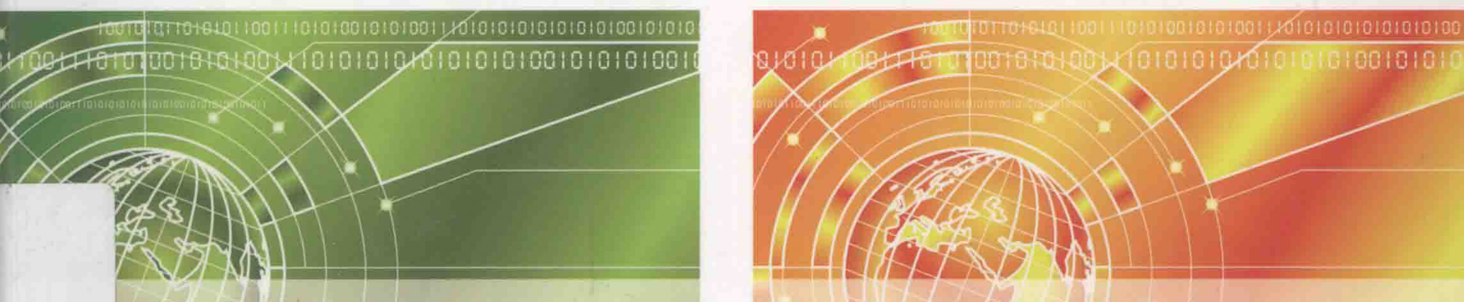
Fifth Edition

COMMUNICATION SYSTEMS

Simon Haykin   Michael Moher



INTERNATIONAL STUDENT VERSION

COMMUNICATION SYSTEMS

5th Edition

International Student Version

Simon Haykin

McMaster University

Michael Moher

Space–Time DSP

John Wiley & Sons, Inc.

ISBN: 978-0-470-16996-4

Printed in Asia

10 9 8 7 6 5 4 3 2 1

In memory of

Colin Campbell
(McMaster University, Hamilton, Ontario)

and

Michael Sablatash
(Communications Research Centre, Ottawa, Ontario)

both of whom passed away in 2008.

PREFACE

In this new edition of "Communication Systems", some major revisions have been made to the layout and contents of the book, as summarized under the following two points:

1. Emphasis has been placed on the treatment of analog communications as the necessary background for understanding digital communications.
2. The organization of the book has been modified extensively. Supplementary material that was previously present may now be found on the book's website www.wiley.com/go/global/haykin. The 10 chapters in this new edition may be summarized as follows:
   - Chapter 1 provides a short motivational introduction to communication systems.
   - Chapter 2 provides a thorough treatment of Fourier analysis of signals and systems and introduces the complex baseband representation of their band-pass versions.
   - Chapters 3 and 4 cover the theory and practical aspects of amplitude modulation and angle modulation, respectively.
   - Chapter 5 reviews aspects of probability theory and random processes that are essential to the treatment of noise in communication systems (be they of an analog or digital kind) at an introductory level.
   - Chapter 6 addresses the issue of how channel noise affects the performance of continuous-wave modulation (i.e., amplitude and angle modulation) systems.
   - In Chapter 7, we begin to shift attention from analog to digital communications by describing the issues involved in the digital representation of analog signals. In effect, this chapter represents the transition from analog to digital communications.
   - Chapter 8 introduces digital baseband communications and discusses the effect of two important impairments: noise and intersymbol interference. These two impairments are considered separately. The key assumptions in this treatment are that the noise is white, and that the channel is linear and time invariant.
   - Chapter 9 introduces digital band-pass communications. The treatment is a combination of band-pass and complex baseband analysis. The discussion of the effects of channel noise on performance shows the importance of the latter representation, in particular, its embodiment in the signal-space representation of different modulation techniques.
   - Finally, Chapter 10 provides an introductory treatment of information theory and coding. In particular, the use of coding provides a powerful tool to bring the degrading effects of channel noise in a digital communication system (i.e., errors incurred in signal detection at the receiver output) under the designer's control.

Every effort has been made to make the book readable and easy to follow in mathematical terms. Moreover, in order to provide a historical account of communication systems, each chapter includes at least one sidebar, highlighting the contribution of a pioneer who has made a significant difference to the subject matter of the chapter in question.

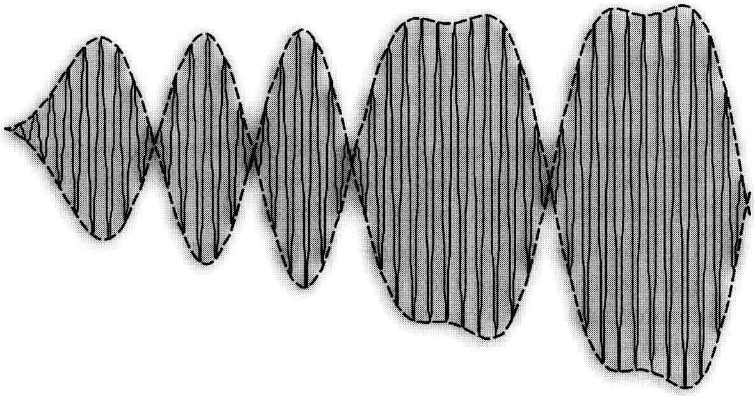
Other distinctive features of the book include Theme Examples, which are intended to focus on some practical aspects of the subject matter/theory covered therein.

Last, but by no means least, worked-out examples, computer experiments, and an abundance of end-of-chapter problems are included to strengthen a reader's understanding of the book. A Solutions Manual, obtainable from the publisher, is available only for instructors who use the book for adoption in an undergraduate course on Communication Systems.

SIMON HAYKIN

MICHAEL MOHER

CONTENTS



1

PROLOGUE

1.1 THE COMMUNICATION PROCESS

The term *communication* covers a very broad area and encompasses a large number of fields of study, ranging from the use of symbols to the social implications and effects. The meaning of the term communication in this book shall focus narrowly on the *transmission of information* from one point to another. At one time, it would have correctly been called *telecommunications*, using the Greek prefix *tele-*, meaning far. However, many modern applications of the techniques described in this book may have short ranges, such as hands-free headsets using Bluetooth or local area networks such as WiFi.

Communication in this sense enters our daily lives in so many different ways that it is ever so easy to overlook many of its facets. With telephones in our hands, radios and televisions in our living rooms, and with computer terminals providing access to the Internet in our offices and homes, we are able to communicate with every corner of the globe. Communication provides information to ships on the high seas, aircraft in flight, and rockets and satellites in space. Communication keeps a weather forecaster informed of environmental communications measured by a multitude of sensors. Indeed, the list of applications involving the use of communication in one way or another is almost endless.

HOW IS A COMMUNICATION SYSTEM ORGANIZED?

In the above sense of the word, a communication system may be divided into a small number of components as shown in Figure 1.1.

- First is the *source* of information. Some obvious examples of information that we may wish to communicate are: voice, music, pictures, videos, or data files.
- The second unshaded component in Figure 1.1 represents the *transmitter*. Transmitter is a generic term for the processing of information in the form provided by the source into a form that is suitable for transmitting over the *channel*. A simple example of this occurs when a music signal is converted to frequency modulation (FM) for radio transmission.

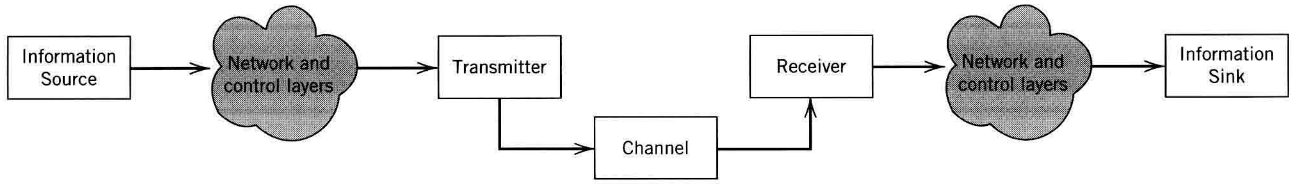


FIGURE 1.1 Elements of a communication system.

- The third unshaded component in Figure 1.1 represents the channel or transmission medium. The transmission medium may be a cable, an optical fiber, or free space if using radio or infrared communication.
- The fourth unshaded component in Figure 1.1 represents the *receiver*. Receiver is again a generic term for the process of converting the signal transmitted over the channel back to a form that may be understood at the intended destination. The receiver's function is typically greater than simply being the inverse of the transmitter; the receiver may also have to compensate for distortions introduced by the channel and perform other functions, such as the synchronization of the receiver to the transmitter.
- The final component is the destination or sink for the information.

Figure 1.1 also shows two shaded areas that are labeled *network* and *control layers*. With simple communications having one transmitter and one receiver, the network and control are likely to be absent. However, most communication systems, such as the Internet and cellular telephone systems, have a large number of transmitters and receivers that must share the same physical medium. The network and control layers permit the multiplicity of terminals to reliably and efficiently share the same physical medium.

1.2 THE LAYERED APPROACH

Modern communication systems are analyzed as a sequence of layers. This concept of layering in communication systems is best illustrated by the Open Systems Interconnect (OSI) for *computer communications*.[1] This seven-layer model is illustrated in Figure 1.2; for our purposes, it is not important that the reader understand the function of each layer in this OSI model. The important points include recognizing the left- and right-hand

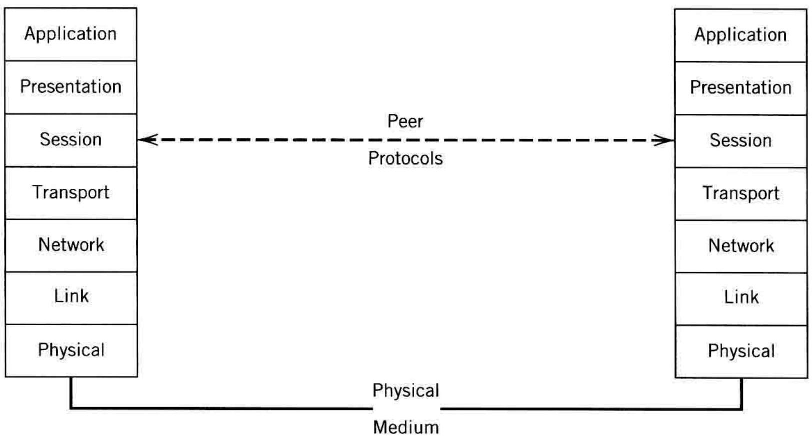


FIGURE 1.2 Peer processes in seven-layer OSI model for computer communications.

layered *stacks* of Figure 1.2, which represent two communication nodes, for example, sender and receiver. Each layer of the stack represents a *protocol*. This protocol has a well-defined interface between the layers above and below it, but the functions it carries out are only pertinent to the corresponding *peer* layer on the receiving side. Peer layers communicate virtually by sending messages down the stack on one side, across the physical medium, and up the *stack* on the other side. Only the physical layer communicates directly with its corresponding layer. In this manner, we could replace or modify the protocol at a particular layer, and not affect the rest of the OSI model.

An important attribute of the OSI model is that it simplifies the design of communication systems and permits independent development of different functions. This layered model is best suited to the communication of digital information, and less so to analog information.[2] Many digital systems use fewer than the seven layers shown in Figure 1.2.

The three central boxes of Figure 1.1, transmitter, channel, and receiver, are often referred to as the *physical layer* of the communication system or simply just the PHY. This book will focus almost entirely on the physical layer of the communications process. The network and control layers are sophisticated in their own respective ways and are the subject of other communication textbooks.

In this book, we study communication methods for analog and digital information sources. These two are often distinguished by the terms *analog communications* and *digital communications*. The term digital communications may be viewed as a misnomer. Due to practical realities, all communications are by means of continuous signals and are thus analog in nature. It is the information which is to be transmitted that has an analog or digital nature. Since most modern communications are "digital," the amount of emphasis placed on analog communications is steadily decreasing. However, some exposure to analog techniques is warranted for three reasons: (a) understanding of legacy systems; (b) many digital communication techniques are motivated from their analog counterparts; and (c) many of the distortions observed in "digital transmission" systems can be characterized as analog in nature. Most importantly, a thorough understanding of analog modulation systems leads to insight in identifying and compensating these distortions.

To summarize, this book focuses on the physical layer of the telecommunications process. With analog information, the boundary between the physical layer and other layers may be somewhat blurred.

Claude E. Shannon (1916–2001)

Shannon is known as the father of information theory, primarily due to several papers that he published in the late 1940s and early 1950s. These papers were so seminal, they effectively created the field. In 1948, he laid the theoretical foundations of digital communications in a paper entitled "A Mathematical Theory of Communications." It is noteworthy that prior to this paper's publication, it was believed that increasing the rate of information transmission over a channel would increase the probability of error. The communication theory community was taken by surprise when Shannon proved that this was not true, provided that the transmission rate was below the channel capacity.

Prior to 1948, Shannon made significant contributions to the area of digital circuit design where he is often credited with introduction of sampling theory to electrical systems, moving circuit design from the analog to the digital world. He was the first to show that Boolean algebra could be used to model and simplify the design of digital circuits.

Shannon was also famous for his hobbies of juggling, unicycling, and chess, as well as many clever inventions related to these hobbies. One invention was an electromechanical mouse named Theseus, that would search a maze to find a target. Shannon's mouse, created in 1950, appears to be first learning device of its kind.

1.3 THEME EXAMPLE—WIRELESS COMMUNICATIONS[3]

In this first theme example of the book, we consider wireless communications as an example of a communication system. Our description applies to general wireless systems but, when appropriate, we provide details of specific systems. In Figure 1.3 we show a simplified block diagram of a transmitter that consists of four major components:

- The first component of the block diagram is the *protocol stack* which we have described earlier. It packages the data so that it can reliably get to the desired destination once it crosses the radio link. In a *point-to-point* radio system or *broadcast* radio system, this component may not exist because no explicit address information is included. Indeed, many early systems worked in this manner. AM and FM radio are

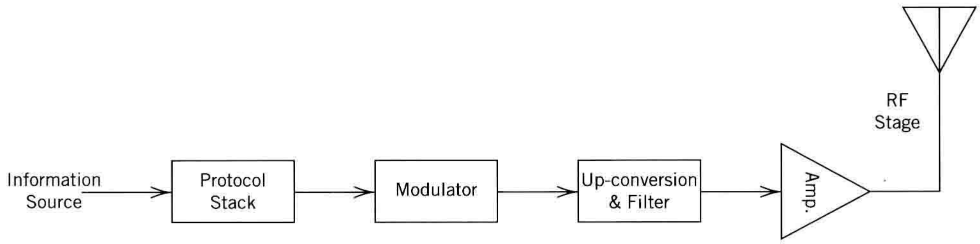


FIGURE 1.3 Illustration of basic components of a radio transmitter.

examples of wireless systems that still work in this manner. To improve the efficiency with which the radio frequencies are used, most modern systems *share radio channels* in one form or another. This *multiplexing* of multiple signals unto the same radio channel requires the use of appropriate protocols.

- The second component of the block diagram is the *modulator*. In this component, the information is impressed upon a carrier frequency in a manner that can be suitably recovered at the receiving end.
- The third component is the *up-conversion* stage. In this stage the modulated signal is converted to the final radio frequency (RF), at which it will be transmitted. A radio may be capable of transmitting at a number of frequencies, so modulating at a common frequency and converting the result to the final desired frequency is often the better approach. However, with improvements in digital signal processing and associated technology, this up-conversion stage may be replaced with a modulator that works in a manner called *direct-to-RF*.
- The fourth component is the *RF stage*. Once at the appropriate RF, the signal is amplified to an appropriate power level and then emitted via an antenna, that is, the electrical signal representing the modulated signal is converted to an electromagnetic wave. The power output will typically depend upon the desired transmission range and can vary from less than a milliwatt for short-range impulse radio applications to an effective radiated power of over a megawatt for some television transmitters. The type of antenna used will depend upon the frequency of operation and application; the possibilities includes whip, parabolic, horn, dipole, and patch antennas.

In modern systems, the modulator is typically implemented using digital signal-processing technology. This technology may be a digital signal processor, a field-programmable gate array (FPGA), or as an integrated circuit for high-volume applications. The components following the modulator are typically implemented in analog although, as mentioned previously, the digital implementation of the up-conversion stage is becoming increasingly practical.

The RF components of the radio system are often highly specific to the intended application. A handheld device typically requires a low-power amplifier and a small antenna, while a broadcast transmitter will typically be high-power and may have an antenna on a tower that is hundreds of feet high. Other systems may have power amplifiers and antennas somewhere in the middle. However, the same modulation technique could potentially be used with any of these applications. Furthermore, a well-designed up-conversion and RF stage could potentially transmit any one of a number of different modulation techniques. This is the basis of the so-called *software-defined radio*.[4] Consequently, the modulation technique is, in a sense, generic to a wide variety of applications. In the past, one of the main considerations in the choice of modulation was the ease of implementation. With the current state of technology, the main consideration is performance and the capability of the modulation to combat channel impairments, which we discuss next.

The illustration of a channel in Figure 1.4 is intended to convey a number of properties of communication channels. In particular, we have:

- *Propagation loss*. Communication usually implies transporting information over distance and inevitably there is a loss of signal strength with increasing distance. With radio channels, the fundamental loss mechanism, due to free-space propagation, causes the received power to decrease with the square of the distance from the receiver. On the other hand, with other channels, such as optical fiber, the loss of signal power only increases linearly with distance.
- *Frequency selectivity*. Communication channels operate over a medium. Many media conduct well only over a relatively small range of frequencies. For example, an optical fiber conducts a small band of optical frequencies well, but is never considered for radio waves. Even within the normal transmission band of a medium, there may be a variation in how well one frequency is transmitted compared to another. This variation is referred to as *frequency selectivity*.
- *Time-varying*. Some channels are time-varying (i.e., their characteristics vary with time). Mobile radio channels are a prime example of this phenomenon. Terrestrial radio-wave propagation depends upon the intervening terrain, buildings, and vegetation between the transmitter and the receiver. When the transmitter or receiver moves, this channel changes and affects performance; common examples of this phenomenon are known as *shadowing* and *fading*.
- *Nonlinear*. Ideally a channel should be linear in order to minimize the distortion of the transmitted signal. However, a channel may include nonlinear elements such as a repeater that includes an amplifier, operated close to or at saturation. A situation where this may occur is a satellite channel where a signal from an Earth station is amplified by the satellite before being rebroadcast over the satellite's field of view.
- *Shared usage*. To make efficient use of communications channels, they are often shared between different users. This leads to a variety of different *multiplexing* schemes that determine how the channel is shared. A common example of this is cellphone users who share the same radio channel in time and frequency in a variety of ways. Multiplexing also leads to potential *interference* between different users if the multiplexing strategies do not provide perfect isolation between users.
- *Noise*. The bane of any communication system that wants to achieve the maximum transmission distance for the minimum transmitter power is the unavoidable presence of noise. The most common source of noise is the *random motion of electrons* in receiver circuits at the point where the signal is weakest, and this usually provides a fundamental limit to performance.

All of these properties are considerations in the selection of modulation strategy. In fact, for almost any one of the above *impairments* to signal transmission, we can find a modulation strategy that has been designed to perform well in the presence of that impairment.

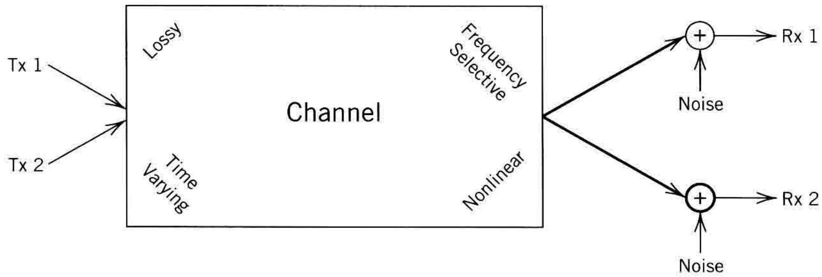


FIGURE 1.4 Illustration of channel impairments.

In practice, these impairments often occur in various combinations and the system designer must be acquainted with a variety of techniques to choose the modulation strategy best suited to the situation.

Figure 1.5 illustrates the last element of the communications link, the *receiver*. Many of the components of the receiver perform inverse functions of their counterpart in the transmitter. In particular, we have:

- *RF stage*. The antenna collects RF energy in the desired frequency band and, depending on its properties, it may collect energy from unwanted as well as wanted sources. The first amplifier in the RF stage, often called a *low-noise amplifier*, is critical to boosting the signal power to a level where it can be processed easily while minimizing the noise introduced.
- *Down-conversion*. This stage filters and translates the RF signal to a frequency where the message signal may be more easily demodulated. With many modern receivers, the signal is translated directly to *baseband*, referred to as *direct-IQ down-conversion*.
- *Demodulation*. This is the stage where the transmitted message signal is recovered. In classical receivers, demodulation often consisted of a sequence of linear filters. In modern receivers, with the advent of digital signal processing and advanced electronics, demodulation is often more complex in order to improve performance.
- *Synchronization*. Almost all communication systems require some form of synchronization circuit, due to differences between the *time* and *frequency clocks* used at the transmitter and the receiver. Depending upon the modulation and multiplexing strategy used, the methods for obtaining synchronization can be quite sophisticated. However, a circuit called the *phase-lock loop* and its variants play a fundamental role in many of these strategies.
- *Channel compensation*. The objective of this stage is to counteract some of the impairments that were encountered in the channel. While the modulation strategy may be designed to counteract a given impairment, additional processing at the receiver will often improve performance. Channel compensation techniques tend to be rather advanced and include *equalization* for frequency-selective channels and *forward error correction* for noisy channels.
- *Protocol stack*. In digital systems, it is often only at this stage where the receiver determines that the detected message was intended for it or not.

From this discussion, it is apparent that the communications receiver tends to be much more complicated than the transmitter, simply because it has more unknowns to deal with and the signal strength is much weaker than what was transmitted. Similar to the transmitter, the design of the RF and down-conversion stages is often dependent on the

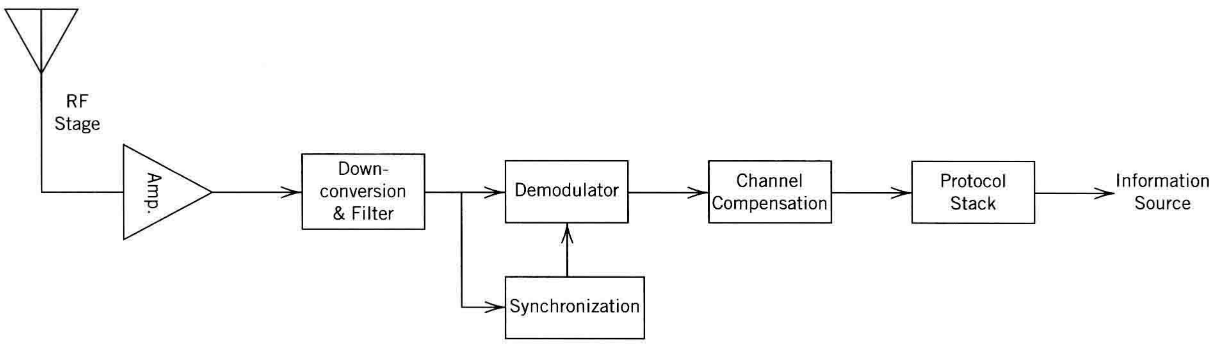


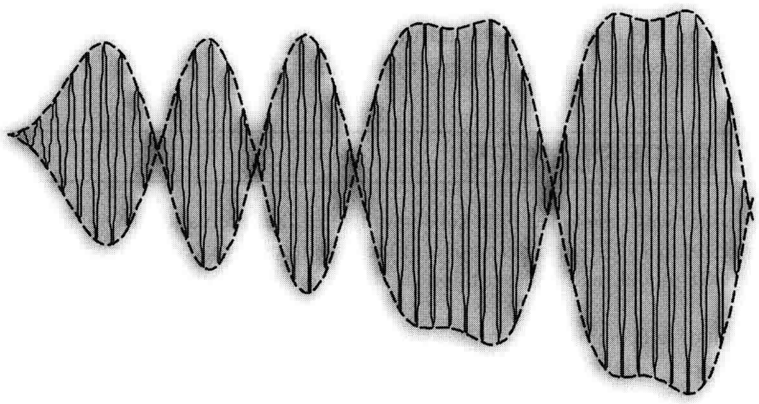
FIGURE 1.5 Illustration of radio receiver.

specific application. The choice of modulation and the corresponding demodulator are clearly the main elements for combating the impairments of the radio channel. With the advances in digital signal-processing technology, our capabilities in this area are improving. For this reason, modulation and demodulation play a key role in the study of communication systems, as presented in Chapters 3 through 7. However, before moving into the subject of modulation, we require a detailed understanding of the representation of signals and systems, which is the subject of interest in Chapter 2.

NOTES AND REFERENCES

1. The OSI reference model was developed by a subcommittee of the International Organization for Standardization (ISO) in 1977. For a discussion of the principles involved in arriving at the seven layers of the OSI model, and a description of the layers themselves, see Tanenbaum (2005).
2. For a historical account of telecommunications, see the second edition of *Introduction to Analog and Digital Communications*, Haykin and Moher (2007).
3. For further information on wireless communication, see Haykin and Moher (2005).
4. Software-defined radio (SDR) is a communication system consisting of programmable hardware under software control. Different software loads give the device different functionalities, for example, a different modulation type and different capabilities. For a detailed treatment of SDR, see the book by Reed (2002).

2



FOURIER THEORY AND COMMUNICATION SIGNALS

2.1 INTRODUCTION

We identify *deterministic signals* as a class of signals whose waveforms are defined exactly as functions of time. In this chapter we study the mathematical description of such signals using the *Fourier transform* that provides the link between the time-domain and frequency-domain descriptions of a signal. The waveform of a signal and its spectrum (i.e., frequency content) are two natural vehicles to understand the signal.

Another related issue that we study in this chapter is the representation of linear time-invariant systems. Here also we find that the Fourier transform plays a key role. Filters of different kinds and certain communication channels are important examples of this class of systems.

We begin the study by presenting a formal definition of the Fourier transform, followed by a discussion of its important properties.

2.2 THE FOURIER TRANSFORM[1]

Let $g(t)$ denote a *nonperiodic deterministic signal*, expressed as some function of time t. By definition, the *Fourier transform* of the signal $g(t)$ is given by the integral

$$G(f) = \int_{-\infty}^{\infty} g(t) \exp(-j2\pi ft)\, dt \tag{2.1}$$

where $j = \sqrt{-1}$, and the variable f denotes *frequency*. Given the Fourier transform $G(f)$, the original signal $g(t)$ is recovered exactly using the formula for the *inverse Fourier transform*:

$$g(t) = \int_{-\infty}^{\infty} G(f) \exp(j2\pi ft)\, df \tag{2.2}$$