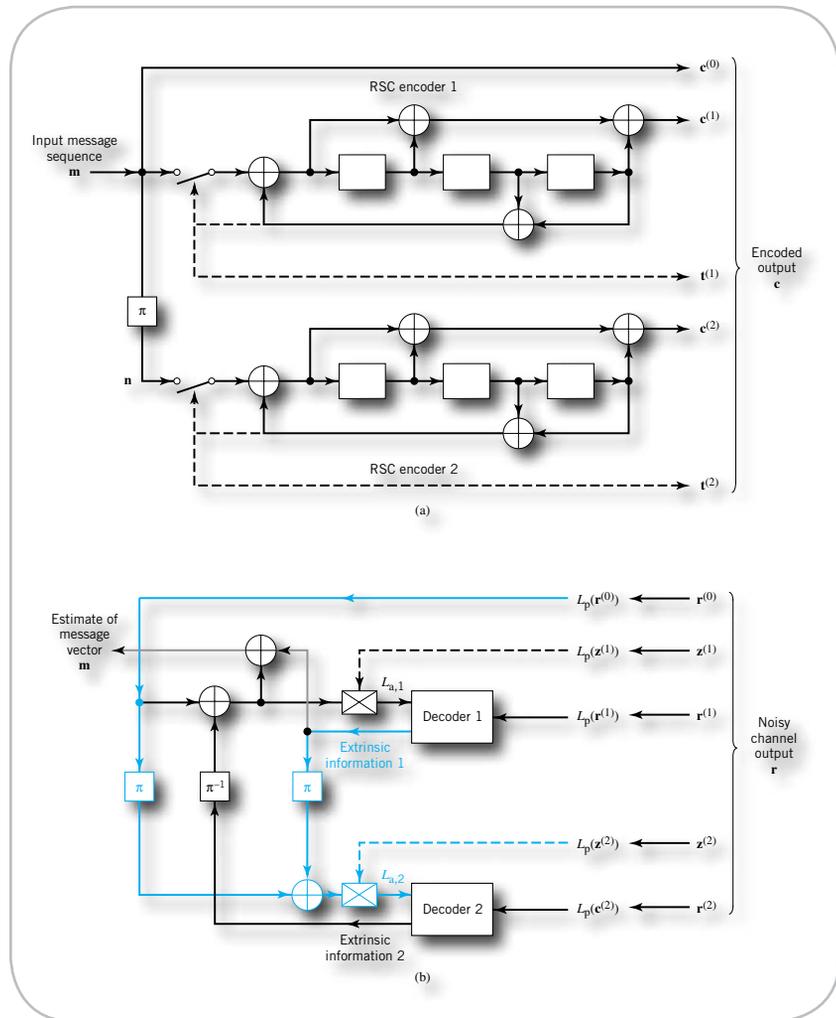


DIGITAL COMMUNICATION SYSTEMS



Simon Haykin

DIGITAL
COMMUNICATION
SYSTEMS

Simon Haykin
McMaster University

WILEY

ASSOCIATE PUBLISHER	Daniel Sayre
EDITORIAL ASSISTANT	Jessica Knecht
MARKETING MANAGER	Christopher Ruel
PRODUCTION MANAGEMENT SERVICES	Publishing Services
CREATIVE DIRECTOR	Harry Nolan
COVER DESIGNER	Kristine Carney

Cover Image: The figure on the cover, depicting the UMTS-turbo code, is adapted from the doctoral thesis of Dr. Liang Li, Department of Electronics and Computer Science, University of Southampton, United Kingdom, with the permission of Dr. Li, his Supervisor Dr. Robert Maunder, and Professor Lajos Hanzo; the figure also appears on page 654 of the book.

This book was set in Times by Publishing Services and printed and bound by RRD Von Hoffmann. The cover was printed by RRD Von Hoffmann.

This book is printed on acid free paper. ∞

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship.

Copyright © 2014 John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, website www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201) 748-6011, fax (201) 748-6008, website www.wiley.com/go/permissions.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return mailing label are available at www.wiley.com/go/returnlabel. If you have chosen to adopt this textbook for use in your course, please accept this book as your complimentary desk copy. Outside of the United States, please contact your local sales representative.

ISBN: 978-0-471-64735-5

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

*In loving memory of
Vera*

Preface

The study of digital communications is an essential element of the undergraduate and postgraduate levels of present-day electrical and computer engineering programs. This book is appropriate for both levels.

A Tour of the Book

The introductory chapter is motivational, beginning with a brief history of digital communications, and continuing with sections on the communication process, digital communications, multiple-access and multiplexing techniques, and the Internet. Four themes organize the remaining nine chapters of the book.

Theme 1 *Mathematics of Digital Communications*

The first theme of the book provides a detailed exposé of the mathematical underpinnings of digital communications, with *continuous mathematics* aimed at the communication channel and interfering signals, and *discrete mathematics* aimed at the transmitter and receiver:

- Chapter 2, *Fourier Analysis of Signals and Systems*, lays down the fundamentals for the representation of signals and linear time-invariant systems, as well as analog modulation theory.
- Chapter 3, *Probability Theory and Bayesian Inference*, presents the underlying mathematics for dealing with uncertainty and the Bayesian paradigm for probabilistic reasoning.
- Chapter 4, *Stochastic Processes*, focuses on weakly or wide-sense stationary processes, their statistical properties, and their roles in formulating models for Poisson, Gaussian, Rayleigh, and Rician distributions.
- Chapter 5, *Information Theory*, presents the notions of entropy and mutual information for discrete as well continuous random variables, leading to Shannon's celebrated theorems on source coding, channel coding, and information capacity, as well as rate-distortion theory.

Theme 2 *From Analog to Digital Communications*

The second theme of the book, covered in Chapter 6, describes how analog waveforms are transformed into coded pulses. It addresses the challenge of performing the transformation with robustness, bandwidth preservation, or minimal computational complexity.

Theme 3 *Signaling Techniques*

Three chapters address the third theme, each focusing on a specific form of *channel impairment*:

- In Chapter 7, *Signaling over Additive White Gaussian Noise (AWGN) Channels*, the impairment is the unavoidable presence of *channel noise*, which is modeled as

additive white Gaussian noise (AWGN). This model is well-suited for the *signal-space diagram*, which brings insight into the study of phase-shift keying (PSK), quadrature-amplitude modulation (QAM), and frequency-shift keying (FSK) as different ways of accommodating the transmission and reception of binary data.

- In Chapter 8, *Signaling over Band-Limited Channels*, bandwidth limitation assumes center stage, with *intersymbol interference* (ISI) as the source of channel impairment.
- Chapter 9, *Signaling over Fading Channels*, focuses on *fading channels* in wireless communications and the practical challenges they present. The channel impairment here is attributed to the *multipath phenomenon*, so called because the transmitted signal reaches the receiver via a multiplicity of paths.

Theme 4 *Error-control Coding*

Chapter 10 addresses the practical issue of *reliable communications*. To this end, various techniques of the feedforward variety are derived therein, so as to satisfy Shannon's celebrated *coding theorem*.

Two families of error-correcting codes are studied in the chapter:

- *Legacy (classic) codes*, which embody linear block codes, cyclic codes, and convolutional codes. Although different in their structural compositions, they look to algebraic mathematics as the procedure for approaching the *Shannon limit*.
- *Probabilistic compound codes*, which embody turbo codes and low-density parity-check (LDPC) codes. What is remarkable about these two codes is that they both approach the Shannon limit with doable computational complexity in a way that was not feasible until 1993. The trick behind this powerful information-processing capability is the adoption of *random codes*, the origin of which could be traced to Shannon's 1948 classic paper.

Features of the Book

Feature 1 *Analog in Digital Communication*

When we think of digital communications, we must not overlook the fact that such a system is of a *hybrid nature*. The channel across which data are transmitted is analog, exemplified by traditional telephone and wireless channels, and many of the sources responsible for the generation of data (e.g., speech and video) are of an analog kind. Moreover, certain principles of analog modulation theory, namely double sideband-suppressed carrier (DSB-SC) and vestigial sideband (VSB) modulation schemes, include binary phase-shift keying (PSK) and offset QPSK as special cases, respectively.

It is with these points in mind that Chapter 2 includes

- detailed discussion of communication channels as examples of linear systems,
- analog modulation theory, and
- phase and group delays.

Feature 2 *Hilbert Transform*

The Hilbert transform, discussed in Chapter 2, plays a key role in the complex representation of signals and systems, whereby

- a band-pass signal, formulated around a sinusoidal carrier, is transformed into an equivalent complex low-pass signal;

- a band-pass system, be it a linear channel or filter with a midband frequency, is transformed into an equivalent complex low-pass system.

Both transformations are performed without loss of information, and their use changes a difficult task into a much simpler one in mathematical terms, suitable for simulation on a computer. However, one must accommodate the use of complex variables.

The Hilbert transform also plays a key role in Chapter 7. In formulating the *method of orthogonal modulation*, we show that one can derive the well-known formulas for the noncoherent detection of binary frequency-shift keying (FSK) and differential phase-shift keying (DPSK) signals, given unknown phase, in a much simpler manner than following traditional approaches that involve the use of Rician distribution.

Feature 3 *Discrete-time Signal Processing*

In Chapter 2, we briefly review *finite-direction impulse response (FIR)* or *tapped-delay line (TDL) filters*, followed by the *discrete Fourier transform (DFT)* and a well-known *fast Fourier transform (FFT)* algorithm for its computational implementations. FIR filters and FFT algorithms feature prominently in:

- Modeling of the *raised-cosine spectrum (RCS)* and its square-root version (SQRCS), which are used in Chapter 8 to mitigate the ISI in band-limited channels;
- Implementing the *Jakes model* for fast fading channels, demonstrated in Chapter 9;
- Using FIR filtering to simplify the mathematical exposition of the most difficult form of channel fading, namely, the *doubly spread channel* (in Chapter 9).

Another topic of importance in discrete-time signal processing is *linear adaptive filtering*, which appears:

- In Chapter 6, dealing with *differential pulse-code modulation (DPCM)*, where an adaptive predictor constitutes a key functional block in both the transmitter and receiver. The motivation here is to preserve channel bandwidth at the expense of increased computational complexity. The algorithm described therein is the widely used *least mean-square (LMS) algorithm*.
- In Chapter 7, dealing with the need for *synchronizing* the receiver to the transmitter, where two algorithms are described, one for recursive estimation of the group delay (essential for timing recovery) and the other for recursive estimation of the unknown carrier phase (essential for carrier recovery). Both algorithms build on the LMS principle so as to maintain linear computational complexity.

Feature 4 *Digital Subscriber Lines*

Digital subscriber lines (DSLs), covered in Chapter 8, have established themselves as an essential tool for transforming a linear wideband channel, exemplified by the twisted-wire pair, into a *discrete multitone (DMT) channel* that is capable of accommodating data transmission at multiple megabits per second. Moreover, the transformation is afforded practical reality by exploiting the FFT algorithm, with the inverse FFT used in the transmitter and the FFT used in the receiver.

Feature 5 *Diversity Techniques*

As already mentioned, the wireless channel is one of the most challenging media for digital communications. The difficulty of reliable data transmission over a wireless

channel is attributed to the multipath phenomenon. Three diversity techniques developed to get around this practical difficulty are covered in Chapter 9:

- *Diversity on receive*, the traditional approach, whereby an array of multiple antennas operating independently is deployed at the receiving end of a wireless channel.
- *Diversity on transmit*, which operates by deploying two or more independent antennas at the transmit end of the wireless channel.
- *Multiple-input multiple-output (MIMO) channels*, where multiple antennas (again operating independently) are deployed at both ends of the wireless channel.

Among these three forms of diversity, the MIMO channel is naturally the most powerful in information-theoretic terms: an advantage gained at the expense of increased computational complexity.

Feature 6 *Turbo Codes*

Error-control coding has established itself as the most commonly used technique for *reliable* data transmission over a noisy channel. Among the challenging legacies bestowed by Claude Shannon was how to design a code that would closely approach the so-called *Shannon limit*. For over four decades, increasingly more powerful coding algorithms were described in the literature; however it was the *turbo code* that had the honor of closely approaching the Shannon limit, and doing so in a computationally feasible manner.

Turbo codes, together with the associated *maximum a posteriori (MAP) decoding algorithm*, occupy a large portion of Chapter 10, which also includes:

- Detailed derivation of the MAP algorithm and an illustrative example of how it operates;
- The *extrinsic information transfer (EXIT) chart*, which provides an experimental tool for the design of turbo codes;
- *Turbo equalization*, for demonstrating applicability of the turbo principle beyond error-control coding.

Feature 7 *Placement of Information Theory*

Typically, information theory is placed just before the chapter on error-control coding. In this book, it is introduced early because:

Information theory is not only of basic importance to error-control coding but also other topics in digital communications.

To elaborate:

- Chapter 6 presents the relevance of *source coding* to pulse-code modulation (PCM), differential pulse-code modulation (DPCM), and delta modulation.
- Comparative evaluation of M -ary PSK versus M -ary FSK, done in Chapter 7, requires knowledge of *Shannon's information capacity law*.
- Analysis and design of *DSL*, presented in Chapter 8, also builds on Shannon's information capacity law.
- *Channel capacity* in Shannon's coding theorem is important to diversity techniques, particularly of the MIMO kind, discussed in Chapter 9.

Examples, Computer Experiments, and Problems

Except for Chapter 1, each of the remaining nine chapters offers the following:

- Illustrative examples are included to strengthen the understanding of a theorem or topic in as much detail as possible. Some of the examples are in the form of computer experiments.
- An extensive list of end-of-chapter problems are grouped by section to fit the material covered in each chapter. The problems range from relatively easy ones all the way to more challenging ones.
- In addition to the computer-oriented examples, nine computer-oriented experiments are included in the end-of-chapter problems.

The Matlab codes for all of the computer-oriented examples in the text, as well as other calculations performed on the computer, are available at www.wiley.com/college/haykin.

Appendices

Eleven appendices broaden the scope of the theoretical as well as practical material covered in the book:

- Appendix A, *Advanced Probabilistic Models*, covers the chi-square distribution, log-normal distribution, and Nakagami distribution that includes the Rayleigh distribution as a special case and is somewhat similar to the Rician distribution. Moreover, an experiment is included therein that demonstrates, in a step-by-step manner, how the Nakagami distribution evolves into the log-normal distribution in an approximate manner, demonstrating its adaptive capability.
- Appendix B develops tight bounds on the *Q-function*.
- Appendix C discusses the ordinary Bessel function and its modified form.
- Appendix D describes the method of Lagrange multipliers for solving constrained optimization problems.
- Appendix E derives the formula for the *channel capacity of the MIMO channel* under two scenarios: one that assumes no knowledge of the channel by the transmitter, and the other that assumes this knowledge is available to the transmitter via a narrowband feedback link.
- Appendix F discusses the idea of *interleaving*, which is needed for dealing with bursts of interfering signals experienced in wireless communications.
- Appendix G addresses the *peak-to-average power reduction (PAPR) problem*, which arises in the use of orthogonal frequency-division multiplexing (OFDM) for both wireless and DSL applications.
- Appendix H discusses *solid-state nonlinear power amplifiers*, which play a critical role in the limited life of batteries in wireless communications.
- Appendix I presents a short exposé of *Monte Carlo integration*: a theorem that deals with mathematically intractable problems.
- Appendix J studies *maximal-length sequences*, also called *m-sequences*, which are used for implementing linear feedback shift registers (LFSRs). An important application of maximal-length sequences (viewed as pseudo-random noise) is in

designing direct-sequence spread-spectrum communications for code-division multiple access (CDMA).

- Finally, Appendix K provides a useful list of *mathematical formulas and functions*.

Two Noteworthy Symbols

Typically, the square-root of minus one is denoted by the italic symbol j , and the differential operator (used in differentiation as well as integration) is denoted by the italic symbol d . In reality, however, both of these terms are *operators*, each one in its own way: it is therefore incorrect to use italic symbols for their notations. Furthermore, italic j and italic d are also frequently used as indices or to represent other matters, thereby raising the potential for confusion. According, throughout the book, *roman j* and *roman d* are used to denote the square root of minus one and the differential operator, respectively.

Concluding Remarks

In writing this book every effort has been made to present the material in the manner easiest to read so as to enhance understanding of the topics covered. Moreover, cross-references within a chapter as well as from chapter to chapter have been included wherever the need calls for it.

Finally, every effort has been made by the author as well as compositor of the book to make it as error-free as humanly possible. In this context, the author would welcome receiving notice of any errors discovered after publication of the book.

Acknowledgements

In writing this book I have benefited enormously from technical input, persistent support, and permissions provided by many.

I am grateful to colleagues around the world for technical inputs that have made a significant difference in the book; in alphabetical order, they are:

- Dr. Daniel Costello, Jr., *University of Notre Dame*, for reading and providing useful comments on the maximum likelihood decoding and maximum a posteriori decoding materials in Chapter 10.
- Dr. Dimitri Bertsekas, *MIT*, for permission to use Table 3.1 on the Q -function in Chapter 3, taken from his co-authored book on the theory of probability.
- Dr. Lajos Hanzo, *University of Southampton*, UK, for many useful comments on turbo codes as well as low-density parity-check codes in Chapter 10. I am also indebted to him for putting me in touch with his colleagues at the University of Southampton, Dr. R. G. Maunder and Dr. L. Li, who were extremely helpfully in performing the insightful computer experiments on UMTS-turbo codes and EXIT charts in Chapter 10.
- Dr. Phillip Regalia, *Catholic University*, Washington DC, for contributing a section on serial-concatenated turbo codes in Chapter 10. This section has been edited by myself to follow the book's writing style, and for its inclusion I take full responsibility.
- Dr. Sam Shanmugan, *University of Kansas*, for his insightful inputs on the use of FIR filters and FFT algorithms for modeling the raised-cosine spectrum (RCS) and

its square-root version (SQ RCS) in Chapter 8, implementing the Jakes model in Chapter 9, as well as other simulation-oriented issues.

- Dr. Yanbo Xue, *University of Alberta*, Canada, for performing computer-oriented experiments and many other graphical computations throughout the book, using well-developed Matlab codes.
- Dr. Q. T. Zhang, *The City University of Hong Kong*, for reading through an early version of the manuscript and offering many valuable suggestions for improving it. I am also grateful to his student, Jiayi Chen, for performing the graphical computations on the Nakagami distribution in Appendix A.

I'd also like to thank the reviewers who read drafts of the manuscript and provided valuable commentary:

- Ender Ayanoglu, *University of California, Irvine*
- Tolga M. Duman, *Arizona State University*
- Bruce A. Harvey, *Florida State University*
- Bing W. Kwan, *FAMU-FSU College of Engineering*
- Chung-Chieh Lee, *Northwestern University*
- Heung-No Lee, *University of Pittsburgh*
- Michael Rice, *Brigham Young University*
- James Ritcey, *University of Washington*
- Lei Wei, *University of Central Florida*

Production of the book would not have been possible without the following:

- Daniel Sayre, Associate Publisher at John Wiley & Sons, who maintained not only his faith in this book but also provided sustained support for it over the past few years. I am deeply indebted to Dan for what he has done to make this book a reality.
- Cindy Johnson, Publishing Services, Newburyport, MA, for her dedicated commitment to the beautiful layout and composition of the book. I am grateful for her tireless efforts to print the book in as errorless manner as humanly possible.

I salute everyone, and others too many to list, for their individual and collective contributions, without which this book would not have been a reality.

Simon Haykin
Ancaster, Ontario
Canada
December, 2012

Contents

- 1 Introduction 1**
 - 1.1 Historical Background 1
 - 1.2 The Communication Process 2
 - 1.3 Multiple-Access Techniques 4
 - 1.4 Networks 6
 - 1.5 Digital Communications 9
 - 1.6 Organization of the Book 11

- 2 Fourier Analysis of Signals and Systems 13**
 - 2.1 Introduction 13
 - 2.2 The Fourier Series 13
 - 2.3 The Fourier Transform 16
 - 2.4 The Inverse Relationship between Time-Domain and Frequency-Domain Representations 25
 - 2.5 The Dirac Delta Function 28
 - 2.6 Fourier Transforms of Periodic Signals 34
 - 2.7 Transmission of Signals through Linear Time-Invariant Systems 37
 - 2.8 Hilbert Transform 42
 - 2.9 Pre-envelopes 45
 - 2.10 Complex Envelopes of Band-Pass Signals 47
 - 2.11 Canonical Representation of Band-Pass Signals 49
 - 2.12 Complex Low-Pass Representations of Band-Pass Systems 52
 - 2.13 Putting the Complex Representations of Band-Pass Signals and Systems All Together 54
 - 2.14 Linear Modulation Theory 58
 - 2.15 Phase and Group Delays 66
 - 2.16 Numerical Computation of the Fourier Transform 69
 - 2.17 Summary and Discussion 78

- 3 Probability Theory and Bayesian Inference 87**
 - 3.1 Introduction 87
 - 3.2 Set Theory 88
 - 3.3 Probability Theory 90
 - 3.4 Random Variables 97
 - 3.5 Distribution Functions 98
 - 3.6 The Concept of Expectation 105

3.7	Second-Order Statistical Averages	108
3.8	Characteristic Function	111
3.9	The Gaussian Distribution	113
3.10	The Central Limit Theorem	118
3.11	Bayesian Inference	119
3.12	Parameter Estimation	122
3.13	Hypothesis Testing	126
3.14	Composite Hypothesis Testing	132
3.15	Summary and Discussion	133
4	Stochastic Processes	145
4.1	Introduction	145
4.2	Mathematical Definition of a Stochastic Process	145
4.3	Two Classes of Stochastic Processes: Strictly Stationary and Weakly Stationary	147
4.4	Mean, Correlation, and Covariance Functions of Weakly Stationary Processes	149
4.5	Ergodic Processes	157
4.6	Transmission of a Weakly Stationary Process through a Linear Time-invariant Filter	158
4.7	Power Spectral Density of a Weakly Stationary Process	160
4.8	Another Definition of the Power Spectral Density	170
4.9	Cross-spectral Densities	172
4.10	The Poisson Process	174
4.11	The Gaussian Process	176
4.12	Noise	179
4.13	Narrowband Noise	183
4.14	Sine Wave Plus Narrowband Noise	193
4.15	Summary and Discussion	195
5	Information Theory	207
5.1	Introduction	207
5.2	Entropy	207
5.3	Source-coding Theorem	214
5.4	Lossless Data Compression Algorithms	215
5.5	Discrete Memoryless Channels	223
5.6	Mutual Information	226
5.7	Channel Capacity	230
5.8	Channel-coding Theorem	232
5.9	Differential Entropy and Mutual Information for Continuous Random Ensembles	237

5.10	Information Capacity Law	240
5.11	Implications of the Information Capacity Law	244
5.12	Information Capacity of Colored Noisy Channel	248
5.13	Rate Distortion Theory	253
5.14	Summary and Discussion	256
6	Conversion of Analog Waveforms into Coded Pulses	267
6.1	Introduction	267
6.2	Sampling Theory	268
6.3	Pulse-Amplitude Modulation	274
6.4	Quantization and its Statistical Characterization	278
6.5	Pulse-Code Modulation	285
6.6	Noise Considerations in PCM Systems	290
6.7	Prediction-Error Filtering for Redundancy Reduction	294
6.8	Differential Pulse-Code Modulation	301
6.9	Delta Modulation	305
6.10	Line Codes	309
6.11	Summary and Discussion	312
7	Signaling over AWGN Channels	323
7.1	Introduction	323
7.2	Geometric Representation of Signals	324
7.3	Conversion of the Continuous AWGN Channel into a Vector Channel	332
7.4	Optimum Receivers Using Coherent Detection	337
7.5	Probability of Error	344
7.6	Phase-Shift Keying Techniques Using Coherent Detection	352
7.7	M -ary Quadrature Amplitude Modulation	370
7.8	Frequency-Shift Keying Techniques Using Coherent Detection	375
7.9	Comparison of M -ary PSK and M -ary FSK from an Information-Theoretic Viewpoint	398
7.10	Detection of Signals with Unknown Phase	400
7.11	Noncoherent Orthogonal Modulation Techniques	404
7.12	Binary Frequency-Shift Keying Using Noncoherent Detection	410
7.13	Differential Phase-Shift Keying	411
7.14	BER Comparison of Signaling Schemes over AWGN Channels	415
7.15	Synchronization	418
7.16	Recursive Maximum Likelihood Estimation for Synchronization	419
7.17	Summary and Discussion	431

8	Signaling over Band-Limited Channels	445
8.1	Introduction	445
8.2	Error Rate Due to Channel Noise in a Matched-Filter Receiver	446
8.3	Intersymbol Interference	447
8.4	Signal Design for Zero ISI	450
8.5	Ideal Nyquist Pulse for Distortionless Baseband Data Transmission	450
8.6	Raised-Cosine Spectrum	454
8.7	Square-Root Raised-Cosine Spectrum	458
8.8	Post-Processing Techniques: The Eye Pattern	463
8.9	Adaptive Equalization	469
8.10	Broadband Backbone Data Network: Signaling over Multiple Baseband Channels	474
8.11	Digital Subscriber Lines	475
8.12	Capacity of AWGN Channel Revisited	477
8.13	Partitioning Continuous-Time Channel into a Set of Subchannels	478
8.14	Water-Filling Interpretation of the Constrained Optimization Problem	484
8.15	DMT System Using Discrete Fourier Transform	487
8.16	Summary and Discussion	494
9	Signaling over Fading Channels	501
9.1	Introduction	501
9.2	Propagation Effects	502
9.3	Jakes Model	506
9.4	Statistical Characterization of Wideband Wireless Channels	511
9.5	FIR Modeling of Doubly Spread Channels	520
9.6	Comparison of Modulation Schemes: Effects of Flat Fading	525
9.7	Diversity Techniques	527
9.8	“Space Diversity-on-Receive” Systems	528
9.9	“Space Diversity-on-Transmit” Systems	538
9.10	“Multiple-Input, Multiple-Output” Systems: Basic Considerations	546
9.11	MIMO Capacity for Channel Known at the Receiver	551
9.12	Orthogonal Frequency Division Multiplexing	556
9.13	Spread Spectrum Signals	557
9.14	Code-Division Multiple Access	560
9.15	The RAKE Receiver and Multipath Diversity	564
9.16	Summary and Discussion	566

- 10 Error-Control Coding 577**
 - 10.1 Introduction 577
 - 10.2 Error Control Using Forward Error Correction 578
 - 10.3 Discrete Memoryless Channels 579
 - 10.4 Linear Block Codes 582
 - 10.5 Cyclic Codes 593
 - 10.6 Convolutional Codes 605
 - 10.7 Optimum Decoding of Convolutional Codes 613
 - 10.8 Maximum Likelihood Decoding of Convolutional Codes 614
 - 10.9 Maximum a Posteriori Probability Decoding of Convolutional Codes 623
 - 10.10 Illustrative Procedure for Map Decoding in the Log-Domain 638
 - 10.11 New Generation of Probabilistic Compound Codes 644
 - 10.12 Turbo Codes 645
 - 10.13 EXIT Charts 657
 - 10.14 Low-Density Parity-Check Codes 666
 - 10.15 Trellis-Coded Modulation 675
 - 10.16 Turbo Decoding of Serial Concatenated Codes 681
 - 10.17 Summary and Discussion 688

Appendices

- A Advanced Probabilistic Models A1**
 - A.1 The Chi-Square Distribution A1
 - A.2 The Log-Normal Distribution A3
 - A.3 The Nakagami Distribution A6
- B Bounds on the Q-Function A11**
- C Bessel Functions A13**
 - C.1 Series Solution of Bessel's Equation A13
 - C.2 Properties of the Bessel Function A14
 - C.3 Modified Bessel Function A16
- D Method of Lagrange Multipliers A19**
 - D.1 Optimization Involving a Single Equality Constraint A19
- E Information Capacity of MIMO Channels A21**
 - E.1 Log-Det Capacity Formula of MIMO Channels A21
 - E.2 MIMO Capacity for Channel Known at the Transmitter A24
- F Interleaving A29**
 - F.1 Block Interleaving A30
 - F.2 Convolutional Interleaving A32
 - F.3 Random Interleaving A33

G	The Peak-Power Reduction Problem in OFDM	A35
G.1	PAPR Properties of OFDM Signals	A35
G.2	Maximum PAPR in OFDM Using M -ary PSK	A36
G.3	Clipping-Filtering: A Technique for PAPR Reduction	A37
H	Nonlinear Solid-State Power Amplifiers	A39
H.1	Power Amplifier Nonlinearities	A39
H.2	Nonlinear Modeling of Band-Pass Power Amplifiers	A42
I	Monte Carlo Integration	A45
J	Maximal-Length Sequences	A47
J.1	Properties of Maximal-Length Sequences	A47
J.2	Choosing a Maximal-Length Sequence	A50
K	Mathematical Tables	A55
	Glossary	G1
	Bibliography	B1
	Index	I1
	Credits	C1

Introduction

1.1 Historical Background

In order to provide a sense of motivation, this introductory treatment of digital communications begins with a historical background of the subject, brief but succinct as it may be. In this first section of the introductory chapter we present some historical notes that identify the pioneering contributors to digital communications specifically, focusing on three important topics: information theory and coding, the Internet, and wireless communications. In their individual ways, these three topics have impacted digital communications in revolutionary ways.

Information Theory and Coding

In 1948, the theoretical foundations of digital communications were laid down by Claude Shannon in a paper entitled “A mathematical theory of communication.” Shannon’s paper was received with immediate and enthusiastic acclaim. It was perhaps this response that emboldened Shannon to amend the title of his classic paper to “The mathematical theory of communication” when it was reprinted later in a book co-authored with Warren Weaver. It is noteworthy that, prior to the publication of Shannon’s 1948 classic paper, it was believed that increasing the rate of transmission over a channel would increase the probability of error; the communication theory community was taken by surprise when Shannon proved that this was not true, provided the transmission rate was below the channel capacity.

Shannon’s 1948 paper was followed by three ground-breaking advances in coding theory, which include the following:

1. Development of the first nontrivial error-correcting code by Golay in 1949 and Hamming in 1950.
2. Development of turbo codes by Berrou, Glavieux and Thitimjshima in 1993; turbo codes provide near-optimum error-correcting coding and decoding performance in additive white Gaussian noise.
3. Rediscovery of *low-density parity-check (LDPC) codes*, which were first described by Gallager in 1962; the rediscovery occurred in 1981 when Tanner provided a new interpretation of LDPC codes from a graphical perspective. Most importantly, it was the discovery of turbo codes in 1993 that reignited interest in LDPC codes.

The Internet

From 1950 to 1970, various studies were made on computer networks. However, the most significant of them all in terms of impact on computer communications was the Advanced Research Project Agency Network (ARPANET), which was put into service in 1971. The development of ARPANET was sponsored by the Advanced Research Projects Agency (ARPA) of the United States Department of Defense. The pioneering work in *packet switching* was done on the ARPANET. In 1985, ARPANET was renamed the *Internet*. However, the turning point in the evolution of the Internet occurred in 1990 when Berners-Lee proposed a hypermedia software interface to the Internet, which he named the *World Wide Web*. Thereupon, in the space of only about 2 years, the Web went from nonexistence to worldwide popularity, culminating in its commercialization in 1994. The Internet has dramatically changed the way in which we communicate on a daily basis, using a wired network.

Wireless Communications

In 1864, James Clerk Maxwell formulated the *electromagnetic theory of light* and predicted the existence of radio waves; the set of four equations that connect electric and magnetic quantities bears his name. Later on in 1887, Heinrich Herz demonstrated the existence of radio waves experimentally.

However, it was on December 12, 1901, that Guglielmo Marconi received a radio signal at Signal Hill in Newfoundland; the radio signal had originated in Cornwall, England, 2100 miles away across the Atlantic. Last but by no means least, in the early days of wireless communications, it was Fessenden, a self-educated academic, who in 1906 made history by conducting the first radio broadcast, transmitting music and voice using a technique that came to be known as *amplitude modulation (AM) radio*.

In 1988, the first digital cellular system was introduced in Europe; it was known as the *Global System for Mobile (GSM) Communications*. Originally, GSM was intended to provide a pan-European standard to replace the myriad of incompatible analog wireless communication systems. The introduction of GSM was soon followed by the North American IS-54 digital standard. As with the Internet, wireless communication has also dramatically changed the way we communicate on a daily basis.

What we have just described under the three headings, namely, information theory and coding, the Internet, and wireless communications, have collectively not only made communications essentially digital, but have also changed the world of communications and made it global.

1.2 The Communication Process

Today, *communication* enters our daily lives in so many different ways that it is very easy to overlook the multitude of its facets. The telephones as well as mobile smart phones and devices at our hands, the radios and televisions in our living rooms, the computer terminals with access to the Internet in our offices and homes, and our newspapers are all capable of providing rapid communications from every corner of the globe. Communication provides the senses for ships on the high seas, aircraft in flight, and rockets and satellites in space. Communication through a wireless telephone keeps a car driver in touch with the office or

home miles away, no matter where. Communication provides the means for social networks to engage in different ways (texting, speaking, visualizing), whereby people are brought together around the world. Communication keeps a weather forecaster informed of conditions measured by a multitude of sensors and satellites. Indeed, the list of applications involving the use of communication in one way or another is almost endless.

In the most fundamental sense, communication involves implicitly the transmission of *information* from one point to another through a succession of processes:

1. The generation of a *message signal* – voice, music, picture, or computer data.
2. The description of that message signal with a certain measure of precision, using a set of *symbols* – electrical, aural, or visual.
3. The *encoding* of those symbols in a suitable form for transmission over a physical medium of interest.
4. The *transmission* of the encoded symbols to the desired destination.
5. The *decoding* and *reproduction* of the original symbols.
6. The *re-creation* of the original message signal with some definable degradation in quality, the degradation being caused by unavoidable imperfections in the system.

There are, of course, many other forms of communication that do not directly involve the human mind in real time. For example, in *computer communications* involving communication between two or more computers, human decisions may enter only in setting up the programs or commands for the computer, or in monitoring the results.

Irrespective of the form of communication process being considered, there are three basic elements to every communication system, namely, *transmitter*, *channel*, and *receiver*, as depicted in Figure 1.1. The transmitter is located at one point in space, the receiver is located at some other point separate from the transmitter, and the channel is the physical medium that connects them together as an integrated communication system. The purpose of the transmitter is to convert the *message signal* produced by the *source of information* into a form suitable for transmission over the channel. However, as the transmitted signal propagates along the channel, it is distorted due to channel imperfections. Moreover, noise and interfering signals (originating from other sources) are added to the channel output, with the result that the *received signal* is a corrupted version of the *transmitted signal*. The receiver has the task of operating on the received signal so as to reconstruct a recognizable form of the original message signal for an end user or information sink.

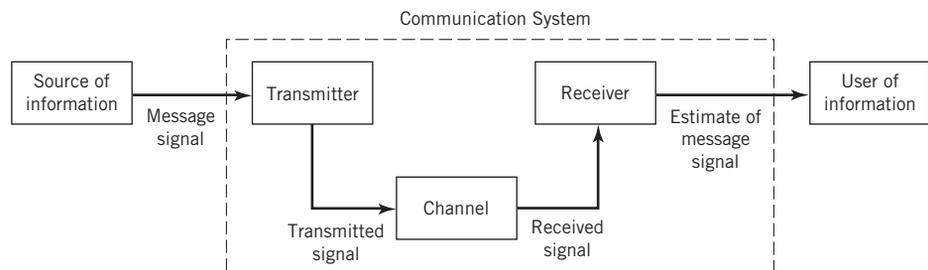


Figure 1.1 Elements of a communication system.

There are two basic modes of communication:

1. *Broadcasting*, which involves the use of a single powerful transmitter and numerous receivers that are relatively inexpensive to build. Here, information-bearing signals flow only in one direction.
2. *Point-to-point communication*, in which the communication process takes place over a link between a single transmitter and a receiver. In this case, there is usually a bidirectional flow of information-bearing signals, which requires the combined use of a transmitter and receiver (i.e., a *transceiver*) at each end of the link.

The underlying communication process in every communication system, irrespective of its kind, is *statistical* in nature. Indeed, it is for this important reason that much of this book is devoted to the statistical underpinnings of digital communication systems. In so doing, we develop a wealth of knowledge on the fundamental issues involved in the study of digital communications.

1.3 Multiple-Access Techniques

Continuing with the communication process, *multiple-access* is a technique whereby many subscribers or local stations can share the use of a communication channel at the same time or nearly so, despite the fact that their individual transmissions may originate from widely different locations. Stated in another way, a multiple-access technique permits the communication resources of the channel to be shared by a large number of users seeking to communicate with each other.

There are subtle differences between multiple access and multiplexing that should be noted:

- Multiple access refers to the remote sharing of a communication channel such as a satellite or radio channel by users in highly dispersed locations. On the other hand, multiplexing refers to the sharing of a channel such as a telephone channel by users confined to a local site.
- In a multiplexed system, user requirements are ordinarily fixed. In contrast, in a multiple-access system user requirements can change dynamically with time, in which case provisions are necessary for dynamic channel allocation.

For obvious reasons it is desirable that in a multiple-access system the sharing of resources of the channel be accomplished without causing serious interference between users of the system. In this context, we may identify four basic types of multiple access:

1. *Frequency-division multiple access (FDMA)*.

In this technique, disjoint subbands of frequencies are allocated to the different users on a continuous-time basis. In order to reduce interference between users allocated adjacent channel bands, *guard bands* are used to act as buffer zones, as illustrated in Figure 1.2a. These guard bands are necessary because of the impossibility of achieving ideal filtering or separating the different users.

2. *Time-division multiple access (TDMA)*.

In this second technique, each user is allocated the full spectral occupancy of the channel, but only for a short duration of time called a *time slot*. As shown in Figure 1.2b, buffer zones in the form of *guard times* are inserted between the assigned time

slots. This is done to reduce interference between users by allowing for time uncertainty that arises due to system imperfections, especially in synchronization schemes.

3. *Code-division multiple access (CDMA).*

In FDMA, the resources of the channel are shared by dividing them along the frequency coordinate into disjoint frequency bands, as illustrated in Figure 1.2a. In TDMA, the resources are shared by dividing them along the time coordinate into disjoint time slots, as illustrated in Figure 1.2b. In Figure 1.2c, we illustrate another technique for sharing the channel resources by using a hybrid combination of FDMA and TDMA, which represents a specific form of code-division multiple access (CDMA). For example, *frequency hopping* may be employed to ensure that during each successive time slot, the frequency bands assigned to the users are reordered in an essentially random manner. To be specific, during time slot 1, user 1 occupies frequency band 1, user 2 occupies frequency band 2, user 3 occupies frequency band 3, and so on. During time slot 2, user 1 hops to frequency band 3, user 2 hops to frequency band 1, user 3 hops to frequency band 2, and so on. Such an arrangement has the appearance of the users playing a game of musical chairs. An important advantage of CDMA over both FDMA and TDMA is that it can provide for *secure* communications. In the type of CDMA illustrated in Figure 1.2c, the frequency hopping mechanism can be implemented through the use of a pseudo-noise (PN) sequence.

4. *Space-division multiple access (SDMA).*

In this multiple-access technique, resource allocation is achieved by exploiting the spatial separation of the individual users. In particular, *multibeam antennas* are used to separate radio signals by pointing them along different directions. Thus, different users are enabled to access the channel simultaneously on the same frequency or in the same time slot.

These multiple-access techniques share a common feature: allocating the communication resources of the channel through the use of disjointness (or orthogonality in a loose sense) in time, frequency, or space.

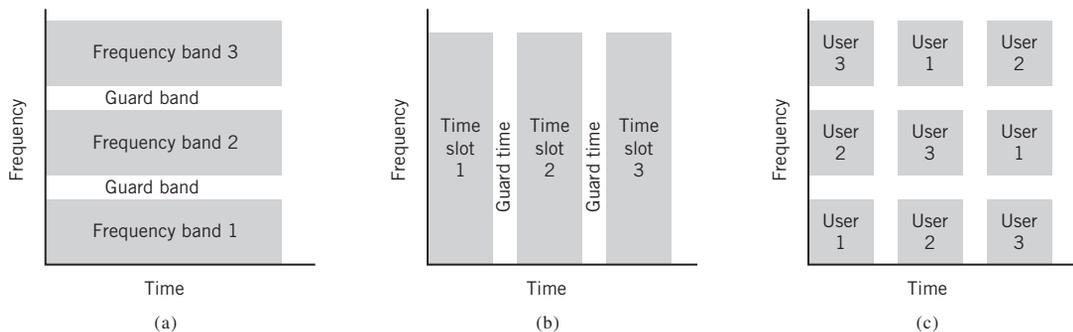


Figure 1.2 Illustrating the ideas behind multiple-access techniques. (a) Frequency-division multiple access. (b) Time-division multiple access. (c) Frequency-hop multiple access.

1.4 Networks

A *communication network* or simply *network*¹, illustrated in Figure 1.3, consists of an interconnection of a number of *nodes* made up of intelligent processors (e.g., microcomputers). The primary purpose of these nodes is to route data through the network. Each node has one or more *stations* attached to it; stations refer to devices wishing to communicate. The network is designed to serve as a shared resource for moving data exchanged between stations in an efficient manner and also to provide a framework to support new applications and services. The traditional telephone network is an example of a communication network in which *circuit switching* is used to provide a dedicated communication path or *circuit* between two stations. The circuit consists of a connected sequence of links from source to destination. The links may consist of time slots in a time-division multiplexed (TDM) system or frequency slots in a frequency-division multiplexed (FDM) system. The circuit, once in place, remains uninterrupted for the entire duration of transmission. Circuit switching is usually controlled by a centralized hierarchical control mechanism with knowledge of the network's organization. To establish a circuit-switched connection, an available path through the network is seized and then dedicated to the exclusive use of the two stations wishing to communicate. In particular, a call-request signal must propagate all the way to the destination, and be acknowledged, before transmission can begin. Then, the network is effectively transparent to the users. This means that, during the connection time, the bandwidth and resources allocated to the circuit are essentially “owned” by the two stations, until the circuit is disconnected. The circuit thus represents an efficient use of resources only to the extent that the allocated bandwidth is properly utilized. Although the telephone network is used to transmit data, voice constitutes the bulk of the network's traffic. Indeed, circuit switching is well suited to the transmission of voice signals, since voice conversations tend to be of long duration (about 2 min on average) compared with the time required for setting up the circuit (about 0.1–0.5 s). Moreover, in most voice conversations, there is information flow for a relatively large percentage of the connection time, which makes circuit switching all the more suitable for voice conversations.

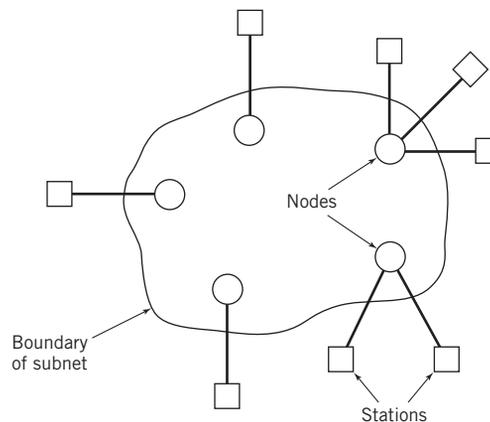


Figure 1.3 Communication network.

In circuit switching, a communication link is shared between the different sessions using that link on a *fixed* allocation basis. In *packet switching*, on the other hand, the sharing is done on a *demand* basis and, therefore, it has an advantage over circuit switching in that when a link has traffic to send, the link may be more fully utilized.

The basic network principle of packet switching is “store and forward.” Specifically, in a *packet-switched network*, any message larger than a specified size is subdivided prior to transmission into segments not exceeding the specified size. The segments are commonly referred to as *packets*. The original message is reassembled at the destination on a packet-by-packet basis. The network may be viewed as a distributed pool of *network resources* (i.e., channel bandwidth, buffers, and switching processors) whose capacity is *shared dynamically* by a community of competing users (stations) wishing to communicate. In contrast, in a circuit-switched network, resources are dedicated to a pair of stations for the entire period they are in session. Accordingly, packet switching is far better suited to a computer-communication environment in which “bursts” of data are exchanged between stations on an occasional basis. The use of packet switching, however, requires that careful *control* be exercised on user demands; otherwise, the network may be seriously abused.

The design of a *data network* (i.e., a network in which the stations are all made up of computers and terminals) may proceed in an orderly way by looking at the network in terms of a *layered architecture*, regarded as a hierarchy of nested layers. A *layer* refers to a process or device inside a computer system, designed to perform a specific function. Naturally, the designers of a layer will be intimately familiar with its internal details and operation. At the system level, however, a user views the layer merely as a “black box” that is described in terms of the inputs, the outputs, and the functional relationship between outputs and inputs. In a layered architecture, each layer regards the next lower layer as one or more black boxes with some given functional specification to be used by the given higher layer. Thus, the highly complex communication problem in data networks is resolved as a manageable set of well-defined interlocking functions. It is this line of reasoning that has led to the development of the *open systems interconnection (OSI)*² *reference model* by a subcommittee of the International Organization for Standardization. The term “open” refers to the ability of any two systems conforming to the reference model and its associated standards to interconnect.

In the OSI reference model, the communications and related-connection functions are organized as a series of *layers* or *levels* with well-defined *interfaces*, and with each layer built on its predecessor. In particular, each layer performs a related subset of primitive functions, and it relies on the next lower layer to perform additional primitive functions. Moreover, each layer offers certain services to the next higher layer and shields the latter from the implementation details of those services. Between each pair of layers, there is an *interface*. It is the interface that defines the services offered by the lower layer to the upper layer.

The OSI model is composed of seven layers, as illustrated in Figure 1.4; this figure also includes a description of the functions of the individual layers of the model. Layer k on system A , say, communicates with layer k on some other system B in accordance with a set of rules and conventions, collectively constituting the *layer k protocol*, where $k = 1, 2, \dots, 7$. (The term “protocol” has been borrowed from common usage, describing conventional social behavior between human beings.) The entities that comprise the corresponding layers on different systems are referred to as *peer processes*. In other words,

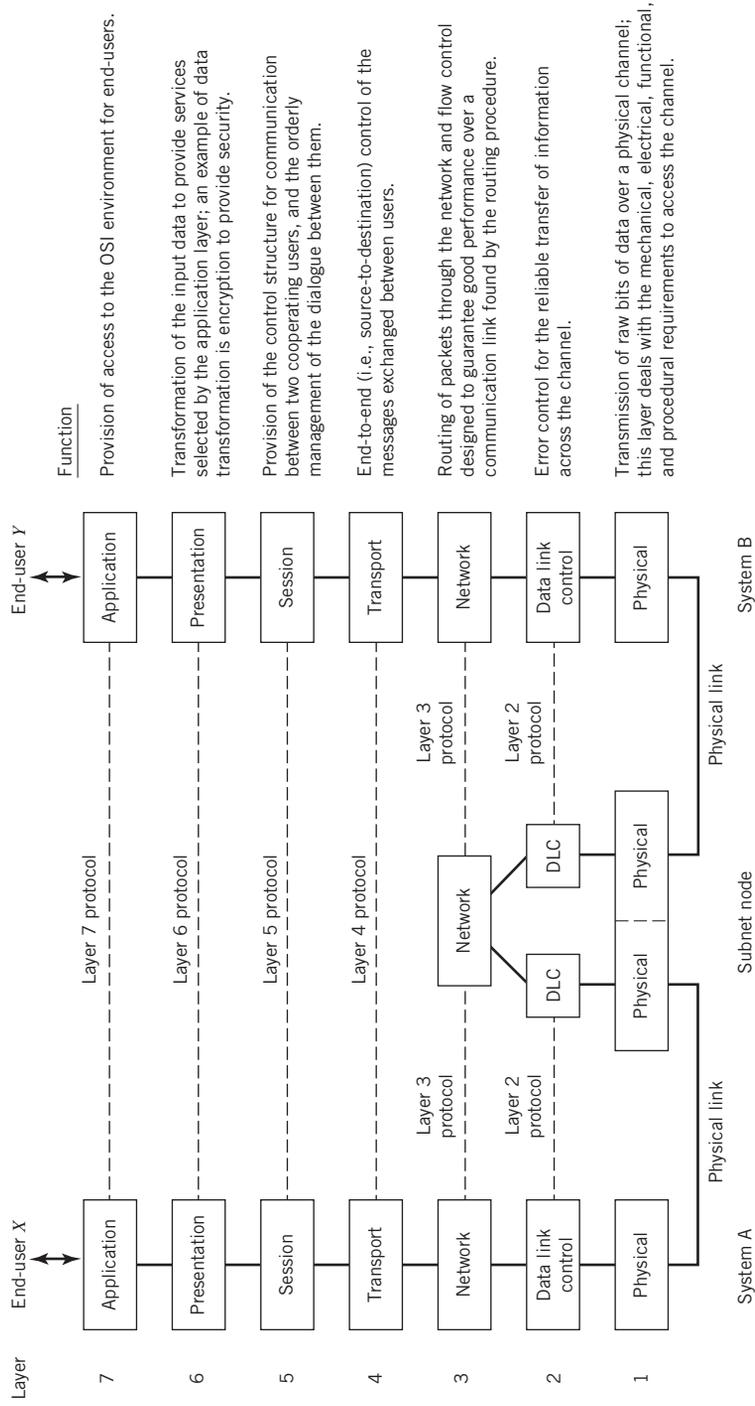


Figure 1.4 OSI model; DLC stands for data link control.

communication is achieved by having the peer processes in two different systems communicate via a protocol, with the protocol itself being defined by a set of rules of procedure. Physical communication between peer processes exists only at layer 1. On the other hand, layers 2 through 7 are in *virtual communication* with their distant peers. However, each of these six layers can exchange data and control information with its neighboring layers (below and above) through layer-to-layer interfaces. In Figure 1.4, physical communication is shown by solid lines and virtual communication by dashed lines. The major principles involved in arriving at seven layers of the OSI reference model are as follows:

1. Each layer performs well-defined functions.
2. A boundary is created at a point where the description of services offered is small and the number of interactions across the boundary is the minimum possible.
3. A layer is created from easily localized functions, so that the architecture of the model may permit modifications to the layer protocol to reflect changes in technology without affecting the other layers.
4. A boundary is created at some point with an eye toward standardization of the associated interface.
5. A layer is created only when a different level of abstraction is needed to handle the data.
6. The number of layers employed should be large enough to assign distinct functions to different layers, yet small enough to maintain a manageable architecture for the model.

Note that the OSI reference model is not a network architecture; rather, it is an international standard for computer communications, which just tells what each layer should do.

1.5 Digital Communications

Today's public communication networks are highly complicated systems. Specifically, public switched telephone networks (collectively referred to as PSTNs), the Internet, and wireless communications (including satellite communications) provide seamless connections between cities, across oceans, and between different countries, languages, and cultures; hence the reference to the world as a "global village."

There are three layers of the OSI model where it can affect the design of digital communication systems, which is the subject of interest of this book:

1. *Physical layer.* This lowest layer of the OSI model embodies the physical mechanism involved in transmitting *bits* (i.e., *binary digits*) between any pair of nodes in the communication network. Communication between the two nodes is accomplished by means of modulation in the transmitter, transmission across the channel, and demodulation in the receiver. The module for performing *modulation* and *demodulation* is often called a *modem*.
2. *Data-link layer.* Communication links are nearly always corrupted by the unavoidable presence of noise and interference. One purpose of the data-link layer, therefore, is to perform *error correction* or *detection*, although this function is also shared with the physical layer. Often, the data-link layer will retransmit packets that are received in error but, for some applications, it discards them. This layer is also

responsible for the way in which different users share the transmission medium. A portion of the data-link layer, called the *medium access control (MAC)* sublayer, is responsible for allowing frames to be sent over the shared transmission media without undue interference with other nodes. This aspect is referred to as *multiple-access* communications.

3. *Network layer.* This layer has several functions, one of which is to determine the *routing* of information, to get it from the source to its ultimate destination. A second function is to determine the *quality of service*. A third function is *flow control*, to ensure that the network does not become congested.

These are three layers of a seven-layer model for the functions that occur in the communications process. Although the three layers occupy a subspace within the OSI model, the functions that they perform are of critical importance to the model.

Block Diagram of Digital Communication System

Typically, in the design of a digital communication system the information source, communication channel, and information sink (end user) are all specified. The challenge is to design the transmitter and the receiver with the following guidelines in mind:

- Encode/modulate the message signal generated by the source of information, transmit it over the channel, and produce an “estimate” of it at the receiver output that satisfies the requirements of the end user.
- Do all of this at an affordable cost.

In a *digital communication* system represented by the block diagram of Figure 1.6, the rationale for which is rooted in information theory, the functional blocks of the transmitter and the receiver starting from the far end of the channel are paired as follows:

- source encoder–decoder;
- channel encoder–decoder;
- modulator–demodulator.

The source encoder removes redundant information from the message signal and is responsible for efficient use of the channel. The resulting sequence of symbols is called the *source codeword*. The data stream is processed next by the channel encoder, which produces a new sequence of symbols called the *channel codeword*. The channel codeword is longer than the source code word by virtue of the *controlled* redundancy built into its construction. Finally, the modulator represents each symbol of the channel codeword by a corresponding analog symbol, appropriately selected from a finite set of possible analog symbols. The sequence of analog symbols produced by the modulator is called a *waveform*, which is suitable for transmission over the channel. At the receiver, the channel output (received signal) is processed in reverse order to that in the transmitter, thereby reconstructing a recognizable version of the original message signal. The reconstructed message signal is finally delivered to the user of information at the destination. From this description it is apparent that the design of a digital communication system is rather complex in conceptual terms but easy to build. Moreover, the system is *robust*, offering greater tolerance of physical effects (e.g., temperature variations, aging, mechanical vibrations) than its analog counterpart; hence the ever-increasing use of digital communications.

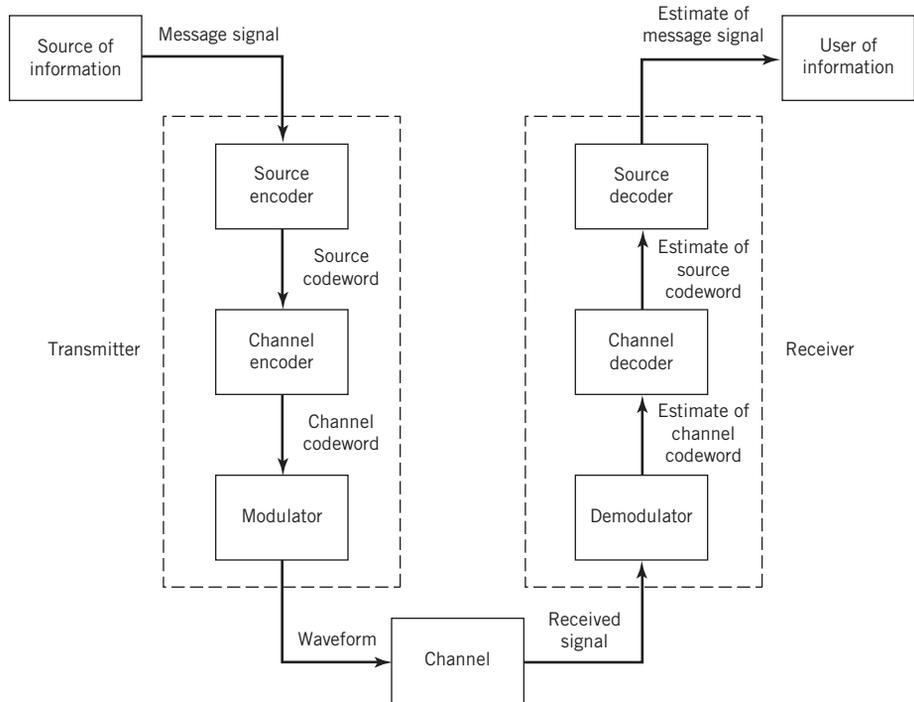


Figure 1.6 Block diagram of a digital communication system.

1.6 Organization of the Book

The main part of the book is organized in ten chapters, which, after this introductory chapter, are organized into five parts of varying sizes as summarized herein.

1. Mathematical Background

Chapter 2 presents a detailed treatment of the Fourier transform, its properties and algorithmic implementations. This chapter also includes two important related topics:

- The Hilbert transform, which provides the mathematical basis for transforming real-valued band-pass signals and systems into their low-pass equivalent representations without loss of information.
- Overview of analog modulation theory, thereby facilitating an insightful link between analog and digital communications.

Chapter 3 presents a mathematical review of probability theory and Bayesian inference, the understanding of which is essential to the study of digital communications.

Chapter 4 is devoted to the study of stochastic processes, the theory of which is basic to the characterization of sources of information and communication channels.

Chapter 5 discusses the fundamental limits of information theory, postulated in terms of source coding, channel capacity, and rate-distortion theory.

2. Transition from Analog to Digital Communications

This material is covered in Chapter 6. Simply put, the study therein discusses the different ways in which analog waveforms are converted into digitally encoded sequences.

3. Signaling Techniques

This third part of the book includes three chapters:

- Chapter 7 discusses the different techniques for signaling over additive white Gaussian noise (AWGN) channels.
- Chapter 8 discusses signaling over band-limited channels, as in data transmission over telephonic channels and the Internet.
- Chapter 9 is devoted to signaling over fading channels, as in wireless communications.

4. Error-Control Coding

The reliability of data transmission over a communication channel is of profound practical importance. Chapter 10 studies the different methods for the encoding of message sequences in the transmitter and decoding them in the receiver. Here, we cover two classes of error-control coding techniques:

- classic codes rooted in algebraic mathematics, and
- new generation of probabilistic compound codes, exemplified by turbo codes and LDPC codes.

5. Appendices

Last but by no means least, the book includes appendices to provide back-up material for different chapters in the book, as they are needed.

Notes

1. For a detailed discussion on communication networks, see the classic book by Tanenbaum, entitled *Computer Networks* (2003).
2. The OSI reference model was developed by a subcommittee of the International Organization for Standardization (ISO) in 1977. For a discussion of the principles involved in arriving at the seven layers of the OSI model and a description of the layers themselves, see Tanenbaum (2003).

Fourier Analysis of Signals and Systems

2.1 Introduction

The study of communication systems involves:

- the processing of a modulated message signal generated at the transmitter output so as to facilitate its transportation across a physical channel and
- subsequent processing of the received signal in the receiver so as to deliver an estimate of the original message signal to a user at the receiver output.

In this study, the *representation of signals and systems* features prominently. More specifically, the *Fourier transform* plays a key role in this representation.

The Fourier transform provides the mathematical link between the time-domain representation (i.e., waveform) of a signal and its frequency-domain description (i.e., spectrum). Most importantly, we can go back and forth between these two descriptions of the signal with no loss of information. Indeed, we may invoke a similar transformation in the representation of linear systems. In this latter case, the time-domain and frequency-domain descriptions of a linear time-invariant system are defined in terms of its impulse response and frequency response, respectively.

In light of this background, it is in order that we begin a mathematical study of communication systems by presenting a review of Fourier analysis. This review, in turn, paves the way for the formulation of simplified representations of band-pass signals and systems to which we resort in subsequent chapters. We begin the study by developing the transition from the Fourier series representation of a periodic signal to the Fourier transform representation of a nonperiodic signal; this we do in the next two sections.

2.2 The Fourier Series

Let $g_{T_0}(t)$ denote a *periodic signal*, where the subscript T_0 denotes the duration of periodicity. By using a *Fourier series expansion* of this signal, we are able to resolve it into an infinite sum of sine and cosine terms, as shown by

$$g_{T_0}(t) = a_0 + 2 \sum_{n=1}^{\infty} [a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t)] \quad (2.1)$$

where

$$f_0 = \frac{1}{T_0} \quad (2.2)$$

is the *fundamental frequency*. The coefficients a_n and b_n represent the amplitudes of the cosine and sine terms, respectively. The quantity nf_0 represents the n th harmonic of the fundamental frequency f_0 . Each of the terms $\cos(2\pi nf_0 t)$ and $\sin(2\pi nf_0 t)$ is called a *basis function*. These basis functions form an *orthogonal set* over the interval T_0 , in that they satisfy three conditions:

$$\int_{-T_0/2}^{T_0/2} \cos(2\pi mf_0 t) \cos(2\pi nf_0 t) dt = \begin{cases} T_0/2, & m = n \\ 0, & m \neq n \end{cases} \quad (2.3)$$

$$\int_{-T_0/2}^{T_0/2} \cos(2\pi mf_0 t) \sin(2\pi nf_0 t) dt = 0, \quad \text{for all } m \text{ and } n \quad (2.4)$$

$$\int_{-T_0/2}^{T_0/2} \sin(2\pi mf_0 t) \sin(2\pi nf_0 t) dt = \begin{cases} T_0/2, & m = n \\ 0, & m \neq n \end{cases} \quad (2.5)$$

To determine the coefficient a_0 , we integrate both sides of (2.1) over a complete period. We thus find that a_0 is the *mean value* of the periodic signal $g_{T_0}(t)$ over one period, as shown by the *time average*

$$a_0 = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g_{T_0}(t) dt \quad (2.6)$$

To determine the coefficient a_n , we multiply both sides of (2.1) by $\cos(2\pi nf_0 t)$ and integrate over the interval $-T_0/2$ to $T_0/2$. Then, using (2.3) and (2.4), we find that

$$a_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g_{T_0}(t) \cos(2\pi nf_0 t) dt, \quad n = 1, 2, \dots \quad (2.7)$$

Similarly, we find that

$$b_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g_{T_0}(t) \sin(2\pi nf_0 t) dt, \quad n = 1, 2, \dots \quad (2.8)$$

A basic question that arises at this point is the following:

Given a periodic signal $g_{T_0}(t)$ of period T_0 , how do we know that the Fourier series expansion of (2.1) is *convergent* in that the infinite sum of terms in this expansion is exactly equal to $g_{T_0}(t)$?

To resolve this fundamental issue, we have to show that, for the coefficients a_0 , a_n , and b_n calculated in accordance with (2.6) to (2.8), this series will indeed converge to $g_{T_0}(t)$. In general, for a periodic signal $g_{T_0}(t)$ of arbitrary waveform, there is no guarantee that the series of (2.1) will converge to $g_{T_0}(t)$ or that the coefficients a_0 , a_n , and b_n will even exist. In a rigorous sense, we may say that a periodic signal $g_{T_0}(t)$ can be expanded in a Fourier

series if the signal $g_{T_0}(t)$ satisfies the *Dirichlet conditions*:¹

1. The function $g_{T_0}(t)$ is single valued within the interval T_0 .
2. The function $g_{T_0}(t)$ has at most a finite number of discontinuities in the interval T_0 .
3. The function $g_{T_0}(t)$ has a finite number of maxima and minima in the interval T_0 .
4. The function $g_{T_0}(t)$ is absolutely integrable; that is,

$$\int_{-T_0/2}^{T_0/2} |g_{T_0}(t)| dt < \infty$$

From an engineering perspective, however, it suffices to say that the Dirichlet conditions are satisfied by the periodic signals encountered in communication systems.

Complex Exponential Fourier Series

The Fourier series of (2.1) can be put into a much simpler and more elegant form with the use of complex exponentials. We do this by substituting into (2.1) the exponential forms for the cosine and sine, namely:

$$\cos(2\pi n f_0 t) = \frac{1}{2} [\exp(j2\pi n f_0 t) + \exp(-j2\pi n f_0 t)]$$

$$\sin(2\pi n f_0 t) = \frac{1}{2j} [\exp(j2\pi n f_0 t) - \exp(-j2\pi n f_0 t)]$$

where $j = \sqrt{-1}$. We thus obtain

$$g_{T_0}(t) = a_0 + \sum_{n=1}^{\infty} [(a_n - j b_n) \exp(j2\pi n f_0 t) + (a_n + j b_n) \exp(-j2\pi n f_0 t)] \quad (2.9)$$

Let c_n denote a complex coefficient related to a_n and b_n by

$$c_n = \begin{cases} a_n - j b_n, & n > 0 \\ a_0, & n = 0 \\ a_n + j b_n, & n < 0 \end{cases} \quad (2.10)$$

Then, we may simplify (2.9) into

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} c_n \exp(j2\pi n f_0 t) \quad (2.11)$$

where

$$c_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g_{T_0}(t) \exp(-j2\pi n f_0 t) dt, \quad n = 0, \pm 1, \pm 2, \dots \quad (2.12)$$

The series expansion of (2.11) is referred to as the *complex exponential Fourier series*. The c_n themselves are called the *complex Fourier coefficients*.

Given a periodic signal $g_{T_0}(t)$, (2.12) states that we may determine the complete set of complex Fourier coefficients. On the other hand, (2.11) states that, given this set of coefficients, we may reconstruct the original periodic signal $g_{T_0}(t)$ exactly.

The integral on the right-hand side of (2.12) is said to be an *inner product* of the signal $g_{T_0}(t)$ with the *basis functions* $\exp(-j2\pi n f_0 t)$, by whose linear combination all square integrable functions can be expressed as in (2.11).

According to this representation, a periodic signal contains all frequencies (both positive and negative) that are harmonically related to the fundamental frequency f_0 . The presence of negative frequencies is simply a result of the fact that the mathematical model of the signal as described by (2.11) requires the use of negative frequencies. Indeed, this representation also requires the use of complex-valued basis functions, namely $\exp(j2\pi n f_0 t)$, which have no physical meaning either. The reason for using complex-valued basis functions and negative frequency components is merely to provide a compact mathematical description of a periodic signal, which is well-suited for both theoretical and practical work.

2.3 The Fourier Transform

In the previous section, we used the Fourier series to represent a periodic signal. We now wish to develop a similar representation for a signal $g(t)$ that is nonperiodic. In order to do this, we first construct a periodic function $g_{T_0}(t)$ of period T_0 in such a way that $g(t)$ defines exactly one cycle of this periodic function, as illustrated in Figure 2.1. In the limit, we let the period T_0 become infinitely large, so that we may express $g(t)$ as

$$g(t) = \lim_{T_0 \rightarrow \infty} g_{T_0}(t) \quad (2.13)$$

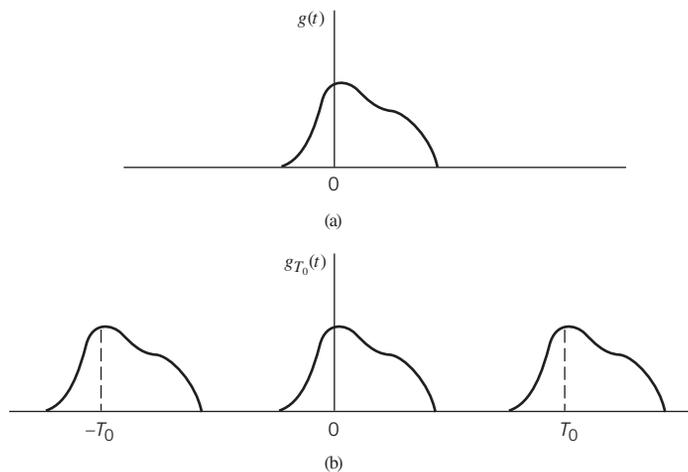


Figure 2.1 Illustrating the use of an arbitrarily defined function of time to construct a periodic waveform. (a) Arbitrarily defined function of time $g(t)$. (b) Periodic waveform $g_{T_0}(t)$ based on $g(t)$.

Representing the periodic function $g_{T_0}(t)$ in terms of the complex exponential form of the Fourier series, we write

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{j2\pi nt}{T_0}\right)$$

where

$$c_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g_{T_0}(t) \exp\left(-\frac{j2\pi nt}{T_0}\right) dt$$

Here, we have purposely replaced f_0 with $1/T_0$ in the exponents. Define

$$\Delta f = \frac{1}{T_0}$$

$$f_n = \frac{n}{T_0}$$

and

$$G(f_n) = c_n T_0$$

We may then go on to modify the original Fourier series representation of $g_{T_0}(t)$ given in (2.11) into a new form described by

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} G(f_n) \exp(j2\pi f_n t) \Delta f \quad (2.14)$$

where

$$G(f_n) = \int_{-T_0/2}^{T_0/2} g_{T_0}(t) \exp(-j2\pi f_n t) dt \quad (2.15)$$

Equations (2.14) and (2.15) apply to a periodic signal $g_{T_0}(t)$. What we would like to do next is to go one step further and develop a corresponding pair of formulas that apply to a nonperiodic signal $g(t)$. To do this transition, we use the defining equation (2.13). Specifically, two things happen:

1. The discrete frequency f_n in (2.14) and (2.15) approaches the continuous frequency variable f .
2. The discrete sum of (2.14) becomes an integral defining the area under the function $G(f) \exp(j2\pi ft)$, integrated with respect to time t .

Accordingly, piecing these points together, we may respectively rewrite the limiting forms of (2.15) and (2.14) as

$$G(f) = \int_{-\infty}^{\infty} g(t) \exp(-j2\pi ft) dt \quad (2.16)$$

and

$$g(t) = \int_{-\infty}^{\infty} G(f) \exp(j2\pi ft) df \quad (2.17)$$

In words, we may say:

- the *Fourier transform* of the nonperiodic signal $g(t)$ is defined by (2.16);
- given the Fourier transform $G(f)$, the original signal $g(t)$ is recovered exactly from the inverse *Fourier transform* of (2.17).

Figure 2.2 illustrates the interplay between these two formulas, where we see that the frequency-domain description based on (2.16) plays the role of *analysis* and the time-domain description based on (2.17) plays the role of *synthesis*.

From a notational point of view, note that in (2.16) and (2.17) we have used a lowercase letter to denote the time function and an uppercase letter to denote the corresponding frequency function. Note also that these two equations are of identical mathematical form, except for changes in the algebraic signs of the exponents.

For the Fourier transform of a signal $g(t)$ to exist, it is sufficient but not necessary that the nonperiodic signal $g(t)$ satisfies three *Dirichlet's conditions* of its own:

1. The function $g(t)$ is single valued, with a finite number of maxima and minima in any finite time interval.
2. The function $g(t)$ has a finite number of discontinuities in any finite time interval.
3. The function $g(t)$ is absolutely integrable; that is,

$$\int_{-\infty}^{\infty} |g(t)| dt < \infty$$

In practice, we may safely ignore the question of the existence of the Fourier transform of a time function $g(t)$ when it is an accurately specified description of a physically realizable signal. In other words, physical realizability is a sufficient condition for the existence of a Fourier transform. Indeed, we may go one step further and state:

All energy signals are Fourier transformable.

A signal $g(t)$ is said to be an *energy signal* if the condition

$$\int_{-\infty}^{\infty} |g(t)|^2 dt < \infty \quad (2.18)$$

holds.²

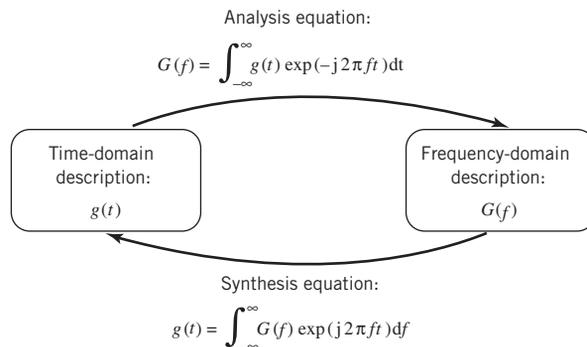


Figure 2.2 Sketch of the interplay between the synthesis and analysis equations embodied in Fourier transformation.

The Fourier transform provides the mathematical tool for measuring the frequency content, or spectrum, of a signal. For this reason, the terms *Fourier transform* and *spectrum* are used interchangeably. Thus, given a signal $g(t)$ with Fourier transform $G(f)$, we may refer to $G(f)$ as the spectrum of the signal $g(t)$. By the same token, we refer to $|G(f)|$ as the *magnitude spectrum* of the signal $g(t)$, and refer to $\arg[G(f)]$ as its *phase spectrum*.

If the signal $g(t)$ is real valued, then the magnitude spectrum of the signal is an even function of frequency f , while the phase spectrum is an odd function of f . In such a case, knowledge of the spectrum of the signal for positive frequencies uniquely defines the spectrum for negative frequencies.

Notations

For convenience of presentation, it is customary to express (2.17) in the short-hand form

$$G(f) = \mathbf{F}[g(t)]$$

where \mathbf{F} plays the role of an *operator*. In a corresponding way, (2.18) is expressed in the short-hand form

$$g(t) = \mathbf{F}^{-1}[G(f)]$$

where \mathbf{F}^{-1} plays the role of an *inverse operator*.

The time function $g(t)$ and the corresponding frequency function $G(f)$ are said to constitute a *Fourier-transform pair*. To emphasize this point, we write

$$g(t) \rightleftharpoons G(f)$$

where the top arrow indicates the forward transformation from $g(t)$ to $G(f)$ and the bottom arrow indicates the inverse transformation. One other notation: the asterisk is used to denote complex conjugation.

Tables of Fourier Transformations

To assist the user of this book, two tables of Fourier transformations are included:

1. Table 2.1 on page 23 summarizes the properties of Fourier transforms; proofs of them are presented as end-of-chapter problems.
2. Table 2.2 on page 24 presents a list of Fourier-transform pairs, where the items listed on the left-hand side of the table are time functions and those in the center column are their Fourier transforms.

EXAMPLE 1

Binary Sequence for Energy Calculations

Consider the five-digit binary sequence 10010. This sequence is represented by two different waveforms, one based on the rectangular function $\text{rect}(t)$, and the other based on the sinc function $\text{sinc}(t)$. Despite this difference, both waveforms are denoted by $g(t)$, which implies they both have exactly the same total energy, to be demonstrated next.

Case 1: $\text{rect}(t)$ as the basis function.

Let binary symbol 1 be represented by $+\text{rect}(t)$ and binary symbol 0 be represented by $-\text{rect}(t)$. Accordingly, the binary sequence 10010 is represented by the waveform

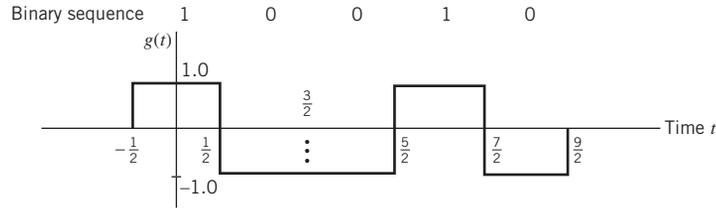


Figure 2.3 Waveform of binary sequence 10010, using $\text{rect}(t)$ for symbol 1 and $-\text{rect}(t)$ for symbol 0. See Table 2.2 for the definition of $\text{rect}(t)$.

shown in Figure 2.3. From this figure, we readily see that, regardless of the representation $\pm\text{rect}(t)$, each symbol contributes a single unit of energy; hence the total energy for Case 1 is five units.

Case 2: $\text{sinc}(t)$ as the basis function.

Consider next the representation of symbol 1 by $+\text{sinc}(t)$ and the representation of symbol 0 by $-\text{sinc}(t)$, which do not interfere with each other in constructing the waveform for the binary sequence 10010. Unfortunately, this time around, it is difficult to calculate the total waveform energy in the time domain. To overcome this difficulty, we do the calculation in the frequency domain.

To this end, in parts a and b of Figure 2.4, we display the waveform of the sinc function in the time domain and its Fourier transform, respectively. On this basis, Figure 2.5 displays the frequency-domain representation of the binary sequence 10010, with part a of the figure displaying the magnitude response $|G(f)|$, and part b displaying the corresponding phase response $\arg[G(f)]$ expressed in radians. Then, applying Rayleigh's energy theorem, described in Property 14 in Table 2.2, to part a of Figure 2.5, we readily find that the energy of the pulse, $\pm\text{sinc}(t)$, is equal to one unit, regardless of its amplitude. The total energy of the sinc-based waveform representing the given binary sequence is also exactly five units, confirming what was said at the beginning of this example.

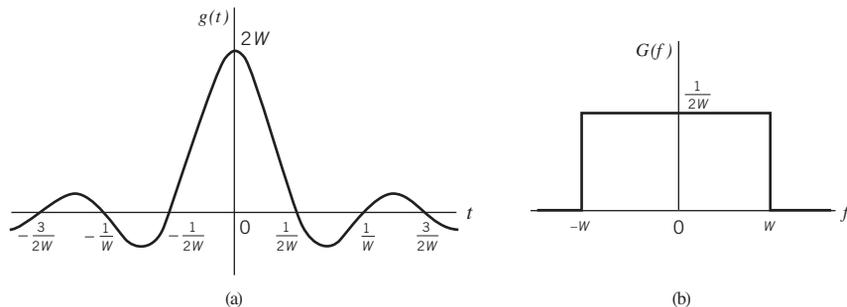


Figure 2.4 (a) Sinc pulse $g(t)$. (b) Fourier transform $G(f)$.

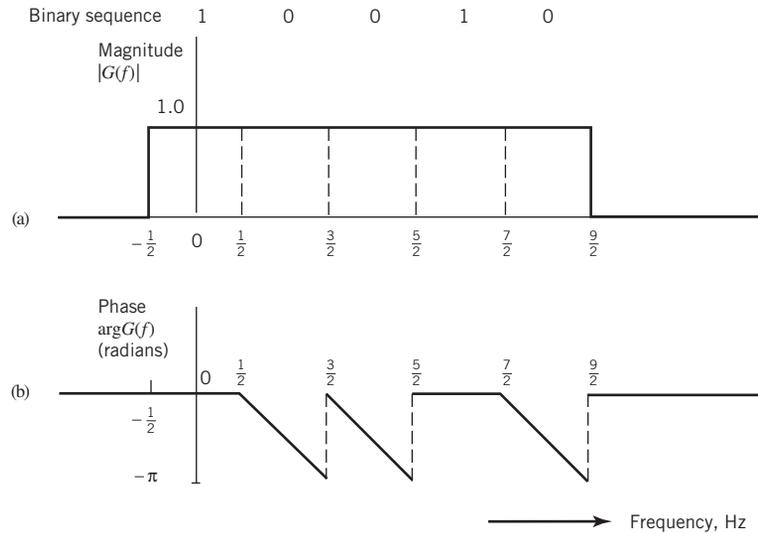


Figure 2.5 (a) Magnitude spectrum of the sequence 10010. (b) Phase spectrum of the sequence.

Observations

1. The dual basis functions, $\text{rect}(t)$ and $\text{sinc}(t)$, are *dilated* to their simplest forms, each of which has an energy of one unit, hence the equality of the results presented under Cases 1 and 2.
2. Examining the waveform $g(t)$ in Figure 2.3, we clearly see the discrimination between binary symbols 1 and 0. On the other hand, it is the phase response $\arg[G(f)]$ in part b of Figure 2.5 that shows the discrimination between binary symbols 1 and 0.

EXAMPLE 2

Unit Gaussian Pulse

Typically, a pulse signal $g(t)$ and its Fourier transform $G(f)$ have different mathematical forms. This observation is illustrated by the Fourier-transform pair studied in Example 1. In this second example, we consider an exception to this observation. In particular, we use the differentiation property of the Fourier transform to derive the particular form of a *pulse signal that has the same mathematical form as its own Fourier transform*.

Let $g(t)$ denote the pulse signal expressed as a function of time t and $G(f)$ denote its Fourier transform. Differentiating the Fourier transform formula of (2.6) with respect to frequency f yields

$$-j2\pi t g(t) \Leftrightarrow \frac{d}{df} G(f)$$

or, equivalently,

$$2\pi t g(t) \Leftrightarrow j \frac{d}{df} G(f) \quad (2.19)$$

Use of the Fourier-transform property on differentiation in the time domain listed in Table 2.1 yields

$$\frac{d}{dt}g(t) \Leftrightarrow j2\pi fG(f) \quad (2.20)$$

Suppose we now impose the equality condition on the left-hand sides of (2.19) and (2.20):

$$\frac{d}{dt}g(t) = 2\pi tg(t) \quad (2.21)$$

Then, in a corresponding way, it follows that the right-hand sides of these two equations must (after canceling the common multiplying factor j) satisfy the condition

$$\frac{d}{df}G(f) = 2\pi fG(f) \quad (2.22)$$

Equations (2.21) and (2.22) show that the pulse signal $g(t)$ and its Fourier transform $G(f)$ have exactly the same mathematical form. In other words, provided that the pulse signal $g(t)$ satisfies the differential equation (2.21), then $G(f) = g(f)$, where $g(f)$ is obtained from $g(t)$ simply by substituting f for t . Solving (2.21) for $g(t)$, we obtain

$$g(t) = \exp(-\pi t^2) \quad (2.23)$$

which has a bell-shaped waveform, as illustrated in Figure 2.6. Such a pulse is called a *Gaussian pulse*, the name of which follows from the similarity of the function $g(t)$ to the Gaussian probability density function of probability theory, to be discussed in Chapter 3. By applying the Fourier-transform property on the area under $g(t)$ listed in Table 2.1, we have

$$\int_{-\infty}^{\infty} \exp(-\pi t^2) dt = 1 \quad (2.24)$$

When the central ordinate and the area under the curve of a pulse are both unity, as in (2.23) and (2.24), we say that the Gaussian pulse is a *unit pulse*. Therefore, we may state that the unit Gaussian pulse is its own Fourier transform, as shown by

$$\exp(-\pi t^2) \Leftrightarrow \exp(-\pi f^2) \quad (2.25)$$

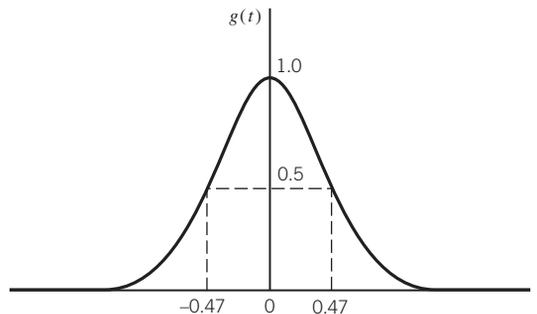


Figure 2.6 Gaussian pulse.

Table 2.1 Fourier-transform theorems

Property	Mathematical description
1. Linearity	$ag_1(t) + bg_2(t) \Leftrightarrow aG_1(f) + bG_2(f)$ where a and b are constants
2. Dilation	$g(at) \Leftrightarrow \frac{1}{ a }G\left(\frac{f}{a}\right)$ where a is a constant
3. Duality	If $g(t) \Leftrightarrow G(f)$, then $G(t) \Leftrightarrow g(-f)$
4. Time shifting	$g(t - t_0) \Leftrightarrow G(f) \exp(-j2\pi ft_0)$
5. Frequency shifting	$g(t) \exp(-j2\pi f_0 t) \Leftrightarrow G(f - f_0)$
6. Area under $g(t)$	$\int_{-\infty}^{\infty} g(t) dt = G(0)$
7. Area under $G(f)$	$g(0) = \int_{-\infty}^{\infty} G(f) df$
8. Differentiation in the time domain	$\frac{d}{dt}g(t) \Leftrightarrow j2\pi fG(f)$
9. Integration in the time domain	$\int_{-\infty}^f g(\tau) d\tau \Leftrightarrow \frac{1}{j2\pi f}G(f) + \frac{G(0)}{2}\delta(f)$
10. Conjugate functions	If $g(t) \Leftrightarrow G(f)$, then $g^*(t) \Leftrightarrow G^*(-f)$
11. Multiplication in the time domain	$g_1(t)g_2(t) \Leftrightarrow \int_{-\infty}^{\infty} G_1(\lambda)G_2(f - \lambda)d\lambda$
12. Convolution in the time domain	$\int_{-\infty}^f g_1(\tau)g_2(t - \tau)d\tau \Leftrightarrow G_1(f)G_2(f)$
13. Correlation theorem	$\int_{-\infty}^{\infty} g_1(t)g_2^*(t - \tau)d\tau \Leftrightarrow G_1(f)G_2^*(f)$
14. Rayleigh's energy theorem	$\int_{-\infty}^{\infty} g(t) ^2 dt = \int_{-\infty}^{\infty} G(f) ^2 df$
15. Parseval's power theorem for periodic signal of period T_0	$\frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g(t) ^2 dt = \sum_{n=-\infty}^{\infty} G(f_n) ^2, \quad f_n = n/T_0$

Table 2.2 Fourier-transform pairs and commonly used time functions

Time function	Fourier transform	Definitions
1. $\text{rect}\left(\frac{t}{T}\right)$	$T \text{sinc}(fT)$	Unit step function:
2. $\text{sinc}(2Wt)$	$\frac{1}{2W} \text{rect}\left(\frac{f}{2W}\right)$	$u(t) = \begin{cases} 1, & t > 0 \\ \frac{1}{2}, & t = 0 \\ 0, & t < 0 \end{cases}$
3. $\exp(-at)u(t), \quad a > 0$	$\frac{1}{a + j2\pi f}$	
4. $\exp(-a t), \quad a > 0$	$\frac{2a}{a^2 + (2\pi f)^2}$	Dirac delta function: $\delta(t) = 0$ for $t \neq 0$ and $\int_{-\infty}^{\infty} \delta(t) dt = 1$
5. $\exp(-\pi t^2)$	$\exp(-\pi f^2)$	
6. $\begin{cases} 1 - \frac{ t }{T}, & t < T \\ 0, & t \geq T \end{cases}$	$T \text{sinc}^2(fT)$	Rectangular function: $\text{rect}(t) = \begin{cases} 1, & -\frac{1}{2} < t \leq \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}$
7. $\delta(t)$	1	
8. 1	$\delta(f)$	Signum function:
9. $\delta(t - t_0)$	$\exp(-j2\pi f t_0)$	$\text{sgn}(t) = \begin{cases} +1, & t > 0 \\ 0, & t = 0 \\ -1, & t < 0 \end{cases}$
10. $\exp(j2\pi f_c t)$	$\delta(f - f_c)$	
11. $\cos(2\pi f_c t)$	$\frac{1}{2}[\delta(f - f_c) + \delta(f + f_c)]$	Sinc function: $\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$
12. $\sin(2\pi f_c t)$	$\frac{1}{2j}[\delta(f - f_c) - \delta(f + f_c)]$	
13. $\text{sgn}(t)$	$\frac{1}{j\pi f}$	Gaussian function: $g(t) = \exp(-\pi t^2)$
14. $\frac{1}{\pi t}$	$-j \text{sgn}(f)$	
15. $u(t)$	$\frac{1}{2}\delta(f) + \frac{1}{j2\pi f}$	
16. $\sum_{i=-\infty}^{\infty} \delta(t - iT_0)$	$f_0 \sum_{n=-\infty}^{\infty} \delta(f - nf_0), \quad f_0 = \frac{1}{T_0}$	

2.4 The Inverse Relationship between Time-Domain and Frequency-Domain Representations

The time-domain and frequency-domain descriptions of a signal are *inversely* related. In this context, we may make four important statements:

1. If the time-domain description of a signal is changed, the frequency-domain description of the signal is changed in an *inverse* manner, and vice versa. This inverse relationship prevents arbitrary specifications of a signal in both domains. In other words:

We may specify an arbitrary function of time or an arbitrary spectrum, but we cannot specify them both together.

2. If a signal is strictly limited in frequency, then the time-domain description of the signal will trail on indefinitely, even though its amplitude may assume a progressively smaller value. To be specific, we say:

A signal is strictly limited in frequency (i.e., strictly band limited) if its Fourier transform is exactly zero outside a finite band of frequencies.

Consider, for example, the band-limited sinc pulse defined by

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$$

whose waveform and spectrum are respectively shown in Figure 2.4: part a shows that the sinc pulse is *asymptotically limited in time* and part b of the figure shows that the sinc pulse is indeed *strictly band limited*, thereby confirming statement 2.

3. In a dual manner to statement 2, we say:

If a signal is strictly limited in time (i.e., the signal is exactly zero outside a finite time interval), then the spectrum of the signal is infinite in extent, even though the magnitude spectrum may assume a progressively smaller value.

This third statement is exemplified by a *rectangular pulse*, the waveform and spectrum of which are defined in accordance with item 1 in Table 2.2.

4. In light of the duality described under statements 2 and 3, we now make the final statement:

A signal cannot be strictly limited in both time and frequency.

The Bandwidth Dilemma

The statements we have just made have an important bearing on the *bandwidth* of a signal, which provides a measure of the *extent of significant spectral content of the signal for positive frequencies*. When the signal is strictly band limited, the bandwidth is well defined. For example, the sinc pulse $\text{sinc}(2Wt)$ has a bandwidth equal to W . However, when the signal is not strictly band limited, as is often the case, we encounter difficulty in defining the bandwidth of the signal. The difficulty arises because the meaning of “significant” attached to the spectral content of the signal is mathematically imprecise. Consequently, there is no universally accepted definition of bandwidth. It is in this sense that we speak of the “bandwidth dilemma.”

Nevertheless, there are some commonly used definitions for bandwidth, as discussed next. When the spectrum of a signal is symmetric with a main lobe bounded by well-defined nulls (i.e., frequencies at which the spectrum is zero), we may use the main lobe as the basis for defining the bandwidth of the signal. Specifically:

If a signal is low-pass (i.e., its spectral content is centered around the origin $f = 0$), the bandwidth is defined as one-half the total width of the main spectral lobe, since only one-half of this lobe lies inside the positive frequency region.

For example, a rectangular pulse of duration T seconds has a main spectral lobe of total width $(2/T)$ hertz centered at the origin. Accordingly, we may define the bandwidth of this rectangular pulse as $(1/T)$ hertz.

If, on the other hand, the signal is *band-pass* with main spectral lobes centered around $\pm f_c$, where f_c is large enough, the bandwidth is defined as the width of the main lobe for positive frequencies. This definition of bandwidth is called the *null-to-null bandwidth*. Consider, for example, a radio-frequency (RF) pulse of duration T seconds and frequency f_c , shown in Figure 2.7. The spectrum of this pulse has main spectral lobes of width $(2/T)$ hertz centered around $\pm f_c$, where it is assumed that f_c is large compared with $(1/T)$. Hence, we define the null-to-null bandwidth of the RF pulse of Figure 2.7 as $(2/T)$ hertz.

On the basis of the definitions presented here, we may state that shifting the spectral content of a low-pass signal by a sufficiently large frequency has the effect of doubling the bandwidth of the signal; this frequency translation is attained by using the process of modulation. Basically, the modulation moves the spectral content of the signal for negative frequencies into the positive frequency region, whereupon the negative frequencies become physically measurable.

Another popular definition of bandwidth is the *3 dB bandwidth*. Specifically, if the signal is low-pass, we say:

The 3 dB bandwidth of a low-pass signal is defined as the separation between zero frequency, where the magnitude spectrum attains its peak value, and the positive frequency at which the amplitude spectrum drops to $1/\sqrt{2}$ of its peak value.

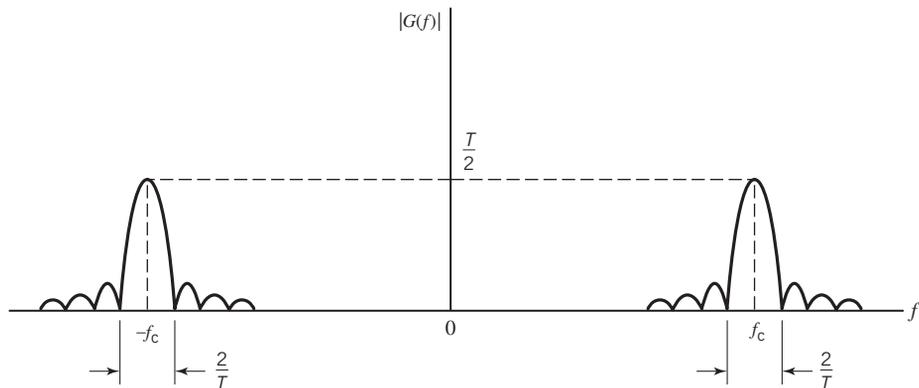


Figure 2.7 Magnitude spectrum of the RF pulse, showing the null-to-null bandwidth to be $2/T$, centered on the mid-band frequency f_c .

For example, the decaying exponential function $\exp(-at)$ has a 3 dB bandwidth of $(a/2\pi)$ hertz.

If, on the other hand, the signal is of a band-pass kind, centered at $\pm f_c$, the 3 dB bandwidth is defined as the separation (along the positive frequency axis) between the two frequencies at which the magnitude spectrum of the signal drops to $1/\sqrt{2}$ of its peak value at f_c .

Regardless of whether we have a low-pass or band-pass signal, the 3 dB bandwidth has the advantage that it can be read directly from a plot of the magnitude spectrum. However, it has the disadvantage that it may be misleading if the magnitude spectrum has slowly decreasing tails.

Time–Bandwidth Product

For any family of pulse signals that differ by a time-scaling factor, the product of the signal's duration and its bandwidth is always a constant, as shown by

$$\text{duration} \times \text{bandwidth} = \text{constant}$$

This product is called the *time–bandwidth product*. The constancy of the time–bandwidth product is another manifestation of the inverse relationship that exists between the time-domain and frequency-domain descriptions of a signal. In particular, if the duration of a pulse signal is decreased by reducing the time scale by a factor a , the frequency scale of the signal's spectrum, and therefore the bandwidth of the signal is increased by the same factor a . This statement follows from the *dilation property* of the Fourier transform (defined in Property 2 of Table 2.1). The time–bandwidth product of the signal is therefore maintained constant. For example, a rectangular pulse of duration T seconds has a bandwidth (defined on the basis of the positive-frequency part of the main lobe) equal to $(1/T)$ hertz; in this example, the time–bandwidth product of the pulse equals unity.

The important point to take from this discussion is that whatever definitions we use for the bandwidth and duration of a signal, the time–bandwidth product remains constant over certain classes of pulse signals; the choice of particular definitions for bandwidth and duration merely change the value of the constant.

Root-Mean-Square Definitions of Bandwidth and Duration

To put matters pertaining to the bandwidth and duration of a signal on a firm mathematical basis, we first introduce the following definition for bandwidth:

The root-mean-square (rms) bandwidth is defined as the square root of the second moment of a normalized form of the squared magnitude spectrum of the signal about a suitably chosen frequency.

To be specific, we assume that the signal $g(t)$ is of a low-pass kind, in which case the second moment is taken about the origin $f = 0$. The squared magnitude spectrum of the signal is denoted by $|G(f)|^2$. To formulate a nonnegative function, the total area under whose curve is unity, we use the normalizing function

$$\int_{-\infty}^{\infty} |G(f)|^2 df$$

We thus mathematically define the rms bandwidth of a low-pass signal $g(t)$ with Fourier transform $G(f)$ as

$$W_{\text{rms}} = \left(\frac{\int_{-\infty}^{\infty} f^2 |G(f)|^2 df}{\int_{-\infty}^{\infty} |G(f)|^2 df} \right)^{1/2} \quad (2.26)$$

which describes the dispersion of the spectrum $G(f)$ around $f = 0$. An attractive feature of the rms bandwidth W_{rms} is that it lends itself readily to mathematical evaluation. But, it is not as easily measurable in the laboratory.

In a manner corresponding to the rms bandwidth, the *rms duration* of the signal $g(t)$ is mathematically defined by

$$T_{\text{rms}} = \left(\frac{\int_{-\infty}^{\infty} t^2 |g(t)|^2 dt}{\int_{-\infty}^{\infty} |g(t)|^2 dt} \right)^{1/2} \quad (2.27)$$

where it is assumed that the signal $g(t)$ is centered around the origin $t = 0$. In Problem 2.7, it is shown that, using the rms definitions of (2.26) and (2.27), the time–bandwidth product takes the form

$$T_{\text{rms}} W_{\text{rms}} \geq \frac{1}{4\pi} \quad (2.28)$$

In Problem 2.7, it is also shown that the Gaussian pulse $\exp(-\pi t^2)$ satisfies this condition exactly with the equality sign.

2.5 The Dirac Delta Function

Strictly speaking, the theory of the Fourier transform, presented in Section 2.3, is applicable only to time functions that satisfy the Dirichlet conditions. As mentioned previously, such functions naturally include energy signals. However, it would be highly desirable to extend this theory in two ways:

1. To combine the Fourier series and Fourier transform into a unified theory, so that the Fourier series may be treated as a special case of the Fourier transform.
2. To include power signals in the list of signals to which we may apply the Fourier transform. A signal $g(t)$ is said to be a *power signal* if the condition

$$\frac{1}{T} \int_{-T/2}^{T/2} |g(t)|^2 dt < \infty$$

holds, where T is the observation interval.

It turns out that both of these objectives can be met through the “proper use” of the *Dirac delta function*, or *unit impulse*.

The Dirac delta function³ or just delta function, denoted by $\delta(t)$, is defined as having zero amplitude everywhere except at $t = 0$, where it is infinitely large in such a way that it contains unit area under its curve; that is,

$$\delta(t) = 0, \quad t \neq 0 \quad (2.29)$$

and

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (2.30)$$

An implication of this pair of relations is that the delta function $\delta(t)$ is an even function of time t , centered at the origin $t = 0$. Perhaps, the simplest way of describing the Dirac delta function is to view it as the rectangular pulse

$$g(t) = \frac{1}{T} \operatorname{rect}\left(\frac{t}{T}\right)$$

whose duration is T and amplitude is $1/T$, as illustrated in Figure 2.8. As T approaches zero, the rectangular pulse $g(t)$ approaches the Dirac delta function $\delta(t)$ in the limit.

For the delta function to have meaning, however, it has to appear as a factor in the integrand of an integral with respect to time, and then, strictly speaking, only when the other factor in the integrand is a continuous function of time. Let $g(t)$ be such a function, and consider the product of $g(t)$ and the time-shifted delta function $\delta(t - t_0)$. In light of the two defining equations (2.29) and (2.30), we may express the integral of this product as

$$\int_{-\infty}^{\infty} g(t) \delta(t - t_0) dt = g(t_0) \quad (2.31)$$

The operation indicated on the left-hand side of this equation sifts out the value $g(t_0)$ of the function $g(t)$ at time $t = t_0$, where $-\infty < t < \infty$. Accordingly, (2.31) is referred to as the *sifting property* of the delta function. This property is sometimes used as the defining equation of a delta function; in effect, it incorporates (2.29) and (2.30) into a single relation.

Noting that the delta function $\delta(t)$ is an even function of t , we may rewrite (2.31) so as to emphasize its resemblance to the convolution integral, as shown by

$$\int_{-\infty}^{\infty} g(\tau) \delta(t - \tau) d\tau = g(t) \quad (2.32)$$

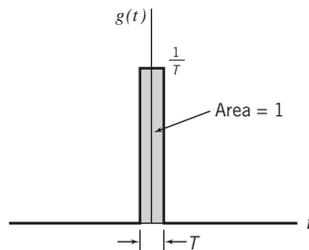


Figure 2.8 Illustrative example of the Dirac delta function as the limiting form of rectangular pulse $\frac{1}{T} \operatorname{rect}\left(\frac{t}{T}\right)$ as T approaches zero.

In words, the convolution of any function with the delta function leaves that function unchanged. We refer to this statement as the *replication property* of the delta function.

It is important to realize that no function in the ordinary sense has the two properties of (2.29) and (2.30) or the equivalent sifting property of (2.31). However, we can imagine a sequence of functions that have progressively taller and thinner peaks at $t = 0$, with the area under the curve consistently remaining equal to unity; as this progression is being performed, the value of the function tends to zero at every point except $t = 0$, where it tends to infinity, as illustrated in Figure 2.8, for example. We may therefore say:

The delta function may be viewed as the limiting form of a pulse of unit area as the duration of the pulse approaches zero.

It is immaterial what sort of pulse shape is used, so long as it is symmetric with respect to the origin; this symmetry is needed to maintain the “even” function property of the delta function.

Two other points are noteworthy:

1. Applicability of the delta function is not confined to the time domain. Rather, it can equally well be applied in the frequency domain; all that we have to do is to replace time t by frequency f in the defining equations (2.29) and (2.30).
2. The area covered by the delta function defines its “strength.” As such, the units, in terms of which the strength is measured, are determined by the specifications of the two coordinates that define the delta function.

EXAMPLE 3

The Sinc Function as a Limiting Form of the Delta Function in the Time Domain

As another illustrative example, consider the scaled sinc function $2W\text{sinc}(2Wt)$, whose waveform covers an area equal to unity for all W .

Figure 2.9 displays the evolution of this time function toward the delta function as the parameter W is varied in three stages: $W = 1$, $W = 2$, and $W = 5$. Referring back to Figure 2.4, we may infer that as the parameter W characterizing the sinc pulse is increased, the amplitude of the pulse at time $t = 0$ increases linearly, while at the same time the duration of the main lobe of the pulse decreases inversely. With this objective in mind, as the parameter W is progressively increased, Figure 2.9 teaches us two important things:

1. The scaled sinc function becomes more like a delta function.
2. The constancy of the function’s spectrum is maintained at unity across an increasingly wider frequency band, in accordance with the constraint that the area under the function is to remain constant at unity; see Property 6 of Table 2.1 for a validation of this point.

Based on the trend exhibited in Figure 2.9, we may write

$$\delta(t) = \lim_{W \rightarrow \infty} 2W \text{sinc}(2Wt) \quad (2.33)$$

which, in addition to the rectangular pulse considered in Figure 2.8, is another way of realizing a delta function in the time domain.

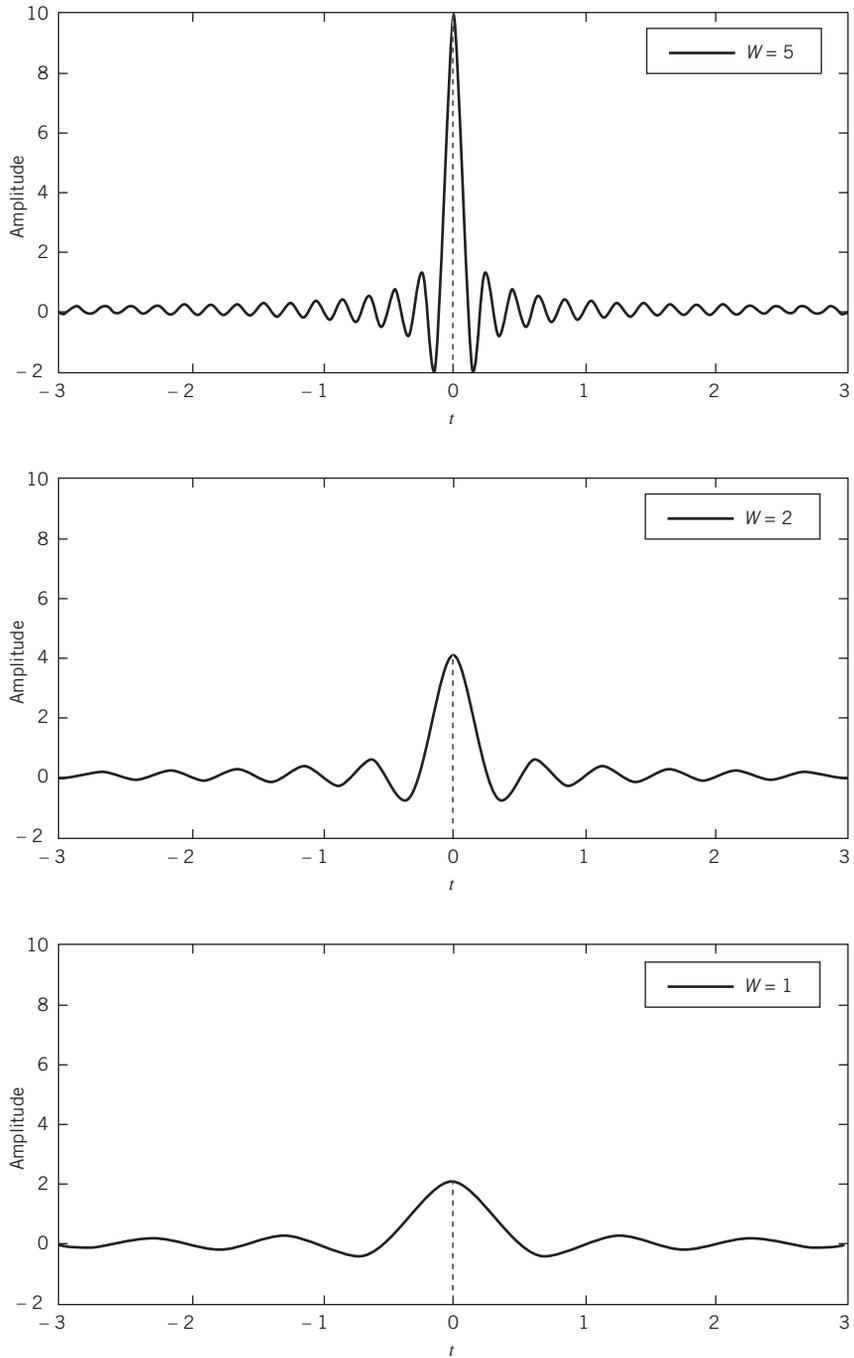


Figure 2.9 Evolution of the sinc function $2W \operatorname{sinc}(2Wi)$ toward the delta function as the parameter W progressively increases.

EXAMPLE 4

Evolution of the Sum of Complex Exponentials toward the Delta Function in the Frequency Domain

For yet another entirely different example, consider the infinite summation term

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mf) \text{ over the interval } -1/2 \leq f < 1/2. \text{ Using Euler's formula}$$

$$\exp(j2\pi mf) = \cos(2\pi mf) + j \sin(2\pi mf)$$

we may express the given summation as

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mf) = \sum_{m=-\infty}^{\infty} \cos(2\pi mf) + j \sum_{m=-\infty}^{\infty} \sin(2\pi mf)$$

The imaginary part of the summation is zero for two reasons. First, $\sin(2\pi mf)$ is zero for $m = 0$. Second, since $\sin(-2\pi mf) = -\sin(2\pi mf)$, the remaining imaginary terms cancel each other. Therefore,

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mf) = \sum_{m=-\infty}^{\infty} \cos(2\pi mf)$$

Figure 2.10 plots this real-valued summation versus frequency f over the interval $-1/2 \leq f < 1/2$ for three ranges of m :

1. $-5 \leq m \leq 5$
2. $-10 \leq m \leq 10$
3. $-20 \leq m \leq 20$

Building on the results exhibited in Figure 2.10, we may go on to say

$$\delta(f) = \sum_{m=-\infty}^{\infty} \cos(2\pi mf), \quad -\frac{1}{2} \leq f < \frac{1}{2} \quad (2.34)$$

which is one way of realizing a delta function in the frequency domain. Note that the area under the summation term on the right-hand side of (2.34) is equal to unity; we say so because

$$\begin{aligned} \int_{-1/2}^{1/2} \sum_{m=-\infty}^{\infty} \cos(2\pi mf) \, df &= \sum_{m=-\infty}^{\infty} \int_{-1/2}^{1/2} \cos(2\pi mf) \, df \\ &= \sum_{m=-\infty}^{\infty} \left[\frac{\sin(2\pi mf)}{2\pi m} \right]_{f=-1/2}^{1/2} \\ &= \sum_{m=-\infty}^{\infty} \left[\frac{\sin(\pi m)}{\pi m} \right] \\ &= \begin{cases} 1 & \text{for } m = 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

This result, formulated in the frequency domain, confirms (2.34) as one way of defining the delta function $\delta(f)$.

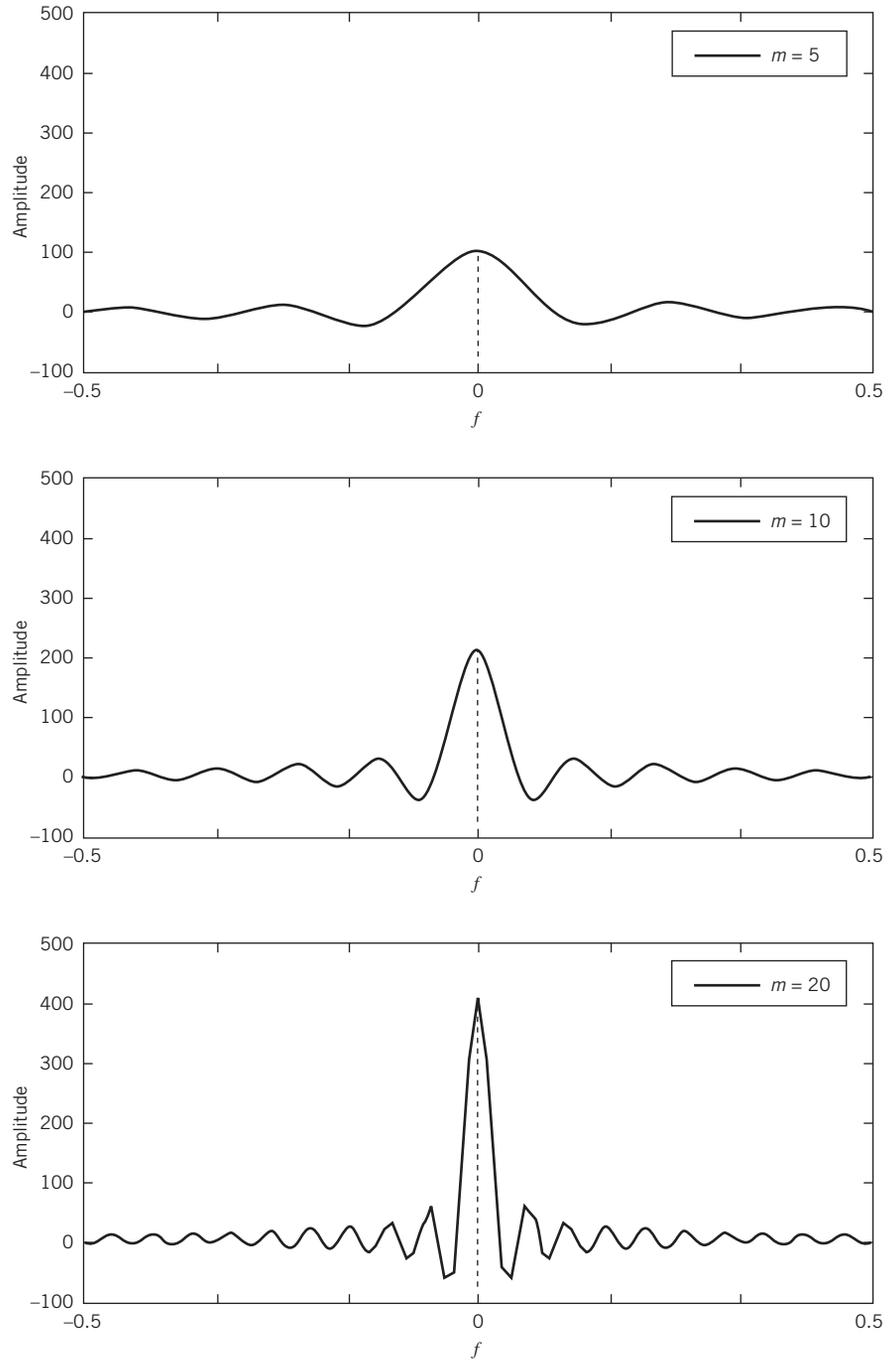


Figure 2.10 Evolution of the sum of m complex exponentials toward a delta function in the frequency domain as m becomes increasingly larger.

2.6 Fourier Transforms of Periodic Signals

We began the study of Fourier analysis by reviewing the Fourier series expansion of periodic signals, which, in turn, paved the way for the formulation of the Fourier transform. Now that we have equipped ourselves with the Dirac delta function, we would like to revisit the Fourier series and show that it can indeed be treated as a special case of the Fourier transform.

To this end, let $g(t)$ be a pulse-like function, which equals a periodic signal $g_{T_0}(t)$ over one period T_0 of the signal and is zero elsewhere, as shown by

$$g(t) = \begin{cases} g_{T_0}(t), & -\frac{T_0}{2} < t \leq \frac{T_0}{2} \\ 0, & \text{elsewhere} \end{cases} \quad (2.35)$$

The periodic signal $g_{T_0}(t)$ itself may be expressed in terms of the function $g(t)$ as an infinite summation, as shown by

$$g_{T_0}(t) = \sum_{m=-\infty}^{\infty} g(t - mT_0) \quad (2.36)$$

In light of the definition of the pulselike function $g(t)$ in (2.35), we may view this function as a *generating function*, so called as it generates the periodic signal $g_{T_0}(t)$ in accordance with (2.36).

Clearly, the generating function $g(t)$ is Fourier transformable; let $G(f)$ denote its Fourier transform. Correspondingly, let $G_{T_0}(f)$ denote the Fourier transform of the periodic signal $g_{T_0}(t)$. Hence, taking the Fourier transforms of both sides of (2.36) and applying the time-shifting property of the Fourier transform (Property 4 of Table 2.1), we may write

$$G_{T_0}(f) = G(f) \sum_{m=-\infty}^{\infty} \exp(-j2\pi mfT_0), \quad -\infty < f < \infty \quad (2.37)$$

where we have taken $G(f)$ outside the summation because it is independent of m .

In Example 4, we showed that

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mf) = \sum_{m=-\infty}^{\infty} \cos(j2\pi mf) = \delta(f), \quad -\frac{1}{2} \leq f < \frac{1}{2}$$

Let this result be expanded to cover the entire frequency range, as shown by

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mf) = \sum_{n=-\infty}^{\infty} \delta(f - n), \quad -\infty < f < \infty \quad (2.38)$$

Equation (2.38) (see Problem 2.8c) represents a *Dirac comb*, consisting of an infinite sequence of uniformly spaced delta functions, as depicted in Figure 2.11.

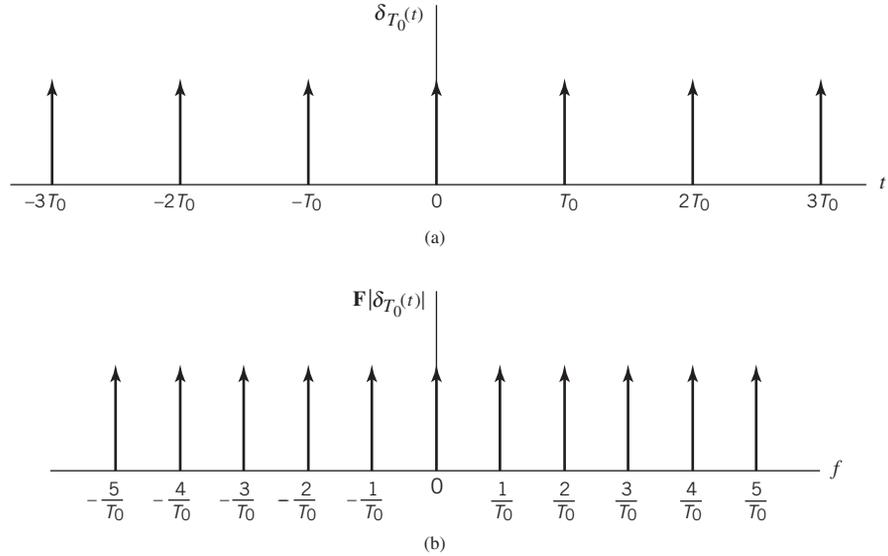


Figure 2.11 (a) Dirac comb. (b) Spectrum of the Dirac comb.

Next, introducing the frequency-scaling factor $f_0 = 1/T_0$ into (2.38), we correspondingly write

$$\sum_{m=-\infty}^{\infty} \exp(j2\pi mfT_0) = f_0 \sum_{n=-\infty}^{\infty} \delta(f - nf_0), \quad -\infty < f < \infty \quad (2.39)$$

Hence, substituting (2.39) into the right-hand side of (2.37), we get

$$\begin{aligned} G_{T_0}(f) &= f_0 G(f) \sum_{n=-\infty}^{\infty} \delta(f - nf_0) \\ &= f_0 \sum_{n=-\infty}^{\infty} G(f_n) \delta(f - f_n), \quad -\infty < f < \infty \end{aligned} \quad (2.40)$$

where $f_n = nf_0$.

What we have to show next is that the inverse Fourier transform of $G_{T_0}(f)$ defined in (2.40) is exactly the same as in the Fourier series formula of (2.14). Specifically, substituting (2.40) into the inverse Fourier transform formula of (2.17), we get

$$g_{T_0}(t) = f_0 \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} G(f_n) \delta(f - f_n) \right] \exp(j2\pi ft) df$$

Interchanging the order of summation and integration, and then invoking the sifting property of the Dirac delta function (this time in the frequency domain), we may go on to write

$$\begin{aligned} g_{T_0}(t) &= f_0 \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} G(f_n) \exp(j2\pi f t) \delta(f - f_n) df \\ &= f_0 \sum_{n=-\infty}^{\infty} G(f_n) \exp(j2\pi f_n t) \end{aligned}$$

which is an exact rewrite of (2.14) with $f_0 = \Delta f$. Equivalently, in light of (2.36), we may formulate the Fourier transform pair

$$\sum_{m=-\infty}^{\infty} g(t - mT_0) = f_0 \sum_{n=-\infty}^{\infty} G(f_n) \exp(j2\pi f_n t) \quad (2.41)$$

The result derived in (2.41) is one form of *Poisson's sum formula*.

We have thus demonstrated that the Fourier series representation of a periodic signal is embodied in the Fourier transformation of (2.16) and (2.17), provided, of course, we permit the use of the Dirac delta function. In so doing, we have closed the “circle” by going from the Fourier series to the Fourier transform, and then back to the Fourier series.

Consequences of Ideal Sampling

Consider a Fourier transformable pulselike signal $g(t)$ with its Fourier transform denoted by $G(f)$. Setting $f_n = nf_0$ in (2.41) and using (2.38), we may express Poisson's sum formula

$$\sum_{m=-\infty}^{\infty} g(t - mT_0) \Leftrightarrow f_0 \sum_{n=-\infty}^{\infty} G(nf_0) \delta(f - nf_0) \quad (2.42)$$

where $f_0 = 1/T_0$. The summation on the left-hand side of this Fourier-transform pair is a periodic signal with period T_0 . The summation on the right-hand side of the pair is a uniformly sampled version of the spectrum $G(f)$. We may therefore make the following statement:

Uniform sampling of the spectrum $G(f)$ in the frequency domain introduces periodicity of the function $g(t)$ in the time domain.

Applying the duality property of the Fourier transform (Property 3 of Table 2.1) to (2.42), we may also write

$$T_0 \sum_{m=-\infty}^{\infty} g(mT_0) \delta(t - mT_0) \Leftrightarrow \sum_{n=-\infty}^{\infty} G(f - nf_0) \quad (2.43)$$

in light of which we may make the following dual statement:

Uniform sampling of the Fourier transformable function $g(t)$ in the time domain introduces periodicity of the spectrum $G(f)$ in the frequency domain.

2.7 Transmission of Signals through Linear Time-Invariant Systems

A *system* refers to any physical entity that produces an output signal in response to an input signal. It is customary to refer to the input signal as the *excitation* and to the output signal as the *response*. In a linear system, the *principle of superposition* holds; that is, the response of a linear system to a number of excitations applied simultaneously is equal to the sum of the responses of the system when each excitation is applied individually.

In the time domain, a linear system is usually described in terms of its *impulse response*, which is formally defined as follows:

The impulse response of a linear system is the response of the system (with zero initial conditions) to a unit impulse or delta function $\delta(t)$ applied to the input of the system at time $t = 0$.

If the system is also *time invariant*, then the shape of the impulse response is the same no matter when the unit impulse is applied to the system. Thus, with the unit impulse or delta function applied to the system at time $t = 0$, the impulse response of a linear time-invariant system is denoted by $h(t)$.

Suppose that a system described by the impulse response $h(t)$ is subjected to an arbitrary excitation $x(t)$, as depicted in Figure 2.12. The resulting response of the system $y(t)$, is defined in terms of the impulse response $h(t)$ by

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau) d\tau \quad (2.44)$$

which is called the *convolution integral*. Equivalently, we may write

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t - \tau) d\tau \quad (2.45)$$

Equations (2.44) and (2.45) state that convolution is *commutative*.

Examining the convolution integral of (2.44), we see that three different time scales are involved: *excitation time* τ , *response time* t , and *system-memory time* $t - \tau$. This relation is the basis of time-domain analysis of linear time-invariant systems. According to (2.44), the present value of the response of a linear time-invariant system is an integral over the past history of the input signal, weighted according to the impulse response of the system. Thus, the impulse response acts as a *memory function* of the system.

Causality and Stability

A linear system with impulse response $h(t)$ is said to be *causal* if its impulse response $h(t)$ satisfies the condition

$$h(t) = 0 \quad \text{for } t < 0$$

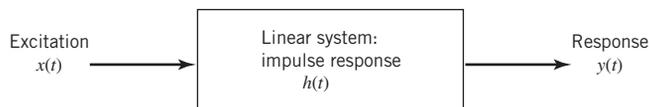


Figure 2.12 Illustrating the roles of excitation $x(t)$, impulse response $h(t)$, and response $y(t)$ in the context of a linear time-invariant system.

The essence of causality is that no response can appear at the output of the system before an excitation is applied to its input. Causality is a necessary requirement for on-line operation of the system. In other words, for a system operating in *real time* to be physically realizable, it has to be causal.

Another important property of a linear system is *stability*. A necessary and sufficient condition for the system to be stable is that its impulse response $h(t)$ must satisfy the inequality

$$\int_{-\infty}^{\infty} |h(t)| dt < \infty$$

This requirement follows from the commonly used criterion of *bounded input–bounded output*. Basically, for the system to be stable, its impulse response must be *absolutely integrable*.

Frequency Response

Let $X(f)$, $H(f)$, and $Y(f)$ denote the Fourier transforms of the excitation $x(t)$, impulse response $h(t)$, and response $y(t)$, respectively. Then, applying Property 12 of the Fourier transform in Table 2.1 to the convolution integral, be it written in the form of (2.44) or (2.45), we get

$$Y(f) = H(f)X(f) \quad (2.46)$$

Equivalently, we may write

$$H(f) = \frac{Y(f)}{X(f)} \quad (2.47)$$

The new frequency function $H(f)$ is called the *transfer function* or *frequency response* of the system; these two terms are used interchangeably. Based on (2.47), we may now formally say:

The frequency response of a linear time-invariant system is defined as the ratio of the Fourier transform of the response of the system to the Fourier transform of the excitation applied to the system.

In general, the frequency response $H(f)$ is a complex quantity, so we may express it in the form

$$H(f) = |H(f)| \exp[j\beta(f)] \quad (2.48)$$

where $|H(f)|$ is called the *magnitude response*, and $\beta(f)$ is the *phase response*, or simply *phase*. When the impulse response of the system is real valued, the frequency response exhibits conjugate symmetry, which means that

$$|H(f)| = |H(-f)|$$

and

$$\beta(f) = -\beta(-f)$$

That is, the magnitude response $|H(f)|$ of a linear system with real-valued impulse response is an even function of frequency, whereas the phase $\beta(f)$ is an odd function of frequency.

In some applications it is preferable to work with the logarithm of $H(f)$ expressed in polar form, rather than with $H(f)$ itself. Using \ln to denote the natural logarithm, let

$$\ln H(f) = \alpha(f) + j\beta(f) \quad (2.49)$$

where

$$\alpha(f) = \ln|H(f)| \quad (2.50)$$

The function $\alpha(f)$ is called the *gain* of the system; it is measured in *neper*s. The phase $\beta(f)$ is measured in *radians*. Equation (2.49) indicates that the gain $\alpha(f)$ and phase $\beta(f)$ are, respectively, the real and imaginary parts of the (natural) logarithm of the transfer function $H(f)$. The gain may also be expressed in *decibels* (dB) by using the definition

$$\alpha'(f) = 20\log_{10}|H(f)|$$

The two gain functions $\alpha(f)$ and $\alpha'(f)$ are related by

$$\alpha'(f) = 8.69\alpha(f)$$

That is, 1 neper is equal to 8.69 dB.

As a means of specifying the constancy of the magnitude response $|H(f)|$ or gain $\alpha(f)$ of a system, we use the notion of *bandwidth*. In the case of a low-pass system, the bandwidth is customarily defined as the frequency at which the magnitude response $|H(f)|$ is $1/\sqrt{2}$ times its value at zero frequency or, equivalently, the frequency at which the gain $\alpha'(f)$ drops by 3 dB below its value at zero frequency, as illustrated in Figure 2.13a. In the case of a band-pass system, the bandwidth is defined as the range of frequencies over which the magnitude response $|H(f)|$ remains within $1/\sqrt{2}$ times its value at the mid-band frequency, as illustrated in Figure 2.13b.

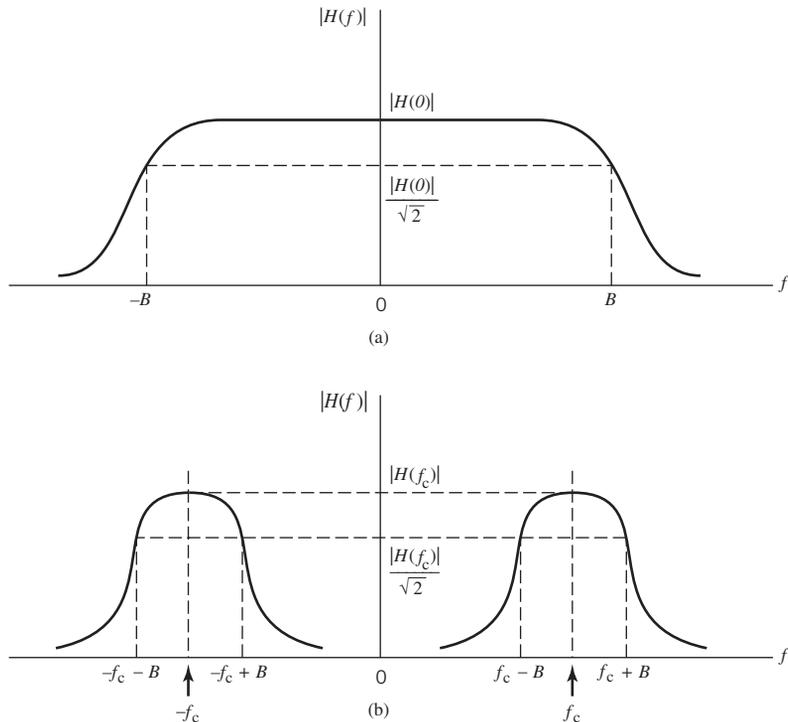


Figure 2.13 Illustrating the definition of system bandwidth. (a) Low-pass system. (b) Band-pass system.

Paley–Wiener Criterion: Another Way of Assessing Causality

A necessary and sufficient condition for a function $\alpha(f)$ to be the gain of a causal filter is the convergence of the integral

$$\int_{-\infty}^{\infty} \frac{|\alpha(f)|}{1+f^2} df < \infty \quad (2.51)$$

This condition is known as the *Paley–Wiener criterion*.⁴ The criterion states that provided the gain $\alpha(f)$ satisfies the condition of (2.51), then we may associate with this gain a suitable phase $\beta(f)$, such that the resulting filter has a causal impulse response that is zero for negative time. In other words, the Paley–Wiener criterion is the frequency-domain equivalent of the causality requirement. A system with a realizable gain characteristic may have infinite attenuation for a discrete set of frequencies, but it cannot have infinite attenuation over a band of frequencies; otherwise, the Paley–Wiener criterion is violated.

Finite-Duration Impulse Response (FIR) Filters

Consider next a linear time-invariant filter with impulse response $h(t)$. We make two assumptions:

1. *Causality*, which means that the impulse response $h(t)$ is zero for $t < 0$.
2. *Finite support*, which means that the impulse response of the filter is of some finite duration T_f , so that we may write $h(t) = 0$ for $t \geq T_f$.

Under these two assumptions, we may express the filter output $y(t)$ produced in response to the input $x(t)$ as

$$y(t) = \int_0^{T_f} h(\tau)x(t - \tau) d\tau \quad (2.52)$$

Let the input $x(t)$, impulse response $h(t)$, and output $y(t)$ be *uniformly sampled* at the rate $(1/\Delta\tau)$ samples per second, so that we may put

$$t = n\Delta\tau$$

and

$$\tau = k\Delta\tau$$

where k and n are integers and $\Delta\tau$ is the *sampling period*. Assuming that $\Delta\tau$ is small enough for the product $h(\tau)x(t - \tau)$ to remain essentially constant for $k\Delta\tau \leq \tau \leq (k + 1)\Delta\tau$ for all values of k and τ , we may approximate (2.52) by the *convolution sum*

$$y(n\Delta\tau) = \sum_{k=0}^{N-1} h(k\Delta\tau)x(n\Delta\tau - k\Delta\tau)\Delta\tau$$

where $N\Delta\tau = T_f$. To simplify the notations used in this summation formula, we introduce three definitions:

$$w_k = h(k\Delta\tau)\Delta\tau$$

$$x(n\Delta\tau) = x_n$$

$$y(n\Delta\tau) = y_n$$

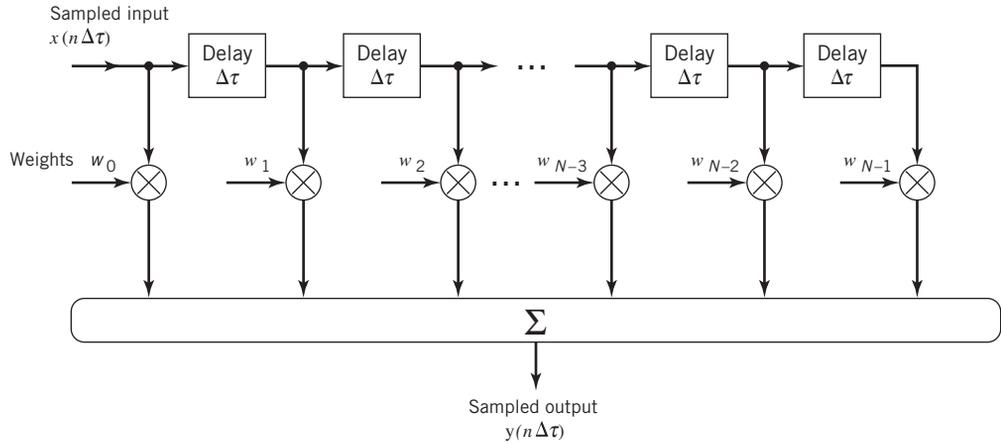


Figure 2.14 Tapped-delay-line (TDL) filter; also referred to as FIR filter.

We may then rewrite the formula for $y(n\Delta\tau)$ in the compact form

$$y_n = \sum_{k=0}^{N-1} w_k x_{n-k}, \quad n = 0, \pm 1, \pm 2, \dots \quad (2.53)$$

Equation (2.53) may be realized using the structure shown in Figure 2.14, which consists of a set of *delay elements* (each producing a delay of $\Delta\tau$ seconds), a set of *multipliers* connected to the *delay-line taps*, a corresponding set of *weights* supplied to the multipliers, and a *summer* for adding the multiplier outputs. The sequences x_n and y_n , for integer values of n as described in (2.53), are referred to as the *input* and *output sequences*, respectively.

In the digital signal-processing literature, the structure of Figure 2.14 is known as a *finite-duration impulse response (FIR) filter*. This filter offers some highly desirable practical features:

1. The filter is inherently *stable*, in the sense that a bounded input sequence produces a bounded output sequence.
2. Depending on how the weights $\{w_k\}_{k=0}^{N-1}$ are designated, the filter can perform the function of a low-pass filter or band-pass filter. Moreover, the phase response of the filter can be configured to be a linear function of frequency, which means that there will be no delay distortion.
3. In a digital realization of the filter, the filter assumes a *programmable* form whereby the application of the filter can be changed merely by making appropriate changes to the weights, leaving the structure of the filter completely unchanged; this kind of flexibility is not available with analog filters.

We will have more to say on the FIR filter in subsequent chapters of the book.

2.8 Hilbert Transform

The Fourier transform is particularly useful for evaluating the frequency content of an energy signal or, in a limiting sense, that of a power signal. As such, it provides the mathematical basis for analyzing and designing *frequency-selective filters* for the separation of signals on the basis of their frequency content. Another method of separating signals is based on *phase selectivity*, which uses phase shifts between the pertinent signals to achieve the desired separation. A phase shift of special interest in this context is that of $\pm 90^\circ$. In particular, when the phase angles of all components of a given signal are shifted by $\pm 90^\circ$, the resulting function of time is known as the *Hilbert transform* of the signal. The Hilbert transform is called a *quadrature filter*; it is so called to emphasize its distinct property of providing a $\pm 90^\circ$ phase shift.

To be specific, consider a Fourier transformable signal $g(t)$ with its Fourier transform denoted by $G(f)$. The *Hilbert transform* of $g(t)$, which we denote by $\hat{g}(t)$, is defined by⁵

$$\hat{g}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(\tau)}{t - \tau} d\tau \quad (2.54)$$

Table 2.3 Hilbert-transform pairs*

Time function	Hilbert transform
1. $m(t)\cos(2\pi f_c t)$	$m(t)\sin(2\pi f_c t)$
2. $m(t)\sin(2\pi f_c t)$	$-m(t)\cos(2\pi f_c t)$
3. $\cos(2\pi f_c t)$	$\sin(2\pi f_c t)$
4. $\sin(2\pi f_c t)$	$-\cos(2\pi f_c t)$
5. $\frac{\sin t}{t}$	$\frac{1 - \cos t}{t}$
6. $\text{rect}(t)$	$-\frac{1}{\pi} \ln \left \frac{t - 1/2}{t + 1/2} \right $
7. $\delta(t)$	$\frac{1}{\pi t}$
8. $\frac{1}{1 + t^2}$	$\frac{t}{1 + t^2}$
9. $\frac{1}{t}$	$-\pi \delta(t)$

Notes: $\delta(t)$ denotes Dirac delta function; $\text{rect}(t)$ denotes rectangular function; \ln denotes natural logarithm.

* In the first two pairs, it is assumed that $m(t)$ is band limited to the interval $-W \leq f \leq W$, where $W < f_c$.

Clearly, Hilbert transformation is a linear operation. The *inverse Hilbert transform*, by means of which the original signal $g(t)$ is linearly recovered from $\hat{g}(t)$, is defined by

$$g(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\hat{g}(\tau)}{t - \tau} d\tau \quad (2.55)$$

The functions $g(t)$ and $\hat{g}(t)$ are said to constitute a *Hilbert-transform pair*. A short table of Hilbert-transform pairs is given in Table 2.3 on page 42.

The definition of the Hilbert transform $\hat{g}(t)$ given in (2.54) may be interpreted as the convolution of $g(t)$ with the time function $1/(\pi t)$. We know from the convolution theorem listed in Table 2.1 that the convolution of two functions in the time domain is transformed into the multiplication of their Fourier transforms in the frequency domain.

For the time function $1/(\pi t)$, we have the Fourier-transform pair (see Property 14 in Table 2.2)

$$\frac{1}{\pi t} \Leftrightarrow -j \operatorname{sgn}(f)$$

where $\operatorname{sgn}(f)$ is the *signum function*, defined in the frequency domain as

$$\operatorname{sgn}(f) = \begin{cases} 1, & f > 0 \\ 0, & f = 0 \\ -1, & f < 0 \end{cases} \quad (2.56)$$

It follows, therefore, that the Fourier transform $\hat{G}(f)$ of $\hat{g}(t)$ is given by

$$\hat{G}(f) = -j \operatorname{sgn}(f) G(f) \quad (2.57)$$

Equation (2.57) states that given a Fourier transformable signal $g(t)$, we may obtain the Fourier transform of its Hilbert transform $\hat{g}(t)$ by passing $g(t)$ through a linear time-invariant system whose frequency response is equal to $-j \operatorname{sgn}(f)$. This system may be considered as one that produces a phase shift of -90° for all positive frequencies of the input signal and $+90^\circ$ degrees for all negative frequencies, as in Figure 2.15. The amplitudes of all frequency components in the signal, however, are unaffected by transmission through the device. Such an ideal system is referred to as a *Hilbert transformer*, or *quadrature filter*.

Properties of the Hilbert Transform

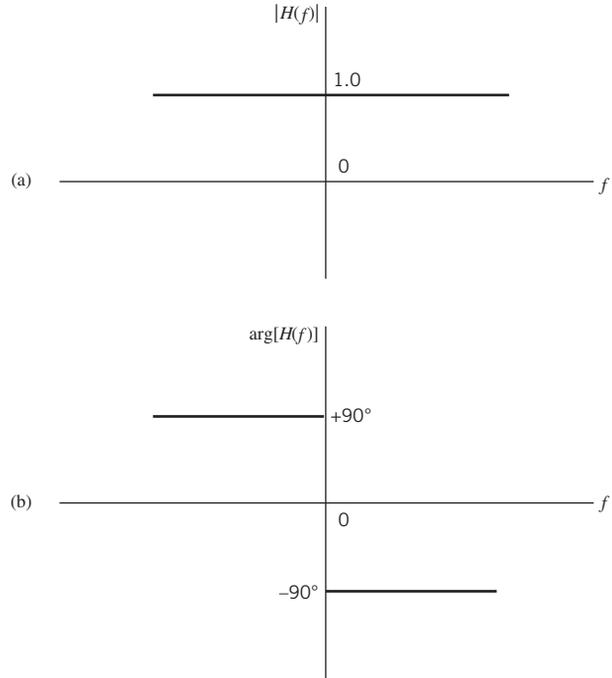
The Hilbert transform differs from the Fourier transform in that it operates exclusively in the time domain. It has a number of useful properties of its own, some of which are listed next. The signal $g(t)$ is assumed to be real valued, which is the usual domain of application of the Hilbert transform. For this class of signals, the Hilbert transform has the following properties.

PROPERTY 1 A signal $g(t)$ and its Hilbert transform $\hat{g}(t)$ have the same magnitude spectrum.

That is to say,

$$|G(f)| = |\hat{G}(f)|$$

Figure 2.15
 (a) Magnitude response and
 (b) phase response of Hilbert
 transform.



PROPERTY 2 If $\hat{g}(t)$ is the Hilbert transform of $g(t)$, then the Hilbert transform of $\hat{g}(t)$ is $-g(t)$.

Another way of stating this property is to write

$$\arg[G(f)] = -\arg\{\hat{G}(f)\}$$

PROPERTY 3 A signal $g(t)$ and its Hilbert transform $\hat{g}(t)$ are orthogonal over the entire time interval $(-\infty, \infty)$.

In mathematical terms, the orthogonality of $g(t)$ and $\hat{g}(t)$ is described by

$$\int_{-\infty}^{\infty} g(t)\hat{g}(t)dt = 0$$

Proofs of these properties follow from (2.54), (2.55), and (2.57).

EXAMPLE 5 Hilbert Transform of Low-Pass Signal

Consider Figure 2.16a that depicts the Fourier transform of a low-pass signal $g(t)$, whose frequency content extends from $-W$ to W . Applying the Hilbert transform to this signal yields a new signal $\hat{g}(t)$ whose Fourier transform, $\hat{G}(f)$, is depicted in Figure 2.16b. This figure illustrates that the frequency content of a Fourier transformable signal can be radically changed as a result of Hilbert transformation.

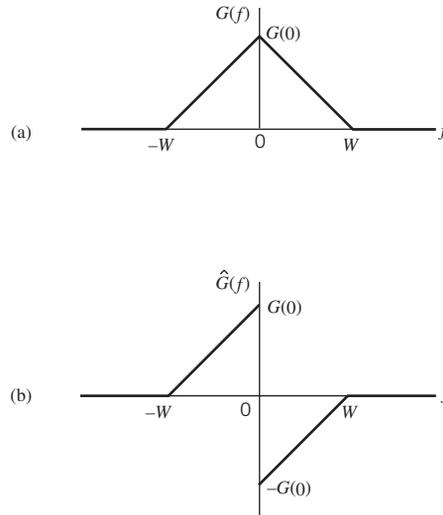


Figure 2.16 Illustrating application of the Hilbert transform to a low-pass signal: (a) Spectrum of the signal $g(t)$; (b) Spectrum of the Hilbert transform $\hat{g}(t)$.

2.9 Pre-envelopes

The Hilbert transform of a signal is defined for both positive and negative frequencies. In light of the spectrum shaping illustrated in Example 5, a question that begs itself is:

How can we modify the frequency content of a real-valued signal $g(t)$ such that all negative frequency components are completely eliminated?

The answer to this fundamental question lies in the idea of a complex-valued signal called the *pre-envelope*⁶ of $g(t)$, formally defined as

$$g_+(t) = g(t) + j\hat{g}(t) \quad (2.58)$$

where $\hat{g}(t)$ is the Hilbert transform of $g(t)$. According to this definition, the given signal $g(t)$ is the real part of the pre-envelope $g_+(t)$, and the Hilbert transform $\hat{g}(t)$ is the imaginary part of the pre-envelope. An important feature of the pre-envelope $g_+(t)$ is the behavior of its Fourier transform. Let $G_+(f)$ denote the Fourier transform of $g_+(t)$. Then, using (2.57) and (2.58) we may write

$$G_+(f) = G(f) + \text{sgn}(f)G(f) \quad (2.59)$$

Next, invoking the definition of the signum function given in (2.56), we may rewrite (2.59) in the equivalent form

$$G_+(f) = \begin{cases} 2G(f), & f > 0 \\ G(0), & f = 0 \\ 0, & f < 0 \end{cases} \quad (2.60)$$

where $G(0)$ is the value of $G(f)$ at the origin $f = 0$. Equation (2.60) clearly shows that the pre-envelope of the signal $g(t)$ has no frequency content (i.e., its Fourier transform vanishes) for all negative frequencies, and the question that was posed earlier has indeed been answered. Note, however, in order to do this, we had to introduce the complex-valued version of a real-valued signal as described in (2.58).

From the foregoing analysis it is apparent that for a given signal $g(t)$ we may determine its pre-envelope $g_+(t)$ in one of two equivalent procedures.

1. *Time-domain procedure.* Given the signal $g(t)$, we use (2.58) to compute the pre-envelope $g_+(t)$.
2. *Frequency-domain procedure.* We first determine the Fourier transform $G(f)$ of the signal $g(t)$, then use (2.60) to determine $G_+(f)$, and finally evaluate the inverse Fourier transform of $G_+(f)$ to obtain

$$g_+(t) = 2 \int_0^{\infty} G(f) \exp(j2\pi ft) df \quad (2.61)$$

Depending on the description of the signal, procedure 1 may be easier than procedure 2, or vice versa.

Equation (2.58) defines the pre-envelope $g_+(t)$ for positive frequencies. Symmetrically, we may define the pre-envelope for *negative frequencies* as

$$g_-(t) = g(t) - j\hat{g}(t) \quad (2.62)$$

The two pre-envelopes $g_+(t)$ and $g_-(t)$ are simply the complex conjugate of each other, as shown by

$$g_-(t) = g_+^*(t) \quad (2.63)$$

where the asterisk denotes complex conjugation. The spectrum of the pre-envelope $g_+(t)$ is nonzero only for *positive* frequencies; hence the use of a plus sign as the subscript. On the other hand, the use of a minus sign as the subscript is intended to indicate that the spectrum of the other pre-envelope $g_-(t)$ is nonzero only for *negative* frequencies, as shown by the Fourier transform

$$G_-(f) = \begin{cases} 0, & f > 0 \\ G(0), & f = 0 \\ 2G(f), & f < 0 \end{cases} \quad (2.64)$$

Thus, the pre-envelope $g_+(t)$ and $g_-(t)$ constitute a complementary pair of complex-valued signals. Note also that the sum of $g_+(t)$ and $g_-(t)$ is exactly twice the original signal $g(t)$.

Given a real-valued signal, (2.60) teaches us that the pre-envelope $g_+(t)$ is uniquely defined by the spectral content of the signal for positive frequencies. By the same token, (2.64) teaches us that the other pre-envelope $g_-(t)$ is uniquely defined by the spectral content of the signal for negative frequencies. Since $g_-(t)$ is simply the complex conjugate of $g_+(t)$ as indicated in (2.63), we may now make the following statement:

The spectral content of a Fourier transformable real-valued signal for positive frequencies uniquely defines that signal.

In other words, given the spectral content of such a signal for positive frequencies, we may uniquely define the spectral content of the signal for negative frequencies. Here then is the mathematical justification for basing the bandwidth of a Fourier transformable signal on its spectral content exclusively for positive frequencies, which is exactly what we did in Section 2.4, dealing with bandwidth.

EXAMPLE 6 Pre-envelopes of Low-Pass Signal

Continuing with the low-pass signal $g(t)$ considered in Example 5, Figure 2.17a and b depict the corresponding spectra of the pre-envelope $g_+(t)$ and the second pre-envelope $g_-(t)$, both of which belong to $g(t)$. Whereas the spectrum of $g(t)$ is defined for $-W \leq f \leq W$ as in Figure 2.16a, we clearly see from Figure 2.17 that the spectral content of $g_+(t)$ is confined entirely to $0 \leq f \leq W$, and the spectral content of $g_-(t)$ is confined entirely to $-W \leq f \leq 0$.

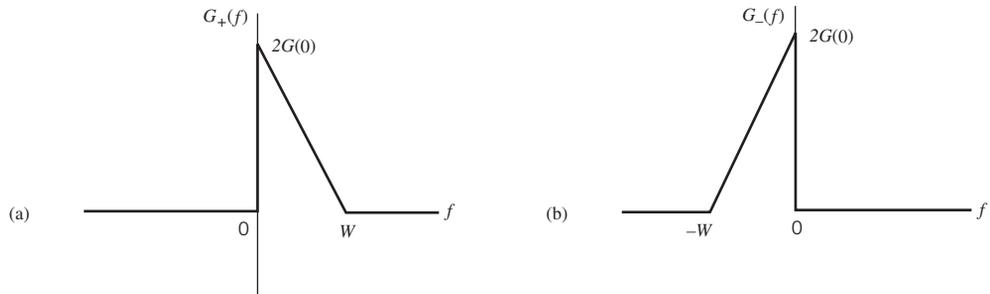


Figure 2.17 Another illustrative application of the Hilbert transform to a low-pass signal: (a) Spectrum of the pre-envelope $g_+(t)$; (b) Spectrum of the other pre-envelope $g_-(t)$.

Practical Importance of the Hilbert Transformation

An astute reader may see an analogy between the use of *phasors* and that of *pre-envelopes*. In particular, just as the use of phasors simplifies the manipulations of alternating currents and voltages in the study of circuit theory, so we find the pre-envelope simplifies the analysis of band-pass signals and band-pass systems in signal theory.

More specifically, by applying the concept of pre-envelope to a band-pass signal, the signal is transformed into an equivalent low-pass representation. In a corresponding way, a band-pass filter is transformed into its own equivalent low-pass representation. Both transformations, rooted in the Hilbert transform, play a key role in the formulation of modulated signals and their demodulation, as demonstrated in what follows in this and subsequent chapters.

2.10 Complex Envelopes of Band-Pass Signals

The idea of pre-envelopes introduced in Section 2.9 applies to any real-valued signal, be it of a low-pass or band-pass kind; the only requirement is that the signal be Fourier transformable. From this point on and for the rest of the chapter, we will restrict attention to band-pass signals. Such signals are exemplified by signals modulated onto a sinusoidal

carrier. In a corresponding way, when it comes to systems we restrict attention to band-pass systems. The primary reason for these restrictions is that the material so presented is directly applicable to analog modulation theory, to be covered in Section 2.14, as well as other digital modulation schemes covered in subsequent chapters of the book. With this objective in mind and the desire to make a consistent use of notation with respect to material to be presented in subsequent chapters, henceforth we will use $s(t)$ to denote a modulated signal. When such a signal is applied to the input of a band-pass system, such as a communication channel, we will use $x(t)$ to denote the resulting system (e.g., channel) output. However, as before, we will use $h(t)$ as the impulse response of the system.

To proceed then, let the band-pass signal of interest be denoted by $s(t)$ and its Fourier transform be denoted by $S(f)$. We assume that the Fourier transform $S(f)$ is essentially confined to a band of frequencies of total extent $2W$, centered about some frequency $\pm f_c$, as illustrated in Figure 2.18a. We refer to f_c as the *carrier frequency*; this terminology is

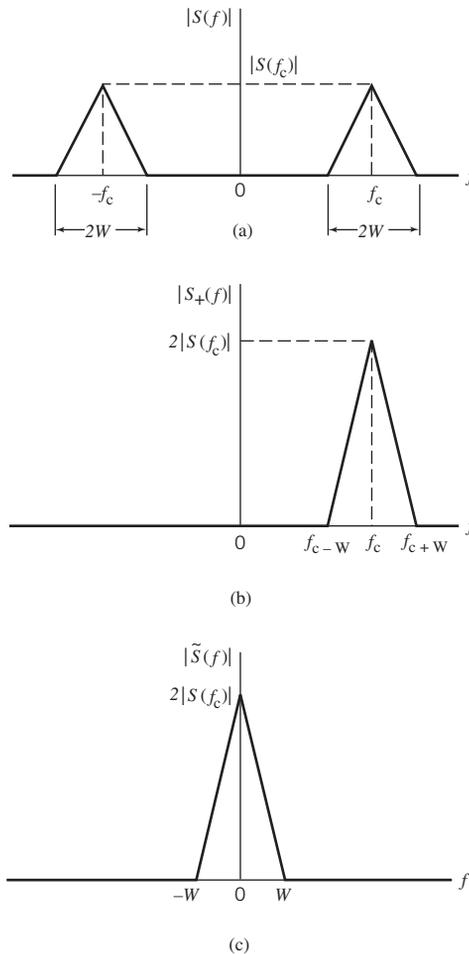


Figure 2.18 (a) Magnitude spectrum of band-pass signal $s(t)$; (b) Magnitude spectrum of pre-envelope $s_+(t)$; (c) Magnitude spectrum of complex envelope $\tilde{s}(t)$.

borrowed from modulation theory. In the majority of communication signals encountered in practice, we find that the bandwidth $2W$ is small compared with f_c , so we may refer to the signal $s(t)$ as a *narrowband signal*. However, a precise statement about how small the bandwidth must be for the signal to be considered narrowband is not necessary for our present discussion. Hereafter, the terms band-pass and narrowband are used interchangeably.

Let the pre-envelope of the narrowband signal $s(t)$ be expressed in the form

$$s_+(t) = \tilde{s}(t) \exp(j2\pi f_c t) \quad (2.65)$$

We refer to $\tilde{s}(t)$ as the *complex envelope* of the band-pass signal $s(t)$. Equation (2.65) may be viewed as the basis of a definition for the complex envelope $\tilde{s}(t)$ in terms of the pre-envelope $s_+(t)$. In light of the narrowband assumption imposed on the spectrum of the band-pass signal $s(t)$, we find that the spectrum of the pre-envelope $s_+(t)$ is limited to the positive frequency band $f_c - W \leq f \leq f_c + W$, as illustrated in Figure 2.18b. Therefore, applying the frequency-shifting property of the Fourier transform to (2.65), we find that the spectrum of the complex envelope $\tilde{s}(t)$ is correspondingly limited to the band $-W \leq f \leq W$ and centered at the origin $f = 0$, as illustrated in Figure 2.18c. In other words, the complex envelope $\tilde{s}(t)$ of the band-pass signal $s(t)$ is a *complex low-pass signal*. The essence of the *mapping* from the band-pass signal $s(t)$ to the complex low-pass signal $\tilde{s}(t)$ is summarized in the following threefold statement:

- The information content of a modulated signal $s(t)$ is fully preserved in the complex envelope $\tilde{s}(t)$.
- Analysis of the band-pass signal $s(t)$ is complicated by the presence of the carrier frequency f_c ; in contrast, the complex envelope $\tilde{s}(t)$ dispenses with f_c , making its analysis simpler to deal with.
- The use of $\tilde{s}(t)$ requires having to handle complex notations.

2.11 Canonical Representation of Band-Pass Signals

By definition, the real part of the pre-envelope $s_+(t)$ is equal to the original band-pass signal $s(t)$. We may therefore express the band-pass signal $s(t)$ in terms of its corresponding complex envelope $\tilde{s}(t)$ as

$$s(t) = \operatorname{Re}[\tilde{s}(t) \exp(j2\pi f_c t)] \quad (2.66)$$

where the operator $\operatorname{Re}[\cdot]$ denotes the real part of the quantity enclosed inside the square brackets. Since, in general, $\tilde{s}(t)$ is a complex-valued quantity, we emphasize this property by expressing it in the Cartesian form

$$\tilde{s}(t) = s_I(t) + js_Q(t) \quad (2.67)$$

where $s_I(t)$ and $s_Q(t)$ are both real-valued low-pass functions; their low-pass property is inherited from the complex envelope $\tilde{s}(t)$. We may therefore use (2.67) in (2.66) to express the original band-pass signal $s(t)$ in the *canonical* or *standard* form

$$s(t) = s_I(t) \cos(2\pi f_c t) - s_Q(t) \sin(2\pi f_c t) \quad (2.68)$$

We refer to $s_I(t)$ as the *in-phase component* of the band-pass signal $s(t)$ and refer to $s_Q(t)$ as the *quadrature-phase component* or simply the *quadrature component* of the signal $s(t)$.

This nomenclature follows from the following observation: if $\cos(2\pi f_c t)$, the multiplying factor of $s_I(t)$, is viewed as the reference sinusoidal carrier, then $\sin(2\pi f_c t)$, the multiplying factor of $s_Q(t)$, is in phase quadrature with respect to $\cos(2\pi f_c t)$.

According to (2.66), the complex envelope $\tilde{s}(t)$ may be pictured as a *time-varying phasor* positioned at the origin of the (s_I, s_Q) -plane, as indicated in Figure 2.19a. With time t varying continuously, the end of the phasor moves about in the plane. Figure 2.19b depicts the phasor representation of the complex exponential $\exp(j2\pi f_c t)$. In the definition given in (2.66), the complex envelope $\tilde{s}(t)$ is multiplied by the complex exponential $\exp(j2\pi f_c t)$. The angles of these two phasors, therefore, add and their lengths multiply, as shown in Figure 2.19c. Moreover, in this latter figure, we show the (s_I, s_Q) -plane rotating with an angular velocity equal to $2\pi f_c$ radians per second. Thus, in the picture portrayed in the figure, the phasor representing the complex envelope $\tilde{s}(t)$ moves in the (s_I, s_Q) -plane, while at the very same time the plane itself rotates about the origin. The original band-pass signal $s(t)$ is the projection of this time-varying phasor on a *fixed line* representing the real axis, as indicated in Figure 2.19c.

Since both $s_I(t)$ and $s_Q(t)$ are low-pass signals limited to the band $-W \leq f \leq W$, they may be extracted from the band-pass signal $s(t)$ using the scheme shown in Figure 2.20a. Both low-pass filters in this figure are designed identically, each with a bandwidth equal to W .

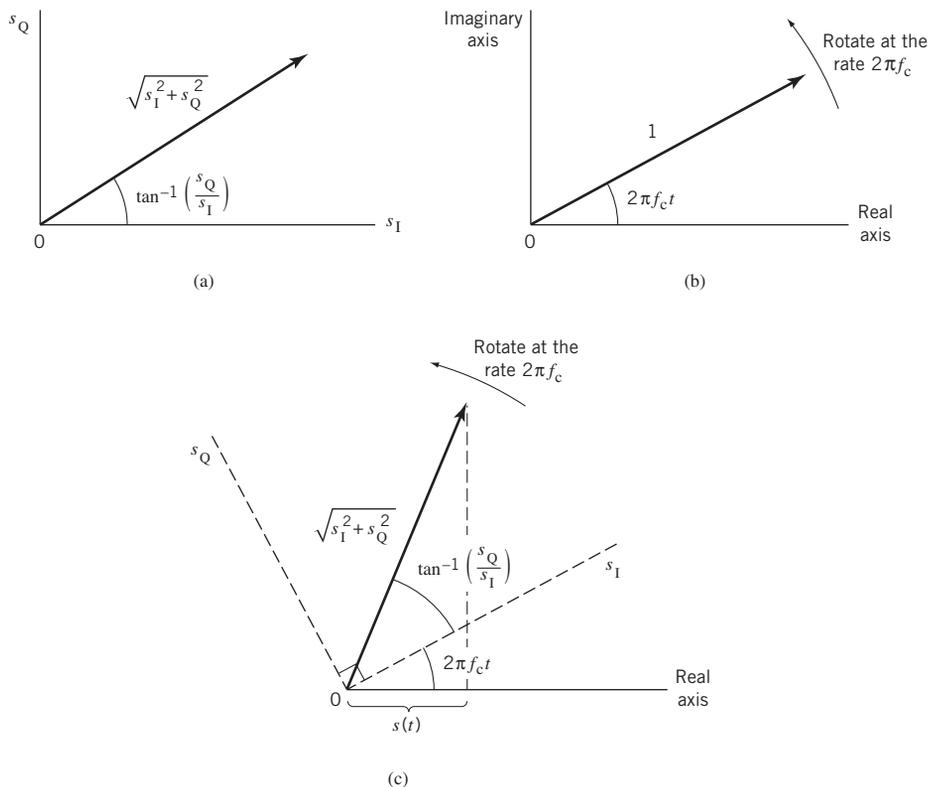


Figure 2.19 Illustrating an interpretation of the complex envelope $\tilde{s}(t)$ and its multiplication by $\exp(j2\pi f_c t)$.

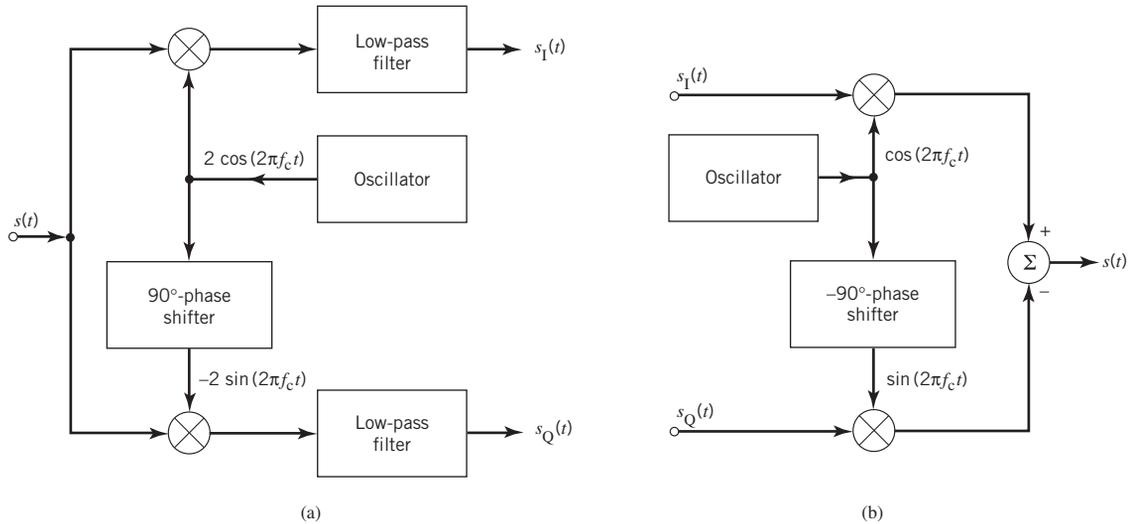


Figure 2.20 (a) Scheme for deriving the in-phase and quadrature components of a band-pass signal $g(t)$. (b) Scheme for reconstructing the band-pass signal from its in-phase and quadrature components.

To reconstruct $s(t)$ from its in-phase and quadrature components, we may use the scheme shown in Figure 2.20b. In light of these statements, we may refer to the scheme in Figure 2.20a as an *analyzer*, in the sense that it extracts the in-phase and quadrature components, $s_I(t)$ and $s_Q(t)$, from the band-pass signal $s(t)$. By the same token, we may refer to the second scheme in Figure 2.20b as a *synthesizer*, in the sense it reconstructs the band-pass signal $s(t)$ from its in-phase and quadrature components, $s_I(t)$ and $s_Q(t)$.

The two schemes shown in Figure 2.20 are basic to the study of *linear modulation schemes, be they of an analog or digital kind*. Multiplication of the low-pass in-phase component $s_I(t)$ by $\cos(2\pi f_c t)$ and multiplication of the quadrature component $s_Q(t)$ by $\sin(2\pi f_c t)$ represent linear forms of modulation. Provided that the carrier frequency f_c is larger than the low-pass bandwidth W , the resulting band-pass function $s(t)$ defined in (2.68) is referred to as a *passband signal waveform*. Correspondingly, the mapping from $s_I(t)$ and $s_Q(t)$ combined into $s(t)$ is known as *passband modulation*.

Polar Representation of Band-Pass Signals

Equation (2.67) is the Cartesian form of defining the complex envelope $\tilde{s}(t)$ of the band-pass signal $s(t)$. Alternatively, we may define $\tilde{s}(t)$ in the *polar form* as

$$\tilde{s}(t) = a(t) \exp[j\phi(t)] \quad (2.69)$$

where $a(t)$ and $\phi(t)$ are both real-valued low-pass functions. Based on the polar representation of (2.69), the original band-pass signal $s(t)$ is itself defined by

$$\tilde{s}(t) = a(t) \cos[2\pi f_c t + \phi(t)] \quad (2.70)$$

We refer to $a(t)$ as the *natural envelope* or simply the *envelope* of the band-pass signal $s(t)$ and refer to $\phi(t)$ as the *phase* of the signal. We now see why the term “pre-envelope” was used in referring to (2.58), the formulation of which *preceded* that of (2.70).

Relationship Between Cartesian and Polar Representations of Band-Pass Signal

The envelope $a(t)$ and phase $\phi(t)$ of a band-pass signal $s(t)$ are respectively related to the in-phase and quadrature components $s_I(t)$ and $s_Q(t)$ as follows (see the time-varying phasor representation of Figure 2.19a):

$$a(t) = \sqrt{s_I^2(t) + s_Q^2(t)} \quad (2.71)$$

and

$$\phi(t) = \tan^{-1}\left(\frac{s_Q(t)}{s_I(t)}\right) \quad (2.72)$$

Conversely, we may write

$$s_I(t) = a(t)\cos[\phi(t)] \quad (2.73)$$

and

$$s_Q(t) = a(t)\sin[\phi(t)] \quad (2.74)$$

Thus, both the in-phase and quadrature components of a band-pass signal contain amplitude and phase information, both of which are uniquely defined for a prescribed phase $\phi(t)$, modulo 2π .

2.12 Complex Low-Pass Representations of Band-Pass Systems

Now that we know how to handle the complex low-pass representation of band-pass signals, it is logical that we develop a corresponding procedure for handling the representation of linear time-invariant band-pass systems. Specifically, we wish to show that the analysis of band-pass systems is greatly simplified by establishing an *analogy*, more precisely an *isomorphism*, between band-pass and low-pass systems. For example, this analogy would help us to facilitate the computer simulation of a wireless communication channel driven by a sinusoidally modulated signal, which otherwise could be a difficult proposition.

Consider a narrowband signal $s(t)$, with its Fourier transform denoted by $S(f)$. We assume that the spectrum of the signal $s(t)$ is limited to frequencies within $\pm W$ hertz of the carrier frequency f_c . We also assume that $W < f_c$. Let the signal $s(t)$ be applied to a linear time-invariant band-pass system with impulse response $h(t)$ and frequency response $H(f)$. We assume that the frequency response of the system is limited to frequencies within $\pm B$ of the carrier frequency f_c . The *system bandwidth* $2B$ is usually narrower than or equal to the input *signal bandwidth* $2W$. We wish to represent the band-pass impulse response $h(t)$ in terms of two quadrature components, denoted by $h_I(t)$ and $h_Q(t)$. In particular, by analogy to the representation of band-pass signals, we express $h(t)$ in the form

$$h(t) = h_I(t)\cos(2\pi f_c t) - h_Q(t)\sin(2\pi f_c t) \quad (2.75)$$

Correspondingly, we define the *complex impulse response* of the band-pass system as

$$\tilde{h}(t) = h_I(t) + jh_Q(t) \quad (2.76)$$

Hence, following (2.66), we may express $h(t)$ in terms of $\tilde{h}(t)$ as

$$h(t) = \operatorname{Re}[\tilde{h}(t)\exp(j2\pi f_c t)] \quad (2.77)$$

Note that $h_I(t)$, $h_Q(t)$, and $\tilde{h}(t)$ are all low-pass functions, limited to the frequency band $-B \leq f \leq B$.

We may determine the complex impulse response $\tilde{h}(t)$ in terms of the in-phase and quadrature components $h_I(t)$ and $h_Q(t)$ of the band-pass impulse response $h(t)$ by building on (2.76). Alternatively, we may determine it from the band-pass frequency response $H(f)$ in the following way. We first use (2.77) to write

$$2h(t) = \tilde{h}(t)\exp(j2\pi f_c t) + \tilde{h}^*(t)\exp(-j2\pi f_c t) \quad (2.78)$$

where $\tilde{h}^*(t)$ is the *complex conjugate* of $\tilde{h}(t)$; the rationale for introducing the factor of 2 on the left-hand side of (2.78) follows from the fact that if we add a complex signal and its complex conjugate, the sum adds up to twice the real part and the imaginary parts cancel. Applying the Fourier transform to both sides of (2.78) and using the complex-conjugation property of the Fourier transform, we get

$$2H(f) = \tilde{H}(f-f_c) + \tilde{H}^*(-f-f_c) \quad (2.79)$$

where $H(f) \Leftrightarrow h(t)$ and $\tilde{H}(f) \Leftrightarrow \tilde{h}(t)$. Equation (2.79) satisfies the requirement that $H^*(f) = H(-f)$ for a real-valued impulse response $h(t)$. Since $\tilde{H}(f)$ represents a low-pass frequency response limited to $|f| \leq B$ with $B < f_c$, we infer from (2.79) that

$$\tilde{H}(f-f_c) = 2H(f), \quad f > 0 \quad (2.80)$$

Equation (2.80) states:

For a specified band-pass frequency response $H(f)$, we may determine the corresponding complex low-pass frequency response $\tilde{H}(f)$ by taking the part of $H(f)$ defined for positive frequencies, shifting it to the origin, and scaling it by the factor 2.

Having determined the complex frequency response $\tilde{H}(f)$, we decompose it into its in-phase and quadrature components, as shown by

$$\tilde{H}(f) = \tilde{H}_I(f) + j\tilde{H}_Q(f) \quad (2.81)$$

where the *in-phase component* is defined by

$$\tilde{H}_I(f) = \frac{1}{2}[\tilde{H}(f) + \tilde{H}^*(-f)] \quad (2.82)$$

and the *quadrature component* is defined by

$$\tilde{H}_Q(f) = \frac{1}{2j}[\tilde{H}(f) - j\tilde{H}^*(-f)] \quad (2.83)$$

Finally, to determine the complex impulse response $\tilde{h}(t)$ of the band-pass system, we take the inverse Fourier transform of $\tilde{H}(f)$, obtaining

$$\tilde{h}(t) = \int_{-\infty}^{\infty} \tilde{H}(f)\exp(j2\pi ft) df \quad (2.84)$$

which is the formula we have been seeking.

2.13 Putting the Complex Representations of Band-Pass Signals and Systems All Together

Examining (2.66) and (2.77), we immediately see that these two equations share a common multiplying factor: the exponential $\exp(j2\pi f_c t)$. In practical terms, the inclusion of this factor accounts for a sinusoidal carrier of frequency f_c , which facilitates transmission of the modulated (band-pass) signal $s(t)$ across a band-pass channel of midband frequency f_c . In analytic terms, however, the presence of this exponential factor in both (2.66) and (2.77) complicates the analysis of the band-pass system driven by the modulated signal $s(t)$. This analysis can be simplified through the combined use of complex low-pass equivalent representations of both the modulated signal $s(t)$ and the band-pass system characterized by the impulse response $h(t)$. The simplification can be carried out in the time domain or frequency domain, as discussed next.

The Time-Domain Procedure

Equipped with the complex representations of band-pass signals and systems, we are ready to derive an analytically efficient method for determining the output of a band-pass system driven by a corresponding band-pass signal. To proceed with the derivation, assume that $S(f)$, denoting the spectrum of the input signal $s(t)$, and $H(f)$, denoting the frequency response of the system, are both centered around the same frequency f_c . In practice, there is no need to consider a situation in which the carrier frequency of the input signal is not aligned with the midband frequency of the band-pass system, since we have considerable freedom in choosing the carrier or midband frequency. Thus, changing the carrier frequency of the input signal by an amount Δf_c , for example, simply corresponds to absorbing (or removing) the factor $\exp(\pm j2\pi \Delta f_c t)$ in the complex envelope of the input signal or the complex impulse response of the band-pass system. We are therefore justified in proceeding on the assumption that $S(f)$ and $H(f)$ are both centered around the same carrier frequency f_c .

Let $x(t)$ denote the output signal of the band-pass system produced in response to the incoming band-pass signal $s(t)$. Clearly, $x(t)$ is also a band-pass signal, so we may represent it in terms of its own low-pass complex envelope $\tilde{x}(t)$ as

$$x(t) = \operatorname{Re}[\tilde{x}(t)\exp(j2\pi f_c t)] \quad (2.85)$$

The output signal $x(t)$ is related to the input signal $s(t)$ and impulse response $h(t)$ of the system in the usual way by the convolution integral

$$x(t) = \int_{-\infty}^{\infty} h(\tau)s(t-\tau) d\tau \quad (2.86)$$

In terms of pre-envelopes, we have $h(t) = \operatorname{Re}[h_+(t)]$ and $s(t) = \operatorname{Re}[s_+(t)]$. We may therefore rewrite (2.86) in terms of the pre-envelopes $s_+(t)$ and $h_+(t)$ as

$$x(t) = \int_{-\infty}^{\infty} \operatorname{Re}[h_+(\tau)]\operatorname{Re}[s_+(t-\tau)] d\tau \quad (2.87)$$

To proceed further, we make use of a basic property of pre-envelopes that is described by the following relation:

$$\int_{-\infty}^{\infty} \operatorname{Re}[h_+(\tau)] \operatorname{Re}[s_+(\tau)] d\tau = \frac{1}{2} \operatorname{Re} \left[\int_{-\infty}^{\infty} h_+(\tau) s_+^*(\tau) d\tau \right] \quad (2.88)$$

where we have used τ as the integration variable to be consistent with that in (2.87); details of (2.88) are presented in Problem 2.20. Next, from Fourier-transform theory we note that using $s(-\tau)$ in place of $s(\tau)$ has the effect of removing the complex conjugation on the right-hand side of (2.88). Hence, bearing in mind the algebraic difference between the argument of $s_+(\tau)$ in (2.88) and that of $s_+(t-\tau)$ in (2.87), and using the relationship between the pre-envelope and complex envelope of a band-pass signal, we may express (2.87) in the equivalent form

$$\begin{aligned} x(t) &= \frac{1}{2} \operatorname{Re} \left[\int_{-\infty}^{\infty} h_+(\tau) s_+(t-\tau) d\tau \right] \\ &= \frac{1}{2} \operatorname{Re} \left\{ \int_{-\infty}^{\infty} \tilde{h}(\tau) \exp(j2\pi f_c \tau) \tilde{s}(t-\tau) \exp[j2\pi f_c(t-\tau)] d\tau \right\} \\ &= \frac{1}{2} \operatorname{Re} \left[\exp(j2\pi f_c t) \int_{-\infty}^{\infty} \tilde{h}(\tau) \tilde{s}(t-\tau) d\tau \right] \end{aligned} \quad (2.89)$$

Thus, comparing the right-hand sides of (2.85) and (2.89), we readily find that for a large enough carrier frequency f_c , the complex envelope $\tilde{x}(t)$ of the output signal is simply defined in terms of the complex envelope $\tilde{s}(t)$ of the input signal and the complex impulse response $\tilde{h}(t)$ of the band-pass system as follows:

$$\tilde{x}(t) = \frac{1}{2} \int_{-\infty}^{\infty} \tilde{h}(t) \tilde{s}(t-\tau) d\tau \quad (2.90)$$

This important relationship is the result of the *isomorphism* between a band-pass function and the corresponding complex low-pass function, in light of which we may now make the following summarizing statement:

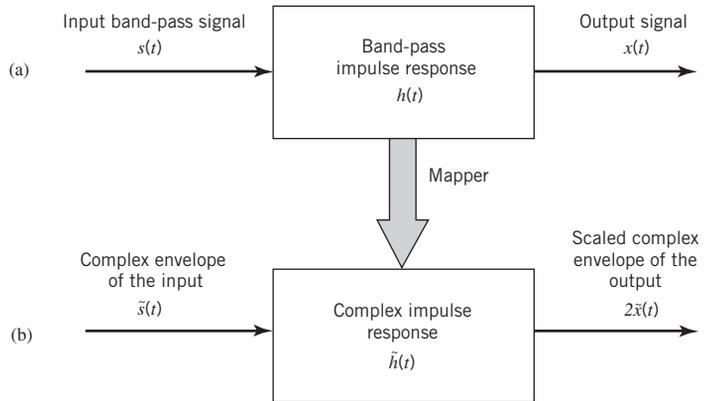
Except for the scaling factor 1/2, the complex envelope $\tilde{x}(t)$ of the output signal of a band-pass system is obtained by convolving the complex impulse response $\tilde{h}(t)$ of the system with the complex envelope $\tilde{s}(t)$ of the input band-pass signal.

In computational terms, the significance of this statement is profound. Specifically, in dealing with band-pass signals and systems, we need only concern ourselves with the functions $\tilde{s}(t)$, $\tilde{x}(t)$, and $\tilde{h}(t)$, representing the complex low-pass equivalents of the excitation applied to the input of the system, the response produced at the output of the system, and the impulse response of the system respectively, as illustrated in Figure 2.21. The essence of the filtering process performed in the original system of Figure 2.21a is completely retained in the complex low-pass equivalent representation depicted in Figure 2.21b.

The complex envelope $\tilde{s}(t)$ of the input band-pass signal and the complex impulse response $\tilde{h}(t)$ of the band-pass system are defined in terms of their respective in-phase

Figure 2.21

(a) Input–output description of a band-pass system; (b) Complex low-pass equivalent model of the band-pass system.



and quadrature components by (2.67) and (2.76), respectively. Substituting these relations into (2.90), we get

$$\begin{aligned} 2\tilde{x}(t) &= \tilde{h}(t) \star \tilde{s}(t) \\ &= [h_I(t) + jh_Q(t)] \star [s_I(t) + js_Q(t)] \end{aligned} \quad (2.91)$$

where the symbol \star denotes convolution. Because convolution is *distributive*, we may rewrite (2.91) in the equivalent form

$$2\tilde{x}(t) = [h_I(t) \star s_I(t) - h_Q(t) \star s_Q(t)] + j[h_Q(t) \star s_I(t) + h_I(t) \star s_Q(t)] \quad (2.92)$$

Let the complex envelope $\tilde{x}(t)$ of the response be defined in terms of its in-phase and quadrature components as

$$\tilde{x}(t) = x_I(t) + jx_Q(t) \quad (2.93)$$

Then, comparing the real and imaginary parts in (2.92) and (2.93), we find that the in-phase component $x_I(t)$ is defined by the relation

$$2x_I(t) = h_I(t) \star s_I(t) - h_Q(t) \star s_Q(t) \quad (2.94)$$

and its quadrature component $x_Q(t)$ is defined by the relation

$$2x_Q(t) = h_Q(t) \star s_I(t) + h_I(t) \star s_Q(t) \quad (2.95)$$

Thus, for the purpose of evaluating the in-phase and quadrature components of the complex envelope $\tilde{x}(t)$ of the system output, we may use the *low-pass equivalent model* shown in Figure 2.22. All the signals and impulse responses shown in this model are real-valued low-pass functions; hence a time-domain procedure for simplifying the analysis of band-pass systems driven by band-pass signals.

The Frequency-Domain Procedure

Alternatively, Fourier-transforming the convolution integral of (2.90) and recognizing that convolution in the time domain is changed into multiplication in the frequency domain, we get

$$\tilde{X}(f) = \frac{1}{2} \tilde{H}(f) \tilde{S}(f) \quad (2.96)$$

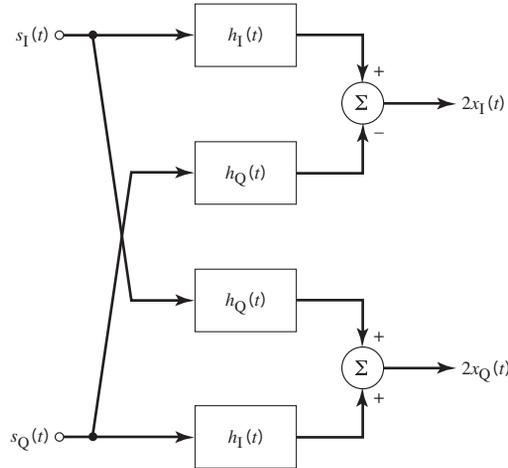


Figure 2.22 Block diagram illustrating the relationship between the in-phase and quadrature components of the response of a band-pass filter and those of the input signal.

where $\tilde{s}(t) \Leftrightarrow \tilde{S}(f)$, $\tilde{h}(t) \Leftrightarrow \tilde{H}(f)$, and $\tilde{x}(t) \Leftrightarrow \tilde{X}(f)$. The $\tilde{H}(f)$ is itself related to the frequency response $H(f)$ of the band-pass system by (2.80). Thus, assuming that $H(f)$ is known, we may use the frequency-domain procedure summarized in Table 2.4 for computing the system output $x(t)$ in response to the system input $s(t)$.

In actual fact, the procedure of Table 2.4 is the frequency-domain representation of the low-pass equivalent to the band-pass system, depicted in Figure 2.21b. In computational terms, this procedure is of profound practical significance. We say so because its use alleviates the analytic and computational difficulty encountered in having to include the carrier frequency f_c in the pertinent calculations.

As discussed earlier in the chapter, the theoretical formulation of the low-pass equivalent in Figure 2.21b is rooted in the Hilbert transformation, the evaluation of which poses a practical problem of its own, because of the wideband 90° -phase shifter involved in its theory. Fortunately, however, we do not need to invoke the Hilbert transform in constructing the low-pass equivalent. This is indeed so, when a message signal modulated onto a sinusoidal carrier is processed by a band-pass filter, as explained here:

1. Typically, the message signal is band limited for all practical purposes. Moreover, the carrier frequency is larger than the highest frequency component of the signal; the modulated signal is therefore a band-pass signal with a well-defined passband. Hence, the in-phase and quadrature components of the modulated signal $s(t)$, represented respectively by $s_I(t)$ and $s_Q(t)$, are readily obtained from the canonical representation of $s(t)$, described in (2.68).
2. Given the well-defined frequency response $H(f)$ of the band-pass system, we may readily evaluate the corresponding complex low-pass frequency response $\tilde{H}(f)$; see (2.80). Hence, we may compute the system output $x(t)$ produced in response to the carrier-modulated input $s(t)$ without invoking the Hilbert transform.

Table 2.4 Procedure for the computational analysis of a band-pass system driven by a band-pass signal

Given the frequency response $H(f)$ of a band-pass system, computation of the output signal $x(t)$ of the system in response to an input band-pass signal $s(t)$ is summarized as follows:

1. Use (2.80), namely $\tilde{H}(f - f_c) = 2H(f)$, for $f > 0$ to determine $\tilde{H}(f)$.
2. Expressing the input band-pass signal $s(t)$ in the canonical form of (2.68), evaluate the complex envelope $\tilde{s}(t) = s_I(t) + js_Q(t)$ where $s_I(t)$ is the in-phase component of $s(t)$ and $s_Q(t)$ is its quadrature component. Hence, compute the Fourier transform $\tilde{S}(f) = \mathbf{F}[\tilde{s}(t)]$
3. Using (2.96), compute $\tilde{X}(f) = \frac{1}{2}\tilde{H}(f)\tilde{S}(f)$, which defines the Fourier transform of the complex envelope $\tilde{x}(t)$ of the output signal $x(t)$.
4. Compute the inverse Fourier transform of $\tilde{X}(f)$, yielding $\tilde{x}(t) = \mathbf{F}^{-1}[\tilde{X}(f)]$
5. Use (2.85) to compute the desired output signal $x(t) = \text{Re}[\tilde{x}(t)\exp(j2\pi f_c t)]$

Procedure for Efficient Simulation of Communication Systems

To summarize, the frequency-domain procedure described in Table 2.4 is well suited for the efficient simulation of communication systems on a computer for two reasons:

1. The low-pass equivalents of the incoming band-pass signal and the band-pass system work by eliminating the exponential factor $\exp(j2\pi f_c t)$ from the computation without loss of information.
2. The *fast Fourier transform (FFT) algorithm*, discussed later in the chapter, is used for numerical computation of the Fourier transform. This algorithm is used twice in Table 2.4, once in step 2 to perform Fourier transformation, and then again in step 4 to perform inverse Fourier transformation.

The procedure of this table, rooted largely in the frequency domain, assumes availability of the band-pass system's frequency response $H(f)$. If, however, it is the system's impulse response $h(t)$ that is known, then all we need is an additional step to Fourier transform $h(t)$ into $H(f)$ before initiating the procedure of Table 2.4.

2.14 Linear Modulation Theory

The material presented in Sections 2.8–2.13 on the complex low-pass representation of band-pass signals and systems is of profound importance in the study of communication theory. In particular, we may use the canonical formula of (2.68) as the mathematical basis for a unified treatment of linear modulation theory, which is the subject matter of this section.

We start this treatment with a formal definition:

Modulation is a process by means of which one or more parameters of a sinusoidal carrier are varied in accordance with a message signal so as to facilitate transmission of that signal over a communication channel.

The message signal (e.g., voice, video, data sequence) is referred to as the *modulating signal*, and the result of the modulation process is referred to as the *modulated signal*. Naturally, in a communication system, modulation is performed in the transmitter. The reverse of modulation, aimed at recovery of the original message signal in the receiver, is called *demodulation*.

Consider the block diagram of Figure 2.23, depicting a modulator, where $m(t)$ is the message signal, $\cos(2\pi f_c t)$ is the carrier, and $s(t)$ is the modulated signal. To apply (2.68) to this modulator, the in-phase component $s_I(t)$ in that equation is treated simply as a scaled version of the message signal denoted by $m(t)$. As for the quadrature component $s_Q(t)$, it is defined by a *spectrally shaped* version of $m(t)$ that is performed linearly. In such a scenario, it follows that a modulated signal $s(t)$ defined by (2.68) is a *linear function* of the message signal $m(t)$; hence the reference to this equation as the mathematical basis of *linear modulation theory*.

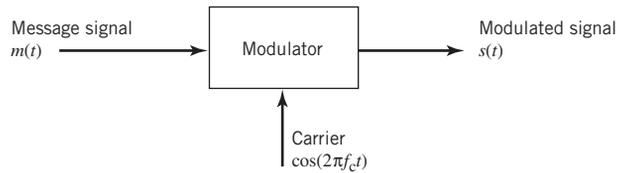


Figure 2.23 Block diagram of a modulator.

To recover the original message signal $m(t)$ from the modulated signal $s(t)$, we may use a demodulator, the block diagram of which is depicted in Figure 2.24. An elegant feature of linear modulation theory is that demodulation of $s(t)$ is also achieved using linear operations. However, for linear demodulation of $s(t)$ to be feasible, the locally generated carrier in the demodulator of Figure 2.24 has to be synchronous with the original sinusoidal carrier used in the modulator of Figure 2.23. Accordingly, we speak of *synchronous demodulation* or *coherent detection*.

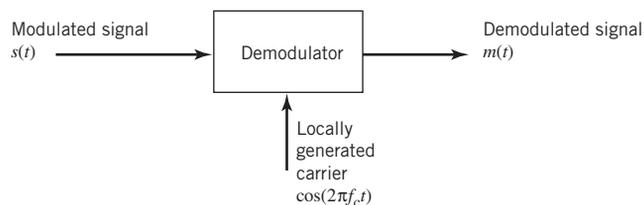


Figure 2.24 Block diagram of a demodulator.

Depending on the spectral composition of the modulated signal, we have three kinds of linear modulation in analog communications:

- double sideband-suppressed carrier (DSB-SC) modulation;
- vestigial sideband (VSB) modulation;
- single sideband (SSB) modulation.

These three methods of modulation are discussed in what follows and in this order.

DSB-SC Modulation

DSB-SC modulation is the simplest form of linear modulation, which is obtained by setting

$$s_I(t) = m(t)$$

and

$$s_Q(t) = 0$$

Accordingly, (2.68) is reduced to

$$s(t) = m(t) \cos(2\pi f_c t) \quad (2.97)$$

the implementation of which simply requires a *product modulator* that multiplies the message signal $m(t)$ by the carrier $\cos(2\pi f_c t)$, assumed to be of unit amplitude.

For a frequency-domain description of the DSB-SC-modulated signal defined in (2.97), suppose that the message signal $m(t)$ occupies the frequency band $-W \leq f \leq W$, as depicted in Figure 2.25a; hereafter, W is referred to as the *message bandwidth*. Then, provided that the carrier frequency satisfies the condition $f_c > W$, we find that the spectrum of the DSB-SC-modulated signal consists of an *upper sideband* and *lower sideband*, as depicted in Figure 2.25b. Comparing the two parts of this figure, we immediately see that the *channel bandwidth*, B , required to support the transmission of the DSB-SC-modulated signal from the transmitter to the receiver is twice the message bandwidth.

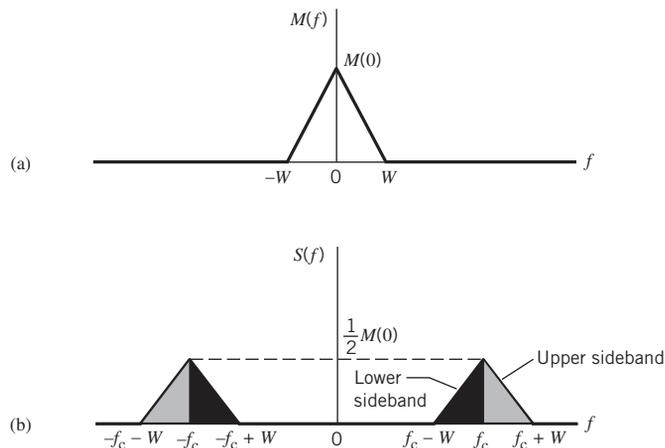


Figure 2.25 (a) Message spectrum. (b) Spectrum of DSB-SC modulated wave $s(t)$, assuming $f_c > W$.

One other interesting point apparent from Figure 2.25b is that the spectrum of the DSB-SC modulated signal is entirely void of delta functions. This statement is further testimony to the fact that the carrier is suppressed from the generation of the modulated signal $s(t)$ of (2.97).

Summarizing the useful features of DSB-SC modulation:

- *suppression of the carrier*, which results in saving of transmitted power;
- *desirable spectral characteristics*, which make it applicable to the modulation of band-limited message signals;
- ease of synchronizing the receiver to the transmitter for coherent detection.

On the downside, DSB-SC modulation is wasteful of channel bandwidth. We say so for the following reason. The two sidebands, constituting the spectral composition of the modulated signal $s(t)$, are actually the *image* of each other with respect to the carrier frequency f_c ; hence, the transmission of either sideband is sufficient for transporting $s(t)$ across the channel.

VSB Modulation

In VSB modulation, one sideband is partially suppressed and a *vestige* of the other sideband is configured in such a way to compensate for the partial sideband suppression by exploiting the fact that the two sidebands in DSB-SC modulation are the image of each other. A popular method of achieving this design objective is to use the *frequency discrimination method*. Specifically, a DSB-SC-modulated signal is first generated using a product modulator, followed by a band-pass filter, as shown in Figure 2.26. The desired spectral shaping is thereby realized through the appropriate design of the band-pass filter.

Suppose that a vestige of the lower sideband is to be transmitted. Then, the frequency response of the band-pass filter, $H(f)$, takes the form shown in Figure 2.27; to simplify matters, only the frequency response for positive frequencies is shown in the figure. Examination of this figure reveals two characteristics of the band-pass filter:

1. *Normalization* of the frequency response, which means that

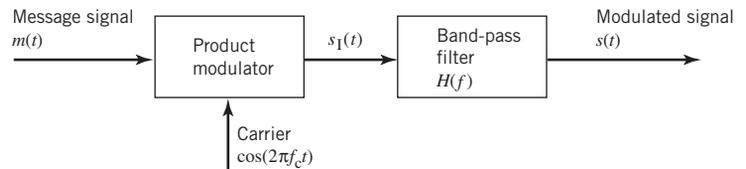
$$H(f) = \begin{cases} 1 & \text{for } f_c + f_v \leq |f| < f_c + W \\ \frac{1}{2} & \text{for } |f| = f_c \end{cases} \quad (2.98)$$

where f_v is the *vestigial bandwidth* and the other parameters are as previously defined.

2. *Odd symmetry of the cutoff portion inside the transition interval* $f_c - f_v \leq |f| \leq f_c + f_v$, which means that values of the frequency response $H(f)$ at any two frequencies equally spaced above and below the carrier frequency add up to unity.

Figure 2.26

Frequency-discrimination method for producing VSB modulation where the intermediate signal $s_1(t)$ is DSB-SC modulated.



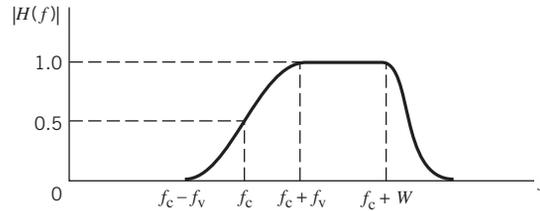


Figure 2.27 Magnitude response of VSB filter; only the positive-frequency portion is shown

Consequently, we find that shifted versions of the frequency response $H(f)$ satisfy the condition

$$H(f - f_c) + H(f + f_c) = 1 \quad \text{for } -W \leq |f| \leq W \quad (2.99)$$

Outside the frequency band of interest defined by $|f| \geq f_c + W$, the frequency response $H(f)$ can assume arbitrary values. We may thus express the channel bandwidth required for the transmission of VSB-modulated signals as

$$B = W + f_v \quad (2.100)$$

With this background, we now address the issue of how to specify $H(f)$. We first use the canonical formula of (2.68) to express the VSB-modulated signal $s_1(t)$, containing a vestige of the lower sideband, as

$$s_1(t) = \frac{1}{2}m(t)\cos(2\pi f_c t) - \frac{1}{2}m_Q(t)\sin(2\pi f_c t) \quad (2.101)$$

where $m(t)$ is the message signal, as before, and $m_Q(t)$ is the spectrally shaped version of $m(t)$; the reason for the factor $1/2$ will become apparent later. Note that if $m_Q(t)$ is set equal to zero, (2.101) reduces to DSB-SC modulation. It is therefore in the *quadrature signal* $m_Q(t)$ that VSB modulation distinguishes itself from DSB-SC modulation. In particular, the role of $m_Q(t)$ is to interfere with the message signal $m(t)$ in such a way that power in one of the sidebands of the VSB-modulated signal $s(t)$ (e.g., the lower sideband in Figure 2.27) is appropriately reduced.

To determine $m_Q(t)$, we examine two different procedures:

1. *Phase-discrimination*, which is rooted in the time-domain description of (2.101); transforming this equation into the frequency domain, we obtain

$$S_1(f) = \frac{1}{4}[M(f - f_c) + M(f + f_c)] - \frac{1}{4j}[M_Q(f - f_c) - M_Q(f + f_c)] \quad (2.102)$$

where

$$M(f) = \mathbf{F}[m(t)] \quad \text{and} \quad M_Q(f) = \mathbf{F}[m_Q(t)]$$

2. *Frequency-discrimination*, which is structured in the manner described in Figure 2.26; passing the DSB-SC-modulated signal (i.e., the intermediate signal $s_1(t)$ in Figure 2.26) through the band-pass filter, we write

$$S_1(f) = \frac{1}{2}[M(f - f_c) + M(f + f_c)]H(f) \quad (2.103)$$

In both (2.102) and (2.103), the spectrum $S_1(f)$ is defined in the frequency interval

$$f_c - W \leq |f| \leq f_c + W$$

Equating the right-hand sides of these two equations, we get (after canceling common terms)

$$\begin{aligned} \frac{1}{2}[M(f-f_c) + M(f+f_c)] - \frac{1}{2j}[M_Q(f-f_c) - M_Q(f+f_c)] \\ = [M(f-f_c) + M(f+f_c)]H(f) \end{aligned} \quad (2.104)$$

Shifting both sides of (2.104) to the left by the amount f_c , we get (after canceling common terms)

$$\frac{1}{2}M(f) - \frac{1}{2j}M_Q(f) = M(f)H(f+f_c), \quad -W \leq |f| \leq W \quad (2.105)$$

where the terms $M(f+2f_c)$ and $M_Q(f+2f_c)$ are ignored as they both lie outside the interval $-W \leq |f| \leq W$. Next, shifting both sides of (2.104) by the amount f_c , but this time to the *right*, we get (after canceling common terms)

$$\frac{1}{2}M(f) + \frac{1}{2j}M_Q(f) = M(f)H(f-f_c), \quad -W \leq |f| \leq W \quad (2.106)$$

where, this time, the terms $M(f-2f_c)$ and $M_Q(f-2f_c)$ are ignored as they both lie outside the interval $-W \leq |f| \leq W$.

Given (2.105) and (2.106), all that remains to be done now is to follow two simple steps:

1. Adding these two equations and then factoring out the common term $M(f)$, we get the condition of (2.99) previously imposed on $H(f)$; indeed, it is with this condition in mind that we introduced the scaling factor 1/2 in (2.101).
2. Subtracting (2.105) from (2.106) and rearranging terms, we get the desired relationship between $M_Q(f)$ and $M(f)$:

$$M_Q(f) = j[H(f-f_c) - H(f+f_c)]M(f), \quad -W \leq |f| \leq W \quad (2.107)$$

Let $H_Q(f)$ denote the frequency response of a *quadrature filter* that operates on the message spectrum $M(f)$ to produce $M_Q(f)$. In light of (2.107), we may readily define $H_Q(f)$ in terms of $H(f)$ as

$$\begin{aligned} H_Q(f) &= \frac{M_Q(f)}{M(f)} \\ &= j[H(f-f_c) - H(f+f_c)], \quad -W \leq |f| \leq W \end{aligned} \quad (2.108)$$

Equation (2.108) provides the frequency-domain basis for the *phase-discrimination method* for generating the VSB-modulated signal $s_1(t)$, where only a vestige of the lower sideband is retained. With this equation at hand, it is instructive to plot the frequency response $H_Q(f)$. For the frequency interval $-W \leq f \leq W$, the term $H(f-f_c)$ is defined by the response $H(f)$ for negative frequencies shifted to the right by f_c , whereas the term $H(f+f_c)$ is defined by the response $H(f)$ for positive frequencies shifted to the left by f_c . Accordingly, building on the positive frequency response plotted in Figure 2.27, we find that the corresponding plot of $H_Q(f)$ is shaped as shown in Figure 2.28.

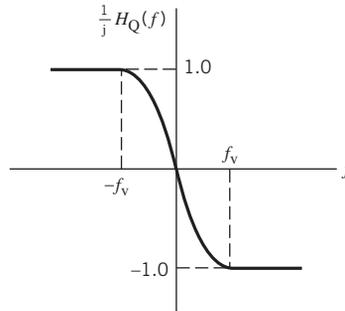


Figure 2.28 Frequency response of the quadrature filter for producing the quadrature component of the VSB wave.

The discussion on VSB modulation has thus far focused on the case where a vestige of the lower sideband is transmitted. For the alternative case when a vestige of the upper sideband is transmitted, we find that the corresponding VSB-modulated wave is described by

$$s_2(t) = \frac{1}{2}m(t)\cos(2\pi f_c t) + \frac{1}{2}m_Q(t)\sin(2\pi f_c t) \quad (2.109)$$

where the quadrature signal $m_Q(t)$ is constructed from the message signal $m(t)$ in exactly the same way as before.

Equations (2.101) and (2.109) are of the same mathematical form, except for an algebraic difference; they may, therefore, be combined into the single formula

$$s(t) = \frac{1}{2}m(t)\cos(2\pi f_c t) \mp \frac{1}{2}m_Q(t)\sin(2\pi f_c t) \quad (2.110)$$

where the minus sign applies to a VSB-modulated signal containing a vestige of the lower sideband and the plus sign applies to the alternative case when the modulated signal contains a vestige of the upper sideband.

The formula of (2.110) for VSB modulation includes DSB-SC modulation as a special case. Specifically, setting $m_Q(t) = 0$, this formula reduces to that of (2.97) for DSB-SC modulation, except for the trivial scaling factor of 1/2.

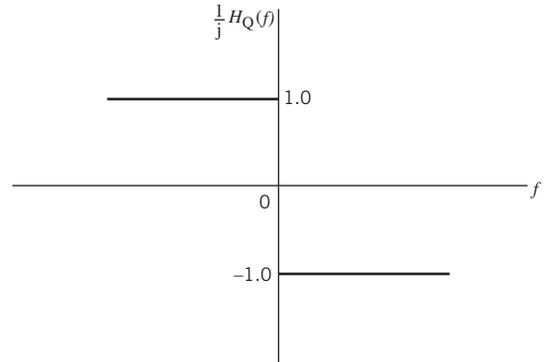
SSB Modulation

Next, considering *SSB modulation*, we may identify two choices:

1. The carrier and the lower sideband are both suppressed, leaving the upper sideband for transmission in its full spectral content; this first SSB-modulated signal is denoted by $s_{\text{USB}}(t)$.
2. The carrier and the upper sideband are both suppressed, leaving the lower sideband for transmission in its full spectral content; this second SSB-modulated signal is denoted by $s_{\text{LSB}}(t)$.

The Fourier transforms of these two modulated signals are the *image* of each other with respect to the carrier frequency f_c , which, as mentioned previously, emphasizes that the transmission of either sideband is actually sufficient for transporting the message signal $m(t)$ over the communication channel. In practical terms, both $s_{\text{USB}}(t)$ and $s_{\text{LSB}}(t)$ require

Figure 2.29
Frequency response of the quadrature filter in SSB modulation.



the smallest feasible channel bandwidth, $B = W$, without compromising the perfect recovery of the message signal under noiseless conditions. It is for these reasons that we say SSB modulation is the *optimum form of linear modulation* for analog communications, preserving both the transmitted power and channel bandwidth in the best manner possible.

SSB modulation may be viewed as a special case of VSB modulation. Specifically, setting the vestigial bandwidth $f_v = 0$, we find that the frequency response of the quadrature filter plotted in Figure 2.28 takes the limiting form of the *signum function* shown in Figure 2.29. In light of the material presented in (2.60) on Hilbert transformation, we therefore find that for $f_v = 0$ the quadrature component $m_Q(t)$ becomes the Hilbert transform of the message signal $m(t)$, denoted by $\hat{m}(t)$. Accordingly, using $\hat{m}(t)$ in place of $m_Q(t)$ in (2.110) yields the SSB formula

$$s(t) = \frac{1}{2}m(t)\cos(2\pi f_c t) \mp \frac{1}{2}\hat{m}(t)\sin(2\pi f_c t) \quad (2.111)$$

where the minus sign applies to the SSB-modulated signal $s_{\text{USB}}(t)$ and the plus sign applies to the alternative SSB-modulated signal $s_{\text{LSB}}(t)$.

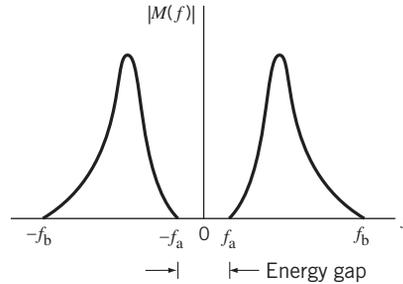
Unlike DSB-SC and VSB methods of modulation, SSB modulation is of limited applicability. Specifically, we say:

For SSB modulation to be feasible in practical terms, the spectral content of the message signal $m(t)$ must have an energy gap centered on the origin.

This requirement, illustrated in Figure 2.30, is imposed on the message signal $m(t)$ so that the band-pass filter in the frequency-discrimination method of Figure 2.26 has a *finite transition band* for the filter to be physically realizable. With the transition band separating the pass-band from the stop-band, it is only when the transition band is finite that the undesired sideband can be suppressed. An example of message signals for which the energy-gap requirement is satisfied is voice signals; for such signals, the energy gap is about 600 Hz, extending from -300 to $+300$ Hz.

In contrast, the spectral contents of television signals and wideband data extend practically to a few hertz, thereby ruling out the applicability of SSB modulation to this second class of message signals. It is for this reason that VSB modulation is preferred over SSB modulation for the transmission of wideband signals.

Figure 2.30
Spectrum of a message signal $m(t)$ with an energy gap centered around the origin.



Summary of Linear Modulation Methods

Equation (2.97) for DSB-SC modulation, (2.110) for VSB modulation, and (2.111) for SSB modulation are summarized in Table 2.5 as special cases of the canonical formula of (2.68). Correspondingly, we may treat the time-domain generations of these three linearly modulated signals as special cases of the “synthesizer” depicted in Figure 2.20b.

Table 2.5 Summary of linear modulation methods viewed as special cases of the canonical formula $s(t) = s_I(t)\cos(2\pi f_c t) - s_Q(t)\sin(2\pi f_c t)$

Type of modulation	In-phase component, $s_I(t)$	Quadrature component, $s_Q(t)$	Comments
DSB-SC	$m(t)$	zero	$m(t)$ = message signal
VSB	$\frac{1}{2}m(t)$	$\pm\frac{1}{2}m_Q(t)$	Plus sign applies to using vestige of lower sideband and minus sign applies to using vestige of upper sideband
SSB	$\frac{1}{2}m(t)$	$\pm\frac{1}{2}\hat{m}(t)$	Plus sign applies to transmission of upper sideband and minus sign applies to transmission of lower sideband

2.15 Phase and Group Delays

A discussion of signal transmission through linear time-invariant systems is incomplete without considering the phase and group delays involved in the signal transmission process.

Whenever a signal is transmitted through a dispersive system, exemplified by a communication channel (or band-pass filter), some *delay* is introduced into the output signal, the delay being measured with respect to the input signal. In an ideal channel, the phase response varies *linearly* with frequency inside the passband of the channel, in which case the filter introduces a constant delay equal to t_0 , where the parameter t_0 controls the slope of the linear phase response of the channel. Now, what if the phase response of the channel is a nonlinear function of frequency, which is frequently the case in practice? The purpose of this section is to address this practical issue.

To begin the discussion, suppose that a steady sinusoidal signal at frequency f_c is transmitted through a dispersive channel that has a phase-shift of $\beta(f_c)$ radians at that frequency. By using two phasors to represent the input signal and the received signal, we see that the received signal phasor lags the input signal phasor by $\beta(f_c)$ radians. The time taken by the received signal phasor to sweep out this phase lag is simply equal to the ratio $\beta(f_c)/(2\pi f_c)$ seconds. This time is called the *phase delay* of the channel.

It is important to realize, however, that the phase delay is not necessarily the true signal delay. This follows from the fact that a steady sinusoidal signal does *not* carry information, so it would be incorrect to deduce from the above reasoning that the phase delay is the true signal delay. To substantiate this statement, suppose that a slowly varying signal, over the interval $-(T/2) \leq t \leq (T/2)$, is multiplied by the carrier, so that the resulting modulated signal consists of a narrow group of frequencies centered around the carrier frequency; the DSB-SC waveform of Figure 2.31 illustrates such a modulated signal. When this modulated signal is transmitted through a communication channel, we find that there is indeed a delay between the envelope of the input signal and that of the received signal. This delay, called the *envelope* or *group delay* of the channel, represents the true signal delay insofar as the information-bearing signal is concerned.

Assume that the dispersive channel is described by the transfer function

$$H(f) = K \exp[j\beta(f)] \quad (2.112)$$

where the amplitude K is a constant scaling factor and the phase $\beta(f)$ is a nonlinear function of frequency f ; it is the nonlinearity of $\beta(f)$ that is responsible for the dispersive

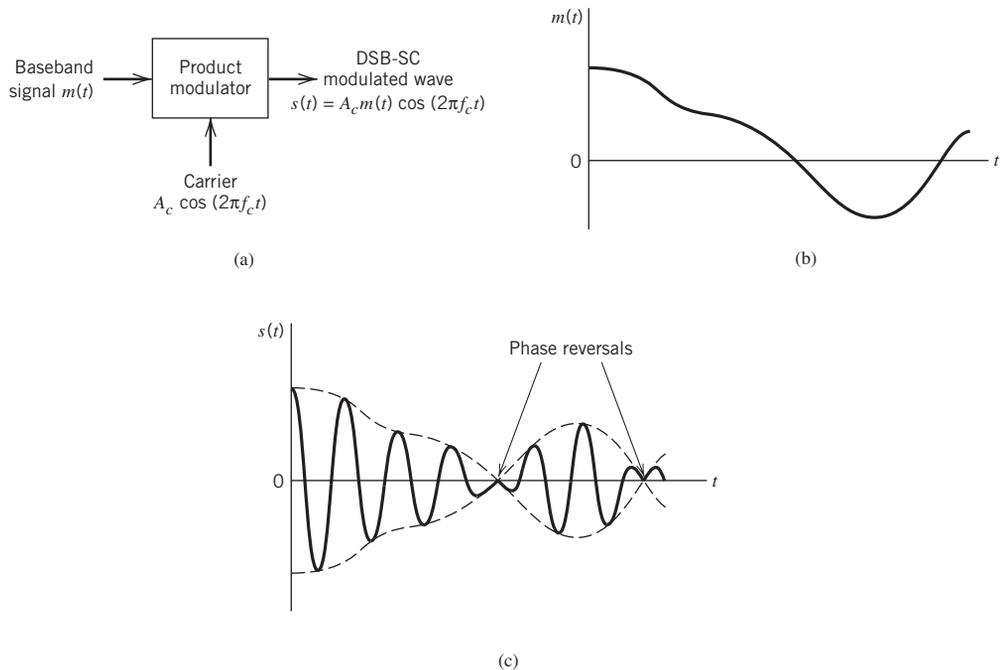


Figure 2.31 (a) Block diagram of product modulator; (b) Baseband signal; (c) DSB-SC modulated wave.

nature of the channel. The input signal $s(t)$ is assumed to be of the kind displayed in Figure 2.31; that is, the DSB-SC-modulated signal

$$s(t) = m(t) \cos(2\pi f_c t) \quad (2.113)$$

where $m(t)$ is the message signal, assumed to be of a low-pass kind and limited to the frequency interval $|f| \leq W$. Moreover, we assume that the carrier frequency $f_c > W$. By expanding the phase $\beta(f)$ in a *Taylor series* about the point $f = f_c$ and retaining only the first two terms, we may approximate $\beta(f)$ as

$$\beta(f) \approx \beta(f_c) + (f - f_c) \left. \frac{\partial \beta(f)}{\partial f} \right|_{f=f_c} \quad (2.114)$$

Define two new terms:

$$\tau_p = -\frac{\beta(f_c)}{2\pi f_c} \quad (2.115)$$

and

$$\tau_g = -\frac{1}{2\pi} \left. \frac{\partial \beta(f)}{\partial f} \right|_{f=f_c} \quad (2.116)$$

Then, we may rewrite (2.114) in the equivalent form

$$\beta(f) \approx -2\pi f_c \tau_p - 2\pi(f - f_c) \tau_g \quad (2.117)$$

Correspondingly, the transfer function of the channel takes the approximate form

$$H(f) \approx K \exp[-j2\pi f_c \tau_p - j2\pi(f - f_c) \tau_g] \quad (2.118)$$

Following the band-pass-to-low-pass transformation described in Section 2.12, in particular using (2.80), we may replace the band-pass channel described by $H(f)$ by an equivalent low-pass filter whose transfer function is approximately given by

$$\tilde{H}(f) \approx 2K \exp(-j2\pi f_c \tau_p - j2\pi f \tau_g), \quad f > f_c \quad (2.119)$$

Correspondingly, using (2.67) we may replace the modulated signal $s(t)$ of (2.113) by its low-pass complex envelope, which, for the DSB-SC example at hand, is simply defined by

$$\tilde{s}(t) = m(t) \quad (2.120)$$

Transforming $\tilde{s}(t)$ into the frequency domain, we may write

$$\tilde{S}(f) = M(f) \quad (2.121)$$

Therefore, in light of (2.96), the Fourier transform of the complex envelope of the signal received at the channel output is given by

$$\begin{aligned} \tilde{X}(f) &= \frac{1}{2} \tilde{H}(f) \tilde{S}(f) \\ &\approx K \exp(-j2\pi f_c \tau_p) \exp(-j2\pi f_c \tau_g) M(f) \end{aligned} \quad (2.122)$$

We note that the multiplying factor $K \exp(-j2\pi f_c \tau_p)$ is a constant for fixed values of f_c and τ_p . We also note from the time-shifting property of the Fourier transform that the term $\exp(-j2\pi f_c \tau_g) M(f)$ represents the Fourier transform of the delayed signal $m(t - \tau_g)$. Accordingly, the complex envelope of the channel output is

$$\tilde{x}(t) = K \exp(-j2\pi f_c \tau_p) m(t - \tau_g) \quad (2.123)$$

Finally, using (2.66) we find that the actual channel output is itself given by

$$\begin{aligned} x(t) &= \operatorname{Re}[\tilde{x}(t)\exp(j2\pi f_c t)] \\ &= Km(t - \tau_g)\cos[2\pi f_c(t - \tau_p)] \end{aligned} \tag{2.124}$$

Equation (2.124) reveals that, as a result of transmitting the modulated signal $s(t)$ through the dispersive channel, two different delay effects occur at the channel output:

1. The sinusoidal carrier wave $\cos(2\pi f_c t)$ is delayed by τ_p seconds; hence, τ_p represents the *phase delay*; sometimes τ_p is referred to as the *carrier delay*.
2. The envelope $m(t)$ is delayed by τ_g seconds; hence, τ_g represents the *envelope* or *group delay*.

Note that τ_g is related to the slope of the phase $\beta(f)$, measured at $f = f_c$. Note also that when the phase response $\beta(f)$ varies linearly with frequency f and $\beta(f_c)$ is zero, the phase delay and group delay assume a common value. It is only then that we can think of these two delays being equal.

2.16 Numerical Computation of the Fourier Transform

The material presented in this chapter clearly testifies to the importance of the Fourier transform as a theoretical tool for the representation of deterministic signals and linear time-invariant systems, be they of the low-pass or band-pass kind. The importance of the Fourier transform is further enhanced by the fact that there exists a class of algorithms called FFT algorithms⁶ for numerical computation of the Fourier transform in an efficient manner.

The FFT algorithm is derived from the discrete Fourier transform (DFT) in which, as the name implies, both time and frequency are represented in discrete form. The DFT provides an *approximation* to the Fourier transform. In order to properly represent the information content of the original signal, we have to take special care in performing the sampling operations involved in defining the DFT. A detailed treatment of the sampling process is presented in Chapter 6. For the present, it suffices to say that, given a band-limited signal, the sampling rate should be greater than twice the highest frequency component of the input signal. Moreover, if the samples are uniformly spaced by T_s seconds, the spectrum of the signal becomes periodic, repeating every $f_s = (1/T_s)$ hz in accordance with (2.43). Let N denote the number of frequency samples contained in the interval f_s . Hence, the *frequency resolution* involved in numerical computation of the Fourier transform is defined by

$$\Delta f = \frac{f_s}{N} = \frac{1}{NT_s} = \frac{1}{T} \tag{2.125}$$

where T is the total duration of the signal.

Consider then a *finite data sequence* $\{g_0, g_1, \dots, g_{N-1}\}$. For brevity, we refer to this sequence as g_n , in which the subscript is the *time index* $n = 0, 1, \dots, N-1$. Such a sequence may represent the result of sampling an analog signal $g(t)$ at times $t = 0, T_s, \dots, (N-1)T_s$, where T_s is the sampling interval. The ordering of the data sequence defines the sample

time in that g_0, g_1, \dots, g_{N-1} denote samples of $g(t)$ taken at times $0, T_s, \dots, (N-1)T_s$, respectively. Thus we have

$$g_n = g(nT_s) \quad (2.126)$$

We formally define the DFT of g_n as

$$G_k = \sum_{n=0}^{N-1} g_n \exp\left(-\frac{j2\pi}{N}kn\right) \quad k = 0, 1, \dots, N-1 \quad (2.127)$$

The sequence $\{G_0, G_1, \dots, G_{N-1}\}$ is called the *transform sequence*. For brevity, we refer to this second sequence simply as G_k , in which the subscript is the *frequency index* $k = 0, 1, \dots, N-1$.

Correspondingly, we define the *inverse discrete Fourier transform* (IDFT) of G_k as

$$g_n = \frac{1}{N} \sum_{k=0}^{N-1} G_k \exp\left(\frac{j2\pi}{N}kn\right) \quad n = 0, 1, \dots, N-1 \quad (2.128)$$

The DFT and the IDFT form a discrete transform pair. Specifically, given a data sequence g_n , we may use the DFT to compute the transform sequence G_k ; and given the transform sequence G_k , we may use the IDFT to recover the original data sequence g_n . A distinctive feature of the DFT is that, for the finite summations defined in (2.127) and (2.128), there is no question of convergence.

When discussing the DFT (and algorithms for its computation), the words “sample” and “point” are used interchangeably to refer to a sequence value. Also, it is common practice to refer to a sequence of length N as an *N -point sequence* and to refer to the DFT of a data sequence of length N as an *N -point DFT*.

Interpretation of the DFT and the IDFT

We may visualize the DFT process described in (2.127) as a collection of N *complex heterodyning* and *averaging* operations, as shown in Figure 2.32a. We say that the heterodyning is complex in that samples of the data sequence are multiplied by *complex exponential sequences*. There is a total of N complex exponential sequences to be considered, corresponding to the frequency index $k = 0, 1, \dots, N-1$. Their periods have been selected in such a way that each complex exponential sequence has precisely an integer number of cycles in the total interval 0 to $N-1$. The zero-frequency response, corresponding to $k = 0$, is the only exception.

For the interpretation of the IDFT process, described in (2.128), we may use the scheme shown in Figure 2.32b. Here we have a collection of N *complex signal generators*, each of which produces the complex exponential sequence

$$\begin{aligned} \exp\left(\frac{j2\pi}{N}kn\right) &= \cos\left(\frac{2\pi}{N}kn\right) + j \sin\left(\frac{2\pi}{N}kn\right) \\ &= \left\{ \cos\left(\frac{2\pi}{N}kn\right), \sin\left(\frac{2\pi}{N}kn\right) \right\}_{k=0}^{N-1} \end{aligned} \quad (2.129)$$

Thus, in reality, each complex signal generator consists of a pair of generators that output a cosinusoidal and a sinusoidal sequence of k cycles per observation interval. The output

of each complex signal generator is weighted by the complex Fourier coefficient G_k . At each time index n , an output is formed by summing the weighted complex generator outputs.

It is noteworthy that although the DFT and the IDFT are similar in their mathematical formulations, as described in (2.127) and (2.128), their interpretations as depicted in Figure 2.32a and b are so completely different.

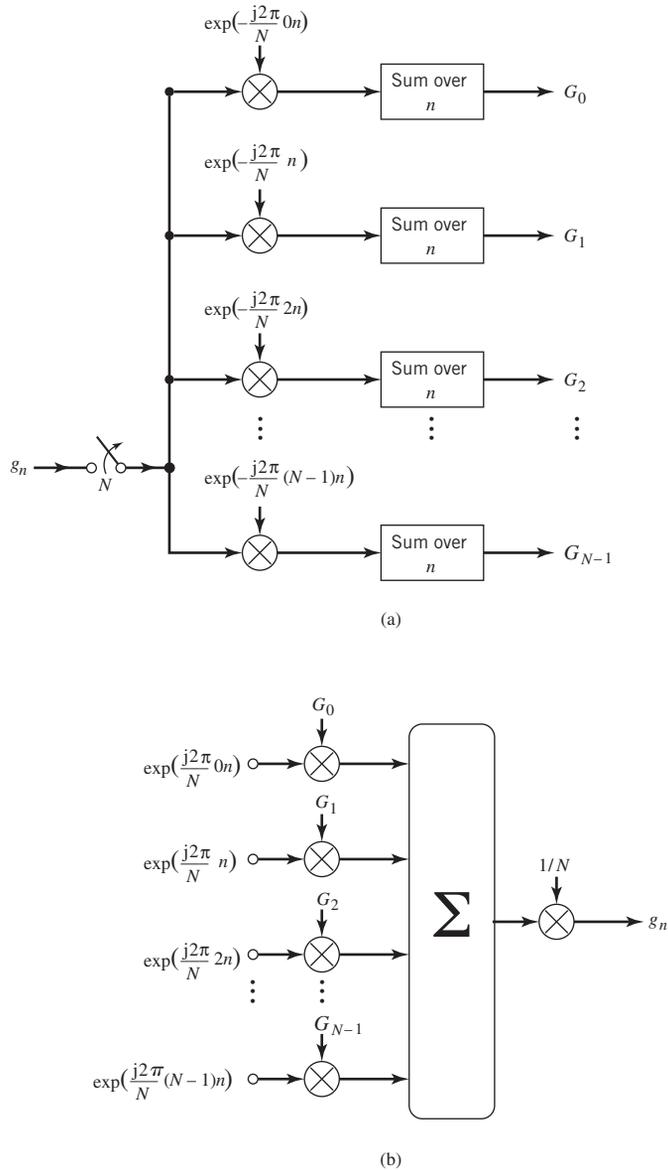


Figure 2.32 Interpretations of (a) the DFT and (b) the IDFT.

Also, the addition of harmonically related periodic signals, involved in these two parts of the figure, suggests that their outputs G_k and g_n must be both periodic. Moreover, the processors shown in Figure 2.32 are linear, suggesting that the DFT and IDFT are both linear operations. This important property is also obvious from the defining equations (2.127) and (2.128).

FFT Algorithms

In the DFT both the input and the output consist of sequences of numbers defined at uniformly spaced points in time and frequency, respectively. This feature makes the DFT ideally suited for direct numerical evaluation on a computer. Moreover, the computation can be implemented most efficiently using a class of algorithms, collectively called *FFT algorithms*. An algorithm refers to a “recipe” that can be written in the form of a computer program.

FFT algorithms are efficient because they use a greatly reduced number of arithmetic operations as compared with the brute force (i.e., direct) computation of the DFT. Basically, an FFT algorithm attains its computational efficiency by following the engineering strategy of “divide and conquer,” whereby the original DFT computation is decomposed successively into smaller DFT computations. In this section, we describe one version of a popular FFT algorithm, the development of which is based on such a strategy.

To proceed with the development, we first rewrite (2.127), defining the DFT of g_n in the convenient mathematical form

$$G_k = \sum_{n=0}^{N-1} g_n W^{kn}, \quad k = 0, 1, \dots, N-1 \quad (2.130)$$

where we have introduced the complex parameter

$$W = \exp\left(-\frac{j2\pi}{N}\right) \quad (2.131)$$

From this definition, we readily see that

$$\begin{aligned} W^N &= 1 \\ W^{N/2} &= -1 \\ W^{(l+1N)(n+mN)} &= W^{kn}, \quad (m, l) = 0, \pm 1, \pm 2, \dots \end{aligned}$$

That is, W^{kn} is periodic with period N . The periodicity of W^{kn} is a key feature in the development of FFT algorithms.

Let N , the number of points in the data sequence, be an integer power of two, as shown by

$$N = 2^L$$

where L is an integer; the rationale for this choice is explained later. Since N is an even integer, $N/2$ is an integer, and so we may divide the data sequence into the first half and last half of the points.

Thus, we may rewrite (2.130) as

$$\begin{aligned}
 G_k &= \sum_{n=0}^{(N/2)-1} g_n W^{kn} + \sum_{n=N/2}^{N-1} g_n W^{kn} \\
 &= \sum_{n=0}^{(N/2)-1} g_n W^{kn} + \sum_{n=0}^{(N/2)-1} g_{n+N/2} W^{k(n+N/2)} \\
 &= \sum_{n=0}^{(N/2)-1} (g_n + g_{n+N/2} W^{kN/2}) W^{kn} \quad k = 0, 1, \dots, N-1
 \end{aligned} \tag{2.132}$$

Since $W^{N/2} = -1$, we have

$$W^{kN/2} = (-1)^k$$

Accordingly, the factor $W^{kN/2}$ in (2.132) takes on only one of two possible values, namely $+1$ or -1 , depending on whether the frequency index k is even or odd, respectively. These two cases are considered in what follows.

First, let k be even, so that $W^{kN/2} = 1$. Also let

$$k = 2l, \quad l = 0, 1, \dots, \frac{N}{2} - 1$$

and define

$$x_n = g_n + g_{n+N/2} \tag{2.133}$$

Then, we may put (2.132) into the new form

$$\begin{aligned}
 G_{2l} &= \sum_{n=0}^{(N/2)-1} x_n W^{2ln} \\
 &= \sum_{n=0}^{(N/2)-1} x_n (W^2)^{ln} \quad l = 0, 1, \dots, \frac{N}{2} - 1
 \end{aligned} \tag{2.134}$$

From the definition of W given in (2.131), we readily see that

$$\begin{aligned}
 W^2 &= \exp\left(-\frac{j4\pi}{N}\right) \\
 &= \exp\left(-\frac{j2\pi}{N/2}\right)
 \end{aligned}$$

Hence, we recognize the sum on the right-hand side of (2.134) as the $(N/2)$ -point DFT of the sequence x_n .

Next, let k be odd so that $W^{kN/2} = -1$. Also, let

$$k = 2l + 1, \quad l = 0, 1, \dots, \frac{N}{2} - 1$$

and define

$$y_n = g_n - g_{n+N/2} \quad (2.135)$$

Then, we may put (2.132) into the corresponding form

$$\begin{aligned} G^{2l+1} &= \sum_{n=0}^{(N/2)-1} y_n W^{(2l+1)n} \\ &= \sum_{n=0}^{(N/2)-1} [y_n W^n] (W^2)^{ln} \quad l = 0, 1, \dots, \frac{N}{2} - 1 \end{aligned} \quad (2.136)$$

We recognize the sum on the right-hand side of (2.136) as the $(N/2)$ -point DFT of the sequence $y_n W^n$. The parameter W^n associated with y_n is called the *twiddle factor*.

Equations (2.134) and (2.136) show that the even- and odd-valued samples of the transform sequence G_k can be obtained from the $(N/2)$ -point DFTs of the sequences x_n and $y_n W^n$, respectively. The sequences x_n and y_n are themselves related to the original data sequence g_n by (2.133) and (2.135), respectively. Thus, the problem of computing an N -point DFT is reduced to that of computing two $(N/2)$ -point DFTs. The procedure just described is repeated a second time, whereby an $(N/2)$ -point DFT is decomposed into two $(N/4)$ -point DFTs. The decomposition procedure is continued in this fashion until (after $L = \log_2 N$ stages) we reach the trivial case of N single-point DFTs.

Figure 2.33 illustrates the computations involved in applying the formulas of (2.134) and (2.136) to an eight-point data sequence; that is, $N = 8$. In constructing left-hand portions of the figure, we have used signal-flow graph notation. A *signal-flow graph* consists of an interconnection of *nodes* and *branches*. The *direction* of signal transmission along a branch is indicated by an arrow. A branch multiplies the variable at a node (to which it is connected) by the branch *transmittance*. A node sums the outputs of all incoming branches. The convention used for branch transmittances in Figure 2.33 is as follows. When no coefficient is indicated on a branch, the transmittance of that branch is assumed to be unity. For other branches, the transmittance of a branch is indicated by -1 or an integer power of W , placed alongside the arrow on the branch.

Thus, in Figure 2.33a the computation of an eight-point DFT is reduced to that of two four-point DFTs. The procedure for the eight-point DFT may be mimicked to simplify the computation of the four-point DFT. This is illustrated in Figure 2.33b, where the computation of a four-point DFT is reduced to that of two two-point DFTs. Finally, the computation of a two-point DFT is shown in Figure 2.33c.

Combining the ideas described in Figure 2.33, we obtain the complete signal-flow graph of Figure 2.34 for the computation of the eight-point DFT. A repetitive structure, called the *butterfly* with two inputs and two outputs, can be discerned in the FFT algorithm of Figure 2.34. Examples of butterflies (for the three stages of the algorithm) are shown by the bold-faced lines in Figure 2.34.

For the general case of $N = 2^L$, the algorithm requires $L = \log_2 N$ stages of computation. Each stage requires $(N/2)$ butterflies. Each butterfly involves one complex multiplication and two complex additions (to be precise, one addition and one subtraction). Accordingly, the FFT structure described here requires $(N/2)\log_2 N$ complex multiplications and $N\log_2 N$

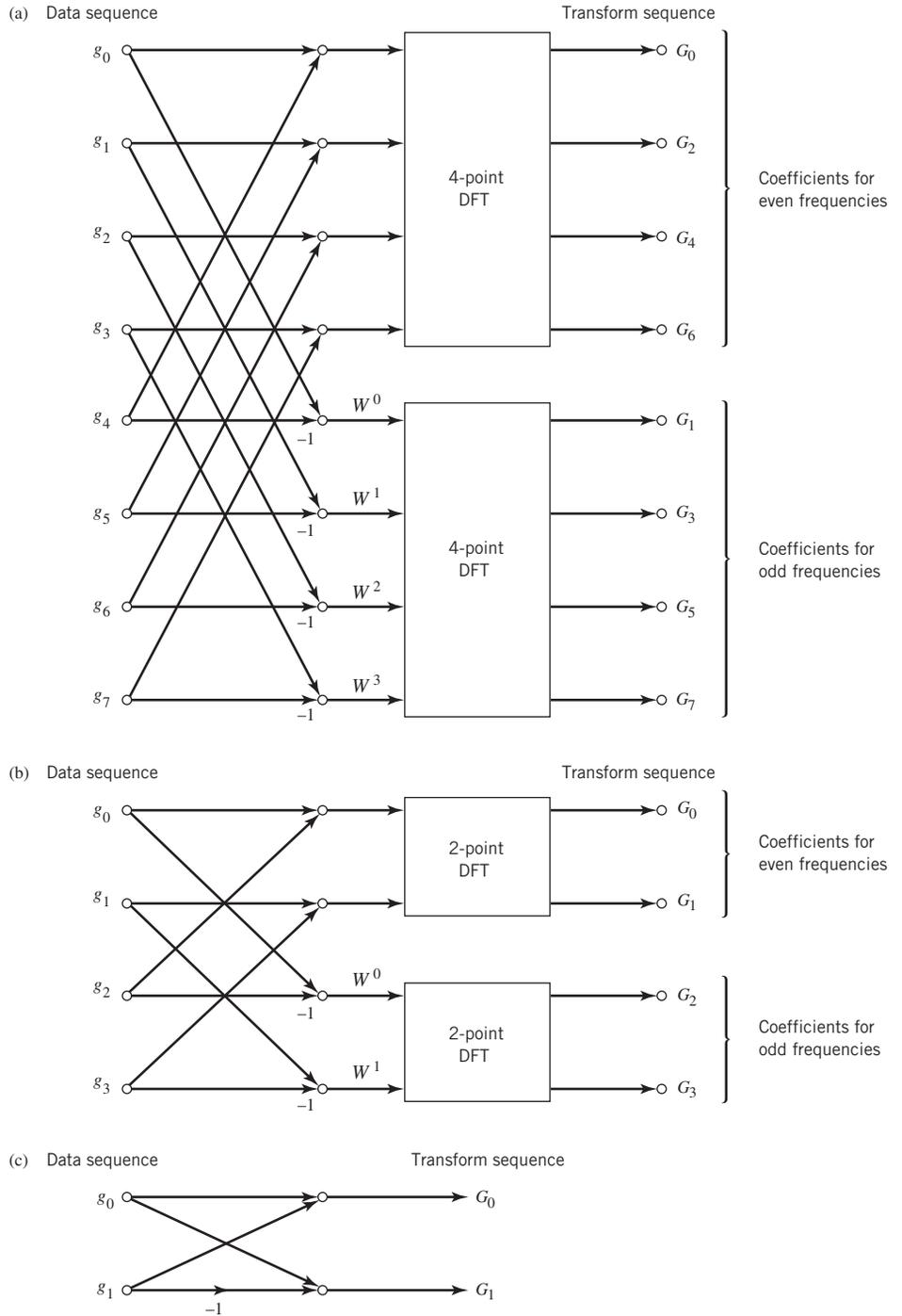
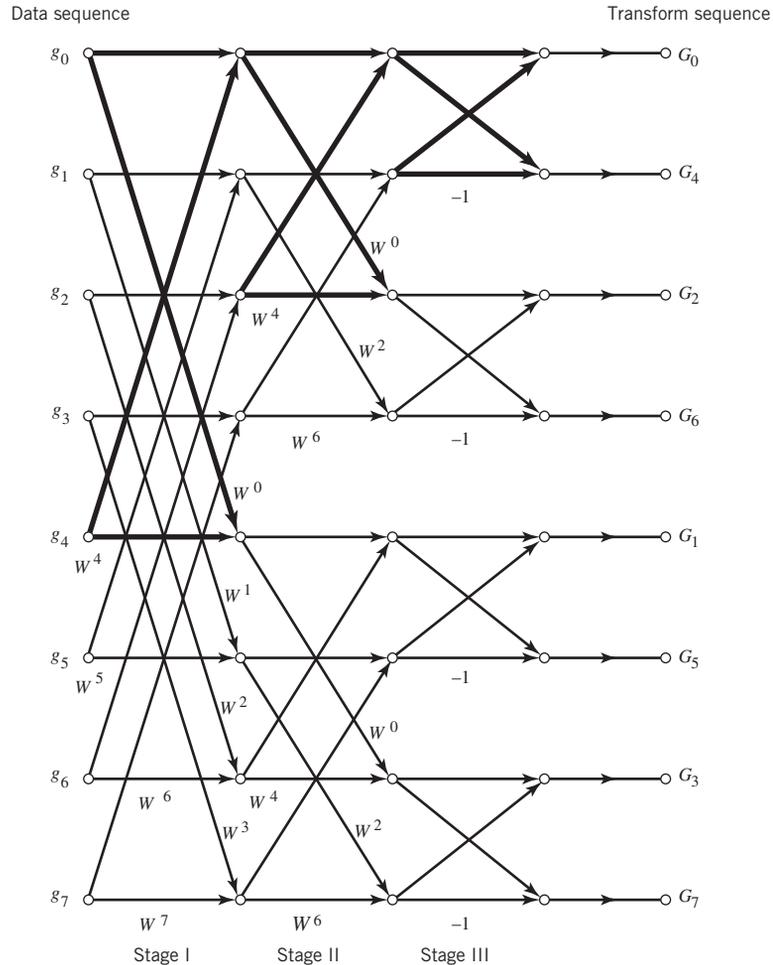


Figure 2.33 (a) Reduction of eight-point DFT into two four-point DFTs. (b) Reduction of four-point DFT into two two-point DFTs. (c) Trivial case of two-point DFT.

Figure 2.34
Decimation-in-frequency
FFT algorithm.



complex additions; actually, the number of multiplications quoted is pessimistic, because we may omit all twiddle factors $W^0 = 1$ and $W^{N/2} = -1$, $W^{N/4} = j$, $W^{3N/4} = -j$. This computational complexity is significantly smaller than that of the N^2 complex multiplications and $N(N - 1)$ complex additions required for *direct* computation of the DFT. The computational savings made possible by the FFT algorithm become more substantial as we increase the data length N . For example, for $N = 8192 = 2^{11}$, the direct approach requires approximately 630 times as many arithmetic operations as the FFT algorithm, hence the popular use of the FFT algorithm in computing the DFT.

We may establish two other important features of the FFT algorithm by carefully examining the signal-flow graph shown in Figure 2.34:

1. At each stage of the computation, the new set of N complex numbers resulting from the computation can be stored in the same memory locations used to store the previous set. This kind of computation is referred to as *in-place computation*.

2. The samples of the transform sequence G_k are stored in a bit-reversed order. To illustrate the meaning of this terminology, consider Table 2.6 constructed for the case of $N = 8$. At the left of the table, we show the eight possible values of the frequency index k (in their natural order) and their 3-bit binary representations. At the right of the table, we show the corresponding bit-reversed binary representations and indices. We observe that the bit-reversed indices in the rightmost column of Table 2.6 appear in the same order as the indices at the output of the FFT algorithm in Figure 2.34.

Table 2.6 Illustrating bit reversal

Frequency index, k	Binary representation	Bit-reversed binary representation	Bit-reversed index
0	000	000	0
1	001	100	4
2	010	010	2
3	011	110	6
4	100	001	1
5	101	101	5
6	110	011	3
7	111	111	7

The FFT algorithm depicted in Figure 2.34 is referred to as a *decimation-in-frequency algorithm*, because the transform (frequency) sequence G_k is divided successively into smaller subsequences. In another popular FFT algorithm, called a *decimation-in-time algorithm*, the data (time) sequence g_n is divided successively into smaller subsequences. Both algorithms have the same computational complexity. They differ from each other in two respects. First, for decimation-in-frequency, the input is in natural order, whereas the output is in bit-reversed order; the reverse is true for decimation-in-time. Second, the butterfly for decimation-in-time is slightly different from that for decimation-in-frequency. The reader is invited to derive the details of the decimation-in-time algorithm using the divide-and-conquer strategy that led to the development of the algorithm described in Figure 2.34.

In devising the FFT algorithm presented herein, we placed the factor $1/N$ in the formula for the forward DFT, as shown in (2.128). In some other FFT algorithms, location of the factor $1/N$ is reversed. In yet other formulations, the factor $1/\sqrt{N}$ is placed in the formulas for both the forward and inverse DFTs for the sake of symmetry.

Computation of the IDFT

The IDFT of the transform G_k is defined by (2.128). We may rewrite this equation in terms of the complex parameter W as

$$g_n = \frac{1}{N} \sum_{k=0}^{N-1} G_k W^{-kn}, \quad n = 0, 1, \dots, N-1 \quad (2.137)$$

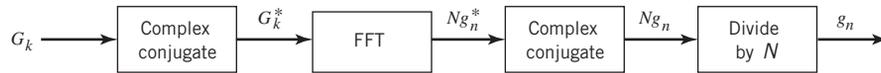


Figure 2.35 Use of the FFT algorithm for computing the IDFT.

Taking the complex conjugate of (2.137) and multiplying by N , we get

$$N g_n^* = \sum_{k=0}^{N-1} G_k^* W^{-kn}, \quad n = 0, 1, \dots, N-1 \quad (2.138)$$

The right-hand side of (2.138) is recognized as the N -point DFT of the complex-conjugated sequence G_k^* . Accordingly, (2.138) suggests that we may compute the desired sequence g_n using the scheme shown in Figure 2.35, based on an N -point FFT algorithm. Thus, the same FFT algorithm can be used to handle the computation of both the IDFT and the DFT.

2.17 Summary and Discussion

In this chapter we have described the Fourier transform as a fundamental tool for relating the time-domain and frequency-domain descriptions of a deterministic signal. The signal of interest may be an energy signal or a power signal. The Fourier transform includes the exponential Fourier series as a special case, provided that we permit the use of the Dirac delta function.

An inverse relationship exists between the time-domain and frequency-domain descriptions of a signal. Whenever an operation is performed on the waveform of a signal in the time domain, a corresponding modification is applied to the spectrum of the signal in the frequency domain. An important consequence of this inverse relationship is the fact that the time–bandwidth product of an energy signal is a constant; the definitions of signal duration and bandwidth merely affect the value of the constant.

An important signal-processing operation frequently encountered in communication systems is that of linear filtering. This operation involves the convolution of the input signal with the impulse response of the filter or, equivalently, the multiplication of the Fourier transform of the input signal by the transfer function (i.e., Fourier transform of the impulse response) of the filter. Low-pass and band-pass filters represent two commonly used types of filters. Band-pass filtering is usually more complicated than low-pass filtering. However, through the combined use of a complex envelope for the representation of an input band-pass signal and the complex impulse response for the representation of a band-pass filter, we may formulate a complex low-pass equivalent for the band-pass filtering problem and thereby replace a difficult problem with a much simpler one. It is also important to note that there is no loss of information in establishing this equivalence. A rigorous treatment of the concepts of complex envelope and complex impulse response as presented in this chapter is rooted in Hilbert transformation.

The material on Fourier analysis, as presented in this chapter, deals with signals whose waveforms can be nonperiodic or periodic, and whose spectra can be continuous or discrete functions of frequency. In this sense, the material has general appeal.

Building on the canonical representation of a band-pass signal involving the in-phase and quadrature components of the signal, we showed that this representation provides an elegant way of describing the three basic forms of linear modulation, namely DSB-SC, VSB, and SSB.

With the Fourier transform playing such a pervasive role in the study of signals and linear systems, we finally described the FFT algorithm as an efficient tool for numerical computation of the DFT that represents the uniformly sampled versions of the forward and inverse forms of the ordinary Fourier transform.

Problems

The Fourier Transform

- 2.1 Prove the dilation property of the Fourier transform, listed as Property 2 in Table 2.1.
- 2.2
- Prove the duality property of the Fourier transform, listed as Property 3 in Table 2.1.
 - Prove the time-shifting property, listed as Property 4; and then use the duality property to prove the frequency-shifting property, listed as Property 5 in the table.
 - Using the frequency-shifting property, determine the Fourier transform of the radio frequency RF pulse

$$g(t) = \text{Arect}\left(\frac{t}{T}\right) \cos(2\pi f_c t)$$

assuming that f_c is larger than $(1/T)$.

- 2.3
- Prove the multiplication-in-the-time-domain property of the Fourier transform, listed as Property 11 in Table 2.1.
 - Prove the convolution in the time-domain property, listed as Property 12.
 - Using the result obtained in part b, prove the correlation theorem, listed as Property 13.
- 2.4 Prove Rayleigh's energy theorem listed as Property 14 in Table 2.1.
- 2.5 The following expression may be viewed as an approximate representation of a pulse with finite rise time:

$$g(t) = \frac{1}{\tau} \int_{t-T}^{t+T} \exp\left(-\frac{\pi u^2}{\tau^2}\right) du$$

where it is assumed that $T \gg \tau$. Determine the Fourier transform of $g(t)$. What happens to this transform when we allow τ to become zero? *Hint*: Express $g(t)$ as the superposition of two signals, one corresponding to integration from $t - T$ to 0, and the other from 0 to $t + T$.

- 2.6 The Fourier transform of a signal $g(t)$ is denoted by $G(f)$. Prove the following properties of the Fourier transform:
- If a real signal $g(t)$ is an even function of time t , the Fourier transform $G(f)$ is purely real. If a real signal $g(t)$ is an odd function of time t , the Fourier transform $G(f)$ is purely imaginary.

b.

$$t^n g(t) \Leftrightarrow \left(\frac{j}{2\pi}\right)^n G^{(n)}(f)$$

where $G^{(n)}(f)$ is the n th derivative of $G(f)$ with respect to f .

c.

$$\int_{-\infty}^{\infty} t^n g(t) dt = \left(\frac{j}{2\pi}\right)^n G^{(n)}(0)$$

- d. Assuming that both $g_1(t)$ and $g_2(t)$ are complex signals, show that:

$$g_1(t)g_2^*(t) \Leftrightarrow \int_{-\infty}^{\infty} G_1(\lambda)G_2^*(\lambda-f) d\lambda$$

and

$$\int_{-\infty}^{\infty} g_1(t)g_2^*(t)dt = \int_{-\infty}^{\infty} G_1(f)G_2^*(f)df$$

- 2.7 a. The *root mean-square (rms) bandwidth* of a low-pass signal $g(t)$ of finite energy is defined by

$$W_{\text{rms}} = \left[\frac{\int_{-\infty}^{\infty} f^2 |G(f)|^2 df}{\int_{-\infty}^{\infty} |G(f)|^2 df} \right]^{1/2}$$

where $|G(f)|^2$ is the energy spectral density of the signal. Correspondingly, the *root mean-square (rms) duration* of the signal is defined by

$$T_{\text{rms}} = \left[\frac{\int_{-\infty}^{\infty} t^2 |g(t)|^2 dt}{\int_{-\infty}^{\infty} |g(t)|^2 dt} \right]^{1/2}$$

Using these definitions, show that

$$T_{\text{rms}} W_{\text{rms}} \geq \frac{1}{4\pi}$$

Assume that $|g(t)| \rightarrow 0$ faster than $1/\sqrt{|t|}$ as $|t| \rightarrow \infty$.

- b. Consider a Gaussian pulse defined by

$$g(t) = \exp(-\pi t^2)$$

Show that for this signal the equality

$$T_{\text{rms}} W_{\text{rms}} = \frac{1}{4\pi}$$

is satisfied.

Hint: Use Schwarz's inequality

$$\left(\int_{-\infty}^{\infty} [g_1^*(t)g_2(t) + g_1(t)g_2^*(t)] dt \right)^2 \leq 4 \int_{-\infty}^{\infty} |g_1(t)|^2 dt \int_{-\infty}^{\infty} |g_2(t)|^2 dt$$

in which we set

$$g_1(t) = tg(t)$$

and

$$g_2(t) = \frac{dg(t)}{dt}$$

- 2.8 The *Dirac comb*, formulated in the time domain, is defined by

$$\delta_{T_0}(t) = \sum_{m=-\infty}^{\infty} \delta(t - mT_0)$$

where T_0 is the period.

- a. Show that the Dirac comb is its own Fourier transform. That is, the Fourier transform of $\delta_{T_0}(t)$ is also an infinitely long periodic train of delta functions, weighted by the factor $f_0 = (1/T_0)$ and regularly spaced by f_0 along the frequency axis.
- b. Hence, prove the pair of dual relations:

$$\sum_{m=-\infty}^{\infty} \delta(t - mT_0) = f_0 \sum_{n=-\infty}^{\infty} \exp(j2\pi n f_0 t)$$

$$T_0 \sum_{m=-\infty}^{\infty} \exp(j2\pi m f T_0) = \sum_{n=-\infty}^{\infty} \delta(f - n f_0)$$

- c. Finally, prove the validity of (2.38).

Signal Transmission through Linear Time-invariant Systems

- 2.9 The periodic signal

$$x(t) = \sum_{m=-\infty}^{\infty} x(nT_0) \delta(t - nT_0)$$

is applied to a linear system of impulse response $h(t)$. Show that the average power of the signal $y(t)$ produced at the system output is defined by

$$P_{av,y} = \sum_{n=-\infty}^{\infty} |x(nT_0)|^2 |H(nf_0)|^2$$

where $H(f)$ is the frequency response of the system, and $f_0 = 1/T_0$.

- 2.10 According to the bounded input–bounded output stability criterion, the impulse response $h(t)$ of a linear-invariant system must be absolutely integrable; that is,

$$\int_{-\infty}^{\infty} |h(t)| dt < \infty$$

Prove that this condition is both necessary and sufficient for stability of the system.

Hilbert Transform and Pre-envelopes

- 2.11 Prove the three properties of the Hilbert transform itemized on pages 43 and 44.
- 2.12 Let $\hat{g}(t)$ denote the Hilbert transform of $g(t)$. Derive the set of Hilbert-transform pairs listed as items 5 to 8 in Table 2.3.
- 2.13 Evaluate the inverse Fourier transform $g(t)$ of the one-sided frequency function:

$$G(f) = \begin{cases} \exp(-f), & f > 0 \\ \frac{1}{2}, & f = 0 \\ 0, & f < 0 \end{cases}$$

Show that $g(t)$ is complex, and that its real and imaginary parts constitute a Hilbert-transform pair.

- 2.14 Let $\hat{g}(t)$ denote the Hilbert transform of a Fourier transformable signal $g(t)$. Show that $\frac{d}{dt} \hat{g}(t)$ is equal to the Hilbert transform of $\frac{d}{dt} g(t)$.

- 2.15 In this problem, we revisit Problem 2.14, except that this time we use integration rather than differentiation. Doing so, we find that, in general, the integral $\int_{-\infty}^{\infty} \hat{g}(t) dt$ is *not* equal to the Hilbert transform of the integral $\int_{-\infty}^{\infty} g(t) dt$.
- Justify this statement.
 - Find the condition for which exact equality holds.
- 2.16 Determine the pre-envelope $g_+(t)$ corresponding to each of the following two signals:
- $g(t) = \text{sinc}(t)$
 - $g(t) = [1 + k \cos(2\pi f_m t)] \cos(2\pi f_c t)$

Complex Envelope

- 2.17 Show that the complex envelope of the sum of two narrowband signals (with the same carrier frequency) is equal to the sum of their individual complex envelopes.
- 2.18 The definition of the complex envelope $\tilde{s}(t)$ of a band-pass signal given in (2.65) is based on the pre-envelope $s_+(t)$ for positive frequencies. How is the complex envelope defined in terms of the pre-envelope $s_-(t)$ for negative frequencies? Justify your answer.
- 2.19 Consider the signal

$$s(t) = c(t)m(t)$$

whose $m(t)$ is a low-pass signal whose Fourier transform $M(f)$ vanishes for $|f| > W$, and $c(t)$ is a high-pass signal whose Fourier transform $C(f)$ vanishes for $|f| < W$. Show that the Hilbert transform of $s(t)$ is $\hat{s}(t) = \hat{c}(t)m(t)$, where $\hat{c}(t)$ is the Hilbert transform of $c(t)$.

- 2.20 a. Consider two real-valued signals $s_1(t)$ and $s_2(t)$ whose pre-envelopes are denoted by $s_{1+}(t)$ and $s_{2+}(t)$, respectively. Show that

$$\int_{-\infty}^{\infty} \text{Re}[s_{1+}(t)] \text{Re}[s_{2+}(t)] dt = \frac{1}{2} \text{Re} \left[\int_{-\infty}^{\infty} s_{1+}(t) s_{2+}^*(t) dt \right]$$

- Suppose that $s_2(t)$ is replaced by $s_2(-t)$. Show that this modification has the effect of removing the complex conjugation in the right-hand side of the formula given in part a.
- Assuming that $s(t)$ is a narrowband signal with complex envelope $\tilde{s}(t)$ and carrier frequency f_c , use the result of part a to show that

$$\int_{-\infty}^{\infty} s^2(t) dt = \frac{1}{2} \int_{-\infty}^{\infty} |\tilde{s}(t)|^2 dt$$

- 2.21 Let a narrow-band signal $s(t)$ be expressed in the form

$$s(t) = s_1(t) \cos(2\pi f_c t) - s_Q(t) \sin(2\pi f_c t)$$

Using $S_+(f)$ to denote the Fourier transform of the pre-envelope of $s_+(t)$, show that the Fourier transforms of the in-phase component $s_1(t)$ and quadrature component $s_Q(t)$ are given by

$$S_1(f) = \frac{1}{2} [S_+(f + f_c) + S_+^*(-f + f_c)]$$

$$S_Q(f) = \frac{1}{2j} [S_+(f + f_c) - S_+^*(-f + f_c)]$$

respectively, where the asterisk denotes complex conjugation.

- 2.22 The block diagram of Figure 2.20a illustrates a method for extracting the in-phase component $s_1(t)$ and quadrature component $s_Q(t)$ of a narrowband signal $s(t)$. Given that the spectrum of $s(t)$ is limited to the interval $f_c - W \leq |f| \leq f_c + W$, demonstrate the validity of this method. Hence, show that

$$S_I(f) = \begin{cases} S(f-f_c) + S(f+f_c), & -W \leq f \leq W \\ 0, & \text{elsewhere} \end{cases}$$

and

$$S_Q(f) = \begin{cases} j[S(f-f_c) - S(f+f_c)], & -W \leq f \leq W \\ 0, & \text{elsewhere} \end{cases}$$

where $S_I(f)$, $S_Q(f)$, and $S(f)$ are the Fourier transforms of $s_I(t)$, $s_Q(t)$, and $s(t)$, respectively.

Low-Pass Equivalent Models of Band-Pass Systems

- 2.23** Equations (2.82) and (2.83) define the in-phase component $\tilde{H}_I(f)$ and the quadrature component $\tilde{H}_Q(f)$ of the frequency response $\tilde{H}(f)$ of the complex low-pass equivalent model of a band-pass system of impulse response $h(t)$. Prove the validity of these two equations.
- 2.24** Explain what happens to the low-pass equivalent model of Figure 2.21b when the amplitude response of the corresponding bandpass filter has even symmetry and the phase response has odd symmetry with respect to the mid-band frequency f_c .
- 2.25** The rectangular RF pulse

$$x(t) = \begin{cases} A \cos(2\pi f_c t), & 0 \leq t \leq T \\ 0, & \text{elsewhere} \end{cases}$$

is applied to a linear filter with impulse response

$$h(t) = x(T-t)$$

Assume that the frequency f_c equals a large integer multiple of $1/T$. Determine the response of the filter and sketch it.

- 2.26** Figure P2.26 depicts the frequency response of an idealized band-pass filter in the receiver of a communication system, namely $H(f)$, which is characterized by a bandwidth of $2B$ centered on the carrier frequency f_c . The signal applied to the band-pass filter is described by the modulated sinc function:

$$x(t) = 4A_c B \operatorname{sinc}(2Bt) \cos[2\pi(f_c \pm \Delta f)t]$$

where Δf is *frequency misalignment* introduced due to the receiver's imperfections, measured with respect to the carrier $A_c \cos(2\pi f_c t)$.

- a.** Find the complex low-pass equivalent models of the signal $x(t)$ and the frequency response $H(f)$.

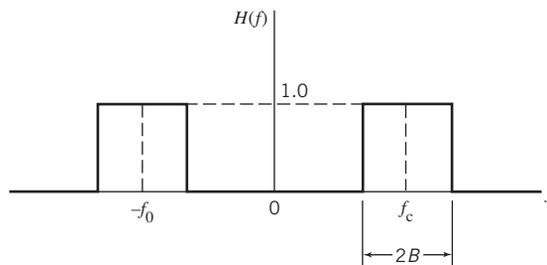


Figure P2.26

- b. Then, go on to find the complex low-pass response of the filter output, denoted by $\tilde{y}(t)$, which includes distortion due to $\pm\Delta f$.
- c. Building on the formula derived for $\tilde{y}(t)$ obtained in part b, explain how you would mitigate the misalignment distortion in the receiver.

Nonlinear Modulations

- 2.27 In analog communications, amplitude modulation is defined by

$$s_{\text{AM}}(t) = A_c [1 + k_a m(t)] \cos(2\pi f_c t)$$

where $A_c \cos(2\pi f_c t)$ is the carrier, $m(t)$ is the message signal, and k_a is a constant called *amplitude sensitivity* of the modulator. Assume that $|k_a m(t)| < 1$ for all time t .

- a. Justify the statement that, in a strict sense, $s_{\text{AM}}(t)$ violates the principle of superposition.
- b. Formulate the complex envelope $\tilde{s}_{\text{AM}}(t)$ and its spectrum.
- c. Compare the result obtained in part b with the complex envelope of DSB-SC. Hence, comment on the advantages and disadvantages of amplitude modulation.

- 2.28 Continuing on with analog communications, *frequency modulation* (FM) is defined by

$$s_{\text{FM}}(t) = A_c \left[\cos(2\pi f_c t) + k_f \int_0^t m(\tau) d\tau \right]$$

where $A_c \cos(2\pi f_c t)$ is the carrier, $m(t)$ is the message signal, and k_f is a constant called the *frequency sensitivity* of the modulator.

- a. Show that frequency modulation is nonlinear in that it violates the principle of superposition.
- b. Formulate the complex envelope of the FM signal, namely $\tilde{s}_{\text{FM}}(t)$.
- c. Consider the message signal to be in the form of a square wave as shown in Figure P2.28. The modulation frequencies used for the positive and negative amplitudes of the square wave, namely f_1 and f_2 , are defined as follows:

$$f_1 + f_2 = \frac{2}{T_b}$$

$$f_1 - f_2 = \frac{1}{T_b}$$

where T_b is the duration of each positive or negative amplitude in the square wave. Show that under these conditions the complex envelope $\tilde{s}_{\text{FM}}(t)$ maintains *continuity* for all time t , including the switching times between positive and negative amplitudes.

- d. Plot the real and imaginary parts of $\tilde{s}_{\text{FM}}(t)$ for the following values:

$$T_b = \frac{1}{3} \text{ s}$$

$$f_1 = 4\frac{1}{2} \text{ Hz}$$

$$f_2 = 1\frac{1}{2} \text{ Hz}$$

Phase and Group Delays

- 2.29 The phase response of a band-pass communication channel is defined by.

$$\phi(f) = -\tan^{-1} \left(\frac{f^2 - f_c^2}{ff_c} \right)$$

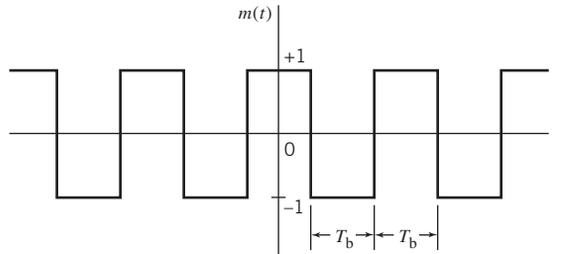


Figure P2.28

A sinusoidally modulated signal defined by

$$s(t) = A_c \cos(2\pi f_m t) \cos(2\pi f_c t)$$

is transmitted through the channel; f_c is the carrier frequency and f_m is the modulation frequency.

- Determine the phase delay τ_p .
- Determine the group delay τ_g .
- Display the waveform produced at the channel output; hence, comment on the results obtained in parts a and b.

Notes

- For a proof of convergence of the Fourier series, see Kammler (2000).
- If a time function $g(t)$ is such that the value of the energy $\int_{-\infty}^{\infty} |g(t)|^2 dt$ is defined and finite, then the Fourier transform $G(f)$ of the function $g(t)$ exists and

$$\lim_{A \rightarrow \infty} \left[\int_{-\infty}^{\infty} \left| g(t) - \int_{-A}^A G(f) \exp(j2\pi ft) df \right|^2 dt \right] = 0$$

This result is known as *Plancherel's theorem*. For a proof of this theorem, see Titchmarsh (1950).

- The notation $\delta(t)$ for a delta function was first introduced into quantum mechanics by Dirac. This notation is now in general use in the signal processing literature. For detailed discussions of the delta function, see Bracewell (1986).

In a rigorous sense, the Dirac delta function is a distribution, not a function; for a rigorous treatment of the subject, see the book by Lighthill (1958).

- The Paley–Wiener criterion is named in honor of the authors of the paper by Paley and Wiener (1934).
- The integral in (2.54), defining the Hilbert transform of a signal, is an *improper* integral in that the integrand has a singularity at $\tau = t$. To avoid this singularity, the integration must be carried out in a symmetrical manner about the point $\tau = t$. For this purpose, we use the definition

$$\text{P} \int_{-\infty}^{\infty} \frac{g(\tau)}{t - \tau} d\tau = \lim_{\epsilon \rightarrow 0} \left[\int_{-\infty}^{t-\epsilon} \frac{g(\tau)}{t - \tau} d\tau + \int_{t+\epsilon}^{\infty} \frac{g(\tau)}{t - \tau} d\tau \right]$$

where the symbol P denotes Cauchy's principal value of the integral and ϵ is incrementally small. For notational simplicity, the symbol P has been omitted from (2.54) and (2.55).

- The complex representation of an arbitrary signal defined in (2.58) was first described by Gabor (1946). Gabor used the term “analytic signal.” The term “pre-envelope” was used in Arens (1957) and Dugundji (1958). For a review of the different envelopes, see the paper by Rice (1982).

- The FFT is *ubiquitous* in that it is applicable to a great variety of unrelated fields. For a detailed mathematical treatment of this widely used tool and its applications, the reader is referred to Brigham (1988).

Probability Theory and Bayesian Inference

3.1 Introduction

The idea of a *mathematical model* used to describe a physical phenomenon is well established in the physical sciences and engineering. In this context, we may distinguish two classes of mathematical models: deterministic and probabilistic. A model is said to be *deterministic* if there is no uncertainty about its time-dependent behavior at any instant of time; linear time-invariant systems considered in Chapter 2 are examples of a deterministic model. However, in many real-world problems, the use of a deterministic model is inappropriate because the underlying physical phenomenon involves too many unknown factors. In such situations, we resort to a *probabilistic model* that accounts for uncertainty in mathematical terms.

Probabilistic models are needed for the design of systems that are reliable in performance in the face of uncertainty, efficient in computational terms, and cost effective in building them. Consider for example, a digital communication system that is required to provide practically error-free communication across a wireless channel. Unfortunately, the wireless channel is subject to *uncertainties*, the sources of which include:

- *noise*, internally generated due to thermal agitation of electrons in the conductors and electronic devices at the front-end of the receiver;
- *fading* of the channel, due to the multipath phenomenon—an inherent characteristic of wireless channels;
- *interference*, representing spurious electromagnetic waves emitted by other communication systems or microwave devices operating in the vicinity of the receiver.

To account for these uncertainties in the design of a wireless communication system, we need a probabilistic model of the wireless channel.

The objective of this chapter, devoted to probability theory, is twofold:

- the formulation of a logical basis for the mathematical description of probabilistic models and
- the development of probabilistic reasoning procedures for handling uncertainty.

Since the probabilistic models are intended to assign probabilities to the collections (sets) of possible outcomes of random experiments, we begin the study of probability theory with a review of set theory, which we do next.

3.2 Set Theory

Definitions

The objects constituting a set are called the *elements* of the set. Let A be a set and x be an element of the set A . To describe this statement, we write $x \in A$; otherwise, we write $x \notin A$. If the set A is empty (i.e., it has no elements), we denote it by \emptyset .

If x_1, x_2, \dots, x_N are all elements of the set A , we write

$$A = \{x_1, x_2, \dots, x_N\}$$

in which case we say that the set A is countably finite. Otherwise, the set is said to be countably infinite. Consider, for example, an *experiment* involving the throws of a die. In this experiment, there are six possible outcomes: the showing of one, two, three, four, five, and six dots on the upper surface of the die; the set of possible *outcomes* of the experiment is therefore countably finite. On the other hand, the set of all possible odd integers, written as $\{\pm 1, \pm 3, \pm 5, \dots\}$, is countably infinite.

If every element of the set A is also an element of another set B , we say that A is a *subset* of B , which we describe by writing $A \subset B$.

If two sets A and B satisfy the conditions $A \subset B$ and $B \subset A$, then the two sets are said to be *identical* or *equal*, in which case we write $A = B$.

In a discussion of set theory, we also find it expedient to think of a *universal set*, denoted by S . Such a set contains every possible element that could occur in the context of a random experiment.

Boolean Operations on Sets

To illustrate the validity of Boolean operations on sets, the use of *Venn diagrams* can be helpful, as shown in what follows.

Unions and Intersections

The *union* of two sets A and B is defined by the set of elements that belong to A or B , or to both. This operation, written as $A \cup B$, is illustrated in the Venn diagram of Figure 3.1.

The *intersection* of two sets A and B is defined by the particular set of elements that belong to both A and B , for which we write $A \cap B$. The shaded part of the Venn diagram in Figure 3.1 represents this second operation.

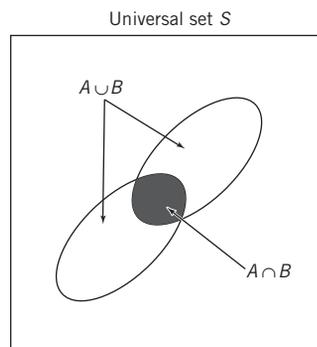
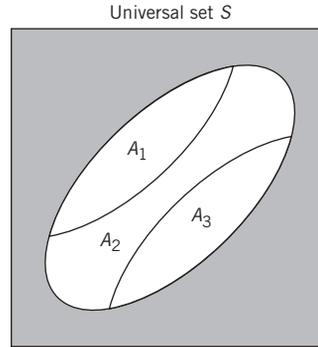


Figure 3.1

Illustrating the union and intersection of two sets, A and B .

Figure 3.2

Illustrating the partition of set A into three subsets: A_1 , A_2 , and A_3 .



Let x be an element of interest. Mathematically, the operations of union and intersection are respectively described by

$$A \cup B = \{x | x \in A \text{ or } x \in B\}$$

and

$$A \cap B = \{x | x \in A \text{ and } x \in B\}$$

where the symbol $|$ is shorthand for “such that.”

Disjoint and Partition Sets

Two sets A and B are said to be *disjoint* if their intersection is empty; that is, they have *no* common elements.

The partition of a set A refers to a collection of disjoint subsets A_1, A_2, \dots, A_N of the set A , the union of which equals A ; that is,

$$A = A_1 \cup A_2 \dots \cup A_N$$

The Venn diagram illustrating the partition operation is depicted in Figure 3.2 for the example of $N = 3$.

Complements

The set A^c is said to be the *complement* of the set A , with respect to the universal set S , if it is made up of all the elements of S that do not belong to A , as depicted in Figure 3.3.

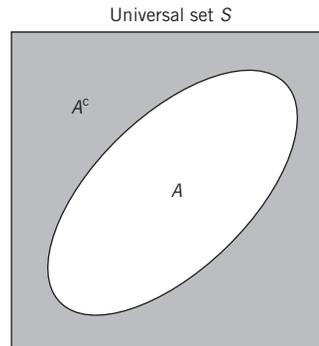


Figure 3.3

Illustrating the complement A^c of set A .

The Algebra of Sets

Boolean operations on sets have several properties, summarized here:

1. *Idempotence property*

$$(A^c)^c = A$$

2. *Commutative property*

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

3. *Associative property*

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

4. *Distributive property*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Note that the commutative and associative properties apply to both the union and intersection, whereas the distributive property applies only to the intersection.

5. *De Morgan's laws*

The complement of the union of two sets A and B is equal to the intersection of their respective complements; that is

$$(A \cup B)^c = A^c \cap B^c$$

The complement of the intersection of two sets A and B is equal to the union of their respective complements; that is,

$$(A \cap B)^c = A^c \cup B^c$$

For illustrations of these five properties and their confirmation, the reader is referred to Problem 3.1.

3.3 Probability Theory

Probabilistic Models

The mathematical description of an experiment with uncertain outcomes is called a *probabilistic model*,¹ the formulation of which rests on three fundamental ingredients:

1. *Sample space* or *universal set* S , which is the set of all conceivable outcomes of a random experiment under study.
2. A *class* E of events that are subsets of S .
3. *Probability law*, according to which a nonnegative measure or number $\mathbb{P}[A]$ is assigned to an *event* A . The measure $\mathbb{P}[A]$ is called the *probability of event* A . In a sense, $\mathbb{P}[A]$ encodes our *belief* in the likelihood of event A occurring when the experiment is conducted.

Throughout the book, we will use the symbol $\mathbb{P}[\cdot]$ to denote the probability of occurrence of the event that appears inside the square brackets.

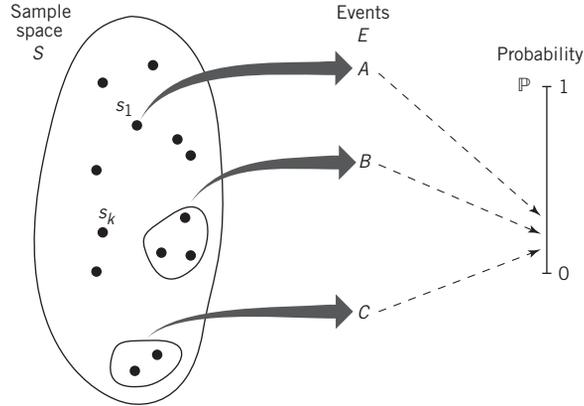


Figure 3.4 Illustration of the relationship between sample space, events, and probability

As illustrated in Figure 3.4, an event may involve a single outcome or a subset of possible outcomes in the sample space S . These possibilities are exemplified by the way in which three events, A , B , and C , are pictured in Figure 3.4. In light of such a reality, we identify two extreme cases:

- *Sure event*, which embodies all the possible outcomes in the sample space S .
- *Null or impossible event*, which corresponds to the empty set or empty space \emptyset .

Axioms of Probability

Fundamentally, the probability measure $\mathbb{P}[A]$, assigned to event A in the class E , is governed by three axioms:

Axiom I Nonnegativity The first axiom states that *the probability of event A is a nonnegative number bounded by unity*, as shown by

$$0 \leq \mathbb{P}[A] \leq 1 \quad \text{for any event } A \quad (3.1)$$

Axiom II Additivity The second axiom states that *if A and B are two disjoint events, then the probability of their union satisfies the equality*

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] \quad (3.2)$$

In general, if the sample space has N elements and A_1, A_2, \dots, A_N is a sequence of disjoint events, then the probability of the union of these N events satisfies the equality

$$\mathbb{P}[A_1 \cup A_2 \cup \dots \cup A_N] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \dots + \mathbb{P}[A_N]$$

Axiom III Normalization The third and final axiom states that *the probability of the entire sample space S is equal to unity*, as shown by

$$\mathbb{P}[S] = 1 \quad (3.3)$$

These three axioms provide an implicit definition of probability. Indeed, we may use them to develop some other basic properties of probability, as described next.

PROPERTY 1 *The probability of an impossible event is zero.*

To prove this property, we first use the axiom of normalization, then express the sample space S as the union of itself with the empty space \emptyset , and then use the axiom of additivity. We thus write

$$\begin{aligned} 1 &= \mathbb{P}[S] \\ &= \mathbb{P}[S \cup \emptyset] \\ &= \mathbb{P}[S] + \mathbb{P}[\emptyset] \\ &= 1 + \mathbb{P}[\emptyset] \end{aligned}$$

from which the property $\mathbb{P}[\emptyset] = 0$ follows immediately.

PROPERTY 2 *Let A^c denote the complement of event A ; we may then write*

$$\mathbb{P}[A^c] = 1 - \mathbb{P}[A] \quad \text{for any event } A \quad (3.4)$$

To prove this property, we first note that the sample space S is the union of the two mutually exclusive events A and A^c . Hence, the use of the additivity and normalization axioms yields

$$\begin{aligned} 1 &= \mathbb{P}[S] \\ &= \mathbb{P}[A \cup A^c] \\ &= \mathbb{P}[A] + \mathbb{P}[A^c] \end{aligned}$$

from which, after rearranging terms, (3.4) follows immediately.

PROPERTY 3 *If event A lies within the subspace of another event B , then*

$$\mathbb{P}[A] \leq \mathbb{P}[B] \quad \text{for } A \subset B \quad (3.5)$$

To prove this third property, consider the *Venn diagram* depicted in Figure 3.5. From this diagram, we observe that event B may be expressed as the union of two disjoint events, one defined by A and the other defined by the intersection of B with the complement of A ; that is,

$$B = A \cup (B \cap A^c)$$

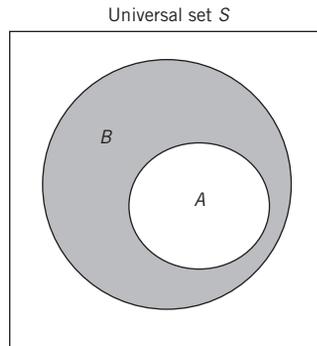


Figure 3.5
The Venn diagram for proving (3.5).

Therefore, applying the additivity axiom to this relation, we get

$$\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}[B \cap A^c]$$

Next, invoking the nonnegativity axiom, we immediately find that the probability of event B must be equal to or greater than the probability of event A , as indicated in (3.5).

PROPERTY 4 *Let N disjoint events A_1, A_2, \dots, A_N satisfy the condition*

$$A_1 \cup A_2 \cup \dots \cup A_N = S \quad (3.6)$$

then

$$\mathbb{P}[A_1] + \mathbb{P}[A_2] + \dots + \mathbb{P}[A_N] = 1 \quad (3.7)$$

To prove this fourth property, we first apply the normalization axiom to (3.6) to write

$$\mathbb{P}[A_1 \cup A_2 \cup \dots \cup A_N] = 1$$

Next, recalling the generalized form of the additivity axiom

$$\mathbb{P}[A_1 \cup A_2 \cup \dots \cup A_N] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \dots + \mathbb{P}[A_N]$$

From these two relations, (3.7) follows immediately.

For the special case of N equally probable events, (3.7) reduces to

$$\mathbb{P}[A_i] = \frac{1}{N} \quad \text{for } i = 1, 2, \dots, N \quad (3.8)$$

PROPERTY 5 *If two events A and B are not disjoint, then the probability of their union event is defined by*

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B] \quad \text{for any two events } A \text{ and } B \quad (3.9)$$

where $\mathbb{P}[A \cap B]$ is called the joint probability of A and B .

To prove this last property, consider the Venn diagram of Figure 3.6. From this figure, we first observe that the union of A and B may be expressed as the union of two disjoint events: A itself and $A^c \cap B$, where A^c is the complement of A . We may therefore apply the additivity axiom to write

$$\begin{aligned} \mathbb{P}[A \cup B] &= \mathbb{P}[A \cup (A^c \cap B)] \\ &= \mathbb{P}[A] + \mathbb{P}[A^c \cap B] \end{aligned} \quad (3.10)$$

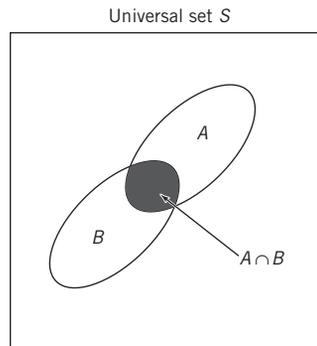


Figure 3.6
The Venn diagram for proving (3.9).

From the Venn diagram of Figure 3.6, we next observe that the event B may be expressed as

$$\begin{aligned} B &= S \cap B \\ &= (A \cup A^c) \cap B \\ &= (A \cap B) \cup (A^c \cap B) \end{aligned}$$

That is, B is the union of two disjoint events: $A \cap B$ and $A^c \cap B$; therefore, applying the additivity axiom to this second relation yields

$$\mathbb{P}[B] = \mathbb{P}[A \cap B] + \mathbb{P}[A^c \cap B] \quad (3.11)$$

Subtracting (3.11) from (3.10), canceling the common term $\mathbb{P}[A^c \cap B]$ and rearranging terms, (3.9) follows and Property 4 is proved.

It is of interest to note that the joint probability $\mathbb{P}[A \cap B]$ accounts for that part of the sample space S where the events A and B coincide. If these two events are disjoint, then the joint probability $\mathbb{P}[A \cap B]$ is zero, in which case (3.9) reduces to the additivity axiom of (3.2).

Conditional Probability

When an experiment is performed and we only obtain *partial information* on the outcome of the experiment, we may *reason* about that particular outcome by invoking the notion of conditional probability. Stated the other way round, we may make the statement:

Conditional probability provides the premise for probabilistic reasoning.

To be specific, suppose we perform an experiment that involves a pair of events A and B . Let $\mathbb{P}[A|B]$ denote the probability of event A given that event B has occurred. The probability $\mathbb{P}[A|B]$ is called the *conditional probability of A given B* . Assuming that B has nonzero probability, the conditional probability $\mathbb{P}[A|B]$ is formally defined by

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \quad (3.12)$$

where $\mathbb{P}[A \cap B]$ is the joint probability of events A and B , and $\mathbb{P}[B]$ is nonzero.

For a fixed event B , the conditional probability $\mathbb{P}[A|B]$ is a legitimate probability law as it satisfies all three axioms of probability:

1. Since by definition, $\mathbb{P}[A|B]$ is a probability, the nonnegativity axiom is clearly satisfied.
2. Viewing the entire sample space S as event A and noting that $S \cup B = B$, we may use (3.12) to write

$$\mathbb{P}[S|B] = \frac{\mathbb{P}[S \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B]}{\mathbb{P}[B]} = 1$$

Hence, the normalization axiom is also satisfied.

3. Finally, to verify the additivity axiom, assume that A_1 and A_2 are two mutually exclusive events. We may then use (3.12) to write

$$\mathbb{P}[A_1 \cup A_2|B] = \frac{\mathbb{P}[(A_1 \cup A_2) \cap B]}{\mathbb{P}[B]}$$

Applying the distributive property to the numerator on the right-hand side, we have

$$\mathbb{P}[A_1 \cup A_2 | B] = \frac{\mathbb{P}[(A_1 \cup B) \cup (A_2 \cap B)]}{\mathbb{P}[B]}$$

Next, recognizing that the two events $A_1 \cap B$ and $A_2 \cap B$ are actually disjoint, we may apply the additivity axiom to write

$$\begin{aligned} \mathbb{P}[A_1 \cup A_2 | B] &= \frac{\mathbb{P}[A_1 \cap B] + \mathbb{P}[A_2 \cap B]}{\mathbb{P}[B]} \\ &= \frac{\mathbb{P}[A_1 \cap B]}{\mathbb{P}[B]} + \frac{\mathbb{P}[A_2 \cap B]}{\mathbb{P}[B]} \end{aligned} \tag{3.13}$$

which proves that the conditional probability also satisfies the additivity axiom.

We therefore conclude that all three axioms of probability (and therefore all known properties of probability laws) are equally valid for the conditional probability $\mathbb{P}[A|B]$. In a sense, this conditional probability captures the *partial information that the occurrence of event B provides about event A*; we may therefore view the conditional probability $\mathbb{P}[A|B]$ as a probability law concentrated on event B .

Bayes' Rule

Suppose we are confronted with a situation where the conditional probability $\mathbb{P}[A|B]$ and the individual probabilities $\mathbb{P}[A]$ and $\mathbb{P}[B]$ are all easily determined directly, but the conditional probability $\mathbb{P}[B|A]$ is desired. To deal with this situation, we first rewrite (3.12) in the form

$$\mathbb{P}[A \cap B] = \mathbb{P}[A|B]\mathbb{P}[B]$$

Clearly, we may equally write

$$\mathbb{P}[A \cap B] = \mathbb{P}[B|A]\mathbb{P}[A]$$

The left-hand parts of these two relations are identical; we therefore have

$$\mathbb{P}[A|B]\mathbb{P}[B] = \mathbb{P}[B|A]\mathbb{P}[A]$$

Provided that $\mathbb{P}[A]$ is nonzero, we may determine the desired conditional probability $\mathbb{P}[B|A]$ by using the relation

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\mathbb{P}[B]}{\mathbb{P}[A]} \tag{3.14}$$

This relation is known as *Bayes' rule*.

As simple as it looks, Bayes' rule provides the correct language for describing *inference*, the formulation of which cannot be done without making assumptions.² The following example illustrates an application of Bayes' rule.

EXAMPLE 1

Radar Detection

Radar, a remote sensing system, operates by transmitting a sequence of pulses and has its receiver listen to echoes produced by a target (e.g., aircraft) that could be present in its surveillance area.

Let the events A and B be defined as follows:

$A = \{\text{a target is present in the area under surveillance}\}$

$A^c = \{\text{there is no target in the area}\}$

$B = \{\text{the radar receiver detects a target}\}$

In the radar detection problem, there are three probabilities of particular interest:

$\mathbb{P}[A]$ probability that a target is present in the area; this probability is called the *prior probability*.

$\mathbb{P}[B|A]$ probability that the radar receiver detects a target, given that a target is actually present in the area; this second probability is called the *probability of detection*.

$\mathbb{P}[B|A^c]$ probability that the radar receiver detects a target in the area, given that there is no target in the surveillance area; this third probability is called the *probability of false alarm*.

Suppose these three probabilities have the following values:

$$\mathbb{P}[A] = 0.02$$

$$\mathbb{P}[B|A] = 0.99$$

$$\mathbb{P}[B|A^c] = 0.01$$

The problem is to calculate the conditional probability $\mathbb{P}[A|B]$ which defines the probability that a target is present in the surveillance area given that the radar receiver has made a target detection.

Applying Bayes' rule, we write

$$\begin{aligned} \mathbb{P}[A|B] &= \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]} \\ &= \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|A^c]\mathbb{P}[A^c]} \\ &= \frac{0.99 \times 0.02}{0.99 \times 0.02 + 0.01 \times 0.98} \\ &= \frac{0.0198}{0.0296} \\ &\approx 0.69 \end{aligned}$$

Independence

Suppose that the occurrence of event A provides no information whatsoever about event B ; that is,

$$\mathbb{P}[B|A] = \mathbb{P}[B]$$

Then, (3.14) also teaches us that

$$\mathbb{P}[A|B] = \mathbb{P}[A]$$

In this special case, we see that knowledge of the occurrence of either event, A or B , tells us no more about the probability of occurrence of the other event than we knew without that knowledge. Events A and B that satisfy this condition are said to be *independent*.

From the definition of conditional probability given in (3.12), namely,

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

we see that the condition $\mathbb{P}[A|B] = \mathbb{P}[A]$ is equivalent to

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$$

We therefore adopt this latter relation as the formal definition of independence. The important point to note here is that the definition still holds even if the probability $\mathbb{P}[B]$ is zero, in which case the conditional probability $\mathbb{P}[A|B]$ is undefined. Moreover, the definition has a *symmetric* property, in light of which we can say the following:

If an event A is independent of another event B , then B is independent of A , and A and B are therefore independent events.

3.4 Random Variables

It is customary, particularly when using the language of sample space pertaining to an experiment, to describe the outcome of the experiment by using one or more real-valued quantities or measurements that help us think in probabilistic terms. These quantities are called *random variables*, for which we offer the following definition:

The random variable is a function whose domain is a sample space and whose range is some set of real numbers.

The following two examples illustrate the notion of a random variable embodied in this definition.

Consider, for example, the sample space that represents the integers 1, 2, ..., 6, each one of which is the number of dots that shows uppermost when a die is thrown. Let the sample point k denote the event that k dots show in one throw of the die. The random variable used to describe the probabilistic event k in this experiment is said to be a *discrete random variable*.

For an entirely different experiment, consider the noise being observed at the front end of a communication receiver. In this new situation, the random variable, representing the amplitude of the noise voltage at a particular instant of time, occupies a continuous range of values, both positive and negative. Accordingly, the random variable representing the noise amplitude is said to be a *continuous random variable*.

The concept of a continuous random variable is illustrated in Figure 3.7, which is a modified version of Figure 3.4. Specifically, for the sake of clarity, we have suppressed the events but show subsets of the sample space S being mapped directly to a subset of a real line representing the random variable. The notion of the random variable depicted in Figure 3.7 applies in exactly the same manner as it applies to the underlying events. The benefit of random variables, pictured in Figure 3.7, is that probability analysis can now be developed in terms of real-valued quantities, regardless of the form or shape of the underlying events of the random experiment under study.

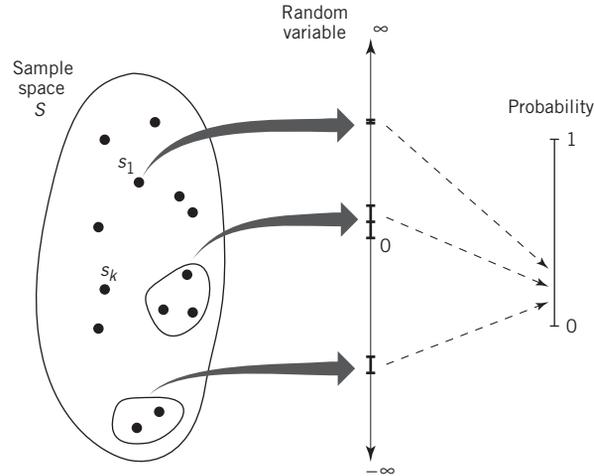


Figure 3.7 Illustration of the relationship between sample space, random variables, and probability.

One last comment is in order before we proceed further. Throughout the whole book, we will be using the following notation:

Uppercase characters denote random variables and lowercase characters denote real values taken by random variables.

3.5 Distribution Functions

To proceed with the probability analysis in mathematical terms, we need a probabilistic description of random variables that works equally well for discrete and continuous random variables. Let us consider the random variable X and the probability of the event $X \leq x$. We denote this probability by $\mathbb{P}[X \leq x]$. It is apparent that this probability is a function of the *dummy variable* x . To simplify the notation, we write

$$F_X(x) = \mathbb{P}[X \leq x] \quad \text{for all } x \quad (3.15)$$

The function $F_X(x)$ is called the *cumulative distribution function* or simply the *distribution function* of the random variable X . Note that $F_X(x)$ is a function of x , not of the random variable X . For any point x in the sample space, the distribution function $F_X(x)$ expresses the probability of an event.

The distribution function $F_X(x)$, applicable to both continuous and discrete random variables, has two fundamental properties:

PROPERTY 1 Boundedness of the Distribution

The distribution function $F_X(x)$ is a bounded function of the dummy variable x that lies between zero and one.

Specifically, $F_X(x)$ tends to zero as x tends to $-\infty$, and it tends to one as x tends to ∞ .

PROPERTY 2 Monotonicity of the Distribution

The distribution function $F_X(x)$ is a monotone nondecreasing function of x .

In mathematical terms, we write

$$F_X(x_1) \leq F_X(x_2) \quad \text{for } x_1 < x_2$$

Both of these properties follow directly from (3.15).

The random variable X is said to be *continuous* if the distribution function $F_X(x)$ is differentiable with respect to the dummy variable x everywhere, as shown by

$$f_X(x) = \frac{d}{dx}F_X(x) \quad \text{for all } x \quad (3.16)$$

The new function $f_X(x)$ is called the *probability density function* of the random variable X . The name, density function, arises from the fact that the probability of the event $x_1 < X \leq x_2$ is

$$\begin{aligned} \mathbb{P}[x_1 < X \leq x_2] &= \mathbb{P}[X \leq x_2] - \mathbb{P}[X \leq x_1] \\ &= F_X(x_2) - F_X(x_1) \\ &= \int_{x_1}^{x_2} f_X(x) dx \end{aligned} \quad (3.17)$$

The probability of an interval is therefore the area under the probability density function in that interval. Putting $x_1 = -\infty$ in (3.17) and changing the notation somewhat, we readily see that the distribution function is defined in terms of the probability density function as

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi \quad (3.18)$$

where ξ is a dummy variable. Since $F_X(\infty) = 1$, corresponding to the probability of a sure event, and $F_X(-\infty) = 0$, corresponding to the probability of an impossible event, we readily find from (3.17) that

$$\int_{-\infty}^{\infty} f_X(x) dx = 1 \quad (3.19)$$

Earlier we mentioned that a distribution function must always be a monotone nondecreasing function of its argument. It follows, therefore, that the probability density function must always be nonnegative. Accordingly, we may now formally make the statement:

The probability density function $f_X(x)$ of a continuous random variable X has two defining properties: nonnegativity and normalization.

PROPERTY 3 Nonnegativity

The probability density function $f_X(x)$ is a nonnegative function of the sample value x of the random variable X .

PROPERTY 4 Normalization

The total area under the graph of the probability density function $f_X(x)$ is equal to unity.

An important point that should be stressed here is that the probability density function $f_X(x)$ contains all the conceivable information needed for statistical characterization of the random variable X .

EXAMPLE 2 Uniform Distribution

To illustrate the properties of the distribution function $F_X(x)$ and the probability density function $f_X(x)$ for a continuous random variable, consider a *uniformly distributed random variable*, described by

$$f_X(x) = \begin{cases} 0, & x \leq a \\ \frac{1}{b-a}, & a < x \leq b \\ 0, & x > b \end{cases} \quad (3.20)$$

Integrating $f_X(x)$ with respect to x yields the associated distribution function

$$F_X(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & x > b \end{cases} \quad (3.21)$$

Plots of these two functions versus the dummy variable x are shown in Figure 3.8.

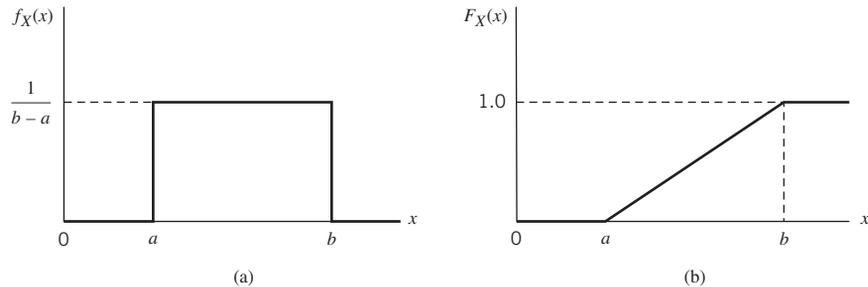


Figure 3.8 Uniform distribution.

Probability Mass Function

Consider next the case of a *discrete* random variable, X , which is a real-valued function of the outcome of a probabilistic experiment that can take a finite or countably infinite number of values. As mentioned previously, the distribution function $F_X(x)$ defined in (3.15) also applies to discrete random variables. However, unlike a continuous random variable, the distribution function of a discrete random variable is not differentiable with respect to its dummy variable x .

To get around this mathematical difficulty, we introduce the notion of the *probability mass function* as another way of characterizing discrete random variables. Let X denote a discrete random variable and let x be any possible value of X taken from a set of real numbers. We may then make the statement:

The probability mass function of x , denoted by $p_X(x)$, is defined as the probability of the event $X = x$, which consists of all possible outcomes of an experiment that lead to a value of X equal to x .

Stated in mathematical terms, we write

$$p_X(x) = \mathbb{P}[X = x] \quad (3.22)$$

which is illustrated in the next example.

EXAMPLE 3 The Bernoulli Random Variable

Consider a probabilistic experiment involving the discrete random variable X that takes one of two possible values:

- the value 1 with probability p ;
- the value 0 with probability $1 - p$.

Such a random variable is called the *Bernoulli random variable*, the probability mass function of which is defined by

$$p_X(x) = \begin{cases} 1 - p & x = 0 \\ p, & x = 1 \\ 0, & \text{otherwise} \end{cases} \quad (3.23)$$

This probability mass function is illustrated in Figure 3.9. The two delta functions, each of weight $1/2$, depicted in Figure 3.9 represent the probability mass function at each of the sample points $x = 0$ and $x = 1$.

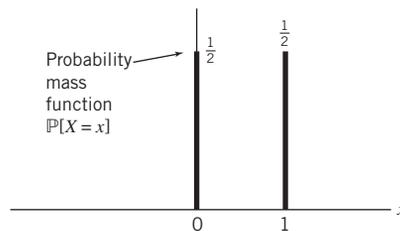


Figure 3.9 Illustrating the probability mass function for a fair coin-tossing experiment.

From here on, we will, largely but not exclusively, focus on the characterization of continuous random variables. A parallel development and similar concepts are possible for discrete random variables as well.³

Multiple Random Variables

Thus far we have focused attention on situations involving a single random variable. However, we frequently find that the outcome of an experiment requires several random variables for its description. In what follows, we consider situations involving two random variables. The probabilistic description developed in this way may be readily extended to any number of random variables.

Consider two random variables X and Y . In this new situation, we say:

The joint distribution function $F_{X,Y}(x,y)$ is the probability that the random variable X is less than or equal to a specified value x , and that the random variable Y is less than or equal to another specified value y .

The variables X and Y may be two separate one-dimensional random variables or the components of a single two-dimensional random vector. In either case, the joint sample space is the xy -plane. The *joint distribution function* $F_{X,Y}(x,y)$ is the probability that the outcome of an experiment will result in a sample point lying inside the quadrant $(-\infty < X \leq x, -\infty < Y \leq y)$ of the joint sample space. That is,

$$F_{X,Y}(x,y) = \mathbb{P}[X \leq x, Y \leq y] \quad (3.24)$$

Suppose that the joint distribution function $F_{X,Y}(x,y)$ is continuous everywhere and that the second-order partial derivative

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \quad (3.25)$$

exists and is continuous everywhere too. We call the new function $f_{X,Y}(x,y)$ the *joint probability density function* of the random variables X and Y . The joint distribution function $F_{X,Y}(x,y)$ is a monotone nondecreasing function of both x and y . Therefore, from (3.25) it follows that the joint probability density function $f_{X,Y}(x,y)$ is always nonnegative. Also, the total volume under the graph of a joint probability density function must be unity, as shown by the double integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \, dx \, dy = 1 \quad (3.26)$$

The so-called *marginal* probability density functions, $f_X(x)$ and $f_Y(y)$, are obtained by differentiating the corresponding marginal distribution functions

$$F_X(x) = f_{X,Y}(x, \infty)$$

and

$$F_Y(y) = f_{X,Y}(\infty, y)$$

with respect to the dummy variables x and y , respectively. We thus write

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \frac{d}{dx} \int_{-\infty}^x \left[\int_{-\infty}^{\infty} f_{X,Y}(\xi, y) \, dy \right] d\xi \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy \end{aligned} \quad (3.27)$$

Similarly, we write

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx \quad (3.28)$$

In words, the first marginal probability density function $f_X(x)$, defined in (3.27), is obtained from the joint probability density function $f_{X,Y}(x,y)$ by simply integrating it over all possible values of the undesired random variable Y . Similarly, the second marginal probability density function $f_Y(y)$, defined in (3.28), is obtained from $f_{X,Y}(x,y)$ by integrating it over all possible values of the undesired random variable; this time, the undesirable random variable is X . Henceforth, we refer to $f_X(x)$ and $f_Y(y)$, obtained in the manner described herein, as the *marginal densities* of the random variables X and Y , whose joint probability density function is $f_{X,Y}(x,y)$. Here again, we conclude the discussion on a pair of random variables with the following statement:

The joint probability density function $f_{X,Y}(x,y)$ contains all the conceivable information on the two continuous random variables X and Y that is needed for the probability analysis of joint random variables.

This statement can be generalized to cover the joint probability density function of many random variables.

Conditional Probability Density Function

Suppose that X and Y are two continuous random variables with their joint probability density function $f_{X,Y}(x,y)$. The *conditional probability density function* of Y , such that $X = x$, is defined by

$$f_Y(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \quad (3.29)$$

provided that $f_X(x) > 0$, where $f_X(x)$ is the marginal density of X ; $f_Y(y|x)$ is a shortened version of $f_{Y|X}(y|x)$, both of which are used interchangeably. The function $f_Y(y|x)$ may be thought of as a function of the variable Y , with the variable x arbitrary but fixed; accordingly, it satisfies all the requirements of an ordinary probability density function for any x , as shown by

$$f_Y(y|x) \geq 0$$

and

$$\int_{-\infty}^{\infty} f_Y(y|x) dy = 1 \quad (3.30)$$

Cross-multiplying terms in (3.29) yields

$$f_{X,Y}(x,y) = f_Y(y|x)f_X(x)$$

which is referred to as the *multiplication rule*.

Suppose that knowledge of the outcome of X can, in no way, affect the distribution of Y . Then, the conditional probability density function $f_Y(y|x)$ reduces to the marginal density $f_Y(y)$, as shown by

$$f_Y(y|x) = f_Y(y)$$

In such a case, we may express the joint probability density function of the random variables X and Y as the product of their respective marginal densities; that is,

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

On the basis of this relation, we may now make the following statement on the *independence* of random variables:

If the joint probability density function of the random variables X and Y equals the product of their marginal densities, then X and Y are statistically independent.

Sum of Independent Random Variables: Convolution

Let X and Y be two continuous random variables that are statistically independent; their respective probability density functions are denoted by $f_X(x)$ and $f_Y(y)$. Define the sum

$$Z = X + Y$$

The issue of interest is to find the probability density function of the new random variable Z , which is denoted by $f_Z(z)$.

To proceed with this evaluation, we first use probabilistic arguments to write

$$\begin{aligned}\mathbb{P}[Z \leq z | X = x] &= \mathbb{P}[X + Y \leq z | X = x] \\ &= \mathbb{P}[x + Y \leq z | X = x]\end{aligned}$$

where, in the second line, the given value x is used for the random variable X . Since X and Y are statistically independent, we may simplify matters by writing

$$\begin{aligned}\mathbb{P}[Z \leq z | X = x] &= \mathbb{P}[x + Y \leq z] \\ &= \mathbb{P}[Y \leq z - x]\end{aligned}$$

Equivalently, in terms of the pertinent distribution functions, we may write

$$F_Z(z|x) = F_Y(z - x)$$

Hence, differentiating both sides of this equation, we get the corresponding probability density functions

$$f_z(z|x) = f_Y(z - x)$$

Using the multiplication rule described in (3.30), we have

$$f_{Z,X}(z, x) = f_Y(z - x)f_X(x) \tag{3.31}$$

Next, adapting the definition of the marginal density given in (3.27) to the problem at hand, we write

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,X}(z, x) dx \tag{3.32}$$

Finally, substituting (3.31) into (3.32), we find that the desired $f_Z(z)$ is equal to the convolution of $f_X(x)$ and $f_Y(y)$, as shown by

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z - x) dx \tag{3.33}$$

In words, we may therefore state:

The summation of two independent continuous random variables leads to the convolution of their respective probability density functions.

Note, however, that no assumptions were made in arriving at this statement except for the random variables X and Y being continuous random variables.

3.6 The Concept of Expectation

As pointed out earlier, the probability density function $f_X(x)$ provides a complete statistical description of a continuous random variable X . However, in many instances, we find that this description includes more detail than is deemed to be essential for practical applications. In situations of this kind, simple *statistical averages* are usually considered to be adequate for the statistical characterization of the random variable X .

In this section, we focus attention on the *first-order* statistical average, called the expected value or mean of a random variable; *second-order* statistical averages are studied in the next section. The rationale for focusing attention on the mean of a random variable is its practical importance in statistical terms, as explained next.

Mean

The *expected value* or *mean* of a continuous random variable X is formally defined by

$$\mu_X = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx \quad (3.34)$$

where \mathbb{E} denotes the *expectation* or *averaging operator*. According to this definition, the expectation operator \mathbb{E} , applied to a continuous random variable x , produces a single number that is derived uniquely from the probability density function $f_X(x)$.

To describe the meaning of the defining equation (3.34), we may say the following:

The mean μ_X of a random variable X , defined by the expectation $\mathbb{E}[x]$, locates the center of gravity of the area under the probability density curve of the random variable X .

To elaborate on this statement, we write the integral in (3.34) as the limit of an approximating sum formulated as follows. Let $\{x_k | k = 0, \pm 1, \pm 2, \dots\}$ denote a set of uniformly spaced points on the real line

$$x_k = \left(k + \frac{1}{2}\right)\Delta, \quad k = 0, \pm 1, \pm 2, \dots \quad (3.35)$$

where Δ is the spacing between adjacent points on the line. We may thus rewrite (3.34) in the form of a limit as follows:

$$\begin{aligned} \mathbb{E}[X] &= \lim_{\Delta \rightarrow 0} \sum_{k=-\infty}^{\infty} \int_{k\Delta}^{(k+1)\Delta} x_k f_X(x) dx \\ &= \lim_{\Delta \rightarrow 0} \sum_{k=-\infty}^{\infty} x_k \mathbb{P}\left[x_k - \frac{\Delta}{2} < X \leq x_k + \frac{\Delta}{2}\right] \end{aligned}$$

For a physical interpretation of the sum in the second line of the right-hand side of this equation, suppose that we make n independent observations of the random variable X . Let $N_n(k)$ denote the number of times that the random variable X falls inside the k th bin, defined by

$$x_k - \frac{\Delta}{2} < X \leq x_k + \frac{\Delta}{2}, \quad k = 0, \pm 1, \pm 2, \dots$$

Arguing heuristically, we may say that, as the number of observations n is made large, the ratio $N_n(k)/n$ approaches the probability $\mathbb{P}[x_k - \Delta/2 < X \leq x_k + \Delta/2]$. Accordingly, we may approximate the expected value of the random variable X as

$$\begin{aligned} \mathbb{E}[X] &\approx \sum_{k=-\infty}^{\infty} x_k \left(\frac{N_n(k)}{n} \right) \\ &= \frac{1}{n} \sum_{k=-\infty}^{\infty} x_k N_n(k), \quad \text{for large } n \end{aligned} \tag{3.36}$$

We now recognize the quantity on the right-hand side of (3.36) simply as the “sample average.” The sum is taken over all the values x_k , each of which is weighted by the number of times it occurs; the sum is then divided by the total number of observations to give the sample average. Indeed, (3.36) provides the basis for computing the expectation $\mathbb{E}[X]$.

In a loose sense, we may say that the *discretization*, introduced in (3.35), has changed the expectation of a continuous random variable to the sample averaging over a discrete random variable. Indeed, in light of (3.36), we may formally define the expectation of a discrete random variable X as

$$\mathbb{E}[X] = \sum_x x p_X(x) \tag{3.37}$$

where $p_X(x)$ is the probability mass function of X , defined in (3.22), and where the summation extends over all possible discrete values of the dummy variable x . Comparing the summation in (3.37) with that of (3.36), we see that, roughly speaking, the ratio $N_n(x)/n$ plays a role similar to that of the probability mass function $p_X(x)$, which is intuitively satisfying.

Just as in the case of a continuous random variable, here again we see from the defining equation (3.37) that the expectation operator \mathbb{E} , applied to a discrete random variable X , produces a single number derived uniquely from the probability mass function $p_X(x)$.

Simply put, the expectation operator \mathbb{E} applies equally well to discrete and continuous random variables.

Properties of the Expectation Operator

The expectation operator \mathbb{E} plays a dominant role in the statistical analysis of random variables (as well as random processes studied in Chapter 4). It is therefore befitting that we study two important properties of this operation in this section; other properties are addressed in the end-of-chapter Problem 3.13.

PROPERTY 1 Linearity

Consider a random variable Z , defined by

$$Z = X + Y$$

where X and Y are two continuous random variables whose probability density functions are respectively denoted by $f_X(x)$ and $f_Y(y)$. Extending the definition of expectation introduced in (3.34) to the random variable Z , we write

$$\mathbb{E}[Z] = \int_{-\infty}^{\infty} z f_Z(z) dz$$

where $f_Z(z)$ is defined by the convolution integral of (3.33). Accordingly, we may go on to express the expectation $\mathbb{E}[Z]$ as the double integral

$$\begin{aligned} \mathbb{E}[Z] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z f_X(x) f_Y(z-x) dx dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z f_{X,Y}(x, z-x) dx dz \end{aligned}$$

where the joint probability density function

$$f_{X,Y}(x, z-x) = f_X(x) f_Y(z-x)$$

Making the one-to-one change of variables

$$y = z - x$$

and

$$x = x$$

we may now express the expectation $\mathbb{E}[Z]$ in the expanded form

$$\begin{aligned} \mathbb{E}[Z] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) f_{X,Y}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x, y) dx dy \end{aligned}$$

Next, we recall from (3.27) that the first marginal density of the random variable X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

and, similarly, for the second marginal density

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

The formula for the expectation $\mathbb{E}[Z]$ is therefore simplified as follows:

$$\begin{aligned} \mathbb{E}[Z] &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= \mathbb{E}[X] + \mathbb{E}[Y] \end{aligned}$$

We may extend this result to the sum of many random variables by *the method of induction* and thus write that, in general,

$$\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] \quad (3.38)$$

In words, we may therefore state:

The expectation of a sum of random variables is equal to the sum of the individual expectations.

This statement proves the linearity property of the *expectation operator*, which makes this operator all the more appealing.

PROPERTY 2 Statistical Independence

Consider next the random variable Z , defined as the product of two independent random variables X and Y , whose probability density functions are respectively denoted by $f_X(x)$ and $f_Y(y)$. As before, the expectation of Z is defined by

$$\mathbb{E}[Z] = \int_{-\infty}^{\infty} z f_Z(z) \, dz$$

except that, this time, we have

$$\begin{aligned} f_Z(z) &= f_{X, Y}(x, y) \\ &= f_X(x) f_Y(y) \end{aligned}$$

where, in the second line, we used the statistical independence of X and Y . With $Z = XY$, we may therefore recast the expectation $\mathbb{E}[Z]$ as

$$\begin{aligned} \mathbb{E}[XY] &= \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) \, dx \, dy \\ &= \int_{-\infty}^{\infty} x f_X(x) \, dx \int_{-\infty}^{\infty} y f_Y(y) \, dy \\ &= \mathbb{E}[X] \mathbb{E}[Y] \end{aligned} \quad (3.39)$$

In words, we may therefore state:

The expectation of the product of two statistically independent random variables is equal to the product of their individual expectations.

Here again, by induction, we may extend this statement to the product of many independent random variables.

3.7 Second-Order Statistical Averages

Function of a Random Variable

In the previous section we studied the mean of random variables in some detail. In this section, we expand on the mean by studying different second-order statistical averages.

These statistical averages, together with the mean, complete the *partial characterization* of random variables.

To this end, let X denote a random variable and let $g(X)$ denote a real-valued function of X defined on the real line. The quantity obtained by letting the argument of the function $g(X)$ be a random variable is also a random variable, which we denote as

$$Y = g(X) \quad (3.40)$$

To find the expectation of the random variable Y , we could, of course, find the probability density function $f_Y(y)$ and then apply the standard formula

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy$$

A simpler procedure, however, is to write

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad (3.41)$$

Equation (3.41) is called the *expected value rule*; validity of this rule for a continuous random variable is addressed in Problem 3.14.

EXAMPLE 4 The Cosine Transformation of a Random Variable

Let

$$Y = g(X) = \cos(X)$$

where X is a random variable uniformly distributed in the interval $(-\pi, \pi)$; that is,

$$f_X(x) = \begin{cases} \frac{1}{2\pi}, & -\pi \leq x \leq \pi \\ 0, & \text{otherwise} \end{cases}$$

According to (3.41), the expected value of Y is

$$\begin{aligned} \mathbb{E}[Y] &= \int_{-\pi}^{\pi} (\cos x) \left(\frac{1}{2\pi} \right) dx \\ &= -\frac{1}{2\pi} \sin x \Big|_{x=-\pi}^{\pi} \\ &= 0 \end{aligned}$$

This result is intuitively satisfying in light of what we know about the dependence of a cosine function on its argument.

Second-order Moments

For the special case of $g(X) = X^n$, the application of (3.41) leads to the *nth moment* of the probability distribution of a random variable X ; that is,

$$\mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n f_X(x) dx \quad (3.42)$$

From an engineering perspective, however, the most important moments of X are the first two moments. Putting $n = 1$ in (3.42) gives the mean of the random variable, which was discussed in Section 3.6. Putting $n = 2$ gives the *mean-square value* of X , defined by

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx \quad (3.43)$$

Variance

We may also define *central moments*, which are simply the moments of the difference between a random variable X and its mean μ_X . Thus, the n th central moment of X is

$$\mathbb{E}[(X - \mu_X)^n] = \int_{-\infty}^{\infty} (x - \mu_X)^n f_X(x) dx \quad (3.44)$$

For $n = 1$, the central moment is, of course, zero. For $n = 2$, the second central moment is referred to as the *variance* of the random variable X , defined by

$$\begin{aligned} \text{var}[X] &= \mathbb{E}(X - \mu_X)^2 \\ &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \end{aligned} \quad (3.45)$$

The variance of a random variable X is commonly denoted by σ_X^2 . The square root of the variance, namely σ_X , is called the *standard deviation* of the random variable X .

In a sense, the variance σ_X^2 of the random variable X is a measure of the variable's "randomness" or "volatility." By specifying the variance σ_X^2 we essentially constrain the effective width of the probability density function $f_X(x)$ of the random variable X about the mean μ_X . A precise statement of this constraint is contained in the *Chebyshev inequality*, which states that for any positive number ε , we have the probability

$$\mathbb{P}[|X - \mu_X| \geq \varepsilon] \leq \frac{\sigma_X^2}{\varepsilon^2} \quad (3.46)$$

From this inequality we see that the mean and variance of a random variable provide a *weak description* of its probability distribution; hence the practical importance of these two statistical averages.

Using (3.43) and (3.45), we find that the variance σ_X^2 and the mean-square value $\mathbb{E}[X^2]$ are related by

$$\begin{aligned} \sigma_X^2 &= \mathbb{E}[X^2 - 2\mu_X X + \mu_X^2] \\ &= \mathbb{E}[X^2] - 2\mu_X \mathbb{E}[X] + \mu_X^2 \\ &= \mathbb{E}[X^2] - \mu_X^2 \end{aligned} \quad (3.47)$$

where, in the second line, we used the linearity property of the statistical expectation operator \mathbb{E} . Equation (3.47) shows that if the mean μ_X is zero, then the variance σ_X^2 and the mean-square value $\mathbb{E}[X^2]$ of the random variable X are equal.

Covariance

Thus far, we have considered the characterization of a single random variable. Consider next a pair of random variables X and Y . In this new setting, a set of statistical averages of importance is the *joint moments*, namely the expectation of $X^i Y^k$, where i and k may assume any positive integer values. Specifically, by definition, we have

$$\mathbb{E}[X^i Y^k] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^i y^k f_{X, Y}(x, y) \, dx \, dy \quad (3.48)$$

A joint moment of particular importance is the *correlation*, defined by $\mathbb{E}[XY]$, which corresponds to $i = k = 1$ in this equation.

More specifically, the correlation of the centered random variables $(X - \mathbb{E}[X])$ and $(Y - \mathbb{E}[Y])$, that is, the joint moment

$$\text{cov}[XY] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \quad (3.49)$$

is called the *covariance* of X and Y . Let $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$; we may then expand (3.49) to obtain the result

$$\text{cov}[XY] = \mathbb{E}[XY] - \mu_X \mu_Y \quad (3.50)$$

where we have made use of the linearity property of the expectation operator \mathbb{E} . Let σ_X^2 and σ_Y^2 denote the variances of X and Y , respectively. Then, the covariance of X and Y , normalized with respect to the product $\sigma_X \sigma_Y$, is called the *correlation coefficient of X and Y* , expressed as

$$\rho(X, Y) = \frac{\text{cov}[XY]}{\sigma_X \sigma_Y} \quad (3.51)$$

The two random variables X and Y are said to be *uncorrelated* if, and only if, their covariance is zero; that is,

$$\text{cov}[XY] = 0$$

They are said to be *orthogonal* if and only if their correlation is zero; that is,

$$\mathbb{E}[XY] = 0$$

In light of (3.50), we may therefore make the following statement:

If one of the random variables X and Y or both have zero means, and if they are orthogonal, then they are uncorrelated, and vice versa.

3.8 Characteristic Function

In the preceding section we showed that, given a continuous random variable X , we can formulate the probability law defining the expectation of X^n (i.e., n th moment of X) in terms of the probability density function $f_X(x)$, as shown in (3.42). We now introduce another way of formulating this probability law; we do so through the *characteristic function*.

For a formal definition of this new concept, we say:

The characteristic function of a continuous random variable X , denoted by $\Phi_X(\nu)$, is defined as the expectation of the complex exponential function $\exp(j\nu X)$, that is

$$\begin{aligned}\Phi_X(\nu) &= \mathbb{E}[\exp(j\nu X)] \\ &= \int_{-\infty}^{\infty} f_X(x) \exp(j\nu x) dx\end{aligned}\tag{3.52}$$

where ν is real and $j = \sqrt{-1}$.

According to the second expression on the right-hand side of (3.52), we may also view the characteristic function $\Phi_X(\nu)$ of the random variable X as the Fourier transform of the associated probability density function $f_X(x)$, except for a sign change in the exponent. In this interpretation of the characteristic function we have used $\exp(j\nu x)$ rather than $\exp(-j\nu x)$ so as to conform with the convention adopted in probability theory.

Recognizing that ν and x play roles analogous to the variables $2\pi f$ and t respectively in the Fourier-transform theory, we may appeal to the Fourier transform theory of Chapter 2 to recover the probability density function $f_X(x)$ of the random variable X given the characteristic function $\Phi_X(\nu)$. Specifically, we may use the *inversion formula* to write

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_X(\nu) \exp(-j\nu x) dx\tag{3.53}$$

Thus, with $f_X(x)$ and $\Phi_X(\nu)$ forming a Fourier-transform pair, we may obtain the moments of the random variable X from the function $\Phi_X(\nu)$. To pursue this issue, we differentiate both sides of (3.52) with respect to ν a total of n times, and then set $\nu = 0$; we thus get the result

$$\frac{d^n}{d\nu^n} \Phi_X(\nu) \Big|_{\nu=0} = (j)^n \int_{-\infty}^{\infty} x^n f_X(x) dx\tag{3.54}$$

The integral on the right-hand side of this relation is recognized as the n th moment of the random variable X . Accordingly, we may recast (3.54) in the equivalent form

$$\mathbb{E}[X^n] = (-j)^n \frac{d^n}{d\nu^n} \Phi_X(\nu) \Big|_{\nu=0}\tag{3.55}$$

This equation is a mathematical statement of the so-called *moment theorem*. Indeed, it is because of (3.55) that the characteristic function $\Phi_X(\nu)$ is also referred to as a *moment-generating function*.

EXAMPLE 5 Exponential Distribution

The exponential distribution is defined by

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x), & x \geq 0 \\ 0, & \text{otherwise} \end{cases}\tag{3.56}$$

where λ is the only parameter of the distribution. The characteristic function of the distribution is therefore

$$\begin{aligned}\Phi(\nu) &= \int_0^{\infty} \lambda \exp(-\lambda x) \exp(j\nu x) dx \\ &= \frac{\lambda}{\lambda - j\nu}\end{aligned}$$

We wish to use this result to find the mean of the exponentially distributed random variable X . To do this evaluation, we differentiate the characteristic function $\Phi(\nu)$ with respect to ν once, obtaining

$$\Phi'_X(\nu) = \frac{\lambda j}{(\lambda - j\nu)^2}$$

where the prime in $\Phi'_X(\nu)$ signifies first-order differentiation with respect to the argument ν . Hence, applying the moment theorem of (3.55), we get the desired result

$$\begin{aligned}\mathbb{E}[X] &= -j\Phi'_X(\nu) \Big|_{\nu=0} \\ &= \frac{1}{\lambda}\end{aligned}\tag{3.57}$$

3.9 The Gaussian Distribution

Among the many distributions studied in the literature on probability theory, the *Gaussian distribution* stands out, by far, as the most commonly used distribution in the statistical analysis of communications systems, for reasons that will become apparent in Section 3.10. Let X denote a continuous random variable; the variable X is said to be *Gaussian distributed* if its probability density function has the general form

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\tag{3.58}$$

where μ and σ are two scalar parameters that characterize the distribution. The parameter μ can assume both positive and negative values (including zero), whereas the parameter σ is always positive. Under these two conditions, the $f_X(x)$ of (3.58) satisfies all the properties of a probability density function, including the normalization property; namely,

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = 1\tag{3.59}$$

Properties of the Gaussian Distribution

A Gaussian random variable has many important properties, four of which are summarized on the next two pages.

PROPERTY 1 Mean and Variance

In the defining (3.58), the parameter μ is the mean of the Gaussian random variable X and σ^2 is its variance. We may therefore state:

A Gaussian random variable is uniquely defined by specifying its mean and variance.

PROPERTY 2 Linear Function of a Gaussian Random Variable

Let X be a Gaussian random variable with mean μ and variance σ^2 . Define a new random variable

$$Y = aX + b$$

where a and b are scalars and $a \neq 0$. Then Y is also Gaussian with mean

$$\mathbb{E}[Y] = a\mu + b$$

and variance

$$\text{var}[Y] = a^2\sigma^2$$

In words, we may state:

Gaussianity is preserved by a linear transformation.

PROPERTY 3 Sum of Independent Gaussian Random Variables

Let X and Y be independent Gaussian random variables with means μ_X and μ_Y , respectively, and variances σ_X^2 and σ_Y^2 , respectively. Define a new random variable

$$Z = X + Y$$

The random variable Z is also Gaussian with mean

$$\mathbb{E}[Z] = \mu_X + \mu_Y \tag{3.60}$$

and variance

$$\text{var}[Z] = \sigma_X^2 + \sigma_Y^2 \tag{3.61}$$

In general, we may therefore state:

The sum of independent Gaussian random variables is also a Gaussian random variable, whose mean and variance are respectively equal to the sum of the means and the sum of the variances of the constituent random variables.

PROPERTY 4 Jointly Gaussian Random Variables

Let X and Y be a pair of jointly Gaussian random variables with zero means and variances σ_X^2 and σ_Y^2 , respectively. The joint probability density function of X and Y is completely determined by σ_X , σ_Y , and ρ , where ρ is the correlation coefficient defined in (3.51). Specifically, we have

$$f_{X,Y}(x,y) = c \exp(-q(x,y)) \tag{3.62}$$

where the normalization constant c is defined by

$$c = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_X\sigma_Y} \tag{3.63}$$

and the exponential term is defined by

$$q(x, y) = \frac{1}{2(1-\rho^2)} \left(\frac{x^2}{\sigma_X^2} - 2\rho \frac{xy}{\sigma_X \sigma_Y} + \frac{y^2}{\sigma_Y^2} \right) \quad (3.64)$$

In the special case where the correlation coefficient ρ is zero, the joint probability density function of X and Y assumes the simple form

$$\begin{aligned} f_{X, Y}(x, y) &= \frac{1}{2\pi \sigma_X \sigma_Y} \exp\left(-\frac{x^2}{2\sigma_X^2} - \frac{y^2}{2\sigma_Y^2}\right) \\ &= f_X(x)f_Y(y) \end{aligned} \quad (3.65)$$

Accordingly, we may make the statement:

If the random variables X and Y are both Gaussian with zero mean and if they are also orthogonal (that is, $\mathbb{E}[XY] = 0$), then they are statistically independent.

By virtue of Gaussianity, this statement is stronger than the last statement made at the end of the subsection on covariance.

Commonly Used Notation

In light of Property 1, the notation $\mathcal{N}(\mu, \sigma^2)$ is commonly used as the shorthand description of a Gaussian distribution parameterized in terms of its mean μ and variance σ^2 . The symbol \mathcal{N} is used in recognition of the fact that the Gaussian distribution is also referred to as the *normal distribution*, particularly in the mathematics literature.

The Standard Gaussian Distribution

When $\mu = 0$ and $\sigma^2 = 1$, the probability density function of (3.58) reduces to the special form:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (3.66)$$

A Gaussian random variable X so described is said to be in its *standard form*.⁴ Correspondingly, the distribution function of the standard Gaussian random variable is defined by

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt \quad (3.67)$$

Owing to the frequent use of integrals of the type described in (3.67), several related functions have been defined and tabulated in the literature. The related function commonly used in the context of communication systems is the *Q-function*, which is formally defined as

$$\begin{aligned} Q(x) &= 1 - F_X(x) \\ &= \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{t^2}{2}\right) dt \end{aligned} \quad (3.68)$$

In words, we may describe the Q -function as follows:

The Q -function, $Q(x)$, is equal to the area covered by the tail of the probability density function of the standard Gaussian random variable X , extending from x to infinity.

Unfortunately, the integral of (3.67) defining the standard Gaussian distribution $F_X(x)$ does not have a closed-form solution. Rather, with accuracy being an issue of importance, $F_X(x)$ is usually presented in the form of a table for varying x . Table 3.1 is one such recording. To utilize this table for calculating the Q -function, we build on two defining equations:

1. For nonnegative values of x , the first line of (3.68) is used.
2. For negative values of x , use is made of the *symmetric property* of the Q -function:

$$Q(-x) = 1 - Q(x) \quad (3.69)$$

Standard Gaussian Graphics

To visualize the graphical formats of the commonly used standard Gaussian functions, $F_X(x)$, $f_X(x)$, and $Q(x)$, three plots are presented at the bottom of this page:

1. Figure 3.10a plots the distribution function, $F_X(x)$, defined in (3.67).
2. Figure 3.10b plots the density function, $f_X(x)$, defined in (3.66).
3. Figure 3.11 plots the Q -function defined in (3.68).

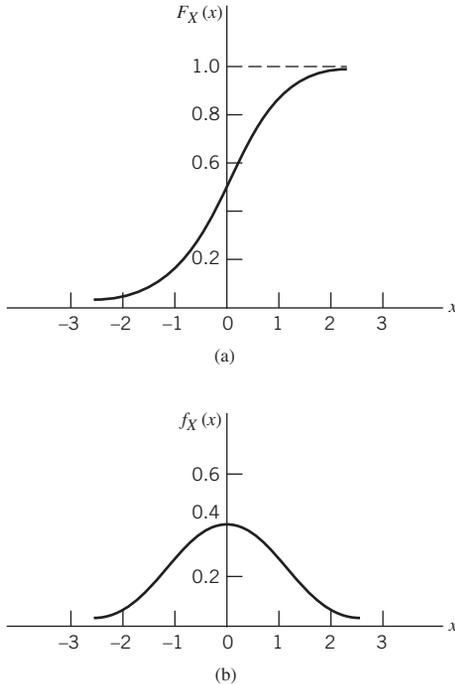


Figure 3.10 The normalized Gaussian (a) distribution function and (b) probability density function.

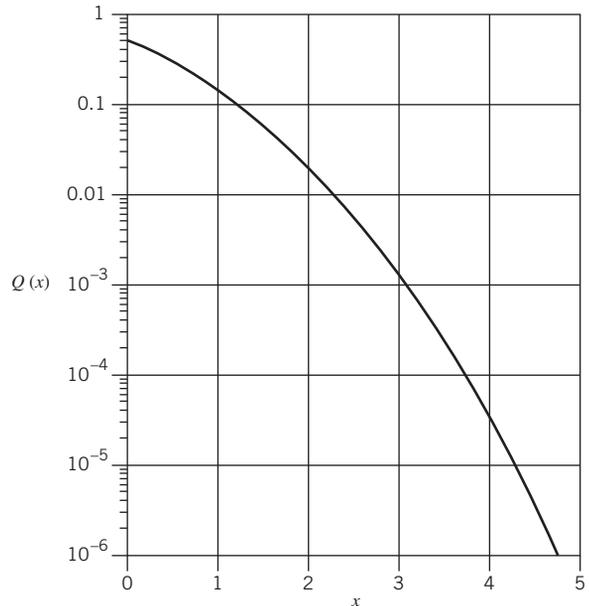


Figure 3.11 The Q -function.

Table 3.1 The standard Gaussian distribution (Q -function) table⁵

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6460	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7485	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9149	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

1. The entries in this table, x say, occupy the range $[0.0, 3.49]$; the x is sample value of the random variable X .
2. For each value of x , the table provides the corresponding value of the Q -function:

$$Q(x) = 1 - F_x(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-t^2/2) dt$$

3.10 The Central Limit Theorem

The *central limit theorem* occupies an important place in probability theory: it provides the mathematical justification for using the Gaussian distribution as a *model* for an observed random variable that is known to be the result of a large number of random events.

For a formal statement of the central limit theorem, let X_1, X_2, \dots, X_n denote a sequence of *independently and identically distributed (iid) random variables* with common mean μ and variance σ^2 . Define the related random variable

$$Y_n = \frac{1}{\sigma\sqrt{n}} \left(\sum_{i=1}^n X_i - n\mu \right) \quad (3.70)$$

The subtraction of the product term $n\mu$ from the sum $\sum_{i=1}^n X_i$ ensures that the random variable Y_n has zero mean; the division by the factor $\sigma\sqrt{n}$ ensures that Y_n has unit variance.

Given the setting described in (3.70), the central limit theorem formally states:

As the number of random variables n in (3.70) approaches infinity, the normalized random variable Y_n converges to the standard Gaussian random variable with the distribution function

$$F_Y(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y \exp\left(-\frac{x^2}{2}\right) dx$$

in the sense that

$$\lim_{n \rightarrow \infty} \mathbb{P}(Y_n \leq y) = Q(y) \quad (3.71)$$

where $Q(y)$ is the Q -function.

To appreciate the practical importance of the central limit theorem, suppose that we have a physical phenomenon whose occurrence is attributed to a large number of random events. The theorem, embodying (3.67)–(3.71), permits us to calculate certain probabilities simply by referring to a Q -function table (e.g., Table 3.1). Moreover, to perform the calculation, all that we need to know are means and variances.

However, a word of caution is in order here. The central limit theorem gives only the “limiting” form of the probability distribution of the standardized random variable Y_n as n approaches infinity. When n is finite, it is sometimes found that the Gaussian limit provides a relatively poor approximation for the actual probability distribution of Y_n , even though n may be large.

EXAMPLE 6 Sum of Uniformly Distributed Random Variables

Consider the random variable

$$Y_n = \sum_{i=1}^n X_i$$

where the X_i are independent and uniformly distributed random variables on the interval from -1 to $+1$. Suppose that we generate 20000 samples of the random variable Y_n for $n = 10$, and then compute the probability density function of Y_n by forming a histogram of the results. Figure 3.11a compares the computed histogram (scaled for unit area) with the probability density function of a Gaussian random variable with the same mean and variance. The figure clearly illustrates that in this particular example the number of independent distributions n does not have to be large for the sum Y_n to closely approximate a Gaussian distribution. Indeed, the results of this example confirm how powerful the central limit theorem is. Moreover, the results explain why Gaussian models are so ubiquitous in the analysis of random signals not only in the study of communication systems, but also in so many other disciplines.

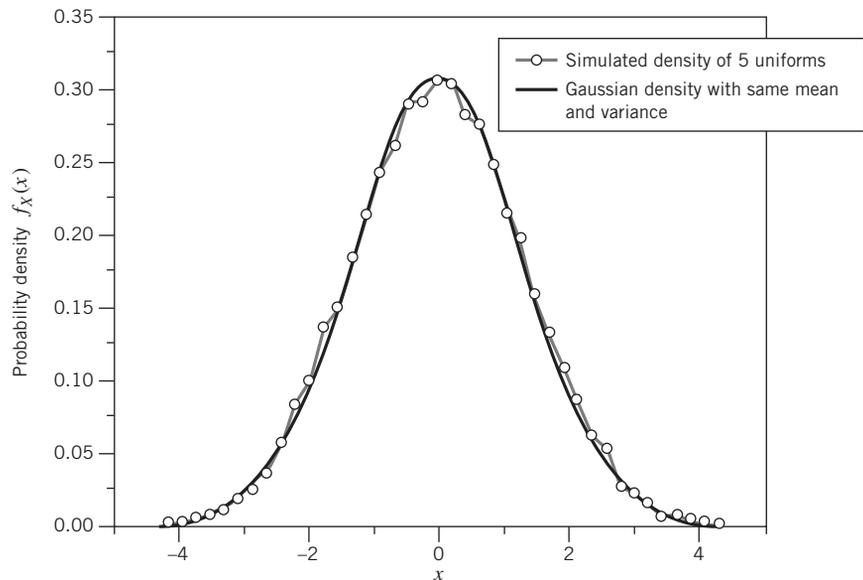


Figure 3.11 Simulation supporting validity of the central limit theorem.

3.11 Bayesian Inference

The material covered up to this point in the chapter has largely addressed issues involved in the mathematical description of probabilistic models. In the remaining part of the chapter we will study the role of probability theory in probabilistic reasoning based on the Bayesian⁵ paradigm, which occupies a central place in statistical communication theory.

To proceed with the discussion, consider Figure 3.12, which depicts two finite-dimensional spaces: a *parameter space* and an *observation space*, with the parameter space being hidden from the *observer*. A parameter vector θ , drawn from the parameter space, is mapped probabilistically onto the observation space, producing the observation vector \mathbf{x} . The vector \mathbf{x} is the sample value of a random vector \mathbf{X} , which provides the

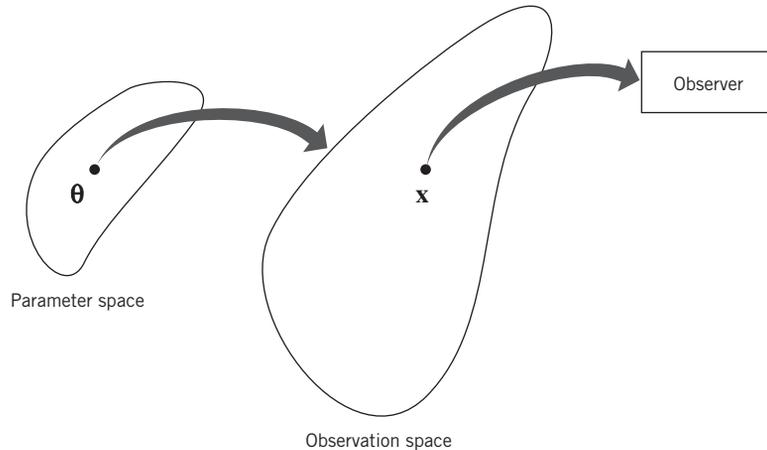


Figure 3.12 Probabilistic model for Bayesian inference.

observer information about θ . Given the probabilistic scenario depicted in Figure 3.12, we may identify two different operations that are the dual of each other.⁶

1. *Probabilistic modeling.* The aim of this operation is to formulate the conditional probability density function $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$, which provides an adequate description of the underlying physical behavior of the observation space.
2. *Statistical analysis.* The aim of this second operation is the *inverse of probabilistic modeling*, for which we need the conditional probability density function $f_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$.

In a fundamental sense, statistical analysis is more profound than probabilistic modeling. We may justify this assertion by viewing the unknown parameter vector θ as the *cause* for the physical behavior of the observation space and viewing the observation vector \mathbf{x} as the *effect*. In essence, statistical analysis solves an *inverse problem* by retrieving the causes (i.e., the parameter vector θ) from the effects (i.e., the observation vector \mathbf{x}). Indeed, we may go on to say that whereas *probabilistic modeling* helps us to characterize the *future behavior* of \mathbf{x} conditional on θ , statistical analysis permits us to make *inference* about θ given \mathbf{x} .

To formulate the conditional probability density function of $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$, we recast Bayes' theorem of (3.14) in its *continuous version*, as shown by

$$f_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)f_{\Theta}(\theta)}{f_{\mathbf{X}}(\mathbf{x})} \quad (3.72)$$

The denominator is itself defined in terms of the numerator as

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \int_{\Theta} f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)f_{\Theta}(\theta) \, d\theta \\ &= \int_{\Theta} f_{\mathbf{X}, \Theta}(\mathbf{x}, \theta) \, d\theta \end{aligned} \quad (3.73)$$

which is the marginal density of \mathbf{X} , obtained by integrating out the dependence of the joint probability density function $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$. In words, $f_{\mathbf{X}}(\mathbf{x})$ is a marginal density of the joint probability density function $f_{\mathbf{X}, \Theta}(\mathbf{x}, \theta)$. The inversion formula of (3.72) is sometimes referred to as the *principle of inverse probability*.

In light of this principle, we may now introduce four notions:

1. *Observation density*. This stands for the conditional probability density function $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$, referring to the “observation” vector \mathbf{x} given the parameter vector θ .
2. *Prior*. This stands for the probability density function $f_{\Theta}(\theta)$, referring to the parameter vector θ “prior” to receiving the observation vector \mathbf{x} .
3. *Posterior*. This stands for the conditional probability density function $f_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$, referring to the parameter vector θ “after” receiving the observation vector \mathbf{x} .
4. *Evidence*. This stands for the probability density function $f_{\mathbf{X}}(\mathbf{x})$, referring to the “information” contained in the observation vector \mathbf{X} for statistical analysis.

The posterior $f_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$ is central to Bayesian inference. In particular, we may view it as the updating of information available on the parameter vector θ in light of the information contained in the observation vector \mathbf{x} , while the prior $f_{\Theta}(\theta)$ is the information available on θ prior to receiving the observation vector \mathbf{x} .

Likelihood

The inversion aspect of statistics manifests itself in the notion of the *likelihood function*.⁷ In a formal sense, the likelihood, denoted by $l(\theta|\mathbf{x})$, is just the observation density $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$ reformulated in a different order, as shown by

$$l(\theta|\mathbf{x}) = f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \quad (3.74)$$

The important point to note here is that the likelihood and the observation density are both governed by exactly the same function that involves the parameter vector θ and the observation vector \mathbf{x} . There is, however, a difference in interpretation: the likelihood function $l(\theta|\mathbf{x})$ is treated as a function of the parameter vector θ given \mathbf{x} , whereas the observation density $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$ is treated as a function of the observation vector \mathbf{x} given θ .

Note, however, unlike $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$, the likelihood $l(\theta|\mathbf{x})$ is *not* a distribution; rather, it is a function of the parameter vector θ , given \mathbf{x} .

In light of the terminologies introduced, namely the posterior, prior, likelihood, and evidence, we may now express Bayes’ rule of (3.72) in words as follows:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

The Likelihood Principle

For convenience of presentation, let

$$\pi(\theta) = f_{\Theta}(\theta) \quad (3.75)$$

Then, recognizing that the evidence defined in (3.73) plays merely the role of a normalizing function that is independent of θ , we may now sum up (3.72) on the principle of inverse probability succinctly as follows:

The Bayesian statistical model is essentially made up of two components: the likelihood function $l(\boldsymbol{\theta}|\mathbf{x})$ and the prior $\pi(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is an unknown parameter vector and \mathbf{x} is the observation vector.

To elaborate on the significance of the defining equation (3.74), consider the likelihood functions $l(\boldsymbol{\theta}|\mathbf{x}_1)$ and $l(\boldsymbol{\theta}|\mathbf{x}_2)$ on parameter vector $\boldsymbol{\theta}$. If, for a prescribed prior $\pi(\boldsymbol{\theta})$, these two likelihood functions are scaled versions of each other, then the corresponding posterior densities of $\boldsymbol{\theta}$ are essentially identical, the validity of which is a straightforward consequence of Bayes' theorem. In light of this result we may now formulate the so-called *likelihood principle*⁸ as follows:

If \mathbf{x}_1 and \mathbf{x}_2 are two observation vectors depending on an unknown parameter vector $\boldsymbol{\theta}$, such that

$$l(\boldsymbol{\theta}|\mathbf{x}_1) = c l(\boldsymbol{\theta}|\mathbf{x}_2) \quad \text{for all } \boldsymbol{\theta}$$

where c is a scaling factor, then these two observation vectors lead to an identical inference on $\boldsymbol{\theta}$ for any prescribed prior $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$.

Sufficient Statistic

Consider a model, parameterized by the vector $\boldsymbol{\theta}$ and given the observation vector \mathbf{x} . In statistical terms, the model is described by the posterior density $f_{\boldsymbol{\theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x})$. In this context, we may now introduce a function $\mathbf{t}(\mathbf{x})$, which is said to be a *sufficient statistic* if the probability density function of the parameter vector $\boldsymbol{\theta}$ given $\mathbf{t}(\mathbf{x})$ satisfies the condition

$$f_{\boldsymbol{\theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x}) = f_{\boldsymbol{\theta}|\mathbf{T}(\mathbf{x})}(\boldsymbol{\theta}|\mathbf{t}(\mathbf{x})) \quad (3.76)$$

This condition imposed on $\mathbf{t}(\mathbf{x})$, for it to be a sufficient statistic, appears intuitively appealing, as evidenced by the following statement:

The function $\mathbf{t}(\mathbf{x})$ provides a sufficient summary of the whole information about the unknown parameter vector $\boldsymbol{\theta}$, which is contained in the observation vector \mathbf{x} .

We may thus view the notion of sufficient statistic as a tool for “data reduction,” the use of which results in considerable simplification in analysis.⁹ The data reduction power of the sufficient statistic $\mathbf{t}(\mathbf{x})$ is well illustrated in Example 7.

3.12 Parameter Estimation

As pointed out previously, the posterior density $f_{\boldsymbol{\theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x})$ is central to the formulation of a Bayesian probabilistic model, where $\boldsymbol{\theta}$ is an unknown parameter vector and \mathbf{x} is the observation vector. It is logical, therefore, that we use this conditional probability density function for parameter estimation.¹⁰ Accordingly, we define the *maximum a posteriori (MAP) estimate* of $\boldsymbol{\theta}$ as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\text{MAP}} &= \arg \max_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x}) \\ &= \arg \max_{\boldsymbol{\theta}} l(\boldsymbol{\theta}|\mathbf{x})\pi(\boldsymbol{\theta}) \end{aligned} \quad (3.77)$$

where $l(\boldsymbol{\theta}|\mathbf{x})$ is the likelihood function defined in (3.74), and $\pi(\boldsymbol{\theta})$ is the prior defined in (3.75). To compute the estimate $\hat{\boldsymbol{\theta}}_{\text{MAP}}$, we require availability of the prior $\pi(\boldsymbol{\theta})$.

In words, the right-hand side of (3.77) reads as follows:

Given the observation vector \mathbf{x} , the estimate $\hat{\boldsymbol{\theta}}_{\text{MAP}}$ is that particular value of the parameter vector $\boldsymbol{\theta}$ in the argument of the posterior density $f_{\boldsymbol{\theta}|\mathbf{x}}(\boldsymbol{\theta}|\mathbf{x})$, for which this density attains its maximum value.

Generalizing the statement made at the end of the discussion on multiple random variables in Section 3.5, we may now go on to say that, for the problem at hand, the conditional probability density function $f_{\boldsymbol{\theta}|\mathbf{x}}(\boldsymbol{\theta}|\mathbf{x})$ contains all the conceivable information about the multidimensional parameter vector $\boldsymbol{\theta}$ given the observation vector \mathbf{x} . The recognition of this fact leads us to make the follow-up important statement, illustrated in Figure 3.13 for the simple case of a one-dimensional parameter vector:

The maximum a posteriori estimate $\hat{\boldsymbol{\theta}}_{\text{MAP}}$ of the unknown parameter vector $\boldsymbol{\theta}$ is the globally optimal solution to the parameter-estimation problem, in the sense that there is no other estimator that can do better.

In referring to $\hat{\boldsymbol{\theta}}_{\text{MAP}}$ as the MAP estimate, we have made a slight change in our terminology: we have, in effect, referred to $f_{\boldsymbol{\theta}|\mathbf{x}}(\boldsymbol{\theta}|\mathbf{x})$ as the *a posteriori density* rather than the *posterior density* of $\boldsymbol{\theta}$. We have made this minor change so as to conform to the MAP terminology that is well and truly embedded in the literature on statistical communication theory.

In another approach to parameter estimation, known as *maximum likelihood estimation*, the parameter vector $\boldsymbol{\theta}$ is estimated using the formula

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \arg \sup_{\boldsymbol{\theta}} l(\boldsymbol{\theta}|\mathbf{x}) \quad (3.78)$$

That is, the *maximum likelihood estimate* $\hat{\boldsymbol{\theta}}_{\text{ML}}$ is that value of the parameter vector $\boldsymbol{\theta}$ that maximizes the conditional distribution $f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})$ at the observation vector \mathbf{x} . Note that this second estimate ignores the prior $\pi(\boldsymbol{\theta})$ and, therefore, lies at the fringe of the Bayesian paradigm. Nevertheless, maximum likelihood estimation is widely used in the literature on statistical communication theory, largely because in ignoring the prior $\pi(\boldsymbol{\theta})$, it is less demanding than maximum posterior estimation in computational complexity.

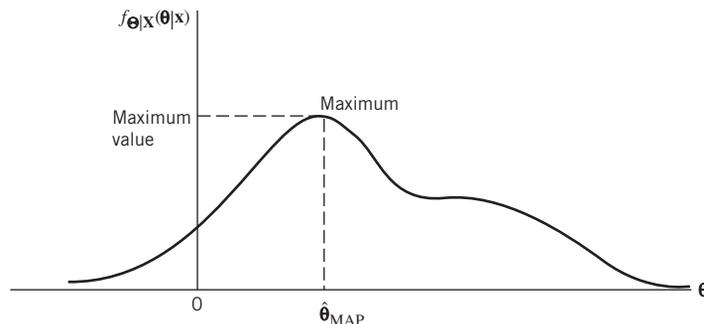


Figure 3.13 Illustrating the a posteriori $f_{\boldsymbol{\theta}|\mathbf{x}}(\boldsymbol{\theta}|\mathbf{x})$ for the case of a one-dimensional parameter space.

The MAP and ML estimates do share a common possibility, in that the maximizations in (3.77) and (3.78) may lead to more than one global maximum. However, they do differ in one important result: the maximization indicated in (3.78) may *not* always be possible; that is, the procedure used to perform the maximization may *diverge*. To overcome this difficulty, the solution to (3.78) has to be *stabilized* by incorporating prior information on the parameter space, exemplified by the distribution $\pi(\boldsymbol{\theta})$, into the solution, which brings us back to the Bayesian approach and, therefore, (3.77). The most critical part in the Bayesian approach to statistical modeling and parameter estimation is how to choose the prior $\pi(\boldsymbol{\theta})$. There is also the possibility of the Bayesian approach requiring high-dimensional computations. We should not, therefore, underestimate the challenges involved in applying the Bayesian approach, on which note we may say the following:

There is no free lunch: for every gain made, there is a price to be paid.

EXAMPLE 7 Parameter Estimation in Additive Noise

Consider a set N of scalar observations, defined by

$$x_i = \theta + n_i, \quad i = 1, 2, \dots, N \quad (3.79)$$

where the unknown parameter θ is drawn from the Gaussian distribution $\mathcal{N}(0, \sigma_\theta^2)$; that is,

$$f_\Theta(\theta) = \frac{1}{\sqrt{2\pi}\sigma_\theta} \exp\left(-\frac{\theta^2}{2\sigma_\theta^2}\right) \quad (3.80)$$

Each n_i is drawn from another Gaussian distribution $\mathcal{N}(0, \sigma_n^2)$; that is,

$$f_{N_i}(n_i) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{n_i^2}{2\sigma_n^2}\right), \quad i = 1, 2, \dots, N$$

It is assumed that the random variables N_i are all independent of each other, and also independent from Θ . The issue of interest is to find the MAP of the parameter θ .

To find the distribution of the random variable X_i , we invoke Property 2 of the Gaussian distribution, described in Section 3.9, in light of which we may say that X_i is also Gaussian with mean θ and variance σ_n^2 . Furthermore, since the N_i are independent, by assumption, it follows that the X_i are also independent. Hence, using the vector \mathbf{x} to denote the N observations, we express the observation density of \mathbf{x} as

$$\begin{aligned} f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{(x_i - \theta)^2}{2\sigma_n^2}\right] \\ &= \frac{1}{(\sqrt{2\pi}\sigma_n)^N} \exp\left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (x_i - \theta)^2\right] \end{aligned} \quad (3.81)$$

The problem is to determine the MAP estimate of the unknown parameter θ .

To solve this problem, we need to know the posterior density $f_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$. Applying (3.72), we write

$$f_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = c(\mathbf{x}) \exp \left[-\frac{1}{2} \left(\frac{\theta^2}{\sigma_\theta^2} + \frac{\sum_{i=1}^N (x_i - \theta)^2}{\sigma_n^2} \right) \right] \quad (3.82)$$

where

$$c(\mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma_\theta} \times \frac{1}{(\sqrt{2\pi}\sigma_n)^N} \frac{1}{f_{\mathbf{X}}(\mathbf{x})} \quad (3.83)$$

The normalization factor $c(\mathbf{x})$ is independent of the parameter θ and, therefore, has no relevance to the MAP of θ . We therefore need only pay attention to the exponent in (3.82).

Rearranging terms and completing the square in the exponent in (3.82), and introducing a new normalization factor $c'(\mathbf{x})$ that absorbs all the terms involving x_i , we get

$$f_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = c'(\mathbf{x}) \exp \left\{ -\frac{1}{2\sigma_p^2} \left(\frac{\sigma_n^2}{\sigma_\theta^2 + (\sigma_n^2/N)} \left(\frac{1}{N} \sum_{i=1}^N x_i \right) - \theta \right)^2 \right\} \quad (3.84)$$

where

$$\sigma_p^2 = \frac{\sigma_\theta^2 \sigma_n^2}{N\sigma_\theta^2 + \sigma_n^2} \quad (3.85)$$

Equation (3.84) shows that the posterior density of the unknown parameter θ is Gaussian with mean θ and variance σ_p^2 . We therefore readily find that the MAP estimate of θ is

$$\hat{\theta}_{\text{MAP}} = \frac{\sigma_n^2}{\sigma_\theta^2 + (\sigma_n^2/N)} \left(\frac{1}{N} \sum_{i=1}^N x_i \right) \quad (3.86)$$

which is the desired result.

Examining (3.84), we also see that the N observations enter the posterior density of θ only through the sum of the x_i . It follows, therefore, that

$$t(\mathbf{x}) = \sum_{i=1}^N x_i \quad (3.87)$$

is a sufficient statistic for the example at hand. This statement merely confirms that (3.84) and (3.87) satisfy the condition of (3.76) for a sufficient statistic.

3.13 Hypothesis Testing

The Bayesian paradigm discussed in Section 3.11 focused on two basic issues: predictive modeling of the observation space and statistical analysis aimed at parameter estimation. As mentioned previously in that section, these two issues are the dual of each other. In this section we discuss another facet of the Bayesian paradigm, aimed at *hypothesis testing*,¹¹ which is basic to signal detection in digital communications, and beyond.

Binary Hypotheses

To set the stage for the study of hypothesis testing, consider the model of Figure 3.14. A *source of binary data* emits a sequence of 0s and 1s, which are respectively denoted by hypotheses H_0 and H_1 . The source (e.g., digital communication transmitter) is followed by a *probabilistic transition mechanism* (e.g., communication channel). According to some probabilistic law, the transition mechanism generates an *observation vector* \mathbf{x} that defines a specific point in the observation space.

The mechanism responsible for probabilistic transition is *hidden* from the observer (e.g., digital communication receiver). Given the observation vector \mathbf{x} and knowledge of the probabilistic law characterizing the transition mechanism, the observer chooses whether hypothesis H_0 or H_1 is true. Assuming that a *decision* must be made, the observer has to have a *decision rule* that works on the observation vector \mathbf{x} , thereby dividing the observation space Z into two regions: Z_0 corresponding to H_0 being true and Z_1 corresponding to H_1 being true. To simplify matters, the decision rule is not shown in Figure 3.14.

In the context of a digital communication system, for example, the channel plays the role of the probabilistic transition mechanism. The observation space of some finite

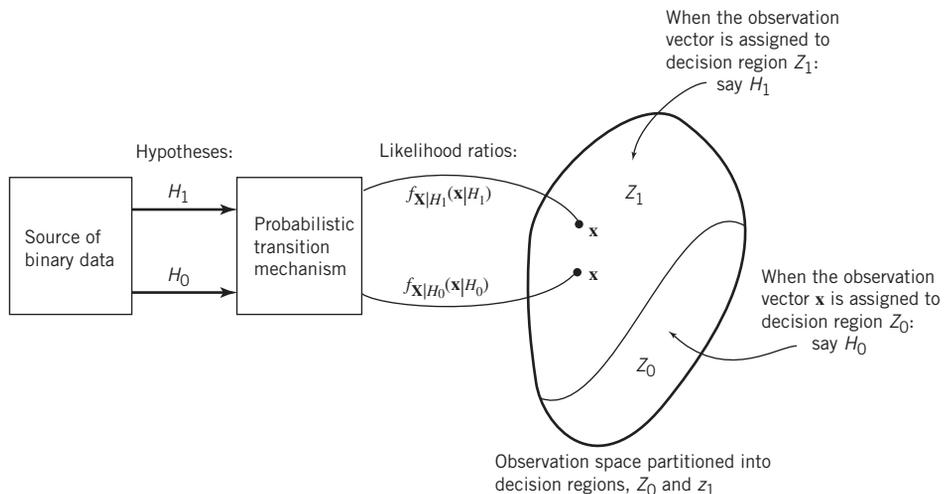


Figure 3.14 Diagram illustrating the binary hypothesis-testing problem. *Note:* according to the likelihood ratio test, the bottom observation vector \mathbf{x} is incorrectly assigned to Z_1 .

dimension corresponds to the ensemble of channel outputs. Finally, the receiver performs the decision rule.

Likelihood Receiver

To proceed with the solution to the binary hypothesis-testing problem, we introduce the following notations:

1. $f_{\mathbf{X}|H_0}(\mathbf{x}|H_0)$, which denotes the conditional density of the observation vector \mathbf{x} given that hypothesis H_0 is true.
2. $f_{\mathbf{X}|H_1}(\mathbf{x}|H_1)$, denotes the conditional density of \mathbf{x} given that the other hypothesis H_1 is true.
3. π_0 and π_1 denote the priors of hypotheses H_0 and H_1 , respectively.

In the context of hypothesis testing, the two conditional probability density functions $f_{\mathbf{X}|H_0}(\mathbf{x}|H_0)$ and $f_{\mathbf{X}|H_1}(\mathbf{x}|H_1)$ are referred to as *likelihood functions*, or just simply *likelihoods*.

Suppose we perform a measurement on the transition mechanism's output, obtaining the observation vector \mathbf{x} . In processing \mathbf{x} , there are two kinds of errors that can be made by the decision rule:

1. *Error of the first kind.* This arises when hypothesis H_0 is true but the rule makes a decision in favor of H_1 , as illustrated in Figure 3.14.
2. *Error of the second kind.* This arises when hypothesis H_1 is true but the rule makes a decision in favor of H_0 .

The conditional probability of an error of the first kind is

$$\int_{Z_1} f_{\mathbf{X}|H_0}(\mathbf{x}|H_0) \, d\mathbf{x}$$

where Z_1 is part of the observation space that corresponds to hypothesis H_1 . Similarly, the conditional probability of an error of the second kind is

$$\int_{Z_0} f_{\mathbf{X}|H_1}(\mathbf{x}|H_1) \, d\mathbf{x}$$

By definition, an *optimum* decision rule is one for which a prescribed *cost function* is minimized. A logical choice for the cost function in digital communications is the *average probability of symbol error*, which, in a Bayesian context, is referred to as the *Bayes risk*. Thus, with the probable occurrence of the two kinds of errors identified above, we define the Bayes risk for the binary hypothesis-testing problem as

$$\mathcal{R} = \pi_0 \int_{Z_1} f_{\mathbf{X}|H_0}(\mathbf{x}|H_0) \, d\mathbf{x} + \pi_1 \int_{Z_0} f_{\mathbf{X}|H_1}(\mathbf{x}|H_1) \, d\mathbf{x} \quad (3.88)$$

where we have accounted for the prior probabilities for which hypotheses H_0 and H_1 are known to occur. Using the language of set theory, let the union of the disjoint subspaces Z_0 and Z_1 be

$$Z = Z_0 \cup Z_1 \quad (3.89)$$

Then, recognizing that the subspace Z_1 is the complement of the subspace Z_0 with respect to the total observation space Z , we may rewrite (3.88) in the equivalent form:

$$\begin{aligned}\mathcal{R} &= \pi_0 \int_{Z-Z_0} f_{\mathbf{X}|H_0}(\mathbf{x}|H_0) d\mathbf{x} + \pi_1 \int_{Z_0} f_{\mathbf{X}|H_1}(\mathbf{x}|H_1) d\mathbf{x} \\ &= \pi_0 \int_Z f_{\mathbf{X}|H_0}(\mathbf{x}|H_0) d\mathbf{x} + \int_{Z_0} [\pi_1 f_{\mathbf{X}|H_1}(\mathbf{x}|H_1) - \pi_0 f_{\mathbf{X}|H_0}(\mathbf{x}|H_0)] d\mathbf{x}\end{aligned}\quad (3.90)$$

The integral $\int_Z f_{\mathbf{X}|H_0}(\mathbf{x}|H_0) d\mathbf{x}$ represents the total volume under the conditional density $f_{\mathbf{X}|H_0}(\mathbf{x}|H_0)$, which, by definition, equals unity. Accordingly, we may reduce (3.90) to

$$\mathcal{R} = \pi_0 + \int_{Z_0} [\pi_1 f_{\mathbf{X}|H_1}(\mathbf{x}|H_1) - \pi_0 f_{\mathbf{X}|H_0}(\mathbf{x}|H_0)] d\mathbf{x}\quad (3.91)$$

The term π_0 on its own on the right-hand side of (3.91) represents a *fixed* cost. The integral term represents the cost controlled by how we assign the observation vector \mathbf{x} to Z_0 . Recognizing that the two terms inside the square brackets are both positive, we must therefore insist on the following plan of action for the average risk \mathcal{R} to be minimized:

Make the integrand in (3.91) negative for the observation vector \mathbf{x} to be assigned to Z_0 .

In light of this statement, the optimum decision rule proceeds as follows:

1. If

$$\pi_0 f_{\mathbf{X}|H_0}(\mathbf{x}|H_0) > \pi_1 f_{\mathbf{X}|H_1}(\mathbf{x}|H_1)$$

then the observation vector \mathbf{x} should be assigned to Z_0 , because these two terms contribute a negative amount to the integral in (3.91). In this case, we say H_0 is true.

2. If, on the other hand,

$$\pi_0 f_{\mathbf{X}|H_0}(\mathbf{x}|H_0) < \pi_1 f_{\mathbf{X}|H_1}(\mathbf{x}|H_1)$$

then the observation vector \mathbf{x} should be excluded from Z_0 (i.e., assigned to Z_1), because these two terms would contribute a positive amount to the integral in (3.91). In this second case, H_1 is true.

When the two terms are equal, the integral would clearly have no effect on the average risk \mathcal{R} ; in such a situation, the observation vector \mathbf{x} may be assigned arbitrarily.

Thus, combining points (1) and (2) on the action plan into a *single* decision rule, we may write

$$\frac{f_{\mathbf{X}|H_1}(\mathbf{x}|H_1)}{f_{\mathbf{X}|H_0}(\mathbf{x}|H_0)} \underset{H_0}{\overset{H_1}{>}} \frac{\pi_0}{\pi_1}\quad (3.92)$$

The observation-dependent quantity on the left-hand side of (3.92) is called the *likelihood ratio*; it is defined by

$$\Lambda(\mathbf{x}) = \frac{f_{\mathbf{X}|H_1}(\mathbf{x}|H_1)}{f_{\mathbf{X}|H_0}(\mathbf{x}|H_0)}\quad (3.93)$$

From this definition, we see that $\Lambda(\mathbf{x})$ is the ratio of two functions of a random variable; therefore, it follows that $\Lambda(\mathbf{x})$ is itself a random variable. Moreover, it is a one-

dimensional variable, which holds regardless of the dimensionality of the observation vector \mathbf{x} . Most importantly, the likelihood ratio is a sufficient statistic.

The scalar quantity on the right-hand side of (3.92), namely,

$$\eta = \frac{\pi_0}{\pi_1} \quad (3.94)$$

is called the *threshold* of the test. Thus, minimization of the Bayes risk \mathcal{R} leads to the *likelihood ratio test*, described by the combined form of two decisions:

$$\Lambda(\mathbf{x}) \begin{matrix} H_1 \\ \gtrless \\ H_0 \end{matrix} \eta \quad (3.95)$$

Correspondingly, the hypothesis testing structure built on (3.93)–(3.95) is called the *likelihood receiver*; it is shown in the form of a block diagram in Figure 3.15a. An elegant characteristic of this receiver is that all the necessary data processing is confined to computing the likelihood ratio $\Lambda(\mathbf{x})$. This characteristic is of considerable practical importance: adjustments to our knowledge of the priors π_0 and π_1 are made simply through the assignment of an appropriate value to the threshold η .

The natural logarithm is known to be a monotone function of its argument. Moreover, both sides of the likelihood ratio test in (3.95) are positive. Accordingly, we may express the test in its *logarithmic* form, as shown by

$$\ln \Lambda(\mathbf{x}) \begin{matrix} H_1 \\ \gtrless \\ H_0 \end{matrix} \ln \eta \quad (3.96)$$

where \ln is the symbol for the natural logarithm. Equation (3.96) leads to the equivalent *log-likelihood ratio receiver*, depicted in Figure 3.15b.

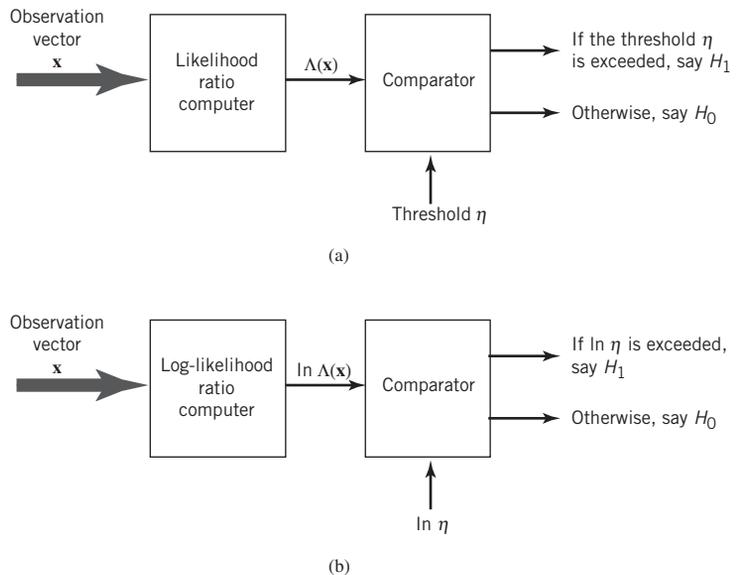


Figure 3.15 Two versions of the likelihood receiver: (a) based on the likelihood ratio $\Lambda(\mathbf{x})$; (b) based on the log-likelihood ratio $\ln \Lambda(\mathbf{x})$.

EXAMPLE 8 Binary Hypothesis Testing

Consider a binary hypothesis testing problem, described by the pair of equations:

$$\begin{aligned} \text{Hypothesis } H_1: x_i &= m + n_i, & i &= 1, 2, \dots, N \\ \text{Hypothesis } H_0: x_i &= n_i, & i &= 1, 2, \dots, N \end{aligned} \quad (3.97)$$

The term m is a constant that is nonzero only under hypothesis H_1 . As in Example 7, the n_i are independent and Gaussian $\mathcal{N}(0, \sigma_n^2)$. The requirement is to formulate a likelihood ratio test for this example to come up with a decision rule.

Following the discussion presented in Example 7, under hypothesis H_1 we write

$$f_{x_i|H_1}(x_i|H_1) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{(x_i - m)^2}{2\sigma_n^2}\right] \quad (3.98)$$

As in Example 7, let the vector \mathbf{x} denote the set of N observations x_i for $i = 1, 2, \dots, N$. Then, invoking the independence of the n_i , we may express the joint density of the x_i under hypothesis H_1 as

$$\begin{aligned} f_{\mathbf{x}|H_1}(\mathbf{x}|H_1) &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{(x_i - m)^2}{2\sigma_n^2}\right] \\ &= \frac{1}{(\sqrt{2\pi}\sigma_n)^N} \exp\left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (x_i - m)^2\right] \end{aligned} \quad (3.99)$$

Setting m to zero in (3.99), we get the corresponding joint density of the x_i under hypothesis H_0 as

$$f_{\mathbf{x}|H_0}(\mathbf{x}|H_0) = \frac{1}{(\sqrt{2\pi}\sigma_n)^N} \exp\left(-\frac{1}{2\sigma_n^2} \sum_{i=1}^N x_i^2\right) \quad (3.100)$$

Hence, substituting (3.99) and (3.100) into the likelihood ratio of (3.93), we get (after canceling common terms)

$$\Lambda(\mathbf{x}) = \exp\left(\frac{m}{\sigma_n^2} \sum_{i=1}^N x_i - \frac{Nm^2}{2\sigma_n^2}\right) \quad (3.101)$$

Equivalently, we may express the likelihood ratio in its logarithmic form

$$\ln \Lambda(\mathbf{x}) = \frac{m}{\sigma_n^2} \sum_{i=1}^N x_i - \frac{Nm^2}{2\sigma_n^2} \quad (3.102)$$

Using (3.102) in the log-likelihood ratio test of (3.96), we get

$$\left(\frac{m}{\sigma_n^2} \sum_{i=1}^N x_i - \frac{Nm^2}{2\sigma_n^2}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta$$

Dividing both sides of this test by (m/σ_n^2) and rearranging terms, we finally write

$$\sum_{i=1}^N x_i \underset{H_0}{\overset{H_1}{\geq}} \left(\frac{\sigma_n^2}{m} \ln \eta + \frac{Nm^2}{2} \right) \quad (3.103)$$

where the threshold η is itself defined by the ratio of priors, namely π_0/π_1 . Equation (3.103) is the desired formula for the decision rule to solve the binary hypothesis-testing problem of (3.97).

One last comment is in order. As with Example 7, the sum of the x_i over the N observations; that is,

$$t(\mathbf{x}) = \sum_{i=1}^N x_i$$

is a sufficient statistic for the problem at hand. We say so because the only way in which the observations can enter the likelihood ratio $\Lambda(\mathbf{x})$ is in the sum; see (3.101).

Multiple Hypotheses

Now that we understand binary hypothesis testing, we are ready to consider the more general scenario where we have M possible source outputs to deal with. As before, we assume that a decision must be made as to which one of the M possible source outputs was actually emitted, given an observation vector \mathbf{x} .

To develop insight into how to construct a decision rule for testing multiple hypotheses, we consider first the case of $M = 3$ and then generalize the result. Moreover, in formulating the decision rule, we will use *probabilistic reasoning* that builds on the findings of the binary hypothesis-testing procedure. In this context, however, we find it more convenient to work with likelihood functions rather than likelihood ratios.

To proceed then, suppose we make a measurement on the probabilistic transition mechanism's output, obtaining the observation vector \mathbf{x} . We use this observation vector and knowledge of the probability law characterizing the transition mechanism to construct three likelihood functions, one for each of the three possible hypotheses. For the sake of illustrating what we have in mind, suppose further that in formulating the three possible probabilistic inequalities, each with its own inference, we get the following three results:

1. $\pi_1 f_{\mathbf{X}|H_1}(\mathbf{x}|H_1) < \pi_0 f_{\mathbf{X}|H_0}(\mathbf{x}|H_0)$
from which we infer that hypothesis H_0 or H_2 is true.
2. $\pi_2 f_{\mathbf{X}|H_2}(\mathbf{x}|H_2) < \pi_0 f_{\mathbf{X}|H_0}(\mathbf{x}|H_0)$
from which we infer that hypothesis H_0 or H_1 is true.
3. $\pi_2 f_{\mathbf{X}|H_2}(\mathbf{x}|H_2) < \pi_1 f_{\mathbf{X}|H_1}(\mathbf{x}|H_1)$
from which we infer that hypothesis H_1 or H_0 is true.

Examining these three possible results for $M = 3$, we immediately see that hypothesis H_0 is the only one that shows up in all three inferences. Accordingly, for the particular scenario we have picked, the decision rule *should* say that hypothesis H_0 is true. Moreover, it is a straightforward matter for us to make similar statements pertaining to hypothesis H_1

or H_2 . The rationale just described for arriving at this test is an *example* of what we mean by probabilistic reasoning: the use of multiple inferences to reach a specific decision.

For an equivalent test, let both sides of each inequality under points 1, 2, and 3 be divided by the evidence $f_{\mathbf{X}}(\mathbf{x})$. Let H_i , $i = 1, 2, 3$, denote the three hypotheses. We may then use the definition of joint probability density function to write

$$\begin{aligned} \frac{\pi_i f_{\mathbf{X}|H_i}(\mathbf{x}|H_i)}{f_{\mathbf{X}}(\mathbf{x})} &= \frac{\mathbb{P}(H_i) f_{\mathbf{X}|H_i}(\mathbf{x}|H_i)}{f_{\mathbf{X}}(\mathbf{x})} \quad \text{where } \mathbb{P}(H_i) = p_i \\ &= \frac{\mathbb{P}(H_i, \mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{\mathbb{P}[H_i|\mathbf{x}] f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \mathbb{P}[H_i|\mathbf{x}] \quad \text{for } i = 0, 1, \dots, M-1 \end{aligned} \tag{3.104}$$

Hence, recognizing that the conditional probability $\mathbb{P}[H_i|\mathbf{x}]$ is actually the posterior probability of hypothesis H_i *after* receiving the observation vector \mathbf{x} , we may now go on to generalize the equivalent test for M possible source outputs as follows:

Given an observation vector \mathbf{x} in a multiple hypothesis test, the average probability of error is minimized by choosing the hypothesis H_i for which the posterior probability $\mathbb{P}[H_i|\mathbf{x}]$ has the largest value for $i = 0, 1, \dots, M-1$.

A processor based on this decision rule is frequently referred to as the *MAP probability computer*. It is with this general hypothesis testing rule that earlier we made the supposition embodied under points 1, 2, and 3.

3.14 Composite Hypothesis Testing

Throughout the discussion presented in Section 3.13, the hypotheses considered therein were all *simple*, in that the probability density function for each hypothesis was completely specified. However, in practice, it is common to find that one or more of the probability density functions are *not* simple due to imperfections in the probabilistic transition mechanism. In situations of this kind, the hypotheses are said to be *composite*.

As an illustrative example, let us revisit the binary hypothesis-testing problem considered in Example 8. This time, however, we treat the mean m of the observable x_i under hypothesis H_1 not as a constant, but as a variable inside some interval $[m_a, m_b]$. If, then, we were to use the likelihood ratio test of (3.93) for simple binary hypothesis testing, we would find that the likelihood ratio $\Lambda(x_i)$ involves the unknown mean m . We cannot therefore compute $\Lambda(x_i)$, thereby negating applicability of the simple likelihood ratio test.

The message to take from this illustrative example is that we have to modify the likelihood ratio test to make it applicable to composite hypotheses. To this end, consider the model depicted in Figure 3.16, which is similar to that of Figure 3.14 for the simple case except for one difference: the transition mechanism is now characterized by the conditional probability density function $f_{\mathbf{X}|\Theta, H_i}(\mathbf{x}|\boldsymbol{\theta}, H_i)$, where $\boldsymbol{\theta}$ is a realization of the unknown parameter vector Θ , and the index $i = 0, 1$. It is the conditional dependence on $\boldsymbol{\theta}$ that makes

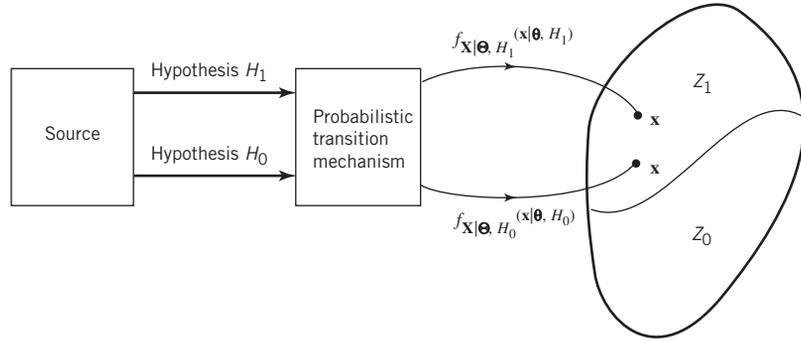


Figure 3.16 Model of composite hypothesis-testing for a binary scenario.

the hypotheses H_0 and H_1 to be of the composite kind. Unlike the simple model of Figure 3.14, we now have two spaces to deal with: an observation space and a parameter space. It is assumed that the conditional probability density function of the unknown parameter vector Θ , that is, $f_{\Theta|H_i}(\theta, H_i)$, is known for $i = 0, 1$.

To formulate the likelihood ratio for the composite hypotheses described in the model of Figure 3.16, we require the likelihood function $f_{\mathbf{X}|H_i}(\mathbf{x}|H_i)$ for $i = 1, 2$. We may satisfy this requirement by reducing the composite hypothesis-testing problem to a simple one by integrating over θ , as shown by

$$f_{\mathbf{X}|H_i}(\mathbf{x}|H_i) = \int_{\Theta} f_{\mathbf{X}|\Theta, H_i}(\mathbf{x}|\theta, H_i) f_{\Theta|H_i}(\theta|H_i) d\theta \quad (3.105)$$

the evaluation of which is contingent on knowing the conditional probability density function of θ given the H_i for $i = 1, 2$. With this specification at hand, we may now formulate the likelihood ratio for composite hypotheses as

$$\Lambda(\mathbf{x}) = \frac{\int_{\Theta} f_{\mathbf{X}|\Theta, H_1}(\mathbf{x}|\theta, H_1) f_{\Theta|H_1}(\theta|H_1) d\theta}{\int_{\Theta} f_{\mathbf{X}|\Theta, H_0}(\mathbf{x}|\theta, H_0) f_{\Theta|H_0}(\theta|H_0) d\theta} \quad (3.106)$$

Accordingly, we may now extend applicability of the likelihood ratio test described in (3.95) to composite hypotheses.

From this discussion, it is clearly apparent that hypothesis testing for composite hypotheses is computationally more demanding than it is for simple hypotheses. Chapter 7 presents applications of composite hypothesis testing to noncoherent detection, in the course of which the phase information in the received signal is accounted for.

3.15 Summary and Discussion

The material presented in this chapter on probability theory is another mathematical pillar in the study of communication systems. Herein, the emphasis has been on how to deal with *uncertainty*, which is a natural feature of every communication system in one form or

another. Typically, uncertainties affect the behavior of channels connecting the transmitter of a communication system to its receiver. Sources of uncertainty include noise, generated internally and externally, and interference from other transmitters.

In this chapter, the emphasis has been on probabilistic modeling, in the context of which we did the following:

1. Starting with set theory, we went on to state the three axioms of probability theory. This introductory material set the stage for the calculation of probabilities and conditional probabilities of events of interest. When partial information is available on the outcome of an experiment, conditional probabilities permit us to reason in a probabilistic sense and thereby enrich our understanding of a random experiment.
2. We discussed the notion of random variables, which provide the natural tools for formulating probabilistic models of random experiments. In particular, we characterized continuous random variables in terms of the cumulative distribution function and probability density function; the latter contains all the conceivable information about a random variable. Through focusing on the mean of a random variable, we studied the expectation or averaging operator, which occupies a dominant role in probability theory. The mean and the variance, considered in that order, provide a weak characterization of a random variable. We also introduced the characteristic function as another way of describing the statistics of a random variable. Although much of the material in the early part of the chapter focused on continuous random variables, we did emphasize important aspects of discrete random variables by describing the concept of the probability mass function (unique to discrete random variables) and the parallel development and similar concepts that embody these two kinds of random variables.
3. Table 3.2 on page 135 summarizes the probabilistic descriptions of some important random variances under two headings: discrete and random. Except for the Rayleigh random variable, these random variables were discussed in the text or are given as end-of-chapter problems; the Rayleigh random variable is discussed in Chapter 4. Appendix A presents advanced probabilistic models that go beyond the contents of Table 3.2.
4. We discussed the characterization of a pair of random variables and introduced the basic concepts of covariance and correlation, and the independence of random variables.
5. We provided a detailed description of the Gaussian distribution and discussed its important properties. Gaussian random variables play a key role in the study of communication systems.

The second part of the chapter focused on the Bayesian paradigm, wherein inference may take one of two forms:

- Probabilistic modeling, the aim of which is to develop a model for describing the physical behavior of an observation space.
- Statistical analysis, the aim of which is the inverse of probabilistic modeling.

In a fundamental sense, statistical analysis is more profound than probabilistic modeling, hence the focused attention on it in the chapter.

Table 3.2 Some important random variables

Discrete random variables	
1. Bernoulli	$p_X(x) = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$ $\mathbb{E}[X] = p$ $\text{var}[X] = p(1-p)$
2. Poisson	$p_X(k) = \frac{\lambda^k}{k!} \exp(-\lambda), \quad k = 0, 1, 2, \dots, \text{ and } \lambda > 0$ $\mathbb{E}[X] = \lambda$ $\text{var}[X] = \lambda$
Continuous random variables	
1. Uniform	$f_X(x) = \frac{1}{b-a}, \quad a \leq x \leq b$ $\mathbb{E}[X] = \frac{1}{2}(a+b)$ $\text{var}[X] = \frac{1}{12}(b-a)^2$
2. Exponential	$f_X(x) = \lambda \exp(-\lambda x), \quad x \geq 0 \text{ and } \lambda > 0$ $\mathbb{E}[X] = 1/\lambda$ $\text{var}[X^2] = 1/\lambda^2$
3. Gaussian	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-(x-\mu)^2/2\sigma^2], \quad -\infty < x < \infty$ $\mathbb{E}[X] = \mu$ $\text{var}[X] = \sigma^2$
4. Rayleigh	$f_X(x) = \frac{x}{\sigma^2} \exp(-x^2/2\sigma^2), \quad x \geq 0 \text{ and } \sigma > 0$ $\mathbb{E}[X] = \sigma\sqrt{\pi/2}$ $\text{var}[X] = \left(2 - \frac{\pi}{2}\right) \sigma^2$
5. Laplacian	$f_X(x) = \frac{\lambda}{2} \exp(-\lambda x), \quad -\infty < x < \infty \text{ and } \lambda > 0$ $\mathbb{E}[X] = 0$ $\text{var}[X] = 2/\lambda^2$

Under statistical analysis, viewed from a digital communications perspective, we discussed the following:

1. Parameter estimation, where the requirement is to estimate an unknown parameter given an observation vector; herein we covered:
 - the maximum a posteriori (MAP) rule that requires prior information, and
 - the maximum likelihood procedure that by-passes the need for the prior and therefore sits on the fringe of the Bayesian paradigm.
2. Hypothesis testing, where in a simple but important scenario, we have two hypotheses to deal with, namely H_1 and H_0 . In this case, the requirement is to make an optimal decision in favor of hypothesis H_1 or hypothesis H_0 given an observation vector. The likelihood ratio test plays the key role here.

To summarize, the material on probability theory sets the stage for the study of stochastic processes in Chapter 4. On the other hand, the material on Bayesian inference plays a key role in Chapters 7, 8, and 9 in one form or another.

Problems

Set Theory

- 3.1 Using Venn diagrams, justify the five properties of the algebra of sets, which were stated (without proofs) in Section 3.1:
 - a. idempotence property
 - b. commutative property
 - c. associative property
 - d. distributive property
 - e. De Morgan's laws.
- 3.2 Let A and B denote two different sets. Validate the following three equalities:
 - a. $A^c = (A^c \cap B) \cup (A^c \cap B^c)$
 - b. $B^c = (A \cap B^c) \cup (A^c \cap B^c)$
 - c. $(A \cap B)^c = (A^c \cap B) \cup (A^c \cap B^c) \cup (A \cap B^c)$

Probability Theory

- 3.3 Using the Bernoulli distribution of Table 3.2, develop an experiment that involves three independent tosses of a fair coin. Irrespective of whether the toss is a head or tail, the probability of every toss is to be conditioned on the results of preceding tosses. Display graphically the sequential evolution of the results.
- 3.4 Use Bayes' rule to convert the conditioning of event B given event A_i into the conditioning of event A_i given event B for the $i = 1, 2, \dots, N$.
- 3.5 A discrete memoryless channel is used to transmit binary data. The channel is discrete in that it is designed to handle discrete messages and it is memoryless in that at any instant of time the channel output depends on the channel input only at that time. Owing to the unavoidable presence of noise in the channel, errors are made in the received binary data stream. The channel is symmetric in that the probability of receiving symbol 1 when symbol 0 is sent is the same as the probability of receiving symbol 0 when symbol 1 is sent.

The transmitter sends 0s across the channel with probability p_0 and 1s with probability p_1 . The receiver occasionally makes random decision errors with probability p ; that is, when symbol 0 is sent across the channel, the receiver makes a decision in favor of symbol 1, and vice versa.

Referring to Figure P3.5, determine the following a posteriori probabilities:

- a. The conditional probability of sending symbol A_0 given that symbol B_0 was received.
- b. The conditional probability of sending symbol A_1 given that symbol B_1 was received.

Hint: Formulate expressions for the probability of receiving event B_0 , and likewise for event B_1 .

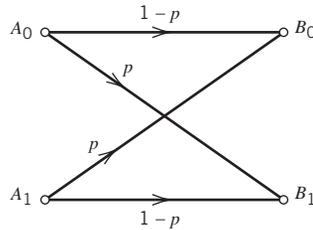


Figure P3.5

- 3.6 Let B_1, B_2, \dots, B_n denote a set of joint events whose union equals the sample space S , and assume that $\mathbb{P}[B_i] > 0$ for all i . Let A be any event in the sample space S .

- a. Show that

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n)$$

- b. The *total probability theorem* states:

$$\mathbb{P}[A] = \mathbb{P}[A|B_1]\mathbb{P}[B_1] + \mathbb{P}[A|B_2]\mathbb{P}[B_2] + \dots + \mathbb{P}[A|B_n]\mathbb{P}[B_n]$$

This theorem is useful for finding the probability of event B when the conditional probabilities $\mathbb{P}[A|B_i]$ are known or easy to find for all i . Justify the theorem.

- 3.7 Figure P3.7 shows the connectivity diagram of a computer network that connects node A to node B along different possible paths. The labeled branches of the diagram display the probabilities for which the links in the network are up; for example, 0.8 is the probability that the link from node A to intermediate node C is up, and so on for the other links. Link failures in the network are assumed to be independent of each other.

- a. When all the links in the network are up, find the probability that there is a path connecting node A to node B.
- b. What is the probability of complete failure in the network, with no connection from node A to node B?

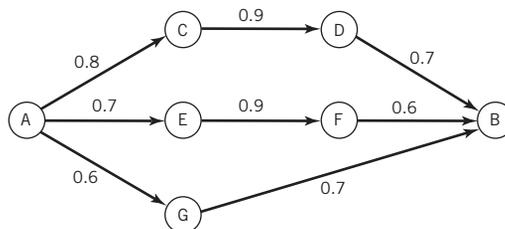


Figure P3.7

Distribution Functions

- 3.8 The probability density function of a continuous random variable X is defined by

$$f_X(x) = \begin{cases} \frac{c}{\sqrt{x}} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Despite the fact that this function becomes infinitely large as x approaches zero, it may qualify to be a legitimate probability density function. Find the value of scalar c for which this condition is satisfied.

- 3.9 The joint probability density function of two random variables X and Y is defined by the *two-dimensional uniform distribution*

$$f_{X,Y}(x,y) = \begin{cases} c & \text{for } a \leq x \leq b \text{ and } a \leq y \leq b \\ 0 & \text{otherwise} \end{cases}$$

Find the scalar c for which $f_{X,Y}(x,y)$ satisfies the normalization property of a two-dimensional probability density function.

- 3.10 In Table 3.2, the probability density function of a Rayleigh random variable is defined by

$$f_X(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad \text{for } x \geq 0 \text{ and } \sigma > 0$$

- a. Show that the mean of X is

$$\mathbb{E}[X] = \sigma \sqrt{\frac{\pi}{2}}$$

- b. Using the result of part a, show that the variance of X is

$$\text{var}[X] = \left(2 - \frac{\pi}{2}\right) \sigma^2$$

- c. Use the results of a and b to determine the Rayleigh cumulative distribution function.

- 3.11 The probability density function of an exponentially distributed random variable X is defined by

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x), & \text{for } 0 \leq x < \infty \\ 0, & \text{otherwise} \end{cases}$$

where λ is a positive parameter.

- a. Show that $f_X(x)$ is a legitimate probability density function.
b. Determine the cumulative distribution function of X .

- 3.12 Consider the one-sided conditional exponential distribution

$$f_X(x|\lambda) = \begin{cases} \frac{\lambda}{Z(\lambda)} \exp(-\lambda x), & 1 \leq x < 20 \\ 0, & \text{otherwise} \end{cases}$$

where $\lambda > 0$ and $Z(\lambda)$ is the normalizing constant required to make the area under $f_X(x|\lambda)$ equal unity.

- a. Determine the normalizing constant $Z(\lambda)$.
b. Given N independent values of x , namely x_1, x_2, \dots, x_N , use Bayes' rule to formulate the conditional probability density function of the parameter λ , given this data set.

Expectation Operator

3.13 In Section 3.6 we described two properties of the expectation operator \mathbb{E} , one on linearity and the other on statistical independence. In this problem, we address two other important properties of the expectation operator.

a. *Scaling property*: Show that

$$\mathbb{E}(ax) = a\mathbb{E}[X]$$

where a is a constant scaling factor.

b. *Linearity of conditional expectation*: Show that

$$\mathbb{E}[X_1 + X_2|Y] = \mathbb{E}[X_1|Y] + \mathbb{E}[X_2|Y]$$

3.14 Validate the expected value rule of (3.41) by building on two expressions:

a. $g(x) = \max[g(x), 0] - \max[-g(x), 0]$

b. For any $a \geq 0$, $g(x) > a$ provided that $\max[g(x), 0] > a$

3.15 Let X be a discrete random variable with probability mass function $p_X(x)$ and let $g(X)$ be a function of the random variable X . Prove the following rule:

$$\mathbb{E}[g(X)] = \sum_x g(x)p_X(x)$$

where the summation is over all possible discrete values of X .

3.16 Continuing with the Bernoulli random variable X in (3.23), find the mean and variance of X .

3.17 The mass probability function of the *Poisson random variable* X is defined by

$$p_X(k) = \frac{1}{k!} \lambda^k \exp(-\lambda), \quad k = 0, 1, 2, \dots, \text{ and } \lambda > 0$$

Find the mean and variance of X .

3.18 Find the mean and variance of the exponentially distributed random variable X in Problem 3.11.

3.19 The probability density function of the *Laplacian random variable* X in Table 3.2 is defined by

$$f_X(x) = \begin{cases} \frac{1}{2} \lambda \exp(-\lambda x) & \text{for } x \geq 0 \\ \frac{1}{2} \lambda \exp(\lambda x) & \text{for } x < 0 \end{cases}$$

for the parameter $\lambda > 0$. Find the mean and variance of X .

3.20 In Example 5 we used the characteristic function $\Phi(j\nu)$ to calculate the mean of an exponentially distributed random variable X . Continuing with that example, calculate the variance of X and check your result against that found in Problem 3.18.

3.21 The characteristic function of a continuous random variable X , denoted by $\Phi(\nu)$, has some important properties of its own:

a. The transformed version of the random variable X , namely, $aX + b$, has the following characteristic function

$$\mathbb{E}[\exp(j\nu(aX + b))] = \exp(jb\nu) \cdot \Phi_X(a\nu)$$

where a and b are constants.

b. The characteristic function $\Phi(\nu)$ is real if, and only if, the distribution function $F_X(x)$, pertaining to the random variable X , is symmetric.

Prove the validity of these two properties, and demonstrate that property b is satisfied by the two-sided exponential distribution described in Problem 3.19.

- 3.22 Let X and Y be two continuous random variables. One version of the *total expectation theorem* states

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X|Y=y]f_Y(y) dy$$

Justify this theorem.

Inequalities and Theorems

- 3.23 Let X be a continuous random variable that can only assume nonnegative values. The *Markov inequality* states

$$\mathbb{P}[X \geq a] \leq \frac{1}{a}\mathbb{E}[X], \quad a > 0$$

Justify this inequality.

- 3.24 In (3.46) we stated the Chebyshev inequality without proof. Justify this inequality. *Hint:* consider the probability $\mathbb{P}[(X - \mu)^2 \geq \varepsilon^2]$ and then apply the Markov inequality, considered in Problem 3.23, with $a = \varepsilon^2$.

- 3.25 Consider a sequence X_1, X_2, \dots, X_n of independent and identically distributed random variables with mean μ and variance σ^2 . The sample mean of this sequence is defined by

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i$$

The *weak law of large numbers* states

$$\lim_{n \rightarrow \infty} \mathbb{P}[|M_n - \mu| < \varepsilon] = 0 \quad \text{for } \varepsilon > 0$$

Justify this law. *Hint:* use the Chebyshev inequality.

- 3.26 Let event A denote one of the possible outcomes of a random experiment. Suppose that in n independent trials of the experiment the event A occurs n_A times. The ratio

$$M_n = \frac{n_A}{n}$$

is called the *relative frequency* or *empirical frequency* of the event A . Let $p = \mathbb{P}[A]$ denote the probability of the event A . The experiment is said to exhibit “statistical regularity” if the relative frequency M_n is most likely to be within ε of p for large n . Use the weak law of large numbers, considered in Problem 3.25, to justify this statement.

The Gaussian Distribution

- 3.27 In the literature on signaling over additive white Gaussian noise (AWGN) channels, formulas are derived for probabilistic error calculations using the *complementary error function*

$$\operatorname{erfc}(x) = 1 - \frac{1}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$$

Show that the $\operatorname{erfc}(x)$ is related to the Q -function as follows

- $Q(x) = \frac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right)$
- $\operatorname{erfc}(x) = 2Q(\sqrt{2}x)$

- 3.28 Equation (3.58) defines the probability density function of a Gaussian random variable X . Show that the area under this function is unity, in accordance with the normalization property described in (3.59).
- 3.29 Continuing with Problem 3.28, justify the four properties of the Gaussian distribution stated in Section 3.8 without proofs.
- 3.30 a. Show that the characteristic function of a Gaussian random variable X of mean μ_X and variance σ_X^2 is

$$\phi_X(v) = \exp\left(jv\mu_X - \frac{1}{2}v^2\sigma_X^2\right)$$

- b. Using the result of part a, show that the n th central moment of this Gaussian random variable is as follows:

$$\mathbb{E}[(X - \mu_X)^n] = \begin{cases} 1 \times 3 \times 5 \dots (n-1)\sigma_X^n & \text{for } n \text{ even} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

- 3.31 A Gaussian-distributed random variable X of zero mean and variance σ_X^2 is transformed by a piecewise-linear rectifier characterized by the input–output relation (see Figure P3.31):

$$Y = \begin{cases} X, & X \geq 0 \\ 0, & X < 0 \end{cases}$$

The probability density function of the new random variable Y is described by

$$f_Y(y) = \begin{cases} 0, & y < 0 \\ k\delta(y), & y = 0 \\ \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{y^2}{2\sigma_X^2}\right), & y > 0 \end{cases}$$

- a. Explain the physical reasons for the functional form of this result.
- b. Determine the value of the constant k by which the delta function $\delta(y)$ is weighted.

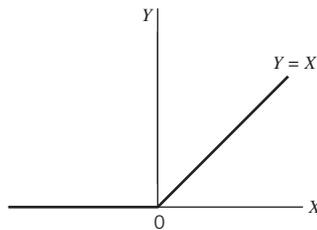


Figure P3.31

- 3.32 In Section 3.9 we stated the central limit theorem embodied in (3.71) without proof. Justify this theorem.

Bayesian Inference

- 3.33 Justify the likelihood principle stated (without proof) in Section 3.11.
- 3.34 In this problem we address a procedure for estimating the mean of the random variable; the procedure was discussed in Section 3.6.

Consider a Gaussian-distributed variable X with unknown mean μ_X and unit variance. The mean μ_X is itself a random variable, uniformly distributed over the interval $[a, b]$. To do the estimation, we are given N independent observations of the random variable X . Justify the estimator of (3.36).

- 3.35 In this problem, we address the issue of estimating the standard deviation σ of a Gaussian-distributed random variable X of zero mean. The standard deviation itself is uniformly distributed inside the interval $[\sigma_1, \sigma_2]$. For the estimation, we have N independent observations of the random variable X , namely, x_1, x_2, \dots, x_N .

- Derive a formula for the estimator $\hat{\sigma}$ using the MAP rule.
- Repeat the estimation using the maximum likelihood criterion.
- Comment on the results of parts a and b.

- 3.36 A binary symbol X is transmitted over a noisy channel. Specifically, symbol $X = 1$ is transmitted with probability p and symbol $X = 0$ is transmitted with probability $(1 - p)$. The received signals at the channel output are defined by

$$Y = X + N$$

The random variable N represents channel noise, modeled as a Gaussian-distributed random variable with zero mean and unit variance. The random variables X and N are independent.

- Describe how the conditional probability $\mathbb{P}[X = 0|Y = y]$ varies with increasing y , all the way from $-\infty$ to $+\infty$.
 - Repeat the problem for the conditional probability $\mathbb{P}[X = 1|Y = y]$.
- 3.37 Consider an experiment involving the Poisson distribution, whose parameter λ is unknown. Given that the distribution of λ follows the exponential law

$$f_n(\lambda) = \begin{cases} a \exp(-a\lambda), & \lambda \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

where $a > 0$, show that the MAP estimate of the parameter λ is given by

$$\hat{\lambda}_{\text{MAP}}(k) = \frac{k}{1 + a}$$

where k is the number of events used in the observation.

- 3.38 In this problem we investigate the use of analytic arguments to justify the optimality of the MAP estimate for the simple case of a one-dimensional parameter vector.

Define the estimation error

$$e_\theta(\mathbf{x}) = \theta - \hat{\theta}(\mathbf{x})$$

where θ is the value of an unknown parameter, $\hat{\theta}(\mathbf{x})$ is the estimator to be optimized, and \mathbf{x} is the observation vector. Figure P3.38 shows a uniform cost function, $C(e)$, for this problem, with zero cost being incurred only when the absolute value of the estimation error $e_\theta(\mathbf{x})$ is less than or equal to $\Delta/2$.

- Formulate the Bayes' risk \mathcal{R} for this parameter estimation problem, accounting for the joint probability density function $f_{\mathbf{A}, \mathbf{X}}(\theta, \mathbf{x})$.
- Hence, determine the MAP estimate $\hat{\theta}_{\text{MAP}}$ by minimizing the risk \mathcal{R} with respect to $\hat{\theta}(\mathbf{x})$. For this minimization, assume that Δ is an arbitrarily small number but nonzero.

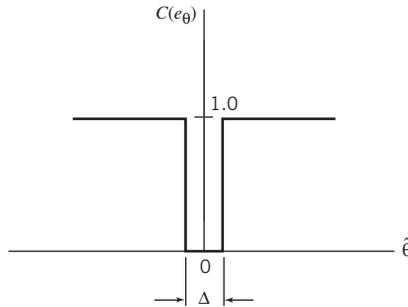


Figure P3.38

- 3.39 In this problem we generalize the likelihood ratio test for simple binary hypotheses by including costs incurred in the decision-making process. Let C_{ij} denote the cost incurred in deciding in favor of hypothesis H_i when hypothesis H_j is true. Hence, show that the likelihood ratio test of (3.95) still holds, except for the fact that the threshold of the test is now defined by

$$\eta = \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})}$$

- 3.40 Consider a binary hypothesis-testing procedure where the two hypotheses H_0 and H_1 are described by different Poisson distributions, characterized by the parameters λ_0 and λ_1 , respectively. The observation is simply a number of events k , depending on whether H_0 or H_1 is true. Specifically, for these two hypotheses, the probability mass functions are defined by

$$p_{X_i}(k) = \frac{(\lambda_i)^k}{k!} \exp(-\lambda_i), \quad k = 0, 1, 2, \dots,$$

where $i = 0$ for hypothesis H_0 and $i = 1$ for hypothesis H_1 . Determine the log-likelihood ratio test for this problem.

- 3.41 Consider the binary hypothesis-testing problem

$$H_1 : X = M + N$$

$$H_0 : X = N$$

The M and N are independent exponentially distributed random variables, as shown by

$$p_M(m) = \begin{cases} \lambda_m \exp(-\lambda_m), & m \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$p_N(n) = \begin{cases} \lambda_n \exp(-\lambda_n), & n \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Determine the likelihood ratio test for this problem.

- 3.42 In this problem we revisit Example 8. But this time we assume that the mean m under hypothesis H_1 is Gaussian distributed, as shown by

$$f_{M|H_1}(m|H_1) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(-\frac{m^2}{2\sigma_m^2}\right)$$

- Derive the likelihood ratio test for the composite hypothesis scenario just described.
- Compare your result with that derived in Example 8.

Notes

1. For a readable account of probability theory, see Bertsekas and Tsitsiklis (2008). For an advanced treatment of probability theory aimed at electrical engineering, see the book by Fine (2006). For an advanced treatment of probability theory, see the two-volume book by Feller (1968, 1971).
2. For an interesting account of inference, see the book by MacKay (2003).
3. For a detailed treatment of the characterization of discrete random variables, see Chapter 2 of the book by Bertsekas and Tsitsiklis (2008).
4. Indeed, we may readily transform the probability density function of (3.58) into the standard form by using the linear transformation

$$Y = \frac{1}{\sigma}(X - \mu)$$

In so doing, (3.58) is simplified as follows:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)$$

which has exactly the same mathematical form as (3.65), except for the use of y in place of x .

5. Calculations based on Bayes' rule, presented previously as (3.14), are referred to as "Bayesian." In actual fact, Bayes provided a continuous version of the rule; see (3.72). In a historical context, it is also of interest to note that the full generality of (3.72) was not actually perceived by Bayes; rather, the task of generalization was left to Laplace.
6. It is because of this duality that the Bayesian paradigm is referred to as a *principle of duality*; see Robert (2001). Robert's book presents a detailed and readable treatment of the Bayesian paradigm. For a more advanced treatment of the subject, see Bernardo and Smith (1998).
7. In a paper published in 1912, R.A. Fisher moved away from the Bayesian approach. Then, in a classic paper published in 1922, he introduced the likelihood.
8. In Appendix B of their book, Bernardo and Smith (1998) show that many non-Bayesian inference procedures do not lead to identical inferences when applied to such proportional likelihoods.
9. For detailed discussion of the sufficient statistic, see Bernardo and Smith (1998).
10. A more detailed treatment of parameter-estimation theory is presented in the classic book by Van Trees (1968); the notation used by Van Trees is somewhat different from that used in this chapter. See also the book by McDonough and Whalen (1995).
11. For a more detailed treatment and readable account of hypothesis testing, see the classic book by Van Trees (1968). See also the book by McDonough and Whalen (1995).

CHAPTER 4

Stochastic Processes

4.1 Introduction

Stated in simple terms, we may say:

A stochastic process is a set of random variables indexed in time.

Elaborating on this succinct statement, we find that in many of the real-life phenomena encountered in practice, *time* features prominently in their description. Moreover, their actual behavior has a random appearance. Referring back to the example of wireless communications briefly described in Section 3.1, we find that the received signal at the wireless channel output varies randomly with time. Processes of this kind are said to be *random* or *stochastic*;¹ hereafter, we will use the term “stochastic.” Although probability theory does not involve time, the study of stochastic processes naturally builds on probability theory.

The way to think about the relationship between probability theory and stochastic processes is as follows. When we consider the statistical characterization of a stochastic process at a particular instant of time, we are basically dealing with the characterization of a *random variable* sampled (i.e., observed) at that instant of time. When, however, we consider a single realization of the process, we have a *random waveform* that evolves across time. The study of stochastic processes, therefore, embodies two approaches: one based on *ensemble averaging* and the other based on *temporal averaging*. Both approaches and their characterizations are considered in this chapter.

Although it is not possible to predict the exact value of a signal drawn from a stochastic process, it is possible to characterize the process in terms of *statistical parameters* such as average power, correlation functions, and power spectra. This chapter is devoted to the mathematical definitions, properties, and measurements of these functions, and related issues.

4.2 Mathematical Definition of a Stochastic Process

To summarize the introduction: stochastic processes have two properties. First, they are functions of time. Second, they are random in the sense that, before conducting an experiment, it is not possible to define the waveforms that will be observed in the future exactly.

In describing a stochastic process, it is convenient to think in terms of a sample space. Specifically, each realization of the process is associated with a *sample point*. The totality of sample points corresponding to the aggregate of all possible realizations of the stochastic process is called the *sample space*. Unlike the sample space in probability

theory, each sample point of the sample space pertaining to a stochastic process is a function of time. We may therefore think of a stochastic process as the sample space or ensemble composed of functions of time. As an integral part of this way of thinking, we assume the existence of a probability distribution defined over an appropriate class of sets in the sample space, so that we may speak with confidence of the probability of various events observed at different points of time.²

Consider, then, a stochastic process specified by

- a. outcomes s observed from some *sample space* S ;
- b. events defined on the sample space S ; and
- c. probabilities of these events.

Suppose that we assign to each sample point s a function of time in accordance with the rule

$$X(t, s), \quad -T \leq t \leq T$$

where $2T$ is the *total observation interval*. For a fixed sample point s_j , the graph of the function $X(t, s_j)$ versus time t is called a *realization* or *sample function* of the stochastic process. To simplify the notation, we denote this sample function as

$$x_j(t) = X(t, s_j), \quad -T \leq t \leq T \quad (4.1)$$

Figure 4.1 illustrates a set of sample functions $\{x_j(t) | j = 1, 2, \dots, n\}$. From this figure, we see that, for a fixed time t_k inside the observation interval, the set of numbers

$$\{x_1(t_k), x_2(t_k), \dots, x_n(t_k)\} = \{X(t_k, s_1), X(t_k, s_2), \dots, X(t_k, s_n)\}$$

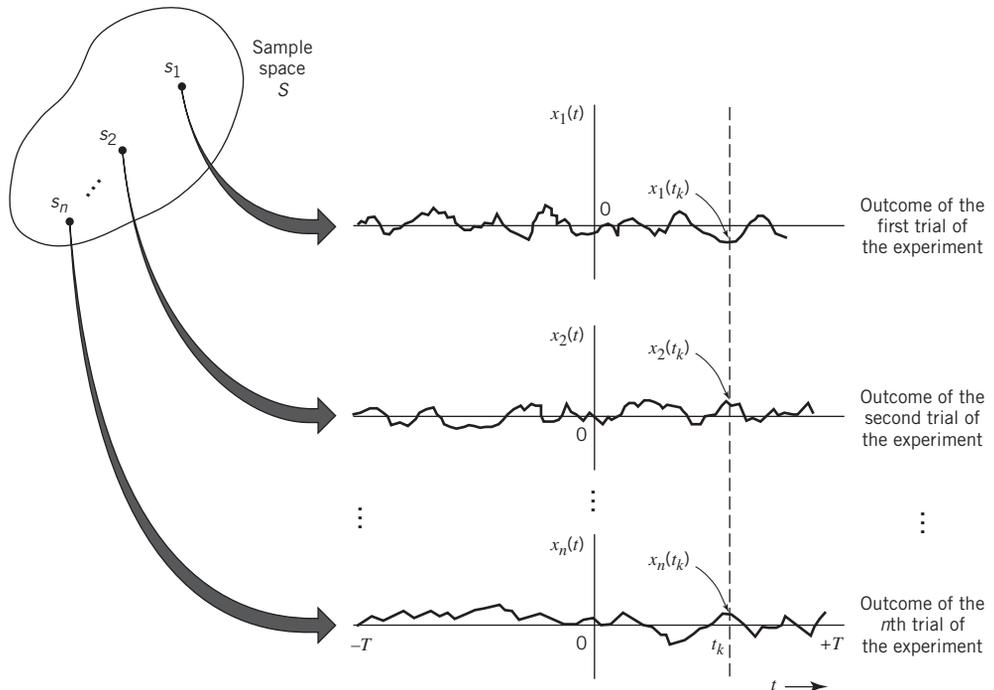


Figure 4.1 An ensemble of sample functions.

constitutes a *random variable*. Thus, a stochastic process $X(t, s)$ is represented by the time-indexed ensemble (family) of random variables $\{X(t, s)\}$. To simplify the notation, the customary practice is to suppress the s and simply use $X(t)$ to denote a stochastic process. We may now formally introduce the definition:

A stochastic process $X(t)$ is an ensemble of time functions, which, together with a probability rule, assigns a probability to any meaningful event associated with an observation of one of the sample functions of the stochastic process.

Moreover, we may distinguish between a random variable and a random process as follows. For a random variable, the outcome of a stochastic experiment is mapped into a number. On the other hand, for a stochastic process, the outcome of a stochastic experiment is mapped into a waveform that is a function of time.

4.3 Two Classes of Stochastic Processes: Strictly Stationary and Weakly Stationary

In dealing with stochastic processes encountered in the real world, we often find that the statistical characterization of a process is independent of the time at which observation of the process is initiated. That is, if such a process is divided into a number of time intervals, the various sections of the process exhibit essentially the same statistical properties. Such a stochastic process is said to be *stationary*. Otherwise, it is said to be *nonstationary*. Generally speaking, we may say:

A stationary process arises from a stable phenomenon that has evolved into a steady-state mode of behavior, whereas a nonstationary process arises from an unstable phenomenon.

To be more precise, consider a stochastic process $X(t)$ that is initiated at $t = -\infty$. Let $X(t_1), X(t_2), \dots, X(t_k)$ denote the random variables obtained by sampling the process $X(t)$ at times t_1, t_2, \dots, t_k , respectively. The joint (cumulative) distribution function of this set of random variables is $F_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k)$. Suppose next we shift all the sampling times by a fixed amount τ denoting the *time shift*, thereby obtaining the new set of random variables: $X(t_1 + \tau), X(t_2 + \tau), \dots, X(t_k + \tau)$. The joint distribution function of this latter set of random variables is $F_{X(t_1 + \tau), \dots, X(t_k + \tau)}(x_1, \dots, x_k)$. The stochastic process $X(t)$ is said to be *stationary in the strict sense*, or *strictly stationary*, if the invariance condition

$$F_{X(t_1 + \tau), \dots, X(t_k + \tau)}(x_1, \dots, x_k) = F_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k) \quad (4.2)$$

holds for all values of time shift τ , all positive integers k , and any possible choice of sampling times t_1, \dots, t_k . In other words, we may state:

A stochastic process $X(t)$, initiated at time $t = -\infty$, is strictly stationary if the joint distribution of any set of random variables obtained by observing the process $X(t)$ is invariant with respect to the location of the origin $t = 0$.

Note that the finite-dimensional distributions in (4.2) depend on the relative time separation between random variables, but not on their absolute time. That is, the stochastic process has the same probabilistic behavior throughout the global time t .

Similarly, we may say that two stochastic processes $X(t)$ and $Y(t)$ are *jointly strictly stationary* if the joint finite-dimensional distributions of the two sets of stochastic variables $X(t_1), \dots, X(t_k)$ and $Y(t'_1), \dots, Y(t'_j)$ are invariant with respect to the origin $t = 0$ for all positive integers k and j , and all choices of the sampling times t_1, \dots, t_k and t'_1, \dots, t'_j .

Returning to (4.2), we may identify two important properties:

1. For $k = 1$, we have

$$F_{X(t)}(x) = F_{X(t+\tau)}(x) = F_X(x) \quad \text{for all } t \text{ and } \tau \quad (4.3)$$

In words, *the first-order distribution function of a strictly stationary stochastic process is independent of time t .*

2. For $k = 2$ and $\tau = -t_2$, we have

$$F_{X(t_1), X(t_2)}(x_1, x_2) = F_{X(0), X(t_1-t_2)}(x_1, x_2) \quad \text{for all } t_1 \text{ and } t_2 \quad (4.4)$$

In words, *the second-order distribution function of a strictly stationary stochastic process depends only on the time difference between the sampling instants and not on the particular times at which the stochastic process is sampled.*

These two properties have profound practical implications for the statistical parameterization of a strictly stationary stochastic process, as discussed in Section 4.4.

EXAMPLE 1

Multiple Spatial Windows for Illustrating Strict Stationarity

Consider Figure 4.2, depicting three spatial windows located at times t_1, t_2, t_3 . We wish to evaluate the probability of obtaining a sample function $x(t)$ of a stochastic process $X(t)$ that passes through this set of windows; that is, the probability of the joint event

$$\mathbb{P}(A) = F_{X(t_1), X(t_2), X(t_3)}(b_1, b_2, b_3) = F_{X(t_1), X(t_2), X(t_3)}(a_1, a_2, a_3)$$

Suppose now the stochastic process $X(t)$ is known to be strictly stationary. An implication of strict stationarity is that the probability of the set of sample functions of this process passing through the windows of Figure 4.3a is equal to the probability of the set of sample functions passing through the corresponding time-shifted windows of Figure 4.3b. Note, however, that it is not necessary that these two sets consist of the same sample functions.

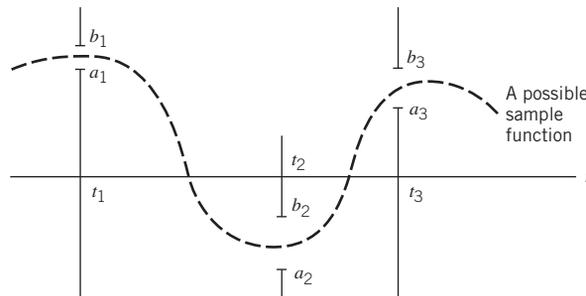
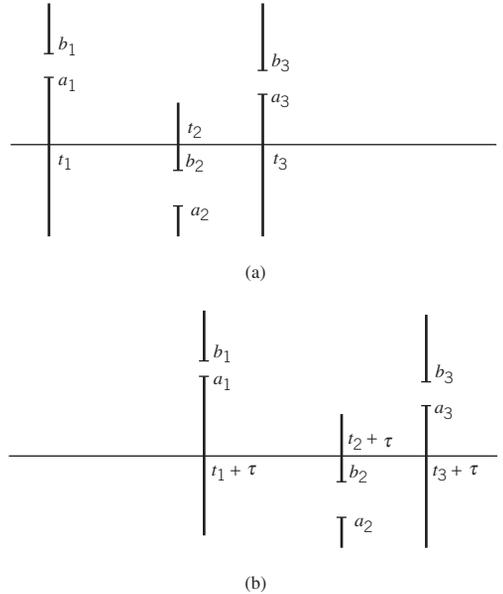


Figure 4.2 Illustrating the probability of a joint event.

Figure 4.3
Illustrating the concept of stationarity in Example 1.



Another important class of stochastic processes is the so-called *weakly stationary processes*. To be specific, a stochastic process $X(t)$ is said to be weakly stationary if its second-order moments satisfy the following two conditions:

1. The mean of the process $X(t)$ is constant for all time t .
2. The autocorrelation function of the process $X(t)$ depends solely on the difference between any two times at which the process is sampled; the “auto” in autocorrelation refers to the correlation of the process with itself.

In this book we focus on weakly stationary processes whose second-order statistics satisfy conditions 1 and 2; both of them are easy to measure and considered to be adequate for practical purposes. Such processes are also referred to as *wide-sense stationary processes* in the literature. Henceforth, both terminologies are used interchangeably.

4.4 Mean, Correlation, and Covariance Functions of Weakly Stationary Processes

Consider a real-valued stochastic process $X(t)$. We define the *mean* of the process $X(t)$ as the expectation of the random variable obtained by sampling the process at some time t , as shown by

$$\begin{aligned}\mu_X(t) &= \mathbb{E}[X(t)] \\ &= \int_{-\infty}^{\infty} x f_{X(t)}(x) dx\end{aligned}\tag{4.5}$$

where $f_{X(t)}(x)$ is the first-order probability density function of the process $X(t)$, observed at time t ; note also that the use of single X as subscript in $\mu_X(t)$ is intended to emphasize the fact that $\mu_X(t)$ is a first-order moment. For the mean $\mu_X(t)$ to be a constant for all time t so that the process $X(t)$ satisfies the first condition of weak stationarity, we require that $f_{X(t)}(x)$ be independent of time t . Consequently, (4.5) simplifies to

$$\mu_X(t) = \mu_X \quad \text{for all } t \quad (4.6)$$

We next define the *autocorrelation function* of the stochastic process $X(t)$ as the expectation of the product of two random variables, $X(t_1)$ and $X(t_2)$, obtained by sampling the process $X(t)$ at times t_1 and t_2 , respectively. Specifically, we write

$$\begin{aligned} M_{XX}(t_1, t_2) &= \mathbb{E}[X(t_1)X(t_2)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X(t_1), X(t_2)}(x_1, x_2) dx_1 dx_2 \end{aligned} \quad (4.7)$$

where $f_{X(t_1), X(t_2)}(x_1, x_2)$ is the joint probability density function of the process $X(t)$ sampled at times t_1 and t_2 ; here, again, note that the use of the double X subscripts is intended to emphasize the fact that $M_{XX}(t_1, t_2)$ is a second-order moment. For $M_{XX}(t_1, t_2)$ to depend only on the time difference $t_2 - t_1$ so that the process $X(t)$ satisfies the second condition of weak stationarity, it is necessary for $f_{X(t_1), X(t_2)}(x_1, x_2)$ to depend only on the time difference $t_2 - t_1$. Consequently, (4.7) reduces to

$$\begin{aligned} M_{XX}(t_1, t_2) &= \mathbb{E}[X(t_1)X(t_2)] \\ &= R_{XX}(t_2 - t_1) \quad \text{for all } t_1 \text{ and } t_2 \end{aligned} \quad (4.8)$$

In (4.8) we have purposely used two different symbols for the autocorrelation function: $M_{XX}(t_1, t_2)$ for any stochastic process $X(t)$ and $R_{XX}(t_2 - t_1)$ for a stochastic process that is weakly stationary.

Similarly, the *autocovariance function* of a weakly stationary process $X(t)$ is defined by

$$\begin{aligned} C_{XX}(t_1, t_2) &= \mathbb{E}[(X(t_1) - \mu_X)(X(t_2) - \mu_X)] \\ &= R_{XX}(t_2 - t_1) - \mu_X^2 \end{aligned} \quad (4.9)$$

Equation (4.9) shows that, like the autocorrelation function, the autocovariance function of a weakly stationary process $X(t)$ depends only on the time difference $(t_2 - t_1)$. This equation also shows that if we know the mean and the autocorrelation function of the process $X(t)$, we can uniquely determine the autocovariance function. The mean and autocorrelation function are therefore sufficient to describe the first two moments of the process.

However, two important points should be carefully noted:

1. The mean and autocorrelation function only provide a *weak description* of the distribution of the stochastic process $X(t)$.
2. The conditions involved in defining (4.6) and (4.8) are *not* sufficient to guarantee the stochastic process $X(t)$ to be strictly stationary, which emphasizes a remark that was made in the preceding section.

Nevertheless, practical considerations often dictate that we simply limit ourselves to a weak description of the process given by the mean and autocorrelation function because the computation of higher order moments can be computationally intractable.

Henceforth, the treatment of stochastic processes is confined to weakly stationary processes, for which the definitions of the second-order moments in (4.6), (4.8), and (4.9) hold.

Properties of the Autocorrelation Function

For convenience of notation, we reformulate the definition of the autocorrelation function of a weakly stationary process $X(t)$, presented in (4.8), as

$$R_{XX}(\tau) = \mathbb{E}[X(t + \tau)X(t)] \quad \text{for all } t \quad (4.10)$$

where τ denotes a *time shift*; that is, $t = t_2$ and $\tau = t_1 - t_2$. This autocorrelation function has several important properties.

PROPERTY 1 Mean-square Value

The mean-square value of a weakly stationary process $X(t)$ is obtained from $R_{XX}(\tau)$ simply by putting $\tau = 0$ in (4.10), as shown by

$$R_{XX}(0) = \mathbb{E}[X^2(t)] \quad (4.11)$$

PROPERTY 2 Symmetry

The autocorrelation function $R_{XX}(\tau)$ of a weakly stationary process $X(t)$ is an even function of the time shift τ ; that is,

$$R_{XX}(\tau) = R_{XX}(-\tau) \quad (4.12)$$

This property follows directly from (4.10). Accordingly, we may also define the autocorrelation function $R_{XX}(\tau)$ as

$$R_{XX}(\tau) = \mathbb{E}[X(t)X(t - \tau)]$$

In words, we may say that a graph of the autocorrelation function $R_{XX}(\tau)$, plotted versus τ , is symmetric about the origin.

PROPERTY 3 Bound on the Autocorrelation Function

The autocorrelation function $R_{XX}(\tau)$ attains its maximum magnitude at $\tau = 0$; that is,

$$|R_{XX}(\tau)| \leq R_{XX}(0) \quad (4.13)$$

To prove this property, consider the nonnegative quantity

$$\mathbb{E}[(X(t + \tau) \pm X(t))^2] \geq 0$$

Expanding terms and taking their individual expectations, we readily find that

$$\mathbb{E}[X^2(t + \tau)] \pm 2\mathbb{E}[X(t + \tau)] + \mathbb{E}[X^2(t)] \geq 0$$

which, in light of (4.11) and (4.12), reduces to

$$2R_{XX}(0) \pm 2R_{XX}(\tau) \geq 0$$

Equivalently, we may write

$$-R_{XX}(0) \leq R_{XX}(\tau) \leq R_{XX}(0)$$

from which (4.13) follows directly.

PROPERTY 4 Normalization

Values of the normalized autocorrelation function

$$\rho_{XX}(\tau) = \frac{R_{XX}(\tau)}{R_{XX}(0)} \quad (4.14)$$

are confined to the range $[-1, 1]$.

This last property follows directly from (4.13).

Physical Significance of the Autocorrelation Function

The autocorrelation function $R_{XX}(\tau)$ is significant because it provides a means of describing the interdependence of two random variables obtained by sampling the stochastic process $X(t)$ at times τ seconds apart. It is apparent, therefore, that the more rapidly the stochastic process $X(t)$ changes with time, the more rapidly will the autocorrelation function $R_{XX}(\tau)$ decrease from its maximum $R_{XX}(0)$ as τ increases, as illustrated in Figure 4.4. This behavior of the autocorrelation function may be characterized by a *decorrelation time* τ_{dec} , such that, for $\tau > \tau_{\text{dec}}$, the magnitude of the autocorrelation function $R_{XX}(\tau)$ remains below some prescribed value. We may thus introduce the following definition:

The decorrelation time τ_{dec} of a weakly stationary process $X(t)$ of zero mean is the time taken for the magnitude of the autocorrelation function $R_{XX}(t)$ to decrease, for example, to 1% of its maximum value $R_{XX}(0)$.

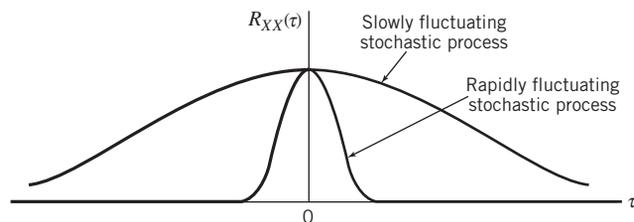
For the example used in this definition, the parameter τ_{dec} is referred to as the *one-percent decorrelation time*.

EXAMPLE 2 Sinusoidal Wave with Random Phase

Consider a sinusoidal signal with random phase, defined by

$$X(t) = A \cos(2\pi f_c t + \Theta) \quad (4.15)$$

Figure 4.4
Illustrating the autocorrelation functions of slowly and rapidly fluctuating stochastic processes.



where A and f_c are constants and Θ is a random variable that is *uniformly distributed* over the interval $[-\pi, \pi]$; that is,

$$f_{\Theta}(\theta) = \begin{cases} \frac{1}{2\pi}, & -\pi \leq \theta \leq \pi \\ 0, & \text{elsewhere} \end{cases} \quad (4.16)$$

According to (4.16), the random variable Θ is equally likely to have any value θ in the interval $[-\pi, \pi]$. Each value of θ corresponds to a point in the sample space S of the stochastic process $X(t)$.

The process $X(t)$ defined by (4.15) and (4.16) may represent a locally generated carrier in the receiver of a communication system, which is used in the demodulation of a received signal. In such an application, the random variable Θ in (4.15) accounts for uncertainties experienced in the course of signal transmission across the communication channel.

The autocorrelation function of $X(t)$ is

$$\begin{aligned} R_{XX}(\tau) &= \mathbb{E}[X(t + \tau)X(t)] \\ &= \mathbb{E}[A^2 \cos(2\pi f_c t + 2\pi f_c \tau + \Theta) \cos(2\pi f_c t + \Theta)] \\ &= \frac{A^2}{2} \mathbb{E}[\cos(4\pi f_c t + 2\pi f_c \tau + 2\Theta)] + \frac{A^2}{2} \mathbb{E}[\cos(2\pi f_c \tau)] \\ &= \frac{A^2}{2} \int_{-\pi}^{\pi} \cos(4\pi f_c t + 2\pi f_c \tau + 2\theta) d\theta + \frac{A^2}{2} \cos(2\pi f_c \tau) \end{aligned}$$

The first term integrates to zero, so we simply have

$$R_{XX}(\tau) = \frac{A^2}{2} \cos(2\pi f_c \tau) \quad (4.17)$$

which is plotted in Figure 4.5. From this figure we see that the autocorrelation function of a sinusoidal wave with random phase is another sinusoid at the same frequency in the “local time domain” denoted by the time shift τ rather than the global time domain denoted by t .

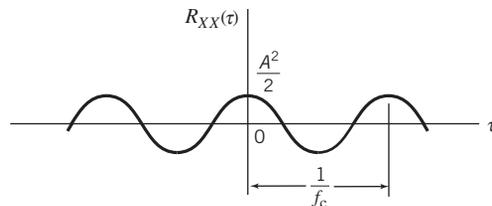


Figure 4.5 Autocorrelation function of a sine wave with random phase.

EXAMPLE 3 Random Binary Wave

Figure 4.6 shows the sample function $x(t)$ of a weakly stationary process $X(t)$ consisting of a random sequence of *binary symbols* 1 and 0. Three assumptions are made:

1. The symbols 1 and 0 are represented by pulses of amplitude $+A$ and $-A$ volts respectively and duration T seconds.
2. The pulses are not synchronized, so the starting time t_d of the first complete pulse for positive time is equally likely to lie anywhere between zero and T seconds. That is, t_d is the sample value of a uniformly distributed random variable T_d , whose probability density function is defined by

$$f_{T_d}(t_d) = \begin{cases} \frac{1}{T}, & 0 \leq t_d \leq T \\ 0, & \text{elsewhere} \end{cases}$$

3. During any time interval $(n-1)T < t - t_d < nT$, where n is a positive integer, the presence of a 1 or a 0 is determined by tossing a fair coin. Specifically, if the outcome is heads, we have a 1; if the outcome is tails, we have a 0. These two symbols are thus equally likely, and the presence of a 1 or 0 in any one interval is independent of all other intervals.

Since the amplitude levels $-A$ and $+A$ occur with equal probability, it follows immediately that $\mathbb{E}[X(t)] = 0$ for all t and the mean of the process is therefore zero.

To find the autocorrelation function $R_{XX}(t_k, t_i)$, we have to evaluate the expectation $\mathbb{E}[X(t_k)X(t_i)]$, where $X(t_k)$ and $X(t_i)$ are random variables obtained by sampling the stochastic process $X(t)$ at times t_k and t_i respectively. To proceed further, we need to consider two distinct conditions:

Condition 1: $|t_k - t_i| > T$

Under this condition, the random variables $X(t_k)$ and $X(t_i)$ occur in different pulse intervals and are therefore independent. We thus have

$$\mathbb{E}[X(t_k)X(t_i)] = \mathbb{E}[X(t_k)]\mathbb{E}[X(t_i)] = 0, \quad |t_k - t_i| > T$$

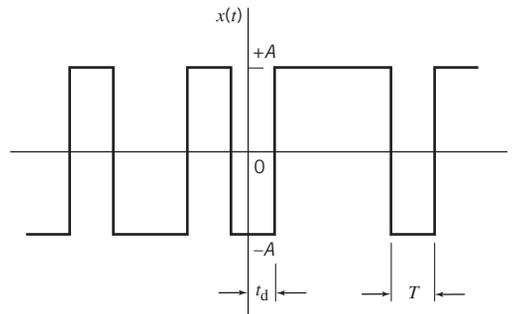


Figure 4.6 Sample function of random binary wave.

Condition 2: $|t_k - t_i| > T$, with $t_k = 0$ and $t_i < t_k$

Under this second condition, we observe from Figure 4.6 that the random variables $X(t_k)$ and $X(t_i)$ occur in the same pulse interval if, and only if, the delay t_d satisfies the condition $t_d < T - |t_k - t_i|$. We thus have the conditional expectation

$$\mathbb{E}[X(t_k)X(t_i)|t_d] = \begin{cases} A^2, & t_d < T - |t_k - t_i| \\ 0, & \text{elsewhere} \end{cases}$$

Averaging this result over all possible values of t_d , we get

$$\begin{aligned} \mathbb{E}[X(t_k)X(t_i)] &= \int_0^{T-|t_k-t_i|} A^2 f_{T_d}(t_d) dt_d \\ &= \int_0^{T-|t_k-t_i|} \frac{A^2}{T} dt_d \\ &= A^2 \left(1 - \frac{|t_k - t_i|}{T}\right), \quad |t_k - t_i| < T \end{aligned}$$

By similar reasoning for any other value of t_k , we conclude that the autocorrelation function of a random binary wave, represented by the sample function shown in Figure 4.6, is only a function of the time difference $\tau = t_k - t_i$, as shown by

$$R_{XX}(\tau) = \begin{cases} A^2 \left(1 - \frac{|\tau|}{T}\right), & |\tau| < T \\ 0, & |\tau| \geq T \end{cases} \quad (4.18)$$

This triangular result, described in (4.18), is plotted in Figure 4.7.

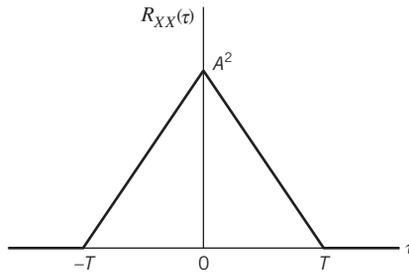


Figure 4.7 Autocorrelation function of random binary wave.

Cross-correlation Functions

Consider next the more general case of two stochastic processes $X(t)$ and $Y(t)$ with autocorrelation functions $M_{XX}(t, u)$ and $M_{YY}(t, u)$ respectively. There are two possible *cross-correlation functions* of $X(t)$ and $Y(t)$ to be considered.

Specifically, we have

$$M_{XY}(t, u) = \mathbb{E}[X(t)Y(u)] \quad (4.19)$$

and

$$M_{YX}(t, u) = \mathbb{E}[Y(t)X(u)] \quad (4.20)$$

where t and u denote two values of the global time at which the processes are observed. All four correlation parameters of the two stochastic processes $X(t)$ and $Y(t)$ may now be displayed conveniently in the form of the two-by-two matrix

$$\mathbf{M}(t, u) = \begin{bmatrix} M_{XX}(t, u) & M_{XY}(t, u) \\ M_{YX}(t, u) & M_{YY}(t, u) \end{bmatrix}$$

which is called the *cross-correlation matrix* of the stochastic processes $X(t)$ and $Y(t)$. If the stochastic processes $X(t)$ and $Y(t)$ are each weakly stationary and, in addition, they are jointly stationary, then the correlation matrix can be expressed by

$$\mathbf{R}(\tau) = \begin{bmatrix} R_{XX}(\tau) & R_{XY}(\tau) \\ R_{YX}(\tau) & R_{YY}(\tau) \end{bmatrix} \quad (4.21)$$

where the time shift $\tau = u - t$.

In general, the cross-correlation function is *not* an even function of the time-shift τ as was true for the autocorrelation function, nor does it have a maximum at the origin. However, it does obey a certain symmetry relationship, described by

$$R_{XY}(\tau) = R_{YX}(-\tau) \quad (4.22)$$

EXAMPLE 4

Quadrature-Modulated Processes

Consider a pair of quadrature-modulated processes $X_1(t)$ and $X_2(t)$ that are respectively related to a weakly stationary process $X(t)$ as follows:

$$X_1(t) = X(t) \cos(2\pi f_c t + \Theta)$$

$$X_2(t) = X(t) \sin(2\pi f_c t + \Theta)$$

where f_c is a carrier frequency and the random variable Θ is uniformly distributed over the interval $[0, 2\pi]$. Moreover, Θ is independent of $X(t)$. One cross-correlation function of $X_1(t)$ and $X_2(t)$ is given by

$$\begin{aligned} R_{12}(\tau) &= \mathbb{E}[X_1(t)X_2(t-\tau)] \\ &= \mathbb{E}[X(t)X(t-\tau) \cos(2\pi f_c t + \Theta) \sin(2\pi f_c t - 2\pi f_c \tau + \Theta)] \\ &= \mathbb{E}[X(t)X(t-\tau)] \mathbb{E}[\cos(2\pi f_c t + \Theta) \sin(2\pi f_c t - 2\pi f_c \tau + \Theta)] \\ &= \frac{1}{2} R_{XX}(\tau) \mathbb{E}[\sin(4\pi f_c \tau - 2\pi f_c t + 2\Theta) - \sin(2\pi f_c \tau)] \\ &= -\frac{1}{2} R_{XX}(\tau) \sin(2\pi f_c \tau) \end{aligned} \quad (4.23)$$

where, in the last line, we have made use of the uniform distribution of the random variable Θ , representing phase. Invoking (4.22), we find that the other cross-correlation function of $X_1(t)$ and $X_2(t)$ is given by

$$\begin{aligned} R_{21}(\tau) &= \frac{1}{2} R_{XX}(-\tau) \sin(2\pi f_c \tau) \\ &= \frac{1}{2} R_{XX}(\tau) \sin(2\pi f_c \tau) \end{aligned}$$

At $\tau = 0$, the factor $\sin(2\pi f_c \tau)$ is zero, in which case we have

$$R_{12}(0) = R_{21}(0) = 0$$

This result shows that the random variables obtained by simultaneously sampling the quadrature-modulated processes $X_1(t)$ and $X_2(t)$ at some fixed value of time t are orthogonal to each other.

4.5 Ergodic Processes

Ergodic processes are subsets of weakly stationary processes. Most importantly, from a practical perspective, the *property of ergodicity* permits us to substitute time averages for ensemble averages.

To elaborate on these two succinct statements, we know that the expectations or ensemble averages of a stochastic process $X(t)$ are averages “across the process.” For example, the mean of a stochastic process $X(t)$ at some fixed time t_k is the expectation of the random variable $X(t_k)$ that describes *all possible values* of sample functions of the process $X(t)$ sampled at time $t = t_k$. Naturally, we may also define *long-term sample averages* or *time averages* that are averages “along the process.” Whereas in ensemble averaging we consider a set of independent realizations of the process $X(t)$ sampled at some fixed time t_k , in time averaging we focus on a single waveform evolving across time t and representing one waveform realization of the process $X(t)$.

With time averages providing the basis of a practical method for possible *estimation* of ensemble averages of a stochastic process, we would like to explore the conditions under which this estimation is justifiable. To address this important issue, consider the sample function $x(t)$ of a weakly stationary process $X(t)$ observed over the interval $-T \leq t \leq T$. The time-average value of the sample function $x(t)$ is defined by the definite integral

$$\mu_x(T) = \frac{1}{2T} \int_{-T}^T x(t) dt \quad (4.24)$$

Clearly, the time average $\mu_x(T)$ is a random variable, as its value depends on the observation interval and which particular sample function of the process $X(t)$ is picked for use in (4.24). Since the process $X(t)$ is assumed to be weakly stationary, the mean of the time average $\mu_x(T)$ is given by (after interchanging the operations of expectation and integration, which is permissible because both operations are linear)

$$\begin{aligned}
\mathbb{E}[\mu_x(T)] &= \frac{1}{2T} \int_{-T}^T \mathbb{E}[x(t)] dt \\
&= \frac{1}{2T} \int_{-T}^T \mu_X dt \\
&= \mu_X
\end{aligned} \tag{4.25}$$

where μ_X is the mean of the process $X(t)$. Accordingly, the time average $\mu_x(T)$ represents an *unbiased* estimate of the ensemble-averaged mean μ_X . Most importantly, we say that the process $X(t)$ is *ergodic in the mean* if two conditions are satisfied:

1. The time average $\mu_x(T)$ approaches the ensemble average μ_X in the limit as the observation interval approaches infinity; that is,

$$\lim_{T \rightarrow \infty} \mu_x(T) = \mu_X$$

2. The variance of $\mu_x(T)$, treated as a random variable, approaches zero in the limit as the observation interval approaches infinity; that is,

$$\lim_{T \rightarrow \infty} \text{var}[\mu_x(T)] = 0$$

The other time average of particular interest is the autocorrelation function $R_{xx}(\tau, T)$, defined in terms of the sample function $x(t)$ observed over the interval $-T \leq t \leq T$. Following (4.24), we may formally define the *time-averaged autocorrelation function* of $x(t)$ as

$$R_{xx}(\tau, T) = \frac{1}{2T} \int_{-T}^T x(t + \tau)x(t) dt \tag{4.26}$$

This second time average should also be viewed as a random variable with a mean and variance of its own. In a manner similar to ergodicity of the mean, we say that the process $x(t)$ is *ergodic in the autocorrelation function* if the following two limiting conditions are satisfied:

$$\lim_{T \rightarrow \infty} R_{xx}(\tau, T) = R_{XX}(\tau)$$

$$\lim_{T \rightarrow \infty} \text{var}[R_{xx}(\tau, T)] = 0$$

With the property of ergodicity confined to the mean and autocorrelation functions, it follows that ergodic processes are subsets of weakly stationary processes. In other words, all ergodic processes are weakly stationary; however, the converse is not necessarily true.

4.6 Transmission of a Weakly Stationary Process through a Linear Time-invariant Filter

Suppose that a stochastic process $X(t)$ is applied as input to a linear time-invariant filter of impulse response $h(t)$, producing a new stochastic process $Y(t)$ at the filter output, as depicted in Figure 4.8. In general, it is difficult to describe the probability distribution of the output stochastic process $Y(t)$, even when the probability distribution of the input stochastic process $X(t)$ is completely specified for the entire time interval $-\infty < t < \infty$.

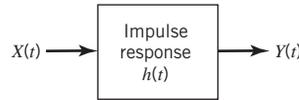


Figure 4.8 Transmission of a stochastic process through a linear time-invariant filter.

For the sake of mathematical tractability, we limit the discussion in this section to the time-domain form of the input–output relations of the filter for defining the mean and autocorrelation functions of the output stochastic process $Y(t)$ in terms of those of the input $X(t)$, assuming that $X(t)$ is a weakly stationary process.

The transmission of a process through a linear time-invariant filter is governed by the *convolution integral*, which was discussed in Chapter 2. For the problem at hand, we may thus express the output stochastic process $Y(t)$ in terms of the input stochastic process $X(t)$ as

$$Y(t) = \int_{-\infty}^{\infty} h(\tau_1)X(t - \tau_1) d\tau_1$$

where τ_1 is a local time. Hence, the mean of $Y(t)$ is

$$\begin{aligned} \mu_Y(t) &= \mathbb{E}[Y(t)] \\ &= \mathbb{E}\left[\int_{-\infty}^{\infty} h(\tau_1)X(t - \tau_1) d\tau_1\right] \end{aligned} \quad (4.27)$$

Provided that the expectation $\mathbb{E}[X(t)]$ is finite for all t and the filter is stable, we may interchange the order of expectation and integration in (4.27), in which case we obtain

$$\begin{aligned} \mu_Y(t) &= \int_{-\infty}^{\infty} h(\tau_1)\mathbb{E}[X(t - \tau_1)] d\tau_1 \\ &= \int_{-\infty}^{\infty} h(\tau_1)\mu_X(t - \tau_1) d\tau_1 \end{aligned} \quad (4.28)$$

When the input stochastic process $X(t)$ is weakly stationary, the mean $\mu_X(t)$ is a constant μ_X ; therefore, we may simplify (4.28) as

$$\begin{aligned} \mu_Y &= \mu_X \int_{-\infty}^{\infty} h(\tau_1) d\tau_1 \\ &= \mu_X H(0) \end{aligned} \quad (4.29)$$

where $H(0)$ is the zero-frequency response of the system. Equation (4.29) states:

The mean of the stochastic process $Y(t)$ produced at the output of a linear time-invariant filter in response to a weakly stationary process $X(t)$, acting as the input process, is equal to the mean of $X(t)$ multiplied by the zero-frequency response of the filter.

This result is intuitively satisfying.

Consider next the autocorrelation function of the output stochastic process $Y(t)$. By definition, we have

$$M_{YY}(t, u) = \mathbb{E}[Y(t)Y(u)]$$

where t and u denote two values of the time at which the output process $Y(t)$ is sampled. We may therefore apply the convolution integral twice to write

$$M_{YY}(t, u) = \mathbb{E}\left[\int_{-\infty}^{\infty} h(\tau_1)X(t-\tau_1) d\tau_1 \int_{-\infty}^{\infty} h(\tau_2)X(u-\tau_2) d\tau_2\right] \quad (4.30)$$

Here again, provided that the mean-square value $\mathbb{E}[X^2(t)]$ is finite for all t and the filter is stable, we may interchange the order of the expectation and the integrations with respect to τ_1 and τ_2 in (4.30), obtaining

$$\begin{aligned} M_{YY}(t, u) &= \int_{-\infty}^{\infty} \left[h(\tau_1) \int_{-\infty}^{\infty} d\tau_2 h(\tau_2) \mathbb{E}[X(t-\tau_1)X(u-\tau_2)] \right] d\tau_1 \\ &= \int_{-\infty}^{\infty} \left[h(\tau_1) \int_{-\infty}^{\infty} d\tau_2 h(\tau_2) M_{XX}(t-\tau_1, u-\tau_2) \right] d\tau_1 \end{aligned} \quad (4.31)$$

When the input $X(t)$ is a weakly stationary process, the autocorrelation function of $X(t)$ is only a function of the difference between the sampling times $t-\tau_1$ and $u-\tau_2$. Thus, putting $\tau = u-t$ in (4.31), we may go on to write

$$R_{YY}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_{XX}(\tau+\tau_1-\tau_2) d\tau_1 d\tau_2 \quad (4.32)$$

which depends only on the time difference τ .

On combining the result of (4.32) with that involving the mean μ_Y in (4.29), we may now make the following statement:

If the input to a stable linear time-invariant filter is a weakly stationary process, then the output of the filter is also a weakly stationary process.

By definition, we have $R_{YY}(0) = \mathbb{E}[Y^2(t)]$. In light of Property 1 of the autocorrelation function $R_{YY}(\tau)$, it follows, therefore, that the *mean-square value* of the output process $Y(t)$ is obtained by putting $\tau=0$ in (4.32), as shown by

$$\mathbb{E}[Y^2(t)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_{XX}(\tau_1-\tau_2) d\tau_1 d\tau_2 \quad (4.33)$$

which, of course, is a constant.

4.7 Power Spectral Density of a Weakly Stationary Process

Thus far we have considered the time-domain characterization of a weakly stationary process applied to a linear filter. We next study the characterization of linearly filtered weakly stationary processes by using frequency-domain ideas. In particular, we wish to derive the frequency-domain equivalent to the result of (4.33), defining the mean-square value of the filter output $Y(t)$. The term “filter” used here should be viewed in a generic sense; for example, it may represent the channel of a communication system.

From Chapter 2, we recall that the impulse response of a linear time-invariant filter is equal to the inverse Fourier transform of the frequency response of the filter. Using $H(f)$ to denote the *frequency response* of the filter, we may thus write

$$h(\tau_1) = \int_{-\infty}^{\infty} H(f) \exp(j2\pi f \tau_1) df \quad (4.34)$$

Substituting this expression for $h(\tau_1)$ into (4.33) and then changing the order of integrations, we get the triple integral

$$\begin{aligned} \mathbb{E}[Y^2(t)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} H(f) \exp(j2\pi f \tau_1) df \right] h(\tau_2) R_{XX}(\tau_1 - \tau_2) d\tau_1 d\tau_2 \\ &= \int_{-\infty}^{\infty} \left[H(f) \int_{-\infty}^{\infty} d\tau_2 h(\tau_2) \int_{-\infty}^{\infty} R_{XX}(\tau_1 - \tau_2) \exp(j2\pi f \tau_1) d\tau_1 \right] df \end{aligned} \quad (4.35)$$

At first, the expression on the right-hand side of (4.35) looks rather overwhelming. However, we may simplify it considerably by first introducing the variable

$$\tau = \tau_1 - \tau_2$$

Then, we may rewrite (4.35) in the new form

$$\mathbb{E}[Y^2(t)] = \int_{-\infty}^{\infty} H(f) \left[\int_{-\infty}^{\infty} h(\tau_2) \exp(j2\pi f \tau_2) d\tau_2 \int_{-\infty}^{\infty} R_{XX}(\tau) \exp(-j2\pi f \tau) d\tau \right] df \quad (4.36)$$

The middle integral involving the variable τ_2 inside the square brackets on the right-hand side in (4.36) is simply $H^*(f)$, the complex conjugate of the frequency response of the filter. Hence, using $|H(f)|^2 = H(f)H^*(f)$, where $|H(f)|$ is the *magnitude response* of the filter, we may simplify (4.36) as

$$\mathbb{E}[Y^2(t)] = \int_{-\infty}^{\infty} |H(f)|^2 \left[\int_{-\infty}^{\infty} R_{XX}(\tau) \exp(-j2\pi f \tau) d\tau \right] df \quad (4.37)$$

We may further simplify (4.37) by recognizing that the integral inside the square brackets in this equation with respect to the variable τ is simply the Fourier transform of the autocorrelation function $R_{XX}(\tau)$ of the input process $X(t)$. In particular, we may now define a new function

$$S_{XX}(f) = \int_{-\infty}^{\infty} R_{XX}(\tau) \exp(-j2\pi f \tau) d\tau \quad (4.38)$$

The new function $S_{XX}(f)$ is called the *power spectral density*, or *power spectrum*, of the weakly stationary process $X(t)$. Thus, substituting (4.38) into (4.37), we obtain the simple formula

$$\mathbb{E}[Y^2(t)] = \int_{-\infty}^{\infty} |H(f)|^2 S_{XX}(f) df \quad (4.39)$$

which is the desired frequency-domain equivalent to the time-domain relation of (4.33). In words, (4.39) states:

The mean-square value of the output of a stable linear time-invariant filter in response to a weakly stationary process is equal to the integral over all

frequencies of the power spectral density of the input process multiplied by the squared magnitude response of the filter.

Physical Significance of the Power Spectral Density

To investigate the physical significance of the power spectral density, suppose that the weakly stationary process $X(t)$ is passed through an ideal narrowband filter with a magnitude response $|H(f)|$ centered about the frequency f_c , depicted in Figure 4.9; we may thus write

$$|H(f)| = \begin{cases} 1, & |f \pm f_c| < \frac{1}{2}\Delta f \\ 0, & |f \pm f_c| > \frac{1}{2}\Delta f \end{cases} \quad (4.40)$$

where Δf is the *bandwidth* of the filter. From (4.39) we readily find that if the bandwidth Δf is made sufficiently small compared with the midband frequency f_c of the filter and $S_{XX}(f)$ is a continuous function of the frequency f , then the mean-square value of the filter output is approximately given by

$$\mathbb{E}[Y^2(t)] \approx (2\Delta f)S_{XX}(f) \quad \text{for all } f \quad (4.41)$$

where, for the sake of generality, we have used f in place of f_c . According to (4.41), however, the filter passes only those frequency components of the input random process $X(t)$ that lie inside the narrow frequency band of width Δf . We may, therefore, say that $S_X(f)$ represents the density of the average power in the weakly stationary process $X(t)$, evaluated at the frequency f . The power spectral density is therefore measured in *watts per hertz* (W/Hz).

The Wiener–Khinchine Relations

According to (4.38), the power spectral density $S_{XX}(f)$ of a weakly stationary process $X(t)$ is the Fourier transform of its autocorrelation function $R_{XX}(\tau)$. Building on what we know about Fourier theory from Chapter 2, we may go on to say that the autocorrelation function $R_{XX}(\tau)$ is the inverse Fourier transform of the power spectral density $S_{XX}(f)$.

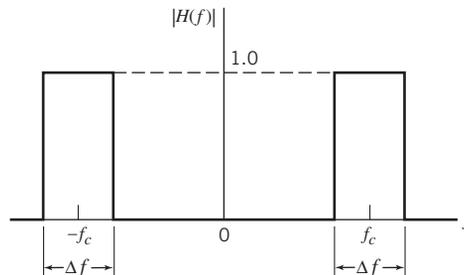


Figure 4.9 Magnitude response of ideal narrowband filter.

Simply put, $R_{XX}(\tau)$ and $S_{XX}(f)$ form a *Fourier-transform pair*, as shown by the following pair of related equations:

$$S_{XX}(f) = \int_{-\infty}^{\infty} R_{XX}(\tau) \exp(-j2\pi f\tau) d\tau \quad (4.42)$$

$$R_{XX}(\tau) = \int_{-\infty}^{\infty} S_{XX}(f) \exp(j2\pi f\tau) df \quad (4.43)$$

These two equations are known as the *Wiener–Khinchine relations*,³ which play a fundamental role in the spectral analysis of weakly stationary processes.

The Wiener–Khinchine relations show that if either the autocorrelation function or power spectral density of a weakly stationary process is known, then the other can be found exactly. Naturally, these functions display different aspects of correlation-related information about the process. Nevertheless, it is commonly accepted that, for practical purposes, the power spectral density is the more useful function of the two for reasons that will become apparent as we progress forward in this chapter and the rest of the book.

Properties of the Power Spectral Density

PROPERTY 1 Zero Correlation among Frequency Components

The individual frequency components of the power spectral density $S_{XX}(f)$ of a weakly stationary process $X(t)$ are uncorrelated with each other.

To justify this property, consider Figure 4.10, which shows two adjacent narrow bands of the power spectral density $S_{XX}(f)$, with the width of each band being denoted by Δf . From this figure, we see that there is no overlap, and therefore no correlation, between the contents of these two bands. As Δf approaches zero, the two narrow bands will correspondingly evolve into two adjacent frequency components of $S_{XX}(f)$, remaining uncorrelated with each other. This important property of the power spectral density $S_{XX}(f)$ is attributed to the weak stationarity assumption of the stochastic process $X(t)$.

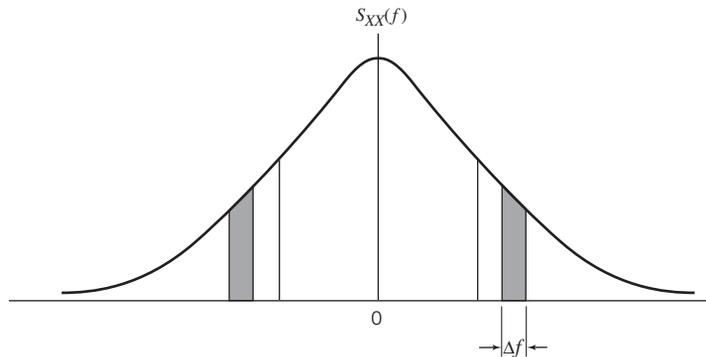


Figure 4.10 Illustration of zero correlation between two adjacent narrow bands of an example power spectral density.

PROPERTY 2 Zero-frequency Value of Power Spectral Density

The zero-frequency value of the power spectral density of a weakly stationary process equals the total area under the graph of the autocorrelation function; that is,

$$S_{XX}(0) = \int_{-\infty}^{\infty} R_{XX}(\tau) d\tau \quad (4.44)$$

This second property follows directly from (4.42) by putting $f = 0$.

PROPERTY 3 Mean-square Value of Stationary Process

The mean-square value of a weakly stationary process $X(t)$ equals the total area under the graph of the power spectral density of the process; that is,

$$\mathbb{E}[X^2(t)] = \int_{-\infty}^{\infty} S_{XX}(f) df \quad (4.45)$$

This third property follows directly from (4.43) by putting $\tau = 0$ and using Property 1 of the autocorrelation function described in (4.11) namely $R_X(0) = \mathbb{E}[X^2(t)]$ for all t .

PROPERTY 4 Nonnegativeness of Power Spectral Density

The power spectral density of a stationary process $X(t)$ is always nonnegative; that is,

$$S_{XX}(f) \geq 0 \quad \text{for all } f \quad (4.46)$$

This property is an immediate consequence of the fact that, since the mean-square value $\mathbb{E}[Y^2(t)]$ is always nonnegative in accordance with (4.41), it follows that $S_{XX}(f) \approx \mathbb{E}[Y^2(t)] / (2\Delta f)$ must also be nonnegative.

PROPERTY 5 Symmetry

The power spectral density of a real-valued weakly stationary process is an even function of frequency; that is,

$$S_{XX}(-f) = S_{XX}(f) \quad (4.47)$$

This property is readily obtained by first substituting $-f$ for the variable f in (4.42):

$$S_{XX}(-f) = \int_{-\infty}^{\infty} R_{XX}(\tau) \exp(j2\pi f\tau) d\tau$$

Next, substituting $-\tau$ for τ , and recognizing that $R_{XX}(-\tau) = R_{XX}(\tau)$ in accordance with Property 2 of the autocorrelation function described in (4.12), we get

$$S_{XX}(-f) = \int_{-\infty}^{\infty} R_{XX}(\tau) \exp(-j2\pi f\tau) d\tau = S_{XX}(f)$$

which is the desired result. It follows, therefore, that the graph of the power spectral density $S_{XX}(f)$, plotted versus frequency f , is symmetric about the origin.

PROPERTY 6 Normalization

The power spectral density, appropriately normalized, has the properties associated with a probability density function in probability theory.

The normalization we have in mind here is with respect to the total area under the graph of the power spectral density (i.e., the mean-square value of the process). Consider then the function

$$p_{XX}(f) = \frac{S_{XX}(f)}{\int_{-\infty}^{\infty} S_{XX}(f) df} \quad (4.48)$$

In light of Properties 3 and 4, we note that $p_{XX}(f) \geq 0$ for all f . Moreover, the total area under the function $p_{XX}(f)$ is unity. Hence, the normalized power spectral density, as defined in (4.48), behaves in a manner similar to a probability density function.

Building on Property 6, we may go on to define the *spectral distribution function* of a weakly stationary process $X(t)$ as

$$F_{XX}(f) = \int_{-\infty}^f p_{XX}(\nu) d\nu \quad (4.49)$$

which has the following properties:

1. $F_{XX}(-\infty) = 0$
2. $F_{XX}(\infty) = 1$
3. $F_{XX}(f)$ is a nondecreasing function of the frequency f .

Conversely, we may state that every nondecreasing and bounded function $F_{XX}(f)$ is the spectral distribution function of a weakly stationary process.

Just as important, we may also state that the spectral distribution function $F_{XX}(f)$ has all the properties of the cumulative distribution function in probability theory, discussed in Chapter 3.

EXAMPLE 5 Sinusoidal Wave with Random Phase (continued)

Consider the stochastic process $X(t) = A\cos(2\pi f_c t + \Theta)$, where Θ is a uniformly distributed random variable over the interval $[-\pi, \pi]$. The autocorrelation function of this stochastic process is given by (4.17), which is reproduced here for convenience:

$$R_{XX}(\tau) = \frac{A^2}{2} \cos(2\pi f_c \tau)$$

Let $\delta(f)$ denote the delta function at $f=0$. Taking the Fourier transform of both sides of the formula defining $R_{XX}(\tau)$, we find that the power spectral density of the sinusoidal process $X(t)$ is

$$S_{XX}(f) = \frac{A^2}{4} [\delta(f-f_c) + \delta(f+f_c)] \quad (4.50)$$

which consists of a pair of delta functions weighted by the factor $A^2/4$ and located at $\pm f_c$, as illustrated in Figure 4.11. Since the total area under a delta function is one, it follows that the total area under $S_{XX}(f)$ is equal to $A^2/2$, as expected.

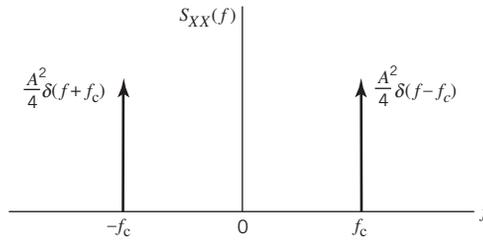


Figure 4.11 Power spectral density of sine wave with random phase; $\delta(f)$ denotes the delta function at $f=0$.

EXAMPLE 6 Random Binary Wave (continued)

Consider again a random binary wave consisting of a sequence of 1s and 0s represented by the values $+A$ and $-A$ respectively. In Example 3 we showed that the autocorrelation function of this random process has the triangular form

$$R_{XX}(\tau) = \begin{cases} A^2 \left(1 - \frac{|\tau|}{T}\right), & |\tau| < T \\ 0, & |\tau| \geq T \end{cases}$$

The power spectral density of the process is therefore

$$S_{XX}(f) = \int_{-T}^T A^2 \left(1 - \frac{|\tau|}{T}\right) \exp(-j2\pi f\tau) d\tau$$

Using the Fourier transform of a triangular function (see Table 2.2 of Chapter 2), we obtain

$$S_{XX}(f) = A^2 T \operatorname{sinc}^2(fT) \quad (4.51)$$

which is plotted in Figure 4.12. Here again we see that the power spectral density is non-negative for all f and that it is an even function of f . Noting that $R_{XX}(0) = A^2$ and using

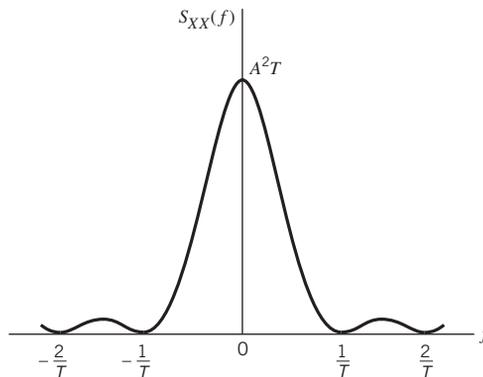


Figure 4.12 Power spectral density of random binary wave.

Property 2 of power spectral density, we find that the total area under $S_{XX}(f)$, or the average power of the random binary wave described here, is A^2 , which is intuitively satisfying.

Generalization of Equation (4.51)

It is informative to generalize (4.51) so that it assumes a more broadly applicable form. With this objective in mind, we first note that the energy spectral density (i.e., the squared magnitude of the Fourier transform) of a rectangular pulse $g(t)$ of amplitude A and duration T is given by

$$E_g(f) = A^2 T^2 \text{sinc}^2(fT) \quad (4.52)$$

We may therefore express (4.51) in terms of $E_g(f)$ simply as

$$S_{XX}(f) = \frac{E_g(f)}{T} \quad (4.53)$$

In words, (4.53) states:

For a random binary wave $X(t)$ in which binary symbols 1 and 0 are represented by pulses $g(t)$ and $-g(t)$ respectively, the power spectral density $S_{XX}(f)$ is equal to the energy spectral density $E_g(f)$ of the symbol-shaping pulse $g(t)$ divided by the symbol duration T .

EXAMPLE 7

Mixing of a Random Process with a Sinusoidal Process

A situation that often arises in practice is that of *mixing* (i.e., multiplication) of a weakly stationary process $X(t)$ with a sinusoidal wave $\cos(2\pi f_c t + \Theta)$, where the phase Θ is a random variable that is uniformly distributed over the interval $[0, 2\pi]$. The addition of the random phase Θ in this manner merely recognizes the fact that the time origin is arbitrarily chosen when both $X(t)$ and $\cos(2\pi f_c t + \Theta)$ come from physically independent sources, as is usually the case in practice. We are interested in determining the power spectral density of the stochastic process

$$Y(t) = X(t) \cos(2\pi f_c t + \Theta) \quad (4.54)$$

Using the definition of autocorrelation function of a weakly stationary process and noting that the random variable Θ is independent of $X(t)$, we find that the autocorrelation function of the process $Y(t)$ is given by

$$\begin{aligned} R_{YY}(\tau) &= \mathbb{E}[Y(t+\tau)Y(t)] \\ &= \mathbb{E}[X(t+\tau) \cos(2\pi f_c t + 2\pi f_c \tau + \Theta) X(t) \cos(2\pi f_c t + \Theta)] \\ &= \mathbb{E}[X(t+\tau)x(t)] \mathbb{E}[\cos(2\pi f_c t + 2\pi f_c \tau + \Theta) \cos(2\pi f_c t + \Theta)] \\ &= \frac{1}{2} R_{XX}(\tau) \mathbb{E}[\cos(2\pi f_c t) + \cos(4\pi f_c t + 2\pi f_c \tau + 2\Theta)] \\ &= \frac{1}{2} R_{XX}(\tau) \cos(2\pi f_c \tau) \end{aligned} \quad (4.55)$$

Since the power spectral density of a weakly stationary process is the Fourier transform of its autocorrelation function, we may go on to express the relationship between the power spectral densities of the processes $X(t)$ and $Y(t)$ as follows:

$$S_{YY}(f) = \frac{1}{4}[S_{XX}(f-f_c) + S_{XX}(f+f_c)] \quad (4.56)$$

Equation (4.56) teaches us that the power spectral density of the stochastic process $Y(t)$ defined in (4.54) can be obtained as follows:

Shift the given power spectral density $S_{XX}(f)$ of the weakly stationary process $X(t)$ to the right by f_c , shift it to the left by f_c , add the two shifted power spectra, and then divide the result by 4, thereby obtaining the desired power spectral density $S_{YY}(f)$.

Relationship between the Power Spectral Densities of Input and Output Weakly Stationary Processes

Let $S_{YY}(f)$ denote the power spectral density of the output stochastic processes $Y(t)$ obtained by passing the weakly stationary process $X(t)$ through a linear time-invariant filter of frequency response $H(f)$. Then, by definition, recognizing that the power spectral density of a weakly stationary process is equal to the Fourier transform of its autocorrelation function and using (4.32), we obtain

$$\begin{aligned} S_{YY}(f) &= \int_{-\infty}^{\infty} R_{YY}(\tau) \exp(-j2\pi f\tau) d\tau \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_{XX}(\tau + \tau_1 - \tau_2) \exp(-j2\pi f\tau) d\tau_1 d\tau_2 d\tau \end{aligned} \quad (4.57)$$

Let $\tau + \tau_1 - \tau_2 = \tau_0$, or equivalently $\tau = \tau_0 - \tau_1 + \tau_2$. By making this substitution into (4.57), we find that $S_{YY}(f)$ may be expressed as the product of three terms:

- the frequency response $H(f)$ of the filter;
- the complex conjugate of $H(f)$; and
- the power spectral density $S_{XX}(f)$ of the input process $X(t)$.

We may thus simplify (4.57) as shown by

$$S_{YY}(f) = H(f)H^*(f)S_{XX}(f) \quad (4.58)$$

Since $|H(f)|^2 = H(f)H^*(f)$, we finally find that the relationship among the power spectral densities of the input and output processes is expressed in the frequency domain by

$$S_{YY}(f) = |H(f)|^2 S_{XX}(f) \quad (4.59)$$

Equation (4.59) states:

The power spectral density of the output process $Y(t)$ equals the power spectral density of the input process $X(t)$, multiplied by the squared magnitude response of the filter.

By using (4.59), we can therefore determine the effect of passing a weakly stationary process through a stable, linear time-invariant filter. In computational terms, (4.59) is

obviously easier to handle than its time-domain counterpart of (4.32) that involves the autocorrelation function.

The Wiener–Khintchine Theorem

At this point in the discussion, a basic question that comes to mind is the following:

Given a function $\rho_{XX}(\tau)$ whose argument is some time shift τ , how do we know that $\rho_{XX}(\tau)$ is the legitimate normalized autocorrelation function of a weakly stationary process $X(t)$?

The answer to this question is embodied in a theorem that was first proved by Wiener (1930) and at a later date by Khintchine (1934). Formally, the *Wiener–Khintchine theorem*⁴ states:

A necessary and sufficient condition for $\rho_{XX}(\tau)$ to be the normalized autocorrelation function of a weakly stationary process $X(t)$ is that there exists a distribution function $F_{XX}(f)$ such that for all possible values of the time shift τ , the function $\rho_{XX}(\tau)$ may be expressed in terms of the well-known *Fourier–Stieltjes theorem*, defined by

$$\rho_{XX}(\tau) = \int_{-\infty}^{\infty} \exp(j2\pi f\tau) dF_{XX}(f) \quad (4.60)$$

The Wiener–Khintchine theorem described in (4.60) is of fundamental importance to a theoretical treatment of weakly stationary processes.

Referring back to the definition of the spectral distribution function $F_{XX}(f)$ given in (4.49), we may express the *integrated spectrum* $dF_{XX}(f)$ as

$$dF_{XX}(f) = p_{XX}(f) df \quad (4.61)$$

which may be interpreted as the probability of $X(t)$ contained in the frequency interval $[f, f + df]$. Hence, we may rewrite (4.60) in the equivalent form

$$\rho_{XX}(\tau) = \int_{-\infty}^{\infty} p_{XX}(f) \exp(j2\pi f\tau) df \quad (4.62)$$

which expresses $\rho_{XX}(\tau)$ as the inverse Fourier transform of $p_{XX}(f)$. At this point, we proceed by taking three steps:

1. Substitute (4.14) for $\rho_{XX}(\tau)$ on the left-hand side of (4.62).
2. Substitute (4.48) for $p_{XX}(\tau)$ inside the integral on the right-hand side of (4.62).
3. Use Property 3 of power spectral density in Section 4.7.

The end result of these three steps is the reformulation of (4.62) as shown by

$$\frac{R_{XX}(\tau)}{R_{XX}(0)} = \int_{-\infty}^{\infty} \frac{S_{XX}(f)}{R_{XX}(0)} \exp(j2\pi f\tau) df$$

Hence, canceling out the common term $R_{XX}(0)$, we obtain

$$R_{XX}(\tau) = \int_{-\infty}^{\infty} S_{XX}(f) \exp(j2\pi f\tau) df \quad (4.63)$$

which is a rewrite of (4.43). We may argue, therefore, that basically the two Wiener–Khinchine equations follow from either one of the following two approaches:

1. The definition of the power spectral density as the Fourier transform of the autocorrelation function, which was first derived in (4.38).
2. The Wiener–Khinchine theorem described in (4.60).

4.8 Another Definition of the Power Spectral Density

Equation (4.38) provides one definition of the power spectral density $S_{XX}(f)$ of a weakly stationary process $X(t)$; that is, $S_{XX}(f)$ is the Fourier transform of the autocorrelation function $R_{XX}(\tau)$ of the process $X(t)$. We arrived at this definition by working on the mean-square value (i.e., average power) of the process $Y(t)$ produced at the output of a linear time-invariant filter, driven by a weakly stationary process $X(t)$. In this section, we provide another definition of the power spectral density by working on the process $X(t)$ directly. The definition so developed is not only mathematically satisfying, but it also provides another way of interpreting the power spectral density.

Consider, then, a stochastic process $X(t)$, which is known to be weakly stationary. Let $x(t)$ represent a *sample function* of the process $X(t)$. For the sample function to be Fourier transformable, it must be absolutely integrable; that is,

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty$$

This condition can never be satisfied by any sample function $x(t)$ of infinite duration. To get around this problem, we consider a truncated segment of $x(t)$ defined over the observation interval $-T \leq t \leq T$, as illustrated in Figure 4.13, as shown by

$$x_T(t) = \begin{cases} x(t), & -T \leq t \leq T \\ 0, & \text{otherwise} \end{cases} \quad (4.64)$$

Clearly, the truncated signal $x_T(t)$ has finite energy; therefore, it is Fourier transformable. Let $X_T(f)$ denote the Fourier transform of $x_T(t)$, as shown by the transform pair:

$$x_T(t) \Leftrightarrow X_T(f)$$

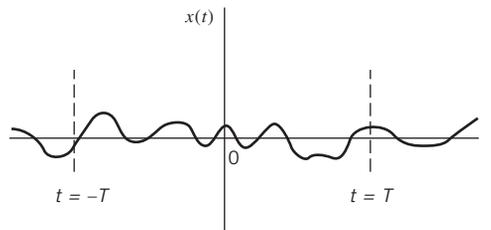


Figure 4.13 Illustration of the truncation of a sample $x(t)$ for Fourier transformability; the actual function $x(t)$ extends beyond the observation interval $(-T, T)$ as shown by the dashed lines.

in light of which we may invoke *Rayleigh's energy theorem* (Property 14 in Table 2.1) to write

$$\int_{-\infty}^{\infty} |x_T(t)|^2 dt = \int_{-\infty}^{\infty} |X_T(f)|^2 df$$

Since (4.64) implies that

$$\int_{-\infty}^{\infty} |x_T(t)|^2 dt = \int_{-T}^T |x(t)|^2 dt$$

we may also apply Rayleigh's energy theorem to the problem at hand as follows:

$$\int_{-T}^T |x(t)|^2 dt = \int_{-\infty}^{\infty} |X_T(f)|^2 df \quad (4.65)$$

With the two sides of (4.65) based on a single realization of the process $X(t)$, they are both subject to numerical variability (i.e., instability) as we go from one sample function of the process $X(t)$ to another. To mitigate this difficulty, we take the ensemble average of (4.65), and thus write

$$\mathbb{E}\left[\int_{-T}^T |x(t)|^2 dt\right] = \mathbb{E}\left[\int_{-\infty}^{\infty} |X_T(f)|^2 df\right] \quad (4.66)$$

What we have in (4.66) are two energy-based quantities. However, in the weakly stationary process $X(t)$, we have a process with some finite power. To put matters right, we multiply both sides of (4.66) by the scaling factor $1/(2T)$ and take the limiting form of the equation as the observation interval T approaches infinity. In so doing, we obtain

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \mathbb{E}\left[\int_{-T}^T |x(t)|^2 dt\right] = \lim_{T \rightarrow \infty} \mathbb{E}\left[\int_{-\infty}^{\infty} \frac{|X_T(f)|^2}{2T} df\right] \quad (4.67)$$

The quantity on the left-hand side of (4.67) is now recognized as the average power of the process $X(t)$, denoted by P_{av} , which applies to all possible sample functions of the process $X(t)$. We may therefore recast (4.67) in the equivalent form

$$P_{\text{av}} = \lim_{T \rightarrow \infty} \mathbb{E}\left[\int_{-\infty}^{\infty} \frac{|X_T(f)|^2}{2T} df\right] \quad (4.68)$$

In (4.68), we next recognize that there are two mathematical operations of fundamental interest:

1. Integration with respect to the frequency f .
2. Limiting operation with respect to the total observation interval $2T$ followed by ensemble averaging.

These two operations, viewed in a composite manner, result in a statistically stable quantity defined by P_{av} . Therefore, it is permissible for us to interchange the order of the two operations on the right-hand side of (4.68), recasting this equation in the desired form:

$$P_{\text{av}} = \int_{-\infty}^{\infty} \left\{ \lim_{T \rightarrow \infty} \mathbb{E}\left[\frac{|X_T(f)|^2}{2T}\right] \right\} df \quad (4.69)$$

With (4.69) at hand, we are now ready to formulate another definition for the power spectral density as⁵

$$S_{XX}(f) = \lim_{T \rightarrow \infty} \mathbb{E} \left[\frac{|X_T(f)|^2}{2T} \right] \quad (4.70)$$

This new definition has the following interpretation:

$S_{XX}(f) df$ is the average of the contributions to the total power from components in a weakly stationary process $X(t)$ with frequencies extending from f to $f + df$, and the average is taken over all possible realizations of the process $X(t)$.

This new interpretation of the power spectral density is all the more satisfying when (4.70) is substituted into (4.68), yielding

$$P_{\text{av}} = \int_{-\infty}^{\infty} S_{XX}(f) df \quad (4.71)$$

which is immediately recognized as another way of describing Property 3 of the power spectral density (i.e., (4.45)). End-of-chapter Problem 4.8 invites the reader to prove other properties of the power spectral density, using the definition of (4.70).

One last comment must be carefully noted: in the definition of the power spectral density given in (4.70), it is *not* permissible to let the observation interval T approach infinity before taking the expectation; in other words, these two operations are *not* commutative.

4.9 Cross-spectral Densities

Just as the power spectral density provides a measure of the frequency distribution of a single weakly stationary process, cross-spectral densities provide measures of the frequency interrelationships between two such processes. To be specific, let $X(t)$ and $Y(t)$ be two jointly weakly stationary processes with their cross-correlation functions denoted by $R_{XY}(\tau)$ and $R_{YX}(\tau)$. We define the corresponding *cross-spectral densities* $S_{XY}(f)$ and $S_{YX}(f)$ of this pair of processes to be the Fourier transforms of their respective cross-correlation functions, as shown by

$$S_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau) \exp(-j2\pi f\tau) d\tau \quad (4.72)$$

and

$$S_{YX}(f) = \int_{-\infty}^{\infty} R_{YX}(\tau) \exp(-j2\pi f\tau) d\tau \quad (4.73)$$

The cross-correlation functions and cross-spectral densities form Fourier-transform pairs. Accordingly, using the formula for inverse Fourier transformation, we may also respectively write

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} S_{XY}(f) \exp(j2\pi f\tau) df \quad (4.74)$$

and

$$R_{YX}(\tau) = \int_{-\infty}^{\infty} S_{YX}(f) \exp(j2\pi f\tau) df \quad (4.75)$$

The cross-spectral densities $S_{XY}(f)$ and $S_{YX}(f)$ are not necessarily real functions of the frequency f . However, substituting the following relationship (i.e., Property 2 of the autocorrelation function)

$$R_{XY}(\tau) = R_{YX}(-\tau)$$

into (4.72) and then using (4.73), we find that $S_{XY}(f)$ and $S_{YX}(f)$ are related as follows:

$$S_{XY}(f) = S_{YX}(-f) = S_{YX}^*(f) \quad (4.76)$$

where the asterisk denotes complex conjugation.

EXAMPLE 8

Sum of Two Weakly Stationary Processes

Suppose that the stochastic processes $X(t)$ and $Y(t)$ have zero mean and let their sum be denoted by

$$Z(t) = X(t) + Y(t)$$

The problem is to determine the power spectral density of the process $Z(t)$.

The autocorrelation function of $Z(t)$ is given by the second-order moment

$$\begin{aligned} M_{ZZ}(t, u) &= \mathbb{E}[Z(t)Z(u)] \\ &= \mathbb{E}[(X(t) + Y(t))(X(u) + Y(u))] \\ &= \mathbb{E}[X(t)X(u)] + \mathbb{E}[X(t)Y(u)] + \mathbb{E}[Y(t)X(u)] + \mathbb{E}[Y(t)Y(u)] \\ &= M_{XX}(t, u) + M_{XY}(t, u) + M_{YX}(t, u) + M_{YY}(t, u) \end{aligned}$$

Defining $\tau = t - u$ and assuming the joint weakly stationarity of the two processes, we may go on to write

$$R_{ZZ}(\tau) = R_{XX}(\tau) + R_{XY}(\tau) + R_{YX}(\tau) + R_{YY}(\tau) \quad (4.77)$$

Accordingly, taking the Fourier transform of both sides of (4.77), we get

$$S_{ZZ}(f) = S_{XX}(f) + S_{XY}(f) + S_{YX}(f) + S_{YY}(f) \quad (4.78)$$

This equation shows that the cross-spectral densities $S_{XY}(f)$ and $S_{YX}(f)$ represent the spectral components that must be added to the individual power spectral densities of a pair of correlated weakly stationary processes in order to obtain the power spectral density of their sum.

When the stationary processes $X(t)$ and $Y(t)$ are uncorrelated, the cross-spectral densities $S_{XY}(f)$ and $S_{YX}(f)$ are zero, in which case (4.78) reduces to

$$S_{ZZ}(f) = S_{XX}(f) + S_{YY}(f) \quad (4.79)$$

We may generalize this latter result by stating:

When there is a multiplicity of zero-mean weakly stationary processes that are uncorrelated with each other, the power spectral density of their sum is equal to the sum of their individual power spectral densities.

EXAMPLE 9 Filtering of Two Jointly Weakly Stationary Processes

Consider next the problem of passing two jointly weakly stationary processes through a pair of separate, stable, linear time-invariant filters, as shown in Figure 4.14. The stochastic process $X(t)$ is the input to the filter of impulse response $h_1(t)$, and the stochastic process $Y(t)$ is the input to the filter of the impulse response $h_2(t)$. Let $V(t)$ and $Z(t)$ denote the processes at the respective filter outputs. The cross-correlation function of the output processes $V(t)$ and $Z(t)$ is therefore defined by the second-order moment

$$\begin{aligned}
 M_{VZ}(t, u) &= \mathbb{E}[V(t)Z(u)] \\
 &= \mathbb{E}\left[\int_{-\infty}^{\infty} h_1(\tau_1)X(t-\tau_1) d\tau_1 \int_{-\infty}^{\infty} h_2(\tau_2)Y(u-\tau_2) d\tau_2\right] \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau_1)h_2(\tau_2)\mathbb{E}[X(t-\tau_1)Y(u-\tau_2)] d\tau_1 d\tau_2 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau_1)h_2(\tau_2)M_{XY}(t-\tau_1, u-\tau_2) d\tau_1 d\tau_2
 \end{aligned} \tag{4.80}$$

where $M_{XY}(t, u)$ is the cross-correlation function of $X(t)$ and $Y(t)$. Because the input stochastic processes are jointly weakly stationary, by hypothesis, we may set $\tau = t - u$, and thereby rewrite (4.80) as

$$R_{VZ}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau_1)h_2(\tau_2)R_{XY}(\tau-\tau_1+\tau_2) d\tau_1 d\tau_2 \tag{4.81}$$

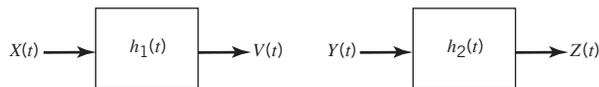
Taking the Fourier transform of both sides of (4.81) and using a procedure similar to that which led to the development of (4.39), we finally get

$$S_{VZ}(f) = H_1(f)H_2^*(f)S_{XY}(f) \tag{4.82}$$

where $H_1(f)$ and $H_2(f)$ are the frequency responses of the respective filters in Figure 4.14 and $H_2^*(f)$ is the complex conjugate of $H_2(f)$. This is the desired relationship between the cross-spectral density of the output processes and that of the input processes. Note that (4.82) includes (4.59) as a special case.

Figure 4.14

A pair of separate linear time-invariant filters.



4.10 The Poisson Process

Having covered the basics of stochastic process theory, we now turn our attention to different kinds of stochastic processes that are commonly encountered in the study of communication systems. We begin the study with the Poisson process,⁶ which is the simplest process dealing with the issue of counting the number of occurrences of random events.

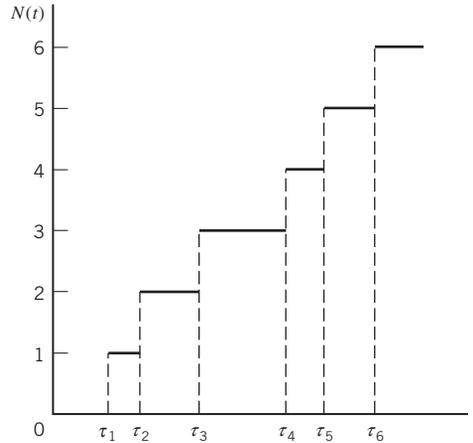


Figure 4.15 Sample function of a Poisson counting process.

Consider, for example, a situation in which events occur at random instants of time, such that the *average rate* of events per second is equal to λ . The sample path of such a random process is illustrated in Figure 4.15, where τ_i denotes the occurrence time of the i th event with $i = 1, 2, \dots$. Let $N(t)$ be the number of event occurrences in the time interval $[0, t]$. As illustrated in Figure 4.15, we see that $N(t)$ is a nondecreasing, integer-valued, continuous process. Let $p_{k,\tau}$ denote the probability that exactly k events occur during an interval of duration τ ; that is,

$$p_{k,\tau} = \mathbb{P}[N(t, t + \tau) = k] \quad (4.83)$$

With this background, we may now formally define the Poisson process:

A random counting process is said to be a Poisson process with average rate λ if it satisfies the three basic properties listed below.

PROPERTY 1 Time Homogeneity

The probability $p_{k,\tau}$ of k event occurrences is the same for all intervals of the same duration τ .

The essence of Property 1 is that the events are equally likely at all times.

PROPERTY 2 Distribution Function

The number of event occurrences, $N_{0,t}$ in the interval $[0, t]$ has a distribution function with mean λt , defined by

$$\mathbb{P}[N(t) = k] = \frac{(\lambda t)^k}{k!} \exp(-\lambda t), \quad k = 0, 1, 2, \dots \quad (4.84)$$

That is, the time between events is *exponentially distributed*.

From Chapter 3, this distribution function is recognized to be the *Poisson distribution*. It is for this reason that $N(t)$ is called the *Poisson process*.

PROPERTY 3 Independence

The numbers of events in nonoverlapping time intervals are statistically independent, regardless of how small or large the intervals happen to be and no matter how close or distant they could be.

Property 3 is the most distinguishing property of the Poisson process. To illustrate the significance of this property, let $[t_i, u_i]$ for $i = 1, 2, \dots, k$ denote k disjoint intervals on the line $[0, \infty]$. We may then write

$$\mathbb{P}[N(t_1, u_1) = n_1; N(t_2, u_2) = n_2; \dots; N(t_k, u_k) = n_k] = \prod_{i=1}^k \mathbb{P}[N(t_i, u_i) = n_i] \quad (4.85)$$

The important point to take from this discussion is that these three properties provide a complete characterization of the Poisson process.

This kind of stochastic process arises, for example, in the statistical characterization of a special kind of noise called *shot noise* in electronic devices (e.g., diodes and transistors), which arises due to the discrete nature of current flow.

4.11 The Gaussian Process

The second stochastic process of interest is the *Gaussian process*, which builds on the Gaussian distribution discussed in Chapter 3. The Gaussian process is by far the most frequently encountered random process in the study of communication systems. We say so for two reasons: practical applicability and mathematical tractability.⁷

Let us suppose that we observe a stochastic process $X(t)$ for an interval that starts at time $t = 0$ and lasts until $t = T$. Suppose also that we weight the process $X(t)$ by some function $g(t)$ and then integrate the product $g(t)X(t)$ over the observation interval $[0, T]$, thereby obtaining the random variable

$$Y = \int_0^T g(t)X(t) dt \quad (4.86)$$

We refer to Y as a *linear functional* of $X(t)$. The distinction between a function and a functional should be carefully noted. For example, the sum $Y = \sum_{i=1}^N a_i X_i$, where the a_i are constants and the X_i are random variables, is a linear function of the X_i ; for each observed set of values for the random variable X_i , we have a corresponding value for the random variable Y . On the other hand, the value of the random variable Y in (4.86) depends on the course of the *integrand function* $g(t)X(t)$ over the entire observation interval from 0 to T . Thus, a functional is a quantity that depends on the entire course of one or more functions rather than on a number of discrete variables. In other words, the domain of a functional is a space of admissible functions rather than a region of coordinate space.

If, in (4.86), the weighting function $g(t)$ is such that the mean-square value of the random variable Y is finite and if the random variable Y is a *Gaussian-distributed* random variable for every $g(t)$ in this class of functions, then the process $X(t)$ is said to be a *Gaussian process*. In words, we may state:

A process $X(t)$ is said to be a Gaussian process if every linear functional of $X(t)$ is a Gaussian random variable.

From Chapter 3 we recall that the random variable Y has a *Gaussian distribution* if its probability density function has the form

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \quad (4.87)$$

where μ is the mean and σ^2 is the variance of the random variable Y . The distribution of a Gaussian process $X(t)$, sampled at some fixed time t_k , say, satisfies (4.87).

From a theoretical as well as practical perspective, a Gaussian process has two main virtues:

1. The Gaussian process has many properties that make analytic results possible; we will discuss these properties later in the section.
2. The stochastic processes produced by physical phenomena are often such that a Gaussian model is appropriate. Furthermore, the use of a Gaussian model to describe physical phenomena is often confirmed by experiments. Last, but by no means least, the central limit theorem (discussed in Chapter 3) provides mathematical justification for the Gaussian distribution.

Thus, the frequent occurrence of physical phenomena for which a Gaussian model is appropriate and the ease with which a Gaussian process is handled mathematically make the Gaussian process very important in the study of communication systems.

Properties of a Gaussian Process

PROPERTY 1 Linear Filtering

If a Gaussian process $X(t)$ is applied to a stable linear filter, then the stochastic process $Y(t)$ developed at the output of the filter is also Gaussian.

This property is readily derived by using the definition of a Gaussian process based on (4.86). Consider the situation depicted in Figure 4.8, where we have a linear time-invariant filter of impulse response $h(t)$, with the stochastic process $X(t)$ as input and the stochastic process $Y(t)$ as output. We assume that $X(t)$ is a Gaussian process. The process $Y(t)$ is related to $X(t)$ by the convolution integral

$$Y(t) = \int_0^T h(t-\tau)X(\tau) d\tau, \quad 0 \leq t < \infty \quad (4.88)$$

We assume that the impulse response $h(t)$ is such that the mean-square value of the output random process $Y(t)$ is finite for all time t in the range $0 \leq t < \infty$, for which the process $Y(t)$ is defined. To demonstrate that the output process $Y(t)$ is Gaussian, we must show that any linear functional of it is also a Gaussian random variable. That is, if we define the random variable

$$Z = \int_0^\infty g_Y(t) \left[\int_0^T h(t-\tau)X(\tau) d\tau \right] dt \quad (4.89)$$

then Z must be a Gaussian random variable for every function $g_Y(t)$, such that the mean-square value of Z is finite. The two operations performed in the right-hand side of (4.89)

are both linear; therefore, it is permissible to interchange the order of integrations, obtaining

$$Z = \int_0^T g(t)X(\tau) dt \quad (4.90)$$

where the new function

$$g(\tau) = \int_0^T g_Y(t)h(t-\tau) dt \quad (4.91)$$

Since $X(t)$ is a Gaussian process by hypothesis, it follows from (4.91) that Z must also be a Gaussian random variable. We have thus shown that if the input $X(t)$ to a linear filter is a Gaussian process, then the output $Y(t)$ is also a Gaussian process. Note, however, that although our proof was carried out assuming a time-invariant linear filter, this property is also true for any arbitrary stable linear filter.

PROPERTY 2 Multivariate Distribution

Consider the set of random variables $X(t_1), X(t_2), \dots, X(t_n)$, obtained by sampling a stochastic process $X(t)$ at times t_1, t_2, \dots, t_n . If the process $X(t)$ is Gaussian, then this set of random variables is jointly Gaussian for any n , with their n -fold joint probability density function being completely determined by specifying the set of means

$$\mu_{X(t_i)} = \mathbb{E}[X(t_i)], \quad i = 1, 2, \dots, n \quad (4.92)$$

and the set of covariance functions

$$C_X(t_k, t_i) = \mathbb{E}[(X(t_k) - \mu_{X(t_k)})(X(t_i) - \mu_{X(t_i)})], \quad k, i = 1, 2, \dots, n \quad (4.93)$$

Let the n -by-1 vector \mathbf{X} denote the set of random variables $X(t_1), X(t_2), \dots, X(t_n)$ derived from the Gaussian process $X(t)$ by sampling it at times t_1, t_2, \dots, t_n . Let the vector \mathbf{x} denote a sample value of \mathbf{X} . According to Property 2, the random vector \mathbf{X} has a *multivariate Gaussian distribution*, defined in matrix form as

$$f_{X(t_1), X(t_2), \dots, X(t_n)}(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \Delta^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right] \quad (4.94)$$

where the superscript T denotes matrix transposition, the mean vector

$$\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]^T$$

the covariance matrix

$$\boldsymbol{\Sigma} = \{C_X(t_k, t_i)\}_{k, i=1}^n$$

$\boldsymbol{\Sigma}^{-1}$ is the inverse of the covariance matrix $\boldsymbol{\Sigma}$, and Δ is the determinant of the covariance matrix $\boldsymbol{\Sigma}$.

Property 2 is frequently used as the definition of a Gaussian process. However, this definition is more difficult to use than that based on (4.86) for evaluating the effects of filtering on a Gaussian process.

Note also that the covariance matrix $\boldsymbol{\Sigma}$ is a symmetric nonnegative definite matrix. For a nondegenerate Gaussian process, $\boldsymbol{\Sigma}$ is positive definite, in which case the covariance matrix is invertible.

PROPERTY 3 Stationarity

If a Gaussian process is weakly stationary, then the process is also strictly stationary.

This follows directly from Property 2.

PROPERTY 4 Independence

If the random variables $X(t_1), X(t_2), \dots, X(t_n)$, obtained by respectively sampling a Gaussian process $X(t)$ at times t_1, t_2, \dots, t_n are uncorrelated, that is

$$\mathbb{E}[(X(t_k) - \mu_{X(t_k)})(X(t_i) - \mu_{X(t_i)})] = 0 \quad i \neq k \quad (4.95)$$

then these random variables are statistically independent.

The uncorrelatedness of $X(t_1), \dots, X(t_n)$ means that the covariance matrix Σ is reduced to a diagonal matrix, as shown by

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \mathbf{0} \\ & \ddots \\ \mathbf{0} & \sigma_n^2 \end{bmatrix} \quad (4.96)$$

where the $\mathbf{0}$ s denote two sets of elements whose values are all zero, and the diagonal terms

$$\sigma_i^2 = \mathbb{E}[X(t_i) - \mathbb{E}[X(t_i)]]^2, \quad i = 1, 2, \dots, n \quad (4.97)$$

Under this special condition, the multivariate Gaussian distribution described in (4.94) simplifies to

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n f_{X_i}(x_i) \quad (4.98)$$

where $X_i = X(t_i)$ and

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x_i - \mu_{X_i})^2}{2\sigma_i^2}\right], \quad i = 1, 2, \dots, n \quad (4.99)$$

In words, if the Gaussian random variables $X(t_1), X(t_2), \dots, X(t_n)$ are uncorrelated, then they are statistically independent, which, in turn, means that the joint probability density function of this set of random variables is expressed as the product of the probability density functions of the individual random variables in the set.

4.12 Noise

The term *noise* is used customarily to designate unwanted signals that tend to disturb the transmission and processing of signals in communication systems, and over which we have incomplete control. In practice, we find that there are many potential sources of noise in a communication system. The sources of noise may be external to the system (e.g.,

atmospheric noise, galactic noise, man-made noise) or internal to the system. The second category includes an important type of noise that arises from the phenomenon of *spontaneous fluctuations* of current flow that is experienced in all electrical circuits. In a physical context, the most common examples of the spontaneous fluctuation phenomenon are *shot noise*, which, as stated in Section 4.10, arises because of the discrete nature of current flow in electronic devices; and *thermal noise*, which is attributed to the random motion of electrons in a conductor.⁸ However, insofar as the noise analysis of communication systems is concerned, be they analog or digital, the analysis is customarily based on a source of noise called white-noise, which is discussed next.

White Noise

This source of noise is *idealized*, in that its power spectral density is assumed to be constant and, therefore, independent of the operating frequency. The adjective “white” is used in the sense that white light contains equal amounts of all frequencies within the visible band of electromagnetic radiation. We may thus make the statement:

White noise, denoted by $W(t)$, is a stationary process whose power spectral density $S_W(f)$ has a constant value across the entire frequency interval $-\infty < f < \infty$.

Clearly, white-noise can only be meaningful as an abstract mathematical concept; we say so because a constant power spectral density corresponds to an unbounded spectral distribution function and, therefore, infinite average power, which is physically nonrealizable. Nevertheless, the utility of white-noise is justified in the study of communication theory by virtue of the fact that it is used to model *channel noise* at the front end of a receiver. Typically, the receiver includes a filter whose frequency response is essentially zero outside a frequency band of some finite value. Consequently, when white-noise is applied to the model of such a receiver, there is no need to describe how the power spectral density $S_{WW}(f)$ falls off outside the usable frequency band of the receiver.⁹

Let

$$S_{WW}(f) = \frac{N_0}{2} \quad \text{for all } f \quad (4.100)$$

as illustrated in Figure 4.16a. Since the autocorrelation function is the inverse Fourier transform of the power spectral density in accordance with the Wiener–Khinchine relations, it follows that for white-noise the autocorrelation function is

$$R_{WW}(\tau) = \frac{N_0}{2} \delta(\tau) \quad (4.101)$$

Hence, the autocorrelation function of white noise consists of a delta function weighted by the factor $N_0/2$ and occurring at the time shift $\tau = 0$, as shown in Figure 4.16b.

Since $R_{WW}(\tau)$ is zero for $\tau \neq 0$, it follows that any two different samples of white noise are uncorrelated no matter how closely together in time those two samples are taken. If the white noise is also Gaussian, then the two samples are statistically independent in accordance with Property 4 of the Gaussian process. In a sense, then, white Gaussian noise represents the ultimate in “randomness.”

The utility of a white-noise process in the noise analysis of communication systems is parallel to that of an impulse function or delta function in the analysis of linear systems.

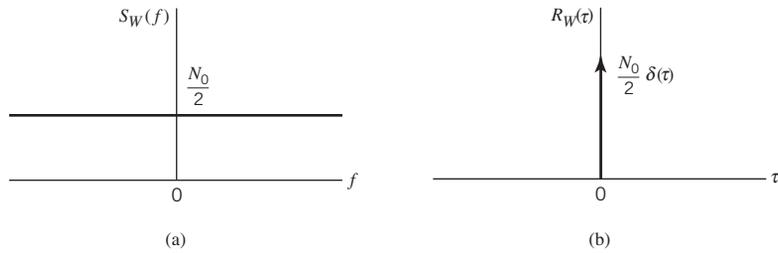


Figure 4.16 Characteristics of white-noise: (a) power spectral density; (b) autocorrelation function.

Just as we may observe the effect of an impulse only after it has been passed through a linear system with a finite bandwidth, so it is with white noise whose effect is observed only after passing through a similar system. We may therefore state:

As long as the bandwidth of a noise process at the input of a system is appreciably larger than the bandwidth of the system itself, then we may model the noise process as white noise.

EXAMPLE 10 Ideal Low-pass Filtered White Noise

Suppose that a white Gaussian noise of zero mean and power spectral density $N_0/2$ is applied to an ideal low-pass filter of bandwidth B and passband magnitude response of one. The power spectral density of the noise $N(t)$ appearing at the filter output, as shown in Figure 4.17a, is therefore

$$S_{NN}(f) = \begin{cases} \frac{N_0}{2}, & -B < f < B \\ 0, & |f| > B \end{cases} \quad (4.102)$$

Since the autocorrelation function is the inverse Fourier transform of the power spectral density, it follows that

$$\begin{aligned} R_{NN}(\tau) &= \int_{-B}^B \frac{N_0}{2} \exp(j2\pi f\tau) df \\ &= N_0 B \operatorname{sinc}(2B\tau) \end{aligned} \quad (4.103)$$

whose dependence on τ is plotted in Figure 4.17b. From this figure, we see that $R_{NN}(\tau)$ has the maximum value $N_0 B$ at the origin and it passes through zero at $\tau = \pm k/(2B)$, where $k = 1, 2, 3, \dots$

Since the input noise $W(t)$ is Gaussian (by hypothesis), it follows that the band-limited noise $N(t)$ at the filter output is also Gaussian. Suppose, then, that $N(t)$ is sampled at the rate of $2B$ times per second. From Figure 4.17b, we see that the resulting noise samples are uncorrelated and, being Gaussian, they are statistically independent. Accordingly, the joint probability density function of a set of noise samples obtained in this way is equal to the product of the individual probability density functions. Note that each such noise sample has a mean of zero and variance of $N_0 B$.

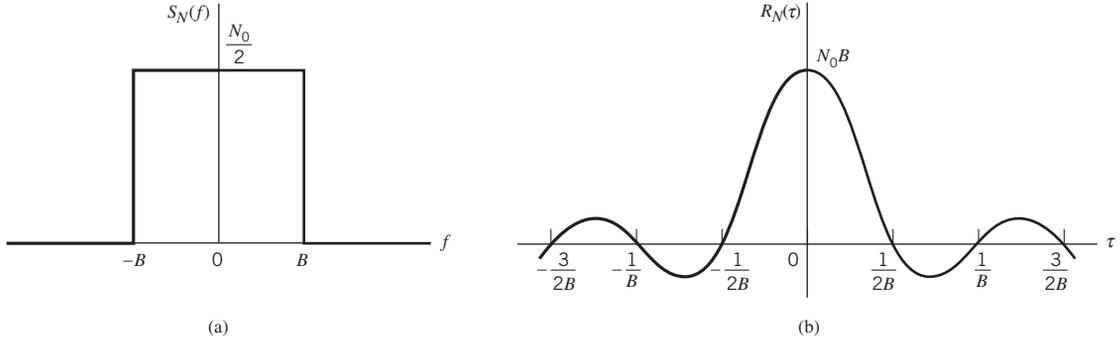


Figure 4.17 Characteristics of low-pass filtered white noise; (a) power spectral density; (b) autocorrelation function.

EXAMPLE 11 Correlation of White Noise with Sinusoidal Wave

Consider the sample function

$$w'(t) = \frac{\sqrt{2}}{\sqrt{T}} \int_0^T w(t) \cos(2\pi f_c t) dt \quad (4.104)$$

which is the output of a correlator with white Gaussian noise sample function $w(t)$ and sinusoidal wave $\sqrt{2/T} \cos(2\pi f_c t)$ as its two inputs; the scaling factor $\sqrt{2/T}$ is included in (4.104) to make the sinusoidal wave input have unit energy over the interval $0 \leq t \leq T$. With $w(t)$ having zero mean, it immediately follows that the correlator output $w'(t)$ has zero mean too. The variance of the correlator output is therefore defined by

$$\begin{aligned} \sigma_{w'}^2 &= \mathbb{E} \left[\frac{2}{T} \int_0^T \int_0^T w(t_1) \cos(2\pi f_c t_1) w(t_2) \cos(2\pi f_c t_2) dt_1 dt_2 \right] \\ &= \frac{2}{T} \int_0^T \int_0^T \mathbb{E}[w(t_1)w(t_2)] \cos(2\pi f_c t_1) \cos(2\pi f_c t_2) dt_1 dt_2 \\ &= \frac{2}{T} \int_0^T \int_0^T \frac{N_0}{2} \delta(t_1 - t_2) \cos(2\pi f_c t_1) \cos(2\pi f_c t_2) dt_1 dt_2 \end{aligned} \quad (4.105)$$

where, in the last line, we made use of (4.101). We now invoke the *sifting property* of the delta function, namely

$$\int_{-\infty}^{\infty} g(t) \delta(t) dt = g(0) \quad (4.106)$$

where $g(t)$ is a continuous function of time that has the value $g(0)$ at time $t = 0$. Hence, we may further simplify the expression for the noise variance as

$$\begin{aligned} \sigma_{w'}^2 &= \frac{N_0 2}{2 T} \int_{-T}^T \cos^2(2\pi f_c t) dt \\ &= \frac{N_0}{2T} \int_0^T [1 + \cos(4\pi f_c t)] dt \\ &= \frac{N_0}{2} \end{aligned} \quad (4.107)$$

where, in the last line, it is assumed that the frequency f_c of the sinusoidal wave input is an integer multiple of the reciprocal of T for mathematical convenience.

4.13 Narrowband Noise

The receiver of a communication system usually includes some provision for *preprocessing* the received signal. Typically, the preprocessing takes the form of a *narrowband filter* whose bandwidth is just large enough to pass the modulated component of the received signal essentially undistorted, so as to limit the effect of channel noise passing through the receiver. The noise process appearing at the output of such a filter is called *narrowband noise*. With the spectral components of narrowband noise concentrated about some midband frequency $\pm f_c$ as in Figure 4.18a, we find that a sample function $n(t)$ of such a process appears somewhat similar to a sine wave of frequency f_c . The sample function $n(t)$ may, therefore, undulate slowly in both amplitude and phase, as illustrated in Figure 4.18b.

Consider, then, the $n(t)$ produced at the output of a narrowband filter in response to the sample function $w(t)$ of a white Gaussian noise process of zero mean and unit power spectral density applied to the filter input; $w(t)$ and $n(t)$ are sample functions of the respective processes $W(t)$ and $N(t)$. Let $H(f)$ denote the transfer function of this filter. Accordingly, we may express the power spectral density $S_{NN}(f)$ of the noise $N(t)$ in terms of $H(f)$ as

$$S_{NN}(f) = |H(f)|^2 \quad (4.108)$$

On the basis of this equation, we may now make the following statement:

Any narrowband noise encountered in practice may be modeled by applying a white-noise to a suitable filter in the manner described in (4.108).

In this section we wish to represent the narrowband noise $n(t)$ in terms of its in-phase and quadrature components in a manner similar to that described for a narrowband signal in Section 2.10. The derivation presented here is based on the idea of pre-envelope and related concepts, which were discussed in Chapter 2 on Fourier analysis of signals and systems.

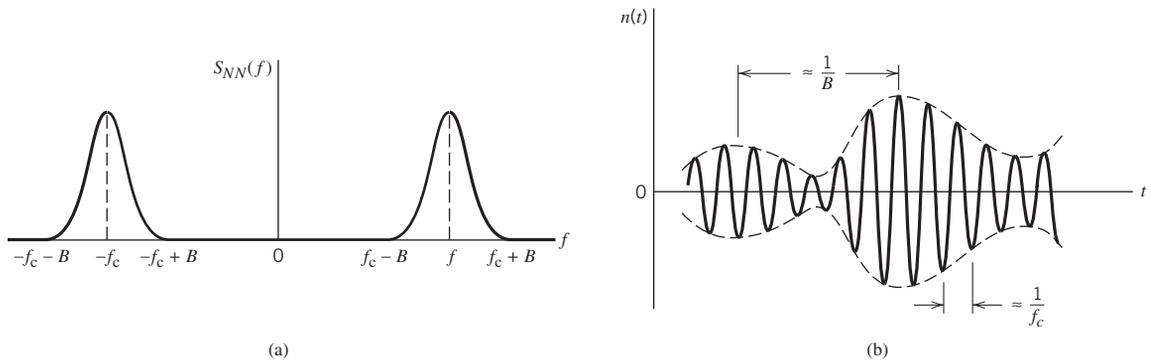


Figure 4.18 (a) Power spectral density of narrowband noise. (b) Sample function of narrowband noise.

Let $n_+(t)$ and $\tilde{n}(t)$, respectively, denote the pre-envelope and complex envelope of the narrowband noise $n(t)$. We assume that the power spectrum of $n(t)$ is centered about the frequency f_c . Then we may write

$$n_+(t) = n(t) + j\hat{n}(t) \quad (4.109)$$

and

$$\hat{n}(t) = n_+(t) \exp(-j2\pi f_c t) \quad (4.110)$$

where $\hat{n}(t)$ is the Hilbert transform of $n(t)$. The complex envelope $\tilde{n}(t)$ may itself be expressed as

$$\tilde{n}(t) = n_I(t) + jn_Q(t) \quad (4.111)$$

Hence, combining (4.109) through (4.111), we find that the *in-phase component* $n_I(t)$ and the *quadrature component* $n_Q(t)$ of the narrowband noise $n(t)$ are

$$n_I(t) = n(t) \cos(2\pi f_c t) + \hat{n}(t) \sin(2\pi f_c t) \quad (4.112)$$

and

$$n_Q(t) = \hat{n}(t) \cos(2\pi f_c t) - n(t) \sin(2\pi f_c t) \quad (4.113)$$

respectively. Eliminating $\hat{n}(t)$ between (4.112) and (4.113), we get the desired *canonical form* for representing the narrowband noise $n(t)$, as shown by

$$n(t) = n_I(t) \cos(2\pi f_c t) - n_Q(t) \sin(2\pi f_c t) \quad (4.114)$$

Using (4.112) to (4.114), we may now derive some important properties of the in-phase and quadrature components of a narrowband noise, as described next.

PROPERTY 1 *The in-phase component $n_I(t)$ and quadrature component $n_Q(t)$ of narrowband noise $n(t)$ have zero mean.*

To prove this property, we first observe that the noise $\hat{n}(t)$ is obtained by passing $n(t)$ through a linear filter (i.e., Hilbert transformer). Accordingly, $\hat{n}(t)$ will have zero mean because $n(t)$ has zero mean by virtue of its narrowband nature. Furthermore, from (4.112) and (4.113), we see that $n_I(t)$ and $n_Q(t)$ are weighted sums of $n(t)$ and $\hat{n}(t)$. It follows, therefore, that the in-phase and quadrature components, $n_I(t)$ and $n_Q(t)$, both have zero mean.

PROPERTY 2 *If the narrowband noise $n(t)$ is Gaussian, then its in-phase component $n_I(t)$ and quadrature component $n_Q(t)$ are jointly Gaussian.*

To prove this property, we observe that $\hat{n}(t)$ is derived from $n(t)$ by a linear filtering operation. Hence, if $n(t)$ is Gaussian, the Hilbert transform $\hat{n}(t)$ is also Gaussian, and $n(t)$ and $\hat{n}(t)$ are jointly Gaussian. It follows, therefore, that the in-phase and quadrature components, $n_I(t)$ and $n_Q(t)$, are jointly Gaussian, since they are weighted sums of jointly Gaussian processes.

PROPERTY 3 *If the narrowband noise $n(t)$ is weakly stationary, then its in-phase component $n_I(t)$ and quadrature component $n_Q(t)$ are jointly weakly stationary.*

If $n(t)$ is weakly stationary, so is its Hilbert transform $\hat{n}(t)$. However, since the in-phase and quadrature components, $n_I(t)$ and $n_Q(t)$, are both weighted sums of $n(t)$ and $\hat{n}(t)$

and the weighting functions, $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$, vary with time, we cannot directly assert that $n_I(t)$ and $n_Q(t)$ are weakly stationary. To prove Property 3, we have to evaluate their correlation functions.

Using (4.112) and (4.113), we find that the in-phase and quadrature components, $n_I(t)$ and $n_Q(t)$, of a narrowband noise $n(t)$ have the same autocorrelation function, as shown by

$$R_{N_I N_I}(\tau) = R_{N_Q N_Q}(\tau) = R_{NN}(\tau) \cos(2\pi f_c \tau) + \hat{R}_{NN}(\tau) \sin(2\pi f_c \tau) \quad (4.115)$$

and their cross-correlation functions are given by

$$R_{N_I N_Q}(\tau) = -R_{N_Q N_I}(\tau) = R_{NN}(\tau) \sin(2\pi f_c \tau) - \hat{R}_{NN}(\tau) \cos(2\pi f_c \tau) \quad (4.116)$$

where $R_{NN}(\tau)$ is the autocorrelation function of $n(t)$, and $\hat{R}_{NN}(\tau)$ is the Hilbert transform of $R_{NN}(\tau)$. From (4.115) and (4.116), we readily see that the correlation functions $R_{N_I N_I}(\tau)$, $R_{N_Q N_Q}(\tau)$, and $R_{N_I N_Q}(\tau)$ of the in-phase and quadrature components $n_I(t)$ and $n_Q(t)$ depend only on the time shift τ . This dependence, in conjunction with Property 1, proves that $n_I(t)$ and $n_Q(t)$ are weakly stationary if the original narrowband noise $n(t)$ is weakly stationary.

PROPERTY 4 *Both the in-phase noise $n_I(t)$ and quadrature noise $n_Q(t)$ have the same power spectral density, which is related to the power spectral density $S_{NN}(f)$ of the original narrowband noise $n(t)$ as follows:*

$$S_{N_I N_I}(f) = S_{N_Q N_Q}(f) = \begin{cases} S_{NN}(f-f_c) + S_{NN}(f+f_c), & -B \leq f \leq B \\ 0, & \text{otherwise} \end{cases} \quad (4.117)$$

where it is assumed that $S_{NN}(f)$ occupies the frequency interval $f_c - B \leq |f| \leq f_c + B$ and $f_c > B$.

To prove this fourth property, we take the Fourier transforms of both sides of (4.115), and use the fact that

$$\begin{aligned} \mathbf{F}[\hat{R}_{NN}(\tau)] &= -j \operatorname{sgn}(f) \mathbf{F}[R_{NN}(\tau)] \\ &= -j \operatorname{sgn}(f) S_{NN}(f) \end{aligned} \quad (4.118)$$

We thus obtain the result

$$\begin{aligned} S_{N_I N_I}(f) &= S_{N_Q N_Q}(f) \\ &= \frac{1}{2} [S_{NN}(f-f_c) + S_{NN}(f+f_c)] \\ &\quad - \frac{1}{2} [S_{NN}(f-f_c) \operatorname{sgn}(f-f_c) - S_{NN}(f+f_c) \operatorname{sgn}(f+f_c)] \\ &= \frac{1}{2} S_{NN}(f-f_c) [1 - \operatorname{sgn}(f-f_c)] + \frac{1}{2} S_{NN}(f+f_c) [1 + \operatorname{sgn}(f+f_c)] \end{aligned} \quad (4.119)$$

Now, with the power spectral density $S_{NN}(f)$ of the original narrowband noise $n(t)$ occupying the frequency interval $f_c - B \leq |f| \leq f_c + B$, where $f_c > B$, as illustrated in Figure 4.19, we find that the corresponding shapes of $S_{NN}(f-f_c)$ and $S_{NN}(f+f_c)$ are as in Figures 4.19b and 4.19c respectively. Figures 4.19d, 4.19e, and 4.19f show the shapes of

$\text{sgn}(f)$, $\text{sgn}(f - f_c)$, and $\text{sgn}(f + f_c)$ respectively. Accordingly, we may make the following observation from Figure 4.19:

1. For frequencies defined by $-B \leq f \leq B$, we have

$$\text{sgn}(f - f_c) = -1$$

and

$$\text{sgn}(f + f_c) = +1$$

Hence, substituting these results into (4.119), we obtain

$$\begin{aligned} S_{N_I N_I}(f) &= S_{N_Q N_Q}(f) \\ &= S_{NN}(f - f_c) + S_{NN}(f + f_c), \quad -B \leq f \leq B \end{aligned}$$

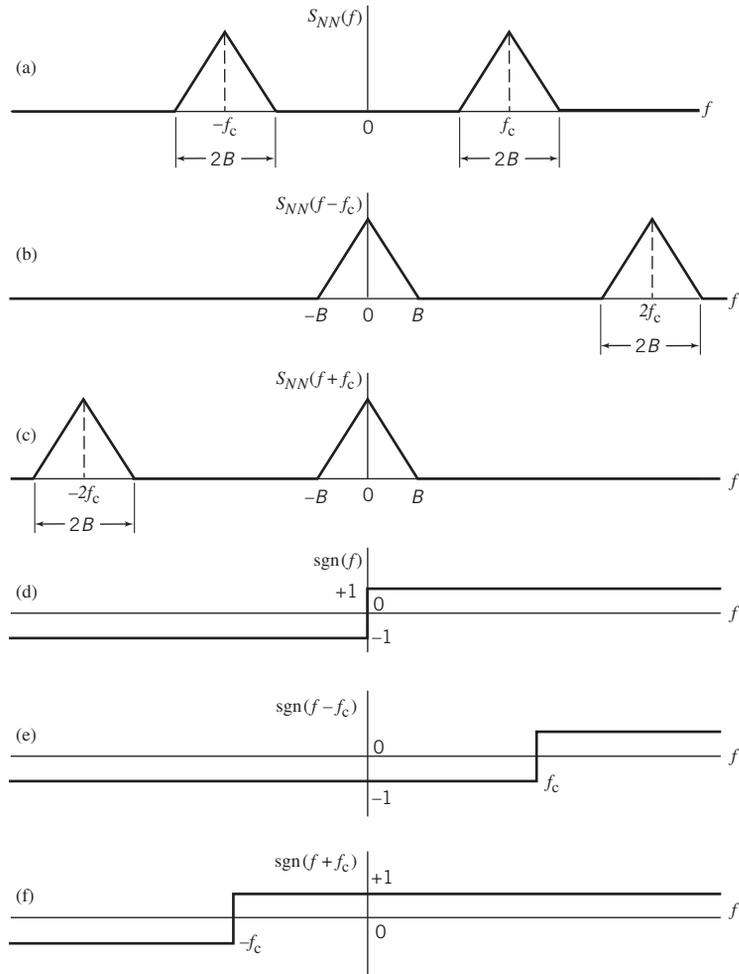


Figure 4.19

- (a) Power spectral density $S_{NN}(f)$ pertaining to narrowband noise $n(t)$.
- (b), (c) Frequency-shifted versions of $S_{NN}(f)$ in opposite directions.
- (d) Signum function $\text{sgn}(f)$.
- (e), (f) Frequency-shifted versions of $\text{sgn}(f)$ in opposite directions.

2. For $2f_c - B \leq f \leq 2f_c + B$, we have

$$\text{sgn}(f - f_c) = 1$$

and

$$\text{sgn}(f + f_c) = 0$$

with the result that $S_{N_I N_I}(f)$ and $S_{N_Q N_Q}(f)$ are both zero.

3. For $-2f_c - B \leq f \leq -2f_c + B$, we have

$$\text{sgn}(f - f_c) = 0$$

and

$$\text{sgn}(f + f_c) = -1$$

with the result that, here also, $S_{N_I N_I}(f)$ and $S_{N_Q N_Q}(f)$ are both zero.

4. Outside the frequency intervals defined in points 1, 2, and 3, both $S_{NN}(f - f_c)$ and $S_{NN}(f + f_c)$ are zero, and in a corresponding way, $S_{N_I N_I}(f - f_c)$ and $S_{N_I N_I}(f + f_c)$ are also zero.

Combining these results, we obtain the simple relationship defined in (4.117).

As a consequence of this property, we may extract the in-phase component $n_I(t)$ and quadrature component $n_Q(t)$, except for scaling factors, from the narrowband noise $n(t)$ by using the scheme shown in Figure 4.20a, where both low-pass filters have a cutoff frequency at B . The scheme shown in Figure 4.20a may be viewed as an *analyzer*. Given the in-phase component $n_I(t)$ and the quadrature component $n_Q(t)$, we may generate the narrowband noise $n(t)$ using the scheme shown in Figure 4.20b, which may be viewed as a *synthesizer*.

PROPERTY 5 *The in-phase and quadrature components $n_I(t)$ and $n_Q(t)$ have the same variance as the narrowband noise $n(t)$.*

This property follows directly from (4.117), according to which the total area under the power spectral density curve $n_I(t)$ or $n_Q(t)$ is the same as the total area under the power spectral density curve of $n(t)$. Hence, $n_I(t)$ and $n_Q(t)$ have the same mean-square value as $n(t)$. Earlier we showed that since $n(t)$ has zero mean, then $n_I(t)$ and $n_Q(t)$ have zero mean, too. It follows, therefore, that $n_I(t)$ and $n_Q(t)$ have the same variance as the narrowband noise $n(t)$.

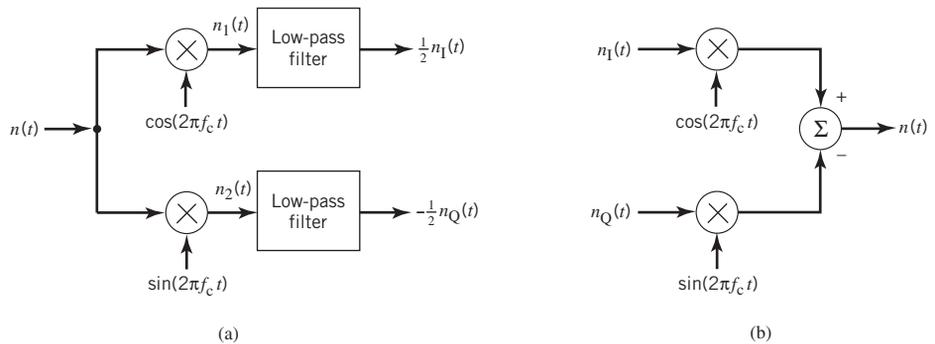


Figure 4.20 (a) Extraction of in-phase and quadrature components of a narrowband process. (b) Generation of a narrowband process from its in-phase and quadrature components.

PROPERTY 6 *The cross-spectral densities of the in-phase and quadrature components of a narrowband noise are purely imaginary, as shown by*

$$\begin{aligned} S_{N_I N_Q}(f) &= -S_{N_Q N_I}(f) \\ &= \begin{cases} j[S_N(f+f_c) - S_N(f-f_c)], & -B \leq f \leq B \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (4.120)$$

To prove this property, we take the Fourier transforms of both sides of (4.116), and use the relation of (4.118), obtaining

$$\begin{aligned} S_{N_I N_Q}(f) &= -S_{N_Q N_I}(f) \\ &= -\frac{j}{2}[S_{NN}(f-f_c) - S_{NN}(f+f_c)] \\ &\quad + \frac{j}{2}[S_{NN}(f-f_c) \operatorname{sgn}(f-f_c) + S_{NN}(f+f_c) \operatorname{sgn}(f+f_c)] \\ &= \frac{j}{2}S_{NN}(f+f_c)[1 + \operatorname{sgn}(f+f_c)] - \frac{j}{2}S_{NN}(f-f_c)[1 - \operatorname{sgn}(f-f_c)] \end{aligned} \quad (4.121)$$

Following a procedure similar to that described for proving Property 4, we find that (4.121) reduces to the form shown in (4.120).

PROPERTY 7 *If a narrowband noise $n(t)$ is Gaussian with zero mean and a power spectral density $S_{NN}(f)$ that is locally symmetric about the midband frequency $\pm f_c$, then the in-phase noise $n_I(t)$ and the quadrature noise $n_Q(t)$ are statistically independent.*

To prove this property, we observe that if $S_{NN}(f)$ is locally symmetric about $\pm f_c$, then

$$S_{NN}(f-f_c) = S_{NN}(f+f_c), \quad -B \leq f \leq B \quad (4.122)$$

Consequently, we find from (4.120) that the cross-spectral densities of the in-phase and quadrature components, $n_I(t)$ and $n_Q(t)$, are zero for all frequencies. This, in turn, means that the cross-correlation functions $S_{N_I N_Q}(f)$ and $S_{N_Q N_I}(f)$ are zero for all τ , as shown by

$$\mathbb{E}[N_I(t_k + \tau)N_Q(t_k + \tau)] = 0 \quad (4.123)$$

which implies that the random variables $N_I(t_k + \tau)$ and $N_Q(t_k)$ (obtained by observing the in-phase component at time $t_k + \tau$ and observing the quadrature component at time t_k respectively) are orthogonal for all τ .

The narrowband noise $n(t)$ is assumed to be Gaussian with zero mean; hence, from Properties 1 and 2 it follows that both $N_I(t_k + \tau)$ and $N_Q(t_k)$ are also Gaussian with zero mean. We thus conclude that because $N_I(t_k + \tau)$ and $N_Q(t_k)$ are orthogonal and have zero mean, they are uncorrelated, and being Gaussian, they are statistically independent for all τ . In other words, the in-phase component $n_I(t)$ and the quadrature component $n_Q(t)$ are statistically independent.

In light of Property 7, we may express the joint probability density function of the random variables $N_I(t_k + \tau)$ and $N_Q(t_k)$ (for any time shift τ) as the product of their individual probability density functions, as shown by

$$\begin{aligned}
f_{N_I(t_k + \tau), N_Q(t_k)}(n_I, n_Q) &= f_{N_I(t_k + \tau)}(n_I) f_{N_Q(t_k)}(n_Q) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{n_I^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{n_Q^2}{2\sigma^2}\right) \\
&= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{n_I^2 + n_Q^2}{2\sigma^2}\right)
\end{aligned} \tag{4.124}$$

where σ^2 is the variance of the original narrowband noise $n(t)$. Equation (4.124) holds if, and only if, the spectral density $S_{NN}(f)$ or $n(t)$ is locally symmetric about $\pm f_c$. Otherwise, this relation holds only for $\tau = 0$ or those values of τ for which $n_I(t)$ and $n_Q(t)$ are uncorrelated.

Summarizing Remarks

To sum up, if the narrowband noise $n(t)$ is zero mean, weakly stationary, and Gaussian, then its in-phase and quadrature components $n_I(t)$ and $n_Q(t)$ are both zero mean, jointly stationary, and jointly Gaussian. To evaluate the power spectral density of $n_I(t)$ or $n_Q(t)$, we may proceed as follows:

1. Shift the positive frequency portion of the power spectral density $S_{NN}(f)$ of the original narrowband noise $n(t)$ to the left by f_c .
2. Shift the negative frequency portion of $S_{NN}(f)$ to the right by f_c .
3. Add these two shifted spectra to obtain the desired $S_{N_I N_I}(f)$ or $S_{N_Q N_Q}(f)$.

EXAMPLE 12 Ideal Band-pass Filtered White Noise

Consider a white Gaussian noise of zero mean and power spectral density $N_0/2$, which is passed through an ideal band-pass filter of passband magnitude response equal to one, midband frequency f_c , and bandwidth $2B$. The power spectral density characteristic of the filtered noise $n(t)$ is, therefore, as shown in Figure 4.21a. The problem is to determine the autocorrelation functions of $n(t)$ and those of its in-phase and quadrature components.

The autocorrelation function of $n(t)$ is the inverse Fourier transform of the power spectral density characteristic shown in Figure 4.21a, as shown by

$$\begin{aligned}
R_{NN}(\tau) &= \int_{-f_c-B}^{-f_c+B} \frac{N_0}{2} \exp(j2\pi f\tau) df + \int_{f_c-B}^{f_c+B} \frac{N_0}{2} \exp(j2\pi f\tau) df \\
&= N_0 B \operatorname{sinc}(2B\tau) [\exp(-j2\pi f_c\tau) + \exp(j2\pi f_c\tau)] \\
&= 2N_0 B \operatorname{sinc}(2B\tau) \cos(2\pi f_c\tau)
\end{aligned} \tag{4.125}$$

which is plotted in Figure 4.21b.

The spectral density characteristic of Figure 4.21a is symmetric about $\pm f_c$. The corresponding spectral density characteristics of the in-phase noise component $n_I(t)$ and the quadrature noise component $n_Q(t)$ are equal, as shown in Figure 4.21c. Scaling the result of Example 10 by a factor of two in accordance with the spectral characteristics of

Figure 4.21a and 4.21c, we find that the autocorrelation function of $n_I(t)$ or $n_Q(t)$ is given by

$$R_{N_I N_I}(\tau) = R_{N_Q N_Q}(\tau) = 2N_0 B \operatorname{sinc}(2B\tau) \quad (4.126)$$

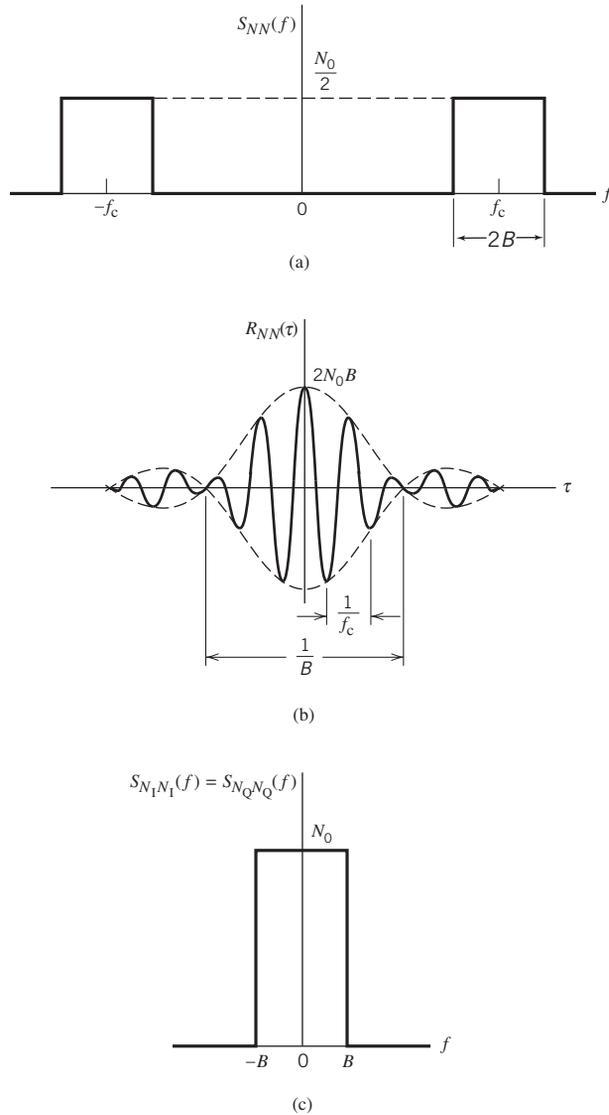


Figure 4.21 Characteristics of ideal band-pass filtered white noise: (a) power spectral density, (b) autocorrelation function, (c) power spectral density of in-phase and quadrature components.

Representation of Narrowband Noise in Terms of Envelope and Phase Components

In the preceding subsection we used the Cartesian representation of a narrowband noise $n(t)$ in terms of its in-phase and quadrature components. In this subsection we use the polar representation of the noise $n(t)$ in terms of its envelope and phase components, as shown by

$$n(t) = r(t) \cos[2\pi f_c t + \psi(t)] \quad (4.127)$$

where

$$r(t) = [n_I^2(t) + n_Q^2(t)]^{1/2} \quad (4.128)$$

and

$$\psi(t) = \tan^{-1} \left[\frac{n_Q(t)}{n_I(t)} \right] \quad (4.129)$$

The function $r(t)$ is the *envelope* of $n(t)$ and the function $\psi(t)$ is the *phase* of $n(t)$.

The probability density functions of $r(t)$ and $\psi(t)$ may be obtained from those of $n_I(t)$ and $n_Q(t)$ as follows. Let N_I and N_Q denote the random variables obtained by sampling (at some fixed time) the stochastic processes represented by the sample functions $n_I(t)$ and $n_Q(t)$ respectively. We note that N_I and N_Q are independent Gaussian random variables of zero mean and variance σ^2 , so we may express their joint probability density function as

$$f_{N_I, N_Q}(n_I, n_Q) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{n_I^2 + n_Q^2}{2\sigma^2}\right) \quad (4.130)$$

Accordingly, the probability of the joint event that N_I lies between n_I and $n_I + dn_I$ and N_Q lies between n_Q and $n_Q + dn_Q$ (i.e., the pair of random variables N_I and N_Q lies jointly inside the shaded area of Figure 4.22a) is given by

$$f_{N_I, N_Q}(n_I, n_Q) dn_I dn_Q = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{n_I^2 + n_Q^2}{2\sigma^2}\right) dn_I dn_Q \quad (4.131)$$

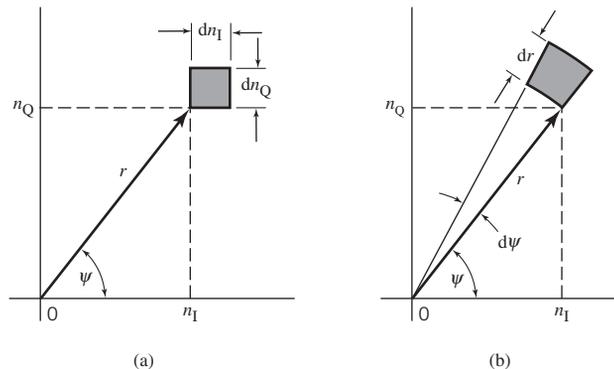


Figure 4.22

Illustrating the coordinate system for representation of narrowband noise: (a) in terms of in-phase and quadrature components; (b) in terms of envelope and phase.

where dn_I and dn_Q are incrementally small. Now, define the transformations (see Figure 4.22b)

$$n_I = r \cos \psi \quad (4.132)$$

$$n_Q = r \sin \psi \quad (4.133)$$

In a limiting sense, we may equate the two incremental areas shown shaded in parts a and b of Figure 4.22 and thus write

$$dn_I dn_Q = r dr d\psi \quad (4.134)$$

Now, let R and Ψ denote the random variables obtained by observing (at some fixed time t) the stochastic processes represented by the envelope $r(t)$ and phase $\psi(t)$ respectively. Then substituting (4.132)–(4.134) into (4.131), we find that the probability of the random variables R and Ψ lying jointly inside the shaded area of Figure 4.22b is equal to the expression

$$\frac{r}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr d\psi$$

That is, the joint probability density function of R and Ψ is given by

$$f_{R, \Psi}(r, \psi) = \frac{r}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (4.135)$$

This probability density function is independent of the angle ψ , which means that the random variables R and Ψ are *statistically independent*. We may thus express $f_{R, \Psi}(r, \psi)$ as the product of the two probability density functions: $f_R(r)$ and $f_\Psi(\psi)$. In particular, the random variable Ψ representing the phase is *uniformly distributed* inside the interval $[0, 2\pi]$, as shown by

$$f_\Psi(\psi) = \begin{cases} \frac{1}{2\pi}, & 0 \leq \psi \leq 2\pi \\ 0, & \text{elsewhere} \end{cases} \quad (4.136)$$

This result leaves the probability density function of the random variable R as

$$f_R(r) = \begin{cases} \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), & r \geq 0 \\ 0, & \text{elsewhere} \end{cases} \quad (4.137)$$

where σ^2 is the variance of the original narrowband noise $n(t)$. A random variable having the probability density function of (4.137) is said to be *Rayleigh distributed*.¹⁰

For convenience of graphical presentation, let

$$v = \frac{r}{\sigma} \quad (4.138)$$

$$f_V(v) = \sigma f_R(r) \quad (4.139)$$

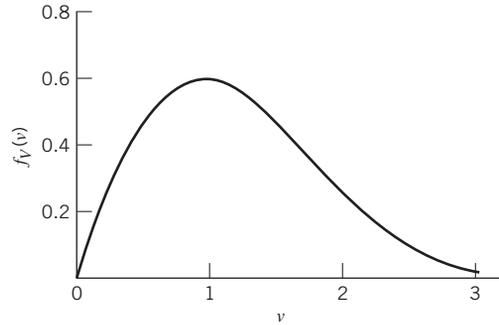


Figure 4.23 Normalized Rayleigh distribution.

Then, we may rewrite the Rayleigh distribution of (4.137) in the *normalized* form

$$f_V(v) = \begin{cases} v \exp\left(-\frac{v^2}{2}\right), & v \geq 0 \\ 0, & \text{elsewhere} \end{cases} \quad (4.140)$$

Equation (4.140) is plotted in Figure 4.23. The peak value of the distribution $f_V(v)$ occurs at $v = 1$ and is equal to 0.607. Note also that, unlike the Gaussian distribution, the Rayleigh distribution is zero for negative values of v , which follows naturally from the fact that the envelope $r(t)$ of the narrowband noise $n(t)$ can only assume nonnegative values.

4.14 Sine Wave Plus Narrowband Noise

Suppose next that we add the sinusoidal wave $A \cos(2\pi f_c t)$ to the narrowband noise $n(t)$, where A and f_c are both constants. We assume that the frequency of the sinusoidal wave is the same as the nominal carrier frequency of the noise. A sample function of the sinusoidal wave plus noise is then expressed by

$$x(t) = A \cos(2\pi f_c t) + n(t) \quad (4.141)$$

Representing the narrowband noise $n(t)$ in terms of its in-phase and quadrature components, we may write

$$x(t) = n'_1(t) \cos(2\pi f_c t) - n_Q(t) \sin(2\pi f_c t) \quad (4.142)$$

where

$$n'_1(t) = A + n_I(t) \quad (4.143)$$

We assume that $n(t)$ is Gaussian with zero mean and variance σ^2 . Accordingly, we may state the following:

1. Both $n'_1(t)$ and $n_Q(t)$ are Gaussian and statistically independent.
2. The mean of $n'_1(t)$ is A and that of $n_Q(t)$ is zero.
3. The variance of both $n'_1(t)$ and $n_Q(t)$ is σ^2 .

We may, therefore, express the joint probability density function of the random variables N'_1 and N_Q , corresponding to $n'_1(t)$ and $n_Q(t)$, as follows:

$$f_{N_1, N_Q}(n'_1, n_Q) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(n'_1 - A)^2 + n_Q^2}{2\sigma^2}\right] \quad (4.144)$$

Let $r(t)$ denote the envelope of $x(t)$ and $\psi(t)$ denote its phase. From (4.142), we thus find that

$$r(t) = \left\{ [n'_1(t)]^2 + n_Q^2(t) \right\}^{1/2} \quad (4.145)$$

and

$$\psi(t) = \tan^{-1}\left[\frac{n_Q(t)}{n'_1(t)}\right] \quad (4.146)$$

Following a procedure similar to that described in Section 4.12 for the derivation of the Rayleigh distribution, we find that the joint probability density function of the random variables R and ψ , corresponding to $r(t)$ and $\psi(t)$ for some fixed time t , is given by

$$f_{R, \psi}(r, \psi) = \frac{r}{2\pi\sigma^2} \exp\left(-\frac{r^2 + A^2 - 2Ar \cos \psi}{2\sigma^2}\right) \quad (4.147)$$

We see that in this case, however, we cannot express the joint probability density function $f_{R, \psi}(r, \psi)$ as a product $f_R(r)f_\psi(\psi)$, because we now have a term involving the values of both random variables multiplied together as $r \cos \psi$. Hence, R and ψ are *dependent* random variables for nonzero values of the amplitude A of the sinusoidal component.

We are interested, in particular, in the probability density function of R . To determine this probability density function, we integrate (4.147) over all possible values of ψ , obtaining the desired marginal density

$$\begin{aligned} f_R(r) &= \int_0^{2\pi} f_{R, \psi}(r, \psi) \, d\psi \\ &= \frac{r}{2\pi\sigma^2} \exp\left(-\frac{r^2 + A^2}{2\sigma^2}\right) \int_0^{2\pi} \exp\left(\frac{Ar \cos \psi}{\sigma^2}\right) \, d\psi \end{aligned} \quad (4.148)$$

An integral similar to that in the right-hand side of (4.148) is referred to in the literature as the *modified Bessel function of the first kind of zero order* (see Appendix C); that is,

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} \exp(x \cos \psi) \, d\psi \quad (4.149)$$

Thus, letting $x = Ar/\sigma^2$, we may rewrite (4.148) in the compact form

$$f_R(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + A^2}{2\sigma^2}\right) I_0\left(\frac{Ar}{\sigma^2}\right), \quad r \geq 0 \quad (4.150)$$

This new distribution is called the *Rician distribution*.¹¹

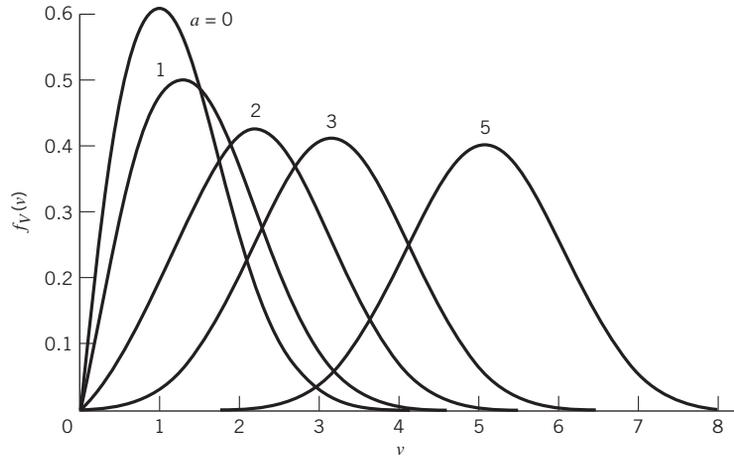


Figure 4.24 Normalized Rician distribution.

As with the Rayleigh distribution, the graphical presentation of the Rician distribution is simplified by putting

$$v = \frac{r}{\sigma} \quad (4.151)$$

$$a = \frac{A}{\sigma} \quad (4.152)$$

$$f_V(v) = \sigma f_R(r) \quad (4.153)$$

Then we may express the Rician distribution of (4.150) in the *normalized* form

$$f_V(v) = v \exp\left(-\frac{v^2 + a^2}{2}\right) I_0(av) \quad (4.154)$$

which is plotted in Figure 4.24 for the values 0, 1, 2, 3, 5, of the parameter a .¹² Based on these curves, we may make two observations:

1. When the parameter $a = 0$, and therefore $I_0(0) = 1$, the Rician distribution reduces to the Rayleigh distribution.
2. The envelope distribution is approximately Gaussian in the vicinity of $v = a$ when a is large; that is, when the sine-wave amplitude A is large compared with σ , the square root of the average power of the noise $n(t)$.

4.15 Summary and Discussion

Much of the material presented in this chapter has dealt with the characterization of a particular class of stochastic processes known to be weakly stationary. The implication of “weak” stationarity is that we may develop a partial description of a stochastic process in terms of two ensemble-averaged parameters: (1) a mean that is independent of time and (2) an autocorrelation function that depends only on the difference between the times at which two samples of the process are drawn. We also discussed ergodicity, which enables

us to use time averages as “estimates” of these parameters. The time averages are computed using a sample function (i.e., single waveform realization) of the stochastic process, evolving as a function of time.

The autocorrelation function $R_{XX}(\tau)$, expressed in terms of the time shift τ , is one way of describing the second-order statistic of a weakly (wide-sense) stationary process $X(t)$. Another equally important parameter, if not more so, for describing the second-order statistic of $X(t)$ is the power spectral density $S_{XX}(f)$, expressed in terms of the frequency f . The Fourier transform and the inverse Fourier transform formulas that relate these two parameters to each other constitute the celebrated Wiener–Khinchine equations. The first of these two equations, namely (4.42), provides the basis for a definition of the power spectral density $S_{XX}(f)$ as the Fourier transform of the autocorrelation function $R_{XX}(\tau)$, given that $R_{XX}(\tau)$ is known. This definition was arrived at by working on the output of a linear time-invariant filter, driven by a weakly stationary process $X(t)$. We also described another definition for the power spectral density $S_{XX}(f)$, described in (4.70); this second definition was derived by working directly on the process $X(t)$.

Another celebrated theorem discussed in the chapter is the Wiener–Khinchine theorem, which provides the necessary and sufficient condition for confirming the function $\rho_{XX}(\tau)$ as the normalized autocorrelation function of a weakly stationary process $X(t)$, provided that it satisfies the Fourier–Stieltjes transform, described in (4.60).

The stochastic-process theory described in this chapter also included the topic of cross-power spectral densities $S_{XY}(f)$ and $S_{YX}(f)$, involving a pair of jointly weakly stationary processes $X(t)$ and $Y(t)$, and how these two frequency-dependent parameters are related to the respective cross-correlation functions $R_{XY}(\tau)$ and $R_{YX}(\tau)$.

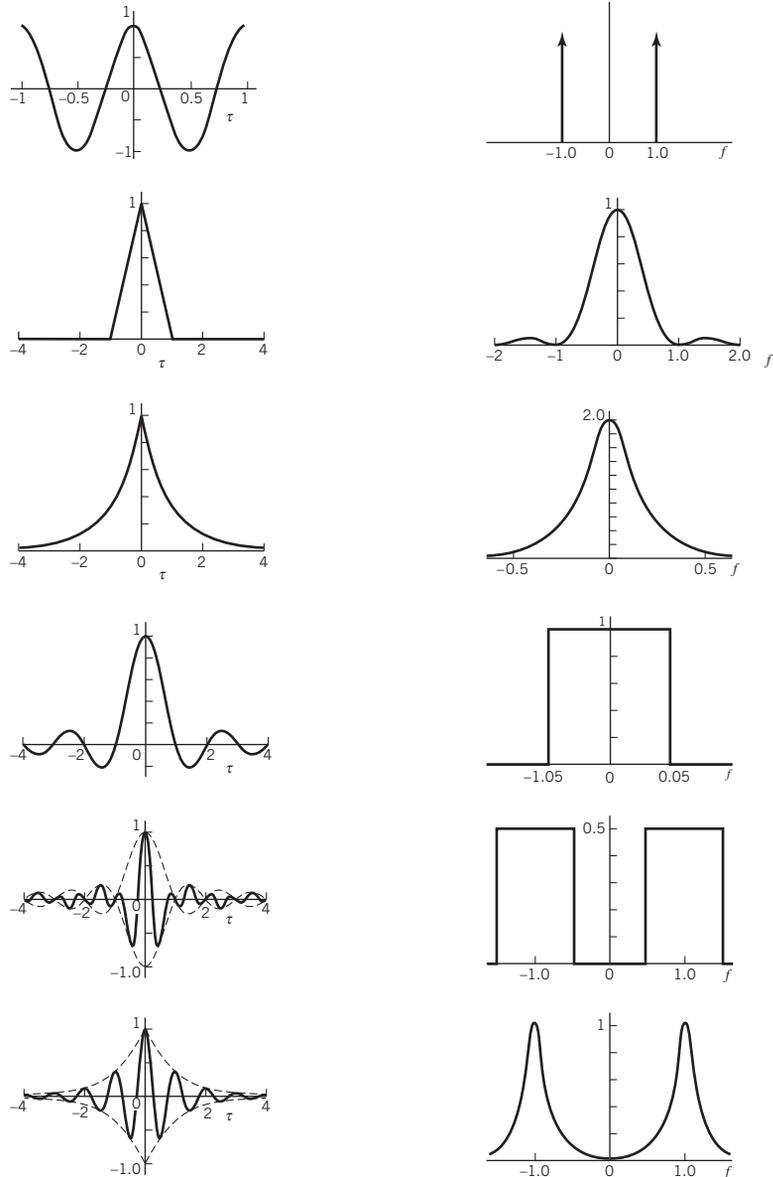
The remaining part of the chapter was devoted to the statistical characterization of different kinds of stochastic processes:

- The Poisson process, which is well-suited for the characterization of random-counting processes.
- The ubiquitous Gaussian process, which is widely used in the statistical study of communication systems.
- The two kinds of electrical noise, namely shot noise and thermal noise.
- White noise, which plays a fundamental role in the noise analysis of communication systems similar to that of the impulse function in the study of linear systems.
- Narrowband noise, which is produced by passing white noise through a linear band-pass filter. Two different methods for the description of narrowband noise were presented: one in terms of the in-phase and quadrature components and the other in terms of the envelope and phase.
- The Rayleigh distribution, which is described by the envelope of a narrowband noise process.
- The Rician distribution, which is described by the envelope of narrowband noise plus a sinusoidal component, with the midband frequency of the narrowband noise and the frequency of the sinusoidal component being coincident.

We conclude this chapter on stochastic processes by including Table 4.1, where we present a graphical summary of the autocorrelation functions and power spectral densities of important stochastic processes. All the processes described in this table are assumed to have zero mean and unit variance. This table should give the reader a feeling for (1) the

interplay between the autocorrelation function and power spectral density of a stochastic process and (2) the role of linear filtering in shaping the autocorrelation function or, equivalently, the power spectral density of a white-noise process.

Table 4.1 Graphical summary of autocorrelation functions and power spectral densities of random processes of zero mean and unit variance



Problems

Stationarity and Ergodicity

- 4.1 Consider a pair of stochastic processes $X(t)$ and $Y(t)$. In the strictly stationary world of stochastic processes, the statistical independence of $X(t)$ and $Y(t)$ corresponds to their uncorrelatedness in the world of weakly stationary processes. Justify this statement.
- 4.2 Let X_1, X_2, \dots, X_k denote a sequence obtained by uniformly sampling a stochastic process $X(t)$. The sequence consists of statistically independent and identically distributed (iid) random variables, with a common cumulative distribution function $F_X(x)$, mean μ , and variance σ^2 . Show that this sequence is strictly stationary.
- 4.3 A stochastic process $X(t)$ is defined by

$$X(t) = A \cos(2\pi f_c t)$$

where A is a Gaussian-distributed random variable of zero mean and variance σ_A^2 . The process $X(t)$ is applied to an ideal integrator, producing the output

$$Y(t) = \int_0^t X(\tau) d\tau$$

- a. Determine the probability density function of the output $Y(t)$ at a particular time t_k .
- b. Determine whether or not $Y(t)$ is strictly stationary.
- 4.4 Continuing with Problem 4.3, determine whether or not the integrator output $Y(t)$ produced in response to the input process $X(t)$ is ergodic.

Autocorrelation Function and Power Spectral Density

- 4.5 The square wave $x(t)$ of Figure P4.5, having constant amplitude A , period T_0 , and time shift t_d , represents the sample function of a stochastic process $X(t)$. The time shift t_d is a random variable, described by the probability density function

$$f_{T_d}(t_d) = \begin{cases} \frac{1}{T_0}, & -\frac{1}{2}T_0 \leq t_d \leq \frac{1}{2}T_0 \\ 0, & \text{otherwise} \end{cases}$$

- a. Determine the probability density function of the random variable $X(t_k)$, obtained by sampling the stochastic process $X(t)$ at time t_k .
- b. Determine the mean and autocorrelation function of $X(t)$ using ensemble averaging.
- c. Determine the mean and autocorrelation function of $X(t)$ using time averaging.
- d. Establish whether or not $X(t)$ is weakly stationary. In what sense is it ergodic?

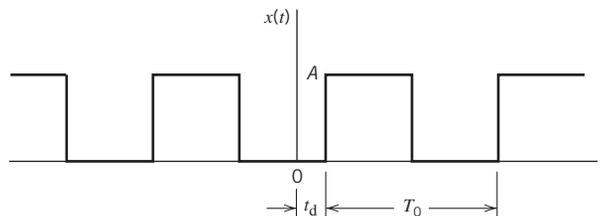


Figure P4.5

- 4.6 A binary wave consists of a random sequence of symbols 1 and 0, similar to that described in Example 6, with one basic difference: symbol 1 is now represented by a pulse of amplitude A volts,

and symbol 0 is represented by zero volts. All other parameters are the same as before. Show that this new random binary wave $X(t)$ is characterized as follows:

- a. The autocorrelation function is

$$R_{XX}(\tau) = \begin{cases} \frac{A^2}{4} + \frac{A^2}{4} \left(1 - \frac{|\tau|}{T}\right), & |\tau| < T \\ \frac{A^2}{4}, & |\tau| \geq T \end{cases}$$

- b. The power spectral density is

$$S_{XX}(f) = \frac{A^2}{4} \delta(f) + \frac{A^2 T}{4} \text{sinc}^2(fT)$$

What is the percentage power contained in the dc component of the binary wave?

- 4.7 The output of an oscillator is described by

$$X(t) = A \cos(\pi Ft + \Theta)$$

where the amplitude A is constant, and F and Θ are independent random variables. The probability density function of Θ is defined by

$$f_{\Theta}(\theta) = \begin{cases} \frac{1}{2\pi}, & 0 \leq \theta \leq 2\pi \\ 0, & \text{otherwise} \end{cases}$$

Find the power spectral density of $X(t)$ in terms of the probability density function of the frequency F . What happens to this power spectral density when the randomized frequency F assumes a constant value?

- 4.8 Equation (4.70) presents the second of two definitions introduced in the chapter for the power spectral density function, $S_{XX}(f)$, pertaining to a weakly stationary process $X(t)$. This definition reconfirms Property 3 of $S_{XX}(f)$, as shown in (4.71).
- a. Using (4.70), prove the other properties of $S_{XX}(f)$: zero correlation among frequency components, zero-frequency value, nonnegativity, symmetry, and normalization, which were discussed in Section 4.8.
- b. Starting with (4.70), derive (4.43) that defines the autocorrelation function $R_{XX}(\tau)$ of the stationary process $X(t)$ in terms of $S_{XX}(f)$.
- 4.9 In the definition of (4.70) for the power spectral density of a weakly stationary process $X(t)$, it is not permissible to interchange the order of expectation and limiting operations. Justify the validity of this statement.

The Wiener–Khinchine Theorem

In the next four problems we explore the application of the Wiener–Khinchine theorem of (4.60) to see whether a given function $\rho(\tau)$, expressed in terms of the time shift τ , is a legitimate normalized autocorrelation function or not.

- 4.10 Consider the Fourier transformable function

$$f(\tau) = \frac{A^2}{2} \sin(2\pi f_c \tau) \quad \text{for all } \tau$$

By inspection, we see that $f(\tau)$ is an odd function of τ . It cannot, therefore, be a legitimate autocorrelation function as it violates a fundamental property of the autocorrelation function. Apply the Wiener–Khinchine theorem to arrive at this same conclusion.

- 4.11 Consider the infinite series

$$f(\tau) = \frac{A^2}{2} \left[1 - \frac{1}{2!} (2\pi f_c \tau)^2 + \frac{1}{4!} (2\pi f_c \tau)^4 - \dots \right] \quad \text{for all } \tau$$

which is an even function of τ , thereby satisfying the symmetry property of the autocorrelation function. Apply the Wiener–Khinchine theorem to confirm that $f(\tau)$ is indeed a legitimate autocorrelation function of a weakly stationary process.

- 4.12 Consider the Gaussian function

$$f(\tau) = \exp(-\pi \tau^2) \quad \text{for all } \tau$$

which is Fourier transformable. Moreover, it is an even function of τ , thereby satisfying the symmetry property of the autocorrelation function around the origin $\tau = 0$. Apply the Wiener–Khinchine theorem to confirm that $f(\tau)$ is indeed a legitimate normalized autocorrelation function of a weakly stationary process.

- 4.13 Consider the Fourier transformable function

$$f(\tau) = \begin{cases} \delta\left(\tau - \frac{1}{2}\right), & \tau = \frac{1}{2} \\ -\delta\left(\tau + \frac{1}{2}\right), & \tau = -\frac{1}{2} \\ 0, & \text{otherwise} \end{cases}$$

which is an odd function of τ . It cannot, therefore, be a legitimate autocorrelation function. Apply the Wiener–Khinchine theorem to arrive at this same conclusion.

Cross-correlation Functions and Cross-spectral Densities

- 4.14 Consider a pair of weakly stationary processes $X(t)$ and $Y(t)$. Show that the cross-correlations $R_{XY}(\tau)$ and $R_{YX}(\tau)$ of these two processes have the following properties:

- $R_{XY}(\tau) = R_{YX}(-\tau)$
- $|R_{XY}(\tau)| \leq \frac{1}{2} [R_{XX}(0) + R_{YY}(0)]$

where $R_{XX}(\tau)$ and $R_{YY}(\tau)$ are the autocorrelation functions of $X(t)$ and $Y(t)$ respectively.

- 4.15 A weakly stationary process $X(t)$, with zero mean and autocorrelation function $R_{XX}(\tau)$, is passed through a differentiator, yielding the new process

$$Y(t) = \frac{d}{dt} X(t)$$

- Determine the autocorrelation function of $Y(t)$.
 - Determine the cross-correlation function between $X(t)$ and $Y(t)$.
- 4.16 Consider two linear filters connected in cascade as in Figure P4.16. Let $X(t)$ be a weakly stationary process with autocorrelation function $R_{XX}(\tau)$. The weakly stationary process appearing at the first filter output is denoted by $V(t)$ and that at the second filter output is denoted by $Y(t)$.
- Find the autocorrelation function of $Y(t)$.
 - Find the cross-correlation function $R_{VY}(\tau)$ of $V(t)$ and $Y(t)$.

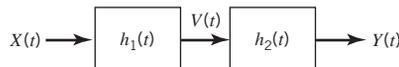


Figure P4.16

4.17 A weakly stationary process $X(t)$ is applied to a linear time-invariant filter of impulse response $h(t)$, producing the output $Y(t)$.

- a. Show that the cross-correlation function $R_{YX}(\tau)$ of the output $Y(t)$ and the input $X(t)$ is equal to the impulse response $h(\tau)$ convolved with the autocorrelation function $R_{XX}(\tau)$ of the input, as shown by

$$R_{YX}(\tau) = \int_{-\infty}^{\infty} h(u)R_{XX}(\tau - u) du$$

Show that the second cross-correlation function $R_{XY}(\tau)$ is

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h(-u)R_{XX}(\tau - u) du$$

- b. Find the cross-spectral densities $S_{YX}(f)$ and $S_{XY}(f)$.
- c. Assuming that $X(t)$ is a white-noise process with zero mean and power spectral density $N_0/2$, show that

$$R_{YX}(\tau) = \frac{N_0}{2}h(\tau)$$

Comment on the practical significance of this result.

Poisson Process

4.18 The sample function of a stochastic process $X(t)$ is shown in Figure P4.18a, where we see that the sample function $x(t)$ assumes the values ± 1 in a random manner. It is assumed that at time $t = 0$, the values $X(0) = -1$ and $X(1) = +1$ are equiprobable. From there on, the changes in $X(t)$ occur in accordance with a Poisson process of average rate λ . The process $X(t)$, described herein, is sometimes referred to as a *telegraph signal*.

- a. Show that, for any time $t > 0$, the values $X(t) = -1$ and $X(t) = +1$ are equiprobable.
- b. Building on the result of part a, show that the mean of $X(t)$ is zero and its variance is unity.
- c. Show that the autocorrelation function of $X(t)$ is given by

$$R_{XX}(\tau) = \exp(-2\lambda\tau)$$

- d. The process $X(t)$ is applied to the simple low-pass filter of Figure P4.18b. Determine the power spectral density of the process $Y(t)$ produced at the filter output.

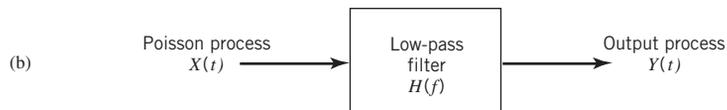
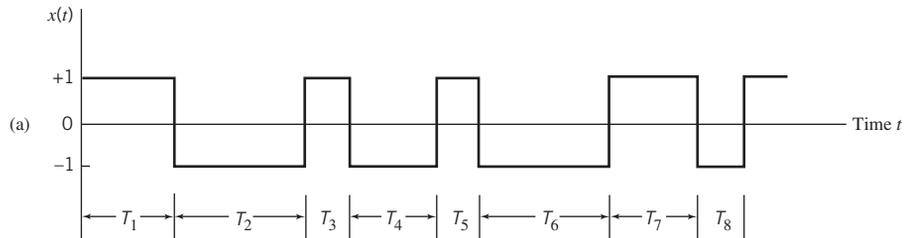


Figure P4.18

Gaussian Process

4.19 Consider the pair of integrals

$$Y_1 = \int_{-\infty}^{\infty} h_1(t)X(t) dt$$

and

$$Y_2 = \int_{-\infty}^{\infty} h_2(t)X(t) dt$$

where $X(t)$ is a Gaussian process and $h_1(t)$ and $h_2(t)$ are two different weighting functions. Show that the two random variables Y_1 and Y_2 , resulting from the integrations, are jointly Gaussian.

4.20 A Gaussian process $X(t)$, with zero mean and variance σ_X^2 , is passed through a full-wave rectifier, which is described by the input–output relationship of Figure P4.20. Show that the probability density function of the random variable $Y(t_k)$, obtained by observing the stochastic process $Y(t)$ produced at the rectifier output at time t_k , is one sided, as shown by

$$f_{Y(t_k)}(y) = \begin{cases} \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_X} \exp\left(-\frac{y^2}{2\sigma_X^2}\right), & y \geq 0 \\ 0, & y < 0 \end{cases}$$

Confirm that the total area under the graph of $f_{Y(t_k)}(y)$ is unity.

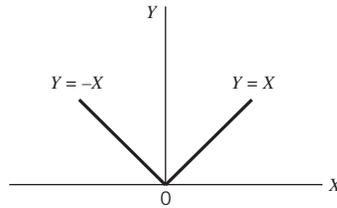


Figure P4.20

4.21 A stationary Gaussian process $X(t)$, with mean μ_X and variance σ_X^2 , is passed through two linear filters with impulse responses $h_1(t)$ and $h_2(t)$, yielding the processes $Y(t)$ and $Z(t)$, as shown in Figure P4.21. Determine the necessary and sufficient conditions, for which $Y(t_1)$ and $Z(t_2)$ are statistically independent Gaussian processes.

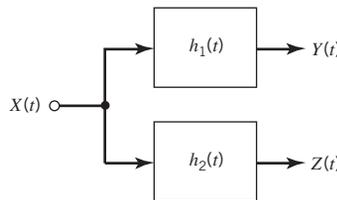


Figure P4.21

White Noise

4.22 Consider the stochastic process

$$X(t) = W(t) + aW(t - t_0)$$

where $W(t)$ is a white-noise process of power spectral density $N_0/2$ and the parameters a and t_0 are constants.

- a. Determine the autocorrelation function of the process $X(t)$, and sketch it.
- b. Determine the power spectral density of the process $X(t)$, and sketch it.

4.23 The process

$$X(t) = A \cos(2\pi f_0 t + \Theta) + W(t)$$

describes a sinusoidal process that is corrupted by an additive white-noise process $W(t)$ of known power spectral density $N_0/2$. The phase of the sinusoidal process, denoted by Θ , is a uniformly distributed random variable, defined by

$$f_{\Theta}(\theta) = \begin{cases} \frac{1}{2\pi} & \text{for } -\pi \leq \theta \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

The amplitude A and frequency f_0 are both constant but unknown.

- a. Determine the autocorrelation function of the process $X(t)$ and its power spectral density.
 - b. How would you use the two results of part a to measure the unknown parameters A and f_0 ?
- 4.24 A white Gaussian noise process of zero mean and power spectral density $N_0/2$ is applied to the filtering scheme shown in Figure P4.24. The noise at the low-pass filter output is denoted by $n(t)$.
- a. Find the power spectral density and the autocorrelation function of $n(t)$.
 - b. Find the mean and variance of $n(t)$.
 - c. What is the maximum rate at which $n(t)$ can be sampled so that the resulting samples are essentially uncorrelated?

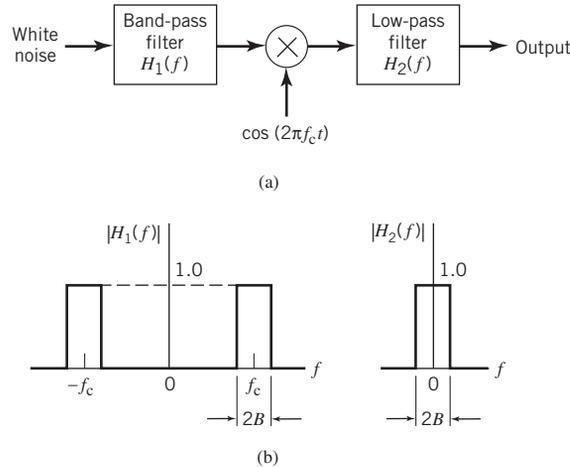


Figure P4.24

- 4.25 Let $X(t)$ be a weakly stationary process with zero mean, autocorrelation function $R_{XX}(\tau)$, and power spectral density $S_{XX}(f)$. We are required to find a linear filter with impulse response $h(t)$, such that the filter output is $X(t)$ when the input is white-noise of power spectral density $N_0/2$.
- a. Determine the condition that the impulse response $h(t)$ must satisfy in order to achieve this requirement.
 - b. What is the corresponding condition on the transfer function $H(f)$ of the filter?
 - c. Using the Paley–Wiener criterion discussed in Chapter 2, find the requirement on $S_{XX}(f)$ for the filter to be causal.

Narrowband Noise

4.26 Consider a narrowband noise $n(t)$ with its Hilbert transform denoted by $\hat{n}(t)$.

a. Show that the cross-correlation functions of $n(t)$ and $\hat{n}(t)$ are given by

$$R_{N\hat{N}}(\tau) = -\hat{R}_{NN}(\tau)$$

and

$$R_{\hat{N}N}(\tau) = \hat{R}_{NN}(\tau)$$

where $\hat{R}_{NN}(\tau)$ is the Hilbert transform of the autocorrelation function $R_{NN}(\tau)$ of $n(t)$.

Hint: use the formula

$$\hat{n}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{n(\lambda)}{t - \lambda} d\lambda$$

b. Show that, for $\tau = 0$, we have $R_{N\hat{N}}(0) = R_{\hat{N}N} = 0$.

4.27 A narrowband noise $n(t)$ has zero mean and autocorrelation function $R_{NN}(\tau)$. Its power spectral density $S_{NN}(f)$ is centered about $\pm f_c$. The in-phase and quadrature components, $n_I(t)$ and $n_Q(t)$, of $n(t)$ are defined by the weighted sums

$$n_I(t) = n(t) \cos(2\pi f_c t) + \hat{n}(t) \sin(2\pi f_c t)$$

and

$$n_Q(t) = \hat{n}(t) \cos(2\pi f_c t) - n(t) \sin(2\pi f_c t)$$

where $\hat{n}(t)$ is the Hilbert transform of the noise $n(t)$. Using the result obtained in part a of Problem 4.26, show that $n_I(t)$ and $n_Q(t)$ have the following autocorrelation functions:

$$R_{N_I N_I}(\tau) = R_{N_Q N_Q}(\tau) = R_{NN}(\tau) \cos(2\pi f_c \tau) + \hat{R}_{NN}(\tau) \sin(2\pi f_c \tau)$$

and

$$R_{N_I N_Q}(\tau) = -R_{N_Q N_I}(\tau) = R_{NN}(\tau) \sin(2\pi f_c \tau) - \hat{R}_{NN}(\tau) \cos(2\pi f_c \tau)$$

Rayleigh and Rician Distributions

4.28 Consider the problem of propagating signals through so-called *random* or *fading communications channels*. Examples of such channels include the *ionosphere* from which short-wave (high-frequency) signals are reflected back to the earth producing long-range radio transmission, and *underwater communications*. A simple model of such a channel is shown in Figure P4.28, which consists of a large collection of *random scatterers*, with the result that a single incident beam is converted into a correspondingly large number of scattered beams at the receiver. The transmitted signal is equal to $A \exp(j2\pi f_c t)$. Assume that all scattered beams travel at the same mean velocity. However, each scattered beam differs in amplitude and phase from the incident beam, so that the k th scattered beam is given by $A_k \exp(j2\pi f_c t + j\Theta_k)$, where the amplitude A_k and the phase Θ_k vary slowly and randomly with time. In particular, assume that the Θ_k are all independent of one another and uniformly distributed random variables.

a. With the received signal denoted by

$$x(t) = r(t) \exp[j2\pi f_c t + \psi(t)]$$

show that the random variable R , obtained by observing the envelope of the received signal at time t , is Rayleigh-distributed, and that the random variable Ψ , obtained by observing the phase at some fixed time, is uniformly distributed.

b. Assuming that the channel includes a line-of-sight path, so that the received signal contains a sinusoidal component of frequency f_c , show that in this case the envelope of the received signal is Rician distributed.

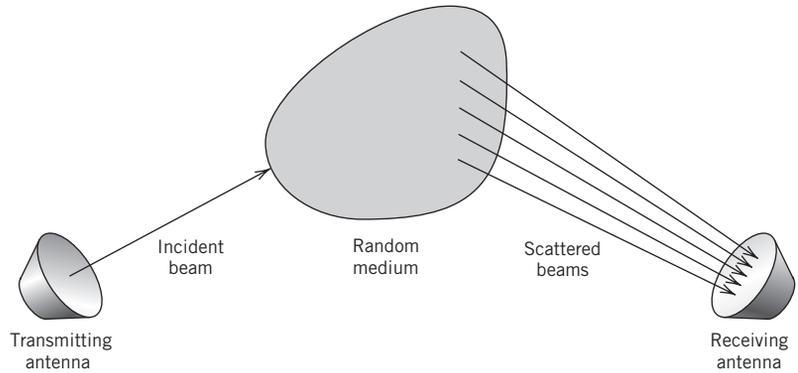


Figure P4.28

- 4.29 Referring back to the graphical plots of Figure 4.23, describing the Rician envelope distribution for varying parameter a , we see that for the parameter $a = 5$, this distribution is approximately Gaussian. Justify the validity of this statement.

Notes

1. Stochastic is of Greek origin.
2. For rigorous treatment of stochastic processes, see the classic books by Doob (1953), Loève (1963), and Cramér and Leadbetter (1967).
3. Traditionally, (4.42) and (4.43) have been referred to in the literature as the Wiener–Khintchine relations in recognition of pioneering work done by Norbert Wiener and A.I. Khintchine; for their original papers, see Wiener (1930) and Khintchine (1934). The discovery of a forgotten paper by Albert Einstein on time-series analysis (delivered at the Swiss Physical Society's February 1914 meeting in Basel) reveals that Einstein had discussed the autocorrelation function and its relationship to the spectral content of a time series many years before Wiener and Khintchine. An English translation of Einstein's paper is reproduced in the IEEE ASSP Magazine, vol. 4, October 1987. This particular issue also contains articles by W.A. Gardner and A.M. Yaglom, which elaborate on Einstein's original work.
4. For a mathematical proof of the Wiener–Khintchine theorem, see Priestley (1981).
5. Equation (4.70) provides the mathematical basis for estimating the power spectral density of a weakly stationary process. There is a plethora of procedures that have been formulated for performing this estimation. For a detailed treatment of reliable procedures to do the estimation, see the book by Percival and Walden (1993).
6. The Poisson process is named in honor of S.D. Poisson. The distribution bearing his name first appeared in an exposition by Poisson on the role of probability in the administration of justice. The classic book on Poisson processes is Snyder (1975). For an introductory treatment of the subject, see Bertsekas and Tsitsiklis (2008: Chapter 6).
7. The Gaussian distribution and the associated Gaussian process are named after the great mathematician C.F. Gauss. At age 18, Gauss invented *the method of least squares* for finding the best value of a sequence of measurements of some quantity. Gauss later used the method of least squares in fitting orbits of planets to data measurements, a procedure that was published in 1809 in his book entitled *Theory of Motion of the Heavenly Bodies*. In connection with the error of observation, he developed the *Gaussian distribution*.

8. Thermal noise was first studied experimentally by J.B. Johnson in 1928, and for this reason it is sometimes referred to as the *Johnson noise*. Johnson's experiments were confirmed theoretically by Nyquist (1928a).
9. For further insight into white noise, see Appendix I on generalized random processes in the book by Yaglom (1962).
10. The Rayleigh distribution is named in honor of the English physicist J.W. Strutt, Lord Rayleigh.
11. The Rician distribution is named in honor of S.O. Rice (1945).
12. In mobile wireless communications to be covered in Chapter 9, the sinusoidal term $A \cos(2\pi f_c t)$ in (4.141) is viewed as a *line-of-sight (LOS) component* of average power $A^2/2$ and the additive noise term $n(t)$ is viewed as a *Gaussian diffuse component* of average power σ^2 , with both being assumed to have zero mean. In such an environment, it is the *Rice factor* K that is used to characterize the Rician distribution. Formally, we write

$$\begin{aligned}
 K &= \frac{\text{Average power of the LOS component}}{\text{Average power of the diffuse component}} \\
 &= \frac{A^2}{2\sigma^2}
 \end{aligned}$$

In effect, $K = \frac{a^2}{2}$. Thus for the graphical plots of Figure 4.23, the running parameter K would assume the values 0, 0.5, 2, 4.5, 12.5.

CHAPTER 5

Information Theory

5.1 Introduction

As mentioned in Chapter 1 and reiterated along the way, the purpose of a communication system is to facilitate the transmission of signals generated by a source of information over a communication channel. But, in basic terms, what do we mean by the term information? To address this important issue, we need to understand the fundamentals of information theory.¹

The rationale for studying the fundamentals of information theory at this early stage in the book is threefold:

1. Information theory makes extensive use of probability theory, which we studied in Chapter 3; it is, therefore, a logical follow-up to that chapter.
2. It adds meaning to the term “information” used in previous chapters of the book.
3. Most importantly, information theory paves the way for many important concepts and topics discussed in subsequent chapters.

In the context of communications, information theory deals with mathematical modeling and analysis of a communication system rather than with physical sources and physical channels. In particular, it provides answers to two fundamental questions (among others):

1. What is the irreducible complexity, below which a signal cannot be compressed?
2. What is the ultimate transmission rate for reliable communication over a noisy channel?

The answers to these two questions lie in the entropy of a source and the capacity of a channel, respectively:

1. *Entropy* is defined in terms of the probabilistic behavior of a source of information; it is so named in deference to the parallel use of this concept in thermodynamics.
2. *Capacity* is defined as the intrinsic ability of a channel to convey information; it is naturally related to the noise characteristics of the channel.

A remarkable result that emerges from information theory is that if the entropy of the source is less than the capacity of the channel, then, ideally, error-free communication over the channel can be achieved. It is, therefore, fitting that we begin our study of information theory by discussing the relationships among uncertainty, information, and entropy.

5.2 Entropy

Suppose that a *probabilistic experiment* involves observation of the output emitted by a discrete source during every signaling interval. The source output is modeled as a

stochastic process, a sample of which is denoted by the discrete random variable S . This random variable takes on symbols from the fixed finite *alphabet*

$$\mathcal{S} = \{s_0, s_1, \dots, s_{K-1}\} \quad (5.1)$$

with probabilities

$$\mathbb{P}(S=s_k) = p_k, \quad k = 0, 1, \dots, K-1 \quad (5.2)$$

Of course, this set of probabilities must satisfy the normalization property

$$\sum_{k=0}^{K-1} p_k = 1, \quad p_k \geq 0 \quad (5.3)$$

We assume that the symbols emitted by the source during successive signaling intervals are statistically independent. Given such a scenario, can we find a *measure* of how much information is produced by such a source? To answer this question, we recognize that the idea of information is closely related to that of uncertainty or surprise, as described next.

Consider the event $S = s_k$, describing the emission of symbol s_k by the source with probability p_k , as defined in (5.2). Clearly, if the probability $p_k = 1$ and $p_i = 0$ for all $i \neq k$, then there is no “surprise” and, therefore, no “information” when symbol s_k is emitted, because we know what the message from the source must be. If, on the other hand, the source symbols occur with different probabilities and the probability p_k is low, then there is more surprise and, therefore, information when symbol s_k is emitted by the source than when another symbol s_i , $i \neq k$, with higher probability is emitted. Thus, the words *uncertainty*, *surprise*, and *information* are all related. Before the event $S = s_k$ occurs, there is an amount of uncertainty. When the event $S = s_k$ occurs, there is an amount of surprise. After the occurrence of the event $S = s_k$, there is gain in the amount of information, the essence of which may be viewed as the *resolution of uncertainty*. Most importantly, the amount of information is related to the inverse of the probability of occurrence of the event $S = s_k$.

We define the *amount of information* gained after observing the event $S = s_k$, which occurs with probability p_k , as the logarithmic function²

$$I(s_k) = \log\left(\frac{1}{p_k}\right) \quad (5.4)$$

which is often termed “self-information” of the event $S = s_k$. This definition exhibits the following important properties that are intuitively satisfying:

PROPERTY 1
$$I(s_k) = 0 \quad \text{for } p_k = 1 \quad (5.5)$$

Obviously, if we are absolutely *certain* of the outcome of an event, even before it occurs, there is *no* information gained.

PROPERTY 2
$$I(s_k) \geq 0 \quad \text{for } 0 \leq p_k \leq 1 \quad (5.6)$$

That is to say, the occurrence of an event $S = s_k$ either provides some or no information, but never brings about a *loss* of information.

PROPERTY 3
$$I(s_k) > I(s_i) \quad \text{for } p_k < p_i \quad (5.7)$$

That is, the less probable an event is, the more information we gain when it occurs.

PROPERTY 4

$I(s_k, s_l) = I(s_k) + I(s_l)$ if s_k and s_l are statistically independent

This additive property follows from the logarithmic definition described in (5.4).

The base of the logarithm in (5.4) specifies the units of information measure. Nevertheless, it is standard practice in information theory to use a logarithm to base 2 with binary signaling in mind. The resulting unit of information is called the *bit*, which is a contraction of the words binary digit. We thus write

$$\begin{aligned} I(s_k) &= \log_2\left(\frac{1}{p_k}\right) \\ &= -\log_2 p_k \quad \text{for } k = 0, 1, \dots, K-1 \end{aligned} \tag{5.8}$$

When $p_k = 1/2$, we have $I(s_k) = 1$ bit. We may, therefore, state:

One bit is the amount of information that we gain when one of two possible and equally likely (i.e., equiprobable) events occurs.

Note that the information $I(s_k)$ is positive, because the logarithm of a number less than one, such as a probability, is negative. Note also that if p_k is zero, then the self-information I_{s_k} assumes an unbounded value.

The amount of information $I(s_k)$ produced by the source during an arbitrary signaling interval depends on the symbol s_k emitted by the source at the time. The self-information $I(s_k)$ is a discrete random variable that takes on the values $I(s_0), I(s_1), \dots, I(s_{K-1})$ with probabilities p_0, p_1, \dots, p_{K-1} respectively. The *expectation* of $I(s_k)$ over all the probable values taken by the random variable S is given by

$$\begin{aligned} H(S) &= \mathbb{E}[I(s_k)] \\ &= \sum_{k=0}^{K-1} p_k I(s_k) \\ &= \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) \end{aligned} \tag{5.9}$$

The quantity $H(S)$ is called the *entropy*,³ formally defined as follows:

The entropy of a discrete random variable, representing the output of a source of information, is a measure of the average information content per source symbol.

Note that the entropy $H(S)$ is independent of the alphabet \mathcal{S} ; it depends only on the probabilities of the symbols in the alphabet \mathcal{S} of the source.

Properties of Entropy

Building on the definition of entropy given in (5.9), we find that entropy of the discrete random variable S is bounded as follows:

$$0 \leq H(S) \leq \log_2 K \tag{5.10}$$

where K is the number of symbols in the alphabet \mathcal{S} .

Elaborating on the two bounds on entropy in (5.10), we now make two statements:

1. $H(S) = 0$, if, and only if, the probability $p_k = 1$ for some k , and the remaining probabilities in the set are all zero; this lower bound on entropy corresponds to *no uncertainty*.
2. $H(S) = \log K$, if, and only if, $p_k = 1/K$ for all k (i.e., all the symbols in the source alphabet \mathcal{S} are equiprobable); this upper bound on entropy corresponds to maximum uncertainty.

To prove these properties of $H(S)$, we proceed as follows. First, since each probability p_k is less than or equal to unity, it follows that each term $p_k \log_2(1/p_k)$ in (5.9) is always nonnegative, so $H(S) \geq 0$. Next, we note that the product term $p_k \log_2(1/p_k)$ is zero if, and only if, $p_k = 0$ or 1. We therefore deduce that $H(S) = 0$ if, and only if, $p_k = 0$ or 1 for some k and all the rest are zero. This completes the proofs of the lower bound in (5.10) and statement 1.

To prove the upper bound in (5.10) and statement 2, we make use of a property of the natural logarithm:

$$\log_e x \leq x - 1, \quad x \geq 0 \quad (5.11)$$

where \log_e is another way of describing the *natural logarithm*, commonly denoted by \ln ; both notations are used interchangeably. This inequality can be readily verified by plotting the functions $\ln x$ and $(x - 1)$ versus x , as shown in Figure 5.1. Here we see that the line $y = x - 1$ always lies above the curve $y = \log_e x$. The equality holds only at the point $x = 1$, where the line is tangential to the curve.

To proceed with the proof, consider first any two different probability distributions denoted by p_0, p_1, \dots, p_{K-1} and q_0, q_1, \dots, q_{K-1} on the alphabet $\mathcal{S} = \{s_0, s_1, \dots, s_{K-1}\}$ of a discrete source. We may then define the *relative entropy* of these two distributions:

$$D(p||q) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{p_k}{q_k}\right) \quad (5.12)$$

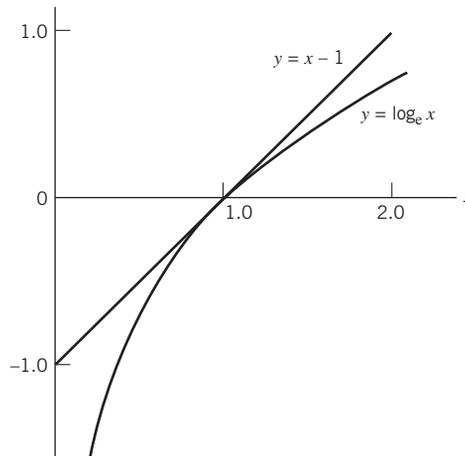


Figure 5.1 Graphs of the functions $x - 1$ and $\log x$ versus x .

Hence, changing to the natural logarithm and using the inequality of (5.11), we may express the summation on the right-hand side of (5.12) as follows:

$$\begin{aligned}
 \sum_{k=0}^{K-1} p_k \log_2 \left(\frac{p_k}{q_k} \right) &= - \sum_{k=0}^{K-1} p_k \log_2 \left(\frac{q_k}{p_k} \right) \\
 &\geq \frac{1}{\ln 2} \sum_{k=0}^{K-1} p_k \left(\frac{q_k}{p_k} - 1 \right) \\
 &= \frac{1}{\log_e 2} \sum_{k=0}^{K-1} (q_k - p_k) \\
 &= 0
 \end{aligned}$$

where, in the third line of the equation, it is noted that the sums over p_k and q_k are both equal to unity in accordance with (5.3). We thus have the *fundamental property* of probability theory:

$$D(p||q) \geq 0 \tag{5.13}$$

In words, (5.13) states:

The relative entropy of a pair of different discrete distributions is always nonnegative; it is zero only when the two distributions are identical.

Suppose we next put

$$q_k = \frac{1}{K} \quad k = 0, 1, \dots, K-1$$

which corresponds to a source alphabet \mathcal{S} with *equiprobable* symbols. Using this distribution in (5.12) yields

$$\begin{aligned}
 D(p||q) &= \sum_{k=0}^{K-1} p_k \log_2 p_k + \log_2 K \sum_{k=0}^{K-1} p_k \\
 &= -H(S) + \log_2 K
 \end{aligned}$$

where we have made use of (5.3) and (5.9). Hence, invoking the fundamental inequality of (5.13), we may finally write

$$H(S) \leq \log_2 K \tag{5.14}$$

Thus, $H(S)$ is always less than or equal to $\log_2 K$. The equality holds if, and only if, the symbols in the alphabet \mathcal{S} are equiprobable. This completes the proof of (5.10) and with it the accompanying statements 1 and 2.

EXAMPLE 1 Entropy of Bernoulli Random Variable

To illustrate the properties of $H(S)$ summed up in (5.10), consider the Bernoulli random variable for which symbol 0 occurs with probability p_0 and symbol 1 with probability $p_1 = 1 - p_0$.

The entropy of this random variable is

$$\begin{aligned} H(S) &= -p_0 \log_2 p_0 - p_1 \log_2 p_1 \\ &= -p_0 \log_2 p_0 - (1 - p_0) \log_2 (1 - p_0) \text{ bits} \end{aligned} \quad (5.15)$$

from which we observe the following:

1. When $p_0 = 0$, the entropy $H(S) = 0$; this follows from the fact that $x \log_e x \rightarrow 0$ as $x \rightarrow 0$.
2. When $p_0 = 1$, the entropy $H(S) = 0$.
3. The entropy $H(S)$ attains its maximum value $H_{\max} = 1$ bit when $p_1 = p_0 = 1/2$; that is, when symbols 1 and 0 are equally probable.

In other words, $H(S)$ is symmetric about $p_0 = 1/2$.

The function of p_0 given on the right-hand side of (5.15) is frequently encountered in information-theoretic problems. It is customary, therefore, to assign a special symbol to this function. Specifically, we define

$$H(p_0) = -p_0 \log_2 p_0 - (1 - p_0) \log_2 (1 - p_0) \quad (5.16)$$

We refer to $H(p_0)$ as the *entropy function*. The distinction between (5.15) and (5.16) should be carefully noted. The $H(S)$ of (5.15) gives the entropy of the Bernoulli random variable S . The $H(p_0)$ of (5.16), on the other hand, is a function of the prior probability p_0 defined on the interval $[0, 1]$. Accordingly, we may plot the entropy function $H(p_0)$ versus p_0 , defined on the interval $[0, 1]$, as shown in Figure 5.2. The curve in Figure 5.2 highlights the observations made under points 1, 2, and 3.

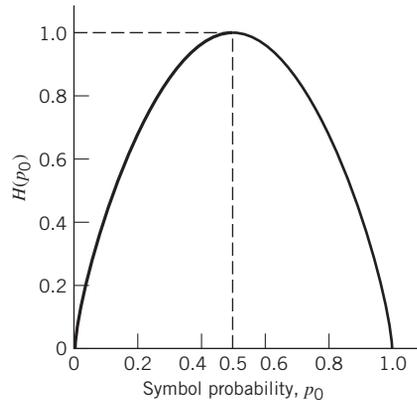


Figure 5.2
Entropy function $H(p_0)$.

Extension of a Discrete Memoryless Source

To add specificity to the discrete source of symbols that has been the focus of attention up until now, we now assume it to be *memoryless* in the sense that the symbol emitted by the source at any time is independent of previous and future emissions.

In this context, we often find it useful to consider *blocks* rather than individual symbols, with each block consisting of n successive source symbols. We may view each such block

as being produced by an *extended source* with a source alphabet described by the Cartesian product of a set S^n that has K^n distinct blocks, where K is the number of distinct symbols in the source alphabet S of the original source. With the source symbols being statistically independent, it follows that the probability of a source symbol in S^n is equal to the product of the probabilities of the n source symbols in S that constitute a particular source symbol of S^n . We may thus intuitively expect that $H(S^n)$, the entropy of the extended source, is equal to n times $H(S)$, the entropy of the original source. That is, we may write

$$H(S^{(n)}) = nH(S) \quad (5.17)$$

We illustrate the validity of this relationship by way of an example.

EXAMPLE 2 Entropy of Extended Source

Consider a discrete memoryless source with source alphabet $\mathcal{S} = \{s_0, s_1, s_2\}$, whose three distinct symbols have the following probabilities:

$$p_0 = \frac{1}{4}$$

$$p_1 = \frac{1}{4}$$

$$p_2 = \frac{1}{2}$$

Hence, the use of (5.9) yields the entropy of the discrete random variable S representing the source as

$$\begin{aligned} H(S) &= p_0 \log_2\left(\frac{1}{p_0}\right) + p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right) \\ &= \frac{1}{4} \log_2(4) + \frac{1}{4} \log_2(4) + \frac{1}{2} \log_2(2) \\ &= \frac{3}{2} \text{ bits} \end{aligned}$$

Consider next the second-order extension of the source. With the source alphabet \mathcal{S} consisting of three symbols, it follows that the source alphabet of the extended source $S^{(2)}$ has nine symbols. The first row of Table 5.1 presents the nine symbols of $S^{(2)}$, denoted by $\sigma_0, \sigma_1, \dots, \sigma_8$. The second row of the table presents the composition of these nine symbols in terms of the corresponding sequences of source symbols s_0, s_1 , and s_2 , taken two at a

Table 5.1 Alphabets of second-order extension of a discrete memoryless source

Symbols of $S^{(2)}$	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8
Corresponding sequences of symbols of S	s_0s_0	s_0s_1	s_0s_2	s_1s_0	s_1s_1	s_1s_2	s_2s_0	s_2s_1	s_2s_2
Probability $\mathbb{P}(\sigma_i)$, $i = 0, 1, \dots, 8$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

time. The probabilities of the nine source symbols of the extended source are presented in the last row of the table. Accordingly, the use of (5.9) yields the entropy of the extended source as

$$\begin{aligned}
 H(S^{(2)}) &= \sum_{i=0}^8 p(\sigma_i) \log_2\left(\frac{1}{p(\sigma_i)}\right) \\
 &= \frac{1}{16} \log_2(16) + \frac{1}{16} \log_2(16) + \frac{1}{8} \log_2(8) + \frac{1}{16} \log_2(16) \\
 &\quad + \frac{1}{16} \log_2(16) + \frac{1}{8} \log_2(8) + \frac{1}{8} \log_2(8) + \frac{1}{8} \log_2(8) + \frac{1}{4} \log_2(4) \\
 &= 3 \text{ bits}
 \end{aligned}$$

We thus see that $H(S^{(2)}) = 2H(S)$ in accordance with (5.17).

5.3 Source-coding Theorem

Now that we understand the meaning of entropy of a random variable, we are equipped to address an important issue in communication theory: the representation of data generated by a discrete source of information.

The process by which this representation is accomplished is called *source encoding*. The device that performs the representation is called a *source encoder*. For reasons to be described, it may be desirable to know the statistics of the source. In particular, if some source symbols are known to be more probable than others, then we may exploit this feature in the generation of a *source code* by assigning *short* codewords to *frequent* source symbols, and *long* codewords to *rare* source symbols. We refer to such a source code as a *variable-length code*. The *Morse code*, used in telegraphy in the past, is an example of a variable-length code. Our primary interest is in the formulation of a source encoder that satisfies two requirements:

1. The codewords produced by the encoder are in *binary* form.
2. The source code is *uniquely decodable*, so that the original source sequence can be reconstructed perfectly from the encoded binary sequence.

The second requirement is particularly important: it constitutes the basis for a *perfect source code*.

Consider then the scheme shown in Figure 5.3 that depicts a discrete memoryless source whose output s_k is converted by the source encoder into a sequence of 0s and 1s, denoted by b_k . We assume that the source has an alphabet with K different symbols and that the k th symbol s_k occurs with probability p_k , $k = 0, 1, \dots, K-1$. Let the binary

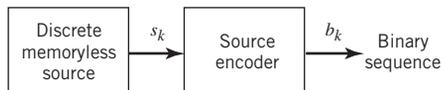


Figure 5.3 Source encoding.

codeword assigned to symbol s_k by the encoder have length l_k , measured in bits. We define the *average codeword length* \bar{L} of the source encoder as

$$\bar{L} = \sum_{k=0}^{K-1} p_k l_k \quad (5.18)$$

In physical terms, the parameter \bar{L} represents the *average number of bits per source symbol* used in the source encoding process. Let L_{\min} denote the *minimum* possible value of L . We then define the *coding efficiency* of the source encoder as

$$\eta = \frac{L_{\min}}{\bar{L}} \quad (5.19)$$

With $\bar{L} \geq L_{\min}$, we clearly have $\eta \leq 1$. The source encoder is said to be *efficient* when η approaches unity.

But how is the minimum value L_{\min} determined? The answer to this fundamental question is embodied in Shannon's first theorem: the *source-coding theorem*,⁴ which may be stated as follows:

Given a discrete memoryless source whose output is denoted by the random variable S , the entropy $H(S)$ imposes the following bound on the average codeword length \bar{L} for any source encoding scheme:

$$\bar{L} \geq H(S) \quad (5.20)$$

According to this theorem, the entropy $H(S)$ represents a *fundamental limit* on the average number of bits per source symbol necessary to represent a discrete memoryless source, in that it can be made as small as but no smaller than the entropy $H(S)$. Thus, setting $L_{\min} = H(S)$, we may rewrite (5.19), defining the efficiency of a source encoder in terms of the entropy $H(S)$ as shown by

$$\eta = \frac{H(S)}{\bar{L}} \quad (5.21)$$

where as before we have $\eta \leq 1$.

5.4 Lossless Data Compression Algorithms

A common characteristic of signals generated by physical sources is that, in their natural form, they contain a significant amount of *redundant* information, the transmission of which is therefore wasteful of primary communication resources. For example, the output of a computer used for business transactions constitutes a redundant sequence in the sense that any two adjacent symbols are typically correlated with each other.

For efficient signal transmission, the redundant information should, therefore, be removed from the signal prior to transmission. This operation, with no loss of information, is ordinarily performed on a signal in digital form, in which case we refer to the operation as *lossless data compression*. The code resulting from such an operation provides a representation of the source output that is not only efficient in terms of the average number of bits per symbol, but also exact in the sense that the original data can be reconstructed with no loss of information. The entropy of the source establishes the fundamental limit on the removal of redundancy from the data. Basically, lossless data compression is achieved

by assigning short descriptions to the most frequent outcomes of the source output and longer descriptions to the less frequent ones.

In this section we discuss some source-coding schemes for lossless data compression. We begin the discussion by describing a type of source code known as a prefix code, which not only is uniquely decodable, but also offers the possibility of realizing an average codeword length that can be made arbitrarily close to the source entropy.

Prefix Coding

Consider a discrete memoryless source of alphabet $\{s_0, s_1, \dots, s_{K-1}\}$ and respective probabilities $\{p_0, p_1, \dots, p_{K-1}\}$. For a source code representing the output of this source to be of practical use, the code has to be uniquely decodable. This restriction ensures that, for each finite sequence of symbols emitted by the source, the corresponding sequence of codewords is different from the sequence of codewords corresponding to any other source sequence. We are specifically interested in a special class of codes satisfying a restriction known as the *prefix condition*. To define the prefix condition, let the codeword assigned to source symbol s_k be denoted by $(m_{k_1}, m_{k_2}, \dots, m_{k_n})$, where the individual elements m_{k_1}, \dots, m_{k_n} are 0s and 1s and n is the codeword length. The initial part of the codeword is represented by the elements m_{k_1}, \dots, m_{k_i} for some $i \leq n$. Any sequence made up of the initial part of the codeword is called a *prefix* of the codeword. We thus say:

A prefix code is defined as a code in which no codeword is the prefix of any other codeword.

Prefix codes are distinguished from other uniquely decodable codes by the fact that the end of a codeword is always recognizable. Hence, the decoding of a prefix can be accomplished as soon as the binary sequence representing a source symbol is fully received. For this reason, prefix codes are also referred to as *instantaneous codes*.

EXAMPLE 3 Illustrative Example of Prefix Coding

To illustrate the meaning of a prefix code, consider the three source codes described in Table 5.2. Code I is not a prefix code because the bit 0, the codeword for s_0 , is a prefix of 00, the codeword for s_2 . Likewise, the bit 1, the codeword for s_1 , is a prefix of 11, the codeword for s_3 . Similarly, we may show that code III is not a prefix code but code II is.

Table 5.2 Illustrating the definition of a prefix code

Symbol source	Probability of occurrence	Code I	Code II	Code III
s_0	0.5	0	0	0
s_1	0.25	1	10	01
s_2	0.125	00	110	011
s_3	0.125	11	111	0111

Decoding of Prefix Code

To decode a sequence of codewords generated from a prefix source code, the *source decoder* simply starts at the beginning of the sequence and decodes one codeword at a time. Specifically, it sets up what is equivalent to a *decision tree*, which is a graphical portrayal of the codewords in the particular source code. For example, Figure 5.4 depicts the decision tree corresponding to code II in Table 5.2. The tree has an *initial state* and four *terminal states* corresponding to source symbols s_0 , s_1 , s_2 , and s_3 . The decoder always starts at the initial state. The first received bit moves the decoder to the terminal state s_0 if it is 0 or else to a second decision point if it is 1. In the latter case, the second bit moves the decoder one step further down the tree, either to terminal state s_1 if it is 0 or else to a third decision point if it is 1, and so on. Once each terminal state emits its symbol, the decoder is reset to its initial state. Note also that each bit in the received encoded sequence is examined only once. Consider, for example, the following encoded sequence:

$$\underbrace{10}_{s_1} \quad \underbrace{111}_{s_3} \quad \underbrace{110}_{s_2} \quad \underbrace{0}_{s_0} \quad \underbrace{0}_{s_0} \quad \dots$$

This sequence is readily decoded as the source sequence $s_1 s_3 s_2 s_0 s_0 \dots$. The reader is invited to carry out this decoding.

As mentioned previously, a prefix code has the important property that it is instantaneously decodable. But the converse is not necessarily true. For example, code III in Table 5.2 does not satisfy the prefix condition, yet it is uniquely decodable because the bit 0 indicates the beginning of each codeword in the code.

To probe more deeply into prefix codes, exemplified by that in Table 5.2, we resort to an inequality, which is considered next.

Kraft Inequality

Consider a discrete memoryless source with source alphabet $\{s_0, s_1, \dots, s_{K-1}\}$ and source probabilities $\{p_0, p_1, \dots, p_{K-1}\}$, with the codeword of symbol s_k having length l_k , $k = 0, 1,$

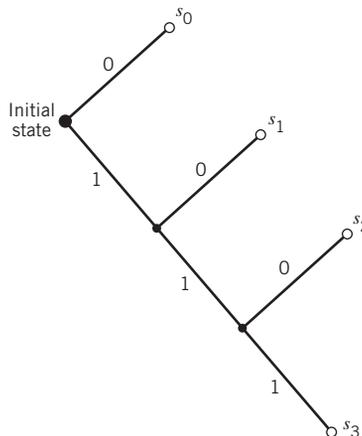


Figure 5.4 Decision tree for code II of Table 5.2.

..., $K-1$. Then, according to the *Kraft inequality*,⁵ the codeword lengths always satisfy the following inequality:

$$\sum_{k=0}^{K-1} 2^{-l_k} \leq 1 \quad (5.22)$$

where the factor 2 refers to the number of symbols in the binary alphabet. The Kraft inequality is a necessary but not sufficient condition for a source code to be a prefix code. In other words, the inequality of (5.22) is merely a condition on the codeword lengths of a prefix code and not on the codewords themselves. For example, referring to the three codes listed in Table 5.2, we see:

- Code I violates the Kraft inequality; it cannot, therefore, be a prefix code.
- The Kraft inequality is satisfied by both codes II and III, but only code II is a prefix code.

Given a discrete memoryless source of entropy $H(S)$, a prefix code can be constructed with an average codeword length \bar{L} , which is bounded as follows:

$$H(S) \leq \bar{L} < H(S) + 1 \quad (5.23)$$

The left-hand bound of (5.23) is satisfied with equality under the condition that symbol s_k is emitted by the source with probability

$$p_k = 2^{-l_k} \quad (5.24)$$

where l_k is the length of the codeword assigned to source symbol s_k . A distribution governed by (5.24) is said to be a *dyadic distribution*. For this distribution, we naturally have

$$\sum_{k=0}^{K-1} 2^{-l_k} = \sum_{k=0}^{K-1} p_k = 1$$

Under this condition, the Kraft inequality of (5.22) confirms that we can construct a prefix code, such that the length of the codeword assigned to source symbol s_k is $-\log_2 p_k$. For such a code, the average codeword length is

$$\bar{L} = \sum_{k=0}^{K-1} \frac{l_k}{2^{l_k}} \quad (5.25)$$

and the corresponding entropy of the source is

$$\begin{aligned} H(S) &= \sum_{k=0}^{K-1} \left(\frac{1}{2^{l_k}} \right) \log_2(2^{l_k}) \\ &= \sum_{k=0}^{K-1} \left(\frac{l_k}{2^{l_k}} \right) \end{aligned} \quad (5.26)$$

Hence, in this special (rather meretricious) case, we find from (5.25) and (5.26) that the prefix code is *matched* to the source in that $\bar{L} = H(S)$.

But how do we match the prefix code to an arbitrary discrete memoryless source? The answer to this basic problem lies in the use of an *extended code*. Let \bar{L}_n denote the

average codeword length of the extended prefix code. For a uniquely decodable code, \bar{L}_n is the smallest possible. From (5.23), we find that

$$nH(S) \leq \bar{L}_n < nH(S) + 1 \quad (5.27)$$

or, equivalently,

$$H(S) \leq \frac{\bar{L}_n}{n} < H(S) + \frac{1}{n} \quad (5.28)$$

In the limit, as n approaches infinity, the lower and upper bounds in (5.28) converge as shown by

$$\lim_{n \rightarrow \infty} \frac{1}{n} \bar{L}_n = H(S) \quad (5.29)$$

We may, therefore, make the statement:

By making the order n of an extended prefix source encoder large enough, we can make the code faithfully represent the discrete memoryless source S as closely as desired.

In other words, the average codeword length of an extended prefix code can be made as small as the entropy of the source, provided that the extended code has a high enough order in accordance with the source-coding theorem. However, the price we have to pay for decreasing the average codeword length is increased decoding complexity, which is brought about by the high order of the extended prefix code.

Huffman Coding

We next describe an important class of prefix codes known as Huffman codes. The basic idea behind *Huffman coding*⁶ is the construction of a simple algorithm that computes an *optimal* prefix code for a given distribution, optimal in the sense that the code has the *shortest expected length*. The end result is a source code whose average codeword length approaches the fundamental limit set by the entropy of a discrete memoryless source, namely $H(S)$. The essence of the *algorithm* used to synthesize the Huffman code is to replace the prescribed set of source statistics of a discrete memoryless source with a simpler one. This *reduction* process is continued in a step-by-step manner until we are left with a final set of only two source statistics (symbols), for which (0, 1) is an optimal code. Starting from this trivial code, we then work backward and thereby construct the Huffman code for the given source.

To be specific, the Huffman *encoding algorithm* proceeds as follows:

1. The source symbols are listed in order of decreasing probability. The two source symbols of lowest probability are assigned 0 and 1. This part of the step is referred to as the *splitting stage*.
2. These two source symbols are then *combined* into a new source symbol with probability equal to the sum of the two original probabilities. (The list of source symbols, and, therefore, source statistics, is thereby *reduced* in size by one.) The probability of the new symbol is placed in the list in accordance with its value.
3. The procedure is repeated until we are left with a final list of source statistics (symbols) of only two for which the symbols 0 and 1 are assigned.

The code for each (original) source is found by working backward and tracing the sequence of 0s and 1s assigned to that symbol as well as its successors.

EXAMPLE 4 Huffman Tree

To illustrate the construction of a Huffman code, consider the five symbols of the alphabet of a discrete memoryless source and their probabilities, which are shown in the two leftmost columns of Figure 5.5b. Following through the Huffman algorithm, we reach the end of the computation in four steps, resulting in a *Huffman tree* similar to that shown in Figure 5.5; the Huffman tree is not to be confused with the decision tree discussed previously in Figure 5.4. The codewords of the Huffman code for the source are tabulated in Figure 5.5a. The average codeword length is, therefore,

$$\begin{aligned} \bar{L} &= 0.4(2) + 0.2(2) + 0.2(2) + 0.1(3) + 0.1(3) \\ &= 2.2 \text{ binary symbols} \end{aligned}$$

The entropy of the specified discrete memoryless source is calculated as follows (see (5.9)):

$$\begin{aligned} H(S) &= 0.4 \log_2\left(\frac{1}{0.4}\right) + 0.2 \log_2\left(\frac{1}{0.2}\right) + 0.2 \log_2\left(\frac{1}{0.2}\right) + 0.1 \log_2\left(\frac{1}{0.1}\right) + 0.1 \log_2\left(\frac{1}{0.1}\right) \\ &= 0.529 + 0.464 + 0.464 + 0.332 + 0.332 \\ &= 2.121 \text{ bits} \end{aligned}$$

For this example, we may make two observations:

1. The average codeword length \bar{L} exceeds the entropy $H(S)$ by only 3.67%.
2. The average codeword length \bar{L} does indeed satisfy (5.23).

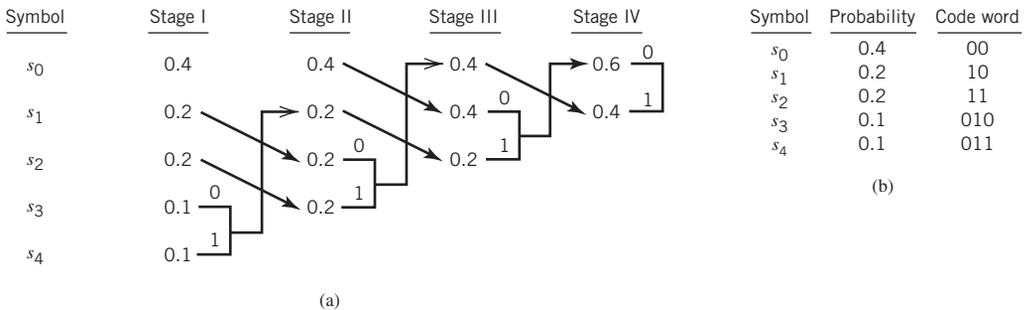


Figure 5.5 (a) Example of the Huffman encoding algorithm. (b) Source code.

It is noteworthy that the Huffman encoding process (i.e., the Huffman tree) is not unique. In particular, we may cite two variations in the process that are responsible for the nonuniqueness of the Huffman code. First, at each splitting stage in the construction of a Huffman code, there is arbitrariness in the way the symbols 0 and 1 are assigned to the last two source symbols. Whichever way the assignments are made, however, the resulting differences are trivial. Second, ambiguity arises when the probability of a *combined*

symbol (obtained by adding the last two probabilities pertinent to a particular step) is found to equal another probability in the list. We may proceed by placing the probability of the new symbol as *high* as possible, as in Example 4. Alternatively, we may place it as *low* as possible. (It is presumed that whichever way the placement is made, high or low, it is consistently adhered to throughout the encoding process.) By this time, noticeable differences arise in that the codewords in the resulting source code can have different lengths. Nevertheless, the average codeword length remains the same.

As a measure of the variability in codeword lengths of a source code, we define the *variance* of the average codeword length \bar{L} over the ensemble of source symbols as

$$\sigma^2 = \sum_{k=0}^{K-1} p_k (l_k - \bar{L})^2 \quad (5.30)$$

where p_0, p_1, \dots, p_{K-1} are the source statistics and l_k is the length of the codeword assigned to source symbol s_k . It is usually found that when a combined symbol is moved as high as possible, the resulting Huffman code has a significantly smaller variance σ^2 than when it is moved as low as possible. On this basis, it is reasonable to choose the former Huffman code over the latter.

Lempel–Ziv Coding

A drawback of the Huffman code is that it requires knowledge of a probabilistic model of the source; unfortunately, in practice, source statistics are not always known a priori. Moreover, in the modeling of text we find that storage requirements prevent the Huffman code from capturing the higher-order relationships between words and phrases because the codebook grows exponentially fast in the size of each super-symbol of letters (i.e., grouping of letters); the efficiency of the code is therefore compromised. To overcome these practical limitations of Huffman codes, we may use the *Lempel–Ziv algorithm*,⁷ which is intrinsically *adaptive* and simpler to implement than Huffman coding.

Basically, the idea behind encoding in the Lempel–Ziv algorithm is described as follows:

The source data stream is parsed into segments that are the shortest subsequences not encountered previously.

To illustrate this simple yet elegant idea, consider the example of the binary sequence

000101110010100101 ...

It is assumed that the binary symbols 0 and 1 are already stored in that order in the code book. We thus write

Subsequences stored: 0, 1
Data to be parsed: 000101110010100101 ...

The encoding process begins at the left. With symbols 0 and 1 already stored, the *shortest subsequence* of the data stream encountered for the first time and not seen before is 00; so we write

Subsequences stored: 0, 1, 00
Data to be parsed: 0101110010100101 ...

The second shortest subsequence not seen before is 01; accordingly, we go on to write

```
Subsequences stored:  0, 0, 00, 01
Data to be parsed:   01110010100101 ...
```

The next shortest subsequence not encountered previously is 011; hence, we write

```
Subsequences stored:  0, 1, 00, 01, 011
Data to be parsed:   10010100101 ...
```

We continue in the manner described here until the given data stream has been completely parsed. Thus, for the example at hand, we get the *code book* of binary subsequences shown in the second row of Figure 5.6.⁸

The first row shown in this figure merely indicates the numerical positions of the individual subsequences in the code book. We now recognize that the first subsequence of the data stream, 00, is made up of the concatenation of the *first* code book entry, 0, with itself; it is, therefore, represented by the number 11. The second subsequence of the data stream, 01, consists of the *first* code book entry, 0, concatenated with the *second* code book entry, 1; it is, therefore, represented by the number 12. The remaining subsequences are treated in a similar fashion. The complete set of numerical representations for the various subsequences in the code book is shown in the third row of Figure 5.6. As a further example illustrating the composition of this row, we note that the subsequence 010 consists of the concatenation of the subsequence 01 in position 4 and symbol 0 in position 1; hence, the numerical representation is 41. The last row shown in Figure 5.6 is the binary encoded representation of the different subsequences of the data stream.

The last symbol of each subsequence in the code book (i.e., the second row of Figure 5.6) is an *innovation symbol*, which is so called in recognition of the fact that its appendage to a particular subsequence distinguishes it from all previous subsequences stored in the code book. Correspondingly, the last bit of each uniform block of bits in the binary encoded representation of the data stream (i.e., the fourth row in Figure 5.6) represents the innovation symbol for the particular subsequence under consideration. The remaining bits provide the equivalent binary representation of the “pointer” to the *root subsequence* that matches the one in question, except for the innovation symbol.

The *Lempel–Ziv decoder* is just as simple as the encoder. Specifically, it uses the pointer to identify the root subsequence and then appends the innovation symbol. Consider, for example, the binary encoded block 1101 in position 9. The last bit, 1, is the innovation symbol. The remaining bits, 110, point to the root subsequence 10 in position 6. Hence, the block 1101 is decoded into 101, which is correct.

From the example described here, we note that, in contrast to Huffman coding, the Lempel–Ziv algorithm uses fixed-length codes to represent a variable number of source symbols; this feature makes the Lempel–Ziv code suitable for synchronous transmission.

Numerical positions	1	2	3	4	5	6	7	8	9
Subsequences	0	1	00	01	011	10	010	100	101
Numerical representations			11	12	42	21	41	61	62
Binary encoded blocks			0010	0011	1001	0100	1000	1100	1101

Figure 5.6 Illustrating the encoding process performed by the Lempel–Ziv algorithm on the binary sequence 000101110010100101 ...

In practice, fixed blocks of 12 bits long are used, which implies a code book of $2^{12} = 4096$ entries.

For a long time, Huffman coding was unchallenged as the algorithm of choice for lossless data compression; Huffman coding is still optimal, but in practice it is hard to implement. It is on account of practical implementation that the Lempel–Ziv algorithm has taken over almost completely from the Huffman algorithm. The Lempel–Ziv algorithm is now the standard algorithm for file compression.

5.5 Discrete Memoryless Channels

Up to this point in the chapter we have been preoccupied with discrete memoryless sources responsible for *information generation*. We next consider the related issue of *information transmission*. To this end, we start the discussion by considering a discrete memoryless channel, the counterpart of a discrete memoryless source.

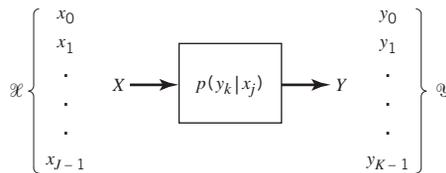
A *discrete memoryless channel* is a statistical model with an input X and an output Y that is a *noisy* version of X ; both X and Y are random variables. Every unit of time, the channel accepts an input symbol X selected from an alphabet \mathcal{X} and, in response, it emits an output symbol Y from an alphabet \mathcal{Y} . The channel is said to be “discrete” when both of the alphabets \mathcal{X} and \mathcal{Y} have *finite* sizes. It is said to be “memoryless” when the current output symbol depends *only* on the current input symbol and *not* any previous or future symbol.

Figure 5.7a shows a view of a discrete memoryless channel. The channel is described in terms of an *input alphabet*

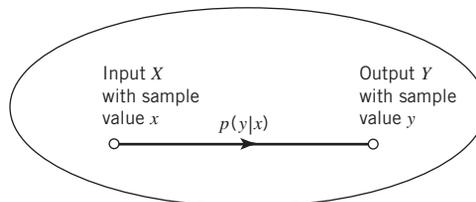
$$\mathcal{X} = \{x_0, x_1, \dots, x_{J-1}\} \quad (5.31)$$

and an *output alphabet*

$$\mathcal{Y} = \{y_0, y_1, \dots, y_{K-1}\} \quad (5.32)$$



(a)



(b)

Figure 5.7 (a) Discrete memoryless channel; (b) Simplified graphical representation of the channel.

The *cardinality* of the alphabets \mathcal{X} and \mathcal{Y} , or any other alphabet for that matter, is defined as the number of elements in the alphabet. Moreover, the channel is characterized by a set of *transition probabilities*

$$p(y_k|x_j) = \mathbb{P}(Y = y_k|X = x_j) \quad \text{for all } j \text{ and } k \quad (5.33)$$

for which, according to probability theory, we naturally have

$$0 \leq p(y_k|x_j) \leq 1 \quad \text{for all } j \text{ and } k \quad (5.34)$$

and

$$\sum_k p(y_k|x_j) = 1 \quad \text{for fixed } j \quad (5.35)$$

When the number of input symbols J and the number of output symbols K are not large, we may depict the discrete memoryless channel graphically in another way, as shown in Figure 5.7b. In this latter depiction, each input–output symbol pair (x, y) , characterized by the transition probability $p(y|x) > 0$, is joined together by a line labeled with the number $p(y|x)$.

Also, the input alphabet \mathcal{X} and output alphabet \mathcal{Y} need not have the same size; hence the use of J for the size of \mathcal{X} and K for the size of \mathcal{Y} . For example, in channel coding, the size K of the output alphabet \mathcal{Y} may be larger than the size J of the input alphabet \mathcal{X} ; thus, $K \geq J$. On the other hand, we may have a situation in which the channel emits the same symbol when either one of two input symbols is sent, in which case we have $K \leq J$.

A convenient way of describing a discrete memoryless channel is to arrange the various transition probabilities of the channel in the form of a *matrix*

$$\mathbf{P} = \begin{bmatrix} p(y_0|x_0) & p(y_1|x_0) & \cdots & p(y_{K-1}|x_0) \\ p(y_0|x_1) & p(y_1|x_1) & \cdots & p(y_{K-1}|x_1) \\ \vdots & \vdots & \ddots & \vdots \\ p(y_0|x_{J-1}) & p(y_1|x_{J-1}) & \cdots & p(y_{K-1}|x_{J-1}) \end{bmatrix} \quad (5.36)$$

The J -by- K matrix \mathbf{P} is called the *channel matrix*, or *stochastic matrix*. Note that each *row* of the channel matrix \mathbf{P} corresponds to a *fixed channel input*, whereas each *column* of the matrix corresponds to a *fixed channel output*. Note also that a fundamental property of the channel matrix \mathbf{P} , as defined here, is that the sum of the elements along any row of the stochastic matrix is always equal to one, according to (5.35).

Suppose now that the inputs to a discrete memoryless channel are selected according to the probability distribution $\{p(x_j), j = 0, 1, \dots, J-1\}$. In other words, the event that the channel input $X = x_j$ occurs with probability

$$p(x_j) = \mathbb{P}(X = x_j) \quad \text{for } j = 0, 1, \dots, J-1 \quad (5.37)$$

Having specified the random variable X denoting the channel input, we may now specify the second random variable Y denoting the channel output. The *joint probability distribution* of the random variables X and Y is given by

$$\begin{aligned} p(x_j, y_k) &= \mathbb{P}(X = x_j, Y = y_k) \\ &= \mathbb{P}(Y = y_k|X = x_j)\mathbb{P}(X = x_j) \\ &= p(y_k|x_j)p(x_j) \end{aligned} \quad (5.38)$$

The *marginal probability distribution* of the output random variable Y is obtained by averaging out the dependence of $p(x_j, y_k)$ on x_j , obtaining

$$\begin{aligned} p(y_k) &= \mathbb{P}(Y = y_k) \\ &= \sum_{j=0}^{J-1} \mathbb{P}(Y = y_k | X = x_j) \mathbb{P}(X = x_j) \\ &= \sum_{j=0}^{J-1} p(y_k | x_j) p(x_j) \quad \text{for } k = 0, 1, \dots, K-1 \end{aligned} \tag{5.39}$$

The probabilities $p(x_j)$ for $j = 0, 1, \dots, J-1$, are known as the *prior probabilities* of the various input symbols. Equation (5.39) states:

If we are given the input prior probabilities $p(x_j)$ and the stochastic matrix (i.e., the matrix of transition probabilities $p(y_k | x_j)$), then we may calculate the probabilities of the various output symbols, the $p(y_k)$.

EXAMPLE 5 Binary Symmetric Channel

The *binary symmetric channel* is of theoretical interest and practical importance. It is a special case of the discrete memoryless channel with $J = K = 2$. The channel has two input symbols ($x_0 = 0, x_1 = 1$) and two output symbols ($y_0 = 0, y_1 = 1$). The channel is symmetric because the probability of receiving 1 if 0 is sent is the same as the probability of receiving 0 if 1 is sent. This conditional probability of error is denoted by p (i.e., the probability of a bit flipping). The *transition probability diagram* of a binary symmetric channel is as shown in Figure 5.8. Correspondingly, we may express the stochastic matrix as

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

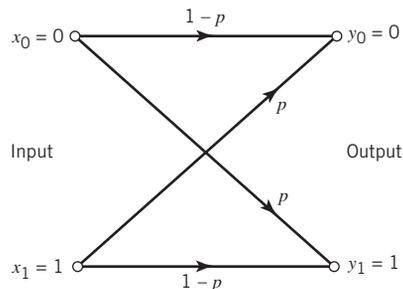


Figure 5.8 Transition probability diagram of binary symmetric channel.

5.6 Mutual Information

Given that we think of the channel output Y (selected from alphabet \mathcal{Y}) as a noisy version of the channel input X (selected from alphabet \mathcal{X}) and that the entropy $H(X)$ is a measure of the prior uncertainty about X , how can we measure the uncertainty about X after observing Y ? To answer this basic question, we extend the ideas developed in Section 5.2 by defining the *conditional entropy* of X selected from alphabet \mathcal{X} , given $Y = y_k$. Specifically, we write

$$H(X|Y = y_k) = \sum_{j=0}^{J-1} p(x_j|y_k) \log_2 \left(\frac{1}{p(x_j|y_k)} \right) \quad (5.40)$$

This quantity is itself a random variable that takes on the values $H(X|Y = y_0)$, \dots , $H(X|Y = y_{K-1})$ with probabilities $p(y_0)$, \dots , $p(y_{K-1})$, respectively. The expectation of entropy $H(X|Y = y_k)$ over the output alphabet \mathcal{Y} is therefore given by

$$\begin{aligned} H(X|Y) &= \sum_{k=0}^{K-1} H(X|Y = y_k) p(y_k) \\ &= \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} p(x_j|y_k) p(y_k) \log_2 \left(\frac{1}{p(x_j|y_k)} \right) \\ &= \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} p(x_j, y_k) \log_2 \left(\frac{1}{p(x_j|y_k)} \right) \end{aligned} \quad (5.41)$$

where, in the last line, we used the definition of the probability of the joint event ($X = x_j$, $Y = y_k$) as shown by

$$p(x_j, y_k) = p(x_j|y_k) p(y_k) \quad (5.42)$$

The quantity $H(X|Y)$ in (5.41) is called the *conditional entropy*, formally defined as follows:

The conditional entropy, $H(X|Y)$, is the average amount of uncertainty remaining about the channel input after the channel output has been observed.

The conditional entropy $H(X|Y)$ relates the channel output Y to the channel input X . The entropy $H(X)$ defines the entropy of the channel input X by itself. Given these two entropies, we now introduce the definition

$$I(X;Y) = H(X) - H(X|Y) \quad (5.43)$$

which is called the *mutual information* of the channel. To add meaning to this new concept, we recognize that the entropy $H(X)$ accounts for the uncertainty about the channel input *before* observing the channel output and the conditional entropy $H(X|Y)$ accounts for the uncertainty about the channel input *after* observing the channel output. We may, therefore, go on to make the statement:

The mutual information $I(X;Y)$ is a measure of the uncertainty about the channel input, which is resolved by observing the channel output.

Equation (5.43) is not the only way of defining the mutual information of a channel. Rather, we may define it in another way, as shown by

$$I(Y;X) = H(Y) - H(Y|X) \quad (5.44)$$

on the basis of which we may make the next statement:

The mutual information $I(Y;X)$ is a measure of the uncertainty about the channel output that is resolved by *sending* the channel input.

On first sight, the two definitions of (5.43) and (5.44) look different. In reality, however, they embody equivalent statements on the mutual information of the channel that are worded differently. More specifically, they could be used interchangeably, as demonstrated next.

Properties of Mutual Information

PROPERTY 1 Symmetry

The mutual information of a channel is symmetric in the sense that

$$I(X;Y) = I(Y;X) \quad (5.45)$$

To prove this property, we first use the formula for entropy and then use (5.35) and (5.38), in that order, obtaining

$$\begin{aligned} H(X) &= \sum_{j=0}^{J-1} p(x_j) \log_2 \left(\frac{1}{p(x_j)} \right) \\ &= \sum_{j=0}^{J-1} p(x_j) \log_2 \left(\frac{1}{p(x_j)} \right) \sum_{k=0}^{K-1} p(y_k|x_j) \\ &= \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(y_k|x_j)p(x_j) \log_2 \left(\frac{1}{p(x_j)} \right) \\ &= \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{1}{p(x_j)} \right) \end{aligned} \quad (5.46)$$

where, in going from the third to the final line, we made use of the definition of a joint probability. Hence, substituting (5.41) and (5.46) into (5.43) and then combining terms, we obtain

$$I(X;Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{p(x_j|y_k)}{p(x_j)} \right) \quad (5.47)$$

Note that the double summation on the right-hand side of (5.47) is *invariant* with respect to swapping the x and y . In other words, the symmetry of the mutual information $I(X;Y)$ is already evident from (5.47).

To further confirm this property, we may use *Bayes' rule* for conditional probabilities, previously discussed in Chapter 3, to write

$$\frac{p(x_j|y_k)}{p(x_j)} = \frac{p(y_k|x_j)}{p(y_k)} \quad (5.48)$$

Hence, substituting (5.48) into (5.47) and interchanging the order of summation, we get

$$\begin{aligned} I(X;Y) &= \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} p(x_j, y_k) \log_2 \left(\frac{p(y_k|x_j)}{p(y_k)} \right) \\ &= I(Y;X) \end{aligned} \quad (5.49)$$

which proves Property 1.

PROPERTY 2 Nonnegativity

The mutual information is always nonnegative; that is;

$$I(X;Y) \geq 0 \quad (5.50)$$

To prove this property, we first note from (5.42) that

$$p(x_j|y_k) = \frac{p(x_j, y_k)}{p(y_k)} \quad (5.51)$$

Hence, substituting (5.51) into (5.47), we may express the mutual information of the channel as

$$I(X;Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{p(x_j, y_k)}{p(x_j)p(y_k)} \right) \quad (5.52)$$

Next, a direct application of the fundamental inequality of (5.12) on relative entropy confirms (5.50), with equality if, and only if,

$$p(x_j, y_k) = p(x_j)p(y_k) \quad \text{for all } j \text{ and } k \quad (5.53)$$

In words, Property 2 states the following:

We cannot lose information, on the average, by observing the output of a channel.

Moreover, the mutual information is zero if, and only if, the input and output symbols of the channel are statistically independent; that is, when (5.53) is satisfied.

PROPERTY 3 Expansion of the Mutual Information

The mutual information of a channel is related to the joint entropy of the channel input and channel output by

$$I(X;Y) = H(X) + H(Y) - H(X, Y) \quad (5.54)$$

where the joint entropy $H(X, Y)$ is defined by

$$H(X, Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{1}{p(x_j, y_k)} \right) \quad (5.55)$$

To prove (5.54), we first rewrite the joint entropy in the equivalent form

$$H(X, Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{p(x_j)p(y_k)}{p(x_j, y_k)} \right) + \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{1}{p(x_j)p(y_k)} \right) \quad (5.56)$$

The first double summation term on the right-hand side of (5.56) is recognized as the negative of the mutual information of the channel, $I(X;Y)$, previously given in (5.52). As for the second summation term, we manipulate it as follows:

$$\begin{aligned} \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{1}{p(x_j)p(y_k)} \right) &= \sum_{j=0}^{J-1} \log_2 \left(\frac{1}{p(x_j)} \right) \sum_{k=0}^{K-1} p(x_j, y_k) \\ &\quad + \sum_{k=0}^{K-1} \log_2 \left(\frac{1}{p(y_k)} \right) \sum_{j=0}^{J-1} p(x_j, y_k) \\ &= \sum_{j=0}^{J-1} p(x_j) \log_2 \left(\frac{1}{p(x_j)} \right) + \sum_{k=0}^{K-1} p(y_k) \log_2 \left(\frac{1}{p(y_k)} \right) \\ &= H(X) + H(Y) \end{aligned} \quad (5.57)$$

where, in the first line, we made use of the following relationship from probability theory:

$$\sum_{k=0}^{K-1} p(x_j, y_k) = p(x_j)$$

and a similar relationship holds for the second line of the equation.

Accordingly, using (5.52) and (5.57) in (5.56), we get the result

$$H(X, Y) = -I(X;Y) + H(X) + H(Y) \quad (5.58)$$

which, on rearrangement, proves Property 3.

We conclude our discussion of the mutual information of a channel by providing a diagrammatic interpretation in Figure 5.9 of (5.43), (5.44), and (5.54).

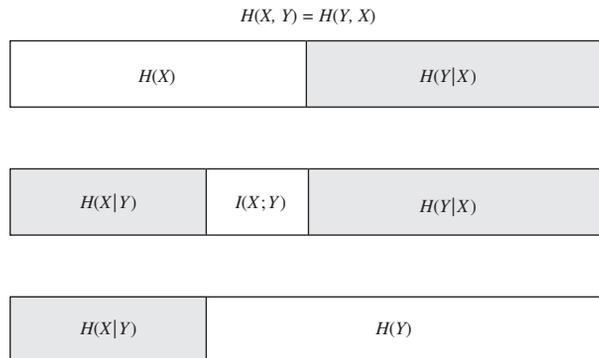


Figure 5.9 Illustrating the relations among various channel entropies.

5.7 Channel Capacity

The concept of entropy introduced in Section 5.2 prepared us for formulating Shannon's first theorem: the source-coding theorem. To set the stage for formulating Shannon's second theorem, namely the channel-coding theorem, this section introduces the concept of *capacity*, which, as mentioned previously, defines the intrinsic ability of a communication channel to convey information.

To proceed, consider a discrete memoryless channel with input alphabet \mathcal{X} , output alphabet \mathcal{Y} , and transition probabilities $p(y_k|x_j)$, where $j = 0, 1, \dots, J-1$ and $k = 0, 1, \dots, K-1$. The mutual information of the channel is defined by the first line of (5.49), which is reproduced here for convenience:

$$I(X;Y) = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} p(x_j, y_k) \log_2 \left(\frac{p(y_k|x_j)}{p(y_k)} \right)$$

where, according to (5.38),

$$p(x_j, y_k) = p(y_k|x_j)p(x_j)$$

Also, from (5.39), we have

$$p(y_k) = \sum_{j=0}^{J-1} p(y_k|x_j)p(x_j)$$

Putting these three equations into a single equation, we write

$$I(X;Y) = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} p(y_k|x_j)p(x_j) \log_2 \left(\frac{p(y_k|x_j)}{\sum_{j=0}^{J-1} p(y_k|x_j)p(x_j)} \right)$$

Careful examination of the double summation in this equation reveals two different probabilities, on which the essence of mutual information $I(X;Y)$ depends:

- the probability distribution $\{p(x_j)\}_{j=0}^{J-1}$ that characterizes the channel input and
- the transition probability distribution $\{p(y_k|x_j)\}_{j=0}^{J-1}, k=0, \dots, K-1$ that characterizes the channel itself.

These two probability distributions are obviously independent of each other. Thus, given a channel characterized by the transition probability distribution $\{p(y_k|x_j)\}$, we may now introduce the *channel capacity*, which is formally defined in terms of the mutual information between the channel input and output as follows:

$$C = \max_{\{p(x_j)\}} I(X;Y) \quad \text{bits per channel use} \quad (5.59)$$

The maximization in (5.59) is performed, subject to two input probabilistic constraints:

$$p(x_j) \geq 0 \quad \text{for all } j$$

and

$$\sum_{j=0}^{J-1} p(x_j) = 1$$

Accordingly, we make the following statement:

The channel capacity of a discrete memoryless channel, commonly denoted by C , is defined as the maximum mutual information $I(X;Y)$ in any single use of the channel (i.e., signaling interval), where the maximization is over all possible input probability distributions $\{p(x_j)\}$ on X .

The channel capacity is clearly an intrinsic property of the channel.

EXAMPLE 6

Binary Symmetric Channel (Revisited)

Consider again the *binary symmetric channel*, which is described by the *transition probability diagram* of Figure 5.8. This diagram is uniquely defined by the conditional probability of error p .

From Example 1 we recall that the entropy $H(X)$ is maximized when the channel input probability $p(x_0) = p(x_1) = 1/2$, where x_0 and x_1 are each 0 or 1. Hence, invoking the defining equation (5.59), we find that the mutual information $I(X;Y)$ is similarly maximized and thus write

$$C = I(X;Y)|_{p(x_0) = p(x_1) = 1/2}$$

From Figure 5.8 we have

$$p(y_0|x_1) = p(y_1|x_0) = p$$

and

$$p(y_0|x_0) = p(y_1|x_1) = 1 - p$$

Therefore, substituting these channel transition probabilities into (5.49) with $J = K = 2$ and then setting the input probability $p(x_0) = p(x_1) = 1/2$ in (5.59), we find that the capacity of the binary symmetric channel is

$$C = 1 + p \log_2 p + (1 - p) \log_2 (1 - p) \quad (5.60)$$

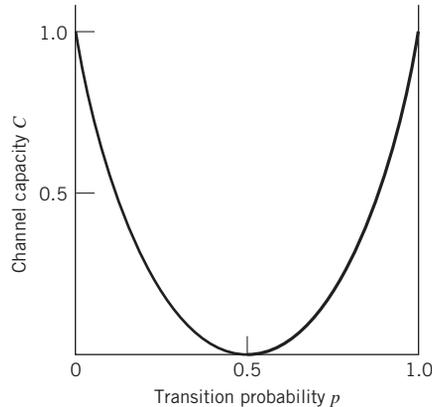
Moreover, using the definition of the entropy function introduced in (5.16), we may reduce (5.60) to

$$C = 1 - H(p)$$

The channel capacity C varies with the probability of error (i.e., transition probability) p in a convex manner as shown in Figure 5.10, which is symmetric about $p = 1/2$. Comparing the curve in this figure with that in Figure 5.2, we make two observations:

1. When the channel is *noise free*, permitting us to set $p = 0$, the channel capacity C attains its maximum value of one bit per channel use, which is exactly the information in each channel input. At this value of p , the entropy function $H(p)$ attains its minimum value of zero.
2. When the conditional probability of error $p = 1/2$ due to channel noise, the channel capacity C attains its minimum value of zero, whereas the entropy function $H(p)$

Figure 5.10
Variation of channel capacity of a binary symmetric channel with transition probability p .



attains its maximum value of unity; in such a case, the channel is said to be *useless* in the sense that the channel input and output assume statistically independent structures.

5.8 Channel-coding Theorem

With the entropy of a discrete memoryless source and the corresponding capacity of a discrete memoryless channel at hand, we are now equipped with the concepts needed for formulating Shannon's second theorem: the channel-coding theorem.

To this end, we first recognize that the inevitable presence of *noise* in a channel causes discrepancies (errors) between the output and input data sequences of a digital communication system. For a relatively noisy channel (e.g., wireless communication channel), the probability of error may reach a value as high as 10^{-1} , which means that (on the average) only 9 out of 10 transmitted bits are received correctly. For many applications, this *level of reliability* is utterly unacceptable. Indeed, a probability of error equal to 10^{-6} or even lower is often a necessary practical requirement. To achieve such a high level of performance, we resort to the use of channel coding.

The design goal of channel coding is to increase the resistance of a digital communication system to channel noise. Specifically, *channel coding* consists of *mapping* the incoming data sequence into a channel input sequence and *inverse mapping* the channel output sequence into an output data sequence in such a way that the overall effect of channel noise on the system is minimized. The first mapping operation is performed in the transmitter by a *channel encoder*, whereas the inverse mapping operation is performed in the receiver by a *channel decoder*, as shown in the block diagram of Figure 5.11; to simplify the exposition, we have not included source encoding (before channel encoding) and source decoding (after channel decoding) in this figure.⁹

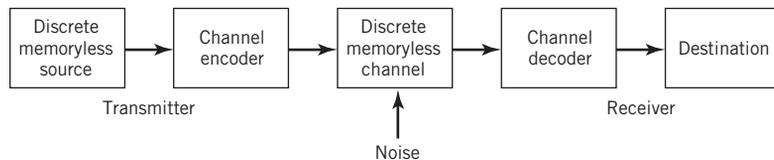


Figure 5.11
Block diagram of digital communication system.

The channel encoder and channel decoder in Figure 5.11 are both under the designer's control and should be designed to optimize the overall reliability of the communication system. The approach taken is to introduce *redundancy* in the channel encoder in a controlled manner, so as to reconstruct the original source sequence as accurately as possible. In a rather loose sense, we may thus view channel coding as the *dual* of source coding, in that the former introduces controlled redundancy to improve reliability whereas the latter reduces redundancy to improve efficiency.

Treatment of the channel-coding techniques is deferred to Chapter 10. For the purpose of our present discussion, it suffices to confine our attention to *block codes*. In this class of codes, the message sequence is subdivided into sequential blocks each k bits long, and each k -bit block is *mapped* into an n -bit block, where $n > k$. The number of redundant bits added by the encoder to each transmitted block is $n - k$ bits. The ratio k/n is called the *code rate*. Using r to denote the code rate, we write

$$r = \frac{k}{n} \quad (5.61)$$

where, of course, r is less than unity. For a prescribed k , the code rate r (and, therefore, the system's coding efficiency) approaches zero as the block length n approaches infinity.

The accurate reconstruction of the original source sequence at the destination requires that the *average probability of symbol error* be arbitrarily low. This raises the following important question:

Does a channel-coding scheme exist such that the probability that a message bit will be in error is less than any positive number ε (i.e., as small as we want it), and yet the channel-coding scheme is efficient in that the code rate need not be too small?

The answer to this fundamental question is an emphatic "yes." Indeed, the answer to the question is provided by Shannon's second theorem in terms of the channel capacity C , as described in what follows.

Up until this point, *time* has not played an important role in our discussion of channel capacity. Suppose then the discrete memoryless source in Figure 5.11 has the source alphabet \mathcal{S} and entropy $H(S)$ bits per source symbol. We assume that the source emits symbols once every T_s seconds. Hence, the *average information rate* of the source is $H(S)/T_s$ bits per second. The decoder delivers decoded symbols to the destination from the source alphabet \mathcal{S} and at the same source rate of one symbol every T_s seconds. The discrete memoryless channel has a channel capacity equal to C bits per use of the channel. We assume that the channel is capable of being used once every T_c seconds. Hence, the *channel capacity per unit time* is C/T_c bits per second, which represents the maximum rate of information transfer over the channel. With this background, we are now ready to state Shannon's second theorem, the *channel-coding theorem*,¹⁰ in two parts as follows:

1. Let a discrete memoryless source with an alphabet \mathcal{S} have entropy $H(S)$ for random variable S and produce symbols once every T_s seconds. Let a discrete memoryless channel have capacity C and be used once every T_c seconds. Then, if

$$\frac{H(S)}{T_s} \leq \frac{C}{T_c} \quad (5.62)$$

there exists a coding scheme for which the source output can be transmitted over the channel and be reconstructed with an arbitrarily small probability of error. The parameter C/T_c is called the *critical rate*; when (5.62) is satisfied with the equality sign, the system is said to be signaling at the critical rate.

2. Conversely, if

$$\frac{H(S)}{T_s} > \frac{C}{T_c}$$

it is not possible to transmit information over the channel and reconstruct it with an arbitrarily small probability of error.

The channel-coding theorem is the single most important result of information theory. The theorem specifies the channel capacity C as a *fundamental limit* on the rate at which the transmission of reliable error-free messages can take place over a discrete memoryless channel. However, it is important to note two limitations of the theorem:

1. The channel-coding theorem does not show us how to construct a good code. Rather, the theorem should be viewed as an *existence proof* in the sense that it tells us that if the condition of (5.62) is satisfied, then good codes do exist. Later, in Chapter 10, we describe good codes for discrete memoryless channels.
2. The theorem does not have a precise result for the probability of symbol error after decoding the channel output. Rather, it tells us that the probability of symbol error tends to zero as the length of the code increases, again provided that the condition of (5.62) is satisfied.

Application of the Channel-coding Theorem to Binary Symmetric Channels

Consider a discrete memoryless source that emits equally likely binary symbols (0s and 1s) once every T_s seconds. With the source entropy equal to one bit per source symbol (see Example 1), the information rate of the source is $(1/T_s)$ bits per second. The source sequence is applied to a channel encoder with code rate r . The channel encoder produces a symbol once every T_c seconds. Hence, the encoded symbol transmission rate is $(1/T_c)$ symbols per second. The channel encoder engages a binary symmetric channel once every T_c seconds. Hence, the channel capacity per unit time is (C/T_c) bits per second, where C is determined by the prescribed channel transition probability p in accordance with (5.60). Accordingly, part (1) of the channel-coding theorem implies that if

$$\frac{1}{T_s} \leq \frac{C}{T_c} \tag{5.63}$$

then the probability of error can be made arbitrarily low by the use of a suitable channel-encoding scheme. But the ratio T_c/T_s equals the code rate of the channel encoder:

$$r = \frac{T_c}{T_s} \tag{5.64}$$

Hence, we may restate the condition of (5.63) simply as

$$r \leq C$$

That is, for $r \leq C$, there exists a code (with code rate less than or equal to channel capacity C) capable of achieving an arbitrarily low probability of error.

EXAMPLE 7 Repetition Code

In this example we present a graphical interpretation of the channel-coding theorem. We also bring out a surprising aspect of the theorem by taking a look at a simple coding scheme.

Consider first a binary symmetric channel with transition probability $p = 10^{-2}$. For this value of p , we find from (5.60) that the channel capacity $C = 0.9192$. Hence, from the channel-coding theorem, we may state that, for any $\varepsilon > 0$ and $r \leq 0.9192$, there exists a code of large enough length n , code rate r , and an appropriate decoding algorithm such that, when the coded bit stream is sent over the given channel, the average probability of channel decoding error is less than ε . This result is depicted in Figure 5.12 for the limiting value $\varepsilon = 10^{-8}$.

To put the significance of this result in perspective, consider next a simple coding scheme that involves the use of a *repetition code*, in which each bit of the message is repeated several times. Let each bit (0 or 1) be repeated n times, where $n = 2m + 1$ is an odd integer. For example, for $n = 3$, we transmit 0 and 1 as 000 and 111, respectively.

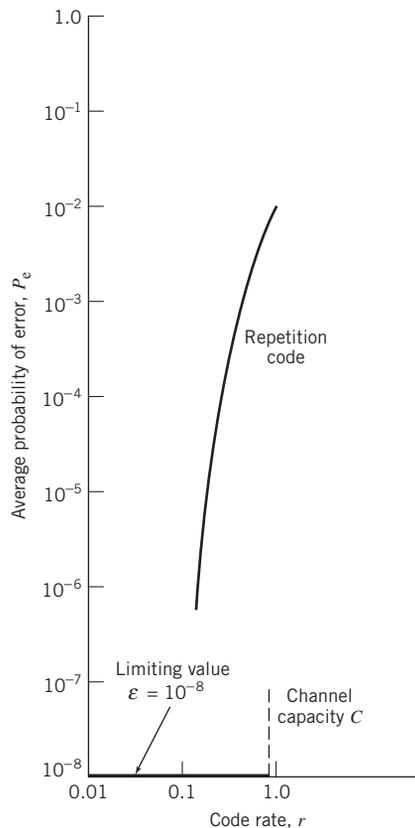


Figure 5.12 Illustrating the significance of the channel-coding theorem.

Intuitively, it would seem logical to use a *majority rule* for decoding, which operates as follows:

If in a block of n repeated bits (representing one bit of the message) the number of 0s exceeds the number of 1s, the decoder decides in favor of a 0; otherwise, it decides in favor of a 1.

Hence, an error occurs when $m + 1$ or more bits out of $n = 2m + 1$ bits are received incorrectly. Because of the assumed symmetric nature of the channel, the *average probability of error*, denoted by P_e , is independent of the *prior* probabilities of 0 and 1. Accordingly, we find that P_e is given by

$$P_e = \sum_{i=m+1}^n \binom{n}{i} p^i (1-p)^{n-i} \quad (5.65)$$

where p is the transition probability of the channel.

Table 5.3 gives the average probability of error P_e for a repetition code that is calculated by using (5.65) for different values of the code rate r . The values given here assume the use of a binary symmetric channel with transition probability $p = 10^{-2}$. The improvement in reliability displayed in Table 5.3 is achieved at the cost of decreasing code rate. The results of this table are also shown plotted as the curve labeled “repetition code” in Figure 5.12. This curve illustrates the *exchange of code rate for message reliability*, which is a characteristic of repetition codes.

This example highlights the unexpected result presented to us by the channel-coding theorem. The result is that it is not necessary to have the code rate r approach zero (as in the case of repetition codes) to achieve more and more reliable operation of the communication link. The theorem merely requires that the code rate be less than the channel capacity C .

Table 5.3 Average probability of error for repetition code

Code rate, $r = 1/n$	Average probability of error, P_e
1	10^{-2}
$\frac{1}{3}$	3×10^{-4}
$\frac{1}{5}$	10^{-6}
$\frac{1}{7}$	4×10^{-7}
$\frac{1}{9}$	10^{-8}
$\frac{1}{11}$	5×10^{-10}

5.9 Differential Entropy and Mutual Information for Continuous Random Ensembles

The sources and channels considered in our discussion of information-theoretic concepts thus far have involved ensembles of random variables that are *discrete* in amplitude. In this section, we extend these concepts to *continuous* random variables. The motivation for doing so is to pave the way for the description of another fundamental limit in information theory, which we take up in Section 5.10.

Consider a continuous random variable X with the *probability density function* $f_X(x)$. By analogy with the entropy of a discrete random variable, we introduce the following definition:

$$h(X) = \int_{-\infty}^{\infty} f_X(x) \log_2 \left[\frac{1}{f_X(x)} \right] dx \quad (5.66)$$

We refer to the new term $h(X)$ as the *differential entropy* of X to distinguish it from the ordinary or absolute entropy. We do so in recognition of the fact that, although $h(X)$ is a useful mathematical quantity to know, it is *not* in any sense a measure of the randomness of X . Nevertheless, we justify the use of (5.66) in what follows. We begin by viewing the continuous random variable X as the limiting form of a discrete random variable that assumes the value $x_k = k\Delta x$, where $k = 0, \pm 1, \pm 2, \dots$, and Δx approaches zero. By definition, the continuous random variable X assumes a value in the interval $[x_k, x_k + \Delta x]$ with probability $f_X(x_k)\Delta x$. Hence, permitting Δx to approach zero, the ordinary entropy of the continuous random variable X takes the limiting form

$$\begin{aligned} H(X) &= \lim_{\Delta x \rightarrow 0} \sum_{k=-\infty}^{\infty} f_X(x_k) \Delta x \log_2 \left(\frac{1}{f_X(x_k) \Delta x} \right) \\ &= \lim_{\Delta x \rightarrow 0} \left(\sum_{k=-\infty}^{\infty} f_X(x_k) \log_2 \left(\frac{1}{f_X(x_k)} \right) \Delta x - \log_2 \Delta x \sum_{k=-\infty}^{\infty} f_X(x_k) \Delta x \right) \\ &= \int_{-\infty}^{\infty} f_X(x) \log_2 \left(\frac{1}{f_X(x)} \right) dx - \lim_{\Delta x \rightarrow 0} \left(\log_2 \Delta x \int_{-\infty}^{\infty} f_X(x_k) dx \right) \\ &= h(X) - \lim_{\Delta x \rightarrow 0} \log_2 \Delta x \end{aligned} \quad (5.67)$$

In the last line of (5.67), use has been made of (5.66) and the fact that the total area under the curve of the probability density function $f_X(x)$ is unity. In the limit as Δx approaches zero, the term $-\log_2 \Delta x$ approaches infinity. This means that the entropy of a continuous random variable is infinitely large. Intuitively, we would expect this to be true because a continuous random variable may assume a value anywhere in the interval $(-\infty, \infty)$; we may, therefore, encounter uncountable infinite numbers of probable outcomes. To avoid the problem associated with the term $\log_2 \Delta x$, we adopt $h(X)$ as a *differential entropy*, with the term $-\log_2 \Delta x$ serving merely as a reference. Moreover, since the information transmitted over a channel is actually the difference between two entropy terms that have a common reference, the information will be the same as the difference between the corresponding

differential entropy terms. We are, therefore, perfectly justified in using the term $h(X)$, defined in (5.66), as the differential entropy of the continuous random variable X .

When we have a continuous random vector \mathbf{X} consisting of n random variables X_1, X_2, \dots, X_n , we define the differential entropy of \mathbf{X} as the n -fold integral

$$h(\mathbf{X}) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) \log_2 \left[\frac{1}{f_{\mathbf{X}}(\mathbf{x})} \right] d\mathbf{x} \quad (5.68)$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the joint probability density function of \mathbf{X} .

EXAMPLE 8 Uniform Distribution

To illustrate the notion of differential entropy, consider a random variable X uniformly distributed over the interval $(0, a)$. The probability density function of X is

$$f_X(x) = \begin{cases} \frac{1}{a}, & 0 < x < a \\ 0, & \text{otherwise} \end{cases}$$

Applying (5.66) to this distribution, we get

$$\begin{aligned} h(X) &= \int_0^a \frac{1}{a} \log(a) dx \\ &= \log a \end{aligned} \quad (5.69)$$

Note that $\log a < 0$ for $a < 1$. Thus, this example shows that, unlike a discrete random variable, the differential entropy of a continuous random variable can assume a negative value.

Relative Entropy of Continuous Distributions

In (5.12) we defined the relative entropy of a pair of different discrete distributions. To extend that definition to a pair of continuous distributions, consider the continuous random variables X and Y whose respective probability density functions are denoted by $f_X(x)$ and $f_Y(x)$ for the same sample value (argument) x . The *relative entropy*¹¹ of the random variables X and Y is defined by

$$D(f_Y || f_X) = \int_{-\infty}^{\infty} f_Y(x) \log_2 \left(\frac{f_Y(x)}{f_X(x)} \right) dx \quad (5.70)$$

where $f_X(x)$ is viewed as the “reference” distribution. In a corresponding way to the fundamental property of (5.13), we have

$$D(f_Y || f_X) \geq 0 \quad (5.71)$$

Combining (5.70) and (5.71) into a single inequality, we may thus write

$$\int_{-\infty}^{\infty} f_Y(x) \log_2 \left(\frac{1}{f_Y(x)} \right) dx \leq \int_{-\infty}^{\infty} f_Y(x) \log_2 \left(\frac{1}{f_X(x)} \right) dx$$

The expression on the left-hand side of this inequality is recognized as the differential entropy of the random variable Y , namely $h(Y)$. Accordingly,

$$h(Y) \leq \int_{-\infty}^{\infty} f_Y(x) \log_2 \left(\frac{1}{f_Y(x)} \right) dx \quad (5.72)$$

The next example illustrates an insightful application of (5.72).

EXAMPLE 9

Gaussian Distribution

Suppose two random variables, X and Y , are described as follows:

- the random variables X and Y have the common mean μ and variance σ^2 ;
- the random variable X is *Gaussian distributed* (see Section 3.9) as shown by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right] \quad (5.73)$$

Hence, substituting (5.73) into (5.72) and changing the base of the logarithm from 2 to $e = 2.7183$, we get

$$h(Y) \leq -\log_2 e \int_{-\infty}^{\infty} f_Y(x) \left[-\frac{(x-\mu)^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma) \right] dx \quad (5.74)$$

where e is the base of the natural logarithm. We now recognize the following characterizations of the random variable Y (given that its mean is μ and its variance is σ^2):

$$\int_{-\infty}^{\infty} f_Y(x) dx = 1$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 f_Y(x) dx = \sigma^2$$

We may, therefore, simplify (5.74) as

$$h(Y) \leq \frac{1}{2} \log_2(2\pi e \sigma^2) \quad (5.75)$$

The quantity on the right-hand side of (5.75) is, in fact, the differential entropy of the Gaussian random variable X :

$$h(X) = \frac{1}{2} \log_2(2\pi e \sigma^2) \quad (5.76)$$

Finally, combining (5.75) and (5.76), we may write

$$h(Y) \leq h(X), \quad \begin{cases} X: \text{Gaussian random variable} \\ Y: \text{nonGaussian random variable} \end{cases} \quad (5.77)$$

where equality holds if, and only if, $Y = X$.

We may now summarize the results of this important example by describing two entropic properties of a random variable:

PROPERTY 1 For any finite variance, a Gaussian random variable has the largest differential entropy attainable by any other random variable.

PROPERTY 2 The entropy of a Gaussian random variable is uniquely determined by its variance (i.e., the entropy is independent of the mean).

Indeed, it is because of Property 1 that the Gaussian channel model is so widely used as a conservative model in the study of digital communication systems.

Mutual Information

Continuing with the information-theoretic characterization of continuous random variables, we may use analogy with (5.47) to define the *mutual information* between the pair of continuous random variables X and Y as follows:

$$I(X;Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \log_2 \left[\frac{f_X(x|y)}{f_X(x)} \right] dx dy \quad (5.78)$$

where $f_{X,Y}(x,y)$ is the joint probability density function of X and Y and $f_X(x|y)$ is the conditional probability density function of X given $Y = y$. Also, by analogy with (5.45), (5.50), (5.43), and (5.44), we find that the mutual information between the pair of Gaussian random variables has the following properties:

$$I(X;Y) = I(Y;X) \quad (5.79)$$

$$I(X;Y) \geq 0 \quad (5.80)$$

$$\begin{aligned} I(X;Y) &= h(X) - h(X|Y) \\ &= h(Y) - h(Y|X) \end{aligned} \quad (5.81)$$

The parameter $h(X)$ is the differential entropy of X ; likewise for $h(Y)$. The parameter $h(X|Y)$ is the *conditional differential entropy* of X given Y ; it is defined by the double integral (see (5.41))

$$h(X|Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \log_2 \left[\frac{1}{f_X(x|y)} \right] dx dy \quad (5.82)$$

The parameter $h(Y|X)$ is the conditional differential entropy of Y given X ; it is defined in a manner similar to $h(X|Y)$.

5.10 Information Capacity Law

In this section we use our knowledge of probability theory to expand Shannon's channel-coding theorem, so as to formulate the information capacity for a *band-limited, power-limited Gaussian channel*, depicted in Figure 5.13. To be specific, consider a zero-mean stationary process $X(t)$ that is band-limited to B hertz. Let X_k , $k = 1, 2, \dots, K$, denote the continuous random variables obtained by uniform sampling of the process $X(t)$ at a rate of $2B$ samples per second. The rate $2B$ samples per second is the smallest permissible rate for a bandwidth B that would not result in a loss of information in accordance with the sampling theorem; this is discussed in Chapter 6. Suppose that these samples are

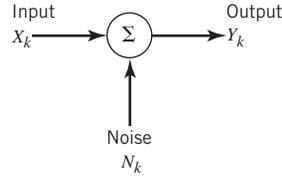


Figure 5.13 Model of discrete-time, memoryless Gaussian channel.

transmitted in T seconds over a noisy channel, also band-limited to B hertz. Hence, the total number of samples K is given by

$$K = 2BT \quad (5.83)$$

We refer to X_k as a sample of the *transmitted signal*. The channel output is perturbed by *additive white Gaussian noise* (AWGN) of zero mean and power spectral density $N_0/2$. The noise is band-limited to B hertz. Let the continuous random variables Y_k , $k = 1, 2, \dots, K$, denote the corresponding samples of the channel output, as shown by

$$Y_k = X_k + N_k, \quad k = 1, 2, \dots, K \quad (5.84)$$

The noise sample N_k in (5.84) is Gaussian with zero mean and variance

$$\sigma^2 = N_0B \quad (5.85)$$

We assume that the samples Y_k , $k = 1, 2, \dots, K$, are statistically independent.

A channel for which the noise and the received signal are as described in (5.84) and (5.85) is called a *discrete-time, memoryless Gaussian channel*, modeled as shown in Figure 5.13. To make meaningful statements about the channel, however, we have to assign a *cost* to each channel input. Typically, the transmitter is *power limited*; therefore, it is reasonable to define the cost as

$$\mathbb{E}[X_k^2] \leq P, \quad k = 1, 2, \dots, K \quad (5.86)$$

where P is the *average transmitted power*. The *power-limited Gaussian channel* described herein is not only of theoretical importance but also of practical importance, in that it models many communication channels, including line-of-sight radio and satellite links.

The *information capacity* of the channel is defined as the maximum of the mutual information between the channel input X_k and the channel output Y_k over all distributions of the input X_k that satisfy the power constraint of (5.86). Let $I(X_k; Y_k)$ denote the mutual information between X_k and Y_k . We may then define the *information capacity* of the channel as

$$C = \max_{f_{X_k}(x)} I(X_k; Y_k), \quad \text{subject to the constraint } \mathbb{E}[X_k^2] = P \quad \text{for all } k \quad (5.87)$$

In words, maximization of the mutual information $I(X_k; Y_k)$ is done with respect to all probability distributions of the channel input X_k , satisfying the power constraint $\mathbb{E}[X_k^2] = P$.

The mutual information $I(X_k; Y_k)$ can be expressed in one of the two equivalent forms shown in (5.81). For the purpose at hand, we use the second line of this equation to write

$$I(X_k; Y_k) = h(Y_k) - h(Y_k|X_k) \quad (5.88)$$

Since X_k and N_k are independent random variables and their sum equals Y_k in accordance with (5.84), we find that the conditional differential entropy of Y_k given X_k is equal to the differential entropy of N_k , as shown by

$$h(Y_k|X_k) = h(N_k) \quad (5.89)$$

Hence, we may rewrite (5.88) as

$$I(X_k; Y_k) = h(Y_k) - h(N_k) \quad (5.90)$$

With $h(N_k)$ being independent of the distribution of X_k , it follows that maximizing $I(X_k; Y_k)$ in accordance with (5.87) requires maximizing the differential entropy $h(Y_k)$. For $h(Y_k)$ to be maximum, Y_k has to be a Gaussian random variable. That is to say, samples of the channel output represent a noiselike process. Next, we observe that since N_k is Gaussian by assumption, the sample X_k of the channel input must be Gaussian too. We may therefore state that the maximization specified in (5.87) is attained by choosing samples of the channel input from a noiselike Gaussian-distributed process of average power P . Correspondingly, we may reformulate (5.87) as

$$C = I(X_k; Y_k): \text{ for Gaussian } X_k \text{ and } \mathbb{E}[X_k^2] = P \quad \text{for all } k \quad (5.91)$$

where the mutual information $I(X_k; Y_k)$ is defined in accordance with (5.90).

For evaluation of the information capacity C , we now proceed in three stages:

1. The variance of sample Y_k of the channel output equals $P + \sigma^2$, which is a consequence of the fact that the random variables X and N are statistically independent; hence, the use of (5.76) yields the differential entropy

$$h(Y_k) = \frac{1}{2} \log_2 [2\pi e(P + \sigma^2)] \quad (5.92)$$

2. The variance of the noisy sample N_k equals σ^2 ; hence, the use of (5.76) yields the differential entropy

$$h(N_k) = \frac{1}{2} \log_2 [2\pi e\sigma^2] \quad (5.93)$$

3. Substituting (5.92) and (5.93) into (5.90), and recognizing the definition of information capacity given in (5.91), we get the formula:

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma^2} \right) \text{ bits per channel use} \quad (5.94)$$

With the channel used K times for the transmission of K samples of the process $X(t)$ in T seconds, we find that the information capacity per unit time is (K/T) times the result given in (5.94). The number K equals $2BT$, as in (5.83). Accordingly, we may express the information capacity of the channel in the following equivalent form:

$$C = B \log_2 \left(1 + \frac{P}{N_0 B} \right) \text{ bits per second} \quad (5.95)$$

where $N_0 B$ is the total noise power at the channel output, defined in accordance with (5.85).

Based on the formula of (5.95), we may now make the following statement

The information capacity of a continuous channel of bandwidth B hertz, perturbed by AWGN of power spectral density $N_0/2$ and limited in bandwidth

to B , is given by the formula

$$C = B \log_2 \left(1 + \frac{P}{N_0 B} \right) \text{ bits per second}$$

where P is the average transmitted power.

The *information capacity law*¹² of (5.95) is one of the most remarkable results of Shannon's information theory. In a single formula, it highlights most vividly the interplay among three key system parameters: channel bandwidth, average transmitted power, and power spectral density of channel noise. Note, however, that the dependence of information capacity C on channel bandwidth B is *linear*, whereas its dependence on signal-to-noise ratio $P/(N_0 B)$ is *logarithmic*. Accordingly, we may make another insightful statement:

It is easier to increase the information capacity of a continuous communication channel by expanding its bandwidth than by increasing the transmitted power for a prescribed noise variance.

The information capacity formula implies that, for given average transmitted power P and channel bandwidth B , we can transmit information at the rate of C bits per second, as defined in (5.95), with arbitrarily small probability of error by employing a sufficiently complex encoding system. It is not possible to transmit at a rate higher than C bits per second by any encoding system without a definite probability of error. Hence, the channel capacity law defines the *fundamental limit* on the permissible rate of error-free transmission for a power-limited, band-limited Gaussian channel. To approach this limit, however, the transmitted signal must have statistical properties approximating those of white Gaussian noise.

Sphere Packing

To provide a plausible argument supporting the information capacity law, suppose that we use an encoding scheme that yields K codewords, one for each sample of the transmitted signal. Let n denote the length (i.e., the number of bits) of each codeword. It is presumed that the coding scheme is designed to produce an acceptably low probability of symbol error. Furthermore, the codewords satisfy the power constraint; that is, the average power contained in the transmission of each codeword with n bits is nP , where P is the average power per bit.

Suppose that any codeword in the code is transmitted. The received vector of n bits is Gaussian distributed with a mean equal to the transmitted codeword and a variance equal to $n\sigma^2$, where σ^2 is the noise variance. With a high probability, we may say that the received signal vector at the channel output lies inside a sphere of radius $\sqrt{n\sigma^2}$; that is, centered on the transmitted codeword. This sphere is itself contained in a larger sphere of radius $\sqrt{n(P + \sigma^2)}$, where $n(P + \sigma^2)$ is the average power of the received signal vector.

We may thus visualize the sphere packing¹³ as portrayed in Figure 5.14. With everything inside a small sphere of radius $\sqrt{n\sigma^2}$ assigned to the codeword on which it is

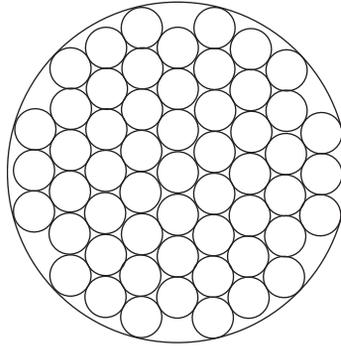


Figure 5.14 The sphere-packing problem.

centered. It is therefore reasonable to say that, when this particular codeword is transmitted, the probability that the received signal vector will lie inside the correct “decoding” sphere is high. The key question is:

How many decoding spheres can be packed inside the larger sphere of received signal vectors? In other words, how many codewords can we in fact choose?

To answer this question, we want to eliminate the overlap between the decoding spheres as depicted in Figure 5.14. Moreover, expressing the volume of an n -dimensional sphere of radius r as $A_n r^n$, where A_n is a scaling factor, we may go on to make two statements:

1. The volume of the sphere of received signal vectors is $A_n [n(P + \sigma^2)]^{n/2}$.
2. The volume of the decoding sphere is $A_n (n\sigma^2)^{n/2}$.

Accordingly, it follows that the maximum number of *nonintersecting* decoding spheres that can be packed inside the sphere of possible received signal vectors is given by

$$\begin{aligned} \frac{A_n [n(P + \sigma^2)]^{n/2}}{A_n (n\sigma^2)^{n/2}} &= \left(1 + \frac{P}{\sigma^2}\right)^{n/2} \\ &= 2^{(n/2)\log_2(1 + P/\sigma^2)} \end{aligned} \tag{5.96}$$

Taking the logarithm of this result to base 2, we readily see that the maximum number of bits per transmission for a low probability of error is indeed as defined previously in (5.94).

A final comment is in order: (5.94) is an idealized manifestation of Shannon’s channel-coding theorem, in that it provides an upper bound on the physically realizable information capacity of a communication channel.

5.11 Implications of the Information Capacity Law

Now that we have a good understanding of the information capacity law, we may go on to discuss its implications in the context of a Gaussian channel that is limited in both power

and bandwidth. For the discussion to be useful, however, we need an ideal framework against which the performance of a practical communication system can be assessed. To this end, we introduce the notion of an *ideal system*, defined as a system that transmits data at a bit rate R_b equal to the information capacity C . We may then express the average transmitted power as

$$P = E_b C \quad (5.97)$$

where E_b is the *transmitted energy per bit*. Accordingly, the ideal system is defined by the equation

$$\frac{C}{B} = \log_2 \left(1 + \frac{E_b C}{N_0 B} \right) \quad (5.98)$$

Rearranging this formula, we may define the *signal energy-per-bit to noise power spectral density ratio*, E_b/N_0 , in terms of the ratio C/B for the ideal system as follows:

$$\frac{E_b}{N_0} = \frac{2^{C/B} - 1}{C/B} \quad (5.99)$$

A plot of the bandwidth efficiency R_b/B versus E_b/N_0 is called the *bandwidth-efficiency diagram*. A generic form of this diagram is displayed in Figure 5.15, where the curve labeled “capacity boundary” corresponds to the ideal system for which $R_b = C$.

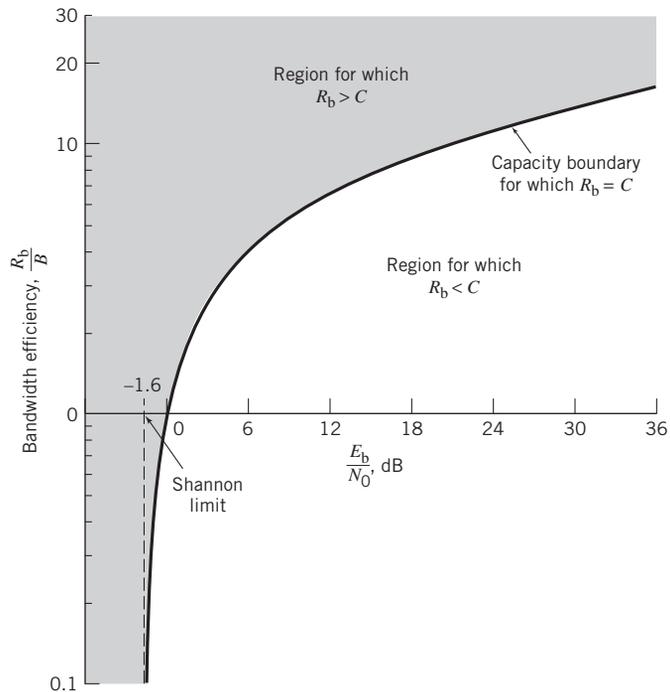


Figure 5.15 Bandwidth-efficiency diagram.

Based on Figure 5.15, we can make three observations:

1. For *infinite channel bandwidth*, the ratio E_b/N_0 approaches the limiting value

$$\begin{aligned} \left(\frac{E_b}{N_0}\right)_\infty &= \lim_{B \rightarrow \infty} \left(\frac{E_b}{N_0}\right) \\ &= \log_e 2 = 0.693 \end{aligned} \quad (5.100)$$

where \log_e stands for the natural logarithm \ln . The value defined in (5.100) is called the *Shannon limit* for an AWGN channel, assuming a code rate of zero. Expressed in decibels, the Shannon limit equals -1.6 dB. The corresponding limiting value of the channel capacity is obtained by letting the channel bandwidth B in (5.95) approach infinity, in which case we obtain

$$\begin{aligned} C_\infty &= \lim_{B \rightarrow \infty} C \\ &= \left(\frac{P}{N_0}\right) \log_2 e \end{aligned} \quad (5.101)$$

2. The *capacity boundary* is defined by the curve for the critical bit rate $R_b = C$. For any point on this boundary, we may flip a fair coin (with probability of $1/2$) whether we have error-free transmission or not. As such, the boundary separates combinations of system parameters that have the potential for supporting error-free transmission ($R_b < C$) from those for which error-free transmission is not possible ($R_b > C$). The latter region is shown shaded in Figure 5.15.
3. The diagram highlights potential *trade-offs* among three quantities: the E_b/N_0 , the ratio R_b/B , and the probability of symbol error P_e . In particular, we may view movement of the operating point along a horizontal line as trading P_e versus E_b/N_0 for a fixed R_b/B . On the other hand, we may view movement of the operating point along a vertical line as trading P_e versus R_b/B for a fixed E_b/N_0 .

EXAMPLE 10 Capacity of Binary-Input AWGN Channel

In this example, we investigate the capacity of an AWGN channel using *encoded* binary antipodal signaling (i.e., levels -1 and $+1$ for binary symbols 0 and 1, respectively). In particular, we address the issue of determining the minimum achievable bit error rate as a function of E_b/N_0 for varying code rate r . It is assumed that the binary symbols 0 and 1 are equiprobable.

Let the random variables X and Y denote the channel input and channel output respectively; X is a discrete variable, whereas Y is a continuous variable. In light of the second line of (5.81), we may express the mutual information between the channel input and channel output as

$$I(X;Y) = h(Y) - h(Y|X)$$

The second term, $h(Y|X)$, is the conditional differential entropy of the channel output Y , given the channel input X . By virtue of (5.89) and (5.93), this term is just the entropy of a Gaussian distribution. Hence, using σ^2 to denote the variance of the channel noise, we write

$$h(Y|X) = \frac{1}{2} \log_2(2\pi e \sigma^2)$$

Next, the first term, $h(Y)$, is the differential entropy of the channel output Y . With the use of binary antipodal signaling, the probability density function of Y , given $X = x$, is a mixture of two Gaussian distributions with common variance σ^2 and mean values -1 and $+1$, as shown by

$$f_Y(y_i|x) = \frac{1}{2} \left\{ \frac{\exp[-(y_i + 1)^2/2\sigma^2]}{\sqrt{2\pi}\sigma} + \frac{\exp[-(y_i - 1)^2/2\sigma^2]}{\sqrt{2\pi}\sigma} \right\} \quad (5.102)$$

Hence, we may determine the differential entropy of Y using the formula

$$h(Y) = - \int_{-\infty}^{\infty} f_Y(y_i|x) \log_2[f_Y(y_i|x)] dy_i$$

where $f_Y(y_i|x)$ is defined by (5.102). From the formulas of $h(Y|X)$ and $h(Y)$, it is clear that the mutual information is solely a function of the noise variance σ^2 . Using $M(\sigma^2)$ to denote this functional dependence, we may thus write

$$I(X;Y) = M(\sigma^2)$$

Unfortunately, there is no closed formula that we can derive for $M(\sigma^2)$ because of the difficulty of determining $h(Y)$. Nevertheless, the differential entropy $h(Y)$ can be well approximated using *Monte Carlo integration*; see Appendix E for details.

Because symbols 0 and 1 are equiprobable, it follows that the channel capacity C is equal to the mutual information between X and Y . Hence, for error-free data transmission over the AWGN channel, the code rate r must satisfy the condition

$$r < M(\sigma^2) \quad (5.103)$$

A robust measure of the ratio E_b/N_0 , is

$$\frac{E_b}{N_0} = \frac{P}{N_0 r} = \frac{P}{2\sigma^2 r}$$

where P is the average transmitted power and $N_0/2$ is the two-sided power spectral density of the channel noise. Without loss of generality, we may set $P = 1$. We may then express the noise variance as

$$\sigma^2 = \frac{N_0}{2E_b r} \quad (5.104)$$

Substituting Equation (5.104) into (5.103) and rearranging terms, we get the desired relation:

$$\frac{E_b}{N_0} = \frac{1}{2rM^{-1}(r)} \quad (5.105)$$

where $M^{-1}(r)$ is the *inverse* of the mutual information between the channel input and output, expressed as a function of the code rate r .

Using the Monte Carlo method to estimate the differential entropy $h(Y)$ and therefore $M^{-1}(r)$, the plots of Figure 5.16 are computed.¹⁴ Figure 5.16a plots the minimum E_b/N_0 versus the code rate r for error-free transmission. Figure 5.16b plots the minimum

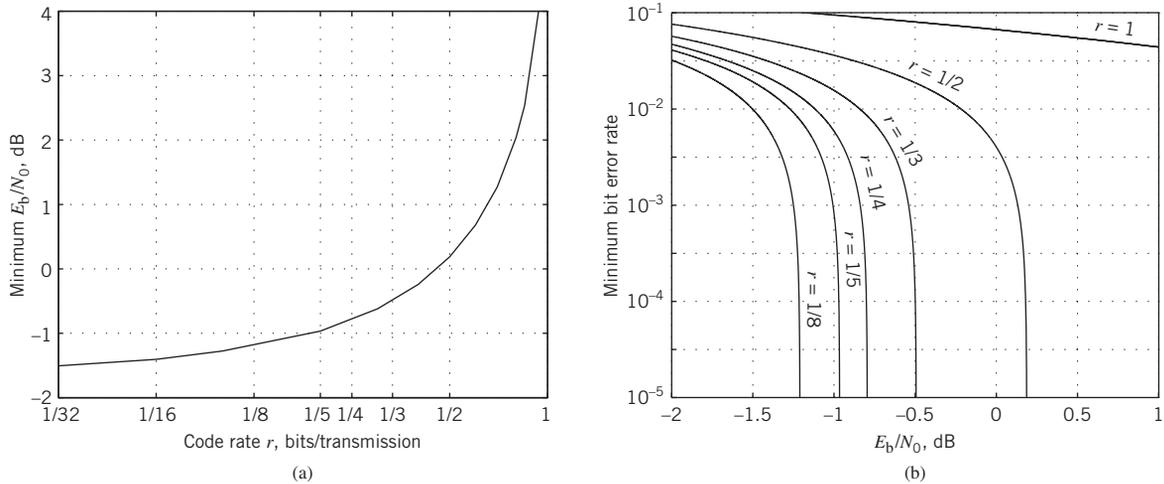


Figure 5.16 Binary antipodal signaling over an AWGN channel. (a) Minimum E_b/N_0 versus the code rate r . (b) Minimum bit error rate versus E_b/N_0 for varying code rate r .

achievable bit error rate versus E_b/N_0 with the code rate r as a running parameter. From Figure 5.16 we may draw the following conclusions:

- For uncoded binary signaling (i.e., $r = 1$), an infinite E_b/N_0 is required for error-free communication, which agrees with what we know about uncoded data transmission over an AWGN channel.
- The minimum E_b/N_0 , decreases with decreasing code rate r , which is intuitively satisfying. For example, for $r = 1/2$, the minimum value of E_b/N_0 is slightly less than 0.2 dB.
- As r approaches zero, the minimum E_b/N_0 approaches the limiting value of -1.6 dB, which agrees with the Shannon limit derived earlier; see (5.100).

5.12 Information Capacity of Colored Noisy Channel

The information capacity theorem as formulated in (5.95) applies to a band-limited white noise channel. In this section we extend Shannon's information capacity law to the more general case of a *nonwhite*, or *colored*, *noisy channel*.¹⁵ To be specific, consider the channel model shown in Figure 5.17a where the transfer function of the channel is denoted by $H(f)$. The channel noise $n(t)$, which appears additively at the channel output, is modeled as the sample function of a stationary Gaussian process of zero mean and power spectral density $S_N(f)$. The requirement is twofold:

1. Find the input ensemble, described by the power spectral density $S_{xx}(f)$, that maximizes the mutual information between the channel output $y(t)$ and the channel input $x(t)$, subject to the constraint that the average power of $x(t)$ is fixed at a constant value P .
2. Hence, determine the optimum information capacity of the channel.

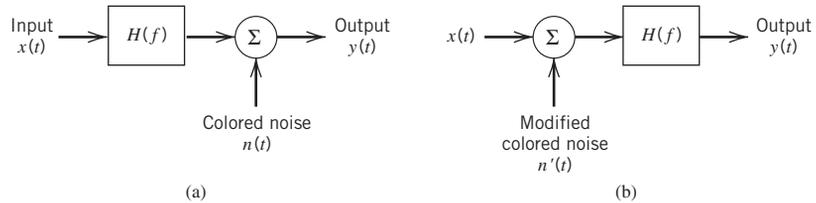


Figure 5.17 (a) Model of band-limited, power-limited noisy channel. (b) Equivalent model of the channel.

This problem is a constrained optimization problem. To solve it, we proceed as follows:

- Because the channel is linear, we may replace the model of Figure 5.17a with the equivalent model shown in Figure 5.17b. From the viewpoint of the spectral characteristics of the signal plus noise measured at the channel output, the two models of Figure 5.17 are equivalent, provided that the power spectral density of the noise $n'(t)$ in Figure 5.17b is defined in terms of the power spectral density of the noise $n(t)$ in Figure 5.17a as

$$S_{N'N'}(f) = \frac{S_{NN}(f)}{|H(f)|^2} \quad (5.106)$$

where $|H(f)|$ is the magnitude response of the channel.

- To simplify the analysis, we use the “principle of divide and conquer” to approximate the continuous $|H(f)|$ described as a function of frequency f in the form of a staircase, as illustrated in Figure 5.18. Specifically, the channel is divided into a large number of adjoining frequency slots. The smaller we make the incremental frequency interval Δf of each subchannel, the better this approximation is.

The net result of these two points is that the original model of Figure 5.17a is replaced by the parallel combination of a finite number of subchannels, N , each of which is corrupted essentially by “band-limited white Gaussian noise.”

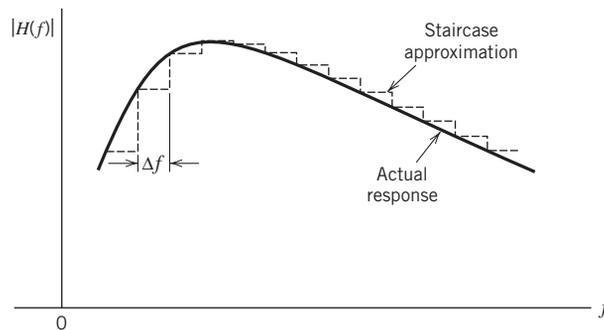


Figure 5.18 Staircase approximation of an arbitrary magnitude response $|H(f)|$; only the positive frequency portion of the response is shown.

The k th subchannel in the approximation to the model of Figure 5.17b is described by

$$y_k(t) = x_k(t) + n_k(t), \quad k = 1, 2, \dots, N \quad (5.107)$$

The average power of the signal component $x_k(t)$ is

$$P_k = S_{XX}(f_k)\Delta f, \quad k = 1, 2, \dots, N \quad (5.108)$$

where $S_X(f_k)$ is the power spectral density of the input signal evaluated at the frequency $f = f_k$. The variance of the noise component $n_k(t)$ is

$$\sigma_k^2 = \frac{S_{NN}(f_k)}{|H(f_k)|^2}\Delta f, \quad k = 1, 2, \dots, N \quad (5.109)$$

where $S_N(f_k)$ and $|H(f_k)|$ are the noise spectral density and the channel's magnitude response evaluated at the frequency f_k , respectively. The information capacity of the k th subchannel is

$$C_k = \frac{1}{2}\Delta f \log_2\left(1 + \frac{P_k}{\sigma_k^2}\right), \quad k = 1, 2, \dots, N \quad (5.110)$$

where the factor $1/2$ accounts for the fact that Δf applies to both positive and negative frequencies. All the N subchannels are independent of one another. Hence, the total capacity of the overall channel is approximately given by the summation

$$\begin{aligned} C &\approx \sum_{k=1}^N C_k \\ &= \frac{1}{2} \sum_{k=1}^N \Delta f \log_2\left(1 + \frac{P_k}{\sigma_k^2}\right) \end{aligned} \quad (5.111)$$

The problem we have to address is to maximize the overall information capacity C subject to the constraint

$$\sum_{k=1}^N P_k = P = \text{constant} \quad (5.112)$$

The usual procedure to solve a constrained optimization problem is to use the *method of Lagrange multipliers* (see Appendix D for a discussion of this method). To proceed with this optimization, we first define an objective function that incorporates both the information capacity C and the constraint (i.e., (5.111) and (5.112)), as shown by

$$J(P_k) = \frac{1}{2} \sum_{k=1}^N \Delta f \log_2\left(1 + \frac{P_k}{\sigma_k^2}\right) + \lambda \left(P - \sum_{k=1}^N P_k\right) \quad (5.113)$$

where λ is the Lagrange multiplier. Next, differentiating the objective function $J(P_k)$ with respect to P_k and setting the result equal to zero, we obtain

$$\frac{\Delta f \log_2 e}{P_k + \sigma_k^2} - \lambda = 0$$

To satisfy this optimizing solution, we impose the following requirement:

$$P_k + \sigma_k^2 = K\Delta f \quad \text{for } k = 1, 2, \dots, N \quad (5.114)$$

where K is a constant that is the same for all k . The constant K is chosen to satisfy the average power constraint.

Inserting the defining values of (5.108) and (5.109) in the optimizing condition of (5.114), simplifying, and rearranging terms we get

$$S_{XX}(f_k) = K - \frac{S_{NN}(f_k)}{|H(f_k)|^2}, \quad k = 1, 2, \dots, N \quad (5.115)$$

Let \mathcal{F}_A denote the frequency range for which the constant K satisfies the condition

$$K \geq \frac{S_{NN}(f_k)}{|H(f_k)|^2}$$

Then, as the incremental frequency interval Δf is allowed to approach zero and the number of subchannels N goes to infinity, we may use (5.115) to formally state that the power spectral density of the input ensemble that achieves the optimum information capacity is a nonnegative quantity defined by

$$S_{XX}(f) = \begin{cases} K - \frac{S_{NN}(f)}{|H(f)|^2} & f \in \mathcal{F}_A \\ 0, & \text{otherwise} \end{cases} \quad (5.116)$$

Because the average power of a random process is the total area under the curve of the power spectral density of the process, we may express the average power of the channel input $x(t)$ as

$$P = \int_{f \in \mathcal{F}_A} \left(K - \frac{S_{NN}(f)}{|H(f)|^2} \right) df \quad (5.117)$$

For a prescribed P and specified $S_N(f)$ and $H(f)$, the constant K is the solution to (5.117).

The only thing that remains for us to do is to find the optimum information capacity. Substituting the optimizing solution of (5.114) into (5.111) and then using the defining values of (5.108) and (5.109), we obtain

$$C \approx \frac{1}{2} \sum_{k=1}^N \Delta f \log_2 \left(K \frac{|H(f_k)|^2}{S_{NN}(f_k)} \right)$$

When the incremental frequency interval Δf is allowed to approach zero, this equation takes the limiting form

$$C = \frac{1}{2} \int_{-\infty}^{\infty} \log_2 \left(K \frac{|H(f)|^2}{S_{NN}(f)} \right) df \quad (5.118)$$

where the constant K is chosen as the solution to (5.117) for a prescribed input signal power P .

Water-filling Interpretation of the Information Capacity Law

Equations (5.116) and (5.117) suggest the picture portrayed in Figure 5.19. Specifically, we make the following observations:

- The appropriate input power spectral density $S_X(f)$ is described as the bottom regions of the function $S_N(f)/|H(f)|^2$ that lie below the constant level K , which are shown shaded.
- The input power P is defined by the total area of these shaded regions.

The spectral-domain picture portrayed here is called the *water-filling (pouring) interpretation*, in the sense that the process by which the input power is distributed across the function $S_N(f)/|H(f)|^2$ is identical to the way in which water distributes itself in a vessel.

Consider now the idealized case of a band-limited signal in AWGN channel of power spectral density $N(f) = N_0/2$. The transfer function $H(f)$ is that of an ideal band-pass filter defined by

$$H(f) = \begin{cases} 1, & 0 \leq f_c - \frac{B}{2} \leq |f| \leq f_c + \frac{B}{2} \\ 0, & \text{otherwise} \end{cases}$$

where f_c is the midband frequency and B is the channel bandwidth. For this special case, (5.117) and (5.118) reduce respectively to

$$P = 2B \left(K - \frac{N_0}{2} \right)$$

and

$$C = B \log_2 \left(\frac{2K}{N_0} \right)$$

Hence, eliminating K between these two equations, we get the standard form of Shannon's capacity theorem, defined by (5.95).

EXAMPLE 11 Capacity of NEXT-Dominated Channel

Digital subscriber lines (DSLs) refer to a family of different technologies that operate over a closed transmission loop; they will be discussed in Chapter 8, Section 8.11. For the

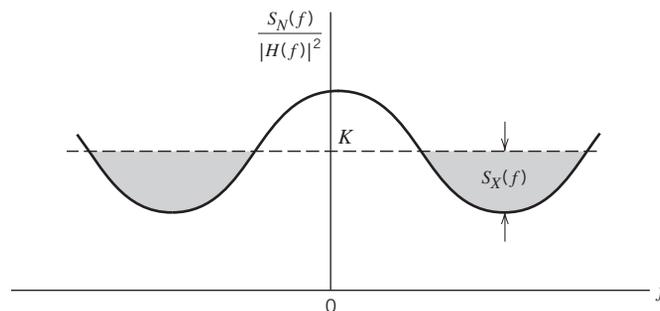


Figure 5.19 Water-filling interpretation of information-capacity theorem for a colored noisy channel.

present, it suffices to say that a DSL is designed to provide for data transmission between a user terminal (e.g., computer) and the central office of a telephone company. A major channel impairment that arises in the deployment of a DSL is the near-end cross-talk (NEXT). The power spectral density of this crosstalk may be taken as

$$S_N(f) = |H_{\text{NEXT}}(f)|^2 S_X(f) \quad (5.119)$$

where $S_X(f)$ is the power spectral density of the transmitted signal and $H_{\text{NEXT}}(f)$ is the transfer function that couples adjacent twisted pairs. The only constraint we have to satisfy in this example is that the power spectral density function $S_X(f)$ be *nonnegative for all f* . Substituting (5.119) into (5.116), we readily find that this condition is satisfied by solving for K as

$$K = \left(1 + \frac{|H_{\text{NEXT}}(f)|^2}{|H(f)|^2} \right) S_X(f)$$

Finally, using this result in (5.118), we find that the capacity of the NEXT-dominated digital subscriber channel is given by

$$C = \frac{1}{2} \int_{\mathcal{F}_A} \log_2 \left(1 + \frac{|H(f)|^2}{|H_{\text{NEXT}}(f)|^2} \right) df$$

where \mathcal{F}_A is the set of positive and negative frequencies for which $S_X(f) > 0$.

5.13 Rate Distortion Theory

In Section 5.3 we introduced the source-coding theorem for a discrete memoryless source, according to which the average codeword length must be at least as large as the source entropy for perfect coding (i.e., perfect representation of the source). However, in many practical situations there are constraints that force the coding to be imperfect, thereby resulting in unavoidable *distortion*. For example, constraints imposed by a communication channel may place an upper limit on the permissible code rate and, therefore, on average codeword length assigned to the information source. As another example, the information source may have a continuous amplitude as in the case of speech, and the requirement is to quantize the amplitude of each sample generated by the source to permit its representation by a codeword of finite length as in pulse-code modulation to be discussed in Chapter 6. In such cases, the problem is referred to as *source coding with a fidelity criterion*, and the branch of information theory that deals with it is called *rate distortion theory*.¹⁶ Rate distortion theory finds applications in two types of situations:

- Source coding where the permitted coding alphabet cannot exactly represent the information source, in which case we are forced to do lossy *data compression*.
- Information transmission at a rate greater than channel capacity.

Accordingly, rate distortion theory may be viewed as a natural extension of Shannon's coding theorem.

Rate Distortion Function

Consider a discrete memoryless source defined by an M -ary alphabet $\mathcal{X}: \{x_i | i = 1, 2, \dots, M\}$, which consists of a set of statistically independent symbols together with the associated symbol probabilities $\{p_i | i = 1, 2, \dots, M\}$. Let R be the average code rate in bits per codeword. The representation codewords are taken from another alphabet $\mathcal{Y}: \{y_j | j = 1, 2, \dots, N\}$. The source-coding theorem states that this second alphabet provides a perfect representation of the source provided that $R > H$, where H is the source entropy. But if we are forced to have $R < H$, then there is unavoidable distortion and, therefore, loss of information.

Let $p(x_i, y_j)$ denote the joint probability of occurrence of source symbol x_i and representation symbol y_j . From probability theory, we have

$$p(x_i, y_j) = p(y_j|x_i)p(x_i) \quad (5.120)$$

where $p(y_j|x_i)$ is a transition probability. Let $d(x_i, y_j)$ denote a measure of the cost incurred in representing the source symbol x_i by the symbol y_j ; the quantity $d(x_i, y_j)$ is referred to as a *single-letter distortion measure*. The statistical average of $d(x_i, y_j)$ over all possible source symbols and representation symbols is given by

$$\bar{d} = \sum_{i=1}^M \sum_{j=1}^N p(x_i)p(y_j|x_i)d(x_i|y_j) \quad (5.121)$$

Note that the average distortion \bar{d} is a nonnegative continuous function of the transition probabilities $p(y_j|x_i)$ that are determined by the source encoder–decoder pair.

A conditional probability assignment $p(y_j|x_i)$ is said to be *D-admissible* if, and only if, the average distortion \bar{d} is less than or equal to some acceptable value D . The set of all *D*-admissible conditional probability assignments is denoted by

$$\mathcal{P}_D = \{p(y_j|x_i): \bar{d} \leq D\} \quad (5.122)$$

For each set of transition probabilities, we have a mutual information

$$I(X;Y) = \sum_{i=1}^M \sum_{j=1}^N p(x_i)p(y_j|x_i) \log \left(\frac{p(y_j|x_i)}{p(y_j)} \right) \quad (5.123)$$

A *rate distortion function* $R(D)$ is defined as *the smallest coding rate possible for which the average distortion is guaranteed not to exceed D*. Let \mathcal{P}_D denote the set to which the conditional probability $p(y_j|x_i)$ belongs for a prescribed D . Then, for a fixed D we write¹⁷

$$R(D) = \min_{p(y_j|x_i) \in \mathcal{P}_D} I(X;Y) \quad (5.124)$$

subject to the constraint

$$\sum_{j=1}^N p(y_j|x_i) = 1 \quad \text{for } i = 1, 2, \dots, M \quad (5.125)$$

The rate distortion function $R(D)$ is measured in units of bits if the base-2 logarithm is used in (5.123). Intuitively, we expect the distortion D to decrease as the rate distortion function $R(D)$ is increased. We may say conversely that tolerating a large distortion D permits the use of a smaller rate for coding and/or transmission of information.

Figure 5.20
Summary of rate distortion theory.

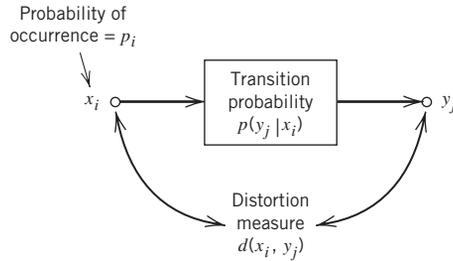


Figure 5.20 summarizes the main parameters of rate distortion theory. In particular, given the source symbols $\{x_i\}$ and their probabilities $\{p_i\}$, and given a definition of the single-letter distortion measure $d(x_i, y_j)$, the calculation of the rate distortion function $R(D)$ involves finding the conditional probability assignment $p(y_j | x_i)$ subject to certain constraints imposed on $p(y_j | x_i)$. This is a variational problem, the solution of which is unfortunately not straightforward in general.

EXAMPLE 12 Gaussian Source

Consider a discrete-time, memoryless Gaussian source with zero mean and variance σ^2 . Let x denote the value of a sample generated by such a source. Let y denote a quantized version of x that permits a finite representation of it. The *square-error distortion*

$$d(x, y) = (x - y)^2$$

provides a distortion measure that is widely used for continuous alphabets. The rate distortion function for the Gaussian source with square-error distortion, as described herein, is given by

$$R(D) = \begin{cases} \frac{1}{2} \log\left(\frac{\sigma^2}{D}\right), & 0 \leq D \leq \sigma^2 \\ 0, & D > \sigma^2 \end{cases} \quad (5.126)$$

In this case, we see that $R(D) \rightarrow \infty$ as $D \rightarrow 0$, and $R(D) = 0$ for $D = \sigma^2$.

EXAMPLE 13 Set of Parallel Gaussian Sources

Consider next a set of N independent Gaussian random variables $\{X_i\}_{i=1}^N$, where X_i has zero mean and variance σ_i^2 . Using the distortion measure

$$d = \sum_{i=1}^N (x_i - \hat{x}_i)^2, \quad \hat{x}_i = \text{estimate of } x_i$$

and building on the result of Example 12, we may express the rate distortion function for the set of parallel Gaussian sources described here as

$$R(D) = \sum_{i=1}^N \frac{1}{2} \log\left(\frac{\sigma_i^2}{D_i}\right) \quad (5.127)$$

where D_i is itself defined by

$$D_i = \begin{cases} \lambda, & \lambda < \sigma_i^2 \\ \sigma_i^2, & \lambda \geq \sigma_i^2 \end{cases} \quad (5.128)$$

and the constant λ is chosen so as to satisfy the condition

$$\sum_{i=1}^N D_i = D \quad (5.129)$$

Compared to Figure 5.19, (5.128) and (5.129) may be interpreted as a kind of “water-filling in reverse,” as illustrated in Figure 5.21. First, we choose a constant λ and only the subset of random variables whose variances exceed the constant λ . No bits are used to describe the remaining subset of random variables whose variances are less than the constant λ .

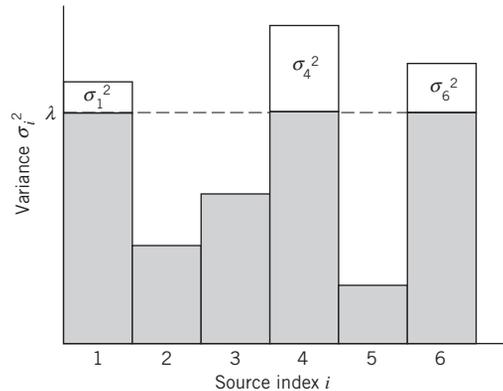


Figure 5.21 Reverse water-filling picture for a set of parallel Gaussian processes.

5.14 Summary and Discussion

In this chapter we established two fundamental limits on different aspects of a communication system, which are embodied in the source-coding theorem and the channel-coding theorem.

The *source-coding theorem*, Shannon’s first theorem, provides the mathematical tool for assessing *data compaction*; that is, *lossless compression* of data generated by a discrete memoryless source. The theorem teaches us that we can make the average number of binary code elements (bits) per source symbol as small as, but no smaller than, the entropy of the source measured in bits. The *entropy* of a source is a function of the probabilities of the source symbols that constitute the alphabet of the source. Since

entropy is a measure of uncertainty, the entropy is maximum when the associated probability distribution generates maximum uncertainty.

The *channel-coding theorem*, Shannon's second theorem, is both the most surprising and the single most important result of information theory. For a *binary symmetric channel*, the channel-coding theorem teaches us that, for any *code rate* r less than or equal to the *channel capacity* C , codes do exist such that the average probability of error is as small as we want it. A binary symmetric channel is the simplest form of a discrete memoryless channel. It is symmetric, because the probability of receiving symbol 1 if symbol 0 is sent is the same as the probability of receiving symbol 0 if symbol 1 is sent. This probability, the probability that an error will occur, is termed a *transition probability*. The transition probability p is determined not only by the additive noise at the channel output, but also by the kind of receiver used. The value of p uniquely defines the channel capacity C .

The *information capacity law*, an application of the channel-coding theorem, teaches us that there is an upper limit to the rate at which any communication system can operate reliably (i.e., free of errors) when the system is constrained in power. This maximum rate, called the *information capacity*, is measured in bits per second. When the system operates at a rate greater than the information capacity, it is condemned to a high probability of error, regardless of the choice of signal set used for transmission or the receiver used for processing the channel output.

When the output of a source of information is compressed in a lossless manner, the resulting data stream usually contains redundant bits. These redundant bits can be removed by using a lossless algorithm such as Huffman coding or the Lempel–Ziv algorithm for data compaction. We may thus speak of data compression followed by data compaction as two constituents of the *dissection of source coding*, which is so called because it refers exclusively to the sources of information.

We conclude this chapter on Shannon's information theory by pointing out that, in many practical situations, there are constraints that force source coding to be imperfect, thereby resulting in unavoidable *distortion*. For example, constraints imposed by a communication channel may place an upper limit on the permissible code rate and, therefore, average codeword length assigned to the information source. As another example, the information source may have a continuous amplitude, as in the case of speech, and the requirement is to *quantize* the amplitude of each sample generated by the source to permit its representation by a codeword of finite length, as in pulse-code modulation discussed in Chapter 6. In such cases, the information-theoretic problem is referred to as *source coding with a fidelity criterion*, and the branch of information theory that deals with it is called *rate distortion theory*, which may be viewed as a natural extension of Shannon's coding theorem.

Problems

Entropy

- 5.1 Let p denote the probability of some event. Plot the amount of information gained by the occurrence of this event for $0 \leq p \leq 1$.

- 5.2 A source emits one of four possible symbols during each signaling interval. The symbols occur with the probabilities

$$\begin{aligned} p_0 &= 0.4 \\ p_1 &= 0.3 \\ p_2 &= 0.2 \\ p_3 &= 0.1 \end{aligned}$$

which sum to unity as they should. Find the amount of information gained by observing the source emitting each of these symbols.

- 5.3 A source emits one of four symbols $s_0, s_1, s_2,$ and s_3 with probabilities $1/3, 1/6, 1/4$ and $1/4$, respectively. The successive symbols emitted by the source are statistically independent. Calculate the entropy of the source.
- 5.4 Let X represent the outcome of a single roll of a fair die. What is the entropy of X ?
- 5.5 The sample function of a Gaussian process of zero mean and unit variance is uniformly sampled and then applied to a uniform quantizer having the input–output amplitude characteristic shown in Figure P5.5. Calculate the entropy of the quantizer output.

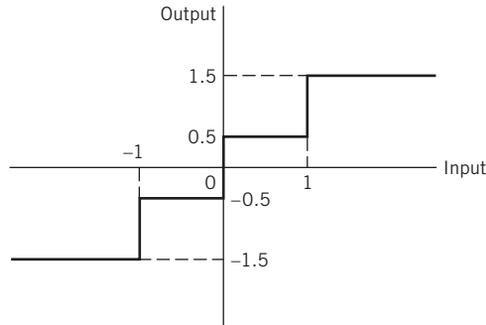


Figure P5.5

- 5.6 Consider a discrete memoryless source with source alphabet $S = \{s_0, s_1, \dots, s_{K-1}\}$ and source statistics $\{p_0, p_1, \dots, p_{K-1}\}$. The n th extension of this source is another discrete memoryless source with source alphabet $S^{(n)} = \{\sigma_0, \sigma_1, \dots, \sigma_{M-1}\}$, where $M = K^n$. Let $P(\sigma_i)$ denote the probability of σ_i .
- a. Show that, as expected,

$$\sum_{i=0}^{M-1} P(\sigma_i) = 1$$

- b. Show that

$$\sum_{i=0}^{M-1} P(\sigma_i) \log_2 \left(\frac{1}{P(\sigma_i)} \right) = H(S), \quad k = 1, 2, \dots, n$$

where p_{i_k} is the probability of symbol s_{i_k} and $H(S)$ is the entropy of the original source.

- c. Hence, show that

$$\begin{aligned} H(S^{(n)}) &= \sum_{i=0}^{M-1} P(\sigma_i) \log_2 \left(\frac{1}{P(\sigma_i)} \right) \\ &= nH(S) \end{aligned}$$

- 5.7 Consider a discrete memoryless source with source alphabet $S = \{s_0, s_1, s_2\}$ and source statistics $\{0.7, 0.15, 0.15\}$.
- Calculate the entropy of the source.
 - Calculate the entropy of the second-order extension of the source.
- 5.8 It may come as a surprise, but the number of bits needed to store text is much less than that required to store its spoken equivalent. Can you explain the reason for this statement?
- 5.9 Let a discrete random variable X assume values in the set $\{x_1, x_2, \dots, x_n\}$. Show that the entropy of X satisfies the inequality

$$H(X) \leq \log n$$

and with equality if, and only if, the probability $p_i = 1/n$ for all i .

Lossless Data Compression

- 5.10 Consider a discrete memoryless source whose alphabet consists of K equiprobable symbols.
- Explain why the use of a fixed-length code for the representation of such a source is about as efficient as any code can be.
 - What conditions have to be satisfied by K and the codeword length for the coding efficiency to be 100%?
- 5.11 Consider the four codes listed below:

Symbol	Code I	Code II	Code III	Code IV
s_0	0	0	0	00
s_1	10	01	01	01
s_2	110	001	011	10
s_3	1110	0010	110	110
s_4	1111	0011	111	111

- Two of these four codes are prefix codes. Identify them and construct their individual decision trees.
 - Apply the Kraft inequality to codes I, II, III, and IV. Discuss your results in light of those obtained in part a.
- 5.12 Consider a sequence of letters of the English alphabet with their probabilities of occurrence

Letter	a	i	l	m	n	o	p	y
Probability	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.1

Compute two different Huffman codes for this alphabet. In one case, move a combined symbol in the coding procedure as high as possible; in the second case, move it as low as possible. Hence, for each of the two codes, find the average codeword length and the variance of the average codeword length over the ensemble of letters. Comment on your results.

- 5.13 A discrete memoryless source has an alphabet of seven symbols whose probabilities of occurrence are as described here:

Symbol	s_0	s_1	s_2	s_3	s_4	s_5	s_6
Probability	0.25	0.25	0.125	0.125	0.125	0.0625	0.0625

Compute the Huffman code for this source, moving a “combined” symbol as high as possible. Explain why the computed source code has an efficiency of 100%.

- 5.14 Consider a discrete memoryless source with alphabet $\{s_0, s_1, s_2\}$ and statistics $\{0.7, 0.15, 0.15\}$ for its output.
- Apply the Huffman algorithm to this source. Hence, show that the average codeword length of the Huffman code equals 1.3 bits/symbol.
 - Let the source be extended to order two. Apply the Huffman algorithm to the resulting extended source and show that the average codeword length of the new code equals 1.1975 bits/symbol.
 - Extend the order of the extended source to three and reapply the Huffman algorithm; hence, calculate the average codeword length.
 - Compare the average codeword length calculated in parts b and c with the entropy of the original source.
- 5.15 Figure P5.15 shows a Huffman tree. What is the codeword for each of the symbols A, B, C, D, E, F, and G represented by this Huffman tree? What are their individual codeword lengths?

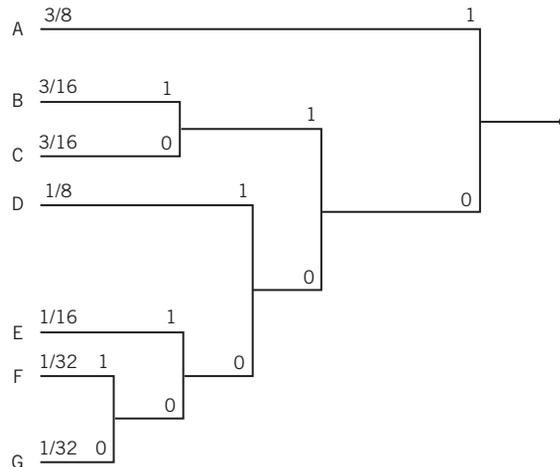


Figure P5.15

- 5.16 A computer executes four instructions that are designated by the codewords (00, 01, 10, 11). Assuming that the instructions are used independently with probabilities $(1/2, 1/8, 1/8, 1/4)$, calculate the percentage by which the number of bits used for the instructions may be reduced by the use of an optimum source code. Construct a Huffman code to realize the reduction.
- 5.17 Consider the following binary sequence

11101001100010110100 ...

Use the Lempel–Ziv algorithm to encode this sequence, assuming that the binary symbols 0 and 1 are already in the cookbook.

Binary Symmetric Channel

- 5.18 Consider the transition probability diagram of a binary symmetric channel shown in Figure 5.8. The input binary symbols 0 and 1 occur with equal probability. Find the probabilities of the binary symbols 0 and 1 appearing at the channel output.
- 5.19 Repeat the calculation in Problem 5.18, assuming that the input binary symbols 0 and 1 occur with probabilities $1/4$ and $3/4$, respectively.

Mutual Information and Channel Capacity

5.20 Consider a binary symmetric channel characterized by the transition probability p . Plot the mutual information of the channel as a function of p_1 , the a priori probability of symbol 1 at the channel input. Do your calculations for the transition probability $p = 0, 0.1, 0.2, 0.3, 0.5$.

5.21 Revisiting (5.12), express the mutual information $I(X;Y)$ in terms of the relative entropy

$$D(p(x,y)||p(x)p(y))$$

5.22 Figure 5.10 depicts the variation of the channel capacity of a binary symmetric channel with the transition probability p . Use the results of Problem 5.19 to explain this variation.

5.23 Consider the binary symmetric channel described in Figure 5.8. Let p_0 denote the probability of sending binary symbol $x_0 = 0$ and let $p_1 = 1 - p_0$ denote the probability of sending binary symbol $x_1 = 1$. Let p denote the transition probability of the channel.

a. Show that the mutual information between the channel input and channel output is given by

$$I(X;Y) = H(z) - H(p)$$

where the two entropy functions

$$H(z) = z \log_2\left(\frac{1}{z}\right) + (1 - z) \log_2\left(\frac{1}{1 - z}\right)$$

$$z = p_0p + (1 - p_0)(1 - p)$$

and

$$H(p) = p \log_2\left(\frac{1}{p}\right) + (1 - p) \log_2\left(\frac{1}{1 - p}\right)$$

b. Show that the value of p_0 that maximizes $I(X;Y)$ is equal to $1/2$.

c. Hence, show that the channel capacity equals

$$C = 1 - H(p)$$

5.24 Two binary symmetric channels are connected in cascade as shown in Figure P5.24. Find the overall channel capacity of the cascaded connection, assuming that both channels have the same transition probability diagram of Figure 5.8.

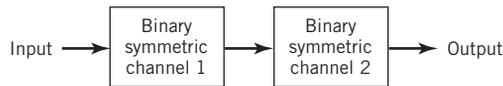


Figure P5.24

5.25 The *binary erasure channel* has two inputs and three outputs as described in Figure P5.25. The inputs are labeled 0 and 1 and the outputs are labeled 0, 1, and e . A fraction α of the incoming bits is erased by the channel. Find the capacity of the channel.

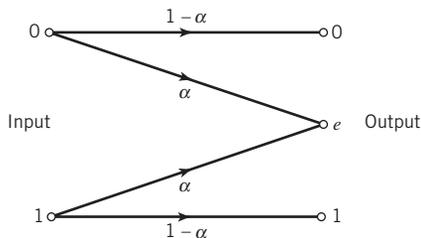


Figure P5.25

5.26 Consider a digital communication system that uses a *repetition code* for the channel encoding/decoding. In particular, each transmission is repeated n times, where $n = 2m + 1$ is an odd integer. The decoder operates as follows. If in a block of n received bits the number of 0s exceeds the number of 1s, then the decoder decides in favor of a 0; otherwise, it decides in favor of a 1. An error occurs when $m + 1$ or more transmissions out of $n = 2m + 1$ are incorrect. Assume a binary symmetric channel.

a. For $n = 3$, show that the average probability of error is given by

$$P_e = 3p^2(1-p) + p^3$$

where p is the transition probability of the channel.

b. For $n = 5$, show that the average probability of error is given by

$$P_e = 10p^3(1-p)^2 + 5p^4(1-p) + p^5$$

c. Hence, for the general case, deduce that the average probability of error is given by

$$P_e = \sum_{i=m+1}^n \binom{n}{i} p^i (1-p)^{n-i}$$

5.27 Let X , Y , and Z be three discrete random variables. For each value of the random variable Z , represented by sample z , define

$$A(z) = \sum_x \sum_y p(y) p(z|x, y)$$

Show that the conditional entropy $H(X|Y)$ satisfies the inequality

$$H(X|Y) \leq H(Z) + \mathbb{E}[\log A]$$

where \mathbb{E} is the expectation operator.

5.28 Consider two correlated discrete random variables X and Y , each of which takes a value in the set $\{x_i\}_{i=1}^n$. Suppose that the value taken by Y is known. The requirement is to guess the value of X . Let P_e denote the probability of error, defined by

$$P_e = \mathbb{P}[X \neq Y]$$

Show that P_e is related to the conditional entropy of X given Y by the inequality

$$H(X|Y) \leq H(P_e) + P_e \log(n-1)$$

This inequality is known as *Fano's inequality*. *Hint*: Use the result derived in Problem 5.27.

5.29 In this problem we explore the *convexity* of the mutual information $I(X;Y)$, involving the pair of discrete random variables X and Y .

Consider a discrete memoryless channel, for which the transition probability $p(y|x)$ is fixed for all x and y . Let X_1 and X_2 be two input random variables, whose input probability distributions are respectively denoted by $p(x_1)$ and $p(x_2)$. The corresponding probability distribution of X is defined by the convex combination

$$p(x) = a_1 p(x_1) + a_2 p(x_2)$$

where a_1 and a_2 are arbitrary constants. Prove the inequality

$$I(X;Y) \geq a_1 I(X_1;Y_1) + a_2 I(X_2;Y_2)$$

where X_1 , X_2 , and X are the channel inputs, and Y_1 , Y_2 , and Y are the corresponding channel outputs.

For the proof, you may use the following form of *Jensen's inequality*:

$$\sum_x \sum_y p_1(x, y) \log \left(\frac{p(y)}{p_1(y)} \right) \leq \log \left[\sum_x \sum_y p_1(x, y) \left(\frac{p(y)}{p_1(y)} \right) \right]$$

Differential Entropy

5.30 The differential entropy of a continuous random variable X is defined by the integral of (5.66). Similarly, the differential entropy of a continuous random vector \mathbf{X} is defined by the integral of (5.68). These two integrals may not exist. Justify this statement.

5.31 Show that the differential entropy of a continuous random variable X is invariant to translation; that is,

$$h(X + c) = h(X)$$

for some constant c .

5.32 Let X_1, X_2, \dots, X_n denote the elements of a Gaussian vector \mathbf{X} . The X_i are independent with mean m_i and variance σ_i^2 , $i = 1, 2, \dots, n$. Show that the differential entropy of the vector \mathbf{X} is given by

$$h(\mathbf{X}) = \frac{n}{2} \log_2 [2\pi e (\sigma_1^2 \sigma_2^2 \dots \sigma_n^2)^{1/n}]$$

where e is the base of the natural logarithm. What does $h(\mathbf{X})$ reduce to if the variances are all equal?

5.33 A continuous random variable X is constrained to a peak magnitude M ; that is,

$$-M < X < M$$

a. Show that the differential entropy of X is maximum when it is uniformly distributed, as shown by

$$f_X(x) = \begin{cases} 1/(2M), & -M < x \leq M \\ 0, & \text{otherwise} \end{cases}$$

b. Determine the maximum differential entropy of X .

5.34 Referring to (5.75), do the following:

a. Verify that the differential entropy of a Gaussian random variable of mean μ and variance σ^2 is given by $1/2 \log_2(2\pi e \sigma^2)$, where e is the base of the natural logarithm.

b. Hence, confirm the inequality of (5.75).

5.35 Demonstrate the properties of symmetry, nonnegativity, and expansion of the mutual information $I(X; Y)$ described in Section 5.6.

5.36 Consider the continuous random variable Y , defined by

$$Y = X + N$$

where the random variables X and N are statistically independent. Show that the conditional differential entropy of Y , given X , equals

$$h(Y | X) = h(N)$$

where $h(N)$ is the differential entropy of N .

Information Capacity Law

5.37 A voice-grade channel of the telephone network has a bandwidth of 3.4 kHz.

- Calculate the information capacity of the telephone channel for a signal-to-noise ratio of 30 dB.
- Calculate the minimum signal-to-noise ratio required to support information transmission through the telephone channel at the rate of 9600 bits/s.

5.38 Alphanumeric data are entered into a computer from a remote terminal through a voice-grade telephone channel. The channel has a bandwidth of 3.4 kHz and output signal-to-noise ratio of 20 dB. The terminal has a total of 128 symbols. Assume that the symbols are equiprobable and the successive transmissions are statistically independent.

- Calculate the information capacity of the channel.
- Calculate the maximum symbol rate for which error-free transmission over the channel is possible.

- 5.39 A black-and-white television picture may be viewed as consisting of approximately 3×10^5 elements, each of which may occupy one of 10 distinct brightness levels with equal probability. Assume that (1) the rate of transmission is 30 picture frames per second and (2) the signal-to-noise ratio is 30 dB.

Using the information capacity law, calculate the minimum bandwidth required to support the transmission of the resulting video signal.

- 5.40 In Section 5.10 we made the statement that it is easier to increase the information capacity of a communication channel by expanding its bandwidth B than increasing the transmitted power for a prescribed noise variance N_0B . This statement assumes that the noise spectral density N_0 varies inversely with B . Why is this inverse relationship the case?

- 5.41 In this problem, we revisit Example 5.10, which deals with coded binary antipodal signaling over an additive white Gaussian noise (AWGN) channel. Starting with (5.105) and the underlying theory, develop a software package for computing the minimum E_b/N_0 required for a given bit error rate, where E_b is the signal energy per bit, and $N_0/2$ is the noise spectral density. Hence, compute the results plotted in parts *a* and *b* of Figure 5.16.

As mentioned in Example 5.10, the computation of the mutual information between the channel input and channel output is well approximated using Monte Carlo integration. To explain how this method works, consider a function $g(y)$ that is difficult to sample randomly, which is indeed the case for the problem at hand. (For this problem, the function $g(y)$ represents the complicated integrand in the formula for the differential entropy of the channel output.) For the computation, proceed as follows:

- Find an area A that includes the region of interest and that is easily sampled.
- Choose N points, uniformly randomly inside the area A .

Then the *Monte Carlo integration theorem* states that the integral of the function $g(y)$ with respect to y is approximately equal to the area A multiplied by the fraction of points that reside below the curve of g , as illustrated in Figure P5.41. The accuracy of the approximation improves with increasing N .

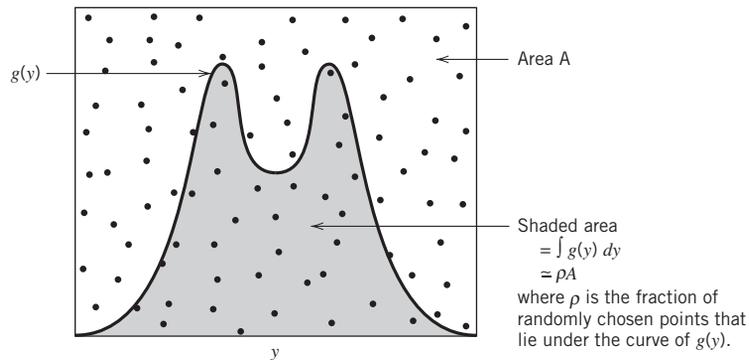


Figure P5.41

Notes

1. According to Lucky (1989), the first mention of the term *information theory* by Shannon occurred in a 1945 memorandum entitled “A mathematical theory of cryptography.” It is rather curious that the term was never used in Shannon’s (1948) classic paper, which laid down the foundations of information theory. For an introductory treatment of information theory, see Part I of the book by McEliece (2004), Chapters 1–6. For an advanced treatment of this subject, viewed in a rather broad context and treated with rigor, and clarity of presentation, see Cover and Thomas (2006).

For a collection of papers on the development of information theory (including the 1948 classic paper by Shannon), see Slepian (1974). For a collection of the original papers published by Shannon, see Sloane and Wyner (1993).

2. The use of a logarithmic measure of information was first suggested by Hartley (1928); however, Hartley used logarithms to base 10.

3. In statistical physics, the entropy of a physical system is defined by (Rief, 1965: 147)

$$L = k_B \ln \Omega$$

where k_B is *Boltzmann's constant*, Ω is the number of states accessible to the system, and \ln denotes the natural logarithm. This entropy has the dimensions of energy, because its definition involves the constant k_B . In particular, it provides a *quantitative measure of the degree of randomness of the system*. Comparing the entropy of statistical physics with that of information theory, we see that they have a similar form.

4. For the original proof of the source coding theorem, see Shannon (1948). A general proof of the source coding theorem is also given in Cover and Thomas (2006). The source coding theorem is also referred to in the literature as the *noiseless coding theorem*, noiseless in the sense that it establishes the condition for error-free encoding to be possible.

5. For proof of the Kraft inequality, see Cover and Thomas (2006). The Kraft inequality is also referred to as the Kraft–McMillan inequality in the literature.

6. The Huffman code is named after its inventor D.A. Huffman (1952). For a detailed account of Huffman coding and its use in data compaction, see Cover and Thomas (2006).

7. The original papers on the Lempel–Ziv algorithm are Ziv and Lempel (1977, 1978). For detailed treatment of the algorithm, see Cover and Thomas (2006).

8. It is also of interest to note that once a “parent” subsequence is joined by its two children, that parent subsequence can be replaced in constructing the Lempel–Ziv algorithm. To illustrate this nice feature of the algorithm, suppose we have the following example sequence:

01, 010, 011, ...

where 01 plays the role of a parent and 010 and 011 play the roles of the parent's children. In this example, the algorithm removes the 01, thereby reducing the length of the table through the use of a pointer.

9. In Cover and Thomas (2006), it is proved that the two-stage method, where the source coding and channel coding are considered separately as depicted in Figure 5.11, is as good as any other method of transmitting information across a noisy channel. This result has practical implications, in that the design of a communication system may be approached in two separate parts: source coding followed by channel coding. Specifically, we may proceed as follows:

- Design a source code for the most efficient representation of data generated by a discrete memoryless source of information.
- Separately and independently, design a channel code that is appropriate for a discrete memoryless channel.

The combination of source coding and channel coding designed in this manner will be as efficient as anything that could be designed by considering the two coding problems jointly.

10. To prove the channel-coding theorem, Shannon used several ideas that were new at the time; however, it was some time later when the proof was made rigorous (Cover and Thomas, 2006: 199).

Perhaps the most thoroughly rigorous proof of this basic theorem of information theory is presented in Chapter 7 of the book by Cover and Thomas (2006). Our statement of the theorem, though slightly different from that presented by Cover and Thomas, in essence is the same.

11. In the literature, the relative entropy is also referred to as the *Kullback–Leibler divergence* (KLD).

12. Equation (5.95) is also referred to in the literature as the *Shannon–Hartley law* in recognition of the early work by Hartley on information transmission (Hartley, 1928). In particular, Hartley showed that the amount of information that can be transmitted over a given channel is proportional to the product of the channel bandwidth and the time of operation.
13. A lucid exposition of sphere packing is presented in Cover and Thomas (2006); see also Wozencraft and Jacobs (1965).
14. Parts a and b of Figure 5.16 follow the corresponding parts of Figure 6.2 in the book by Frey (1998).
15. For a rigorous treatment of information capacity of a colored noisy channel, see Gallager (1968). The idea of replacing the channel model of Figure 5.17a with that of Figure 5.17b is discussed in Gitlin, Hayes, and Weinstein (1992).
16. For a complete treatment of rate distortion theory, see the classic book by Berger (1971); this subject is also treated in somewhat less detail in Cover and Thomas (1991), McEliece (1977), and Gallager (1968).
17. For the derivation of (5.124), see Cover and Thomas (2006). An algorithm for computation of the rate distortion function $R(D)$ defined in (5.124) is described in Blahut (1987) and Cover and Thomas (2006).

Conversion of Analog Waveforms into Coded Pulses

6.1 Introduction

In *continuous-wave (CW) modulation*, which was studied briefly in Chapter 2, some parameter of a sinusoidal carrier wave is varied continuously in accordance with the message signal. This is in direct contrast to *pulse modulation*, which we study in this chapter. In pulse modulation, some parameter of a pulse train is varied in accordance with the message signal. On this basis, we may distinguish two families of pulse modulation:

1. *Analog pulse modulation*, in which a periodic pulse train is used as the carrier wave and some characteristic feature of each pulse (e.g., amplitude, duration, or position) is varied in a continuous manner in accordance with the corresponding *sample* value of the message signal. Thus, in analog pulse modulation, information is transmitted basically in analog form but the transmission takes place at discrete times.
2. *Digital pulse modulation*, in which the message signal is represented in a form that is discrete in both time and amplitude, thereby permitting transmission of the message in digital form as a sequence of *coded pulses*; this form of signal transmission has *no* CW counterpart.

The use of coded pulses for the transmission of analog information-bearing signals represents a basic ingredient in digital communications. In this chapter, we focus attention on digital pulse modulation, which, in basic terms, is described as the *conversion of analog waveforms into coded pulses*. As such, the conversion may be viewed as the transition from analog to digital communications.

Three different kinds of digital pulse modulation are studied in the chapter:

1. *Pulse-code modulation (PCM)*, which has emerged as the most favored scheme for the digital transmission of analog information-bearing signals (e.g., voice and video signals). The important advantages of PCM are summarized thus:
 - *robustness* to channel noise and interference;
 - efficient *regeneration* of the coded signal along the transmission path;
 - efficient *exchange* of increased channel bandwidth for improved signal-to-quantization noise ratio, obeying an exponential law;
 - a *uniform format* for the transmission of different kinds of baseband signals, hence their integration with other forms of digital data in a common network;

- comparative *ease* with which message sources may be dropped or reinserted in a multiplex system;
- *secure* communication through the use of special modulation schemes or encryption.

These advantages, however, are attained at the cost of increased system complexity and increased transmission bandwidth. Simply stated:

There is no free lunch.

For every gain we make, there is a price to pay.

2. *Differential pulse-code modulation* (DPCM), which exploits the use of *lossy data compression* to remove the redundancy inherent in a message signal, such as voice or video, so as to reduce the bit rate of the transmitted data without serious degradation in overall system response. In effect, increased system complexity is traded off for reduced bit rate, therefore reducing the bandwidth requirement of PCM.
3. *Delta modulation* (DM), which addresses another practical limitation of PCM: the need for simplicity of implementation when it is a necessary requirement. DM satisfies this requirement by intentionally “oversampling” the message signal. In effect, increased transmission bandwidth is traded off for reduced system complexity. DM may therefore be viewed as the dual of DPCM.

Although, indeed, these three methods of analog-to-digital conversion are quite different, they do share two basic signal-processing operations, namely sampling and quantization:

- the process of sampling, followed by
- pulse-amplitude modulation (PAM) and finally
- amplitude quantization

are studied in what follows in this order.

6.2 Sampling Theory

The *sampling process* is usually described in the time domain. As such, it is an operation that is basic to digital signal processing and digital communications. Through use of the sampling process, an analog signal is converted into a corresponding sequence of samples that are usually spaced uniformly in time. Clearly, for such a procedure to have practical utility, it is necessary that we choose the sampling rate properly in relation to the bandwidth of the message signal, so that the sequence of samples uniquely defines the original analog signal. This is the essence of the sampling theorem, which is derived in what follows.

Frequency-Domain Description of Sampling

Consider an arbitrary signal $g(t)$ of finite energy, which is specified for all time t . A segment of the signal $g(t)$ is shown in Figure 6.1a. Suppose that we sample the signal $g(t)$ instantaneously and at a uniform rate, once every T_s seconds. Consequently, we obtain an infinite sequence of samples spaced T_s seconds apart and denoted by $\{g(nT_s)\}$, where n takes on all possible integer values, positive as well as negative. We refer to T_s as the *sampling period*, and to its reciprocal $f_s = 1/T_s$ as the *sampling rate*. For obvious reasons, this ideal form of sampling is called *instantaneous sampling*.

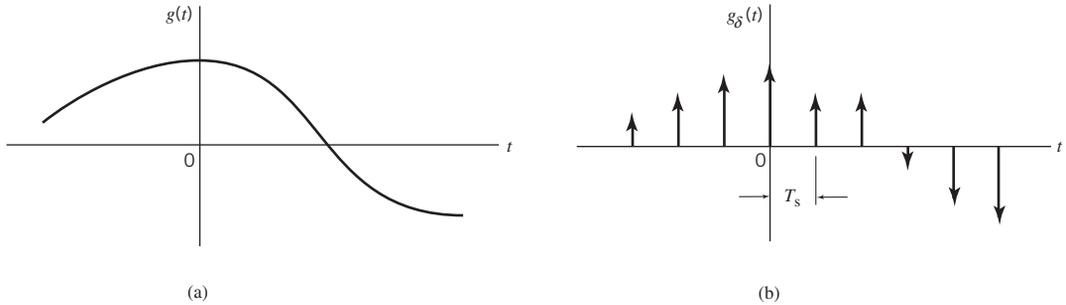


Figure 6.1 The sampling process. (a) Analog signal. (b) Instantaneously sampled version of the analog signal.

Let $g_\delta(t)$ denote the signal obtained by individually weighting the elements of a periodic sequence of delta functions spaced T_s seconds apart by the sequence of numbers $\{g(nT_s)\}$, as shown by (see Figure 6.1b):

$$g_\delta(t) = \sum_{n=-\infty}^{\infty} g(nT_s)\delta(t - nT_s) \quad (6.1)$$

We refer to $g_\delta(t)$ as the *ideal sampled signal*. The term $\delta(t - nT_s)$ represents a delta function positioned at time $t = nT_s$. From the definition of the delta function, we recall from Chapter 2 that such an idealized function has unit area. We may therefore view the multiplying factor $g(nT_s)$ in (6.1) as a “mass” assigned to the delta function $\delta(t - nT_s)$. A delta function weighted in this manner is closely approximated by a rectangular pulse of duration Δt and amplitude $g(nT_s)/\Delta t$; *the smaller we make Δt the better the approximation will be.*

Referring to the table of Fourier-transform pairs in Table 2.2, we have

$$g_\delta(t) \rightleftharpoons_{f_s} \sum_{m=-\infty}^{\infty} G(f - mf_s) \quad (6.2)$$

where $G(f)$ is the Fourier transform of the original signal $g(t)$ and f_s is the sampling rate. Equation (6.2) states:

The process of uniformly sampling a continuous-time signal of finite energy results in a periodic spectrum with a frequency equal to the sampling rate.

Another useful expression for the Fourier transform of the ideal sampled signal $g_\delta(t)$ may be obtained by taking the Fourier transform of both sides of (6.1) and noting that the Fourier transform of the delta function $\delta(t - nT_s)$ is equal to $\exp(-j2\pi n f T_s)$. Letting $G_\delta(f)$ denote the Fourier transform of $g_\delta(t)$, we may write

$$G_\delta(f) = \sum_{n=-\infty}^{\infty} g(nT_s)\exp(-j2\pi n f T_s) \quad (6.3)$$

Equation (6.3) describes the *discrete-time Fourier transform*. It may be viewed as a complex Fourier series representation of the periodic frequency function $G_\delta(f)$, with the sequence of samples $\{g(nT_s)\}$ defining the coefficients of the expansion.

The discussion presented thus far applies to any continuous-time signal $g(t)$ of finite energy and infinite duration. Suppose, however, that the signal $g(t)$ is *strictly band limited*, with no frequency components higher than W hertz. That is, the Fourier transform $G(f)$ of the signal $g(t)$ has the property that $G(f)$ is zero for $|f| \geq W$, as illustrated in Figure 6.2a; the shape of the spectrum shown in this figure is merely intended for the purpose of illustration. Suppose also that we choose the sampling period $T_s = 1/2W$. Then the corresponding spectrum $G_\delta(f)$ of the sampled signal $g_\delta(t)$ is as shown in Figure 6.2b. Putting $T_s = 1/2W$ in (6.3) yields

$$G_\delta(f) = \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \exp\left(-\frac{j\pi n f}{W}\right) \quad (6.4)$$

Isolating the term on the right-hand side of (6.2), corresponding to $m = 0$, we readily see that the Fourier transform of $g_\delta(t)$ may also be expressed as

$$G_\delta(f) = f_s G(f) + f_s \sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} G(f - mf_s) \quad (6.5)$$

Suppose, now, we impose the following two conditions:

1. $G(f) = 0$ for $|f| \geq W$.
2. $f_s = 2W$.

We may then reduce (6.5) to

$$G(f) = \frac{1}{2W} G_\delta(f), \quad -W < f < W \quad (6.6)$$

Substituting (6.4) into (6.6), we may also write

$$G(f) = \frac{1}{2W} \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \exp\left(-\frac{j\pi n f}{W}\right), \quad -W < f < W \quad (6.7)$$

Equation (6.7) is the desired formula for the frequency-domain description of sampling. This formula reveals that if the sample values $g(n/2W)$ of the signal $g(t)$ are specified for all n , then the Fourier transform $G(f)$ of that signal is uniquely determined. Because $g(t)$ is related to $G(f)$ by the inverse Fourier transform, it follows, therefore, that $g(t)$ is itself uniquely determined by the sample values $g(n/2W)$ for $-\infty < n < \infty$. In other words, the sequence $\{g(n/2W)\}$ has all the information contained in the original signal $g(t)$.

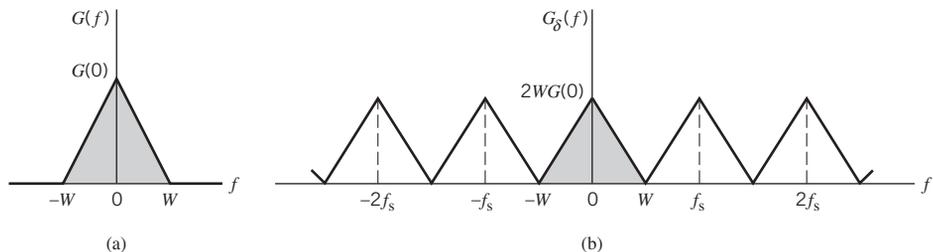


Figure 6.2 (a) Spectrum of a strictly band-limited signal $g(t)$. (b) Spectrum of the sampled version of $g(t)$ for a sampling period $T_s = 1/2W$.

Consider next the problem of reconstructing the signal $g(t)$ from the sequence of sample values $\{g(n/2W)\}$. Substituting (6.7) in the formula for the inverse Fourier transform

$$g(t) = \int_{-\infty}^{\infty} G(f) \exp(j2\pi ft) df$$

and interchanging the order of summation and integration, which is permissible because both operations are linear, we may go on to write

$$g(t) = \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \frac{1}{2W} \int_{-W}^W \exp\left[j2\pi f\left(t - \frac{n}{2W}\right)\right] df \quad (6.8)$$

The definite integral in (6.8), including the multiplying factor $1/2W$, is readily evaluated in terms of the sinc function, as shown by

$$\begin{aligned} \frac{1}{2W} \int_{-W}^W \exp\left[j2\pi f\left(t - \frac{n}{2W}\right)\right] df &= \frac{\sin(2\pi Wt - n\pi)}{2\pi Wt - n\pi} \\ &= \text{sinc}(2Wt - n) \end{aligned}$$

Accordingly, (6.8) reduces to the infinite-series expansion

$$g(t) = \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \text{sinc}(2Wt - n), \quad -\infty < t < \infty \quad (6.9)$$

Equation (6.9) is the desired *reconstruction formula*. This formula provides the basis for reconstructing the original signal $g(t)$ from the sequence of sample values $\{g(n/2W)\}$, with the sinc function $\text{sinc}(2Wt)$ playing the role of a *basis function* of the expansion. Each sample, $g(n/2W)$, is multiplied by a delayed version of the *basis function*, $\text{sinc}(2Wt - n)$, and all the resulting individual waveforms in the expansion are added to reconstruct the original signal $g(t)$.

The Sampling Theorem

Equipped with the frequency-domain description of sampling given in (6.7) and the reconstruction formula of (6.9), we may now state the *sampling theorem* for strictly band-limited signals of finite energy in two equivalent parts:

1. A band-limited signal of finite energy that has no frequency components higher than W hertz is completely described by specifying the values of the signal instants of time separated by $1/2W$ seconds.
2. A band-limited signal of finite energy that has no frequency components higher than W hertz is completely recovered from a knowledge of its samples taken at the rate of $2W$ samples per second.

Part 1 of the theorem, following from (6.7), is performed in the transmitter. Part 2 of the theorem, following from (6.9), is performed in the receiver. For a signal bandwidth of W hertz, the sampling rate of $2W$ samples per second, for a signal bandwidth of W hertz, is called the *Nyquist rate*; its reciprocal $1/2W$ (measured in seconds) is called the *Nyquist interval*; see the classic paper (Nyquist, 1928b).

Aliasing Phenomenon

Derivation of the sampling theorem just described is based on the assumption that the signal $g(t)$ is strictly band limited. In practice, however, a message signal is *not* strictly band limited, with the result that some degree of undersampling is encountered, as a consequence of which *aliasing* is produced by the sampling process. Aliasing refers to the phenomenon of a high-frequency component in the spectrum of the signal seemingly taking on the identity of a lower frequency in the spectrum of its sampled version, as illustrated in Figure 6.3. The aliased spectrum, shown by the solid curve in Figure 6.3b, pertains to the undersampled version of the message signal represented by the spectrum of Figure 6.3a.

To combat the effects of aliasing in practice, we may use two corrective measures:

1. Prior to sampling, a low-pass *anti-aliasing filter* is used to attenuate those high-frequency components of the signal that are not essential to the information being conveyed by the message signal $g(t)$.
2. The filtered signal is sampled at a rate slightly higher than the Nyquist rate.

The use of a sampling rate higher than the Nyquist rate also has the beneficial effect of easing the design of the *reconstruction filter* used to recover the original signal from its sampled version. Consider the example of a message signal that has been anti-alias (low-pass) filtered, resulting in the spectrum shown in Figure 6.4a. The corresponding spectrum of the instantaneously sampled version of the signal is shown in Figure 6.4b, assuming a sampling rate higher than the Nyquist rate. According to Figure 6.4b, we readily see that design of the reconstruction filter may be specified as follows:

- The reconstruction filter is low-pass with a passband extending from $-W$ to W , which is itself determined by the anti-aliasing filter.
- The reconstruction filter has a transition band extending (for positive frequencies) from W to $(f_s - W)$, where f_s is the sampling rate.

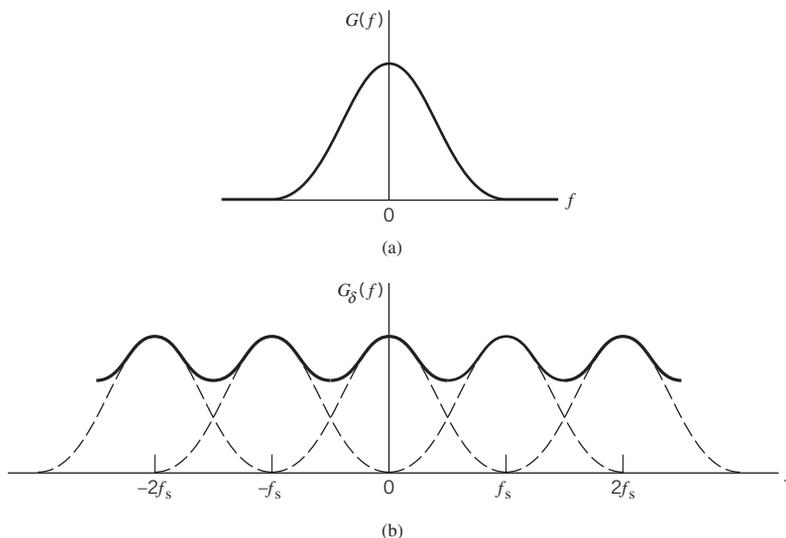


Figure 6.3 (a) Spectrum of a signal. (b) Spectrum of an under-sampled version of the signal exhibiting the aliasing phenomenon.

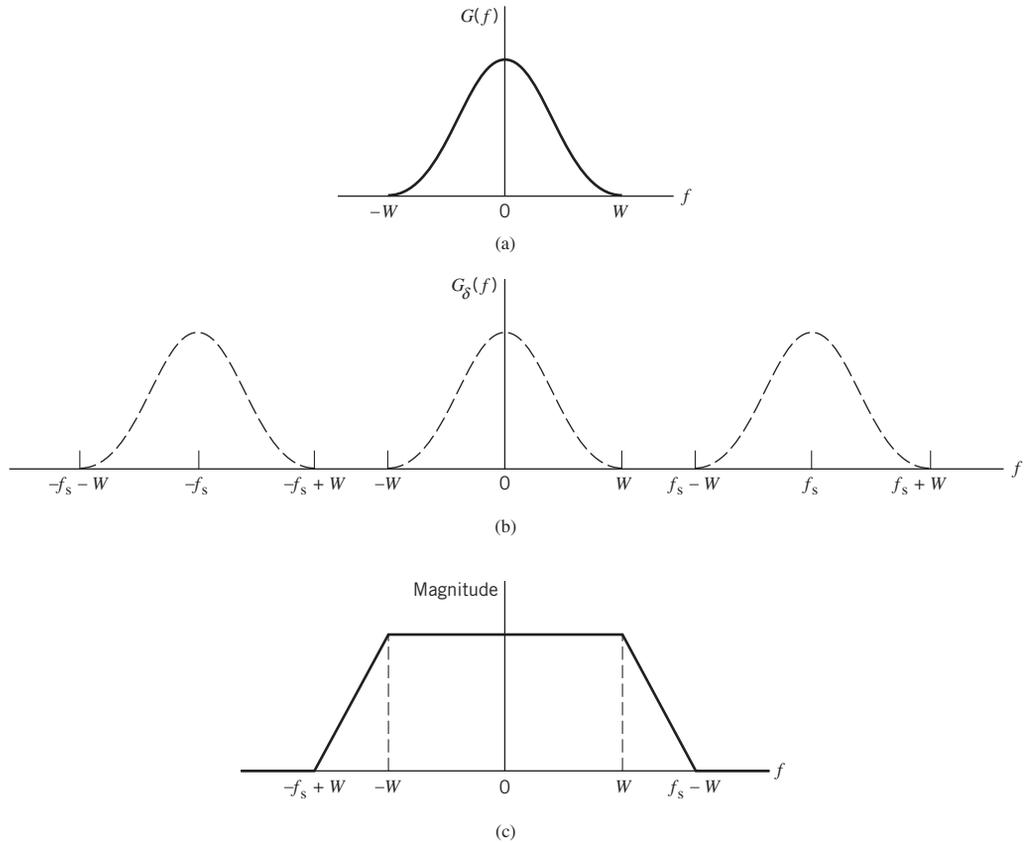


Figure 6.4 (a) Anti-alias filtered spectrum of an information-bearing signal. (b) Spectrum of instantaneously sampled version of the signal, assuming the use of a sampling rate greater than the Nyquist rate. (c) Magnitude response of reconstruction filter.

EXAMPLE 1 Sampling of Voice Signals

As an illustrative example, consider the sampling of voice signals for waveform coding. Typically, the frequency band, extending from 100 Hz to 3.1 kHz, is considered to be adequate for telephonic communication. This limited frequency band is accomplished by passing the voice signal through a low-pass filter with its cutoff frequency set at 3.1 kHz; such a filter may be viewed as an anti-aliasing filter. With such a cutoff frequency, the Nyquist rate is $f_s = 2 \times 3.1 = 6.2$ kHz. The standard sampling rate for the waveform coding of voice signals is 8 kHz. Putting these numbers together, design specifications for the reconstruction (low-pass) filter in the receiver are as follows:

Cutoff frequency	3.1 kHz
Transition band	6.2 to 8 kHz
Transition-band width	1.8 kHz.

6.3 Pulse-Amplitude Modulation

Now that we understand the essence of the sampling process, we are ready to formally define PAM, which is the simplest and most basic form of analog pulse modulation. It is formally defined as follows:

PAM is a linear modulation process where the amplitudes of regularly spaced pulses are varied in proportion to the corresponding sample values of a continuous message signal.

The pulses themselves can be of a rectangular form or some other appropriate shape.

The waveform of a PAM signal is illustrated in Figure 6.5. The dashed curve in this figure depicts the waveform of a message signal $m(t)$, and the sequence of amplitude-modulated rectangular pulses shown as solid lines represents the corresponding PAM signal $s(t)$. There are two operations involved in the generation of the PAM signal:

1. *Instantaneous sampling* of the message signal $m(t)$ every T_s seconds, where the sampling rate $f_s = 1/T_s$ is chosen in accordance with the sampling theorem.
2. *Lengthening* the duration of each sample so obtained to some constant value T .

In digital circuit technology, these two operations are jointly referred to as “sample and hold.” One important reason for intentionally lengthening the duration of each sample is to avoid the use of an excessive channel bandwidth, because bandwidth is inversely proportional to pulse duration. However, care has to be exercised in how long we make the sample duration T , as the following analysis reveals.

Let $s(t)$ denote the sequence of flat-top pulses generated in the manner described in Figure 6.5. We may express the PAM signal as a *discrete convolution sum*:

$$s(t) = \sum_{n=-\infty}^{\infty} m(nT_s)h(t-nT_s) \quad (6.10)$$

where T_s is the *sampling period* and $m(nT_s)$ is the sample value of $m(t)$ obtained at time $t = nT_s$. The $h(t)$ is a Fourier-transformal pulse. With spectral analysis of $s(t)$ in mind, we would like to recast (6.10) in the form of a convolution integral. To this end, we begin by invoking the sifting property of a delta function (discussed in Chapter 2) to express the delayed version of the pulse shape $h(t)$ in (6.10) as

$$h(t-nT_s) = \int_{-\infty}^{\infty} h(t-\tau)\delta(t-nT_s) d\tau \quad (6.11)$$

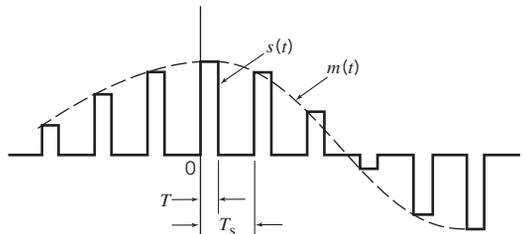


Figure 6.5 Flat-top samples, representing an analog signal.

Hence, substituting (6.11) into (6.10), and interchanging the order of summation and integration, we get

$$s(t) = \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} m(nT_s) \delta(t - nT_s) \right] h(t - \tau) d\tau \quad (6.12)$$

Referring to (6.1), we recognize that the expression inside the brackets in (6.12) is simply the instantaneously sampled version of the message signal $m(t)$, as shown by

$$m_{\delta}(t) = \sum_{n=-\infty}^{\infty} m(nT_s) \delta(t - nT_s) \quad (6.13)$$

Accordingly, substituting (6.13) into (6.12), we may reformulate the PAM signal $s(t)$ in the desired form

$$\begin{aligned} s(t) &= \int_{-\infty}^{\infty} m_{\delta}(t) h(t - \tau) d\tau \\ &= m_{\delta}(t) \star h(t) \end{aligned} \quad (6.14)$$

which is the convolution of the two time functions; $m_{\delta}(t)$ and $h(t)$.

The stage is now set for taking the Fourier transform of both sides of (6.14) and recognizing that the convolution of two time functions is transformed into the multiplication of their respective Fourier transforms; we get the simple result

$$S(f) = M_{\delta}(f)H(f) \quad (6.15)$$

where $S(f) = \mathbf{F}[s(t)]$, $M_{\delta}(f) = \mathbf{F}[m_{\delta}(t)]$, and $H(f) = \mathbf{F}[h(t)]$. Adapting (6.2) to the problem at hand, we note that the Fourier transform $M_{\delta}(f)$ is related to the Fourier transform $M(f)$ of the original message signal $m(t)$ as follows:

$$M_{\delta}(f) = f_s \sum_{k=-\infty}^{\infty} M(f - kf_s) \quad (6.16)$$

where f_s is the sampling rate. Therefore, the substitution of (6.16) into (6.15) yields the desired formula for the Fourier transform of the PAM signal $s(t)$, as shown by

$$S(f) = f_s \sum_{k=-\infty}^{\infty} M(f - kf_s)H(f) \quad (6.17)$$

Given this formula, how do we recover the original message signal $m(t)$? As a first step in this reconstruction, we may pass $s(t)$ through a low-pass filter whose frequency response is defined in Figure 6.4c; here, it is assumed that the message signal is limited to bandwidth W and the sampling rate f_s is larger than the Nyquist rate $2W$. Then, from (6.17) we find that the spectrum of the resulting filter output is equal to $M(f)H(f)$. This output is equivalent to passing the original message signal $m(t)$ through another low-pass filter of frequency response $H(f)$.

Equation (6.17) applies to any Fourier-transformable pulse shape $h(t)$.

Consider now the special case of a rectangular pulse of unit amplitude and duration T , as shown in Figure 6.6a; specifically:

$$h(t) = \begin{cases} 1, & 0 < t < T \\ \frac{1}{2}, & t = 0, t = T \\ 0, & \text{otherwise} \end{cases} \quad (6.18)$$

Correspondingly, the Fourier transform of $h(t)$ is given by

$$H(f) = T \operatorname{sinc}(fT) \exp(-j\pi fT) \quad (6.19)$$

which is plotted in Figure 6.6b. We therefore find from (6.17) that by using flat-top samples to generate a PAM signal we have introduced *amplitude distortion* as well as a *delay* of $T/2$. This effect is rather similar to the variation in transmission with frequency that is caused by the finite size of the scanning aperture in television. Accordingly, the distortion caused by the use of PAM to transmit an analog information-bearing signal is referred to as the *aperture effect*.

To correct for this distortion, we connect an *equalizer* in cascade with the low-pass reconstruction filter, as shown in Figure 6.7. The equalizer has the effect of decreasing the in-band loss of the reconstruction filter as the frequency increases in such a manner as to

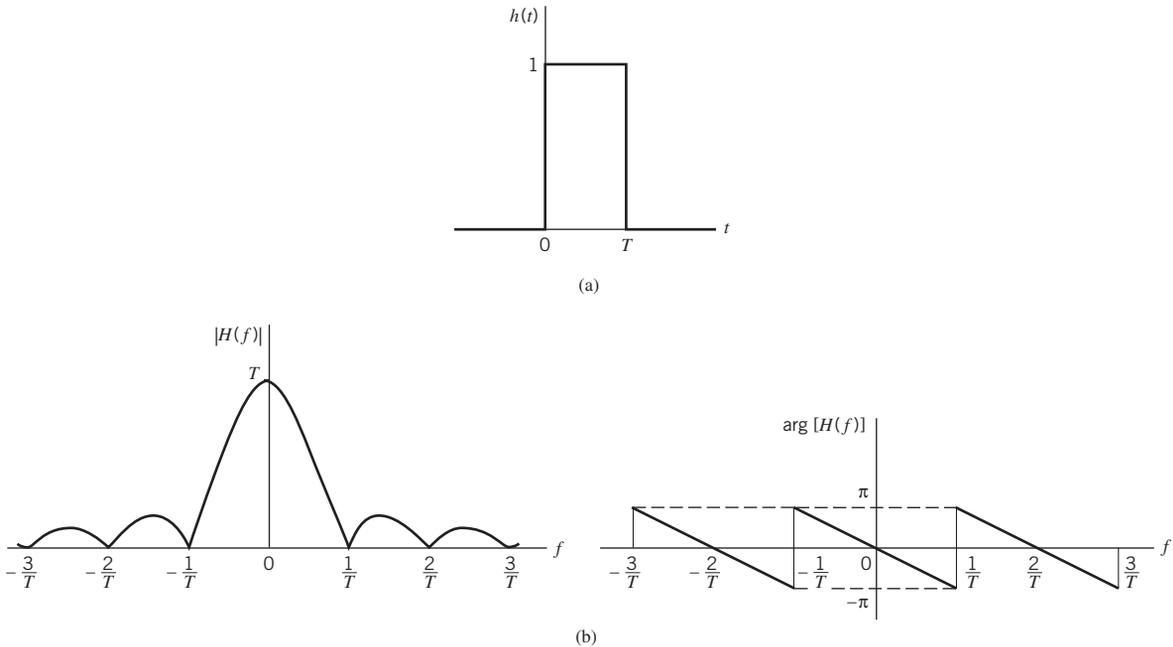


Figure 6.6 (a) Rectangular pulse $h(t)$. (b) Transfer function $H(f)$, made up of the magnitude $|H(f)|$ and phase $\arg[H(f)]$.

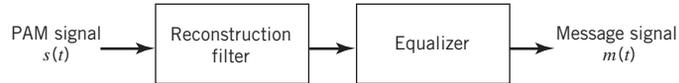


Figure 6.7 System for recovering message signal $m(t)$ from PAM signal $s(t)$.

compensate for the aperture effect. In light of (6.19), the magnitude response of the equalizer should ideally be

$$\frac{1}{|H(f)|} = \frac{1}{T \operatorname{sinc}(fT)} = \frac{\pi f}{\sin(\pi f T)}$$

The amount of equalization needed in practice is usually small. Indeed, for a duty cycle defined by the ratio $T/T_s \leq 0.1$, the amplitude distortion is less than 0.5%. In such a situation, the need for equalization may be omitted altogether.

Practical Considerations

The transmission of a PAM signal imposes rather stringent requirements on the frequency response of the channel, because of the relatively short duration of the transmitted pulses. One other point that should be noted: relying on amplitude as the parameter subject to modulation, the noise performance of a PAM system can never be better than baseband-signal transmission. Accordingly, in practice, we find that for transmission over a communication channel PAM is used only as the preliminary means of message processing, whereafter the PAM signal is changed to some other more appropriate form of pulse modulation.

With analog-to-digital conversion as the aim, what would be the appropriate form of modulation to build on PAM? Basically, there are three potential candidates, each with its own advantages and disadvantages, as summarized here:

1. *PCM*, which, as remarked previously in Section 6.1, is robust but demanding in both transmission bandwidth and computational requirements. Indeed, PCM has established itself as the standard method for the conversion of speech and video signals into digital form.
2. *DPCM*, which provides a method for the reduction in transmission bandwidth but at the expense of increased computational complexity.
3. *DM*, which is relatively simple to implement but requires a significant increase in transmission bandwidth.

Before we go on, a comment on terminology is in order. The term “modulation” used herein is a *misnomer*. In reality, PCM, DM, and DPCM are different forms of source coding, with source coding being understood in the sense described in Chapter 5 on information theory. Nevertheless, the terminologies used to describe them have become embedded in the digital communications literature, so much so that we just have to live with them.

Despite their basic differences, PCM, DPCM and DM do share an important feature: the message signal is represented in discrete form in both time and amplitude. PAM takes care of the discrete-time representation. As for the discrete-amplitude representation, we resort to a process known as quantization, which is discussed next.

6.4 Quantization and its Statistical Characterization

Typically, an analog message signal (e.g., voice) has a continuous range of amplitudes and, therefore, its samples have a continuous amplitude range. In other words, within the finite amplitude range of the signal, we find an infinite number of amplitude levels. In actual fact, however, it is not necessary to transmit the exact amplitudes of the samples for the following reason: any human sense (the ear or the eye) as ultimate receiver can detect only finite intensity differences. This means that the message signal may be *approximated* by a signal constructed of discrete amplitudes selected on a minimum error basis from an available set. The existence of a finite number of discrete amplitude levels is a basic condition of waveform coding exemplified by PCM. Clearly, if we assign the discrete amplitude levels with sufficiently close spacing, then we may make the approximated signal practically indistinguishable from the original message signal. For a formal definition of *amplitude quantization*, or just *quantization* for short, we say:

Quantization is the process of transforming the sample amplitude $m(nT_s)$ of a message signal $m(t)$ at time $t = nT_s$ into a discrete amplitude $v(nT_s)$ taken from a finite set of possible amplitudes.

This definition assumes that the *quantizer* (i.e., the device performing the quantization process) is *memoryless and instantaneous*, which means that the transformation at time $t = nT_s$ is not affected by earlier or later samples of the message signal $m(t)$. This simple form of scalar quantization, though not optimum, is commonly used in practice.

When dealing with a memoryless quantizer, we may simplify the notation by dropping the time index. Henceforth, the symbol m_k is used in place of $m(kT_s)$, as indicated in the block diagram of a quantizer shown in Figure 6.8a. Then, as shown in Figure 6.8b, the signal amplitude m is specified by the index k if it lies inside the *partition cell*

$$J_k: \{m_k < m \leq m_{k+1}\}, \quad k = 1, 2, \dots, L \quad (6.20)$$

where

$$m_k = m(kT_s) \quad (6.21)$$

and L is the total number of amplitude levels used in the quantizer. The discrete amplitudes m_k , $k = 1, 2, \dots, L$, at the quantizer input are called *decision levels* or *decision thresholds*. At the quantizer output, the index k is transformed into an amplitude v_k that represents all amplitudes of the cell J_k ; the discrete amplitudes v_k , $k = 1, 2, \dots, L$, are called *representation levels* or *reconstruction levels*. The spacing between two adjacent representation levels is called a *quantum* or *step-size*. Thus, given a quantizer denoted by $g(\cdot)$, the quantized output v equals v_k if the input sample m belongs to the interval J_k . In effect, the mapping (see Figure 6.8a)

$$v = g(m) \quad (6.22)$$

defines the *quantizer characteristic*, described by a staircase function.



Figure 6.8
Description of a
memoryless quantizer.

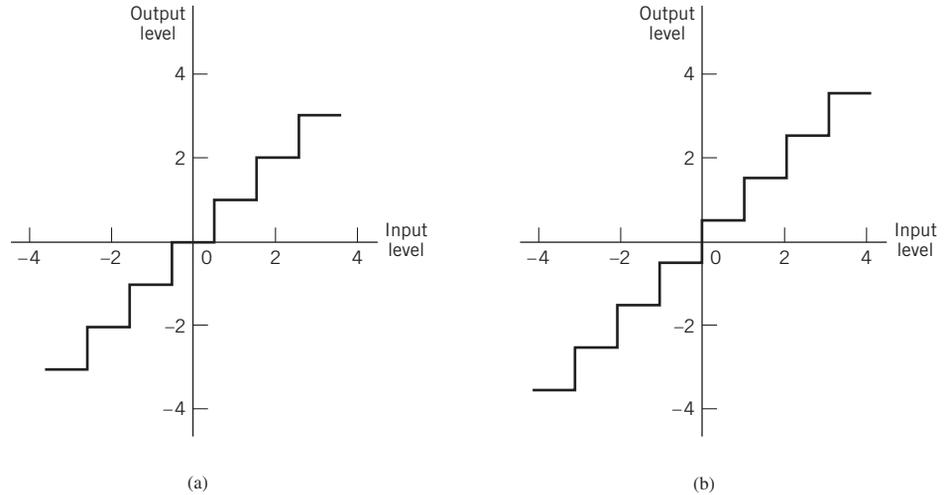


Figure 6.9 Two types of quantization: (a) midtread and (b) midrise.

Quantizers can be of a *uniform* or *nonuniform* type. In a uniform quantizer, the representation levels are uniformly spaced; otherwise, the quantizer is nonuniform. In this section, we consider only uniform quantizers; nonuniform quantizers are considered in Section 6.5. The quantizer characteristic can also be of *midtread* or *midrise* type. Figure 6.9a shows the input–output characteristic of a uniform quantizer of the midtread type, which is so called because the origin lies in the middle of a tread of the staircaselike graph. Figure 6.9b shows the corresponding input–output characteristic of a uniform quantizer of the midrise type, in which the origin lies in the middle of a rising part of the staircaselike graph. Despite their different appearances, both the midtread and midrise types of uniform quantizers illustrated in Figure 6.9 are *symmetric* about the origin.

Quantization Noise

Inevitably, the use of quantization introduces an error defined as the difference between the continuous input sample m and the quantized output sample v . The error is called *quantization noise*.¹ Figure 6.10 illustrates a typical variation of quantization noise as a function of time, assuming the use of a uniform quantizer of the midtread type.

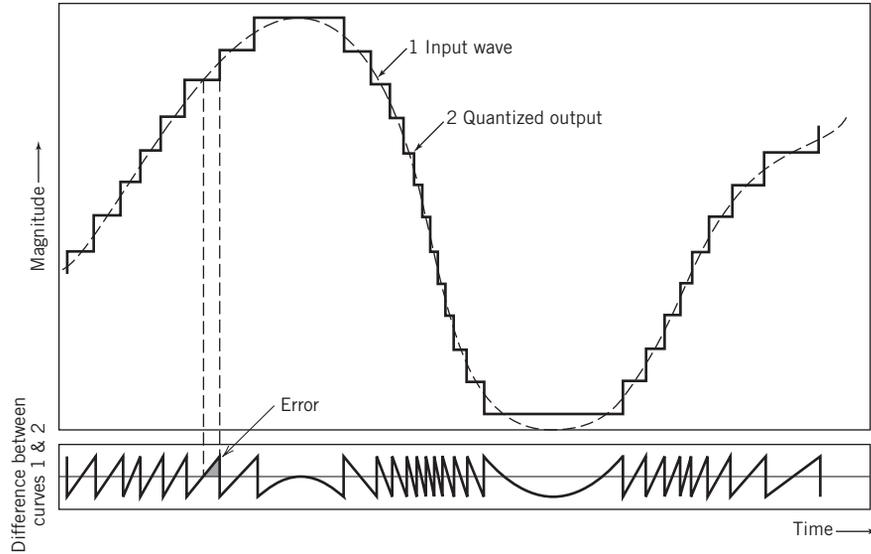
Let the quantizer input m be the sample value of a zero-mean random variable M . (If the input has a nonzero mean, we can always remove it by subtracting the mean from the input and then adding it back after quantization.) A quantizer, denoted by $g(\cdot)$, maps the input random variable M of continuous amplitude into a discrete random variable V ; their respective sample values m and v are related by the nonlinear function $g(\cdot)$ in (6.22). Let the quantization error be denoted by the random variable Q of sample value q . We may thus write

$$q = m - v \quad (6.23)$$

or, correspondingly,

$$Q = M - V \quad (6.24)$$

Figure 6.10
Illustration of the
quantization process.



With the input M having zero mean and the quantizer assumed to be symmetric as in Figure 6.9, it follows that the quantizer output V and, therefore, the quantization error Q will also have zero mean. Thus, for a partial statistical characterization of the quantizer in terms of output signal-to-(quantization) noise ratio, we need only find the mean-square value of the quantization error Q .

Consider, then, an input m of continuous amplitude, which, symmetrically, occupies the range $[-m_{\max}, m_{\max}]$. Assuming a uniform quantizer of the midrise type illustrated in Figure 6.9b, we find that the step size of the quantizer is given by

$$\Delta = \frac{2m_{\max}}{L} \quad (6.25)$$

where L is the total number of representation levels. For a uniform quantizer, the quantization error Q will have its sample values bounded by $-\Delta/2 \leq q \leq \Delta/2$. If the step size Δ is sufficiently small (i.e., the number of representation levels L is sufficiently large), it is reasonable to assume that the quantization error Q is a *uniformly distributed* random variable and the interfering effect of the quantization error on the quantizer input is similar to that of thermal noise, hence the reference to quantization error as *quantization noise*. We may thus express the probability density function of the quantization noise as

$$f_Q(q) = \begin{cases} \frac{1}{\Delta}, & -\frac{\Delta}{2} < q \leq \frac{\Delta}{2} \\ 0, & \text{otherwise} \end{cases} \quad (6.26)$$

For this to be true, however, we must ensure that the incoming continuous sample does *not* overload the quantizer. Then, with the mean of the quantization noise being zero, its variance σ_Q^2 is the same as the mean-square value; that is,

$$\begin{aligned}\sigma_Q^2 &= \mathbb{E}[Q^2] \\ &= \int_{-\Delta/2}^{\Delta/2} q^2 f_Q(q) \, dq\end{aligned}\tag{6.27}$$

Substituting (6.26) into (6.27), we get

$$\begin{aligned}\sigma_Q^2 &= \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} q^2 \, dq \\ &= \frac{\Delta^2}{12}\end{aligned}\tag{6.28}$$

Typically, the L -ary number k , denoting the k th representation level of the quantizer, is transmitted to the receiver in binary form. Let R denote the *number of bits per sample* used in the construction of the binary code. We may then write

$$L = 2^R\tag{6.29}$$

or, equivalently,

$$R = \log_2 L\tag{6.30}$$

Hence, substituting (6.29) into (6.25), we get the step size

$$\Delta = \frac{2m_{\max}}{2^R}\tag{6.31}$$

Thus, the use of (6.31) in (6.28) yields

$$\sigma_Q^2 = \frac{1}{3} m_{\max}^2 2^{-2R}\tag{6.32}$$

Let P denote the average power of the original message signal $m(t)$. We may then express the *output signal-to-noise ratio* of a uniform quantizer as

$$\begin{aligned}(\text{SNR})_O &= \frac{P}{\sigma_Q^2} \\ &= \left(\frac{3P}{m_{\max}^2} \right) 2^{2R}\end{aligned}\tag{6.33}$$

Equation (6.33) shows that the output signal-to-noise ratio of a uniform quantizer $(\text{SNR})_O$ increases *exponentially* with increasing number of bits per sample R , which is intuitively satisfying.

EXAMPLE 2

Sinusoidal Modulating Signal

Consider the special case of a full-load sinusoidal modulating signal of amplitude A_m , which utilizes all the representation levels provided. The average signal power is (assuming a load of $1 \, \Omega$)

$$P = \frac{A_m^2}{2}$$

The total range of the quantizer input is $2A_m$, because the modulating signal swings between $-A_m$ and A_m . We may, therefore, set $m_{\max} = A_m$, in which case the use of (6.32) yields the average power (variance) of the quantization noise as

$$\sigma_Q^2 = \frac{1}{3}A_m^2 2^{-2R}$$

Thus, the output signal-to-noise of a uniform quantizer, for a full-load test tone, is

$$(\text{SNR})_O = \frac{A_m^2/2}{A_m^2 2^{-2R}/3} = \frac{3}{2}(2^{2R}) \quad (6.34)$$

Expressing the signal-to-noise (SNR) in decibels, we get

$$10 \log_{10}(\text{SNR})_O = 1.8 + 6R \quad (6.35)$$

The corresponding values of signal-to-noise ratio for various values of L and R , are given in Table 6.1. For sinusoidal modulation, this table provides a basis for making a quick estimate of the number of bits per sample required for a desired output signal-to-noise ratio.

Table 6.1 Signal-to-(quantization) noise ratio for varying number of representation levels for sinusoidal modulation

No. of representation levels L	No. of bits per sample R	SNR (dB)
32	5	31.8
64	6	37.8
128	7	43.8
256	8	49.8

Conditions of Optimality of Scalar Quantizers

In designing a scalar quantizer, the challenge is how to select the representation levels and surrounding partition cells so as to minimize the average quantization power for a fixed number of representation levels.

To state the problem in mathematical terms: consider a message signal $m(t)$ drawn from a stationary process and whose dynamic range, denoted by $-A \leq m \leq A$, is partitioned into a set of L cells, as depicted in Figure 6.11. The boundaries of the partition cells are defined by a set of real numbers m_1, m_2, \dots, m_{L-1} that satisfy the following three conditions:

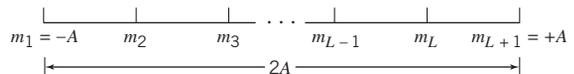
$$m_1 = -A$$

$$m_{L-1} = A$$

$$m_k \leq m_{k-1} \text{ for } k = 1, 2, \dots, L$$

Figure 6.11

Illustrating the partitioning of the dynamic range $-A \leq m \leq A$ of a message signal $m(t)$ into a set of L cells.



The k th partition cell is defined by (6.20), reproduced here for convenience:

$$J_k: m_k < m < m_{k-1} \text{ for } k = 1, 2, \dots, L \quad (6.36)$$

Let the representation levels (i.e., quantization values) be denoted by v_k , $k = 1, 2, \dots, L$. Then, assuming that $d(m, v_k)$ denotes a *distortion measure* for using v_k to represent all those values of the input m that lie inside the partition cell J_k , the goal is to find the two sets $\{v_k\}_{k=1}^L$ and $\{J_k\}_{k=1}^L$ that minimize the *average distortion*

$$D = \sum_{k=1}^L \int_{m \in v_k} d(m, v_k) f_M(m) dm \quad (6.37)$$

where $f_M(m)$ is the probability density function of the random variable M with sample value m .

A commonly used distortion measure is defined by

$$d(m, v_k) = (m - v_k)^2 \quad (6.38)$$

in which case we speak of the *mean-square distortion*. In any event, the optimization problem stated herein is nonlinear, defying an explicit, closed-form solution. To get around this difficulty, we resort to an *algorithmic approach* for solving the problem in an *iterative manner*.

Structurally speaking, the quantizer consists of two components with interrelated design parameters:

- An encoder characterized by the set of partition cells $\{J_k\}_{k=1}^L$; this is located in the transmitter.
- A decoder characterized by the set of representation levels $\{v_k\}_{k=1}^L$; this is located in the receiver.

Accordingly, we may identify two critically important conditions that provide the mathematical basis for all algorithmic solutions to the optimum quantization problem. One condition assumes that we are given a decoder and the problem is to find the optimum encoder in the transmitter. The other condition assumes that we are given an encoder and the problem is to find the optimum decoder in the receiver. Henceforth, these two conditions are referred to as condition I and II, respectively.

Condition I: Optimality of the Encoder for a Given Decoder

The availability of a decoder means that we have a certain *codebook* in mind. Let the codebook be defined by

$$\mathcal{C}: \{v_k\}_{k=1}^L \quad (6.39)$$

Given the codebook \mathcal{C} , the problem is to find the set of partition cells $\{J_k\}_{k=1}^L$ that minimizes the mean-square distortion D . That is, we wish to find the encoder defined by the nonlinear mapping

$$g(m) = v_k, \quad k = 1, 2, \dots, L \quad (6.40)$$

such that we have

$$D = \int_{-A}^A d(m, g(m)) f_M(m) dM \geq \sum_{k=1}^L \int_{m \in J_k} [\min d(m, v_k)] f_M(m) dm \quad (6.41)$$

For the lower bound specified in (6.41) to be attained, we require that the nonlinear mapping of (6.40) be satisfied only if the condition

$$d(m, v_k) \leq d(m, v_j) \quad \text{holds for all } j \neq k \quad (6.42)$$

The necessary condition described in (6.42) for optimality of the encoder for a specified codebook \mathcal{C} is recognized as the *nearest-neighbor condition*. In words, the nearest neighbor condition requires that the partition cell J_k should embody all those values of the input m that are closer to v_k than any other element of the codebook \mathcal{C} . This optimality condition is indeed intuitively satisfying.

Condition II: Optimality of the Decoder for a Given Encoder

Consider next the reverse situation to that described under condition I, which may be stated as follows: optimize the codebook $\mathcal{C} = \{v_k\}_{k=1}^L$ for the decoder, given that the set of partition cells $\{J_k\}_{k=1}^L$ characterizing the encoder is fixed. The criterion for optimization is the average (mean-square) distortion:

$$D = \sum_{k=1}^L \int_{m \in J_k} (m - v_k)^2 f_M(m) dm \quad (6.43)$$

The probability density function $f_M(m)$ is clearly independent of the codebook \mathcal{C} . Hence, differentiating D with respect to the representation level v_k , we readily obtain

$$\frac{\partial D}{\partial v_k} = -2 \sum_{k=1}^L \int_{m \in J_k} (m - v_k) f_M(m) dm \quad (6.44)$$

Setting $\partial D / \partial v_k$ equal to zero and then solving for v_k , we obtain the optimum value

$$v_{k, \text{opt}} = \frac{\int_{m \in J_k} m f_M(m) dm}{\int_{m \in J_k} f_M(m) dm} \quad (6.45)$$

The denominator in (6.45) is just the probability p_k that the random variable M with sample value m lies in the partition cell J_k , as shown by

$$\begin{aligned} p_k &= \mathbb{P}(m_k < M \leq m_k + 1) \\ &= \int_{m \in J_k} f_M(m) dm \end{aligned} \quad (6.46)$$

Accordingly, we may interpret the optimality condition of (6.45) as choosing the representation level v_k to equal the *conditional mean* of the random variable M , given that M lies in the partition cell J_k . We can thus formally state that the condition for optimality of the decoder for a given encoder as follows:

$$v_{k, \text{opt}} = \mathbb{E}[M | m_k < M \leq m_k + 1] \quad (6.47)$$

where \mathbb{E} is the expectation operator. Equation (6.47) is also intuitively satisfying.

Note that the nearest neighbor condition (I) for optimality of the encoder for a given decoder was proved for a generic average distortion. However, the conditional mean requirement (condition II) for optimality of the decoder for a given encoder was proved for

the special case of a mean-square distortion. In any event, these two conditions are necessary for optimality of a scalar quantizer. Basically, the algorithm for designing the quantizer consists of alternately optimizing the encoder in accordance with condition I, then optimizing the decoder in accordance with condition II, and continuing in this manner until the average distortion D reaches a minimum. The optimum quantizer designed in this manner is called the *Lloyd–Max quantizer*.²

6.5 Pulse-Code Modulation

With the material on sampling, PAM, and quantization presented in the preceding sections, the stage is set for describing PCM, for which we offer the following definition:

PCM is a discrete-time, discrete-amplitude waveform-coding process, by means of which an analog signal is directly represented by a sequence of coded pulses.

Specifically, the transmitter consists of two components: a *pulse-amplitude modulator* followed by an *analog-to-digital (A/D) converter*. The latter component itself embodies a *quantizer* followed by an *encoder*. The receiver performs the inverse of these two operations: *digital-to-analog (D/A) conversion* followed by *pulse-amplitude demodulation*. The communication channel is responsible for transporting the encoded pulses from the transmitter to the receiver.

Figure 6.12, a block diagram of the PCM, shows the transmitter, the transmission path from the transmitter output to the receiver input, and the receiver.

It is important to realize, however, that once distortion in the form of quantization noise is introduced into the encoded pulses, there is absolutely nothing that can be done at the receiver to compensate for that distortion. The only design precaution that can be taken is to choose a number of representation levels in the receiver that is large enough to ensure that the quantization noise is imperceptible for human use at the receiver output.

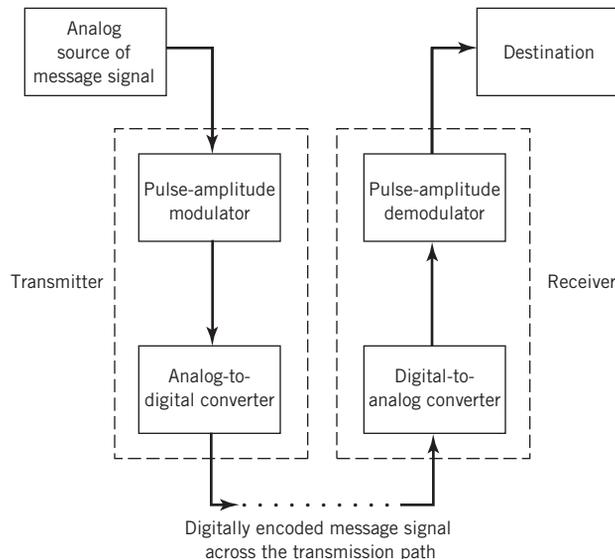


Figure 6.12 Block diagram of PCM system.

Sampling in the Transmitter

The incoming message signal is sampled with a train of rectangular pulses short enough to closely approximate the instantaneous sampling process. To ensure perfect reconstruction of the message signal at the receiver, the sampling rate must be greater than twice the highest frequency component W of the message signal in accordance with the sampling theorem. In practice, a low-pass anti-aliasing filter is used at the front end of the pulse-amplitude modulator to exclude frequencies greater than W before sampling and which are of negligible practical importance. Thus, the application of sampling permits the reduction of the continuously varying message signal to a limited number of discrete values per second.

Quantization in the Transmitter

The PAM representation of the message signal is then quantized in the analog-to-digital converter, thereby providing a new representation of the signal that is discrete in both time and amplitude. The quantization process may follow a uniform law as described in Section 6.4. In telephonic communication, however, it is preferable to use a variable separation between the representation levels for efficient utilization of the communication channel. Consider, for example, the quantization of voice signals. Typically, we find that the range of voltages covered by voice signals, from the peaks of loud talk to the weak passages of weak talk, is on the order of 1000 to 1. By using a *nonuniform quantizer* with the feature that the step size increases as the separation from the origin of the input–output amplitude characteristic of the quantizer is increased, the large end-steps of the quantizer can take care of possible excursions of the voice signal into the large amplitude ranges that occur relatively infrequently. In other words, the weak passages needing more protection are favored at the expense of the loud passages. In this way, a nearly uniform percentage precision is achieved throughout the greater part of the amplitude range of the input signal. The end result is that fewer steps are needed than would be the case if a uniform quantizer were used; hence the improvement in channel utilization.

Assuming memoryless quantization, the use of a nonuniform quantizer is equivalent to passing the message signal through a *compressor* and then applying the compressed signal to a *uniform quantizer*, as illustrated in Figure 6.13a. A particular form of *compression law* that is used in practice is the so-called μ -law,³ which is defined by

$$|v| = \frac{\ln(1 + \mu|m|)}{\ln(1 + \mu)} \quad (6.48)$$

where \ln , i.e., \log_e , denotes the natural logarithm, m and v are the input and output voltages of the *compressor*, and μ is a positive constant. It is assumed that m and,

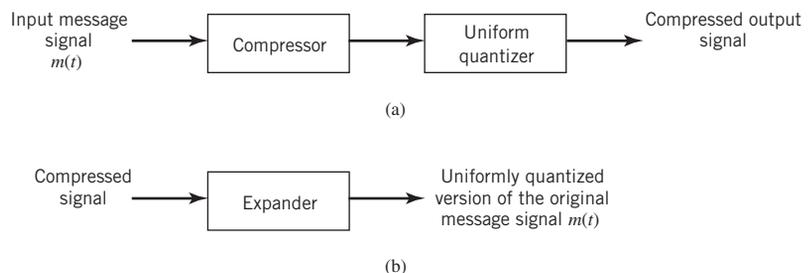
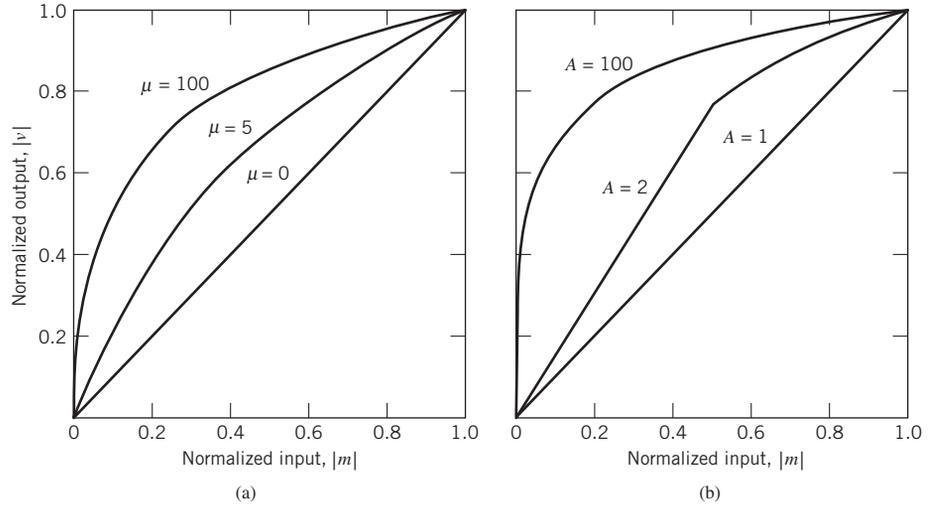


Figure 6.13

(a) Nonuniform quantization of the message signal in the transmitter. (b) Uniform quantization of the original message signal in the receiver.

Figure 6.14
Compression laws:
(a) μ -law;
(b) A -law.



therefore, v are scaled so that they both lie inside the interval $[-1, 1]$. The μ -law is plotted for three different values of μ in Figure 6.14a. The case of uniform quantization corresponds to $\mu = 0$. For a given value of μ , the reciprocal slope of the compression curve that defines the quantum steps is given by the derivative of the absolute value $|m|$ with respect to the corresponding absolute value $|v|$; that is,

$$\frac{d|m|}{d|v|} = \frac{\ln(1+\mu)}{\mu}(1+\mu|m|) \quad (6.49)$$

From (6.49) it is apparent that the μ -law is neither strictly linear nor strictly logarithmic. Rather, it is approximately linear at low input levels corresponding to $\mu|m| \ll 1$ and approximately logarithmic at high input levels corresponding to $\mu|m| \gg 1$.

Another compression law that is used in practice is the so-called A -law, defined by

$$|v| = \begin{cases} \frac{A|m|}{1 + \ln A}, & 0 \leq |m| \leq \frac{1}{A} \\ \frac{1 + \ln(A|m|)}{1 + \ln A}, & \frac{1}{A} \leq |m| \leq 1 \end{cases} \quad (6.50)$$

where A is another positive constant. Equation (6.50) is plotted in Figure 6.14b for varying A . The case of uniform quantization corresponds to $A = 1$. The reciprocal slope of this second compression curve is given by the derivative of $|m|$ with respect to $|v|$, as shown by

$$\frac{d|m|}{d|v|} = \begin{cases} \frac{1 + \ln A}{A}, & 0 \leq |m| \leq \frac{1}{A} \\ (1 + \ln A)|m|, & \frac{1}{A} \leq |m| \leq 1 \end{cases} \quad (6.51)$$

To restore the signal samples to their correct relative level, we must, of course, use a device in the receiver with a characteristic complementary to the compressor. Such a device is called an *expander*. Ideally, the compression and expansion laws are exactly the inverse of each other. With this provision in place, we find that, except for the effect of quantization, the expander output is equal to the compressor input. The cascade combination of a *compressor* and an *expander*, depicted in Figure 6.13, is called a *componder*.

For both the μ -law and A -law, the dynamic range capability of the compander improves with increasing μ and A , respectively. The SNR for low-level signals increases at the expense of the SNR for high-level signals. To accommodate these two conflicting requirements (i.e., a reasonable SNR for both low- and high-level signals), a compromise is usually made in choosing the value of parameter μ for the μ -law and parameter A for the A -law. The typical values used in practice are $\mu = 255$ for the μ -law and $A = 87.6$ for the A -law.⁴

Encoding in the Transmitter

Through the combined use of sampling and quantization, the specification of an analog message signal becomes limited to a discrete set of values, but not in the form best suited to transmission over a telephone line or radio link. To exploit the advantages of sampling and quantizing for the purpose of making the transmitted signal more robust to noise, interference, and other channel impairments, we require the use of an *encoding process* to translate the discrete set of sample values to a more appropriate form of signal. Any plan for representing each of this discrete set of values as a particular arrangement of discrete events constitutes a *code*. Table 6.2 describes the one-to-one correspondence between representation levels and codewords for a binary number system for $R = 4$ bits per sample. Following the terminology of Chapter 5, the two symbols of a binary code are customarily denoted as 0 and 1. In practice, the binary code is the preferred choice for encoding for the following reason:

The maximum advantage over the effects of noise encountered in a communication system is obtained by using a binary code because a binary symbol withstands a relatively high level of noise and, furthermore, it is easy to regenerate.

The last signal-processing operation in the transmitter is that of *line coding*, the purpose of which is to represent each binary codeword by a sequence of pulses; for example, symbol 1 is represented by the presence of a pulse and symbol 0 is represented by absence of the pulse. Line codes are discussed in Section 6.10. Suppose that, in a binary code, each codeword consists of R bits. Then, using such a code, we may represent a total of 2^R distinct numbers. For example, a sample quantized into one of 256 levels may be represented by an 8-bit codeword.

Inverse Operations in the PCM Receiver

The first operation in the receiver of a PCM system is to *regenerate* (i.e., reshape and clean up) the received pulses. These clean pulses are then regrouped into codewords and decoded (i.e., mapped back) into a quantized pulse-amplitude modulated signal. The *decoding* process involves generating a pulse the amplitude of which is the linear sum of all the pulses in the codeword. Each pulse is weighted by its place value ($2^0, 2^1, 2^2, \dots, 2^{R-1}$) in the code, where R is the number of bits per sample. Note, however, that whereas the analog-to-digital

Table 6.2 Binary number system for $T = 4$ bits/sample

Ordinal number of representation level	Level number expressed as sum of powers of 2	Binary number
0		0000
1	2^0	0001
2	2^1	0010
3	$2^1 + 2^0$	0011
4	2^2	0100
5	$2^2 + 2^0$	0101
6	$2^2 + 2^1$	0110
7	$2^2 + 2^1 + 2^0$	0111
8	2^3	1000
9	$2^3 + 2^0$	1001
10	$2^3 + 2^1$	1010
11	$2^3 + 2^1 + 2^0$	1011
12	$2^3 + 2^2$	1100
13	$2^3 + 2^2 + 2^0$	1101
14	$2^3 + 2^2 + 2^1$	1110
15	$2^3 + 2^2 + 2^1 + 2^0$	1111

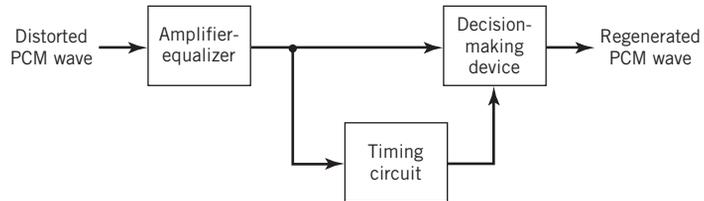
converter in the transmitter involves both quantization and encoding, the digital-to-analog converter in the receiver involves decoding only, as illustrated in Figure 6.12.

The final operation in the receiver is that of *signal reconstruction*. Specifically, an estimate of the original message signal is produced by passing the decoder output through a *low-pass reconstruction filter* whose cutoff frequency is equal to the message bandwidth W . Assuming that the transmission link (connecting the receiver to the transmitter) is error free, the reconstructed message signal includes no noise with the exception of the initial distortion introduced by the quantization process.

PCM Regeneration along the Transmission Path

The most important feature of a PCM systems is its ability to control the effects of distortion and noise produced by transmitting a PCM signal through the channel, connecting the receiver to the transmitter. This capability is accomplished by reconstructing the PCM signal through a chain of *regenerative repeaters*, located at sufficiently close spacing along the transmission path.

Figure 6.15
Block diagram of
regenerative repeater.



As illustrated in Figure 6.15, three basic functions are performed in a regenerative repeater: *equalization*, *timing*, and *decision making*. The equalizer shapes the received pulses so as to compensate for the effects of amplitude and phase distortions produced by the non-ideal transmission characteristics of the channel. The timing circuitry provides a periodic pulse train, derived from the received pulses, for sampling the equalized pulses at the instants of time where the SNR ratio is a maximum. Each sample so extracted is compared with a predetermined *threshold* in the decision-making device. In each bit interval, a decision is then made on whether the received symbol is 1 or 0 by observing whether the threshold is exceeded or not. If the threshold is exceeded, a clean new pulse representing symbol 1 is transmitted to the next repeater; otherwise, another clean new pulse representing symbol 0 is transmitted. In this way, it is possible for the accumulation of distortion and noise in a repeater span to be almost completely removed, provided that the disturbance is not too large to cause an error in the decision-making process. Ideally, except for delay, the regenerated signal is exactly the same as the signal originally transmitted. In practice, however, the regenerated signal departs from the original signal for two main reasons:

1. The unavoidable presence of channel noise and interference causes the repeater to make wrong decisions occasionally, thereby introducing *bit errors* into the regenerated signal.
2. If the spacing between received pulses deviates from its assigned value, a *jitter* is introduced into the regenerated pulse position, thereby causing distortion.

The important point to take from this subsection on PCM is the fact that regeneration along the transmission path is provided across the spacing between individual regenerative repeaters (including the last stage of regeneration at the receiver input) provided that the spacing is short enough. If the transmitted SNR ratio is high enough, then the regenerated PCM data stream is the same as the transmitted PCM data stream, except for a practically negligibly small *bit error rate* (BER). In other words, under these operating conditions, performance degradation in the PCM system is essentially confined to quantization noise in the transmitter.

6.6 Noise Considerations in PCM Systems

The performance of a PCM system is influenced by two major sources of noise:

1. *Channel noise*, which is introduced anywhere between the transmitter output and the receiver input; channel noise is always present, once the equipment is switched on.
2. *Quantization noise*, which is introduced in the transmitter and is carried all the way along to the receiver output; unlike channel noise, quantization noise is *signal dependent*, in the sense that it disappears when the message signal is switched off.

Naturally, these two sources of noise appear simultaneously once the PCM system is in operation. However, the traditional practice is to consider them separately, so that we may develop insight into their individual effects on the system performance.

The main effect of channel noise is to introduce *bit errors* into the received signal. In the case of a binary PCM system, the presence of a bit error causes symbol 1 to be mistaken for symbol 0, or vice versa. Clearly, the more frequently bit errors occur, the more dissimilar the receiver output becomes compared with the original message signal. The fidelity of information transmission by PCM in the presence of channel noise may be measured in terms of the *average probability of symbol error*, which is defined as the probability that the reconstructed symbol at the receiver output differs from the transmitted binary symbol on the average. The average probability of symbol error, also referred to as the BER, assumes that all the bits in the original binary wave are of equal importance. When, however, there is more interest in restructuring the analog waveform of the original message signal, different symbol errors may be *weighted* differently; for example, an error in the most significant bit in a codeword (representing a quantized sample of the message signal) is more harmful than an error in the least significant bit.

To optimize system performance in the presence of channel noise, we need to minimize the average probability of symbol error. For this evaluation, it is customary to model the channel noise as an ideal *additive white Gaussian noise* (AWGN) channel. The effect of channel noise can be made practically negligible by using an adequate signal energy-to-noise density ratio through the provision of short-enough spacing between the regenerative repeaters in the PCM system. In such a situation, the performance of the PCM system is essentially limited by quantization noise acting alone.

From the discussion of quantization noise presented in Section 6.4, we recognize that quantization noise is essentially under the designer's control. It can be made negligibly small through the use of an adequate number of representation levels in the quantizer and the selection of a companding strategy matched to the characteristics of the type of message signal being transmitted. We thus find that the use of PCM offers the possibility of building a communication system that is *rugged* with respect to channel noise on a scale that is beyond the capability of any analog communication system; hence its use as a *standard* against which other waveform coders (e.g., DPCM and DM) are compared.

Error Threshold

The underlying theory of BER calculation in a PCM system is deferred to Chapter 8. For the present, it suffices to say that the average probability of symbol error in a binary encoded PCM receiver due to AWGN depends solely on E_b/N_0 , which is defined as *the ratio of the transmitted signal energy per bit E_b , to the noise spectral density N_0* . Note that the ratio E_b/N_0 is dimensionless even though the quantities E_b and N_0 have different physical meaning. In Table 6.3, we present a summary of this dependence for the case of a binary PCM system, in which symbols 1 and 0 are represented by rectangular pulses of equal but opposite amplitudes. The results presented in the last column of the table assume a bit rate of 10^5 bits/s.

From Table 6.3 it is clear that there is an *error threshold* (at about 11 dB). For E_b/N_0 below the error threshold the receiver performance involves significant numbers of errors, and above it the effect of channel noise is practically negligible. In other words, provided that the ratio E_b/N_0 exceeds the error threshold, channel noise has virtually no effect on

Table 6.3 Influence of E_b/N_0 on the probability of error

E_b/N_0 (dB)	Probability of error P_e	For a bit rate of 10^5 bits/s, this is about one error every
4.3	10^{-2}	10^{-3} s
8.4	10^{-4}	10^{-1} s
10.6	10^{-6}	10 s
12.0	10^{-8}	20 min
13.0	10^{-10}	1 day
14.0	10^{-12}	3 months

the receiver performance, which is precisely the goal of PCM. When, however, E_b/N_0 drops below the error threshold, there is a sharp increase in the rate at which errors occur in the receiver. Because decision errors result in the construction of incorrect codewords, we find that when the errors are frequent, the reconstructed message at the receiver output bears little resemblance to the original message signal.

An important characteristic of a PCM system is its *ruggedness to interference*, caused by impulsive noise or cross-channel interference. The combined presence of channel noise and interference causes the error threshold necessary for satisfactory operation of the PCM system to increase. If, however, an adequate margin over the error threshold is provided in the first place, the system can withstand the presence of relatively large amounts of interference. In other words, a PCM system is *robust* with respect to channel noise and interference, providing further confirmation to the point made in the previous section that performance degradation in PCM is essentially confined to quantization noise in the transmitter.

PCM Noise Performance Viewed in Light of the Information Capacity Law

Consider now a PCM system that is known to operate above the error threshold, in which case we would be justified to ignore the effect of channel noise. In other words, the noise performance of the PCM system is essentially determined by quantization noise acting alone. Given such a scenario, how does the PCM system fare compared with the information capacity law, derived in Chapter 5?

To address this question of practical importance, suppose that the system uses a codeword consisting of n symbols with each symbol representing one of M possible discrete amplitude levels; hence the reference to the system as an “ M -ary” PCM system. For this system to operate above the error threshold, there must be provision for a large enough noise margin.

For the PCM system to operate above the error threshold as proposed, the requirement for a noise margin that is sufficiently large to maintain a negligible error rate due to channel noise. This, in turn, means there must be a certain separation between the M discrete amplitude levels. Call this separation $c\sigma$, where c is a constant and $\sigma^2 = N_0B$ is the

noise variance measured in a channel bandwidth B . The number of amplitude levels M is usually an integer power of 2. The average transmitted power will be least if the amplitude range is symmetrical about zero. Then, the discrete amplitude levels, normalized with respect to the separation $c\sigma$, will have the values $\pm 1/2, \pm 3/2, \dots, \pm(M-1)/2$. We assume that these M different amplitude levels are equally likely. Accordingly, we find that the average transmitted power is given by

$$\begin{aligned} P &= \frac{2}{M} \left[\left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + \dots + \left(\frac{M-1}{2}\right)^2 \right] (c\sigma)^2 \\ &= c^2 \sigma^2 \left(\frac{M^2 - 1}{12} \right) \end{aligned} \quad (6.52)$$

Suppose that the M -ary PCM system described herein is used to transmit a message signal with its highest frequency component equal to W hertz. The signal is sampled at the Nyquist rate of $2W$ samples per second. We assume that the system uses a quantizer of the midrise type, with L equally likely representation levels. Hence, the probability of occurrence of any one of the L representation levels is $1/L$. Correspondingly, the amount of information carried by a single sample of the signal is $\log_2 L$ bits. With a maximum sampling rate of $2W$ samples per second, the maximum rate of information transmission of the PCM system measured in bits per second is given by

$$R_b = 2W \log_2 L \text{ bits/s} \quad (6.53)$$

Since the PCM system uses a codeword consisting of n code elements with each one having M possible discrete amplitude values, we have M^n different possible codewords. For a unique encoding process, therefore, we require

$$L = M^n \quad (6.54)$$

Clearly, the rate of information transmission in the system is unaffected by the use of an encoding process. We may, therefore, eliminate L between (6.53) and (6.54) to obtain

$$R_b = 2Wn \log_2 M \text{ bits/s} \quad (6.55)$$

Equation (6.52) defines the average transmitted power required to maintain an M -ary PCM system operating above the error threshold. Hence, solving this equation for the number of discrete amplitude levels, we may express the number M in terms of the average transmitted power P and channel noise variance $\sigma^2 = N_0B$ as follows:

$$M = \left(1 + \frac{12P}{c^2 N_0 B} \right)^{1/2} \quad (6.56)$$

Therefore, substituting (6.56) into (6.55), we obtain

$$R_b = Wn \log_2 \left(1 + \frac{12P}{c^2 N_0 B} \right) \quad (6.57)$$

The channel bandwidth B required to transmit a rectangular pulse of duration $1/(2nW)$, representing a symbol in the codeword, is given by

$$B = \kappa n W \quad (6.58)$$

where κ is a constant with a value lying between 1 and 2. Using the minimum possible value $\kappa = 1$, we find that the channel bandwidth $B = nW$. We may thus rewrite (6.57) as

$$R_b = B \log_2 \left(1 + \frac{12P}{c^2 N_0 B} \right) \text{ bits/s} \quad (6.59)$$

which defines the upper bound on the information capacity realizable by an M -ary PCM system.

From Chapter 5 we recall that, in accordance with Shannon's information capacity law, the *ideal transmission system* is described by the formula

$$C = B \log_2 \left(1 + \frac{P}{N_0 B} \right) \text{ bits/s} \quad (6.60)$$

The most interesting point derived from the comparison of (6.59) with (6.60) is the fact that (6.59) is of the right mathematical form in an information-theoretic context. To be more specific, we make the following statement:

Power and bandwidth in a PCM system are exchanged on a logarithmic basis, and the information capacity of the system is proportional to the channel bandwidth B .

As a corollary, we may go on to state:

When the SNR ratio is high, the bandwidth-noise trade-off follows an exponential law in PCM.

From the study of noise in analog modulation systems,⁵ it is known that the use of frequency modulation provides the best improvement in SNR ratio. To be specific, when the carrier-to-noise ratio is high enough, the bandwidth-noise trade-off follows a *square law* in frequency modulation (FM). Accordingly, in comparing the noise performance of FM with that of PCM we make the concluding statement:

PCM is more efficient than FM in trading off an increase in bandwidth for improved noise performance.

Indeed, this statement is further testimony for the PCM being viewed as a standard for waveform coding.

6.7 Prediction-Error Filtering for Redundancy Reduction

When a voice or video signal is sampled at a rate slightly higher than the Nyquist rate, as usually done in PCM, the resulting sampled signal is found to exhibit a high degree of *correlation* between adjacent samples. The meaning of this high correlation is that, in an average sense, the signal does not change rapidly from one sample to the next. As a result, the difference between adjacent samples has a variance that is smaller than the variance of the original signal. When these highly correlated samples are encoded, as in the standard PCM system, the resulting encoded signal contains *redundant information*. This kind of signal structure means that symbols that are not absolutely essential to the transmission of

information are generated as a result of the conventional encoding process described in Section 6.5. By reducing this redundancy before encoding, we obtain a *more efficient* coded signal, which is the basic idea behind DPCM. Discussion of this latter form of waveform coding is deferred to the next section. In this section we discuss prediction-error filtering, which provides a method for reduction and, therefore, improved waveform coding.

Theoretical Considerations

To elaborate, consider the block diagram of Figure 6.16a, which includes:

- a direct forward path from the input to the output;
- a predictor in the forward direction as well; and
- a comparator for computing the difference between the input signal and the predictor output.

The difference signal, so computed, is called the *prediction error*. Correspondingly, a filter that operates on the message signal to produce the prediction error, illustrated in Figure 6.16a, is called a *prediction-error filter*.

To simplify the presentation, let

$$m_n = m(nT_s) \quad (6.61)$$

denote a sample of the message signal $m(t)$ taken at time $t = nT_s$. Then, with \hat{m}_n denoting the corresponding predictor output, the prediction error is defined by

$$e_n = m_n - \hat{m}_n \quad (6.62)$$

where e_n is the amount by which the predictor fails to predict the input sample m_n exactly. In any case, the objective is to design the predictor so as to *minimize the variance* of the prediction error e_n . In so doing, we effectively end up using a smaller number of bits to represent e_n than the original message sample m_n ; hence, the need for a smaller transmission bandwidth.

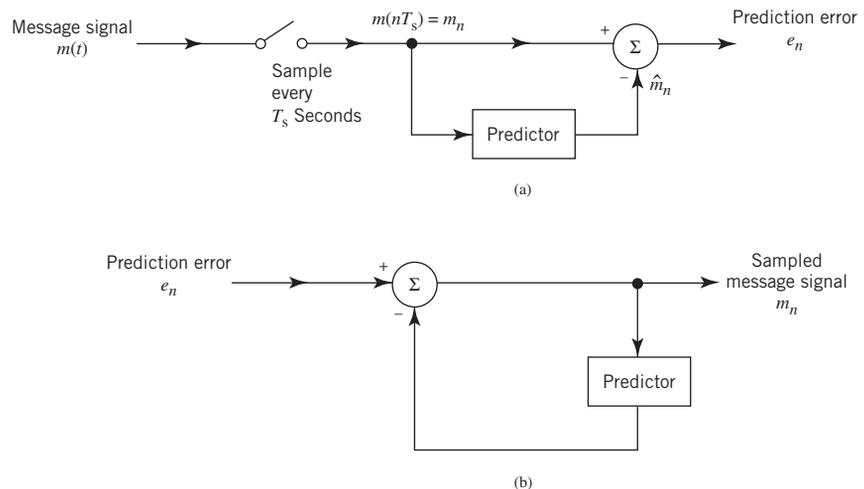


Figure 6.16 Block diagram of (a) prediction-error filter and (b) its inverse.

The prediction-error filter operates on the message signal on a sample-by-sample basis to produce the prediction error. With such an operation performed in the transmitter, how do we recover the original message signal from the prediction error at the receiver? To address this fundamental question in a simple-minded and yet practical way, we invoke the use of *linearity*. Let the *operator* \mathbf{L} denote the action of the predictor, as shown by

$$\hat{m}_n = \mathbf{L}[m_n] \quad (6.63)$$

Accordingly, we may rewrite (6.62) in operator form as follows:

$$\begin{aligned} e_n &= m_n - \mathbf{L}[m_n] \\ &= (1 - \mathbf{L})[m_n] \end{aligned} \quad (6.64)$$

Under the assumption of linearity, we may invert (6.64) to recover the message sample from the prediction error, as shown by

$$m_n = \left(\frac{1}{1 - \mathbf{L}} \right) [e_n] \quad (6.65)$$

Equation (6.65) is immediately recognized as the equation of a *feedback system*, as illustrated in Figure 6.16b. Most importantly, in functional terms, this feedback system may be viewed as the *inverse of prediction-error filtering*.

Discrete-Time Structure for Prediction

To simplify the design of the linear predictor in Figure 6.16, we propose to use a discrete-time structure in the form of a *finite-duration impulse response (FIR) filter*, which is well known in the digital signal-processing literature. The FIR filter was briefly discussed in Chapter 2.

Figure 6.17 depicts an FIR filter, consisting of two functional components:

- a set of p *unit-delay elements*, each of which is represented by z^{-1} ; and
- a corresponding set of *adders* used to sum the scaled versions of the delayed inputs,

$$m_{n-1}, m_{n-2}, \dots, m_{n-p}$$

The overall linearly predicted output is thus defined by the *convolution sum*

$$\hat{m}_n = \sum_{k=1}^p w_k m_{n-k} \quad (6.66)$$

where p is called the *prediction order*. Minimization of the prediction-error variance is achieved by a proper choice of the FIR filter-coefficients as described next.

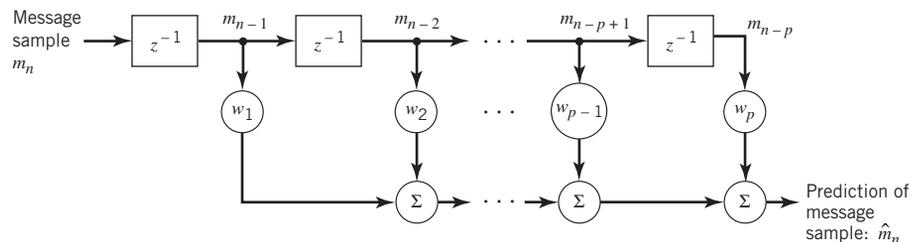


Figure 6.17 Block diagram of an FIR filter of order p .

First, however, we make the following assumption:

The message signal $m(t)$ is drawn from a stationary stochastic processor $M(t)$ with zero mean.

This assumption may be satisfied by processing the message signal on a block-by-block basis, with each block being just long enough to satisfy the assumption in a *pseudo-stationary manner*. For example, a block duration of 40 ms is considered to be adequate for voice signals.

With the random variable M_n assumed to have zero mean, it follows that the variance of the prediction error e_n is the same as its mean-square value. We may thus define

$$J = \mathbb{E}[e^2(n)] \quad (6.67)$$

as the *index of performance*. Substituting (6.65) and (6.66) into (6.67) and then expanding terms, the index of performance is expressed as follows:

$$J(\mathbf{w}) = \mathbb{E}[m_n^2] - 2 \sum_{k=1}^p w_k \mathbb{E}[m_n m_{n-k}] + \sum_{j=1}^p \sum_{k=1}^p w_j w_k \mathbb{E}[m_{n-j} m_{n-k}] \quad (6.68)$$

Moreover, under the above assumption of pseudo-stationarity, we may go on to introduce the following second-order statistical parameters for m_n treated as a sample of the stochastic process $M(t)$ at $t = nT_s$:

1. Variance

$$\begin{aligned} \sigma_M^2 &= \mathbb{E}[(m_n - \mathbb{E}[m_n])^2] \\ &= \mathbb{E}[m_n^2] \text{ for } \mathbb{E}[m_n] = 0 \end{aligned} \quad (6.69)$$

2. Autocorrelation function

$$R_{M, k-j} = \mathbb{E}[m_{n-j} m_{n-k}] \quad (6.70)$$

Note that to simplify the notation in (6.67) to (6.70), we have applied the expectation operator \mathbb{E} to samples rather than the corresponding random variables.

In any event, using (6.69) and (6.70), we may reformulate the index of performance of (6.68) in the new form involving statistical parameters:

$$J(\mathbf{w}) = \sigma_M^2 - 2 \sum_{k=1}^p w_k R_{M, k} + \sum_{j=1}^p \sum_{k=1}^p w_j w_k R_{M, k-j} \quad (6.71)$$

Differentiating this index of performance with respect to the filter coefficients, setting the resulting expression equal to zero, and then rearranging terms, we obtain the following system of simultaneous equations:

$$\sum_{j=1}^p w_{o,j} R_{M, k-j} = R_{M, k}, \quad k = 1, 2, \dots, p \quad (6.72)$$

where $w_{o,j}$ is the optimal value of the j th filter coefficient w_j . This optimal set of equations is the discrete-time version of the celebrated *Wiener–Hopf equations* for linear prediction.

With compactness of mathematical exposition in mind, we find it convenient to formulate the Wiener–Hopf equations in matrix form, as shown by

$$\mathbf{R}_M \mathbf{w}_o = \mathbf{r}_M \quad (6.73)$$

where

$$\mathbf{w}_o = [w_{o,1}, w_{o,2}, \dots, w_{o,p}]^T \quad (6.74)$$

is the p -by-1 *optimum coefficient vector* of the FIR predictor,

$$\mathbf{r}_M = [R_{M,1}, R_{M,2}, \dots, R_{M,p}]^T \quad (6.75)$$

is the p -by-1 *autocorrelation vector* of the original message signal, excluding the mean-square value represented by $R_{M,0}$, and

$$\mathbf{R}_M = \begin{bmatrix} R_{M,0} & R_{M,1} & \cdots & R_{M,p-1} \\ R_{M,1} & R_{M,0} & \cdots & R_{M,p-2} \\ \cdots & \cdots & \cdots & \cdots \\ R_{M,p-1} & R_{M,p-2} & \cdots & R_{M,0} \end{bmatrix} \quad (6.76)$$

is the p -by- p *correlation matrix* of the original message signal, including $R_{M,0}$.⁶

Careful examination of (6.76) reveals the *Toeplitz property* of the autocorrelation matrix \mathbf{R}_M , which embodies two distinctive characteristics:

1. All the elements on the main diagonal of the matrix \mathbf{R}_M are equal to the mean-square value or, equivalently under the zero-mean assumption, the variance of the message sample m_n , as shown by

$$R_{M,0} = \sigma_M^2$$

2. The matrix is *symmetric* about the main diagonal.

This Toeplitz property is a direct consequence of the assumption that message signal $m(t)$ is the sample function of a stationary stochastic process. From a practical perspective, the Toeplitz property of the autocorrelation matrix \mathbf{R}_M is important in that all of its elements are uniquely defined by the *autocorrelation sequence* $\{R_{M,k}\}_{k=0}^{p-1}$. Moreover, from the defining equation (6.75), it is clear that the autocorrelation vector \mathbf{r}_M is uniquely defined by the autocorrelation sequence $\{R_{M,k}\}_{k=1}^p$. We may therefore make the following statement:

The p filter coefficients of the optimized linear predictor, configured in the form of an FIR filter, are uniquely defined by the variance $\sigma_M^2 = R_{M,0}$ and the autocorrelation sequence $\{R_{M,k}\}_{k=0}^{p-1}$, which pertain to the message signal $m(t)$ drawn from a weakly stationary process.

Typically, we have

$$|R_{M,k}| < R_{M,0} \quad \text{for } k = 1, 2, \dots, p$$

Under this condition, we find that the autocorrelation matrix \mathbf{R}_M is also invertible; that is, the inverse matrix \mathbf{R}_M^{-1} exists. We may therefore solve (6.73) for the unknown value of the optimal coefficient vector \mathbf{w}_o using the formula⁷

$$\mathbf{w}_o = \mathbf{R}_M^{-1} \mathbf{r}_M \quad (6.77)$$

Thus, given the variance σ_M^2 and autocorrelation sequence $\{R_{M,k}\}_{k=1}^p$, we may uniquely determine the optimized coefficient vector of the linear predictor, \mathbf{w}_o , defining an FIR filter of order p ; and with it our design objective is satisfied.

To complete the linear prediction theory presented herein, we need to find the minimum mean-square value of prediction error, resulting from the use of the optimized predictor. We do this by first reformulating (6.71) in the matrix form:

$$J(\mathbf{w}_o) = \sigma_M^2 - 2\mathbf{w}_o^T \mathbf{r}_M + \mathbf{w}_o^T \mathbf{R}_M \mathbf{w}_o \quad (6.78)$$

where the superscript T denotes *matrix transposition*, $\mathbf{w}_o^T \mathbf{r}_M$ is the *inner product* of the p -by-1 vectors \mathbf{w}_o and \mathbf{r}_M , and the matrix product $\mathbf{w}_o^T \mathbf{R}_M \mathbf{w}_o$ is a *quadratic form*. Then, substituting the optimum formula of (6.77) into (6.78), we find that the *minimum mean-square value of prediction error* is given by

$$\begin{aligned} J_{\min} &= \sigma_M^2 - 2(\mathbf{R}_M^{-1} \mathbf{r}_M)^T \mathbf{r}_M + (\mathbf{R}_M^{-1} \mathbf{r}_M)^T \mathbf{R}_M (\mathbf{R}_M^{-1} \mathbf{r}_M) \\ &= \sigma_M^2 - 2\mathbf{r}_M^T \mathbf{R}_M^{-1} \mathbf{r}_M + \mathbf{r}_M^T \mathbf{R}_M^{-1} \mathbf{r}_M \\ &= \sigma_M^2 - \mathbf{r}_M^T \mathbf{R}_M^{-1} \mathbf{r}_M \end{aligned} \quad (6.79)$$

where we have used the property that the autocorrelation matrix of a weakly stationary process is *symmetric*; that is,

$$\mathbf{R}_M^T = \mathbf{R}_M \quad (6.80)$$

By definition, the quadratic form $\mathbf{r}_M^T \mathbf{R}_M^{-1} \mathbf{r}_M$ is always positive. Accordingly, from (6.79) it follows that the minimum value of the mean-square prediction error J_{\min} is always smaller than the variance σ_M^2 of the zero-mean message sample m_n that is being predicted. Through the use of linear prediction as described herein, we have thus satisfied the objective:

To design a prediction-error filter the output of which has a smaller variance than the variance of the message sample applied to its input, we need to follow the optimum formula of (6.77).

This statement provides the rationale for going on to describe how the bandwidth requirement of the standard PCM can be reduced through redundancy reduction. However, before proceeding to do so, it is instructive that we consider an adaptive implementation of the linear predictor.

Linear Adaptive Prediction

The use of (6.77) for calculating the optimum weight vector of a linear predictor requires knowledge of the autocorrelation function $R_{m,k}$ of the message signal sequence $\{m_k\}_{k=0}^p$ where p is the prediction order. What if knowledge of this sequence is not available? In situations of this kind, which occur frequently in practice, we may resort to the use of an *adaptive predictor*.

The predictor is said to be adaptive in the following sense:

- Computation of the tap weights w_k , $k = 1, 2, \dots, p$, proceeds in an iterative manner, starting from some arbitrary initial values of the tap weights.
- The algorithm used to adjust the tap weights (from one iteration to the next) is “self-designed,” operating solely on the basis of available data.

The aim of the algorithm is to find the minimum point of the *bowl-shaped error surface* that describes the dependence of the cost function J on the tap weights. It is, therefore, intuitively reasonable that successive adjustments to the tap weights of the predictor be made in the direction of the steepest descent of the error surface; that is, in a direction opposite to the *gradient vector* whose elements are defined by

$$g_k = \frac{\partial J}{\partial w_k}, \quad k = 1, 2, \dots, p \quad (6.81)$$

This is indeed the idea behind the *method of deepest descent*. Let $w_{k,n}$ denote the value of the k th tap weight at iteration n . Then, the updated value of this weight at iteration $n + 1$ is defined by

$$w_{k,n+1} = w_{k,n} - \frac{1}{2}\mu g_k, \quad k = 1, 2, \dots, p \quad (6.82)$$

where μ is a *step-size parameter* that controls the speed of adaptation and the factor $1/2$ is included for convenience of presentation. Differentiating the cost function J of (6.68) with respect to w_k , we readily find that

$$g_k = -2\mathbb{E}[m_n m_{n-k}] + \sum_{j=1}^p w_j \mathbb{E}[m_{n-j} m_{n-k}] \quad (6.83)$$

From a practical perspective, the formula for the gradient g_k in (6.83) could do with further simplification that ignores the expectation operator. In effect, *instantaneous values are used as estimates of autocorrelation functions*. The motivation for this simplification is to permit the adaptive process to proceed forward on a step-by-step basis in a self-organized manner. Clearly, by ignoring the expectation operator in (6.83), the gradient g_k takes on a time-dependent value, denoted by $g_{k,n}$. We may thus write

$$g_{k,n} = -2m_n m_{n-k} + 2m_{n-k} \sum_{j=1}^p \hat{w}_{j,n} m_{n,j}, \quad k = 1, 2, \dots, p \quad (6.84)$$

where $\hat{w}_{j,n}$ is an estimate of the filter coefficient $w_{j,n}$ at time n .

The stage is now set for substituting (6.84) into (6.82), where in the latter equation $\hat{w}_{k,n}$ is substituted for $w_{k,n}$; this change is made to account for dispensing with the expectation operator:

$$\begin{aligned}
\hat{w}_{k,n+1} &= \hat{w}_{k,n} - \frac{1}{2} \mu g_{k,n} \\
&= \hat{w}_{k,n} + \mu \left(m_n m_{n-k} - \sum_{j=1}^p \hat{w}_{j,n} m_{n-j} m_{n-k} \right) \\
&= \hat{w}_{k,n} + \mu m_{n-k} \left(m_n - \sum_{j=1}^p \hat{w}_{j,n} m_{n-j} \right) \\
&= \hat{w}_{k,n} + \mu m_{n-k} e_n
\end{aligned} \tag{6.85}$$

where e_n is the *new prediction error* defined by

$$e_n = m_n - \sum_{j=1}^p \hat{w}_{j,n} m_{n-j} \tag{6.86}$$

Note that the current value of the message signal, m_n , plays a role as the *desired response for predicting* the value of m_n given the past values of the message signal: m_{n-1} , m_{n-2} , ..., m_{n-p} .

In words, we may express the adaptive filtering algorithm of (6.85) as follows:

$$\left(\begin{array}{c} \text{Updated value of the } k\text{th} \\ \text{filter coefficient at time } n+1 \end{array} \right) = \left(\begin{array}{c} \text{Old value of the same} \\ \text{filter coefficient at time } n \end{array} \right) + \left(\begin{array}{c} \text{Step-size} \\ \text{parameter} \end{array} \right) \times \left(\begin{array}{c} \text{Message signal } m_n \\ \text{delayed by } k \text{ time steps} \end{array} \right) \left(\begin{array}{c} \text{Prediction error} \\ \text{computed at time } n \end{array} \right)$$

The algorithm just described is the popular *least-mean-square (LMS) algorithm*, formulated for the purpose of linear prediction. The reason for popularity of this adaptive filtering algorithm is the simplicity of its implementation. In particular, the computational complexity of the algorithm, measured in terms of the number of additions and multiplications, is *linear* in the prediction order p . Moreover, the algorithm is not only *computationally efficient* but it is also *effective in performance*.

The LMS algorithm is a *stochastic* adaptive filtering algorithm, stochastic in the sense that, starting from the *initial condition* defined by $\{w_{k,0}\}_{k=1}^p$, it seeks to find the minimum point of the error surface by following a zig-zag path. However, it never finds this minimum point exactly. Rather, it continues to execute a random motion around the minimum point of the error surface (Haykin, 2013).

6.8 Differential Pulse-Code Modulation

DPCM, the scheme to be considered for *channel-bandwidth conservation*, exploits the idea of linear prediction theory with a practical difference:

In the transmitter, the linear prediction is performed on a quantized version of the message sample instead of the message sample itself, as illustrated in Figure 6.18.

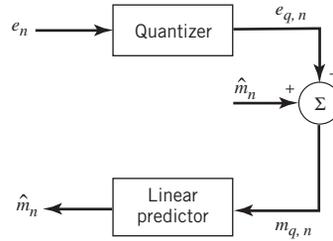


Figure 6.18 Block diagram of a differential quantizer.

The resulting process is referred to as *differential quantization*. The motivation behind the use of differential quantization follows from two practical considerations:

1. Waveform encoding in the transmitter requires the use of quantization.
2. Waveform decoding in the receiver, therefore, has to process a quantized signal.

In order to cater to both requirements in such a way that the *same structure* is used for predictors in both the transmitter and the receiver, the transmitter has to perform prediction-error filtering on the quantized version of the message signal rather than the signal itself, as shown in Figure 6.19a. Then, assuming a noise-free channel, the predictors in the transmitter and receiver operate on exactly the same sequence of quantized message samples.

To demonstrate this highly desirable and distinctive characteristic of differential PCM, we see from Figure 6.19a that

$$e_{q,n} = e_n + q_n \quad (6.87)$$

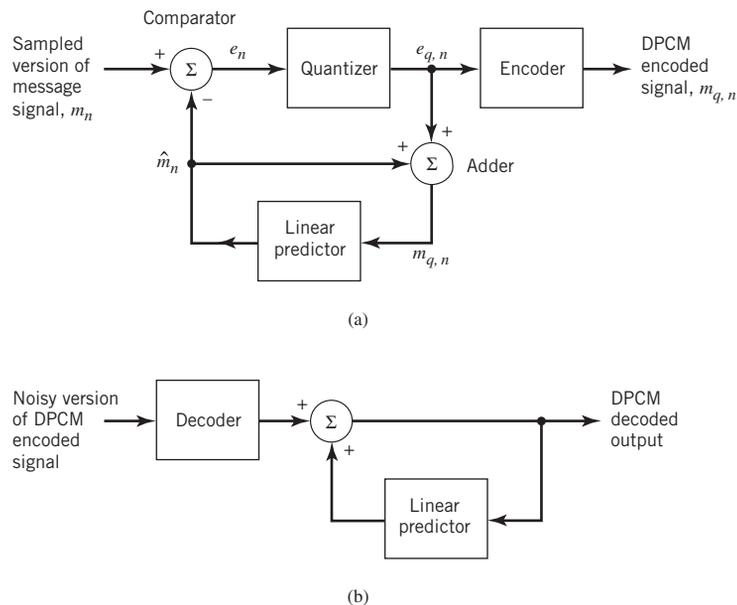


Figure 6.19 DPCM system: (a) transmitter; (b) receiver.

where q_n is the quantization noise produced by the quantizer operating on the prediction error e_n . Moreover, from Figure 6.19a, we readily see that

$$m_{q,n} = \hat{m}_n + e_{q,n} \quad (6.88)$$

where \hat{m}_n is the predicted value of the original message sample m_n ; thus, (6.88) is in perfect agreement with Figure 6.18. Hence, the use of (6.87) in (6.88) yields

$$m_{q,n} = \hat{m}_n + e_n + q_n \quad (6.89)$$

We may now invoke (6.88) of linear prediction theory to rewrite (6.89) in the equivalent form:

$$m_{q,n} = m_n + q_n \quad (6.90)$$

which describes a quantized version of the original message sample m_n .

With the differential quantization scheme of Figure 6.19a at hand, we may now expand on the structures of the transmitter and receiver of DPCM.

DPCM Transmitter

Operation of the DPCM transmitter proceeds as follows:

1. Given the predicted message sample \hat{m}_n , the comparator at the transmitter input computes the prediction error e_n , which is quantized to produce the quantized version of e_n in accordance with (6.87).
2. With \hat{m}_n and $e_{q,n}$ at hand, the adder in the transmitter produces the quantized version of the original message sample m_n , namely $m_{q,n}$, in accordance with (6.88).
3. The required one-step prediction \hat{m}_n is produced by applying the sequence of quantized samples $\{m_{q,k}\}_{k=1}^p$ to a linear FIR predictor of order p .

This multistage operation is clearly *cyclic*, encompassing three steps that are repeated at each time step n . Moreover, at each time step, the encoder operates on the quantized prediction error $e_{q,n}$ to produce the DPCM-encoded version of the original message sample m_n . The DPCM code so produced is a *lossy-compressed* version of the PCM code; it is “lossy” because of the prediction error.

DPCM Receiver

The structure of the receiver is much simpler than that of the transmitter, as depicted in Figure 6.19b. Specifically, first, the decoder reconstructs the quantized version of the prediction error, namely $e_{q,n}$. An estimate of the original message sample m_n is then computed by applying the decoder output to the same predictor used in the transmitter of Figure 6.19a. In the absence of channel noise, the encoded signal at the receiver input is identical to the encoded signal at the transmitter output. Under this ideal condition, we find that the corresponding receiver output is equal to $m_{q,n}$, which differs from the original signal sample m_n only by the quantization error q_n incurred as a result of quantizing the prediction error e_n .

From the foregoing analysis, we thus observe that, in a noise-free environment, the linear predictors in the transmitter and receiver of DPCM operate on the same sequence of samples, $m_{q,n}$. It is with this point in mind that a feedback path is appended to the quantizer in the transmitter of Figure 6.19a.

Processing Gain

The output SNR of the DPCM system, shown in Figure 6.19, is, by definition,

$$(\text{SNR})_O = \frac{\sigma_M^2}{\sigma_Q^2} \quad (6.91)$$

where σ_M^2 is the variance of the original signal sample m_n , assumed to be of zero mean, and σ_Q^2 is the variance of the quantization error q_n , also of zero mean. We may rewrite (6.91) as the product of two factors, as shown by

$$\begin{aligned} (\text{SNR})_O &= \left(\frac{\sigma_M^2}{\sigma_E^2} \right) \left(\frac{\sigma_E^2}{\sigma_Q^2} \right) \\ &= G_p (\text{SNR})_Q \end{aligned} \quad (6.92)$$

where, in the first line, σ_E^2 is the variance of the prediction error e_n . The factor $(\text{SNR})_Q$ introduced in the second line is the *signal-to-quantization noise ratio*, which is itself defined by

$$(\text{SNR})_Q = \frac{\sigma_E^2}{\sigma_Q^2} \quad (6.93)$$

The other factor G_p is the *processing gain* produced by the differential quantization scheme; it is formally defined by

$$G_p = \frac{\sigma_M^2}{\sigma_E^2} \quad (6.94)$$

The quantity G_p , when it is greater than unity, represents a *gain in signal-to-noise ratio*, which is due to the differential quantization scheme of Figure 6.19. Now, for a given message signal, the variance σ_M^2 is fixed, so that G_p is maximized by minimizing the variance σ_M^2 of the prediction error e_n . Accordingly, the objective in implementing the DPCM should be to design the prediction filter so as to minimize the prediction-error variance, σ_E^2 .

In the case of voice signals, it is found that the optimum signal-to-quantization noise advantage of the DPCM over the standard PCM is in the neighborhood of 4–11 dB. Based on experimental studies, it appears that the greatest improvement occurs in going from no prediction to first-order prediction, with some additional gain resulting from increasing the order p of the prediction filter up to 4 or 5, after which little additional gain is obtained. Since 6 dB of quantization noise is equivalent to 1 bit per sample by virtue of the results presented in Table 6.1 for sinusoidal modulation, the advantage of DPCM may also be expressed in terms of bit rate. For a constant signal-to-quantization noise ratio, and assuming a sampling rate of 8 kHz, the use of DPCM may provide a saving of about 8–16 kHz (i.e., 1 to 2 bits per sample) compared with the standard PCM.

6.9 Delta Modulation

In choosing DPCM for waveform coding, we are, in effect, economizing on transmission bandwidth by increasing system complexity, compared with standard PCM. In other words, DPCM exploits the *complexity–bandwidth tradeoff*. However, in practice, the need may arise for reduced system complexity compared with the standard PCM. To achieve this other objective, transmission bandwidth is traded off for reduced system complexity, which is precisely the motivation behind DM. Thus, whereas DPCM exploits the *complexity–bandwidth tradeoff*, DM exploits the *bandwidth–complexity tradeoff*. We may, therefore, differentiate between the standard PCM, the DPCM, and the DM along the lines described in Figure 6.20. With the bandwidth–complexity tradeoff being at the heart of DM, the incoming message signal $m(t)$ is *oversampled*, which requires the use of a sampling rate higher than the Nyquist rate. Accordingly, the correlation between adjacent samples of the message signal is purposely increased so as to permit the use of a *simple* quantizing strategy for constructing the encoded signal.

DM Transmitter

In the DM transmitter, system complexity is reduced to the minimum possible by using the combination of two strategies:

1. *Single-bit quantizer*, which is the simplest quantizing strategy; as depicted in Figure 6.21, the quantizer acts as a hard limiter with only two decision levels, namely, $\pm\Delta$.
2. *Single unit-delay element*, which is the most primitive form of a predictor; in other words, the only component retained in the FIR predictor of Figure 6.17 is the front-end block labeled z^{-1} , which acts as an *accumulator*.

Thus, replacing the multilevel quantizer and the FIR predictor in the DPCM transmitter of Figure 6.19a in the manner described under points 1 and 2, respectively, we obtain the block diagram of Figure 6.21a for the DM transmitter.

From this figure, we may express the equations underlying the operation of the DM transmitter by the following set of equations (6.95)–(6.97):

$$\begin{aligned} e_n &= m_n - \hat{m}_n \\ &= m_n - m_{q, n-1} \end{aligned} \tag{6.95}$$

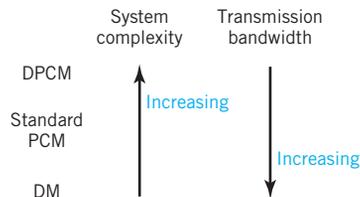


Figure 6.20 Illustrating the tradeoffs between standard PCM, DPCM, and DM.

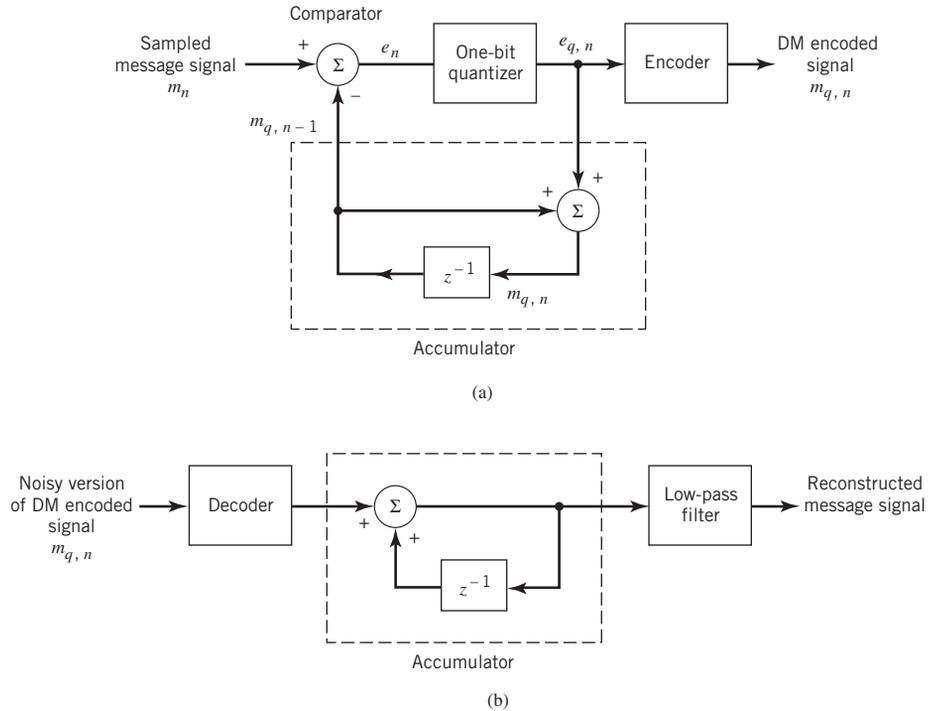


Figure 6.21 DM system: (a) transmitter; (b) receiver.

$$e_{q,n} = \Delta \operatorname{sgn}[e_n]$$

$$= \begin{cases} +\Delta & \text{if } e_n > 0 \\ -\Delta & \text{if } e_n < 0 \end{cases} \quad (6.96)$$

$$m_{q,n} = m_{q,n-1} + e_{q,n} \quad (6.97)$$

According to (6.95) and (6.96), two possibilities may naturally occur:

1. The error signal e_n (i.e., the difference between the message sample m_n and its approximation \hat{m}_n) is positive, in which case the approximation $\hat{m}_n = m_{q,n-1}$ is increased by the amount Δ ; in this first case, the encoder sends out symbol 1.
2. The error signal e_n is negative, in which case the approximation $\hat{m}_n = m_{q,n-1}$ is reduced by the amount Δ ; in this second case, the encoder sends out symbol 0.

From this description it is apparent that the delta modulator produces a staircase approximation to the message signal, as illustrated in Figure 6.22a. Moreover, the rate of data transmission in DM is equal to the sampling rate $f_s = 1/T_s$, as illustrated in the binary sequence of Figure 6.22b.

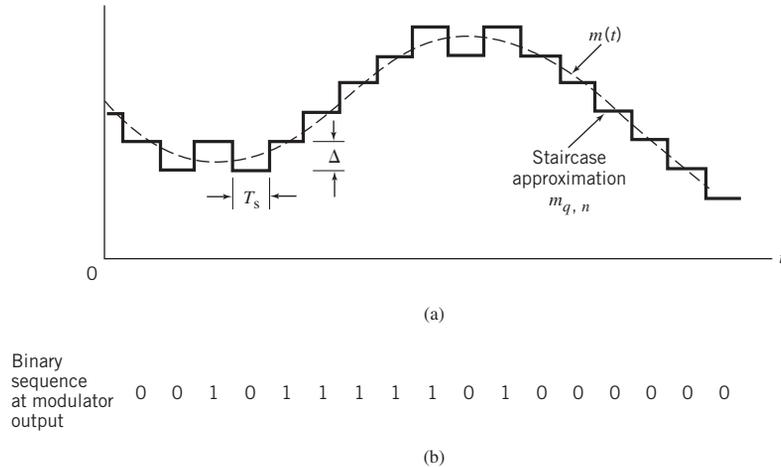


Figure 6.22 Illustration of DM.

DM Receiver

Following a procedure similar to the way in which we constructed the DM transmitter of Figure 6.21a, we may construct the DM receiver of Figure 6.21b as a special case of the DPCM receiver of Figure 6.19b. Working through the operation of the DM receiver, we find that reconstruction of the staircase approximation to the original message signal is achieved by passing the sequence of positive and negative pulses (representing symbols 1 and 0, respectively) through the block labeled “accumulator.”

Under the assumption that the channel is distortionless, the accumulated output is the desired $m_{q,n}$ given that the decoded channel output is $e_{q,n}$. The out-of-band quantization noise in the high-frequency staircase waveform in the accumulator output is suppressed by passing it through a low-pass filter with a cutoff frequency equal to the message bandwidth.

Quantization Errors in DM

DM is subject to two types of quantization error: slope overload distortion and granular noise. We will discuss the case of slope overload distortion first.

Starting with (6.97), we observe that this equation is the *digital equivalent of integration*, in the sense that it represents the *accumulation* of positive and negative increments of magnitude Δ . Moreover, denoting the quantization error applied to the message sample m_n by q_n , we may express the quantized message sample as

$$m_{q,n} = m_n + q_n \quad (6.98)$$

With this expression for $m_{q,n}$ at hand, we find from (6.98) that the quantizer input is

$$e_n = m_n - (m_{n-1} + q_{n-1}) \quad (6.99)$$

Thus, except for the delayed quantization error q_{n-1} , the quantizer input is a *first backward difference* of the original message sample. This difference may be viewed as a

digital approximation to the quantizer input or, equivalently, as the *inverse* of the digital integration process carried out in the DM transmitter. If, then, we consider the maximum slope of the original message signal $m(t)$, it is clear that in order for the sequence of samples $\{m_{q,n}\}$ to increase as fast as the sequence of message samples $\{m_n\}$ in a region of maximum slope of $m(t)$, we require that the condition

$$\frac{\Delta}{T_s} \geq \max \left| \frac{dm(t)}{dt} \right| \quad (6.100)$$

be satisfied. Otherwise, we find that the step-size Δ is too small for the staircase approximation $m_q(t)$ to follow a steep segment of the message signal $m(t)$, with the result that $m_q(t)$ falls behind $m(t)$, as illustrated in Figure 6.23. This condition is called *slope overload*, and the resulting quantization error is called *slope-overload distortion (noise)*. Note that since the maximum slope of the staircase approximation $m_q(t)$ is fixed by the step size Δ , increases and decreases in $m_q(t)$ tend to occur along straight lines. For this reason, a delta modulator using a fixed step size is often referred to as a *linear delta modulator*.

In contrast to slope-overload distortion, *granular noise* occurs when the step size Δ is too large relative to the local slope characteristics of the message signal $m(t)$, thereby causing the staircase approximation $m_q(t)$ to hunt around a relatively flat segment of $m(t)$; this phenomenon is also illustrated in the tail end of Figure 6.23. Granular noise is analogous to quantization noise in a PCM system.

Adaptive DM

From the discussion just presented, it is appropriate that we need to have a large step size to accommodate a wide dynamic range, whereas a small step size is required for the accurate representation of relatively low-level signals. It is clear, therefore, that the choice of the optimum step size that minimizes the mean-square value of the quantization error in a linear delta modulator will be the result of a compromise between slope-overload distortion and granular noise. To satisfy such a requirement, we need to make the delta modulator “adaptive,” in the sense that the step size is made to vary in accordance with the input signal. The step size is thereby made variable, such that it is enlarged during intervals when the slope-overload distortion is dominant and reduced in value when the granular (quantization) noise is dominant.

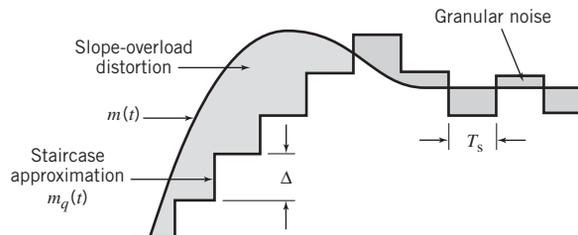


Figure 6.23 Illustration of the two different forms of quantization error in DM.

6.10 Line Codes

In this chapter, we have described three basic waveform-coding schemes: PCM, DPCM, and DM. Naturally, they differ from each other in several ways: transmission–bandwidth requirement, transmitter–receiver structural composition and complexity, and quantization noise. Nevertheless, all three of them have a common need: *line codes* for electrical representation of the encoded binary streams produced by their individual transmitters, so as to facilitate transmission of the binary streams across the communication channel.

Figure 6.24 displays the waveforms of five important line codes for the example data stream 01101001. Figure 6.25 displays their individual power spectra (for positive frequencies) for randomly generated binary data, assuming that first, symbols 0 and 1 are equiprobable, second, the average power is normalized to unity, and third, the frequency f is normalized with respect to the bit rate $1/T_b$. In what follows, we describe the five line codes involved in generating the coded waveforms of Figure 6.24.

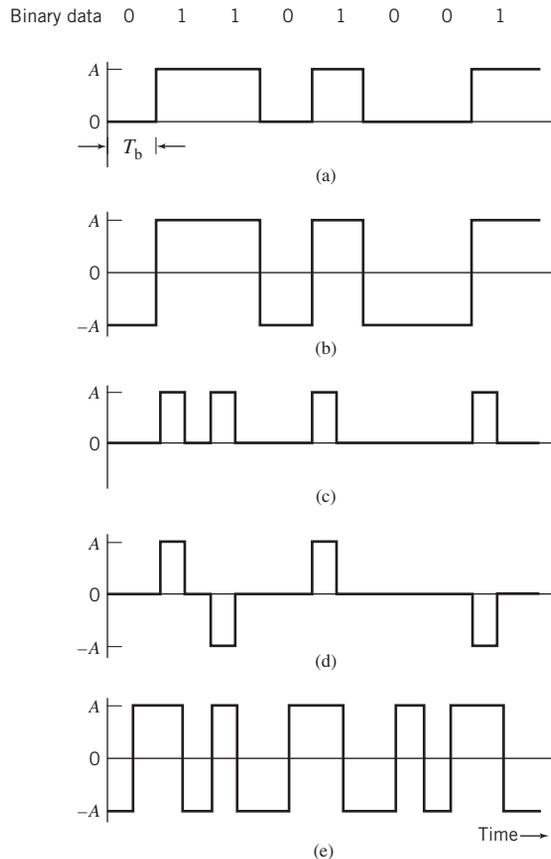


Figure 6.24 Line codes for the electrical representations of binary data: (a) unipolar nonreturn-to-zero (NRZ) signaling; (b) polar NRZ signaling; (c) unipolar return-to-zero (RZ) signaling; (d) bipolar RZ signaling; (e) split-phase or Manchester code.

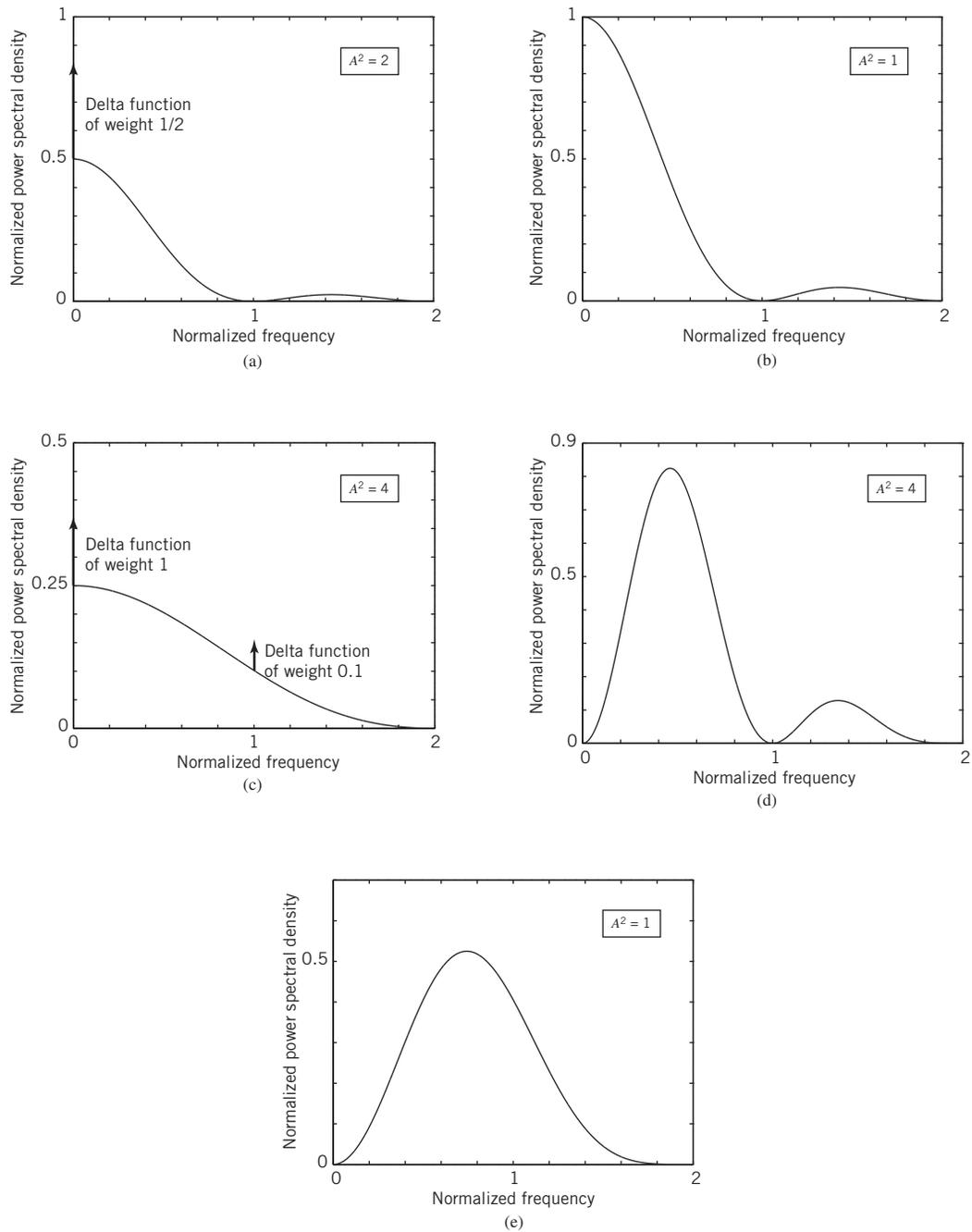


Figure 6.25 Power spectra of line codes: (a) unipolar NRZ signal; (b) polar NRZ signal; (c) unipolar RZ signal; (d) bipolar RZ signal; (e) Manchester-encoded signal. The frequency is normalized with respect to the bit rate $1/T_b$, and the average power is normalized to unity.

Unipolar NRZ Signaling

In this line code, symbol 1 is represented by transmitting a pulse of amplitude A for the duration of the symbol, and symbol 0 is represented by switching off the pulse, as in Figure 6.24a. The unipolar NRZ line code is also referred to as *on-off signaling*. Disadvantages of on-off signaling are the waste of power due to the transmitted DC level and the fact that the power spectrum of the transmitted signal does not approach zero at zero frequency.

Polar NRZ Signaling

In this second line code, symbols 1 and 0 are represented by transmitting pulses of amplitudes $+A$ and $-A$, respectively, as illustrated in Figure 6.24b. The polar NRZ line code is relatively easy to generate, but its disadvantage is that the power spectrum of the signal is large near zero frequency.

Unipolar RZ Signaling

In this third line code, symbol 1 is represented by a rectangular pulse of amplitude A and half-symbol width and symbol 0 is represented by transmitting *no* pulse, as illustrated in Figure 6.24c. An attractive feature of the unipolar RZ line code is the presence of delta functions at $f = 0, \pm 1/T_b$ in the power spectrum of the transmitted signal; the delta functions can be used for *bit-timing recovery* at the receiver. However, its disadvantage is that it requires 3 dB more power than polar RZ signaling for the same probability of symbol error.

Bipolar RZ Signaling

This line code uses three amplitude levels, as indicated in Figure 6.24(d). Specifically, positive and negative pulses of equal amplitude (i.e., $+A$ and $-A$) are used alternately for symbol 1, with each pulse having a half-symbol width; no pulse is always used for symbol 0. A useful property of the bipolar RZ signaling is that the power spectrum of the transmitted signal has no DC component and relatively insignificant low-frequency components for the case when symbols 1 and 0 occur with equal probability. The bipolar RZ line code is also called *alternate mark inversion* (AMI) signaling.

Split-Phase (Manchester Code)

In this final method of signaling, illustrated in Figure 6.24e, symbol 1 is represented by a positive pulse of amplitude A followed by a negative pulse of amplitude $-A$, with both pulses being half-symbol wide. For symbol 0, the polarities of these two pulses are reversed. A unique property of the Manchester code is that it suppresses the DC component and has relatively insignificant low-frequency components, *regardless of the signal statistics*. This property is essential in some applications.

6.11 Summary and Discussion

In this chapter we introduced two fundamental and complementary processes:

- *Sampling*, which operates in the time domain; the sampling process is the link between an analog waveform and its discrete-time representation.
- *Quantization*, which operates in the amplitude domain; the quantization process is the link between an analog waveform and its discrete-amplitude representation.

The sampling process builds on the *sampling theorem*, which states that a strictly band-limited signal with no frequency components higher than W Hz is represented uniquely by a sequence of samples taken at a uniform rate equal to or greater than the Nyquist rate of $2W$ samples per second. The quantization process exploits the fact that any human sense, as ultimate receiver, can only detect finite intensity differences.

The sampling process is basic to the operation of all pulse modulation systems, which may be classified into analog pulse modulation and digital pulse modulation. The distinguishing feature between them is that analog pulse modulation systems maintain a continuous amplitude representation of the message signal, whereas digital pulse modulation systems also employ quantization to provide a representation of the message signal that is discrete in both time and amplitude.

Analog pulse modulation results from varying some parameter of the transmitted pulses, such as amplitude, duration, or position, in which case we speak of PAM, pulse-duration modulation, or pulse-position modulation, respectively. In this chapter we focused on PAM, as it is used in all forms of digital pulse modulation.

Digital pulse modulation systems transmit analog message signals as a sequence of coded pulses, which is made possible through the combined use of sampling and quantization. PCM is an important form of digital pulse modulation that is endowed with some unique system advantages, which, in turn, have made it the standard method of modulation for the transmission of such analog signals as voice and video signals. The advantages of PCM include robustness to noise and interference, efficient regeneration of the coded pulses along the transmission path, and a uniform format for different kinds of baseband signals.

Indeed, it is because of this list of advantages unique to PCM that it has become the method of choice for the construction of public switched telephone networks (PSTNs). In this context, the reader should carefully note that the telephone channel viewed from the PSTN by an Internet service provider, for example, is *nonlinear* due to the use of companding and, most importantly, it is *entirely digital*. This observation has a significant impact on the design of high-speed modems for communications between a computer user and server, which will be discussed in Chapter 8.

DM and DPCM are two other useful forms of digital pulse modulation. The principal advantage of DM is the simplicity of its circuitry, which is achieved at the expense of increased transmission bandwidth. In contrast, DPCM employs increased circuit complexity to reduce channel bandwidth. The improvement is achieved by using the idea of prediction to reduce redundant symbols from an incoming data stream. A further improvement in the operation of DPCM can be made through the use of adaptivity to account for statistical variations in the input data. By so doing, bandwidth requirement may be reduced significantly without serious degradation in system performance.⁸

Problems

Sampling Process

- 6.1 In natural sampling, an analog signal $g(t)$ is multiplied by a periodic train of rectangular pulses $c(t)$, each of unit area. Given that the pulse repetition frequency of this periodic train is f_s and the duration of each rectangular pulse is T (with $f_s T \ll 1$), do the following:
- Find the spectrum of the signal $s(t)$ that results from the use of natural sampling; you may assume that time $t = 0$ corresponds to the midpoint of a rectangular pulse in $c(t)$.
 - Show that the original signal $g(t)$ may be recovered exactly from its naturally sampled version, provided that the conditions embodied in the sampling theorem are satisfied.
- 6.2 Specify the Nyquist rate and the Nyquist interval for each of the following signals:
- $g(t) = \text{sinc}(200t)$.
 - $g(t) = \text{sinc}^2(200t)$.
 - $g(t) = \text{sinc}(200t) + \text{sinc}^2(200t)$.
- 6.3 Discussion of the sampling theorem presented in Section 6.2 was confined to the time domain. Describe how the sampling theorem can be applied in the frequency domain.

Pulse-Amplitude Modulation

- 6.4 Figure P6.4 shows the idealized spectrum of a message signal $m(t)$. The signal is sampled at a rate equal to 1 kHz using flat-top pulses, with each pulse being of unit amplitude and duration 0.1 ms. Determine and sketch the spectrum of the resulting PAM signal.

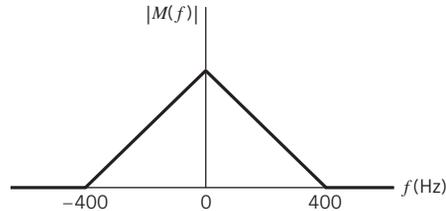


Figure P6.4

- 6.5 In this problem, we evaluate the equalization needed for the aperture effect in a PAM system. The operating frequency $f = f_s/2$, which corresponds to the highest frequency component of the message signal for a sampling rate equal to the Nyquist rate. Plot $1/\text{sinc}(0.5T/T_s)$ versus T/T_s , and hence find the equalization needed when $T/T_s = 0.1$.
- 6.6 Consider a PAM wave transmitted through a channel with white Gaussian noise and minimum bandwidth $B_T = 1/2T_s$, where T_s is the sampling period. The noise is of zero mean and power spectral density $N_0/2$. The PAM signal uses a standard pulse $g(t)$ with its Fourier transform defined by

$$G(f) = \begin{cases} \frac{1}{2B_T}, & |f| < B_T \\ 0, & |f| > B_T \end{cases}$$

By considering a full-load sinusoidal modulating wave, show that PAM and baseband-signal transmission have equal SNRs for the same average transmitted power.

- 6.7 Twenty-four voice signals are sampled uniformly and then time-division multiplexed (TDM). The sampling operation uses flat-top samples with $1 \mu\text{s}$ duration. The multiplexing operation includes

provision for synchronization by adding an extra pulse of sufficient amplitude and also 1 μ s duration. The highest frequency component of each voice signal is 3.4 kHz.

- a. Assuming a sampling rate of 8 kHz, calculate the spacing between successive pulses of the multiplexed signal.
 - b. Repeat your calculation assuming the use of Nyquist rate sampling.
- 6.8 Twelve different message signals, each with a bandwidth of 10 kHz, are to be multiplexed and transmitted. Determine the minimum bandwidth required if the multiplexing/modulation method used is time-division multiplexing (TDM), which was discussed in Chapter 1.

Pulse-Code Modulation

- 6.9 A speech signal has a total duration of 10 s. It is sampled at the rate of 8 kHz and then encoded. The signal-to-(quantization) noise ratio is required to be 40 dB. Calculate the minimum storage capacity needed to accommodate this digitized speech signal.
- 6.10 Consider a uniform quantizer characterized by the input-output relation illustrated in Figure 6.9a. Assume that a Gaussian-distributed random variable with zero mean and unit variance is applied to this quantizer input.
- a. What is the probability that the amplitude of the input lies outside the range -4 to $+4$?
 - b. Using the result of part a, show that the output SNR of the quantizer is given by

$$(\text{SNR})_{\text{O}} = 6R - 7.2 \text{ dB}$$

where R is the number of bits per sample. Specifically, you may assume that the quantizer input extends from -4 to $+4$. Compare the result of part b with that obtained in Example 2.

- 6.11 A PCM system uses a uniform quantizer followed by a 7-bit binary encoder. The bit rate of the system is equal to 50×10^6 bits/s.
- a. What is the maximum message bandwidth for which the system operates satisfactorily?
 - b. Determine the output signal-to-(quantization) noise when a full-load sinusoidal modulating wave of frequency 1 MHz is applied to the input.
- 6.12 Show that with a nonuniform quantizer the mean-square value of the quantization error is approximately equal to $(1/12)\sum_i \Delta_i^2 p_i$, where Δ_i is the i th step size and p_i is the probability that the input signal amplitude lies within the i th interval. Assume that the step size Δ_i is small compared with the excursion of the input signal.
- 6.13
- a. A sinusoidal signal with an amplitude of 3.25 V is applied to a uniform quantizer of the midtread type whose output takes on the values 0, ± 1 , ± 2 , ± 3 V. Sketch the waveform of the resulting quantizer output for one complete cycle of the input.
 - b. Repeat this evaluation for the case when the quantizer is of the midrise type whose output takes on the values 0.5, ± 1.5 , ± 2.5 , ± 3.5 V.

- 6.14 The signal

$$m(t) \text{ (volts)} = 6 \sin(2\pi t)$$

is transmitted using a 40-bit binary PCM system. The quantizer is of the midrise type, with a step size of 1 V. Sketch the resulting PCM wave for one complete cycle of the input. Assume a sampling rate of four samples per second, with samples taken at $t(\text{s}) = \pm 1/8, \pm 3/8, \pm 5/8, \dots$

- 6.15 Figure P6.15 shows a PCM signal in which the amplitude levels of $+1$ V and -1 V are used to represent binary symbols 1 and 0, respectively. The codeword used consists of three bits. Find the sampled version of an analog signal from which this PCM signal is derived.

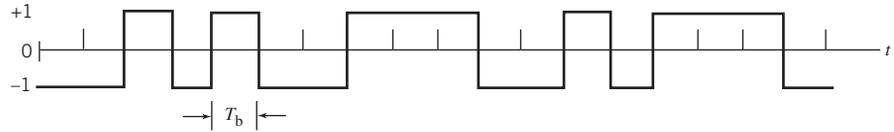


Figure P6.15

- 6.16 Consider a chain of $(n - 1)$ regenerative repeaters, with a total of n sequential decisions made on a binary PCM wave, including the final decision made at the receiver. Assume that any binary symbol transmitted through the system has an independent probability p_1 of being inverted by any repeater. Let p_n represent the probability that a binary symbol is in error after transmission through the complete system.

a. Show that

$$p_n = \frac{1}{2}[1 - (1 - 2p_1)^n]$$

b. If p_1 is very small and n is not too large, what is the corresponding value of p_n ?

- 6.17 Discuss the basic issues involved in the design of a regenerative repeater for PCM.

Linear Prediction

- 6.18 A one-step linear predictor operates on the sampled version of a sinusoidal signal. The sampling rate is equal to $10f_0$, where f_0 is the frequency of the sinusoid. The predictor has a single coefficient denoted by w_1 .

- a. Determine the optimum value of w_1 required to minimize the prediction-error variance.
 b. Determine the minimum value of the prediction error variance.

- 6.19 A stationary process $X(t)$ has the following values for its autocorrelation function:

$$R_X(0) = 1$$

$$R_X(1) = 0.8$$

$$R_X(2) = 0.6$$

$$R_X(3) = 0.4$$

- a. Calculate the coefficients of an optimum linear predictor involving the use of three unit-time delays.
 b. Calculate the variance of the resulting prediction error.
- 6.20 Repeat the calculations of Problem 6.19, but this time use a linear predictor with two unit-time delays. Compare the performance of this second optimum linear predictor with that considered in Problem 6.19.

Differential Pulse-Code Modulation

- 6.21 A DPCM system uses a linear predictor with a single tap. The normalized autocorrelation function of the input signal for a lag of one sampling interval is 0.75. The predictor is designed to minimize the prediction-error variance. Determine the processing gain attained by the use of this predictor.
- 6.22 Calculate the improvement in processing gain of a DPCM system using the optimized three-tap linear predictor. For this calculation, use the autocorrelation function values of the input signal specified in Problem 6.19.
- 6.23 In this problem, we compare the performance of a DPCM system with that of an ordinary PCM system using companding.

For a sufficiently large number of representation levels, the signal-to-(quantization) noise ratio of PCM systems, in general, is defined by

$$10 \log_{10}(\text{SNR})_O \text{ (dB)} = \alpha + 6n$$

where 2^n is the number of representation levels. For a companded PCM system using the μ -law, the constant α is itself defined by

$$\alpha \text{ (dB)} \approx 4.77 - 20 \log_{10} \log(1 + \mu)$$

For a DPCM system, on the other hand, the constant α lies in the range $-3 < \alpha < 15$ dBs. The formulas quoted herein apply to telephone-quality speech signals.

Compare the performance of the DPCM system against that of the μ -companded PCM system with $\mu = 255$ for each of the following scenarios:

- The improvement in $(\text{SNR})_O$ realized by DPCM over companded PCM for the same number of bits per sample.
- The reduction in the number of bits per sample required by DPCM, compared with the companded PCM for the same $(\text{SNR})_O$.

6.24 In the DPCM system depicted in Figure P6.24, show that in the absence of channel noise, the transmitting and receiving prediction filters operate on slightly different input signals.

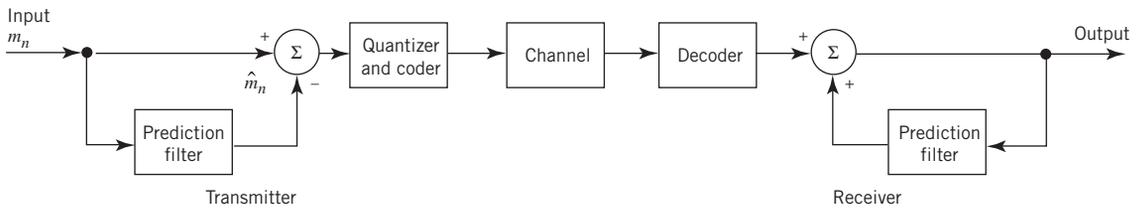


Figure P6.24

6.25 Figure P6.25 depicts the block diagram of adaptive quantization for DPCM. The quantization is of a backward estimation kind because samples of the quantization output and prediction errors are used to continuously derive backward estimates of the variance of the message signal. This estimate computed at time n is denoted by $\hat{\sigma}_{m,n}^2$. Given this estimate, the step size is varied so as to match the actual variance of the message sample m_n , as shown by

$$\Delta_n = \phi \hat{\sigma}_{m,n}$$

where $\hat{\sigma}_{m,n}^2$ is the estimate of the standard deviation and ϕ is a constant. An attractive feature of the adaptive scheme in Figure P6.25 is that samples of the quantization output and the prediction error are used to compute the predictor's coefficients.

Modify the block diagram of the DPCM transmitter in Figure 6.19a so as to accommodate *adaptive prediction with backward estimation*.

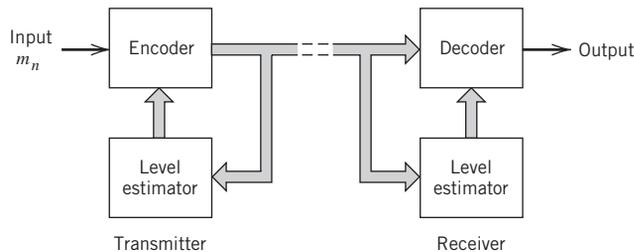


Figure P6.25

Delta Modulation

- 6.26 Consider a test signal $m(t)$ defined by a hyperbolic tangent function:

$$m(t) = A \tanh(\beta t)$$

where A and β are constants. Determine the minimum step size Δ for DM of this signal, which is required to avoid slope-overload distortion.

- 6.27 Consider a sine wave of frequency f_m and amplitude A_m , which is applied to a delta modulator of step size Δ . Show that slope-overload distortion will occur if

$$A_m > \frac{\Delta}{2\pi f_m T_s}$$

where T_s is the sampling period. What is the maximum power that may be transmitted without slope-overload distortion?

- 6.28 A linear delta modulator is designed to operate on speech signals limited to 3.4 kHz. The specifications of the modulator are as follows:

- Sampling rate = $10f_{\text{Nyquist}}$, where f_{Nyquist} is the Nyquist rate of the speech signal.
- Step size $\Delta = 100$ mV.

The modulator is tested with a 1 kHz sinusoidal signal. Determine the maximum amplitude of this test signal required to avoid slope-overload distortion.

- 6.29 In this problem, we derive an empirical formula for the average signal-to-(quantization) noise ratio of a DM system with a sinusoidal signal of amplitude A and frequency f_m as the test signal. Assume that the power spectral density of the granular noise generated by the system is governed by the formula

$$S_N(f) = \frac{\Delta^2}{6f_s}$$

where f_s is the sampling rate and Δ is the step size. (Note that this formula is basically the same as that for the power spectral density of quantization noise in a PCM system with $\Delta/2$ for PCM being replaced by Δ for DM.) The DM system is designed to handle analog message signals limited to bandwidth W .

- a. Show that the average quantization noise power produced by the system is

$$N = \frac{4\pi^2 A^2 f_m^2 W}{3f_s^3}$$

where it is assumed that the step size Δ has been chosen in accordance with the formula used in Problem 6.28 so as to avoid slope-overload distortion.

- b. Hence, determine the signal-to-(quantization) noise ratio of the DM system for a sinusoidal input.
- 6.30 Consider a DM system designed to accommodate analog message signals limited to bandwidth $W = 5$ kHz. A sinusoidal test signal of amplitude $A = 1$ V and frequency $f_m = 1$ kHz is applied to the system. The sampling rate of the system is 50 kHz.

- a. Calculate the step size Δ required to minimize slope overload distortion.
- b. Calculate the signal-to-(quantization) noise ratio of the system for the specified sinusoidal test signal.

For these calculations, use the formula derived in Problem 6.29.

- 6.31 Consider a low-pass signal with a bandwidth of 3 kHz. A linear DM system with step size $\Delta = 0.1$ V is used to process this signal at a sampling rate 10 times the Nyquist rate.

- a. Evaluate the maximum amplitude of a test sinusoidal signal of frequency 1 kHz, which can be processed by the system without slope-overload distortion.

- b. For the specifications given in part a, evaluate the output SNR under (i) prefiltered and (ii) postfiltered conditions.

6.32 In the conventional form of DM, the quantizer input may be viewed as an approximate to the *derivative* of the incoming message signal $m(t)$. This behavior leads to a drawback of DM: transmission disturbances (e.g., noise) result in an accumulation error in the demodulated signal. This drawback can be overcome by *integrating* the message signal $m(t)$ prior to DM, resulting in three beneficial effects:

- Low frequency content of $m(t)$ is pre-emphasized.
- Correlation between adjacent samples of $m(t)$ is increased, tending to improve overall system performance by reducing the variance of the error signal at the quantizer input.
- Design of the receiver is simplified.

Such a DM scheme is called *delta-sigma modulation*.

Construct a block diagram of the delta-sigma modulation system in such a way that it provides an interpretation of the system as a “smoothed” version of 1-bit PCM in the following composite sense:

- smoothness implies that the comparator output is integrated prior to quantization, and
- 1-bit modulation merely restates that the quantizer consists of a hard limiter with only two representation levels.

Explain how the receiver of the delta-sigma modulation system is simplified, compared with conventional DM.

Line Codes

6.33 In this problem, we derive the formulas used to compute the power spectra of Figure 6.25 for the five line codes described in Section 6.10. In the case of each line code, the bit duration is T_b and the pulse amplitude A is conditioned to normalize the average power of the line code to unity as indicated in Figure 6.25. Assume that the data stream is randomly generated and symbols 0 and 1 are equally likely.

Derive the power spectral densities of these line codes as summarized here:

- a. Unipolar NRZ signals:

$$S(f) = \frac{A^2 T_b}{4} \text{sinc}^2(fT_b) \left(1 + \frac{1}{T_b} \delta(f)\right)$$

- b. Polar NRZ signals:

$$S(f) = A^2 T_b \text{sinc}^2(fT_b)$$

- c. Unipolar RZ signals:

$$S(f) = \frac{A^2 T_b}{16} \text{sinc}^2\left(\frac{fT_b}{2}\right) \left[1 + \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right)\right]$$

- d. Bipolar RZ signals:

$$S(f) = \frac{A^2 T_b}{4} \text{sinc}^2\left(\frac{fT_b}{2}\right) \sin^2(\pi f T_b)$$

- e. Manchester-encoded signals:

$$S(f) = \frac{A^2 T_b}{4} \text{sinc}^2\left(\frac{fT_b}{2}\right) \sin^2\left(\frac{\pi f T_b}{2}\right)$$

Hence, confirm the spectral plots displayed in Figure 6.25.

- 6.34 A randomly generated data stream consists of equiprobable binary symbols 0 and 1. It is encoded into a polar NRZ waveform with each binary symbol being defined as follows:

$$s(t) = \begin{cases} \cos\left(\frac{\pi t}{T_b}\right), & -\frac{T_b}{2} < t \leq \frac{T_b}{2} \\ 0, & \text{otherwise} \end{cases}$$

- Sketch the waveform so generated, assuming that the data stream is 00101110.
 - Derive an expression for the power spectral density of this signal and sketch it.
 - Compare the power spectral density of this random waveform with that defined in part b of Problem 6.33.
- 6.35 Given the data stream 1110010100, sketch the transmitted sequence of pulses for each of the following line codes:
- unipolar NRZ
 - polar NRZ
 - unipolar RZ
 - bipolar RZ
 - Manchester code.

Computer Experiments

- **6.36 A sinusoidal signal of frequency $f_0 = 10^4/2\pi$ Hz is sampled at the rate of 8 kHz and then applied to a sample-and-hold circuit to produce a flat-topped PAM signal $s(t)$ with pulse duration $T = 500 \mu\text{s}$.
- Compute the waveform of the PAM signal $s(t)$.
 - Compute $|S(f)|$, denoting the magnitude spectrum of the PAM signal $s(t)$.
 - Compute the envelope of $|S(f)|$. Hence confirm that the frequency at which this envelope goes through zero for the first time is equal to $(1/T) = 20$ kHz.
- **6.37 In this problem, we use computer simulation to compare the performance of a companded PCM system using the μ -law against that of the corresponding system using a uniform quantizer. The simulation is to be performed for a sinusoidal input signal of varying amplitude. With a companded PCM system in mind, Table 6.4 describes the 15-segment *pseudo-linear* characteristic that consists of 15 linear segments configured to approximate the logarithmic μ -law

Table 6.4 The 15-segment companding characteristic ($\mu = 255$)

Linear segment number	Step-size	Projections of segment end points onto the horizontal axis
0	2	± 31
1a, 1b	4	± 95
2a, 2b	8	± 223
3a, 3b	16	± 479
4a, 4b	32	± 991
5a, 5b	64	± 2015
6a, 6b	128	± 4063
7a, 7b	256	± 8159

of (6.48), with $\mu = 255$. This approximation is constructed in such a way that the segment endpoints in Table 6.4 lie on the compression curve computed from (6.48).

- Using the μ -law described in Table 6.4, plot the output signal-to-noise ratio as a function of the input signal-to-noise ratio, both ratios being expressed in decibels.
- Compare the results of your computation in part (a) with a uniform quantizer having 256 representation levels.

****6.38** In this experiment we study the linear adaptive prediction of a signal x_n governed by the following recursion:

$$x_n = 0.8x_{n-1} - 0.1x_{n-2} + 0.1v_n$$

where v_n is drawn from a discrete-time white noise process of zero mean and unit variance. (A process generated in this manner is referred to as an *autoregressive process of order two*.) Specifically, the adaptive prediction is performed using the *normalized LMS algorithm* defined by

$$\hat{x}_n = \sum_{k=1}^p w_{k,n} x_{n-k}$$

$$e_n = x_n - \hat{x}_n$$

$$w_{k,n+1} = w_{k,n} + \mu \left(\sum_{k=1}^p x_{n-k}^2 \right) x_{n-k} e_n \quad k = 1, 2, \dots, p$$

where p is the prediction order and μ is the normalized step-size parameter. The important point to note here is that μ is dimensionless and stability of the algorithm is assured by choosing it in accordance with the formula

$$0 < \mu < 2$$

The algorithm is initiated by setting

$$w_{k,0} = 0 \quad \text{for all } k$$

The *learning curve* of the algorithm is defined as a plot of the mean-square error versus the number of iterations n for specified parameter values, which is obtained by averaging the plot of e_n^2 versus n over a large number of different realizations of the algorithm.

- Plot the learning curves for the adaptive prediction of x_n for a fixed prediction order $p = 5$ and three different values of step-size parameter: $\mu = 0.0075, 0.05, \text{ and } 0.5$.
- What observations can you make from the learning curves of part a?

****6.39** In this problem, we study adaptive delta modulation, the underlying principle of which is two-fold:

- If successive errors are of opposite polarity, then the delta modulator is operating in the granular mode, in which case the step size Δ is reduced.
- If, on the other hand, the successive errors are of the same polarity, then the delta modulator is operating in the slope-overload mode, in which case the step size Δ is increased.

Parts a and b of Figure P6.39 depict the block diagrams of the transmitter and receiver of the adaptive delta modulator, respectively, in which the step size, Δ , is increased or decreased by a factor of 50% at each iteration of the adaptive process, as shown by:

$$\Delta_n = \begin{cases} \frac{\Delta_{n-1}}{m_{q,n}} (m_{q,n} + 0.5m_{q,n-1}) & \text{if } \Delta_{n-1} \geq \Delta_{\min} \\ \Delta_{\min} & \text{if } \Delta_{n-1} < \Delta_{\min} \end{cases}$$

where Δ_n is the step size at iteration (time step) n of the adaptation algorithm, and $m_{q,n}$ is the 1-bit quantizer output that equals ± 1 .

Specifications: The input signal applied to the transmitter is sinusoidal as shown by

$$m_t = A \sin(2\pi f_m t)$$

where $A = 10$ and $f_m = f_s / 100$ where f_s is the sampling frequency; the step size $\Delta_n = 1$ for all n ; $\Delta_{\min} = 1/8$.

- Using the above-described adaptation algorithm, use a computer to plot the resulting waveform for one complete cycle of the sinusoidal modulating signal, and also display the coded modulator output in the transmitter.
- For the same specifications, repeat the computation using linear modulation.
- Comment on the results obtained in parts a and b of the problem.

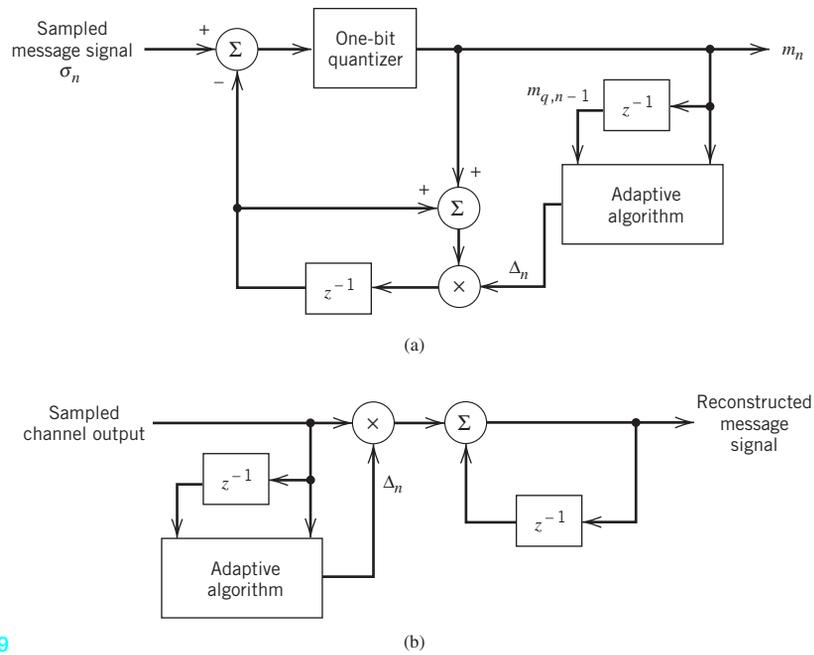


Figure P6.39

Notes

1. For an exhaustive study of quantization noise in signal processing and communications, see Widrow and Kollar (2008).
2. The two necessary conditions of (3.42) and (3.47) for optimality of a scalar quantizer were reported independently by Lloyd (1957) and Max (1960), hence the name “Lloyd–Max quantizer.” The derivation of these two optimality conditions presented in this chapter follows the book by Gersho and Gray (1992).
3. The μ -law is used in the USA, Canada, and Japan. On the other hand, in Europe, the A-law is used for signal compression.
4. In actual PCM systems, the companding circuitry does not produce an exact replica of the nonlinear compression curves shown in Figure 6.14. Rather, it provides a *piecewise linear* approximation to the desired curve. By using a large enough number of linear segments, the approximation can approach the true compression curve very closely; for detailed discussion of this issue, see Bellamy (1991).
5. For a discussion of noise in analog modulation systems with particular reference to FM, see Chapter 4 of *Communication Systems* (Haykin, 2001).
6. To simplify notational matters, \mathbf{R}_M is used to denote the autocorrelation matrix in (6.70) rather than \mathbf{R}_{MM} as in Chapter 4 on Stochastic Processes. To see the rationale for this simplification, the reader is referred to (6.79) for simplicity. For the same reason, henceforth the practice adopted in this chapter will be continued for the rest of the book, dealing with autocorrelation matrices and power spectral density.
7. An optimum predictor that follows (6.77) is said to be a special case of the *Wiener filter*.
8. For a detailed discussion of adaptive DPCM involving the use of adaptive quantization with forward estimation as well as backward estimation, the reader is referred to the classic book (Jayant and Noll, 1984).