X(t) &= \mu_X \\ \phi_{XX}(t_1, t_2) &= \phi_{XX}(\tau), \quad \tau = t_2 - t_1, \end{aligned}$$

then it is said to be wide-sense stationary. Note that if a random process is strictly stationary, then it is wide-sense stationary; however, the converse may not be true. A notable exception is the Gaussian random process which is strictly stationary if and only 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 $X(t_i)$.

A.5.1.1 Properties of $\phi_{XX}(\tau)$

The autocorrelation function, $\phi_{XX}(\tau)$, of a stationary random process satisfies the following properties.

- $\phi_{XX}(0) = E[X^2(t)]$. This is the total power in the random process.
- $\phi_{XX}(\tau) = \phi_{XX}(-\tau)$. The autocorrelation function must be an even function.
- $|\phi_{XX}(\tau)| \leq \phi_{XX}(0)$. This is a variant of the Cauchy-Schwartz inequality.
- $\phi_{XX}(\infty) = E^2[X(t)] = \mu_X^2$. This holds if $X(t)$ contains no periodic components and is equal to the d.c. power.

Example A.2: This example shows that $|\phi_{XX}(\tau)| \leq \phi_{XX}(0)$. This inequality can be established through the following steps.

$$\begin{aligned} 0 &\leq E[X(t) \pm (X(t + \tau))^2] \\ &= E[X^2(t) + X^2(t + \tau) \pm 2X(t)X(t + \tau)] \\ &= E[X^2(t)] + E[X^2(t + \tau)] \pm 2E[X(t)X(t + \tau)] \\ &= 2E[X^2(t)] \pm 2E[X(t)X(t + \tau)] \\ &= 2\phi_{XX}(0) \pm 2\phi_{XX}(\tau). \end{aligned}$$

(continued)

Example A.2 (continued)

Therefore,

$$\begin{aligned}\pm\phi_{XX}(\tau) &\leq \phi_{XX}(0) \\ |\phi_{XX}(\tau)| &\leq \phi_{XX}(0).\end{aligned}$$

A.5.1.2 Ergodic Random Processes

A random process is ergodic if for all $g(\mathbf{X})$ and \mathbf{X}

$$\begin{aligned}E[g(\mathbf{X})] &= \int_{-\infty}^{\infty} g(\mathbf{X})p_{\mathbf{X}(t)}(\mathbf{x})d\mathbf{x} \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T g[\mathbf{X}(t)]dt \\ &= \langle g[\mathbf{X}(t)] \rangle.\end{aligned}\tag{A.80}$$

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 $\langle X(t) \rangle = \mu_X$ and ergodic in the autocorrelation if $\langle X(t)X(t + \tau) \rangle = \phi_{XX}(\tau)$.

Example A.3: Consider the random process

$$X(t) = A \cos(2\pi f_c t + \Theta)$$

where A and f_c are constants, and

$$p_{\Theta}(\theta) = \begin{cases} 1/(2\pi), & 0 \leq \theta \leq 2\pi \\ 0, & \text{elsewhere} \end{cases}.$$

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

$$\begin{aligned}\phi_{XX}(t_1, t_2) &= E_{\Theta}[X(t_1)X(t_2)] \\ &= E_{\Theta}[A^2 \cos(2\pi f_c t_1 + \theta) \cos(2\pi f_c t_2 + \theta)] \\ &= \frac{A^2}{2} E_{\Theta}[\cos(2\pi f_c t_1 + 2\pi f_c t_2 + 2\theta)] + \frac{A^2}{2} E_{\Theta}[\cos(2\pi f_c(t_2 - t_1))] \\ &= \frac{A^2}{2} \cos(2\pi f_c(t_2 - t_1)) \\ &= \frac{A^2}{2} \cos(2\pi f_c \tau), \quad \tau = t_2 - t_1.\end{aligned}$$

It is clear that this random process is wide-sense stationary.

(continued)

Example A.3 (continued)

The time average mean of $X(t)$ is

$$\langle X(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T A \cos(2\pi f_c t + \theta) dt = 0$$

and the time average autocorrelation of $X(t)$ is

$$\begin{aligned} \langle X(t + \tau)X(t) \rangle &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T A^2 \cos(2\pi f_c t + \theta) \cos(2\pi f_c t + 2\pi f_c \tau + \theta) dt \\ &= \lim_{T \rightarrow \infty} \frac{A^2}{4T} \int_{-T}^T A^2 [\cos(2\pi f_c \tau) + \cos(4\pi f_c t + 2\pi f_c \tau + 2\theta)] dt \\ &= \frac{A^2}{2} \cos(2\pi f_c \tau). \end{aligned}$$

By comparing the ensemble and time average mean and autocorrelation, it can be concluded that this random process is ergodic in the mean and ergodic in the autocorrelation.

Example A.4: Consider the random process

$$Y(t) = X \cos t, \quad X \sim \mathcal{N}(0, 1).$$

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 strictly stationary.

1. To find the probability density function of $Y(0)$,

$$Y(0) = X \cos 0 = X.$$

Therefore,

$$p_{Y(0)}(y_0) = \frac{1}{\sqrt{2\pi}} e^{-y_0^2/2}.$$

2. To find the joint density of $Y(0)$ and $Y(\pi)$,

$$Y(0) = X = -Y(\pi).$$

Therefore,

$$p_{Y(0)|Y(\pi)}(y_0|y_\pi) = \delta(y_0 + y_\pi)$$

and

$$\begin{aligned} p_{Y(0)Y(\pi)}(y_0, y_\pi) &= p_{Y(0)|Y(\pi)}(y_0|y_\pi) p_{Y(\pi)}(y_\pi) \\ &= \frac{1}{\sqrt{2\pi}} e^{-y_\pi^2/2} \delta(y_0 + y_\pi). \end{aligned}$$

(continued)

Example A.4 (continued)

3. To determine whether or not $Y(t)$ is strictly 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 and, hence, the pdf of this random process varies with time, the random process is not strictly stationary.

A.5.2 Cross-correlation and Crosscovariance

Consider two random processes $X(t)$ and $Y(t)$. The cross-correlation of $X(t)$ and $Y(t)$ is

$$\phi_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] \quad (\text{A.81})$$

$$\phi_{YX}(t_1, t_2) = E[Y(t_1)X(t_2)]. \quad (\text{A.82})$$

The correlation matrix of $X(t)$ and $Y(t)$ is

$$\Phi(t_1, t_2) = \begin{bmatrix} \phi_{XX}(t_1, t_2) & \phi_{XY}(t_1, t_2) \\ \phi_{YX}(t_1, t_2) & \phi_{YY}(t_1, t_2) \end{bmatrix}. \quad (\text{A.83})$$

The crosscovariance of $X(t)$ and $Y(t)$ is

$$\begin{aligned} \lambda_{XY}(t_1, t_2) &= E[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))] \\ &= \phi_{XY}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2). \end{aligned} \quad (\text{A.84})$$

The covariance matrix of $X(t)$ and $Y(t)$ is

$$\Lambda(t_1, t_2) = \begin{bmatrix} \lambda_{XX}(t_1, t_2) & \lambda_{XY}(t_1, t_2) \\ \lambda_{YX}(t_1, t_2) & \lambda_{YY}(t_1, t_2) \end{bmatrix}. \quad (\text{A.85})$$

If $X(t)$ and $Y(t)$ are each wide-sense stationary and jointly wide-sense stationary, then

$$\Phi(t_1, t_2) = \Phi(t_2 - t_1) = \Phi(\tau) \quad (\text{A.86})$$

$$\Lambda(t_1, t_2) = \Lambda(t_2 - t_1) = \Lambda(\tau), \quad (\text{A.87})$$

where $\tau = t_2 - t_1$.

A.5.2.1 Properties of $\phi_{XY}(\tau)$

Consider two random processes $X(t)$ and $Y(t)$ are each wide-sense stationary and jointly wide-sense stationary. The cross-correlation function $\phi_{XY}(\tau)$ has the following properties.

- $\phi_{XY}(\tau) = \phi_{YX}(-\tau)$
- $|\phi_{XY}(\tau)| \leq \frac{1}{2}[\phi_{XX}(0) + \phi_{YY}(0)]$
- $|\phi_{XY}(\tau)|^2 \leq \phi_{XX}(0)\phi_{YY}(0)$ if $X(t)$ and $Y(t)$ have zero mean.

A.5.2.2 Classifications of Random Processes

Two random processes $X(t)$ and $Y(t)$ are said to be

- uncorrelated if and only if $\lambda_{XY}(\tau) = 0$.
- 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 the random processes are also orthogonal if they are uncorrelated. Statistically independent random processes are always uncorrelated, however, not all uncorrelated random processes are statistically independent. In the special case of Gaussian random processes, if the processes are uncorrelated, then they are also statistically independent.

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 of $Z(t)$ is

$$\begin{aligned}\phi_{ZZ}(\tau) &= E[Z(t)Z(t+\tau)] \\ &= E[(X(t) + Y(t))(X(t+\tau) + Y(t+\tau))] \\ &= \phi_{XX}(\tau) + \phi_{YX}(\tau) + \phi_{XY}(\tau) + \phi_{YY}(\tau).\end{aligned}$$

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 at least one has 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?

$$\Phi(\tau) = \begin{bmatrix} \phi_{XX}(\tau) & \phi_{XY}(\tau) \\ \phi_{YX}(\tau) & \phi_{YY}(\tau) \end{bmatrix} = \begin{bmatrix} A^2 \cos(\tau) & 2A^2 \cos(3\tau/2) \\ 2A^2 \cos(3\tau/2) & A^2 \sin(2\tau) \end{bmatrix}.$$

(continued)

Example A.6 (continued)

The answer is no, because 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.

A.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-valued random processes.

A.5.3.1 Autocorrelation Function

The autocorrelation function of a complex-valued random process is

$$\begin{aligned} \phi_{ZZ}(t_1, t_2) &= \frac{1}{2}E[Z^*(t_1)Z(t_2)] \\ &= \frac{1}{2}E[(X(t_1) - jY(t_1))(X(t_2) + jY(t_2))] \\ &= \frac{1}{2}(\phi_{XX}(t_1, t_2) + \phi_{YY}(t_1, t_2) + j(\phi_{XY}(t_1, t_2) - \phi_{YX}(t_1, t_2))). \end{aligned} \quad (\text{A.88})$$

The factor of 1/2 is included for convenience, when $Z(t)$ is a complex-valued Gaussian random process. If $Z(t)$ is wide-sense stationary, then

$$\phi_{ZZ}(t_1, t_2) = \phi_{ZZ}(t_2 - t_1) = \phi_{ZZ}(\tau), \quad \tau = t_2 - t_1.$$

A.5.3.2 Cross-correlation Function

Consider two complex-valued random processes

$$\begin{aligned} Z(t) &= X(t) + jY(t) \\ W(t) &= U(t) + jV(t). \end{aligned}$$

The cross-correlation function of $Z(t)$ and $W(t)$ is

$$\begin{aligned} \phi_{ZW}(t_1, t_2) &= \frac{1}{2}E[Z^*(t_1)W(t_2)] \\ &= \frac{1}{2}(\phi_{XU}(t_1, t_2) + \phi_{YV}(t_1, t_2) + j(\phi_{XV}(t_1, t_2) - \phi_{YU}(t_1, t_2))). \end{aligned} \quad (\text{A.89})$$

If $X(t)$, $Y(t)$, $U(t)$ and $V(t)$ are pairwise wide-sense stationary random processes, then

$$\phi_{ZW}(t_1, t_2) = \phi_{ZW}(t_2 - t_1) = \phi_{ZW}(\tau). \quad (\text{A.90})$$

The cross-correlation of a complex wide-sense stationary random process satisfies the following property

$$\begin{aligned}
 \phi_{ZW}^*(\tau) &= \frac{1}{2} \mathbb{E}[Z(t)W^*(t + \tau)] \\
 &= \frac{1}{2} \mathbb{E}[Z(\hat{t} - \tau)W^*(\hat{t})] \\
 &= \frac{1}{2} \mathbb{E}[W^*(\hat{t})Z(\hat{t} - \tau)] \\
 &= \phi_{WZ}(-\tau),
 \end{aligned} \tag{A.91}$$

where the second line uses the change of variable $\hat{t} = t + \tau$. For a complex-valued random process $Z(t)$, it also follows that

$$\phi_{ZZ}^*(\tau) = \phi_{ZZ}(-\tau). \tag{A.92}$$

A.5.4 Power Spectral Density

The power spectral density (psd) of a wide-sense stationary random process $X(t)$ is the Fourier transform of the autocorrelation function, i.e.,

$$S_{XX}(f) = \int_{-\infty}^{\infty} \phi_{XX}(\tau) e^{-j2\pi f\tau} d\tau \tag{A.93}$$

$$\phi_{XX}(\tau) = \int_{-\infty}^{\infty} S_{XX}(f) e^{j2\pi f\tau} df. \tag{A.94}$$

If $X(t)$ is a real-valued wide-sense stationary random process, then its autocorrelation function $\phi_{XX}(\tau)$ is real and even. Therefore, $S_{XX}(-f) = S_{XX}(f)$ meaning that the power spectrum $S_{XX}(f)$ is also real and even. If $Z(t)$ is a complex-valued wide-sense stationary random process, then $\phi_{ZZ}(\tau) = \phi_{ZZ}^*(-\tau)$, and $S_{ZZ}^*(f) = S_{ZZ}(f)$ meaning that the power spectrum $S_{ZZ}(f)$ is real but not necessarily even.

The power, P , in a wide-sense stationary random process $X(t)$ is

$$\begin{aligned}
 P &= \mathbb{E}[X^2(t)] \\
 &= \phi_{XX}(0) \\
 &= \int_{-\infty}^{\infty} S_{XX}(f) df
 \end{aligned}$$

a result known as Parseval's theorem.

The cross power spectral density between two random processes $X(t)$ and $Y(t)$ is

$$S_{XY}(f) = \int_{-\infty}^{\infty} \phi_{XY}(\tau) e^{-j2\pi f\tau} d\tau. \tag{A.95}$$

If $X(t)$ and $Y(t)$ are both real-valued random processes, then

$$\phi_{XY}(\tau) = \phi_{YX}(-\tau)$$

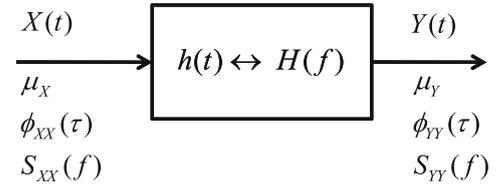
and

$$S_{XY}(f) = S_{YX}(-f).$$

If $X(t)$ and $Y(t)$ are complex-valued random processes, then

$$\phi_{XY}^*(\tau) = \phi_{YX}(-\tau)$$

Fig. A.2 Random process through a linear system



and

$$S_{XY}^*(f) = S_{YX}(f).$$

A.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 real-valued 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

$$\mu_Y = \int_{-\infty}^{\infty} h(\tau)E[X(t - \tau)]d\tau = \mu_X \int_{-\infty}^{\infty} h(\tau)d\tau = \mu_X H(0),$$

which is equal to the input mean multiplied by the d.c. gain of the filter.

The output autocorrelation function is

$$\begin{aligned} \phi_{YY}(\tau) &= E[Y(t)Y(t + \tau)] \\ &= E \left[\int_{-\infty}^{\infty} h(\beta)X(t - \beta)d\beta \int_{-\infty}^{\infty} h(\alpha)X(t + \tau - \alpha)d\alpha \right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\beta)h(\alpha)\phi_{XX}(\tau - \alpha + \beta)d\beta d\alpha \\ &= \int_{-\infty}^{\infty} h(\alpha) \int_{-\infty}^{\infty} h(\beta)\phi_{XX}(\tau + \beta - \alpha)d\beta d\alpha \\ &= \left(\int_{-\infty}^{\infty} h(\beta)\phi_{XX}(\tau + \beta)d\beta \right) * h(\tau) \\ &= h(-\tau) * \phi_{XX}(\tau) * h(\tau). \end{aligned}$$

Taking the Fourier transform of both sides, the power density spectrum of the output process $Y(t)$ is

$$\begin{aligned} S_{YY}(f) &= H(f)H^*(f)S_{XX}(f) \\ &= |H(f)|^2 S_{XX}(f). \end{aligned}$$

Example A.7: Consider the linear system shown in Fig. A.2. In this example we will find the cross-correlation between the input process $X(t)$ and the output $Y(t)$. The cross-correlation $\phi_{XY}(\tau)$ is given by

$$\begin{aligned}\phi_{XY}(\tau) &= E[X(t)Y(t + \tau)] \\ &= E\left[X(t) \int_{-\infty}^{\infty} h(\alpha)X(t + \tau - \alpha)d\alpha\right] \\ &= \int_{-\infty}^{\infty} h(\alpha)E[X(t)X(t + \tau - \alpha)]d\alpha \\ &= \int_{-\infty}^{\infty} h(\alpha)\phi_{XX}(\tau - \alpha)d\alpha \\ &= h(\tau) * \phi_{XX}(\tau).\end{aligned}$$

Also,

$$S_{XY}(f) = H(f)S_{XX}(f).$$

Example A.8: Suppose that a real-valued Gaussian random process $X(t)$ with mean μ_X and covariance function $\lambda_{XX}(\tau)$ is passed through the linear filter shown in Fig. A.2. In this example, the joint density of the random variables $X_1 = X(t_1)$ and $X_2 = Y(t_2)$ is of interest. If a Gaussian random process is passed through a linear filter, then the output process will also be a Gaussian random process. This is due to the fact that a sum of Gaussian random variables will yield another Gaussian random variable. Hence, X_1 and X_2 have a joint Gaussian density function as defined in (A.43) 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

$$\begin{aligned}\lambda_{X_1X_2}(\tau) &= E[(X(t) - \mu_X)(Y(t + \tau) - \mu_Y)] \\ &= E[X(t)Y(t + \tau)] - \mu_Y\mu_X.\end{aligned}$$

Now $\mu_Y = H(0)\mu_X$. Also, from the previous example

$$\begin{aligned}E[X(t)Y(t + \tau)] &= \int_{-\infty}^{\infty} h(\alpha)\phi_{XX}(\tau - \alpha)d\alpha \\ &= \int_{-\infty}^{\infty} h(\alpha)[\lambda_{XX}(\tau - \alpha) + \mu_X^2]d\alpha \\ &= \int_{-\infty}^{\infty} h(\alpha)\lambda_{XX}(\tau - \alpha)d\alpha + H(0)\mu_X^2.\end{aligned}$$

Therefore,

$$\lambda_{X_1X_2}(\tau) = \int_{-\infty}^{\infty} h(\alpha)\lambda_{XX}(\tau - \alpha)d\alpha = h(\tau) * \lambda_{XX}(\tau).$$

Also

$$\begin{aligned}\lambda_{X_2X_1}(\tau) &= \lambda_{X_1X_2}(-\tau) = \lambda_{X_1X_2}(\tau) \\ \lambda_{X_1X_1}(\tau) &= \lambda_{XX}(\tau)\end{aligned}$$

(continued)

Example A.8 (continued)

$$\lambda_{X_2X_2}(\tau) = h(\tau) * h(-\tau) * \lambda_{XX}(\tau),$$

where the first line follows from the even property of the autocovariance function. Hence, the covariance matrix is

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{X_1X_1}(0) & \lambda_{X_1X_2}(\tau) \\ \lambda_{X_2X_1}(\tau) & \lambda_{X_2X_2}(0) \end{bmatrix} = \begin{bmatrix} \lambda_{XX}(0) & h(\tau) * \lambda_{XX}(\tau) \\ h(\tau) * \lambda_{XX}(\tau) & h(\tau) * h(-\tau) * \lambda_{XX}(\tau) |_{\tau=0} \end{bmatrix}$$

Step 2: Write the joint density function of X_1 and X_2 .

Let

$$\mathbf{X} = (X_1, X_2)^T$$

$$\mathbf{x} = (x_1, x_2)^T$$

$$\boldsymbol{\mu}_X = (\mu_X, \mu_Y)^T = (\mu_X, H(0)\mu_X)^T.$$

Then

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi|\mathbf{\Lambda}|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_X)^T \mathbf{\Lambda}^{-1}(\mathbf{z} - \boldsymbol{\mu}_X) \right\}.$$

A.5.6 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 $\phi_{aa}(n) = \frac{1}{2}E[a_k^* a_{k+n}]$, and $\psi(t)$ is a real-valued pulse having finite energy. Note that the mean of $X(t)$

$$\mu_X(t) = \mu_a \sum_{n=-\infty}^{\infty} \psi(t - nT)$$

is periodic in t with period T . The autocorrelation function of $X(t)$ is

$$\begin{aligned} \phi_{XX}(t, t + \tau) &= \frac{1}{2}E[X^*(t)X(t + \tau)] \\ &= \frac{1}{2}E \left[\sum_{n=-\infty}^{\infty} a_n^* \psi(t - nT) \sum_{m=-\infty}^{\infty} a_m \psi(t + \tau - mT) \right] \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \phi_{aa}(m - n) \psi(t - nT) \psi(t + \tau - mT). \end{aligned}$$

It is relatively straightforward to show that

$$\phi_{XX}(t + kT, t + \tau + kT) = \phi_{XX}(t, t + \tau).$$

Therefore, the autocorrelation function $\phi_{XX}(t, t + \tau)$ is periodic in t with period T . Such a process with a periodic mean and autocorrelation function is said to be cyclostationary or periodic wide-sense stationary.

The power spectrum of a cyclostationary random process $X(t)$ can be computed by first determining the time-average autocorrelation

$$\phi_{XX}(\tau) = \langle \phi_{XX}(t, t + \tau) \rangle = \frac{1}{T} \int_T \phi_{XX}(t, t + \tau) dt$$

and then taking the Fourier transform in (A.93).

A.5.7 Discrete-Time Random Processes

Let $X_n \equiv X(n)$, where n is an integer time variable, be a complex-valued discrete-time random process. Then the m th moment of X_n is

$$E[X_n^m] = \int_{-\infty}^{\infty} x_n^m p_X(x_n) dx_n. \quad (\text{A.96})$$

The autocorrelation function of X_n is

$$\phi_{XX}(n, k) = \frac{1}{2} E[X_n^* X_k] = \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_n^* x_k p_{X_n, X_k}(x_n, x_k) dx_n dx_k \quad (\text{A.97})$$

and the autocovariance function is

$$\lambda_{XX}(n, k) = \phi(n, k) - \frac{1}{2} E[X_n^*] E[X_k]. \quad (\text{A.98})$$

If X_n is a wide-sense stationary discrete-time random process, then

$$\phi_{XX}(n, k) = \phi_{XX}(k - n) \quad (\text{A.99})$$

$$\lambda_{XX}(n, k) = \lambda_{XX}(k - n) = \phi_{XX}(k - n) - \frac{1}{2} |\mu_X|^2. \quad (\text{A.100})$$

From Parseval's theorem, the total power in the process X_n is

$$P = \frac{1}{2} E[|X_n|^2] = \phi_{XX}(0). \quad (\text{A.101})$$

The power spectrum of a discrete-time random process X_n is the discrete-time Fourier transform of the autocorrelation function

$$S_{XX}(f) = \sum_{n=-\infty}^{\infty} \phi_{XX}(n) e^{-j2\pi f n} \quad (\text{A.102})$$

and

$$\phi_{XX}(n) = \int_{-1/2}^{1/2} S_{XX}(f) e^{j2\pi f n} df. \quad (\text{A.103})$$

Note that $S_{XX}(f)$ is periodic in f with a period of unity, i.e., $S_{XX}(f) = S_{XX}(f + k)$ for any integer k . This is a characteristic of any discrete-time random process. For example, one obtained by sampling a continuous time random process $X_n = x(nT)$, where T is the sample period.

Suppose that a wide-sense stationary complex-valued discrete-time random process X_n is input to a discrete-time linear time-invariant system with impulse response h_n . The process is assumed to have mean μ_X and autocorrelation function $\phi_{XX}(n)$. The transfer function of the filter is

$$H(f) = \sum_{n=-\infty}^{\infty} h_n e^{-j2\pi fn}. \quad (\text{A.104})$$

The input, X_n , and output, Y_n , are related by the convolution sum

$$Y_n = \sum_{k=-\infty}^{\infty} h_k X_{n-k}. \quad (\text{A.105})$$

The output mean is

$$\mu_Y = E[Y_n] = \sum_{k=-\infty}^{\infty} h_k E[X_{n-k}] = \mu_X \sum_{k=-\infty}^{\infty} h_k = \mu_X H(0). \quad (\text{A.106})$$

The output autocorrelation is

$$\begin{aligned} \phi_{YY}(k) &= \frac{1}{2} E[Y_n^* Y_{n+k}] \\ &= \frac{1}{2} E \left[\sum_{\ell=-\infty}^{\infty} h_{\ell}^* X_{n-\ell}^* \sum_{m=-\infty}^{\infty} h_m X_{n+k-m} \right] \\ &= \sum_{\ell=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h_{\ell}^* h_m \frac{1}{2} E[X_{n-\ell}^* X_{n+k-m}] \\ &= \sum_{\ell=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h_{\ell}^* h_m \phi_{XX}(k + \ell - m) \\ &= \sum_{m=-\infty}^{\infty} h_m \sum_{\ell=-\infty}^{\infty} h_{\ell}^* \phi_{XX}(k + \ell - m) \\ &= h_k * \left\{ \sum_{\ell=-\infty}^{\infty} h_{\ell}^* \phi_{XX}(k + \ell) \right\} \\ &= h_k * \phi_{XX}(k) * h_{-k}^*. \end{aligned} \quad (\text{A.107})$$

where the convolution operation is understood to be a discrete-time convolution. The output psd can be obtained by taking the discrete-time Fourier transform of the autocorrelation function, resulting in

$$\begin{aligned} S_{YY}(f) &= H(f) S_{XX}(f) H^*(f) \\ &= |H(f)|^2 S_{XX}(f). \end{aligned} \quad (\text{A.108})$$

Once again, $S_{YY}(f)$ is periodic in f with a period of unity, i.e., $S_{YY}(f) = S_{YY}(f + k)$ for any integer k .

References

1. 3GPP, Study on 3D channel model for LTE (2015)
2. V.A. Aalo, J. Zhang, Performance of antenna array systems with optimum combining in a Rayleigh fading environment. *IEEE Commun. Lett.* **4**, 387–389 (2000)
3. A. Abdi, J.A. Barger, M. Kaveh, A parametric model for the distribution of the angle of arrival and the associated correlation function and power spectrum at the mobile station. *IEEE Trans. Veh. Technol.* **51**, 425–434 (2002)
4. M. Abramowitz, I.A. Stegun (ed.), *Handbook of Mathematical Functions* (Dover, New York, 1965)
5. F. Abrishamkar, E. Biglieri, Suboptimum detection of trellis-coded CPM for transmission on bandwidth- and power-limited channels. *IEEE Trans. Commun.* **39**, 1065–1074 (1991)
6. A. Abu-Dayya, N. Beaulieu, Outage probabilities of cellular mobile radio systems with multiple Nakagami interferers. *IEEE Trans. Veh. Technol.* **40**, 757–768 (1991)
7. A.A. Abu-Dayya, N.C. Beaulieu, Micro- and macrodiversity NCFSK (DPSK) on shadowed Nakagami-fading channels. *IEEE Trans. Commun.* **42**, 2693–2702 (1994)
8. A.A. Abu-Dayya, N.C. Beaulieu, Micro- and macrodiversity MDPSK on shadowed frequency-selective channels. *IEEE Trans. Commun.* **43**, 2334–2343 (1995)
9. G. Acosta, K. Tokuda, M.A. Ingram, Measured joint Doppler-delay power profiles for vehicle-to-vehicle communications at 2.4 GHz, in *IEEE Global Communications Conference*, Dallas, TX, November 2004, pp. 3813–3817
10. S. Agarwal, J. Holtzman, Modelling and analysis of handoff algorithms in multi-cellular systems, in *IEEE Vehicular Technology Conference*, Phoenix, AZ, May 1997, pp. 300–304
11. Y. Akaiwa, A conceptual design of microcellular radio communication system, in *IEEE Vehicular Technology Conference*, Orlando, FL, May 1990, pp. 156–160
12. Y. Akaiwa, H. Andoh, Channel segregation - a self organized dynamic channel allocation method: application to TDMA/FDMA microcellular systems. *IEEE J. Sel. Areas Commun.* **11**, 949–954 (1993)
13. A.S. Akki, Statistical properties of mobile-to-mobile land communication channels. *IEEE Trans. Veh. Technol.* **43**, 826–831 (1994)
14. A.S. Akki, F. Haber, A statistical model of mobile-to-mobile land communication channel. *IEEE Trans. Veh. Technol.* **35**, 2–7 (1986)
15. S.M. Alamouti, A simple transmit diversity technique for wireless communications. *IEEE J. Sel. Areas Commun.* **16**, 1451–1458 (1998)
16. L. Anderson, A simulation study of some dynamic channel assignment algorithms in a high capacity mobile telecommunications system. *IEEE Trans. Veh. Technol.* **22**, 210–217 (1973)
17. J.B. Anderson, T. Aulin, C.E. Sundberg, *Digital Phase Modulation* (Plenum, New York, 1986)
18. J.B. Andersen, T. Rappaport, S. Yoshida, Propagation measurements and models for wireless communications channels. *IEEE Commun. Mag.* **33**, 42–49 (1995)
19. M.R. Andrews, P.P. Mitra, R. deCarvalho, Tripling the capacity of wireless communications using electromagnetic polarization. *Nature* **409**, 316–318 (2001)
20. S. Ariyavisitakul, SIR-based power control in a CDMA system, in *IEEE Global Communication Conference*, Orlando, FL, December 1992, pp. 868–873
21. S. Ariyavisitakul, L.F. Chang, Signal and interference statistics of a CMDA system with feedback power control. *IEEE Trans. Commun.* **41**, 1626–1634 (1993)
22. T. Aulin, A modified model for the fading signal at a mobile radio channel. *IEEE Trans. Veh. Technol.* **28**, 182–203 (1979)
23. M.E. Austin, Decision-feedback equalization for digital communication over dispersive channels, Technical Report, MIT Lincoln Lab., Lexington, MA, August 1967
24. M.D. Austin, G.L. Stüber, Velocity adaptive handoff algorithms for microcellular systems, in *IEEE Conference on Universal Personal Communications*, Ottawa, Canada, October 1993, pp. 793–797
25. M.D. Austin, G.L. Stüber, Co-channel interference modeling for signal strength based handoff analysis. *Electron. Lett.* **30**, 1914–1915 (1994)
26. M.D. Austin, G.L. Stüber, Direction biased handoff algorithms for urban microcells, in *IEEE Vehicular Technology Conference*, Stockholm, Sweden, June 1994, pp. 101–105
27. M.D. Austin, G.L. Stüber, Exact co-channel interference analysis for log-normal shadowed Rician fading channels. *Electron. Lett.* **30**, 748–749 (1994)
28. M.D. Austin, G.L. Stüber, Velocity adaptive handoff algorithms for microcellular systems. *IEEE Trans. Veh. Technol.* **43**, 549–561 (1994)
29. M.D. Austin, G.L. Stüber, In-service signal quality estimation for TDMA cellular systems, in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Toronto, Canada, September 1995, pp. 836–840

30. M.D. Austin, G.L. Stüber, In-service signal quality estimation for TDMA cellular systems. *Kluwer J. Wireless Personal Commun.* **2**, 245–254 (1996)
31. L. Bahl, J. Cocke, F. Jelinek, J. Raviv, Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans. Inf. Theory* **20**, 284–287 (1974)
32. J.C.A. Baird, C.L. Zahm, Performance criteria for narrowband array processing, in *IEEE Conference on Decision and Control*, Miami, FL, December 1971, pp. 564–565
33. G. Battail, A conceptual framework for understanding Turbo codes. *IEEE J. Sel. Areas Commun.* **16**, 245–254 (1998)
34. N.C. Beaulieu, A.A. Abu-Dayya, P.J. McLane, Comparison of methods of computing lognormal sum distributions and outages for digital wireless applications, in *IEEE International Conference on Communications*, New Orleans, LA, May 1994, pp. 1270–1275
35. P. Bello, Characterization of random time-variant linear channels. *IEEE Trans. Commun.* **11**, 360–393 (1963)
36. S. Benedetto, G. Montorsi, Design of parallel concatenated convolutional codes. *IEEE Trans. Commun.* **44**, 591–600 (1996)
37. S. Benedetto, G. Montorsi, Unveiling turbo codes: some results on parallel concatenated coding schemes. *IEEE Trans. Inf. Theory* **42**, 409–428 (1996)
38. S. Benedetto, G. Montorsi, D. Divsalar, F. Pollara, Serial concatenation of interleaved codes: performance analysis, design and iterative decoding. JPL-TDA Progress Report 42–126, (1996), pp. 42–126
39. S. Benedetto, G. Montorsi, D. Divsalar, F. Pollara, Soft-output decoding algorithms in iterative decoding of turbo codes. JPL-TDA Progress Report 42–124 (1996), pp. 63–84
40. J.-E. Berg, R. Bownds, F. Lotse, Path loss and fading models for microcells at 900 MHz, in *IEEE Vehicular Technology Conference*, Denver, CO, May 1992, pp. 666–671
41. R.C. Bernhardt, Macroscopic diversity in frequency reuse radio systems. *IEEE J. Sel. Areas Commun.* **5**, 862–870 (1987)
42. C. Berrou, A. Glavieux, P. Thitimajshima, Near Shannon limit error-correcting coding and decoding, in *IEEE International Conference on Communications*, Geneva, Switzerland, June 1993, pp. 1064–1070
43. E. Biglieri, D. Divsalar, P. McLane, M. Simon, *Introduction to Trellis-Coded Modulation with Applications* (McMillan, New York, 1991)
44. M.A. Birchler, S.C. Jasper, A 64 kbps digital land mobile radio system employing M-16QAM, in *Proceedings of 5th Nordic Seminar on Land Mobile Radio*, Helsinki, Finland, December 1992, pp. 237–241
45. V.M. Bogachev, I.G. Kieslev, Optimum combining of signals in space-diversity reception. *Telecommun. Radio Eng.* **34/35**, 83 (1980)
46. G.P. Bortnyk, Combining signal images in accordance with signal-to-noise ratios, U.S. Patent No. 6,801,867, October 2004
47. G. Bortnyk, E. Franke, An improved quad diversity combining technique for effectively combating multipath, antenna gain and shadowing, in *IEEE Military Communications Conference*, October 2001, pp. 641–646
48. G.E. Bottomley, K. Jamal, Adaptive arrays and MLSE equalization, in *IEEE Vehicular Technology Conference*, Chicago, IL, July 1995, pp. 50–54
49. A. Burr, Capacity bounds and estimates for finite scatterers MIMO wireless channel. *IEEE J. Sel. Areas Commun.* **21**, 812–818 (2003)
50. G.J. Byers, F. Takawira, Spatially and temporally correlated MIMO channels: modeling and capacity analysis. *IEEE Trans. Veh. Technol.* **53**, 634–643 (2004)
51. A.R. Calderbank, J. Mazo, A new description of trellis codes. *IEEE Trans. Inf. Theory* **30**, 784–791 (1984)
52. A.R. Calderbank, N.J. Sloane, New trellis codes based on lattices and cosets. *IEEE Trans. Inf. Theory* **33**, 177–195 (1987)
53. J. Cavers, P. Ho, Analysis of the error performance of trellis-coded modulation in Rayleigh-fading channels. *IEEE Trans. Commun.* **40**, 74–83 (1992)
54. C. Chandra, T. Jeanes, W.H. Leung, Determination of optimal handover boundaries in a cellular network based on traffic distribution analysis of mobile measurement reports, in *IEEE Vehicular Technology Conference*, Phoenix, AZ, May 1997, pp. 305–309
55. R.W. Chang, Synthesis of band-limited orthogonal signals for multi-channel data transmission. *Bell Syst. Tech. J.* **46**, 1775–1796 (1966)
56. C.J. Chang, F.C. Ren, Downlink power control in DS/CDMA cellular mobile radio networks, in *IEEE Conference on Universal Personal Communications* (1994), pp. 89–93
57. U. Charash, Reception through Nakagami fading multipath channels with random delays. *IEEE Trans. Commun.* **27**, 657–670 (1979)
58. Y. Chau, S.-H. Yu, Space modulation on wireless fading channels, in *IEEE Vehicular Technology Conference*, p. 1668
59. D.K. Cheng, *Field and Wave Electromagnetics*, 2nd edn. (Addison-Wesley, Reading, 1989)
60. P. Chevillat, E. Eleftheriou, Decoding of trellis-encoded signal in the presence of intersymbol interference and noise. *IEEE Trans. Commun.* **37**, 669–676 (1989)
61. S. Chia, R. Steele, E. Green, A. Baran, Propagation and bit-error ratio measurements for a microcellular system. *J. Inst. Electron. Radio Eng. (UK)* **57**, 255–266 (1987)
62. S.T.S. Chia, R.J. Warburton, Handover criteria for city microcellular systems, in *IEEE Vehicular Technology Conference*, Orlando, FL, May 1990, pp. 276–281
63. C. Chong, C. Tan, D. Laurenson, M. Beach, A. Nix, A new statistical wideband spatio-temporal channel model for 5 GHz band WLAN systems. *IEEE J. Sel. Areas Commun.* **21**, 139–150 (2003)
64. D. Chu, Polyphase codes with good periodic correlation properties. *IEEE Trans. Inf. Theory* **18**, 531–532 (1972)
65. K.M. Chugg, A. Polydoros, MLSE for an unknown channel – part I: optimality considerations. *IEEE Trans. Commun.* **44**, 836–846 (1996)
66. L.J. Cimini, Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. *IEEE Trans. Commun.* **33**, 665–675 (1985)
67. J.L. Cimini, G. Foschini, Distributed algorithms for dynamic channel allocation in microcellular systems, in *IEEE Vehicular Technology Conference*, Denver, CO, May 1992, pp. 641–644
68. L.J. Cimini, G.J. Foschini, C.-L. I, Z. Miljanic, Call blocking performance of distributed algorithms for dynamic channel allocation in microcells. *IEEE Trans. Commun.* **42**, 2600–2607 (1994)
69. L.J. Cimini, G.J. Foschini, L. Shepp, Single-channel user-capacity calculations for self-organizing cellular systems. *IEEE Trans. Commun.* **42**, 3137–3143 (1994)
70. J.M. Cioffi, A.C. Bingham, A data-driven multitone echo canceller. *IEEE Trans. Commun.* **42**, 2853–2869 (1994)
71. A.P. Clark (ed.), *Advanced Data Transmission Systems* (Pentech Press, London, 1977)

72. A.P. Clark, R. Harun, Assessment of Kalman-filter channel estimator for an HF radio link. *IEE Proc.* **133**, 513–521 (1986)
73. G.C. Clark, Jr., J.B. Cain, *Error-Correction Coding for Digital Communications* (Plenum, New York, 1981)
74. R. Clarke, A statistical theory of mobile radio reception. *Bell System Tech. J.* **47**, 957–1000 (1968)
75. B. Conolly, I.J. Goods, A table of discrete Fourier transform pairs. *SIAM J. Appl. Math.* **32**, 810–822 (1977)
76. G.R. Cooper, R.W. Nettleton, A spread-spectrum technique for high-capacity mobile communications. *IEEE Trans. Veh. Technol.* **27**, 264–275 (1978)
77. K.G. Cornett, S.B. Wicker, Bit error rate estimation techniques for digital land mobile radios, in *IEEE Vehicular Technology Conference*, Saint Louis, MO, May 1991, pp. 543–548
78. *COST 207 Digital Land Mobile Radio Communications* (Commission of the European Communities, Brussels, 1989)
79. M. Failli, COST 207 Management Committee, TD(86)51-REV 3 (WG1), Proposal on channel transfer functions to be used in GSM tests late 1986 (1986)
80. COST 231 TD(91)109, 1800 MHz mobile net planning based on 900 MHz measurements (1991)
81. COST 231 TD(97)119-REV 2 (WG2), Urban transmission loss models for mobile radio in the 900- and 1,800-MHz bands (1991)
82. D.J. Costello, Construction of convolutional codes for sequential decoding. Technical Report EE-692, University of Notre Dame (1969)
83. D.J. Costello, G. Meyerhans, Concatenated turbo codes, in *IEEE Information Theory Symposium*, October 1996, pp. 571–574
84. D.C. Cox, Cochannel interference considerations in frequency reuse small-coverage-area radio systems. *IEEE Trans. Commun.* **30**, 135–142 (1982)
85. D.C. Cox, D.O. Reudnik, A comparison of some channel assignment strategies in large-scale mobile communication systems. *IEEE Trans. Commun.* **20**, 190–195 (1972)
86. D.C. Cox, D.O. Reudnik, Increasing channel occupancy in large-scale mobile radio systems: dynamic channel reassignment. *IEEE Trans. Veh. Technol.* **22**, 218–222 (1973)
87. J. Cui, D. Falconer, A.U. Sheikh, Performance evaluation of optimum combining and maximal ratio combining in the presence of co-channel interference and channel correlation for wireless communication systems. *Mob. Netw. Appl.* **2**, 315–324 (1997)
88. K. Daikoku, H. Ohdate, Optimal channel reuse in cellular land mobile radio systems. *IEEE Trans. Veh. Technol.* **32**, 217–224 (1983)
89. W.B. Davenport, W.L. Root, *An Introduction to the Theory of Random Signals and Noise* (McGraw-Hill, New York, 1987)
90. G.W. Davidson, D.D. Falconer, A.U.H. Sheikh, An investigation of block adaptive decision feedback equalization for frequency selective fading channels, in *IEEE International Conference on Communications*, Philadelphia, PA, June 1988, pp. 360–365
91. F. de Jager, C.B. Dekker, Tamed frequency modulation, a novel method to achieve spectrum economy in digital transmission. *IEEE Trans. Commun.* **26**, 534–542 (1978)
92. M. Debbah, R. Müller, MIMO channel modeling and the principle of maximum entropy. *IEEE Trans. Inf. Theory* **51**, 1667–1690 (2005)
93. P. Dent, G.E. Bottomley, T. Croft, Jakes fading model revisited. *Electron. Lett.* **7**, 1162–1163 (1993)
94. D. Divsalar, M. Simon, Trellis coded modulation for 4800–9600 bit/s transmission over a fading mobile satellite channel. *IEEE J. Sel. Areas Commun.* **5**, 162–174 (1987)
95. D. Divsalar, M. Simon, The design of trellis coded MPSK for fading channels: performance criteria. *IEEE Trans. Commun.* **36**, 1004–1012 (1988)
96. D. Divsalar, M. Simon, The design of trellis coded MPSK for fading channels: set partitioning for optimum code design. *IEEE Trans. Commun.* **36**, 1013–1022 (1988)
97. R.C. Dixon, *Spread Spectrum Techniques* (IEEE Press, New York, 1976)
98. S. Dolinar, D. Divsalar, Weight distributions for turbo codes using random and nonrandom permutations. *JPL-TDA Progress Report*, 42–121 (1995)
99. J. Driscoll, N. Karia, Detection process for V32 modems using trellis coding. *Proc. IEEE* **135**, 143–154 (1988)
100. A. Duel-Hallen, Detection for channels with intersymbol interference. Ph.D. Thesis, Cornell University, Ithaca, NY, 1987
101. A. Duel-Hallen, C. Heegard, Delayed decision feedback sequence estimation. *IEEE Trans. Commun.* **37**, 428–436 (1989)
102. A. Duel-Hallen, A family of multiuser decision-feedback detectors for asynchronous code-division multiple-access channels. *IEEE Trans. Commun.* **43**, 421–434 (1995)
103. F. Edbauer, Performance of interleaved trellis-coded differential 8-PSK modulation over fading channels. *IEEE J. Sel. Areas Commun.* **7**, 1340–1346 (1989)
104. EIA/TIA IS-54, Cellular system dual-mode mobile station – base station compatibility standard, Revision B (1992)
105. EIA/TIA IS-95, Mobile station – base station compatibility standard for dual-mode wideband spread spectrum cellular system, Revision B (1999)
106. E. Eleftheriou, D.D. Falconer, Tracking properties and steady-state performance of RLS adaptive filter algorithms. *IEEE Trans. Acoust. Speech Signal Process.* **34**, 1097–1110 (1986)
107. E. Eleftheriou, D.D. Falconer, Adaptive equalization techniques for HF channels. *IEEE J. Sel. Areas Commun.* **5**, 238–247 (1987)
108. S. Elnoubi, R. Singh, S. Gupta, A new frequency channel assignment algorithm in high capacity mobile communications systems. *IEEE Trans. Veh. Technol.* **31**, 125–131 (1982)
109. J. Engel, M. Peritsky, Statistically-optimum dynamic server assignment in systems with interfering servers. *IEEE Trans. Veh. Technol.* **22**, 203–209 (1973)
110. V. Erceg, P. Soma, D. Baum, S. Catreux, Multiple-input multiple-output fixed wireless radio channel measurements and modeling using dual-polarized antennas at 2.5 GHz. *IEEE Trans. Wirel. Commun.* **3**, 2288–2298 (2004)
111. V. Erceg, H. Sampath, S. Catreux-Erceg, Dual-polarization versus single-polarization MIMO channel measurement results and modeling. *IEEE Trans Wireless Commun.* **5**, 28–33 (2006)
112. ETSI – European Telecommunications Standards Institute, GSM Recommendation 05.08, January 1991
113. ETSI TR 125 943, Universal Mobile Telecommunications System (UMTS), Deployment aspects 3GPP TR 25.943 version 7.0.0 Release 7 (2007)
114. W.V. Etten, Maximum likelihood receiver for multiple channel transmission systems. *IEEE Trans. Commun.* **24**, 276–283 (1976)
115. D. Everitt, Traffic capacity of cellular mobile communications systems. *Comput. Netw. ISDN Syst.* **20**, 447–454 (1990)

116. D.E. Everitt, N.W. MacFadyen, Analysis of multicellular mobile radio-telephone systems: a model and evaluation. *British Telecom Technol. J.* **1**, 37–45 (1983)
117. D. Everitt, D. Manfield, Performance analysis of cellular mobile communication systems with dynamic channel assignment. *IEEE J. Sel. Areas Commun.* **7**, 1172–1179 (1989)
118. V.M. Eyuboğlu, Detection of coded modulation signals on linear, severely distorted channels using decision-feedback noise prediction with interleaving. *IEEE Trans. Commun.* **36**, 401–409 (1988)
119. V.M. Eyuboğlu, S.U. Qureshi, Reduced-state sequence estimation with set partitioning and decision feedback. *IEEE Trans. Commun.* **36**, 13–20 (1988)
120. G. Falciasecca, M. Frullone, G. Riva, M. Sentinelli, A.M. Serra, Investigation on a dynamic channel allocation for high capacity mobile radio systems, in *IEEE Vehicular Technology Conference*, Philadelphia, PA, (1988), pp. 176–181
121. D. Falconer, S. Ariyavisitakul, A. Benyamin-Seeyar, B. Edison, Frequency domain equalization for single-carrier broadband wireless systems. *IEEE Commun. Mag.* **40**, 58–66 (2002)
122. K. Feher, *Advanced Digital Communications* (Prentice-Hall, Englewood Cliffs, 1987)
123. K. Feher, *Wireless Digital Communications* (Prentice-Hall, Englewood Cliffs, 1995)
124. L.F. Fenton, The sum of log-normal probability distributions in scatter transmission systems. *IRE Trans. Commun.* **8**, 57–67 (1960)
125. F. Fernandes, A. Ashikhmin, T.L. Marzetta, Inter-cell interference in noncooperative TDD large scale antenna systems. *IEEE J. Sel. Areas Commun.* **31**, 192–201 (2013)
126. E.A. Frech, C.L. Mesquida, Cellular models and hand-off criteria, in *IEEE Vehicular Technology Conference*, San Francisco, CA, May 1989, pp. 128–135
127. R.C. French, Error rate predictions and measurements in the mobile radio data channel. *IEEE Trans. Veh. Technol.* **27**, 214–220 (1978)
128. R.C. French, The effect of fading and shadowing on channel reuse in mobile radio. *IEEE Trans. Veh. Technol.* **28**, 171–181 (1979)
129. G.D. Forney, Jr., Convolutional codes I: algebraic structure. *IEEE Trans. Inf. Theory* **16**, 720–738 (1970)
130. G.D. Forney, Jr., Burst-correcting codes for the classic bursty channel. *IEEE Trans. Commun. Technol.* **19**, 772–781 (1971)
131. G.D. Forney, Jr., Maximum likelihood sequence estimation of digital sequence in the presence of intersymbol interference. *IEEE Trans. Inform. Theory* **18**, 363–378 (1972)
132. G.D. Forney, Jr., Coset codes – part I: introduction to geometrical classification. *IEEE Trans. Inf. Theory* **34**, 1123–1151 (1988)
133. H. Furukawa, A. Yoshihiko, Self-organized reuse partitioning, a dynamic channel assignment method in cellular systems, in *IEEE Vehicular Technology Conference*, Secaucus, NJ, May 1993, pp. 524–527
134. P. Gaasvik, M. Cornefjord, V. Svenson, Different methods of giving priority to handoff traffic in a mobile telephone system with directed retry, in *IEEE Vehicular Technology Conference*, Saint Louis, MO, May 1991, pp. 549–553
135. G.J. Garrison, A power spectral density analysis for digital FM. *IEEE Trans. Commun.* **23**, 1228–1243 (1975)
136. E.A. Geraniotis, Direct-sequence spread-spectrum multiple-access communications over nonselective and frequency-selective Rician fading channels. *IEEE Trans. Commun.* **34**, 756–764 (1986)
137. E.A. Geraniotis, R. Mani, Throughput analysis of a random access tree protocol for direct-sequence spread-spectrum packet radio networks, in *IEEE Military Communications Conference*, Washington, D.C., October 1987, pp. 23.7.1–23.7.6
138. A. Gersho, Adaptive equalization of highly dispersive channels. *Bell Syst. Tech. J.* **48**, 55–70 (1969)
139. K.S. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver, Jr., C.E. Wheatley III, On the capacity of a cellular CDMA system. *IEEE Trans. Veh. Technol.* **40**, 303–312 (1991)
140. A.A. Giordano, F.M. Hsu (eds.), *Least Square Estimation with Applications to Digital Signal Processing* (Wiley, New York, 1985)
141. R. Gold, Optimum binary sequences for spread-spectrum multiplexing. *IEEE Trans. Inf. Theory* **13**, 619–621 (1967)
142. A.J. Goldsmith, L.J. Greenstein, A Measurement-based model for predicting coverage areas of urban microcells. *IEEE J. Sel. Areas Commun.* **11**, 1013–1023 (1993)
143. A.J. Goldsmith, L.J. Greenstein, G.J. Foschini, Error statistics of real time power measurements in cellular channels with multipath and shadowing, in *IEEE Vehicular Technology Conference*, Secaucus, NJ (1993), pp. 108–110
144. A.J. Goldsmith, L.J. Greenstein, G.J. Foschini, Error statistics of real-time power measurements in cellular channels with multipath and shadowing. *IEEE Trans. Veh. Technol.* **43**, 439–446 (1994)
145. D.J. Goodman, S.A. Grandhi, R. Vijayan, Distributed dynamic channel assignment schemes, in *IEEE Vehicular Technology Conference*, Secaucus, NJ, May 1993, pp. 532–535
146. J.-L. Gorricho, A. Rojas, J. Paradells, Power control at the combiner output to maximize the uplink capacity on a cellular spread spectrum system. *IEEE Commun. Lett.* **2**, 273–275 (1998)
147. I. Gradshteyn, I. Ryzhik, *Tables of Integrals, Series, and Products* (Academic Press, San Diego, 1980)
148. J. Granlund, Topics in the design of antennas for scatter. Technical Report 135, Lincoln Lab., Massachusetts Institute of Technology, Cambridge, MA (1956)
149. O. Grimlund, B. Gudmundson, Handoff strategies in microcellular systems, in *IEEE Vehicular Technology Conference*, Saint Louis, MO, May 1991, pp. 505–510
150. M. Gudmundson, Analysis of handover algorithms, in *IEEE Vehicular Technology Conference*, Saint Louis, MO, May 1991, pp. 537–541
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. *Electron. Lett.* **27**, 2145–2146 (1991)
153. J.C. Haartsen, The Bluetooth radio system. *IEEE Pers. Commun. Mag.* **7**, 28–36 (2000)
154. J. Hagenauer, E. Offer, L. Papke, Iterative decoding of binary block and convolutional codes. *IEEE Trans. Inf. Theory* **42**, 429–445 (1996)
155. S.W. Halpern, Reuse partitioning in cellular systems, in *IEEE Vehicular Technology Conference*, Toronto, Ontario, Canada (1983), pp. 322–327
156. K. Hamied, G.L. Stüber, A fractionally-spaced MLSE Receiver, in *IEEE International Conference on Communications*, Seattle, WA, 1995, pp. 7–11

157. K. Hanabe, V. Tetsuro, T. Otsu, Distributed adaptive channel allocation scheme with variable C/I threshold in cellular systems, in *IEEE Vehicular Technology Conference*, Secaucus, NJ, May 1993, pp. 164–167
158. K. Haneda, J. Zhang, L. Tian, G. Liu, Y. Zheng, H. Asplund, J. Li, Y. Wang, D. Steer, C. Li, T. Balercia, S. Lee, Y. Kim, A. Ghosh, T. Thomas, T. Nakamura, Y. Kakishima, T. Imai, H. Papadopoulos, T.S. Rappaport, G.R. MacCartney Jr., M.K. Samimi, S. Sun, O. Koymen, S. Hur, J. Park, C. Zhang, E. Mellios, A.F. Molisch, S.S. Ghassemzadeh, A. Ghosh, 5G 3GPP-like channel models for outdoor urban microcellular and macrocellular environments, in *IEEE Vehicular Technology Conference*, Nanjing, China, May 2016
159. S.V. Hanly, Capacity and power control in spread spectrum macrodiversity radio networks. *IEEE Trans. Commun.* **44**, 247–256 (1996)
160. P. Harley, Short distance attenuation measurements at 900 MHz and 1.8 GHz using low antenna heights for microcells. *IEEE J. Sel. Areas Commun.* **7**, 5–11 (1989)
161. M. Hata, T. Nagatsu, Mobile location using signal strength measurements in cellular systems. *IEEE Trans. Veh. Technol.* **29**, 245–251 (1980)
162. S. Haykin (ed.), *Adaptive Filter Theory* (Prentice-Hall, Englewood Cliff, 1986)
163. F. Heliot, M.A. Imran, R. Tafazolli, On the energy efficiency-spectral efficiency trade-off over the MIMO Rayleigh fading channel. *IEEE Trans. Commun.* **60**, 1345–1356 (2012)
164. M.-J. Ho, G.L. Stüber, Co-channel interference of microcellular systems on shadowed Nakagami fading channels, in *IEEE Vehicular Technology Conference*, Secaucus, NJ, May 1993, pp. 568–571
165. P. Hoeher, A statistical discrete-time model for the WSSUS multipath channel. *IEEE Trans. Veh. Technol.* **41**, 461–468 (1992)
166. J.M. Holtzman, A simple, accurate method to calculate spread-spectrum multiple-access error probabilities. *IEEE Trans. Commun.* **40**, 461–464 (1992)
167. J. Holtzman, Adaptive measurement intervals for handoffs, in *IEEE International Conference on Communications*, Chicago, IL, June 1992, pp. 1032–1036
168. S.Y.K. Hamabe, A. Ushirokawa, Forward-link power control utilizing neighboring-cell pilot power for DS-CDMA cellular systems, in *IEEE Vehicular Technology Conference*, Phoenix, AZ (1997), pp. 939–943
169. D. Hong, S.S. Rappaport, Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures. *IEEE Trans. Veh. Technol.* **35**, 77–92 (1986)
170. K. Imamura, A. Murase, Mobile communication control using multi-transmitter simul/sequential casting (MSSC), in *IEEE Vehicular Technology Conference*, Dallas, TX, May 1986, pp. 334–341
171. W.C. Jakes, *Microwave Mobile Communication* (IEEE Press, New York, 1993)
172. H. Jiang, S.S. Rappaport, CBWL: a new channel assignment and sharing method for cellular communication systems. *IEEE Trans. Veh. Technol.* **43**, 313–322 (1994)
173. M. Kaji, A. Akeyama, UHF-band propagation characteristics for land mobile radio, in *International Symposium on Antennas and Propagation*, University of British Columbia, Canada, June 1985, pp. 835–838
174. T. Kanai, Autonomous reuse partitioning in cellular systems, in *IEEE Vehicular Technology Conference*, Denver, CO, May 1992, pp. 782–785
175. T. Kasami, Weight distribution of Bose-Chaudhuri-Hocquenghem codes, in *Combinatorial Mathematics and its Applications* (University of North Carolina Press, Chapel Hill, 1967), pp. 335–357
176. T. Kasami, S. Lin, W. Peterson, Some results on cyclic codes which are invariant under the affine group and their applications. *Inf. Control* **11**, 475–496 (1968)
177. F.P. Kelly, *Reversibility and Stochastic Networks* (Wiley, New York, 1979)
178. F.P. Kelly, Blocked probabilities in large circuit-switched networks. *Adv. Appl. Probab.* **18**, 473–505 (1986)
179. J. Keramoal, L. Schumacher, K. Pedersen, P. Mogensen, F. Frederiksen, A stochastic MIMO radio channel model with experimental verification. *IEEE J. Sel. Areas Commun.* **20**, 1211–1226 (2002)
180. R.W. Kerr, P. McLane, Coherent detection of interleaved trellis encoded CPFSK on shadowed mobile radio channels. *IEEE Trans. Commun.* **41**, 159–169 (1992)
181. C. Komninakis, A fast and accurate Rayleigh fading simulator, in *IEEE Communications Conference*, San Francisco, CA, December 2003, pp. 3306–3310
182. I.Z. Kovacs, P.C.F. Eggers, K. Olesen, L.G. Petersen, Investigations of outdoor-to-indoor mobile-to-mobile radio communication channels, in *IEEE Vehicular Technology Conference*, Vancouver, Canada (2002), pp. 430–434
183. S. Kozono, Co-channel interference measurement method for mobile communication. *IEEE Trans. Veh. Technol.* **36**, 7–13 (1987)
184. S. Kozono, T. Suruhara, M. Sakamoto, Base station polarization diversity reception for mobile radio. *IEEE Trans. Veh. Technol.* **33**, 301–306 (1984)
185. A. Kuchar, J.P. Rossi, E. Bonek, Directional macro-cell channel characterization from urban measurements. *IEEE Trans. Antennas Propag.* **48**, 137–146 (2000)
186. E. Kudoh, T. Matsumoto, Effects of power control error on the system user capacity of DS/CDMA cellular mobile radios. *IEICE Trans.* **E75-B**, 524–529 (1992)
187. S.S. Kuek, W.C. Wong, Ordered dynamic channel assignment scheme with reassignment in highway microcells. *IEEE Trans. Commun.* **41**, 271–276 (1992)
188. J. Kunisch, J. Pamp, Wideband Car-to-Car Radio Channel Measurements and Model at 5.9 GHz, in *IEEE Vehicular Technology Conference*, Calgary, Canada (2008)
189. S.-C. Kwon, G.L. Stüber, Polarized channel model for body area networks using reflection coefficients
190. G. Labeledz, K. Felix, V. Lev, D. Schaeffer, Handover control issues in very high capacity cellular systems using small cells, in *International Conference on Digital Land Mobile Radio Communication*, University of Warwick, Coventry, UK (1987)
191. M. Landmann, K. Sivasondivat, J.-I. Takada, I. Ida, R. Thoma, Polarisation behaviour of discrete multipath and diffuse scattering in urban environments at 4.5 GHz. *EURASIP J. Wirel. Commun. Netw.*, **2007**, 60 (2007), Article ID 57980
192. B. Larsson, B. Gudmundson, K. Raith, Receiver performance for the North American digital cellular system, in *IEEE Vehicular Technology Conference*, Saint Louis, MO, May 1991, pp. 1–6
193. P.A. Laurent, Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses (AMP). *IEEE Trans. Commun.* **34**, 150–160 (1986)

194. W.C.Y. Lee, *Mobile Communications Engineering* (McGraw Hill, New York, 1982)
195. W.C.Y. Lee, Estimate of local average power of a mobile radio signal. *IEEE Trans. Veh. Technol.* **34**, 22–27 (1985)
196. W.C.Y. Lee, *Mobile Communications Design Fundamentals* (SAMS, Indianapolis, 1986)
197. T.H. Lee, J.C. Lin, A study on the distributed power control for cellular mobile systems, in *IEEE Vehicular Technology Conference*, Atlanta, GA (1996), pp. 1130–1134
198. W. Lee, Y. Yeh, Polarization diversity system for mobile radio. *IEEE Trans. Commun.* **20**, 912–923 (1972)
199. W.C.Y. Lee, Y.S. Yeh, On the estimation of the second-order statistics of log-normal fading in mobile radio environment. *IEEE Trans. Commun.* **22**, 809–873 (1974)
200. J.S. Lehnert, M.B. Pursley, Error probability for binary direct-sequence spread-spectrum communications with random signature sequences. *IEEE Trans. Commun.* **35**, 87–98 (1987)
201. A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering* (Addison-Wesley, Reading, 1989)
202. Y.X. Li, X. Huang, The simulation of independent Rayleigh faders. *IEEE Trans. Commun.* **50**, 1503–1514 (2002)
203. S. Lin, D.J. Costello, Jr., *Error Control Coding: Fundamentals and Applications* (Prentice-Hall, Englewood Cliffs, 1983)
204. J. Lin, F. Ling, J. Proakis, Joint data and channel estimation for TDMA mobile channels, in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Boston, MA (1992), pp. 235–239
205. J. Lin, F. Ling, J. Proakis, Fading channel tracking properties of several adaptive algorithms for the North American digital cellular system, in *IEEE Vehicular Technology Conference* (1993), pp. 273–276
206. F. Ling, J.G. Proakis, Adaptive lattice decision-feedback equalizers – their performance and application to time-variant multipath channels. *IEEE Trans. Commun.* **33**, 348–356 (1985)
207. J.-P.M. Linnartz, Site diversity in land mobile cellular telephony network with discontinuous voice transmission. *Eur. Trans. Telecommun.* **2**, 471–480 (1991)
208. J.-P.M. Linnartz, Exact analysis of the outage probability in multiple-user mobile radio. *IEEE Trans. Commun.* **40**, 20–23 (1992)
209. J.M.G. Linnartz, R.F. Fiesta, Evaluation of radio links and networks. California PATH Program (1996) [Online]. Available: <http://www.path.berkeley.edu/PATH/Publications/PDF/PRR/96/PRR-96-16.pdf>
210. F. Lotse, A. Wejke, Propagation measurements for microcells in central Stockholm, in *IEEE Vehicular Technology Conference*, Orlando, FL (1990), pp. 539–541
211. R.W. Lucky, Automatic equalization for digital communication. *Bell Syst. Tech. J.* **44**, 547–588 (1965)
212. R.W. Lucky, Techniques for adaptive equalization of digital communication systems. *Bell Syst. Tech. J.* **45**, 255–286 (1966)
213. R. Lucky, J. Salz, E. Weldon, *Principles of Data Communication* (McGraw Hill, New York, 1968)
214. M.B. Madayam, P.-C. Chen, J.M. Hotzman, Minimum duration outage for cellular systems: a level crossing analysis, in *IEEE Vehicular Technology Conference*, Atlanta, GA, June 1996, pp. 879–883
215. F.R. Magee, J.G. Proakis, Adaptive maximum-likelihood sequence estimation for digital signaling in the presence of intersymbol interference. *IEEE Trans. Inf. Theory* **19**, 120–124 (1973)
216. M. Marsan, G. Hess, Shadow variability in an urban land mobile radio environment. *Electron. Lett.* **26**, 646–648 (1990)
217. T.L. Marzetta, BLAST training: estimating channel characteristics for high capacity space-time wireless, in *Proceedings 37th Annual Allerton Conference Communications, Control, and Computing* (1999)
218. T.L. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Trans. Wireless Commun.* **9**, 3590–3600 (2010)
219. S.J. Mason, Feedback theory: further properties of signal flow graphs. *Proc. IRE* **44**, 920–926 (1956)
220. J.L. Massey, Coding and modulation in digital communications, in *International Zurich Seminar on Digital Communications*, Zurich, Switzerland (1974), pp. E2(1)–E2(4)
221. I. Mathworks, Filter design toolbox for use with MATLAB, User's Guide (2005)
222. J. Maurer, T. Fügen, W. Wiesbeck, Narrow-band measurement and analysis of the inter-vehicle transmission channel at 5.2 GHz, in *IEEE Vehicular Technology Conference*, Birmingham, AL, pp. 1274–1278, 2002.
223. W.R. Mende, Evaluation of a proposed handover algorithm for the GSM cellular system, in *IEEE Vehicular Technology Conference*, Orlando, FL (1990), pp. 264–269
224. V. Mignone, A. Morello, CD3-OFDM: A novel demodulation scheme for fixed and mobile receivers. *IEEE Trans. Commun.* **44**, 1144–1151 (1996)
225. L.B. Milstein, D.L. Schilling, R.L. Pickholtz, V. Erceg, M. Kullback, E.G. Kanterakis, D. Fishman, W.H. Biederman, D.C. Salerno, On the feasibility of a CDMA overlay for personal communications networks. *IEEE J. Sel. Areas Commun.* **10**, 655–668 (1992)
226. S. Mockford, 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 (1988)
227. S. Mockford, A.M.D. Turkmani, J.D. Parsons, Local mean signal variability in rural areas at 900 MHz, in *IEEE Vehicular Technology Conference*, Orlando, FL, May 1990, pp. 610–615
228. P.E. Mogensen, P. Eggers, C. Jensen, J.B. Andersen, Urban area radio propagation measurements at 955 and 1845 MHz for small and micro cells, in *IEEE Global Communications Conference*, Phoenix, AZ, December 1991, pp. 1297–1302
229. M.A. Mokhtar, S.C. Gupta, Capacity for cellular CDMA PCS's in Nakagami fading log-normal shadowing channels, in *IEEE Conference Universal Personal Communications* (1992), pp. 190–194
230. R.K. Morrow, Jr., J.S. Lehnert, Bit-to-bit error dependence in slotted DS/CDMA packet systems with random signature sequences. *IEEE Trans. Commun.* **37**, 1052–1061 (1989)
231. R. Muammar, Co-channel interference in microcellular mobile radio system, in *IEEE Vehicular Technology Conference*, Saint Louis, MO (1991), pp. 198–203
232. R. Muammar, S.C. Gupta, Cochannel interference in high-capacity mobile radio systems. *IEEE Trans. Commun.* **30**, 1973–1978 (1982)
233. A. Murase, I.C. Symington, E. Green, Handover criterion for macro and microcellular systems, in *IEEE Vehicular Technology Conference*, Saint Louis, MO, May 1991, pp. 524–530
234. K. Murota, K. Hirata, GMSK modulation for digital mobile radio telephony. *IEEE Trans. Commun.* **29**, 1044–1050 (1981)

235. H.G. Myung, J. Lim, D.J. Goodman, in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Helsinki, Finland, September 2006, pp. 1–5
236. H.G. Myung, J. Lim, D.J. Goodman, Single carrier FDMA for uplink wireless transmission. *IEEE Veh. Technol. Mag.* **1**, 30–38 (2006)
237. Y. Nagata, Analysis for spectrum efficiency in single cell trunked and cellular mobile radio. *IEEE Trans. Veh. Technol.* **35**, 100–113 (1987)
238. M. Nakagami, The m distribution; a general formula of intensity distribution of rapid fading, in *Statistical Methods in Radio Wave Propagation*, ed. by W.G. Hoffman (Pergamon Press, New York, 1960), pp. 3–36
239. S. Nanda, Teletraffic models for urban and suburban microcells: cell sizes and handoff rates. *IEEE Trans. Veh. Technol.* **42**, 673–682 (1993)
240. S. Nanda, D.J. Goodman, Dynamic resource acquisition: distributed carrier allocation for TDMA cellular systems, in *Third Generation Wireless Information Networks* (1992), pp. 99–124
241. R.W. Nettleton, G.R. Schloemer, A high capacity assignment method for cellular mobile telephone systems, in *IEEE Vehicular Technology Conference*, San Francisco, CA, May 1989, pp. 359–367
242. P. Newson, M.R. Heath, The capacity of a spread spectrum CDMA system for cellular mobile radio with consideration of system imperfections. *IEEE J. Sel. Areas Commun.* **12**, 673–683 (1994)
243. H.Q. Ngo, E.G. Larsson, T.L. Marzetta, Aspects of favorable propagation in massive MIMO, in *Proceedings of the 22nd European Signal Process Conference (EUSIPCO)*, Lisbon, Portugal (2014)
244. H. Nyquist, Certain topics in telegraph transmission theory. *Trans. Am. Inst. Electr. Eng.* **47**, 617–644 (1928)
245. C. Oestges, V. Erceg, A.J. Paulraj, Propagation modeling of MIMO multipolarized fixed wireless channels. *IEEE Trans. Veh. Technol.* **53**, 644–654 (2004)
246. K. Okada, A dynamic channel assignment strategy using information of speed and moving direction in microcellular systems, in *International Symposium on Circuits and Systems*, Chicago, IL, May 1993, pp. 2212–2215
247. K. Okada, F. Kubota, On dynamic channel assignment in cellular mobile radio systems, in *International Symposium on Circuits and Systems*, Singapore, June 1991, pp. 938–941
248. K. Okada, F. Kubota, A proposal of a dynamic channel assignment strategy with information of moving direction in microcellular systems. *Trans. IEICE E75-A*, 1667–1673 (1992)
249. K. Okanoue, A. Ushirokawa, H. Tomita, Y. Furuya, New MLSE receiver free from sample timing and input level controls, in *IEEE Vehicular Technology Conference*, Secaucus, NJ, June 1993, pp. 408–411
250. Y. Okumura, E. Ohmuri, T. Kawano, K. Fukuda, Field strength and its variability in VHF and UHF land mobile radio service. *Rev. ECL* **16**, 825–873 (1968)
251. A. Oppenheim, R. Schaffer, *Discrete-Time Signal Processing*, 3rd edn. (Prentice-Hall, Upper Saddle River, 2010)
252. W.P. Osborne, M.B. Luntz, Coherent and noncoherent detection of CPFSK. *IEEE Trans. Commun.* **22**, 1023–1036 (1974)
253. H. Panzer, R. Beck, Adaptive resource allocation in metropolitan area cellular mobile radio systems, in *IEEE Vehicular Technology Conference*, Orlando, FL, May 1990, pp. 638–645
254. A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1984)
255. J.D. Parsons, *The Mobile Radio Propagation Channel* (Wiley, New York, 1992)
256. J.D. Parsons, J.G. Gardiner, *Mobile Communication Systems* (Halsted Press, New York, 1989)
257. J.D. Parsons, A.M.D. Turkmani, Characterisation of mobile radio signals: model description, in *IEE Proceedings I, Communications, Speech and Vision*, vol. 138 (1991), pp. 549–556
258. S. Pasupathy, Nyquist's third criterion. *Proc. IEEE* **62**, 860–861 (1974)
259. C.S. Patel, G.L. Stüber, T.G. Pratt, Comparative analysis of statistical models for the simulation of Rayleigh faded cellular channels. *IEEE Trans. Commun.* **53**, 1017–1026 (2005)
260. M. Pätzold, *Mobile Fading Channels* (Wiley, West Sussex, 2002)
261. M. Pätzold, U. Killat, F. Laue, Y. Li, On the statistical properties of deterministic simulation models for mobile fading channels. *IEEE Trans. Veh. Technol.* **47**, 254–269 (1998)
262. W.W. Peterson, E.J. Weldon, *Error Correcting Codes*, 2nd edn. (MIT Press, Cambridge, 1972)
263. R.L. Pickholtz, D.L. Schilling, L.B. Milstein, Theory of spread-spectrum communications – a tutorial. *IEEE Trans. Commun.* **30**, 855–884 (1982)
264. M.F. Pop, N.C. Beaulieu, Limitations of sum-of-sinusoids fading channel simulators. *IEEE Trans. Commun.* **49**, 699–708 (2001)
265. R. Prasad, J.C. Arnbak, Comments on “analysis for spectrum efficiency in single cell trunked and cellular mobile radio”. *IEEE Trans. Veh. Technol.* **37**, 220–222 (1988)
266. R. Prasad, A. Kegel, Improved assessment of interference limits in cellular radio performance. *IEEE Trans. Veh. Technol.* **40**, 412–419 (1991)
267. R. Prasad, A. Kegel, Spectrum efficiency of microcellular systems. *Electron. Lett.* **27**, 423–425 (1991)
268. R. Prasad, A. Kegel, Effects of Rician faded and log-normal shadowed signals on spectrum efficiency in microcellular radio. *IEEE Trans. Veh. Technol.* **42**, 274–281 (1993)
269. R. Prasad, A. Kegel, J.C. Arnbak, Analysis of system performance of high-capacity mobile radio, in *IEEE Vehicular Technology Conference*, San Francisco, CA, May 1989, pp. 306–309
270. R. Price, P.E. Green, A communication technique for multipath channels. *Proc. IEEE* **46**, 555–570 (1958)
271. J.G. Proakis, J. Miller, An adaptive receiver for digital signaling through channels with intersymbol interference. *IEEE Trans. Inf. Theory* **15**, 484–497 (1969)
272. J.G. Proakis, M. Salehi, *Digital Communications*, 5th edn. (McGraw-Hill, New York, 2007)
273. M.B. Pursley, F.D. Garber, J.S. Lehnert, Analysis of generalized quadriphase spread-spectrum communications, in *IEEE International Conference on Communications*, Seattle, WA (1980), pp. 15.3.1–15.3.6
274. D. Qong, T.J. Lim, Soft handoff in CDMA mobile systems. *IEEE Pers. Commun. Mag.* **6**, 6–17 (1997)
275. S.U. Qureshi, Adaptive equalization. *Proc. IEEE* **73**, 1349–1387 (1985)
276. R.I.-R. M.1225, Guidelines for evaluation of radio transmission techniques for IMT-2000 (1997)
277. R. Raheli, A. Polydoros, C.-K. Tzou, The principle of per-survivor processing: a general approach to approximate and adaptive MLSE, in *IEEE Global Communications Conference* (1991), pp. 33.3.1–33.3.6

278. A.H.M. Raitola, R. Wichman, Transmission diversity in wideband CDMA, in *IEEE Vehicular Technology Conference*, Houston, TX (1999), pp. 1545–1549
279. J.L. Ramsey, Realization of optimum interleavers. *IEEE Trans. Inf. Theory* **16**, 338–345 (1970)
280. T. Rappaport, L. Milstein, Effects of path loss and fringe user distribution on CDMA cellular frequency reuse efficiency, in *IEEE Global Communications Conference*, San Diego, CA, December 1990, pp. 500–506
281. P. Raymond, Performance analysis of cellular networks. *IEEE Trans. Commun.* **39**, 1787–1793 (1991)
282. Research & Development Center for Radio Communications (RCR), Digital cellular telecommunication systems, April 1991. RCR STD-27 (1991)
283. S. Rice, Statistical properties of a sine wave plus noise. *Bell Syst. Tech. J.* **27**, 109–157 (1948)
284. M.H.K. Rohani, K. Kuchi, A comparison of base station transmit diversity methods for third generation cellular standards, in *IEEE Vehicular Technology Conference*, Houston, TX (1999), pp. 351–355
285. F. Rusek, D. Persson, B.K. Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, F. Tufvesson, Scaling Up MIMO, opportunities and challenges with very large arrays. *IEEE Signal Proc. Mag.* **30**, 40–60 (2013)
286. A. Rustako, N. Amitay, G. Owens, R. Roman, Radio propagation at microwave frequencies for line-of-sight microcellular mobile and personal communications. *IEEE Trans. Veh. Technol.* **40**, 203–210 (1991)
287. J. Salz, Optimum mean-square decision-feedback equalization. *Bell Syst. Tech. J.* **52**, 1341–1373 (1973)
288. A. Sampath, J. Holtzman, Estimation of maximum Doppler frequency for handoff decisions, in *IEEE Vehicular Technology Conference*, Secaucus, NJ, May 1993, pp. 859–862
289. H. Sari, G. Karam, Orthogonal frequency-division multiple access and its application to CATV networks. *Eur. Trans. Commun.* **45**, 507–516 (1998)
290. H. Sari, G. Karam, I. Jeanclaude, Channel equalization and carrier synchronization in OFDM systems, in *Audio and Video Digital Radio Broadcasting Systems and Techniques*, ed. by R. de Gaudenzi, M. Luise (Elsevier, Amsterdam, 1994), pp. 191–202
291. H. Sari, G. Karam, I. Jeanclaude, Frequency-domain equalization of mobile radio and terrestrial broadcast channels, in *IEEE Global Communications Conference*, San Francisco, CA (1994), pp. 1–5
292. H. Sari, G. Karam, I. Jeanclaude, Transmission techniques for digital terrestrial TV broadcasting. *IEEE Commun. Mag.* **33**, 100–109 (1995)
293. A. Sayeed, Deconstructing multiantenna fading channels. *IEEE Trans. Acoust. Speech Signal Process.* **50**, 2563–2579 (2002)
294. D. Schleher, Generalized Gram-Charlier series with application to the sum of lognormal variates. *IEEE Trans. Inf. Theory* **23**, 275–280 (1977)
295. T.A. Schonhoff, Symbol error probabilities for M-ary CPFSK: Coherent and noncoherent detection. *IEEE Trans. Commun.* **24**, 644–652 (1976)
296. S. Schwartz, Y.S. Yeh, On the distribution function and moments of power sums with log-normal components. *Bell Syst. Tech. J.* **61**, 1441–1462 (1982)
297. M. Serizawa, J. Murakami, Phase tracking Viterbi demodulator. *Electron. Lett.* **40**, 792–794 (1989)
298. N. Seshadri, Joint data and channel estimation using fast blind trellis search techniques, in *IEEE Global Communications Conference*, San Diego, CA (1990), pp. 1659–1663
299. M. Shafi, M. Zhang, A. Moustakas, P. Smith, A. Molisch, F. Tufvesson, S. Simon, Polarized MIMO channels in 3-D: models, measurements, and mutual information. *IEEE J. Sel. Areas Commun.* **24**, 514–527 (2006)
300. C.E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948)
301. W.H. Sheen, G.L. Stüber, MLSE equalization and decoding for multipath-fading channels. *IEEE Trans. Commun.* **39**, 1455–1464 (1991)
302. H. Shiino, N. Yamaguchi, Y. Shoji, Performance of an adaptive maximum-likelihood receiver for fast fading multipath channel, in *IEEE Vehicular Technology Conference*, Denver, CO, May 1992, pp. 380–383
303. M.K. Simon, M.-S. Alouini, *Digital Communication over Fading Channels – A Unified Approach to Performance Analysis* (Wiley, New York, 2000)
304. M.K. Simon, D. Divsalar, Maximum-likelihood block detection of noncoherent continuous phase modulation. *IEEE Trans. Commun.* **41**, 90–98 (1993)
305. M.K. Simon, J.K. Omura, R.A. Scholtz, B.K. Levitt, *Spread Spectrum Communications* (Computer Science Press, Rockville, 1985)
306. P. Soma, D.S. Baum, V. Erceg, R. Krishnamoorthy, A.J. Paulraj, Analysis and modeling of MIMO radio channel based on outdoor measurements conducted at 2.5 GHz for fixed BWA applications, in *IEEE International Conference on Communications*, April 2002, pp. 272–276
307. K. Steiglitz, Computer-aided design of recursive digital filters. *IEEE Trans. Audio Electroacoust.* **18**, 123–129 (1970)
308. G.L. Stüber, Soft decision direct-sequence DPSK receivers. *IEEE Trans. Veh. Technol.* **37**, 151–157 (1988)
309. E.W. Swokowski, *Calculus with Analytical Geometry* (Prindle, Weber, and Schmidt, New York, 1979)
310. J. Tajima, K. Imamura, A strategy for flexible channel assignment in mobile communication systems. *IEEE Trans. Veh. Technol.* **37**, 92–103 (1988)
311. K. Takeo, M. Nishino, Y. Ameazwa, S. Sato, Adaptive traffic control scheme for non-uniform traffic distribution in microcellular mobile communication system, in *IEEE Vehicular Technology Conference*, Orlando, FL, May 1990, pp. 527–531
312. J. Tan, G. Stüber, Analysis and design of interleaver mappings for iteratively decoded BICM. *IEEE Trans. Wirel. Commun.* **4**, 662–672 (2005)
313. J. Tan, G.L. Stüber, Frequency-domain equalization for continuous phase modulation. *IEEE Trans. Wirel. Commun.* **4**, 2479–2490 (2005)
314. V. Tarokh, N. Seshadri, A.R. Calderbank, Space-time codes for high data rate wireless communication: performance analysis and code construction. *IEEE Trans. Inf. Theory* **44**, 744–765 (1998)
315. V. Tarokh, H. Jafarkhani, A.R. Calderbank, Space-time block codes from orthogonal designs. *IEEE Trans. Inf. Theory* **45**, 1456–1467 (1999)
316. V. Tarokh, H. Jafarkhani, A.R. Calderbank, Space-time block coding for wireless: performance results. *IEEE J. Sel. Areas Commun.* **17**, 451–460 (1999)
317. S. Tekinay, B. Jabbari, Handover and channel assignment in mobile cellular networks. *IEEE Commun. Mag.* **29**, 42–46 (1991)
318. H. Thaper, Real-time application of trellis coding to high-speed voiceband data transmission. *IEEE J. Sel. Areas Commun.* **2**, 648–658 (1984)
319. N.D. Tripathi, J.H. Reed, H.F. Vanlandingham, Handoff in cellular systems. *IEEE Pers. Commun. Mag.* **5**, 26–37 (1998)

320. R.J. Tront, J.K. Cavers, M.R. Ito, Performance of Kalman decision-feedback equalization in HF radio modems, in *IEEE International Conference on Communications*, Toronto, Canada, June 1986, pp. 1617–1621
321. K. Tsoukatos, Power control in a mobility environment, in *IEEE Vehicular Technology Conference*, Phoenix, AZ (1997), pp. 740–744
322. A.M.D. Turkmani, Performance evaluation of a composite microscopic plus macroscopic diversity system. *IEE Proc. I*, **138**, 15–20 (1991)
323. A.M.D. Turkmani, Probability of error for M-branch selection diversity. *IEE Proc. I*, **139**, 71–78 (1992)
324. A.M.D. Turkmani, J.D. Parsons, F. Ju, D.G. Lewis, Microcellular radio measurements at 900,1500, and 1800 MHz, in *5th International Conference on Mobile Radio and Personal Communications*, Coventry, UK, December 1989, pp. 65–68
325. G. Ungerboeck, Channel coding with multilevel phase signals. *IEEE Trans. Inf. Theory* **28**, 55–67 (1982)
326. G. Ungerboeck, Trellis coded modulation with redundant signal sets – part I: introduction. *IEEE Commun. Mag.* **25**, 5–11 (1987)
327. G.M. Vachula, J.F.S. Hill, On optimal detection of band-limited PAM signals with excess bandwidth. *IEEE Trans. Commun.* **29**, 886–890 (1981)
328. R. van Nee, G. Awater, M. Morikura, H. Takanashi, M. Webster, K.W. Halford, New high-rate wireless LAN standards. *IEEE Commun. Mag.* **37**, 82–88 (1999)
329. F. Vatalaro, A. Forcella, Doppler spectrum in mobile-to-mobile communications in the presence of three-dimensional multipath scattering. *IEEE Trans. Veh. Technol.* **46**, 213–219 (1997)
330. R.G. Vaughan, Polarization diversity in mobile communications. *IEEE Trans. Veh. Technol.* **39**, 177–185 (1990)
331. R. Vijayan, J. Holtzman, Analysis of handoff algorithms using nonstationary signal strength measurements, in *IEEE Global Communications Conference*, Orlando, FL, December 1992, pp. 1405–1409
332. R. Vijayan, J. Holtzman, Sensitivity of handoff algorithms to variations in the propagation environment, in *IEEE Conference on Universal Personal Communications*, Ottawa, Canada, October 1993, pp. 158–162
333. R. Vijayan, J.M. Holtzman, A model for analyzing handoff algorithms. *IEEE Trans. Veh. Technol.* **42**, 351–356 (1993)
334. A.J. Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory* **13**, 260–269 (1967)
335. A.J. Viterbi, Convolutional codes and their performance in communication systems. *IEEE Trans. Commun.* **19**, 751–772 (1971)
336. A.J. Viterbi, A.M. Viterbi, K. Gilhousen, E. Zehavi, Soft handoff extends CDMA cell coverage and increases reverse channel capacity. *IEEE J. Sel. Areas Commun.* **12**, 1281–1288 (1994)
337. E. Viterbo, K. Fazel, How to combat long echoes in OFDM transmission schemes: subchannel equalization or more powerful channel coding, in *IEEE Global Communications Conference*, Singapore (1995), pp. 2069–2074
338. J.-F. Wagen, Signal strength measurements at 881 MHz for urban microcells in downtown Tampa, in *IEEE Global Communications Conference*, Phoenix, AZ, December 1991, pp. 1313–1317
339. E.H. Walker, Penetration of radio signals into buildings in cellular radio environments. *Bell Syst. Tech. J.* **62**, 2719–2734 (1983)
340. J. Wallace, M. Jensen, Modeling the indoor MIMO wireless channel. *IEEE Trans. Antennas Propag.* **50**, 591–599 (2002)
341. L.-C. Wang, C.-T. Lea, Macrodiversity co-channel interference analysis. *Electron. Lett.* **31**, 614–616 (1995)
342. L.-C. Wang, C.-T. Lea, Performance gain of a S-macrodiversity in a lognormal shadowed Rayleigh fading channel. *Electron. Lett.* **31**, 1785–1787 (1995)
343. L.-C. Wang, G.L. Stüber, C.-T. Lea, Architecture design, frequency planning, and performance analysis for a microcell/macrocell overlaying system. *IEEE Trans. Veh. Technol.* **46**, 836–848 (1997)
344. L.-C. Wang, G.L. Stüber, C.-T. Lea, Effects of Rician fading and branch correlation on a local-mean-based macrodiversity cellular system. *IEEE Trans. Veh. Technol.* **2**, 429–436 (1999)
345. R. Wang, D. Cox, Channel modeling for ad hoc mobile wireless networks, in *IEEE Vehicular Technology Conference*, Birmingham, AL (2002), pp. 21–25
346. L.F. Wei, Trellis-coded modulation with multidimensional constellations. *IEEE Trans. Inf. Theory* **33**, 483–501 (1987)
347. L.F. Wei, Coded M-DPSK with built-in time diversity for fading channels. *IEEE Trans. Inf. Theory* **39**, 1820–1839 (1993)
348. S.B. Weinstein, P.M. Ebert, Data transmission by frequency division multiplexing using the discrete Fourier transform. *IEEE Trans. Commun.* **19**, 628–634 (1971)
349. K.A. West, G.L. Stüber, An aggressive dynamic channel assignment strategy for a microcellular environment. *IEEE Trans. Veh. Technol.* **43**, 1027–1038 (1994)
350. W. Wiechselberger, M. Herdin, H. Özcelik, E. Bonek, A stochastic MIMO channel model with joint correlation at both link ends. *IEEE Trans. Wirel. Commun.* **5**, 90–100 (2006)
351. A. Williamson, B. Egan, J. Chester, Mobile radio propagation in Auckland at 851 MHz. *Electron. Lett.* **20**, 517–518 (1984)
352. J.H. Winters, Optimum combining in digital mobile radio with cochannel interference. *IEEE Trans. Veh. Technol.* **33**, 144–155 (1984)
353. J.H. Winters, Optimum combining in digital mobile radio with cochannel interference. *IEEE J. Sel. Areas Commun.* **4**, 528–539 (1984)
354. J.H. Winters, Signal acquisition and tracking with adaptive arrays in the digital mobile radio system IS-54 with flat fading. *IEEE Trans. Veh. Technol.* **42**, 377–384 (1993)
355. L. Wong, P. McLane, Performance of trellis codes for a class of equalized ISI channels. *IEEE Trans. Commun.* **36**, 1330–1336 (1988)
356. A. Wonjar, Unknown bounds on performance in Nakagami channels. *IEEE Trans. Commun.* **34**, 22–24 (1986)
357. J.M. Wozencraft, I.M. Jacobs, *Principles of Communication Engineering* (Waveland Press, Prospect Heights, 1990)
358. H. Xia, H. Bertoni, L. Maciel, A. Landsay-Stewart, Radio propagation measurements and modeling for line-of-sight microcellular systems, in *IEEE Vehicular Technology Conference*, Denver, CO, May 1992, pp. 349–354
359. C. Xiao, Y.R. Zheng, A statistical simulation model for mobile radio fading channels, in *Wireless Communications and Networking Conference*, New Orleans, LA, March 2003, pp. 144–149
360. C. Xiao, Y.R. Zheng, N. Beaulieu, Statistical simulation models for Rayleigh and Rician fading, in *IEEE International Conference on Communications*, Anchorage, AK, May 2003, pp. 3524–3529
361. Y. Yamada, Y. Ebine, N. Nakajima, Base station/vehicular antenna design techniques employed in high capacity land mobile communications system. *Rev. Electr. Commun. Lab. NTT* **35**, 115–121 (1987)

362. Y. Yao, A. Sheikh, Outage probability analysis for microcell mobile radio systems with cochannel interferers in Rician/Rayleigh fading environment. *Electron. Lett.* **26**, 864–866 (1990)
363. Y. Yao, U. Sheikh, Investigation into cochannel interference in microcellular mobile radio systems. *IEEE Trans. Veh. Technol.* **41**, 114–123 (1992)
364. R.D. Yates, A framework for uplink power control in cellular radio system. *IEEE J. Sel. Areas Commun.* **13**, 1341–1347 (1995)
365. Y. Yeh, S.C. Schwartz, Outage probability in mobile telephony due to multiple log-normal interferers. *IEEE Trans. Commun.* **32**, 380–388 (1984)
366. Y.H. Yeh, J.W. Wilson, S.C. Schwartz, Outage probability in mobile telephone with directive antennas and macrodiversity. *IEEE J. Sel. Areas Commun.* **2**, 507–511 (1984)
367. K.L. Yeung, T.-S.P. Yum, Compact pattern based channel assignment for cellular mobile systems. *IEEE Trans. Veh. Technol.* **43**, 892–896 (1994)
368. L. Yiin, G.L. Stüber, Noncoherently detected trellis-coded partial response CPM on mobile radio channels. *IEEE Trans. Commun.* **44**, 967–975 (1996)
369. D.J. Young, N.C. Beaulieu, The generation of correlated Rayleigh random variates by inverse discrete Fourier transform. *IEEE Trans. Commun.* **48**, 1114–1127 (2000)
370. W.P. Yung, Probability of Bit Error for MPSK modulation with diversity reception in Rayleigh fading and lognormal shadowing channel. *IEEE Trans. Commun.* **38**, 933–937 (1990)
371. A.G. Zajić, G.L. Stüber, A new simulation model for mobile-to-mobile Rayleigh fading channels, in *Wireless Communications and Networking Conference*, Las Vegas, NV, April 2006, pp. 1266–1270
372. A.G. Zajić, G.L. Stüber, Efficient simulation of Rayleigh fading with enhanced de-correlation properties. *IEEE Trans. Wirel. Commun.* **5**, 1866–1875 (2006)
373. N. Zhang, J. Holtzman, Analysis of handoff algorithms using both absolute and relative measurements, in *IEEE Vehicular Technology Conference*, Stockholm, Sweden, June 1994, pp. 82–86
374. N. Zhang, J. Holtzman, Analysis of a CDMA soft handoff algorithm, in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Toronto, Canada, September 1995, pp. 819–823
375. M. Zhang, T. Yum, Comparisons of channel-assignment strategies in cellular mobile telephone systems. *IEEE Trans. Veh. Technol.* **38**, 211–215 (1989)
376. Y.R. Zheng, C. Xiao, Improved models for the generation of multiple uncorrelated Rayleigh fading waveforms. *IEEE Commun. Lett.* **6**, 256–258 (2002)
377. Y.R. Zheng, C. Xiao, Simulation models with correct statistical properties for Rayleigh fading channels. *IEEE Trans. Commun.* **51**, 920–928 (2003)
378. R. Ziemer, R. Peterson, *Digital Communications and Spread Spectrum Systems* (MacMillan, New York, 1985)

Index

A

Active set, 625
Adjacent channel interference, 17
Alamouti's transmit diversity, 310
Analog cellular systems
 AMPS, 2
 NMT900, 2
 NTT, 2
Aulin's model, 45
Average fade duration, 56

B

Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm, 416
 backward recursion, 417
 branch metric, 417
 forward recursion, 416
Bandwidth efficiency, 25
Barker sequences, 459
Binary block codes
 dual code, 399
 encoder
 and decoder, 399
 generator matrix, 399
 parity check matrix, 399
 systematic, 399
Binary orthogonal codes, 183
Bi-orthogonal signals, 184
 error probability, 253
Bit interleaved coded modulation (BICM), 433
Block codes, 399
 error correction, 402
 error detection, 401
 free Hamming distance, 400
 MDS, 400
 probability of undetected error, 40
 singleton bound, 400
 space-time block codes, 403
 standard array decoding, 402
 syndrome, 400
 syndrome decoding, 403
 weight distribution, 401
Block interleaver
 interleaver depth, 428
 interleaver span, 428
Bluetooth, 13

C

Capacity, 24, 28
 GSM, 28
 IS-95, 29
Carrier-to-interference-plus-noise ratio (CINR) estimation, 632
 discrete-time model, 633
 estimation of $(I+N)$, 633
Carrier-to-interference ratio (CIR) estimation
 estimation of $C/(I+N)$, 635
 training sequence based, 636
CDMA. *See* Code division multiple access (CDMA)
 cdma2000, 6, 7
Cell breathing, 19
Cell capacity, 28
Cell sectoring, directional antennas, 529
Cell splitting, 531
 channel segmenting, 532
 overlaid cells, 532
 power reduction, 532
Cellular architectures
 CDMA, 563
 hierarchical, 574
 selective transmit diversity, 584
 cell sectoring, 529
 cell splitting, 531
 cluster planning, 536
 fractional reuse, 533
 hierarchical maximum ratio combining, 575
 macrodiversity, 554
 multi-carrier, 501
 OFDMA, 517, 529
 reuse partitioning, 533
 TDMA, 529
Channel assignment, 643
 autonomous reuse partitioning, (ARP), 657
 carrier groups, 657
 channel borrowing, 652
 channel locking, 653, 654
 channel ordering, 653
 directional lending, 654
 directional locking, 654
 dynamic, 653
 channel change, 644
 channel changing rate, 645
 deadlock, 644
 directed handoff (DH), 656

- Channel assignment, 643 (*cont.*)
 - directed retry (DR), 656
 - hybrid FCA/DCA, 652
 - instability, 644
 - moving direction, 656
 - reduced transceiver coverage, 656
 - self-organized reuse partitioning (SORP), 657
 - service interruption, 644
 - speed and moving direction (SMD), 656
 - Channel borrowing without locking (CBWL), 654
 - Channel polarization functions, 70
 - Classical beam forming, 303
 - Classification of channels
 - power delay profile, 67
 - signal and system correlation functions, 69
 - uncorrelated scattering channel, 66
 - wide sense stationary channel, 66
 - wide sense stationary uncorrelated scattering channel, 67
 - Cluster planning, 536
 - adjacent channel interference, 553
 - performance analysis, 540
 - downlink C/I analysis, 543
 - macrocell performance, 542
 - microcell performance, 547
 - uplink CCI analysis, 544
 - procedure, 536
 - system architecture, 536
 - underlaid microcells, 538
 - Co-channel demodulation, 375
 - channel model, 375
 - discrete-time channel model, 378
 - J-MLSE receiver, 375
 - pairwise error probability, 381
 - $T = 2$ -spaced receiver, 382
 - error probability, 383
 - practical receiver, 386
 - timing phase sensitivity, 384
 - Viterbi algorithm, 380
 - Co-channel interference, 17
 - log-normal interferers, 148
 - Farley's method, 152
 - Fenton-Wilkinson method, 149
 - Schwartz-and-Yeh method, 150
 - multiple log-normal interferers, 153
 - multiple log-normal Nakagami interferers, 157
 - outage, 147
 - Ricean/multiple Rayleigh interferers, 156
 - Code division multiple access (CDMA), 563
 - conventional correlation detector, 485
 - capacity, 565
 - closed-loop power control, 563
 - corner effect, 564
 - error probability, 485
 - Gaussian approximation, 488
 - simplified approximation, 490
 - forward link capacity, 571
 - imperfect power control, 572
 - multiuser detection, 491
 - decorrelator detector, 494
 - MMSE detector, 494
 - optimum detector, 492
 - near-far effect, 4, 563
 - power control, 4, 564
 - forward link, 581
 - reverse link, 576
 - reverse link capacity, 565
 - Code performance, flat fading, 429
 - Coherence bandwidth, 68
 - Coherence time, 68
 - Coherent detection, 234
 - correlation detector, 235
 - MAP receiver, 234
 - matched filter detector, 236
 - maximum likelihood receiver, 234
 - quadrature demodulator, 235
 - complementary codes, 462
 - Continuous phase frequency shift keying (CPFSK), 190
 - power spectrum, 219
 - Continuous phase modulation (CPM), 188
 - CPFSK, 190
 - detection, 265
 - coherent detector, 266
 - non-coherent detector, 267
 - excess phase, 189
 - frequency shaping function, 189
 - full response, 189, 190
 - partial response, 189
 - GMSK, 195
 - Laurent's decomposition, 198
 - linearized GMSK, 198
 - modulation index, 189
 - MSK, 191
 - partial response, 192
 - phase states, 194
 - shaping functions, 192
 - phase shaping function, 189
 - phase tree, 190
 - phase trellis, 191
 - power spectrum, 215
 - TFM, 200
 - GTFM, 202
 - Convolutional codes, 407
 - BCJR algorithm, 416
 - encoder, 407
 - constraint length, 407
 - finite-state machine, 407
 - modified state diagram, 411
 - state, 410
 - transfer function, 411
 - generator polynomials, 409
 - generator sequences, 408
 - recursive systematic codes, 412
 - state diagram, 409
 - systematic, 409
 - total encoder memory, 410
 - trellis diagram, 409
 - union bound, 424
 - union-Chernoff bound, 426
 - Viterbi algorithm, 414
 - Convolutional interleaver, 429
 - Correlation functions, 65
 - COST 207 models, 98, 128
 - COST 259 models, 99, 130
 - Coverage, 23
 - CPM. *See* Continuous phase modulation (CPM)
 - Cross polarization discrimination, 70, 74
 - Cyclostationary random process, 204
- D**
- DCA. *See* Dynamic channel assignment (DCA)
 - Decision feedback equalizer, 352

- adaptive solution, 354
 - performance, 354
 - tap solution, 353
 - Delayed decision feedback sequence estimation (DDFSE), 360
 - Depolarization, 70
 - Differential detection, 258
 - binary DPSK, 258
 - n=4-DQPSK, 260
 - Differential encoding, 248
 - Digital Enhanced Cordless Telephone (DECT), 10
 - Digital modulation
 - CPFSK, 190
 - CPM, 188
 - GMSK, 195
 - linearized GMSK, 198
 - MSK, 191
 - multiresolution modulation, 188
 - n=4-DQPSK, 180
 - Nyquist pulse shaping, 173
 - OFDM, 185
 - ICI, 255
 - OQPSK, 180
 - orthogonal modulation, 182
 - power spectrum, 203
 - PSK, 179
 - QAM, 176
 - signal representation, 166
 - quadrature form, 166
 - signal correlation, 172
 - standard form, 166
 - complex envelope, 166
 - correlation, 173
 - envelope-phase form, 166
 - Euclidean distance, 173
 - generalized shaping function, 166
 - signal energy, 172
 - vector space representation, 167
 - TFM, 200
 - vector-space representation, 231
 - Gram-Schmidt orthonormalization, 167
 - Directional antennas, 529
 - Direct sequence (DS) spread spectrum, 450
 - basic receiver, 452
 - frequency-selective fading, 478
 - RAKE receiver, 481
 - tapped delay line model, 480
 - long code, 450
 - PN chip, 450
 - power spectrum, 464
 - processing gain, 450
 - short code, 450
 - design, 475
 - spreading waveform, 450
 - tone interference, 467
 - long code, 477
 - short code, 470
 - Distributed antenna systems, 301
 - Diversity
 - macrodiversity, 22
 - multipath diversity, 481
 - Diversity techniques, 273
 - distributed antenna systems, 301
 - diversity combining, 274
 - equal gain, 283
 - Granlund combiner, 280
 - maximal ratio, 277
 - optimum combining, 291
 - postdetection equal gain, 286
 - selective, 275
 - square-law combining, 289
 - switched, 285
 - optimum combining, performance, 295
 - transmit diversity, 310
 - types, 273
 - Doppler shift, 35
 - Doppler spectrum, 39
 - bandpass, 41
 - Doppler spreading, 37
 - Duplexer, 1
 - Dynamic channel assignment (DCA), 643
 - aggressive and timid, 651, 662
 - TAKE carrier, 663
 - carrier acquisition criterion, 658
 - carrier orderings, 659
 - carrier release criterion, 658
 - centralized, 644, 646
 - channel segregation (CS), 649
 - compact pattern based DCA (CPDCA), 655
 - CS with variable threshold, 649
 - decentralized, 644, 648
 - dynamic resource acquisition (DRA), 648
 - first available (FA), 648
 - fully decentralized, 644, 649
 - maximum packing (MP), 646
 - MAXMIN, 647
 - mean square (MSQ), 648
 - minimum interference (MI), 650
 - nearest neighbor (NN), 648
 - nearest neighbor+1 (NN+1), 648
 - random MI (RMI), 651
 - sequential MI, 651
 - simple DCA (SDCA), 659
- E**
- Envelope correlation, 37, 48
 - Envelope distribution, 44
 - Envelope fading, 33
 - Envelope phase, multipath-fading, 47
 - Envelope spectrum, 48
 - Equal gain combining (EGC), 283
 - Equalizers
 - sequence estimation, 355
 - symbol-by-symbol, 343
 - decision feedback equalizer, 352
 - minimum mean-square-error, 348
 - linear, 344
 - zero-forcing, 344
 - Erlang, 25
 - Erlang-B formula, 27
 - Erlang capacity, 28
 - Erlang-C formula, 31
 - Error probability, 237
 - biorthogonal signals, 253
 - bit vs. symbol error, 241
 - equally likely symbol errors, 241
 - gray coding, 241
 - differential PSK, 248
 - lower bounds, 240
 - MSK, 265
 - OFDM, 253
 - orthogonal signals, coherent detection, 251

- Error probability, 237 (*cont.*)
 - pairwise error probability, 238
 - PAM, 248
 - PSK, 243
 - Rayleigh fading, 246
 - QAM, 250
 - Rayleigh fading, 251
 - rotational invariance, 242
 - rotations and translations, 242
 - translational invariance, 242
 - union bound, 239
 - upper bounds, 239
- EV-DO, 7
- Events
 - mutually exclusive, 671
 - statistically independent, 671
- F**
- Fading, 33
- Fading simulators, 78
 - Clarke's model, 83
 - filtered Gaussian noise, 79
 - IDFT method, 79
 - IIR filtering method, 82
 - Jakes' model, 84
 - method of exact Doppler spreads, 87
 - mobile-to-mobile channels, 100
 - Akki & Haber model, 100
 - Patel & Stüber deterministic model, 102
 - Patel & Stüber statistical model, 102
 - Zajíc & Stüber statistical model, 103
 - modified Hoeher model, 88
 - multiple envelopes
 - Li & Huang Model, 90
 - Zajíc & Stüber deterministic Model, 90
 - Zajíc & Stüber statistical Model, 92
 - Zheng & Xiao model, 89
 - multiple faded envelopes, 89
 - sum of sinusoids method, 82
 - wide-band channels, 95
 - COST 207 models, 98
 - COST 259 models, 99
 - ITU models, 100
 - symbol-spaced model, 106
 - Zheng & Xiao model, 87
- Fixed channel assignment (FCA), 643
- Flat fading, 37
- Folded spectrum, 174
- Fractional reuse, 533
- Frequency hopped (FH) spread spectrum
 - slow frequency hopped, 453
 - fast frequency hopped, 453
- Frequency reuse, 13, 14
 - co-channel reuse distance, 14
 - co-channel reuse factor, 14
 - frequency plan, 14
 - interference neighborhood, 538, 648
 - microcells, 14
 - outage, 17
 - reuse cluster, 14
 - universal, 565
- Frequency selective fading, 37
 - correlation functions, 65
 - transmission functions, 63
 - coherence bandwidth, 68
 - coherence time, 68
 - scattering function, 69
- Frequency shift keying (FSK), 182
- Frequency spreading, 37
- G**
- Gaussian minimum shift keying (GMSK), 195
 - frequency shaping pulse, 197
 - Gaussian filter, 196
 - power spectrum, 222
- Generalized tamed frequency modulation, 202
- Global System for Mobile Communication (GSM), 3
- Gold sequences, 458
- Grade of service
 - definition, 645
 - forced termination, 645
 - new call blocking, 645
- Granlund combiner, 280
- Gray coding, 241
- H**
- Hadamard matrix, 183, 460
- Handoff algorithms, 593
 - velocity adaptive, performance, 613
 - backward, 594
 - direction biased, 597
 - forward, 594
 - hard, signal strength, 596
 - mobile assisted, 593, 596
 - mobile controlled, 593
 - network-controlled, 593
 - soft, C/I-based, 597
 - velocity adaptive, 611
- Handoffs, 22
 - analysis, co-channel interference, 615
 - gain, 22
 - handoff priority, 645, 657, 660
 - guard channels, 645, 657
 - handoff queueing, 657, 660
 - hard, 594
 - analysis, 614
 - Corner effect, 594
 - hysteresis, 594
 - intercell handoff, 593
 - intracell handoff, 593
 - signal strength averaging, 598
 - soft, 595
 - active set, 623
 - analysis, 618, 623
 - interference analysis, 619
 - power control, 595
 - velocity estimation, 601
- Handovers, 22
- Hard decision decoding, 425
- Hard handoff, 22
- High-speed packet access (HSPA), 8
- I**
- IEEE802.11, 11
- IEEE802.15, 13
- Interference loading, 19
- Interleaving, 427
 - bit interleaver, 427

- block Interleaver, 428
- convolutional Interleaver, 429
- random, 439
- S-random, 439
- symbol interleaver, 427
- International Telecommunications Union (ITU) models, 131
- Intersymbol interference (ISI) channels, 173
 - discrete-time channel model, 336
 - channel impulse response, 340
 - diversity reception, 340
 - minimum energy property, 337
 - minimum phase, 337
 - noise whitening filter, 337
 - discrete-time, white noise channel model, 339
 - fractionally-spaced receiver, 341
 - ISI channel modeling, 333
 - ISI coefficients, 335
 - optimum receiver, 335
 - quasi-static fading, 340
 - vector-space representation, 334
- IS-54/136, 4
- IS-95, 4
- Isotropic scattering, 39

- K**
- Kasami sequences, 459
- Kronecker product, 308

- L**
- Laurent's decomposition, 198
- Level crossing rate, 53
- Linearized GMSK, 198
- Link budget, 18
 - cell breathing, 19
 - handoff gain, 20
 - interference loading, 19
 - interference margin, 19
 - maximum path loss, 19
 - receiver sensitivity, 19
 - shadow margin, 20
- Link imbalance, 26
- Log-normal approximations
 - Farley's method, 152
 - Fenton-Wilkinson method, 149
 - Schwartz & Yeh method, 150

- M**
- Macrodiversity, 554
 - probability of outage, 554
 - shadow correlation, 555
 - massive MIMO, 318
 - favorable propagation, 318, 324
 - forward link data, 323
 - system model, 319
 - pilot contamination, 319
 - reverse link data, 321
 - reverse link pilots, 320
- Maximal ratio combining, 277
- Maximum likelihood sequence estimation (MLSE), 355
 - adaptive receiver, 359
 - branch metric, 357
 - error event, 363
 - error probability, 363
 - fading ISI channels, 367
 - pairwise error probability, 365
 - static ISI channels, 365
 - $T = 2$ -spaced receiver, 370
 - union bound, 364
- fractionally-spaced receiver, 360, 373
- likelihood function, 356
- LMS algorithm, 359
- log-likelihood function, 356
- MIMO receivers, IRC receiver, 387
- per survivor processing, 359
- RLS algorithm, 359
- state diagram, 355
- states, 355
- $T=2$ -spaced receiver
 - Practical receiver, 373
 - Timing phase sensitivity, 372
- Trellis diagram, 355
- Viterbi algorithm, 357
- Microcells, 14
 - highway microcells, 15
 - Manhattan microcells, 15
- Microcellular systems, overlay/underlay, micro area, 538
- MIMO. *See* Multiple-input multiple-output (MIMO) channels
- Minimum mean-square error (MMSE) equalizer, 348
 - adaptive solution, 350
 - performance, 351
 - tap solution, 349
- Minimum shift keying (MSK), 191
 - error probability, 265
 - OQASK equivalent, 191
 - power spectrum, 220
- Mobile-to-mobile channels, reference model, 62
- Modulation
 - bandwidth efficiency, 165
 - desirable properties, 165
- Moments
 - central moment, 673
 - characteristic function, 673
 - generating function, 673
 - variance, 673
- Multi-carrier, 501
- Multipath, 33
- Multipath-fading, 16
 - average fade duration, 56
 - Doppler spectrum, 39
 - flat, Rayleigh, 45
 - statistical characterization, 63
 - envelope correlation, 37, 48
 - envelope distribution, 44
 - flat fading, 37
 - frequency selective fading, 37
 - level crossing rate, 53
 - Nakagami fading, 46
 - phase distribution, 44
 - Ricean fading, 45
 - space-time correlation, 57
 - squared-envelope correlation, 51
 - zero crossing rate, 55
- Multipath propagation, 33
- Multiple-input multiple-output (MIMO) channels, 306
 - analytical models, 307, 308
 - co-channel demodulation, 375
 - detection, 313
 - inverse channel detector (ICD), 315
 - Kronecker model, 308

Multiple-input multiple-output (MIMO) channels, 306 (*cont.*)
 massive, 318
 maximum likelihood detection, 314
 minimum mean-square-error (MMSE) detector, 314
 physical models, 307
 precoding, 315
 Weichselberger model, 309

Multiresolution modulation, 188

Multiuser detection, 491
 decorrelator detector, 494
 MMSE detector, 494
 optimum detector, 492

N

Nakagami fading, 46
 shape factor, 47

Non-coherent detection, 261
 error probability, 263
 square-law detector, 262

Nyquist frequency, 174

Nyquist pulse shaping, 173
 folded spectrum, 174
 ideal Nyquist pulse, 174
 Nyquist first criterion, 174
 raised cosine, 175
 roll-off factor, 176
 root-raised cosine, 176

O

Offset (OQPSK), 180
 power spectrum, 210

Optimum combining, 291

Orthogonal frequency division multiple access (OFDMA), 517
 forward link, 518
 receiver, 521
 transmitter, 518
 frequency planning, 534
 reuse partitioning, 535

PAPR, 523

raised cosine windowing, 520

reverse link, 522

sub-carrier allocation, 520
 clustered carrier, 521
 random interleaving, 521
 spaced carrier, 521

time domain windowing, 518

Orthogonal frequency division multiplexing (OFDM), 185
 adaptive bit loading, 187
 channel estimation, 511
 complex envelope, 185
 cyclic suffix, 186
 error probability, 253
 interchannel interference, 255

FFT implementation, 186

guard interval, 502

ISI channels, 502

power spectrum, 211
 IDFT implementation, 212

residual ISI, 504
 cancellation, 508

Orthogonal modulation
 binary orthogonal codes, 183
 bi-orthogonal signals, 184
 FSK, 182
 orthogonal multipulse modulation, 184

Walsh-Hadamard sequences, 461

Orthogonal multipulse modulation, 184

Orthogonal signals, error probability
 coherent detection, 251
 non-coherent detection, 263

Outage, 17
 co-channel interference, 17
 thermal noise, 17

P

Pairwise error probability, 238

Parseval's theorem, 693

Path loss models, 16, 114
 CCIR model, 117
 COST231-Hata model, 119
 COST231-Walfish-Ikegami model, 120
 empirical models, 117
 flat Earth path loss, 115
 free space path loss, 114
 indoor microcells, 128
 Lee's area-to-area model, 118
 line-of-sight probability, 124
 mm-wave models, 127
 Okumura-Hata model, 117
 path loss exponent, 16
 street microcells, 122
 Corner effect, 123

3GPP 3-D path loss models, 123
 two-slope model, 122

Personal Digital Cellular (PDC), 5

Personal Handy-phone System (PHS), 10

Phase distribution, 44

Phase shift keying (PSK), 179
 error probability, 243
 power spectrum, 210

Polarization, 69
 conservation of polarization, 71
 depolarization, 70
 geometric model, 70
 cross polarization discrimination, 70

Power control, 595

Power delay profile, 67
 average delay, 67
 rms delay spread, 67

Power spectral densities, 693

Power spectrum, 203
 CPFSK, 219
 complex envelope, 204
 linear full response modulation, 207
 linear partial response modulation, 208

DS spread spectrum, 464

full response CPM, 215

GMSK, 222

MSK, 220

n=4-DQPSK, 211

OFDM, 211
 IDFT implementation, 212

OQPSK, 210

PSK, 210

QAM, 209

TFM, 222

Probability
 Bayes' theorem, 672
 cdf, 673
 cdfc, 673

- complementary error function, 676
- conditional, 671
- error function, 676
- pdf, 673
- total probability, 672
- Probability distributions
 - binomial, 675
 - central chi-square, 680
 - complex multivariate Gaussian, 677
 - exponential, 681
 - Gaussian, 676
 - geometric, 675
 - multivariate Gaussian, 677
 - non-central chi-square, 681
 - Poisson, 675
 - Rayleigh, 678
 - Rice, 679
- Pulse amplitude modulation (PAM), 178
 - constellations, 178
 - error probability, 248
- Pulse shaping, partial response, 209

- Q**
- Quadrature amplitude modulation (QAM), 176
 - error probability, 250
 - power spectrum, 209
 - signal constellations, 177

- R**
- Radio propagation
 - diffraction, 15
 - fixed-to-mobile channels, 35
 - MIMO channels, 306
 - mobile-to-mobile channels, 61
 - multipath fading, 16
 - path loss, 16
 - reflections, 15
 - scattering, 16
- Raised cosine pulse, 175
- RAKE receiver, 481
- Random processes, 685
 - autocorrelation, 687
 - autocovariance, 687
 - complex-valued, 692
 - covariance matrix, 690
 - crosscorrelation, 690
 - crosscovariance, 690
 - cyclostationary, 203, 696
 - discrete-time, 697
 - ergodic, 688
 - autocorrelation, 688
 - mean, 688
 - linear systems, 694
 - orthogonal, 691
 - statistically independent, 691
 - strictly stationary, 686
 - uncorrelated, 691
 - wide sense stationary, 687
- Rayleigh fading, 44
- Rayleigh quotient, 366
- Receiver sensitivity, 19
- Recursive systematic convolutional codes, 412
- Reduced state sequence estimation (RSSE)
 - subset-state, 362
 - subset transition, 362
 - subset trellis, 362
- Reuse partitioning, 533
 - cell splitting, 534
- Ricean fading, 45
 - Aulin's model, 45
 - Rice factor, 45
- Rice factor, 45
- Root-raised cosine pulse, 176
- RSSE. *See* Reduced state sequence estimation (RSSE)

- S**
- Scattering function, 69
- SC-FDE. *See* Single-carrier frequency domain equalization (SC-FDE)
- SC-FDMA. *See* Single-carrier frequency division multiple access (SC-FDMA)
- Selective combining, 275
- Sequence estimation
 - DDFSE, 360
 - MLSE, 355
 - RSSE, 362
- Shadowing, 16, 34, 110
 - area mean, 111
 - composite shadow-fading distributions, 112
 - gamma-log-normal, 113
 - local mean, 110
 - location area, 110
 - shadow standard deviation, 16
 - simulation, 111
- Signal and system correlation functions, 69
- Signal strength averaging, 598
 - sample averaging, 600
 - window length, 599
- Single-carrier frequency division multiple access (SC-FDMA), 523
 - frequency domain equalization, 525
 - PAPR, root-raised cosine filtering, 526
 - receiver, 523
 - sub-carrier allocation
 - interleaved, 523
 - localized, 524
 - transmitter, 523
- Single-carrier frequency domain equalization (SC-FDE), 513
 - MMSE, 517
 - zero forcing, 516
- Singleton bound, 400
- Soft decision decoding, 425
- Soft handoff, 4, 22
- Space-time block codes, 403
 - Alamouti code, 403
 - code rate, 403
 - complex orthogonal codes, 405
 - decoding orthogonal codes, 406
 - orthogonal codes, 403
 - real orthogonal codes, 404
- Space-time codes
 - design, 434
 - determinant criterion, 436
 - rank criterion, 436
 - Trellis codes, 436
- Space-time correlation, 57
 - base station, 59
 - mobile station, 57
- Space-time trellis codes, 436
 - decoder, 437
 - encoder description, 436
 - Viterbi algorithm, 438

- Spatial efficiency, 25
 - Spatial modulation, 316
 - Spectral efficiency, 24
 - bandwidth efficiency, 25
 - spatial efficiency, 25
 - trunking efficiency, 27
 - Spreading
 - dual-channel quaternary, 451
 - balanced quaternary, 451
 - complex, 450
 - simple binary, 451
 - Spreading sequences, 454
 - full period autocorrelation, 454
 - aperiodic autocorrelation, 454
 - full period cross-correlation, 454
 - partial period correlation, 454
 - Zadoff–Chu sequences, 462
 - Spreading waveforms, 455
 - autocorrelation, 455
 - Barker sequences, 459
 - complementary codes, 462
 - Gold sequences, 458
 - Kasami sequences, 459
 - m*-sequences, 456
 - variable length orthogonal codes, 461
 - Walsh-Hadamard sequences, 460
 - Spread spectrum, 449
 - Squared-envelope correlation, 51
 - Squared-envelope spectrum, 51
 - Square-law combining, 289
 - Standard array decoding, 402
 - Syndrome decoding, 403
- T**
- Tamed frequency modulation, 200
 - Tone interference, 467
 - Transmission functions, 63
 - delay Doppler-spread function, 65
 - impulse response, 63
 - output Doppler-spread function, 64
 - transfer function, 65
 - Transmit diversity, 310
 - Trellis-coded modulation (TCM), 420
 - asymptotic coding gain, 424
 - design rules, 432
 - encoder, 420
 - mapping by set partitioning, 421
 - pairwise error probability, 424
 - partition chain, 421
 - performance AWGN channel, 423
 - power spectrum, 222
 - symbol interleaving, 429
 - union bound, 424
 - Trellis coding
 - minimum built-in time diversity, 431
 - minimum product squared Euclidean distance, 431
 - Trunking efficiency, 27
 - Erlang-B formula, 27
 - grade of service, 643
 - Turbo codes, 438
 - error floor, 443
 - parallel decoder, 440
 - parallel encoder, 439
 - serial encoder, 441
 - uniform interleaver, 443
 - weight distribution, 442
 - parallel codes, 443
 - serial codes, 445
 - spectral thinning, 442
- U**
- Uncorrelated scattering channel, 66
 - Universal Mobile Telecommunications System (UMTS), 6, 7
 - Upper bounds
 - Chebyshev bound, 682
 - Chernoff bound, 426, 683
 - union-Chernoff bound, 427
- V**
- Variable length orthogonal codes, 461
 - Vector-space representation, 231
 - remainder process, 232
 - sufficient statistics, 234
 - Velocity estimation, 601
 - level crossing rate, 603
 - covariance method, 606
 - envelope, 603
 - zero crossing rate, 603
 - sensitivity, 607
 - Gaussian noise, 609
 - sampling density, 612
 - scattering distribution, 608
 - Viterbi algorithm, 357, 414
 - branch metric, 415
 - path metrics, 357, 414
 - surviving sequences, 357
 - survivors, 414
- W**
- Walsh-Hadamard sequences, 460
 - orthogonal CDMA, 461
 - orthogonal modulation, 461
 - Wideband Code Division Multiple Access (WCDMA), 6
 - Wide sense stationary channel, 66
 - Wide sense stationary uncorrelated scattering channel, 67
 - Wireless systems and standards, 1
 - analog cellular systems, 1
 - cordless telephones
 - DECT, 10
 - PHS, 10
 - fifth generation cellular systems, 10
 - fourth generation cellular systems, 8
 - LTE, 9
 - WiMAX, 9
 - second generation cellular systems
 - GSM, 3
 - IS-54/136, 4
 - IS-95, 4
 - Personal Digital Cellular, 5
 - systems and standards, 2
 - third generation cellular systems, 5
 - cdma2000, 6, 7
 - EV-DO, 7
 - HSPA, 8
 - UMTS, 6, 7
 - W-CDMA, 6
 - wireless LANS, 11
 - IEEE802.11, 11
 - and PANs, Bluetooth, 13

and PANs, IEEE802.15, 13

wireless PANs, IEEE802.15, 13

Worldwide Interoperability for Microwave Access (WiMAX), 9

Z

Zadoff-Chu sequences, 462

Zero crossing rate, 55

Zero-forcing equalizer, 344

adaptive solution, 346

data mode, 346

performance, 346

tap solution, 345

training mode, 346