

## Modulation Techniques for Mobile Radio

**M**odulation is the process of encoding information from a message source in a manner suitable for transmission. It generally involves translating a baseband message signal (called the *source*) to a bandpass signal at frequencies that are very high when compared to the baseband frequency. The bandpass signal is called the *modulated* signal and the baseband message signal is called the *modulating* signal. Modulation may be done by varying the amplitude, phase, or frequency of a high frequency carrier in accordance with the amplitude of the message signal. *Demodulation* is the process of extracting the baseband message from the carrier so that it may be processed and interpreted by the intended receiver (also called the *sink*).

This chapter describes various modulation techniques that are used in mobile communication systems. Analog modulation schemes that are employed in first generation mobile radio systems, as well as digital modulation schemes proposed for use in present and future systems, are covered. Since digital modulation offers numerous benefits and is already being used to replace conventional analog systems, the primary emphasis of this chapter is on digital modulation schemes. However, since analog systems are in widespread use, and will continue to exist, they are treated first.

Modulation is a topic that is covered in great detail in various communications textbooks. Here, the coverage focuses on modulation and demodulation as it applies to mobile radio systems. A large variety of modulation techniques have been studied for use in mobile radio communications systems, and research is ongoing. Given the hostile fading and multipath conditions in the mobile radio channel, designing a modulation scheme that is resistant to mobile channel impairments is a challenging task. Since the ultimate goal of a modulation technique is to transport the message signal through a radio channel with the best

possible quality while occupying the least amount of radio spectrum, new advances in digital signal processing continue to bring about new forms of modulation and demodulation. This chapter describes many practical modulation schemes, receiver architectures, design trade-offs, and their performance under various types of channel impairments.

### 5.1 Frequency Modulation vs. Amplitude Modulation

Frequency modulation (FM) is the most popular analog modulation technique used in mobile radio systems. In FM, the amplitude of the modulated carrier signal is kept constant while its frequency is varied by the modulating message signal. Thus, FM signals have all their information in the *phase or frequency* of the carrier. As shown subsequently, this provides a nonlinear and very rapid improvement in reception quality once a certain minimum received signal level, called the FM threshold, is achieved. In amplitude modulation (AM) schemes, there is a linear relationship between the quality of the received signal and the power of the received signal since AM signals superimpose the exact relative amplitudes of the modulating signal onto the carrier. Thus, AM signals have all their information in the *amplitude* of the carrier. FM offers many advantages over amplitude modulation (AM), which makes it a better choice for many mobile radio applications.

Frequency modulation has better noise immunity when compared to amplitude modulation. Since signals are represented as frequency variations rather than amplitude variations, FM signals are less susceptible to atmospheric and impulse noise, which tend to cause rapid fluctuations in the amplitude of the received radio signal. Also, message amplitude variations do not carry information in FM, so burst noise does not affect FM system performance as much as AM systems, provided that the FM received signal is above the FM threshold. Chapter 4 illustrated how small-scale fading can cause rapid fluctuations in the received signal, thus FM offers superior qualitative performance in fading when compared to AM. Also, in an FM system, it is possible to tradeoff bandwidth occupancy for improved noise performance. Unlike AM, in an FM system, the *modulation index*, and hence bandwidth occupancy, can be varied to obtain greater signal-to-noise performance. It can be shown that, under certain conditions, the FM signal-to-noise ratio improves 6 dB for each doubling of bandwidth occupancy. This ability of an FM system to trade bandwidth for SNR is perhaps the most important reason for its superiority over AM. However, AM signals are able to occupy less bandwidth as compared to FM signals, since the transmission system is linear. In modern AM systems, susceptibility to fading has been dramatically improved through the use of in-band pilot tones which are transmitted along with the standard AM signal. The modern AM receiver is able to monitor the pilot tone and rapidly adjust the receiver gain to compensate for the amplitude fluctuations.

An FM signal is a *constant envelope* signal, due to the fact that the envelope of the carrier does not change with changes in the modulating signal. Hence the transmitted power of an FM signal is constant regardless of the amplitude of the message signal. The constant envelope of the transmitted signal allows efficient Class C power amplifiers to be used for RF power amplification of FM. In AM, however, it is critical to maintain linearity between the applied message and the amplitude of the transmitted signal, thus linear Class A or AB amplifiers, which are not as power efficient, must be used.

The issue of amplifier efficiency is extremely important when designing portable subscriber terminals since the battery life of the portable is tied to the power amplifier efficiency. Typical efficiencies for Class C amplifiers are 70%, meaning that 70% of the applied DC power to the final amplifier circuit is converted into radiated RF power. Class A or AB amplifiers have efficiencies on the order of 30-40%. This implies that for the same battery, constant envelope FM modulation may provide twice as much talk time as AM.

Frequency modulation exhibits a so-called *capture effect* characteristic. The capture effect is a direct result of the rapid nonlinear improvement in received quality for an increase in received power. If two signals in the same frequency band are available at an FM receiver, the one appearing at the higher received signal level is accepted and demodulated, while the weaker one is rejected. This inherent ability to pick up the strongest signal and reject the rest makes FM systems very resistant to co-channel interference and provides excellent subjective received quality. In AM systems, on the other hand, all of the interferers are received at once and must be discriminated after the demodulation process.

While FM systems have many advantages over AM systems, they also have certain disadvantages. FM systems require a wider frequency band in the transmitting media (generally several times as large as that needed for AM) in order to obtain the advantages of reduced noise and capture effect. FM transmitter and receiver equipment is also more complex than that used by amplitude modulation systems. Although frequency modulation systems are tolerant to certain types of signal and circuit nonlinearities, special attention must be given to phase characteristics. Both AM and FM may be demodulated using inexpensive noncoherent detectors. AM is easily demodulated using an envelope detector whereas FM is demodulated using a discriminator or slope detector. AM may be detected coherently with a product detector, and in such cases AM can outperform FM in weak signal conditions since FM must be received above threshold.

## 5.2 Amplitude Modulation

In amplitude modulation, the amplitude of a high frequency carrier signal is varied in accordance to the instantaneous amplitude of the modulating message signal. If  $A_c \cos(2\pi f_c t)$  is the carrier signal and  $m(t)$  is the modulating message signal, the AM signal can be represented as

$$s_{AM}(t) = A_c [1 + m(t)] \cos(2\pi f_c t) \quad (5.1)$$

The *modulation index*  $k$  of an AM signal is defined as the ratio of the peak message signal amplitude to the peak carrier amplitude. For a sinusoidal modulating signal  $m(t) = (A_m/A_c) \cos(2\pi f_m t)$ , the modulation index is given by

$$k = \frac{A_m}{A_c} \quad (5.2)$$

The modulation index is often expressed as a percentage, and is called *percentage modulation*. Figure 5.1 shows a sinusoidal modulating signal and the corresponding AM signal. For the case shown in Figure 5.1,  $A_m = 0.5A_c$ , and the signal is said to be 50% modulated. A percentage of modulation greater than 100% will distort the message signal if detected by an envelope detector. Equation (5.1) may be equivalently expressed as

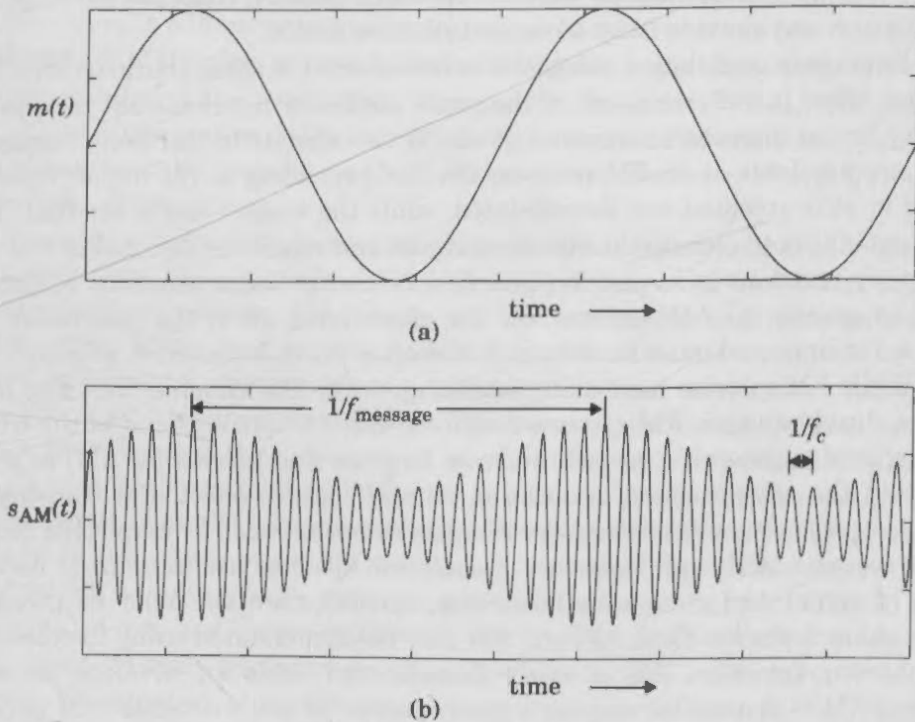


Figure 5.1

(a) A sinusoidal modulating signal.

(b) Corresponding AM signal with modulation index 0.5.

$$s_{AM}(t) = \text{Re} \{ g(t) \exp(j2\pi f_c t) \} \quad (5.3)$$

where  $g(t)$  is the complex envelope of the AM signal given by

$$g(t) = A_c [1 + m(t)] \quad (5.4)$$

The spectrum of an AM signal can be shown to be

$$S_{AM}(f) = \frac{1}{2}A_c [\delta(f-f_c) + M(f-f_c) + \delta(f+f_c) + M(f+f_c)] \quad (5.5)$$

where  $\delta(\cdot)$  is the unit impulse function, and  $M(f)$  is the message signal spectrum. Figure 5.2 shows an AM spectrum for a message signal whose magnitude spectrum is a triangular function. As seen from Figure 5.2, the AM spectrum consists of an impulse at the carrier frequency, and two sidebands which replicate the message spectrum. The sidebands above and below the carrier frequency are called the *upper* and *lower* sidebands, respectively. The bandwidth of an AM signal is equal to

$$B_{AM} = 2f_m \quad (5.6)$$

where  $f_m$  is the maximum frequency contained in the modulating message signal. The total power in an AM signal can be shown to be

$$P_{AM} = \frac{1}{2}A_c^2 [1 + 2\langle m(t) \rangle + \langle m^2(t) \rangle] \quad (5.7)$$

where  $\langle \cdot \rangle$  represents the average value. If the modulating signal is  $m(t) = k \cos(2\pi f_m t)$ , equation (5.7) may be simplified as

$$P_{AM} = \frac{1}{2}A_c^2 [1 + P_m] = P_c \left[ 1 + \frac{k^2}{2} \right] \quad (5.8)$$

where  $P_c = A_c^2/2$  is the power in the carrier signal,  $P_m = \langle m^2(t) \rangle$  is the power in the modulating signal  $m(t)$ , and  $k$  is the modulation index.

### Example 5.1

A zero mean sinusoidal message is applied to a transmitter that radiates an AM signal with 10 kW power. Compute the carrier power if the modulation index is 0.6. What percentage of the total power is in the carrier? Calculate the power in each sideband.

### Solution to Example 5.1

Using equation (5.8) we have

$$P_c = \frac{P_{AM}}{1 + k^2/2} = \frac{10}{1 + 0.6^2/2} = 8.47 \text{ kW}$$

Percentage power in the carrier is

$$\frac{P_c}{P_{AM}} \times 100 = \frac{8.47}{10} \times 100 = 84.7\%$$

Power in each sideband is given by

$$\frac{1}{2}(P_{AM} - P_c) = 0.5 \times (10 - 8.47) = 0.765 \text{ kW}$$

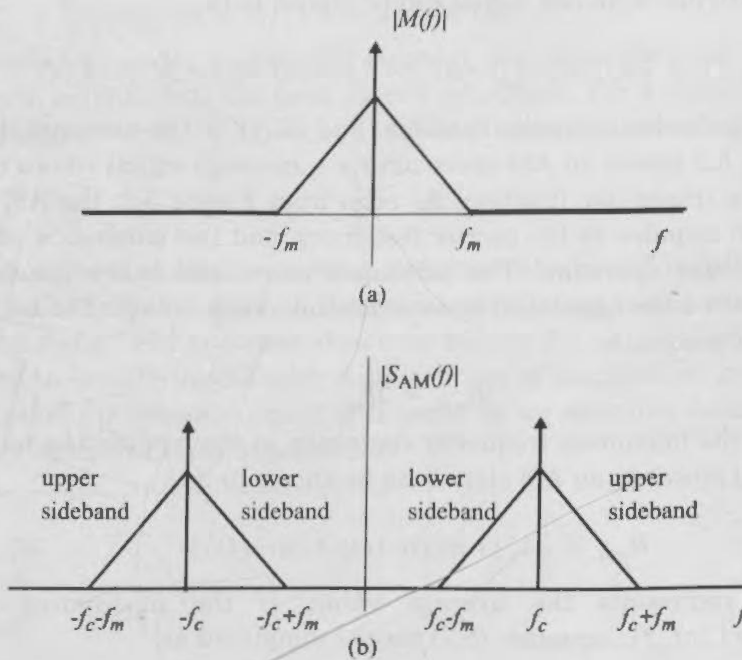


Figure 5.2  
 (a) Spectrum of a message signal.  
 (b) Spectrum of the corresponding AM signal.

### 5.2.1 Single Sideband AM

Since both the sidebands of an AM signal carry the same information, it is possible to remove one of them without losing any information. Single sideband (SSB) AM systems transmit only one of the sidebands (either upper or lower) about the carrier, and hence occupy only half the bandwidth of conventional AM systems. An SSB signal can be mathematically expressed as

$$S_{SSB}(t) = A_c [m(t) \cos(2\pi f_c t) \mp \hat{m}(t) \sin(2\pi f_c t)] \tag{5.9}$$

where the negative sign in equation (5.9) is used for upper sideband SSB and the positive sign is used for lower sideband SSB. The term  $\hat{m}(t)$  denotes the Hilbert transform of  $m(t)$  which is given by

$$\hat{m}(t) = m(t) \otimes h_{HT}(t) = m(t) \otimes \frac{1}{\pi t} \tag{5.10}$$

and  $H_{HT}(f)$ , the Fourier transform of  $h_{HT}(t)$ , corresponds to a  $-90^\circ$  phase shift network

$$H(f) = \begin{cases} -j & f > 0 \\ j & f < 0 \end{cases} \tag{5.11}$$

The two common techniques used for generating an SSB signal are the *filter method* and the *balanced modulator* method. In the filter method, SSB signals are generated by passing a double sideband AM signal through a bandpass filter which removes one of the sidebands. A block diagram of such a modulator is shown in Figure 5.3a. Excellent sideband suppression can be obtained using crystal filters at an intermediate frequency (IF).

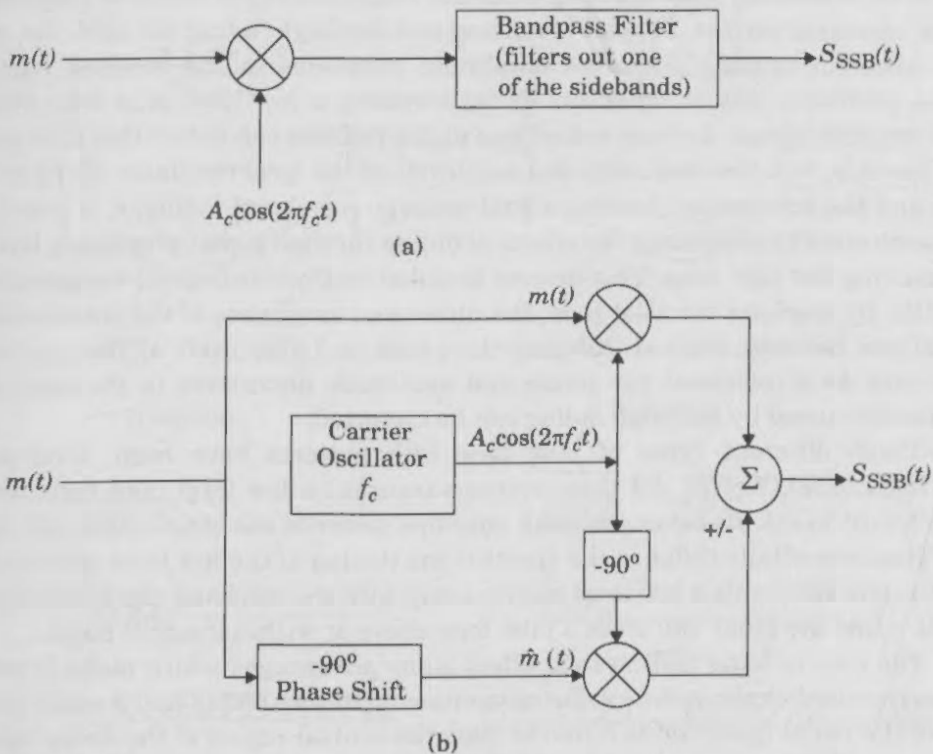


Figure 5.3  
Generation of SSB using (a) a sideband filter, and (b) balanced modulator.

Figure 5.3b shows a block diagram of a balanced modulator which is a direct implementation of equation (5.9). The modulating signal is split into two identical signals, one which modulates the in-phase carrier and the other which is passed through a  $-90^\circ$  phase shifter before modulating a quadrature carrier. The sign used for the quadrature component determines whether USSB or LSSB is transmitted.

### 5.2.2 Pilot Tone SSB

While SSB systems have the advantage of being very bandwidth efficient, their performance in fading channels is very poor. For proper detection of SSB

signals, the frequency of the oscillator at the product detector mixer in the receiver must be the same as that of the incoming carrier frequency. If these two frequencies are not identical, product detection will lead to shifting the demodulated spectrum by an amount equal to the difference in the frequencies between the incoming carrier and local oscillator. This leads to an increase or decrease in the pitch of the received audio signal. In conventional SSB receivers, it is difficult to electronically tune the local oscillator frequency to the identical frequency of the incoming carrier. Doppler spreading and Rayleigh fading can shift the signal spectrum causing pitch and amplitude variations in the received signal. These problems may be overcome by transmitting a low level pilot tone along with the SSB signal. A phase locked loop at the receiver can detect this pilot tone and use it to lock the frequency and amplitude of the local oscillator. If the pilot tone and the information bearing signal undergo correlated fading, it is possible at the receiver to counteract the effects of fading through signal processing based on tracking the pilot tone. This process is called *feedforward signal regeneration* (FFSR). By tracking the pilot tone, the phase and amplitude of the transmitted signal can be reestablished. Keeping the phase and amplitude of the received pilot tone as a reference, the phase and amplitude distortions in the received sidebands caused by Rayleigh fading can be corrected.

Three different types of pilot tone SSB systems have been developed [Gos78],[Lus78],[Wel78]. All three systems transmit a low level pilot tone, usually -7.5 dB to -15 dB below the peak envelope power of the single sideband signal. They essentially differ in the spectral positioning of the low level pilot tone. One system transmits a low level carrier along with the sideband signal (tone-in-band), while the other two place a pilot tone above or within the SSB band.

The *tone-in-band* SSB system offers many advantages which make it particularly suited to the mobile radio environment. In this technique, a small portion of the audio spectrum is removed from the central region of the audio band using a notch filter, and a low level pilot tone is inserted in its place. This has the advantage of maintaining the low bandwidth property of the SSB signal, while at the same time providing good adjacent channel protection. Due to very high correlation between the fades experienced by the pilot tone and the audio signals, a tone-in-band system makes it possible to employ some form of feedforward automatic gain and frequency control to mitigate the effects of multipath induced fading.

For proper operation of tone-in-band SSB, the tone must be transparent to data and be spaced across the band to avoid spectral overlap with audio frequencies. McGeehan and Bateman [McG84] proposed a *Transparent Tone-In-Band* (TTIB) system which satisfies these requirements. Figure 5.4 illustrates the proposed technique. The baseband signal spectrum is split into two approximately equal width segments. The upper frequency band is filtered out separately and upconverted by an amount equal to the required notch width. The low level pilot



tone is added to the center of the resultant notch, and the composite signal is then transmitted. At the receiver, the pilot tone is removed for automatic gain and frequency control purposes, and complementary frequency translation operations are performed to regenerate the audio spectrum. The TTIB system directly trades system bandwidth for notch width. The selection of notch width depends on the maximum Doppler spread induced by the channel, as well as practical filter rolloff factors.

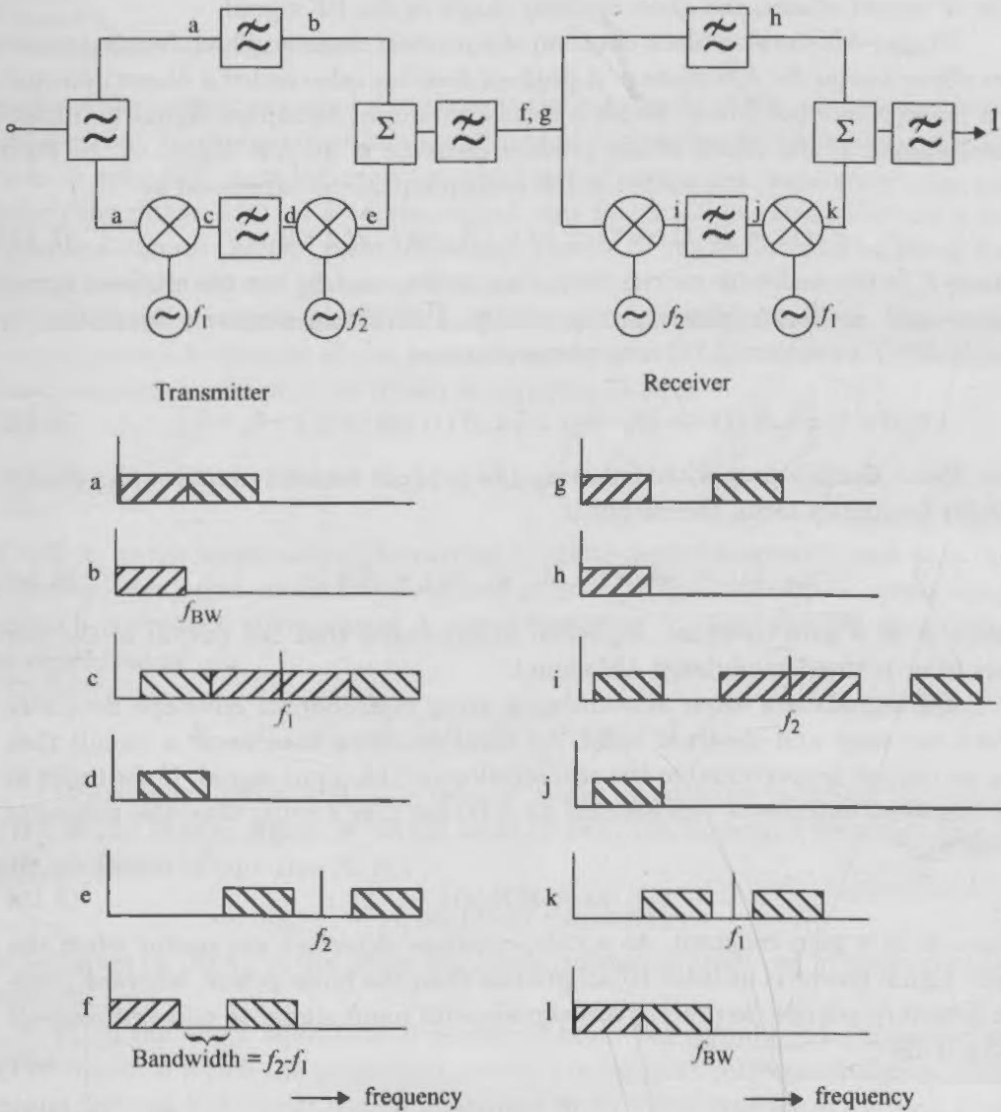


Figure 5.4  
 Illustration of transparent tone-in-band system [From [McG84] © IEEE]. Only positive frequencies are shown, and the two different cross-hatchings denote different spectral bands.

### 5.2.3 Demodulation of AM signals

AM demodulation techniques may be broadly divided into two categories: *coherent* and *noncoherent* demodulation. Coherent demodulation requires knowledge of the transmitted carrier frequency and phase at the receiver, whereas noncoherent detection requires no phase information. In practical AM receivers, the received signal is filtered and amplified at the carrier frequency and then converted to an intermediate frequency (IF) using a superhetrodyne receiver. The IF signal retains the exact spectral shape as the RF signal.

Figure 5.5 shows a block diagram of a *product detector* which forms a coherent demodulator for AM signals. A product detector (also called a phase detector) is a down converter circuit which converts the input bandpass signal to a base-band signal. If the input to the product detector is an AM signal of the form  $R(t) \cos(2\pi f_c t + \theta_r)$ , the output of the multiplier can be expressed as

$$v_1(t) = R(t) \cos(2\pi f_c t + \theta_r) A_0 \cos(2\pi f_c t + \theta_0) \quad (5.12)$$

where  $f_c$  is the oscillator carrier frequency, and  $\theta_r$  and  $\theta_0$  are the received signal phase and oscillator phases, respectively. Using trigonometric identities in Appendix D, equation (5.12) may be rewritten as

$$v_1(t) = \frac{1}{2} A_0 R(t) \cos(\theta_r - \theta_0) + \frac{1}{2} A_0 R(t) \cos[\pi 2f_c t + \theta_r + \theta_0] \quad (5.13)$$

Since the low pass filter following the product detector removes the double carrier frequency term, the output is

$$v_{out}(t) = \frac{1}{2} A_0 R(t) \cos[\theta_r - \theta_0] = KR(t) \quad (5.14)$$

where  $K$  is a gain constant. Equation (5.14) shows that the output of the low pass filter is the demodulated AM signal.

AM signals are often demodulated using noncoherent envelope detectors which are easy and cheap to build. An ideal envelope detector is a circuit that has an output proportional to the real envelope of the input signal. If the input to the envelope detector is represented as  $R(t) \cos(2\pi f_c t + \theta_r)$ , then the output is given by

$$v_{out}(t) = K|R(t)| \quad (5.15)$$

where  $K$  is a gain constant. As a rule, envelope detectors are useful when the input signal power is at least 10 dB greater than the noise power, whereas product detectors are able to process AM signals with input signal-to-noise ratios well below 0 dB.

## 5.3 Angle Modulation

FM is part of a more general class of modulation known as *angle modulation*. Angle modulation varies a sinusoidal carrier signal in such a way that the

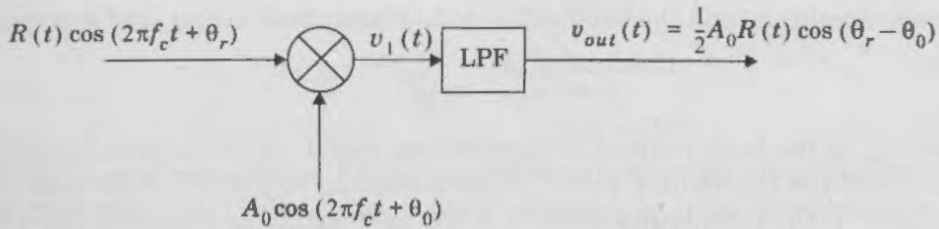


Figure 5.5  
Block diagram of a product detector.

angle of the carrier is varied according to the amplitude of the modulating baseband signal. In this method, the amplitude of the carrier wave is kept constant (this is why FM is called *constant envelope*). There are a number of ways in which the phase  $\theta(t)$  of a carrier signal may be varied in accordance with the baseband signal; the two most important classes of angle modulation being *frequency modulation* and *phase modulation*.

*Frequency modulation* (FM) is a form of angle modulation in which the instantaneous frequency of the carrier signal is varied linearly with the baseband message signal  $m(t)$ , as shown in equation (5.16).

$$S_{\text{FM}}(t) = A_c \cos [2\pi f_c t + \theta(t)] = A_c \cos \left[ 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\eta) d\eta \right] \quad (5.16)$$

where  $A_c$  is the amplitude of the carrier,  $f_c$  is the carrier frequency, and  $k_f$  is the frequency deviation constant (measured in units of Hz/volt). If the modulating signal is a sinusoid of amplitude  $A_m$ , and frequency  $f_m$ , then the FM signal may be expressed as

$$S_{\text{FM}}(t) = A_c \cos \left[ 2\pi f_c t + \frac{k_f A_m}{f_m} \sin(2\pi f_m t) \right] \quad (5.17)$$

*Phase modulation* (PM) is a form of angle modulation in which the angle  $\theta(t)$  of the carrier signal is varied linearly with the baseband message signal  $m(t)$ , as shown in equation (5.18).

$$S_{\text{PM}}(t) = A_c \cos [2\pi f_c t + k_\theta m(t)] \quad (5.18)$$

In equation (5.18)  $k_\theta$  is the phase deviation constant (measured in units of radians/volt).

From the above equations, it is clear that an FM signal can be regarded as a PM signal in which the modulating wave is integrated before modulation. This means that an FM signal can be generated by first integrating  $m(t)$  and then using the result as an input to a phase modulator. Conversely, a PM wave can be generated by first differentiating  $m(t)$  and then using the result as the input to a frequency modulator.

The frequency modulation index  $\beta_f$  defines the relationship between the message amplitude and the bandwidth of the transmitted signal, and is given by

$$\beta_f = \frac{k_f A_m}{W} = \frac{\Delta f}{W} \quad (5.19)$$

where  $A_m$  is the peak value of the modulating signal,  $\Delta f$  is the peak frequency deviation of the transmitter and  $W$  is the maximum bandwidth of the modulating signal. If the modulating signal is a low pass signal, as is usually the case, then  $W$  is equal to the highest frequency component  $f_{max}$  present in the modulating signal.

The phase modulation index  $\beta_p$  is given by

$$\beta_p = k_\theta A_m = \Delta\theta \quad (5.20)$$

where  $\Delta\theta$  is the peak phase deviation of the transmitter.

### Example 5.2

A sinusoidal modulating signal,  $m(t) = 4 \cos 2\pi 4 \times 10^3 t$ , is applied to an FM modulator that has a frequency deviation constant gain of 10 kHz/V. Compute (a) the peak frequency deviation, and (b) the modulation index.

### Solution to Example 5.2

Given:

Frequency deviation constant  $k_f = 10 \text{ kHz/V}$

Modulating frequency,  $f_m = 4 \text{ kHz}$

a) The maximum frequency deviation will occur when the instantaneous value of the input signal is at its maximum. For the given  $m(t)$ , the maximum value is 4 V, and hence the peak deviation is equal to

$$\Delta f = 4 \text{ V} \times 10 \text{ kHz/V} = 40 \text{ kHz}$$

b) The modulation index is given by

$$\beta_f = \frac{k_f A_m}{f_m} = \frac{\Delta f}{f_m} = \frac{40}{4} = 10$$

### 5.3.1 Spectra and Bandwidth of FM Signals

When a sinusoidal test tone is used such that  $m(t) = A_m \cos 2\pi f_m t$ , the spectrum of  $S_{FM}(t)$  contains a carrier component and an infinite number of sidebands located on either side of the carrier frequency, spaced at integer multiples of the modulating frequency  $f_m$ . Since  $S_{FM}(t)$  is a nonlinear function of  $m(t)$ , the spectrum of an FM signal must be evaluated on a case-by-case basis for a particular modulating wave shape of interest. It can be shown that for a sinusoidal message, amplitudes of the spectral components are given by Bessel functions of the modulation index  $\beta_f$ .

An FM signal has 98% of the total transmitted power in a RF bandwidth  $B_T$ , given by

$$B_T = 2(\beta_f + 1)f_m \quad (\text{Upper bound}) \quad (5.21)$$

$$B_T = 2\Delta f \quad (\text{Lower bound}) \quad (5.22)$$

The above approximation of FM bandwidth is called Carson's rule. Carson's rule states that for small values of modulation index ( $\beta_f < 1$ ), the spectrum of an FM wave is effectively limited to the carrier frequency  $f_c$ , and one pair of side band frequencies at  $f_c \pm f_m$ , and that for large values of modulation index, the bandwidth approaches, and is only slightly greater than,  $2\Delta f$ .

As a practical example of quantifying the spectrum of an FM signal, the U.S. AMPS cellular system uses a modulation index  $\beta_f = 3$ , and  $f_m = 4$  kHz. Using Carson's rule, the AMPS channel bandwidth has an upper bound of 32 kHz and a lower bound of 24 kHz. However, in practice, the AMPS standard only specifies that the modulation products outside 20 kHz from the carrier shall not exceed 26 dB below the unmodulated carrier. It is further specified that the modulation products outside  $\pm 45$  kHz from the carrier shall not exceed 45 dB below the unmodulated carrier [EIA90].

### Example 5.3

An 880 MHz carrier signal is frequency modulated using a 100 kHz sinusoidal modulating waveform. The peak deviation of the FM signal is 500 kHz. If this FM signal is received by a superheterodyne receiver having an IF frequency of 5 MHz, determine the IF bandwidth necessary to pass the signal.

### Solution to Example 5.3

Given:

Modulating frequency,  $f_m = 100$  kHz

Frequency deviation,  $\Delta f = 500$  kHz

Therefore modulation index,  $\beta_f = \Delta f / f_m = 500 / 100 = 5$

Using Carson's rule, the bandwidth occupied by the FM signal is given by

$$B_T = 2(\beta_f + 1)f_m = 2(5 + 1)100 \text{ kHz} = 1200 \text{ kHz}$$

The IF filter at the receiver needs to pass all the components in this bandwidth, hence the IF filter should be designed for a bandwidth of 1200 kHz.

### 5.3.2 FM Modulation Methods

There are basically two methods of generating an FM signal: the *direct method* and the *indirect method*. In the direct method, the carrier frequency is directly varied in accordance with the input modulating signal. In the indirect method, a narrowband FM signal is generated using a balanced modulator, and

frequency multiplication is used to increase both the frequency deviation and the carrier frequency to the required level.

### Direct Method

In this method, voltage-controlled oscillators (VCO) are used to vary the frequency of the carrier signal in accordance with the baseband signal amplitude variations. These oscillators use devices with reactance that can be varied by the application of a voltage, where the reactance causes the instantaneous frequency of the VCO to change proportionally. The most commonly used variable reactance device is the voltage-variable capacitor called a *varactor*. The voltage-variable capacitor may be obtained, for example, by using a reverse biased p-n junction diode. The larger the reverse voltage applied to such a diode, the smaller the transition capacitance will be of the diode. By incorporating such a device into a standard Hartley or Colpitts oscillator, FM signals can be generated. Figure 5.6 shows a simple reactance modulator. While VCOs offer a simple way to generate narrowband FM signals, the stability of the center frequency (carrier) of the VCO becomes a major issue when it is used for wideband FM generation. The stability of the VCO can be improved by incorporating a phase locked loop (PLL) which locks the center frequency to a stable crystal reference frequency.

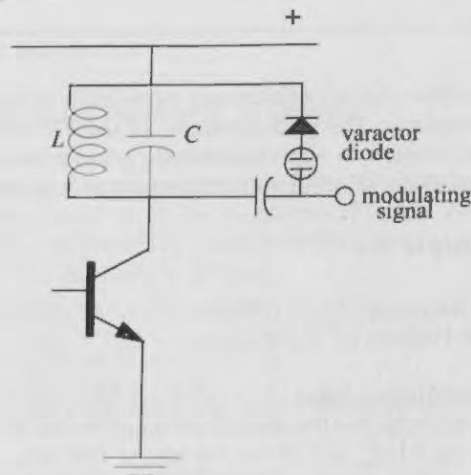


Figure 5.6

A simple reactance modulator in which the capacitance of a varactor diode is changed to vary the frequency of a simple oscillator. This circuit serves as a VCO.

### Indirect Method

The indirect method of generating FM was first proposed by its inventor, Major Edwin Armstrong, in 1936. It is based on approximating a narrowband FM signal as the sum of a carrier signal and a single sideband (SSB) signal where the sideband is  $90^\circ$  out of phase with the carrier. Using a Taylor series for small values of  $\theta(t)$ , equation (5.16) can be expressed as

$$S_{FM}(t) \cong A_c \cos 2\pi f_c t - A_c \theta(t) \sin 2\pi f_c t \quad (5.23)$$

where the first term represents the carrier and the second term represents the sideband.

A simple block diagram of the indirect FM transmitter is shown in Figure 5.7. A narrow band FM signal is generated using a balanced modulator which modulates a crystal controlled oscillator. Figure 5.7 is a direct implementation of equation (5.23). The maximum frequency deviation is kept constant and small in order to maintain the validity of equation (5.23), and hence the output is a narrowband FM signal. A wideband FM signal is then produced by multiplying in frequency the narrowband FM signal using frequency multipliers. A disadvantage of using the indirect method for wideband FM generation is that the phase noise in the system increases with the frequency multiplying factor  $N$ .

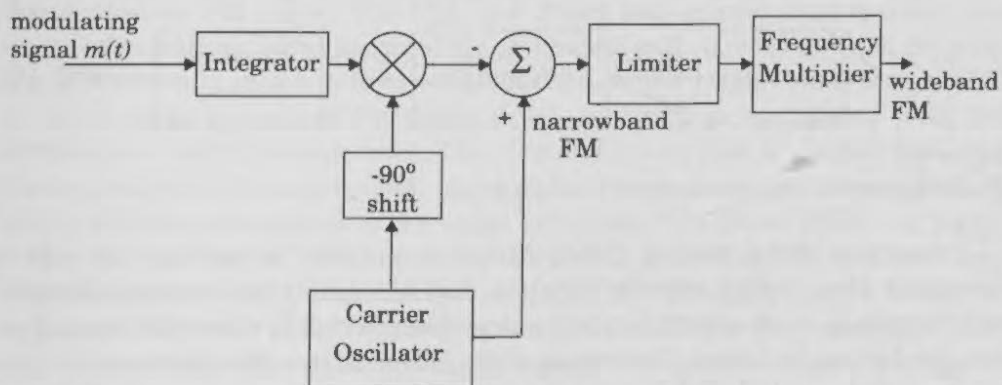


Figure 5.7

Indirect method for generating a wideband FM signal. A narrowband FM signal is generated using a balanced modulator and then frequency multiplied to generate a wideband FM signal.

### 5.3.3 FM Detection Techniques

There are many ways to recover the original information from an FM signal. The objective of all FM demodulators is to produce a transfer characteristic that is the inverse of that of the frequency modulator. That is, a frequency demodulator should produce an output voltage with an instantaneous amplitude that is directly proportional to the instantaneous frequency of the input FM signal. Thus, a frequency-to-amplitude converter circuit is a frequency demodulator. Various techniques such as slope detection, zero-crossing detection, phase locked discrimination and quadrature detection are used to demodulate FM. Devices which perform FM demodulation are often called *frequency discriminators*. In practical receivers, the RF signal is received, amplified, and filtered at the carrier and then converted to an intermediate frequency (IF) which contains the same spectrum as the original received signal.

### Slope Detector

It can be easily shown that FM demodulation can be performed by taking the time derivative (often called *slope detection*) of the FM signal, followed by envelope detection. A block diagram of such an FM demodulator is shown in Figure 5.8. The FM signal is first passed through an amplitude limiter which removes any amplitude perturbations which the signal might have undergone due to fading in the channel, and produces a constant envelope signal. Using equation (5.16) the signal at the output of the limiter can be represented as

$$v_1(t) = V_1 \cos [2\pi f_c t + \theta(t)] = V_1 \cos \left[ 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\eta) d\eta \right] \quad (5.24)$$

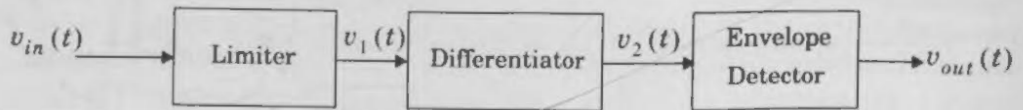


Figure 5.8  
Block diagram of a slope detector type FM demodulator.

Equation (5.24) can be differentiated in practice by passing the signal through a filter with a transfer function that has gain that increases linearly with frequency. Such a filter is called a slope filter (which is where the term *slope detector* derives its name). The output of the differentiator then becomes

$$v_2(t) = -V_1 \left[ 2\pi f_c t + \frac{d\theta}{dt} \right] \sin(2\pi f_c t + \theta(t)) \quad (5.25)$$

and the output of the envelope detector becomes

$$\begin{aligned} v_{out}(t) &= V_1 \left[ 2\pi f_c + \frac{d}{dt} \theta(t) \right] \\ &= V_1 2\pi f_c + V_1 2\pi k_f m(t) \end{aligned} \quad (5.26)$$

The above equation shows that the output of the envelope detector contains a dc term proportional to the carrier frequency and a time-varying term proportional to the original message signal  $m(t)$ . The dc term can be filtered out using a capacitor to obtain the desired demodulated signal.

### Zero-crossing Detector

When linearity is required over a broad range of frequencies, such as for data communications, a zero-crossing detector is used to perform frequency-to-amplitude conversion by directly counting the number of zero crossings in the input FM signal. The rationale behind this technique is to use the output of the zero-crossing detector to generate a pulse train with an average value that is



proportional to the frequency of the input signal. This demodulator is sometimes referred to as a *pulse-averaging discriminator*. A block diagram of a pulse-averaging discriminator is shown in Figure 5.9. The input FM signal is first passed through a limiter circuit which converts the input signal to a frequency modulated pulse train. This pulse train  $v_1(t)$  is then passed through a differentiator whose output is used to trigger a monostable multivibrator (also called a "one-shot"). The output of the one-shot consists of a train of pulses with average duration proportional to the desired message signal. A low pass filter is used to perform the averaging operation by extracting the slowly varying dc component of the signal at the output of the one-shot. The output of the low pass filter is the desired demodulated signal.

#### **PLL for FM Detection**

The phase locked loop (PLL) method is another popular technique to demodulate an FM signal. The PLL is a closed loop control system which can track the variations in the received signal phase and frequency. A block diagram of a PLL circuit is shown in Figure 5.10. It consists of a voltage controlled oscillator  $H(s)$  with an output frequency which is varied in accordance with the demodulated output voltage level. The output of the voltage controlled oscillator is compared with the input signal using a phase comparator, which produces an output voltage proportional to the phase difference. The phase difference signal is then fed back to the VCO to control the output frequency. The feedback loop functions in a manner that facilitates locking of the VCO frequency to the input frequency. Once the VCO frequency is locked to the input frequency, the VCO continues to track the variations in the input frequency. Once this tracking is achieved, the control voltage to the VCO is simply the demodulated FM signal.

#### **Quadrature Detection**

Quadrature detection is one of the more popular detection techniques used in the demodulation of frequency modulated signals. This technique can be easily implemented on an integrated circuit at a very low cost. The detector consists of a network which shifts the phase of the incoming FM signal by an amount proportional to its instantaneous frequency, and uses a product detector (phase detector) to detect the phase difference between the original FM signal and the signal at the output of the phase-shift network. Since the phase shift introduced by the phase-shift network is proportional to the instantaneous frequency of the FM signal, the output voltage of the phase detector will also be proportional to the instantaneous frequency of the input FM signal. In this manner, a frequency-to-amplitude conversion is achieved, and the FM signal is demodulated.

To achieve optimum performance from a quadrature detector, a very small (no more than  $\pm 5$  degree) phase shift should be introduced across the modulated signal bandwidth. The phase-shift network should have a constant amplitude response and a linear phase response over the spectrum occupied by the FM

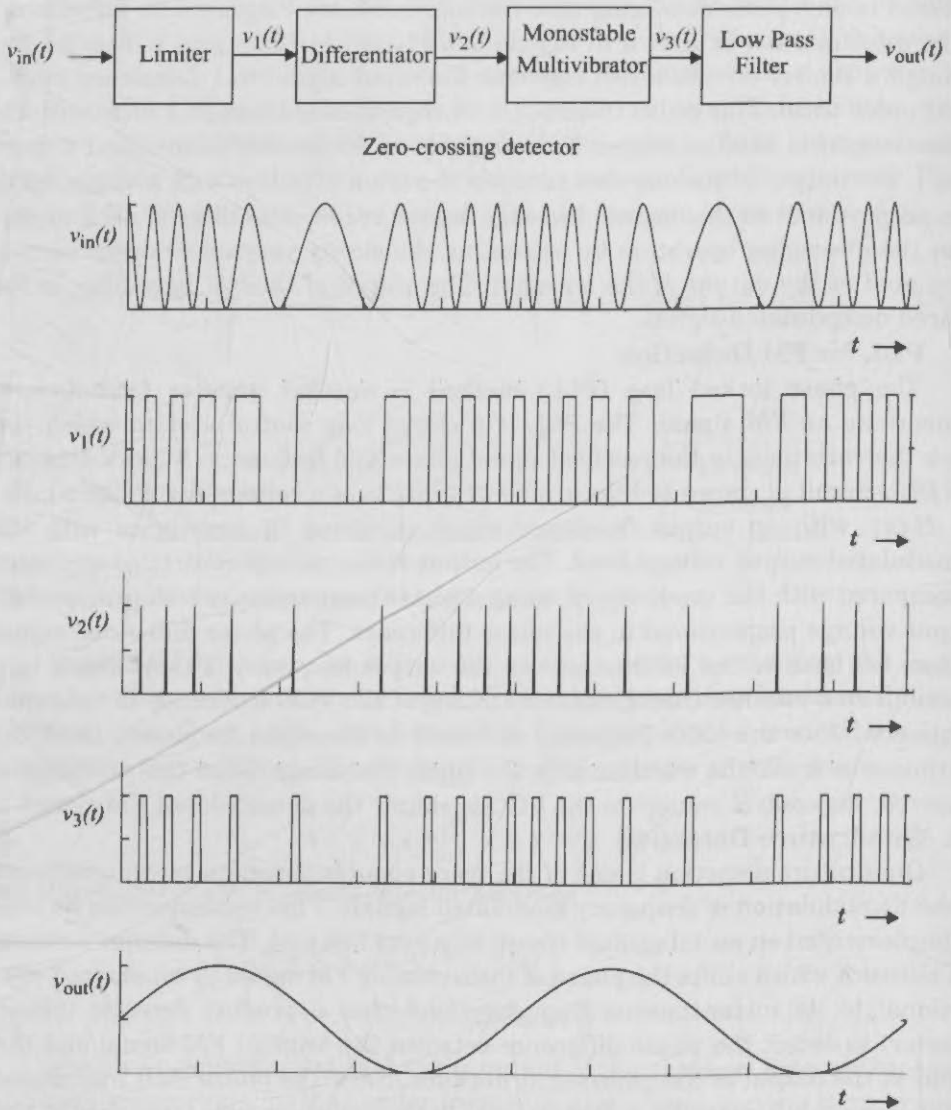


Figure 5.9

Block diagram of a zero-crossing detector and associated waveforms

signal, as shown in Figure 5.11. Further, the network should have a nominal  $90^\circ$  phase shift at the carrier frequency.

Figure 5.12 shows a block diagram of a quadrature detector. The following analysis shows that this circuit functions as an FM demodulator. The phase response function of the phase-shift network can be expressed as

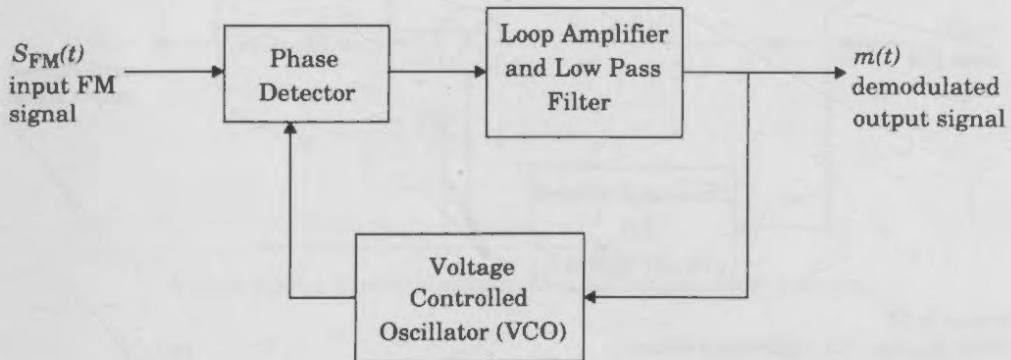


Figure 5.10  
Block diagram of a PLL used as a frequency demodulator.

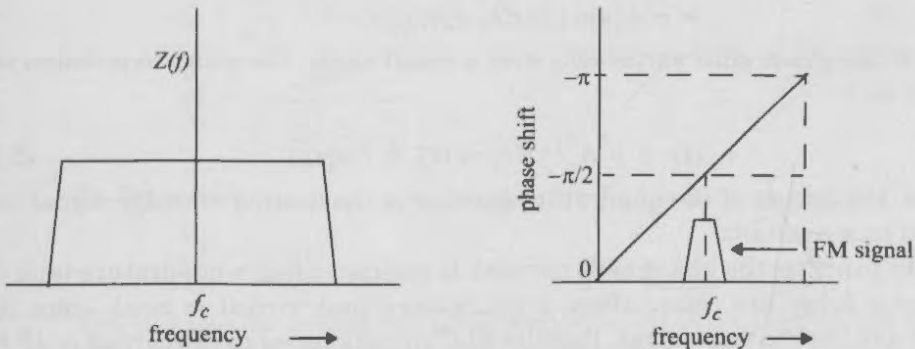


Figure 5.11  
Characteristics of the phase-shift network with constant gain and linear phase.

$$\phi(f) = -\frac{\pi}{2} + 2\pi K(f - f_c) \tag{5.27}$$

where  $K$  is a proportionality constant. When an FM signal (see equation (5.16)) is passed through the phase-shift network, the output can be expressed as

$$v_\phi(t) = \rho A_c \cos \left[ 2\pi f_c t + 2\pi k_f \int m(\eta) d\eta + \phi(f_i(t)) \right] \tag{5.28}$$

where  $\rho$  is a constant, and  $f_i(t)$  is the instantaneous frequency of the input FM signal, which is defined as

$$f_i(t) = f_c + k_f m(t) \tag{5.29}$$

The output of the product detector is proportional to the cosine of the phase difference between  $v_\phi(t)$  and  $S_{FM}(t)$  (see Figure 5.12), and is given by

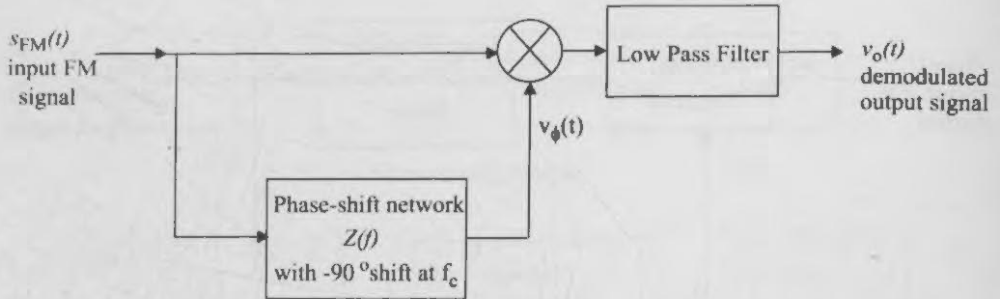


Figure 5.12  
Block diagram of a quadrature detector.

$$\begin{aligned}
 v_o(t) &= \rho^2 A_c^2 \cos(\phi(f_i(t))) \\
 &= \rho^2 A_c^2 \cos(-\pi/2 + 2\pi K[f_i(t) - f_c]) \\
 &= \rho^2 A_c^2 \sin[2\pi K k_f m(t)]
 \end{aligned} \tag{5.30}$$

If the phase shift varies only over a small angle, the above expression simplifies to

$$v_o(t) = \rho^2 A_c^2 2\pi K k_f m(t) = C m(t) \tag{5.31}$$

Hence, the output of the quadrature detector is the desired message signal multiplied by a constant.

In practice, the phase-shift network is realized using a quadrature tank circuit or a delay line. More often, a quadrature tank circuit is used, since it is cheap and easy to implement. Parallel RLC circuits tuned to the carrier or IF frequency can be used to build the quadrature tank circuit.

#### Example 5.4

Design an RLC network that implements an IF quadrature FM detector with  $f_c = 10.7$  MHz, and a 500 kHz symmetrical bandpass spectrum. Also plot the transfer function of the designed network to verify that it will work.

#### Solution to Example 5.4

A quadrature detector is represented by the block diagram in Figure 5.12, and the phase-shift network is implemented by the RLC circuit shown in Figure E5.4.1. Here the phase shift is  $90^\circ$  instead of  $-90^\circ$  for  $f = f_c$ . With reference to Figure E5.4.1

$$\frac{V_g(\omega)}{V_f(\omega)} = \frac{Z_1(\omega)}{Z_1(\omega) + Z_2(\omega)} \tag{E5.4.1}$$

Multiplying and dividing by  $1/(Z_1 Z_2)$ , we get

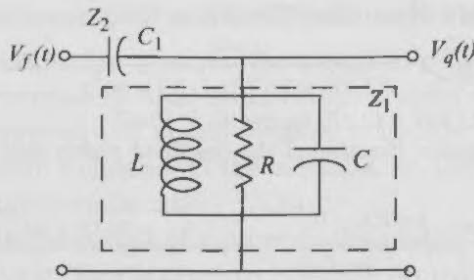


Figure E5.4.1 Circuit diagram of an RLC phase-shift network.

$$\frac{V_g(\omega)}{V_f(\omega)} = \frac{Y_2}{Y_1 + Y_2} = \frac{j\omega C_1}{j\omega C + \frac{1}{R} + \frac{1}{j\omega L} + j\omega C_1} = \frac{j\omega R C_1}{1 + jR\left(\omega(C + C_1) - \frac{1}{\omega L}\right)} \quad (\text{E5.4.2})$$

Let  $\omega_c^2 = 1/(L(C_1 + C))$  for the overall circuit. Then

$$Q = \frac{R}{\omega_c L} = R\omega_c(C_1 + C) \quad (\text{E5.4.3})$$

$$\frac{V_g}{V_f} = \frac{j\omega R C_1}{1 + jQ\left(\frac{\omega}{\omega_c} - \frac{\omega_c}{\omega}\right)}$$

So, for  $\omega = \omega_c$ :

$$\frac{V_g}{V_f} = j\omega_c R C_1$$

This provides the desired  $90^\circ$  phase shift at  $\omega_c$ . At IF frequencies, the phase shift introduced by the network may be expressed as

$$\phi(\omega_i) = \frac{\pi}{2} + \tan^{-1}\left[Q\left(\frac{\omega_i}{\omega_c} - \frac{\omega_c}{\omega_i}\right)\right] = 90^\circ + \eta$$

For a good system we need  $-5^\circ < \phi(\omega_i) < 5^\circ$  (approximately).

Therefore, for  $f_c = 10.7$  MHz and  $B = 500$  kHz, at the largest IF frequency  $f_i = f_c + 250$  kHz. Thus, we require

$$Q\left(\frac{10.7 \times 10^6 + 250 \times 10^3}{10.7 \times 10^6} - \frac{10.7 \times 10^6}{10.7 \times 10^6 + 250 \times 10^3}\right) = \tan 5^\circ$$

Therefore,  $Q = 1.894$ .

Using  $Q = 1.894$ , one may verify the phase shift at the smallest IF frequency  $f_i = f_c - 250$  kHz,

$$\tan^{-1}\left[1.894\left(\frac{10.45}{10.7} - \frac{10.7}{10.45}\right)\right] = -5.12^\circ \approx -5^\circ$$

We have verified that a circuit with  $Q = 1.894$  will satisfy the phase shift requirements.

Now, to compute the values of  $L$ ,  $R$ ,  $C$ , and  $C_1$ .

Choose  $L = 10 \mu\text{H}$ . Using the first part of equation (E5.4.3), the value of  $R$  can be computed as  $1.273 \text{ k}\Omega$

Using the second part of equation (E5.4.3)

$$C_1 + C = \frac{Q}{R\omega_c} = \frac{1.894}{(1.273 \times 10^3) 2\pi (10.7 \times 10^6)} = 22.13 \text{ pF.}$$

Assuming  $C_1 = 12.13 \text{ pF} \cong 12 \text{ pF}$ , we get  $C = 10 \text{ pF}$ .

The magnitude transfer function of the designed phase shift network is given by

$$|H(f)| = \frac{2\pi fRC_1}{\sqrt{1 + Q^2 \left(\frac{f}{f_c} - \frac{f_c}{f}\right)^2}} = \frac{97.02 \times 10^{-9} f}{\sqrt{1 + 3.587 \left(\frac{f}{10.7 \times 10^6} - \frac{10.7 \times 10^6}{f}\right)^2}}$$

and the phase transfer function is given by

$$\angle H(f) = \frac{\pi}{2} + \tan^{-1} \left[ Q \left( \frac{f}{f_c} - \frac{f_c}{f} \right) \right] = \frac{\pi}{2} + \tan \left[ 1.894 \left( \frac{f}{10.7 \times 10^6} - \frac{10.7 \times 10^6}{f} \right) \right]$$

Both the magnitude and phase transfer functions are plotted in Figure E5.4.2. It is clearly seen from the plots that the transfer function satisfies the requirements of the phase-shift network, thus allowing detection of FM.

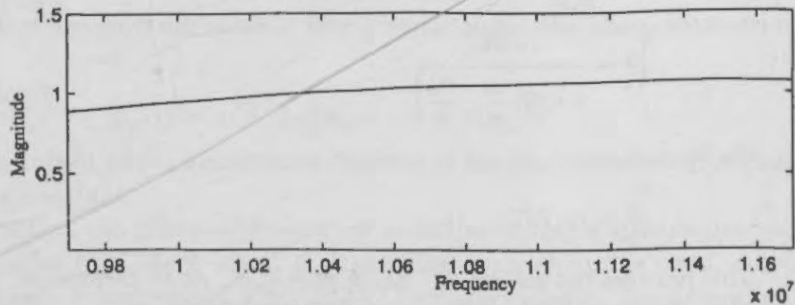


Figure E5.4.2 (a) Magnitude response of the designed phase-shift network.

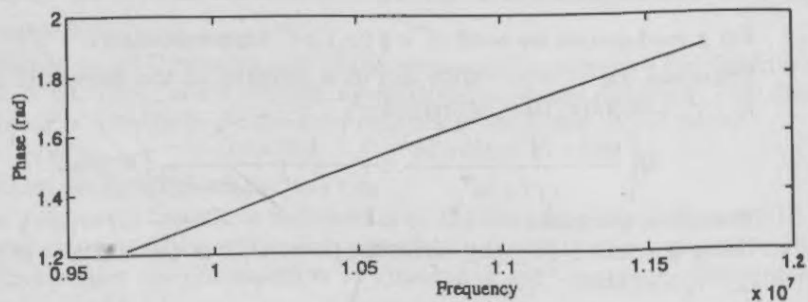


Figure E5.4.2 (b) Phase response of the designed phase-shift network.

### 5.3.4 Tradeoff Between SNR and Bandwidth in an FM Signal

In angle modulation systems, the signal-to-noise ratio *before* detection is a function of the receiver IF filter bandwidth, received carrier power, and received interference. However, the signal-to-noise ratio *after* detection is a function of  $f_{max}$ , the maximum frequency of the message,  $\beta_f$ , the modulation index, and the given input signal-to-noise ratio  $(SNR)_{in}$ .

The SNR at the output of a properly designed FM receiver is dependent on the modulation index and is given by [Cou93]

$$(SNR)_{out} = 6(\beta_f + 1)\beta_f^2 \left( \frac{m(t)}{V_p} \right)^2 (SNR)_{in} \quad (5.32)$$

where  $V_p$  is the peak-to-zero value of the modulating signal  $m(t)$ , and the input signal-to-noise ratio  $(SNR)_{in}$  is given by

$$(SNR)_{in} = \frac{A_c^2/2}{2N_0(\beta_f + 1)B} \quad (5.33)$$

where  $A_c$  is the carrier amplitude,  $N_0$  is the white noise RF power spectral density, and  $B$  is the equivalent RF bandwidth of the bandpass filter at the front end of the receiver. Note that  $(SNR)_{in}$  uses the RF signal bandwidth given by Carson's rule in equation (5.21). For comparison purposes, let  $(SNR)_{in;AM}$  be defined as the input power to a conventional AM receiver having RF bandwidth equal to  $2B$ . That is,

$$SNR_{in;AM} = \frac{A_c^2}{2N_0B} \quad (5.34)$$

Then, for  $m(t) = A_m \sin \omega_m t$ , equation (5.32) can be simplified to

$$[SNR]_{out} = 3\beta_f^2(\beta_f + 1)(SNR)_{in} = 3\beta_f^2(SNR)_{in;AM} \quad (5.35)$$

The above expression for  $(SNR)_{out}$  is valid only if  $(SNR)_{in}$  exceeds the threshold of the FM detector. The minimum received value of  $(SNR)_{in}$  needed to exceed the threshold is typically around 10 dB. When  $(SNR)_{in}$  falls below the threshold, the demodulated signal becomes noisy. In FM mobile radio systems, it is not uncommon to hear *click noise* as a received signal rises and falls about the threshold. Equation (5.35) shows that the SNR at the output of the FM detector can be increased by increasing the modulation index  $\beta_f$  of the transmitted signal. In other words, it is possible to obtain an FM detection gain at the receiver by increasing the modulation index of the FM signal. The increase in modulation index, however, leads to an increased bandwidth and spectral occupancy. For large values of  $\beta_f$ , Carson's rule estimates the channel bandwidth as  $2\beta_f f_{max}$ . As shown in the right hand side of equation (5.35), the SNR at the output of an FM detector is  $3\beta_f^2$  greater than the input SNR for an AM signal with

the same RF bandwidth. Since AM detectors have a linear detection gain, it follows that  $(SNR)_{out}$  for FM is much greater than  $(SNR)_{out}$  for AM.

Equation (5.35) shows the SNR at the output of an FM detector increases as the cube of the bandwidth of the message. This clearly illustrates why FM offers excellent performance for fading signals. As long as  $(SNR)_{in}$  remains above threshold,  $(SNR)_{out}$  is much greater than  $(SNR)_{in}$ . A technique called *threshold extension* is used in FM demodulators to improve detection sensitivity to about  $(SNR)_{in} = 6$  dB.

FM can improve receiver performance through adjustment of the modulation index at the transmitter, and not the transmitted power. This is not the case in AM, since linear modulation techniques do not trade bandwidth for SNR.

### Example 5.5

How much bandwidth is required for an analog frequency modulated signal that has an audio bandwidth of 5 kHz and a modulation index of 3? How much output SNR improvement would be obtained if the modulation index is increased to 5? What is the trade-off bandwidth for this improvement?

### Solution to Example 5.5

From Carson's rule the bandwidth is

$$B_T = 2(\beta_f + 1)f_m = 2(3 + 1)5 \text{ kHz} = 40 \text{ kHz}$$

From equation (5.35) the output SNR improvement factor is approximately  $3\beta_f^3 + 3\beta_f^2$ .

Therefore

$$\text{for } \beta_f = 3, \text{ the output SNR factor is } \approx 3(3)^3 + 3(3)^2 = 108 = 20.33 \text{ dB}$$

$$\text{for } \beta_f = 5, \text{ the output SNR factor is } \approx 3(5)^3 + 3(5)^2 = 450 = 26.53 \text{ dB}$$

The improvement in output SNR by increasing the modulation index from 3 to 5 is therefore  $26.53 - 20.33 = 6.2$  dB

This improvement is achieved at the expense of bandwidth. For  $\beta_f = 3$ , a bandwidth of 40 kHz is needed, while  $\beta_f = 5$  requires a bandwidth of 60 kHz.

## 5.4 Digital Modulation — an Overview

Modern mobile communication systems use digital modulation techniques. Advancements in very large-scale integration (VLSI) and digital signal processing (DSP) technology have made digital modulation more cost effective than analog transmission systems. Digital modulation offers many advantages over analog modulation. Some advantages include greater noise immunity and robustness to channel impairments, easier multiplexing of various forms of information (e.g., voice, data, and video), and greater security. Furthermore, digital transmissions accommodate digital error-control codes which detect and/or correct transmission errors, and support complex signal conditioning and processing techniques such as source coding, encryption, and equalization to



improve the performance of the overall communication link. New multipurpose programmable digital signal processors have made it possible to implement digital modulators and demodulators completely in software. Instead of having a particular modem design permanently frozen as hardware, embedded software implementations now allow alterations and improvements without having to redesign or replace the modem.

In digital wireless communication systems, the modulating signal (e.g., the message) may be represented as a time sequence of symbols or pulses, where each symbol has  $m$  finite states. Each symbol represents  $n$  bits of information, where  $n = \log_2 m$  bits/symbol. Many digital modulation schemes are used in modern wireless communication systems, and many more are sure to be introduced. Some of these techniques have subtle differences between one another, and each technique belongs to a family of related modulation methods. For example, phase shift keying (PSK) may be either coherently or differentially detected; and may have two, four, eight or more possible levels (e.g.,  $n = 1, 2, 3$ , or more bits) per symbol, depending on the manner in which information is transmitted within a single symbol.

#### 5.4.1 Factors That Influence the Choice of Digital Modulation

Several factors influence the choice of a digital modulation scheme. A desirable modulation scheme provides low bit error rates at low received signal-to-noise ratios, performs well in multipath and fading conditions, occupies a minimum of bandwidth, and is easy and cost-effective to implement. Existing modulation schemes do not simultaneously satisfy all of these requirements. Some modulation schemes are better in terms of the bit error rate performance, while others are better in terms of bandwidth efficiency. Depending on the demands of the particular application, trade-offs are made when selecting a digital modulation.

The performance of a modulation scheme is often measured in terms of its *power efficiency* and *bandwidth efficiency*. Power efficiency describes the ability of a modulation technique to preserve the fidelity of the digital message at low power levels. In a digital communication system, in order to increase noise immunity, it is necessary to increase the signal power. However, the amount by which the signal power should be increased to obtain a certain level of fidelity (i.e., an acceptable bit error probability) depends on the particular type of modulation employed. The *power efficiency*,  $\eta_p$  (sometimes called energy efficiency) of a digital modulation scheme is a measure of how favorably this tradeoff between fidelity and signal power is made, and is often expressed as the ratio of the *signal energy per bit to noise power spectral density* ( $E_b/N_0$ ) required at the receiver input for a certain probability of error (say  $10^{-5}$ ).

*Bandwidth efficiency* describes the ability of a modulation scheme to accommodate data within a limited bandwidth. In general, increasing the data rate

implies decreasing the pulse width of a digital symbol, which increases the bandwidth of the signal. Thus, there is an unavoidable relationship between data rate and bandwidth occupancy. However, some modulation schemes perform better than the others in making this tradeoff. Bandwidth efficiency reflects how efficiently the allocated bandwidth is utilized and is defined as the ratio of the *throughput data rate per Hertz* in a given bandwidth. If  $R$  is the data rate in bits per second, and  $B$  is the bandwidth occupied by the modulated RF signal, then bandwidth efficiency  $\eta_B$  is expressed as

$$\eta_B = \frac{R}{B} \text{ bps/Hz} \quad (5.36)$$

The system capacity of a digital mobile communication system is directly related to the bandwidth efficiency of the modulation scheme, since a modulation with a greater value of  $\eta_B$  will transmit more data in a given spectrum allocation.

There is a fundamental upper bound on achievable bandwidth efficiency. Shannon's channel coding theorem states that for an arbitrarily small probability of error, the maximum possible bandwidth efficiency is limited by the noise in the channel, and is given by the channel capacity formula [Sha48]

$$\eta_{Bmax} = \frac{C}{B} = \log_2 \left( 1 + \frac{S}{N} \right) \quad (5.37)$$

where  $C$  is the channel capacity (in bps),  $B$  is the RF bandwidth, and  $S/N$  is the signal-to-noise ratio.

In the design of a digital communication system, very often there is a tradeoff between bandwidth efficiency and power efficiency. For example, as shown in Chapter 6, adding error control coding to a message increases the bandwidth occupancy (and this, in turn, reduces the bandwidth efficiency), but at the same time reduces the required received power for a particular bit error rate, and hence trades bandwidth efficiency for power efficiency. On the other hand, higher level modulation schemes ( $M$ -ary keying) decrease bandwidth occupancy but increase the required received power, and hence trade power efficiency for bandwidth efficiency.

While power and bandwidth efficiency considerations are very important, other factors also affect the choice of a digital modulation scheme. For example, for all personal communication systems which serve a large user community, the cost and complexity of the subscriber receiver must be minimized, and a modulation which is simple to detect is most attractive. The performance of the modulation scheme under various types of channel impairments such as Rayleigh and Rician fading and multipath time dispersion, given a particular demodulator implementation, is another key factor in selecting a modulation. In cellular systems where interference is a major issue, the performance of a modulation scheme in an interference environment is extremely important. Sensitivity to

detection of timing jitter, caused by time-varying channels, is also an important consideration in choosing a particular modulation scheme. In general, the modulation, interference, and implementation of the time-varying effects of the channel as well as the performance of the specific demodulator are analyzed as a complete system using simulation to determine relative performance and ultimate selection.

### Example 5.6

If the SNR of a wireless communication link is 20 dB and the RF bandwidth is 30 kHz, determine the maximum theoretical data rate that can be transmitted. Compare this rate to the U.S. Digital Cellular Standard described in Chapter 1.

### Solution to Example 5.6

Given:

$$S/N = 20 \text{ dB} = 100$$

$$\text{RF Bandwidth } B = 30000 \text{ Hz}$$

Using Shannon's channel capacity formula (5.37), the maximum possible data rate

$$C = B \log_2 \left( 1 + \frac{S}{N} \right) = 30000 \log_2 (1 + 100) = 199.75 \text{ kbps}$$

The USDC data rate is 48.6 kbps, which is only about one fourth the theoretical limit under 20 dB SNR conditions.

### Example 5.7

What is the theoretical maximum data rate that can be supported in a 200 kHz channel for  $SNR = 10 \text{ dB}$ ,  $30 \text{ dB}$ . How does this compare to the GSM standard described in Chapter 1?

### Solution to Example 5.7

For  $SNR = 10 \text{ dB} = 10$ ,  $B = 200 \text{ kHz}$ .

Using Shannon's channel capacity formula (5.37), the maximum possible data rate

$$C = B \log_2 \left( 1 + \frac{S}{N} \right) = 200000 \log_2 (1 + 10) = 691.886 \text{ kbps}$$

The GSM data rate is 270.833 kbps, which is only about 40% of the theoretical limit for 10 dB SNR conditions.

For  $SNR = 30 \text{ dB} = 1000$ ,  $B = 200 \text{ kHz}$ .

The maximum possible data rate

$$C = B \log_2 \left( 1 + \frac{S}{N} \right) = 200000 \log_2 (1 + 1000) = 1.99 \text{ Mbps.}$$

### 5.4.2 Bandwidth and Power Spectral Density of Digital Signals

The definition of signal *bandwidth* varies with context, and there is no single definition which suits all applications [Amo80]. All definitions, however, are based on some measure on the *power spectral density* (PSD) of the signal. The power spectral density of a random signal  $w(t)$  is defined as [Cou93]

$$P_w(f) = \lim_{T \rightarrow \infty} \left( \frac{\overline{|W_T(f)|^2}}{T} \right) \quad (5.38)$$

where the bar denotes an ensemble average, and  $W_T(f)$  is the Fourier transform of  $w_T(t)$ , which is the truncated version of the signal  $w(t)$ , defined as

$$w_T(t) = \begin{cases} w(t) & -T/2 < t < T/2 \\ 0 & \text{elsewhere} \end{cases} \quad (5.39)$$

The power spectral density of a modulated (bandpass) signal is related to the power spectral density of its baseband complex envelope. If a bandpass signal  $s(t)$  is represented as

$$s(t) = \text{Re} \{ g(t) \exp(j2\pi f_c t) \} \quad (5.40)$$

where  $g(t)$  is the complex baseband envelope, then the PSD of the bandpass signal is given by

$$P_s(f) = \frac{1}{4} [P_g(f-f_c) + P_g(-f-f_c)] \quad (5.41)$$

where  $P_g(f)$  is the PSD of  $g(t)$ .

The *absolute bandwidth* of a signal is defined as the range of frequencies over which the signal has a non-zero power spectral density. For symbols represented as rectangular baseband pulses, the PSD has a  $(\sin f)^2/f^2$  profile which extends over an infinite range of frequencies, and has an absolute bandwidth of infinity. A simpler and more widely accepted measure of bandwidth is the first *null-to-null bandwidth*. The null-to-null bandwidth is equal to the width of the main spectral lobe.

A very popular measure of bandwidth which measures the dispersion of the spectrum is the *half-power bandwidth*. The half-power bandwidth is defined as the interval between frequencies at which the PSD has dropped to half power, or 3 dB below the peak value. Half-power bandwidth is also called the *3 dB bandwidth*.

The definition adopted by the Federal Communications Commission (FCC) defines occupied bandwidth as the band which leaves exactly 0.5 percent of the signal above the upper band limit and exactly 0.5 percent of the signal power below the lower band limit. In other words, 99 percent of the signal power is contained within the occupied bandwidth.

Another commonly used method to specify bandwidth is to state that everywhere outside the specified band, the PSD is below a certain stated level. Typically, 45 dB to 60 dB attenuation is specified.

### 5.4.3 Line Coding

Digital baseband signals often use line codes to provide particular spectral characteristics of a pulse train. The most common codes for mobile communication are *return-to-zero* (RZ), *non-return-to-zero* (NRZ), and *Manchester* codes (see Figure 5.13 and Figure 5.14). All of these may either be unipolar (with voltage levels being either 0 or  $V$ ) or bipolar (with voltage levels being either  $-V$  or  $V$ ). RZ implies that the pulse returns to zero within every bit period. This leads to spectral widening, but improves timing synchronization. NRZ codes, on the other hand, do not return to zero during a bit period — the signal stays at constant levels throughout a bit period. NRZ codes are more spectrally efficient than RZ codes, but offer poorer synchronization capabilities. Because of the large dc component, NRZ line codes are used for data that does not have to be passed through dc blocking circuits such as audio amplifiers or phone lines.

The Manchester code is a special type of NRZ line code that is ideally suited for signaling that must pass through phone lines and other dc blocking circuits, as it has no dc component and offers simple synchronization. Manchester codes use two pulses to represent each binary symbol, and thereby provide easy clock recovery since zero-crossings are guaranteed in every bit period. The power spectral density of these line codes are shown in Figure 5.13 and the time waveforms are given in Figure 5.14.

## 5.5 Pulse Shaping Techniques

When rectangular pulses are passed through a bandlimited channel, the pulses will spread in time, and the pulse for each symbol will smear into the time intervals of succeeding symbols. This causes *intersymbol interference* (ISI) and leads to an increased probability of the receiver making an error in detecting a symbol. One obvious way to minimize intersymbol interference is to increase the channel bandwidth. However, mobile communication systems operate with minimal bandwidth, and techniques that reduce the modulation bandwidth and suppress out-of-band radiation, while reducing intersymbol interference, are highly desirable. Out-of-band radiation in the adjacent channel in a mobile radio system should generally be 40 dB to 80 dB below that in the desired passband. Since it is difficult to directly manipulate the transmitter spectrum at RF frequencies, spectral shaping is done through baseband or IF processing. There are a number of well known pulse shaping techniques which are used to simultaneously reduce the intersymbol effects and the spectral width of a modulated digital signal.

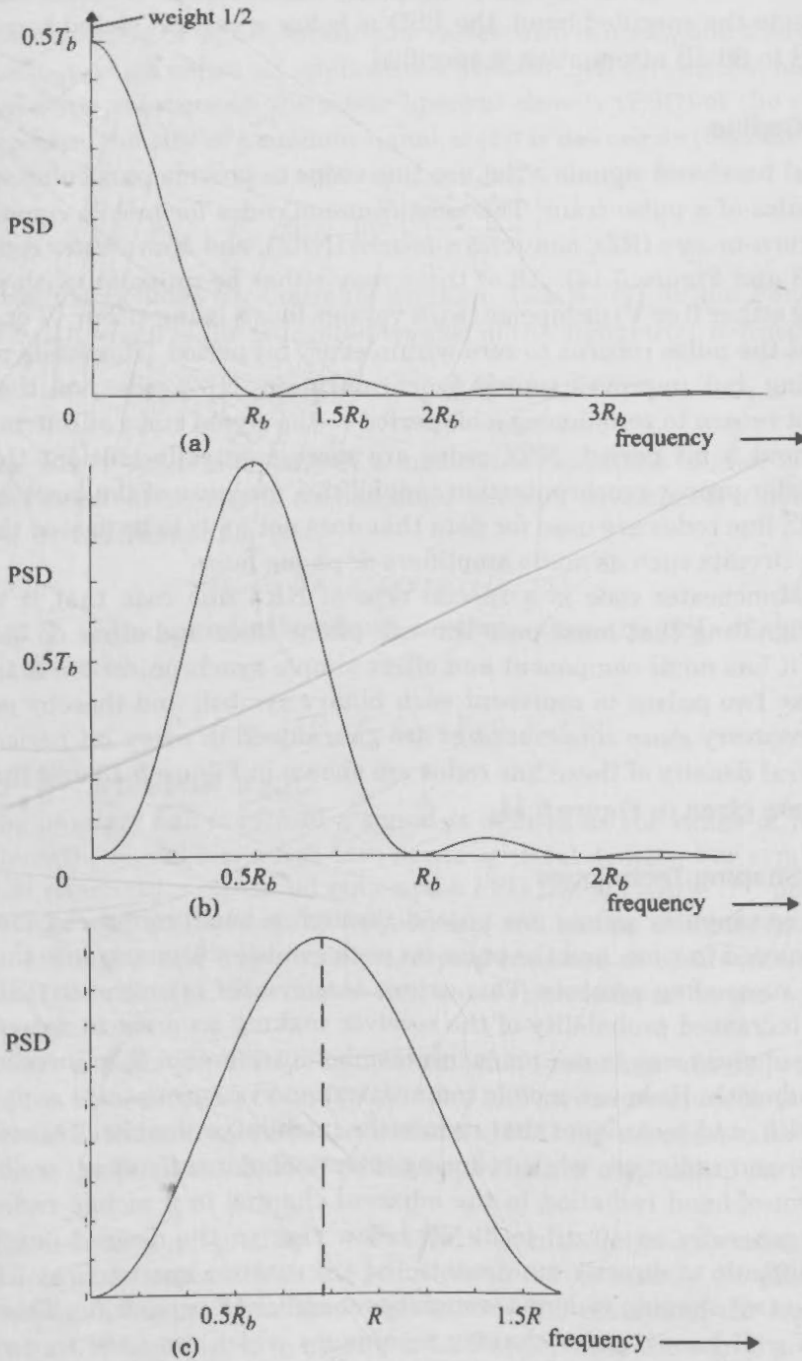


Figure 5.13  
Power spectral density of (a) unipolar NRZ, (b) unipolar RZ, and (c) Manchester line codes.

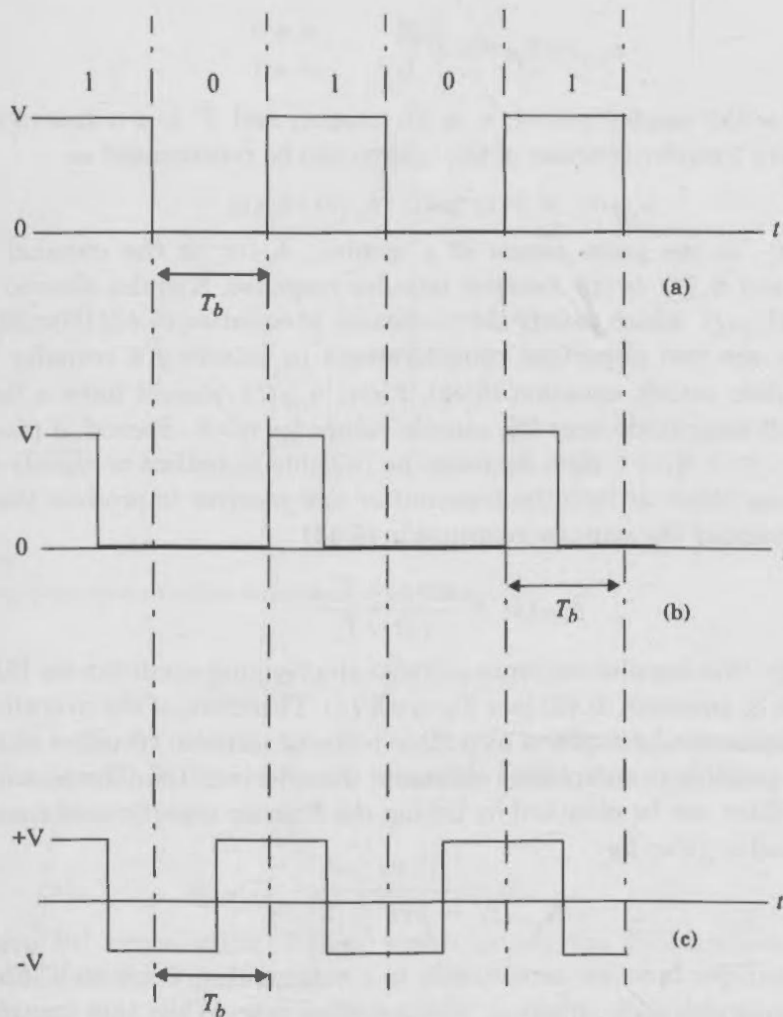


Figure 5.14

Time waveforms of binary line codes (a) Unipolar NRZ (b) Unipolar RZ (c) Manchester NRZ

### 5.5.1 Nyquist Criterion for ISI Cancellation

Nyquist was the first to solve the problem of overcoming intersymbol interference while keeping the transmission bandwidth low [Nyq28]. He observed that the effect of ISI could be completely nullified if the overall response of the communication system (including transmitter, channel, and receiver) is designed so that at every sampling instant at the receiver, the response due to all symbols except the current symbol is equal to zero. If  $h_{eff}(t)$  is the impulse response of the overall communication system, this condition can be mathematically stated as

$$h_{eff}(nT_s) = \begin{cases} K & n = 0 \\ 0 & n \neq 0 \end{cases} \quad (5.42)$$

where  $T_s$  is the symbol period,  $n$  is an integer, and  $K$  is a non-zero constant. The effective transfer function of the system can be represented as

$$h_{eff}(t) = \delta(t) * p(t) * h_c(t) * h_r(t) \quad (5.43)$$

where  $p(t)$  is the pulse shape of a symbol,  $h_c(t)$  is the channel impulse response, and  $h_r(t)$  is the receiver impulse response. Nyquist derived transfer functions  $H_{eff}(f)$  which satisfy the conditions of equation (5.42) [Nyq28].

There are two important considerations in selecting a transfer function  $H_{eff}(f)$  which satisfy equation (5.42). First,  $h_{eff}(t)$  should have a fast decay with a small magnitude near the sample values for  $n \neq 0$ . Second, if the channel is ideal ( $h_c(t) = \delta(t)$ ), then it should be possible to realize or closely approximate shaping filters at both the transmitter and receiver to produce the desired  $H_{eff}(f)$ . Consider the impulse response in (5.44)

$$h_{eff}(t) = \frac{\sin(\pi t/T_s)}{(\pi t)/T_s} \quad (5.44)$$

Clearly, this impulse response satisfies the Nyquist condition for ISI cancellation given in equation (5.42) (see Figure 5.15). Therefore, if the overall communication system can be modeled as a filter with the impulse response of equation (5.44), it is possible to completely eliminate the effects of ISI. The transfer function of the filter can be obtained by taking the Fourier transform of the impulse response, and is given by

$$H_{eff}(f) = \frac{1}{f_s} \Pi\left(\frac{f}{f_s}\right) \quad (5.45)$$

This transfer function corresponds to a rectangular "brick-wall" filter with absolute bandwidth  $f_s/2$ , where  $f_s$  is the symbol rate. While this transfer function satisfies the zero ISI criterion with a minimum of bandwidth, there are practical difficulties in implementing it, since it corresponds to a noncausal system ( $h_{eff}(t)$  exists for  $t < 0$ ) and is thus difficult to approximate. Also, the  $(\sin t)/t$  pulse has a waveform slope that is  $1/t$  at each zero crossing, and is zero only at exact multiples of  $T_s$ , thus any error in the sampling time of zero-crossings will cause significant ISI due to overlapping from adjacent symbols (A slope of  $1/t^2$  or  $1/t^3$  is more desirable to minimize the ISI due to timing jitter in adjacent samples).

Nyquist also proved that any filter with a transfer function having a rectangular filter of bandwidth  $f_0 \geq 1/2T_s$ , convolved with any arbitrary even function  $Z(f)$  with zero magnitude outside the passband of the rectangular filter, satisfies the zero ISI condition. Mathematically, the transfer function of the filter which satisfies the zero ISI condition can be expressed as



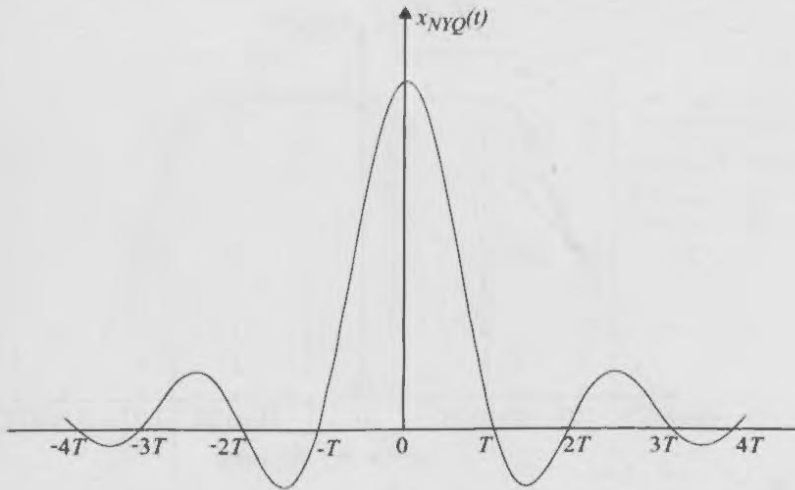


Figure 5.15  
Nyquist ideal pulse shape for zero intersymbol interference.

$$H_{eff}(f) = \Pi\left(\frac{f}{f_0}\right) \otimes Z(f) \quad (5.46)$$

where  $Z(f) = Z(-f)$ , and  $Z(f) = 0$  for  $|f| \geq f_0 \geq 1/2T_s$ . Expressed in terms of the impulse response, the Nyquist criterion states that any filter with an impulse response

$$h_{eff}(t) = \frac{\sin(\pi t/T_s)}{\pi t} z(t) \quad (5.47)$$

can achieve ISI cancellation. Filters which satisfy the Nyquist criterion are called Nyquist filters (Figure 5.16).

Assuming that the distortions introduced in the channel can be completely nullified by using an equalizer which has a transfer function that is equal to the inverse of the channel response, then the overall transfer function  $H_{eff}(f)$  can be approximated as the product of the transfer functions of the transmitter and receiver filters. An effective end-to-end transfer function of  $H_{eff}(f)$  is often achieved by using filters with transfer functions  $\sqrt{H_{eff}(f)}$  at both the transmitter and receiver. This has the advantage of providing a matched filter response for the system, while at the same time minimizing the bandwidth and intersymbol interference.

### 5.5.2 Raised Cosine Rolloff Filter

The most popular pulse shaping filter used in mobile communications is the raised cosine filter. A raised cosine filter belongs to the class of filters which sat-

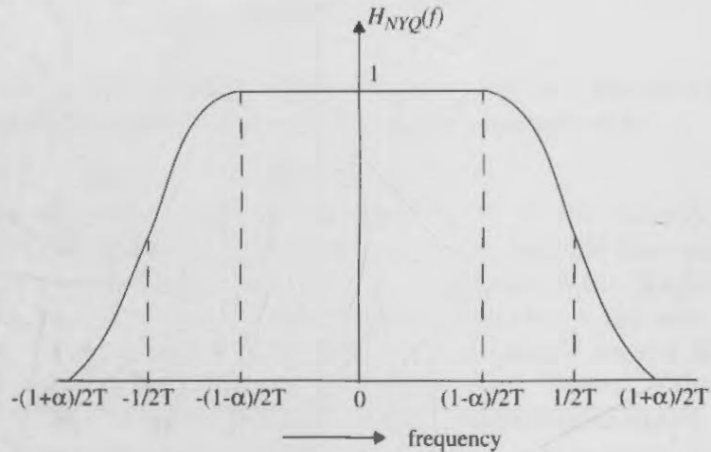


Figure 5.16  
Transfer function of a Nyquist pulse-shaping filter.

isfy the Nyquist criterion. The transfer function of a raised cosine filter is given by

$$H_{RC}(f) = \begin{cases} 1 & 0 \leq |f| \leq (1 - \alpha)/2T_s \\ \frac{1}{2} \left[ 1 + \cos \left( \frac{\pi[(2T_s|f|) - 1 + \alpha]}{2\alpha} \right) \right] & (1 - \alpha)/2T_s < |f| \leq (1 + \alpha)/2T_s \\ 0 & |f| > (1 + \alpha)/2T_s \end{cases} \quad (5.48)$$

where  $\alpha$  is the rolloff factor which ranges between 0 and 1. This transfer function is plotted in Figure 5.17 for various values of  $\alpha$ . When  $\alpha = 0$ , the raised cosine rolloff filter corresponds to a rectangular filter of minimum bandwidth. The corresponding impulse response of the filter can be obtained by taking the inverse Fourier transform of the transfer function, and is given by

$$h_{RC}(t) = \left( \frac{\sin(\pi t/T_s)}{\pi t} \right) \left( \frac{\cos(\pi \alpha t/T_s)}{1 - (4\alpha t/(2T_s))^2} \right) \quad (5.49)$$

The impulse response of the cosine rolloff filter at baseband is plotted in Figure 5.18 for various values of  $\alpha$ . Notice that the impulse response decays much faster at the zero-crossings (approximately as  $1/t^3$  for  $t \gg T_s$ ) when compared to the “brick-wall” filter ( $\alpha=0$ ). The rapid time rolloff allows it to be truncated in time with little deviation in performance from theory. As seen from Figure 5.17, as the rolloff factor  $\alpha$  increases, the bandwidth of the filter also increases, and the time sidelobe levels decrease in adjacent symbol slots. This implies that increasing  $\alpha$  decreases the sensitivity to timing jitter, but increases the occupied bandwidth.

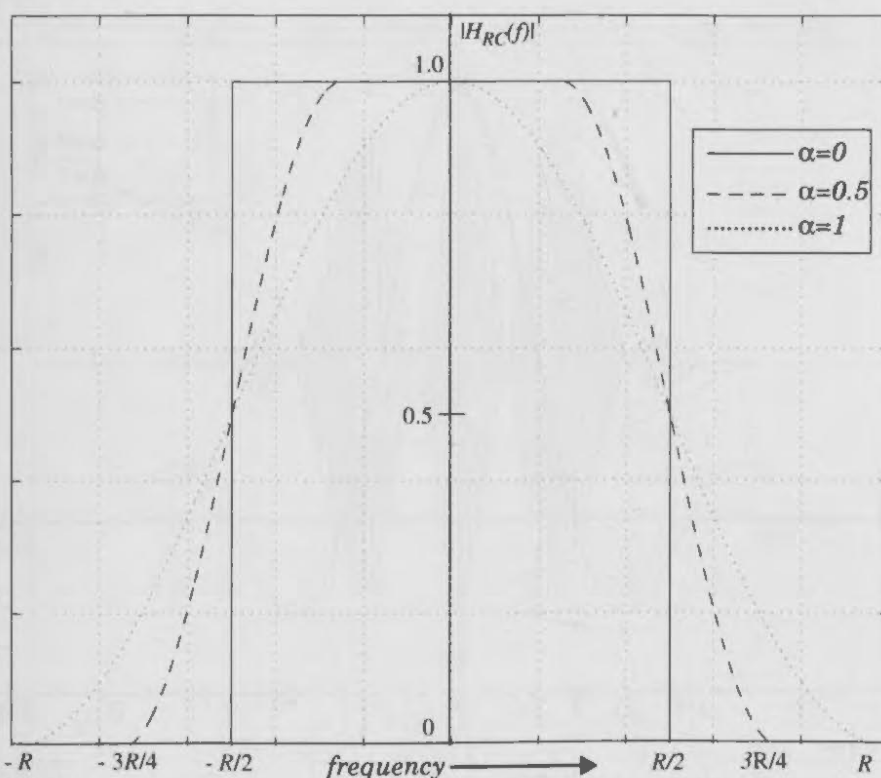


Figure 5.17  
Magnitude transfer function of a raised cosine filter.

The symbol rate  $R_s$  that can be passed through a baseband raised cosine rolloff filter is given by

$$R_s = \frac{1}{T_s} = \frac{2B}{1+\alpha} \quad (5.50)$$

where  $B$  is the absolute filter bandwidth. For RF systems, the RF passband bandwidth doubles and

$$R_s = \frac{B}{1+\alpha} \quad (5.51)$$

The cosine rolloff transfer function can be achieved by using identical  $\sqrt{H_{RC}(f)}$  filters at the transmitter and receiver, while providing a matched filter for optimum performance in a flat fading channel. To implement the filter responses, pulse shaping filters can be used either on the baseband data or at the output of the transmitter. As a rule, pulse shaping filters are implemented in DSP in baseband. Because  $h_{RC}(t)$  is noncausal, it must be truncated, and pulse shaping filters are typically implemented for  $\pm 6T_s$  about the  $t = 0$  point for each symbol. For this reason, digital communication systems which use pulse shaping often store several symbols at a time inside the modulator, and then

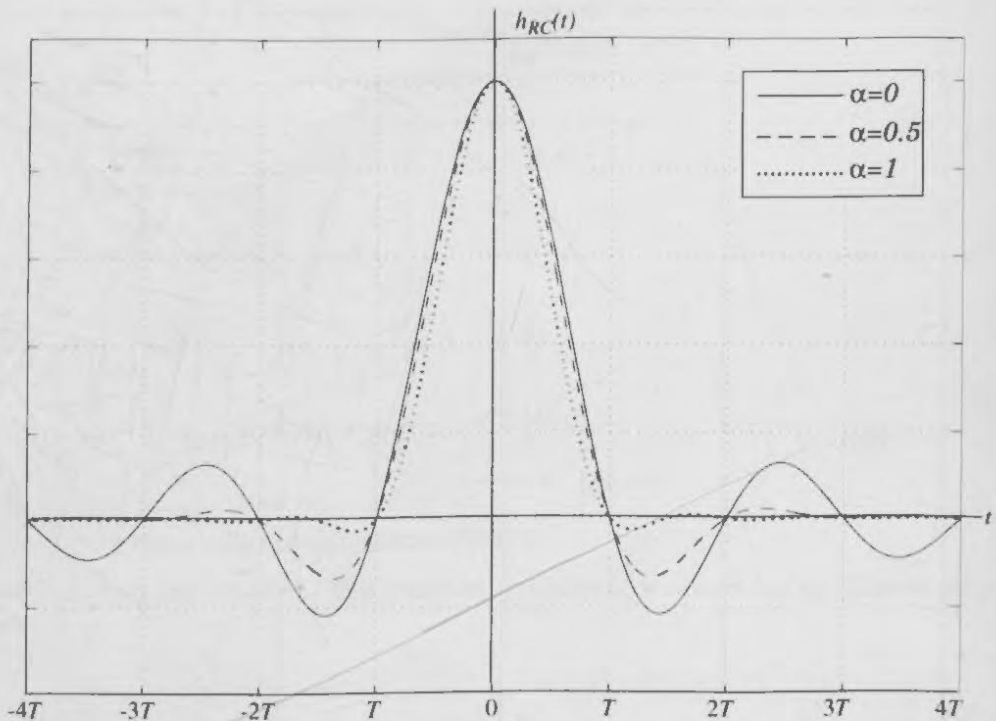


Figure 5.18  
Impulse response of a raised cosine rolloff filter.

clock out a group of symbols by using a look-up table which represents a discrete-time waveform of the stored symbols. As an example, assume binary baseband pulses are to be transmitted using a raised cosine rolloff filter with  $\alpha = 1/2$ . If the modulator stores 3 bits at a time, then there are 8 possible waveform states that may be produced at random for the group. If  $\pm 6T_s$  is used to represent the time span for each symbol (a symbol is the same as a bit in this case), then the time span of the discrete-time waveform will be  $14T_s$ . Figure 5.1 illustrates the RF time waveform for the data sequence 1, 0, 1. The optimal bit decision points occur at  $4T_s$ ,  $5T_s$ , and  $6T_s$ , and the time dispersive nature of pulse shaping can be seen.

The spectral efficiency offered by a raised cosine filter only occurs if the exact pulse shape is preserved at the carrier. This becomes difficult if nonlinear RF amplifiers are used. Small distortions in the baseband pulse shape can dramatically change the spectral occupancy of the transmitted signal. If not properly controlled, this can cause serious adjacent channel interference in mobile communication systems. A dilemma for mobile communication designers is that the reduced bandwidth offered by Nyquist pulse shaping requires linear amplifiers which are not power efficient. An obvious solution to this problem would be to develop linear amplifiers which use real-time feedback to offer more power effi-

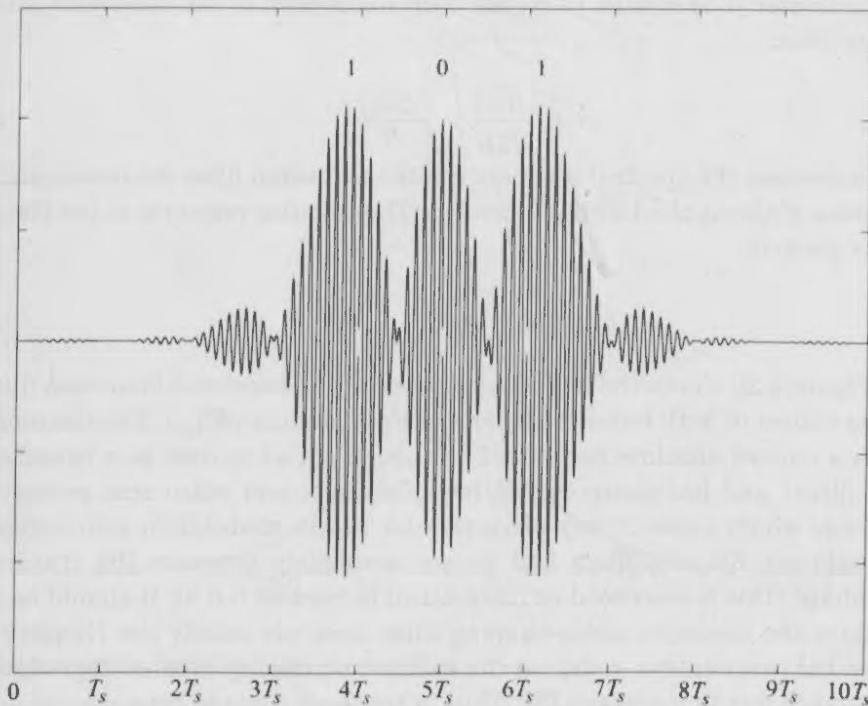


Figure 5.19

Raised cosine filtered ( $\alpha = 0.5$ ) pulses corresponding to 1, 0, 1 data stream for a BPSK signal. Notice that the decision points (at  $4T_s$ ,  $5T_s$ ,  $6T_s$ ) do not always correspond to the maximum values of the RF waveform.

ciency, and this is currently an active research thrust for mobile communications.

### 5.5.3 Gaussian Pulse-shaping Filter

It is also possible to use non-Nyquist techniques for pulse shaping. Prominent among such techniques is the use of a Gaussian pulse-shaping filter which is particularly effective when used in conjunction with Minimum Shift Keying (MSK) modulation, or other modulations which are well suited for power efficient nonlinear amplifiers. Unlike Nyquist filters which have zero-crossings at adjacent symbol peaks and a truncated transfer function, the Gaussian filter has a smooth transfer function with no zero-crossings. The impulse response of the Gaussian filter gives rise to a transfer function that is highly dependent upon the 3-dB bandwidth. The Gaussian lowpass filter has a transfer function given by

$$H_G(f) = \exp(-\alpha^2 f^2) \quad (5.52)$$

The parameter  $\alpha$  is related to  $B$ , the 3-dB bandwidth of the baseband gaussian shaping filter,

$$\alpha = \frac{\sqrt{\ln 2}}{\sqrt{2}B} = \frac{0.5887}{B} \quad (5.53)$$

As  $\alpha$  increases, the spectral occupancy of the Gaussian filter decreases and time dispersion of the applied signal increases. The impulse response of the Gaussian filter is given by

$$h_G(t) = \frac{\sqrt{\pi}}{\alpha} \exp\left(-\frac{\pi^2 t^2}{\alpha^2}\right) \quad (5.54)$$

Figure 5.20 shows the impulse response of the baseband Gaussian filter for various values of 3-dB bandwidth-symbol time product ( $BT_s$ ). The Gaussian filter has a narrow absolute bandwidth (although not as narrow as a raised cosine rolloff filter), and has sharp cut-off, low overshoot, and pulse area preservation properties which make it very attractive for use in modulation techniques that use nonlinear RF amplifiers and do not accurately preserve the transmitted pulse shape (this is discussed in more detail in Section 5.8.3). It should be noted that since the Gaussian pulse-shaping filter does not satisfy the Nyquist criterion for ISI cancellation, reducing the spectral occupancy creates degradation in performance due to increased ISI. Thus, a trade-off is made between the desired RF bandwidth and the irreducible error due to ISI of adjacent symbols when Gaussian pulse shaping is used. Gaussian pulses are used when cost is a major factor and the bit error rates due to ISI are deemed to be lower than what is nominally required.

### Example 5.8

Find the first zero-crossing RF bandwidth of a rectangular pulse which has  $T_s = 41.06 \mu\text{s}$ . Compare this to the bandwidth of a raised cosine filter pulse with  $T_s = 41.06 \mu\text{s}$  and  $\alpha = 0.35$ .

### Solution to Example 5.8

The first zero-crossing (null-to-null) bandwidth of a rectangular pulse is equal to

$$2/T_s = 2/(41.06 \mu\text{s}) = 48.71 \text{ kHz}$$

and that of a raised cosine filter with  $\alpha = 0.35$  is

$$\frac{1}{T_s}(1 + \alpha) = \frac{1}{41.06 \mu\text{s}}(1 + 0.35) = 32.88 \text{ kHz}$$

## 5.6 Geometric Representation of Modulation Signals

Digital modulation involves choosing a particular signal waveform  $s_i(t)$ , from a finite set of possible signal waveforms (or symbols) based on the informa-

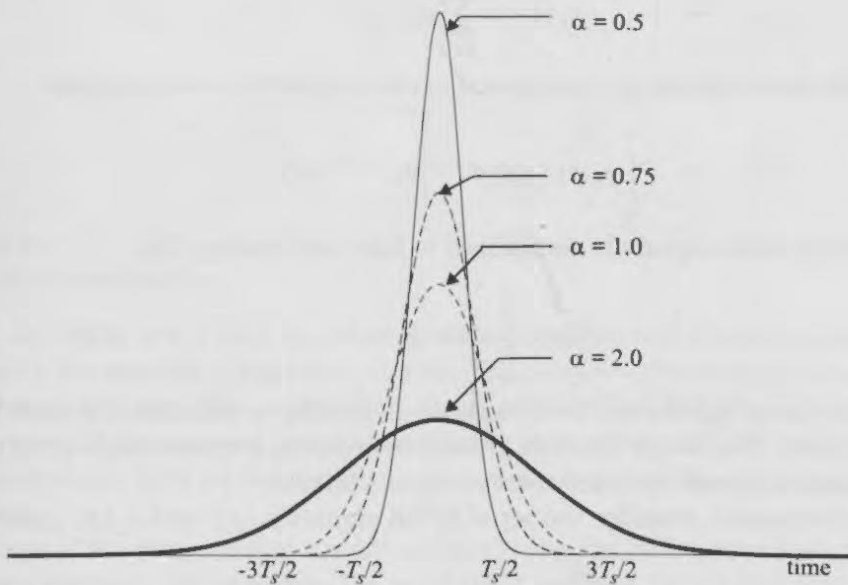


Figure 5.20  
Impulse response of a Gaussian pulse-shaping filter.

tion bits applied to the modulator. If there are a total of  $M$  possible signals, the modulation signal set  $S$  can be represented as

$$S = \{s_1(t), s_2(t), \dots, s_M(t)\} \quad (5.55)$$

For binary modulation schemes, a binary information bit is mapped directly to a signal, and  $S$  will contain only two signals. For higher level modulation schemes ( $M$ -ary keying) the signal set will contain more than two signals, and each signal (or symbol) will represent more than a single bit of information. With a signal set of size  $M$ , it is possible to transmit a maximum of  $\log_2 M$  bits of information per symbol.

It is instructive to view the elements of  $S$  as points in a *vector space*. The vector space representation of modulation signals provides valuable insight into the performance of particular modulation schemes. The vector space concepts are extremely general and can be applied to any type of modulation.

Basic to the geometric viewpoint is the fact that any finite set of physically realizable waveforms in a vector space can be expressed as a linear combination of  $N$  orthonormal waveforms which form the basis of that vector space. To represent the modulation signals on a vector space, one must find a set of signals that form a basis for that vector space. Once a basis is determined, any point in that vector space can be represented as a linear combination of the basis signals  $\{\phi_j(t) | j = 1, 2, \dots, N\}$  such that

$$s_i(t) = \sum_{j=1}^N s_{ij} \phi_j(t) \quad (5.56)$$

The basis signals are orthogonal to one another in time such that

$$\int_{-\infty}^{\infty} \phi_i(t) \phi_j(t) dt = 0 \quad i \neq j \quad (5.57)$$

Each of the basis signals is normalized to have unit energy, i.e.,

$$E = \int_{-\infty}^{\infty} \phi_i^2(t) dt = 1 \quad (5.58)$$

The basis signals can be thought of as forming a coordinate system for the vector space. The Gram-Schmidt procedure provides a systematic way of obtaining the basis signals for a given set of signals [Zie92].

For example, consider the set of BPSK signals  $s_1(t)$  and  $s_2(t)$  given by

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) \quad 0 \leq t \leq T_b \quad (5.59.a)$$

and

$$s_2(t) = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) \quad 0 \leq t \leq T_b \quad (5.59.b)$$

where  $E_b$  is the energy per bit,  $T_b$  is the bit period, and a rectangular pulse shape  $p(t) = \Pi((t - T_b/2)/T_b)$  is assumed.  $\phi_i(t)$  for this signal set simply consists of a single waveform  $\phi_1(t)$  where

$$\phi_1(t) = \sqrt{\frac{2}{T_b}} \cos(2\pi f_c t) \quad 0 \leq t \leq T_b \quad (5.60)$$

Using this basis signal, the BPSK signal set can be represented as

$$S_{\text{BPSK}} = \{ \sqrt{E_b} \phi_1(t), -\sqrt{E_b} \phi_1(t) \} \quad (5.61)$$

This signal set can be shown geometrically in Figure 5.21. Such a representation is called a *constellation diagram* which provides a graphical representation of the complex envelope of each possible symbol state. The  $x$ -axis of a constellation diagram represents the in-phase component of the complex envelope, and the  $y$ -axis represents the quadrature component of the complex envelope. The distance between signals on a constellation diagram relates to how different the modulation waveforms are, and how well a receiver can differentiate between all possible symbols when random noise is present.





Figure 5.21  
BPSK constellation diagram.

It should be noted that the number of basis signals will always be less than or equal to the number of signals in the set. The number of basis signals required to represent the complete modulation signal set is called the *dimension* of the vector space. If there are as many basis signals as there are signals in the modulation signal set, then all the signals in the set are necessarily *orthogonal* to one another.

Some of the properties of a modulation scheme can be inferred from its constellation diagram. For example, the bandwidth occupied by the modulation signals decreases as the number of signal points/dimension increases. Therefore, if a modulation scheme has a constellation that is densely packed, it is more bandwidth efficient than a modulation scheme with a sparsely packed constellation. However, it should be noted that the bandwidth occupied by a modulated signal increases with the dimension  $N$  of the constellation.

The probability of bit error is proportional to the distance between the closest points in the constellation. This implies that a modulation scheme with a constellation that is densely packed is less energy efficient than a modulation scheme that has a sparse constellation.

A simple upper bound for the probability of symbol error in an additive white Gaussian noise channel (AWGN) channel with a noise spectral density  $N_0$  for an arbitrary constellation can be obtained using the union bound [Zie92]. The union bound provides a representative estimate for the average probability of error for a particular modulation signal,  $P_s(\epsilon|s_i)$

$$P_s(\epsilon|s_i) \leq \sum_{j \neq i} Q\left(\frac{d_{ij}}{\sqrt{2N_0}}\right) \quad (5.62)$$

where  $d_{ij}$  is the Euclidean distance between the  $i$ th and  $j$ th signal point in the constellation, and  $Q(x)$  is the  $Q$ -function defined in Appendix D

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx \quad (5.63)$$

If all of the  $M$  modulation waveforms are equally likely to be transmitted, then the average probability of error for a modulation can be estimated by

$$P_s(\epsilon) = P_s(\epsilon|s_i)P(s_i) = \frac{1}{M} \sum_{i=1}^M P_s(\epsilon|s_i) \quad (5.64)$$

For symmetric constellations, the distance between all constellation points are equivalent, and the conditional error probability  $P_s(\epsilon|s_i)$  is the same for all  $i$ . Hence equation (5.62) gives the average probability of symbol error for a particular constellation set.

### 5.7 Linear Modulation Techniques

Digital modulation techniques may be broadly classified as *linear* and *non-linear*. In linear modulation techniques, the amplitude of the transmitted signal,  $s(t)$ , varies linearly with the modulating digital signal,  $m(t)$ . Linear modulation techniques are bandwidth efficient and hence are very attractive for use in wireless communication systems where there is an increasing demand to accommodate more and more users within a limited spectrum.

In a linear modulation scheme, the transmitted signal  $s(t)$  can be expressed as [Zie92]

$$\begin{aligned} s(t) &= \text{Re} [A m(t) \exp(j2\pi f_c t)] \\ &= A [m_R(t) \cos(2\pi f_c t) - m_I(t) \sin(2\pi f_c t)] \end{aligned} \quad (5.65)$$

where  $A$  is the amplitude,  $f_c$  is the carrier frequency, and  $m(t) = m_R(t) + jm_I(t)$  is a complex envelope representation of the modulated signal which is in general complex form. From equation (5.65), it is clear that the amplitude of the carrier varies linearly with the modulating signal. Linear modulation schemes, in general, do not have a constant envelope. As shown subsequently, some nonlinear modulations may have either linear or constant carrier envelopes, depending on whether or not the baseband waveform is pulse shaped.

While linear modulation schemes have very good spectral efficiency, they must be transmitted using linear RF amplifiers which have poor power efficiency [You79]. Using power efficient nonlinear amplifiers leads to the regeneration of filtered sidelobes which can cause severe adjacent channel interference, and results in the loss of all the spectral efficiency gained by linear modulation. However, clever ways have been developed to get around these difficulties. The most popular linear modulation techniques include pulse-shaped QPSK, OQPSK, and  $\pi/4$  QPSK, which are discussed subsequently.

#### 5.7.1 Binary Phase Shift Keying (BPSK)

In binary phase shift keying (BPSK), the phase of a constant amplitude carrier signal is switched between two values according to the two possible signals  $m_1$  and  $m_2$  corresponding to binary 1 and 0, respectively. Normally, the two

phases are separated by  $180^\circ$ . If the sinusoidal carrier has an amplitude  $A_c$  and energy per bit  $E_b = \frac{1}{2}A_c^2 T_b$ , then the transmitted BPSK signal is either

$$S_{\text{BPSK}}(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \theta_c) \quad 0 \leq t \leq T_b \text{ (binary 1)} \quad (5.66.a)$$

or

$$\begin{aligned} S_{\text{BPSK}}(t) &= \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \pi + \theta_c) \\ &= -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \theta_c) \quad 0 \leq t \leq T_b \text{ (binary 0)} \end{aligned} \quad (5.66.b)$$

It is often convenient to generalize  $m_1$  and  $m_2$  as a binary data signal  $m(t)$ , which takes on one of two possible pulse shapes. Then the transmitted signal may be represented as

$$S_{\text{BPSK}}(t) = m(t) \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \theta_c) \quad (5.67)$$

The BPSK signal is equivalent to a double sideband suppressed carrier amplitude modulated waveform, where  $\cos(2\pi f_c t)$  is applied as the carrier, and the data signal  $m(t)$  is applied as the modulating waveform. Hence a BPSK signal can be generated using a balanced modulator.

**Spectrum and Bandwidth of BPSK**

The BPSK signal using a polar baseband data waveform  $m(t)$  can be expressed in complex envelope form as

$$S_{\text{BPSK}} = \text{Re} \{ g_{\text{BPSK}}(t) \exp(j2\pi f_c t) \} \quad (5.68)$$

where  $g_{\text{BPSK}}(t)$  is the complex envelope of the signal given by

$$g_{\text{BPSK}}(t) = \sqrt{\frac{2E_b}{T_b}} m(t) e^{j\theta_c} \quad (5.69)$$

The power spectral density (PSD) of the complex envelope can be shown to be

$$P_{g_{\text{BPSK}}}(f) = 2E_b \left( \frac{\sin \pi f T_b}{\pi f T_b} \right)^2 \quad (5.70)$$

The PSD for the BPSK signal at RF can be evaluated by translating the baseband spectrum to the carrier frequency using the relation given in equation (5.41).

Hence the PSD of a BPSK signal at RF is given by

$$P_{\text{BPSK}} = \frac{E_b}{2} \left[ \left( \frac{\sin \pi (f-f_c) T_b}{\pi (f-f_c) T_b} \right)^2 + \left( \frac{\sin \pi (-f-f_c) T_b}{\pi (-f-f_c) T_b} \right)^2 \right] \quad (5.71)$$

The PSD of the BPSK signal for both rectangular and raised cosine rolloff pulse shapes is plotted in Figure 5.22. The null-to-null bandwidth is found to be equal to twice the bit rate ( $BW = 2R_b = 2/T_b$ ). From the plot, it can also be shown that 90% of the BPSK signal energy is contained within a bandwidth approximately equal to  $1.6R_b$  for rectangular pulses, and all of the energy is within  $1.5R_b$  for pulses with  $\alpha = 0.5$  raised cosine filtering.

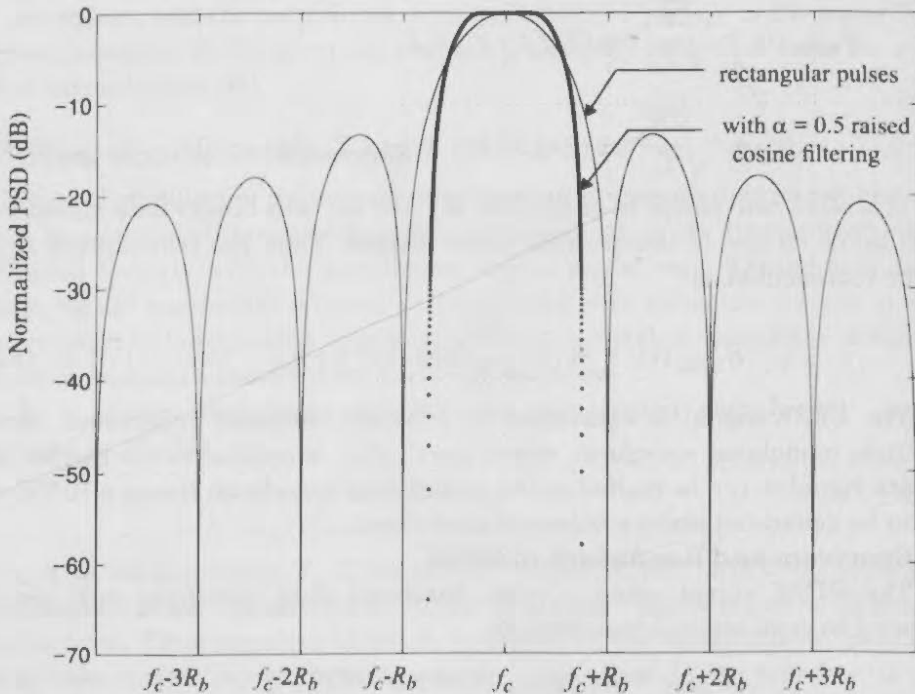


Figure 5.22  
Power Spectral Density (PSD) of a BPSK signal.

### BPSK Receiver

If no multipath impairments are induced by the channel, the received BPSK signal can be expressed as

$$\begin{aligned} S_{\text{BPSK}}(t) &= m(t) \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \theta_c + \theta_{ch}) \\ &= m(t) \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \theta) \end{aligned} \quad (5.72)$$

where  $\theta_{ch}$  is the phase shift corresponding to the time delay in the channel. BPSK uses *coherent* or *synchronous* demodulation, which requires that information about the phase and frequency of the carrier be available at the receiver. If a low level pilot carrier signal is transmitted along with the BPSK signal, then the carrier phase and frequency may be recovered at the receiver using a phase

locked loop (PLL). If no pilot carrier is transmitted, a Costas loop or squaring loop may be used to synthesize the carrier phase and frequency from the received BPSK signal. Figure 5.23 shows the block diagram of a BPSK receiver along with the carrier recovery circuits.

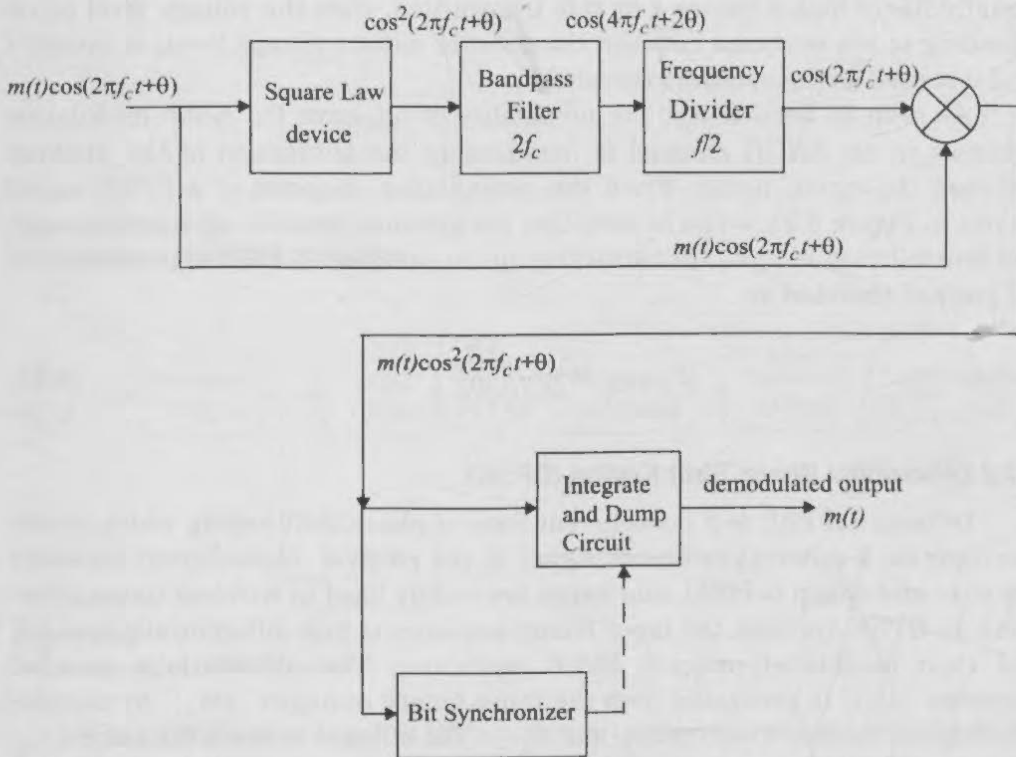


Figure 5.23 BPSK receiver with carrier recovery circuits.

The received signal  $\cos(2\pi f_c t + \theta)$  is squared to generate a dc signal and an amplitude varying sinusoid at twice the carrier frequency. The dc signal is filtered out using a bandpass filter with center frequency tuned to  $2f_c$ . A frequency divider is then used to recreate the waveform  $\cos(2\pi f_c t + \theta)$ . The output of the multiplier after the frequency divider is given by

$$m(t) \sqrt{\frac{2E_b}{T_b}} \cos^2(2\pi f_c t + \theta) = m(t) \sqrt{\frac{2E_b}{T_b}} \left[ \frac{1}{2} + \frac{1}{2} \cos 2(2\pi f_c t + \theta) \right] \quad (5.73)$$

This signal is applied to an integrate and dump circuit which forms the low pass filter segment of a BPSK detector. If the transmitter and receiver pulse shapes are matched, then the detection will be optimum. A bit synchronizer is used to facilitate sampling of the integrator output precisely at the end of each

bit period. At the end of each bit period, the switch at the output of the integrator closes to dump the output signal to the decision circuit. Depending on whether the integrator output is above or below a certain threshold, the decision circuit decides that the received signal corresponds to a binary 1 or 0. The threshold is set at an optimum level such that the probability of error is minimized. If it is equally likely that a binary 1 or 0 is transmitted, then the voltage level corresponding to the midpoint between the detector output voltage levels of binary 1 and 0 is used as the optimum threshold.

As seen in Section 5.6, the probability of bit error for *many* modulation schemes in an AWGN channel is found using the  $Q$ -function of the distance between the signal points. From the constellation diagram of a BPSK signal shown in Figure 5.21, it can be seen that the distance between adjacent points in the constellation is  $2\sqrt{E_b}$ . Substituting this in equation (5.62), the probability of bit error is obtained as

$$P_{e, \text{BPSK}} = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \quad (5.74)$$

### 5.7.2 Differential Phase Shift Keying (DPSK)

Differential PSK is a noncoherent form of phase shift keying which avoids the need for a coherent reference signal at the receiver. Noncoherent receivers are easy and cheap to build, and hence are widely used in wireless communications. In DPSK systems, the input binary sequence is first differentially encoded and then modulated using a BPSK modulator. The differentially encoded sequence  $\{d_k\}$  is generated from the input binary sequence  $\{m_k\}$  by complementing the modulo-2 sum of  $m_k$  and  $d_{k-1}$ . The effect is to leave the symbol  $d_k$  unchanged from the previous symbol if the incoming binary symbol  $m_k$  is 1, and to toggle  $d_k$  if  $m_k$  is 0. Table 5.1 illustrates the generation of a DPSK signal for a sample sequence  $m_k$  which follows the relationship  $d_k = \overline{m_k \oplus d_{k-1}}$ .

Table 5.1 Illustration of the Differential Encoding Process

$\{m_k\}$		1	0	0	1	0	1	1	0
$\{d_{k-1}\}$		1	1	0	1	1	0	0	0
$\{d_k\}$	1	1	0	1	1	0	0	0	1

A block diagram of a DPSK transmitter is shown in Figure 5.24. It consists of a one bit delay element and a logic circuit interconnected so as to generate the differentially encoded sequence from the input binary sequence. The output is passed through a product modulator to obtain the DPSK signal. At the receiver, the original sequence is recovered from the demodulated differentially encoded signal through a complementary process, as shown in Figure 5.25.

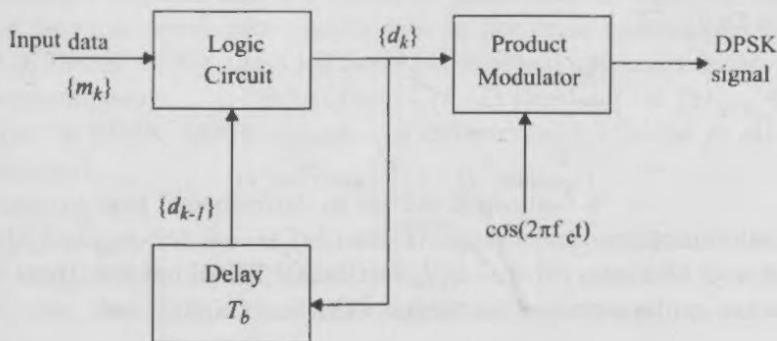


Figure 5.24  
Block diagram of a DPSK transmitter.

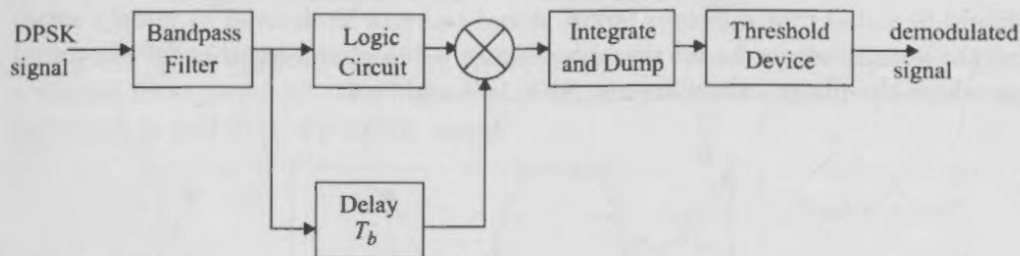


Figure 5.25  
Block diagram of DPSK receiver.

While DPSK signaling has the advantage of reduced receiver complexity, its energy efficiency is inferior to that of coherent PSK by about 3 dB. The average probability of error for DPSK in additive white Gaussian noise is given by

$$P_{e, \text{DPSK}} = \frac{1}{2} \exp\left(-\frac{E_b}{N_0}\right) \tag{5.75}$$

### 5.7.3 Quadrature Phase Shift Keying (QPSK)

Quadrature phase shift keying (QPSK) has twice the bandwidth efficiency of BPSK, since 2 bits are transmitted in a single modulation symbol. The phase of the carrier takes on 1 of 4 equally spaced values, such as 0,  $\pi/2$ ,  $\pi$ , and  $3\pi/2$ , where each value of phase corresponds to a unique pair of message bits. The QPSK signal for this set of symbol states may be defined as

$$S_{\text{QPSK}}(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left[2\pi f_c t + (i-1)\frac{\pi}{2}\right] \quad 0 \leq t \leq T_s \quad i = 1, 2, 3, 4. \tag{5.76}$$

where  $T_s$  is the symbol duration and is equal to twice the bit period.

Using trigonometric identities, the above equations can be rewritten for the interval  $0 \leq t \leq T_s$  as,

$$S_{\text{QPSK}}(t) = \sqrt{\frac{2E_s}{T_s}} \cos \left[ (i-1) \frac{\pi}{2} \right] \cos(2\pi f_c t) - \sqrt{\frac{2E_s}{T_s}} \sin \left[ (i-1) \frac{\pi}{2} \right] \sin(2\pi f_c t) \quad (5.77)$$

If basis functions  $\phi_1(t) = \sqrt{2/T_s} \cos(2\pi f_c t)$ ,  $\phi_2(t) = \sqrt{2/T_s} \sin(2\pi f_c t)$  are defined over the interval  $0 \leq t \leq T_s$  for the QPSK signal set, then the 4 signals in the set can be expressed in terms of the basis signals as

$$S_{\text{QPSK}} = \{ \sqrt{E_s} \cos \left[ (i-1) \frac{\pi}{2} \right] \phi_1(t) - \sqrt{E_s} \sin \left[ (i-1) \frac{\pi}{2} \right] \phi_2(t) \} \quad i = 1, 2, 3, 4 \quad (5.78)$$

Based on this representation, a QPSK signal can be depicted using a two-dimensional constellation diagram with four points as shown in Figure 5.26a. It should be noted that different QPSK signal sets can be derived by simply rotating the constellation. As an example, Figure 5.26b shows another QPSK signal set where the phase values are  $\pi/4$ ,  $3\pi/4$ ,  $5\pi/4$  and  $7\pi/4$ .

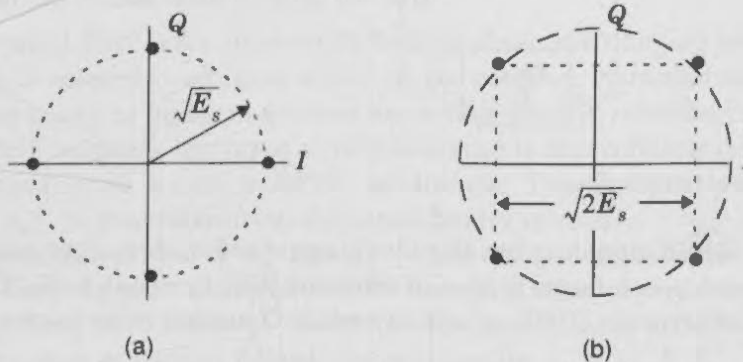


Figure 5.26

- (a) QPSK constellation where the carrier phases are  $0, \pi/2, \pi, 3\pi/2$ .  
 (b) QPSK constellation where the carrier phases are  $\pi/4, 3\pi/4, 5\pi/4, 7\pi/4$ .

From the constellation diagram of a QPSK signal, it can be seen that the distance between adjacent points in the constellation is  $\sqrt{2E_s}$ . Since each symbol corresponds to two bits, then  $E_s = 2E_b$ , thus the distance between two neighboring points in the QPSK constellation is equal to  $2\sqrt{E_b}$ . Substituting this in equation (5.62), the average probability of bit error in the additive white Gaussian noise (AWGN) channel is obtained as

$$P_{e, \text{QPSK}} = Q \left( \sqrt{\frac{2E_b}{N_0}} \right) \quad (5.79)$$



A striking result is that the bit error probability of QPSK is identical to BPSK, but twice as much data can be sent in the same bandwidth. Thus when compared to BPSK, QPSK provides twice the spectral efficiency with exactly the same energy efficiency.

Similar to BPSK, QPSK can also be differentially encoded to allow noncoherent detection.

**Spectrum and Bandwidth of QPSK Signals**

The power spectral density of a QPSK signal can be obtained in a manner similar to that used for BPSK, with the bit periods  $T_b$  replaced by symbol periods  $T_s$ . Hence, the PSD of a QPSK signal using rectangular pulses can be expressed as

$$P_{\text{QPSK}} = \frac{E_s}{2} \left[ \left( \frac{\sin \pi (f-f_c) T_s}{\pi (f-f_c) T_s} \right)^2 + \left( \frac{\sin \pi (-f-f_c) T_s}{\pi (-f-f_c) T_s} \right)^2 \right] \tag{5.80}$$

$$= E_b \left[ \left( \frac{\sin 2\pi (f-f_c) T_b}{2\pi (f-f_c) T_b} \right)^2 + \left( \frac{\sin 2\pi (-f-f_c) T_b}{2\pi (-f-f_c) T_b} \right)^2 \right]$$

The PSD of a QPSK signal for rectangular and raised cosine filtered pulses is plotted in Figure 5.27. The null-to-null RF bandwidth is equal to the bit rate  $R_b$ , which is half that of a BPSK signal.

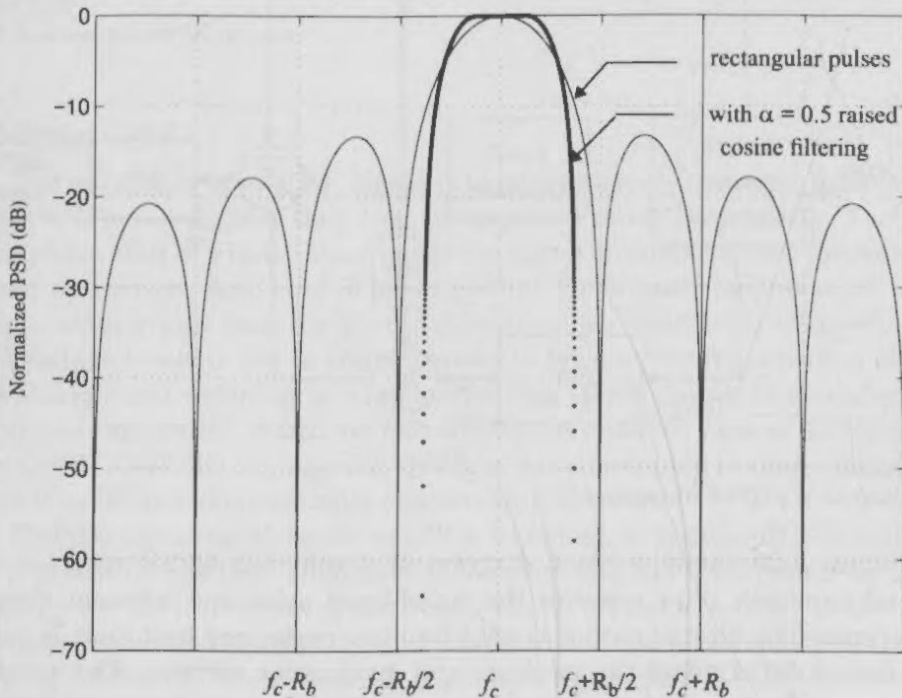


Figure 5.27  
Power spectral density of a QPSK signal.

### 5.7.4 QPSK Transmission and Detection Techniques

Figure 5.28 shows a block diagram of a typical QPSK transmitter. The unipolar binary message stream has bit rate  $R_b$  and is first converted into a bipolar non-return-to-zero (NRZ) sequence using a unipolar to bipolar converter. The bit stream  $m(t)$  is then split into two bit streams  $m_I(t)$  and  $m_Q(t)$  (in-phase and quadrature streams), each having a bit rate of  $R_s = R_b/2$ . The bit stream  $m_I(t)$  is called the “even” stream and  $m_Q(t)$  is called the “odd” stream. The two binary sequences are separately modulated by two carriers  $\phi_1(t)$  and  $\phi_2(t)$ , which are in quadrature. The two modulated signals, each of which can be considered to be a BPSK signal, are summed to produce a QPSK signal. The filter at the output of the modulator confines the power spectrum of the QPSK signal within the allocated band. This prevents spill-over of signal energy into adjacent channels and also removes out-of-band spurious signals generated during the modulation process. In most implementations, pulse shaping is done at baseband to provide proper RF filtering at the transmitter output.

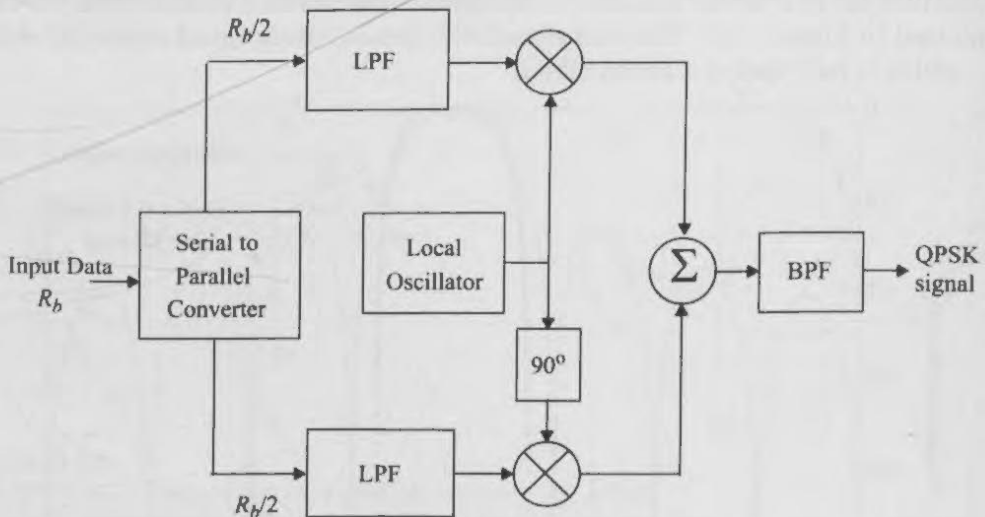


Figure 5.28  
Block diagram of a QPSK transmitter.

Figure 5.29 shows a block diagram of a coherent QPSK receiver. The frontend bandpass filter removes the out-of-band noise and adjacent channel interference. The filtered output is split into two parts, and each part is coherently demodulated using the in-phase and quadrature carriers. The coherent carriers used for demodulation are recovered from the received signal using carrier recovery circuits of the type described in Figure 5.23. The outputs of the demodulators are passed through decision circuits which generate the in-phase

and quadrature binary streams. The two components are then multiplexed to reproduce the original binary sequence.

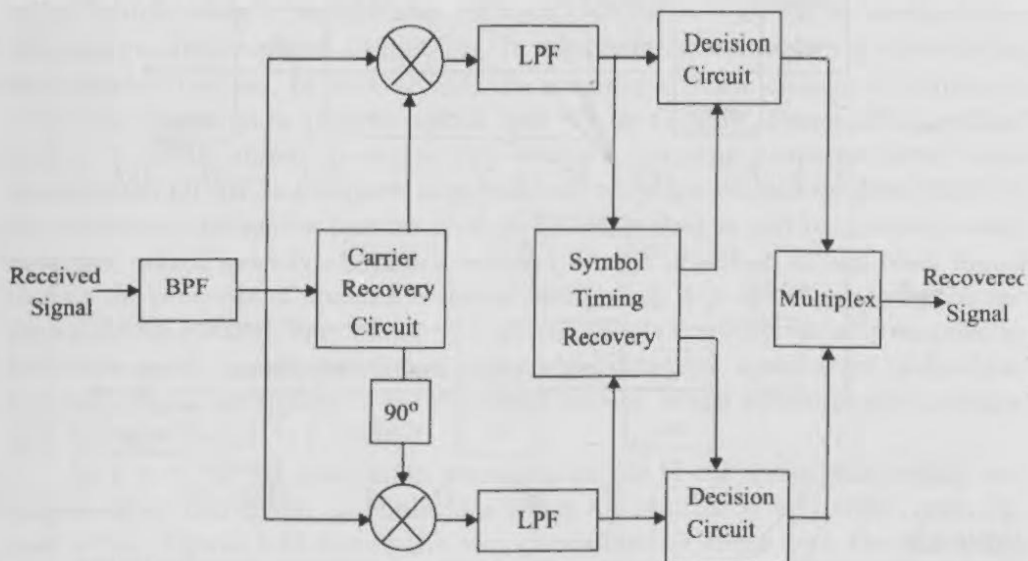


Figure 5.29  
Block diagram of a QPSK receiver.

### 5.7.5 Offset QPSK

The amplitude of a QPSK signal is ideally constant. However, when QPSK signals are pulse shaped, they lose the constant envelope property. The occasional phase shift of  $\pi$  radians can cause the signal envelope to pass through zero for just an instant. Any kind of hardlimiting or nonlinear amplification of the zero-crossings brings back the filtered sidelobes since the fidelity of the signal at small voltage levels is lost in transmission. To prevent the regeneration of sidelobes and spectral widening, it is imperative that QPSK signals be amplified only using linear amplifiers, which are less efficient. A modified form of QPSK, called *offset QPSK (OQPSK)* or *staggered QPSK* is less susceptible to these deleterious effects [Pas79] and supports more efficient amplification.

OQPSK signaling is similar to QPSK signaling, as represented by equation (5.77), except for the time alignment of the even and odd bit streams. In QPSK signaling, the bit transitions of the even and odd bit streams occur at the same time instants, but in OQPSK signaling, the even and odd bit streams,  $m_I(t)$  and  $m_Q(t)$ , are offset in their relative alignment by one bit period (half-symbol period). This is shown in the waveforms of Figure 5.30.

Due to the time alignment of  $m_I(t)$  and  $m_Q(t)$  in standard QPSK, phase transitions occur only once every  $T_s = 2T_b$  s, and will be a maximum of  $180^\circ$  if there is a change in the value of both  $m_I(t)$  and  $m_Q(t)$ . However, in OQPSK

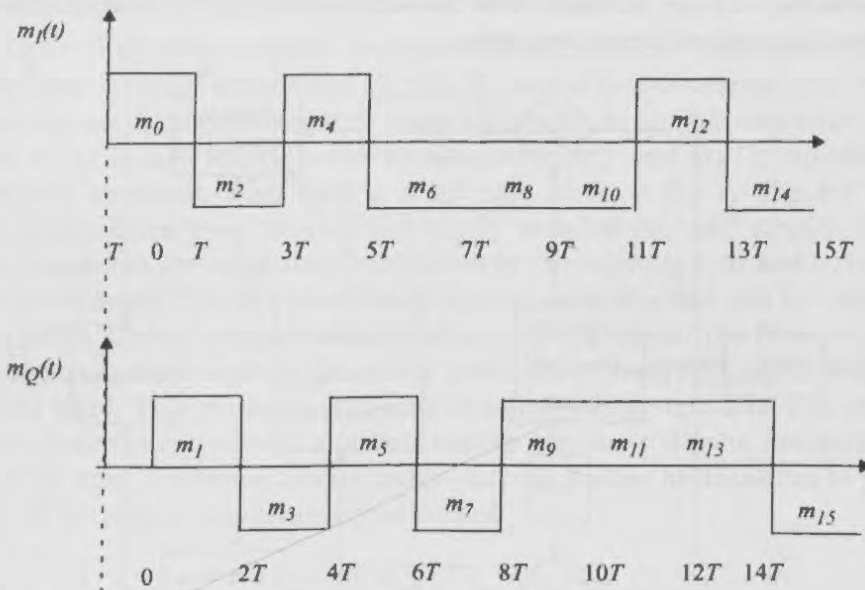


Figure 5.30

The time offset waveforms that are applied to the in-phase and quadrature arms of an OQPSK modulator. Notice that a half-symbol offset is used.

signaling, bit transitions (and hence phase transitions) occur every  $T_b$  s. Since the transitions instants of  $m_I(t)$  and  $m_Q(t)$  are offset, at any given time only one of the two bit streams can change values. This implies that the maximum phase shift of the transmitted signal at any given time is limited to  $\pm 90^\circ$ . Hence, by switching phases more frequently (i.e., every  $T_b$  s instead of  $2T_b$  s) OQPSK signaling eliminates  $180^\circ$  phase transitions.

Since  $180^\circ$  phase transitions have been eliminated, bandlimiting of (i.e., pulse shaping) OQPSK signals does not cause the signal envelope to go to zero. Obviously, there will be some amount of ISI caused by the bandlimiting process, especially at the  $90^\circ$  phase transition points. But the envelope variations are considerably less, and hence bandlimiting or nonlinear amplification of OQPSK signals does not regenerate the high frequency sidelobes as much as in QPSK. Thus, spectral occupancy is significantly reduced, while permitting more efficient RF amplification.

The spectrum of an OQPSK signal is identical to that of a QPSK signal, hence both signals occupy the same bandwidth. The staggered alignment of the even and odd bit streams does not change the nature of the spectrum. OQPSK retains its bandlimited nature even after nonlinear amplification, and therefore is very attractive for mobile communication systems where bandwidth efficiency and efficient nonlinear amplifiers are critical for low power drain. Further, OQPSK signals also appear to perform better than QPSK in the presence of phase jitter due to noisy reference signals at the receiver [Chu87].

### 5.7.6 $\pi/4$ QPSK

The  $\pi/4$  shifted QPSK modulation is a quadrature phase shift keying technique which offers a compromise between OQPSK and QPSK in terms of the allowed maximum phase transitions. It may be demodulated in a coherent or noncoherent fashion. In  $\pi/4$  QPSK, the maximum phase change is limited to  $\pm 135^\circ$ , as compared to  $180^\circ$  for QPSK and  $90^\circ$  for OQPSK. Hence, the bandlimited  $\pi/4$  QPSK signal preserves the constant envelope property better than bandlimited QPSK, but is more susceptible to envelope variations than OQPSK. An extremely attractive feature of  $\pi/4$  QPSK is that it can be noncoherently detected, which greatly simplifies receiver design. Further, it has been found that in the presence of multipath spread and fading,  $\pi/4$  QPSK performs better than OQPSK [Liu89]. Very often,  $\pi/4$  QPSK signals are differentially encoded to facilitate easier implementation of differential detection or coherent demodulation with phase ambiguity in the recovered carrier. When differentially encoded,  $\pi/4$  QPSK is called  $\pi/4$  DQPSK.

In a  $\pi/4$  QPSK modulator, signaling points of the modulated signal are selected from two QPSK constellations which are shifted by  $\pi/4$  with respect to each other. Figure 5.31 shows the two constellations along with the combined constellation where the links between two signal points indicate the possible phase transitions. Switching between two constellations, every successive bit ensures that there is at least a phase shift which is an integer multiple of  $\pi/4$  radians between successive symbols. This ensures that there is a phase transition for every symbol, which enables a receiver to perform timing recovery and synchronization.

### 5.7.7 $\pi/4$ QPSK Transmission Techniques

A block diagram of a generic  $\pi/4$  QPSK transmitter is shown in Figure 5.32. The input bit stream is partitioned by a serial-to-parallel (S/P) converter into two parallel data streams  $m_{I,k}$  and  $m_{Q,k}$ , each with a symbol rate equal to half that of the incoming bit rate. The  $k$ th in-phase and quadrature pulses,  $I_k$  and  $Q_k$ , are produced at the output of the signal mapping circuit over time  $kT \leq t \leq (k+1)T$  and are determined by their previous values,  $I_{k-1}$  and  $Q_{k-1}$ , as well as  $\theta_k$ , which itself is a function of  $\phi_k$ , which is a function of the current input symbols  $m_{I,k}$  and  $m_{Q,k}$ .  $I_k$  and  $Q_k$  represent rectangular pulses over one symbol duration having amplitudes given by

$$I_k = \cos\theta_k = I_{k-1} \cos\phi_k - Q_{k-1} \sin\phi_k \quad (5.81)$$

$$Q_k = \sin\theta_k = I_{k-1} \sin\phi_k + Q_{k-1} \cos\phi_k \quad (5.82)$$

where

$$\theta_k = \theta_{k-1} + \phi_k \quad (5.83)$$

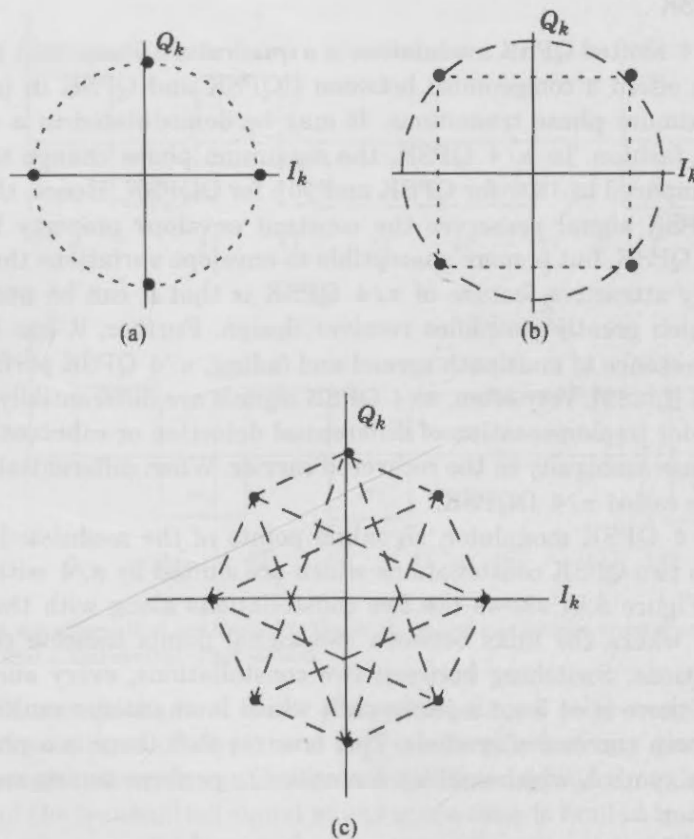


Figure 5.31

Constellation diagram of a  $\pi/4$  QPSK signal: (a) possible states for  $\theta_k$  when  $\theta_{k-1} = n\pi/4$ ; (b) possible states when  $\theta_{k-1} = n\pi/2$ ; (c) all possible states.

and  $\theta_k$  and  $\theta_{k-1}$  are phases of the  $k$ th and  $k-1$ st symbols. The phase shift  $\phi_k$  is related to the input symbols  $m_{I_k}$  and  $m_{Q_k}$  according to Table 5.2.

Table 5.2 Carrier Phase Shifts Corresponding to Various Input Bit Pairs [Feh91], [Rap91b]

Information bits $m_{I_k}, m_{Q_k}$	Phase shift $\phi_k$
1 1	$\pi/4$
0 1	$3\pi/4$
0 0	$-3\pi/4$
1 0	$-\pi/4$

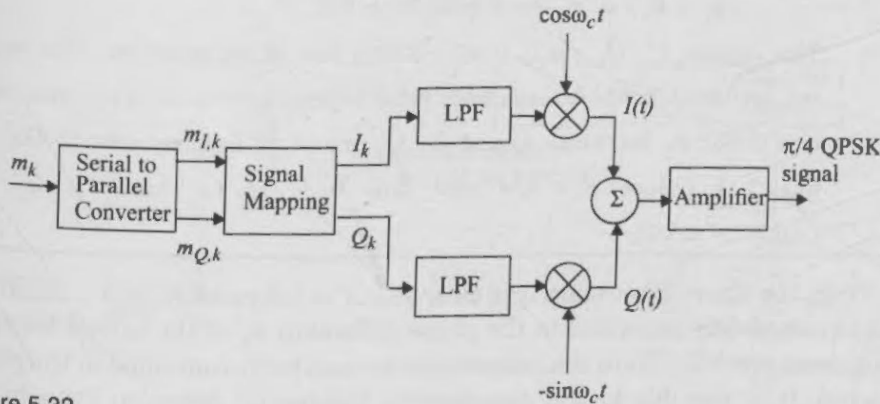


Figure 5.32  
Generic  $\pi/4$  QPSK transmitter.

Just as in a QPSK modulator, the in-phase and quadrature bit streams  $I_k$  and  $Q_k$  are then separately modulated by two carriers which are in quadrature with one another, to produce the  $\pi/4$  QPSK waveform given by

$$s_{\pi/4\text{QPSK}}(t) = I(t) \cos \omega_c t - Q(t) \sin \omega_c t \tag{5.84}$$

where

$$I(t) = \sum_{k=0}^{N-1} I_k p(t - kT_s - T_s/2) = \sum_{k=0}^{N-1} \cos \theta_k p(t - kT_s - T_s/2) \tag{5.85}$$

$$Q(t) = \sum_{k=0}^{N-1} Q_k p(t - kT_s - T_s/2) = \sum_{k=0}^{N-1} \sin \theta_k p(t - kT_s - T_s/2) \tag{5.86}$$

Both  $I_k$  and  $Q_k$  are usually passed through raised cosine rolloff pulse shaping filters before modulation, in order to reduce the bandwidth occupancy. The function  $p(t)$  in equations (5.85) and (5.86) corresponds to the pulse shape, and  $T_s$  is the symbol period. Pulse shaping also reduces the spectral restoration problem which may be significant in fully saturated, nonlinear amplified systems. It should be noted that the values of  $I_k$  and  $Q_k$  and the peak amplitude of the waveforms  $I(t)$  and  $Q(t)$  can take one of the five possible values, 0, +1, -1,  $+1/\sqrt{2}$ ,  $-1/\sqrt{2}$ .

**Example 5.9**

Assume that  $\theta_0 = 0^\circ$ . The bit stream 0 0 1 0 1 1 is to be sent using  $\pi/4$  DQPSK. The leftmost bits are first applied to the transmitter. Determine the phase of  $\theta_k$ , and the values of  $I_k$ ,  $Q_k$  during transmission.

**Solution to Example 5.9**

Given  $\theta_0 = 0^\circ$ , the first two bits are 0 0, which implies that

$$\theta_1 = \theta_0 + \phi_1 = -3\pi/4 \text{ from Table 5.2.}$$

This implies  $I_1, Q_1$  are  $(-0.707, -0.707)$  from (5.81) and (5.82). The second two bits are 1 0, which maps from Table 5.2 into  $\phi_2 = -\pi/4$ . Thus from equation (5.83),  $\theta_2$  becomes  $-\pi$ , and  $I_2, Q_2$  are  $(-1, 0)$  from equation (5.81). The bits 1 1 induce  $\phi_3 = \pi/4$  and thus  $\theta_3 = -3\pi/4$ . Thus,  $I_3, Q_3$  are  $(-0.707, -0.707)$ .

From the above discussion it is clear that the information in a  $\pi/4$  QPSK signal is completely contained in the phase difference  $\phi_k$  of the carrier between two adjacent symbols. Since the information is completely contained in the phase difference, it is possible to use noncoherent differential detection even in the absence of differential encoding.

### 5.7.8 $\pi/4$ QPSK Detection Techniques

Due to ease of hardware implementation, differential detection is often employed to demodulate  $\pi/4$  QPSK signals. In an AWGN channel, the BER performance of a differentially detected  $\pi/4$  QPSK is about 3 dB inferior to QPSK, while coherently detected  $\pi/4$  QPSK has the same error performance as QPSK. In low bit rate, fast Rayleigh fading channels, differential detection offers a lower error floor since it does not rely on phase synchronization [Feh91]. There are various types of detection techniques that are used for the detection of  $\pi/4$  QPSK signals. They include *baseband differential detection*, *IF differential detection*, and *FM discriminator detection*. While both the baseband and IF differential detector determine the cosine and sine functions of the phase difference, and then decide on the phase difference accordingly, the FM discriminator detects the phase difference directly in a noncoherent manner. Interestingly, simulations have shown that all 3 receiver structures offer very similar bit error rate performances, although there are implementation issues which are specific to each technique [Anv91].

#### Baseband Differential Detection

Figure 5.33 shows a block diagram of a baseband differential detector. The incoming  $\pi/4$  QPSK signal is quadrature demodulated using two local oscillator signals that have the same frequency as the unmodulated carrier at the transmitter, but not necessarily the same phase. If  $\phi_k = \tan^{-1}(Q_k/I_k)$  is the phase of the carrier due to the  $k$ th data bit, the output  $w_k$  and  $z_k$  from the two low pass filters in the in-phase and quadrature arms of the demodulator can be expressed as

$$w_k = \cos(\phi_k - \gamma) \quad (5.87)$$

$$z_k = \sin(\phi_k - \gamma) \quad (5.88)$$



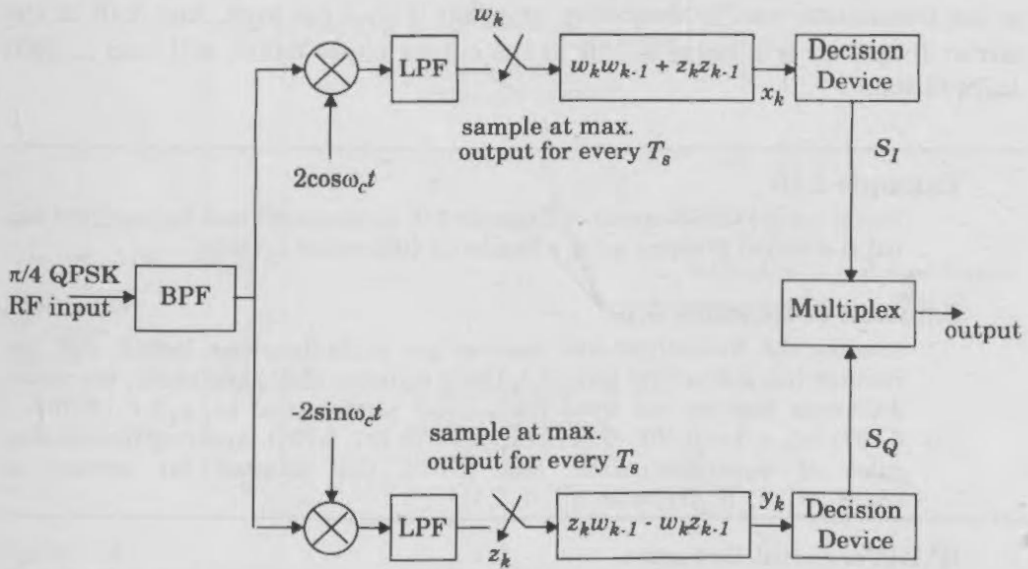


Figure 5.33  
Block diagram of a baseband differential detector [From [Feh91] © IEEE].

where  $\gamma$  is a phase shift due to noise, propagation, and interference. The phase  $\gamma$  is assumed to change much slower than  $\phi_k$  so it is essentially a constant. The two sequences  $w_k$  and  $z_k$  are passed through a differential decoder which operates on the following rule.

$$x_k = w_k w_{k-1} + z_k z_{k-1} \tag{5.89}$$

$$y_k = z_k w_{k-1} - w_k z_{k-1} \tag{5.90}$$

The output of the differential decoder can be expressed as

$$x_k = \cos(\phi_k - \gamma) \cos(\phi_{k-1} - \gamma) + \sin(\phi_k - \gamma) \sin(\phi_{k-1} - \gamma) = \cos(\phi_k - \phi_{k-1}) \tag{5.91}$$

$$y_k = \sin(\phi_k - \gamma) \cos(\phi_{k-1} - \gamma) + \cos(\phi_k - \gamma) \sin(\phi_{k-1} - \gamma) = \sin(\phi_k - \phi_{k-1}) \tag{5.92}$$

The output of the differential decoder is applied to the decision circuit, which uses Table 5.2 to determine

$$S_I = 1, \text{ if } x_k > 0 \quad \text{or} \quad S_I = 0, \text{ if } x_k < 0 \tag{5.93}$$

$$S_Q = 1, \text{ if } y_k > 0 \quad \text{or} \quad S_Q = 0, \text{ if } y_k < 0 \tag{5.94}$$

where  $S_I$  and  $S_Q$  are the detected bits in the in-phase and quadrature arms, respectively.

It is important to ensure the local receiver oscillator frequency is the same as the transmitter carrier frequency, and that it does not drift. Any drift in the carrier frequency will cause a drift in the output phase which will lead to BER degradation.

### Example 5.10

Using the  $\pi/4$  QPSK signal of Example 5.9, demonstrate how the received signal is detected properly using a baseband differential detector.

### Solution to Example 5.10

Assume the transmitter and receiver are perfectly phase locked, and the receiver has a front-end gain of 2. Using equation (5.91) and (5.92), the phase difference between the three transmitted phases yield  $(x_1, y_1) = (-0.707, -0.707)$ ;  $(x_2, y_2) = (0.707, -0.707)$ ;  $(x_3, y_3) = (0.707, 0.707)$ . Applying the decision rules of equations (5.93) and (5.94), the detected bit stream is  $(S_1, S_2, S_3, S_4, S_5, S_6) = (0, 0, 1, 0, 1, 1)$ .

### IF Differential Detector

The IF differential detector shown in Figure 5.34 avoids the need for a local oscillator by using a delay line and two phase detectors. The received signal is converted to IF and is handpass filtered. The bandpass filter is designed to match the transmitted pulse shape, so that the carrier phase is preserved and noise power is minimized. To minimize the effect of ISI and noise, the bandwidth of the filters are chosen to be  $0.57/T_s$  [Liu91]. The received IF signal is differentially decoded using a delay line and two mixers. The bandwidth of the signal at the output of the differential detector is twice that of the baseband signal at the transmitter end.

### FM Discriminator

Figure 5.35 shows a block diagram of an FM discriminator detector for  $\pi/4$  QPSK. The input signal is first filtered using a handpass filter that is matched to the transmitted signal. The filtered signal is then hardlimited to remove any envelope fluctuations. Hardlimiting preserves the phase changes in the input signal and hence no information is lost. The FM discriminator extracts the instantaneous frequency deviation of the received signal which, when integrated over each symbol period gives the phase difference between two sampling instants. The phase difference is then detected by a four level threshold comparator to obtain the original signal. The phase difference can also be detected using a modulo- $2\pi$  phase detector. The modulo- $2\pi$  phase detector improves the BER performance and reduces the effect of click noise [Feh91].

## 5.8 Constant Envelope Modulation

Many practical mobile radio communications systems use nonlinear modulation methods, where the amplitude of the carrier is constant, regardless of the

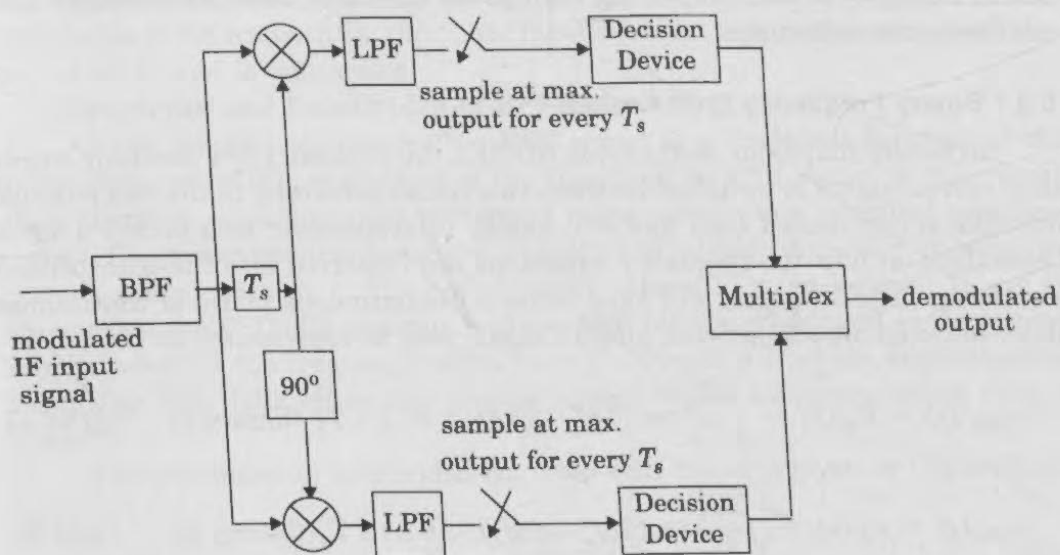


Figure 5.34  
Block diagram of an IF differential detector for  $\pi/4$  QPSK.

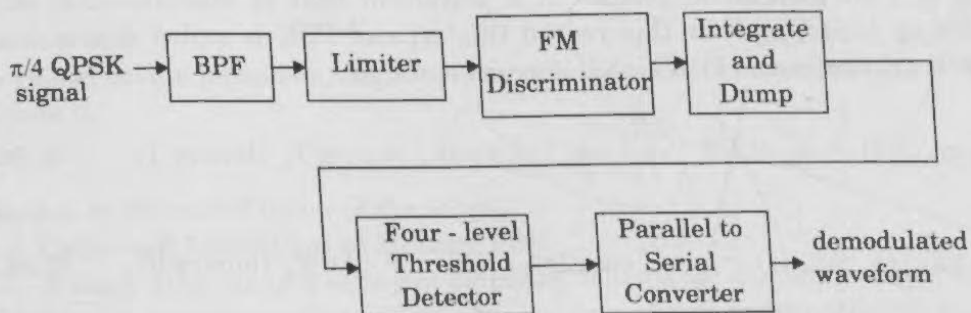


Figure 5.35  
FM discriminator detector for  $\pi/4$  DQPSK demodulation.

variation in the modulating signal. The constant envelope family of modulations has the advantage of satisfying a number of conditions [You79], some of which are:

- Power efficient Class C amplifiers can be used without introducing degradation in the spectrum occupancy of the transmitted signal.
- Low out-of-band radiation of the order of -60 dB to -70 dB can be achieved.
- Limiter-discriminator detection can be used, which simplifies receiver design and provides high immunity against random FM noise and signal fluctuations due to Rayleigh fading.

While constant envelope modulations have many advantages, they occupy a larger bandwidth than linear modulation schemes. In situations where band-

width efficiency is more important than power efficiency, constant envelope modulation is not well-suited.

### 5.8.1 Binary Frequency Shift Keying

In binary frequency shift keying (BFSK), the frequency of a constant amplitude carrier signal is switched between two values according to the two possible message states (called *high* and *low* tones), corresponding to a binary 1 or 0. Depending on how the frequency variations are imparted into the transmitted waveform, the FSK signal will have either a discontinuous phase or continuous phase between bits. In general, an FSK signal may be represented as

$$s_{\text{FSK}}(t) = v_H(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c + 2\pi\Delta f)t \quad 0 \leq t \leq T_b \quad (\text{binary 1}) \quad (5.95.a)$$

$$s_{\text{FSK}}(t) = v_L(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c - 2\pi\Delta f)t \quad 0 \leq t \leq T_b \quad (\text{binary 0}) \quad (5.95.b)$$

where  $2\pi\Delta f$  is a constant offset from the nominal carrier frequency.

One obvious way to generate an FSK signal is to switch between two independent oscillators according to whether the data bit is a 0 or a 1. Normally, this form of FSK generation results in a waveform that is discontinuous at the switching times, and for this reason this type of FSK is called *discontinuous* FSK. A discontinuous FSK signal is represented as

$$s_{\text{FSK}}(t) = v_H(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_H t + \theta_1) \quad 0 \leq t \leq T_b \quad (\text{binary 1}) \quad (5.96.a)$$

$$s_{\text{FSK}}(t) = v_L(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_L t + \theta_2) \quad 0 \leq t \leq T_b \quad (\text{binary 0}) \quad (5.96.b)$$

Since the phase discontinuities pose several problems, such as spectral spreading and spurious transmissions, this type of FSK is generally not used in highly regulated wireless systems.

The more common method for generating an FSK signal is to frequency modulate a single carrier oscillator using the message waveform. This type of modulation is similar to analog FM generation, except that the modulating signal  $m(t)$  is a binary waveform. Therefore, FSK may be represented as

$$\begin{aligned} s_{\text{FSK}}(t) &= \sqrt{\frac{2E_b}{T_b}} \cos[2\pi f_c t + \theta(t)] \\ &= \sqrt{\frac{2E_b}{T_b}} \cos \left[ 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\eta) d\eta \right] \end{aligned} \quad (5.97)$$

It should be noted that even though the modulating waveform  $m(t)$  is discontinuous at bit transitions, the phase function  $\theta(t)$  is proportional to the integral of  $m(t)$  and is continuous.

### Spectrum and Bandwidth of BFSK signals

As the complex envelope of an FSK signal is a nonlinear function of the message signal  $m(t)$ , evaluation of the spectra of an FSK signal is, in general, quite involved, and is usually performed using actual time averaged measurements. The power spectral density of a binary FSK signal consists of discrete frequency components at  $f_c$ ,  $f_c + n\Delta f$ ,  $f_c - n\Delta f$ , where  $n$  is an integer. It can be shown that the PSD of a continuous phase FSK ultimately falls off as the inverse fourth power of the frequency offset from  $f_c$ . However, if phase discontinuities exist, the PSD falls off as the inverse square of the frequency offset from  $f_c$  [Cou93].

The transmission bandwidth  $B_T$  of an FSK signal is given by Carson's rule as

$$B_T = 2\Delta f + 2B \quad (5.98)$$

where  $B$  is the bandwidth of the digital baseband signal. Assuming that first null bandwidth is used, the bandwidth of rectangular pulses is  $B = R$ . Hence, the FSK transmission bandwidth becomes

$$B_T = 2(\Delta f + R) \quad (5.99)$$

If a raised cosine pulse-shaping filter is used, then the transmission bandwidth reduces to

$$B_T = 2\Delta f + (1 + \alpha)R \quad (5.100)$$

where  $\alpha$  is the rolloff factor of the filter.

### Coherent Detection of Binary FSK

A block diagram of a coherent detection scheme for demodulation of binary FSK signals is shown in Figure 5.36. The receiver shown is the optimum detector for coherent binary FSK in the presence of additive white Gaussian noise. It consists of two correlators which are supplied with locally generated coherent reference signals. The difference of the correlator outputs is then compared with a threshold comparator. If the difference signal has a value greater than the threshold, the receiver decides in favor of a 1, otherwise it decides in favor of a 0.

It can be shown that the probability of error for a coherent FSK receiver is given by

$$P_{e, \text{FSK}} = Q\left(\sqrt{\frac{E_b}{N_0}}\right) \quad (5.101)$$

### Noncoherent Detection of Binary FSK

Unlike phase-shift keying, it is possible to detect FSK signals in the presence of noise without using a coherent carrier reference. A block diagram of a

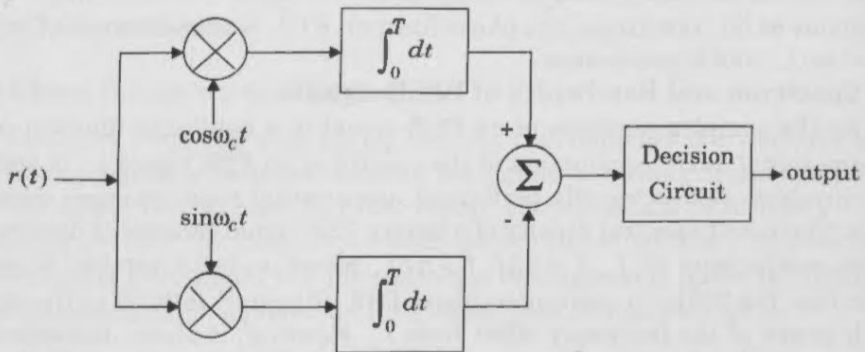


Figure 5.36  
Coherent detection of FSK signals.

noncoherent FSK receiver is shown in Figure 5.37. The receiver consists of a pair of matched filters followed by envelope detectors. The filter in the upper path is matched to the FSK signal of frequency  $f_H$  and the filter in the lower path is matched to the signal of frequency  $f_L$ . These matched filters function as band-pass filters centered at  $f_H$  and  $f_L$ , respectively. The outputs of the envelope detectors are sampled at every  $t = kT_b$ , where  $k$  is an integer, and their values compared. Depending on the magnitude of the envelope detector output, the comparator decides whether the data bit was a 1 or 0.

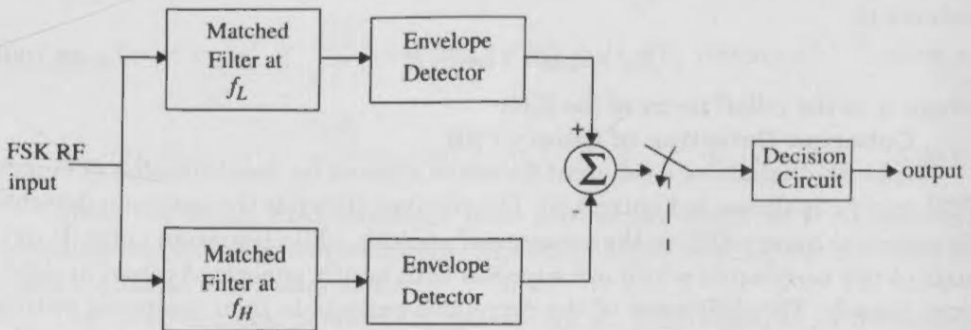


Figure 5.37  
Block diagram of noncoherent FSK receiver.

The average probability of error of an FSK system employing noncoherent detection is given by

$$P_{e, \text{FSK, NC}} = \frac{1}{2} \exp\left(-\frac{E_b}{2N_0}\right) \quad (5.102)$$

### 5.8.2 Minimum Shift Keying (MSK)

Minimum shift keying (MSK) is a special type of *continuous phase-frequency shift keying* (CPFSK) wherein the peak frequency deviation is equal to 1/4 the bit rate. In other words, MSK is continuous phase FSK with a modulation index of 0.5. The modulation index of an FSK signal is similar to the FM modulation index, and is defined as  $k_{\text{FSK}} = (2\Delta F)/R_b$ , where  $\Delta F$  is the peak RF frequency deviation and  $R_b$  is the bit rate. A modulation index of 0.5 corresponds to the minimum frequency spacing that allows two FSK signals to be coherently orthogonal, and the name minimum shift keying implies the minimum frequency separation (i.e. bandwidth) that allows orthogonal detection. Two FSK signals  $v_H(t)$  and  $v_L(t)$  are said to be orthogonal if

$$\int_0^T v_H(t)v_L(t)dt = 0 \tag{5.103}$$

MSK is sometimes referred to as *fast FSK*, as the frequency spacing used is only half as much as that used in conventional noncoherent FSK [Xio94].

MSK is a spectrally efficient modulation scheme and is particularly attractive for use in mobile radio communication systems. It possesses properties such as constant envelope, spectral efficiency, good BER performance, and self-synchronizing capability.

An MSK signal can be thought of as a special form of OQPSK where the baseband rectangular pulses are replaced with half-sinusoidal pulses [Pas79]. These pulses have shapes like the St. Louis arch during a period of  $2T_b$ . Consider the OQPSK signal with the bit streams offset as shown in Figure 5.30. If half-sinusoidal pulses are used instead of rectangular pulses, the modified signal can be defined as MSK and for an N-bit stream is given by

$$S_{\text{MSK}}(t) = \sum_{i=0}^{N-1} m_I(t)p(t - 2iT_b)\cos 2\pi f_c t + \sum_{i=0}^{N-1} m_Q(t)p(t - 2iT_b - T_b)\sin 2\pi f_c t \tag{5.104}$$

$$\text{where } p(t) = \begin{cases} \sin\left(\frac{\pi t}{2T_b}\right) & 0 \leq t \leq 2T_b \\ 0 & \text{elsewhere} \end{cases}$$

and where  $m_I(t)$  and  $m_Q(t)$  are the “odd” and “even” bits of the bipolar data stream which have values of  $\pm 1$  and which feed the in-phase and quadrature arms of the modulator at a rate of  $R_b/2$ . It should be noted that there are a number of variations of MSK that exist in the literature [Sun86]. For example, while one version of MSK uses only positive half-sinusoids as the basic pulse shape, another version uses alternating positive and negative half-sinusoids as the

basic pulse shape. However, all variations of MSK are continuous phase FSK employing different techniques to achieve spectral efficiency [Sun86].

The MSK waveform can be seen as a special type of a continuous phase FSK if equation (5.97) is rewritten using trigonometric identities as

$$S_{\text{MSK}}(t) = \sqrt{\frac{2E_b}{T_b}} \cos \left[ 2\pi f_c t - m_I(t)m_Q(t) \frac{\pi t}{2T_b} + \phi_k \right] \quad (5.105)$$

where  $\phi_k$  is 0 or  $\pi$  depending on whether  $m_I(t)$  is 1 or -1. From equation (5.105) it can be deduced that MSK has a constant amplitude. Phase continuity at the bit transition periods is ensured by choosing the carrier frequency to be an integral multiple of one fourth the bit rate,  $1/4T$ . Comparing equation (5.105) with equation (5.97), it can be concluded that the MSK signal is an FSK signal with binary signaling frequencies of  $f_c + 1/4T$  and  $f_c - 1/4T$ . It can further be seen from equation (5.105) that the phase of the MSK signal varies linearly during the course of each bit period [Pro94, Chapter 9].

### MSK Power Spectrum

From equation (5.41), the RF power spectrum is obtained by frequency shifting the magnitude squared of the Fourier transform of the baseband pulse-shaping function. For MSK, the baseband pulse shaping function is given by

$$p(t) = \begin{cases} \cos\left(\frac{\pi t}{2T}\right) & |t| < T \\ 0 & \text{elsewhere} \end{cases} \quad (5.106)$$

Thus the normalized power spectral density for MSK is given by [Pas79]

$$P_{\text{MSK}} = \frac{16}{\pi^2} \left( \frac{\cos 2\pi(f + f_c)T}{1.16f^2 T^2} \right)^2 + \frac{16}{\pi^2} \left( \frac{\cos 2\pi(f - f_c)T}{1.16f^2 T^2} \right)^2 \quad (5.107)$$

Figure 5.38 shows the power spectral density of an MSK signal. The spectral density of QPSK and OQPSK are also drawn for comparison. From Figure 5.38 it is seen that the MSK spectrum has lower sidelobes than QPSK and OQPSK. Ninety-nine percent of the MSK power is contained within a bandwidth  $B = 1.2/T$ , while for QPSK and OQPSK, the 99 percent bandwidth  $B$  is equal to  $8/T$ . The faster rolloff of the MSK spectrum is due to the fact that smoother pulse functions are used. Figure 5.38 also shows that the main lobe of MSK is wider than that of QPSK and OQPSK, and hence when compared in terms of first null bandwidth, MSK is less spectrally efficient than the phase-shift keying techniques [Pas79].

Since there is no abrupt change in phase at bit transition periods, bandlimiting the MSK signal to meet required out-of-band specifications does not cause the envelope to go through zero. The envelope is kept more or less constant even after bandlimiting. Any small variations in the envelope level can be removed by hardlimiting at the receiver without raising the out-of-band radiation levels. Since the amplitude is kept constant, MSK signals can be amplified using efficient nonlinear amplifiers. The continuous phase property makes it highly desir-



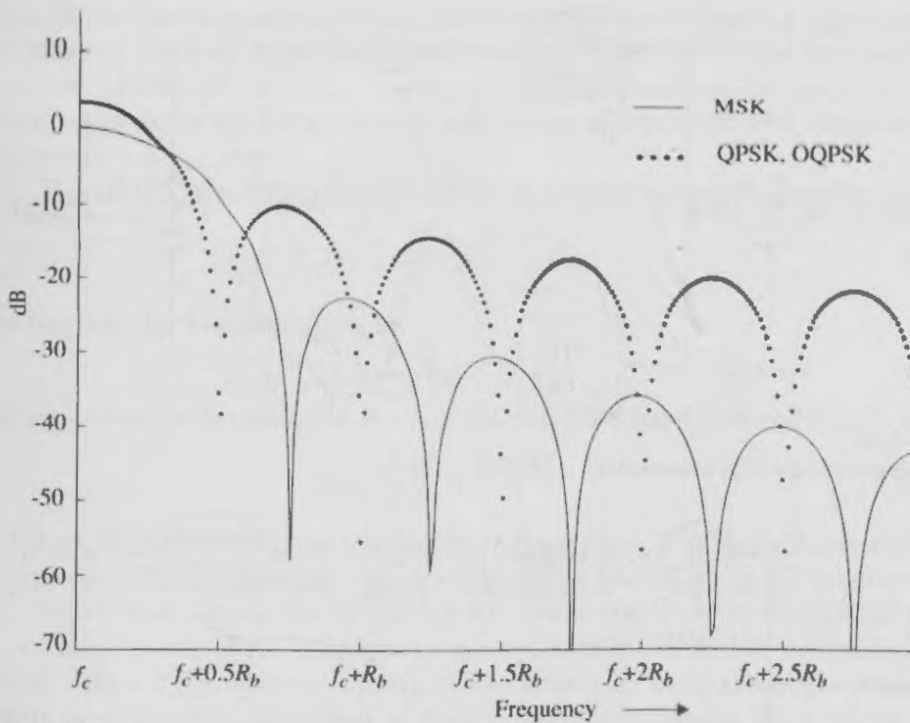


Figure 5.38

Power spectral density of MSK signals as compared to QPSK and OQPSK signals.

able for highly reactive loads. In addition to these advantages, MSK has simple demodulation and synchronization circuits. It is for these reasons that MSK is a popular modulation scheme for mobile radio communications.

#### MSK Transmitter and Receiver

Figure 5.39 shows a typical MSK modulator. Multiplying a carrier signal with  $\cos[\pi t/2T]$  produces two phase-coherent signals at  $f_c + 1/4T$  and  $f_c - 1/4T$ . These two FSK signals are separated using two narrow bandpass filters and appropriately combined to form the in-phase and quadrature carrier components  $x(t)$  and  $y(t)$ , respectively. These carriers are multiplied with the odd and even bit streams,  $m_I(t)$  and  $m_Q(t)$ , to produce the MSK modulated signal  $s_{\text{MSK}}(t)$ .

The block diagram of an MSK receiver is shown in Figure 5.40. The received signal  $s_{\text{MSK}}(t)$  (in the absence of noise and interference) is multiplied by the respective in-phase and quadrature carriers  $x(t)$  and  $y(t)$ . The output of the multipliers are integrated over two bit periods and dumped to a decision circuit at the end of each two bit periods. Based on the level of the signal at the output of the integrator, the threshold detector decides whether the signal is a 0 or a 1. The output data streams correspond to  $m_I(t)$  and  $m_Q(t)$ , which are offset combined to obtain the demodulated signal.

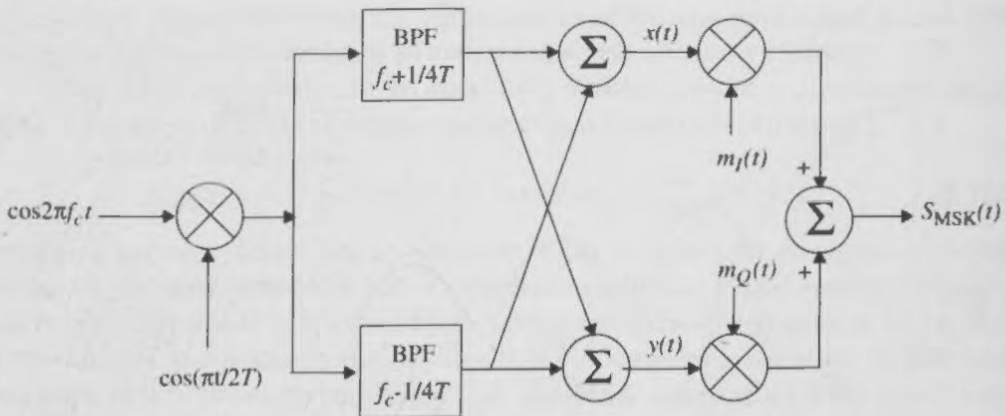


Figure 5.39  
Block diagram of an MSK transmitter.

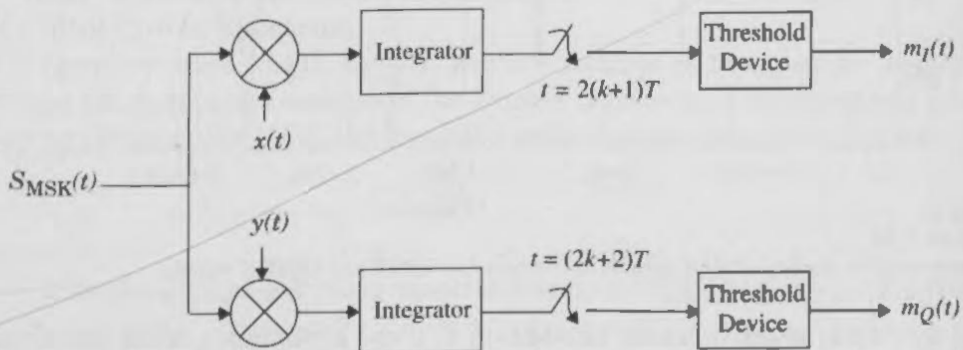


Figure 5.40  
Block diagram of an MSK receiver.

### 5.8.3 Gaussian Minimum Shift Keying (GMSK)

GMSK is a simple binary modulation scheme which may be viewed as a derivative of MSK. In GMSK, the sidelobe levels of the spectrum are further reduced by passing the modulating NRZ data waveform through a premodulation Gaussian pulse-shaping filter [Mur81] (see section 5.5.3). Baseband Gaussian pulse shaping smooths the phase trajectory of the MSK signal and hence stabilizes the instantaneous frequency variations over time. This has the effect of considerably reducing the sidelobe levels in the transmitted spectrum.

Premodulation Gaussian filtering converts the full response message signal (where each baseband symbol occupies a single bit period  $T$ ) into a partial response scheme where each transmitted symbol spans several bit periods. However, since pulse shaping does not cause the pattern-averaged phase trajectory to deviate from that of simple MSK, GMSK can be coherently detected just as an MSK signal, or noncoherently detected as simple FSK. In practice, GMSK is most attractive for its excellent power efficiency (due to the constant envelope) and its excellent spectral efficiency. The premodulation Gaussian filtering intro-

duces ISI in the transmitted signal, but it can be shown that the degradation is not severe if the 3 dB-bandwidth-bit duration product ( $BT$ ) of the filter is greater than 0.5. GMSK sacrifices the irreducible error rate caused by partial response signaling in exchange for extremely good spectral efficiency and constant envelope properties.

The GMSK premodulation filter has an impulse response given by

$$h_G(t) = \frac{\sqrt{\pi}}{\alpha} \exp\left(-\frac{\pi^2}{\alpha^2} t^2\right) \quad (5.108)$$

and the transfer function given by

$$H_G(f) = \exp(-\alpha^2 f^2) \quad (5.109)$$

The parameter  $\alpha$  is related to  $B$ , the 3 dB baseband bandwidth of  $H_G(f)$ , by

$$\alpha = \frac{\sqrt{\ln 2}}{\sqrt{2}B} = \frac{0.5887}{B} \quad (5.110)$$

and the GMSK filter may be completely defined from  $B$  and the baseband symbol duration  $T$ . It is therefore customary to define GMSK by its  $BT$  product.

Figure 5.41 shows the simulated RF power spectrum of the GMSK signal for various values of  $BT$ . The power spectrum of MSK, which is equivalent to GMSK with a  $BT$  product of infinity, is also shown for comparison purposes. It is clearly seen from the graph that as the  $BT$  product decreases, the sidelobe levels fall off very rapidly. For example, for a  $BT=0.5$ , the peak of the second lobe is more than 30dB below the main lobe, whereas for simple MSK, the second lobe is only 20 dB below main lobe. However, reducing  $BT$  increases the irreducible error rate produced by the low pass filter due to ISI. As shown in Section 5.11, mobile radio channels induce an irreducible error rate due to mobile velocity, so *as long as the GMSK irreducible error rate is less than that produced by the mobile channel, there is no penalty in using GMSK*. Table 5.3 shows occupied bandwidth containing a given percentage of power in a GMSK signal as a function of the  $BT$  product [Mur81].

Table 5.3 Occupied RF Bandwidth (for GMSK and MSK as a fraction of  $R_b$ ) Containing a Given Percentage of Power [Mur81]. Notice that GMSK is spectrally tighter than MSK.

$BT$	90%	99%	99.9%	99.99%
0.2 GMSK	0.52	0.79	0.99	1.22
0.25 GMSK	0.57	0.86	1.09	1.37
0.5 GMSK	0.69	1.04	1.33	2.08
MSK	0.78	1.20	2.76	6.00

While the GMSK spectrum becomes more and more compact with decreasing  $BT$  value, the degradation due to ISI increases. It was shown by Ishizuka [Ish80] that the BER degradation due to ISI caused by filtering is minimum for a

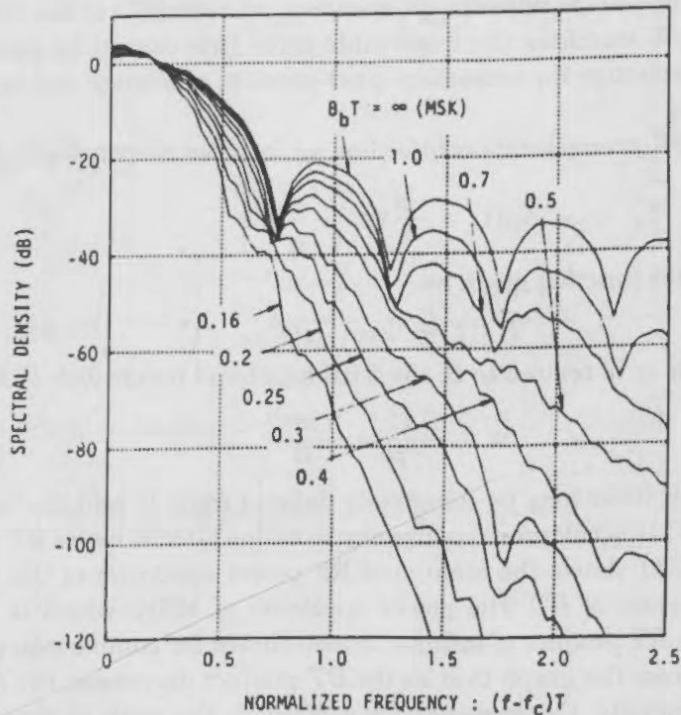


Figure 5.41  
Power spectral density of a GMSK signal [From [Mur81] © IEEE].

$BT$  value of 0.5887, where the degradation in the required  $E_b/N_0$  is only 0.14 dB from the case of no ISI.

#### GMSK Bit Error Rate

The bit error rate for GMSK was first found in [Mur81] for AWGN channels, and was shown to offer performance within 1 dB of optimum MSK when  $BT=0.25$ . The bit error probability is a function of  $BT$ , since the pulse shaping impacts ISI. The bit error probability for GMSK is given by

$$P_e = Q \left\{ \sqrt{\frac{2\gamma E_b}{N_0}} \right\} \quad (5.111.a)$$

where  $\gamma$  is a constant related to  $BT$  by

$$\gamma \cong \begin{cases} 0.68 & \text{for GMSK with } BT = 0.25 \\ 0.85 & \text{for simple MSK } (BT = \infty) \end{cases} \quad (5.111.b)$$

#### GMSK Transmitter and Receiver

The simplest way to generate a GMSK signal is to pass a NRZ message bit stream through a Gaussian baseband filter having an impulse response given in

equation (5.108), followed by an FM modulator. This modulation technique is shown in Figure 5.42 and is currently used in a variety of analog and digital implementations for the U.S. Cellular Digital Packet Data (CDPD) system as well as for the Global System for Mobile (GSM) system. Figure 5.42 may also be implemented digitally using a standard  $I/Q$  modulator.

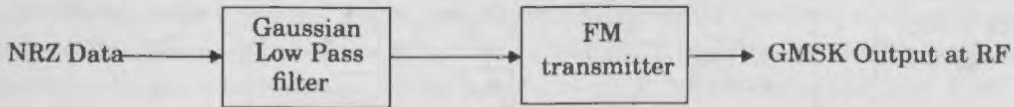


Figure 5.42

Block diagram of a GMSK transmitter using direct FM generation.

GMSK signals can be detected using orthogonal coherent detectors as shown in Figure 5.43, or with simple noncoherent detectors such as standard FM discriminators. Carrier recovery is sometimes performed using a method suggested by de Buda [deB72] where the sum of the two discrete frequency components contained at the output of a frequency doubler is divided by four. De Buda's method is similar to the Costas loop and is equivalent to that of a PLL with a frequency doubler. This type of receiver can be easily implemented using digital logic as shown in Figure 5.44. The two  $D$  flip-flops act as a quadrature product demodulator and the XOR gates act as baseband multipliers. The mutually orthogonal reference carriers are generated using two  $D$  flip-flops, and the VCO center frequency is set equal to four times the carrier center frequency. A nonoptimum, but highly effective method of detecting GMSK signal is to simply sample the output of an FM demodulator.

### Example 5.11

Find the 3-dB bandwidth for a Gaussian low pass filter used to produce 0.25 GMSK with a channel data rate of  $R_b = 270$  kbps. What is the 90% power bandwidth in the RF channel? Specify the Gaussian filter parameter  $\alpha$ .

### Solution to Example 5.11

From the problem statement

$$T = \frac{1}{R_b} = \frac{1}{270 \times 10^3} = 3.7 \mu\text{s}$$

Solving for  $B$ , where  $BT = 0.25$ ,

$$B = \frac{0.25}{T} = \frac{0.25}{3.7 \times 10^{-6}} = 67.567 \text{ kHz}$$

Thus the 3-dB bandwidth is 67.567 kHz. To determine the 90% power bandwidth, use Table 5.3 to find that  $0.57R_b$  is the desired value. Thus, the occupied RF spectrum for a 90% power bandwidth is given by

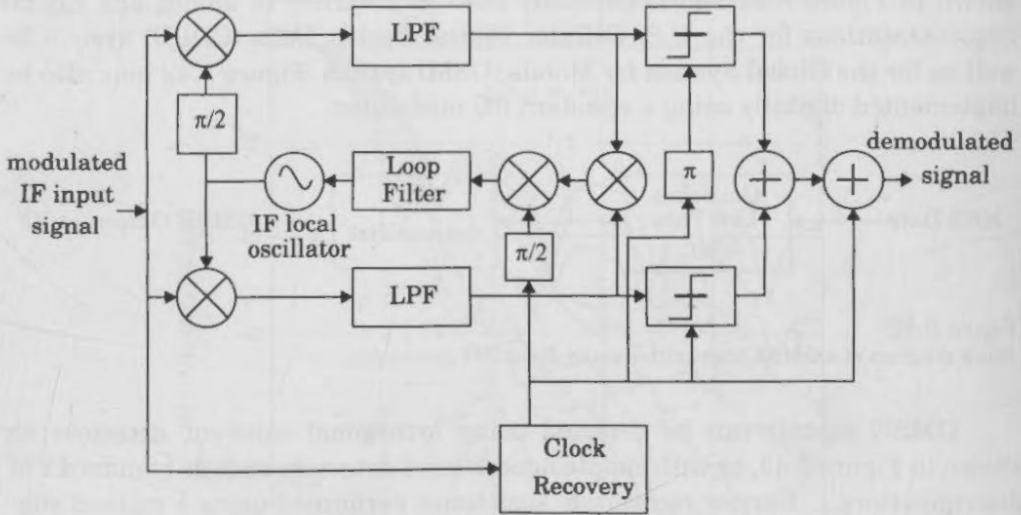


Figure 5.43  
Block diagram of a GMSK receiver.

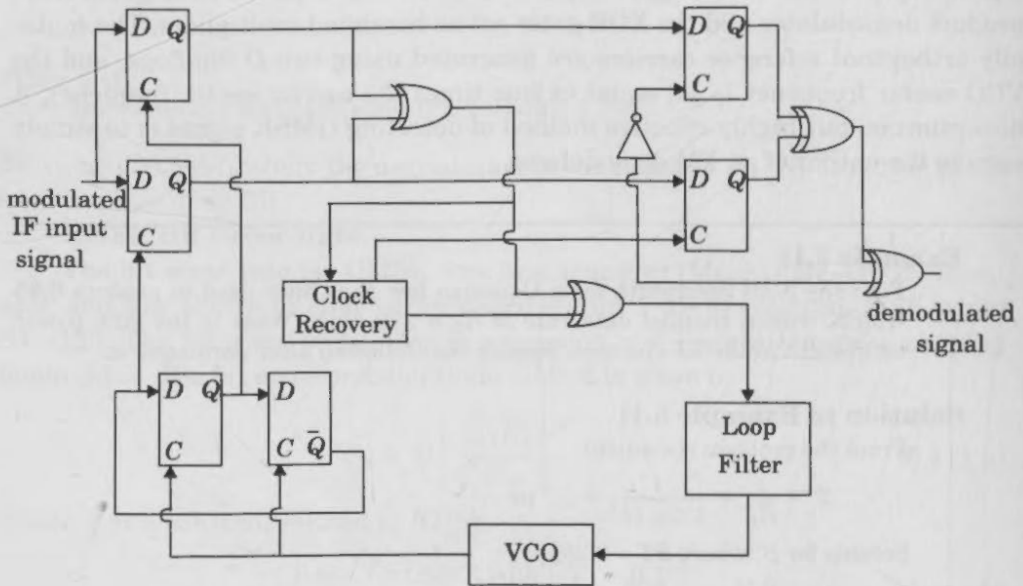


Figure 5.44  
Digital logic circuit for GMSK demodulation [From [deB72] © IEEE].

$$RF\ BW = 0.57R_b = 0.57 \times 270 \times 10^3 = 153.9\text{ kHz}$$

### 5.9 Combined Linear and Constant Envelope Modulation Techniques

Modern modulation techniques exploit the fact that digital baseband data may be sent by varying both the envelope and phase (or frequency) of an RF carrier. Because the envelope and phase offer two degrees of freedom, such modulation techniques map baseband data into four or more possible RF carrier signals. Such modulation techniques are called M-ary modulation, since they can represent more signals than if just the amplitude or phase were varied alone.

In an M-ary signaling scheme, two or more bits are grouped together to form symbols and one of M possible signals,  $s_1(t), s_2(t), \dots, s_M(t)$  is transmitted during each symbol period of duration  $T_s$ . Usually, the number of possible signals is  $M = 2^n$ , where  $n$  is an integer. Depending on whether the amplitude, phase, or frequency of the carrier is varied, the modulation scheme is called M-ary ASK, M-ary PSK, or M-ary FSK. Modulations which alter both the amplitude and phase of the carrier are the subject of active research.

M-ary signaling is particularly attractive for use in bandlimited channels, but are limited in their applications due to sensitivity to timing jitter (i.e., timing errors increase when smaller distances between signals in the constellation diagram are used. This results in poorer error performance).

M-ary modulation schemes achieve better bandwidth efficiency at the expense of power efficiency. For example, an 8-PSK system requires a bandwidth that is  $\log_2 8 = 3$  times smaller than a BPSK system, whereas its BER performance is significantly worse than BPSK since signals are packed more closely in the signal constellation.

#### 5.9.1 M-ary Phase Shift Keying (MPSK)

In M-ary PSK, the carrier phase takes on one of M possible values, namely,  $\theta_i = 2(i-1)\pi/M$ , where  $i = 1, 2, \dots, M$ . The modulated waveform can be expressed as

$$S_i(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left(2\pi f_c t + \frac{2\pi}{M}(i-1)\right), \quad 0 \leq t \leq T_s \quad i = 1, 2, \dots, M \quad (5.112)$$

where  $E_s = (\log_2 M)E_b$  is the energy per symbol and  $T_s = (\log_2 M)T_b$  is the symbol period. The above equation can be rewritten in quadrature form as

$$S_i(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left[(i-1)\frac{2\pi}{M}\right] \cos(2\pi f_c t) - \sqrt{\frac{2E_s}{T_s}} \sin\left[(i-1)\frac{2\pi}{M}\right] \sin(2\pi f_c t) \quad i = 1, 2, \dots, M \quad (5.113)$$

By choosing orthogonal basis signals  $\phi_1(t) = \sqrt{\frac{2}{T_s}} \cos(2\pi f_c t)$ , and  $\phi_2(t) = \sqrt{\frac{2}{T_s}} \sin(2\pi f_c t)$  defined over the interval  $0 \leq t \leq T_s$ , the M-ary PSK signal set can be expressed as

$$S_{\text{M-PSK}}(t) = \left\{ \sqrt{E_s} \cos \left[ (i-1) \frac{\pi}{2} \right] \phi_1(t) - \sqrt{E_s} \sin \left[ (i-1) \frac{\pi}{2} \right] \phi_2(t) \right\} \quad (5.114)$$

$$i = 1, 2, \dots, M$$

Since there are only two basis signals, the constellation of M-ary PSK is two dimensional. The M-ary message points are equally spaced on a circle of radius  $\sqrt{E_s}$  centered at the origin. The constellation diagram of an 8-ary PSK signal set is illustrated in Figure 5.45. It is clear from Figure 5.45 that MPSK is a constant envelope signal when no pulse shaping is used.

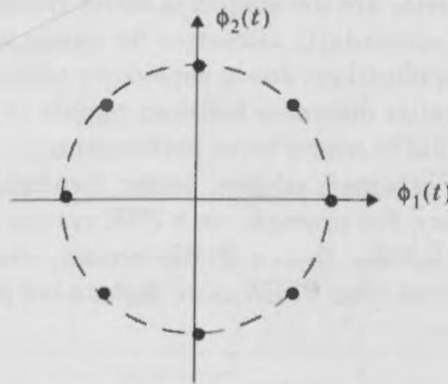


Figure 5.45  
Constellation diagram of an M-ary PSK system (M=8).

Equation (5.62), can be used to compute the probability of symbol error for MPSK systems in an AWGN channel. From the geometry of Figure 5.45, it is easily seen that the distance between adjacent symbols is equal to  $2\sqrt{E_s} \sin\left(\frac{\pi}{M}\right)$ . Hence the average symbol error probability of an M-ary PSK system is given by

$$P_e \leq 2Q\left(\sqrt{\frac{2E_b \log_2 M}{N_0}} \sin\left(\frac{\pi}{M}\right)\right) \quad (5.115)$$

Just as in BPSK and QPSK modulation, M-ary PSK modulation is either coherently detected or differentially encoded for noncoherent differential detection. The symbol error probability of a differential M-ary PSK system in AWGN channel for  $M \geq 4$  is approximated by [Hay94]



$$P_e \approx 2Q\left(\sqrt{\frac{4E_s}{N_0}} \sin\left(\frac{\pi}{2M}\right)\right) \tag{5.116}$$

**Power Spectra of M-ary PSK**

The power spectral density (PSD) of an M-ary PSK signal can be obtained in a manner similar to that described for BPSK and QPSK signals. The symbol duration  $T_s$  of an M-ary PSK signal is related to the bit duration  $T_b$  by

$$T_s = T_b \log_2 M \tag{5.117}$$

The PSD of the M-ary PSK signal with rectangular pulses is given by

$$P_{\text{MPSK}} = \frac{E_s}{2} \left[ \left( \frac{\sin \pi(f - f_c)T_s}{\pi(f - f_c)T_s} \right)^2 + \left( \frac{\sin \pi(-f - f_c)T_s}{\pi(-f - f_c)T_s} \right)^2 \right] \tag{5.118}$$

$$P_{\text{MPSK}} = \frac{E_b \log_2 M}{2} \left[ \left( \frac{\sin \pi(f - f_c)T_b \log_2 M}{\pi(f - f_c)T_b \log_2 M} \right)^2 + \left( \frac{\sin \pi(-f - f_c)T_b \log_2 M}{\pi(-f - f_c)T_b \log_2 M} \right)^2 \right] \tag{5.119}$$

The PSD of M-ary PSK systems for  $M = 8$  and  $M = 16$  are shown in Figure 5.46. As clearly seen from equation (5.119) and Figure 5.46, the first null bandwidth of M-ary PSK signals decrease as  $M$  increases while  $R_b$  is held constant. Therefore, as the value of  $M$  increases, the bandwidth efficiency also increases. That is, for fixed  $R_b$ ,  $\eta_b$  increases and  $B$  decreases as  $M$  is increased. At the same time, increasing  $M$  implies that the constellation is more densely packed, and hence the power efficiency (noise tolerance) is decreased. The bandwidth and power efficiency of M-PSK systems using ideal Nyquist pulse shaping in AWGN for various values of  $M$  are listed in Table 5.4. These values assume no timing jitter or fading, which have a large negative effect on bit error rate as  $M$  increases. In general, simulation must be used to determine bit error values in actual wireless communication channels, since interference and multipath can alter the instantaneous phase of an MPSK signal, thereby creating errors at the detector. Also, the particular implementation of the receiver often impacts performance.

**Table 5.4 Bandwidth and Power Efficiency of M-ary PSK Signals**

$M$	2	4	8	16	32	64
$\eta_B = R_b/B^*$	0.5	1	1.5	2	2.5	3
$E_b/N_0$ for BER= $10^{-6}$	10.5	10.5	14	18.5	23.4	28.5

\*.  $B$ : First null bandwidth of M-ary PSK signals

In practice, pilot symbols or equalization must be used to exploit MPSK in mobile channels, and this has not been a popular commercial practice.

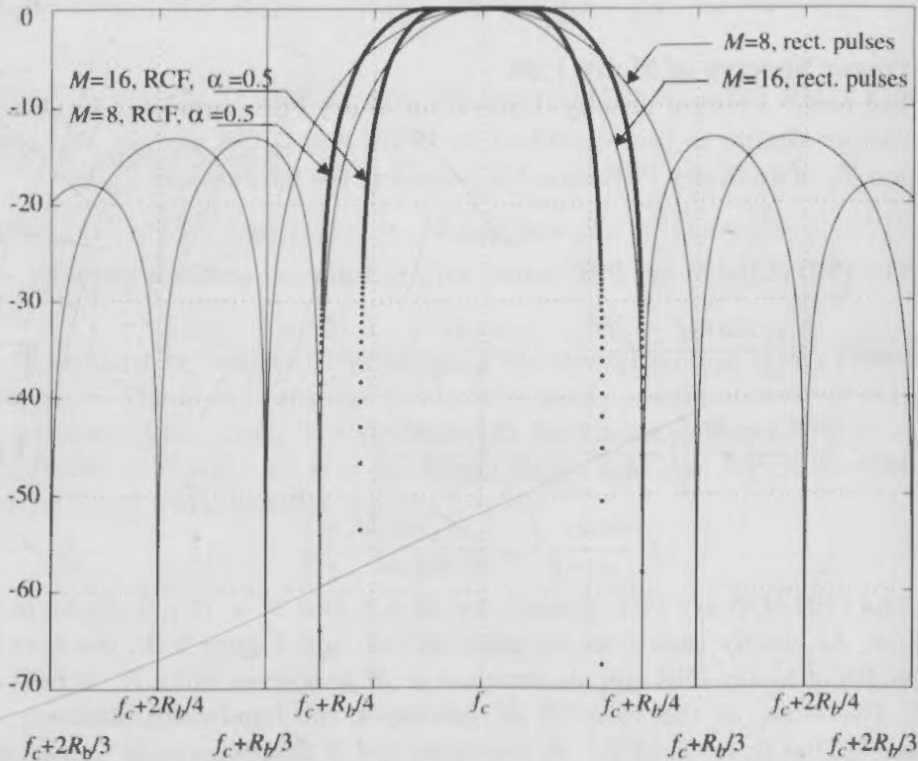


Figure 5.46

M-ary PSK power spectral density, for M=8, 16, (PSD for both rectangular and raised cosine filtered pulses are shown for fixed  $R_b$ ).

### 5.9.2 M-ary Quadrature Amplitude Modulation (QAM)

In M-ary PSK modulation, the amplitude of the transmitted signal was constrained to remain constant, thereby yielding a circular constellation. By allowing the amplitude to also vary with the phase, a new modulation scheme called *quadrature amplitude modulation* (QAM) is obtained. Figure 5.47 shows the constellation diagram of 16-ary QAM. The constellation consists of a square lattice of signal points. The general form of an M-ary QAM signal can be defined as

$$S_i(t) = \sqrt{\frac{2E_{min}}{T_s}} a_i \cos(2\pi f_c t) + \sqrt{\frac{2E_{min}}{T_s}} b_i \sin(2\pi f_c t) \quad (5.120)$$

$$0 \leq t \leq T \quad i = 1, 2, \dots, M$$

where  $E_{min}$  is the energy of the signal with the lowest amplitude, and  $a_i$  and  $b_i$  are a pair of independent integers chosen according to the location of the particular signal point. Note that M-ary QAM does not have constant energy per sym-

bol, nor does it have constant distance between possible symbol states. It reasons that particular values of  $S_i(t)$  will be detected with higher probability than others.

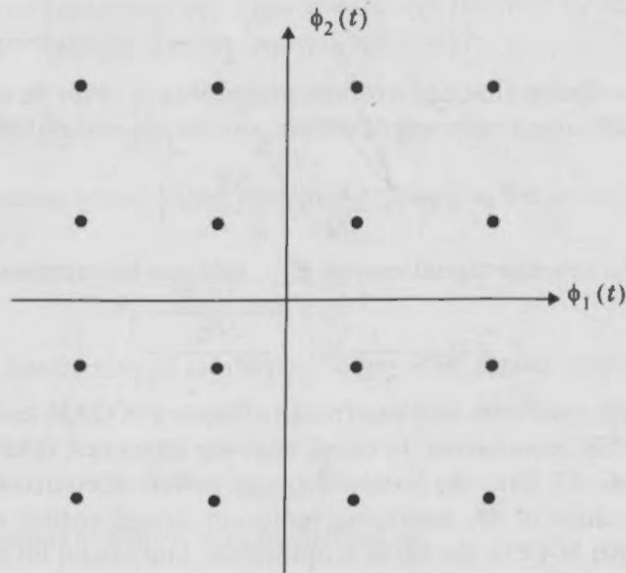


Figure 5.47 Constellation diagram of an M-ary QAM ( $M=16$ ) signal set.

If rectangular pulse shapes are assumed, the signal  $S_i(t)$  may be expanded in terms of a pair of basis functions defined as

$$\phi_1(t) = \sqrt{\frac{2}{T_s}} \cos(2\pi f_c t) \quad 0 \leq t \leq T_s \tag{5.121}$$

$$\phi_2(t) = \sqrt{\frac{2}{T_s}} \sin(2\pi f_c t) \quad 0 \leq t \leq T_s \tag{5.122}$$

The coordinates of the  $i$ th message point are  $a_i \sqrt{E_{min}}$  and  $b_i \sqrt{E_{min}}$  where  $(a_i, b_i)$  is an element of the  $L$  by  $L$  matrix given by

$$\{a_i, b_i\} = \begin{bmatrix} (-L+1, L-1) & (-L+3, L-1) & \dots & (L-1, L-1) \\ (-L+1, L-3) & (-L+3, L-3) & \dots & (L-1, L-3) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ (-L+1, -L+1) & (-L+3, -L+1) & \dots & (L-1, -L+1) \end{bmatrix} \tag{5.123}$$

where  $L = \sqrt{M}$ . For example, for a 16-QAM with signal constellation as shown in Figure 5.47, the  $L$  by  $L$  matrix is

$$\{a_i, b_i\} = \begin{bmatrix} (-3, 3) & (-1, 3) & (1, 3) & (3, 3) \\ (-3, 1) & (-1, 1) & (1, 1) & (3, 1) \\ (-3, -1) & (-1, -1) & (1, -1) & (3, -1) \\ (-3, -3) & (-1, -3) & (1, -3) & (3, -3) \end{bmatrix} \quad (5.124)$$

It can be shown that the average probability of error in an AWGN channel for M-ary QAM, using coherent detection, can be approximated by [Hay94]

$$P_e \cong 4 \left( 1 - \frac{1}{\sqrt{M}} \right) Q \left( \sqrt{\frac{2E_{min}}{N_0}} \right) \quad (5.125)$$

In terms of the average signal energy  $E_{av}$ , this can be expressed as [Zie92]

$$P_e \cong 4 \left( 1 - \frac{1}{\sqrt{M}} \right) Q \left( \sqrt{\frac{3E_{av}}{(M-1)N_0}} \right) \quad (5.126)$$

The power spectrum and bandwidth efficiency of QAM modulation is identical to M-ary PSK modulation. In terms of power efficiency, QAM is superior to M-ary PSK. Table 5.5 lists the bandwidth and power efficiencies of a QAM signal for various values of  $M$ , assuming optimum raised cosine rolloff filtering in AWGN. As with M-PSK, the table is optimistic, and actual bit error probabilities for wireless systems must be determined by simulating the various impairments of the channel and the specific receiver implementation. Pilot tones or equalization must be used for QAM in mobile systems.

Table 5.5 Bandwidth and Power Efficiency of QAM [Zie92]

$M$	4	16	64	256	1024	4096
$\eta_B$	1	2	3	4	5	6
$E_b/N_0$ for BER= $10^{-6}$	10.5	15	18.5	24	28	33.5

### 5.9.3 M-ary Frequency Shift Keying (MFSK)

In M-ary FSK modulation, the transmitted signals are defined by

$$S_i(t) = \sqrt{\frac{2E_s}{T_s}} \cos \left[ \frac{\pi}{T_s} (n_c + i)t \right] \quad 0 \leq t \leq T_s \quad i = 1, 2, \dots, M \quad (5.127)$$

where  $f_c = n_c/2T_s$  for some fixed integer  $n_c$ . The  $M$  transmitted signals are of equal energy and equal duration, and the signal frequencies are separated by  $1/2T_s$  Hertz, making the signals orthogonal to one another.

For coherent M-ary FSK, the optimum receiver consists of a bank of  $M$  correlators, or matched filters. The average probability of error based on the union bound is given by [Zie92]

$$P_e \leq (M-1) Q\left(\sqrt{\frac{E_b \log_2 M}{N_0}}\right) \tag{5.128}$$

For noncoherent detection using matched filters followed by envelope detectors, the average probability of error is given by [Zie92]

$$P_e = \sum_{k=1}^{M-1} \left(\frac{(-1)^{k+1}}{k+1}\right) \binom{M-1}{k} \exp\left(\frac{-kE_s}{(k+1)N_0}\right) \tag{5.129}$$

Using only the leading terms of the binomial expansion, the probability of error can be bounded as

$$P_e \leq \frac{M-1}{2} \exp\left(\frac{-E_s}{2N_0}\right) \tag{5.130}$$

The channel bandwidth of a coherent M-ary FSK signal may be defined as [Zie92]

$$B = \frac{R_b (M+3)}{2 \log_2 M} \tag{5.131}$$

and that of a noncoherent MFSK may be defined as

$$B = \frac{R_b M}{2 \log_2 M} \tag{5.132}$$

This implies that the bandwidth efficiency of an M-ary FSK signal decreases with increasing *M*. Therefore, unlike M-PSK signals, M-FSK signals are bandwidth inefficient. However, since all the *M* signals are orthogonal, there is no crowding in the signal space, and hence the power efficiency increases with *M*. Furthermore, M-ary FSK can be amplified using nonlinear amplifiers with no performance degradation. Table 5.6 provides a listing of bandwidth and power efficiency of M-FSK signals for various values of *M*.

**Table 5.6 Bandwidth and Power Efficiency of Coherent M-ary FSK [Zie92]**

<i>M</i>	2	4	8	16	32	64
$\eta_B$	0.4	0.57	0.55	0.42	0.29	0.18
$E_b/N_0$ for BER= $10^{-6}$	13.5	10.8	9.3	8.2	7.5	6.9

The orthogonality characteristic of MFSK has led researchers to explore Orthogonal Frequency Division Multiplexing (OFDM) as a means of providing power efficient signaling for a large number of users on the same channel. Each frequency in equation (5.127) is modulated with binary data (on/off) to provide a number of parallel carriers each containing a portion of user data.

### 5.10 Spread Spectrum Modulation Techniques

All of the modulation and demodulation techniques described so far strive to achieve greater power and/or bandwidth efficiency in a stationary additive white Gaussian noise channel. Since bandwidth is a limited resource, one of the primary design objectives of all the modulation schemes detailed thus far is to minimize the required transmission bandwidth. Spread spectrum techniques, on the other hand, employ a transmission bandwidth that is several orders of magnitude *greater* than the minimum required signal bandwidth. While this system is very bandwidth inefficient for a single user, the advantage of spread spectrum is that many users can simultaneously use the same bandwidth without significantly interfering with one another. In a multiple-user, *multiple access interference* (MAI) environment, spread spectrum systems become very bandwidth efficient.

Apart from occupying a very large bandwidth, spread spectrum signals are *pseudorandom* and have noise-like properties when compared with the digital information data. The spreading waveform is controlled by a *pseudo-noise (PN) sequence* or *pseudo-noise code*, which is a binary sequence that appears random but can be reproduced in a deterministic manner by intended receivers. Spread spectrum signals are demodulated at the receiver through crosscorrelation with a locally-generated version of the pseudorandom carrier. Crosscorrelation with the correct PN sequence *despreads* the spread spectrum signal and restores the modulated message in the same narrow band as the original data, whereas crosscorrelating the signal from an undesired user results in a very small amount of wideband noise at the receiver output.

Spread spectrum modulation has many properties that make it particularly well-suited for use in the mobile radio environment. The most important advantage is its inherent interference rejection capability. Since each user is assigned a unique PN code which is approximately orthogonal to the codes of other users, the receiver can separate each user based on their codes, even though they occupy the same spectrum at all times. This implies that, up to a certain number of users, interference between spread spectrum signals using the same frequency is negligible. Not only can a particular spread spectrum signal be recovered from a number of other spread spectrum signals, it is also possible to completely recover a spread spectrum signal even when it is jammed by a narrowband interferer. Since narrowband interference effects only a small portion of the spread spectrum signal, it can easily be removed through notch filtering without much loss of information. Since all users are able to share the same spectrum, spread spectrum may eliminate frequency planning, since all cells can use the same channels.

Resistance to multipath fading is another fundamental reason for considering spread spectrum systems for wireless communications. Chapter 4 showed that wideband signals are frequency selective. Since spread spectrum signals

have uniform energy over a very large bandwidth, at any given time only a small portion of the spectrum will undergo fading (recall the comparison between wideband and narrowband signal responses in multipath channels in Chapter 4). Viewed in the time domain, the multipath resistance properties are due to the fact that the delayed versions of the transmitted PN signal will have poor correlation with the original PN sequence, and will thus appear as another uncorrelated user which is ignored by the receiver. Spread spectrum systems are not only resistant to multipath fading, but they can also exploit the multipath components to improve the performance of the system. This can be done using a RAKE receiver which combines the information obtained from several resolvable multipath components. A RAKE receiver consists of a bank of correlators, each of which correlate to a particular multipath component of the desired signal. The correlator outputs may be weighted according to their relative strengths and summed to obtain the final signal estimate [Pri58]. RAKE receivers are described in Chapter 6.

### 5.10.1 Pseudo-noise (PN) Sequences

A pseudo-noise (PN) or pseudorandom sequence is a binary sequence with an autocorrelation that resembles, over a period, the autocorrelation of a random binary sequence. Its autocorrelation also roughly resembles the autocorrelation of band-limited white noise. Although it is deterministic, a pseudonoise sequence has many characteristics that are similar to those of random binary sequences, such as having a nearly equal number of 0s and 1s, very low correlation between shifted versions of the sequence, very low crosscorrelation between any two sequences, etc. The PN sequence is usually generated using sequential logic circuits. A feedback shift register, which is diagrammed in Figure 5.48, consists of consecutive stages of two state memory devices and feedback logic. Binary sequences are shifted through the shift registers in response to clock pulses, and the output of the various stages are logically combined and fed back as the input to the first stage. When the feedback logic consists of exclusive-OR gates, which is usually the case, the shift register is called a linear PN sequence generator.

The initial contents of the memory stages and the feedback logic circuit determine the successive contents of the memory. If a linear shift register reaches zero state at some time, it would always remain in the zero state, and the output would subsequently be all 0's. Since there are exactly  $2^m - 1$  nonzero states for an  $m$ -stage feedback shift register, the period of a PN sequence produced by a linear  $m$ -stage shift register cannot exceed  $2^m - 1$  symbols. A sequence of period  $2^m - 1$  generated by a linear feedback register is called a *maximal length* (ML) sequence. An excellent treatment of PN codes is given in [Coo86h].

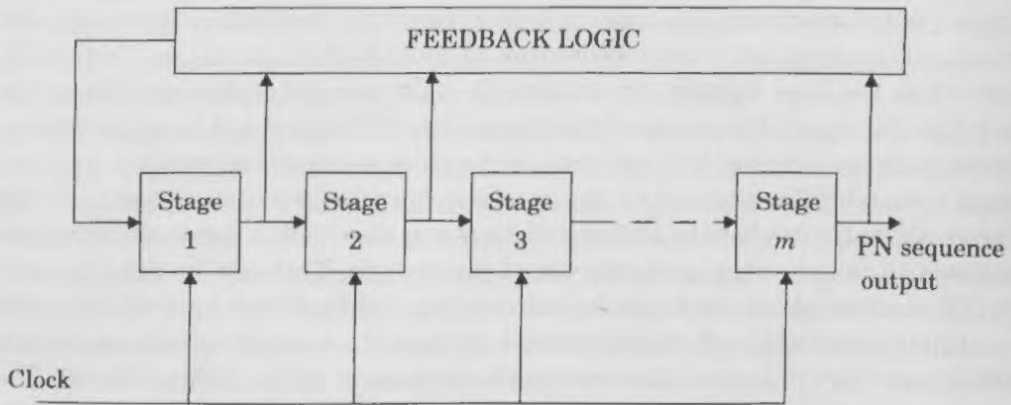


Figure 5.48  
Block diagram of a generalized feedback shift register with  $m$  stages.

### 5.10.2 Direct Sequence Spread Spectrum (DS-SS)

A *direct sequence spread spectrum* (DS-SS) system spreads the baseband data by directly multiplying the baseband data pulses with a pseudo-noise sequence that is produced by a pseudo-noise code generator. A single pulse or symbol of the PN waveform is called a *chip*. Figure 5.49 shows a functional block diagram of a DS system with binary phase modulation. This system is one of the most widely used direct sequence implementations. Synchronized data symbols, which may be information bits or binary channel code symbols, are added in modulo-2 fashion to the chips before being phase modulated. A coherent or differentially coherent phase-shift keying (PSK) demodulation may be used in the receiver.

The received spread spectrum signal for a single user can be represented as

$$S_{ss}(t) = \sqrt{\frac{2E_s}{T_s}} m(t) p(t) \cos(2\pi f_c t + \theta) \quad (5.133)$$

where  $m(t)$  is the data sequence,  $p(t)$  is the PN spreading sequence,  $f_c$  is the carrier frequency, and  $\theta$  is the carrier phase angle at  $t = 0$ . The data waveform is a time sequence of nonoverlapping rectangular pulses, each of which has an amplitude equal to +1 or -1. Each symbol in  $m(t)$  represents a data symbol and has duration  $T_s$ . Each pulse in  $p(t)$  represents a chip, is usually rectangular with an amplitude equal to +1 or -1, and has a duration of  $T_c$ . The transitions of the data symbols and chips coincide such that the ratio  $T_s$  to  $T_c$  is an integer. If  $W_{ss}$  is the bandwidth of  $S_{ss}(t)$  and  $B$  is the bandwidth of  $m(t)\cos(2\pi f_c t)$ , the spreading due to  $p(t)$  gives  $W_{ss} \gg B$ .

Figure 5.49(b) illustrates a DS receiver. Assuming that code synchronization has been achieved at the receiver, the received signal passes through the



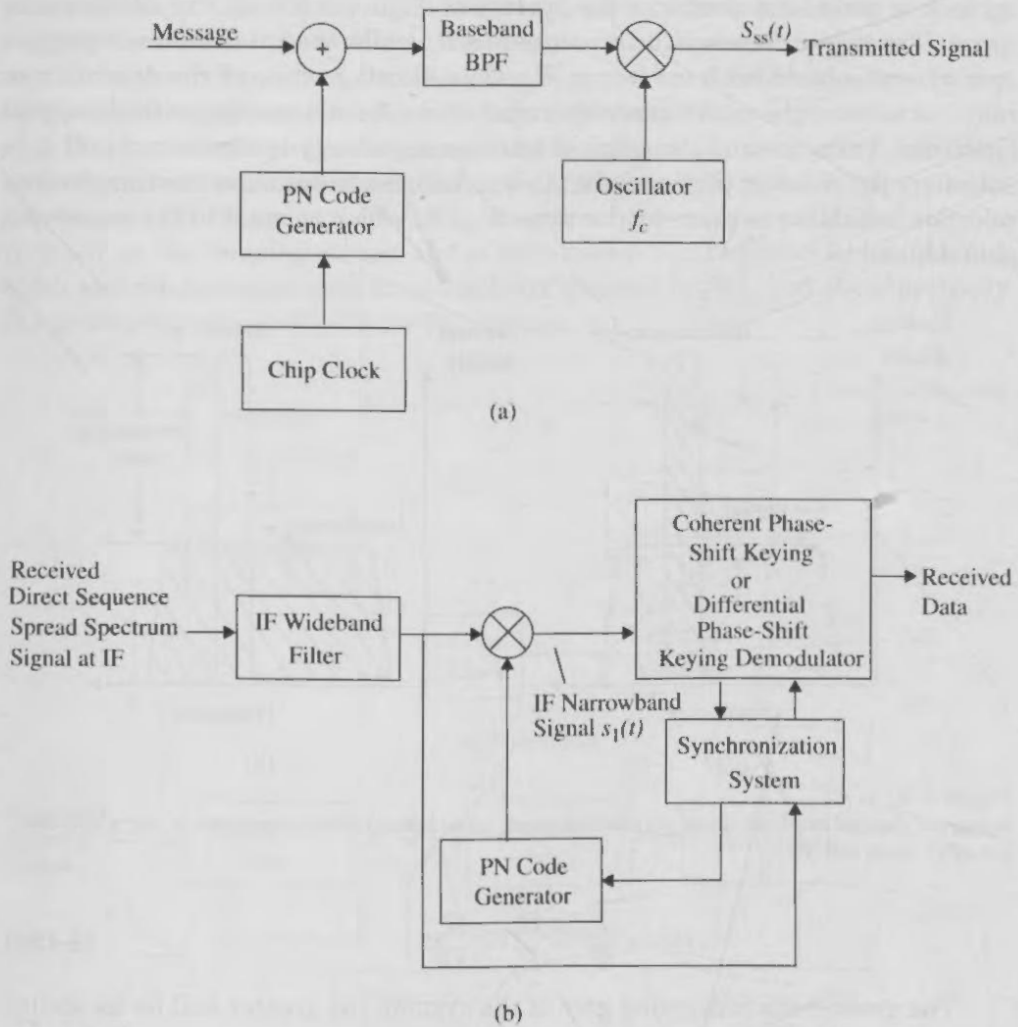


Figure 5.49 Block diagram of a DS-SS system with binary phase modulation: (a) transmitter and (b) receiver

wideband filter and is multiplied by a local replica of the PN code sequence  $p(t)$ . If  $p(t) = \pm 1$ , then  $p^2(t) = 1$ , and this multiplication yields the despread signal  $s(t)$  given by

$$s_1(t) = \sqrt{\frac{2E_s}{T_s}} m(t) \cos(2\pi f_c t + \theta) \tag{5.134}$$

at the input of the demodulator. Because  $s_1(t)$  has the form of a BPSK signal, the corresponding demodulation extracts  $m(t)$ .

Figure 5.50 shows the received spectra of the desired signal and the interference at the output of the receiver wideband filter. Multiplication by the

spreading waveform produces the spectra of Figure 5.50b at the demodulator input. The signal bandwidth is reduced to  $B$ , while the interference energy is spread over a bandwidth exceeding  $W_{ss}$ . The filtering action of the demodulator removes most of the interference spectrum that does not overlap with the signal spectrum. Thus, most of the original interference energy is eliminated and does not affect the receiver performance. An approximate measure of the interference rejection capability is given by the ratio  $W_{ss}/B$ , which is equal to the processing gain defined as

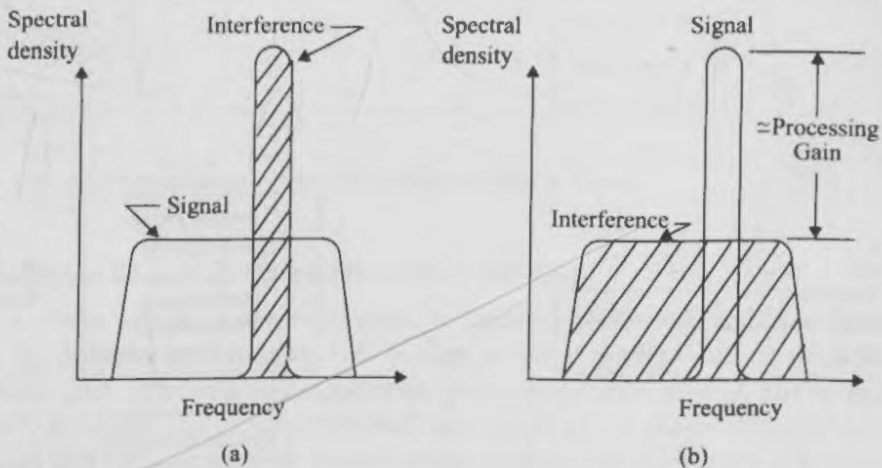


Figure 5.50

Spectra of desired received signal with interference: (a) wideband filter output and (b) correlator output after despreading.

$$PG = \frac{T_s}{T_c} = \frac{R_c}{R_s} = \frac{W_{ss}}{2R_s} \quad (5.135)$$

The greater the processing gain of the system, the greater will be its ability to suppress in-band interference.

### 5.10.3 Frequency Hopped Spread Spectrum (FH-SS)

Frequency hopping involves a periodic change of transmission frequency. A frequency hopping signal may be regarded as a sequence of modulated data bursts with time-varying, pseudorandom carrier frequencies. The set of possible carrier frequencies is called the *hopset*. Hopping occurs over a frequency band that includes a number of channels. Each channel is defined as a spectral region with a central frequency in the hopset and a bandwidth large enough to include most of the power in a narrowband modulation burst (usually FSK) having the corresponding carrier frequency. The bandwidth of a channel used in the hopset is called the *instantaneous bandwidth*. The bandwidth of the spectrum over

which the hopping occurs is called the *total hopping bandwidth*. Data is sent by hopping the transmitter carrier to seemingly random channels which are known only to the desired receiver. On each channel, small bursts of data are sent using conventional narrowband modulation before the transmitter hops again.

If only a single carrier frequency (single channel) is used on each hop, digital data modulation is called *single channel modulation*. Figure 5.51 shows a single channel FH-SS system. The time duration between hops is called the *hop duration* or the *hopping period* and is denoted by  $T_h$ . The total hopping bandwidth and the instantaneous bandwidth are denoted by  $W_{ss}$  and  $B$ , respectively. The processing gain =  $W_{ss}/B$  for FH systems.

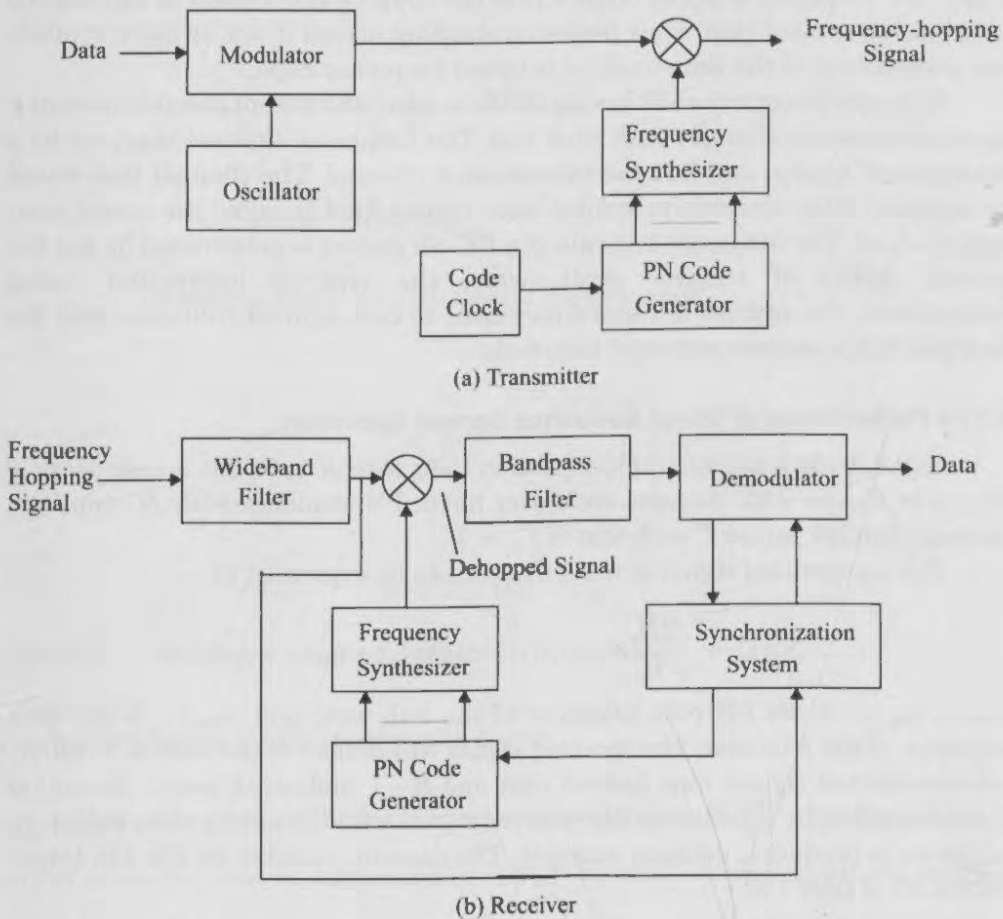


Figure 5.51 Block diagram of frequency hopping (FH) system with single channel modulation.

After frequency hopping has been removed from the received signal, the resulting signal is said to be *dehopped*. If the frequency pattern produced by the

receiver synthesizer in Figure 5.51(b) is synchronized with the frequency pattern of the received signal, then the mixer output is a dehopped signal at a fixed difference frequency. Before demodulation, the dehopped signal is applied to a conventional receiver. In FH, whenever an undesired signal occupies a particular hopping channel, the noise and interference in that channel are translated in frequency so that they enter the demodulator. Thus it is possible to have collisions in a FH system where an undesired user transmits in the same channel at the same time as the desired user.

Frequency hopping may be classified as fast or slow. *Fast frequency hopping* occurs if there is more than one frequency hop during each transmitted symbol. Thus, fast frequency hopping implies that the hopping rate equals or exceeds the information symbol rate. *Slow frequency hopping* occurs if one or more symbols are transmitted in the time interval between frequency hops.

If binary frequency-shift keying (FSK) is used, the pair of possible instantaneous frequencies changes with each hop. The frequency channel occupied by a transmitted symbol is called the *transmission channel*. The channel that would be occupied if the alternative symbol were transmitted is called the *complementary channel*. The frequency hop rate of a FH-SS system is determined by the frequency agility of receiver synthesizers, the type of information being transmitted, the amount of redundancy used to code against collisions, and the distance to the nearest potential interferer.

#### 5.10.4 Performance of Direct Sequence Spread Spectrum

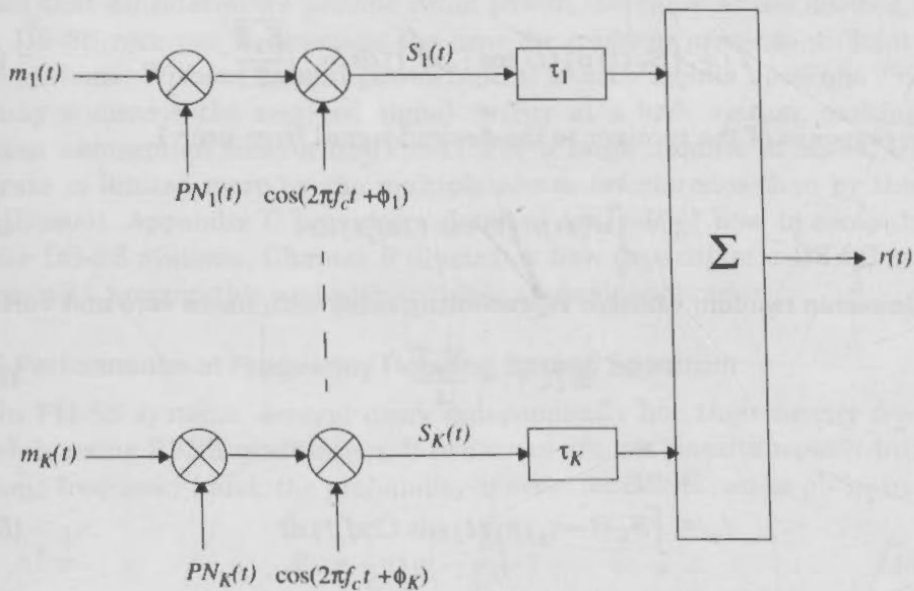
A direct sequence spread spectrum system with  $K$  multiple access users is shown in Figure 5.52. Assume each user has a PN sequence with  $N$  chips per message symbol period  $T$  such that  $NT_c = T$ .

The transmitted signal of the  $k$ th user can be expressed as

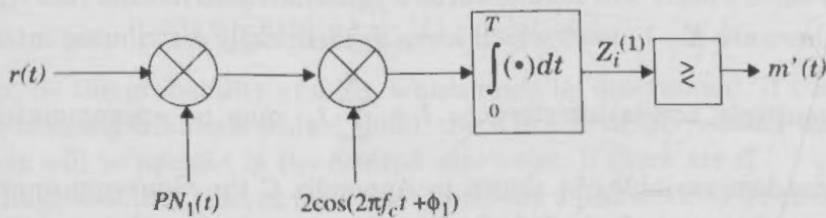
$$S_k(t) = \sqrt{\frac{2E_s}{T_s}} m_k(t) p_k(t) \cos(2\pi f_c t + \phi_k) \quad (5.136)$$

where  $p_k(t)$  is the PN code sequence of the  $k$ th user, and  $m_k(t)$  is the data sequence of the  $k$ th user. The received signal will consist of the sum of  $K$  different transmitted signals (one desired user and  $K - 1$  undesired users). Reception is accomplished by correlating the received signal with the appropriate signature sequence to produce a decision variable. The decision variable for the  $i$ th transmitted bit of user 1 is

$$Z_i^{(1)} = \int_{(i-1)T + \tau_1}^{iT + \tau_1} r(t) p_1(t - \tau_1) \cos[2\pi f_c(t - \tau_1) + \phi_1] dt \quad (5.137)$$



(a) Model of  $K$  users in a CDMA spread spectrum system



(b) Receiver structure for user 1

Figure 5.52  
A simplified diagram of a DS-SS system with  $K$  users.

If  $m_{1,i} = -1$ , then the bit will be received in error if  $Z_i^{(1)} > 0$ . The probability of error can now be calculated as  $Pr[Z_i^{(1)} > 0 | m_{1,i} = -1]$ . Since the received signal  $r(t)$  is a linear combination of signals, equation (5.137) can be rewritten as

$$Z_i^{(1)} = I_1 + \sum_{k=2}^K I_k + \xi \tag{5.138}$$

where

$$I_1 = \int_0^T S_1(t) p_1(t) \cos(2\pi f_c t) dt = \sqrt{\frac{E_s T}{2}} \quad (5.139)$$

is the response of the receiver to the desired signal from user 1.

$$\xi = \int_0^T n(t) p_1(t) \cos(2\pi f_c t) dt \quad (5.140)$$

is a Gaussian random variable representing noise with mean zero and variance

$$E[\xi^2] = \frac{N_0 T}{4} \quad (5.141)$$

and

$$I_k = \int_0^T S_k(t - \tau_k) p_1(t) \cos(2\pi f_c t) dt \quad (5.142)$$

represents the multiple access interference from user  $k$ . Assuming that  $I_k$  is composed of the cumulative effects of  $N$  random chips from the  $k$ th interferer over the integration period  $T$  of one bit, the central limit theorem implies that the sum of these effects will tend towards a Gaussian distribution (see Appendix C). Since there are  $K - 1$  users which serve as identically distributed interferers,

the total multiple access interference  $I = \sum_{k=2}^K I_k$  may be approximated by a

Gaussian random variable. As shown in Appendix C the Gaussian approximation assumes that each  $I_k$  is independent, but in actuality they are not. The Gaussian approximation yields a convenient expression for the average probability of bit error given by

$$P_e = Q \left( \frac{1}{\sqrt{\frac{K-1}{3N} + \frac{N_0}{2E_b}}} \right) \quad (5.143)$$

For a single user,  $K = 1$ , this expression reduces to the BER expression for BPSK modulation. For the interference limited case where thermal noise is not a factor,  $E_b/N_0$  tends to infinity, and the BER expression has a value equal to

$$P_e = Q \left( \sqrt{\frac{3N}{K-1}} \right) \quad (5.144)$$

This is the irreducible error floor due to multiple access interference and assumes that all interferers provide equal power, the same as the desired user, at the DS-SS receiver. In practice, the *near-far problem* presents difficulty for DS-SS systems. Without careful power control of each mobile user, one close-in user may dominate the received signal energy at a base station, making the Gaussian assumption inaccurate [Pic91]. For a large number of users, the bit error rate is limited more by the multiple access interference than by thermal noise [Bue94]. Appendix C provides a detailed analysis of how to compute the BER for DS-SS systems. Chapter 8 illustrates how capacity of a DS-SS system changes with propagation and with multiple access interference.

### 5.10.5 Performance of Frequency Hopping Spread Spectrum

In FH-SS systems, several users independently hop their carrier frequencies while using BFSK modulation. If two users are not simultaneously utilizing the same frequency band, the probability of error for BFSK can be given by

$$P_e = \frac{1}{2} \exp\left(-\frac{E_b}{2N_0}\right) \quad (5.145)$$

However, if two users transmit simultaneously in the same frequency band, a collision, or "hit", occurs. In this case it is reasonable to assume that the probability of error is 0.5. Thus the overall probability of bit error can be modeled as

$$P_e = \frac{1}{2} \exp\left(-\frac{E_b}{2N_0}\right) (1 - p_h) + \frac{1}{2} p_h \quad (5.146)$$

where  $p_h$  is the probability of a hit, which must be determined. If there are  $M$  possible hopping channels (called slots), there is a  $1/M$  probability that a given interferer will be present in the desired user's slot. If there are  $K-1$  interfering users, the probability that at least one is present in the desired frequency slot is

$$p_h = 1 - \left(1 - \frac{1}{M}\right)^{K-1} \approx \frac{K-1}{M} \quad (5.147)$$

assuming  $M$  is large. Substituting this in equation (5.146) gives

$$P_e = \frac{1}{2} \exp\left(-\frac{E_b}{2N_0}\right) \left(1 - \frac{K-1}{M}\right) + \frac{1}{2} \left[\frac{K-1}{M}\right] \quad (5.148)$$

Now consider the following special cases. If  $K = 1$ , the probability of error reduces to equation (5.145), the standard probability of error for BFSK. Also, if  $E_b/N_0$  approaches infinity,

$$\lim_{\frac{E_b}{N_0} \rightarrow \infty} (P_e) = \frac{1}{2} \left[\frac{K-1}{M}\right] \quad (5.149)$$

which illustrates the irreducible error rate due to multiple access interference.

The previous analysis assumes that all users hop their carrier frequencies synchronously. This is called *slotted frequency hopping*. This may not be a realistic scenario for many FH-SS systems. Even when synchronization can be achieved between individual user clocks, radio signals will not arrive synchronously to each user due to the various propagation delays. As described by Geraniotis [Ger82], the probability of a hit in the asynchronous case is

$$P_h = 1 - \left\{ 1 - \frac{1}{M} \left( 1 + \frac{1}{N_b} \right) \right\}^{K-1} \quad (5.150)$$

where  $N_b$  is the number of bits per hop. Comparing equation (5.150) to (5.147) we see that for the asynchronous case, the probability of a hit is increased (this would be expected). Using (5.150) in (5.146), the probability of error for the asynchronous FH-SS case is

$$P_e = \frac{1}{2} \exp\left(-\frac{E_b}{N_0}\right) \left\{ 1 - \frac{1}{M} \left( 1 + \frac{1}{N_b} \right) \right\}^{K-1} + \frac{1}{2} \left[ 1 - \left\{ 1 - \frac{1}{M} \left( 1 + \frac{1}{N_b} \right) \right\}^{K-1} \right] \quad (5.151)$$

FH-SS has an advantage over DS-SS in that it is not as susceptible to the near-far problem. Because signals are generally not utilizing the same frequency simultaneously, the relative power levels of signals are not as critical as in DS-SS. The near-far problem is not totally avoided, however, since there will be some interference caused by stronger signals bleeding into weaker signals due to imperfect filtering of adjacent channels. To combat the occasional hits, error-correction coding is required on all transmissions. By applying strong Reed-Solomon or other burst error correcting codes, performance can be increased dramatically, even with an occasional collision.

### 5.11 Modulation Performance in Fading and Multipath Channels

As discussed in Chapter 3 and 4, the mobile radio channel is characterized by various impairments such as fading, multipath, and Doppler spread. In order to study the effectiveness of any modulation scheme in a mobile radio environment, it is required to evaluate the performance of the modulation scheme over such channel conditions. Although bit error rate (BER) evaluation gives a good indication of the performance of a particular modulation scheme, it does not provide information about the type of errors. For example, it does not give incidents of bursty errors. In a fading mobile radio channel, it is likely that a transmitted signal will suffer deep fades which can lead to outage or a complete loss of the signal.

Evaluating the *probability of outage* is another means to judge the effectiveness of the signaling scheme in a mobile radio channel. An outage event is specified by a specific number of bit errors occurring in a given transmission. Bit error rates and probability of outage for various modulation schemes under various types of channel impairments can be evaluated either through analytical



techniques or through simulations. While simple analytical techniques for computing bit error rates in slow, flat fading channels exist, performance evaluation in frequency selective channels and computation of outage probabilities are often made through computer simulations. Computer simulations are based on convolving the input bit stream with a suitable channel impulse response model and counting the bit errors at the output of the receiver decision circuit [Rap91b], [Fun93].

Before a study of the performance of various modulation schemes in multipath and fading channels is made, it is imperative that a thorough understanding of the channel characteristics be obtained. The channel models described in Chapter 4 may be used in the evaluation of various modulation schemes.

### 5.11.1 Performance of Digital Modulation in Slow, Flat Fading Channels

As discussed in Chapter 4, flat fading channels cause a multiplicative (gain) variation in the transmitted signal  $s(t)$ . Since slow, flat fading channels change much slower than the applied modulation, it can be assumed that the attenuation and phase shift of the signal is constant over at least one symbol interval. Therefore, the received signal  $r(t)$  may be expressed as

$$r(t) = \alpha(t)\exp(-j\theta(t))s(t) + n(t) \quad 0 \leq t \leq T \quad (5.152)$$

where  $\alpha(t)$  is the gain of the channel,  $\theta(t)$  is the phase shift of the channel, and  $n(t)$  is additive Gaussian noise.

Depending on whether it is possible to make an accurate estimate of the phase  $\theta(t)$ , coherent or noncoherent matched filter detection may be employed at the receiver.

To evaluate the probability of error of any digital modulation scheme in a slow, flat fading channel, one must average the probability of error of the particular modulation in AWGN channels over the possible ranges of signal strength due to fading. In other words, the probability of error in AWGN channels is viewed as a conditional error probability, where the condition is that  $\alpha$  is fixed. Hence, the probability of error in slow, flat fading channels can be obtained by averaging the error in AWGN channels over the fading probability density function. In doing so, the probability of error in a slow, flat fading channel can be evaluated as

$$P_e = \int_0^{\infty} P_e(X)p(X)dX \quad (5.153)$$

$P_e(X)$  is the probability of error for an arbitrary modulation at a specific value of signal-to-noise ratio  $X$ , where  $X = \alpha^2 E_b/N_0$  and  $p(X)$  is the probability density function of  $X$  due to the fading channel.  $E_b$  and  $N_0$  are constants, and the random variable  $\alpha$  is used to represent amplitude values of the fading channel, with respect to  $E_b/N_0$ .

For Rayleigh fading channels,  $\alpha$  has a Rayleigh distribution, so  $\alpha^2$  and consequently  $X$  have a chi-square distribution with two degrees of freedom. Therefore,

$$p(X) = \frac{1}{\Gamma} \exp\left(-\frac{X}{\Gamma}\right) \quad X \geq 0 \quad (5.154)$$

where  $\Gamma = \frac{E_b}{N_0} \alpha^{-2}$  is the average value of the signal-to-noise ratio.

By using equation (5.154) and the probability of error of a particular modulation scheme in AWGN, the probability of error in a slow, flat fading channel can be evaluated. It can be shown that for coherent binary PSK and coherent binary FSK, equation (5.153) evaluates to [Ste87]

$$P_{e, \text{PSK}} = \frac{1}{2} \left[ 1 - \sqrt{\frac{\Gamma}{1+\Gamma}} \right] \quad (\text{coherent binary PSK}) \quad (5.155)$$

$$P_{e, \text{FSK}} = \frac{1}{2} \left[ 1 - \sqrt{\frac{\Gamma}{2+\Gamma}} \right] \quad (\text{coherent binary FSK}) \quad (5.156)$$

It can also be shown that the average error probability of DPSK and orthogonal noncoherent FSK in a slow, flat, Rayleigh fading channel are given by

$$P_{e, \text{DPSK}} = \frac{1}{2(1+\Gamma)} \quad (\text{differential binary PSK}) \quad (5.157)$$

$$P_{e, \text{NCFSK}} = \frac{1}{2+\Gamma} \quad (\text{noncoherent orthogonal binary FSK}) \quad (5.158)$$

Figure 5.53 illustrates how the BER for various modulations changes as a function of  $E_b/N_0$  in a Rayleigh, flat fading environment. The figure was produced using simulation instead of analysis, but agrees closely with equations (5.155) to (5.158) [Rap91b].

For large values of  $E_b/N_0$  (i.e., large values of  $X$ ) the error probability equations may be simplified as

$$P_{e, \text{PSK}} = \frac{1}{4\Gamma} \quad (\text{coherent binary PSK}) \quad (5.159)$$

$$P_{e, \text{FSK}} = \frac{1}{2\Gamma} \quad (\text{coherent FSK}) \quad (5.160)$$

$$P_{e, \text{DPSK}} = \frac{1}{2\Gamma} \quad (\text{Differential PSK}) \quad (5.161)$$

$$P_{e, \text{NCFSK}} = \frac{1}{\Gamma} \quad (\text{noncoherent orthogonal binary FSK}) \quad (5.162)$$

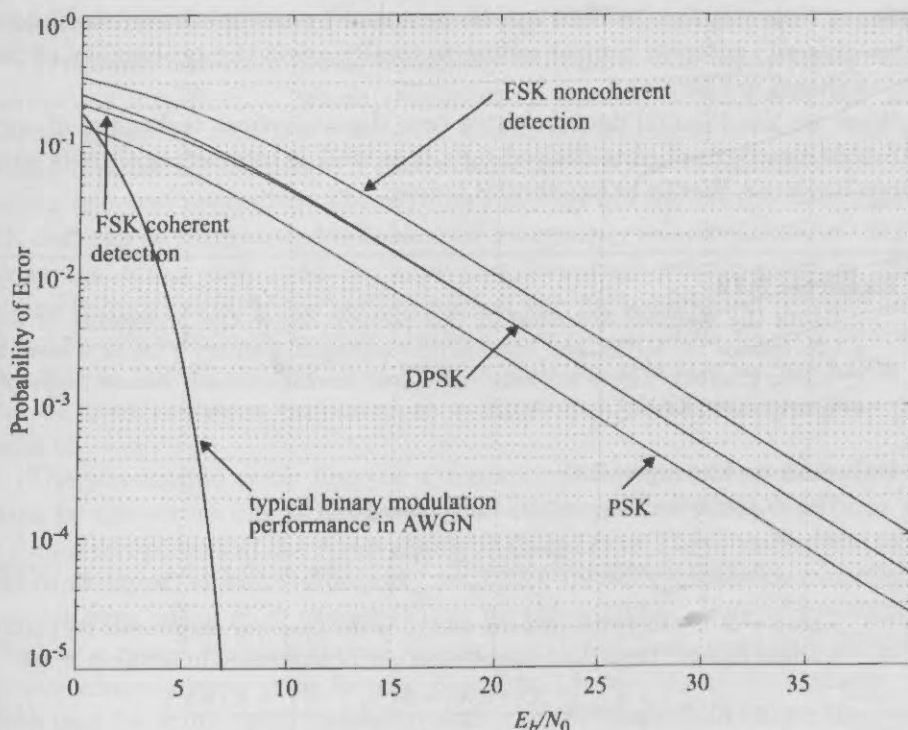


Figure 5.53 Bit error rate performance of binary modulation schemes in a Rayleigh, flat fading channel as compared to a typical performance curve in AWGN.

For GMSK, the expression for BER in the AWGN channel is given in equation (5.111.a) which when evaluated in equation (5.153) yields a Rayleigh fading BER of

$$P_{e, \text{GMSK}} = \frac{1}{2} \left( 1 - \sqrt{\frac{\delta \Gamma}{\delta \Gamma + 1}} \right) \cong \frac{1}{4\delta \Gamma} \quad (\text{coherent GMSK}) \quad (5.163)$$

where 
$$\delta \cong \begin{cases} 0.68 & \text{for } BT = 0.25 \\ 0.85 & \text{for } BT = \infty \end{cases} \quad (5.164)$$

As seen from equations (5.159) to (5.163), for lower error rates all five modulation techniques exhibit an inverse algebraic relation between error rate and mean SNR. This is in contrast with the exponential relationship between error rate and SNR in an AWGN channel. According to these results, it is seen that operating at BERs of  $10^{-3}$  to  $10^{-6}$  requires roughly a 30 dB to 60 dB mean SNR. This is significantly larger than that required when operating over a nonfading Gaussian noise channel (20 dB to 50 dB more link is required). However, it can easily be shown that the poor error performance is due to the non-zero probability of very deep fades, when the instantaneous BER can become as low as 0.5.

Significant improvement in BER can be achieved by using efficient techniques such as diversity or error control coding to totally avoid the probability of deep fades, as shown in Chapter 6.

Work by Yao [Yao92] demonstrates how the analytical technique of equation (5.153) may be applied to desired signals as well as interfering signals which undergo Rayleigh, Rician or log-normal fading.

### Example 5.12

Using the methods described in this section, derive the probability of error expressions for DPSK and noncoherent orthogonal binary FSK in a slow, flat fading channel, where the received signal envelope has a Rician probability distribution [Rob94].

### Solution to Example 5.12

The Rician probability density function is given by

$$p(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + A^2}{2\sigma^2}\right) I_0\left(\frac{Ar}{\sigma^2}\right) \quad \text{for } A \geq 0, r \geq 0 \quad (\text{E5.1})$$

where  $r$  is the Rician amplitude and  $A$  is the specular amplitude. By suitable transformation, the Rician distribution can be expressed in terms of  $X$  as

$$p(X) = \frac{1+K}{\Gamma} \exp\left(-\frac{X(1+K) + K\Gamma}{\Gamma}\right) I_0\left(\sqrt{\frac{4(1+K)KX}{\Gamma}}\right) \quad (\text{E5.2})$$

where  $K = A^2/2\sigma^2$  is the specular-to-random ratio of the Rician distribution. The probability of error of DPSK and noncoherent orthogonal FSK in an AWGN channel can be expressed as

$$P_e(X, k_1, k_2) = k_1 \exp(-k_2 X) \quad (\text{E5.3})$$

where  $k_1 = k_2 = 1/2$ , for FSK and  $k_1 = 1/2, k_2 = 1$  for DPSK.

To obtain the probability of error in a slow, flat fading channel we need to evaluate the expression

$$P_e = \int_0^\infty P_e(X) p(X) dX \quad (\text{E5.4})$$

Substituting equations (E5.2) and (E5.3) in equation (E5.4) and solving the integration we get the probability of error in a Rician, slow, flat fading channel as

$$P_e = \frac{k_1(1+K)}{(k_2\Gamma + 1 + K)} \exp\left(\frac{-k_2K\Gamma}{k_2\Gamma + 1 + K}\right)$$

Substituting  $k_1 = k_2 = 1/2$ , for FSK, the probability of error is given by

$$P_{e, \text{NCFSK}} = \frac{(1+K)}{(\Gamma + 2 + 2K)} \exp\left(\frac{-K\Gamma}{\Gamma + 2 + 2K}\right)$$

Similarly, for DPSK, substituting  $k_1 = 1/2, k_2 = 1$ , we get

$$P_{e, \text{DPSK}} = \frac{(1+K)}{2(\Gamma + 1 + K)} \exp\left(\frac{-K\Gamma}{\Gamma + 1 + K}\right)$$

### 5.11.2 Digital Modulation in Frequency Selective Mobile Channels

Frequency selective fading caused by multipath time delay spread causes intersymbol interference, which results in an irreducible BER floor for mobile systems. However, even if a mobile channel is not frequency selective, the time-varying Doppler spread due to motion creates an irreducible BER floor due to the random spectral spreading. These factors impose bounds on the data rate and BER that can be transmitted reliably over a frequency selective channel. Simulation is the major tool used for analyzing frequency selective fading effects. Chuang [Chu87] studied the performance of various modulation schemes in frequency selective fading channels through simulations. Both filtered and unfiltered BPSK, QPSK, OQPSK, and MSK modulation schemes were studied, and their BER curves were simulated as a function of the normalized rms delay spread ( $d = \sigma_\tau / T_s$ ).

The irreducible error floor in a frequency selective channel is primarily caused by the errors due to the intersymbol interference, which interferes with the signal component at the receiver sampling instants. This occurs when (a) the main (undelayed) signal component is removed through multipath cancellation, (b) a non-zero value of  $d$  causes ISI, or (c) the sampling time of a receiver is shifted as a result of delay spread. Chuang observed that errors in a frequency selective channel tend to be bursty. Based on the results of simulations, it is known that for small delay spreads (relative to the symbol duration), the resulting flat fading is the dominant cause of error bursts. For large delay spread, timing errors and ISI are the dominant error mechanisms.

Figure 5.54 shows the average irreducible BER as a function of  $d$  for different unfiltered modulation schemes using coherent detection. From the figure, it is seen that the BER performance of BPSK is the best among all the modulation schemes compared. This is because symbol offset interference (called *cross-rail* interference due to the fact that the eye diagram has multiple rails) does not exist in BPSK. Both OQPSK and MSK have a  $T/2$  timing offset between two bit sequences, hence the cross-rail ISI is more severe, and their performances are similar to QPSK. Figure 5.55 shows the BER as a function of rms delay spread normalized to the bit period ( $d' = \sigma_\tau / T_b$ ) rather than the symbol period as used in Figure 5.54. By comparing on a bit, rather than symbol basis, it becomes easier to compare different modulations. This is done in Figure 5.55, where it is clear that 4-level modulations (QPSK, OQPSK, and MSK) are more resistant to delay spread than BPSK for constant information throughput. Interestingly, 8-ary keying has been found to be less resistant than 4-ary keying, and this has led to the choice of 4-ary keying for all 3rd generation wireless standards described in Chapter 10.

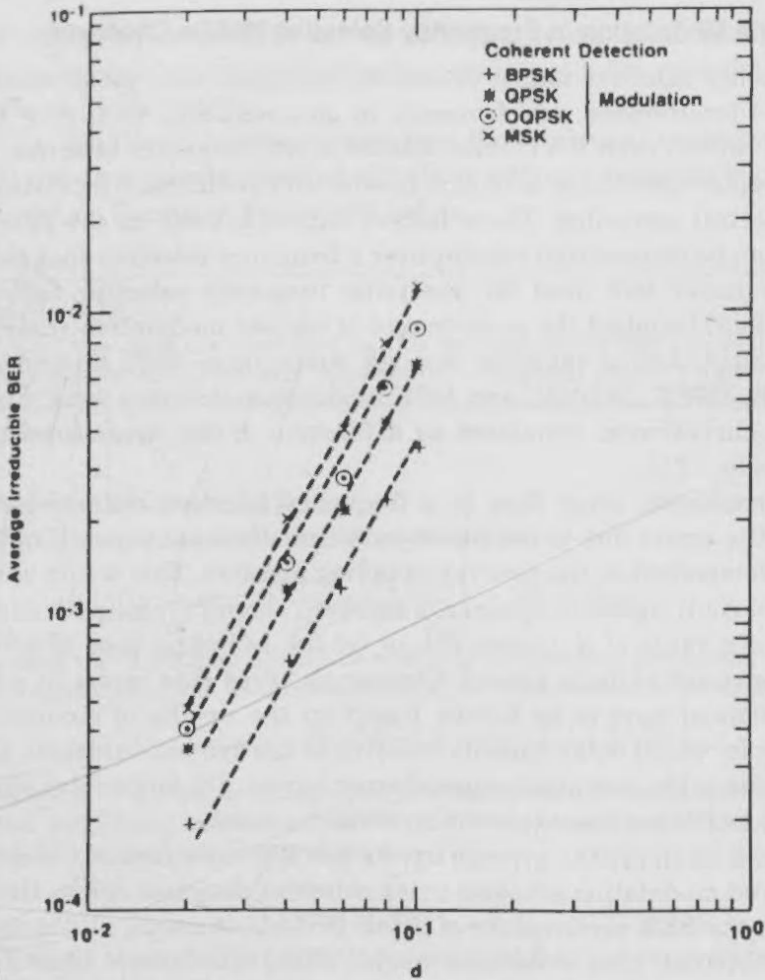


Figure 5.54

The irreducible BER performance for different modulations with coherent detection for a channel with a Gaussian shaped power delay profile. The parameter  $d$  is the rms delay spread normalized by the symbol period [From [Chu87] © IEEE].

### 5.11.3 Performance of $\pi/4$ DQPSK in Fading and Interference

Liu and Feher [Liu89], [Liu91] and Fung, Thoma, and Rappaport [Rap91b], [Fun93] studied the performance of  $\pi/4$  DQPSK in the mobile radio environment. They modeled the channel as a frequency selective, 2-ray, Rayleigh fading channel with additive Gaussian noise and co-channel interference (CCI). In addition, Thoma studied the effects of real-world multipath channel data, and discovered that sometimes such channels induce poorer bit error rates than the 2-ray Rayleigh fading model. Based on the analysis and simulation results, numerical computations were carried out to evaluate the bit error rates at different multipath delays between the two rays, at different vehicle speeds (i.e., different Dop-

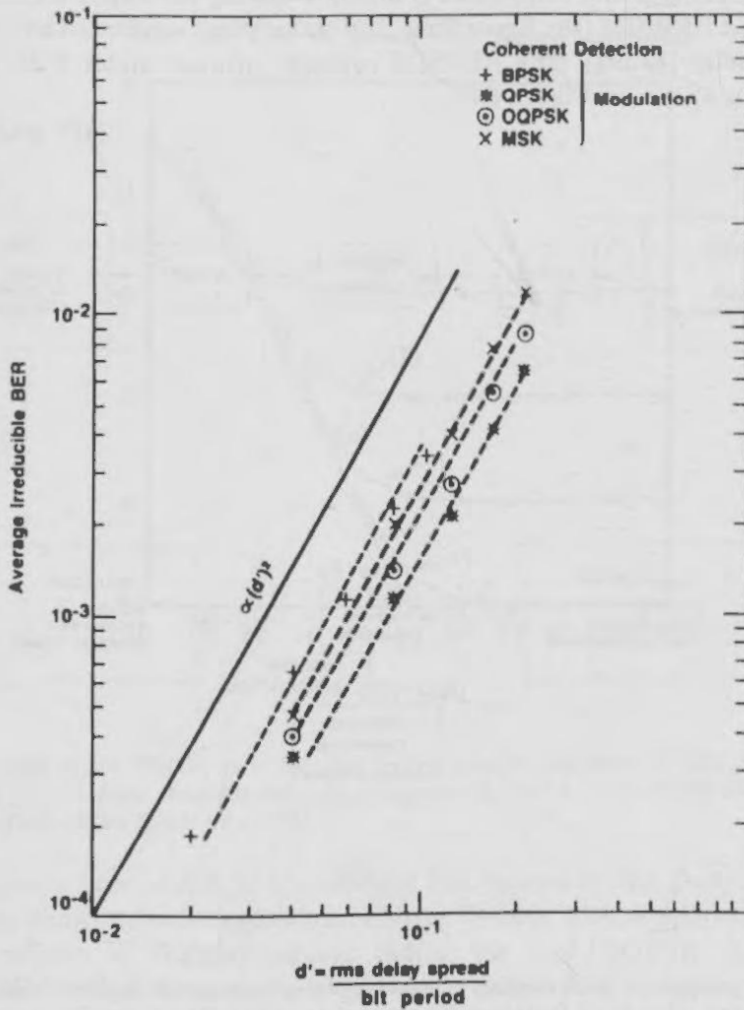


Figure 5.55

The same set of curves as plotted in Figure 5.54 plotted as a function of rms delay spread normalized by bit period [From [Chu87] © IEEE].

pler shifts), and various co-channel interference levels. BER was calculated and analyzed as a function of the following parameters:

- The Doppler spread normalized to the symbol rate:  $B_D T_s$  or  $B_D / R_s$
- Delay of the second multipath  $\tau$ , normalized to the symbol duration:  $\tau / T$
- Ratio of average carrier energy to noise power spectral density in decibels:  $E_b / N_0$  dB
- Average carrier to interference power ratio in decibels:  $C / I$  dB
- Average main-path to delayed-path power ratio:  $C / D$  dB

Fung, Thoma, and Rappaport [Fun93] [Rap91b] developed a computer simulator called BERSIM (Bit Error Rate SIMulator) that confirmed the analysis by Liu and Feher [Liu91]. The BERSIM concept, covered under U.S. patent No. 5,233,628, is shown in Figure 5.56.

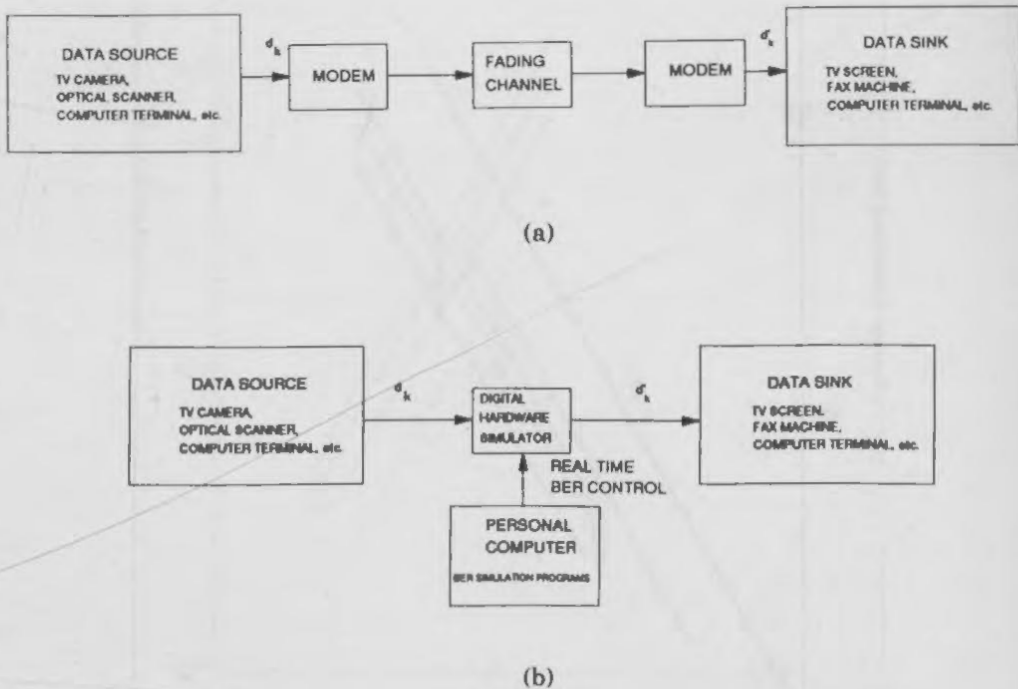


Figure 5.56

The BERSIM concept. (a) Block diagram of actual digital communication system. (b) Block diagram of BERSIM using a baseband digital hardware simulator with software simulation as a driver for real-time BER control.

Figure 5.57 shows a plot of the average probability of error of a U.S. digital cellular  $\pi/4$  DQPSK system as a function of carrier-to-noise ratio ( $C/N$ ) for different co-channel interference levels in a slow, flat, Rayleigh fading channel. In a slow, flat fading channel, the multipath time dispersion and Doppler spread are negligible, and errors are caused mainly by fades and co-channel interference. It is clearly seen that for  $C/I > 20$  dB, the errors are primarily due to fading, and interference has little effect. However, as  $C/I$  drops to below 20 dB, interference dominates the link performance. This is why high-capacity mobile systems are interference limited, not noise limited.

In mobile systems, even if there is no time dispersion and if  $C/N$  is infinitely large, the BER does not decrease below a certain irreducible floor. This



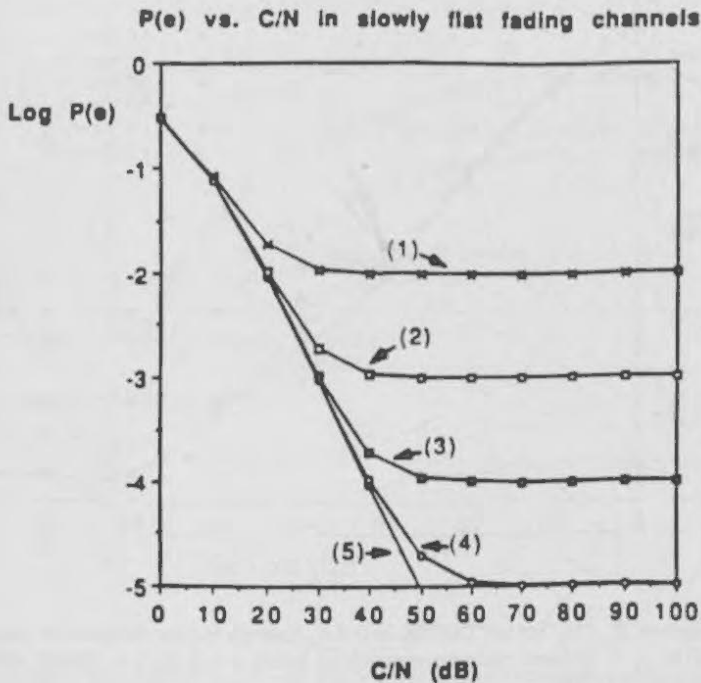


Figure 5.57

BER performance of  $\pi/4$  DQPSK in a flat slow fading channel corrupted by CCI and AWGN.  $f_c = 850\text{MHz}$ ,  $f_s = 24\text{kpsps}$ , raised cosine roll-off factor = 0.2,  $C/I =$  (1) 20 dB, (2) 30dB (3) 40dB (4) 50dB (5) infinity [From [Liu91]. © IEEE].

irreducible error floor is due to the random FM caused by the Doppler spread, and was first shown to exist by Bello and Nelin [Bel62]. Figure 5.58 clearly illustrates the effects of Doppler-induced fading for  $\pi/4$  DQPSK. As velocity increases, the irreducible error floor increases, despite an increase in  $E_b/N_0$ . Thus, once a certain  $E_b/N_0$  is achieved, there can be no further improvement in link performance due to motion.

Figure 5.59 shows the BER of the U.S. digital cellular  $\pi/4$ -DQPSK system in a Rayleigh, 2-ray fading channel for vehicle speeds of 40 km/hr and 120 km/hr, and with extremely large SNR ( $E_b/N_0 = 100$  dB). The curves are plotted for  $C/D = 0$  and 10 dB, and for various values of  $\tau/T_s$ . When the delay between the two multipath rays approaches 20% of the symbol period (i.e.,  $\tau/T = 0.2$ ), the bit error rate due to multipath rises above  $10^{-2}$  and makes the link unusable, even when the delayed ray has an average signal 10 dB below the main ray. The average bit error rate in the channel is important for speech coders. As a general rule,  $10^{-2}$  channel bit error rate is needed for modem speech coders to work properly. Notice also that at  $\tau/T = 0.1$ , the probability of error is well below  $10^{-2}$ , even when the second multipath component is equal in power to the first. Figure

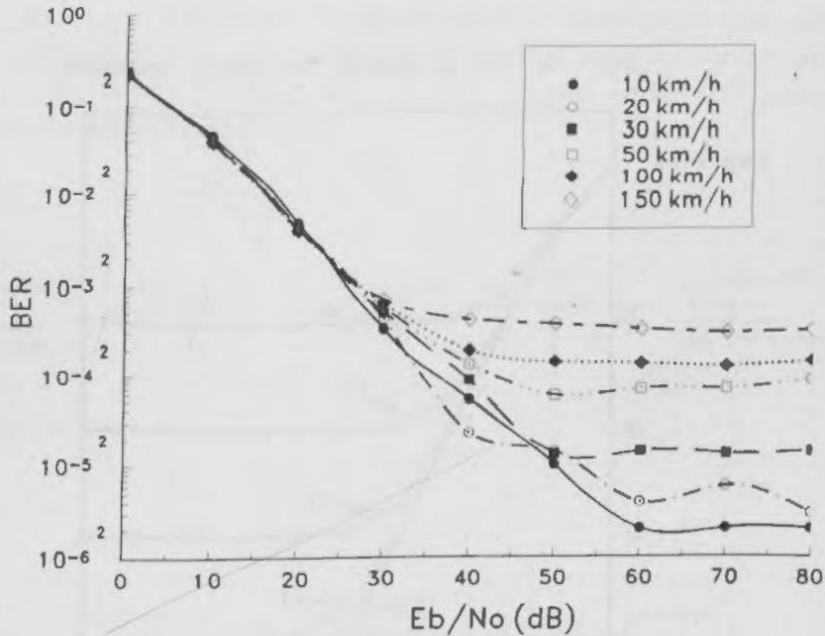


Figure 5.58

BER performance versus  $E_b/N_0$  for  $\pi/4$  DQPSK in a flat, Rayleigh fading channel for various mobile speeds:  $f_c = 850$  MHz,  $f_s = 24$  kbps, raised cosine rolloff factor is 0.2,  $C/I = 100$  dB. Generated by BERSIM [From [Fun93] © IEEE].

5.59 shows how the delay and amplitude of the second ray have strong impact on the average BER.

Simulations such as those conducted by [Chu87], [Liu91], [Rap91b], and [Fun93] provide valuable insight into the bit error mechanisms in a wide range of operating conditions. The mobile velocity, channel delay spread, interference levels, and modulation format all independently impact the raw bit error rate in mobile communications systems, and simulation is a powerful way of designing or predicting the performance of wireless communication links in very complex, time-varying channel conditions.

## 5.12 Problems

- 5.1 A frequency modulated wave has an angular carrier frequency  $\omega_c = 5000$  rad/sec and a modulation index of 10. Compute the bandwidth and the upper and lower sideband frequencies if the modulating signal  $m(t) = 20 \cos(5t)$ .
- 5.2 A 2 MHz carrier with an amplitude of 4 volts is frequency modulated by a modulating signal  $m(t) = \sin(1000\pi t)$ . The amplitude of the modulating signal is 2 volts and the peak frequency deviation was found to be 1 kHz. If the amplitude and frequency of the modulating signal are increased to 8 volts and 2 kHz respectively, write an expression for the new modulated signal.

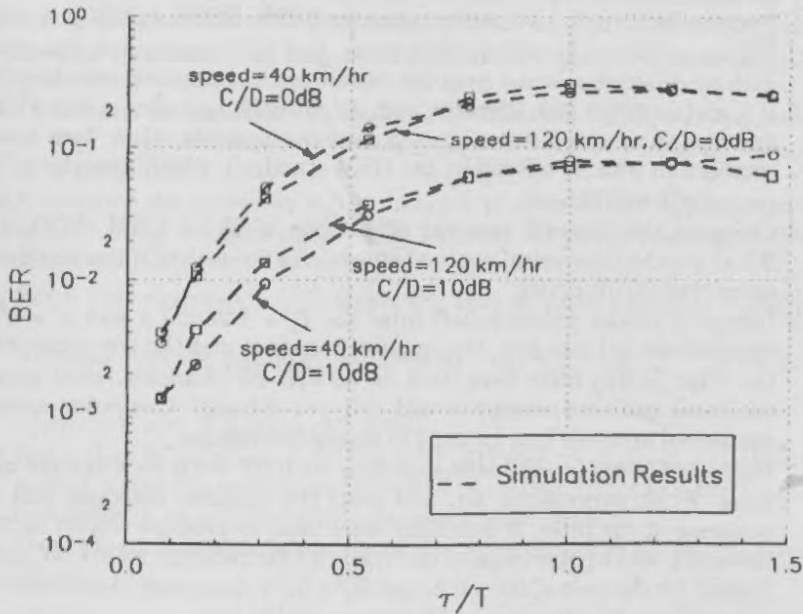


Figure 5.59

BER performance of  $\pi/4$  DQPSK in a 2-ray, Rayleigh fading channel, where the time delay  $\tau$ , and the power ratio  $C/D$  between the first and second ray are varied.  $f_c = 850$  MHz,  $f_s = 24$  kbps, raised cosine rolloff rate is 0.2,  $v = 40$  km/hr, 120 km/hr,  $E_b/N_0 = 100$  dB. Produced by BERSIM [From [Fun93] © IEEE].

- 5.3 (a) Generate and plot the time waveform for a binary baseband communication system which sends the bit string 1, 0, 1 through an ideal channel, where a raised cosine rolloff filter response is used having  $\alpha = 1/2$ . Assume the symbol rate is 50 kbps, and use a truncation time of  $\pm 6$  symbols.  
 (b) Illustrate the ideal sampling points of the waveform.  
 (c) If a receiver has a timing jitter of  $\pm 10^{-6}$  seconds, what is the difference in detected voltage as compared to ideal voltage at each sample point?
- 5.4 If  $f_c = 440$  MHz and vehicles travel at a maximum speed of 80 mph, determine the proper spectrum allocation of voice and tone signals for a TTIB system.
- 5.5 If  $f_d = 12$  kHz and  $W = 4$  kHz, determine the modulation index for an FM transmitter. These are parameters for the AMPS standard.
- 5.6 For AMPS FM transmissions, if the  $SNR_{in} = 10$  dB, determine the  $SNR_{out}$  of the FM detector. If  $SNR_{in}$  is increased by 10 dB, what is the corresponding increase in SNR out of the detector?
- 5.7 Prove that for an FM signal, a quadrature detector is able to detect the message properly.
- 5.8 Design a quadrature detector for an IF frequency of 70 MHz, assuming a 200 kHz IF passband. Pick reasonable circuit values and plot the amplitude and phase response about the IF center frequency.
- 5.9 Using computer simulation, demonstrate that an FM signal can be demodulated using (a) slope detection and (b) a zero-crossing detector.
- 5.10 Verify that the BPSK receiver shown in Figure 5.23 is able to recover a digital

- message  $m(t)$ .
- 5.11 Plot the BER vs.  $E_b/N_0$  performance for BPSK, DPSK, QPSK and noncoherent FSK in additive Gaussian white noise. List advantages and disadvantages of each modulation method from the mobile communications standpoint.
  - 5.12 If a mobile radio link operates with 30 dB SNR and uses a 200 kHz channel, find the theoretic maximum data capacity possible. How does your answer compare to what is offered by the GSM standard, which operates at a channel rate of 270.8333 kbps?
  - 5.13 Compare the channel spectral efficiencies of IS-54, GSM, PDC, and IS-95. What are the theoretical spectral efficiencies for each of these standards if they operate at 20 dB SNR?
  - 5.14 Design a raised cosine rolloff filter for  $T_g = 1/24300$  s and  $\alpha = 0.35$ . Write expressions for, and plot, the impulse response and the frequency response of the filter. If this filter were used for 30 kHz RF channels, what percentage of the total radiated energy would fall out-of-band? Computer simulation or numerical analysis may be used to determine results.
  - 5.15 Design a Gaussian pulse-shaping filter with  $BT = 0.5$  for a symbol rate of 19.2 kbps. Write expressions for, and plot, the impulse response and frequency response of the filter. If this filter were used to produce GMSK in 30 kHz RF channels, what percentage of the total radiated energy would fall out-of-band? Repeat for the case of  $BT = 0.2$ , and  $BT = 0.75$ . Computer simulation or numerical analysis may be used to determine results.
  - 5.16 Derive equation (5.105) for an MSK signal.
  - 5.17 Generate the binary message 01100101 through MSK transmitter and receiver shown in Figure 5.39 and Figure 5.40. Sketch the waveshapes at the inputs, outputs, and through the system. Computer simulation may be used.
  - 5.18 If 63 users share a CDMA system, and each user has a processing gain of 511, determine the average probability of error for each user. What assumptions have you made in determining your result?
  - 5.19 For problem 5.18, determine the percentage increase in the number of users if the probability of error is allowed to increase by an order of magnitude.
  - 5.20 A FH-SS system uses 50 kHz channels over a continuous 20 MHz spectrum. Fast frequency hopping is used, where 2 hops occur for each bit. If binary FSK is the modulation method used in this system, determine (a) the number of hops per second if each user transmits at 25 kbps; (b) the probability of error for a single user operating at an  $E_b/N_0 = 20$  dB; (c) the probability of error for a user operating at  $E_b/N_0 = 20$  dB with 20 other FH-SS users which are independently frequency hopped; (d) the probability of error for a user operating at  $E_b/N_0 = 20$  dB with 200 other FH-SS users which are independently frequency hopped.
  - 5.21 Simulate a GMSK signal and verify that the Gaussian filter bandwidth has a major impact on the spectral shape of the signal. Plot spectral shapes for (a)  $BT = 0.2$ , (b)  $BT = 0.5$ , and (c)  $BT = 1$ .
  - 5.22 Compare the BER and RF bandwidth of a GMSK signal operating in AWGN for the following BT values: (a) 0.25, (b) 0.5, (c) 1, (d) 5. Discuss the practical advantages and disadvantages of these cases.
  - 5.23 Demonstrate mathematically that a  $\pi/4$  QPSK data stream may be detected using an FM receiver (Hint: consider how an FM receiver responds to phase changes).
  - 5.24 Demonstrate mathematically that a  $\pi/4$  QPSK signal can be detected using the

IF and baseband differential detection circuits described in Chapter 5.

- 5.25 Using the expression for probability of error in a flat fading channel, find the average probability of error for DPSK if a channel has an exponential SNR probability density of  $p(x) = e^{-x}$  for  $x > 0$ .
- 5.26 Determine the necessary  $E_b/N_0$  in order to detect DPSK with an average BER of  $10^{-3}$  for a (a) Rayleigh fading channel (b) Ricean fading channel, with  $K = 6dB, 7dB$ .
- 5.27 Determine the necessary  $E_b/N_0$  in order to detect BPSK with an average BER of  $10^{-5}$  for a (a) Rayleigh fading channel (b) Ricean fading channel, with  $K = 6dB, 7dB$ .
- 5.28 Show that equation (5.154) is correct. That is, prove that if  $\alpha$  is Rayleigh distributed then the pdf of  $\alpha^2$  is given by  $p(\alpha^2) = \frac{1}{\alpha^2} e^{-\alpha^2/\alpha^2}$ . This is the Chi-square pdf with 2 degrees of freedom.
- 5.29 Prove that the expression for the BER of GMSK as given in equation (5.163) is correct, using the techniques described in section 5.11.1 and the  $P_e$  expression for GMSK in an additive white Gaussian noise channel.
- 5.30 Prove that a Rayleigh fading signal has an exponential pdf for the power of the signal.

