

TURING

图灵原版电子与电气工程系列

CAMBRIDGE

国际信息领域泰斗Robert G. Gallager博士最新力作

# 数字通信原理 (英文版)

## Principles of Digital Communication

[美] Robert G. Gallager 著



人民邮电出版社  
POSTS & TELECOM PRESS

TURING

图灵原版电子与电气工程系列

# 数字通信原理 (英文版)

Principles of Digital Communication



[美] Robert G. Gallager 著

人民邮电出版社  
北京

## 图书在版编目 (CIP) 数据

数字通信原理=Principles of Digital  
Communication: 英文/(美)加拉格 (Gallager, R. G.)  
著. —北京: 人民邮电出版社, 2010.4  
(图灵原版电子与电气工程系列)  
ISBN 978-7-115-22336-4

I. ①数… II. ①加… III. ①数字通信—英文 IV.  
①TN914.3

中国版本图书馆CIP数据核字 (2010) 第 030008 号

## 内 容 提 要

本书是世界通信权威、信息领域泰斗 Robert G. Gallager 博士新作, 在数字通信原理的基础上精炼, 重点阐述了理论、问题和工程设计之间的关系。内容涉及离散源编码、量化、信道波形、向量空间和信号空间、随机过程和噪声、编码、解码等数字通信基本问题, 最后还简单介绍了无线数字通信。

本书是通信专业高年级本科生和研究生教材, 也可供工程技术人员参考。

图灵原版电子与电气工程系列

## 数字通信原理 (英文版)

- 
- ◆ 著 [美] Robert G. Gallager
  - 责任编辑 朱 巍
  - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号  
邮编 100061 电子函件 315@ptpress.com.cn  
网址 <http://www.ptpress.com.cn>  
北京隆昌伟业印刷有限公司印刷
  - ◆ 开本: 700 × 1000 1/16  
印张: 26.25  
字数: 600 千字 2010 年 4 月第 1 版  
印数: 1 - 2 500 册 2010 年 4 月北京第 1 次印刷

著作权合同登记号 图字: 01-2010-0208 号

ISBN 978-7-115-22336-4

---

定价: 69.00 元

读者服务热线: (010) 51095186 印装质量热线: (010) 67129223

反盗版热线: (010) 67171154

# 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.

# Contents

<b>1</b>	<b>Introduction to digital communication</b>	<b>1</b>
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
<b>2</b>	<b>Coding for discrete sources</b>	<b>16</b>
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



.....

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
<b>3</b>	<b>Quantization</b>	<b>67</b>
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
<b>4</b>	<b>Source and channel waveforms</b>	<b>93</b>
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

4.7	<i>Aliasing and the sinc-weighted sinusoid expansion</i>	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
<b>5</b>	<b>Vector spaces and signal space</b>	<b>153</b>
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
<b>6</b>	<b>Channels, modulation, and demodulation</b>	<b>181</b>
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
<b>7</b>	<b>Random processes and noise</b>	<b>216</b>
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

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
<b>8</b>	<b>Detection, coding, and decoding</b>	<b>268</b>
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	$M$ -ary detection and sequence detection	285
8.4.1	$M$ -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
<b>9</b>	<b>Wireless digital communication</b>	<b>330</b>
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

# 1 Introduction to digital communication

Communication has been one of the deepest needs of the human race throughout recorded history. It is essential to forming social unions, to educating the young, and to expressing a myriad of emotions and needs. Good communication is central to a civilized society.

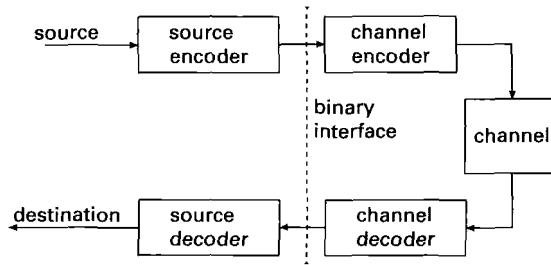
The various communication disciplines in engineering have the purpose of providing technological aids to human communication. One could view the smoke signals and drum rolls of primitive societies as being technological aids to communication, but communication technology as we view it today became important with telegraphy, then telephony, then video, then computer communication, and today the amazing mixture of all of these in inexpensive, small portable devices.

Initially these technologies were developed as separate networks and were viewed as having little in common. As these networks grew, however, the fact that all parts of a given network had to work together, coupled with the fact that different components were developed at different times using different design methodologies, caused an increased focus on the underlying principles and architectural understanding required for continued system evolution.

This need for basic principles was probably best understood at American Telephone and Telegraph (AT&T), where Bell Laboratories was created as the research and development arm of AT&T. The Math Center at Bell Labs became the predominant center for communication research in the world, and held that position until quite recently. The central core of the principles of communication technology were developed at that center.

Perhaps the greatest contribution from the Math Center was the creation of Information Theory by Claude Shannon (Shannon, 1948). For perhaps the first 25 years of its existence, Information Theory was regarded as a beautiful theory but not as a central guide to the architecture and design of communication systems. After that time, however, both the device technology and the engineering understanding of the theory were sufficient to enable system development to follow information theoretic principles.

A number of information theoretic ideas and how they affect communication system design will be explained carefully in subsequent chapters. One pair of ideas, however, is central to almost every topic. The first is to view all communication sources, e.g., speech waveforms, image waveforms, and text files, as being representable by binary sequences. The second is to design communication systems that first convert the



**Figure 1.1.** Placing a binary interface between source and channel. The source encoder converts the source output to a binary sequence and the channel encoder (often called a modulator) processes the binary sequence for transmission over the channel. The channel decoder (demodulator) recreates the incoming binary sequence (hopefully reliably), and the source decoder recreates the source output.

source output into a binary sequence and then convert that binary sequence into a form suitable for transmission over particular physical media such as cable, twisted wire pair, optical fiber, or electromagnetic radiation through space.

*Digital communication systems*, by definition, are communication systems that use such a digital<sup>1</sup> sequence as an interface between the source and the channel input (and similarly between the channel output and final destination) (see Figure 1.1).

The idea of converting an analog source output to a binary sequence was quite revolutionary in 1948, and the notion that this should be done before channel processing was even more revolutionary. Today, with digital cameras, digital video, digital voice, etc., the idea of digitizing any kind of source is commonplace even among the most technophobic. The notion of a binary interface before channel transmission is almost as commonplace. For example, we all refer to the speed of our Internet connection in bits per second.

There are a number of reasons why communication systems now usually contain a binary interface between source and channel (i.e., why digital communication systems are now standard). These will be explained with the necessary qualifications later, but briefly they are as follows.

- Digital hardware has become so cheap, reliable, and miniaturized that digital interfaces are eminently practical.
- A standardized binary interface between source and channel simplifies implementation and understanding, since source coding/decoding can be done independently of the channel, and, similarly, channel coding/decoding can be done independently of the source.

<sup>1</sup> A digital sequence is a sequence made up of elements from a finite alphabet (e.g. the binary digits {0, 1}, the decimal digits {0, 1, ..., 9}, or the letters of the English alphabet). The binary digits are almost universally used for digital communication and storage, so we only distinguish digital from binary in those few places where the difference is significant.

- A standardized binary interface between source and channel simplifies networking, which now reduces to sending binary sequences through the network.
- One of the most important of Shannon's information theoretic results is that if a source can be transmitted over a channel in any way at all, it can be transmitted using a binary interface between source and channel. This is known as the *source/channel separation theorem*.

In the remainder of this chapter, the problems of source coding and decoding and channel coding and decoding are briefly introduced. First, however, the notion of layering in a communication system is introduced. One particularly important example of layering was introduced in Figure 1.1, where source coding and decoding are viewed as one layer and channel coding and decoding are viewed as another layer.

## 1.1 Standardized interfaces and layering

Large communication systems such as the Public Switched Telephone Network (PSTN) and the Internet have incredible complexity, made up of an enormous variety of equipment made by different manufacturers at different times following different design principles. Such complex networks need to be based on some simple architectural principles in order to be understood, managed, and maintained. Two such fundamental architectural principles are *standardized interfaces* and *layering*.

A standardized interface allows the user or equipment on one side of the interface to ignore all details about the other side of the interface except for certain specified interface characteristics. For example, the binary interface<sup>2</sup> in Figure 1.1 allows the source coding/decoding to be done independently of the channel coding/decoding.

The idea of layering in communication systems is to break up communication functions into a string of separate layers, as illustrated in Figure 1.2.

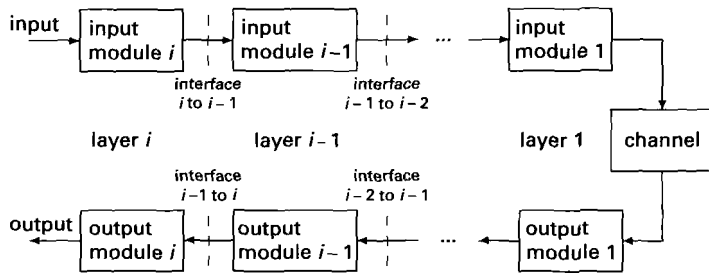
Each layer consists of an input module at the input end of a communication system and a “peer” output module at the other end. The input module at layer  $i$  processes the information received from layer  $i + 1$  and sends the processed information on to layer  $i - 1$ . The peer output module at layer  $i$  works in the opposite direction, processing the received information from layer  $i - 1$  and sending it on to layer  $i$ .

As an example, an input module might receive a voice waveform from the next higher layer and convert the waveform into a binary data sequence that is passed on to the next lower layer. The output peer module would receive a binary sequence from the next lower layer at the output and convert it back to a speech waveform.

As another example, a *modem* consists of an input module (a modulator) and an output module (a demodulator). The modulator receives a binary sequence from the next higher input layer and generates a corresponding modulated waveform for transmission over a channel. The peer module is the remote demodulator at the other end of the channel. It receives a more or less faithful replica of the transmitted

<sup>2</sup> The use of a binary sequence at the interface is not quite enough to specify it, as will be discussed later.

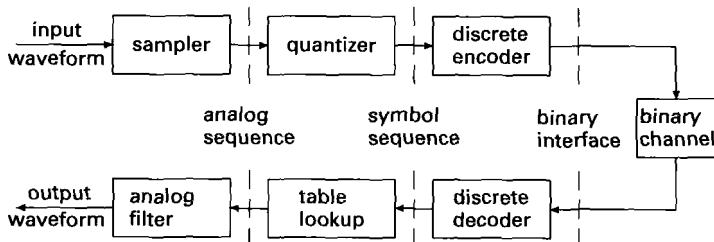




**Figure 1.2.** Layers and interfaces. The specification of the interface between layers  $i$  and  $i-1$  should specify how input module  $i$  communicates with input module  $i-1$ , how the corresponding output modules communicate, and, most important, the input/output behavior of the system to the right of the interface. The designer of layer  $i-1$  uses the input/output behavior of the layers to the right of  $i-1$  to produce the required input/output performance to the right of layer  $i$ . Later examples will show how this multilayer process can simplify the overall system design.

waveform and reconstructs a typically faithful replica of the binary sequence. Similarly, the local demodulator is the peer to a remote modulator (often collocated with the remote demodulator above). Thus a modem is an input module for communication in one direction and an output module for independent communication in the opposite direction. Later chapters consider modems in much greater depth, including how noise affects the channel waveform and how that affects the reliability of the recovered binary sequence at the output. For now, however, it is enough simply to view the modulator as converting a binary sequence to a waveform, with the peer demodulator converting the waveform back to the binary sequence.

As another example, the source coding/decoding layer for a waveform source can be split into three layers, as shown in Figure 1.3. One of the advantages of this layering is that discrete sources are an important topic in their own right (discussed in Chapter 2) and correspond to the inner layer of Figure 1.3. Quantization is also an important topic in its own right (discussed in Chapter 3). After both of these are understood, waveform sources become quite simple to understand.



**Figure 1.3.** Breaking the source coding/decoding layer into three layers for a waveform source. The input side of the outermost layer converts the waveform into a sequence of samples and the output side converts the recovered samples back to the waveform. The quantizer then converts each sample into one of a finite set of symbols, and the peer module recreates the sample (with some distortion). Finally the inner layer encodes the sequence of symbols into binary digits.