CHAPTER 3

PHYSICAL LAYER ALTERNATIVES FOR WIRELESS NETWORKS

3.1 INTRODUCTION

In this chapter we describe transmission technologies, which have been adopted in many of the developing standards and products for wireless networks. In principle these techniques are applicable to all wired and wireless modems because the basic design issues are common to both systems. In general we would like to transmit data with the highest achievable data rate with the minimum expenditure of signal power, channel bandwidth, and transmitter and receiver complexity. In other words we usually want to maximize both bandwidth efficiency and power efficiency and minimize the transmission system complexity. However, the emphasis on these three objectives varies according to the application requirement and medium for transmission, and there are certain details that are specific to particular applications and media of transmission. Also these design objectives are often conflicting and the trade-offs decide what factors are considered more important than others. We start this chapter with a brief description of specific characteristics of the wireless medium that affect the design of transmission techniques. Then we provide a comprehensive overview of applied wireless transmission techniques, followed by a review of coding techniques and a brief description of software implementation of these radios.

3.1.1 Wired Transmission Techniques

In most wired data applications, such as LANs, transmission schemes over TP, coaxial cable, or optical fiber are very simple. The received data from the higher layers are line coded (e.g., Manchester coded on Ethernet) and the voltages (or optical signals) are applied directly to the medium. These transmission techniques are often referred to as baseband transmission schemes. In voice-band modems, DSL, and coaxial cable modem applications, the transmitted signal is modulated over a *carrier.* The amplitude, frequency, phase of the carrier, or a combination of these is used to carry data. These digital modulation schemes are correspondingly called amplitude shift keying (ASK), frequency shift keying (FSK), phase shift keying (PSK), or quadrature amplitude modulation (QAM). In voice-band modems, this carrier is around 1,800 Hz which is the center of the passband of 300–3,300 Hz of the telephone channels. The purpose of modulation here is to eliminate the DC component from the transmission spectrum and to allow the usage of more bandwidth-efficient modulation to support higher data rates over the telephone channel. For DSL services, the spectrum that is utilized is shifted away from the lower frequencies used for voice applications. Discrete multitone transmission, a form of OFDM, is employed there. In cable modems, modulation is employed to shift the spectrum of the signal to a particular frequency channel and to improve the bandwidth efficiency of the channel to support higher data rates. In the data networking industry, cable modems are referred to as broadband modems because they provide a much higher data rate (broader band) than the voice-band modems. High bandwidth efficiency in the voice-band modems has a direct economic advantage to the user, as it can reduce connect time and avoids the necessity for leasing additional circuits to support the application at hand. The typical telephone channel is less hostile than a typical radio channel, providing a fertile environment for examining and employing complex modulation and coding techniques such as QAM and *trellis coded modulation* (TCM) and signal processing algorithms such as equalizers and echo cancellers. Specific impairments seen on telephone channels are amplitude and delay distortion, phase jitter, frequency offset, and effects of nonlinearity. Many of the practical design techniques of wired modems have been developed to efficiently deal with these categories of impairments.

3.1.2 Considerations in the Design of Wireless Modems

Radio channels are characterized by multipath fading and Doppler spread, and a key impediment in the radio environment is the relatively high levels of average signal power needed to overcome fading. However, there are other considerations that impact the selection of a modem technique for a wireless application. For example, in radio systems, bandwidth efficiency is an important consideration, because the radio spectrum is limited, and many operational bands are becoming increasingly crowded. There are a number of considerations that enter into the choice of a modulation technique for use in a wireless application, and here we briefly review the key requirements. These requirements can vary somewhat from one system to another, depending on the type of system, the requirements for delivered services, and the users' equipment constraints.

3.1.2.1 Bandwidth Efficiency

Most wireless networks that support mobile users have a need for bandwidth-efficient modulation, and this requirement steadily grows in importance each year. One of the major incentives of the cellular telephone industry for moving from analog to digital and then from TDMA to CDMA was to increase the bandwidth efficiency and consequently the number of users. A cellular carrier company is assigned a specified amount of licensed bandwidth in which to operate their system, and therefore an increase in system capacity leads directly to increased revenues.

This defines another clear need for modulation techniques that provide efficient utilization of available bandwidth.

An area of wireless communications network development where the modulation bandwidth efficiency is not yet as critical as that of the cellular industry is that of WLANs. Unlike cellular systems, which to date have been used to support circuit-switched services, WLANs are typically used for burst-mode traffic. Due to the bursty nature of the user data, the aggregate data traffic on a WLAN rarely approaches system capacity. Furthermore, almost all WLAN products operate in the unlicensed bands, where the same frequencies are reused again and again in relatively close geographic areas. It is for these reasons that the WLAN industry had placed relatively little emphasis on bandwidth-efficient modulation techniques.

3.1.2.2 Power Efficiency

Power efficiency is another parameter, which is not of major importance for wireless equipment using AC power sources, such as interLAN bridges, but is of crucial importance in battery-oriented applications such as handheld cellular or WLAN cards used in laptops. In these applications power consumption translates into battery size and recharging intervals, and even more important to the mobile user, into the size and weight of portable terminals. Figure 3.1 demonstrates how the power consumption is directly related to device size and weight. Power effi-

Figure 3.1 Power consumption and the size/weight of a mobile terminal.

ciency will become increasingly important as consumers become accustomed to the convenience of small, handheld communication devices.

There are two facets of the power requirement: One is the power needed to operate the electronics in the terminal; the other is the amount of power needed at the output of the power amplifier in order to radiate a given amount of signal power from the antenna. The radiated signal power, of course, translates directly into signal coverage, and it is a function of the data rate and the complexity of the receiver. Higher data rates require higher operating levels of SNR. More complex systems using computationally demanding coding techniques such as TCM or employing adaptive equalization require less transmission power. However, a more complex receiver design increases the power consumed by the electronics and consequently reduces battery life. In some applications a compromise has to be made between the complexity of the receiver and the electronic power consumption. For example, in handheld local communication devices some manufacturers avoid the use of complex speech coding techniques in order to reduce battery consumption. Also, in the design of high-speed data communication networks for laptop or handheld computers, some designers find it difficult to justify the additional electronic power consumption required for the inclusion of adaptive equalization algorithms.

In spread spectrum CDMA systems, power efficiency and overall system bandwith efficiency are closely related. The use of a more power-efficient modulation method allows a system to operate at lower SNR. The performance of a CDMA system is limited by the interference from other users on the system, and an improvement in power efficiency in turn increases the bandwidth efficiency of the system.

3.1.2.3 Out-of-Band Radiation

An important issue in the selection of a modulation technique for a radio modem is the amount of transmitted signal energy lying outside the main lobe of the signal spectrum. In cellular operation the performance is limited by adjacent-channel interference rather than additive noise. Figure 3.2 illustrates the situation: two users—one close to the BS antenna and the other at a larger distance—are operating in two adjacent channels. The out-of-band interference of the mobile closer to the antenna is a serious source of interference for the mobile located in a farther distance.

The adjacent-channel interference (ACI), which is the interference that a transmitting radio presents to the user channels immediately above and below the transmitting user's channel, is a major parameter in the design of cellular systems. The ACI determines the geographic area where mobile users can be served by a single base station. As shown in Figure 3.2, a low level of ACI will permit a distant mobile transmitter to reach the base station with a weak signal while another mobile much closer to the base station is transmitting in an adjacent channel. Thus ACI specifications will indirectly influence system capacity and cost. The characteristics of the transmit and receive channel filters, nonlinearity in the transmitter, and of course the height and roll-off characteristics of the skirts of the transmitted signal spectrum influence the level of ACI. Radio manufacturers strive to design radios that keep ACI below a specified level, typically −60 dB below the main lobe, and the out-of-band spectral power of the modulation scheme is the principal ingredient in achieving that goal. In contrast, the out-of-

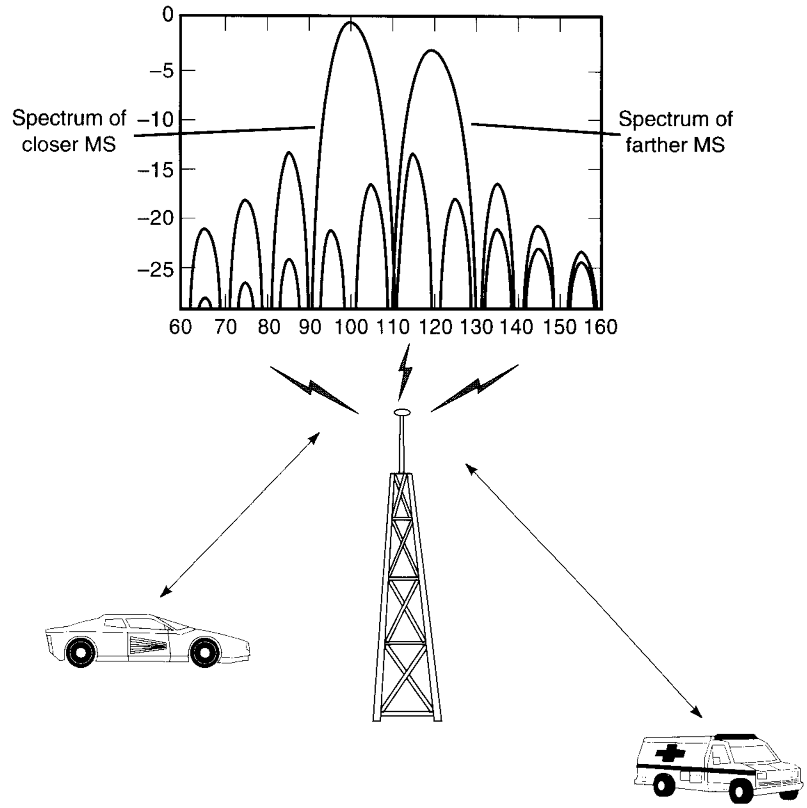


Figure 3.2 Illustration of adjacent channel interference.

band signal power in voice-band modems is not as critical and a voice-band modem manufacturer would be satisfied with an out-of-band power of around -40 dB below the main lobe.

3.1.2.4 Resistance to Multipath

Another important issue in the design of a radio modem is sensitivity to multipath. Various modulation techniques have different degrees of resistance to multipath. This was a major issue in the development of the digital cellular and PCS standards, where it was necessary that each standard be written to accommodate the worst-case multipath conditions likely to be encountered by users over the entire geographic region of usage for that standard.

3.1.2.5 Constant Envelope Modulation

Most mobile radio products are designed with Class-C power amplifiers, which provide the highest power efficiency among the common types of power amplifiers. However, Class-C amplifiers are highly nonlinear, and therefore it is necessary that the signal to be amplified has a constant-envelope or as nearly so as is practical. Though analog frequency modulation (FM) mobile radio systems were originally designed for analog voice, they have been extended to digital service simply by feeding baseband data streams to the frequency modulator. The FM modulation always provides a constant envelope for the transmitted signal, and the information is contained in the variations of the frequency of the transmitted signal. A popular modulation technique developed this way is Gaussian minimum shift keying (GMSK) which will be described in the next section. An FM signal by its very nature has a constant-envelope, but it is not spectrally efficient due to its large side lobes. Thus as the need for greater bandwidth efficiency had grown, efforts have been made to design modulation schemes that are less wasteful of bandwidth while preserving (or nearly so) the constant-envelope nature of FM. The $\pi/4$-quadrature phase shift keying (QPSK) modulation has struck a good balance, and it has emerged as another popular transmission technique in wireless networks. This modulation technique is also described in the next section.

3.2 APPLIED WIRELESS TRANSMISSION TECHNIQUES

As we saw in Chapter 1, 1G wireless cellular and cordless telephone systems used analog FM modulation. With the emergence of 2G wireless networks, digital techniques replaced analog modulation. To increase the capacity, analog voice was source coded into digital format at the mobile terminal for digital transmission, then POTS transmits the analog voice to the network where it is digitized for long haul transmission. Speech coding at the terminal also facilitates the integration of voice and data services in a single terminal. After the emergence of 2G systems, digital transmission has become the dominant choice for wireless communication networks. Therefore, in the rest of this chapter we describe digital transmission techniques applied to modern wireless networks.

Popular digital wireless transmission techniques can be divided into three categories according to their applications. The first category is pulse transmission techniques used mostly in IR applications and more recently applied to the so-called *impulse radio* or ultra wideband (UWB) transmission [SCH00]. The second category is basic modulation techniques widely used in TDMA cellular, as well as a number of mobile data networks. The third category is spread spectrum systems used in the CDMA, as well as wireless LANs operating in ISM bands. More recently, new transmission technologies are emerging to increase the data rate of all these systems to support broadband access for popular Internet access and other data-oriented applications. Variations of the spread spectrum technology and OFDM have been adopted by WLAN standard organizations and are considered for incorporation into the future voice-oriented cellular networks. In the following sections, we provide a detailed description of all these popular modulation techniques to provide the reader with an understanding of the applied wireless physical layer alternatives.

3.3 SHORT DISTANCE BASEBAND TRANSMISSION

In baseband transmission, the digital signal is transmitted without modulating with a carrier. Without a carrier signal, multiple channels cannot be accommodated in an FDM format, and consequently each user occupies the entire available bandwidth of the medium. With one user occupying the entire band, the designer does not need to pay attention to the out-of-band radiation. However, to arrange a multiuser environment, one has to resort to innovative techniques other than the simple, traditional FDM. There are two basic steps in baseband transmission: line coding and pulse modulation. In the first step, the digital data stream is line coded to facilitate synchronization at the receiver and avoid the DC offset during transmission. Baseband line coded signaling is commonly used in short distance wired as well as wireless applications. Wired and IR-based WLANs often only employ line-coded baseband transmission. In pulse modulation, the transmitted information is coded into amplitude, location, or duration of a pulse shape. Pulse-modulation baseband transmission is commonly used for low-speed IR data communications such as remote controls or connections between personal computers (PCs) and the printers or keyboards. More recently UWB pulse modulation is being considered for very low power short-range radio communications.

If the data stream produced by a computer is applied directly to the wires, the receivers will have difficulty in sychronizing with the transmitted symbols. To provide better synchronization at the receiver, the format of the incoming data stream is modified before transmission. This modification process is often referred to as *line coding.*

In wired applications, baseband signaling using differential Manchester line coding is used in the IEEE 802.3 Ethernet, the dominant standard for LANs, as well as IEEE 802.5 Token Ring, the competitor of Ethernet in the early days of the LAN industry. In wireless applications baseband transmission with line coding is popular in high-speed diffused and directed beam IR wireless LANs.

Example 3.1: Manchester Encoding in an IR Transmitter

Figure 3.3 shows the Manchester code implementation of an IR transmitter used in a number of IR-based WLANs. The received nonreturn to zero (NRZ) digital stream is first Manchester encoded. The line coded signal is then intensity modulated by the emitted IR light by simply turning the transmitted light to on and off positions. The receiver consists of a simple photosensitive diode detecting presence and lack of the light and either producing or not producing an electric signal that is then amplified and used as the received signal. The Manchester coded data double the transmission rate but provide one transition per bit. These transitions are important for the receiver because the receiver uses them to synchronize its clock with that of the transmitter clock. In the NRZ transmission if we have a long stream of "0"s of "1"s, the receiver loses the reference timing of the transmitter.

Base band pulse modulation techniques are usually divided into pulse-amplitude-modulation (PAM), pulse-position-modulation (PPM), and pulse-

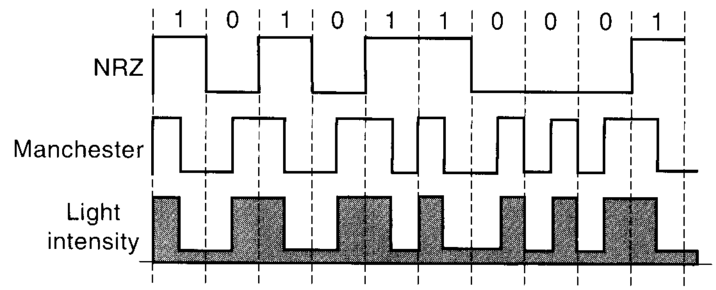


Figure 3.3 Manchester code implementation of an IR transceiver.

duration- or width-duration-modulation (PDM or PWM). As the names suggest, in PAM, PPM, and PDM the transmitted information is signaled through the amplitude, position (location), and duration of a basic pulse shape. Because wireless channels suffer from extensive amplitude fluctuations caused by fading and near-far problems, PAM is not popular in wireless operations, leaving variations of PPM and PDM as favorite choices for pulse modulation over wireless channels.

Example 3.2: Practical Implementation of PPM

A practical implementation of PPM and PDM used in applications such as connecting a keypad to a computer is shown in Fig. 3.4. The data stream is encoded to a pulse at the start of a bit to represent a "1" and a pulse in the middle to represent a "0" to generate the PPM signal. Rather than one single pulse, multiple narrow pulses are transmitted to code the transmission of digitized information. When detected by the photosensitive diode at the receiver, multiple narrow pulses will produce a single continuous pulse that is close to what would have been received if a single pulse was transmitted. However, with multiple narrow pulses, the required transmission power is smaller because the light is on for a shorter period of time. Therefore, multiple pulse transmission saves the life of the battery at the transmitter. In Figure 3.4 one of the options uses three narrow pulses and the other one uses nine narrow pulses. A disadvantage of using several pulses per symbol is the large bandwidth occupied by each pulse. However, this is not too important in IR applications.

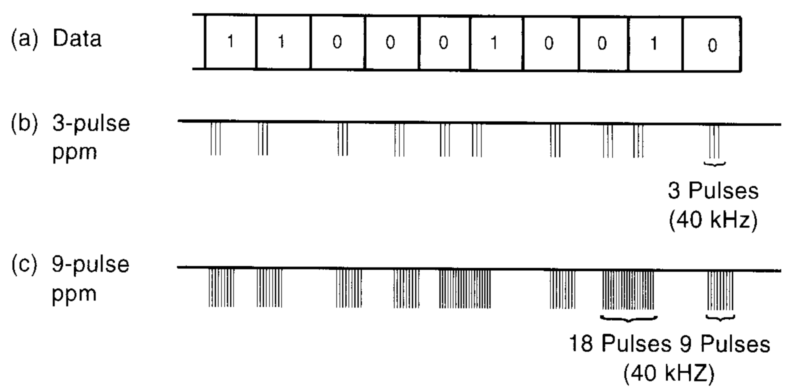


Figure 3.4 Practical implementation of PPM.

1 Mbps

0000 → 0000000000000001
0001 → 0000000000000010
0011 → 0000000000000100
0010 → 0000000000001000
⋮

2 Mbps

00 → 0001
01 → 0010
10 → 0100
11 → 1000

Figure 3.5 PPM using 250 ns pulses for IEEE 802.11.

Example 3.3: The IEEE 802.11 IR Standard

The IEEE 802.11 standard specifies a standard physical layer for high-speed diffused IR medium using PPM. The wavelength range specified by the standard is 850 nm–950 nm. The basic data rates are 1 and 2 Mbps which are consistent with the other two options using spread spectrum technology in ISM bands. As shown in Figure 3.5, the 1 Mbps physical layer uses 16-PPM, and the 2 Mbps uses 4-PPM. The width of each pulse for both cases is 250 ns. In 16-PPM, a 250 ns pulse occupies one of 16 positions, the duration of one symbol being $16 \times 250 =$ 4 microseconds. Each symbol corresponds to four bits (16 symbols $=> \log_2 16 =$ 4 bits/symbol). The result is 4 bits being transmitted every 4 microseconds or a data rate of 1 Mbps. Similarly, in 4-PPM, two data bits are transmitted every 1 microsecond for a data rate of 2 Mbps.

3.4 UWB PULSE TRANSMISSION

Impulse radio [SCH00], [MIT01] has recently attracted considerable attention for short-range communications. In this technique, a very narrow width (on order of a few tenths of a nanosecond) and low power (high-duty cycle of several hundreds of nanoseconds) pulse are used for information transmission. The spectrum of this pulse obviously occupies a very wide band (several GHz), and for that reason this technology is sometimes referred to as ultra wide band (UWB). The spectral height of the UWB signal is very low because a small transmission power is spread over a large bandwidth. Therefore, UWB signals can be designed to coexist with existing radio systems. As we saw in Chapter 2, signal fading is caused by the overlap of the received signal from different paths. Because of its extremely wide bandwidth and large duty cycle, the UWB signal isolates (resolves) multipath components, resulting in a stable received power signal with minimal fading effects. Implementation of the transmitter does not involve modulation, and if carefully designed, it can be much simpler than traditional narrowband or wideband spread spectrum systems.

Designers of UWB systems chose a transmission waveform that (1) has such a high bandwidth and signal processing gain that interference from existing systems into the system is negligible; (2) its spectral height is comparable with background noise so that the FCC allows it to coexist with the existing systems; (3) its implementation is easy; and (4) the spectrum looks like a pass band transmission so that there is no DC component in the spectrum.

Example 3.4: UWB Pulse Shape

Recently, Time Domain Corporation (TDC) has developed a range of UWB systems [TDCweb] trademarked as PulsON technology. TDC's UWB transmitters emit ultra-short "Gaussian" monocycles with tightly controlled pulse-to-pulse intervals. TDC has been working with monocycle pulse widths of between 0.20 and 1.50 nanoseconds and pulse-to-pulse intervals of between 25 and 1,000 nanoseconds. These short monocycles are inherently UWB. The pulse shape used by TDC is mathematically given by:

$$v(t) = 6A\sqrt{\frac{e\pi}{3}}\,\frac{t}{\tau}\,e^{-6\pi\left(\frac{t}{\tau}\right)^2} \tag{3.1}$$

where A represents the peak amplitude of the pulse, τ is a constant determining the width of the pulse, and t is the time. The spectrum of the monocycle pulse in frequency domain is given by:

$$V(f) = -j\frac{2f}{3f_c^2}\sqrt{\frac{e\pi}{2}}\,e^{\frac{\pi}{6}\left(\frac{f}{f_c}\right)^2} \tag{3.2}$$

where $f_c = 1/\tau$ is the center frequency of the pulse. Figure 3.6 shows a typical graph of the pulse for $\tau = 0.5$ ns associated with a center frequency of $f_c = 2$ GHz. The half power (3-dB) bandwidth of the pulse occupies around 2 GHz of bandwidth. The applications suggested for this technology include precision geolocation and higher performance radar.

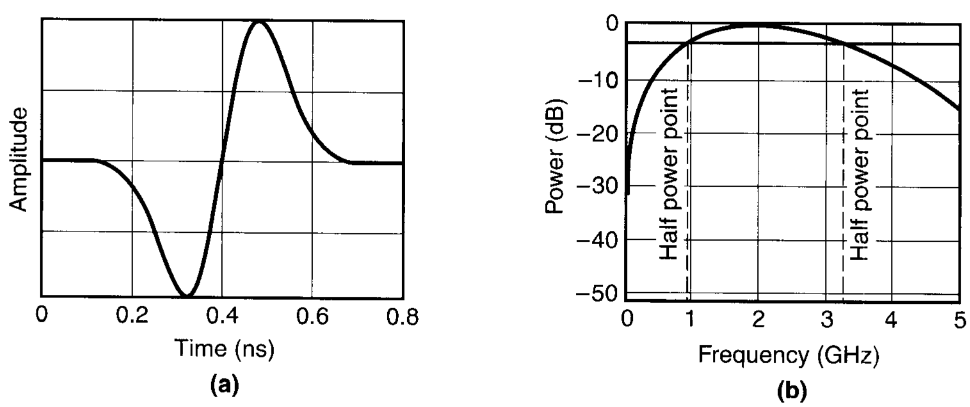


Figure 3.6 (a) The UWB pulse and (b) its spectrum.

3.5 CARRIER MODULATED TRANSMISSION

In broadband signaling, the message signal is mixed with a *carrier signal* at a higher frequency before transmission. The wired computer communication community is dominated by baseband systems, and the word broadband is used for all carrier-modulated systems. In the wireless community, however, the dominant technique is carrier modulation, and the word broadband is used when the transmission rate approaches a very high value. Carrier modulation shifts the spectrum of the transmitted signal to the location of the carrier in the spectrum allowing orderly coexistence of a number of transmissions via FDM. In wired networking, we can always add new wiring for new applications. We have telephone wiring for our telephones, cable for video distribution, and LAN wiring for data communications. In wireless networking, we have only one medium (air) that must be shared among a variety of applications using FDM. Our television, AM/FM radios, cordless and cellular phones, remote controls, and other wireless appliances share the same medium and are separated only by their carrier frequencies. Therefore, carrier modulation and FDM is the cornerstone of multiservice radio communications.

Carrier modulation also shifts the frequency operation to higher values providing better coverage and reducing the length of the antenna to a practical size. The size of the antenna is usually on the order of the transmission wavelength. As the carrier frequency is increased, the wavelength and, consequently, the size of the antenna, reduce. As the carrier frequency increases, there are wider frequency bands available to support higher data rates. However, with increasing carrier frequency, the design of RF circuits becomes more challenging, and also in-building penetration of the signal becomes smaller. Designers of wireless modems need to make a compromise among availability of frequency of operation, required bandwidth, coverage of indoor and outdoor areas, and the cost of implementation. There are two classes of carrier-modulated technologies that are used in a wireless network: traditional radio modem and spread spectrum modems. In the next four sections, we provide a description of traditional and spread spectrum carrier modulated techniques that are applied to voice- and data-oriented wireless networks.

3.6 TRADITIONAL DIGITAL CELLULAR TRANSMISSION

A number of alternatives for digital transmission over radio channel were evaluated for modern wireless communications in the past two decades. In their broadest form, carrier modulation techniques can be divided into three categories of amplitude-, frequency-, and phase-modulation techniques. Radio modems need low side lobes for operation using FDM, and they have to cope with extensive amplitude fluctuations caused by fading. As a result, amplitude modulation techniques are not desirable, and frequency and phase digital modulation techniques have emerged as the traditional techniques in this industry. The two most commonly used modulation techniques in traditional wireless networks are GMSK and $\pi/4$-QPSK modulation techniques. GMSK is adopted by GSM, today's most popular digital cellular standard, and a number of other wireless data networks such as CDPD and Mobitex. The $\pi/4$-QPSK modulation is adopted by the North American TDMA digital cellular standard, IS-136, and the Japanese digital cellular, JDC, as well as the European mobile data service called TETRA. In the next two subsections, we address evolution of digital frequency modulation techniques to GMSK and digital phase modulation techniques to $\pi/4$-QPSK.

3.6.1 Digital Frequency Modulation and GMSK

As we noted earlier in this chapter, FM is the predominant form of analog modulation used in the mobile radio industry. Digital FM modulation is referred to FSK which forms a simple and popular method for wireless communications. Figure 3.7(a) shows the basic concept behind binary FSK modulation. The binary baseband data stream is encoded into two different frequencies before transmission in the channel. To implement this modulation in it simplest form, as shown in Figure 3.7(b), one can input the binary data stream directly to a traditional analog FM transmitter and use an analog FM receiver to demodulate the signal at the receiver. In its ideal form, an analog FM transmitter linearly maps the instantaneous amplitude of the message signal to a constant amplitude sinusoid with varying frequency at the output of the transmitter. A binary input signal takes only two levels of amplitude, so the output would be a constant envelope signal with two frequencies associated with the two different levels of signal. If the baseband input signal has multiple levels, representing PAM symbols, the output would be still a constant envelope signal with as many frequencies as the number of levels in the PAM signal.

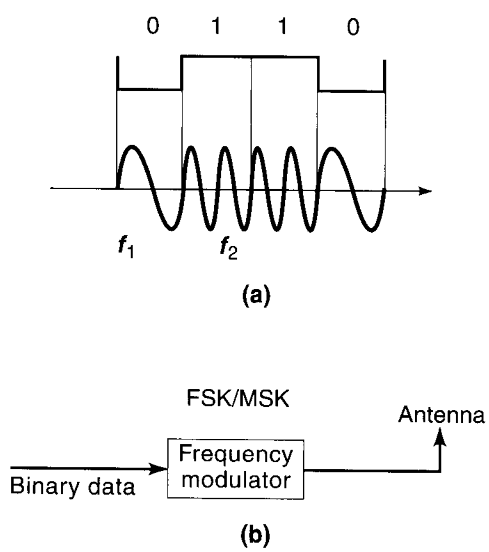


Figure 3.7 (a) Basic concept of FSK; (b) Implementation of FSK using an FM transceiver.

Example 3.5: Four-Level FSK

A 4-FSK-modulation technique was adopted for Altair, the first WLAN product that operated in the 18–19 GHz bands. This pioneering WLAN product operated at a bit rate of 10 Mbps and employed very advanced signal processing and fabrication techniques.

An important parameter in the design of FSK modems is the frequency spacing between the tones. This distance is representative of the occupied bandwidth of an FSK signal, and to maintain optimal detection at the receiver, it should take specific values that ensure orthogonality of the transmitted symbols. For noncoherent detection (when the receiver is not locked to the phase of the transmitted carrier), FSK modems use a minimum distance between the tones of $1/T$ where T is the duration of the transmitted data symbols. For coherent demodulation, the distance between the tones can be reduced to $1/2T$, that is, the minimum acceptable distance between the tones that ensures the orthogonality of the transmitted symbols. The FSK modulation with minimal tone distance of $1/2T$ is referred to as *minimum shift keying* (MSK), which is a very popular transmission technique in radio communications.

To make an MSK signal even more attractive for radio communications, as shown in Figure 3.8, the transmitted baseband signal is filtered before FM modulation to further reduce the side lobes. The most popular filters used for this implementation are Gaussian filters and associated modulation technique is referred to as Gaussian MSK or GMSK, which is one of the most popular modulation techniques in 2G wireless networks. The transmitted signal at the output of the FM modulator is still a constant envelope FM signal that avoids nonlinearities that can be introduced by the power amplifiers. In the time domain, the Gaussian filter smoothes sharp transitions of the voltage levels. As a result, rather than immediate changes of the tone frequency at the output of the FM modulator, we will have a smooth transition from one tone frequency to another that reduces the side lobes of the transmitted FM modulated signal. Because phase represents variations (the derivative) of the carrier frequency in time, modulation with gradual transition of the frequency is referred to as *continuous phase modulation* techniques. Desirable modulation techniques for radio channels are *constant-envelope continuous phase modulation* techniques that are only slightly affected by nonlinearity and have low side lobes. GMSK is an ideal example of a constant envelope continuous phase modulation technique.

An important factor that affects performance in GMSK is the time-bandwidth product B_bT. Here, B_b is the 3-dB bandwidth of the Gaussian filter and T is the symbol duration. Figure 3.9 shows the spectrum of the GMSK signal for va-

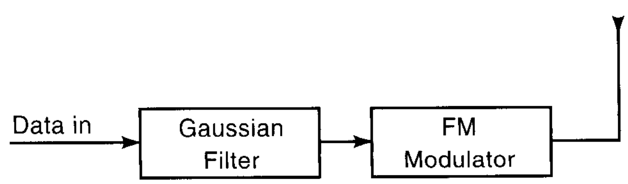


Figure 3.8 Block diagram of a GMSK modulator.

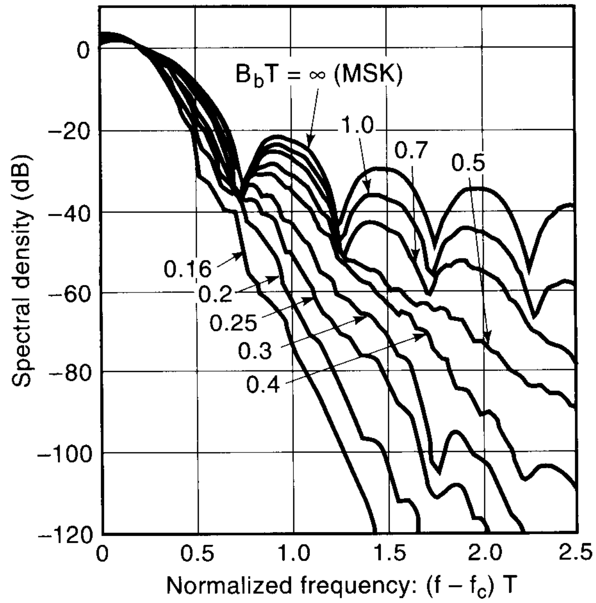


Figure 3.9 Spectra of GMSK signals for different B_bT products.

riety of B_bT, normalized 3-dB bandwidths, of the Gaussian filter. For $B_bT = \infty$, the bandwidth of the Gaussian filter is infinity (no filtering), and the system is indeed an MSK system. As the bandwidth of the filter becomes narrower, the power in the side lobes of the transmitted signal, and consequently adjacent channel interference, reduces. On the other hand, reduction of the bandwidth of the filter further smoothes the transition between levels in the time domain which increases the probability of erroneous detection at the receiver. The designer of the modem should decide on a compromise between the adjacent channel interference and detection error rate. The GSM voice-oriented standard recommends a B_bT value of 0.3, and the CDPD services recommend a B_bT value of 0.5.

Depending on the parameters of the filtering and the type of FSK-based systems, a number of modulation techniques with a variety of normalized bandwidth occupation are adopted by a variety of standards. The following example compares the bandwidth efficiency of these systems.

Example 3.6: Bandwidth Efficiency in Various Technologies

The ARDIS mobile data services use a 4-FSK modulation to support a data rate of 19.2 kbps in 25 KHz channels. The normalized bandwidth occupation or bandwidth efficiency of this system is 19.2/25 = 0.77 bits/sec/Hz. Mobitex mobile data services support 8 kbps in 12.5 kHz of bandwidth using GMSK modulation with a bandwidth efficiency of 0.64 bits/sec/Hz. CDPD uses GMSK to provide 19.2 kbps over a 30 kHz channel with a bandwidth efficiency of 0.64 bits/sec/Hz. The DECT system uses GFSK (Gaussian frequency shift keying) modulation to support 1.152 Mbps over a 1.728 MHz channel with a bandwidth efficiency of 0.67 bits/sec/Hz.

CT-2 cordless telephony uses GFSK to support 72 kbps over a 100 kHz channel with the bandwidth efficiency of 0.72 bits/sec/Hz. GSM uses GMSK with a data rate of 270.833 kbps in a 200 kHz band, giving an efficiency of 1.35 bits/sec/Hz.

Higher bandwidth efficiencies are achieved at the expense of more complex coherent implementations and lower uncoded bit error rate requirement. Data services often have more restrictions on the bit error rate than voice-oriented networks. Cellular phones accept lower qualities than cordless telephones that were designed for wireline quality operation. The diversity in requirements combined with the availability of bandwidth for a particular service has created this diversity in modulation parameters for different standards.

3.6.2 Digital Phase Modulation and $\pi/4$-QPSK

In digital phase modulation or PSK, the baseband information signal is encoded in the phase of the transmitted signal. Figure 3.10 illustrates the basic operation of a BPSK system. In Figure 3.10(a) two symbols used for binary communication are two sinusoids with 180-degree phase difference. The phase shift changes according to the voltage level of the baseband information signal.

It is customary to represent the magnitude and phase of the transmitted symbols in a complex coordinate system referred to as *signal constellation.* The signal constellation shows the unique characteristics of the modem necessary for calculation of the error rate in additive noise channels. The error rate is a function of the distance between the two points in the constellation and the level of the noise disturbing the channel.

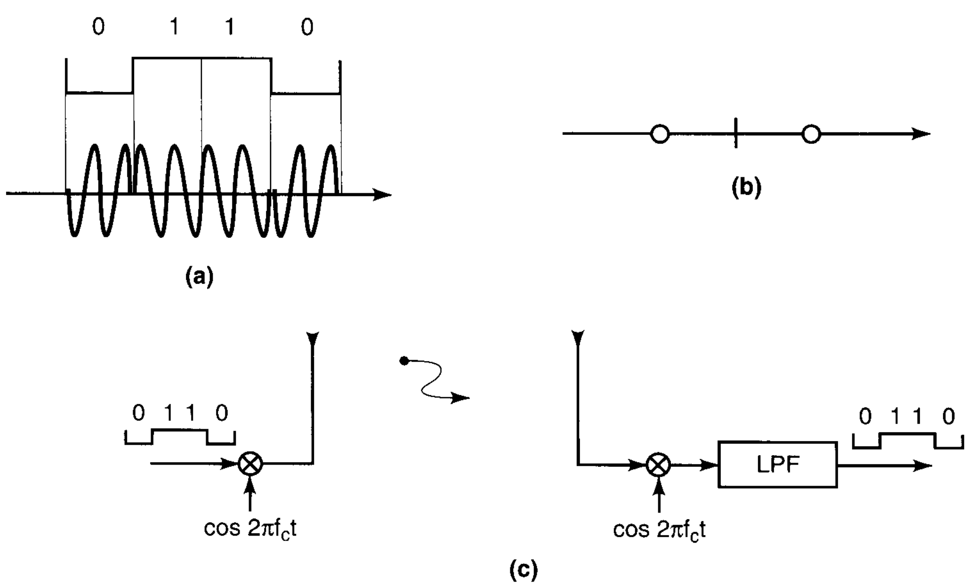


Figure 3.10 (a) Binary phase shift keying signal, (b) BPSK constellation, and (c) a block diagram of a BPSK system.

Figure 3.10(b) shows the signal constellation of the BPSK represented by two equal amplitude symbols with opposite polarity. Figure 3.10(c) represents a simple block diagram for implementation of a basic BPSK modem. The received baseband signal is simply multiplied (mixed) with a carrier frequency, and after filtering it is sent to the antenna. At the receiver the signal is multiplied with the carrier signal that is phase locked to the transmitter carrier (coherent demodulation) and then passed through a low pass filter to eliminate the unnecessary higher frequency components of the signal and discriminate the transmitted waveform. The resulting transmitted signal is a constant envelop signal shown in the lower part of the Figure 3.10(a).

Noncoherent detection of the PSK signal is also possible. The basic idea behind noncoherent detection of the PSK signal is to use the carrier in the current bit as the reference carrier of the following bit. To make this happen, the transmitted bits are differentially encoded. In differential encoding, rather than sending the actual bits the value of the exclusive-OR of the consequent bits are transmitted. Therefore the two phases that are transmitted represent the changes in the polarity of the current bit with respect to the previous bit. If this arrangement is made at the transmitter, the received signal can be detected noncoherently with the circuit shown in Figure 3.11. The value of the delay is equal to the duration of a transmitted bit. This way the carrier of the previous bit is used as the reference for detection of the current bit. The advantage of differential PSK (DPSK) over coherently detected BPSK is that the receiver circuitry does not need to recover the phase of the transmitted carrier. The disadvantage, as shown in Figure 3A.3, is that there is about 1 to 2 dB degradation in performance compared with coherent BPSK.

In a manner similar to multisymbol FSK modulation, one can design multiphase PSK modulation schemes. Four-phase PSK, often referred to as quadrature PSK (QPSK), is commonly used in radio modems. The two-dimensional signal constellation for QPSK is shown in Figure 3.12(a). Four transmitted symbols assume four different phase values of 0°, 90°, 180°, and 270°, each representing a block of two information bits. Figure 3.12(b) illustrates the basic principle of the 2-D modulation techniques used for transmission of the complex symbols. The basic structure is the same as the BPSK system except that here we have two branches of BPSK modems—one modulated over a sine wave and the other modulated over a cosine wave. Because the carriers in the two branches are orthogonal to one another at the receiver, the transmitted sine does not go through the cosine branch and vice versa. In other words, this structure has two independent modems operating over

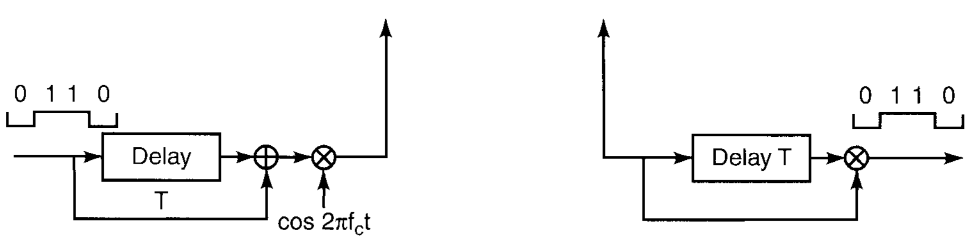
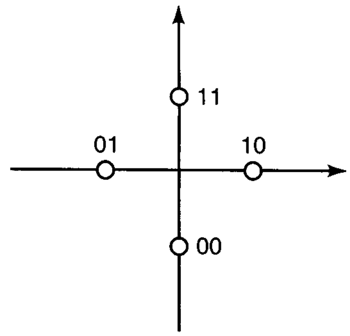
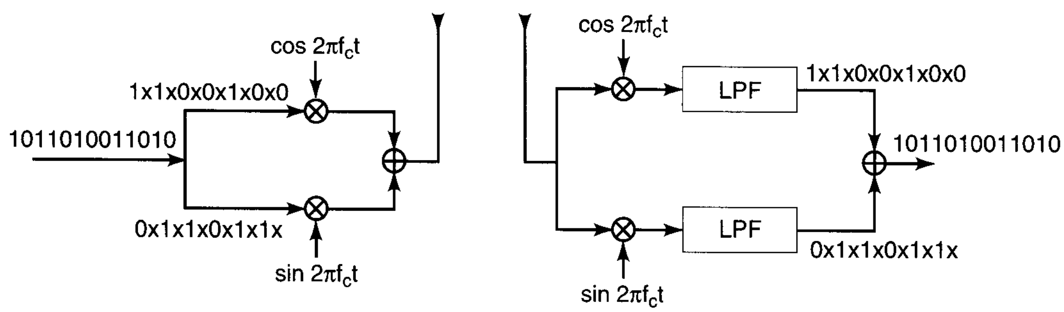


Figure 3.11 Differentially encoded BPSK.



(a)



(b)

Figure 3.12 (a) Signal constellation for QPSK and (b) modulation scheme for transmission of QPSK.

the same bandwidth and separated using their orthogonal[1] carriers. With two independent channels, we can always form a complex transmission system that can implement phase modulation. The advantage of this implementation is that we have two channels sharing the same bandwidth, resulting in a system with twice the bandwidth efficiency. The amplitude of the real part of a symbol is coded in the amplitude of the transmitted pulse over the cosine branch, usually referred to as in-phase or I-channel, and the amplitude of the imaginary part of the same symbol is coded in the amplitude of the transmitted pulse over the sine branch, often referred to as the quadrature-phase or Q-channel.

Figure 3.13 shows the normalized frequency spectrum (the actual spectrum with the abscissa divided by the bit rate of the channel and shifted from the carrier frequency to zero) of the transmitted signal for BPSK, MSK, and QPSK. Two important parameters in evaluating the spectrum of a radio modem are the width of the main lobe (i.e., a measure of the occupied bandwidth) and the peak of the sidelobes, reflecting the level of adjacent channel interference. The bandwidth of

[1]Orthogonality is an important concept in digital communications. Essentially this means that two orthogonal signals cause zero interference *after* processing at the receiver. This processing at the receiver usually implies integration over a symbol period.

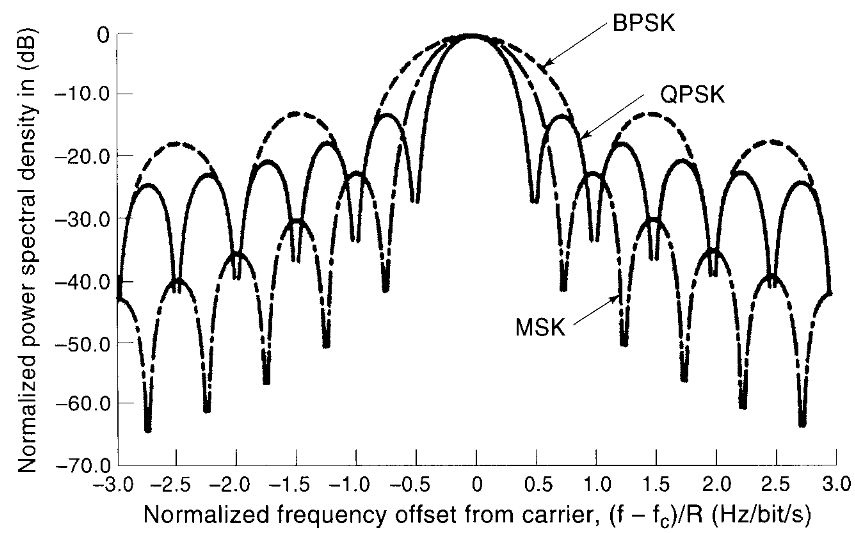


Figure 3.13 Spectra of BPSK, MSK, and QPSK.

QPSK is half of that of BPSK, and MSK lies somewhere in between. The bandwidths here are normalized to the data rate. The sidelobes of MSK, however, are more than 10 dB lower than those of BPSK and QPSK. The discussion provided in this section explains why QPSK and MSK are favored as the basis for radio modems. However, for reliable multiuser, multichannel wireless communications, we need to further reduce the height of the sidelobes. In the previous section, we introduced simple Gaussian filtering and continuous phase GMSK modulation as an improved version of the constant envelope MSK modulation. Filtering techniques can also be applied to QPSK systems to further control the sidelobes. As a matter of fact, if the baseband data stream is kept in its rectangular form, the sidelobes of the QPSK are very high (around 13 dB below the peak), which cannot attract attention for any serious radio application. All traditional QPSK and in general 2-D modulation systems use pulse-shaping filters (PSF) to control the sidelobes.

Figure 3.14 shows the basic implementation of a PSF in BPSK modem. The filter can be placed as a low-pass filter before mixing the signal with the carrier or as a band-pass filter after mixer; in either case the sidelobes of the transmission bandwidth can be controlled by this filter. Also in practice, a pair of identical filters at the receiver and transmitter is used for pulse shaping. Such a pair is referred to as a pair of matched filters. As shown in Figure 3.14, the baseband rectangular pulses are changed to pulses with smoother transitions to control the sidelobes. At the receiver side, the pulse shapes are sampled at their peak values, and then the rectangular transmitted pulses are reconstructed based on the value of the sample. Ideally the best PSF filter is an ideal low-pass filter that passes all the frequencies in the band with equal gain and eliminates all other frequencies and consequently has no sidelobes and adjacent channel interference. However, the time domain pulse associated with an ideal filter is a sinc pulse that has strong sidelobes in time domain

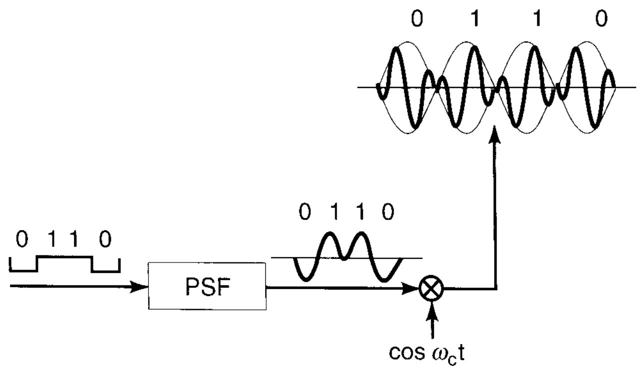


Figure 3.14 Basic implementation of PSFs for BPSK modems.

and is extremely difficult to implement. Sidelobes in time domain cause ISI, which acts as a source of noise to increase the bit error rate of the received signal. *Raised cosine* pulses provide a practical compromise between ISI and the occupied bandwidth.

Figure 3.15 shows the time domain pulse and the spectrum of the raised cosine pulses. The spectrum consists of two half-cosines at the two sides and a flat portion in the middle. The parameter β, referred to as the roll-off factor, takes values between zero, for ideal case, and one as the maximum where the flat part disappears. It controls the expansion of the spectrum. As the spectrum stretches further from $1/2T$ toward $1/T$ the sidelobes of the time domain pulse reduces, resulting in smaller values of ISI. In practical implementations, values of β between 0.2–0.5 are used in a variety of wired and wireless modems that bring the sidelobes 40–60 dB below the main lobe. Pulses shown in Figure 3.15 can be sent every $2T$ seconds where the main lobes of the received pulses do not overlap. For this case, only one sample per pulse is adequate for digital processing at the receiver side. If the receiver can manage several (typically eight) samples per main lobe, as shown in Figure 3.16, the raised cosine pulse can be sent in an overlapping manner every T seconds, reducing the transmission bandwidth by half. Overlapping and nonoverlapping pulses in the time domain in QPSK can be thought to be similar to the spectral overlap of FSK versus MSK. There is a twofold increase in bandwidth efficiency involved that is offset, however, by the complexity of the receiver design. Depending on the roll-off factor and overlapping or nonoverlapping pulses, a variety of QPSK modulation formats occupy different normalized bandwidths. An example will illustrate the situation.

Example 3.7: Bandwidth Efficiency in QPSK-Based Systems

The IS-95 standard uses QPSK modulation for its spread spectrum modulated baseband signal. The transmission bandwidth per carrier is 1.25 MHz, and the chip rate supported by the system is 1.2288 Mcps. The normalized bandwidth occupancy or efficiency of this QPSK system is 0.98 chips/sec/Hz. The IEEE 802.11 system occupies a bandwidth of 26 MHz for a spread spectrum QPSK system with a chip rate of 22 Mcps that yields a bandwidth efficiency of 0.85 chips/sec/Hz. In

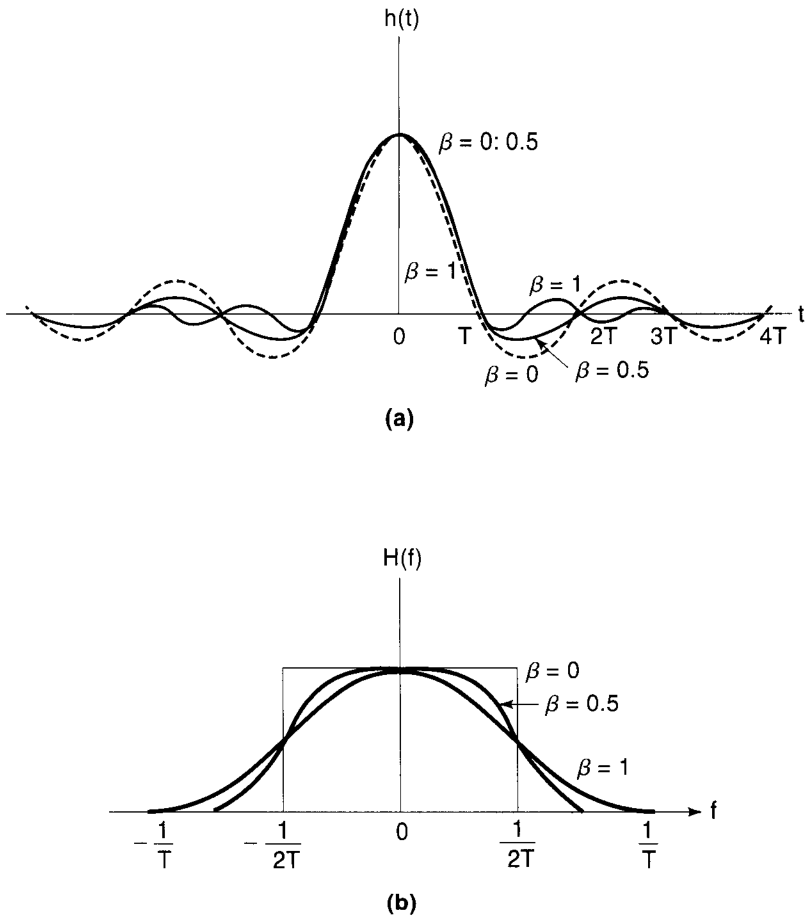


Figure 3.15 Time domain raised cosine pulse and its spectrum.

both of these spread spectrum systems, nonoverlapping pulses are used to implement the system. The overlapping pulse $\pi/4$-QPSK system in IS-136 TDMA standard uses a nominal bandwidth of 30 kHz to support a data rate of 48.6 kbps per channel that has the bandwidth efficiency of 1.62 bits/sec/Hz. The Japanese 2G digital cellular system uses overlapping pulses based on $\pi/4$-DQPSK to support a 42 kbps data rate over a 25 kHz channel resulting in a 1.68 bits/sec/Hz bandwidth efficiency. The PACS system uses $\pi/4$-QDPSK with a bandwidth efficiency of 500 kbps/350 kHz = 1.42 bits/sec/Hz. If we compare this example with Example 6, we can easily see that PSK-based modems are capable of supporting slightly higher bandwidth efficiencies compared with the equivalent FSK-based systems.

So far we have examined the similarities of the effects of filtering to reduce the sidelobes for both QPSK and GMSK modems. However, there is a major difference between the filtered QPSK and GMSK. In QPSK, when we use filters the transmitted signal is no longer a constant envelope signal. Also, QPSK has discontinuities of up to 180 degrees in the phase of the carrier. Therefore, filtered QPSK signals have neither a constant envelope nor a continuous phase. This gives a

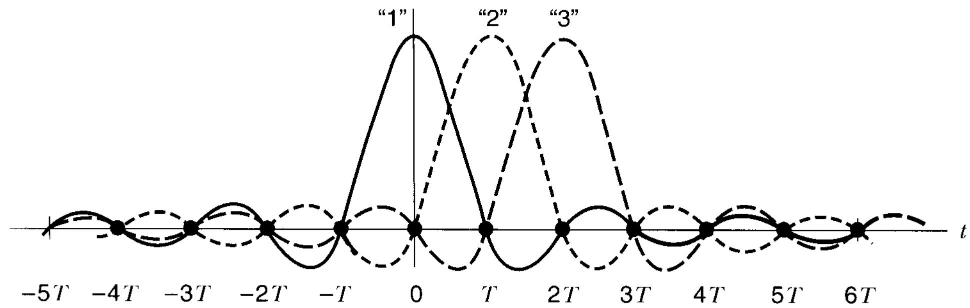


Figure 3.16 Transmission of raised cosine pulses.

practical edge for GMSK, compensating for its lower bandwidth efficiency. In the past few decades, a number of modifications to QPSK have been examined to improve its envelope and phase characteristics. The two most popular systems used in the wireless industry are offset or staggered QPSK (OQPSK or SQPSK) and π-/4-QPSK.

3.6.2.1 Further Improvements to QPSK

In an OQPSK modulator, instead of applying the source data bits to the two branches simultaneously every T seconds, as shown in Figure 3.12, the two branches are offset by $T/2$ seconds. The benefits of this scheme are twofold: (1) The envelopes of the in-phase and quadrature-phase signals overlap one another, resulting in fewer fluctuations of the amplitude of the transmitted signal (the peaks of the envelope in one branch occur in between the peaks of the other branch); and (2) the changes in the phase of the carrier in the two branches occurs every $T/2$ seconds rather than every T seconds. This way, every $T/2$ seconds, the phase is shifted by $\pm$ 90 degrees, and it avoids the $\pm$ 180-degree phase shift that was possible in standard QPSK. Therefore, OQPSK provides a better constancy of envelope and a better continuity in phase than QPSK which improves the power requirements and further reduces the sidelobes. A drawback of OQPSK modulation is that it is difficult to develop noncoherent receivers for this modulation technique, and it is more sensitive to multipath fading channels with large Doppler shifts [PAH85a], [FEH91]. The search for nonstaggered modulation schemes having low postfiltering amplitude variations led to work by Akaiwa and Nagata [AKA87] and others, including [LIU89], [GOO90], [LIU90], on $\pi/4$-QPSK modulation.

Simply described, $\pi/4$-QPSK is a form of QPSK modulation in which the QPSK signal constellation is shifted by 45 degrees each symbol interval T. This means that the phase transitions from one symbol to the next are restricted to $\pm$ 45 degrees and $\pm$ 135 degrees. By eliminating the 180-degree transitions of QPSK, the amplitude variations after filtering are significantly reduced. Figure 3.17(a) shows the eight possible phase states of the signal constellation of the $\pi/4$-QPSK modulation. The eight phases represent the four-phase QPSK constellation in its two shifted positions. The four dotted lines radiating from each point on the circle indicate the allowed phase transitions. In the implementation depicted, the modulation is implemented using a sine wave pulse shaping [FEH91]. Figure 3.17(b) shows the spectra measured for two versions of $\pi/4$-QPSK, nonlinearly amplified. The upper trace is strict-sense $\pi/4$-QPSK, while the lower trace is the spectrum for a sine wave pulse-shaped $\pi/4$-QPSK.

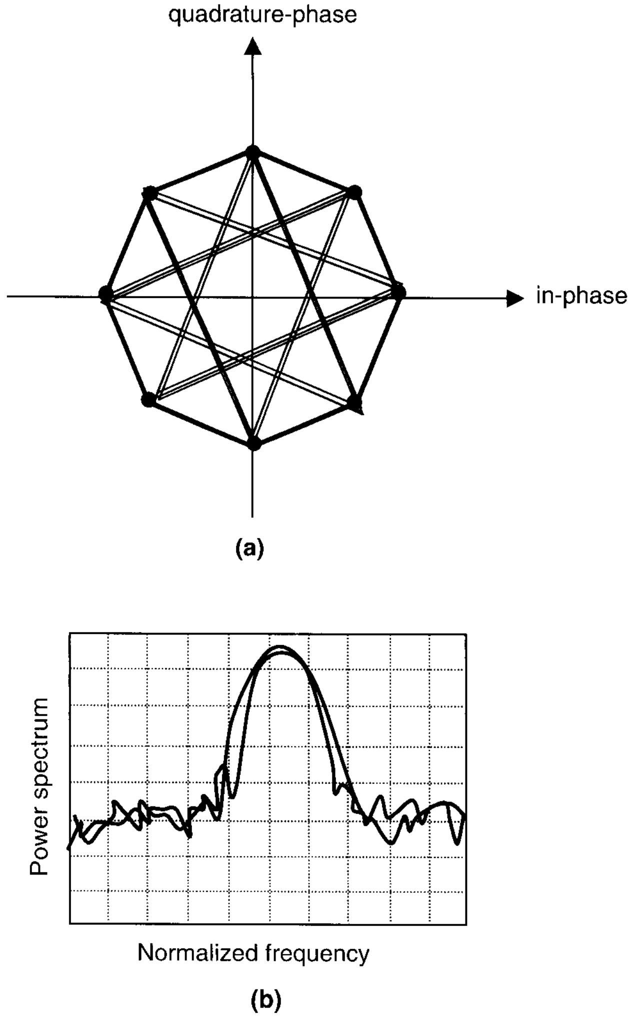


Figure 3.17 (a) Signal constellation and phase transitions of $\pi/4$-QPSK and (b) spectra for two different pulse shapes.

Thus $\pi/4$-QPSK modulation provides the bandwidth efficiency of QPSK together with a diminished range of amplitude fluctuations. Furthermore, the $\pi/4$-QPSK modulation has the advantage that it can be implemented with coherent, differentially coherent, or discriminator detection [LIU90]. These multiple advantages of $\pi/4$-QPSK led to its adoption for the North American Digital Cellular TDMA standard, IS-136, as well as the Japanese Digital Cellular standard [NAK90] and the standards for TETRA [HAI92].

Although it is not essential, $\pi/4$-QPSK modulation is frequently implemented with differential encoding, because this permits the use of differential detection in the receiver, though coherent detection may also be used to achieve optimum performance. The use of differential detection avoids the complexity required to reliably extract a coherent carrier reference under multipath fading conditions. This scheme is termed $\pi/4$-differential QPSK, denoted simply as $\pi/4$-DQPSK.

In summary, GMSK and $\pi/4$-QPSK are the most popular traditional modem technologies used in wireless networks. Overall, considering the bandwidth efficiency, adjacent channel interference, and the ease of implementation, these two techniques are very comparable. It can be shown that it is possible to implement GMSK using the 2-D branches of QPSK modems [PAH95] which reveals a great similarity between the two schemes.

3.7 BROADBAND MODEMS FOR HIGHER SPEEDS

In the previous section, we described popular modulation techniques used in digital cellular networks. These networks are designed for voice or other relatively low-speed modulation techniques for which comprehensive coverage and mobility are the dominant design concerns. In WLANs and point-to-point fixed wireless communications, coverage and mobility are restricted, but achievable data rate is the prime concern. In the past decade, a number of modulation techniques have been examined to support higher speed wireless networks for indoor wireless LANs and outdoor broadband fixed wireless networks, such as local multipoint distribution service (LMDS), used as trunks, providing the last mile access to the PSTN or the Internet. In the late 1990s OFDM emerged as the most popular modulation technique for these applications and is adopted in the IEEE 802.11a and HIPERLAN-2 next-generation WLANs, as well as LMDS and digital audio broadcast (DAB) systems.

3.7.1 Multicarrier, Multisymbol, Multirate OFDM Modulation

What is known as OFDM today combines three transmission principles: multirate, multisymbol, and multicarrier modulation (MCM). However, its name comes from the method of implementation of the MCM, using orthogonality of the adjacent carriers. An MCM system is indeed an FDM system for which a single user uses all the FDM channels together. OFDM is an implementation of MCM that takes advantage of the orthogonality of the channels and develops a computationally efficient implementation based on the Fast Fourier Transform (FFT) algorithm.

3.7.1.1 Multicarrier Modulation

MCM was first evaluated for high-speed voice band modems [HOL63], and it was augmented with FFT implementation for the same application in the early 1980s [PAH88]. It found its way into wireless modems in the early 1990s [BIN90]. The concept here is very simple. Instead of modulating a single carrier at a rate of R_s symbols/sec, we use N carriers spaced by about R_s/N Hz and modulate each of the carriers at the rate R_s/N symbols/sec. The advantage of this scheme is that on a multipath channel the multipath is in effect reduced relative to a symbol interval by the ratio of $1/N$ and thus imposes less distortion in each demodulated symbol. If the symbols are made sufficiently long relative to the multipath spread, reliable demodulation performance can be achieved without the need for any antimultipath signal processing technique. Therefore, in frequency selective multipath fading channels, MCM provides a simpler alternative to single-carrier modulation that needs complex signal processing algorithms at the receiver. A further advantage of MCM is that in frequency-selective fading, the subchannels provide a form of frequency diversity, which can be exploited by applying error-control coding *across symbols* in different subchannels. In the OFDM implementation, this latter technique is referred to as coded OFDM or COFDM. To further improve the performance of an MCM system, one may measure the received signal power in different subchannels and using a feedback channel may change the modulation and/or coding of the transmitted subcarriers to optimize performance. One may adjust the transmitted power per channel to compensate for the frequency selective fading affecting different channels differently. With these features MCM is an ideal solution for broadband communications because increasing the data rate is simply a matter of increasing the number of carriers. The limitations are complexity of implementation and the limitation on the transmitted power. To avoid overlap between consecutive symbols, a time guard is enforced between transmissions of two OFDM pulses that will reduce the effective data rate. Also some of the carriers are dedicated to the synchronization signal, and some are reserved for redundancy.

Example 3.8: MCM in Wireless LANs

Figure 3.18 shows the 64 subchannel implementation of MCM for the IEEE 802.11a and HIPERLAN-2 physical layer specifications. Each channel carries a symbol rate of 250 kilo symbols per second (ksps). We have 48 subcarriers devoted to information transmission, four subcarriers for pilot tones used for synchroniza-

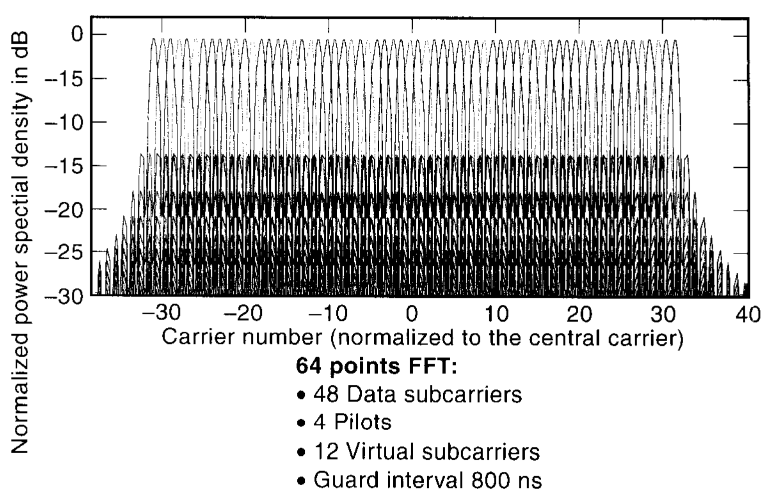


Figure 3.18 OFDM implementation of IEEE 802.11a and HIPERLAN-2 physical layer.

tion, and 12 reserved for other purposes. The occupied bandwidth is 20 MHz, providing a channel occupancy of 20 MHz/64 kHz = 312.5 kHz per subchannel. Therefore, the modulation efficiency is 250 ksps/312.5 kHz = 0.8 symbols/sec/Hz, and the user symbol transmission rate is 48 × 250 ksps = 12 Msps. The bit transmission rate depends on the number of bits per symbol. The guard time between two transmitted symbols is 800 ns compared with the symbol duration of 1/250 ksps = 4000 ns with a time utilization efficiency of 4000/4800 = 83 percent.

3.7.1.2 Multisymbol Modulation

Multisymbol modulation uses multiamplitude and multiphase modulation and coding techniques for increasing the data rate. As we saw in the previous section, traditional radio modems, such as QPSK, are four-symbol systems encoding two bits in one of four transmitted symbols. They have a signal constellation with four points representing the amplitudes and phases of the four distinct symbols. The advantage of these systems is that with the same symbol transmission rate (the same occupied bandwidth) they could double the bit transmission rate compared with BPSK. Multiamplitude, multiphase modulation techniques extend this concept by increasing the number of symbols in the constellation, allowing more encoding of the number of bits per symbols. These symbols are then modulated over a QAM modem that transmits the real and imaginary part of the encoded symbols in the in-phase and quadrature-phase channels of a modem similar to the one shown in Figure 3.12. The number of bits per symbol of a signal constellation represents the increase of the data rate over binary communication systems using a one bit per symbol scheme.

In radio channels, often symbols are encoded with a coding technique with a certain rate that represents the ratio of the actual information rate to the encoded data stream. The coding rate, number of points in the signal constellation, and the symbol rate are used to find the actual information transmission rate of a system.

Example 3.9: Data Rate in QAM Systems

Figure 3.19 shows the 16-QAM (4-bit per symbol) and 64-QAM (6-bit per symbol) signal constellations. If the symbol transmission rate for these constellations are 250 ksps, the bit rate of the 16-QAM is 4 × 250 ksps = 1 Mbps, and the bit

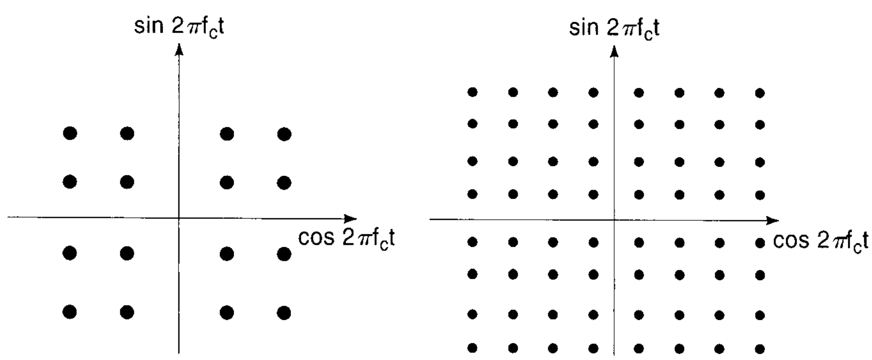


Figure 3.19 QAM constellations for four and six bits per symbol.

rate of the 64-QAM is 6 × 250 ksps = 1.5 Mbps. If the information bits were convolutional encoded with a rate of 3/4, the actual data transmission rates would be 750 kbps and 1.125 Mbps, respectively. These system specifications are used in the IEEE 802.11a/HIPERLAN-2 standards to provide data rates of 36 Mbps and 54 Mbps over the multicarrier structure described in our previous example.

3.7.1.3 Multirate Transmission

Yet another approach to increasing data rate is to use a multirate modem. A multirate modem provides one or more "fallback" modes of operation for increased reliability of communication under degraded channel conditions. The idea here is as follows: If the modulation efficiency is increased (the number of bits per symbol is increased), the required signal-to-noise ratio at the receiver also increases. In voice band modems if the modem is connected to a line with poor characteristics, the data rate is reduced. In the radio modems, as the user moves away from a base or an access point, the received signal-to-noise ratio reduces, and the modem falls to a lower rate, providing reasonable error rates at lower values of the signal to noise ratio. Readers interested in quantitative performance improvements due to multirate transmission in radio channels can refer to [WIN85], [ZHA90], and [PAH95]. Most wireless LAN products and standards have adopted multirate transmission.

Example 3.10: Multirate Transmission in Wireless LANS

The IEEE 802.11/HIPERLAN-2 standard using OFDM modulation that was described in Examples 3.8 and 3.9 uses a number of fallback options. As the distance between the transmitter and the receiver is increased, the data rate is reduced by adjusting the coding rate and the symbol transmission rate (size of the constellation). The fallback data rates are 54 Mbps to 36, 27, 18, 12, 9, and 6 Mbps to cover distances up to 100 meters.

3.8 SPREAD SPECTRUM TRANSMISSIONS

The main difference between the spread spectrum transmission and traditional radio modem technologies is that the transmitted signal in spread spectrum systems occupies a much larger bandwidth than the traditional radio modems where the transmitted signal has a bandwidth of the same order as the information signal at baseband [PIC91]. Compared with UWB, however, the occupied bandwidth by spread spectrum is still restricted enough so that the spread spectrum radio can share the medium with other spread spectrum and traditional radios in an FDM format. There are two basic methods for spread spectrum transmission: direct sequence spread spectrum (DSSS) and frequency hopping spread spectrum (FHSS). Spread spectrum technology was first invented during the Second World War, and it has dominated military communication applications, where it is attractive because of its resistance to interference and interception, as well as its amenability to high-resolution ranging. In the 1980s, commercial applications of spread spectrum technology were investigated, and today it is the transmission choice for emerging 3G cellular as well as IEEE 802.11 wireless LAN standards. The voice-oriented digital cellular and PCS industries have selected spread spectrum technology to support CDMA networks as an alternative to TDMA/FDMA networks. This increases system capacity, provides a more reliable service, and supports soft handoff of cellular connections that are discussed in more detail in subsequent chapters. In the WLAN industry, spread spectrum technology was adopted primarily because the first unlicensed frequency bands suitable for high-speed radio communication were ISM bands, which were released by the FCC under the condition that the devices operating in these bands use spread spectrum. The FCC ruling on the ISM bands was to protect users from interfering with one another [MAR85].

The principle advantages of spread spectrum transmission are as follows:

1. Spread-spectrum signals can be overlaid onto bands where other systems are already operating, with minimal performance impact to both systems.
2. Spread spectrum is a wideband signal that has a superior performance over traditional radios on frequency selective fading multipath channel. Spread spectrum provides a robust and reliable transmission in urban and indoor environments where wireless transmission suffers from heavy multipath conditions.
3. The anti-interference characteristics of spread spectrum are important in some applications, such as networks operating on manufacturing floors, where the signal interference environment can be harsh.
4. Cellular systems designed with CDMA spread spectrum technology offer greater operational flexibility and overall system capacity than systems built on FDMA or TDMA access methods.
5. The convenience of unlicensed spread-spectrum operation in ISM bands in the United States is attractive to manufacturers and users alike.

Many of these features are also shared with UWB transmission discussed earlier in this chapter. UWB can be compared with DSSS transmission; however, from an application point of view, the major difference is that the UWB is for very short-range communications whereas spread spectrum technology can cover wide areas.

3.8.1 Frequency Hopping Spread Spectrum

The FHSS technique, invented by the Austrian-born movie star Hedy Lamarr to protect guided torpedoes from jamming, is a relatively simple concept. In order to avoid a jammer, the transmitter shifts the center frequency of the transmitted signal. The shifts in frequency, or *frequency hops,* occur according to a random pattern that is known only to the transmitter and the receiver. If we move the center frequency randomly among 100 different frequencies, then the required transmission bandwidth is 100 times more than the original transmission bandwidth. We call this new technique a *spread spectrum technique* because the spectrum is spread

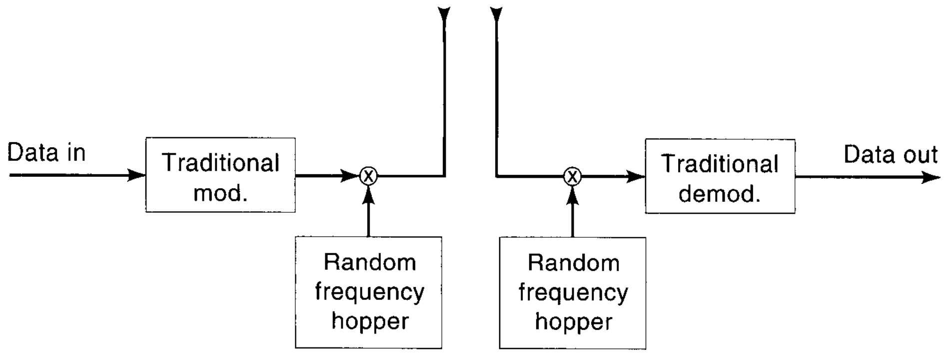


Figure 3.20 The frequency-hopping concept.

over a band that is 100 times larger than original traditional radio. FHSS can be applied to both analog and digital communications, but it has been applied primarily for digital transmissions. Figure 3.20 shows a simple diagram to describe the basic concept of FHSS transmission.

In the first stage, the input data is modulated with a traditional modulator, and in the second stage, the center frequency is changed according to a random hopping pattern generated by a random number generator. Ideally, the random pattern or spreading code is designed so that the occurrence of frequencies is statistically independent of one another. Because the sequenced pattern is coded to only *appear* random, the sequences are referred to as pseudorandom sequences or codes. At the receiver, first a dehopper, synchronized to the transmitter, repeats the hopping pattern of the transmitted signal, and then a traditional demodulator detects the received data. In digital implementation of this system, the sampling rate is the same as the sampling rate of the traditional system, leaving the complexity of the implementation in the same range as traditional modems. As we will see later, DSSS needs much higher sampling rates and consequently a more complex hardware implementation.

Example 3.11: Hopping Sequence of an FHSS System

Figure 3.21 shows the hopping pattern and associated frequencies for a frequency hopping system transfering data packets over the air. Each packet is transmitted using a different frequency. The sequence of frequencies is $f_3, f_5, f_6, f_1, f_4, f_8, f_2, f_7$ before returning to the first frequency, f_3. This is a slow frequency hopping system where a long packet with a number of data bits are transferred at each hop using the same frequency. In a fast frequency hopping system, the frequency hops occur much more rapidly and in each hop a very short packet is transmitted. In fast frequency hopping systems used in military applications, sometimes the same bit is transferred using several frequencies.

In the FHSS, the hopping of the carrier frequency does not affect the performance in additive noise because the noise level in each hop remains the same as the noise level of the traditional modems. Therefore, the performance of the FHSS

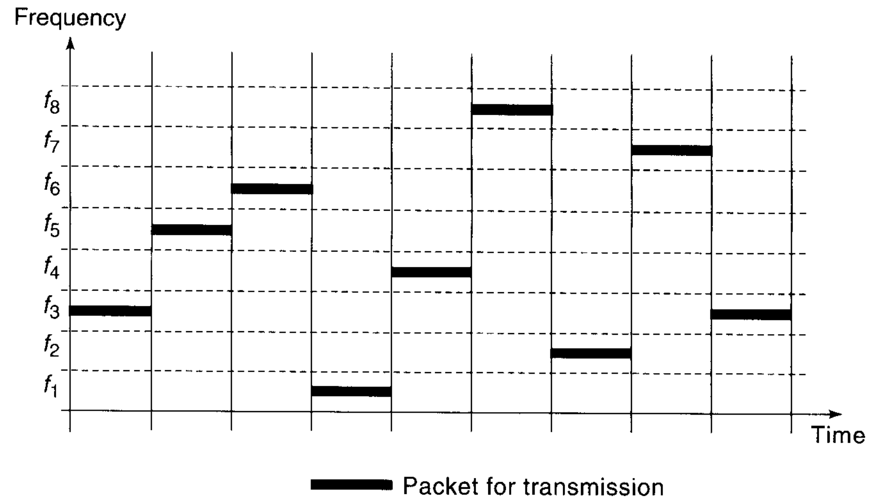


Figure 3.21 Example of an FHSS system.

systems in noninterfering environments remains exactly the same as the performance of the traditional systems without frequency hopping. In the presence of a narrowband interference, the signal-to-interference ratio of a traditional modem operating at the frequency of the interferer becomes very low, corrupting the integrity of the received digital information. The same situation happens in frequency selective fading channels when the center frequency of a traditional system coincides with a deep frequency selective fade. In an FHSS system, because the carrier frequency is constantly changing, the interference or frequency selective fading corrupts only a fraction of the transmitted information, and transmission in the rest of the center frequencies remains unaffected. This feature of the FHSS is exploited in the design of wireless networks to provide a reliable transmission in the presence of interfering signals or when a system works over a frequency selective fading channel.

Example 3.12: FHSS and Retransmissions

In a wireless packet data network, often an ACK mechanism (either at MAC, LLC, or higher layers) is in place to ensure retransmission of corrupted or lost packets. In a traditional system, if the channel is distorted with interference, all the transmissions will be corrupted, and a retransmission mechanism of any sort would not help. In a system using FHSS, the packets transmitted during the hop that coincides with the interference frequency would be corrupted. If we design the system so that we send one or a few packets per hub, when these packets are retransmitted, the hop frequency is changed and the retransmission mechanism works successfully. In IEEE 802.11, the maximum time of each hop is specified as 400 ms, and the maximum length of the packet is around 30 ms. Therefore, a packet corrupted by narrowband interference can be retransmitted during the next hop, around 400 ms later, which is a reasonable delay compared to the maximum duration of the packet.

Example 3.13: Frequency Selective Fading Channels and FHSS

The same feature of the FHSS that was described in the previous example also helps successful transmission in frequency selective fading channels. As we discussed in Chapter 2, when a frequency selective fading occurs, traditional systems operating over center frequencies coinciding with the faded frequencies cannot operate properly. An FHSS system can be designed so that the deep fades in the environment only corrupt a few hops, leaving the rest of the hops for successful retransmission. In an indoor environment, the width of the fade is around several MHz, and the FHSS system used in IEEE 802.11 uses hops that are 1 MHz apart from one another. Therefore, if a hop occurs in a deep fade and the data transmitted in that hop is not reliable, the retransmitted data after around 400 ns (as in the last example) will be successful.

Example 3.14: FHSS and GSM

In voice-oriented networks, often there is no retransmission mechanism. Corrupted packets are either discarded or are retained with a wrong value, in both cases causing distortions in the voice signal at the receiver. In TDMA systems, if the channel coincides with a deep frequency selective fading or when the cochannel interference (CCI) from another cell using the same frequency is excessive, the distortion in the received voice signal will persist until the terminal moves adequately and the frequency selective fading pattern is changed or the CCI is reduced. One method to reduce the duration of the frequency selective fade or excessive CCI situations is to provide for a slow frequency hopping pattern that forces a restriction on the duration of the frequency selective fading or CCI effects. This option is exercised in the GSM system that supports an optional frequency-hopping pattern of 217.6 hops per second.

Frequency hopping spread spectrum allows the coexistence of several transmissions in the same frequency band using CDMA. To implement CDMA we simply assign a different random hopping pattern to each terminal. Then multiuser interference occurs when two different users transmit on the same hop frequency. If the codes are random and independent from one another, the "hits" will occur with some calculable probability. If the codes are synchronized and the hopping patterns are selected so that two users never hop to the same frequency at the same time, multiple-user interference is eliminated. The number of frequency slots in this case limits the number of users.

Example 3.15: Multiple Access-Points in IEEE 802.11 Using FHSS

The IEEE 802.11 FHSS WLAN specifies 78 hopping channels each separated by 1 MHz. These frequencies are divided into three patterns of 26 hops each corresponding to channel numbers (0, 3, 6, 9, . . . , 75), (1, 4, 7, 10, . . . , 76), (2, 5, 8, 11, . . . , 77). These choices are available for three different systems to coexist without any hop collision or "hit." This mechanism allows installation of three APs in the same area in an overlapping format that results in a threefold increase in the capacity of the cell.

3.8.2 Direct Sequence Spread Spectrum

In a manner similar to FHSS, DSSS can be thought of as a two-stage modulation technique, shown in Figure 3.22. In the first stage, each transmitted information bit is spread (mapped) into N smaller pulses referred to as chips. In the second stage, the chips are transmitted over a traditional digital modulator. At the receiver, the transmitted chips are first demodulated and then passed through a correlator that despreads the signal. The despreader correlates the received signal with the duplicated transmitted spreading signal (chip sequence). The peak of the autocorrelation function is used to detect the transmitted bit. The autocorrelation[2] function of a good random code has a very high peak-to-sidelobe ratio that approximately equals N which is usually referred to as the *processing gain* [GAR00] of the receiver.

Example 3.16: DSSS in IEEE 802.11

A Barker code of length 11, used in the IEEE 802.11 as the spreading signal for the DSSS physical layer, is given by $[1, 1, 1, -1, -1, -1, 1, -1, -1, 1, -1]$. Figure 3.23 shows a data bit "1" in a binary communication system, the transmitted Barker code for the data bit, and the autocorrelation function at the receiver with its high peak and low sidelobes.

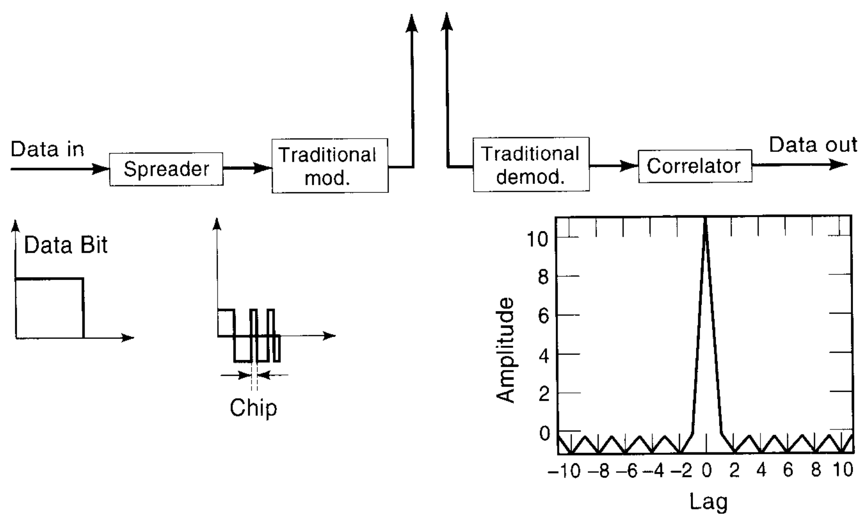


Figure 3.22 A DSSS system.

---

[2]Autocorrelation corresponds to correlating the pulse with itself, and this involves multiplying the pulse with a delayed version of itself and integrating the product over the duration of the pulse. For more details, see Appendix 3B.

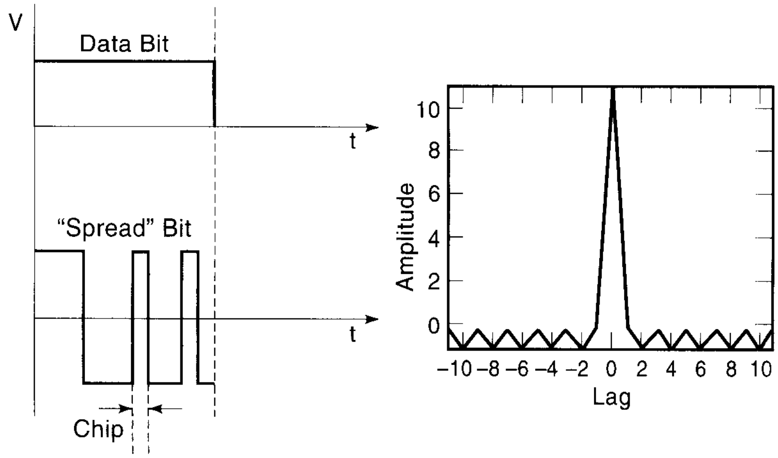


Figure 3.23 Barker code modulated DSSS signal in IEEE 802.11 and its autocorrelation.

The bandwidth of any digital system is inversely proportional to the duration of the transmitted pulse or symbol. Because the transmitted chips are N times narrower than data bits, the bandwidth of the transmitted DSSS signal is N times larger than a traditional system without spreading. As a result, N is also referred to as the *bandwidth expansion factor.* In a manner similar to FHSS, DSSS is also anti-interference and resistant to frequency selective fading. The transmission bandwidth of the DSSS is always wide, whereas the FHSS is a narrowband system hopping over a number of frequencies in a wide spectrum. As a result there is some distinction between the two methods. The DSSS systems provide a robust signal with better coverage area than FHSS. The FHSS system can be implemented with much slower sampling rates saving in the implementation costs and power consumption of the mobile units.

The DSSS can also be employed for code division multiple access. In a multiuser DS-CDMA environment, different codes are assigned to different users. In other words, each user has its own unique "key" code which is used to spread and despread only that user's messages. The codes assigned to other users are selected so that during the despreading process at the receiver, they produce very small signal levels (like noise) that are on the order of the sidelobes of the autocorrelation function. Consequently, they don't interfere with the detection of the peak of the autocorrelation function of the target receiver. In this manner each user is a source of noise for the detection of other users' signals. As the number of users increases, the multiuser interference increases for all of the users. This phenomenon continues up to a point that mutual interference among all terminals stops the proper operation for all of them. We discuss this topic in more detail in Chapter 4 when we compare the capacity of CDMA with TDMA and FDMA systems.

3.9 HIGH-SPEED MODEMS FOR SPREAD SPECTRUM TECHNOLOGY

3.9.1 PPM-DSSS

If we neglect the inner modulation in the DSSS systems and we look at the input and output of Figure 3.23, we see a signaling system in which every transmitted bit is received as an N times narrower pulse. Therefore, from the receiver's point of view it is very similar to a pulse transmission technique that sends a narrow pulse every N slots to signal the transmitted information. However, during the transmission the total energy of the pulse is spread over the entire N chips. In a way, each chip carries a piece of the information related to the received pulse, and the receiver knows the pattern (the spreading code) to put all these pieces together and construct the associated pulse. This feature can be used to implement a pulse transmission system without actually transmitting a narrow pulse. The advantage of this approach is that it refrains from transmitting high-power, narrow, short-time pulses and instead spreads the transmitted energy over time. Pulse transmission techniques of this sort are very popular in measuring the multipath characteristics of the radio channels [PAH95]. In these systems, in order to measure the impulse response of the channel, rather than sending a direct narrow pulse, a long DSSS waveform is sent periodically. After correlation at the receiver, a number of pulses associated with the signal arriving from different paths reveals the multipath characteristics of the medium. To explain this phenomenon, sometimes it is said that spread spectrum systems isolate or resolve the transmission paths. These isolated paths can also be used to provide the so-called RAKE-based time diversity in the received signal. As we will see in the next section, a smart receiver can take advantage of the time diversity of the received spread spectrum signal to improve the performance of a DSSS system.

Because we can think of DSSS as a pulse transmission technique, we might as well consider implementing a pulse modulation technique using the DSSS. The idea of using PPM with DSSS has been used for the design of high-speed wireless LANs. Let's assume that we have no serious multipath and consider a DSSS system similar to Figure 3.23. If at the transmitter we move the position of the transmitted pulse, the peak of the received pulse is also shifted. Therefore we can code our information in the position of the pulse similar to direct PPM techniques.

3.9.2 CCK Modulation

Orthogonal codes provide another method of increasing the bandwidth efficiency of a spread spectrum system [PAH87]. With this approach, each user has a set of orthogonal sequences representing a set of symbols for transmission. The set of orthogonal sequences, or *orthogonal codes,* will typically be a *Hadamard* or *Walsh* code, which are discussed later in this chapter. A stream of information bits is segmented into groups, and each group represents a nonbinary information symbol, which is associated with a particular transmitted code sequence. If there are N bits per group, one of a set of 2^N sequences is transmitted in each symbol interval. The received signal is correlated with a set of 2^N matched codes, each matched to the code sequence of one symbol. The correlator outputs are compared, and the symbol associated with the largest output is declared as the transmitted symbol.

Example 17: Orthogonal Sequences in IEEE 802.11

Recently, while the IEEE 802.11 committee was evaluating a variety of proposals for increasing the data rate of the first standard beyond 2 Mbps, they adopted an orthogonal coding technique called complementary code keying (CCK). A simplified block diagram of the basic principles of the CCK, adopted by IEEE 802.11b for 11 Mbps operation in 2.4 GHz ISM bands, is shown in Figure 3.24. The input data stream is grouped into 256 eight-bit symbols at 11 Mbps/8 bpS = 1.375 MSps. The encoder maps each eight-bit symbol into 8 four-phase coded symbols. The 256 coded symbols are selected from the 4^8 = 65,536 available alternatives, so that they are orthogonal to one another. The elements of the blocks of 8 four-phase symbols are transmitted serially using a QPSK four-phase modulator. At the receiver, each eight received complex waveforms from the QPSK demodulator are grouped in a block and sent to the decoder to find the closest eight-bit symbol associated with the demodulated 8 four-phase signals. The chip rate and occupied bandwidth of this system is the same as that of the original IEEE 802.11 standard, but the data rate is increased to 11 Mbps. The following equation gives the mapping rule for generation of the codes:

$$c = \{e^{j(\varphi 1 + \varphi 2 + \varphi 3 + \varphi 4)}, e^{j(\varphi 1 + \varphi 3 + \varphi 4)}, e^{j(\varphi 1 + \varphi 2 + \varphi 4)}, \\ -e^{j(\varphi 1 + \varphi 4)}, e^{j(\varphi 1 + \varphi 2 + \varphi 3)}, e^{j(\varphi 1 + \varphi 3)}, -e^{j(\varphi 1 + \varphi 2)}, e^{j\varphi 1}\}$$

The eight input bits are further grouped into 4 two-bit complex four-phase symbols. The resulting four phases $\varphi 1$, $\varphi 2$, $\varphi 3$, and $\varphi 4$ are inserted in the equation to find the eight complex-coded waveform that is serially modulated by the QPSK modem.

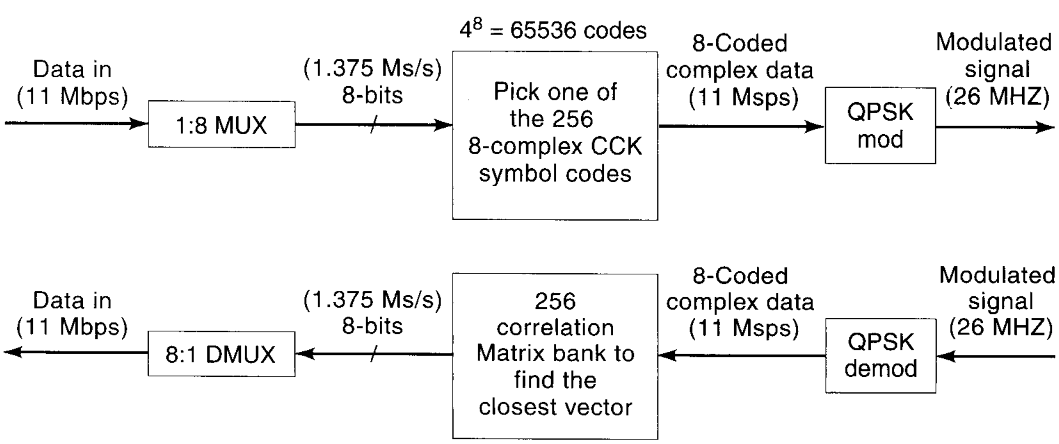


Figure 3.24 Principle of operation of CCK modulation used in IEEE 802.11b.

3.9.3 Multicarrier CDMA

Multicarrier modulation is also applied to 3G voice-oriented cellular networks using CDMA technology. The incentive of using multicarrier modulation in the CDMA systems is to provide for higher data rates for data application, wider bandwidth and consequently better quality for the voice users, and backward compatibility with the existing systems.

Example 3.18: Multicarrier in cdma2000

The cdma2000 standard for 3G cellular systems is based on the IS-95 or cdmaOne standard, which uses 1.25 Mcps for each user to support voice or data applications. The cdma2000 employs a multicarrier operation in which a user is allowed to use 1, 3, 6, or 9 of the cdmaOne channels to support a more reliable voice and variety of data channels.

3.10 DIVERSITY AND SMART RECEIVING TECHNIQUES

As we described earlier in this chapter and in Chapter 2, the main difference between the behavior of the wireless and wired media is the extensive fluctuations of the received signal that is caused by multipath and shadow fading. In particular, during short periods of time, the channel goes into deep fades, causing a significant number of errors that virtually dominate the overall average error rate of the system. In order to compensate for the effects of fading when operating with a fixed-power transmitter, the power must typically be increased by several orders of magnitude relative to the nonfading operation. This increase of power protects the system during the short intervals of time when the channel is deeply faded. The most effective method of counteracting the effects of fading is to use diversity techniques in the transmission and reception of the signal. The concept here is to provide multiple received signals whose fading patterns are different. With the use of diversity, the probability that all the received signals are in a fade at the same time reduces significantly, which in turn can yield a large reduction in the average error rate of the system.

A variety of techniques are available for reception of the diversity signals. With *selection diversity,* one signal is chosen from the set of diversity branches, usually on the basis of received signal strength. With *linear combining,* as the name suggests, the diversity branches are simply summed together before demodulation. In the optimum method of combining, called *maximal-ratio combining,* the diversity branches are weighted prior to summing them, each weight being proportional to the received branch signal strength. It can be shown that maximal-ratio combining provides the optimum performance among all diversity combining techniques [PAH95].

As the data rate of a modem increases beyond fractions (about 20%) of the inverse of the multipath spread, the channel becomes frequency selective for which a deep null may occur in the passband of the channel. In a frequency selective fading multipath transmission, the data rate is large enough, so that the bit duration is on the order of the multipath spread of the channel, resulting in performance degradation due to ISI. The performance degradation caused by ISI forces the performance curves into flat areas where any increase in the transmitted power does not improve the bit error rate performance of the modem. As a result, in a frequency selective fading channel, increase in the transmitted power level is not effective, and diversity techniques remain as the only remedy for implementation of a reliable system over these channels.

Diversity can be provided spatially by using multiple antennas, in frequency by providing signal replicas at different carrier frequencies, or in time by providing signal replicas with different arrival times. In the past several decades, a number of spatial-, frequency-, and time-diversity techniques have emerged for the implementation of advanced wireless modems that are now adopted in many standards for wireless communications.

3.10.1 Time Diversity Techniques

As we discussed earlier, in the wireless medium, due to reflection, diffraction, and scattering the received signal arrives from different paths, and because the length of these paths are not the same, signals arrive at different time delays. Therefore, in the wireless medium we have to deal with the multipath arrival problem. We also explained that for traditional receivers, the multipath arrival of the signal has the harmful effect of causing ISI. Looking into this issue conceptually, the signals arriving from different paths are exposed to different fading patterns, and if a receiver isolates these paths, it can provide a source of diversity for the arriving signal. If the paths are isolated at the receiver, the multipath signals can be regarded as a form of diversity, and a smart receiver can use this diversity to improve its performance.

In this section we describe several applied receivers that take advantage of the time diversity of the multipath arrival in wireless medium. Historically speaking the first receiver taking advantage of the in-band time diversity was the RAKE receiver, which was invented in the 1950s in MIT's Lincoln Laboratory [PRI58]. Today, to provide a robust reception, RAKE receivers are commonly used in DSSS receivers in CDMA cellular telephones. The second class of receivers exploiting time diversity operates with traditional modems. These modems either adaptively estimate the channel multipath characteristic to unfold the effect of multipath, or they use an adaptive filter with the inverse of the characteristics of the channel. These two approaches serve the same purpose of equalizing the effects of the channel, and for that reason they are referred to as *equalization* techniques.

3.10.1.1 DSSS and the RAKE Receiver

The original RAKE system was designed for wireless teletype communications with a symbol rate of 90 cps operating over an ionospheric channel. The envelope of the transmitted binary FSK signal was a 10 kHz PN-sequence. The received signal was passed through two tapped delay lines for mark and space frequencies, and the outputs were compared for decision making. The tap gains of the delay lines were adaptively adjusted by cross-correlating the received signal with both mark and space reference signals at the receiver. Because the data rate was considerably smaller than the transmission bandwidth, the ISI was negligible. More recently, other versions of RAKE receivers have been examined for urban radio [KAM81], HF radio [BEL88], and indoor radio [PAH90a] channels.

As we described in Section 3.9, from the receiver point of view, the operation of a DSSS system is similar to a pulse modulation system. After correlation in the receiver (see Fig. 3.23), the signal is a narrow pulse with a width twice the chip duration that occurs every bit interval. In a multipath environment, the output of the receiver (matched filter) will have several peaks associated with the signal arriving at different path delays. Figure 3.25 illustrates the output of the receiver with three multipath components. By comparing this figure with Figure 3.23, can one observe the effects of multipath on the receiver output in a DSSS system. Figure 3.23 represents the received signal in a single path channel, whereas Figure 3.25 represents the received signal in a three-path multipath channel. In Figure 3.25, the interarrival delay among the multipath signals is greater than the width of the base of the autocorrelation function, and the delay spread is less than the information bit interval. However, these conditions do not apply in every situation. For instance, if the delay spread of the channel is greater than bit duration, we will have ISI. To avoid interference between detected information symbols, the bit duration must be kept larger than the multipath spread of the channel. In other words, the symbol transmission rate should be kept below the coherence bandwidth of the channel. When consecutive signal paths arrive with delay differences greater than the width of the autocorrelation function (i.e., on the order of the chip duration), the receiver output will exhibit separate peaks, as shown in Figure 3.25. If the delay between two consecutive paths is significantly less than the chip duration, the two paths will merge and appear as one path equivalent to the phasor sum of the two actual paths. Thus, as the transmission bandwidth is made smaller, the chip duration becomes correspondingly longer, and fewer isolated paths can be resolved at the receiver output. Of course, as paths merge together, the fluctuations in their amplitudes and phases produce an overall fluctuation in the phasor sum, which we observe as fading.

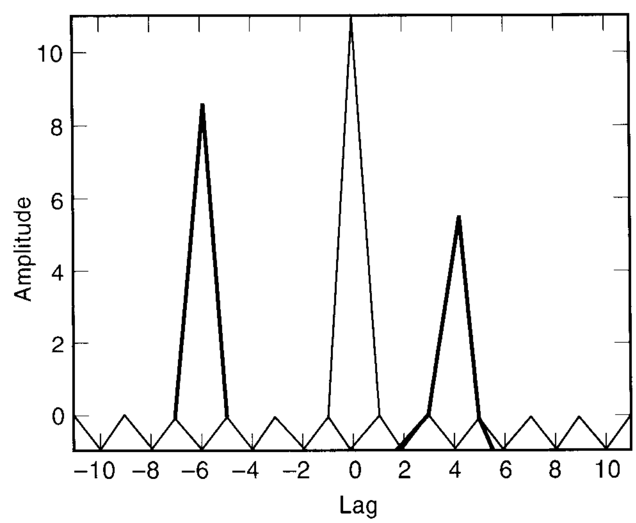


Figure 3.25 The output of a matched filter with signals arriving via multiple paths at the input.

Example 3.19: Multipath Reception in IS-95 CDMA

The IS-95 CDMA system uses a chip rate of 1.25 Mcps and a symbol transmission rate of 4,800 Sps (equivalent to 9,600 bps at two-bits per symbol in QPSK modulation). Therefore, it can resolve multipath components on the order of 1/1.25 Mcps = 800 ns apart, and a multipath spread of up to 1/4800 bps = 2.08 ms cannot cause ISI in the system. The multipath spread in the outdoor microcellular environment is on the order of several tens of microseconds, and in indoor picocell areas, it is on the order of several hundreds of nanoseconds. Therefore, IS-95 does not suffer from ISI in any of the indoor or outdoor environments. However, in an indoor picocellular environment, it is unlikely that the system resolves the multipath components. Some 3G systems offer similar bit rates with chip rates that are up to an order of magnitude shorter (10 times larger bandwidth). With a pulse resolution of around 80 ns for systems, one expects resolving several multipath components even in the indoor picocellular environment.

Example 3.20: Multipath Reception for IEEE 802.11

The IEEE 802.11-based systems have a chip rate of 11 Mcps and a symbol transmission rate of 1 Mbps (2 Mbps for QPSK modem). The resolution of the channel is on the order of 1/11 Mcps = 90 ns, and multipath spread of up to 1/1 MSps = 1000 ns does not cause ISI in the system. Wireless LANs are designed for indoor applications; therefore the IEEE 802.11 receiver can isolate several paths and it does not suffer from ISI.

If we operate with chip durations short enough to resolve individual paths, we can design a smart receiver to take advantage of the multiple paths to provide diversity and enhance the reliability of the decision on each received information symbol. In a DSSS system, a receiver that optimally combines the multipath components as part of the decision process is referred to as a RAKE receiver. A typical RAKE receiver structure for a DSSS system is shown in Figure 3.26. The received signal is passed through a tapped-delay line, and the signal at each tap is passed through a correlator similar to the one used for standard DSSS receivers. The outputs of the correlators are then brought together in a diversity combiner whose output is the estimate of the transmitted information symbol.

In the original RAKE receiver [PRI58], the delays between the consecutive taps or "fingers" of the RAKE receiver were fixed at half of the chip duration to provide two samples of the overall correlation function for each chip period. Using this method for a rectangular chip pulse with triangular correlation function, we will have four samples of each triangle in the correlation function. Because the peaks in general are not aligned precisely at multiples of the sampling rate, it is not possible to capture all the major peaks of the correlation function. But a RAKE receiver implemented with a sufficiently large number of taps will provide a good approximation of all major peaks. Modern digitally implemented receivers typically have only a few RAKE taps and the capability to adjust the tap locations. An algorithm is used to search for a few dominant peaks of the correlation function and then position the taps accordingly.

A RAKE receiver can combine the arriving signal paths using any standard diversity combiner such as a selective, equal-gain, square-law, or maximal ratio

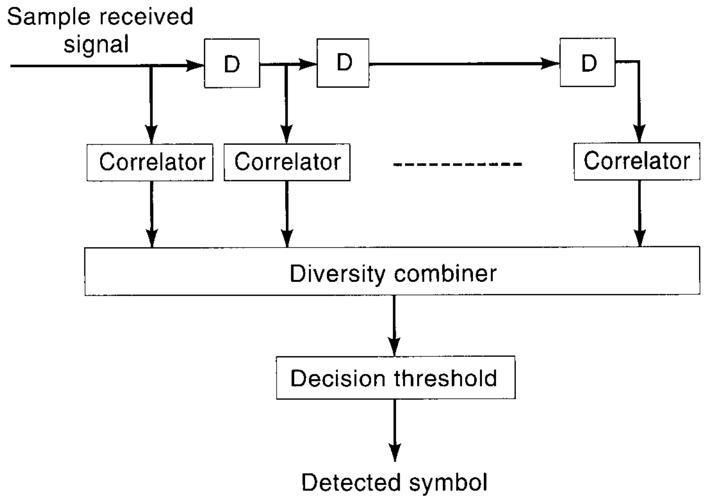


Figure 3.26 The RAKE receiver structure.

combiner [SKL01]. As we discussed earlier in this section, the optimum diversity combiner is the maximal ratio combiner, which weights the received signal from each branch by the signal-to-noise ratio at that branch. A maximal ratio combining RAKE receiver that resolves all the paths and does not introduce ISI provides optimum system performance in the presence of time diversity.

Example 3.21: Rake Reception in IS-95

The QUALCOMM original receiver uses three moving fingers to implement the RAKE receiver. The original WaveLAN that was the model for implementation of transmission system for the IEEE 802.11 standard was not using a RAKE receiver. Later on other manufacturers started implementing RAKE receivers in wireless LANs to support a more robust transmission.

In summary, in the operation of a DSSS system, multipath does not cause ISI unless the information symbol transmission rate approaches the coherence bandwidth of the channel. Also, it is possible to design a receiver which takes advantage of the isolated arriving paths to improve or even optimize system performance. The wider the transmission bandwidth, the greater is the order of implicit time diversity that can be utilized. The isolated paths will provide a source of implicit diversity to a DSSS receiver, which improves the performance of the system. On the other hand, if not utilized in some diversity combining receiver, the signal arriving from each path is a wideband interference to the signals arriving from other paths, which degrades the performance of the DSSS system.

3.10.1.2 Traditional Modems and Equalizers

In the past few decades, other adaptive techniques exploiting time diversity have emerged in the development of radio modems for fading multipath channels that operate in conjunction with traditional modems. Time-gating of the transmitted pulse to avoid ISI, with adaptive matched filtering of each received pulse, was the approach taken for a family of military troposcatter radios [PAH80], [CON78]. The adaptive decision feedback equalizer was another approach investigated for application to troposcatter [GRZ75], [MON84]; microwave LOS, [BEL84], [PAH85]; HF [FAL85]; and more indoor radio [SEX89a] channels. Finally, an adaptive version of the maximum likelihood sequence estimation (MLSE) technique [FOR72] was investigated for troposcatter in [BEL69] and for HF in [CHA75]. More recently, these techniques have found their way into wireless standards and products.

Example 3.22: Equalization in Wireless Networks

The GSM standard recommends using equalization, and to support proper operation of the equalizer, as we will see later, provides a 26-bit training sequence in each transmitted packet. The HIPERLAN-1 wireless LAN standard specification recommends using equalization in indoor radio channels to achieve data rates of up to 23 Mbps.

A standard usually does not specify the design of a receiver, but it provides a training sequence that can be used for the equalization. Therefore, manufacturers have different options for implementation of an equalizer. There are two methods that can be used to take advantage of the training sequence and implement an equalizer. One is to use the training sequence to estimate the channel multipath characteristics and then use the estimate of the channel to eliminate the effects of ISI. This technique was originally named MLSE after the name of the algorithm that uses channel estimates to eliminate the effects of ISI [FOR72]. The second technique is to use the training sequence to design an adaptive filter at the receiver that inverses the channel distortions. In the early literature in this field, this method was referred to as equalization [MON77]. More recently, both techniques are referred to as adaptive equalization. This is partially because of the fact that both techniques use the training sequence reserved for equalization, and partially due to the fact that though they are two different approaches to equalize the distortions caused by the channel, they are both adaptive in nature.

The MLSE receiver is the optimum receiver in the presence of ISI. Given the estimates of the discrete channel impulse response, an MLSE receiver uses a trellis diagram with the Viterbi algorithm to obtain maximum-likelihood estimates of the transmitted symbols. The adaptive MLSE receiver is shown in Figure 3.27. It consists of two parts: the adaptive channel estimator and the MLSE algorithm. The sampled channel impulse response is measured with the adaptive channel estimator, which operates in a way similar to the top part of the RAKE receiver described in the previous section. Using the reference sequence (a PN sequence similar to those used in DSSS systems), the top part of the figure estimates the sampled channel impulse response. In the RAKE receiver, these samples were taken at half the symbol duration, whereas in the MLSE receiver they are taken at symbol intervals. The lower part of the receiver uses the samples of channel impulse response to compare the sequence of the sampled received signal with all possible received sequences and determine the most likely transmitted sequence of symbols. The MLSE procedure [FOR72]

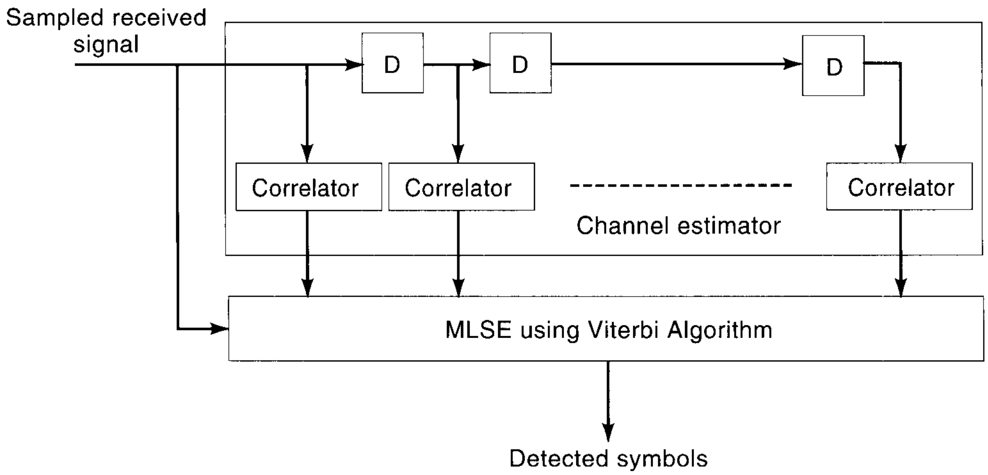


Figure 3.27 The adaptive MLSE receiver.

uses the Viterbi algorithm [VIT67] to minimize the computational complexity of the maximum likelihood selection among all possible transmitted sequences.

The MLSE is the optimal method of canceling the ISI; however, the complexity of this receiver grows exponentially with the length of the channel impulse response, whereas, as we discuss later, the complexity of the other equalizer grows only linearly with the length of the impulse response. For this reason, MLSE is an attractive option for channels with short impulse responses, but for longer impulse responses, equalizers are more practical. In the radio communication literature, MLSE is usually compared with decision feedback equalization (DFE), which is the equalizer of choice for frequency selective fading multipath channels. Comparisons of MLSE versus DFE performance are given for telephone line modems in [FAL76b,c], for HF radio in [FAL85], and for troposcatter radio links in [MON77], [CHA75].

Here we provide a short description of the principles of operation of DFE. For more detailed treatments of equalization techniques, the reader can refer to [PAH95], [PRO00], [QUR85], and [MES87]. To describe the DFE, it is easier to start with the linear transversal equalizer (LTE). Figure 3.28 shows the principal elements of an LTE. In an LTE, similar to the RAKE and channel estimator, the received signal is passed through a tapped delay line. The delayed signal is then multiplied with different tap gains, and the weighted delayed signals are added together to form a sample that is measured against some threshold to detect the value of the transmitted symbol. The weight of the tap gains is determined by variety of algorithms based on the difference between the detected symbol and the output of the adder that was the estimate of the received value of the transmitted symbol. The algorithm indeed determines the best tap weights that somehow minimize the error in the system. The LTE equalizer is a discrete-time filter intended to compensate for the amplitude and phase distortions of the channel. One can see intuitively that for an infinite-tap equalizer, the sampled frequency response of the equalizer should be the inverse of the frequency response of the channel.

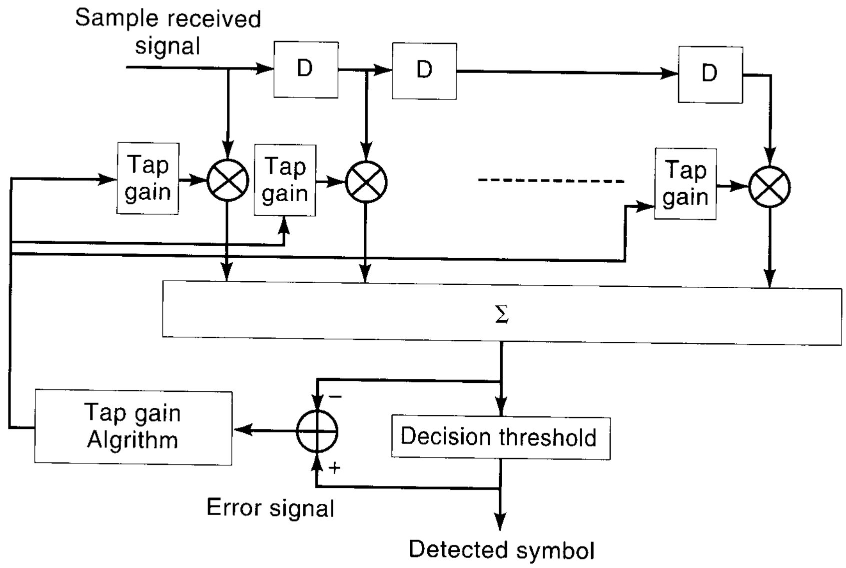


Figure 3.28 The linear transversal equalizer.

Linear equalizers are unable to properly equalize channels having deep nulls in the pass band. For these channels the linear equalizer applies a high gain in its frequency response to compensate for the null, which in turn causes noise enhancement. However, the backward filter of a DFE does not suffer from the noise enhancement problem, because it estimates the channel rather than its inverse. As a result, for channels with deep nulls in the pass band, DFEs are superior to linear equalizers.

In frequency selective fading radio channels, channels occasionally experience deep nulls in the pass band, resulting in an unsatisfactory performance for linear equalizers. On telephone channels, the most significant amplitude and delay distortion is found at the edges of the pass band. As a result a DFE, which is more effective against nulls in the middle of the pass band, has little to offer over a linear equalizer with a large number of taps. Thus linear equalization is still the predominant design choice for voice-band data modems operating over telephone channels.

DFEs [BEL79], [SAL73], shown in Figure 3.29, consist of two TDL filters, referred to as the forward and the backward equalizers. The input to the forward equalizer is the received signal, and it operates similarly to the linear equalizers discussed earlier. The input to the backward equalizer section is the stream of detected symbols. The tap gains of this section are the estimates of the channel sampled impulse response, including the forward equalizer, and this section cancels the ISI due to past samples.

3.10.2 Frequency Diversity Techniques

As we discussed in Chapter 2, a fading multipath channel displays frequency selectivity. In other words, if we sweep a large spectrum of frequencies with the same transmitted power at different frequencies, the received power fluctuates signifi-

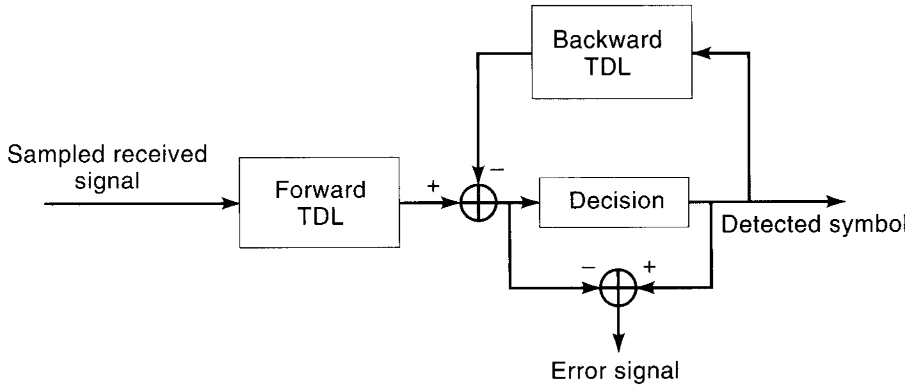


Figure 3.29 Structure of a decision feedback equalizer.

cantly, and in particular certain frequencies go to deep fades of 30–40 dB lower than the average received power at other frequencies. The width of the fades is proportional to the delay spread of the multipath arrival of the signal. In the radio channel modeling literature, this phenomenon is referred to as frequency selective fading.

Example 3.23: Frequency Selective Fading Channel

Figure 3.30 represents a sample measured impulse response and frequency response of an indoor radio channel in the 900–1,100 MHz range. We see a number of fades and, in particular, one very deep fade close to the center of the band. The multipath spread in this experiment is around several hundred ns, resulting in fade durations of around several MHz. The difference between the peak received power and the deepest fade is around 35 dB.

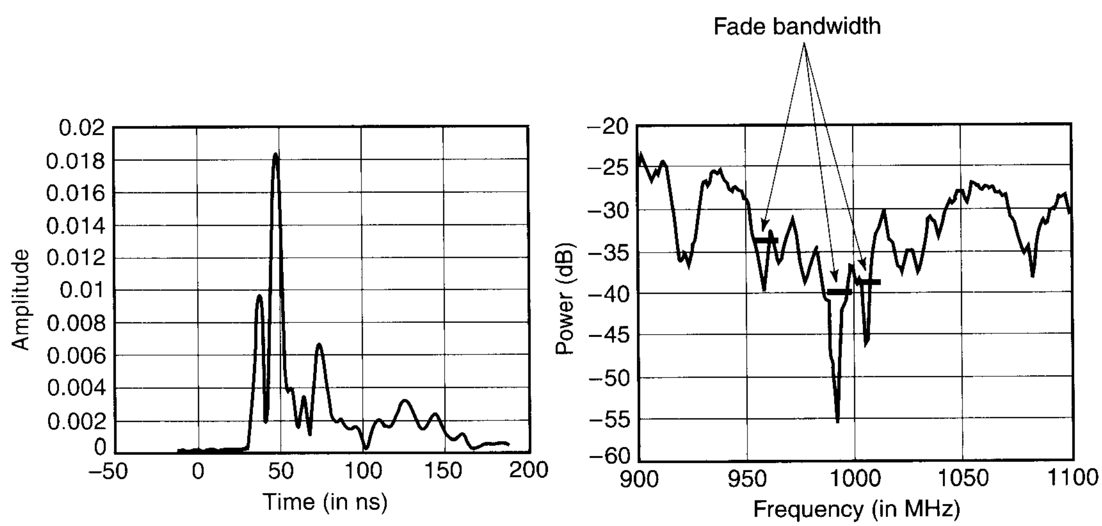


Figure 3.30 Frequency response of the channel over 200 MHz frequency band at 1 GHz.

As a terminal moves or people move around the terminal, the multipath profile and the frequency selectivity pattern both change continually. With no movement of the terminal or movement of the objects close to the terminals, the channel behavior will remain quasi-stationary. This inequality of the received signal at different frequencies provides a source of *frequency diversity* in the received signal that a smart system can use to improve its performance.

As we also saw before, if the symbol transmission rate of a signal is much smaller than coherence bandwidth, the entire signal bandwidth is similarly affected. The good part of this situation is that there is no ISI, and the bad part is that when the signal is hit by fade the received signal strength becomes very low because all the spectrum of the signal goes into a fade. In time diversity techniques, we increased the data rate so that the entire signal did not fade at the same time, and then we used an adaptive technique to take advantage of multipath arrivals. Receivers taking advantage of the frequency diversity intend to send the signal at different frequencies, so that only a portion of the transmitted signal is damaged due to the frequency selective fading. The first class of systems taking advantage of the frequency diversity are the FHSS systems, which are naturally hopping over different random frequencies.

Example 3.24: Frequency Hopping and Bluetooth

The Bluetooth system designed for personal area networking uses a symbol transmission rate of 1 Msps that does not need time diversity receivers to operate in indoor areas. However, it uses a fast frequency hopping technique (1,600 hops per second) that transmits one packet per fixed hop slot of 625 μsec hopping over 79 MHz of bandwidth. If the hop frequency hits the deep fade, the packet is lost, and it will be retransmitted in the next hop whose frequency is selected in a random manner. Bluetooth is a simple system with low power consumption and reliable transmission.

Example 3.25: Frequency Hopping and IEEE 802.11

The IEEE 802.11 FHSS system operates at 1 and 2 Mbps with a transmission bandwidth of 1 MHz using MSK modulation. This system has a variable packet duration of up to 20 ms, and the frequency hopping recommended by this standard is 2.5 hops per second. In a fade, a packet has a chance to be successfully received in a few attempts, and if not successful in the next hop, it will go through the system. The maximum delay jitter for the user would be around 400 ms, which is associated with a deep and long persisting frequency selective fading. FHSS again provides a low power simple implementation as it is compared with the DSSS alternative capable of taking advantage of time diversity.

Example 3.26: Frequency Hopping and GSM

The GSM standard has provisions for the implementation of equalization techniques using the reference (training) signal. The GSM standard also recommends slow frequency hopping transmission to introduce frequency diversity. If a mobile terminal is in a persisting fading situation, the transmitted power controlled by the BS is maximized, and the equalizer must be operating too in order to consume the life of the battery very quickly. Frequency hopping will allow the hop to equalize the performance and battery consumption among all terminals.

Another class of systems taking advantage of frequency diversity is the class of multicarrier systems. If the bandwidth of each carrier is smaller than the coherence bandwidth of the channel and the number of carriers is large enough so that the bandwidth is much larger than the fade durations, we will always have a number of carriers that are not affected by deep frequency selective fading.

Example 3.27: Frequency Diversity and MCM in IEEE 802.11a/HIPERLAN-2

Figure 3.31 shows a fading pattern and the band used by IEEE 802.11a/HIPERLAN-2 standards. As shown in the figure, we always have a number of channels that are not affected by deep null and can provide a reliable error rate.

A smart receiver using frequency diversity of this form has several choices. One is to measure the received power in all subcarriers and improve the performance of the faded subcarrier by reducing the transmission rate of that subcarrier. As a rule of thumb, the reduction of each bit per symbol will reduce the power requirement by about 3 dB.

Example 3.28: SNR Requirements versus Bit Rate

IEEE 802.11a and HIPERLAN-2 have transmission choices from 6 bits per symbol (64-QAM) to 1 bit per symbol (BPSK), which provide a range of 15 dB adjustments for frequency selective fading effects by changing from highest to lowest data rate.

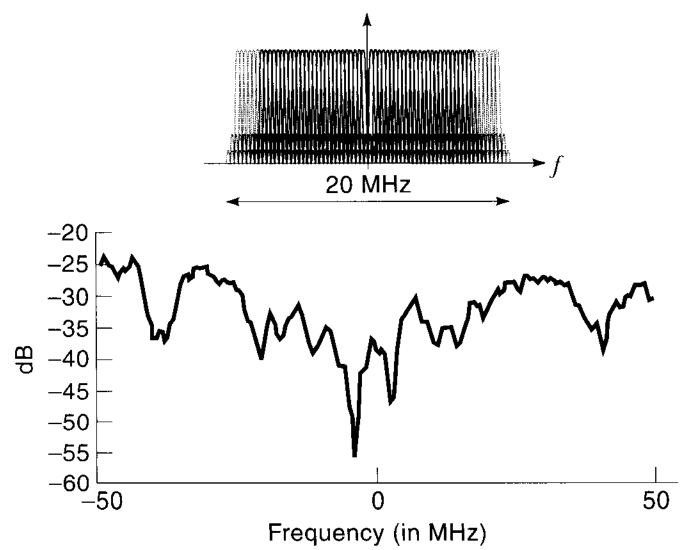


Figure 3.31 Spectrum of an OFDM carrier and the corresponding frequency response.

Similar results can be obtained if the transmit power in a subcarrier is increased as the signal in that subcarrier goes further into fade. These approaches require a feedback channel and channel measurements at the receiver that can be implemented for data applications preferring reliability to delay. Yet another approach is to use strong error correcting codes for the transmitted bits in parallel subcarriers. This way there is no need for a feedback channel and channel measurements, and the low error rate bits in the normal channels are used to correct the high error rate data from subcarriers affected by deep fades. Application of any of these frequency diversity techniques does not prevent the use of others in the same system, but the complexity of the system increases as more features are included. Coding techniques are similarly useful for FHSS systems. It means that the bits over several hops can be encoded to recover the bits lost during the fade. In many FHSS applications, the transmitted data stream is scrambled over several hops so that if a few bits are lost due to the fade, the strong coding techniques help in the recovery of these bits by providing reliable communication. Scramblers will prevent the occurrence of a string of errors in data that is more difficult to correct.

3.10.3 Space Diversity Techniques

In a fading multipath channel, the received signal at the antenna of a receiver is composed of a number of signals arriving through different paths from different spatial angles. Each path is formed after a series of reflections, transmissions, diffractions, and scattering patterns that are unique to that path. As a result, signal strength, polarization of the waveform, delay, and the angle of arrival of each path is quite different from other paths. In addition, as the location of the antenna is changed by about the wavelength of the transmitted signal, all the path structures and its associated parameters will change as well. Therefore, besides time and frequency diversity of the received signal, discussed in the past two sections, there is a significant amount of diversity in the special behavior of the signal. In the same way as time- and frequency-diversity techniques, a number of space-diversity techniques have been developed that allow a smart receiver to take advantage of the diversity in the arriving signal in the space.

Figure 3.32 illustrates the basic concepts behind four approaches to take advantage of the spatial diversity. Spatial diversity can be implemented using multiple antennas located in different locations (Fig. 3.32[a]), by using multiple antennas with different polarization located in the same location (Fig. 3.32[b]), by a sectored antenna limiting the angle of arrival of the signal (Fig. 3.32[c]) or by an antenna array that changes its antenna pattern adaptively (Fig. 3.32[d]). Although all four techniques are smart methods to take advantage of the diversity of the signal in the space, only the fourth technique is referred to as a smart antenna. Using spatially separated antennas is very common in fixed wireless systems. However, installing multiple antennas for mobile terminals is challenging in practice. Polarization diversity, however, has found its way into mobile terminals for cellular voice and wireless LAN applications.

Example 3.29: Array Antennas for Spatial Diversity

The first DFE modem [MON76] designed for fixed troposcatter communications used a quadruple antenna array with antennas spatially separated in the order of

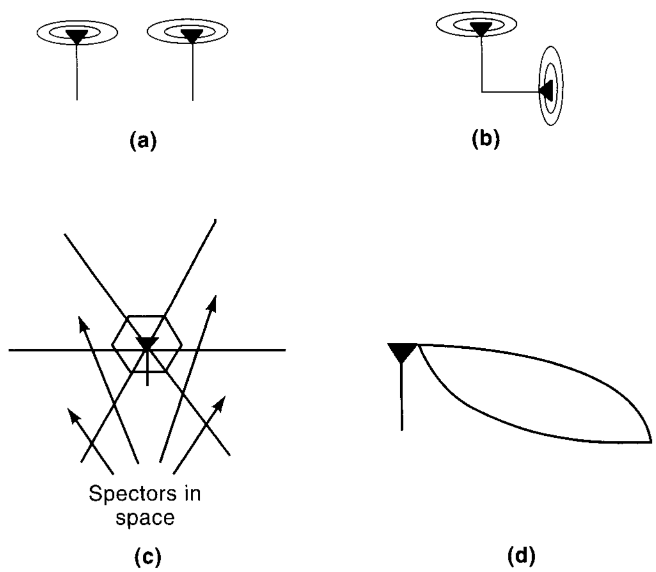


Figure 3.32 Spatial diversity schemes using (a) multiple antennas, (b) polarization diversity, (c) sectored antennas and angle diversity, and (d) adaptive angle diversity.

wavelength of the channel. WaveLAN, the first successful wireless LAN, used polarized receiver antennas for its mobile terminals [TUC91].

A sectored antenna has several sectors each selecting only the signal arriving in their field of view. In other words, the sectored antenna divides the space into several noninterfering zones. The use of sectored antennas has several advantages relative to the first two techniques. A sectored antenna reduces the interference from other users operating in the same band because it restricts the spatial angle of the arriving interference signals. A sectored antenna reduces the multipath delay spread because it only accepts a fraction of the arriving paths that fall in the sector of the antenna pattern. The reduction of multipath delay spread allows increasing the maximum data rate achievable on the channel. Using a sectored antenna, an effective diversity reception can be provided without requiring wide physical separation between antennas, making compact product packaging feasible. Sectored antennas are used in cellular systems to restrict the interference and increase the capacity of the system. In wireless LAN applications, sectored antennas have been used to increase the maximum supportable data rate by controlling the multipath arrival.

Example 3.30: Trisectored Antennas for Capacity in CDMA

Three sectored antennas are commonly used in digital cellular industry, as we will see in Chapter 5. These antennas improve the quality of transmission and increase the capacity of the network. In the next chapter when we calculate the capacity of the cellular CDMA systems, we will quantitatively show the significance of the sectored antennas in increasing the capacity of a network.

Example 3.31: Switched Sector Antennas in Motorola Altair

The Altair wireless LAN product, the first wireless LAN operating at 18–19 GHz licensed bands, used 2 six-sector antennas, one each at the transmitter and receiver, which provided a total of 36 effective sector-pairs. With such a design, the array of all multiple-reflected signal paths arriving at the receiver is divided into 36 subsets, and with appropriate signal processing (say selection diversity), the receiver can extract the subset producing the least signal degradation, discarding all the other signal arrivals. With this approach, Altair was able to support data transmission rates of up to 15 Mbps using simple four-level FSK modulation scheme.

Smart antenna systems use an adaptive antenna array that can form a beam toward a target. These antennas have found a number of applications in the cellular telephone industry that range from interference cancellation to control of eavesdropping. Because the antenna is highly directive, the gain is substantially increased, enabling better building penetration and longer ranges, thereby reducing initial deployment costs. By employing smart receiver antennas, signals from mobile terminals are isolated, reducing the near-far effect discussed earlier.

Example 3.32: Smart Antennas and CDMA

In CDMA, the multiple access interference is the primary cause for capacity limitations. Smart antennas reduce the multiple access interference by focusing the signals in one direction only, thereby increasing the system capacity.

3.11 COMPARISON OF MODULATION SCHEMES

In this section, following [FAL96], we compare the bandwidth and power requirements of various transmission techniques operating in a indoor test area, as shown in Figure 3.33. An indoor area is considered because broadband communications for applications such as WLANs are becoming more important these days. However, the overall conclusions can be extended to any other wireless fading multipath channel. The test area consists of seven rooms in the second floor of the Atwater Kent Laboratories at the Worcester Polytechnic Institute (WPI). Results of more than 600 wideband channel measurements in these rooms [HOW90] are used to calibrate a ray-tracing algorithm [YAN94] that is then used to generate several hundreds of thousands of channel profiles in the area. These profiles are then used for performance evaluation of different modem design technologies operating in this area. The basic modulation for all techniques was QPSK, and acceptable performance for a modem was considered to be a bit error rate that is better than 10^{-5} in 99 percent of location in the area. The purpose of this exercise was to examine the relationship between bandwidth and power requirements to the maximum data rate of each technique.

First we assume a fixed bandwidth of 10 MHz that is appropriate for the unlicensed operation in the PCS bands, and we compare the minimum required radia-

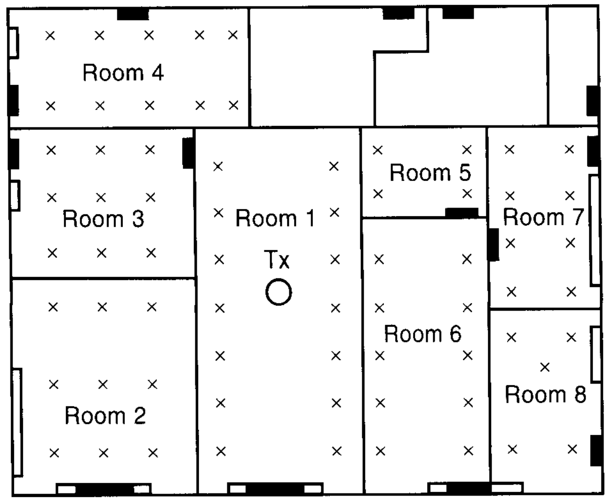


Figure 3.33 Indoor test area in WPI's Atwater Kent Laboratories.

tion power for each transmission technique to cover the test area. This section addresses the important issue of power consumption for battery-operated mobile terminals. It was found that with a maximum transmission power of 100 mW all the transmission techniques discussed here are able to cover the test area. If we assume the transmission power is maintained at 100 mW and there is no restriction on the bandwidth, we can determine the maximum data rate that can be achieved with any of the transmission techniques in the test area. This exercise addresses the transmission technologies in the context of demand for higher data rate, which has been an extremely important factor in the evolution of the LAN industry.

Figure 3.34 shows the minimum radiation power required to achieve a data rate for a 10 MHz channel operating in the test area for DSSS with a four-tap

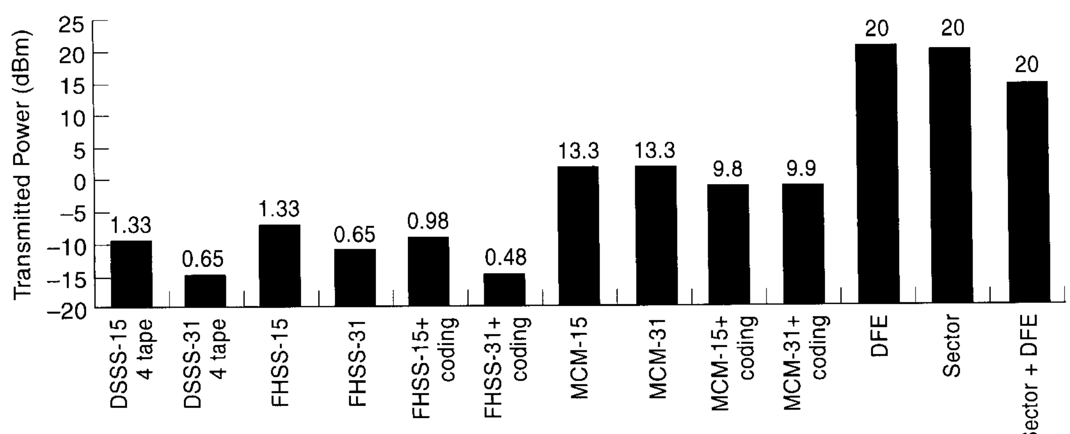


Figure 3.34 Power requirements of different transmission schemes.

RAKE receiver and processing gains of 15 and 31, as well as FHSS and MCM with and without Reed-Solomon coding (COFDM), using 15 and 31 carriers. Also presented are the power requirements to achieve the same outage probability using DFE with three forward and three feedback taps, six sectored antenna systems (SAS), and a DFE/SAS system. The number on top of each bar in the chart represents the data rate supported with that technology. Therefore, using this figure we can find the power requirements and supported data rate for a variety of modem design technologies in a fixed bandwidth (10 MHz) environment.

Example 3.33: Comparison between DFE and SAS Systems

From Figure 3.34, one can find out that in the typical area shown in Figure 3.33, with 10 MHz bandwidth, a maximum supportable data rate of 20 Mbps would be provided by DFE, SAS, or DFE/SAS systems. The DFE, a time-diversity system, and SAS, a space-diversity technique both require approximately 20 dBm (100 mW) to cover 99 percent of the area with a reasonable bit error rate. The DFE/SAS system, taking advantage of both time- and space-diversity, would need around 6 dB (four times) less transmit power at the expense of a more complex receiver that, however, will consume more electronic power for operation.

Example 3.34: Comparison between DSSS, FHSS, and DFE

A DSSS system with a processing gain of 15 (comparable to the value of 11 in IEEE 802.11) and a four-tap RAKE receiver, using time diversity, needs -10 dBm to cover the area with a data rate of 1.33 Mbps. Compared with DFE or SAS operating in the same 10 MHz bandwidth, the DSSS technique consumes about 30 dB (1,000 times) less power and supports a data rate that is about 15 times smaller. A FHSS system with 15 hopping frequencies and Reed-Solmon coding, taking advantage of frequency diversity, provides a 1 Mbps system with almost the same power consumption as DSSS-15 system.

Solomon

O

These two examples lead us to the conclusion that for fixed bandwidth channels, DFE and SAS provide the highest data rates at the expense of considerably higher power consumption. The spread spectrum systems provide a better coverage at the expense of lowering the operating data rate.

The quantitative study of radiated power consumption is important for two reasons: (1) Given the restriction on the maximum radiated power by the FCC, we need to know how the coverage of various transmission techniques compare with one another; (2) considering the increasing expansion of the market for battery-operated wireless LANs, we need to examine the total power consumption of a modem. The results of Figure 3.34 can be used directly to analyze the coverage of various transmission techniques in a typical indoor area. This allows a designer or a standards regulator to justify the selection of a particular transmission technique for a specified system description. Because the power consumption varies substantially with the design and fabrication technology used in the implementation of a WLAN, the designer can use the results of Figure 3.34 with their estimation of the electronic power consumption for their own implementation of the system.

Figure 3.35 presents the maximum attainable data rate for different transmission techniques in the test area. The transmission power is maintained a 100 mW, and there is no constraint on the bandwidth—a situation similar to U-NII bands where several hundreds of MHz of bandwidth are available for implementation of a broadband service. In Figure 3.35, the required bandwidth for each technology is indicated on top of the bar graph representing the highest supportable data rate with that technology. In a manner similar to Figure 3.34, a number of interesting practical conclusions can be drawn out of this figure as well.

Example 3.35: Data Rates with MCM

An MCM modem (OFDM) using frequency diversity with 15 carriers can achieve a data rate close to 40 Mbps, and this data rate doubled when coding (COFDM) is added to the system. Coding is a smart method to combine the frequency diversity of the received signal in an OFDM modem. With seven carriers, the data rate drops to slightly more than half of the data rate for a 15 carriers system, but the coding is not as effective as before. The significant difference between the seven carriers and 15 carriers performance reflects the fact that the bandwidth of the seven-carrier modem is not wide enough to take advantage of the frequency diversity in the received signal.

From Example 3.35 one can observe that by appropriate selection of the bandwidth of the individual carrier and by applying the correct coding technique one may completely eliminate the effects of multipath fading for the system. As a result, if there is no power or bandwidth constraint, one can achieve any data rate with multicarrier modems. This property is not shared by other techniques where the increase of data rate finally reaches to a point that the effects of frequency selectivity of the channel are dominant and an increase in the transmit power is no longer effective. The restriction in MCM is implementation complexity that increases with an increase in the number of carriers. In the practical band- and power-limited applications, this

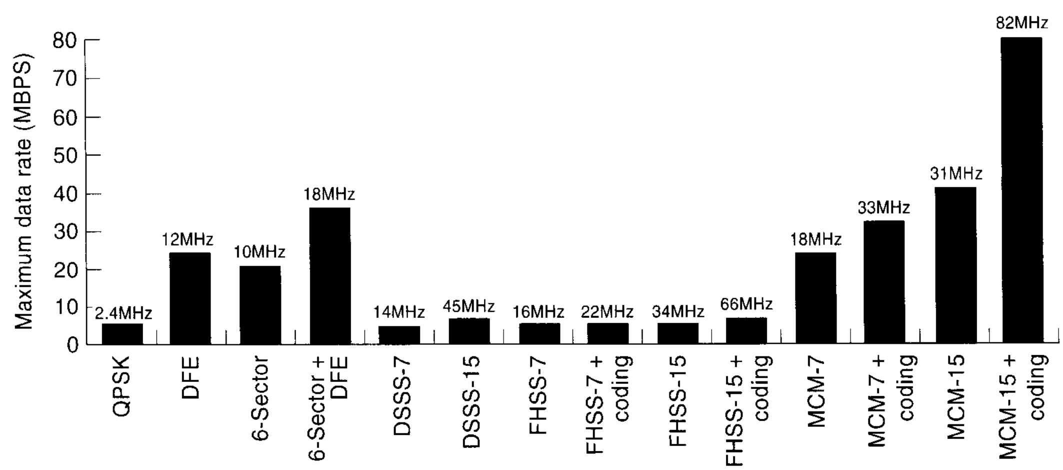


Figure 3.35 Maximum achievable data rates with different transmission schemes.

performance is expected to improve even further if, similar to IEEE 802.11a/HIPERLAN-2, a smart frequency diversity receiver uses multirate transmission or exploits measurement of the channel characteristics to adjust the power in different carriers.

The DFE modem has the advantage of having a single channel, and it can achieve data rates on the order of 20 Mbps that is considered for HIPERLAN-1. Similar results are obtained for sectored antenna systems. However, the complexity of the system for mobile applications has been the drawback for ETSI's RES-10 to adopt SAS for HIPERLAN-1 [WIL95]. Higher data rates on the order of 30 Mbps are obtained using DFE/SAS at the expense of a very complex implementation.

Obviously spread spectrum provides lower data rates. However, bandwidth efficiency is only one side of the equation; the processing gain that spread spectrum provides can serve to increase the fade margin, reducing the transmitter's power requirement. This is an important consideration if the wireless network is being used to connect portable, battery-operated computers. Besides, the results presented here are based on a single-code spread spectrum that is considered by IEEE 802.11. As we saw earlier in this chapter, by using M-ary orthogonal coding the data rate of DSSS used for 802.11 can be increased significantly to support 11 Mbps (the 802.11b standard). Then each spread spectrum code is analogous to a carrier in MCM. In a manner similar to MCM, a CDMA modem of this form has limitations caused by the hardware complexity.

3.12 CODING TECHNIQUES FOR WIRELESS COMMUNICATIONS

In this section, we discuss coding techniques for wireless communications. Coding of bits is a common technique that is employed for a variety of reasons. The most popular reason for coding is error control. Codes are also employed to convert voice into bits. The perceived quality of voice often depends on the speech coding employed, and there is a trade-off between this quality and the bandwidth requirement for digitized voice. In code division multiple access, several different codes are employed to differentiate between multiple users on the same frequency-time channel, between transmissions in various cells, and for error control. Complementary codes employed over regular spread spectrum transmissions can be used to increase the data rate by encoding data symbols with orthogonal codes.

3.12.1 Error Control Coding

In Section 3.9 we discussed a variety of diversity techniques that essentially employ *redundancy* in time, frequency, or space to improve the reliability of reception of transmitted data. In some sense, error control coding is also a diversity scheme because it introduces redundancy in the transmitted bits to correct errors that may be introduced by a channel and if correction is not possible, to provide the capability to detect the occurrence of errors.

Error control coding, as the name suggests, is a technique to *code* the transmitted bits to control the error rate. This becomes increasingly important in radio communications because of the harsh channel conditions. Errors in wireless channels usually occur in bursts. That is, a string of data bits is subject to fading or other harsh impairments such as interference, resulting in several consecutive bits (up to 50 percent of the bits in the burst) arriving at the receiver in error. This is in contrast to wired communications where errors usually occur at random, a bit at a time. Consequently, error control coding schemes are different for wired and wireless channels.

Error control coding is also dependent upon the application under consideration. Voice packets can usually tolerate error rates as high as 1 in 100 bits or 10^{-2}. Such error rates are generally unacceptable for data packets and messaging systems that require error rates as low as 10^{-5}. In some cases, it is impossible to achieve such low error rates, and in such cases, it is reasonable to *retransmit* the data packets that are lost. Such schemes are referred to as *automatic repeat request* (ARQ) schemes. In order to determine whether a packet (or a block of data bits) has been received in error, *block-coding* schemes are employed, and the process is called error detection. Block codes can also be used to correct errors, and this is called *forward error correction* (FEC). Another coding scheme that can be used for FEC is convolutional coding, which employs some memory of previously transmitted bits to determine a *most likely* sequence of transmitted bits (similar to the MLSE receiver in Section 3.10.1.2). These are discussed briefly in Appendix 3B. An interested reader may investigate details of coding schemes in references such as [SKL01], [LIN83], and [MIC85].

3.12.2 Speech Coding

Encoding analog voice into digital format has received much attention as it influences not only the quality of voice, but also the performance and capacity of the system. The important parameters of a speech code are the transmitted bit rate, the speech quality, the robustness in the presence of transmission errors, and the complexity of implementation. Speech quality is usually subjective and is determined by the mean-opinion-score [JAY84]. A low rate speech coder will essentially require less bandwidth for transmission (which is beneficial in wireless environments) but usually compromises on the quality of speech. Table 3.1 includes the details of some important speech coding techniques employed in wireless networks.

Table 3.1 Speech Coders Employed in Some Wireless Systems

System	Application	Voice Coder	Uncoded Rate	Overall Rate
GSM	European digital cellular (2G)	RPE-LTP	13 kbps	22.8 kbps
IS-136	U.S. digital cellular (2G)	VSELP	8 kbps	13 kbps
JDC	Japanese digital cellular (2G)	VSELP	8 kbps	13 kbps
IS-95	U.S. and other digital cellular CDMA	QCELP	9.6, 4.8, 2.4 and 1.2 kbps	28.8 or 19.2 kbps (FEC + repetition)
DCS-1800	PCS in the United States	RPE-LTP	13 kbps	22.8 kbps
PHS	Personal handiphone in Japan	ADPCM	32 kbps	32 kbps
CT-2	European cordless	ADPCM	32 kbps	32 kbps
DECT	Cordless and WPBX	ADPCM	32 kbps	32 kbps

There are two bit rates associated with a speech coder—the uncoded or raw bit rate and the encoded bit rate to account for error correction. Where the bandwidth efficiency is not the most important criterion and the voice quality is, a higher rate encoder such as an adaptive differential pulse code modulation (ADPCM) scheme is employed for speech coding. Resistance to channel errors is also an important issue. Voice codes may give a poor performance when the error rates are as large as 10^{-2}. This is the reason why some error control coding is applied with low-rate voice coders. The voice compression scheme removes redundancies in digitized voice, and the error-coding scheme introduces some structured redundancy to provide better performance. Block interleaving techniques (see Appendix 3B) are sometimes used to improve performance.

The important speech coding techniques are waveform encoding techniques such as pulse code modulation (PCM) and ADPCM; model-based speech coders such as linear predictive coders (LPCs), regular-pulse excitation (RPE), and code-excite linear predictive (CELP) techniques; and hybrid schemes. The complexity of implementation is low in the case of waveform encoding schemes, and the quality of voice is extremely high. The bit rates are also correspondingly larger, making them unattractive for wireless applications. LPC techniques can provide good voice quality at bit rates as low as 2,400 bps compared with the 64 kbps rate of PCM at the expense of burdensome computation. GSM uses a version of RPE, called RPE-LTP, that has an acceptable implementation complexity and delay. It operates at 13 kbps and utilizes a speech frame that lasts for 20 ms. For still further low bit rates, the quality of RPE-LTP coded voice is not adequate, and CELP techniques are preferred. Vector-sum excitation linear prediction (VSELP) and QUALCOMM's CELP (QCELP) are employed in the North American digital cellular TDMA and CDMA standards, respectively, to achieve voice coding rates of around 8 kbps.

prediction

3.12.3 Coding for Spread Spectrum Systems

Spread spectrum systems and, in particular, CDMA systems are heavily dependent on coding schemes for good performance. The literature in this field is exhaustive, and the interested reader is referred to it [PRO00]. Coding is employed for a variety of functions in CDMA systems. These include codes for separation of channels over the same frequency bands, identification of base stations covering an area, error control coding, and so on. We briefly consider these in the following sections.

3.12.3.1 PN Spreading Codes

Pseudonoise (PN) codes are also called pseudorandom sequences, and they are used as codes for spreading bits in a spread spectrum system. Even though they are called pseudorandom, the sequences are not random, but appear random because they contain almost equal numbers of zeros and ones. The most common PN sequences are the maximal-length sequences or M-sequences that are created using maximal-length linear feedback shift registers (LFSRs) of length m. The name arises from the fact that the M-sequences are the longest sequences that can be generated by an m-stage LFSR. The contents of the LFSRs repeat after a cycle of $2^m - 1$ shifts. The length of the PN sequence before it repeats itself is also $2^m - 1$. The maximal length shift registers are represented by polynomials that denote the connections that are active in the LFSR. For example, an LFSR of length $m = 3$ with feedback connections from the first (exponent 0) and second (exponent 1) shift registers (but not the third) is represented by $1 + x + x^3$. The corresponding LFSR is shown in Figure 3.36. The states of the LFSR and the outputs are also shown in this figure.

Example 3.36: PN Sequences in IS-95

On the reverse channel, IS-95 employs two PN sequences, one for the in-phase and the other for the quadrature channel of the QPSK modulations scheme. The two sequences are each of length $2^{15} - 1 = 32{,}767$. The corresponding polynomials are given by:

$$G_I(x^3) = x^{15} + x^{13} + x^9 + x^8 + x^7 + x^5 + 1$$
$$G_Q(x) = x^{15} + x^{12} + x^{11} + x^{10} + x^6 + x^5 + x^4 + x^3 + 1$$

PN sequences are widely employed because of their nice properties. Some of the important properties include the fact that they have nearly equal numbers of zeros and ones and their autocorrelation exhibits a strong peak with low sidelobes. The periodic autocorrelation has a peak value of $2^m - 1$ for zero lag and a value of -1 for all other lags. Consequently it is possible to differentiate between users by computing the correlation between their allocated sequences and a replica of the sequence of the user under consideration. Figure 3.37 shows the periodic autocorrelation of an M-sequence of length 7 generated by the LFSR of Figure 3.36.

Several other sequences exhibit good correlation properties. These include Kasami sequences [DIN98] and Barker sequences. Barker sequences of length 11 are employed in IEEE 802.11 for simply spreading the codes. The autocorrelation of this sequence is shown in Figure 3.23.

C	B	A
1	0	0
0	1	0
1	0	1
1	1	0
1	1	1
0	1	1
0	0	1
1	0	0

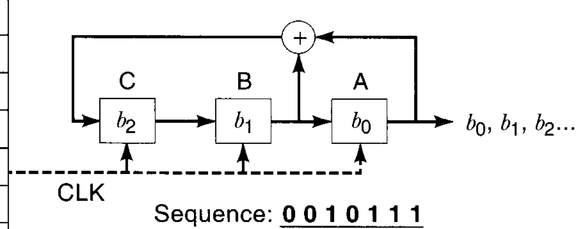


Figure 3.36 An LFSR of length $m = 3$.

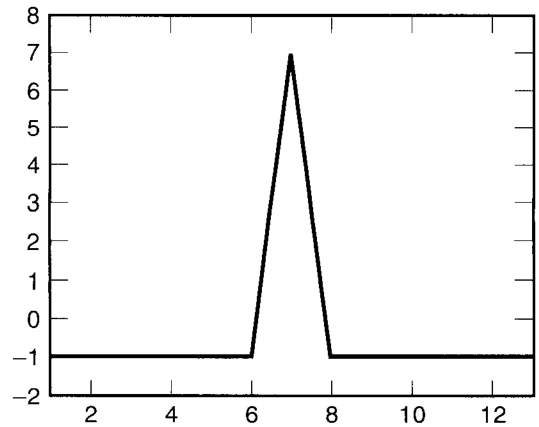


Figure 3.37 Periodic autocorrelation of an M-sequence of length 7.

3.12.3.2 Orthogonal Codes

PN sequences have good correlation properties, but the sidelobes are nonzero and cause interference to other users operating over the same channel. In order to increase capacity, CDMA systems employ orthogonal sequences. The cross-correlation between two orthogonal sequences is zero when synchronized. If two users employ orthogonal sequences for spreading at the receiver, it is possible to completely separate the two signals by correlating them with the signal replicas, without interference. The problem with orthogonal sequences is that the users must be synchronized because orthogonal sequences do not have good correlation properties outside of the zero-lag case. Consequently orthogonal sequences are employed on the downlink where the base station is able to synchronize transmissions. PN sequences are preferred for spreading on the uplink.

Example 37: Orthogonal Sequences in IS-95

IS-95 employs Walsh sequences that are generated by the rows of a Hadamard matrix. A Hadamard matrix is a square matrix of order $n \times n$ where all pairs of rows are orthogonal. The following matrix is an example of a Hadamard matrix.

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

The row [0 0 0 0] corresponds to the all-zero Walsh code. Each row of the Hadamard matrix corresponds to a unique Walsh code. It is easy to construct Hadamard matrices of order $2^m \times 2^m$. In IS-95, Walsh codes of order 64 generated by a Hadamard matrix of order 64×64 are employed on the downlink for spreading the signal.

Complementary codes are also orthogonal codes employed in the IEEE 802.11b standard. Example 3.17 discusses complementary codes.

3.13 A BRIEF OVERVIEW OF SOFTWARE RADIO

The receiver techniques discussed so far have considered traditional ways of implementation. Recently, with the immense popularity of cellular services and the emergence of a large number of standards, software implementation of the mobile terminal that can dynamically adapt itself with time to the radio environment in which it is located is becoming an attractive solution. This concept is generally referred to as *software radio* [MIT00], [BUR00]. Software radio provides the impetus for fast rollout of new services; mix-and-match services offered by a variety of standards provide choice to the customer and increase the hardware lifetime of mobile terminals. In the literature, software radio has several definitions, including (1) a software-controllable and flexible transmitter/receiver architecture, (2) replacement of radio functionality by signal processing as much as possible, (3) the ability to download an air-interface architecture and dynamically reconfigure the user terminal, (4) multimode or multistandard support, and (5) a transceiver that can define the frequency bands, the modulation and coding schemes, radio resource and mobility management, as well as user applications to be used in software. DSP technology and reconfigurable hardware technology is driving efforts toward actual implementation of software radio.

From a mobile terminal's point of view, software radio needs to have limited circuit complexity, low cost, low power consumption, and a small form factor. Ideally the analog components of a software radio are limited, and most of the radio functionality should be implemented digitally to enable software reconfigurability. However, analog-to-digital conversion at RF is extremely difficult, and instead a programmable down converter appears promising. However the limited bandwidth of the converters, the jitter introduced by the digital operations, and intermodulation products are problems with the sampled signal that are yet to be completely addressed. The processing power becomes an issue of significance in mobile terminals operating with battery power. Special-purpose DSPs, which are required for real-time computation, are expensive and complex.

For adaptive reconfigurability several solutions have been considered. Each manufacturer could have proprietary software for a variety of hardware platforms. This provides the ability to differentiate products in the market, but creates a problem for network operators, especially when mobile terminals are required to "download" the air-interface. A standard hardware platform would eliminate the numerous proprietary solutions, but would restrict the differentiability of products. A third solution proposed includes a real-time compiler that would compile a common source code into solutions for different hardware platforms. In [BUR00], Java programming language is suggested as a language for implementing the third option. This is because Java already possesses the ability to have a uniform "bytecode"[3] for all hardware platforms, requiring only an interpreter for the bytecode for each platform.

[3] A bytecode is a compiled format for Java programs that run on any Java virtual machine.

Connecting the terminal to a PC, by using smart cards or over the air, could do the download of the air-interface. Using a PC is not a feasible solution, especially while the user is on the move. Potentially, smart cards could provide a fast solution for changing the air-interface, but there are technology limitations as of today. Over-the-air downloads are preferable and require no effort by the user, and intelligent updates are possible. In this case, there is a suggestion of a *universal control channel* for accessing the radio personalities over the air. The problems with such a solution are the security of such a download, the possibility of the radio channel introducing errors during the download, delay and slowness of the download procedure, and the need for protocols, resources, and bandwidth to assist the procedure. The interested reader is referred to the software radio forum [SDRweb] for details.

APPENDIX 3A PERFORMANCE OF COMMUNICATION SYSTEMS

3A.1 Signal-to-Noise Ratio

Signal-to-noise ratio (SNR) is an important measure of performance in communication systems. Broadly speaking, the SNR is the ratio of the average signal power to the average noise power, either at the input or output of a component of the communication system. The SNR determines quantitatively the *quality* of a signal that is corrupted by noise. In digital communications, where the information is contained in the transmitted bits, the ratio of the signal energy per bit (E_b) to average noise power spectral density (N_0), commonly written as E_b/N_0 or γ_b, is important. A plot of the bit error rate versus γ_b provides the tradeoff between the transmit power requirements of modulation schemes and receiver techniques and the performance in terms of the bit error rate as discussed in the chapter. As we see there, a larger data rate can be supported over the same bandwidth using higher level modulation schemes at the expense of larger transmit powers.

Detailed discussion and calculation of the signal power or energy or the noise power is beyond the scope of this textbook and the reader is referred to [HAY00] or [SKL01]. We do however illustrate a simple case of determining the signal energy per bit E_b. In digital modulation schemes, the information-carrying signal usually consists of a train of symbols selected from a set of M symbols, each of duration T. Since there are M possible symbols, each symbol carries $\log_2 M$ bits. If there are only two symbols, we have the binary case ($M = 2$). In binary phase shift keying (BPSK), there are two symbols: $s_1(t) = \cos(2\pi f_c t)$ and $s_2(t) = \cos(2\pi f_c t + 180°)$, each lasting for a duration T seconds. Suppose $T = n/f_c$ where n is an integer. The energy in one symbol is given by:

$$E_b = \int_0^T s_1^2(t)dt = \int_0^T \cos^2(2\pi f_c t)dt = \frac{1}{2}\left(\int_0^T [1 + \cos(4\pi f_c t)]dt\right) = T/2 \quad (3A.1)$$

As another example, if $M = 8$, each symbol carries 3 bits. Phase shift keying with eight phases (8-PSK) is an example of this case. Each symbol consists of a truncated sinusoid of duration T seconds. There are eight possible truncated sinusoids (each having the same frequency but a different phase). As noted in the chapter, information can be carried either in the phase, frequency, or amplitude of the sinusoid.

The average noise power depends on the type of noise under consideration. Usually, noise arising from thermal effects on electrons is considered additive and its statistical properties are described by a Gaussian uncorrelated random process. In cellular topologies, sometimes the co-channel interference is approximated as additive, white, Gaussian noise (AWGN). In the case of CDMA, the interference from other users in the same cell and neighboring cells is also noise. While this is not Gaussian, sometimes this approximation is made. There have been proposals of schemes that use knowledge of the interference to subtract it from a given signal and thus increase capacity in CDMA systems [MOS96].

3A.2 Performance over Wired Channels

In a band-limited wired or wireless channel, the transmitted waveform is distorted because of the effects of filtering and additive noise. Figure 3A.1 shows a simple square pulse and the received pulse if the effects of filtering and additive noise are considered. In digital communications, the additive noise causes erroneous deci-

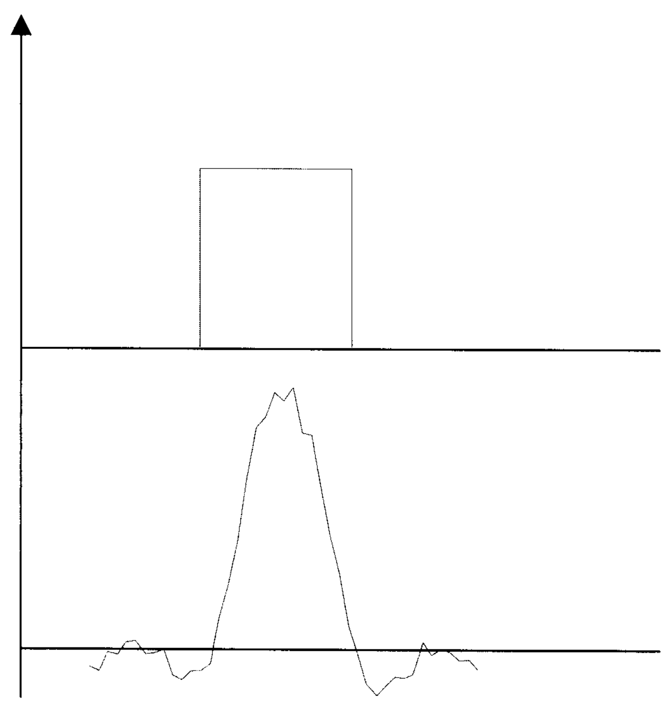


Figure 3A.1 Transmitted pulse shape and the output (received pulse shape) of a bandlimited noisy channel.

sions in detecting the transmitted bits at the receiver. To measure the performance of these transmission techniques, the bit error rate (BER) or probability of bit error (P_e) for transmission of the signal is often plotted against the signal to noise ratio (SNR) in dB for particular implementations of transceivers. The SNR can have several interpretations [PAH95], but the most common interpretation is the ratio of the signal energy *per bit* to the background noise power (γ_b). This is a measure of how much received power is required to detect information in a signal correctly with a given probability of error.

Example 3A.1: Probability of Error in AWGN

The graph in Figure 3A.2 shows the probability of bit error versus SNR for an implementation of BPSK[4] modulation technique The modem designer uses these curves to translate the customer's error rate requirement to the SNR. From the figure, if the customer demands a BER of 10^{-5}, the designer should implement the system so that the ratio of the received signal power to the background noise power at the receiver is higher than 10 dB.

Figure 3A.3 shows the graphs relating the bit error performance to γ_b in AWGN channels representing basic phase and frequency shift modulation techniques. The differential BPSK system (DBPSK) and noncoherent FSK (NC-FSK) do not need a reference of the carrier phase at the receiver, so they can be implemented easier. The purpose of the plots is to demonstrate how much performance degradation is resulted when we design the simpler receivers.

Example 3A.2: Power Requirements of Different Modulation Techniques for a Given Error Rate

As shown in Figure 3A.3, coherent FSK (CFSK) for an error rate 10^{-5} needs roughly 13 dB of γ_b. For the same error rate, an NC-FSK requires around 14 dB. Therefore, a CFSK needs a dB less power to maintain the same error rate as NC-FSK. This means that in practice if the modem is designed so that the main consumption of the battery is due to the transmission power, the lifetime of the battery in a mobile terminal using CFSK modulation is 1.25 times the life of an NC-FSK terminal. As shown in Figure 3A.3, BPSK needs around 10 dB of γ_b for an error rate 10^{-5}. For the same error rate, CFSK requires 13 dB of γ_b. Therefore, a coherent FSK needs 3 dB more power to maintain the same error rate as a BPSK system. Similarly DPSK with simpler implementation needs one more dB transmission power to match the performance of a BPSK.

3A.3 Performance over Wireless Channels

The two main characteristics of the wireless medium affecting the performance of a modem are large fluctuations of the received power level (called fading) and arrival of the received signal via delayed multiple paths referred to as multipath propagation. In the rest of this section, we give an overview of the performance of

[4]BPSK uses the phase of the carrier to carry one information bit per symbol.

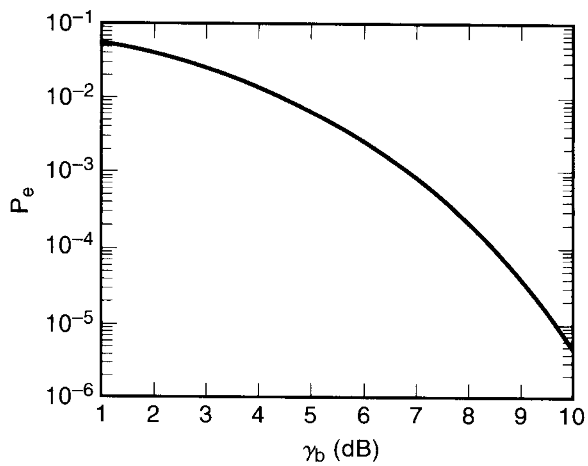


Figure 3A.2 Probability of error as a function of SNR for BPSK signals in an AWGN channel.

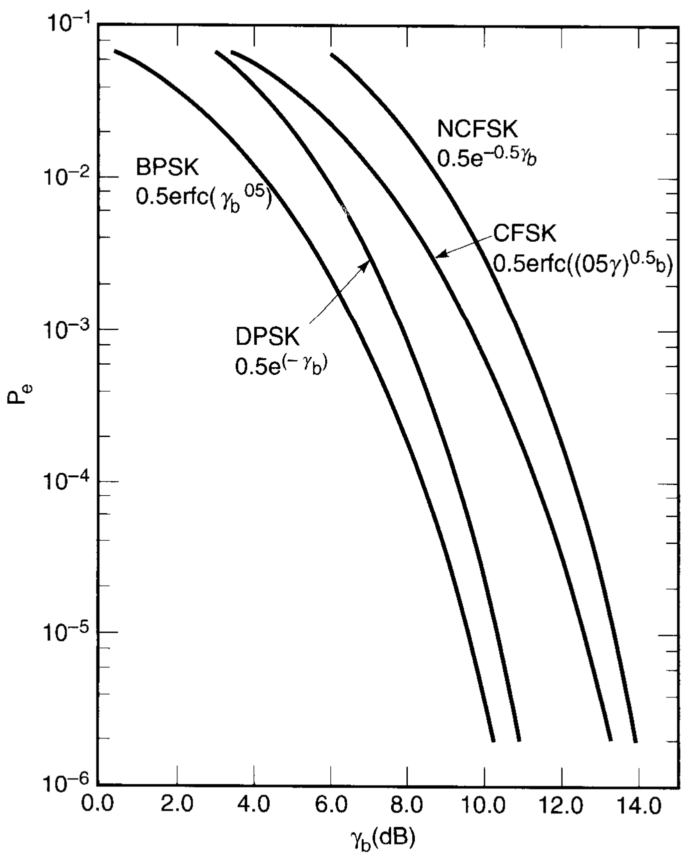


Figure 3A.3 Probability of error as a function of the SNR per bit for various modulation schemes.

basic amplitude, phase, and frequency modulation techniques, and we analyze the performance as Rayleigh fading and simple multipath conditions affect it. This will provide the reader with an intuitive understanding of the measures that are used to evaluate the performance of a modulation technique and how this measure is affected when we increase the power level or when we get exposed to fading or multipath conditions.

3A.3.1 Effects of Fading

As opposed to wired channels, the received signal from wireless channels suffers from strong amplitude fluctuations (of the order of 30–40 dB) which cause fading in the received signal. Figure 3A.4 shows a simple diagram describing the effects of fading on the BER. During periods of signal fading, the error rate of the transmission system increases, and when the system is out of fade, the error rate becomes negligible. To evaluate the performance over a fading radio channel, the *average* BER versus *average* received SNR-γ_b is used.

Example 3A.3: Probability of Error of BPSK in Rayleigh Fading Channel

The graph in Figure 3A.4 shows the average probability of error versus average received SNR over a flat Rayleigh fading channel for the same implementation of BPSK as in Example 3A.1. This time, for an average BER of 10^{-5}, we need an average received SNR of around 40 dB, which is 30 dB more than the required SNR for a nonfading wired channel.

To compensate for the additional SNR requirement and to maintain the average BER during fading, designers of wireless modems increase the SNR re-

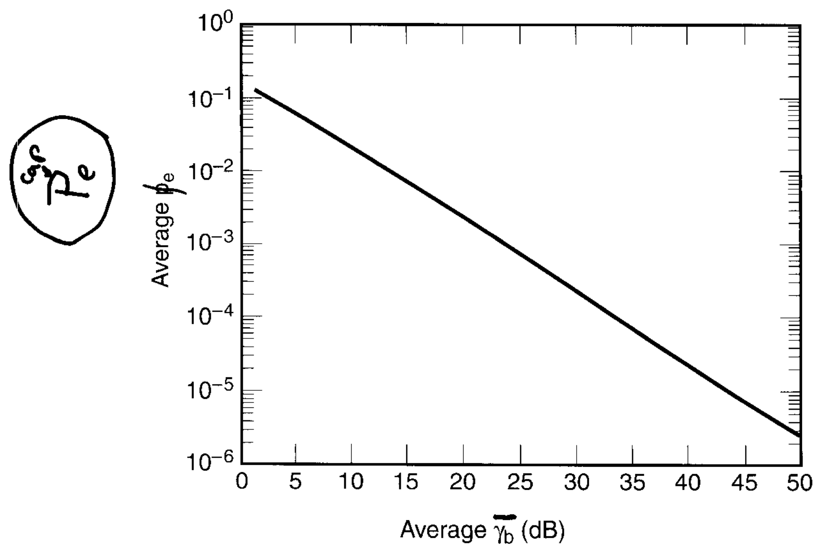


Figure 3A.4 Probability of error of binary phase shift keying in a flat Rayleigh fading channel with additive white Gaussian noise.

quirement with a margin referred to as the *fading margin.* Addition of the fading margin to the SNR requirement maintains the average performance at the required BER.

Table 3A.1 gives the BER versus SNR over a nonfading and flat Rayleigh fading channel with AWGN for basic modulation techniques. In this table, M is the number of symbols, $2^m = M$, and erfc is the complementary error function.

3A.3.2 ISI Effects Due to Multipath

One of the main differences between wireless and wired channels is that the wireless channel suffers from multipath propagation. The shape of the received pulse and its time duration of the signaling pulse are both changed due to multipath arrivals. The difference between the first and the last arriving pulses is the *delay spread* of the channel. If the symbol duration is much larger than the multipath spread, this means that the data rate is much smaller than the coherence bandwidth of the channel, and all pulses received via different paths arrive roughly on top of one another, causing amplitude fluctuations and fading. If the ratio of the delay spread to the pulse duration becomes considerable, the received pulse shape is severely distorted, and it also interferes with neighboring symbols, causing ISI. In addition to SNR fluctuations due to fading effects, the interference power also degrades the performance. However, the ISI effect of multipath degrades the performance in a different manner than fading. The effects of fading can be compensated via an increase in the transmit power by a fading margin. Increase of the transmit power cannot compensate the effects of ISI. This is because an increase in the transmit power increases the signal, as

Table 3A.1 Probabilities of Bit Error for Common Modulation Schemes

Modulation Scheme	BER in AWGN Channels	Average BER in Flat Rayleigh Fading Channels
On-Off Keying or Amplitude Shift Keying	$\frac{1}{2}\,\mathrm{erfc}\sqrt{\frac{\gamma_b}{2}}$	Hard to use in fading channels because of amplitude dependence
Binary Phase Shift Keying	$\frac{1}{2}\,\mathrm{erfc}\sqrt{\gamma_b}$	$\frac{1}{2}\left(1-\sqrt{\frac{\bar{\gamma}_b}{(1+\bar{\gamma}_b)}}\right)$
Differential Phase Shift Keying	$\frac{1}{2}e^{-\gamma_b}$	$\frac{1}{2(1+\bar{\gamma}_b)}$
Frequency Shift Keying (Coherent Detection)	$\frac{1}{2}e^{\frac{\gamma_b}{2}}$	$\frac{1}{2}\left(1-\sqrt{\frac{\bar{\gamma}_b}{(2+\bar{\gamma}_b)}}\right)$
M-ary Phase Shift Keying	$\approx \frac{2^{m-1}}{M-1}\,\mathrm{erfc}\sqrt{\sin^2\left(\frac{\pi}{M}\right)m\gamma_b}$	$\approx \frac{2^{m-1}}{M-1}\left(1-\sqrt{\frac{\sin^2\left(\frac{\pi}{m}\right)m\bar{\gamma}_b}{\left(1+\sin^2\left(\frac{\pi}{m}\right)m\bar{\gamma}_b\right)}}\right)$
M-ary QAM	$\approx \frac{2^{m-1}}{M-1}\overset{2}{\cancel{M}}\,\mathrm{erfc}\sqrt{\frac{3}{2M-1}m\gamma_b}$	Hard to use in fading channels because of amplitude dependence

In this column all "γ_b" changes to "$\bar{\gamma}_b$"

well as the ISI interference power keeping the signal to interference ratio at the same level. Example 3A.4 will clarify the situation.

Example 3A.4: ISI Effects of Multipath

Figure 3A.5(a) represents a simple two-path channel model. The delay between the arrival of the paths is t and the duration of the symbol is T. Figure 3A.5(b) shows the probability of error for a BPSK modem versus the normalized delay spread t/T for different relative powers of the second path to the power of the first path and SNR of 15 dB. In the leftmost corner, we have $t/T = 0$ (no multipath), and the error rate of the system is 10^{-8}. Until a value of t/T of around 10 percent the effects of multipath are negligible. For higher normalized values of the multipath spread, the lowest curve belongs to the case where the ratio of the strength of the second path amplitude to the main path is 10 percent, the next is for a ratio of 50 percent, and the top path belongs to the case where both paths have the same strength. For a 10 percent multipath strength, the error rate remains reasonable even for high ratios of t/T. For 50 percent and 100 percent multipath strengths, the BER degrades drastically to around 10^{-2}.

In practical situations, the multipath delay spread is measured by the RMS multipath delay spread, τ_{rms}. The RMS delay spread τ_{rms}, is the square root of the second central moment of the multipath profile of a channel that is a good measure of both distribution of amplitude strength and arrival times of the multiple signal paths. We referred to the inverse of the RMS multipath spread as the coherence bandwidth of the channel. Usually, the coherence bandwidth of the channel is known and designers try to increase the data rate by reducing the transmitted pulse duration T, while avoiding the effects of multipath arrivals. Experimental work in this field has shown that as long as the ratio τ_{rms}/T is less than 20 percent, the ISI effects of the multipath

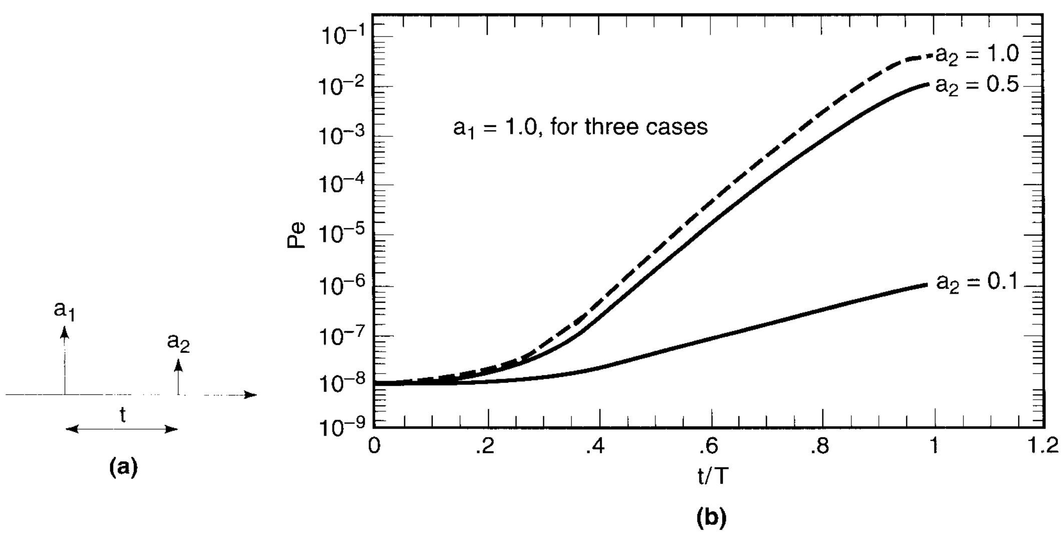


Figure 3A.5 (a) The simple two-path model and its parameters and (b) performance of BPSK in a two-path channel.

are negligible. In other words, radio design engineers try to keep the data rate to below 20 percent of the coherence bandwidth of the channel to avoid ISI caused by multipath. A number of signal processing techniques such as equalization have been developed that allow higher data rates over wireless channels.

APPENDIX 3B CODING AND CORRELATION

3B.1 Block Codes

Block coding, as the name suggests, involves encoding a *block* of bits into another block of bits, with some redundancy to combat errors. Block coding in its simplest form consists of a *parity check* bit. An extra bit is added to each block of k bits, and the extra bit is selected so that each new block of $k + 1$ bits has either an even number of ones or an odd number of ones. The extra bit is called the parity check bit. The result is that if the channel introduces a single bit error in a block of $k + 1$ bits, the number of ones in the block will no longer be even (or odd), and the receiver can *detect* the error. The simplicity of this code is clear because if there is an even number of errors in the block the errors cannot be detected because number of ones is maintained even or odd.

Using a variety of algebraic techniques, efficient encoding rules have been obtained that calculate a set of $n - k$ parity check bits which apply parity checks to a group of bits in the block of k bits. Together with these parity check bits, the size of the encoded block is $k+(n-k) = n$ bits. The block code is called an (n, k) block code, and the code rate is $C = k/n$. This means that if the raw data rate is R bps, only kR/n bps corresponds to actual data. The factor k/n is called the code rate. The rest of the bits do not contain useful information and are included only for error control purposes.

Example 3B.1: Code Rate for the GSM Control Channel

In GSM a block of 184 bits is encoded into 224 bits of codeword on the control channel before it is sent to a convolutional encoder. The number of parity check bits is 40. The code rate of this block encoder is 184/224 = 0.82.

In multilevel modulation schemes, it is possible to encode *symbols* (nonbinary alphabets) in a similar manner. Bose-Chaudhuri-Hocquenghem (BCH) codes are a popular class of nonbinary codes. Reed-Solomon (RS) codes, a subset of the BCH codes, are a good example of nonbinary block codes employed in a variety of wireless systems. The symbols are commonly drawn from a set of 2^m alphabets (each alphabet represents m bits). An RS codeword has length $n = 2^m - 1$. The number of symbols being encodes can vary from 1 to $n - 1$. Depending on the number of symbols encoded, the minimum distance between the codes (see next section), which in the case of RS codes is $n - k + 1$, changes.

Example 3B.2: Nonbinary Block Codes in CDPD

CDPD uses a (63, 47) RS encoder. Instead of encoding bits, CDPD encodes alphabets from a set of 64 symbols, each representing six bits. This means that a block of

47 symbols is encoded into a block of 63 symbols for a symbol code rate of 47/63 = 0.746. The minimum distance between the codes is $d_{min} = 63 - 47 + 1 = 17$. It can thus correct up to eight symbol errors.

3B.1.1 Operation of Block Codes

Block codes use finite-field arithmetic (modern algebraic techniques) properties to encode and decode blocks of bit or symbols. Most operations are based on linear feedback shift registers that are easy to implement and inexpensive. Most block codes are created in *systematic* form where the k data bit are retained as is and the $n-k$ parity check bits are either prepended or appended to them (see Fig. 3B.1). The parity check bit are generated via a generator matrix or generator polynomial. The encoded block of n bits is called the *codeword,* and this is transmitted over the channel. Codes generated by a polynomial are called cyclic codes, and block codes of this nature are called *cyclic redundancy check* (CRC) codes and are employed in a variety of data transmission schemes for error correction and detection.

The received word may be identical to the codeword in the case of error-free transmission and may have been modified due to channel errors. The modifications may result in another valid codeword, in which case, it is not possible either to detect or correct the errors. The probability of such a false detection is upper bounded [WOL82] by:

$$P_{FD} \leq 2^{-(n-k)} \tag{3B.1}$$

The idea behind the design of block codes is to thus have a large *distance* between any pair of codewords. This distance is measured in terms of the number of positions (bits or symbols) in which the codewords differ, and is called the Hamming distance. The *minimum* Hamming distance between the set of all codewords of a block code determines its error detection or correction capability. A block code with a minimum distance of d_{min} can detect blocks of errors that have a "weight"[5] of less than d_{min} and can correct blocks of errors that have a weight up to t_{max} where

$$t_{max} = \left[\frac{d_{min} - 1}{2}\right] \tag{3B.2}$$

Here, $[x]$ refers to the largest integer less than or equal to x. An error block is also represented in a manner similar to a block of data bits. Those bits that are not changed are represented by zeros and those that are ~~represented~~ by ones. The changed weight of the error block corresponds to the number of ones in the block, or the number of bits that are changed and thus in error. Intuitively, we can see why a block code with a minimum distance of d_{min} can correct up to t_{max} errors. Given two codewords in the set, the distance between them is greater than or equal to d_{min}. An error block modifies a codeword into the received word. If its distance from the correct codeword is *less* than half the distance between the correct codeword and

---

[5] "Weight" here refers to the number of ones in the codeword.

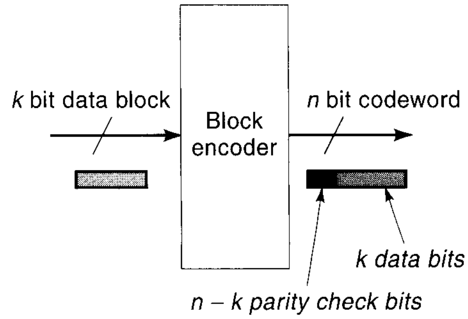


Figure 3B.1 Operation of a block code.

any other codeword, we can associate the received word as being *closest* to the original codeword, and correct it accordingly. If, however, the error block modifies the received word to make it closer to some other codeword, the error correction procedure will not work.

For sensitive data transfer, it is still possible to *detect* errors with weights larger than t_{max}. The detection of errors is performed by determining whether the received word is a valid codeword. This can be done by computing the parity check bits again from the k data bits and comparing it with the parity check bits received over the channel. It could be possible that the data bits were received correctly but the parity checks bits were in error, but the two cases are indistinguishable.

3B.2 Convolutional Codes

Unlike block codes convolutional codes do not map individual blocks of bits into blocks of codewords. Instead they accept a continuous stream of bits and map them into an output stream introducing redundancies in the process. Usually a code rate can be defined for convolutional codes as well. If there are k bits per second input to the convolutional encoder and the output is n bits per second, the code rate is once again k/n. The redundancy is however dependent on not only the incoming k bits, but also several of the preceding k bits. The number of the preceding k bits used in the encoding process is called the constraint length m that is similar to the *memory* in the system. Typically, the values of k, n, and m are 1–2, 2–3, and 4–7 in commonly employed convolutional codes.

Example 3B.3: Convolutional Coding in Wireless Systems

In the IS-95 CDMA standard, a convolutional encoder is employed in both the forward and reverse links. In the forward link, a rate 1/2 convolutional encoder is used that has a constraint length of $m = 9$. On the reverse link, a rate 1/3 convolutional encoder is employed with the same constraint length.

GSM also employs a convolutional encoder. Digitized voice is broken up into 182 class-I bits and 78 class-II bits. The most significant 50 bits of the class-I bits is enhanced with a block code that adds three parity bits. The sum total of 182 class-I bits plus the three parity bits, plus four tail bits (189 bits in all) are passed through a rate 1/2 convolutional coder to produce 378 bits. The 78 class-II bits are added to these 378 bits to produce 456 bits of encoded data.

In general, convolutional codes are more powerful than block codes in terms of FEC, but are not useful for error detection or ARQ schemes. At the receiver, FEC is performed using a maximum-likelihood decoding algorithm that determines what sequence was most likely transmitted given the received sequence of bits. The Viterbi algorithm is the most common algorithm of all, and several VLSI implementations of this algorithm are available.

3B.2.1 Hard Decision versus Soft Decision

Most receivers simply decide whether a received bit is a zero or a one and send this information to the channel decoder. The decoder employs its knowledge of the coding scheme to either detect or correct errors at this stage. Such a procedure is called *hard decision.* In soft decision decoding schemes, the receiver will convert the received signal into one of several *levels* of output. Usually there are Q quantized levels, where Q is larger than the number of alphabets. For example, in a binary system, there may be eight levels of quantized demodulator outputs instead of the usual two levels with hard decision. The decoder will use the additional information now available in order to make a decision on the received block. The Viterbi algorithm can be used for soft decisions in both convolutional encoding and block coding techniques.

3B.3 Automatic Repeat Request Schemes

In voice networks, if a packet is received in error, it is either dropped or replaced with an attenuated version of the previous packet to provide a semblance of continuity, and transmissions of damaged packets are not done because of the sensitivity of voice to delay. ARQ schemes essentially are used in data networks where reliability of received information is of paramount importance and delay is less of a problem compared with real-time multimedia applications. If a block of data is received in error, the receiver requests retransmission of the block of data. This request may be explicit or built into several protocols already operational in the system for flow control or other purposes. Usually an acknowledgment packet is employed to indicate correct reception of one or more transmitted packets. If an acknowledgment is not received in a certain time frame, or a negative acknowledgment is received, the transmitter will retransmit the packet. Such mechanisms are commonly employed with the random-access protocols discussed in Chapter 4.

There are three basic ARQ schemes. The stop-and-wait ARQ scheme waits for an acknowledgment for each individual packet before sending the next one. This is especially inefficient if the round-trip times are large because the transmitter spends a lot of time waiting for the acknowledgment. In order to improve upon this scheme, the Go-back-N ARQ scheme transmit up to N packets at a time and waits for acknowledgments. Multiple packets can be acknowledged with one response. Depending on the receipt of acknowledgments, the transmitter will back up to the last correctly received packet and retransmit the following ones (N or less packets). It is possible that some of the subsequent packets are received correctly, but they will be discarded. In order to eliminate this inefficiency, a selective-repeat ARQ scheme can be employed. Here, only those packets that are received in error are retransmitted.

Example 3B.4: Acknowledgments and Retransmissions in IEEE 802.11

In IEEE 802.11, every transmitted packet is acknowledged because the channel is unreliable. It is similar to a stop-and-wait protocol in that sense. Because round-trip times are small, and both the AP and the mobile stations share the channel, the inefficiency is limited. It is often possible to piggyback the acknowledgments over data packets transmitted in the other direction.

3B.4 Block Interleaving

Block interleaving is a technique used in wireless systems to spread the errors out over a large number of codewords. For instance, consider a Hamming code that can correct single bit errors over codewords of seven bit size. This means that if there is a single error over a block of seven bits, the coding scheme can correct it. On the other hand, a burst of five errors cannot be corrected by this code. If we can, however, spread these errors over five codewords so that each codeword "sees" only one error, it is possible to correct each of the errors. The way this works is shown in Fig. 3B.2. Codewords are arranged one below the other, and bits are transmitted vertically. At the receiver, the codewords are reconstructed, and the bits are decoded horizontally. Because the burst of errors affects the serially transmitted vertical bits that are spread over several codewords, the errors can be corrected. Block interleaving introduces delay because several codewords have to be first received before the voice packet can be reconstructed. There is only so much delay that is acceptable for normal voice conversations, and the interleaving process should not create an unacceptable value of delay in the process.

Example 3B.5: Block Interleaving in GSM

In GSM, the output of the convolutional encoder consists of 456 bits for each input of 228 bits. The 456 bits are split into eight blocks of 57 bits each. The 57 bits are spread over eight frames so that even if one frame out of five is lost, the voice quality is not affected. The delay in reconstructing the codewords corresponds to

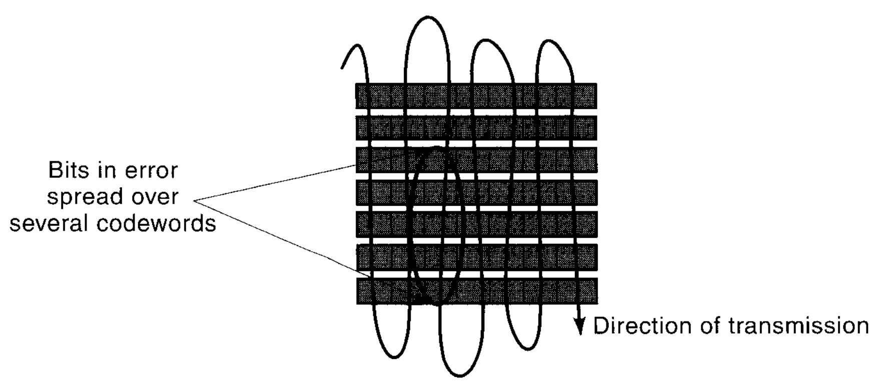


Figure 3B.2 Block interleaving.

the reception of eight frames, which takes 37 ms. A delay of 50 ms is usually tolerable for voice conversations.

3B.5 Correlation

Correlation is a measure of how similar two quantities are. When we talk of correlation between signals, we are measuring their similarity as a function of time or time lag. This essentially provides an idea of how similar a given signal $s_1(t)$ is to another signal $s_2(t)$ or a time shifted version of it [such as $s_2(t + \tau)$]. If the two signals being considered for comparison are the same, we call this the *autocorrelation*. If the two signals being considered are different, we call it the *cross-correlation*. Detailed calculation of autocorrelation and cross-correlation belong to a signals and systems course and such details are readily available in [HAY00]. We only discuss the discrete version of these and briefly.

Consider samples of two signals $a(t)$ and $b(t)$ given by two sequences $\boldsymbol{a} = [a_0, a_1, a_2, a_3, \ldots a_n]$ and $\boldsymbol{b} = [b_0, b_1, b_2, b_3, \ldots b_n]$. The periodic cross-correlation between the two sequences is given by:

$$R_{ab}(l) = \sum_{j=-n}^{n} a_j b_{j-l} \tag{3B.3}$$

where the values a_i and b_i are zero for $i < 1$ and $i > n$.

Example 3B.6: Autocorrelation and Cross-Correlation

For instance, suppose $a = [1, 1, 1, -1]$ and $b = [-1, 1, -1, 1]$, the cross-correlation is: $[1, 0, 1, -2, 1, -2, 1]$. The periodic cross-correlation does the same, but with the sequences wrapping around themselves so that each term in the correlation will be a scalar product of sequence a and a cyclically shifted version of b, i.e., the periodic autocorrelation of a and b will have the following form:

$$R_{per,ab} = [a_0b_0 + a_1b_1 + \ldots + a_nb_n, a_0b_1 + a_1b_2 + \ldots + a_nb_0, a_0b_2 + a_1b_3 + \ldots + a_nb_1, \ldots a_0b_n + a_1b_{n-1} + \ldots + a_nb_0]$$

For the particular example shown here, the periodic autocorrelation is $[-2, 2, -2, 2]$.

Autocorrelation and cross-correlation play an important role in spread spectrum where sequences with good correlation properties need to be used for robustness under interference and multipath.

QUESTIONS

3.1 Name the two most popular modulation techniques used in digital cellular modems and give one example standard that uses each of them.

3.2 What is the difference between GMSK and FSK modulation techniques? What is the difference between $\pi/4$-QPSK and QPSK modulation techniques?

3.3 For a fixed given bandwidth, which transmission technique among DFE, sectored antenna, MCM, DSSS, and FHSS provides the highest data rate and which one consumes the minimum power?

3.4 For a fixed transmission power and unlimited available bandwidth, which transmission technique among DFE, sectored antenna, MCM, DSSS, and FHSS provides the highest data rate.

3.5 In an OFDM modem with 48 channels, each channel uses 16-QAM modulation. If the overall transmission rate is 10 Mbps, what is the symbol transmission rate per channel?

3.6 Explain inter-symbol interference. What are the sources of ISI? What techniques can be used to combat ISI?

3.7 Name five design considerations in selecting a modulation scheme for a wireless network.

3.8 Why is out-of-band radiation an important issue in designing modulation schemes? How is GMSK a good solution for this?

3.9 Why is PPM used with infrared communications instead of PAM?

3.10 Differentiate between frequency hopping and direct-sequence spread spectrum.

3.11 Name four diversity techniques.

3.12 Explain how spread spectrum receivers can exploit multipath diversity using RAKE receivers.

3.13 What are sectored antennas? How are they useful in combatting multipath?

3.14 Differentiate between block codes and convolutional codes.

3.15 What is block interleaving? How is it useful in combating the effects of fast fading?

PROBLEMS

3.1 Use the results of data rate versus power consumption for the central part of the Atwater Kent building (see Figures 3.34 and 3.35) to answer the following questions:

- **a.** For 10MHz of bandwidth, what is the power requirement and maximum data rate supported by DSSS-15, MCM-15, and DFE modulation?
- **b.** What is the maximum data rate and required bandwidth for DSSS-15, MCM-15, and MCM modulations to cover this area with 100 mW power?
- **c.** Name three standards using DSSS, DFE, and MCM for implementation of wireless LANs.

3.2 For a 64-QAM modem:

- **a.** Give the SNR at which the error rate over a telephone line is 10^{-5}.
- **b.** Give the average SNR at which the average error rate over a flat Rayleigh fading radio channel is 10^{-5}.
- **c.** Give the outage rate from the threshold error rate of 10^{-5} if the system operates in a Rayleigh fading radio channel and the receiver uses a single antenna. Assume that the average received SNR per symbol is 14 dB.

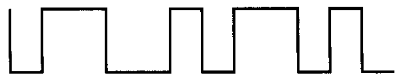
3.3 The IS-136 digital cellular replaces the AMPS analog cellular. The modulation technique for the IS-136 is $\pi/4$-QPSK.

- **a.** What is the minimum required average SNR for the IS-136 modems if the minimum acceptable average error rate is 10^{-3} and the channel is assumed to be flat Rayleigh fading?
- **b.** What is the threshold SNR if the acceptable error is 10^{-3}?
- **c.** With the average SNR of part (a) and the threshold SNR of part (b), what is the outage rate of the system?

3.4 In the following differential encoded Manchester coded signal

- **a.** Show the beginning and the end of each bit.
- **b.** Identify all the bits in the data sequence.

c. Identify the bits if it was non-differential Manchester coded.



3.5 Consider Table 3A.1 that gives the expressions for the probability of error of various modulation schemes in an AWGN channel in terms of the complementary error function.

a. Plot the error rate curves as a function of the SNR per bit. Use $M = 16$ and 64 for M-ary phase shift keying and M-ary QAM. Note that $2^m = M$ and that the SNR per bit should be expressed in absolute values in these expressions. However, plot the error curves as a function of the SNR per bit in dB.

b. Use the Matlab erfinv function (or other software tools you may prefer) to determine the SNR per bit required to obtain an error rate of 10^{-5} for each of the above modulation schemes. Comment on the results. Which modulation scheme(s) is (are) power efficient? Why?

c. Repeat (a) and (b) for Rayleigh fading channels.

3.6 This problem illustrates the concept of using orthogonal waveforms in CDMA. Figure 3.38 shows the waveforms used by three users—A, B, and C—to send messages from a wireless text transmitter device to a base station. The chip duration is 100 ns and the bit duration is 400 ns. Each of them is transmitting two bits each, corresponding to 00, 01, and 11, respectively. They transmit their waveform if the bit is a zero, and the negative of the waveform if it is a one.

a. Draw the signals corresponding to the bits transmitted by each of them.

b. Draw the composite transmitted signal that will be the sum of the individual signals if they are transmitting at the same time and are synchronized perfectly.

c. Let us refer to the transmitted signal in (b). We shall call the part of the signal during the durations 0–400 ns, 400–800 ns, and so on, as "symbols." Compute the cross-correlation at zero lag of each of the symbols in the transmitted waveform with the waveforms of A, B, and C. Comment on the results.

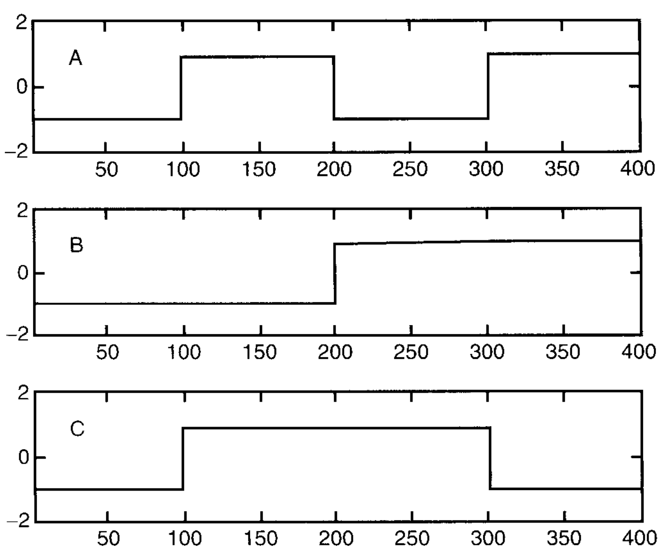


Figure 3.38 Orthogonal waveforms transmitted by three users.

3.7 Assume that instead of the waveforms shown in Figure 3.38, A, B, and C use the following waveforms:

A: $\cos(\pi t/T)$, $0 \leq t \leq T$; B: $\cos(2\pi t/T)$, $0 \leq t \leq T$; C: $\cos(3\pi t/T)$, $0 \leq t \leq T$

Draw these three waveforms for $T = 1$ microsecond. Show that they are orthogonal. Also draw the spectrum of the three waveforms. Note that these waveforms are at baseband. If a single user is transmitting all three waveforms on a single carrier, we have OFDM with three sub-carriers.

3.8 In Matlab, the function conv performs the convolution of two vectors. Samples of a signal can be represented as vectors (as in the case of spread spectrum pulses). Suppose that the M-sequence of Problem 3.1 is used as the basic waveform for transmitting a zero and its negative is used to transmit a one. The matched filter will have the flipped version of the M-sequence as its impulse response, i.e., [1, −1, −1, −1, −1, 1, 1, 1, −1, 1, 1, 1, −1, −1, 1]. You convolve the input to the matched filter with this vector to get the output. Let us suppose we are transmitting four bits: 0, 1, 1, 0. Assume also that the channel is a three-path channel with interpath delays of $5T_c$ and $8T_c$, respectively. Plot the output of the matched filter. What will be the output if you are using NRZ with a matched filter? (Note that in this case, you need to replace the M-sequence by all ones. What will the MF impulse response be?)

3.9 Show that the sequences shown in Example 3.37 are orthogonal to one another. (*Hint:* Represent the zeros by −1s in the sequences. Orthogonality is demonstrated as follows: Multiply the sequences element-wise and sum the resulting elements. If the sum is zero, the sequences are orthogonal.)

3.10 Show the steps to generate the periodic M-sequence of period 7 from the linear feedback shift register shown in Figure 3.36.

3.11 Draw the linear feedback shift register used to generate the PN sequences in the CDMA standard for digital cellular systems. Use the polynomials given in Example 3.36.

3.12 Using Reed-Solomon codes as an example, plot a graph showing the relationship between the code rate n/k and the minimum distance d_{min}. Discuss the tradeoff between reduced data rates and improved performance as the error correcting code is made more powerful.

3.13 Suppose the maximum fade duration over a radio channel is 0.001 ms. Assume that all the bits are in error when a signal encounters a fade. What is the maximum number of consecutive bits that are in error for a transmission through this channel if the data rate is 10 kbps? If the data rate is 11 Mbps?

3.14 Block interleaving is a solution to enable simple error correcting codes to correct long bursts of errors. For both the situations in Problem 3.13, determine the number of codewords over which interleaving has to be performed if the length of the codeword is 7 bits and a single bit error can be corrected.

3.15 If codewords have to be received in sequence for message delivery, determine the delay encountered by the block interleaving scheme of Problem 3.14. How does this impact voice transmission?