MODERN DIGITAL AND ANALOG COMMUNICATION SYSTEMS

4TH EDITION

South Asia Edition

B. P. LATHI

California State University—Sacramento

ZHI DING

University of California—Davis

Adapted by HARI M. GUPTA

Indian Institute of Technology Delhi

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trademark of Oxford University Press in the UK and in certain other countries.

Published in India by
Oxford University Press India
Ground Floor, 2/11, Ansari Road, Daryaganj, New Delhi 110002, India

First Indian Edition published in 2017

Modern Digital and Analog Communication Systems, International Fourth Edition (ISBN: 9780195384932) was originally published in English in 2009 by Oxford University Press USA, 198 Madison Avenue, New York, NY 10016. This adapted edition is published in arrangement with Oxford University Press. Oxford University Press India is solely responsible for this adaptation from the original work. Copyright ©2009 by Oxford University Press

ISBN-13: 978-0-19-947628-2
ISBN-10: 0-19-947628-4

Typeset in Times New Roman
by Cameo Corporate Services Limited, Chennai
Printed in India by Magic International (P) Ltd., Greater Noida

Cover image: vic927 / Shutterstock

PREFACE

This adapted version of *Modern Digital and Analog Communication Systems*, fourth international edition, is designed as a textbook for students of electrical, electronics, and communication engineering. The primary objective of the book is to provide a comprehensive coverage of the basic principles of design and analysis of analog and digital communication systems. Every effort has been made to deliver insights—rather than just understanding—as well as heuristic explanations of theoretical results wherever possible. Numerous examples are provided for further clarification of abstract results.

New to this Edition

Based on the suggestions from users, this edition has been updated and revised to enhance the clarity of presentation where required. Some of the prominent changes are as follows:

- To improve the structural organization of the book, chapters 2 and 3 of the previous edition, related to signal description and signal transmission through systems, as well as chapter 8 on probability theory and chapter 9 on random processes have been merged into single chapters. This makes the evolution of subject material coherent and seamless.
- A new chapter on Noise in Communication Systems has been added to complete the discussion on noise performance of analog communication systems. The material in this chapter is spread in two parts, viz. receiving system noise—both thermal and shot noise—and effect of channel noise on performance of amplitude and angle modulated systems. This chapter clearly establishes superiority of FM systems over AM systems under some specified conditions.
- About 50 solved examples and an equal number of end-of-chapter problems have been added throughout the text. These examples expand the existing concepts and the new problems range from practice-oriented to challenging. Both examples and problems are expected to enhance the understanding and appreciation of the subject matter.
- The discussion at various places in the book has been augmented, by including new sections on graphical interpretation of convolution (chapter 2), discrete random variables (chapter 6), multipath and Doppler spread (chapter 11), and Shannon-Fano coding (chapter 12), and by expanding coverage of material pertaining to superheterodyne receivers (chapter 4). These augmentations are necessary for better understanding of the concepts related to communication techniques and systems.

Organization

The book begins with a traditional review of signal and system fundamentals and proceeds to the core communication topics of analog modulation. It then presents the fundamental tools of probability theory and random processes to be used in the design and analysis of digital communications in the rest of this text. After coverage of the fundamentals of digital communication systems, the last two chapters provide an overview of information theory and the fundamentals of forward error correction codes. Ideally, the subjects covered in the book should be taught in two courses: one on the basic operations of communication systems, and one on the analysis of modern communication systems under noise and other distortions. The former relies heavily on deterministic analytical tools such as Fourier series, Fourier transforms and the sampling theorem, while the latter relies on tools from probability and random processes to tackle the unpredictability of message signals and noises. This book is designed for adoption both as a one-semester course (in which the deterministic aspects of communication systems are emphasized with little consideration of the effects of noise and interference) and for a course that deals with both the deterministic and probabilistic aspects of communication systems. The book is self-contained, providing all the necessary background in probability and random processes.

Chapter 1 introduces a panoramic view of communication systems. All the important concepts of communication theory are explained qualitatively in a heuristic way. This attracts students to communication topics in general. With this momentum, they are motivated to study the tool of signal analysis in **Chapter 2**, where they are encouraged to see a signal as a vector, and to think of the Fourier spectrum as a way of representing a signal in terms of its vector components. **Chapters 3 and 4** discuss amplitude (linear) and angle (nonlinear) modulations, respectively. Many instructors feel that in this digital age, modulation should be deemphasized. We hold that modulation is not so much a method of communication as a basic tool of signal processing; it will always be needed, not only in the area of communication (digital or analog), but also in many other areas of electrical engineering. Hence, neglecting modulation may prove to be rather shortsighted. **Chapter 5**, which serves as the fundamental link between analog and digital communications, describes the process of analog-to-digital conversion (ADC). It provides details of sampling, pulse code modulation (including DPCM), delta modulation, speech coding (vocoder), image/video coding, and compression. **Chapter 6** provides the essential background on theories of probability and random processes. These comprise the tools required for the study of communication systems. Every attempt is made to motivate students and to maintain their interest through these chapters by providing applications to communication problems wherever possible. **Chapter 7**, new to this edition, presents a unified treatment of noise in communication systems. It focuses on noise in receiving systems, both thermal and shot noise, and performance of analog modulated systems in the presence of channel noise. **Chapter 8** discusses the principles and techniques used in digital modulation. It introduces the concept of channel distortion and presents equalization as an effective means of distortion compensation. **Chapter 9** presents the analysis of digital communication systems in the presence of noise. It contains optimum signal detection in digital communication. **Chapter 10** focuses on spread spectrum communications. **Chapter 11** presents various practical techniques that can be used to combat practical channel distortions. This chapter captures both channel equalization and the broadly applied technology of OFDM. **Chapter 12** provides a tutorial on information theory. Finally, the principles and key practical aspects of error control coding are given in **Chapter 13**.

Appendices A–E provide useful material on orthogonality of signal sets, Cauchy–Schwarz inequality, Gram–Schmidt orthogonalization of a vector set, and basic matrix properties and operations. **Appendix F** provides answers to selected end-of-chapter problems.

MATLAB and Laboratory Experience

Since many universities no longer have hardware communication laboratories, MATLAB based communication system exercises are included to enhance the learning experience. Students will be able to design systems and modify their parameters to evaluate the overall effects on the performance of communication systems through computer displays and the measurement of bit error rates. Students will acquire first-hand knowledge on how to design and perform simulations of communication systems.

Online Resources

The following resources are available at https://india.oup.com/orcs/9780199476282 to support the faculty using this book.

- Instructor's Manual
- PowerPoint-based Figure Slides

Acknowledgments

First, we thank all the students we have had over the years. This edition would not have been possible without the much feedback from, and discussions with, our students. We thank all the reviewers for providing invaluable input to improve the text. Finally, we also thank Professor Norman Morrison, University of Cape Town, for suggesting a new problem (P6.2-3) in this edition.

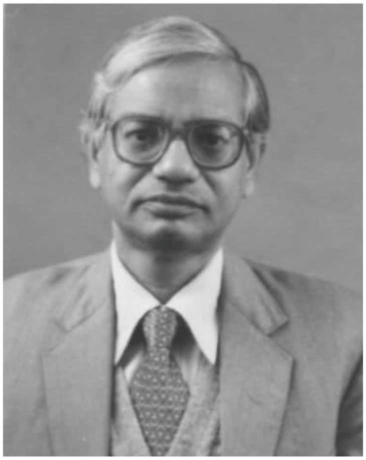
B. P. Lathi
Zhi Ding

I thank my numerous students over more than four decades at IIT Delhi, McGill University, Montreal, Canada, and Drexel University, Philadelphia, USA, who contributed to the instructional process in the general area of Communication Systems which is reflected in the creation of examples and problems in this adapted edition. Among my colleagues, I particularly thank Prof Swades De of IIT Delhi for illuminating discussions on the subject. Finally, I must thank my wife Sneh and my daughters Ira and Neha for supporting me throughout this effort.

Hari M. Gupta

Oxford University Press India would like to thank Prof Animesh Maitra of Institute of Radio Physics and Electronics, University of Calcutta, for helping with developing the adaptation plan of this adapted edition.

ABOUT THE ADAPTATION AUTHOR



Dr Hari Mohan Gupta was a Chair Professor and an Emeritus Professor at Indian Institute of Technology Delhi. He received his B.E. from University of Roorkee (now IIT Roorkee), M.Tech. from IIT Kharagpur, and PhD from IIT Kanpur. He joined the faculty of Department of Electrical Engineering, IIT Delhi in 1973, where he became a full professor in 1986. He has held faculty positions at McGill University, Montreal, Canada, and Drexel University, Philadelphia, USA. Currently he is an Adjunct Professor at IIT Ropar and at Indian Institute of Information Technology and Management (IIITM), Gwalior.

Dr Gupta is a fellow of Institution of Communication Engineers and Information Technologists (ICEIT), Institution of Engineers (India), Computer Society of India, and Institution of Electronics and Telecommunication Engineers (IETE). He has been active in research and has more than 150 publications to his credit in reputed journals and conferences.

His academic and research interests are telecommunication systems and networks, computer communications, photonic information systems, and multimedia communications.

DETAILED CONTENTS

1 INTRODUCTION

Over the last two decades, the rapid expansion of digital communication technologies has been simply astounding. Internet, a word and concept once familiar only to technologists and the scientific community, has permeated every aspect of people's daily lives. It is quite difficult to find any individual in a modern society that has not been touched by new communication technologies ranging from cellular phones to Bluetooth. This book examines the basic principles of communication by electric signals. Before modern times, messages were carried by runners, carrier pigeons, lights, and fires. These schemes were adequate for the distances and "data rates" of the age. In most parts of the world, these modes of communication have been superseded by electrical communication systems,* which can transmit signals over much longer distances (even to distant planets and galaxies) and at the speed of light.

Electrical communication is dependable and economical; communication technologies improve productivity and energy conservation. Increasingly, business meetings are conducted through teleconferences, saving the time and energy formerly expended on travel. Ubiquitous communication allows real-time management and coordination of project participants from around the globe. E-mail is rapidly replacing the more costly and slower "snail mails." E-commerce has also drastically reduced some costs and delays associated with marketing, while customers are also much better informed about new products and product information. Traditional media outlets such as television, radio, and newspapers have been rapidly evolving in the past few years to cope with, and better utilize, the new communication and networking technologies. The goal of this textbook is to provide the fundamental technical knowledge needed by next-generation communication engineers and technologists for designing even better communication systems of the future.

1.1 COMMUNICATION SYSTEMS

Figure 1.1 presents three typical communication systems: a wire-line telephone–cellular phone connection, a TV broadcasting system, and a wireless computer network. Because of the numerous examples of communication systems in existence, it would be unwise to attempt to study the details of all kinds of communication systems in this book. Instead, the most efficient and effective way to learn about communication is by studying the major functional blocks common to practically all communication systems. This way, students are not merely learning the operations of those existing systems they have studied; more importantly, they can acquire the basic knowledge needed to design and analyze new systems never encountered in a textbook. To begin, it is essential to establish a typical communication system model as shown in Fig. 1.2. The key components of a communication system are as follows.

* With the exception of the postal service.

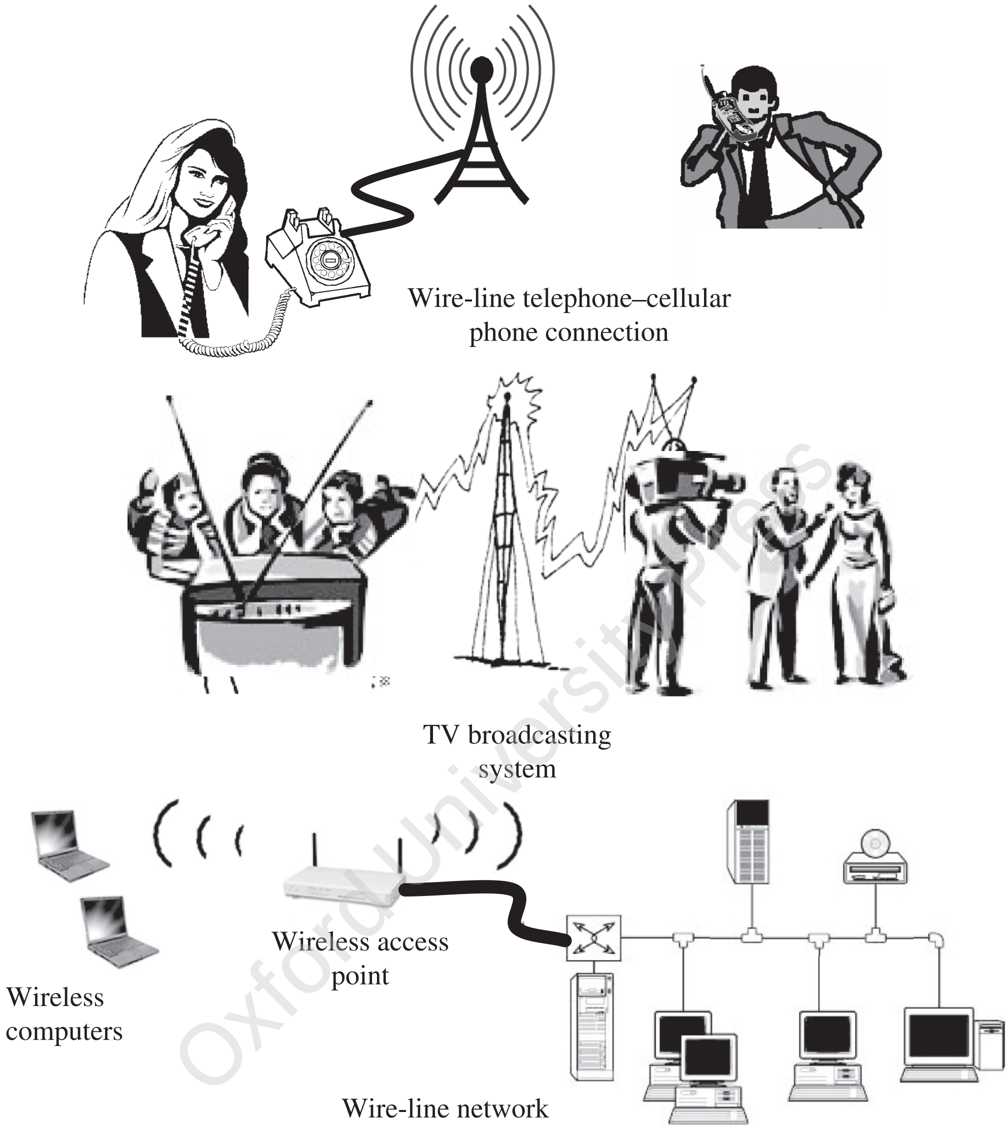


Figure 1.1 Some examples of communications systems.

The **source** originates a message, such as a human voice, a television picture, an e-mail message, or data. If the data is nonelectric (e.g., human voice, e-mail text, television video), it must be converted by an **input transducer** into an electric waveform referred to as the **baseband signal** or **message signal** through physical devices such as a microphone, a computer keyboard, or a CCD camera.

The **transmitter** modifies the baseband signal for efficient transmission. The transmitter may consist of one or more subsystems: an A/D converter, an encoder, and a modulator. Similarly, the receiver may consist of a demodulator, a decoder, and a D/A converter.

The **channel** is a medium of choice that can convey the electric signals at the transmitter output over a distance. A typical channel can be a pair of twisted copper wires (telephone and DSL), coaxial cable (television and internet), an optical fiber, or a radio link. Additionally, a channel can also be a point-to-point connection in a mesh of interconnected channels that form a communication network.

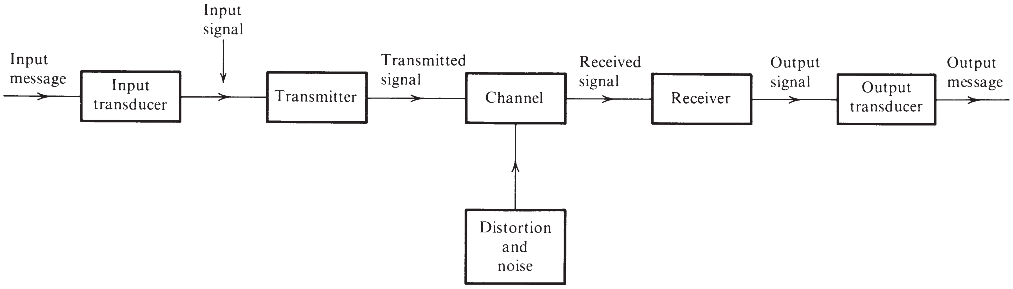


Figure 1.2 Communication system.

The **receiver** reprocesses the signal received from the channel by reversing the signal modifications made at the transmitter and removing the distortions made by the channel. The receiver output is fed to the **output transducer**, which converts the electric signal to its original form—the message.

The **destination** is the unit to which the message is communicated.

A channel is a physical medium that behaves partly like a filter that generally attenuates the signal and distorts the transmitted waveforms. The signal attenuation increases with the length of the channel, varying from a few percent for short distances to orders of magnitude in interplanetary communications. Signal waveforms are distorted because of physical phenomena such as frequency-dependent gains, multipath effects, and Doppler shift. For example, a *frequency-selective* channel causes different amounts of attenuation and phase shift to different frequency components of the signal. A square pulse is rounded or "spread out" during transmission over a lowpass channel. These types of distortion, called **linear distortion**, can be partly corrected at the receiver by an equalizer with gain and phase characteristics complementary to those of the channel. Channels may also cause **nonlinear distortion** through attenuation that varies with the signal amplitude. Such distortions can also be partly corrected by a complementary equalizer at the receiver. Channel distortions, if known, can also be precompensated by transmitters by applying channel-dependent predistortions.

In a practical environment, signals passing through communication channels not only experience channel distortions but also are corrupted along the path by undesirable interferences and disturbances lumped under the broad term **noise**. These interfering signals are random and are unpredictable from sources both external and internal. External noise includes interference signals transmitted on nearby channels, human-made noise generated by faulty contact switches of electrical equipment, automobile ignition radiation, fluorescent lights or natural noise from lightning, microwave ovens, and cellphone emissions, as well as electric storms and solar and intergalactic radiation. With proper care in system design, external noise can be minimized or even eliminated in some cases. Internal noise results from thermal motion of charged particles in conductors, random emission, and diffusion or recombination of charged carriers in electronic devices. Proper care can reduce the effect of internal noise but can never eliminate it. Noise is one of the underlying factors that limit the rate of telecommunications.

Thus, in practical communication systems, the channel distorts the signal, and noise accumulates along the path. Worse yet, the signal strength decreases while the noise level remains steady regardless of the distance from the transmitter. Thus, the signal quality is continuously worsening along the length of the channel. Amplification of the received signal to make up for the attenuation is to no avail because the noise will be amplified by the same proportion, and the quality remains, at best, unchanged.* These are the key challenges that we must face in designing modern communication systems.

* Actually, amplification may further deteriorate the signal because of additional amplifier noise.

1.2 ANALOG AND DIGITAL MESSAGES

Messages are digital or analog. Digital messages are ordered combinations of finite symbols or codewords. For example, printed English consists of 26 letters, 10 numbers, a space, and several punctuation marks. Thus, a text document written in English is a digital message constructed from the ASCII keyboard of 128 symbols. Human speech is also a digital message, because it is made up from a finite vocabulary in a language.* Music notes are also digital, even though the music sound itself is analog. Similarly, a Morse-coded telegraph message is a digital message constructed from a set of only **two** symbols—dash and dot. It is therefore a **binary** message, implying only two symbols. A digital message constructed with M symbols is called an M**-ary** message.

Analog messages, on the other hand, are characterized by data whose values vary over a continuous range and are defined for a continuous range of time. For example, the temperature or the atmospheric pressure of a certain location over time can vary over a continuous range and can assume an (uncountable) infinite number of possible values. A piece of music recorded by a pianist is also an analog signal. Similarly, a particular speech waveform has amplitudes that vary over a continuous range. Over a given time interval, an infinite number of possible different speech waveforms exist, in contrast to only a finite number of possible digital messages.

1.2.1 Noise Immunity of Digital Signals

It is no secret to even a casual observer that every time one looks at the latest electronic communication products, newer and better "digital technology" is replacing the old analog technology. Within the last three decades, cellular phones have completed their transformation from the first-generation analog AMPS to the second-generation (e.g., GSM, CDMA), to the third-generation (e.g., WCDMA), and to the fourth-generation (e.g., MIMO, LTE) digital offspring. Fifth-generation mobile systems are under design and testing and are expected to be deployed in the near future. Digital television continues the digital assault on analog video technology by driving out the last analog holdout of color television. There is every reason to ask: Why are digital technologies better? The answer has to do with both economics and quality. The case for economics is made by noting the ease of adopting versatile, powerful, and inexpensive high-speed digital microprocessors. But more importantly at the quality level, one prominent feature of digital communications is the enhanced immunity of digital signals to noise and interferences. This is especially noticeable in satellite communication where received signal levels are low and the noise introduced in the communication channels is very high.

Digital messages are transmitted as a finite set of electrical waveforms. In other words, a digital message is generated from a finite alphabet, while each character in the alphabet can be represented by one waveform or a sequential combination of such waveforms. For example, in sending messages via Morse code, a dash can be transmitted by an electrical pulse of amplitude $A/2$ and a dot can be transmitted by a pulse of negative amplitude $-A/2$ (Fig 1.3a). In an M-ary case, M distinct electrical pulses (or waveforms) are used; each of the M pulses represents one of the M possible symbols. Once transmitted, the receiver must extract the message from a distorted and noisy signal at the channel output. Message extraction is often easier from digital signals than from analog signals because the digital decision must belong to the finite-sized alphabet. Consider a binary case: two symbols are encoded as rectangular pulses of amplitudes $A/2$ and $-A/2$. The only decision at the receiver is to select between two possible pulses received; the fine details of the pulse shape are not an issue. A finite alphabet leads to noise and interference immunity. The receiver's decision can be made with reasonable certainty even if the pulses have suffered modest distortion and noise (Fig. 1.3). The digital message in Fig. 1.3a is distorted by the channel, as shown in Fig. 1.3b. Yet, if the distortion is not too large, we can recover the data without error because we need make only a simple binary decision: Is the

* Here we imply the information contained in the speech rather than its details such as the pronunciation of words and varying inflections, pitch, and emphasis. The speech signal from a microphone contains all these details and is therefore an analog signal, and its information content is more than a thousand times greater than the information accessible from the written text of the same speech.

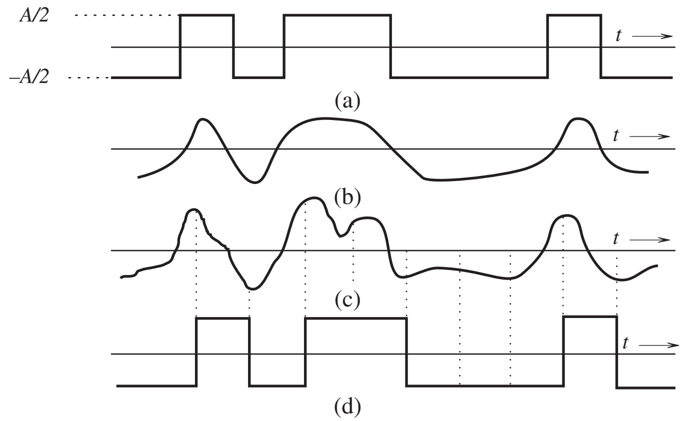


Figure 1.3 (a) Transmitted signal. (b) Received distorted signal (without noise). (c) Received distorted signal (with noise). (d) Regenerated signal (delayed).

received pulse positive or negative? Figure 1.3c shows the same data with channel distortion and noise. Here again, the data can be recovered correctly as long as the distortion and the noise are within limits. In contrast, the waveform shape itself in an analog message carries the needed information, and even a slight distortion or interference in the waveform will show up in the received signal. Clearly, a digital communication system is more rugged than an analog communication system in the sense that it can better withstand noise and distortion (as long as they are within a limit).

1.2.2 Viability of Distortionless Regenerative Repeaters

One main reason for the superior quality of digital systems over analog ones is the viability of **regenerative** repeaters and network nodes in the former. Repeater stations are placed along the communication path of a digital system at distances short enough to ensure that noise and distortion remain within a limit. This allows pulse detection with high accuracy. At each repeater station, or network node, the incoming pulses are detected such that new, "clean" pulses are retransmitted to the next repeater station or node. This process prevents the accumulation of noise and distortion along the path by cleaning the pulses at regular repeater intervals. We can thus transmit messages over longer distances with greater accuracy. There has been widespread application of distortionless regeneration by repeaters in long-haul communication systems and by nodes in a large (possibly heterogeneous) network.

For analog systems, signals and noise within the same bandwidth cannot be separated. Repeaters in analog systems are basically filters plus amplifiers and are not "regenerative." Thus, it is impossible to avoid in-band accumulation of noise and distortion along the path. As a result, the distortion and the noise interference can accumulate over the entire transmission path as a signal traverses through the network. To compound the problem, the signal is attenuated continuously over the transmission path. Thus, with increasing distance the signal becomes weaker, whereas the distortion and the noise accumulate more. Ultimately, the signal, overwhelmed by the distortion and noise, is buried beyond recognition. Amplification is of little help, since it enhances both the signal and the noise equally. Consequently, the distance over which an analog message can be successfully received is limited by the first transmitter power. Despite these limitations, analog communication was used widely and successfully in the past for short- to medium-range communications. Nowadays, because of the advent of optical fiber communications and the dramatic cost reduction achieved in the fabrication of high-speed digital circuitry and digital storage devices, almost all new communication systems being installed are digital. But some old analog communication facilities are still in use, including those for AM and FM radio broadcasting.

1.2.3 Analog-to-Digital (A/D) Conversion

Despite the differences between analog and digital signals, a meeting ground exists between them: conversion of analog signals to digital signals (A/D conversion). A key device in electronics, the analog-to-digital (A/D) converter, enables digital communication systems to convey analog source signals such as audio and video. Generally, analog signals are continuous in time and in range; that is, they have values at every time instant, and their values can be anything within the range. On the other hand, digital signals exist only at discrete points of time, and they can take on only finite values. A/D conversion can never be 100% accurate. Since, however, human perception does not require infinite accuracy, A/D conversion can effectively capture necessary information from the analog source for digital signal transmission.

Two steps take place in A/D conversion: a continuous time signal is first *sampled* into a discrete time signal, whose continuous amplitude is then *quantized* into a discrete level signal. First, the frequency spectrum of a signal indicates relative magnitudes of various frequency components. The **sampling theorem** (Chapter 5) states that if the highest frequency in the signal spectrum is B (in hertz), the signal can be reconstructed from its discrete samples, taken uniformly at a rate not less than $2B$ samples per second. This means that to preserve the information from a continuous-time signal, we need transmit only its samples (Fig. 1.4). However, the sample values are still not digital because they lie in a continuous dynamic range. Here, the second step of **quantization** comes to rescue. Through quantization, each sample is approximated, or "rounded off," to the nearest quantized level, as shown in Fig. 1.4. As human perception has only limited accuracy, quantization with sufficient granularity does not compromise the signal quality. If amplitudes of the message signal $m(t)$ lie in the range $(-m_p,\ m_p)$, the quantizer partitions the signal range into L intervals. Each sample amplitude is approximated by the midpoint of the interval in which the sample value falls. Each sample is now represented by one of the L numbers. The information is thus digitized. Hence, after the two steps of sampling and quantizing, the analog-to-digital (A/D) conversion is completed.

The quantized signal is an approximation of the original one. We can improve the accuracy of the quantized signal to any desired level by increasing the number of levels L. For intelligibility of voice signals, for example, $L = 8$ or 16 is sufficient. For commercial use, $L = 32$ is a minimum, and for telephone communication, $L = 128$ or 256 is commonly used.

A typical distorted binary signal with noise acquired over the channel is shown in Fig. 1.3c. If A is sufficiently large in comparison to typical noise amplitudes, the receiver can still correctly distinguish between the two pulses. The pulse amplitude is typically 5 to 10 times the rms noise amplitude. For such a high signal-to-noise ratio (SNR) the probability of error at the receiver is less than 10^{-6}; that is, on the average, the receiver will make fewer than one error per million pulses. The effect of random channel noise and distortion is thus practically eliminated. Hence, when analog signals are transmitted by digital means, some

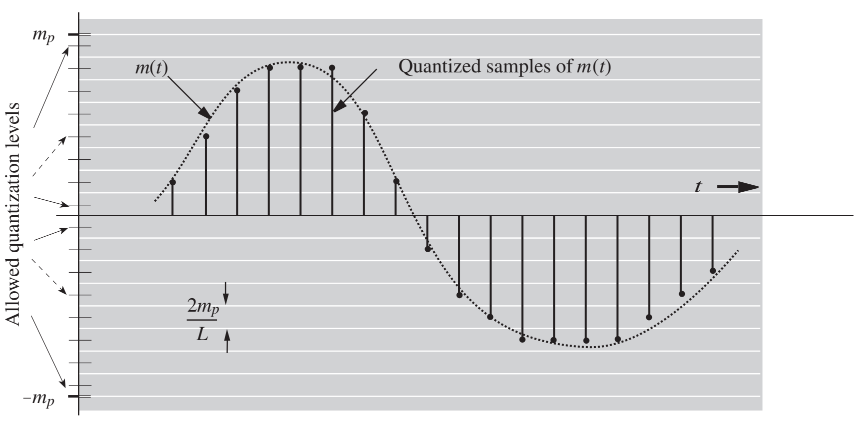


Figure 1.4 Analog-to-digital conversion of a signal.

error, or uncertainty, in the received signal can be caused by quantization, in addition to channel noise and interferences. By increasing L, we can reduce to any desired amount the uncertainty, or error, caused by quantization. At the same time, because of the use of regenerative repeaters, we can transmit signals over a much longer distance than would have been possible for the analog signal. As will be seen later in this text, the price for all these benefits of digital communication is paid in terms of increased processing complexity and bandwidth of transmission.

1.2.4 Pulse-Coded Modulation—A Digital Representation

Once the A/D conversion is over, the original analog message is represented by a sequence of samples, each of which takes on one of the L preset quantization levels. The transmission of this quantized sequence is the task of digital communication systems. For this reason, signal waveforms must be used to represent the quantized sample sequence in the transmission process. Similarly, a digital storage device also would need to represent the samples as signal waveforms. *Pulse code modulation* (PCM) is a very simple and yet common mechanism for this purpose.

First, one information *bit* refers to one *binary digit* of **1** or **0**. The idea of PCM is to represent each quantized sample by an ordered combination of two basic pulses: $p_1(t)$ representing **1** and $p_0(t)$ representing **0**. Because each of the L possible sample values can be written as a bit string of length $\log_2 L$, each sample can therefore also be mapped into a short pulse sequence that represents the binary sequence of bits. For example, if $L = 16$, then, each quantized level can be described uniquely by 4 bits. If we use two basic pulses, $p_1(t) = A/2$ and $p_0(t) = -A/2$. A sequence of four such pulses gives $2 \times 2 \times 2 \times 2 = 16$ distinct patterns, as shown in Fig. 1.5. We can assign one pattern to each of the 16 quantized values to be transmitted. Each quantized sample is now coded into a sequence of four binary pulses. This is the principle of PCM transmission, where signaling is carried out by means of only two basic pulses (or symbols). The binary case is of great practical importance because of its simplicity and ease of detection. Much of today's digital communication is binary.*

Although PCM was invented by P. M. Rainey in 1926 and rediscovered by A. H. Reeves in 1939, it was not until the early 1960s that the Bell System installed the first communication link using PCM for digital voice transmission. The cost and size of vacuum tube circuits were the chief impediments to the use of PCM in the early days before the discovery of semiconductor devices. It was the transistor that made PCM practicable.

From all these discussions on PCM, we arrive at a rather interesting (and to certain extent not obvious) conclusion—that every possible communication can be carried on with a minimum of two symbols. Thus, merely by using a proper sequence of a wink of the eye, one can convey any message, be it a conversation, a book, a movie, or an opera. Every possible detail (such as various shades of colors of the objects and tones of the voice, etc.) that is reproducible on a movie screen or on the high-definition color television can be conveyed with no less accuracy, merely by winks of an eye.†

1.3 CHANNEL EFFECT, SIGNAL-TO-NOISE RATIO, AND CAPACITY

In designing communication systems, it is important to understand and analyze important factors such as the channel and signal characteristics, the relative noise strength, the maximum number of bits that can be sent over a channel per second, and, ultimately, the signal quality.

* An intermediate case exists where we use four basic pulses (quaternary pulses) of amplitudes $\pm A/2$ and $\pm 3A/2$. A sequence of two quaternary pulses can form $4 \times 4 = 16$ distinct levels of values.

† Of course, to convey the information in a movie or a television program in real time, the winking would have to be at an inhumanly high speed. For example, the HDTV signal is represented by 19 million bits (winks) per second.

Digit	Binary equivalent	Pulse code waveform
0	**0000**	
1	**0001**	
2	**0010**	
3	**0011**	
4	**0100**	
5	**0101**	
6	**0110**	
7	**0111**	
8	**1000**	
9	**1001**	
10	**1010**	
11	**1011**	
12	**1100**	
13	**1101**	
14	**1110**	
15	**1111**	

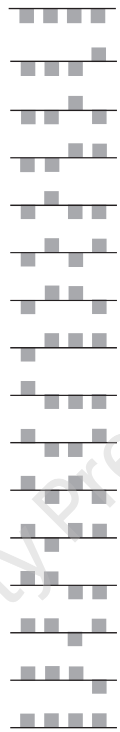


Figure 1.5 Example of PCM encoding.

1.3.1 Signal Bandwidth and Power

In a given (digital) communication system, the fundamental parameters and physical limitations that control the rate and quality are the channel bandwidth B and the signal power P_S. Their precise and quantitative relationships will be discussed in later chapters. Here we shall demonstrate these relationships qualitatively.

The **bandwidth** of a channel is the range of frequencies that it can transmit with reasonable fidelity. For example, if a channel can transmit with reasonable fidelity a signal whose frequency components vary from 0 Hz (dc) up to a maximum of 5000 Hz (5 kHz), the channel bandwidth B is 5 kHz. Likewise, each signal also has a bandwidth that measures the maximum range of its frequency components.

The faster a signal changes, the higher its maximum frequency is, and the larger its bandwidth is. Signals rich in content that changes quickly (such as those for battle scenes in a video) have larger bandwidth than signals that are dull and vary slowly (such as those for a daytime soap opera or a video of sleeping animals). A signal can be successfully sent over a channel if the channel bandwidth exceeds the signal bandwidth.

To understand the role of B, consider the possibility of increasing the speed of information transmission by compressing the signal in time. Compressing a signal in time by a factor of 2 allows it to be transmitted in half the time, and the transmission speed (rate) doubles. Time compression by a factor of 2, however, causes the signal to "wiggle" twice as fast, implying that the frequencies of its components are doubled. Many people have had firsthand experience of this effect when playing a piece of audiotape twice as fast, making the voices of normal people sound like the high-pitched speech of cartoon characters. Now, to transmit this compressed signal without distortion, the channel bandwidth must also be doubled. Thus, the rate of information transmission that a channel can successfully carry is directly proportional to B. More generally, if a channel of bandwidth B can transmit N pulses per second, then to transmit KN pulses per second by means of the same technology, we need a channel of bandwidth KB. To reiterate, the number of pulses per second that can be transmitted over a channel is directly proportional to its bandwidth B.

The **signal power** P_s plays a dual role in information transmission. First, P_s is related to the quality of transmission. Increasing P_s strengthens the signal pulse and diminishes the effect of channel noise and interference. In fact, the quality of either analog or digital communication systems varies with the SNR. In any event, a certain minimum SNR at the receiver is necessary for successful communication. Thus, a larger signal power P_s allows the system to maintain a minimum SNR over a longer distance, thereby enabling successful communication over a longer span.

The second role of the signal power is less obvious, although equally important. From the information theory point of view, the channel bandwidth B and the signal power P_s are, to some extent, exchangeable; that is, to maintain a given rate and accuracy of information transmission, we can trade P_S for B, and vice versa. Thus, one may use less B if one is willing to increase P_s, or one may reduce P_s if one is given bigger B. The rigorous proof of this will be provided in Chapter 12.

In short, the two primary communication resources are the bandwidth and the transmitted power. In a given communication channel, one resource may be more valuable than the other, and the communication scheme should be designed accordingly. A typical telephone channel, for example, has a limited bandwidth (3 kHz), but the power is less restrictive. On the other hand, in space vehicles, huge bandwidth is available but the power is severely limited. Hence, the communication solutions in the two cases are radically different.

1.3.2 Channel Capacity and Data Rate

Channel bandwidth limits the bandwidth of signals that can successfully pass through, whereas signal SNR at the receiver determines the recoverability of the transmitted signals. Higher SNR means that the transmitted signal pulse can use more signal levels, thereby carrying more bits with each pulse transmission. Higher bandwidth B also means that one can transmit more pulses (faster variation) over the channel. Hence, SNR and bandwidth B can both affect the underlying channel "throughput." The peak throughput that can be reliably carried by a channel is defined as the channel capacity.

One of the most commonly encountered channels is known as the additive white Gaussian noise (AWGN) channel. The AWGN channel model assumes no channel distortions except for the additive white Gaussian noise and its finite bandwidth B. This ideal model captures application cases with distortionless channels and provides a performance upper bound for more general distortive channels. The band-limited AWGN channel capacity was dramatically highlighted by Shannon's equation,

$$C = B \log_2(1 + \text{SNR}) \quad \text{bit/s} \tag{1.1}$$

Here the channel capacity C is the upper bound on the rate of information transmission per second. In other words, C is the maximum number of bits that can be transmitted per second with a probability of error arbitrarily close to zero; that is, the transmission is as accurate as one desires. The capacity only points out this *possibility*, however; it does not specify how it is to be realized. Moreover, it is impossible to transmit at a rate higher than this without incurring errors. Shannon's equation clearly brings out the limitation on the rate of communication imposed by B and SNR. If there is no noise on the channel (assuming SNR $= \infty$), then the capacity C would be ∞, and communication rate could be arbitrarily high. We could then transmit any amount of information in the world over one noiseless channel. This can be readily verified. If noise were zero, there would be no uncertainty in the received pulse amplitude, and the receiver would be able to detect any pulse amplitude without error. The minimum pulse amplitude separation can be arbitrarily small, and for any given pulse, we have an infinite number of fine levels available. We can assign one level to every possible message. Because an infinite number of levels are available, it is possible to assign one level to any conceivable message. Cataloging such a code may not be practical, but that is beside the point. Rather, the point is that if the noise is zero, communication ceases to be a problem, at least theoretically. Implementation of such a scheme would be difficult because of the requirement of generation and detection of pulses of precise amplitudes. Such practical difficulties would then set a limit on the rate of communication. It should be remembered that Shannon's result, which represents the upper limit on the rate of communication over a channel, would be achievable only with a system of monstrous and impractical complexity, and with a time delay in reception approaching infinity. Practical systems operate at rates below the Shannon rate.

In conclusion, Shannon's capacity equation demonstrates qualitatively the basic role played by B and SNR in limiting the performance of a communication system. These two parameters then represent the ultimate limitation on the rate of communication. The possibility of resource exchange between these two basic parameters is also demonstrated by the Shannon equation.

As a practical example of trading SNR for bandwidth B, consider the scenario in which we meet a soft-spoken man who speaks a little bit too fast for us to fully understand. This means that as listeners, our bandwidth B is too low and therefore, the capacity C is not high enough to accommodate the rapidly spoken sentences. However, if the man can speak louder (increasing power and hence the SNR), we are likely to understand him much better without changing anything else. This example illustrates the concept of resource exchange between SNR and B. Note, however, that this is not a one-to-one trade. Doubling the speaker volume allows the speaker to talk a little faster, but not twice as fast. This unequal trade effect is fully captured by Shannon's equation [Eq. (1.1)], where doubling the SNR cannot always compensate the loss of B by 50%.

1.4 MODULATION AND DETECTION

Analog signals generated by the message sources or digital signals generated through A/D conversion of analog signals are often referred to as baseband signals because they typically are lowpass in nature. Baseband signals may be directly transmitted over a suitable channel (e.g., telephone, fax). However, depending on the channel and signal frequency domain characteristics, baseband signals produced by various information sources are not always suitable for direct transmission over a given channel. When signal and channel frequency bands do not match exactly, channels cannot be moved. Hence, messages must be moved to the right channel frequency bandwidth. Message signals must therefore be further modified to facilitate transmission. In this conversion process, known as **modulation**, the baseband signal is used to modify (i.e., modulate), some parameter of a radio-frequency (RF) *carrier* signal.

A **carrier** is a sinusoid of high frequency. Through modulation, one of the carrier sinusoidal parameters—such as amplitude, frequency, or phase—is varied in proportion to the baseband signal $m(t)$. Accordingly, we have amplitude modulation (AM), frequency modulation (FM), or phase modulation (PM). Figure 1.6 shows a baseband signal $m(t)$ and the corresponding AM and FM waveforms. In AM, the carrier amplitude varies in proportion to $m(t)$, and in FM, the carrier frequency varies in proportion to $m(t)$. To reconstruct the baseband signal at the receiver, the modulated signal must pass through a reversal process called **demodulation**.

As mentioned earlier, modulation is used to facilitate transmission. Some of the important reasons for modulation are given next.

1.4.1 Ease of Radiation/Transmission

For efficient radiation of electromagnetic energy, the radiating antenna should be on the order of a fraction or more of the wavelength of the driving signal. For many baseband signals, the wavelengths are too large for reasonable antenna dimensions. For example, the power in a speech signal is concentrated at frequencies in the range of 100 to 3000 Hz. The corresponding wavelength is 100 to 3000 km. This long wavelength would necessitate an impractically large antenna. Instead, by modulating a high-frequency carrier, we effectively translate the signal spectrum to the neighborhood of the carrier frequency that corresponds to a much

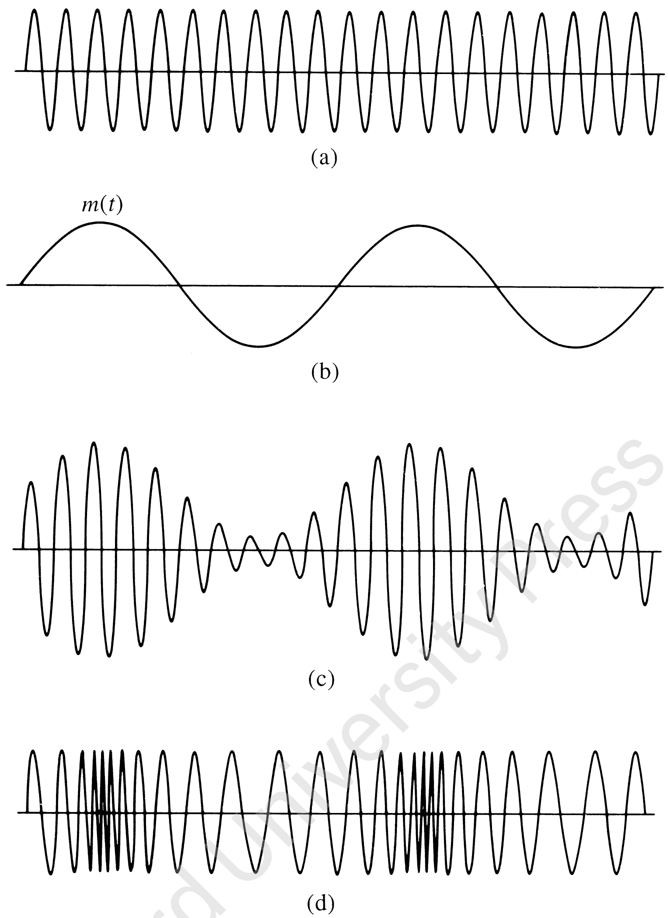


Figure 1.6 Modulation: (a) carrier; (b) modulating (baseband) signal; (c) amplitude-modulated wave; (d) frequency-modulated wave.

smaller wavelength. For example, a 10 MHz carrier has a wavelength of only 30 m, and its transmission can be achieved with an antenna size on the order of 3 m. In this respect, modulation is like letting the baseband signal hitch a ride on a high-frequency sinusoid (carrier). The carrier and the baseband signal may also be compared to a stone and a piece of paper. If we wish to throw a piece of paper, it cannot go too far by itself. But if it is wrapped around a stone (a carrier), it can be thrown over a longer distance.

1.4.2 Simultaneous Transmission of Multiple Signals—Multiplexing

Modulation also allows multiple signals to be transmitted at the same time in the same geographical area without direct mutual interference. This case in point is simply demonstrated by considering the output of multiple television stations carried by the same cable (or over the air) to people's television receivers. Without modulation, multiple video signals will all be interfering with one another because all baseband video signals effectively have the same bandwidth. Thus, cable TV or broadcast TV without modulation would be limited to one station at a time in a given location—a highly wasteful protocol because the channel bandwidth is many times larger than that of the signal.

One way to solve this problem is to use modulation. We can use various TV stations to modulate different carrier frequencies, thus translating each signal to a different frequency range. If the various carriers are chosen sufficiently far apart in frequency, the spectra of the modulated signals (known as TV channels) will not overlap and thus will not interfere with each other. At the receiver (TV set), a tunable bandpass filter can select the desired station or TV channel for viewing. This method of transmitting several signals simultaneously, over nonoverlapping frequency bands, is known as **frequency division multiplexing (FDM)**. A similar approach is also used in AM and FM radio broadcasting. Here the bandwidth of the channel is shared by various signals without any overlapping.

Another method of multiplexing several signals is known as **time division multiplexing (TDM)**. This method is suitable when a signal is in the form of a pulse train (as in PCM). When the pulses are made narrower, the spaces left between pulses of one user signal are used for pulses from other signals. Thus, in effect, the transmission time is shared by a number of signals by interleaving the pulse trains of various signals in a specified order. At the receiver, the pulse trains corresponding to various signals are separated.

1.4.3 Demodulation

Once multiple modulated signals have arrived at the receiver, the desired signal must be detected and recovered into its original baseband form. Note that because of FDM, the first stage of a demodulator typically requires a tunable bandpass filter so that the receiver can select the modulated signal at a predetermined frequency band specified by the transmission station or channel. Once a particular modulated signal has been isolated, the demodulator will then need to convert the carrier variation of amplitude, frequency, or phase, back into the baseband signal voltage.

For the three basic modulation schemes of AM, FM, and PM, the corresponding demodulators must be designed such that the detector output voltage varies in proportion to the input modulated signal's amplitude, frequency, and phase, respectively. Once circuits with such response characteristics have been implemented, the demodulators can downconvert the modulated (RF) signals back into the baseband signals that represent the original source message, be it audio, video, or data.

1.5 DIGITAL SOURCE CODING AND ERROR CORRECTION CODING

As stated earlier, SNR and bandwidth are two factors that determine the performance of a given communication. Unlike analog communication systems, digital systems often adopt aggressive measures to lower the source data rate and to fight against channel noise. In particular, *source coding* is applied to generate the fewest bits possible for a given message without sacrificing its detection accuracy. On the other hand, to combat errors that arise from noise and interferences, *redundancy* needs to be introduced systematically at the transmitter, such that the receivers can rely on the redundancy to correct errors caused by channel distortion and noise. This process is known as error correction coding by the transmitter and decoding by the receiver.

Source coding and error correction coding are two successive stages in a digital communication system that work in a see-saw battle. On one hand, the job of source coding is to remove as much redundancy from the message as possible to shorten the digital message sequence that requires transmission. Source coding aims to use as little bandwidth as possible without considering channel noise and interference. On the other hand, error correction coding intentionally introduces redundancy intelligently, such that if errors occur upon detection, the redundancy can help correct the most likely errors.

Randomness, Redundancy, and Source Coding

To understand source coding, it is important to first discuss the role of *randomness* in communications. As noted earlier, channel noise is a major factor limiting communication performance because it is random and cannot be removed by prediction. On the other hand, randomness is also closely associated with the desired signals in communications. Indeed, randomness is the essence of communication. Randomness means unpredictability, or uncertainty, of a source message. If a source had no unpredictability, like a friend who always wants to repeat the same story on "how I was abducted by an alien," then the information would be known beforehand and would contain no information. Similarly, if a person winks, it conveys some information in a given context. But if a person winks continuously with the regularity of a clock, the winks convey no information. In short, a predictable signal is not random and is fully redundant. Thus, a message contains information only if it is unpredictable. Higher predictability means higher redundancy and, consequently, less information. Conversely, more unpredictable or less likely random signals contain more information.

Source coding reduces redundancy based on the predictability of the message source. The objective of source coding is to use codes that are as short as possible to represent the source signal. Shorter codes are more efficient because they require less time to transmit at a given data rate. Hence, source coding should remove signal redundancy while encoding and transmitting the unpredictable, random part of the signal. The more predictable messages contain more redundancy and require shorter codes, while messages that are less likely contain more information and should be encoded with longer codes. By assigning more likely messages with shorter source codes and less likely messages with longer source codes, one obtains more efficient source coding. Consider the Morse code, for example. In this code, various combinations of dashes and dots (code words) are assigned to each letter. To minimize transmission time, shorter code words are assigned to more frequently occurring (more probable) letters (such as e, t, and a) and longer code words are assigned to rarely occurring (less probable) letters (such as x, q, and z). Thus, on average, messages in English would tend to follow a known letter distribution, thereby leading to shorter code sequences that can be quickly transmitted. This explains why Morse code is a good source code.

It will be shown in Chapter 12 that for digital signals, the overall transmission time is minimized if a message (or symbol) of probability P is assigned a code word with a length proportional to $\log(1/P)$. Hence, from an engineering point of view, the information of a message with probability P is proportional to $\log(1/P)$. This is known as entropy (source) coding.

Error Correction Coding

Error correction coding also plays an important role in communication. While source coding removes redundancy, error correction codes add redundancy. The systematic introduction of redundancy supports reliable communication.[1] Because of redundancy, if certain bits are in error due to noise or interference, other related bits may help them recover, allowing us to decode a message accurately despite errors in the received signal. All languages are redundant. For example, English is about 50% redundant; that is, on the average, we may throw out half the letters or words without losing the meaning of a given message. This also means that in any English message, the speaker or the writer has free choice over half the letters or words, on the average. The remaining half is determined by the statistical structure of the language. If all the redundancy of English were removed, it would take about half the time to transmit a telegram or telephone conversation. If an error occurred at the receiver, however, it would be rather difficult to make sense out of the received message. The redundancy in a message, therefore, plays a useful role in combating channel noises and interferences.

It may appear paradoxical that in source coding we would remove redundancy, only to add more redundancy at the subsequent error correction coding. To explain why this is sensible, consider the removal of all redundancy in English through source coding. This would shorten the message by 50% (for bandwidth saving). However, for error correction, we may restore some systematic redundancy, except that this well-designed redundancy is only half as long as what was removed by source coding while still providing the same amount of error protection. It is therefore clear that a good combination of source coding and error correction coding can remove inefficient redundancy without sacrificing error correction. In fact, a very popular problem in this field is the persistent pursuit of *joint source-channel coding* that can maximally remove signal redundancy without losing error correction.

How redundancy can enable error correction can be seen with an example: to transmit samples with $L = 16$ quantizing levels, we may use a group of four binary pulses, as shown in Fig. 1.5. In this coding scheme, no redundancy exists. If an error occurs in the reception of even one of the pulses, the receiver will produce a wrong value. Here we may use redundancy to eliminate the effect of possible errors caused by channel noise or imperfections. Thus, if we add to each code word one more pulse of such polarity as to make the number of positive pulses even, we have a code that can detect a single error in any place. Thus, to the code word **0001** we add a fifth pulse, of positive polarity, to make a new code word, **00011**. Now the number of positive pulses is 2 (even). If a single error occurs in any position, this parity will be violated. The receiver knows that an error has been made and can request retransmission of the message. This is a very simple coding scheme. It can only detect an error; it cannot locate or correct it. Moreover, it cannot detect an even number of errors. By introducing more redundancy, it is possible not only to detect but also to correct errors. For example, for $L = 16$, it can be shown that properly adding three pulses will not only detect but also correct a single error occurring at any location. Details on the subject of error correcting codes will be discussed in Chapter 13.

1.6 A BRIEF HISTORICAL REVIEW OF MODERN TELECOMMUNICATIONS

Telecommunications (literally: communications at a distance) are always critical to human society. Even in ancient times, governments and military units relied heavily on telecommunications to gather information and to issue orders. The first type was with messengers on foot or on horseback; but the need to convey a short message over a large distance (such as one warning a city of approaching raiders) led to the use of fire and smoke signals. Using signal mirrors to reflect sunlight (heliography) was another effective way of telecommunication. Its first recorded use was in ancient Greece. Signal mirrors were also mentioned in Marco Polo's account of his trip to the Far East.[2] These ancient *visual* communication technologies are, amazingly enough, digital. Fires and smoke in different configurations would form different codewords. On hills or mountains near Greek cities there were also special personnel for such communications, forming a chain of regenerative repeaters. In fact, fire and smoke signal platforms still dot the Great Wall of China. More interestingly, reflectors or lenses, equivalent to the amplifiers and antennas we use today, were used to directionally guide the light farther.

Naturally, these early *visual* communication systems were very tedious to set up and could transmit only several bits of information per hour. A much faster visual communication system was developed just over two centuries ago. In 1793 Claude Chappe of France invented and performed a series of experiments on the concept of "semaphore telegraph." His system was a series of signaling devices called semaphores, which were mounted on towers, typically spaced 10 km apart. (A semaphore looked like a large human figure with signal flags in both hands.) A receiving semaphore operator would transcribe visually, often with the aid of a telescope, and then relay the message from his tower to the next, and so on. This visual telegraph became the government telecommunication system in France and spread to other countries, including the United States. The semaphore telegraph was eventually eclipsed by electric telegraphy. Today, only a few remaining streets and landmarks with the name "Telegraph Hill" remind us of the place of this system in history. Still, visual communications (via Aldis lamps, ship flags, and heliographs) remained an important part of maritime communications well into the twentieth century.

These early telecommunication systems are optical systems based on visual receivers. Thus, they can cover only line-of-sight distance, and human operators are required to decode the signals. An important event that changed the history of telecommunication occurred in 1820, when Hans Christian Oersted of Denmark discovered the interaction between electricity and magnetism.[3] **Michael Faraday** made the next crucial discovery, which **changed the history of both electricity and telecommunications**, when he found that electric current can be induced on a conductor by a changing magnetic field. Thus, electricity generation

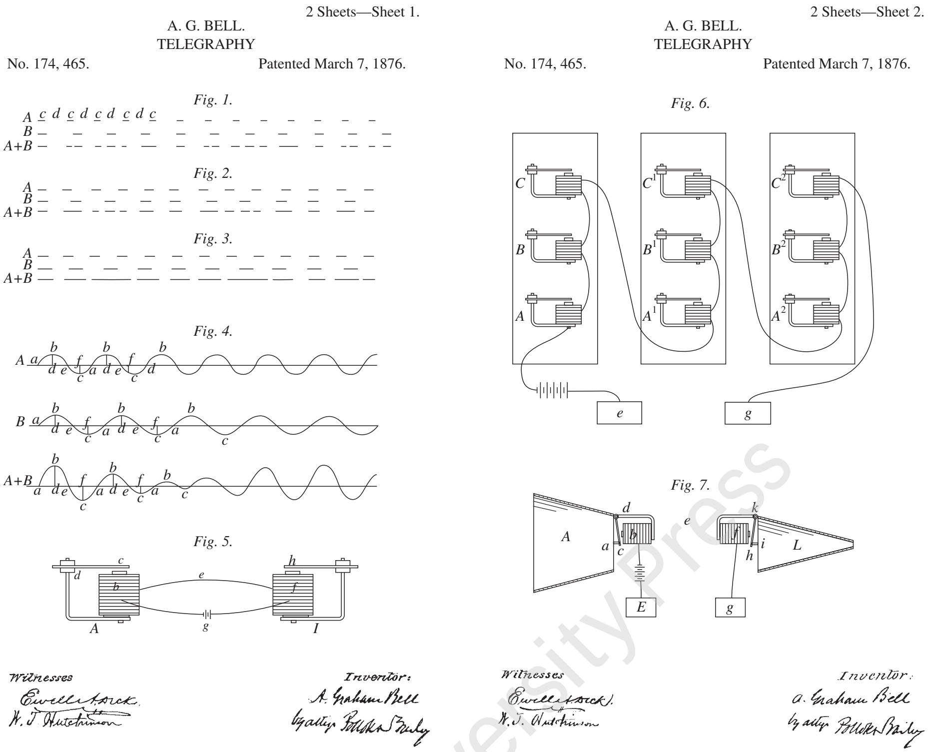


Figure 1.7 Illustration from Bell's U.S. Patent No. 174,465 issued March 7, 1876 (*Source*: U.S. Patent and Trademark Office).

became possible by magnetic field motion. Moreover, the transmission of electric signals became possible by varying an electromagnetic field to induce current change in a distant circuit. The amazing aspect of Faraday's discovery on current induction is that it provides the foundation for wireless telecommunication over distances without line-of-sight, and more importantly, it shows how to generate electricity as an energy source to power such systems. The invention of the electric telegraph soon followed, and the world entered the modern electric telecommunication era.

Modern communication systems have come a long way from their infancy. Since it would be difficult to detail all the historical events that mark the recent development of telecommunication, we shall instead use Table 1.1 to chronicle some of the most notable events in the development of modern communication systems. Since our focus is on electrical telecommunication, we shall refrain from reviewing the equally long history of optical (fiber) communications.

It is remarkable that all the early telecommunication systems are symbol-based digital systems. It was not until Alexander Graham Bell's invention of the telephone system that analog **live signals** were transmitted. Live signals can be instantly heard or seen by the receiving users. The Bell invention that marks the beginning of a new (analog communication) era is therefore a major milestone in the history of telecommunications. Figure 1.7 shows a copy of an illustration from Bell's groundbreaking 1876 telephone patent. Scientific historians often hail this invention as the *most valuable* patent ever issued in history.

The invention of telephone systems also marks the beginning of the analog communication era and live signal transmission. On an exciting but separate path, wireless communication began in 1887, when Heinrich Hertz first demonstrated a way to detect the presence of electromagnetic waves. French scientist Edouard Branly, English physicist Oliver Lodge, and Russian inventor Alexander Popov all made important contributions to the development of radio receivers. Another important contributor to this area was the

TABLE 1.1
Important Events of the Past Two Centuries of Telecommunications

Year	Major Events
1820	**First experiment of electric current causing magnetism (by Hans C. Oersted)**
1831	**Discovery of induced current from electromagnetic radiation (by Michael Faraday)**
1830–32	Birth of telegraph (credited to Joseph Henry and Pavel Schilling)
1837	Invention of Morse code by Samuel F. B. Morse
1864	Theory of electromagnetic waves developed by James C. Maxwell
1866	First transatlantic telegraph cable in operation
1876	**Invention of telephone by Alexander G. Bell**
1878	First telephone exchange in New Haven, Connecticut, United States
1887	**Detection of electromagnetic waves by Heinrich Hertz**
1896	**Wireless telegraphy (radio telegraphy) patented by Guglielmo Marconi**
1901	First transatlantic radio telegraph transmission by Marconi
1906	**First amplitude modulation radio broadcasting (by Reginald A. Fessenden)**
1907	Regular transatlantic radio telegraph service
1915	First transcontinental telephone service
1920	First commercial AM radio stations
1921	Mobile radio adopted by Detroit Police Department
1925	First television system demonstration (by Charles F. Jenkins)
1928	First television station W3XK in the United States
1935	**First FM radio demonstration (by Edwin H. Armstrong)**
1941	NTSC black and white television standard
	First commercial FM radio service
1945	Arthur C. Clarke, a science fiction writer, proposes using space satellites for communication
1947	**Cellular concept first proposed at Bell Labs**
1948	**First major information theory paper published by Claude E. Shannon**
	Invention of transistor by William Shockley, Walter Brattain, and John Bardeen
1949	The construction of Golay code for 3 (or fewer) bit error correction
1950	Hamming codes constructed for simple error corrections
1953	NTSC color television standard
1958	**Integrated circuit proposed by Jack Kilby (Texas Instruments)**
1960	**Construction of the powerful Reed–Solomon error correcting codes**
1962	First computer telephone modem developed: Bell Dataphone 103A (300 bit/s)
	Low-density parity check error correcting codes proposed by Robert G. Gallager
	First active communication satellite Telstar 1 by Telesat
	Digital time-division multiplexing demonstrated at Bell Labs
1968–69	First error correction encoders on board NASA space missions (Pioneer IX and Mariner VI)
1970	Optical fiber for communication demonstrated by Charles K. Kao and George Hockham at Corning Glass Works
1971	**First wireless computer network: AlohaNet**
1973	First portable cellular telephone demonstration to the U.S. Federal Communications Commission, by Motorola
1978	First mobile cellular trial by AT&T
1984	First handheld (analog) AMPS cellular phone service by Motorola
1989	Development of DSL modems for high-speed computer connections
1991	First (digital) GSM cellular service launched (Finland)
	First wireless local area network (LAN) developed (AT&T-NCR)
1993	Digital ATSC standard established
	Turbo codes proposed by Berrou, Glavieux, and Thitimajshima
1996	First commercial CDMA (IS-95) cellular service launched
	First HDTV broadcasting

(*continued*)

TABLE 1.1
Continued

Year	Major Events
1997	IEEE 802.11b wireless LAN standard
1998	Large-scope commercial ADSL deployment
	Cognitive radio mooted by Joseph Mitola at KTH, Sweden
1999	IEEE 802.11a wireless LAN standard
	Internet of Things (IoT) proposed by Kevin Ashton at Auto-ID Center at MIT
2000	First 3G cellular service launched
2003	IEEE 802.11g wireless LAN standard
2008	International Telecommunication Union-Radiocommunication (ITU-R) Sector specifies 4G mobile communication standards
2011	Li-Fi (wireless communication using light) is proposed by Prof. Harald Haas at the University of Edinburgh
2017	5G mobile communication standards are expected to be approved in November 2017; deployments targeted

Croatian-born genius Nikola Tesla. Building upon earlier experiments and inventions, Italian scientist and inventor Guglielmo Marconi developed a wireless telegraphy system in 1895 for which he shared the Nobel Prize in Physics in 1909. Marconi's wireless telegraphy marked a historical event of commercial wireless communications. Soon, the marriage of the inventions of Bell and Marconi allowed analog audio signals to go wireless, thanks to AM technology. Quality music transmission via FM radio broadcast was first demonstrated by American inventor Major Edwin H. Armstrong. Armstrong's FM demonstration in 1935 took place at an IEEE meeting in New York's Empire State Building.

A historic year for both communications and electronics was 1948, the year that witnessed the rebirth of digital communications and the invention of semiconductor transistors. The rebirth of digital communications is owing to the originality and brilliance of Claude E. Shannon, widely known as the father of modern digital communication and information theory. In two seminal articles published in 1948, he first established the fundamental concept of channel capacity and its relation to information transmission rate. Deriving the channel capacity of several important models, Shannon[4] proved that as long as the information is transmitted through a channel at a rate below the channel capacity, error-free communications can be possible. Given noisy channels, Shannon showed the existence of good codes that can make the probability of transmission error arbitrarily small. This noisy channel coding theorem gave rise to the modern field of error correcting codes. Coincidentally, the invention of the first transistor in the same year (by Bill Shockley, Walter Brattain, and John Bardeen) paved the way to the design and implementation of more compact, more powerful, and less noisy circuits to put Shannon's theorems into practical use. The launch of Mariner IX Mars orbiter in March of 1971 was the first NASA mission officially equipped with error correcting codes, which reliably transmitted photos taken from Mars.

Today, we are in an era of digital and multimedia communications, marked by the widespread applications of computer networking and cellular phones. The first telephone modem for home computer connection to a mainframe was developed by AT&T Bell Labs in 1962. It uses an acoustic coupler to interface with a regular telephone handset. The acoustic coupler converts the local computer data into audible tones and uses the regular telephone microphone to transmit the tones over telephone lines. The coupler receives the mainframe computer data via the telephone headphone and converts them into bits for the local computer terminal, typically at rates below 300 bit/s. Rapid advances in integrated circuits (first credited to Jack Kilby in 1958) and digital communication technology dramatically increased the link rate to 56 kbit/s by the 1990s. By 2000, wireless local area network (WLAN) modems were developed to connect computers at speed up to 11 Mbit/s. These commercial WLAN modems, the size of a credit card, were first standardized as IEEE 802.11b.

Technological advances also dramatically reshaped the cellular systems. While the cellular concept was developed in 1947 at Bell Labs, commercial systems were not available until 1983. The "mobile" phones of the 1980s were bulky and expensive, mainly used for business. The world's first cellular phone, developed by Motorola in 1983 and known as DynaTAC 8000X, weighed 28 ounces, earning the nickname of "brick" and costing $3995. These analog phones are basically two-way FM radios for voice only. Today, a cellphone is truly a multimedia, multifunctional device that is useful not only for voice communication but also can send and receive e-mail, access websites, and display videos. Cellular devices are now very small, weighing no more than a few ounces. Unlike in the past, cellular phones are now for the masses. In fact, Europe now has more cellphones than people.

Throughout history, the progress of human civilization has been inseparable from technological advances in telecommunications. Telecommunications played a key role in almost every major historical event. It is not an exaggeration to state that telecommunications helped shape the very world we live in today and will continue to define our future. It is therefore the authors' hope that this text can help stimulate the interest of many students in telecommunication technologies. By providing the fundamental principles of modern digital and analog communication systems, the authors hope to provide a solid foundation for the training of future generations of communication scientists and engineers.

REFERENCES

1. S. Lin and D. J. Costello Jr., *Error Control Coding*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 2004.
2. M. G. Murray, "Aimable Air/Sea Rescue Signal Mirrors," *The Bent of Tau Beta Pi*, pp. 29–32, Fall 2004.
3. B. Bunch and A. Hellemans, Eds., *The History of Science and Technology: A Browser's Guide to the Great Discoveries, Inventions, and the People Who Made Them from the Dawn of Time to Today*, Houghton Mifflin, Boston, MA, 2004.
4. C. E. Shannon, "A Mathematical Theory of Communications," *Bell Syst. Tech. J.*, part I: pp. 379–423; part II: pp. 623–656, July 1948.