

Principles of Digital Communication

The renowned communication theorist Robert Gallager brings his lucid writing style to this first-year graduate textbook on the fundamental system aspects of digital communication. With the clarity and insight that have characterized his teaching and earlier textbooks he develops a simple framework and then combines this with careful proofs to help the reader understand modern systems and simplified models in an intuitive yet precise way. Although many features of various modern digital communication systems are discussed, the focus is always on principles explained using a hierarchy of simple models.

A major simplifying principle of digital communication is to separate source coding and channel coding by a standard binary interface. Data compression, i.e., source coding, is then treated as the conversion of arbitrary communication sources into binary data streams. Similarly digital modulation, i.e., channel coding, becomes the conversion of binary data into waveforms suitable for transmission over communication channels. These waveforms are viewed as vectors in signal space, modeled mathematically as Hilbert space.

A self-contained introduction to random processes is used to model the noise and interference in communication channels. The principles of detection and decoding are then developed to extract the transmitted data from noisy received waveforms. An introduction to coding and coded modulation then leads to Shannon's noisy-channel coding theorem. The final topic is wireless communication. After developing models to explain various aspects of fading, there is a case study of cellular CDMA communication which illustrates the major principles of digital communication.

Throughout, principles are developed with both mathematical precision and intuitive explanations, allowing readers to choose their own mix of mathematics and engineering. An extensive set of exercises ranges from confidence-building examples to more challenging problems. Instructor solutions and other resources are available at www.cambridge.org/9780521879071.

'Prof. Gallager is a legendary figure . . . known for his insights and excellent style of exposition'

Professor Lang Tong, Cornell University

'a compelling read'

Professor Emre Telatar, EPFL

'It is surely going to be a classic in the field'

Professor Hideki Imai, University of Tokyo

Robert G. Gallager has had a profound influence on the development of modern digital communication systems through his research and teaching. As a Professor at M.I.T. since 1960 in the areas of information theory, communication technology, and data networks. He is a member of the U.S. National Academy of Engineering, the U.S. National Academy of Sciences, and, among many honors, received the IEEE Medal of Honor in 1990 and the Marconi prize in 2003. This text has been his central academic passion over recent years.

Cambridge University Press
978-0-521-87907-1 - Principles of Digital Communication
Robert G. Gallager
Frontmatter
[More information](#)

Cambridge University Press
978-0-521-87907-1 - Principles of Digital Communication
Robert G. Gallager
Frontmatter
[More information](#)

Principles of Digital Communication

ROBERT G. GALLAGER

Massachusetts Institute of Technology



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
978-0-521-87907-1 - Principles of Digital Communication
Robert G. Gallager
Frontmatter
[More information](#)

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Published in the United States of America by Cambridge University Press, New York

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9780521879071

© Cambridge University Press 2008

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2008, 2011

Second Edition 2012

Reprinted 2013

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-87907-1 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

	<i>Preface</i>	<i>page xi</i>
	<i>Acknowledgements</i>	<i>xiv</i>
1	Introduction to digital communication	1
	1.1 Standardized interfaces and layering	3
	1.2 Communication sources	5
	1.2.1 Source coding	6
	1.3 Communication channels	7
	1.3.1 Channel encoding (modulation)	10
	1.3.2 Error correction	11
	1.4 Digital interface	12
	1.4.1 Network aspects of the digital interface	12
	1.5 Supplementary reading	14
2	Coding for discrete sources	16
	2.1 Introduction	16
	2.2 Fixed-length codes for discrete sources	18
	2.3 Variable-length codes for discrete sources	19
	2.3.1 Unique decodability	20
	2.3.2 Prefix-free codes for discrete sources	21
	2.3.3 The Kraft inequality for prefix-free codes	23
	2.4 Probability models for discrete sources	26
	2.4.1 Discrete memoryless sources	26
	2.5 Minimum $\bar{L}$ for prefix-free codes	27
	2.5.1 Lagrange multiplier solution for the minimum $\bar{L}$	28
	2.5.2 Entropy bounds on $\bar{L}$	29
	2.5.3 Huffman’s algorithm for optimal source codes	31
	2.6 Entropy and fixed-to-variable-length codes	35
	2.6.1 Fixed-to-variable-length codes	37
	2.7 The AEP and the source coding theorems	38
	2.7.1 The weak law of large numbers	39
	2.7.2 The asymptotic equipartition property	40
	2.7.3 Source coding theorems	43
	2.7.4 The entropy bound for general classes of codes	44
	2.8 Markov sources	46
	2.8.1 Coding for Markov sources	48
	2.8.2 Conditional entropy	48

vi	Contents	
	2.9	Lempel–Ziv universal data compression 51
	2.9.1	The LZ77 algorithm 51
	2.9.2	Why LZ77 works 53
	2.9.3	Discussion 54
	2.10	Summary of discrete source coding 55
	2.11	Exercises 56
3	Quantization	67
	3.1	Introduction to quantization 67
	3.2	Scalar quantization 68
	3.2.1	Choice of intervals for given representation points 69
	3.2.2	Choice of representation points for given intervals 69
	3.2.3	The Lloyd–Max algorithm 70
	3.3	Vector quantization 72
	3.4	Entropy-coded quantization 73
	3.5	High-rate entropy-coded quantization 75
	3.6	Differential entropy 76
	3.7	Performance of uniform high-rate scalar quantizers 78
	3.8	High-rate two-dimensional quantizers 81
	3.9	Summary of quantization 84
	3.10	Appendixes 85
	3.10.1	Nonuniform scalar quantizers 85
	3.10.2	Nonuniform 2D quantizers 87
	3.11	Exercises 88
4	Source and channel waveforms	93
	4.1	Introduction 93
	4.1.1	Analog sources 93
	4.1.2	Communication channels 95
	4.2	Fourier series 96
	4.2.1	Finite-energy waveforms 98
	4.3	$\mathcal{L}_2$ functions and Lebesgue integration over $[-T/2, T/2]$ 101
	4.3.1	Lebesgue measure for a union of intervals 102
	4.3.2	Measure for more general sets 104
	4.3.3	Measurable functions and integration over $[-T/2, T/2]$ 106
	4.3.4	Measurability of functions defined by other functions 108
	4.3.5	$\mathcal{L}_1$ and $\mathcal{L}_2$ functions over $[-T/2, T/2]$ 108
	4.4	Fourier series for $\mathcal{L}_2$ waveforms 109
	4.4.1	The T -spaced truncated sinusoid expansion 111
	4.5	Fourier transforms and $\mathcal{L}_2$ waveforms 114
	4.5.1	Measure and integration over $\mathbb{R}$ 116
	4.5.2	Fourier transforms of $\mathcal{L}_2$ functions 118
	4.6	The DTFT and the sampling theorem 120
	4.6.1	The discrete-time Fourier transform 121
	4.6.2	The sampling theorem 122
	4.6.3	Source coding using sampled waveforms 124
	4.6.4	The sampling theorem for $[\Delta - W, \Delta + W]$ 125

	Contents	vii
4.7	Aliasing and the sinc-weighted sinusoid expansion	126
4.7.1	The T -spaced sinc-weighted sinusoid expansion	127
4.7.2	Degrees of freedom	128
4.7.3	Aliasing – a time-domain approach	129
4.7.4	Aliasing – a frequency-domain approach	130
4.8	Summary	132
4.9	Appendix: Supplementary material and proofs	133
4.9.1	Countable sets	133
4.9.2	Finite unions of intervals over $[-T/2, T/2]$	135
4.9.3	Countable unions and outer measure over $[-T/2, T/2]$	136
4.9.4	Arbitrary measurable sets over $[-T/2, T/2]$	139
4.10	Exercises	143
5	Vector spaces and signal space	153
5.1	Axioms and basic properties of vector spaces	154
5.1.1	Finite-dimensional vector spaces	156
5.2	Inner product spaces	158
5.2.1	The inner product spaces $\mathbb{R}^n$ and $\mathbb{C}^n$	158
5.2.2	One-dimensional projections	159
5.2.3	The inner product space of $\mathcal{L}_2$ functions	161
5.2.4	Subspaces of inner product spaces	162
5.3	Orthonormal bases and the projection theorem	163
5.3.1	Finite-dimensional projections	164
5.3.2	Corollaries of the projection theorem	165
5.3.3	Gram–Schmidt orthonormalization	166
5.3.4	Orthonormal expansions in $\mathcal{L}_2$	167
5.4	Summary	169
5.5	Appendix: Supplementary material and proofs	170
5.5.1	The Plancherel theorem	170
5.5.2	The sampling and aliasing theorems	174
5.5.3	Prolate spheroidal waveforms	176
5.6	Exercises	177
6	Channels, modulation, and demodulation	181
6.1	Introduction	181
6.2	Pulse amplitude modulation (PAM)	184
6.2.1	Signal constellations	184
6.2.2	Channel imperfections: a preliminary view	185
6.2.3	Choice of the modulation pulse	187
6.2.4	PAM demodulation	189
6.3	The Nyquist criterion	190
6.3.1	Band-edge symmetry	191
6.3.2	Choosing $\{p(t - kT); k \in \mathbb{Z}\}$ as an orthonormal set	193
6.3.3	Relation between PAM and analog source coding	194
6.4	Modulation: baseband to passband and back	195
6.4.1	Double-sideband amplitude modulation	195

6.5	Quadrature amplitude modulation (QAM)	196
6.5.1	QAM signal set	198
6.5.2	QAM baseband modulation and demodulation	199
6.5.3	QAM: baseband to passband and back	200
6.5.4	Implementation of QAM	201
6.6	Signal space and degrees of freedom	203
6.6.1	Distance and orthogonality	204
6.7	Carrier and phase recovery in QAM systems	206
6.7.1	Tracking phase in the presence of noise	207
6.7.2	Large phase errors	208
6.8	Summary of modulation and demodulation	208
6.9	Exercises	209
7	Random processes and noise	216
7.1	Introduction	216
7.2	Random processes	217
7.2.1	Examples of random processes	218
7.2.2	The mean and covariance of a random process	220
7.2.3	Additive noise channels	221
7.3	Gaussian random variables, vectors, and processes	221
7.3.1	The covariance matrix of a jointly Gaussian random vector	224
7.3.2	The probability density of a jointly Gaussian random vector	224
7.3.3	Special case of a 2D zero-mean Gaussian random vector	227
7.3.4	$Z = AW$, where A is orthogonal	228
7.3.5	Probability density for Gaussian vectors in terms of principal axes	228
7.3.6	Fourier transforms for joint densities	230
7.4	Linear functionals and filters for random processes	231
7.4.1	Gaussian processes defined over orthonormal expansions	232
7.4.2	Linear filtering of Gaussian processes	233
7.4.3	Covariance for linear functionals and filters	234
7.5	Stationarity and related concepts	235
7.5.1	Wide-sense stationary (WSS) random processes	236
7.5.2	Effectively stationary and effectively WSS random processes	238
7.5.3	Linear functionals for effectively WSS random processes	239
7.5.4	Linear filters for effectively WSS random processes	239
7.6	Stationarity in the frequency domain	242
7.7	White Gaussian noise	244
7.7.1	The sinc expansion as an approximation to WGN	246
7.7.2	Poisson process noise	247
7.8	Adding noise to modulated communication	248
7.8.1	Complex Gaussian random variables and vectors	250
7.9	Signal-to-noise ratio	251

	Contents	ix
7.10	Summary of random processes	254
7.11	Appendix: Supplementary topics	255
7.11.1	Properties of covariance matrices	255
7.11.2	The Fourier series expansion of a truncated random process	257
7.11.3	Uncorrelated coefficients in a Fourier series	259
7.11.4	The Karhunen–Loeve expansion	262
7.12	Exercises	263
8	Detection, coding, and decoding	268
8.1	Introduction	268
8.2	Binary detection	271
8.3	Binary signals in white Gaussian noise	273
8.3.1	Detection for PAM antipodal signals	273
8.3.2	Detection for binary nonantipodal signals	275
8.3.3	Detection for binary real vectors in WGN	276
8.3.4	Detection for binary complex vectors in WGN	279
8.3.5	Detection of binary antipodal waveforms in WGN	281
8.4	<i>M</i> -ary detection and sequence detection	285
8.4.1	<i>M</i> -ary detection	285
8.4.2	Successive transmissions of QAM signals in WGN	286
8.4.3	Detection with arbitrary modulation schemes	289
8.5	Orthogonal signal sets and simple channel coding	292
8.5.1	Simplex signal sets	293
8.5.2	Biorthogonal signal sets	294
8.5.3	Error probability for orthogonal signal sets	294
8.6	Block coding	298
8.6.1	Binary orthogonal codes and Hadamard matrices	298
8.6.2	Reed–Muller codes	300
8.7	Noisy-channel coding theorem	302
8.7.1	Discrete memoryless channels	303
8.7.2	Capacity	304
8.7.3	Converse to the noisy-channel coding theorem	306
8.7.4	Noisy-channel coding theorem, forward part	307
8.7.5	The noisy-channel coding theorem for WGN	311
8.8	Convolutional codes	312
8.8.1	Decoding of convolutional codes	314
8.8.2	The Viterbi algorithm	315
8.9	Summary of detection, coding, and decoding	317
8.10	Appendix: Neyman–Pearson threshold tests	317
8.11	Exercises	322
9	Wireless digital communication	330
9.1	Introduction	330
9.2	Physical modeling for wireless channels	334
9.2.1	Free-space, fixed transmitting and receiving antennas	334
9.2.2	Free-space, moving antenna	337
9.2.3	Moving antenna, reflecting wall	337

9.2.4	Reflection from a ground plane	340
9.2.5	Shadowing	340
9.2.6	Moving antenna, multiple reflectors	341
9.3	Input/output models of wireless channels	341
9.3.1	The system function and impulse response for LTV systems	343
9.3.2	Doppler spread and coherence time	345
9.3.3	Delay spread and coherence frequency	348
9.4	Baseband system functions and impulse responses	350
9.4.1	A discrete-time baseband model	353
9.5	Statistical channel models	355
9.5.1	Passband and baseband noise	358
9.6	Data detection	359
9.6.1	Binary detection in flat Rayleigh fading	360
9.6.2	Noncoherent detection with known channel magnitude	363
9.6.3	Noncoherent detection in flat Rician fading	365
9.7	Channel measurement	367
9.7.1	The use of probing signals to estimate the channel	368
9.7.2	Rake receivers	373
9.8	Diversity	376
9.9	CDMA: the IS95 standard	379
9.9.1	Voice compression	380
9.9.2	Channel coding and decoding	381
9.9.3	Viterbi decoding for fading channels	382
9.9.4	Modulation and demodulation	383
9.9.5	Multiaccess interference in IS95	386
9.10	Summary of wireless communication	388
9.11	Appendix: Error probability for noncoherent detection	390
9.12	Exercises	391
	<i>References</i>	398
	<i>Index</i>	400

Preface

Digital communication is an enormous and rapidly growing industry, roughly comparable in size to the computer industry. The objective of this text is to study those aspects of digital communication systems that are unique. That is, rather than focusing on hardware and software for these systems (which is much like that in many other fields), we focus on the fundamental system aspects of modern digital communication.

Digital communication is a field in which theoretical ideas have had an unusually powerful impact on system design and practice. The basis of the theory was developed in 1948 by Claude Shannon, and is called information theory. For the first 25 years or so of its existence, information theory served as a rich source of academic research problems and as a tantalizing suggestion that communication systems could be made more efficient and more reliable by using these approaches. Other than small experiments and a few highly specialized military systems, the theory had little interaction with practice. By the mid 1970s, however, mainstream systems using information-theoretic ideas began to be widely implemented. The first reason for this was the increasing number of engineers who understood both information theory and communication system practice. The second reason was that the low cost and increasing processing power of digital hardware made it possible to implement the sophisticated algorithms suggested by information theory. The third reason was that the increasing complexity of communication systems required the architectural principles of information theory.

The theoretical principles here fall roughly into two categories – the first provides analytical tools for determining the performance of particular systems, and the second puts fundamental limits on the performance of any system. Much of the first category can be understood by engineering undergraduates, while the second category is distinctly graduate in nature. It is not that graduate students know so much more than undergraduates, but rather that undergraduate engineering students are trained to master enormous amounts of detail and the equations that deal with that detail. They are not used to the patience and deep thinking required to understand abstract performance limits. This patience comes later with thesis research.

My original purpose was to write an undergraduate text on digital communication, but experience teaching this material over a number of years convinced me that I could not write an honest exposition of principles, including both what is possible and what is not possible, without losing most undergraduates. There are many excellent undergraduate texts on digital communication describing a wide variety of systems, and I did not see the need for another. Thus this text is now aimed at graduate students, but is accessible to patient undergraduates.

The relationship between theory, problem sets, and engineering/design in an academic subject is rather complex. The theory deals with relationships and analysis for *models* of real systems. A good theory (and information theory is one of the best) allows for simple analysis of simplified models. It also provides structural principles that allow insights from these simple models to be applied to more complex and

realistic models. Problem sets provide students with an opportunity to analyze these highly simplified models, and, with patience, to start to understand the general principles. Engineering deals with making the approximations and judgment calls to create simple models that focus on the critical elements of a situation, and from there to design workable systems.

The important point here is that engineering (at this level) cannot really be separated from theory. Engineering is necessary to choose appropriate theoretical models, and theory is necessary to find the general properties of those models. To oversimplify, engineering determines what the reality is and theory determines the consequences and structure of that reality. At a deeper level, however, the engineering perception of reality heavily depends on the perceived structure (all of us carry oversimplified models around in our heads). Similarly, the structures created by theory depend on engineering common sense to focus on important issues. Engineering sometimes becomes overly concerned with detail, and theory becomes overly concerned with mathematical niceties, but we shall try to avoid both these excesses here.

Each topic in the text is introduced with highly oversimplified toy models. The results about these toy models are then related to actual communication systems, and these are used to generalize the models. We then iterate back and forth between analysis of models and creation of models. Understanding the performance limits on classes of models is essential in this process.

There are many exercises designed to help the reader understand each topic. Some give examples showing how an analysis breaks down if the restrictions are violated. Since analysis always treats models rather than reality, these examples build insight into how the results about models apply to real systems. Other exercises apply the text results to very simple cases, and others generalize the results to more complex systems. Yet more explore the sense in which theoretical models apply to particular practical problems.

It is important to understand that the purpose of the exercises is not so much to get the “answer” as to acquire understanding. Thus students using this text will learn much more if they discuss the exercises with others and think about what they have learned after completing the exercise. The point is not to manipulate equations (which computers can now do better than students), but rather to understand the equations (which computers cannot do).

As pointed out above, the material here is primarily graduate in terms of abstraction and patience, but requires only a knowledge of elementary probability, linear systems, and simple mathematical abstraction, so it can be understood at the undergraduate level. For both undergraduates and graduates, I feel strongly that learning to reason about engineering material is more important, both in the workplace and in further education, than learning to pattern match and manipulate equations.

Most undergraduate communication texts aim at familiarity with a large variety of different systems that have been implemented historically. This is certainly valuable in the workplace, at least for the near term, and provides a rich set of examples that are valuable for further study. The digital communication field is so vast, however, that learning from examples is limited, and in the long term it is necessary to learn

the underlying principles. The examples from undergraduate courses provide a useful background for studying these principles, but the ability to reason abstractly that comes from elementary pure mathematics courses is equally valuable.

Most graduate communication texts focus more on the analysis of problems, with less focus on the modeling, approximation, and insight needed to see how these problems arise. Our objective here is to use simple models and approximations as a way to understand the general principles. We will use quite a bit of mathematics in the process, but the mathematics will be used to establish general results precisely, rather than to carry out detailed analyses of special cases.

Acknowledgements

This book has evolved from lecture notes for a one-semester course on digital communication given at MIT for the past ten years. I am particularly grateful for the feedback I have received from the other faculty members, namely Professors Amos Lapidoth, Dave Forney, Greg Wornell, and Lizhong Zheng, who have used these notes in the MIT course. Their comments, on both tutorial and technical aspects, have been critically helpful. The notes in the early years were written jointly with Amos Lapidoth and Dave Forney. The notes have been rewritten and edited countless times since then, but I am very grateful for their ideas and wording, which, even after many modifications, have been an enormous help. I am doubly indebted to Dave Forney for reading the entire text a number of times and saving me from many errors, ranging from conceptual to grammatical and stylistic.

I am indebted to a number of others, including Randy Berry, Sanjoy Mitter, Baris Nakiboglu, Emre Telatar, David Tse, Edmund Yeh, and some anonymous reviewers for important help with both content and tutorial presentation.

Emre Koksall, Tengo Saengudomlert, Shan-Yuan Ho, Manish Bhardwaj, Ashish Khisti, Etty Lee, and Emmanuel Abbe have all made major contributions to the text as teaching assistants for the MIT course. They have not only suggested new exercises and prepared solutions for others, but have also given me many insights about why certain material is difficult for some students, and suggested how to explain it better to avoid this confusion. The final test for clarity, of course, comes from the three or four hundred students who have taken the course over the last ten years, and I am grateful to them for looking puzzled when my explanations have failed and asking questions when I have been almost successful.

Finally, I am particularly grateful to my wife, Marie, for making our life a delight, even during the worst moments of writing yet another book.