1 An Overview of Digital Communication Systems

1.1 Introduction

Until the late 1980s, analog modems were most widely used for data transmission over telephone wirelines. The techniques developed in digital communications from the 1960s through the mid-1980s were mainly targeted for wireline modems. Since then, various forms of digital transmission technologies have been developed and become popular for communication over digital subscriber loops, Ethernet, and wireless networks. The basic principles of synchronization techniques developed in the analog modem era have evolved. They are still being used directly or as the foundation of synchronization in other types of digital communication systems.

Since the mid-1980s, wireless has taken over wireline as the main form of communication connecting our society and permeating citizens' everyday life. Because of the widespread deployment of cellular communication systems, wireless technologies have been making impressive progress in all disciplines including synchronization.

Another important area of digital communications is satellite communication, which also experienced rapid progress during the same time period. While satellite communication has its unique properties, it also has many commonalities with wireline and wireless communications including those in the area of synchronization. In this book, we will focus mainly on the theories and techniques of synchronization for wireline and cellular-type wireless communication systems. However, what is discussed is also applicable to satellite communications.

To achieve bidirectional communications, two communication channels are needed. The two channels can be physically independent or can share the same physical media. Over wirelines, communications in the two directions can be either symmetric or asymmetric. For wireless, such as mobile communications, they are most likely asymmetric in the two directions. Communications from wireless base stations to mobile devices are usually called *forward link* communications. Communications in the other direction, i.e., from devices to the base stations, are usually called *reverse link* communications. In this book, most techniques discussed and examples given for wireless communications are assumed for the forward link, although what is discussed can be adapted for the reverse link as well.

In this chapter, an overview of the communication system and its main functional blocks is provided in Section 1.2. The details of the three major components of a typical communication system, i.e., the *transmitter*, the *channel*, and the *receiver*, are presented in Section 1.3. Section 1.4 provides a high-level view of the main synchronization functions in communication systems as an introduction to the subjects on which this book focuses. Sections 1.5 and 1.6 describe the basics of code division multiple access (CDMA) and orthogonal frequency-division multiplexing (OFDM) technologies to facilitate the discussion of synchronizations in communication systems that employ these technologies presented in later chapters. Finally, Section 1.7 summarizes what has been discussed and provides a summary of the common parameter notations used in this book to conclude this chapter.

1.2 A High-Level View of Digital Communications Systems

A digital communication system has three major components: a transmitter (Tx),[1] a receiver (Rx), and a communications channel. A high-level block diagram of such a typical digital communication system is shown in Figure 1.1.

The objective of a digital communication system is to send information from one entity, in which the information resides or is generated, through a communication channel to another entity, which uses the information. A high-level description of the process of how a communication system achieves information transfer from the source to the destination is given below. Details of their realization in different types of communications systems are provided later in this chapter.

A Brief Overview of Transmitter Operations

To be processed by the transmitter, the information from the source must be in binary form, i.e., represented by bits in the form of zeros and ones. If the information exists in nonbinary forms, such as in quantized audio or video waveforms, it is first converted to the digital form by source coding. The source-coded information or information that is already in the binary form may be encrypted and/or compressed. Such preprocessed source information in binary form is referred to as *information bits*, which are ready for further processing in the transmitter.

Inside the transmitter, the information bits are first processed in digital form. This step is commonly referred to as *digital baseband processing*. Its output is baseband signal samples, which are converted to analog baseband waveforms by a digital-to-analog converter (DAC).

The generated analog baseband waveforms are modulated onto a carrier frequency f_c to become *passband signals*. After being filtered and amplified, the passband signal at the carrier frequency is transmitted over a communication channel. The communication channel can be a wireless channel through radio wave propagation, or it can be a wireline channel such as a twisted pair of wires or cables.

[1] In some references, *Tx* is often used as a shorthand notation of transmitter. It can also be used to describe or specify any quantity that is transmitter related. The same usages can also apply to *Rx*.

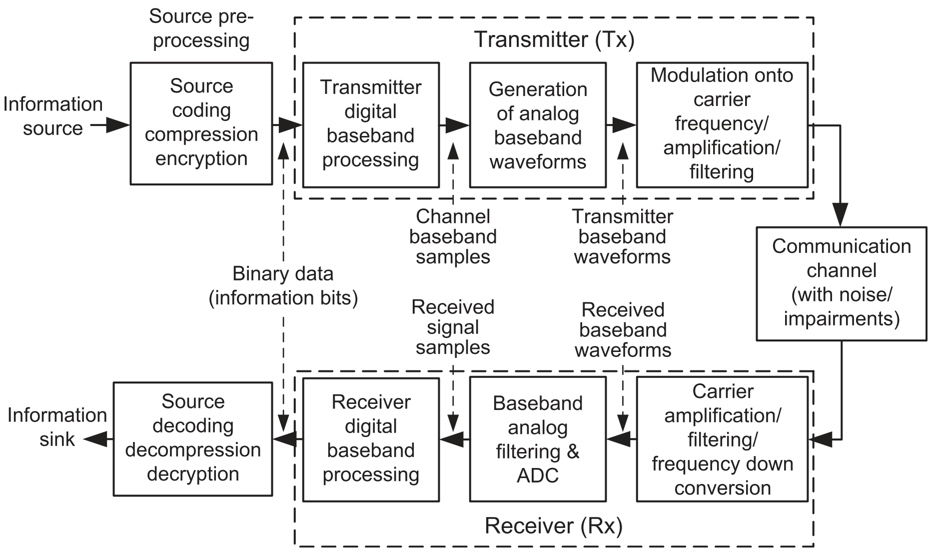

Figure 1.1 High-level block diagram of a digital communication system

A Brief Overview of Receiver Operations

After passing through the communication channel, the transmitted signal reaches the receiver at the other end of the channel. Communication channels always introduce various types of impairments including linear/nonlinear distortions, additive noise, and interference. As a result, the received signal is a distorted version of the transmitted signal. The task of the receiver is to recover the original transmitted information with no or little loss.

The first step of the receiving process is to convert the passband signal at carrier frequency to baseband. After performing passband filtering to reduce as much outband interference and noise as possible, the received signal is *frequency down-converted* or *demodulated* to become a baseband analog waveform.[2] The generated baseband analog signal can be expressed as the convolution of the transmitted baseband waveform and the channel impulse response (CIR) of the equivalent baseband channel.

The baseband analog waveforms of the received signal are analog filtered and are converted to digital samples. Then the baseband digital signal samples are processed by the receiver digital baseband processing block to regenerate the original information bits sent by the transmitter with no or few errors. The original source information is recovered from the regenerated information bits to complete the operations of the communication system.

The details of the operations performed in the blocks as just described are explained in the following.

[2] In digital communication terminology, the term *demodulation* commonly refers to the general operation of recovering the original data symbols from their modulated forms. In a loose sense, it may also mean the operations for data symbol recovery including frequency down-conversion, or carrier phase and/or frequency offset correction.

1.3 Major Components of Typical Digital Communication Systems

As discussed above, a digital communication system consists of a source information processing block, a transmitter, a channel, a receiver, and a source-information recovering block. In this section, we consider its three major components, which perform essential operations for achieving successful digital transmission, i.e., the transmitter, the channel, and the receiver.[3]

The functional blocks of the transmitter and the receiver and their operations described below are in the context of single-carrier digital communication.[4] However, in principle, what is discussed is also applicable to other digital communication systems with additional operations as necessary. Some digital communication technologies that are widely used in modern communication systems will be introduced later in this chapter. For example, details of the direct-sequence code division multiple access (DS-CDMA), which is a type of direct-sequence spread spectrum (DSSS) communication, and OFDM will be presented in Sections 1.5 and 1.6, respectively.

1.3.1 Transmitters and Their Operations

In this section, we describe various transmitter functional blocks and their operations for data transmission in single-carrier digital communication systems. The functional blocks considered are channel coding and interleaving, data symbol mapping, data packet forming and control signal insertion, spectrum/pulse shaping, digital-to-analog conversion, and carrier modulation and filtering. The block diagram of such a transmitter is shown in Figure 1.2.

1.3.1.1 Channel Coding and Interleaving

In order to improve the reliability of data transmission over the communication link, the *information bits* to be transmitted are first coded by a forward error correction (FEC) encoder. This step is called *channel coding*. Channel coding introduces redundancy into the input bits. Such redundant information is used by the decoder in the receiver to detect the information bits and to correct the errors introduced when the information was transmitted over the communication channel. Thus, a system with FEC coding can tolerate higher distortion and noise that are introduced during transmission than a system without it. In other words, a coded system can achieve more reliable communication at a lower signal-to-noise ratio (SNR) of the received signal than an uncoded system. The difference between the SNRs required by the two systems to attain the

[3] The source information processing and recovery blocks are sometimes viewed as associated with the transmitter and receiver, respectively. However, in order to concentrate on the aspects that are most pertinent to the subject matter of this book, their functions will not be considered here.

[4] Direct-sequence spread spectrum communication is also a type of single-carrier digital communication. To avoid confusion, in this book, the term "single-carrier communication" generally refers to non-DSSS single-carrier communication, unless otherwise specified.

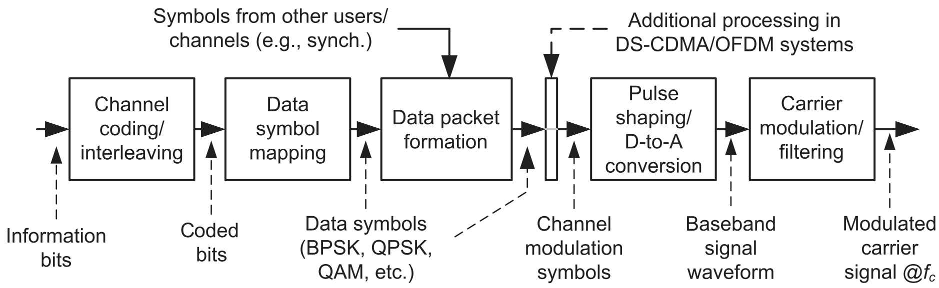

Figure 1.2 Block diagram of a typical transmitter

same error rate is called the *coding gain* with the SNR measured by the ratio of energy per information bit (E_b) divided by noise density (N_0), i.e., E_b/N_0.

Due to the introduced redundancy, the number of coded bits generated by the FEC encoder is always greater than the number of input information bits. Consequently, the ratio of the number of the input information bits divided by the number of the output coded bits, known as the *coding rate R*,

$$R = \frac{\text{number of information bits}}{\text{number of coded bits}} \tag{1.1}$$

is always less than 1 in order to achieve positive coding gain.

Channel coding is an important component in a digital communication system. Its purpose is to improve the performance of digital communication systems so that these systems can approach the theoretical limit given by the Shannon channel capacity [1]. In earlier years, the most commonly used FEC codes were algebraic block codes, convolutional codes, and trellis codes. Since the mid-1990s, turbo and low-density parity-check (LDPC) codes have become popular in communication system designs because of their ability to approach and/or achieve the Shannon channel capacity.

FEC is one of the most important disciplines of digital communications and has a very well developed theoretical foundation. Due to the scope of this book, it is not possible to cover all of its details. We will only mention some of its aspects related to synchronization when appropriate. Interested readers can find abundant information regarding the coding theory, code design, and code implementation from many references and textbooks including [2] and [3].

In many communication systems, especially those intended for communication over fading channels, the coded bits are interleaved before being further processed. Briefly, an interleaver changes the order of the input bits. As a result, the consecutive coded bits are distributed over a longer time period when transmitted. Interleaving makes the effects of impulsive noise and channel deep fading spread over different parts of the coded bit stream, so that the impact of the noise and fading on the decoding in the receiver is reduced.

Because an interleaver only changes the order of the coded bits, the numbers of the input and output bits are the same. Thus, the interleaver can be viewed as a rate one encoder, which provides no coding gain over static channels.

The most popular interleavers used for digital transmissions are *block interleavers* [4, 5] and *convolutional interleavers* [6, 7]. Similar to error correction coding, interleaving does not directly affect synchronization. We will not elaborate on its functions further. The interested reader can find the relevant information in the available references.

1.3.1.2 Data Symbol Mapping

In order to be modulated and transmitted over a communication channel, the binary coded bits are usually grouped to form *data symbols*. Thus, each data symbol can represent one or more coded bits. Depending on the modulation technique used for transmission, the data symbols may be represented by a symbol constellation, as shown in Figure 1.3. The most popular types of modulation are amplitude modulation (AM), binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK), multiple phase-shift keying (MPSK), and quadrature amplitude modulation (QAM).

Note that the bits mapped to the constellation points shown in the figure satisfy a special property such that any two adjacent constellation points differ by only one bit. This is called *Gray mapping* or *Gray coding*. This property is important to achieving good system performance. This is because at the receiver a symbol error is most likely to occur when the adjacent constellation points are misinterpreted as the constellation point of the actually transmitted symbol. With Gray mapping, such a symbol error with high probability will result in only a one-bit error, which is easier to be corrected by channel decoding.

Depending on the symbol constellations used, different numbers of coded bits are mapped to one of the constellation points. The simplest modulation constellations are BPSK and QPSK. The BPSK constellation is used if one bit is mapped to a data symbol. A bit 0 is mapped to the real value 1, and a bit 1 is mapped to the real value -1. The QPSK constellation is used if two coded bits are mapped to a data symbol corresponding to a point in a QPSK constellation. For example, 00 are mapped to the complex point 1+j, 10 to -1+j, etc.

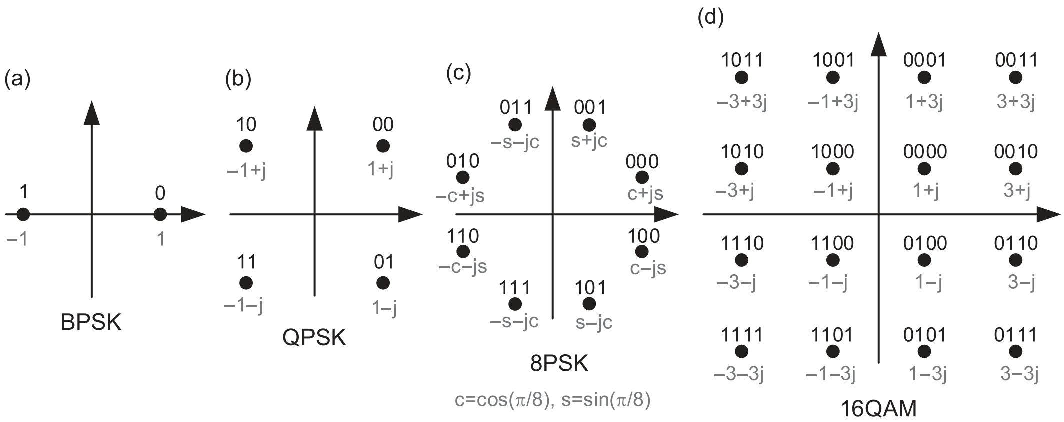

Figure 1.3 BPSK, QPSK, 8PSK, and 16QAM symbol constellations

When each data symbol represents more than two coded bits, MPSK and QAM constellations are often used. An MPSK constellation consists of 2^M points evenly distributed on a circle centered at the origin, and each data symbol corresponds to M coded bits. The 8PSK constellation is shown in Figure 1.3(c). As shown there, the triplet 000 maps to the constellation points of the complex number $\cos(\pi/8) + j\sin(\pi/8)$, 001 maps to $\sin(\pi/8) + j\cos(\pi/8)$, etc.

The most popular QAM constellations have 2^N constellation points on a square lattice grid. Therefore, each constellation point represents a data symbol mapped from an N-tuple of coded bits. The 16QAM constellation with four-bit Gray mapping and the coordinates of its constellation points in the complex plane is shown in Figure 1.3(d).

So far in this section, we have assumed that the output of the channel encoder is coded binary bits, and these bits are mapped to complex-valued data symbols. There are also channel encoders that directly generate complex data symbols from the information bits. One example is trellis coding, which was widely used during the 1980s and early 1990s. However, from the viewpoint of synchronization, only the form of the channel modulation symbols is relevant. Therefore, in the rest of this book, we are interested only in the type of data symbol constellations that is used in the transmission regardless of how the symbols are generated.

In the literature, data symbols as described above are often referred to as data modulation symbols. However, to distinguish the data modulation symbols from the channel modulation symbols to be discussed below, we will refer to them simply as data symbols. For single-carrier communications, data and channel modulation symbols are the same. However, they are different for OFDM and CDMA communication systems.

1.3.1.3 Formation of Data Packets/Streams for Transmission

The data symbol sequence as generated above can be directly used to generate baseband signal waveforms. However, in many digital communication systems, additional information may need to be transmitted together with the basic data. For example, in multiple access systems, there may be data from more than one user that need to be transmitted. Moreover, there are control signals such as signals for synchronization that need to be transmitted as well.

Data symbols carrying signaling/data information are generated in ways similar to those described in Section 1.3.1.2. The data and signaling symbols that are transmitted together may use different symbol constellations. For example, the data symbols may have 16QAM or 64QAM constellation in order to improve the transmission spectrum efficiency. In contrast, the signaling symbols, which are used to facilitate synchronization and for other purposes, such as pilot or reference symbols for performing channel estimation, are often transmitted in BPSK or QPSK. Nonetheless, their symbol rates of transmission are usually the same.

Different types of communication systems may have different data packet/stream transmission formats. For example, in wireline digital transmission, the control signals for training the receiver are usually sent at the beginning of a data communication session. Then regular data are transmitted continuously as a long symbol stream. Because the wireline channel does not change much in a short time duration, no additional training signals are sent during the normal data transmission/receiving session to improve the efficiency of the system.

In contrast, for wireless communication, the channel often experiences fast fading. As a result, the data symbols to be transmitted are usually organized in packets, which are transmitted consecutively to form a long data stream. Each packet has a short time span and is embedded with synchronization signals. In the receiver, these synchronization signals, such as pilot symbols, are used to perform channel estimation for the estimation of data symbols, as well as to achieve carrier and timing synchronization. These signals can also be used to obtain other information from systems that the receiver needs to communicate with.

1.3.1.4 Generation of Baseband Signal Waveform – Spectrum/Pulse Shaping

Transmitter channel modulation symbols, or simply *channel symbols*, are generated from the data symbols described above and converted to analog baseband waveforms. A channel symbol is represented by a complex number belonging to one of the modulation symbol constellations, including BPSK, QPSK, MPSK, QAM, etc. The baseband analog signal is generated by multiplying channel symbols with time-shifted continuous-time pulses that satisfy certain time- and frequency-domain characteristics. The Fourier transform of the most commonly used time pulse, denoted by $g_T(t)$, for generating analog baseband signal $x(t)$, is a square-root raised cosine (SRCOS) function, i.e.,

$$G_T(f) \cong \mathrm{FT}[g_T(t)] = \begin{cases} 1 & |f| \le \dfrac{(1-\beta)}{2T} \\ \sqrt{0.5\cos\left[\dfrac{\pi T}{\beta}\left(|f| - \dfrac{1-\beta}{2T}\right)\right] + 0.5} & \dfrac{(1-\beta)}{2T} < |f| \le \dfrac{(1+\beta)}{2T} \\ 0 & |f| > \dfrac{(1+\beta)}{2T} \end{cases} \tag{1.2}$$

The time pulse $g_T(t)$ in the transmitter is actually generated by multistage processing, which will be mentioned later. It has a large peak at $t = 0$ and decays to zero when t goes to positive or negative infinity.

In single-carrier communication systems, the data symbols are directly used as the channel modulation symbols to generate the analog baseband signals. Thus, no complex conversion is needed, except maybe a single scaling operation. However, in DS-CDMA and OFDM communication systems, additional processing steps are performed to convert the data symbols to the suitable forms of channel symbols.

Generating baseband signal pulses with the desired spectrum from the channel symbols is often called *spectrum shaping* or *pulse shaping*. This process is the same for all three types of communication systems mentioned above. A block diagram of baseband waveform generation is shown in Figure 1.4(a).

Channel modulation symbols can be viewed as complex-valued random variables. They are converted to the analog form by *digital-to-analog converters*. The output of a DAC may be analog voltage impulses or rectangular pulses. Every T second, two

DAC outputs are generated for every complex channel symbol, one from its real part and the other from its imaginary part. To simplify our discussion, we will assume that the DAC outputs are impulses. Practically, it is more common to use the "sample-and-hold" DACs to generate output in the form of rectangular pulses with a width equal to T. The description given below is also applicable to such DACs with minor modifications.

The power spectrum of a channel symbol sequence has a constant magnitude as it is assumed that the channel symbols are, by design, uncorrelated in most systems. For T-spaced symbols, the spectrum is periodic with a period of $1/T$ [8] as shown in Figure 1.4(b).

It is possible to generate the analog baseband pulses by suitable analog low-pass filtering of the analog impulses generated by DACs directly from the symbol sequence as indicated in Figure 1.4(b). Because the analog filter performs both spectrum/pulse shaping and image rejection, it is quite demanding to implement. In practice, it is more efficient to perform spectrum shaping first by digital low-pass filtering. Once converted to the analog domain, the spectral images can be rejected by using a simple analog filter as described below.

To perform digital spectrum shaping, the channel symbol sequence is first up-sampled by inserting zeros between adjacent symbols. As an example, we consider the simplest case that one zero is inserted between any two adjacent channel symbols. The spectrum of the symbol sequence does not change after zeros are inserted. However, because the sample spacing is now equal to $T/2$, the spectrum of the new sequence has a base period of $2/T$ as shown in Figure 1.4(c). The zero-inserted sequence is filtered by a digital low-pass filter, which has the desired frequency response such as SRCOS previously mentioned. Figure 1.4(d) shows the power spectrum of the sequence at the filter output.

DACs are used to convert the digital sample sequence at the digital low-pass filter's output to $T/2$ spaced analog impulses, which have the power spectrum shape shown in Figure 1.4(d). Given that the spectrum has a period of $2/T$, the image bands are located at $2m/T$ Hz, $m = \pm1, \pm2, \ldots$ A simple analog low-pass filter shown in the figure filters the impulse sequence to retain the baseband signal spectrum centered at zero frequency and removes the image spectra. The output of the analog low-pass filter is the desired baseband signal waveform as shown in Figure 1.4(e).

In practical implementations, sample-and-hold circuits are usually incorporated into DACs to generate rectangular pulses. For $T/2$ spaced input digital samples, the widths of the rectangular pulses are equal to $T/2$. Thus, the power spectrum at the DAC output is equal to the spectrum shown in Figure 1.4(d) multiplied by a window of $\sin^2(\pi Tf)/(\pi Tf)^2 = \mathrm{sinc}^2(Tf)$.[5] As a result, the power spectrum of the actual analog waveform is no longer the same as the squared frequency response of the digital spectrum shaping filter.

[5] In this book, the sinc function is defined as $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$, commonly used in information theory and digital signal processing, such as in Matlab. It can also be defined as $\mathrm{sinc}(x) = \sin(x)/x$ in some other disciplines in mathematics.

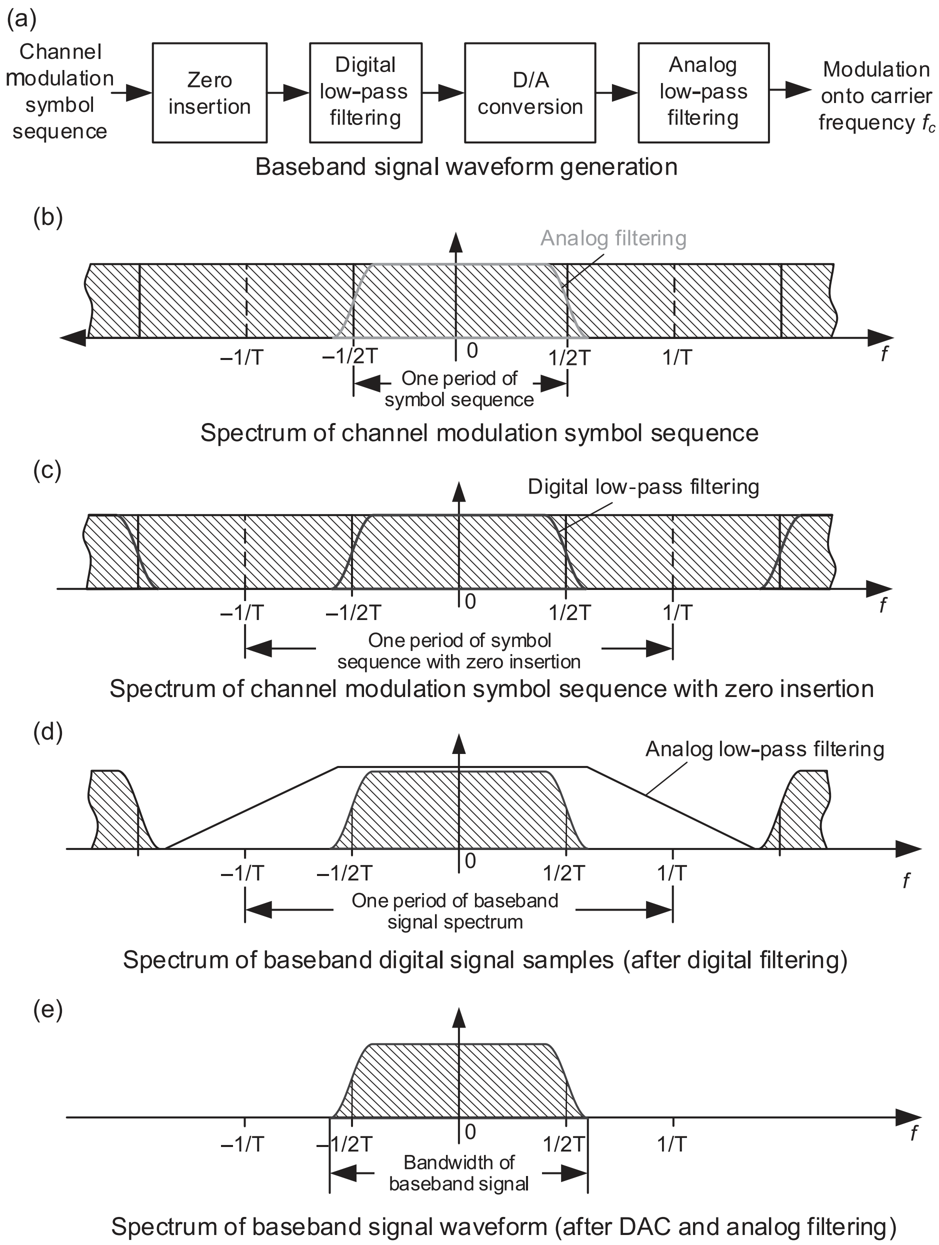

Figure 1.4 Baseband waveform generation

Such distortions in signal spectra can be either precorrected by properly designed spectrum shaping filters or post-corrected by appropriate analog low-pass filters. Practically, the most convenient remedy for the distortion introduced by the rectangular analog pulse is to increase the number of the zeros between the channel symbols, i.e., to increase the upsampling frequency. By such a design, the signal spectrum occupies only