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.

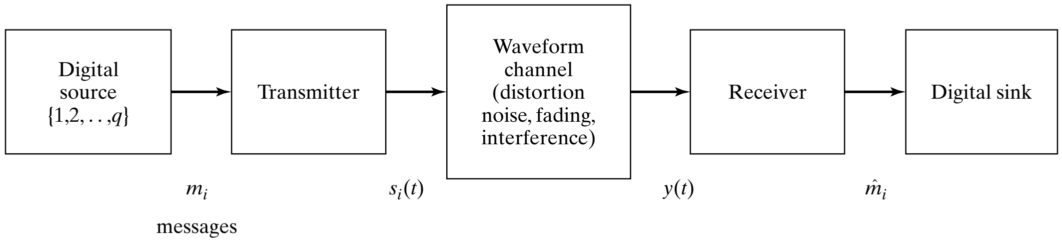

FIGURE 1–1 Simplified block diagram for a digital communication system.

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).

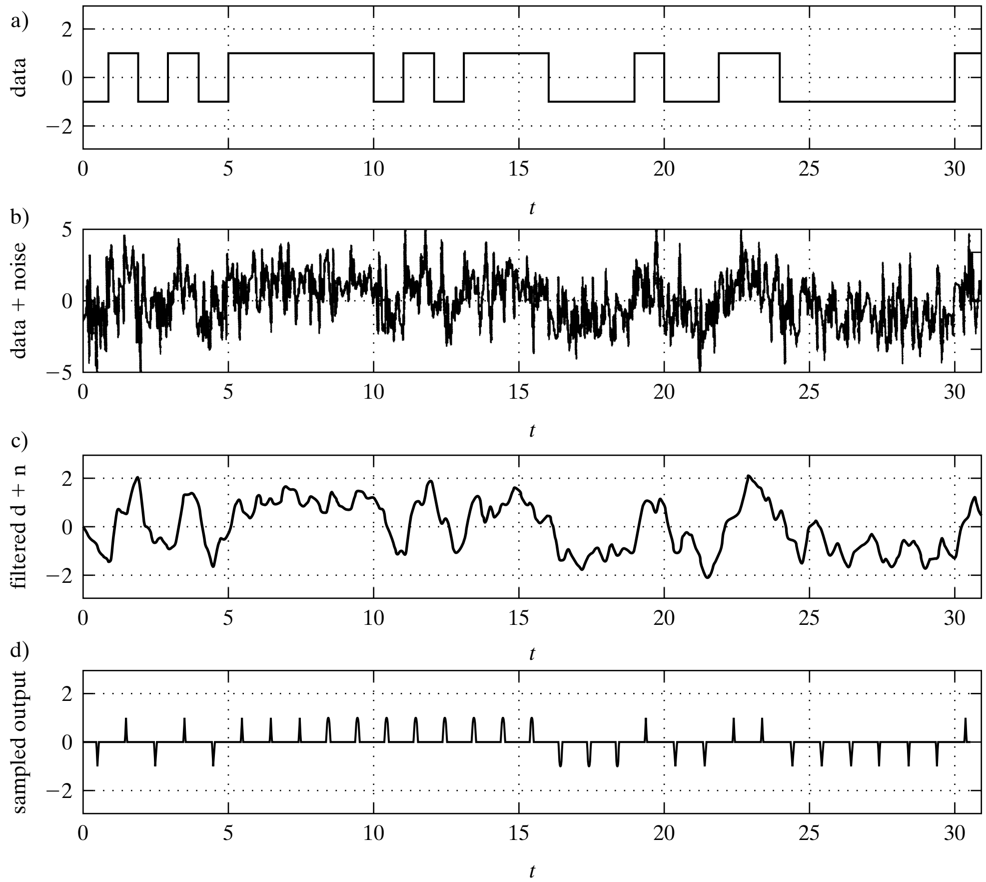

FIGURE 1–2 Typical waveforms in a simple digital communication system that uses a filter/sampler/thresholder for a detector: (a) undistorted digital signal; (b) noise plus signal; (c) filtered noisy signal; (d) hard-limited samples of filtered noisy signal—decision = 1 if sample > 0 and –1 if sample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

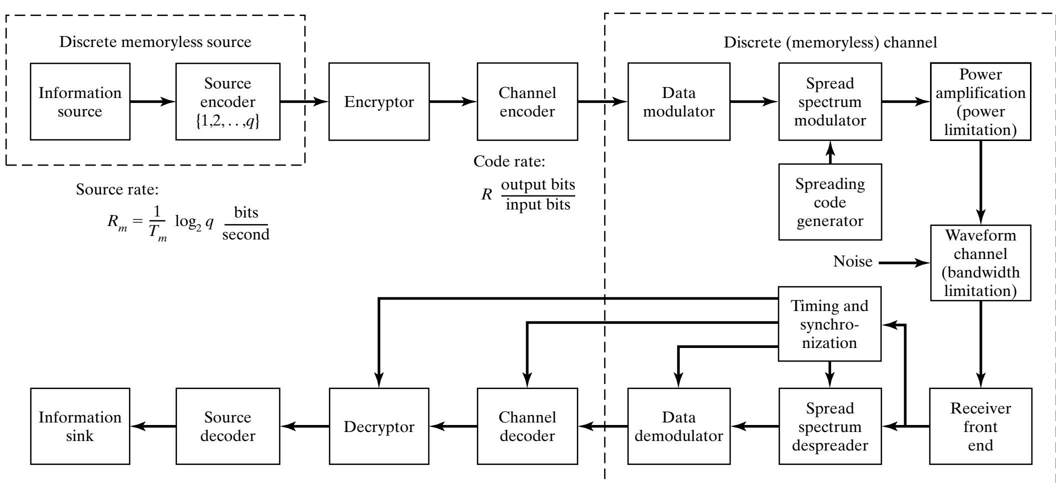

FIGURE 1–3 Block diagram of a typical digital communication system.

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.

error correction. It is therefore difficult to use it in improving the level of system transmission reliability.[9]

The data modulator produces a continuous-time waveform suitable for transmission through the channel, while the data demodulator's function is to extract the data from the received signal, now possibly distorted and noisy. The basic idea involving data detection from a distorted, noisy received signal was illustrated by the discussion given in connection with Figure 1–2. Since it is one of the main functions of this book to characterize the performances of various digital modulation schemes, we will dispense with further discussion here.

The next set of blocks, the spread-spectrum modulator and demodulator, suggests an additional level of modulation beyond the data modulation. Spread-spectrum modulation is not always employed, but there are important reasons for doing so in some cases which will be given shortly. In spread-spectrum communication system design, bandwidth efficiency is not of primary concern (an exception to this statement is when spread spectrum is being used to provide access for multiple users to the same spectrum allocation; in this case the designer wants to accommodate as many users as possible). The term *spread spectrum* refers to any modulation scheme that produces a spectrum for the transmitted signal much wider than and *independent* of the bandwidth of the information to be transmitted. There are many schemes for doing this, and some of them will be discussed in Chapter 9. Why would such a scheme be employed? Among the reasons for doing so are

1. To provide some degree of resistance to interference and jamming (i.e., intentional disruption of communications by an enemy) [referred to as *jam resistance (JR)*].
2. To provide a means for masking the transmitted signal in background noise in order to lower the probability of intercept by an adversary [referred to as *low probability of intercept (LPI)*].[10] It is important to point out that JR and LPI are not achieved simultaneously, for the former implies that one uses the maximum transmitted power available, whereas the latter implies that the power level is *just sufficient* to carry out the communication.
3. To provide resistance to signal interference from multiple transmission paths [commonly referred to as *multipath*].
4. To permit the access of a common communication channel by more than one user [referred to as *multiple access*].
5. To provide a means for measuring range or distance between two points.

[9]Both source and channel encoding are important and comprehensive subject areas with considerable research being done in both. We consider channel coding in Chapters 6 and 7. Source coding is particularly germane to vocoder design, a device that is essential in second and third generation cellular systems. References [6] and [7] provide comprehensive treatments of the subject.

[10]Two levels of security are used in secure communications: (1) *transec* refers to *transmission security* and is the type provided by spread spectrum; (2) *comsec* stands for *communications security* and is the type provided by encripting the message before transmission.

Final operations, such as power amplification and filtering to restrict the spectrum of the transmitted signal, are performed before transmission in many communications systems. Likewise, there are several preliminary operations performed in any receiver, such as amplification, mixing, and filtering. The power amplification and receiver front-end blocks shown in Figure 1–3 incorporate these functions.

The channel can be of many different types. Possibilities include twisted wire pairs, waveguides, free space, optical fiber, and so on. Further discussion of some of these will be given shortly.

1.2.3 Capacity of a Communications Link

It is useful at this point to explore briefly the concept of the *capacity* of a digital communication link. Suppose that the communications system designer is asked to design a digital communication link that transmits no more than P watts and such that the majority of the transmitted power is contained in a bandwidth W. Assume that the only effect of the channel is to add thermal noise (see Appendix B for a short discussion about thermal noise) to the transmitted signal and that the bandwidth of this noise is very wide relative to the signal bandwidth, W. The statistics of this noise are assumed Gaussian; the channel is called the *additive white Gaussian noise (AWGN)* channel. Given these constraints, there exists a maximum rate at which information can be transmitted over the link with arbitrarily high reliability. This rate is called the *error-free capacity* of a communication system. The pioneering work of Claude Shannon [8] in the late 1940s proves that signaling schemes exist such that error-free transmission can be achieved at any rate lower than capacity. Shannon showed that the normalized error-free capacity is given by

$$\frac{C}{W} = \log_2\left(1 + \frac{P}{N_0 W}\right) = \log_2\left(1 + \frac{E_b}{N_0}\frac{R}{W}\right) \text{ bits} \tag{1–3}$$

where

- C = channel capacity, bits/s
- W = transmission bandwidth, hertz
- P = $E_b R$ = signal power, watts
- N_0 = single-sided noise power spectral density, watts/hertz
- E_b = energy per bit of the received signal, joules
- R = data rate in bits/s (not to be confused with the *code rate* to be defined in Chapter 6)

More will be said later about these parameters. For now, an intuitive understanding will be sufficient. For example, the capacity, C, is the maximum rate at which *information* can be put through the channel with arbitrarily high reliability if the source is suitably matched to the channel. Its full significance requires a definition of information content of a message and how the source is to be matched to the channel; both of these topics are addressed in Chapter 6. The rate of information transfer may be conveniently expressed in bits per second (bits/s), which is the number of binary symbols that must be transmitted per second to represent a digital data sequence or to represent an analog signal with a given fidelity. These units are discussed further in Chapter 6.

An ideal communication system can be defined as one in which data is transmitted at the maximum rate $R = C$ bits/s. Thus, setting $R = C$ on the right-hand side of (1–3), we have for the ideal system that

$$\frac{C}{W} = \log_2\left[1 + \frac{E_b}{N_0}\left(\frac{C}{W}\right)\right] \tag{1–4}$$

Solving for E_b/N_0, we obtain an explicit relation between E_b/N_0 and $C/W = R/W$:

$$\frac{E_b}{N_0} = \frac{2^{C/W} - 1}{C/W} \tag{1–5}$$

The graph of this equation is shown in Figure 1–4.

If the information rate, R, at the channel input is less than C, Shannon proved that it is theoretically possible through coding to achieve error-free transmission through the

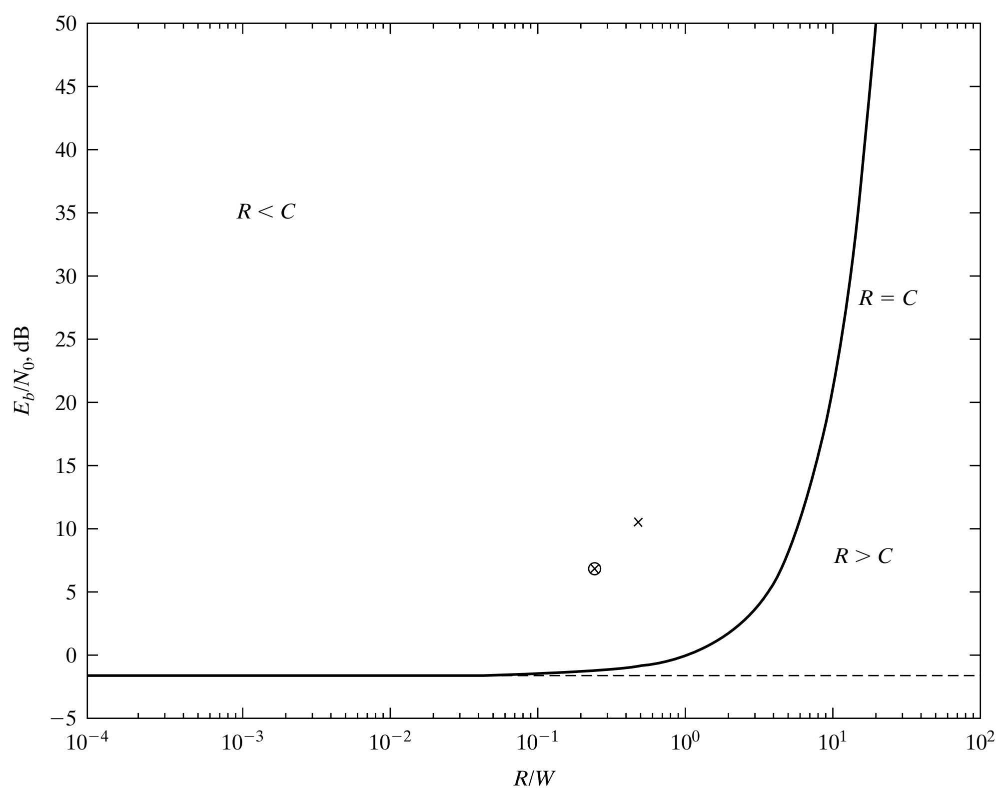

FIGURE 1–4 Power-bandwidth tradeoff for error-free transmission through noisy, bandwidth-limited channels.

channel. This result, sometimes referred to as *Shannon's second theorem,* does not provide a constructive means for finding codes that will achieve error-free transmission, but does provide a yardstick by which the performance of practical communication schemes may be measured. At points below and to the right of the curve shown in Figure 1–4, no amount of coding or complexity will achieve totally reliable transmission. At points above and to the left of the curve, error-free transmission is possible, although perhaps at a very high price in terms of bandwidth, complexity, or transmission delay. Note that data transmission is possible at all points in the plane of Figure 1–4, but some errors are unavoidable at rates above capacity.

This plot can be separated into a *bandwidth-limited region,* where $R/W > 1$, and a *power-limited region*, where $R/W < 1$. That is, if the number of bits/s/Hz is greater than unity, the scheme is efficient in terms of utilizing bandwidth. If the number of bits/s/Hz is less than unity, the scheme is efficient in terms of power utilization. For power limited operation, an interesting behavior is noted: As $R/W \rightarrow 0$ (i.e., infinite bandwidth) the limiting signal-to-noise ratio, E_b/N_0, approaches ln(2) or about –1.6 dB.[11] At any E_b/N_0 greater than –1.6 dB zero probability of making a transmission error is possible at the expense of infinite transmission bandwidth. Even more important to note, however, is that this is simply one point on the graph; *for any given rate-to-bandwidth ratio,* a *signal-to-noise ratio exists above which error-free transmission is possible and below which it is not.* Quite often, practical communication schemes are compared with this ideal by choosing some suitable probability of error, say, 10^{-6}, and finding the signal-to-noise ratio necessary to achieve it. This signal-to-noise ratio is then plotted versus R/W for the system, where W is found according to some suitable definition of bandwidth. Example 1–1 illustrates the concepts just presented.

EXAMPLE 1–1

A certain binary digital communications system can achieve an error probability of $P_E = 10^{-6}$ at an E_b/N_0 of 10.6 dB (power efficiency). Its rate-to-bandwidth ratio is approximately R/W = 1/2 bits/s/Hz (bandwidth efficiency).

(a) Assuming that $P_E = 10^{-6}$ can be viewed as error free, locate the operating point of this system on the plot of Figure 1–4.

(b) Error correction coding is now imposed on the system of part (a). For the coding scheme used, two encoded bits are sent for each source bit. The coding scheme reduces the required E_b/N_0 to achieve $P_E = 10^{-6}$ by 3.6 dB over the uncoded system. Locate the operating point for the coded scheme on Figure 1–4.

(c) Is transmission operation in the bandwidth-limited or power-limited regime?

Solution: **(a)** The point E_b/N_0 = 10.6 dB and R/W = 1/2 bits/s/Hz is shown as the $\times$ in Figure 1–4.

[11]As discussed in Section 1.4.1, a decibel (dB) is 10 times the logarithm to the base 10 of a power ratio. Although E_b/N_0 is the ratio of energy to power per hertz of bandwidth, it is dimensionally equivalent to a ratio of powers.

(b) For this part, we may subtract 3.6 dB from the value of E_b/N_0 used in part (a) to give 7 dB. However, because of the two encoded bits for each source bit, we now have R/W = 1/4 bits/s/Hz (the rate at which *information bits* are sent through the channel is kept constant). This point is shown as the ⊗ in Figure 1–4.

(c) Both schemes operate in the power-limited region. Note that plenty of room is left for improvement in either case.

1.3 COMMUNICATIONS CHANNEL MODELING

1.3.1 Introduction

The *channel* is defined as a single path for transmitting signals either in one direction only or in both directions. If a single direction only, the channel is called half duplex; if transmission can take place simultaneously in both directions, it is referred to as full duplex. The physical means by which the transmission is effected could make use of electromagnetic energy or acoustical energy, for example. If electromagnetic, the type of transmission could be further categorized as taking place in what is normally referred to as the radio spectrum (300 to 3×10^{11} Hz) or in the infrared, visible, or ultraviolet regions. Furthermore, the type of propagation can be guided or free space. Table 1–1 gives a listing of various electromagnetic spectrum bands along with typical applications.

With all these possibilities, it is difficult to say anything of a generally applicable nature that applies to all modes of transmission. Indeed, the mode of transmission employed determines, to a large degree, the perturbations that the transmitted signal experiences in passing through the channel.

1.3.2 Specific Examples of Communication Channels

1.3.2.1 Propagation Channels [9]. Perhaps the type of channel that comes to mind first when discussing communication systems is what will be referred to as the propagation channel. It is worthwhile to point out a few characteristics and current uses. Noise in propagation channels varies in nature and intensity with frequency. At low frequencies, the normal thermal noise of the electronic devices used in the communication subsystems is enhanced by environmental and human-made noise intercepted by the antenna, such as atmospheric noise, or spherics, from lightning discharges, power line corona, and commutator noise [10]. At higher frequencies, depending on the gain characteristics and pointing of the antenna, the communication subsystem thermal noise is often accompanied by galactic noise from our solar system and others. In addition, rain will enhance the noisiness of a communication system through the scattering of the electromagnetic waves from the raindrops, particularly at very high frequencies, where rain greatly attenuates the propagating signal as well.

Radio systems have found a myriad of applications over the 100-plus years that they have been in existence. Three relatively recent application areas are mentioned here. The most obvious one is perhaps that of *cellular mobile radio,* or its more inclusive and

TABLE 1–1 Frequency Bands and Communications Applications

Frequency	Wavelength	Band Designation	Typical Applications
3–30 Hz	10^4–10^5 km	ELF	Survivable communications (military)
30–300 Hz	10^3–10^4 km	SLF	Survivable communications (military)
300–3000 Hz	10^2–10^3 km	ULF	Survivable communications (military)
3–30 kHz	10–100 km	VLF	Survivable communications (military) Sonar (acoustic) Omega navigation (10–14 kHz)
30–300 kHz	1–10 km	LF	Loran C navigation (100 kHz) Amateur radio
300–3000 kHz	0.1–1 km	MF	Commercial AM radio (0.54 – 1.6 MHz)
3–30 MHz	10–100 m	HF	Commercial communications Over-the-horizon radar Citizens band radio Amateur radio
30–300 MHz	1–10 m	VHF	Citizens band radio Television (54–88 MHz, Ch. 2–6; 174–216 MHz, Ch. 7–13) Commercial FM (88–108 MHz) Navigational aids: VOR/ILS Military radio: SINCGARS; HAVE QUICK Phased-array radars Air-ground communications
300–3000 MHz	0.1–1 m	UHF	Air-ground communications Television (420–890 MHz, Ch. 14–83) Navigational aids Cellular radio Tactical air surveillance and control Global positioning systems IFF/TACAN/JTIDS Common carrier microwave
3–30 GHz	1–10 cm	SHF	Common carrier microwave Radio navigation Satellite television Precision approach radar Advanced Communication Technology Satellite Airborne fire control and navigation radar
30–300 GHz	0.1–1 cm	EHF	Millimeter wave seeker/sensor Navy auto carrier landing Artillery location radar Strategic satellite communications
300 GHz–3 THz	0.1–1 mm		Experimental
	7–3 μm	Mid IR	Far infrared at upper end
	3–0.7 μm	Near IR	Laser communications
	0.7–0.4 μm	Visible light	Laser communications
	0.4–0.1 μm	Ultraviolet	

Key: E = extremely; S = super; U = ultra; V = very; L = low; M = medium; H = high; μm = micron = 10^{-6} meters; k = kilo = $\times 10^3$; M = mega = $\times 10^6$; G = giga = $\times 10^9$; T = tera = $\times 10^{12}$.

more recent cousin, *personal communications systems.* Such systems are the topic of Chapter 10.

A by now familiar type of propagation channel is the satellite communications channel. Satellite communications is the subject of Chapter 11. In the mid-1970s, it was thought that all of the exciting applications of communication satellites had been thought of and, hence, further research on satellite communications was unnecessary. This was thought to be even more the case once optical fibers had been extensively laid under the oceans, since the major commercial application of satellite communication systems up to that time had been for long-haul communications. With the installation of the very wideband and low-cost (relative to satellite systems) optical fibers, therefore, it was argued that the need for long-haul satellite communications was essentially dead, except for a few remote and/or low population regions of the earth. However, with the advent of cellular mobile radio systems, it was thought by many research groups that the next logical development in the quest for communications anywhere, anytime should be *satellite mobile personal communications systems.* Several industrial concerns are actively developing these systems, which involve constellations of several satellites (a few tens to hundreds) in low- or medium-earth orbits capable of relaying conversations (and later perhaps video and data) between two arbitrary points on the earth's surface using hand-held devices (i.e., telephones, computers, personal digital assistants or PDAs, etc.). The exact means for accomplishing this vary widely depending on the system being developed. For example, some depend on the use of the terrestrial telephone system and some do not.

Yet another relatively recent development in satellite communications is that of very small aperture earth terminals. These are characterized by the small dishes mounted on the sides of dwellings for television delivery. The advantages of this development for television access are economy (after the initial investment), much greater programming variety, and not having to put up with unsightly large dishes as in the past. Also appearing on the horizon are satellite systems for providing Internet connections for hard-to-access locations.

1.3.2.2 Land Line [11, 12]. Following the propagation channel the most obvious communication medium, perhaps, is land line. The most geographically pervasive example of this is the telephone system. A few years ago, we could have referred to it as wire line, but more and more of the telephone plant is being replaced by optical fibers, which have tremendously more bandwidth than the original wire-line form of this system. While the original form of the telephone system used analog transmission, the move now is to digital transmission for reasons already cited. The installation of more fiber communication paths means that a broader range of services is available, with perhaps the most obvious of these being the Internet. While fiber was not a requisite for development of the Internet, it is definitely more supportive of its wide geographical dispersion and the range of media becoming available over it.

Another example of a land line system, not quite so geographically pervasive as the telephone system, is cable television. The coaxial cable has wider bandwidth than the wireline twisted pair telephone system. The cable was originally intended as a single direction (simplex) connection although future developments may make much of the plant two-way (duplex) to give two-way wideband connections to homes and businesses. Cable television networks have a tree structure. When modified so that signals can be sent in the reverse direction (down the tree, so to speak), it is found that the medium is much noisier because every branch on the tree acts as a noise sensor and funnels noise back to the cable plant (toward the tree trunk). This noise ingress can be controlled by careful plant maintenance to make sure sources for it (bad connections and openings in the cable) are minimized. As with the telephone plant, more and more television cable is being replaced by fiber, although the connections between the backbone and home or business are typically coaxial cable.

1.3.2.3 Compact Disc (CD) Channels [13]. As one final example, we summarize the characteristics of a medium that perhaps, at first, one would not consider a communications medium at all. Yet, no one can deny that the CD, and its derivative, the compact disc-read only memory, or CD-ROM, have taken a leading position in storage and reproduction of software and data as well as delivery of audio and video works for entertainment. The CD was developed in the 1970s by Sony of Japan and Phillips of the Netherlands. The information is stored digitally on the disc material by etching a series of pits along a spiral path (the region between two adjacent pits is called a land) and reading from the disc by illumination with a laser beam. In read-write CDs, the recording process is accomplished by means of moderate laser beam illumination, called *write power*, that provides sufficient heat to allow modification of a layer on the disc such that, when the illuminated spot cools, crystals are formed. When illuminated again with a stronger beam at *erase power*, the material forms an amorphous layer and is then ready to be re-recorded. Reading the data on the disc is accomplished with a laser beam illumination that is weaker than either the write or read powers.[12]

For the communications engineer, the truly amazing accomplishment in development of the CD for information storage is not in the recording/playback process, however. Rather, it is in the coding and decoding done to provide for the reliable reading of data from the disc. The recording process utilizes a process called *eight-to-fourteen modulation*, which takes 8-bit bytes (a byte is eight bits) and converts them to 14-bit blocks for recording in such a manner that there are no 1s in succession. This allows a closer packing of bits in the recording process, wherein 1s are represented by the transition from a land to a pit and vice versa; 0s are then represented by the distances between intensity changes of the radiation detected when the disc is illuminated by the laser beam. The bits recorded on the disc are the code symbols of a two-step encoding process interposed between the data (or sampled audio or video) and the recording process, or between the reading process and the final data destination. Both encoding steps utilize Reed-Solomon block codes (to be considered in Chapter 6). The reason for two encoders (only one could be used with correction capabilities equivalent to the two separate ones) is to allow encoding of symbols *along and across* the recording paths, with scrambling of symbols between these two

[12]Older "Write Once Read Many" recording media employed a magnetic layer whose direction of magnetization could be modified once heated by a laser beam.

steps. The net result of this record/playback encoding process is an extremely high tolerance for errors that may be caused by imperfections in the recording medium, either through the manufacturing process or by accident, such a scratches or smudges on the medium. It is rather ironic that the development of the CD provided the most widespread application (in terms of users and units sold) of coding pioneered by Claude Shannon [14], although this was not at all clear when the CD first appeared on the market.

1.3.3 Approaches to Communication Channel Modeling

In a digital communication system, the channel can be modeled using one of two approaches, which will be referred to as (1) the *discrete channel approach* and (2) the *continuous waveform representation.* In this section, we concentrate on the second approach. The first approach will be dealt with in more detail in Chapter 6.

1.3.3.1 Discrete Channel Approach. In the discrete memoryless model of a channel, attention is focused on discrete input symbols and discrete output symbols and the set of conditional probabilities relating them.[13] The channel is memoryless if successive channel uses are independent. In this example, we assume binary input and output symbol sets. The probability relationship between these two sets can be expressed by the matrix equation

$$\begin{bmatrix} P(Y=0) \\ P(Y=1) \end{bmatrix} = \begin{bmatrix} P(Y=0 \mid X=0) & P(Y=0 \mid X=1) \\ P(Y=1 \mid X=0) & P(Y=1 \mid X=1) \end{bmatrix} \begin{bmatrix} P(X=0) \\ P(X=1) \end{bmatrix} \tag{1–6}$$

where Y is a binary-valued random variable referring to the output and X is a binary-valued random variable referring to the input. The probabilities on the left-hand side of (1–6) are called output probabilities, the ones in the square matrix are called *transition probabilities*, and the ones in the column matrix on the right-hand side are called input probabilities. Written out, the matrix equation (1–6) is equivalent to

$$\begin{aligned} \beta_1 &= p_{11}\alpha_1 + p_{12}\alpha_2 \\ \beta_2 &= p_{21}\alpha_1 + p_{22}\alpha_2 \end{aligned} \tag{1–7}$$

where

$$\begin{aligned} \beta_1 &= P(Y=0) = 1 - \beta_2 = 1 - P(Y=1) \\ \alpha_1 &= P(X=0) = 1 - \alpha_2 = 1 - P(X=1) \\ p_{11} &= P(Y=0 \mid X=0) \\ p_{12} &= P(Y=0 \mid X=1) \\ p_{21} &= P(Y=1 \mid X=0) = 1 - p_{11} \\ p_{22} &= P(Y=1 \mid X=1) = 1 - p_{12} \end{aligned}$$

[13]A model that includes memory could involve probabilities for the source symbols that are conditioned on one or more previous symbols or channel transition probabilities that are conditioned on previously transmitted symbols or both.

The last four probabilities are known as *channel transition probabilities* and can be found by knowing the channel characteristics and the receiver structure.

EXAMPLE 1–2

In this example, suppose $\alpha_1 = 0.6$ and $\alpha_2 = 0.4$. Also, let $p_{11} = p_{22} = 0.9$ and $p_{12} = p_{21} = 0.1$. A channel with equal crossover probabilities, p_{12} and p_{21}, is known as a binary symmetric channel (BSC).

(a) Find the probabilities, β_1 and β_2, of a 0 and a 1, respectively, at the output of the channel.

(b) Given that a 0 was received, what is the probability that this resulted from a 0 being sent?

Solution: **(a)** From (1–7),

$$\beta_1 = (0.9)(0.6) + (0.1)(0.4) = 0.58$$

$$\beta_2 = (0.1)(0.6) + (0.9)(0.4) = 0.42$$

(b) Using Bayes' rule (see Appendix A),

$$P(X = 0 \mid Y = 0) = \frac{P(Y = 0 \mid X = 0)\, P(X = 0)}{P(Y = 0)}$$

$$= \frac{(0.9)\,(0.6)}{0.58} = 0.93$$

where $P(Y = 0) = 0.58$ was obtained from part (a).

1.3.3.2 Waveform Description of Communication Channels. The block diagram for a simplified description of a channel at the waveform level is shown in Figure 1–5. Although not all possible perturbations on the input signal are shown, several

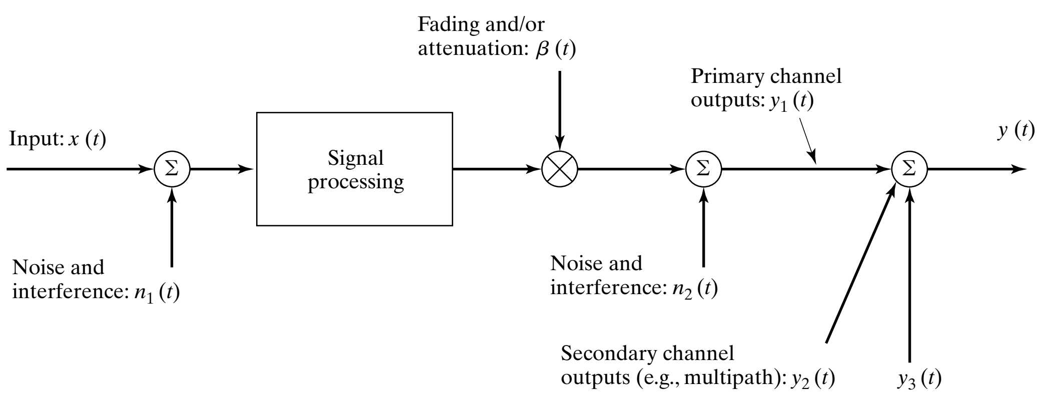

FIGURE 1–5 Channel model at the waveform level showing various perturbations.

representative ones are shown in order to discuss typical conditions that may prevail. The block diagram of Figure 1–5 is suggestive of a propagation channel, although it might be applicable to other types of channels as well.

1.3.4 Interference and Distortion in Communication Channels

There are several types of signal interferences and distortions that may arise in practical channels, as illustrated in Figure 1–5. These are enumerated here for convenience in later discussions.

1. Additive noise, which usually denotes a random waveform due to the chaotic motion of charge carriers.
2. Interference, typically denoting random waveforms due to other communication sources or human-made noise such as power line corona discharge at the channel input.
3. Deterministic signal processing, which may include linear filtering, frequency translation, and so on.
4. Multiplication by an attenuation factor, $\beta(t)$, which may be a function of time that is independent of the signal (often referred to as fading).
5. Additive noise or interference at the channel output.
6. Addition of other channel outputs that are secondary in nature (examples are multipath signals and signals from other sources that may or may not be intentional).

Note that all the perturbations shown in Figure 1–5 result in a *linear* channel as far as the input signal is concerned. That is, noise and interference terms are additive or, if multiplicative, are independent of the signal. Furthermore, the signal processing is assumed to be a linear operation.[14] Although channels that introduce nonlinear perturbations on the transmitted signal are important, the analysis of the effect of such perturbations is difficult, and many practical channels may be modeled as linear. Consequently, linear channel models are focused on in this book. Two specific types of channels will now be discussed to illustrate the applications of the general model shown in Figure 1–5.

EXAMPLE 1–3

As a first example of the application of the channel model shown in Figure 1–5, consider the block diagram of Figure 1–6, which illustrates a satellite relay link. Transmission up to the satellite is effected by a carrier of frequency f_1, and the noise $n_1(t)$ represents noise added in this portion of the transmission, which is usually due primarily to the input stages of the satellite retransmission system. The function of the satellite retransmission, or relay, is to amplify the received signal and translate it in frequency to a new spectral location suitable for

[14] A system is linear if superposition holds; i.e., if x_1 and x_2 are inputs to a system, denoted as $\mathcal{H}(\cdot)$, producing the outputs $y_1 = \mathcal{H}(x_1)$ and $y_2 = \mathcal{H}(x_2)$, superposition holds if the input $\alpha_1 x_1 + \alpha_2 x_2$ produces the output $y = \mathcal{H}(a_1 x_1 + a_2 x_2) = a_1 y_1 + a_2 y_2$.

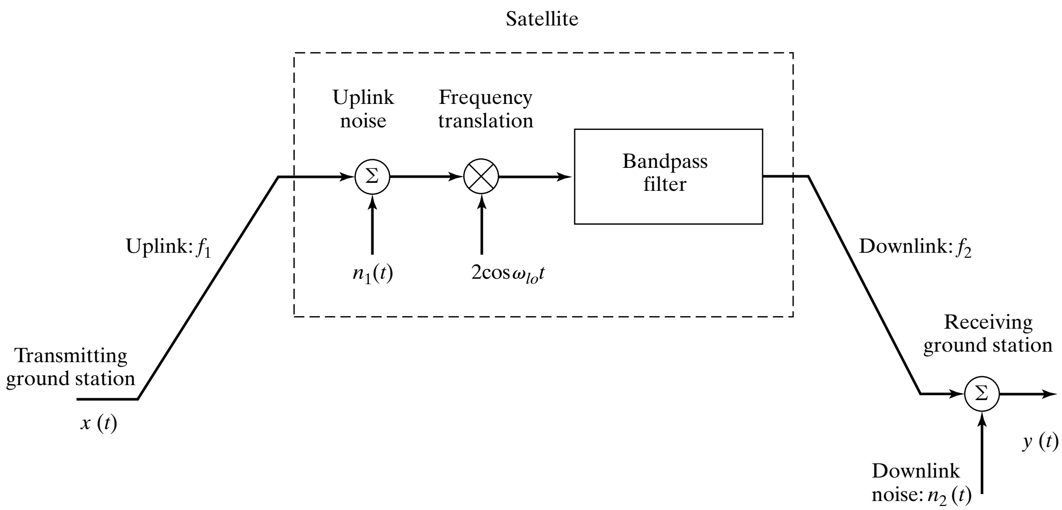

FIGURE 1–6 Model for a satellite communications link.

transmission to the destination earth station. Noise is also added in the downlink of the transmission system and is represented as $n_2(t)$. Other channel perturbations could be included, such as interference due to adjacent channels and/or signals from adjacent satellites in the same orbit. To illustrate the function of the multiplication by $2\cos\omega_{lo}t$, where the subscript *lo* stands for local oscillator, assume that the input is $x(t) = m(t)\cos 2\pi f_1 t = m(t)\cos\omega_1 t$, where $m(t)$ is a slowly varying modulating signal, then, ignoring the noise for now, the input to the bandpass filter onboard the satellite is

$$z(t) = 2m(t)\cos\omega_{lo}t\cos\omega_1 t = m(t)\cos(\omega_1 - \omega_{lo})t + m(t)\cos(\omega_1 + \omega_{lo})t$$

Depending on whether $f_2 > f_1$ or $f_2 < f_1$ is desired, the bandpass filter is centered in frequency to pass the first term or the last term. Usually, for reasons to be discussed in Chapter 11, the downlink frequency is chosen to be lower than the uplink frequency so that the downlink signal without noise is $y(t) = m(t)\cos(\omega_1 - \omega_{lo})t$; i.e., $f_2 = (\omega_1 - \omega_{lo})/2\pi$.

The additive noise and interference mentioned in Example 1–3 can fall into two possible categories: externally generated noise and noise generated internally to the communication system. Examples of the former include solar and galactic noise due to electromagnetic wave emissions from stars, including our sun, and other heavenly bodies; atmospheric noise that results primarily from electromagnetic waves generated by natural electrical discharges within the atmosphere; and human-made noise such as corona discharge from power lines. The modeling of this noise is, in general, difficult and imprecise due to its highly variable nature.

Internally generated noise is due primarily to the random motion and random production and annihilation of charge carriers within electrical components making up a

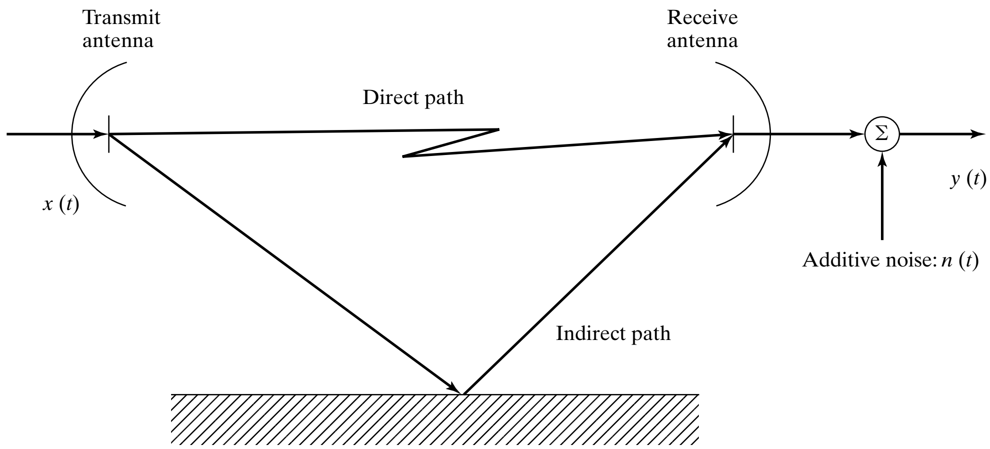

FIGURE 1–7 Model for a line-of-sight microwave relay link.

communication system. There are several treatises on the physical description and characterization of such noise (see, e.g., [10]). A short summary of its statistical description in terms of noise figure and noise temperature is given in Appendix B.

EXAMPLE 1–4

Figure 1–7 illustrates a suitable model for certain types of terrestrial microwave communications links. In addition to additive noise, which is represented by $n(t)$ in the figure, an indirect transmission path, commonly referred to as *multipath*, exists. Thus the equation relating channel input to output is

$$y(t) = a_1 x(t - T_1) + a_2 x(t - T_2) + n(t) \tag{1–8}$$

where a_1 and a_2 are constants referred to as the attenuations of the direct and indirect transmission paths, respectively, and T_1 and T_2 are their respective delays. This channel model is an extremely simple one and yet, by virtue of the multipath term, results in two signal perturbations, known as *intersymbol interference* (*ISI*) and *fading*, either one of which can introduce severe performance degradations.

To illustrate the idea of ISI, consider Figure 1–8, which illustrates received binary data signals from the direct and indirect paths after demodulation. The differential delay $\tau = T_1 - T_2$ is assumed to be less than one bit period although it could be several bit periods in duration. Actually, it would be impossible to observe these separate signals because they are received together at the antenna.[15] However, they are shown separately to illustrate that some bits destructively interfere and others reinforce each other. Unfortunately, those that destructively interfere dominate the probability of error (the average of 10^{-5}, which is a typical bit

[15]The addition of direct and multipath signals takes place at radio frequency but is shown here at baseband for simplicity.

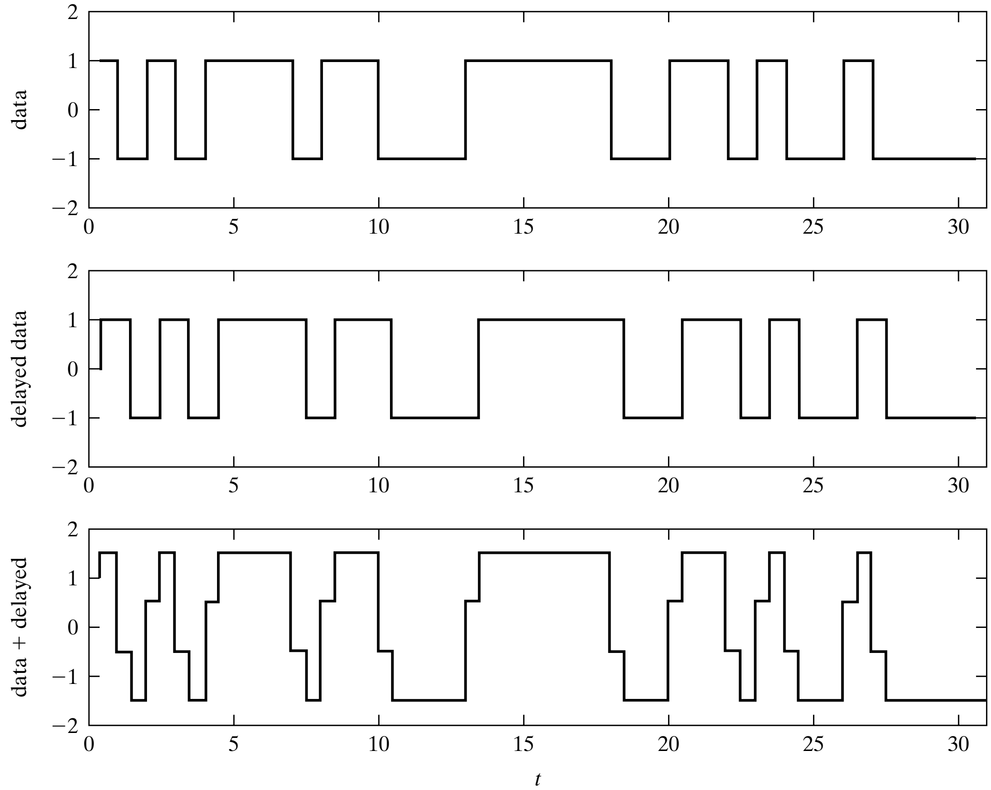

FIGURE 1–8 Waveforms for digital signal showing the effect of intersymbol interference. The undelayed bit stream is shown in the top figure and the delayed bit stream is shown in the middle figure. The delay, τ, is one-half bit. The received waveform (bottom figure) is the sum of the undelayed and one-half of the delayed bit stream.

error probability for a reinforcing situation, and of 10^{-2}, which might be typical for an interfering case, is very nearly 10^{-2}).

The phenomenon of fading is similar to that of ISI except that the destructive and constructive interference takes place with the high-frequency carrier. To illustrate the phenomenon of fading, consider a pure cosinusoidal input to the channel, which is representative of an unmodulated carrier. The output of the channel, ignoring the noise term, is

$$y(t) = A[a_1 \cos 2\pi f_0 t + a_2 \cos 2\pi f_0 (t - \tau)] \tag{1–9}$$

where the amplitude A is the amplitude of the input and where, for convenience, the time reference has been chosen such that the delay of the direct component is 0. Using suitable trigonometric identities, (1–9) can be put into the form

$$y(t) = AB(\tau)\cos[2\pi f_0 t + \theta(\tau)] \tag{1–10}$$

where

$$B(f_0;\tau) = \sqrt{a_1^2 + 2a_1a_2\cos 2\pi f_0\tau + a_2^2} \tag{1–11}$$

and

$$\theta(f_0;\tau) = -\tan^{-1}\left(\frac{a_2\sin 2\pi f_0\tau}{a_1 + a_2\cos 2\pi f_0\tau}\right) \tag{1–12}$$

Equation (1–11) shows that as the differential delay τ changes by an integer multiple of a half-carrier period, the received signal changes from a minimum amplitude of

$$AB_{\min} = |a_1 - a_2|A \tag{1–13}$$

to a maximum amplitude of

$$AB_{\max} = |a_1 + a_2|A \tag{1–14}$$

The carrier frequency in a line-of-sight terrestrial microwave link can be of the order of 10^{10} Hz = 10 GHz or higher. The wavelength at 10 GHz is

$$\lambda = \frac{c}{f_0} = 3 \text{ cm}$$

where f_0 is the carrier frequency in hertz and $c = 3 \times 10^8$ m/s is the free space speed of propagation for electromagnetic waves. Thus a change in differential path length of only 1.5 cm at a carrier frequency of 10 GHz means that conditions change from reinforcement to cancellation for the received carrier. This can impose severe system degradation. Because of this frequency dependence, such channels are referred to as *frequency selective fading*.

A plot of $B(\tau;\, f_0)$ as given by (1–11) as a function of f_0 better illustrates the frequency-dependent nature of the channel. Figure 1–9 shows that the transmitted signal will have "notches" placed in its spectrum each τ^{-1} hertz of bandwidth. The bandlimited nature of the channel is therefore apparent. Any modulated signal with bandwidth of the order of or greater than τ^{-1} hertz will suffer distortion as it propagates through the channel because of this notching effect. The differential delay between the main and secondary paths therefore imposes an upper limit on the bandwidth of the signals that the channel can support unless compensation for the notches is accomplished somehow (such compensation can be achieved partially by an *equalizer*).

Another possibility for minimizing degradation is to lengthen the signal duration so that it is much longer than the multipath delay differential. For binary signaling, this of course implies a lower data rate. There are ways to lengthen the symbol duration without accepting a lower data rate, however. In Chapter 4, one such technique, referred to as orthogonal frequency division multiplexing (OFDM), will be discussed.

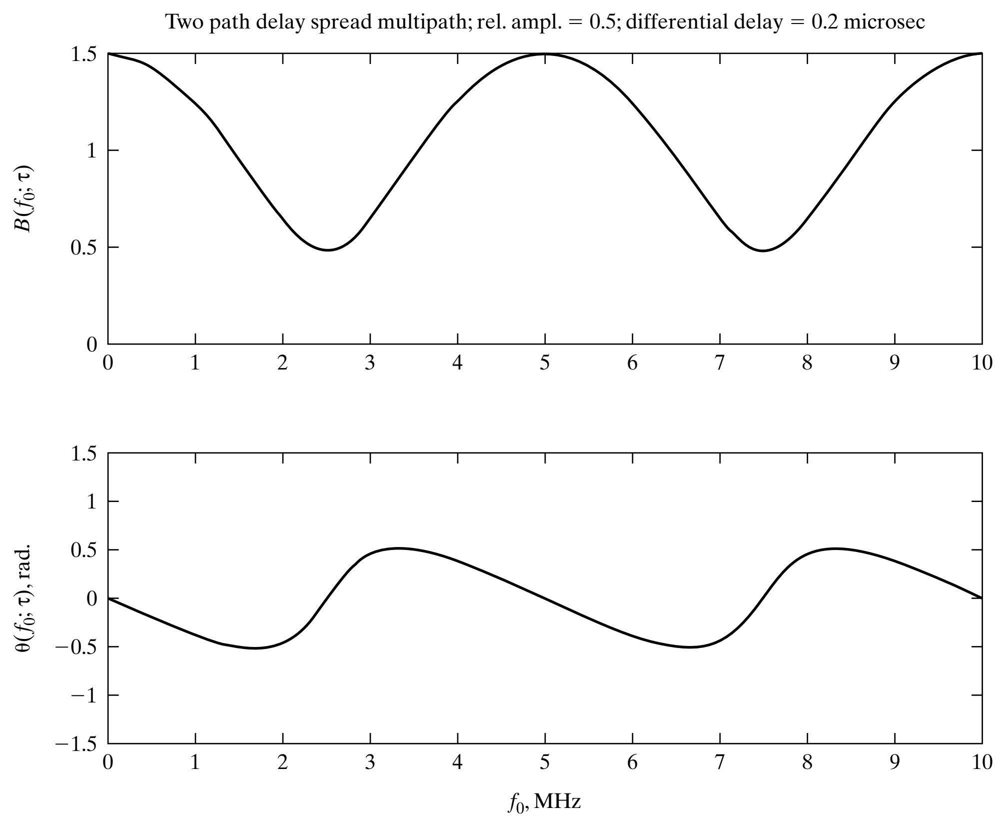

FIGURE 1–9 Received signal amplitude and phase shift versus frequency for a two-path channel.

The dual of the situation where two signals arrive at the receiver with different delays is the situation where they arrive with different Doppler shifted frequencies (in general, such a channel is called *Doppler spread*). To investigate this case, consider

$$\begin{aligned} y_{\text{dopp}}(t) &= A[a_1 \cos 2\pi f_0 t + a_2 \cos 2\pi(f_0 - f_d)t] \\ &= AC(t) \cos[2\pi f_0 t + \phi(t)] \end{aligned} \tag{1–15}$$

where it follows from appropriate algebra and trigonometry that

$$C(t) = A\sqrt{a_1^2 + 2a_1 a_2 \cos 2\pi f_d t + a_2^2} \tag{1–16}$$

and

$$\phi(t) = -\tan^{-1}\left(\frac{a_2 \sin 2\pi f_d t}{a_1 + a_2 \cos 2\pi f_d t}\right) \tag{1–17}$$

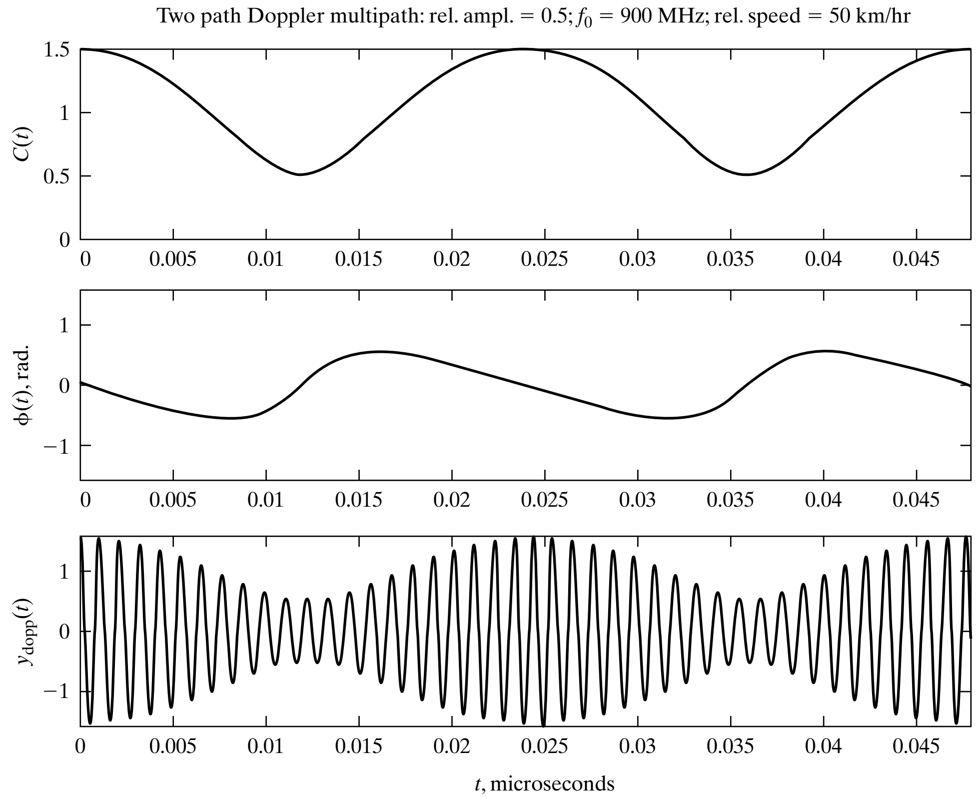

FIGURE 1–10 Time-varying envelope (top), time-varying phase (middle), and composite signal (bottom) for a two-ray Doppler spread channel.

In contrast to the case of the two sinusoidal components being received with different delays, which resulted in a frequency-dependent attenuation and phase shift of the composite signal, the received signal in this case has a time-varying envelope and phase shift as illustrated in Figure 1–10. Because of the time-varying envelope and phase of the received signal, such channels are also referred to as *time selective.*

A word is in order about the physical mechanisms that may give rise to multipath. The model illustrated in Figure 1–7 for the two-path case is, of course, an oversimplified view of the physical channel, as are all models. A two-path multipath channel would arise under conditions where the propagation path is highly stratified due to a temperature inversion or where stray reflections take place off an object, such as a building or airplane. Such multipath is referred to as *specular* because the indirect path results in a component that is essentially a mirror-like reflection, or specular, version of the direct component. When several indirect paths combine to produce a noise-like multipath component, the resulting multipath is referred to as *diffuse.* In the specular case, alleviation of the adverse effects of the multipath can be combated by employing a filter at the receiver, referred to as an equalizer, that has a frequency response function that is approximately the inverse of the frequency response of the channel. In the case of diffuse multipath, the indirect-path signal power essentially acts as a signal-dependent noise.

If destructive interference of the carrier, known as fading, is the predominant channel-induced perturbation, an obvious solution is to change carrier frequency so that reinforcement results from the multipath component. Since the exact path differential is unknown or may change slowly with time, a solution is to transmit on several frequencies simultaneously, called *frequency diversity*. Other types of diversity are also possible, such as *space diversity*, wherein several different transmission paths are used; *polarization diversity*, wherein horizontally or vertically (or counter rotating) polarized carrier waves are used, and *time diversity*, wherein the transmission of a symbol is spread out over time. Channel coding, to be discussed in Chapters 6 and 7, is a form of time diversity.

1.3.5 External Channel Propagation Considerations

When dealing with radio-wave propagation channels, several considerations pertain to conditions within the propagating medium. Often, the earth's atmosphere is involved for a portion or the entire propagation path, and its effects can be very important in design of a communication system. Several of these radio-wave propagation factors are discussed here:

1. *Attenuation* (absorption) caused by atmospheric gases. The principal gaseous constituents of the earth's atmosphere that produce significant absorption are oxygen and water vapor. The first three absorption bands are centered at frequencies of 22.2 GHz (H_2O), 60 GHz (O_2), and 118.8 GHz (O_2) [10].[16] The frequency dependence of the absorption has been found to depend on an empirical line-width constant, which is a function of temperature, pressure, and humidity of the atmosphere. Details of an empirical model to predict attenuation due to atmospheric absorption are presented in Appendix C.
2. *Attenuation* (scattering and absorption) by hydrometeors (rain, hail, wet snow, clouds, etc.). The relationship between rain rate, R (in mm/h measured at the earth's surface), and specific attenuation can be approximated by

$$\alpha = aR^b \text{ dB/km} \tag{1–18}$$

   where a and b are frequency- and temperature-dependent constants [15–17]. An empirically based method for obtaining a and b is discussed in Appendix C.
3. *Depolarization* by hydrometeors, multipath, and Faraday rotation. It is often desirable to use polarization of the transmitted electromagnetic wave as a means to separate two transmitted signals at the same carrier frequency. This is the case, for example, when diversity transmission is used to combat fading—for each carrier frequency or spatial path used it is possible to add another path through

[16]A plot of the attenuation versus frequency due to absorption reveals that these peaks are fairly broad so that significant signal attenuation takes place well on either side of these resonant frequencies.

the use of two perpendicularly polarized waves. When the propagation is through rain or ice crystals, it is possible for a small portion of an electromagnetic wave that is, say, horizontally polarized to be converted to vertical polarization, thereby creating crosstalk between the two transmissions. This effect is generally small at frequencies below 10 to 20 GHz. For further discussion of this effect, see [18].

4. *Noise emission* due to gaseous absorption and hydrometeors. The gaseous constituents of the earth's atmosphere, and clouds or precipitation when present, all act as an absorbing medium to electromagnetic waves; therefore, they are also radiation sources of thermal noise.
5. *Scintillation* (rapid variations) of amplitude and phase caused by turbulence or refractive index irregularities. Scintillation on a radio-wave path describes the phenomenon of rapid fluctuations of the amplitude, phase, or angle of arrival of the wave passing through a medium with small-scale refractive index irregularities that cause changes in the transmission path with time. Scintillation effects, often referred to as *atmospheric multipath fading*, can be produced in both the troposphere and the ionosphere.
6. *Antenna gain degradations* due to phase decorrelation across the antenna aperture. The antenna gain in a communications system is generally defined in terms of the antenna's behavior when illuminated by a uniform plane wave. Amplitude and phase fluctuations induced by the atmosphere can produce perturbations across the physical antenna aperture resulting in a reduction of total power available at the feed. The resulting effect on the antenna will look to the system like a loss of antenna gain, or a gain degradation. This gain degradation increases as the electrical receiving aperture size increases; hence the problem becomes more significant as operating frequency and/or physical aperture size increases.
7. *Bandwidth limitations* of the channel due to multipath have been illustrated by Example 1–4. The effects of such bandwidth limitations and possible cures for them will be discussed further in Chapter 3 where equalization is discussed.

1.4 COMMUNICATION LINK POWER CALCULATIONS

In this section, we look at the computation of power levels in a radio-frequency communications link. We first review the meaning of the decibel unit because of its convenience in such calculations.

1.4.1 Decibels in Communication System Performance Calculations

It is often convenient in calculations involving communications systems to use a decibel (dB) scale because of the tremendous ranges of some variables and parameters encountered. Recall that a decibel is defined as 10 times the logarithm to the base 10 of a ratio of two powers. That is,

$$R_{\mathrm{db}} = 10 \log_{10}\left(\frac{P}{P_{\mathrm{ref}}}\right) \tag{1–19}$$

where P and P_{ref} are two powers having the same units (e.g., watts, milliwatts, etc.). Since power is proportional to the magnitude of a voltage phasor squared, we can also write

$$R_{\mathrm{dB}} = 10 \log_{10}\left(\frac{|V|^2}{|V_{\mathrm{ref}}|^2}\right) = 20 \log_{10}\left(\frac{|V|}{|V_{\mathrm{ref}}|}\right) \tag{1–20}$$

In (1–19), suppose that P_{ref} is a reference level of 1 watt. Then, a power level of P referenced to 1 watt, termed P dBW, is defined by (1–19) with P_{ref} set equal to 1 watt:

$$P_{\mathrm{dBW}} = 10 \log_{10} P \tag{1–21}$$

Sometimes the reference level of 1 milliwatt is used. The resulting power level is then in units of dBm (decibels referenced to one milliwatt). It follows that

$$P_{\mathrm{dBm}} = P_{\mathrm{dBW}} + 30 \tag{1–22}$$

since 1 watt = 10^3 milliwatts and $10 \log_{10}(10^3) = 30$.

It is also sometimes convenient to express bandwidth, temperature, or other quantities with dimensions in terms of decibels. If this is done, the dimension of the referenced quantity is simply attached to the decibel. For example, a 1 MHz bandwidth can be expressed as

$$B_{\mathrm{dB\text{-}Hz}} = 10 \log_{10}\left(\frac{1 \times 10^6\ \mathrm{Hz}}{1\ \mathrm{Hz}}\right) = 60\ \mathrm{dB\text{-}Hz}$$

and a temperature of 300 kelvins can be expressed as

$$T_{\mathrm{dB\text{-}K}} = 10 \log_{10}\left(\frac{300\ \mathrm{K}}{1\ \mathrm{K}}\right) = 24.77\ \mathrm{dB\text{-}K}$$

Although this might seem somewhat strange, it will turn out to be very convenient in the future when we consider calculation of powers in communication systems.

1.4.2 Calculation of Power Levels in Communication Systems; Link Budgets

We now assume the scenario of Figure 1–11. A transmitter in a communication system is transmitting at an average power level of P_T watts. We wish to find the level of the received power at the output terminals of a receiving antenna located a distance d meters away from the transmitter. We attack this problem by first considering the power density provided by a transmitter that isotropically radiates the power. The approach used here is applicable to communications between platforms, which can be viewed as far removed from the earth's surface so that inverse square law propagation is valid. Since a sphere of radius d has surface area $4\pi d^2$, it follows that the *power density* at the receiver is

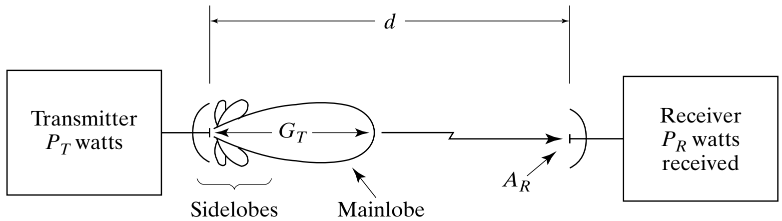

FIGURE 1–11 Hypothetical communications link for computing received signal power.

$$p_R = \frac{P_T}{4\pi d^2} \text{ watts/m}^2 \tag{1–23}$$

However, some transmitters are equipped with directional antennas that radiate more power in a given direction. This directivity, shown schematically in Figure 1–11, is described by a power gain factor, G_T, so that the power density at the receiver site is G_T times (1–23). The amount of power captured by the receiver is p_R times the aperture area, A_R, of the receiving antenna. The aperture area is related to the maximum antenna gain by

$$G_R = \frac{4\pi A_R}{\lambda^2} \tag{1–24}$$

where $\lambda = c/f$ is the wavelength of the radiation. For a parabolic antenna the aperture area A_R in (1–24) is the *effective* area, which is less than the physical area by an efficiency factor ρ_R. Typical values for ρ_R range from 60% to 80%. Putting (1–23) and (1–24) together with the transmitting antenna directivity factor, G_T, yields the result

$$P_R = \frac{P_T G_T G_R \lambda^2}{(4\pi d)^2} \tag{1–25}$$

Equation (1–25) includes only the power loss due to spreading of the transmitted wave. If other losses are also present, such as atmospheric absorption, or ohmic, losses in the waveguides leading to the antennas, (1–25) is modified to

$$P_R = \left(\frac{\lambda}{4\pi d}\right)^2 \frac{P_T G_T G_R}{L_0} \tag{1–26}$$

where L_0 is the loss factor for the additional losses. When expressed in terms of decibels, (1–26) becomes

$$P_{R,\text{dBW}} = 20 \log_{10}\left(\frac{\lambda}{4\pi d}\right) + 10 \log_{10}(P_T G_T) + G_{R,\text{dB}} - L_{0,\text{dB}} \tag{1–27}$$

The product $P_T G_T$ is referred to as the equivalent isotropic radiated power (EIRP), and the term $-20 \log_{10} (\lambda/4\pi d)$ is called the *free-space loss* in decibels (the minus sign is because a *loss* is a positive quantity).

EXAMPLE 1–5

Consider a satellite at an altitude of 800 km transmitting to a mobile receiver. The following parameters relate to this satellite-mobile link:

Satellite EIRP (G_T = 30 dB; P_T = 100 W): 50 dBW

Transmit frequency: 1500 MHz = 1.5 GHz (λ = 0.2 m)

Mobile receiver antenna gain: 3 dB

Total system losses: 6 dB

Find the received signal power at the mobile receiver antenna terminals.

Solution: The free-space *loss* in dB is

$$L_{\text{free space}} = -20\log_{10}\left(\frac{\lambda}{4\pi d}\right) = -20\log_{10}\left[\frac{0.2}{4\pi(800{,}000)}\right] = 154.03 \text{ dB}$$

From (1–27), the received power in dBW is

$$P_{R,\text{dBW}} = -154.03 + 50 + 3 - 6 = -107.03 \text{ dBW} = -77.03 \text{ dBm} = 1.98 \times 10^{-8} \text{ mW}$$

Note that increasing the frequency by a factor of 10 to 15 GHz would increase the free space loss by 20 dB. However, for a constant aperture transmitting antenna, G_T would be increased by 20 dB since $G_T = 4\pi A_T/\lambda^2$. This even tradeoff is optimistic, however, for two reasons. First, losses increase with increasing frequency, particularly atmospheric absorption and absorption due to hydrometeors. Second, antenna efficiency is lower at higher frequencies, and the 20 dB increase in free-space loss is not exactly compensated by a 20 dB increase in transmit antenna gain.

We continue our consideration of communication link power budgets by considering the noise power. A receiver can be characterized by its noise figure, F, or by its effective noise temperature, T_e, as shown in Appendix B. The two are related by

$$T_e = T_0(F - 1) \tag{1–28}$$

where T_0 = 290 K is standard ("room") temperature. The average noise power generated internally to the receiver, referenced to the receiver input, is

$$P_{n,\text{int}} = kT_eB \tag{1–29}$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant and B is the bandwidth of interest. To include the effect of noise seen by the antenna, we add the temperature of the antenna, T_{ant}, to T_e to get

$$P_n = k(T_{\text{ant}} + T_e)B = kTB \tag{1–30}$$

where $T = T_{\text{ant}} + T_e$.

The antenna temperature is not a physical temperature, but depends on where the antenna is pointed and the frequency band of the received signal. If the antenna is on board a satellite and pointed toward the earth, a reasonable value for T_{ant} is 300 K. If the receive antenna is mounted on the ground and pointed at the sky between 10 and 90 degrees with respect to the horizontal and the signal frequency is between 1 and 20 GHz, $T_{\text{ant}} \approx 50$ K. Moderate rainfall rates ($\approx$ 10 mm/h) affect this very little, but severe rainstorms may cause an increase of 10 to 50 K due to scattering of sky background noise into the antenna by the raindrops (see item 4 under Section 1.3.5). If the antenna is pointed at the sun, the antenna temperature becomes very high, and this condition is to be avoided.

Putting (1–30) together with (1–25), the received signal-to-noise (SNR) power ratio in the bandwidth B is

$$\frac{P_R}{P_n} = \left(\frac{\lambda}{4\pi d}\right)^2 \frac{P_T G_T G_R}{L_0 kTB} \tag{1–31}$$

Note that this expression holds at the receiver input or the receiver output (assuming the bandwidth B remains the same), since signal and noise powers are multiplied by the same gain factor for the receiver. The ratio G_R/T serves as a figure of merit for the receiver in that the larger this ratio, the greater the SNR. We will use (1–31) in Chapters 3, 4, and 11 to carry out design examples for communication systems.

EXAMPLE 1–6

Consider the satellite communications link of Example 1–5 with the following additional parameters:

Receiver noise figure: F = 5 dB = 3.162 ratio

Receiver antenna temperature: T_{ant} = 50 K

Receiver bandwidth: B = 1 MHz

Find the received SNR.

Solution: From (1–28),

$$T_e = (3.162 - 1)(290) = 627.1 \text{ K}$$

From (1–30) we find the receiver noise power to be

$$P_n = (1.38 \times 10^{-23})(50 + 627.1) \times 10^6 = 9.343 \times 10^{-15} \text{ watts} = -140.3 \text{ dBW}$$

Using the result of Example 1–5, we find the SNR to be

$$\frac{P_R}{P_n} = -107.03 - (-140.3) = 33.27$$

The free-space loss term in (1–25) resulted from considering ideal propagation in free space. It is a result of the wave spreading in spherical wave fronts. For terrestrial propagation (i.e., propagation along the surface of the earth), the attenuation with distance from the source will not usually obey an inverse square law with distance for a number of reasons. The first, and most obvious, reason is that the wave does not propagate as spherical wave fronts. A second reason is that other losses are present, such as attenuation of a wave front propagating with the ground as one limiting surface, which introduces attenuation losses due to the finite conductivity of the ground. A third reason, and this is most pertinent to cellular radio, is that the propagating wave is the superposition of many waves due to refraction, reflection, and scattering. In view of this, we generalize (1–27) by replacing the free-space loss term with

$$L_{\text{prop}} = L(d_0) + 10n \log_{10}\left(\frac{d}{d_0}\right) = L_0 + 10n \log_{10} d \text{ dB } (d_0 = 1 \text{ m}) \qquad (1\text{–}32)$$

where L_0 and n are constants that depend on carrier frequency, antenna height and type, terrain type, and so on. These constants may be determined for a particular situation analytically in simple environments, empirically, or a combination of the two.

A commonly used model for propagation in cellular systems is the *Hata* model [9], which is empirically derived (i.e., an analytical model was fit to extensive measurements of power levels in terrestrial radio wave propagation in Tokyo). This model will be discussed more fully in Chapter 10.

EXAMPLE 1–7

The following parameters apply to a certain cellular radio system:

Frequency	= 900 MHz
Base station antenna height	= 30 m
Mobile antenna height	= 2 m

For these parameter values, application of the Hata model for urban propagation gives L_0 = 19.53 dB and n = 3.52. Application in a suburban area gives L_0 = 9.59 dB and n = 3.52. Find the loss at 10 km for both the urban and suburban situations.

Solution: For the urban case,

$$L = 19.53 + 35.2 \log_{10} d$$

For d = 10,000 m, we obtain L = 19.53 + (35.2)(4) = 160.33 dB.

For the suburban case, we have

$$L = 9.59 + 35.2 \log_{10} d$$

For d = 10,000 m, the loss is now L = 150.39 dB.

1.5 DRIVING FORCES IN COMMUNICATIONS

Citizens of nations with ready access to information technology seem to demand ever increasing information transfer rates. It is sometimes difficult to ascertain true demand versus apparent demand stimulated by advertising by manufacturers of equipment and suppliers of services. It appears that one result of this "more begets more" mentality will be wideband communications to the home at affordable cost [19–21] in the near future. Just how this delivery will take place is not clear. Indeed, in the United States, the demarcation lines once sharply defined by government regulation between telephone services, broadcast, and entertainment media have all but disappeared. Furthermore, the mode of delivery depends on such factors as location (e.g., mountainous terrain, city streets, etc.) and population density (e.g., higher population density makes wideband delivery more economically feasible, lower population density means that services must be scaled back or government subsidies are necessary).

It is true in the United States, at least, that the invention and development of many modern communication techniques and systems were hastened by the expenditure of large amounts of government research and development funds. Yet, during peacetime, production of commercial communications gear far outstrips the military in terms of units produced and consumed. This is partly due to a larger per unit cost resulting from the necessary ruggedization of military gear and the relatively small numbers produced (unless in time of war). Commercial development and sales, however, are of necessity of affordable cost per unit with far more units produced. In the competitive societies of the developed nations, this not only means a tremendous value to the consumer, but also continually more efficient and better performing units as production of a given product continues. Examples are many—CD players, televisions, video cameras and players, and so on. Consider, for example, the development of cellular telephone systems. When first developed in the 1970s, it was thought that the number of subscribers would number a few 10,000s and the price of a hand-held telephone unit exceeded $1,000. Present-day numbers far exceed this initial estimate, with there being in excess of 75 million United States subscribers of cellular telephones alone (June 1999 statistics); correspondingly, the price of a hand-held unit has dropped below $100 or is even free with the subscription service.

Another example is the development and acceptance of the Internet—predicted by several experts to be the technological advance of the twentieth century that will have the biggest impact on the way we live and work within the next twenty-five years.[17] As an interesting comparison of adoption of specific technologies by the public, consider the following time periods for reaching 50 million users in the United States:

Internet—3 years
Television—13 years
Personal computer—16 years
Radio—38 years

Granted, the U.S. population is far larger now than when the radio was first introduced to the general public, but the rate of acceptance of the Internet has indeed been astounding. References [20–24] provide some very interesting reading on what can be expected of communication trends in the future.

1.6 COMPUTER USE IN COMMUNICATION SYSTEM ANALYSIS AND DESIGN

Generally, analysis and design of any device or system nowadays makes extensive use of computers. In the case of communication systems, computer use can take two forms—computational procedures (with suitable plotting capability) and simulation. The latter makes use of sampling theory and random number generation to represent the signals and noise within a communication systems. Signal processing operations are implemented in terms of suitably designed algorithms. Very complete system simulations are possible with only portable and desktop computers.

Within the past few years, several convenient and powerful computer simulation tools for communication systems have appeared on the market. Some of these are very expensive and comprehensive while others much more affordable but, generally, with less capability than the more expensive ones. Examples of the more expensive variety are MIL3's OPNET for radio channel and network modeling and CACI's COMNET III for network simulation.[18] Examples of the less expensive types are MATLAB's Communications Toolbox [25], Visual Solution's VisSim/Comm, and Elanix's SystemView [26].[19] All of these programs are block-diagram based: The user constructs a block diagram of the system to be simulated from a palette of blocks, which is then implemented in terms of computer software by the mathematical functions represented by the blocks. In this book, we will use MATLAB exclusively for analysis and simulation. Furthermore, we will construct our own MATLAB programs rather than use the Communications Toolbox. This will be done for two reasons. The first is that we want to use the least expensive option possible, and

[17]L. Geppert and W. Sweet, issue editors, "Technology 2000 Analysis and Forecast," *IEEE Spectrum*, Jan. 2000.

[18]http://www.mil3.com/ and http://www.caciasl.com/, respectively.

[19]http://www.mathworks.com/, http://www.vissim.com/, and http://www.elanix.com/, respectively.

MATLAB's Student version [27, 28] can be obtained for \$100.[20] Second, we feel that in a learning environment, such as a course for which this book would be used, should encourage understanding of how and why a particular software program is implemented rather than to simply use the canned capability of one of the simulation packages. A less important reason is that the more complex simulation programs involve a fairly extensive learning time in order to get reasonably proficient with them. It is a toss up whether this learning time is longer or shorter than the time required to write one's own program but, with the latter approach, one achieves the understanding required to design the simulation program. We begin writing simulation programs at the end of Chapter 2.

General papers on computer simulation and modeling of communication systems can be found in several archival publications [29].

1.7 PREVIEW OF THE BOOK

In this chapter, we have attempted to set the context of digital communications. The remainder of the book enlarges upon the theory of digital communications with many design examples used as illustrations.

Chapter 2 is a review of signal and system theory. It is assumed that the student has had a prior course on signals and systems before taking this course. However, very ambitious students can fill a void in this regard with diligent study of Chapter 2. Another purpose of Chapter 2 is to introduce notation that will be used throughout the book. It has been similarly assumed that the student has had a prior course on probability. Those persons that are a little rusty in regard to probability will find a short review of the essentials of this subject in Appendix A. In many undergraduate probability courses, random signals are not studied, or very little time is spent on them. Accordingly, a section on random signals has been included in Chapter 2.

In Chapter 3, the subject of digital data transmission is introduced. The treatment is simplified in that perfect synchronization of the receiver is assumed and the noise is taken as additive and white. In addition, a particular type of receiver structure is assumed, namely, a linear filter followed by a threshold detector. The receiver filter is optimized to maximize peak signal-to-rms (root-mean-square) noise ratio at its output, and the result is the classic matched filter receiver. The data transmission schemes considered are binary. Although the channel is initially considered to be of infinite bandwidth, this is eventually relaxed by considering optimum systems for the strictly bandlimited case. Next in the chapter is a consideration of equalization methods—that is, means for undoing the intersymbol interference introduced by bandlimiting in the channel. The chapter ends with a brief consideration of signal design for bandlimited channels and noise effects in pulse-code modulation systems. This chapter includes all the essentials of the digital data transmission process and can be used in an abbreviated course as the sole consideration of standard digital data modulation systems.

[20]As of Dec. 1999, MathWorks has made student versions of MATLAB and SIMULINK directly available from The MathWorks, Inc., 24 Prime Park Way, Natick, MA 01760, Tel. (508) 647–7632 for \$100. There are no limitations on array sizes and other MathWorks products, such as the Communications Toolbox, can be added on.

The purpose of Chapter 4 is to provide a sound theoretical basis for the digital modulation systems introduced in Chapter 3, as well as to extend the results in several directions. The approach used is that of Bayes' detection couched in the language of signal space. The background noise is still considered to be additive, Gaussian, and white, which allows the use of any orthogonal basis function set that spans the signal space. This is a considerable simplification and one that gives a very clear geometric picture of the digital signal reception process. The main generalization of Chapter 4 over Chapter 3, in addition to putting the signal reception and detection problem on a sound theoretical basis, is to consider M-ary digital data transmission and the explicit treatment of modulation schemes suitable for practical channels. In order to compare M-ary systems on an equivalent basis, the concepts of equivalent bit error probability and bandwidth efficiency in terms of bits per second per hertz of bandwidth are introduced. This chapter ends with several example design problems and a basic introduction to orthogonal frequency division multiplexing.

Chapter 5 takes up several topics that can be considered to be degradation sources for the ideal systems considered in Chapter 4. Synchronization methods at various levels (i.e., carrier, bit, and frame) are discussed, and the degradation imposed by imperfect carrier synchronization is characterized. The detrimental effects of fading channels are characterized, and a means to combat them, namely diversity transmission, is discussed. The chapter ends by considering several gross means of characterizing practical digital communication system performance, including envelope plots, eye diagrams, and phasor plots.

Chapters 6 through 8 take up the subject of coding, with the elements of information theory and block coding considered in Chapter 6, and the elements of convolutional coding covered in Chapter 7.

Coding is a very broad subject, and over the years a rich theory has grown up around it. Wherever possible, the theoretical foundations are provided, but the underlying objective of Chapters 6 and 7 is always one of system applications. Accordingly, the student is reminded again and again of the two fundamental characteristics of any application of digital communications, whether coding is employed or not. These are the concepts of power and bandwidth efficiency as introduced in Chapter 1. Thus all coding techniques considered in Chapters 6 and 7 are characterized in terms of their ability to lower the signal-to-noise ratio required to achieve a desired probability of bit error (power efficiency) and the bits per second that can be supported per hertz of bandwidth (bandwidth efficiency). Chapter 8 provides a brief treatment of another error control scheme called automatic repeat request (ARQ), which utilizes a feedback channel.

Chapter 9 contains an overview of spread-spectrum communications. There is much that can be said of this fascinating subject and entire textbooks have been written on it. Only the high points are given here to introduce the student to the basic ideas. The important concept of multiuser detection is considered where, when signals from multiple users are being received, the detection process takes into account their statistical characteristics and improves the detector performance over what could be obtained if the other-user signals were treated as noise.

Chapter 10 deals with cellular radio communications. Both fading channel models and the analysis of terrestrial communications links under fading conditions are treated in some depth, although several entire books exist on the subject. Although first-generation cellular systems were analog, second-generation systems use digital techniques even though they are limited primarily to voice. With the advent of third-generation systems in the early 2000s, the use of digital modulation techniques is mandatory because such systems must handle voice, video, and data. Only with digital transmission will such mixed types of traffic be accommodated. These ideas are discussed after an introduction to the basics of cellular radio.

Chapter 11 treats satellite communications as an example where digital communications concepts and applications have come into extensive use over the years. The concepts are illustrated with several design examples.

REFERENCES

[1] B. Oliver, J. Pierce, and C. Shannon, "Philosophy of PCM," *Proc. IRE*, Vol. 36, pp. 1324–1331, Nov. 1948.

[2] M. Ristenbatt, "Alternatives in Digital Communications," *Proc. IEEE*, Vol. 61, pp. 703–721, June 1973.

[3] Special Issue on Software Radios, *IEEE Commun. Mag.*, Vol. 33, May 1995.

[4] H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *Trans. AIEE*, Vol. 47, pp. 617–644, April 1928.

[5] W. Diffie and M. Hellman, "New Directions in Cryptography," *IEEE Trans. Inf. Theory*, Vol. IT-22, pp. 644–654, Nov. 1976.

[6] N. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video* (Upper Saddle River, NJ: Prentice-Hall, 1984).

[7] G. Held, *Data Compression*, 2nd ed. (New York: John Wiley, 1987).

[8] C. Shannon, "A Mathematical Theory of Communications," *Bell Syst. Tech. J.*, Vol. 27, pp. 379–423, July 1948, and pp. 623–656, Oct. 1948.

[9] T. S. Rappaport, R. Muhamed, and V. Kapoor, "Propagation Models," in *The Communications Handbook*, J. D. Gibson (ed.) (Boca Raton, FL: CRC Press, 1997)

[10] A. van der Ziel, *Noise in Measurements* (New York: Wiley-Interscience, 1976).

[11] P. E. Green, Jr., *Fiber Optic Networks* (Upper Saddle River, NJ: Prentice Hall, 1993)

[12] R. C. Bray and D. M. Baney, "Optical Networks: Backbones for Universal Connectivity," *Hewlett-Packard Journal*, Vol. 48, pp.19–31, Dec. 1997

[13] I. Lebow, *Understanding Digital Transmission and Recording* (Piscataway, NJ: IEEE Press, 1998)

[14] N. Sloane and A. Wyner (eds.), *Claude Elwood Shannon Collected Papers* (Piscataway, NJ: IEEE Press, 1993)

[15] R. Olsen, et al., "The aR^b Relation in the Calculation of Rain Attenuation," *IEEE Trans. Antenn. Propag.*, Vol. AP-26, pp. 318–329, Mar. 1978.

[16] R. Crane, "Prediction of Attenuation by Rain," *IEEE Trans. Commun.*, Vol. COM-28, pp. 1717–1733, Sept. 1980.

[17] D. Hogg and T. Chu, "The Role of Rain Attenuation in Satellite Communications," *Proc. IEEE*, pp. 1308–1331, Sept. 1975.

[18] L. Ippolito, *Radio Propagation in Satellite Communications* (New York: Van Nostrand Reinhold, 1986).

[19] I. Lebow, *Information Highways and Byways* (Piscataway, NJ: IEEE Press, 1995).

[20] Special Issue on Communications in the 21st Century, *Proc. IEEE*, Vol. 85, Oct. 1997.
[21] Special Issue on The Global Information Infrastructure, *Proc. IEEE*, Vol. 85, Dec. 1997.
[22] V. Li, "Personal Information Service (PIS)—An Application of Wide-Band Communications, 2012 A.D.," *Proc. IEEE*, Vol. 86, pp. 737–740, April 1998.
[23] U. Black, *Mobile and Wireless Networks* (Upper Saddle River, NJ: Prentice Hall, 1996).
[24] E. Wesel, *Wireless Multimedia Networks* (Reading, MA: Addison-Wesley, 1997).
[25] User's Guide for "Communications Toolbox," Natick, MA: The Math Works, Inc., 1996.
[26] Elanix, Inc., Westlake Village, CA 91362.
[27] *MATLAB® student version, Version 5.3 User's Guide* (Natick, MA: The MathWorks, Inc., 1999).
[28] *SIMULINK student version, User's Guide,* version 3.0 (Natick, MA: The MathWorks, Inc., 1999).
[29] Special Issues on Computer Modeling of Communication Systems, *IEEE Journ. on Selected Areas in Commun.,* Jan. 1984, Jan. 1988, and May 1997.

PROBLEMS

1–1. An analog source has a bandwidth of 4 kHz. Plot the minimum source rate in bits per second versus the logarithm to the base 2 of the number of levels to which each sample is quantized for the following cases: 2 levels, 4 levels, 8 levels, 16 levels, 32 levels, 64 levels, 128 levels, 256 levels.

1–2. In a digital telephone system, it is determined that the voice spectrum has significant frequency content up to 3 kHz. The channel can accommodate 18 kilobits per second. What is the maximum number of quantization levels to which each sample can be quantized?

1–3. **(a)** Show the steps in deriving (1–5) from (1–4).
(b) Plot your own curve like Figure 1–4 on a piece of graph paper.
(c) An error correction code is used in conjunction with a digital modulation scheme that employs bits of duration T_b and requires a bandwidth of $W = T_b^{-1}$ hertz. The code imposes a bandwidth expansion factor of 2, that is, each information bit requires 2 code bits. Therefore, the rate out of the encoder must be twice the bit rate into the encoder, necessitating twice the bandwidth. The code provides a bit error probability of 10^{-5} at a signal-to-noise ratio, E_b/N_0, of 6 dB, which is considered error-free for all practical purposes. Accounting for the code bandwidth expansion, plot the performance of this system on your graph constructed in part (b).

1–4. A customer desires a communication system that is capable of conveying 60 kilobits per second through a channel of bandwidth 10 kHz. The customer can achieve a received signal power of 1 picowatt. The channel noise is 10^{-19} watts per hertz of bandwidth. Should your company submit a bid to build the communication system?

1–5. A *Gray* code is sometimes used to represent symbols in a digital communication system. If $b_1b_2b_3 \ldots b_n$ denotes a binary number representation, the Gray encoded number is found using the rules

$$g_1 = b_1$$
$$g_n = b_n \oplus b_{n-1}$$

where ($\oplus$ denotes modulo-2 addition without carry) and the Gray encoded number is represented by $g_1g_2g_3 \ldots g_n$. Construct a table giving the binary code along with the corresponding Gray code representations for the decimal digits 0–7.

1–6. Referring to (1–6), consider a communication system with the following values for the different probabilities:

$$P(Y = 0|X = 0) = 0.7 \qquad P(Y = 0|X = 1) = 0.1$$
$$P(Y = 1|X = 0) = 0.3 \qquad P(Y = 1|X = 1) = 0.9$$
$$P(X = 0) = 0.4; P(X = 1) = 0.6$$

(a) Find the probabilities $P(Y = 0)$ and $P(Y = 1)$ at the channel output.
(b) Find the probability that a 1 was sent given that a 1 was received.
(c) What is the probability that a 0 was sent given that a 1 was received?

1.7. **(a)** Derive the relations (1–11) and (1–12). Show that

$$B_{\max} = A|a_1 + a_2|$$
$$\text{and } B_{\min} = A|a_1 - a_2|$$

(b) Plot a figure like the ones shown in Figure 1–9 for $a_1 = 1$, $a_2 = 0.2$, and $\tau = 10^{-6}$ seconds. Assume $A = 1$.
(c) Comment on the distortion introduced by such a channel to a digital signal of bandwidth 10 kHz, 100 kHz, 1 MHz, and 10 MHz. You can use descriptors like “negligible,” “moderate,” and “extreme.”

1–8. **(a)** Consider Figure 1–8c with the delay of the second bit stream, $\tau = 0$. A modulation scheme is employed for which the probability of bit error is $P_b = 1/2 \exp(-E_b/N_0)$, where E_b is the energy per bit and N_0 is the noise power spectral density. It follows from Figure 1–8c that $E_b = (a_1 + a_2)^2 T_b$ for $\tau = 0$. Assume that $a_2 = 0.5a_1$. Find the constant k, defined as $k = a_1^2 T_b/N_0$ such that $P_b = 10^{-5}$. This part of the problem is that of calibration.
(b) Now let $\tau = 0.5T$ and $a_2 = 0.5a_1$. Find P_b for each bit in Figure 1–8c using the constant k obtained in part (a). Find the average P_b over the first ten bits shown in Figure 1–8c (i.e., the sequence 1, –1, 1, –1, 1, 1, 1, –1, 1, 1).
(c) Obtain P_b averaged over the 10-bit sequence as a function of τ for $0 \le \tau \le T_b$ assuming the amplitude values of part (b). Plot versus τ. This is fairly easy once one deduces the energies of an “interfered” bit and a “reinforced bit”.

Note: The calculations in this problem illustrate the use of a “typical data sequence” to evaluate the degradation due to memory effects in digital communications systems. Normally, the computation would be carried out over a much longer sequence with the aid of a computer.

1–9. Doppler frequency shift is given by $f_d = v/\lambda = vf_c/c$ where v is the velocity of the receiver relative to the transmitter or reflecting object, λ is the wavelength of the transmitter radiation, f_c is the frequency of the radiated signal (usually the carrier frequency in the case of a modulated signal) and $c = 3 \times 10^8$ m/s is the speed of electromagnetic propagation.

a) A sinusoidal signal is radiated from a cellular radio base station to a moving automobile traveling at a speed of 75 km/hr directly away from the base station. There is both a direct propagation path to the automobile and a reflected path from an object some distance in front of it. Accounting for the distance difference and reflection, the relative amplitudes of the direct and reflected waves is 1:0.2. The carrier frequency is 900 MHz. The antenna on the automobile is omnidirectional. Write down expressions and plot the envelope and phase functions given in (1–16) and (1–17) for this situation.

b) If the periods of the functions $C(t)$ and $\phi(t)$ are appreciable fractions of a bit period in a digital communication system, Doppler spread will impose nonnegligible degradation on the system. Consider the following bit rates for a digital communication system communicating with the automobile: 1 kbps; 10 kbps; 50 kbps. Label the degradations for each of these with the descriptors "negligible," "moderate," and "extreme."

1–10. Coefficients for the aR^b relationship for rain attenuation are given in the following table for various frequencies:

Frequency GHz	*a*	*b*
12	0.0215	1.136
15	0.0368	1.118
20	0.0719	1.097
30	0.1860	1.043
40	0.3620	0.972

Compute the rain attenuation per kilometer for the following combinations of rain rate and carrier frequency:

(a) $f = 12$ GHz; $R = 1$ mm/h (light rain)
(b) $f = 40$ GHz; $R = 1$ mm/h
(c) $f = 12$ GHz; $R = 25$ mm/h (heavy rain)
(d) $f = 40$ GHz; $R = 25$ mm/h
(e) $f = 20$ GHz; R 10 mm/h (moderate rain)

1–11. The available noise power per hertz from a resistive source is $P_{\text{avail}} = kT$ watts where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant and T is the temperature in kelvins. Compute the available noise power spectral density (the power per hertz) from a resistor at room temperature ($T = 290$ K) in dBm/Hz and dBW/Hz.

1–12. Consider a geostationary-orbit satellite at an altitude of 35,784 km above the earth's equator (such a satellite has a period equal to one day and appears stationary relative to the earth if it is in an equatorial orbit). The transmit frequency is 12 GHz. The transmit and receive antennas have a 1 m^2 aperture with an aperture efficiency of 70%. The transmit power is 100 watts. Assume total system losses of 6 dB. Find the signal power at the receive antenna terminals in dBW, dBm, and microwatts.

1–13. **(a)** Consider a geostationary satellite (see Problem 1–12 for an explanation of this term) where the transmit frequency is 20 GHz. Transmit and receive antenna apertures of 2 m^2 are used and the aperture efficiency is 75%. Assume 2 dB of hardware losses and 0.4 dB of attenuation due to atmospheric absorption. Also, allow for attenuation due to a moderate rain storm of 10 mm/h, which is 2 km in extent over the transmission path (see Problem 1–10 for the computation of this loss). Find the required transmit power in watts to provide a receive power at the antenna terminals of one picowatt.

(b) Redo part (a) for a frequency of 15 GHz. Assume the aperture efficiency stays the same, but scale the atmospheric absorption inversely proportional to frequency squared. You can recompute the rain attenuation given the information in Problem l–10.

1–14. Consider a geostationary ground-satellite-ground link with the following parameters:

Uplink frequency:	12 GHz
Downlink frequency:	10 GHz
Slant range:	40,000 km
Atmospheric absorption:	0.2 dB
System losses:	3 dB (uplink and downlink)
Rain attenuation:	Negligible
Antenna aperture (satellite and ground):	1.5 m^2
Antenna aperture efficiency:	70%
Receiver noise figure (satellite and ground):	4.5 dB
Satellite receiver antenna temperature:	300 K
Ground receiver antenna temperature:	50 K
Bandwidth (uplink and downlink):	100 kHz

Find the ground and satellite transmitter powers to provide a 20 dB signal-to-noise power ratio in both uplink and downlink. Assume that the same antennas are used for transmit and receive. Assume the effect of noise on the uplink is negligible on the downlink.

1–15. Equation (1–24) gives an expression for one important attribute of an antenna, the maximum gain, or simply gain. Another important parameter is the 3 dB mainlobe beamwidth, which is the angle bounded by the points on the mainlobe gain pattern (see Figure 1–11) where the gain falls to one-half of the maximum value. A useful approximation to this 3 dB beamwidth for an antenna with a circular aperture of diameter d is

$$\phi_{3\text{ dB}} = \frac{\lambda}{d\sqrt{\rho}} \text{ radians}$$

where ρ is the aperture efficiency and λ is the wavelength.

(a) Show that (1–24) in the case of a circular aperture antenna of diameter d and efficiency ρ, becomes

$$G = \rho\left(\frac{\pi d}{\lambda}\right)^2$$

(b) Compute the maximum gain and 3 dB beamwidth for a circular-aperture antenna of diameter 1 meter and efficiency 75% for the following frequencies: 10 GHz, 15 GHz, 20 GHz, 30 GHz.

1–16. It is desired to have a circular-aperture antenna with a gain of 30 dB operating at a frequency of 1.5 GHz. The aperture efficiency is 75%. What should the diameter be? What is the 3 dB beamwidth? (Refer to Problem 1–15.)

1–17. **(a)** For a satellite transmitter operating at 20 GHz, a circular-aperture antenna with a two-degree beamwidth is desired. The aperture efficiency is 65%. What should the antenna diameter be to achieve this beamwidth? What is the corresponding gain? (Refer to Problem 1–15.)

(b) Find the diameter of an antenna that will provide a 150-mile diameter spot at the equator within the 3 dB beamwidth if the platform on which the antenna is mounted is a geostationary-orbit satellite. Assume the operating frequency and efficiency of part (a). What is the gain of the antenna? (Geostationary altitude is 35,784 km above the earth's equator.)