1

Introduction to Digital Data Transmission

1.1 INTRODUCTION

This book is concerned with the transmission of information by electrical means using *digital communication techniques.* Information may be transmitted from one point to another using either digital or analog communication systems. In a digital communication system, the information is processed so that it can be represented by a sequence of discrete messages as shown in Figure 1–1. The digital source in Figure 1–1 may be the result of sampling and quantizing an analog source such as speech, or it may represent a naturally digital source such as an electronic mail file. In either case, each message is one of a finite set containing q messages. If $q = 2$, the source is referred to as a *binary source*, and the two possible digit values are called *bits*, a contraction for *binary digits.* Note also that source outputs, whether discrete or analog, are inherently random. If they were not, there would be no need for a communication system.

For example, expanding on the case where the digital information results from an analog source, consider a sensor whose output voltage at any given time instant may assume a continuum of values. This waveform may be processed by sampling at appropriately spaced time instants, quantizing these samples, and converting each quantized sample to a binary number (i.e., an analog-to-digital converter). Each sample value is therefore represented by a sequence of 1s and 0s, and the communication system associates the message 1 with a transmitted signal $s_1(t)$ and the message 0 with a transmitted signal $s_0(t)$. During each signaling interval either the message 0 or 1 is transmitted with no other possibilities. In practice, the transmitted signals $s_0(t)$ and $s_1(t)$ may be conveyed by the following means (other representations are possible):

1. By two different amplitudes of a sinusoidal signal, say, A_0 and A_1
2. By two different phases of a sinusoidal signal, say, $\pi/2$ and $-\pi/2$ radians
3. By two different frequencies of a sinusoidal signal, say, f_0 and f_1 hertz

In an analog communication system, on the other hand, the sensor output would be used directly to modify some characteristic of the transmitted signal, such as amplitude, phase, or frequency, with the chosen parameter varying over a continuum of values.

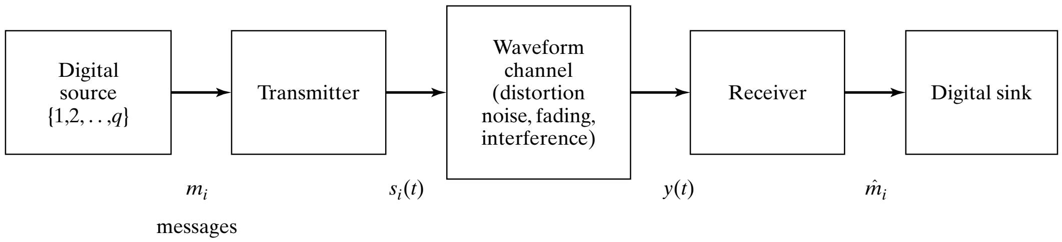

FIGURE 1–1 Simplified block diagram for a digital communication system.

Interestingly, digital transmission of information actually preceded that of analog transmission, having been used for signaling for military purposes since antiquity through the use of signal fires, semaphores, and reflected sunlight. The invention of the telegraph, a device for digital data transmission, preceded the invention of the telephone, an analog communications instrument, by more than thirty-five years.[1]

Following the invention of the telephone, it appeared that analog transmission would become the dominant form of electrical communications. Indeed, this was true for almost a century until today, when digital transmission is replacing even traditionally analog transmission areas. Several reasons may be given for the move toward digital communications:

1. In the late 1940s it was recognized that *regenerative repeaters* could be used to reconstruct the digital signal essentially *error free* at appropriately spaced intervals.[2] That is, the effects of noise and channel-induced distortions in a digital communications link can be almost completely removed, whereas a repeater in an *analog* system (i.e., an amplifier) regenerates the noise and distortion together with the signal.
2. A second advantage of digital representation of information is the flexibility inherent in the processing of digital signals.[3] That is, a digital signal can be processed independently of whether it represents a discrete data source or a digitized analog source. This means that an essentially unlimited range of signal conditioning and processing options is available to the designer. Depending on the origination and intended destination of the information being conveyed, these might include *source coding, compression, encryption, pulse shaping* for spectral control, *forward error correction (FEC) coding*, special modulation

[1]The telegraph was invented by Samuel F. B. Morse in the United States and by Sir Charles Wheatstone in Great Britain in 1837, and the first public telegram was sent in 1844. Alexander Graham Bell invented the telephone in 1876.

[2]See [1] in the references at the end of the chapter.

[3]An excellent overview of terminology, ideas, and mathematical descriptions of digital communications is provided in an article by Ristenbatt [2].

to *spread* the signal spectrum, and *equalization* to compensate for channel distortion. These terms and others will be defined and discussed throughout the book.

3. The third major reason for the increasing popularity of digital data transmission is that it can be used to exploit the cost effectiveness of digital integrated circuits. Special-purpose digital signal-processing functions have been realized as large-scale integrated circuits for several years, and more and more modem[4] functions are being implemented in ever smaller packages (e.g., the modem card in a laptop computer). The development of the microcomputer and of special-purpose programmable digital signal processors mean that data transmission systems can now be implemented as *software.*[5] This is advantageous in that a particular design is not "frozen" as hardware but can be altered or replaced with the advent of improved designs or changed requirements.
4. A fourth reason that digital transmission of information is the format of choice in a majority of applications nowadays is that information represented digitally can be treated the same regardless of its origin, as already pointed out, but more importantly easily intermixed in the process of transmission. An example is the Internet, which initially was used to convey packets or files of information or relatively short text messages. As its popularity exploded in the early 1990s and as transmission speeds dramatically increased, it was discovered that it could be used to convey traditionally analog forms of information, such as audio and video, along with the more traditional forms of packetized information.

In the remainder of this chapter, some of the systems aspects of digital communications are discussed. The simplified block diagram of a digital communications system shown in Figure 1–1 indicates that any communications system consists of a *transmitter*, a *channel* or transmission medium, and a *receiver.*[6]

To illustrate the effect of the channel on the transmitted signal, we return to the binary source case considered earlier. The two possible messages can be represented by the set {0, 1} where the 0s and ls are called bits (for binary digit) as mentioned previously. If a 0 or a 1 is emitted from the source every T seconds, a 1 might be represented by a voltage pulse of A volts T seconds in duration and a 0 by a voltage pulse of $-A$ volts T seconds in duration. The transmitted waveform appears as shown in Figure 1–2a. Assume that noise is added to this waveform by the channel that results in the waveform of Figure 1–2b. The receiver consists of a filter to remove some of the noise followed by a sampler. The filtered output is shown in Figure 1–2c and the samples are shown in Figure 1–2d. If a sample is greater than 0, it is decided that A was sent; if it is less than 0 the decision is

[4] A contraction of modulator/demodulator. *See* J. Sevenhans, B. Verstraeten, and S. Taraborrelli, "Trends in Silicon Radio Large Scale Integration," IEEE Commun. Mag., Vol. 38, pp. 142–147, Jan. 2000 for progress in IC realization of radio functions.

[5] See the *IEEE Communications Magazine* special issue on software radios [3].

[6] This block diagram suggests a *single link* communications system. It is often the case that communication systems are *many-to-one*, *one-to-many*, or *many-to-many* in terms of transmitters (sources) and receivers (sinks).

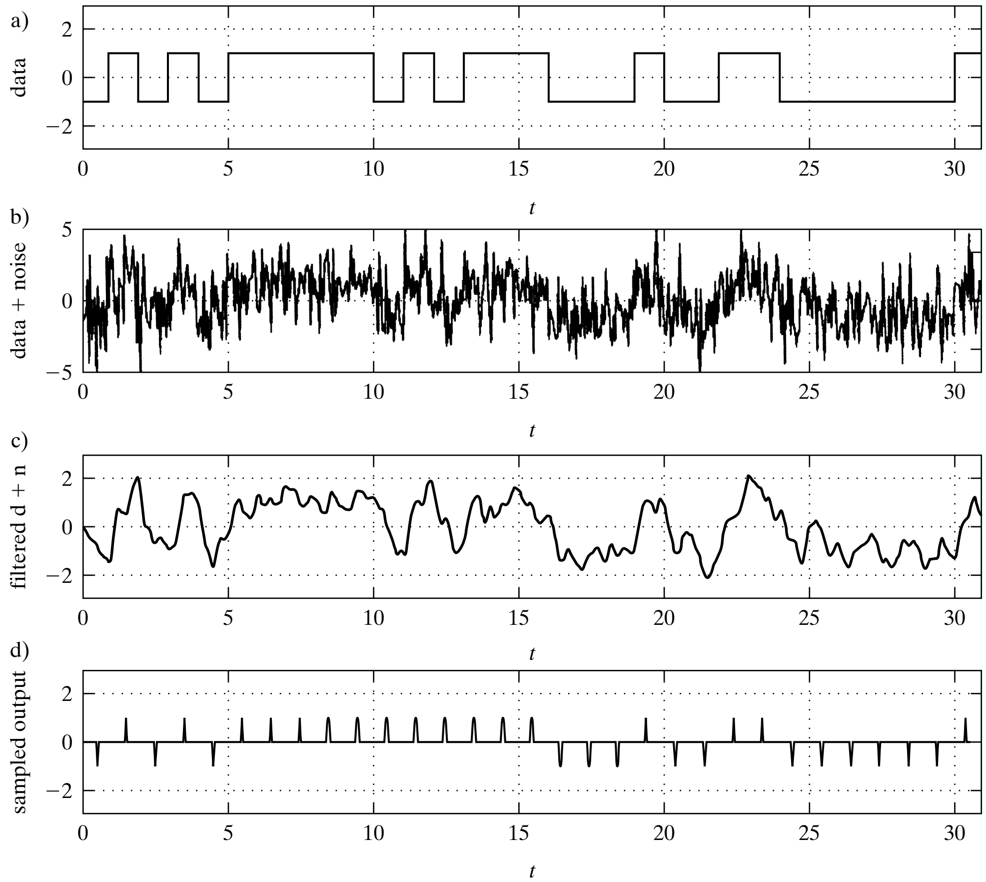

FIGURE 1–2 Typical waveforms in a simple digital communication system that uses a filter/sampler/thresholder for a detector: (a) undistorted digital signal; (b) noise plus signal; (c) filtered noisy signal; (d) hard-limited samples of filtered noisy signal—decision = 1 if sample > 0 and –1 if sample < 0. Note the errors resulting from the fairly high noise level.

that a $-A$ was sent. Because of the noise added in the channel, errors may be made in this decision process. Several are evident in Figure 1–2 upon comparing the top waveform with the samples in the bottom plot. The synchronization required to sample at the proper instant is no small problem, but will be considered to be carried out ideally in this example.

In the next section, we consider a more detailed block diagram than Figure 1–1 and explain the different operations that may be encountered in a digital communications system.

1.2 COMPONENTS OF A DIGITAL COMMUNICATIONS SYSTEM

The mechanization and performance considerations for digital communications systems will now be discussed in more detail. Figure 1–3 shows a system block diagram that is more detailed than that of Figure 1–1. The functions of all the blocks of Figure 1–3 are discussed in this section.

1.2.1 General Considerations

In most communication system designs, a general objective is to use the resources of bandwidth and transmitted power as efficiently as possible. In many applications, one of these resources is scarcer than the other, which results in the classification of most channels as either bandwidth limited or power limited. Thus we are interested in both a transmission scheme's *bandwidth efficiency*, defined as the ratio of data rate to signal bandwidth, and its *power efficiency*, characterized by the probability of making a reception error as a function of signal-to-noise ratio. We give a preliminary discussion of this power-bandwidth efficiency trade-off in Section 1.2.3. Often, secondary restrictions may be imposed in choosing a transmission method, for example, the waveform at the output of the data modulator may be required to have certain properties in order to accommodate nonlinear amplifiers such as a traveling-wave tube amplifier (TWTA).

1.2.2 Subsystems in a Typical Communication System

We now briefly consider each set of blocks in Figure 1–3, one at the transmitting end and its partner at the receiving end. Consider first the source and sink blocks. As previously discussed, the discrete information source can be the result of desiring to transmit a natu-

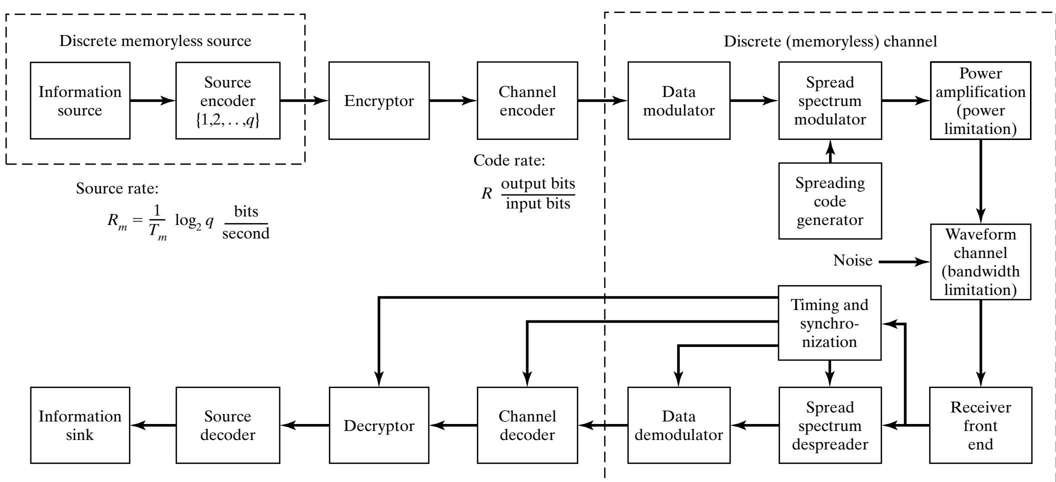

FIGURE 1–3 Block diagram of a typical digital communication system.

rally discrete alphabet of characters or the desire to transmit the output of an analog source digitally. If the latter is the case, the analog source, assumed lowpass of bandwidth W hertz in this discussion, is sampled and each sample quantized. In order to recover the signal from its samples, according to the sampling theorem (Chapter 2), the sampling rate f_s must obey the Nyquist criterion, which is[7]

$$f_s \geq 2W \text{ samples/second} \tag{1–1}$$

Furthermore, if each sample is quantized into q levels, then $\log_2 q$ bits are required to represent each sample value and the minimum source rate in this case is

$$R_m = (f_s)_{\min} \log_2 q = 2W \log_2 q \text{ bits/second} \tag{1–2}$$

Consider next the source encoder and decoder blocks in Figure 1–3. Most sources possess *redundancy*, manifested by dependencies between successive symbols or by the probabilities of occurrence of these symbols not being equal, in their outputs. It is therefore possible to represent a string of symbols, each one being selected from an alphabet of q symbols, from the output of a redundant source by fewer than $\log_2 q$ bits per symbol *on the average*. Means for doing so will be discussed in Chapter 6. Thus the function of the source encoder and decoder blocks in Figure 1–3 is to remove redundancy before transmission and decode the reduced-redundancy symbols at the receiver, respectively.

It is often desirable to make the transmissions *secure* from unwanted interceptors. This is the function of the encryptor and decryptor blocks shown in Figure 1–3. This is true not only in military applications, but many civilian applications as well (consider the undesirability, for example, of a competitor learning the details of a competing bid for a construction project that is being sent to a potential customer by means of a public carrier transmission system). Although much of the literature on this subject is classified, [5] provides an excellent overview.

In many communications systems, it might not be possible to achieve the level of transmission reliability desired with the transmitter and receiver parameters available (e.g., power, bandwidth, receiver sensitivity, and modulation[8] technique). A way to improve performance in many cases is to encode the transmitted data sequence by adding redundant symbols and using this redundancy to detect and correct errors at the receiver output. This is the function of the channel encoder/decoder blocks shown in Figure 1–3. It may seem strange that redundancy is now added after removing redundancy with the source encoder. This is reasonable, however, since the channel encoder *adds controlled redundancy*, which the channel decoder makes use of to correct errors, whereas the redundancy removed by the source encoder is uncontrolled and is difficult to make use of in

[7]To emphasize that communication theory stands on the shoulders of many pioneers, historical references are given in this chapter from time to time; [4] is the one pertaining to Nyquist's development of sampling theory.

[8]Modulation and demodulation denote the imposing of the information-bearing signal on a carrier at the transmitter and the recovery of it at the receiver, respectively. There are several reasons for modulation, among which are ease of radiation by an antenna, the imposition of a specific band of frequencies to a given user by a regulatory body, the sharing of a common frequency resource by many users, and combatting perturbations imposed by the channel.