1 Introduction

If the mind is in tranquility, time and space cease to exist.

Essence of Buddhism

1.1 Brief overview

This book provides the principles of digital communication and studies techniques to design and analyze digital communication systems for point-to-point and point-to-multipoint transmission and reception. Other than for radio broadcasting, modern communication systems are going digital, and in the USA the conversion of analog TV broadcasting into digital HDTV broadcasting at the beginning of 2009 signified the coming end of analog communications. Communications between living beings began with the voice, and the three biggest voice systems in the world are the telephone, and the cellular and radio broadcasting systems.

The dissemination of visual activities then propelled the development of TV broadcasting systems. The pioneer telephone network and radio broadcasting systems employed analog communication techniques, such as AM and FM, for transmission of analog voice, as did the analog TV broadcasting systems, which employed VSB-AM for picture transmission. The quality of the message, such as voice and images, at the analog receiver depends on how well the waveform that carries the message over the physical channel (twisted-pair telephone wires, coaxial and fiber-optic cables, space, and water) can be reproduced. In addition, the fidelity of the received message depends on the signal-to-noise ratio at the receiver input. For good analog communications, the signal-to-noise ratio must be large, and this requires high-power transmitters, such as are used in AM radio and TV broadcasting. For FM radio broadcasting a large frequency spectrum is used, such as 200 kHz for radio broadcasting, which shows that analog communications do not utilize power and bandwidth efficiently. Furthermore, the advent of the Internet requires audio, video, imagery, and text messages to be integrated for transmission over a common channel and this in effect rules out analog communications such as AM and FM.

In analog communications, the message signal requires an *infinite set of continuous-time waveforms* for transmission over a physical channel. This is because the message itself, such as audio or video, must first be converted into a *voltage baseband waveform* with a *continuous range* in amplitude that has countless possible values. When the baseband voltage waveform is used to modulate an RF carrier for transmission, such as in AM or FM, the modulated RF signal transmitted over the physical channel also has countless possible values in both its amplitude and frequency ranges. The only way to recover the message signal is to faithfully reproduce the baseband waveform from the modulated signal. This can be done easily in the case of no noise and no equipment imperfections, but otherwise the fidelity of the message signal may be reduced. Digital communication does not involve the faithful reproduction of the baseband waveform in the presence of noise and equipment imperfections. Digital communication operates instead with a *finite set of continuous-time modulation waveforms* for transmission over a physical channel. This implies that the message signal must be represented by a *finite set of voltage baseband waveforms*. Mathematically, a finite set of waveforms can only represent a finite set of alphabets, commonly referred to as *symbols*. A symbol consists of a fixed number of binary digits or *bits*. For example, the set of four distinct symbols {00, 01, 10, 11} can be represented by four distinct waveforms $\{\pm A\cos 2\pi f_c t, \pm A\sin 2\pi f_c t\}$. The time separation of consecutive waveforms that represent a symbol stream is called the *symbol time*, which is the inverse of the *symbol rate*. If the waveforms are of finite duration then this duration is the symbol time. This begs the question of how to obtain the bits or symbols that represent the message. The process of converting a voltage baseband waveform that represents an audio or video message into bits is referred to as the *analog-to-digital conversion* (or A/D). Text messages generated by computers are inherently in bits, so with A/D conversion, audio, video, text, and imagery can all be integrated into a single digital stream of bits. The process of A/D, bit-symbol mapping, baseband waveform shaping, and modulation is referred to as digital transmission. The process of demodulating the modulated signal, detecting the symbol, symbol–bit mapping, and *digital-to-analog conversion* (or D/A) is called digital reception.

Digital communication makes no attempts to reproduce the finite set of voltage baseband waveforms. Instead, the receiver detects the energy content of each baseband waveform in the presence of noise and equipment imperfections, and then makes a best estimate of which transmitted symbol was received. If the *signal-to noise ratio per symbol* is reasonably large, a symbol will most likely be detected correctly with high probability. If not, a symbol error may occur. This is the essence of digital communication. For a given signal-to-noise ratio, an analog communication receiver attempts to reproduce the voltage baseband waveform with certain *subjective* fidelity. On the other hand, for a given signal-to-noise ratio per symbol, a digital communication receiver produces symbols with a *quantitative* error rate. It is important to know in advance the *lower bound* of the signal-to-noise ratio per symbol for a specified error rate irrespective of the type and size of the set of modulation waveforms. In 1948 Claude Shannon established this lower bound and also provided the channel capacity for reliable transmission [1]. Shannon's work gives the designers of digital communication systems the freedom to choose the set of modulation waveforms that achieve either the best power or bandwidth efficiency, or a trade-off combination of both. As long as the transmission rate is below the channel capacity and the signal-to-noise ratio per symbol is above the *Shannon limit*, reliable communication is possible with an arbitrarily small error rate. Guided by the *Shannon channel capacity theorem* (*main theorem*), the designer can further integrate error-correction codes with modulation techniques to lower the signal-to-noise ratio per symbol to achieve a specified error rate. The first error-correction code, the *Hamming code*, was discovered by Richard W. Hamming in 1950, two years after Shannon published his landmark work [2]. In addition to the main theorem, the *Shannon first theorem*

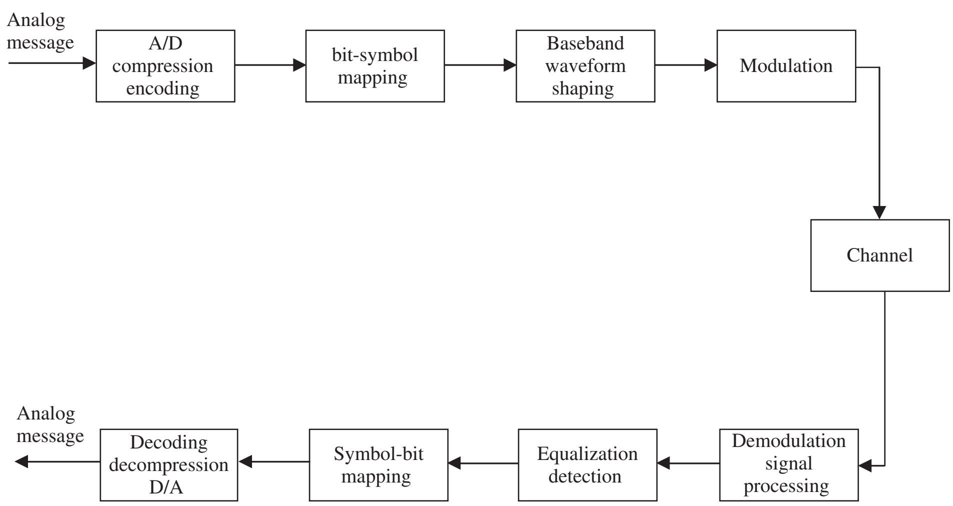

Figure 1.1 Conceptual block diagram of a digital communication system.

provided the framework for encoding a discrete source of a finite set of symbols to minimize the bit rate at the source encoder output. This allows the compression of the A/D samples of the message signal to remove redundancy and any insignificant information not perceptible by the human eye or ear.

The most common compression algorithms in use today are MP3 for music, JPEG for pictures, and MPEG for video. Figure 1.1 shows the conceptual block diagram of a digital communication system. The material in the book is organized to cover the transmitter, receiver, and channel.

1.2 Scope

Chapter 2 provides a general study of deterministic signals that can be analyzed with Fourier transform and Fourier series. Simple classification of signals and the concept of power and energy are reviewed. One important class of signal, namely orthogonal signals, such as the Walsh functions employed in IS-95, CDMA 2000, and WCDMA, is discussed in detail. The majority of continuous-time and finite-energy signals in practice can be conveniently analyzed via their signal spaces. These signal spaces are displays of the signal vectors in their respective constellations. The signal vectors which can be viewed as the A/D versions of a signal set contain all the information about the signal set. Practical communication systems are inherently linear time-invariant systems operating in the small-signal range. They can be analyzed by Fourier series and Fourier transform to provide a frequency-domain snapshot of the signal bandwidth. The concept of autocorrelation and its relationship with energy or power spectral density are discussed for linear time-invariant systems. The sampling theorem that governs the A/D conversion of an analog signal and the

Nyquist–Shannon interpolation for reconstruction of the analog signal are presented. Finally, the representations of a bandpass signal, that is, the signal sent over a physical channel, are discussed. The complex envelope (equivalent lowpass signal) of a bandpass signal that can be employed to simplify the analysis of a communication system is also included.

Chapter 3 studies random signals and their statistics. Although a finite set of deterministic signals is employed to represent a finite set of information symbols, the transmitted symbols are truly random, with each one in the set occuring with a fixed probability. Therefore, the infinite series of signals transmitted over the channel is indeed a random process with finite power. The study of random processes allows the establishment of the Fourier transform relationship between the autocorrelation of the random process and its power spectral density via the *Einstein–Wiener–Khinchine theorem*. The emphasis here is on cyclostationary processes, which encompass all digitally modulated signals. The Gaussian process that is used to represent channel noise is discussed in sufficient detail. Sampling of bandlimited white Gaussian process, sufficient statistics for white Gaussian samples, the *Karhunen–Loeve theorem*, and whitening filter are studied. To study the performance of wireless communication via a fading channel we look at a variety of processes derived from the Gaussian process, such as the Rayleigh, Rice, Nakagami-m, χ^2, and log-normal processes.

Chapter 4 provides a general study of information theory developed by Shannon, and addresses both source and channel coding. The concept of source entropy and prefix code is discussed. *Shannon–Fano* and *Huffman* prefix codes are used as examples. The *Shannon first theorem* is presented with a proof, and the concept of mutual information is presented together with the *Shannon main theorem* for a discrete channel. The concept of differential entropy for a Gaussian channel is introduced, and leads to mutual information and Shannon channel capacity. Vector Gaussian channels and the water filling strategy are presented to highlight the concept of channel coding, which is also interpreted via the sphere packing bound. The channel capacity of a bandlimited Gaussian channel and the channel capacity of a Gaussian channel with discrete inputs are derived. The latter provides a snapshot of how efficient digitally modulated signals perform as compared to the *Shannon capacity*. Channel coding can be done with error-correction codes such as block codes and convolutional codes. Performance of coded digital signals is presented for both block codes and convolutional codes, with the emphasis on low-density parity-check codes (LDPC) and convolutional codes. The decoding of LDPC codes is implemented via the *message passing algorithm*. The decoding of convolutional codes is carried out via the *Viterbi algorithm*, which includes hard decoding, and quantized or unquantized soft decoding.

Chapter 5 examines methods for establishing a communication link between the transmitter and receiver, commonly referred to as link analysis. The link budget involves the allocation of power to the transmitter and noise temperature (or noise figure) to the receiver so that a signal-to-noise ratio is established at the receiver to match a specified error rate range. Given the transmitter power and the channel attenuation, the power of the received signal can then be established. The channel attenuation is unique for each physical medium. The chapter begins with the concept of the noise temperature of a two-port network, which leads to the concept of the *system noise temperature* of a cascade of two-port networks modeling a receiver. The system noise temperature allows the evaluation of noise power in the receiver bandwidth, and hence the *system signal-to-noise ratio*. The physical channels investigated in this chapter are the cellular and satellite channels. For the cellular channel, we adopt the well-known *Hata* model to estimate the median path loss between the transmitter and receiver. The presence of co-channel interference between cells is also taken into account. Both narrowband cellular systems (IS-136, GSM) and wideband CDMA cellular systems (IS-95, CDMA-2000, WCDMA) are covered. For a satellite channel, the communication link is a point-to-point link, consisting of up- and downlinks. The *Friis* formula for free-space attenuation is employed to establish the uplink or downlink attenuation.

Chapter 6 presents modulation techniques for transmitting information over the physical channel. The chapter essentially has two parts, namely binary modulation and M-ary modulation. The structure of each modulation technique is studied via the signal waveform, the power spectral density, and the modulator. For binary modulation we investigate phase shift keying (PSK), differential phase shift keying (DPSK), amplitude shift keying (ASK) (commonly referred to as intensity-modulated on–off keying (OOK), a technique used in fiber optic communication), frequency shift keying (FSK), minimum shift keying (MSK), and Gaussian MSK employed by the GSM cellular standard. Many practical applications require either the higher spectral efficiency or higher power efficiency that binary modulation techniques can provide; M-ary modulation can accommodate both. The second part of this chapter covers M-ary amplitude shift keying (MASK), M-ary phase shift keying (MPSK), offset quadrature phase shift keying (OQPSK), differential M-ary phase shift keying (DMPSK), $\pi/4$ shifted differential quadrature phase shift keying ($\pi/4$-DQPSK), M-ary quadrature amplitude modulation (MQAM), code shift keying (CSK), M-ary frequency shift keying (MFSK), and continuous phase modulation (CPM). The chapter continues with a treatment of the dominant multiplexing-modulation technique, namely *orthogonal frequency division multiplexing* (OFDM), which is used in many wireless standards. The chapter ends with a look at *trellis coded modulation* (TCM) for bandlimited channels. Both Ungerboeck and pragmatic TCM are investigated.

Chapter 7 provides a treatment of digital demodulation. A generic digital demodulator consists of two major subsystems, namely the signal processor and the detector. There are four types of signal processor: the matched filter, the correlator, the noncoherent matched filter, and the noncoherent correlator. The first two types are employed in coherent demodulation while the last two types are used in noncoherent demodulation. For binary demodulation the two fundamental detectors are threshold and maximum detectors. For M-ary demodulation the two fundamental detectors are the minimum Euclidean distance detector and the M-ary maximum detector. Combining the signal processor(s) and the detector in that order produces an L-path demodulator for the set of digital signals with L orthonormal basis functions and an M-path demodulator for the set of M orthogonal signals. The bit error probability analysis is carried out for binary modulation techniques such as coherent PSK, coherent DPSK, direct-detection ASK (for fiber optic communication), coherent FSK, coherent MSK, precoded MSK and GMSK, noncoherent FSK and MSK, and noncoherent DPSK. For M-ary demodulation, the bit error probability analysis is carried out for coherent MASK, coherent MPSK, coherent DMPSK, noncoherent DMPSK, coherent MQAM and DMQAM, coherent CSK and MFSK, noncoherent CSK and MFSK, coherent CPM with sequence detection, coherent CPM with symbol-by-symbol detection, and noncoherent CPM. The chapter continues with OFDM demodulation, with emphasis on the IEEE 802.11a,g standards. Finally, the demodulation and decoding of TCM are studied and performance analysis is investigated. The Viterbi algorithm is again used to illustrate the decoding process.

Chapter 8 investigates two major spread spectrum communication techniques for both commercial and military applications: direct sequence (DS) and frequency hop (FH). The chapter begins with a presentation of the pseudo-noise (PN) sequences needed for spreading the modulated signal. Next the concept of quadrature orthogonal covering using Walsh functions of the same length for multiplexing DS signals with an identical symbol rate is discussed. This concept is then extended to variable-length orthogonal covering for variable symbol rates. IS-95 is used as a real life example for the study of the direct sequence spread spectrum. The demodulation of DS signals in the presence of tone jamming, broadband jamming, and pulse jamming is analyzed. Demodulation of quadrature orthogonal covering (IS-95 forward link) as well as noncoherent DS-CSK (IS-95 reverse link) is presented. The analysis of code division multiple access (CDMA) with random spreading sequences is presented together with a closed form expression and a tight upper bound for bit error probability. For frequency hop signals, three jamming strategies are studied: partial-band jamming, multi-tone jamming, and follower jamming. Both slow and fast hops are considered for follower jamming.

Chapter 9 deals with intersymbol interference (ISI) in a bandlimited channel. The Nyquist criterion for zero ISI is stated together with the corresponding pulse shapes that satisfy it. The design of an optimum demodulator for a bandlimited channel with Gaussian noise is carried out. The optimum demodulator relies on the signal pulse shape implemented at the modulator. The channel is converted to an ideal channel via an equalizer implemented at the modulator (the equalizer is a filter with a transfer function equal to the inverse transfer function of the channel). At the demodulator, a matched filter matched to the signal pulse shape simultaneously achieves both the maximum signal-to-noise ratio and zero ISI as long as the pulse shape at the matched filter output satisfies the Nyquist criterion of zero ISI. In practice, because the channel transfer function is not known or varies with time, ISI removal is instead implemented at the demodulator. The equalizer implemented at the demodulator can be classified into two types: linear and nonlinear. The treatment of linear equalizers covers zero-forcing and mean-square error equalizers. The latter alleviates the noise enhancement effect that severely degrades the former in channels with deep attenuation in the passband. Nonlinear equalizers such as zero-forcing decision-feedback and mean-square error decision-feedback can avoid the noise enhancement effect altogether, although in channels with severe distortion the error propagation due to decision feedback could worsen the performance. To obtain optimum performance, maximum likelihood sequence detection may be employed to mitigate the ISI. The motivation behind sequence detection is to use the symbol energy that resides in the ISI portion of the symbol to aid the detection instead of throwing it away. The Viterbi algorithm is employed in practice for sequence detection. Finally, a fractionally spaced equalizer that can mitigate timing error is presented.

Cambridge University Press & Assessment
978-0-521-76174-1 — Theory and Design of Digital Communication Systems
Tri T. Ha
Excerpt
More Information

Chapter 10 studies the transmission and reception of a digitally modulated signal over a fading channel. Unlike the AWGN channel, where noise is the only problem, a fading channel is a greater challenge, as it may cause signal envelope variations, phase errors, and intersymbol interference, all of which are detrimental to the performance of the signal. Thus, it is necessary to understand the mechanism that causes these unwanted effects and find ways to mitigate them. A fading channel arises from the movement of the transmitter and receiver, commonly referred to as the time-varying effect or Doppler effect. This may cause random amplitude attenuation and random phase rotation of the signal space. In underwater acoustic communication, the medium also changes over time, further compounding the problem. A fading channel also exhibits a space-varying effect, where the locations of the transmitter and receiver and the physical structures in the environment dictate the paths that the signal may travel. A transmitted signal representing an arbitrary symbol may arrive at the receiver via multiple paths. Depending on the time delay between paths, signal echoes of a symbol may overlap the next several symbols causing intersymbol interference. Both time-varying and space-varying effects can be classified into four fading characteristics: *slow fading* with random amplitude attenuation and negligible phase error, *fast fading* with random amplitude attenuation and large phase error, *flat fading* with random amplitude attenuation and negligible intersymbol interference, and *frequency-selective fading* with random amplitude attenuation and intersymbol interference. A mobile wireless channel may have two of these four characteristics where the random amplitude attenuation is described by a special distribution (Rayleigh, Rice, and Nakagami-m). For analysis, it is convenient to model a fading channel with a channel impulse response that includes both time- and space-varying effects. From the channel impulse response, the multipath autocorrelation and Doppler profiles are derived, which lead to the concept of the Doppler power spectrum. Clarke–Doppler and Aulin–Doppler spectra are studied as examples. Using a mathematical model, the performance of a modulated signal in a fading channel is analysed. First, *ideal* coherent demodulation (assuming the carrier phase is always available for symbol-by-symbol detection) is investigated and the fading channel is assumed to produce only random amplitude attenuation. Channel tap estimation as well as the channel tap error effect is studied to reflect real world situations.

Next, the slow fading channel with random amplitude attenuation is investigated for *pilot symbol-aided* demodulation (the pilot symbols are periodically transmitted in the symbol stream; this represents a less desirable situation than ideal coherent demodulation but remains in line with practical applications). These investigations are extended to OFDM (the major waveform that was adopted by IEEE 802.11a-g, 802.16 to name a few) where a slow and frequency-selective fading channel is assumed. The fundamentals of coherent demodulation are extended to noncoherent demodulation, where the Doppler tracking of orthogonal signals is investigated. The next discussion centers on another major waveform that was adopted by IS-95, CDMA 2000, and WCDMA (for use either in the forward channel or reverse channel or both) for their respective cellular systems, namely, orthogonal covering and spread spectrum signals. Complex spreading and despreading as well as Doppler analysis and tracking are presented. For completeness the demodulation of the signal used in the reverse channel of IS-95 is also presented. Once the Doppler phase error resulting from either slow or fast fading is corrected via Doppler tracking, and the ISI resulting from frequency-selective fading is mitigated via OFDM and equalization, only random amplitude attenuation remains to be dealt with. This particular effect can be effectively alleviated via the use of time, frequency, antenna, or multipath delay diversity. Diversity is a signal combining method that makes use of uncorrelated signal redundancy for both transmission and reception to enhance symbol detection in the presence of a deep fade which may destroy a non-diversity symbol. Diversity can be achieved via redundant symbol interleaving for time diversity, or via uncorrelated subcarrier combining in OFDM for frequency diversity. It can also be achieved via multiple transmit antennas for transmit antenna diversity, and via multiple receive antennas for receive antenna diversity or using a Rake receiver for multipath delay diversity. Combinations of these methods are also possible. Three main signal combining methods are studied: maximal ratio combining (MRC), selection combining (SC), and equal gain combining (EGC). MRC is the optimum combining scheme for coherent or pilot symbol-aided demodulation in AWGN and is the most commonly used method for wireless LAN, MAN, WAN, and cellular systems. It is superior to SC and EGC, although it cannot be used for noncoherent demodulation, unlike the other two schemes. Wireless communication in a fading channel favors the use of multiple transmit antennas for performance enhancement. This type of antenna diversity employs orthogonal space-time block codes with rates of ½, ¾, and 1 with MRC at the receiver. The Alamouti code with unity rate was recommended for the IEEE 802.16 family. Integrating both transmit and receive antenna diversity provides the receiver with a powerful method to combat random amplitude attenuation.

The remaining parts of this chapter investigate the capacity of a fading channel. Since a fade causes an outage that may drive the instantaneous channel capacity to zero, the average channel capacity defined for an AWGN channel does not exist for a slow fading channel. Therefore, the outage channel capacity is defined instead. The evaluation of outage capacity for a slow fading channel, as well as slow fading channels with receive antenna diversity (SIMO), transmit antenna diversity (MISO), both receive and transmit antenna diversity (MIMO), and OFDM are presented. For a fast fading channel, the average channel capacity is well defined since symbols fade independently and there exists a coding system that ensures a maximum reliable rate, which is the ensemble average rate. Fast fading is less detrimental than slow fading from the capacity point of view but requires more complex channel coding to deal with both Doppler phase error (virtually eliminated via Doppler tracking in slow fading) and random amplitude attenuation.

1.3 Summary

The structure of the book can be summarized as follows:

- **Acquiring** the prerequisite knowledge of communication signals: Chapters 2 and 3.
- **Packaging** the message and introducing the concept of signal-to-noise ratio and bandwidth: Chapter 4.

- **Measuring and establishing** the required signal-to-noise ratio for a given communication coverage: Chapter 5.
- **Sending** the message based on the required signal-to-noise ratio and bandwidth: Chapter 6.
- **Receiving** the message and **providing** the best detection: Chapter 7.
- **Enhancing** the survivability of narrowband modulation in the presence of interference via bandwidth spreading, i.e., spread spectrum modulation: Chapter 8.
- **Pulse shaping and equalizing** the effect of a bandlimited channel for modulated signals: Chapter 9.
- **Dealing** with sending and receiving signals over a mobile channel for previously discussed modulation techniques: Chapter 10.

Bibliography

1. C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, Vol. **27**, pp. 379–423, pp. 623–56, 1948. Also available at http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html.
2. R. W. Hamming, "Error detecting and error correcting codes," *Bell Syst. Tech. J.*, Vol. **29**, pp. 147–60, 1950.
3. R. W. Hamming, *Coding and Information Theory*, 2nd edition, Hemel Hempstead: Prentice-Hall, 1986.
4. R. G. Gallager, *Principles of Digital Communications*, Cambridge: Cambridge University Press, 2008.

2 Deterministic signal analysis

Introduction

In this chapter we lay the foundation for the analysis and design of communication systems, and digital communication systems in particular. We employ *deterministic signals* to carry information from the transmitter to the receiver. These deterministic signals contain certain *a priori* features sufficiently adequate for the receiver to retrieve the information. Note that the information always appears random to the receiver, that is, it does not know which data it will receive; otherwise, communications would not be needed. Deterministic signals form a very broad class of signals; therefore, the first step is to categorize them so that their characterization can be fully exploited. The categorization leads to the labels *continuous-time*, *discrete-time*, *periodic*, *aperiodic*, *analog*, *digital*, *energy*, and *power signals*. Further study leads us to *orthogonal signals* and the use of *signal space* to represent digital signals as vectors. We also review *linear time-invariant* (LTI) systems and the important *convolution* operation that relates the inputs and outputs of an LTI system.

We then investigate *Fourier series* representation of continuous-time periodic signals, and *Fourier transform* of continuous-time aperiodic signals. The Fourier transform is indispensable in the analysis and design of LTI systems. The *energy spectral density* of an energy signal and the *power spectral density* of a power signal are studied. From here the *autocorrelation* functions of both energy and power signals are examined.

The process of representing a continuous-time signal by its samples is then studied using the *sampling theorem*. We also discuss the process of recovering a continuous-time signal from its samples. Finally, we study various representations of *bandpass signals*, which are commonly used in the analysis of communication systems.

2.1 General description of deterministic signals

A deterministic signal is completely specified at any instant of time t. There is no uncertainty about its value at t. The transmitter employs deterministic signals to carry random information. When the receiver receives a transmitted signal that has been corrupted by noise (a random signal), it attempts to detect the information by stripping away the deterministic signals. A deterministic signal can fall into a number of categories, which are described below.