CHAPTER

FOURTEEN

BANDPASS DIGITAL TRANSMISSION

Long-haul digital transmission usually requires CW modulation to generate a *bandpass* signal suited to the transmission medium — be it radio, cable, or whatever. Just as there are a multitude of modulation methods for analog signals, so too there are many ways of impressing digital information upon a carrier wave. This chapter applies concepts of baseband digital transmission and CW modulation to the study of bandpass digital transmission.

We begin with waveforms and spectral analysis of digital CW modulation for binary and M-ary modulating signals. Then we focus on the demodulation of binary signals in noise to bring out the distinction between *coherent* (synchronous) detection and *noncoherent* (envelope) detection. The last section deals with quadrature-carrier and M-ary systems, leading to a comparison of modulation methods with regard to spectral efficiency, hardware complexity, and system performance in the face of corrupting noise.

14.1 DIGITAL CW MODULATION

A digital signal can modulate the amplitude, frequency, or phase of a sinusoidal carrier wave. If the modulating waveform consists of NRZ rectangular pulses, then the modulated parameter will be switched or *keyed* from one discrete value to another. Figure 14.1-1 illustrates binary *amplitude-shift keying* (ASK), *frequency-shift keying* (FSK), and *phase-shift keying* (PSK). Also shown, for contrast, is the waveform produced by DSB modulation with Nyquist pulse shaping at baseband. Other modulation techniques combine amplitude and phase modulation, with or without baseband pulse shaping.

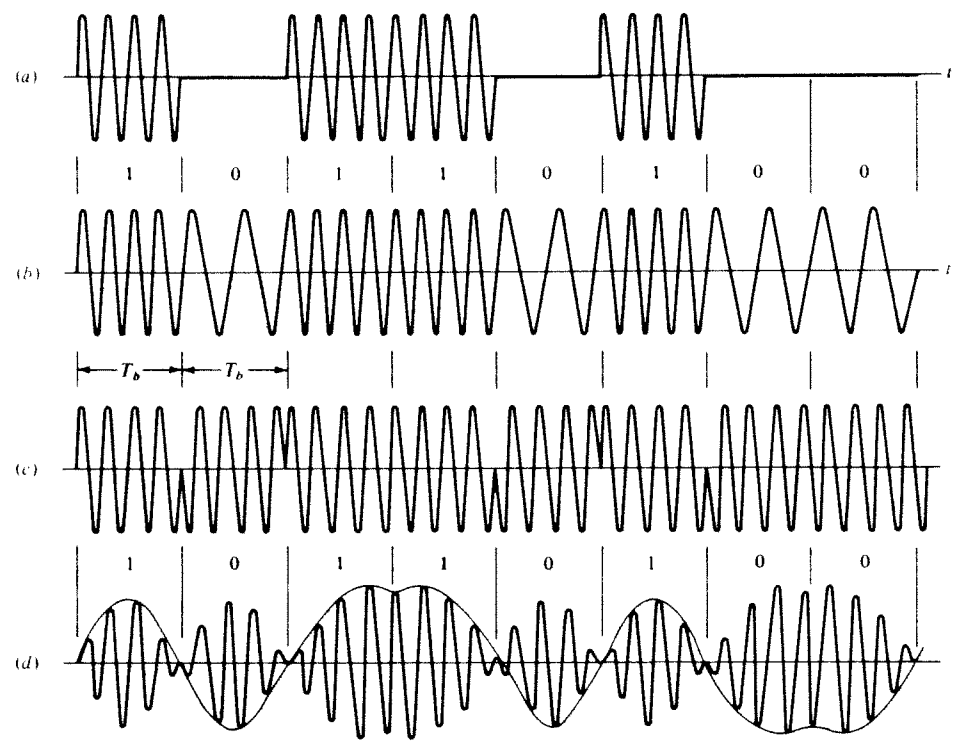


Figure 14.1-1 Binary modulated waveforms. (*a*) ASK; (*b*) FSK; (*c*) PSK; (*d*) DSB with baseband pulse shaping.

In this section we'll define specific types of digital modulation in terms of mathematical models and/or transmitter diagrams. We'll also examine their power spectra and estimate therefrom the transmission bandwidth required for a given digital signaling rate. As preparation, we first develop a technique for spectral analysis of bandpass digital signals.

Spectral Analysis of Bandpass Digital Signals

Any modulated bandpass signal may be expressed in the quadrature-carrier form

$$x_c(t) = A_c[x_i(t)\cos(\omega_c t + \theta) - x_q(t)\sin(\omega_c t + \theta)] \tag{1}$$

The carrier frequency f_c, amplitude A_c, and phase θ are constant, while the time-varying i (in-phase) and q (quadrature) components contain the message. Spectral analysis of $x_c(t)$ is relatively easy when the i and q components are statistically independent signals and at least one has zero mean. Then, from the superposition and modulation relations in Sect. 5.2, the power spectrum of $x_c(t)$ becomes

$$G_c(f) = \frac{A_c^2}{4}[G_i(f - f_c) + G_i(f + f_c) + G_q(f - f_c) + G_q(f + f_c)]$$

where $G_i(f)$ and $G_q(f)$ are the power spectra of the i and q components. For a more compact expression, we define the *equivalent lowpass spectrum*

$$G_{\ell p}(f) \triangleq G_i(f) + G_q(f) \tag{2}$$

so that

$$G_c(f) = \frac{A_c^2}{4}\left[G_{\ell p}(f - f_c) + G_{\ell p}(f + f_c)\right] \tag{3}$$

Thus, the bandpass spectrum is obtained from the equivalent lowpass spectrum by simple frequency translation.

Now suppose that the i component is an M-ary digital signal, say

$$x_i(t) = \sum_k a_k\, p(t - kD) \tag{4a}$$

where a_k represents a sequence of source digits with rate $r = 1/D$. We assume throughout that source digits are equiprobable, statistically independent, and uncorrelated. Consequently, Eq. (11), Sect. 11.1, applies here and

$$G_i(f) = \sigma_a^2\, r\,|P(f)|^2 + (m_a\, r)^2 \sum_{n=-\infty}^{\infty} |P(nr)|^2\ \delta(f - nr) \tag{4b}$$

Similar expressions hold when the q component is another digital waveform.

The pulse shape $p(t)$ in Eq. (4a) depends on the baseband filtering, if any, and on the type of modulation. Keyed modulation involves NRZ rectangular pulses and we'll find it convenient to work with pulses that start at $t = kD$, rather than being centered at $t = kD$ as in Chap. 11. Accordingly, let

$$p_D(t) \triangleq u(t) - u(t - D) = \begin{cases} 1 & 0 < t < D \\ 0 & \text{otherwise} \end{cases} \tag{5a}$$

whose Fourier transform yields

$$|P_D(f)|^2 = D^2\ \text{sinc}^2\, fD = \frac{1}{r^2}\ \text{sinc}^2\, \frac{f}{r} \tag{5b}$$

If $p(t) = p_D(t)$ in Eq. (4a), then the continuous spectral term in Eq. (4b) will be proportional to $|P_D(f)|^2$. Since $\text{sinc}^2\,(f/r)$ is not bandlimited, we conclude from Eqs. (2) and (3) that keyed modulation requires $f_c \gg r$ in order to produce a *bandpass* signal.

Amplitude Modulation Methods

The binary ASK waveform illustrated in Fig. 14.1-1a could be generated simply by turning the carrier on and off, a process described as *on-off keying* (OOK). In general, an M-ary ASK waveform has $M - 1$ discrete "on" amplitudes as well as the "off" state. Since there are no phase reversals or other variations, we can set

the q component of $x_c(t)$ equal to zero and take the i component to be a *unipolar* NRZ signal, namely

$$x_i(t) = \sum_k a_k\, p_D(t - kD) \qquad a_k = 0, 1, \ldots, M - 1 \tag{6a}$$

The mean and variance of the digital sequence are

$$m_a = \overline{a_k} = \frac{M - 1}{2} \qquad \sigma_a^2 = \overline{a_k^2} - m_a^2 = \frac{M^2 - 1}{12} \tag{6b}$$

Hence, the equivalent lowpass spectrum is

$$G_{\ell p}(f) = G_i(f) = \frac{M^2 - 1}{12r}\ \text{sinc}^2\, \frac{f}{r} + \frac{(M - 1)^2}{4}\,\delta(f) \tag{7}$$

obtained with the help of Eqs. (2), (4b), and (5b).

Figure 14.1-2 shows the resulting bandpass spectrum $G_c(f)$ for $f > 0$. Most of the signal power is contained within the range $f_c \pm r/2$, and the spectrum has a *second-order rolloff* proportional to $|f - f_c|^{-2}$ away from the carrier frequency. These considerations suggest the estimated transmission bandwidth to be $B_T \approx r$. If an M-ary ASK signal represents binary data at rate $r_b = r \log_2 M$, then $B_T \approx r_b/\log_2 M$ or

$$r_b/B_T \approx \log_2 M \qquad \text{bps/Hz} \tag{8}$$

This ratio of bit rate to transmission bandwidth serves as our measure of modulation "speed" or spectral efficiency. Binary OOK has the poorest spectral efficiency since $r_b/B_T \approx 1$ bps/Hz when $M = 2$.

Drawing upon the principle of quadrature-carrier multiplexing, *quadrature-carrier* AM (QAM) achieves twice the modulation speed of binary ASK. Figure 14.1-3a depicts the functional blocks of a binary QAM transmitter with a *polar* binary input at rate r_b. The serial-to-parallel converter divides the input into two

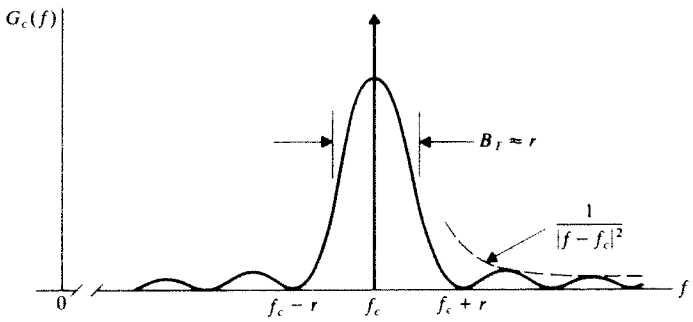


Figure 14.1-2 ASK power spectrum.

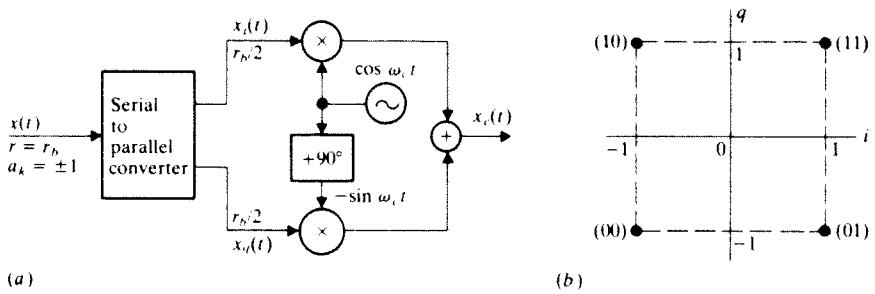


Figure 14.1-3 Binary QAM. (*a*) Transmitter; (*b*) signal constellation.

streams consisting of alternate bits at rate $r = r_b/2$. Thus, the i and q modulating signals are represented by

$$x_i(t) = \sum_k a_{2k}\, p_D(t - kD) \qquad x_q(t) = \sum_k a_{2k+1}\, p_D(t - kD)$$

where $D = 1/r = 2T_b$ and $a_k = \pm 1$. The peak modulating values are $x_i = x_q = \pm 1$ during an arbitrary interval $kD < t < (k + 1)D$. Figure 14.1-3*b* conveys this information as a two-dimensional *signal constellation.* The four signal points have been labeled with the corresponding pairs of source bits, known as *dibits.*

Summing the modulated carriers finally yields the QAM signal in the form of Eq. (1). The i and q components are independent but they have the same pulse shape and the same statistical values, namely, $m_a = 0$ and $\sigma_a^2 = 1$. Thus,

$$G_{c_p}(f) = 2 \times r\,|P_D(f)|^2 = \frac{4}{r_b}\,\text{sinc}^2\,\frac{2f}{r_b} \tag{9}$$

where we've used Eqs. (4*b*) and (5*b*) with $r = r_b/2$. Binary QAM achieves $r_b/B_T \approx 2$ bps/Hz because the dibit rate equals one-half of the input bit rate, reducing the transmission bandwidth to $B_T \approx r_b/2$.

Keep in mind, however, that ASK and QAM spectra actually extend beyond the estimated transmission bandwidth. Such spectral "spillover" outside B_T becomes an important concern in radio transmission and frequency-division multiplexing systems when it creates interference with other signal channels. Bandpass filtering at the output of the modulator controls spillover, but heavy filtering introduces ISI in the modulated signal and should be avoided.

Spectral efficiency without spillover is achieved by the *vestigial-sideband* modulator diagrammed in Fig. 14.1-4*a*. This VSB method applies Nyquist pulse shaping to a polar input signal, as covered in Sect. 11.3, producing a bandlimited modulating signal with $B = (r/2) + \beta_N$. The VSB filter then removes all but a vestige of width β_V from one sideband, so $G_c(f)$ looks something like Fig. 14.1-4*b*—a bandlimited spectrum with $B_T = (r/2) + \beta_N + \beta_V$. Therefore, if $r = r_b/\log_2 M$, then

$$r_b/B_T \le 2 \log_2 M \tag{10}$$

and the upper bound holds when $\beta_N \ll r$ and $\beta_V \ll r$.

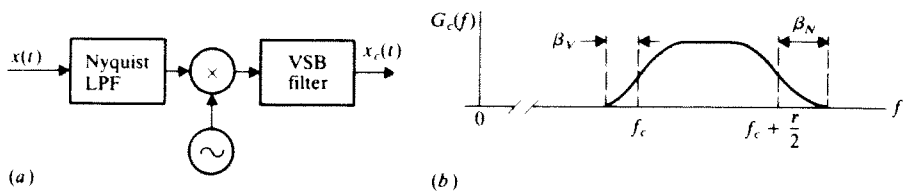


Figure 14.1-4 Digital VSB. (*a*) Transmitter; (*b*) power spectrum.

Exercise 14.1-1 Binary data is to be transmitted on a 1-MHz carrier. Spillover is not a concern, but B_T must satisfy the fractional bandwidth constraint $B_T/f_c \le 0.1$. Estimate the maximum possible bit rate r_b when the modulation is: (*a*) OOK, (*b*) binary QAM, (*c*) VSB with $M = 8$.

Phase Modulation Methods

The binary PSK waveform back in Fig. 14.1-1*c* contains phase shifts of $\pm\pi$ radians, often described as *phase-reversal keying* (PRK). An M-ary PSK signal has phase shift ϕ_k in the time interval $kD < t < (k + 1)D$, expressed in general by

$$x_c(t) = A_c \sum_k \cos\,(\omega_c t + \theta + \phi_k) p_D(t - kD) \tag{11}$$

Trigonometric expansion of the cosine function yields our desired quadrature-carrier form with

$$x_i(t) = \sum_k I_k\, p_D(t - kD) \qquad x_q(t) = \sum_k Q_k\, p_D(t - kD) \tag{12a}$$

where

$$I_k = \cos\,\phi_k \qquad Q_k = \sin\,\phi_k \tag{12b}$$

To ensure the largest possible phase modulation for a given value of M, we'll take the relationship between ϕ_k and a_k to be

$$\phi_k = \pi(2a_k + N)/M \qquad a_k = 0, 1, \ldots, M - 1 \tag{13}$$

in which N is an integer, usually 0 or 1.

Examples of PSK signal constellations are shown in Fig. 14.1-5, including the corresponding binary words in Gray code. The binary words for adjacent signal points therefore differ by just one bit. The PSK signal with $M = 4$ and $N = 0$ represented in Fig. 14.1-5*a* is designated *quaternary* or *quadriphase* PSK (QPSK). Had we taken QPSK with $N = 1$, the signal points would have been identical to binary QAM (Fig. 14.1-3*b*). Indeed, you can think of binary QAM as two PRK signals on quadrature carriers. M-ary PSK differs from M-ary ASK, of course, since an ideal PSK waveform always has a *constant envelope.*

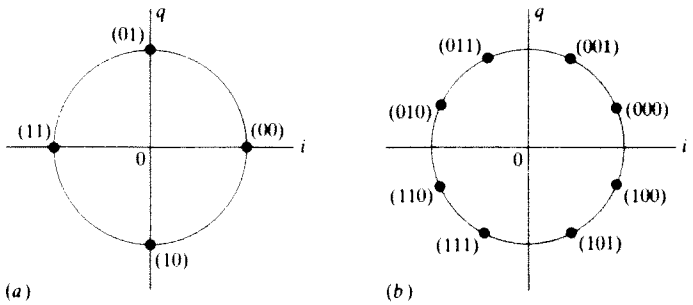


Figure 14.1-5 PSK signal constellations. (*a*) $M = 4$; (*b*) $M = 8$.

PSK spectral analysis becomes a routine task after you note from Eqs. (12*b*) and (13) that

$$\bar{I}_k = \bar{Q}_k = 0 \qquad \overline{I_k^2} = \overline{Q_k^2} = 1/2 \qquad \overline{I_k Q_j} = 0$$

Hence, the i and q components are statistically independent, and

$$G_{\ell p}(f) = 2 \times \frac{r}{2} |P_D(f)|^2 = \frac{1}{r} \operatorname{sinc}^2 \frac{f}{r} \tag{14}$$

Comparison with Eq. (7) reveals that $G_c(f)$ will have the same shape as an ASK spectrum (Fig. 14.1-2) without the carrier-frequency impulse. The absence of a discrete carrier component means that PSK has better power efficiency, but the spectral efficiency is the same as ASK.

Some PSK transmitters include a BPF to control spillover. However, bandpass filtering produces *envelope variations* via the FM-to-AM conversion effect discussed in Sect. 7.2. (Remember that a stepwise phase shift is equivalent to an FM impulse.) The typical nonlinear amplifier used at microwave carrier frequencies will flatten out these envelope variations and restore spillover—largely negating the function of the BPF. A special form of QPSK called *staggered* or *offset-keyed* QPSK (OQPSK) has been devised to combat this problem. The OQPSK transmitter diagrammed in Fig. 14.1-6 delays the quadrature signal such

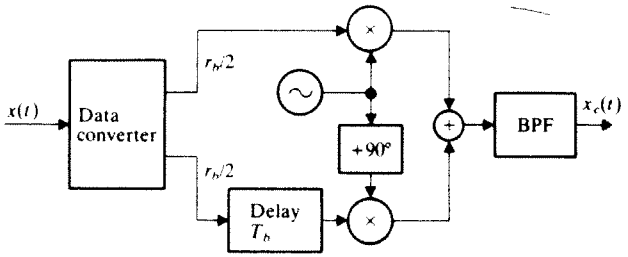


Figure 14.1-6 Offset-keyed QPSK transmitter.

that modulated phase shifts occur every $D/2 = T_b$ seconds but they never exceed $\pm\pi/2$ radians. Cutting the maximum phase shift in half results in much smaller envelope variations after bandpass filtering.

When envelope variations are allowable, combined *amplitude-phase keying* (APK) is an attractive family of modulation methods. APK has essentially the same spectral efficiency as PSK, but it can achieve better performance with respect to noise and errors. Further discussion is postponed to Sect. 14.4.

Exercise 14.1-2 Draw the signal constellation for binary PSK with $\phi_k = \pi(2a_k - 1)/4$ and $a_k = 0, 1$. Then determine the lowpass equivalent spectrum and sketch $G_c(f)$.

Frequency Modulation Methods

There are two basic methods for digital frequency modulation. *Frequency-shift keying* (FSK) is represented conceptually by Fig. 14.1-7*a*, where the digital signal $x(t)$ controls a switch that selects the modulated frequency from a bank of M oscillators. The modulated signal is therefore *discontinuous* at every switching instant $t = kD$ unless the amplitude, frequency, and phase of each oscillator has been carefully adjusted. Discontinuities are avoided in *continuous-phase FSK* (CPFSK) represented by Fig. 14.1-7*b*, where $x(t)$ modulates the frequency of a single oscillator. Both forms of digital frequency modulation pose significant difficulties for spectral analysis, so we'll limit our consideration to some selected cases.

First, consider M-ary FSK. Let all oscillators in Fig. 14.1-7*a* have the same amplitude A_c and phase θ, and let their frequencies be related to a_k by

$$f_k = f_c + f_d a_k \qquad a_k = \pm 1, \pm 3, \ldots, \pm(M-1) \tag{15a}$$

which assumes that M is even, Then

$$x_c(t) = A_c \sum_k \cos(\omega_c t + \theta + \omega_d a_k t) p_D(t - kD) \tag{15b}$$

where $\omega_d = 2\pi f_d$. The parameter f_d equals the frequency shift away from f_c when $a_k = \pm 1$, and adjacent frequencies are spaced by $2f_d$. Continuity of $x_c(t)$ at $t = kD$ is assured if $2\omega_d D = 2\pi N$ where N is an integer.

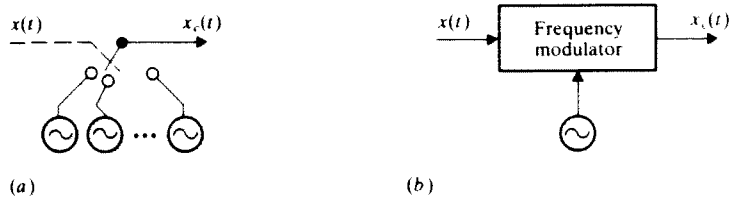


Figure 14.1-7 Digital frequency modulation. (*a*) FSK; (*b*) continuous-phase FSK.

We'll analyze a version of binary FSK known as *Sunde's* FSK, defined by the foregoing relations with $M = 2$, $D = T_b = 1/r_b$, and $N = 1$. Then $p_D(t) = u(t) - u(t - kT_b)$ and

$$f_d = r_b/2 \tag{16}$$

After trigonometric expansion of $x_c(t)$, we use the fact that $a_k = \pm 1$ to write

$$\cos \omega_d a_k t = \cos \omega_d t \qquad \sin \omega_d a_k t = a_k \sin \omega_d t$$

The i component thereby reduces to

$$x_i(t) = \cos \pi r_b t \tag{17a}$$

independent of a_k. The q component contains a_k in the form

$$x_q(t) = \sum_k a_k \sin (\pi r_b t)[u(t - kT_b) - u(t - kT_b - T_b)]$$

$$= \sum_k Q_k p(t - kT_b) \qquad Q_k = (-1)^k a_k \tag{17b}$$

where

$$p(t) = \sin (\pi r_b t)[u(t) - u(t - T_b)] \tag{17c}$$

The intervening manipulations are left to you as an instructive exercise.

Once again, we have independent i and q components. The i component, being a sinusoid, just contributes spectral impulses at $\pm r_b/2$ in the equivalent lowpass spectrum. The power spectrum of the q component contains no impulses since $\bar{Q}_k = 0$, whereas $\overline{Q_k^2} = \overline{a_k^2} = 1$. Thus,

$$G_{cp}(f) = \frac{1}{4}\left[\delta\left(f - \frac{r_b}{2}\right) + \delta\left(f + \frac{r_b}{2}\right)\right] + r_b |P(f)|^2 \tag{18a}$$

where

$$|P(f)|^2 = \frac{1}{4r_b^2}\left[\operatorname{sinc} \frac{f - (r_b/2)}{r_b} + \operatorname{sinc} \frac{f + (r_b/2)}{r_b}\right]^2$$

$$= \frac{4}{\pi^2 r_b^2}\left[\frac{\cos (\pi f/r_b)}{(2f/r_b)^2 - 1}\right]^2 \tag{18b}$$

The resulting bandpass spectrum is shown in Fig. 14.1-8.

Observe that the impulses correspond to the keyed frequencies $f_c \pm f_d = f_c \pm r_b/2$, and that the spectrum has a fourth-order rolloff. This rapid rolloff means that Sunde's FSK has very little spillover for $|f - f_c| > r_b$. We therefore take $B_T \approx r_b$, even though the central lobe of $G_c(f)$ is 50% wider than the central lobe of a binary ASK or PSK spectrum.

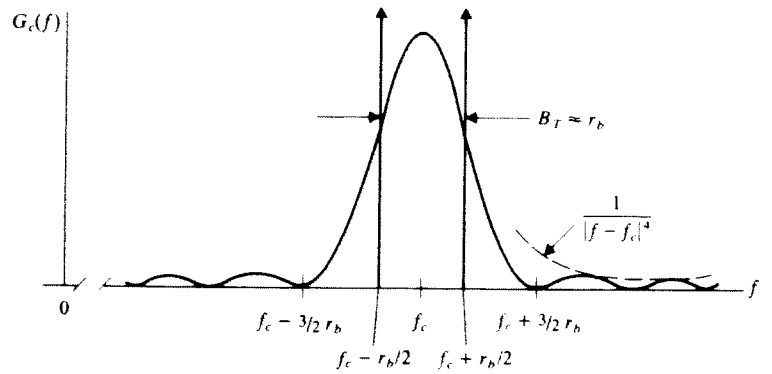


Figure 14.1-8 Power spectrum of binary FSK with $f_d = r_b/2$.

Another special case is *M-ary orthogonal FSK*, in which the M keyed frequencies are equispaced by $2f_d = 1/2D = r/2$. Without attempting the spectral analysis, we can surmise that $B_T \geq M \times 2f_d = Mr/2 = Mr_b/(2 \log_2 M)$. Therefore,

$$r_b/B_T \leq (2 \log_2 M)/M \tag{19}$$

and the modulation speed is less than M-ary ASK or PSK for $M \geq 4$. In other words, orthogonal FSK is a *wideband* modulation method.

Continuous-phase FSK may be wideband or narrowband depending on the frequency deviation. Let $x(t)$ in Fig. 14.1-7*b* start at $t = 0$, so

$$x(t) = \sum_{k=0}^{\infty} a_k p_D(t - kD) \qquad a_k = \pm 1, \pm 2, \ldots, \pm(M - 1)$$

and frequency modulation produces the CPFSK signal

$$x_c(t) = A_c \cos\left[\omega_c t + \theta + \omega_d \int_0^t x(\lambda)\, d\lambda\right] \qquad t \geq 0$$

To bring out the difference between CPFSK and FSK, consider the integral

$$\int_0^t x(\lambda)\, d\lambda = \sum_{k=0}^{\infty} a_k \int_0^t p_D(\lambda - kD)\, d\lambda$$

in which $p_D(\lambda - kD) = 0$ except for $kD < \lambda < (k + 1)D$ when $p_D(\lambda - kD) = 1$. Piecewise integration yields

$$\int_0^t x(\lambda)\, d\lambda = a_0 t \qquad 0 < t < D$$

$$= a_0 D + a_1(t - D) \qquad D < t < 2D$$

$$= \left(\sum_{j=0}^{k-1} a_j\right) D + a_k(t - kD) \qquad kD < t < (k + 1)D$$

Now we can express $x_c(t)$ in the summation form

$$x_c(t) = A_c \sum_{k=0}^{\infty} \cos\,[\omega_c t + \theta + \phi_k + \omega_d a_k(t - kD)]p_D(t - kD) \qquad (20a)$$

where $t \geq 0$ and

$$\phi_k \triangleq \omega_d D \sum_{j=0}^{k-1} a_j \qquad (20b)$$

with the understanding that $\phi_k = 0$ for $k = 0$.

Equation (20) shows that CPSK has a frequency shift $f_d a_k$ in the interval $kD < t < (k + 1)D$, just like FSK. But it also has a phase shift ϕ_k that depends on the *previous* digits. This phase shift results from the frequency-modulation process and ensures phase continuity for all t. Unfortunately, the past history embodied in ϕ_k greatly complicates CPFSK spectral analysis. Proakis (1983, chap. 3) gives further details and plots of $G_c(f)$ for various values of f_d when $M = 2$, 4, and 8. To conclude this section, we'll examine an important special case of binary CPSK called *minimum-shift keying* (MSK).

Exercise 14.1-3 Carry out the details omitted in the derivation of Eqs. (17*a*)–(17*c*). *Hint:* Show that $\sin \omega_d t = \sin\,[\omega_d(t - kT_b) + k\pi] = \cos\,(k\pi) \times \sin\,[\omega_d(t - kT_b)]$.

Minimum-Shift Keying★

Minimum-shift keying, also known as *fast FSK*, is binary CPFSK with

$$f_d = \frac{r_b}{4} \qquad a_k = \pm 1 \qquad \phi_k = \frac{\pi}{2} \sum_{j=0}^{k-1} a_j \qquad (21)$$

Notice that the frequency spacing $2f_d = r_b/2$ is half that of Sunde's FSK. This fact, together with the continuous-phase property, results in a more compact spectrum, free of impulses. Subsequent analysis will prove that $G_i(f) = G_q(f)$ and

$$G_{cp}(f) = \frac{1}{r_b}\left[\operatorname{sinc}\frac{f - (r_b/4)}{(r_b/2)} + \operatorname{sinc}\frac{f + (r_b/4)}{(r_b/2)}\right]^2$$
$$= \frac{16}{\pi^2 r_b}\left[\frac{\cos\,(2\pi f/r_b)}{(4f/r_b)^2 - 1}\right]^2 \qquad (22)$$

The bandpass spectrum $G_c(f)$ plotted in Fig. 14.1-9 has minuscule spillover beyond the central lobe of width $3r_b/2$. The rapid rolloff justifies taking $B_T \approx r_b/2$, so

$$r_b/B_T \approx 2 \text{ bps/Hz}$$

which is twice the modulation speed of Sunde's FSK and accounts for the name "fast" FSK.

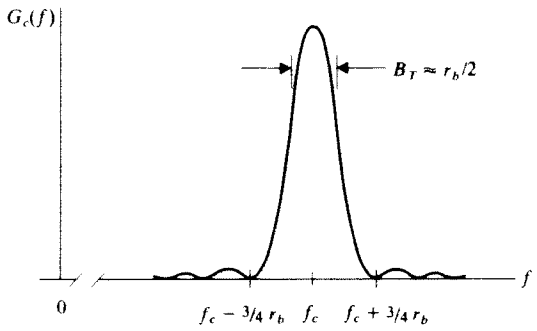


Figure 14.1-9 MSK power spectrum.

Our investigation of MSK starts with the usual trigonometric expansion to put $x_c(t)$ in quadrature-carrier form with

$$x_i(t) = \sum_{k=0}^{\infty} \cos\,(\phi_k + a_k c_k)p_{T_b}(t - kT_b)$$

$$x_q(t) = \sum_{k=0}^{\infty} \sin\,(\phi_k + a_k c_k)p_{T_b}(t - kT_b)$$

where

$$c_k \triangleq \frac{\pi r_b}{2}(t - kT_b) \qquad p_{T_b}(t) = u(t) - u(t - kT_b)$$

We'll also draw upon the behavior of ϕ_k versus k as displayed in the trellis pattern of Fig. 14.1-10. This pattern clearly reveals that $\phi_k = 0, \pm\pi, \pm 2\pi, \ldots$, for even values of k while $\phi_k = \pm\pi/2, \pm 3\pi/2, \ldots$, for odd values of k.

As a specific example, let the input message sequence be 100010111. The resulting phase path ϕ_k is shown in Fig. 14.1-11*a*, taking $a_k = +1$ for input bit 1

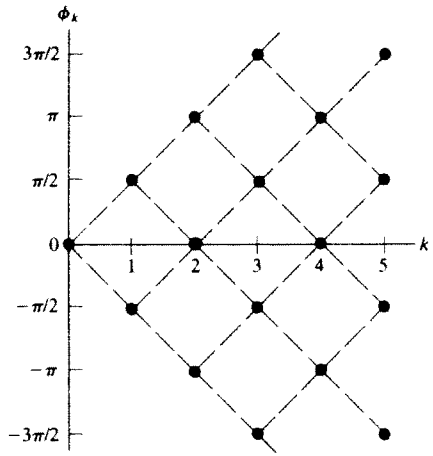


Figure 14.1-10 MSK phase trellis.

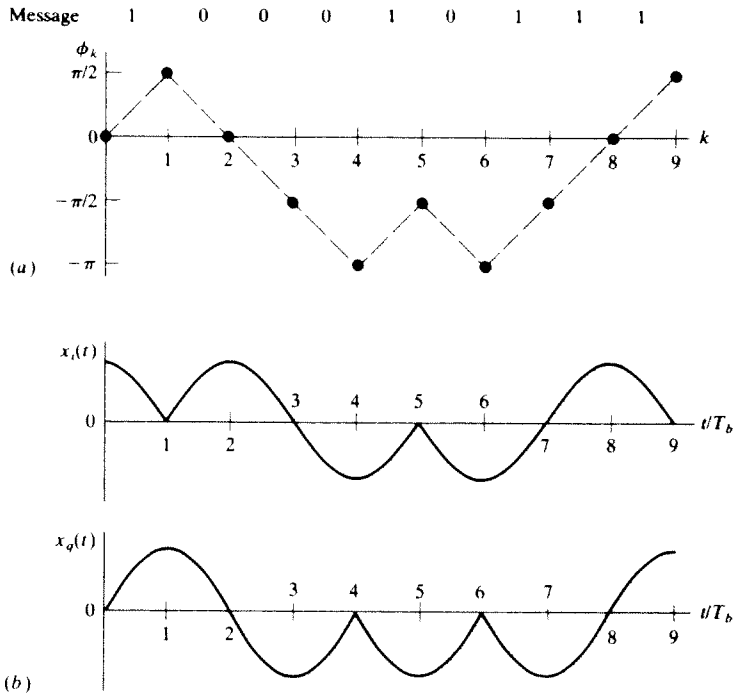


Figure 14.1-11 Illustration of MSK. (*a*) Phase path; (*b*) *i* and *q* waveforms.

and $a_k = -1$ for input bit 0. The corresponding i and q waveforms calculated from the foregoing expressions are sketched in Fig. 14.1-11*b*. We see that both waveforms have zeros spaced by $2T_b$, but staggered such that the zeros of $x_i(t)$ coincide with the peaks of $x_q(t)$, and vice versa. These observations will guide our subsequent work.

Consider an arbitrary time interval between adjacent zeros of the i component, i.e.,

$$(k-1)T_b < t < (k+1)T_b$$

with k being *even*. During this interval,

$$x_i(t) = \cos(\phi_{k-1} + a_{k-1}c_{k-1})p_{T_b}[t-(k-1)T_b] + \cos(\phi_k + a_k c_k)p_{T_b}(t-kT_b)$$

which we seek to combine into a single term. Since k is even, $\sin\phi_k = 0$, and routine trigonometric manipulations yield

$$\cos(\phi_k + a_k c_k) = \cos\phi_k \cos(a_k c_k) = \cos\phi_k \cos c_k$$

Likewise, using

$$\cos\phi_{k-1} = 0 \qquad \phi_{k-1} = \phi_k - a_{k-1}\pi/2 \qquad c_{k-1} = c_k + \pi/2$$

we get

$$\cos(\phi_{k-1} + a_{k-1}c_{k-1}) = -\sin\phi_{k-1}\sin(a_{k-1}c_{k-1}) = a_{k-1}^2\cos\phi_k\cos c_k = \cos\phi_k\cos c_k$$

Thus, for the interval in question,

$$x_i(t) = \cos\phi_k\cos c_k\{p_{T_b}[t-(k-1)T_b] + p_{T_b}(t-kT_b)\}$$
$$= \cos\phi_k\cos[(\pi r_b/2)(t-kT_b)][u(t-kT_b+T_b) - u(t-kT_b-T_b)]$$

Summing intervals to encompass all $t \geq 0$ finally yields

$$x_i(t) = \sum_{k\ \text{even}} I_k\, p(t-kT_b) \qquad I_k = \cos\phi_k \tag{23}$$

where

$$p(t) = \cos(\pi r_b t/2)[u(t+T_b) - u(t-T_b)] \tag{24}$$

This result checks out against the waveform in Fig. 14.1-11*b* since $I_k = \cos\phi_k = \pm 1$ when k is even.

Now, for the q component, we consider the interval $(k-1)T_b < t < (k+1)T_b$ with k *odd*. Similar manipulations as before lead to

$$x_q(t) = \sin\phi_k\cos c_k\{p_{T_b}[t-(k-1)T_b] + p_{T_b}(t-kT_b)\}$$

Thus, for all $t \geq 0$,

$$x_q(t) = \sum_{k\ \text{odd}} Q_k\, p(t-kT_b) \qquad Q_k = \sin\phi_k \tag{25}$$

which also agrees with Fig. 14.1-11*b*. Equation (22) follows from Eqs. (23)–(25) since the i and q components are independent, with $\overline{I_k} = \overline{Q_k} = 0$ and $\overline{I_k^2} = \overline{Q_k^2} = 1$.

14.2 COHERENT BINARY SYSTEMS

Coherent bandpass digital systems employ information about the carrier frequency and phase at the receiver to detect the message—like synchronous analog detection. *Noncoherent* systems don't require synchronization with the carrier phase, but they fall short of the optimum performance made possible by coherent detection.

This section examines coherent binary transmission, starting with a general treatment of optimum binary detection in the presence of additive white gaussian noise (AWGN). The results are then applied to assess the performance of specific binary modulation systems. We'll focus throughout on keyed modulation (OOK, PRK, and FSK), without baseband filtering or transmission distortion that might produce ISI in the modulated signal. The challenging problems of optimum receiver design including ISI effects are addressed in more advanced texts such as Proakis (1983, chap. 6).

Optimum Binary Detection

Any bandpass binary signal with keyed modulation can be expressed in the general quadrature-carrier form

$$x_c(t) = A_c \left\{ \left[\sum_k I_k p_i(t - kT_b) \right] \cos(\omega_c t + \theta) - \left[\sum_k Q_k p_q(t - kT_b) \right] \sin(\omega_c t + \theta) \right\}$$

For practical coherent systems, the carrier wave should be synchronized with the digital modulation. Accordingly, we'll take $\theta = 0$ and impose the condition

$$f_c = N_c/T_b = N_c r_b \tag{1}$$

where N_c is an integer — usually a very large integer. Then

$$x_c(t) = A_c \sum_k [I_k p_i(t - kT_b) \cos \omega_c(t - kT_b) - Q_k p_q(t - kT_b) \sin \omega_c(t - kT_b)]$$

and we can concentrate on a single bit interval by writing

$$x_c(t) = s_m(t - kT_b) \qquad kT_b < t < (k+1)T_b \tag{2}$$

with

$$s_m(t) \triangleq A_c[I_k p_i(t) \cos \omega_c t - Q_k p_q(t) \sin \omega_c t]$$

Here, $s_m(t)$ stands for either of two *signaling waveforms*, $s_0(t)$ and $s_1(t)$, representing the message bits $m = 0$ and $m = 1$.

Now consider the received signal $x_c(t)$ corrupted by white gaussian noise. We showed in Sect. 11.2 that an optimum baseband receiver minimizes error probability with the help of a filter matched to the baseband pulse shape. But binary CW modulation involves two different signaling waveforms, as in Eq. (2), rather than one pulse shape with two different amplitudes. Consequently, we must redo our previous analysis in terms of $s_0(t)$ and $s_1(t)$.

Figure 14.2-1 shows the proposed receiver structure labeled with the relevant signals and noise for the interval under consideration. This bandpass receiver is just like a baseband receiver with a BPF in place of an LPF. The filtered signal

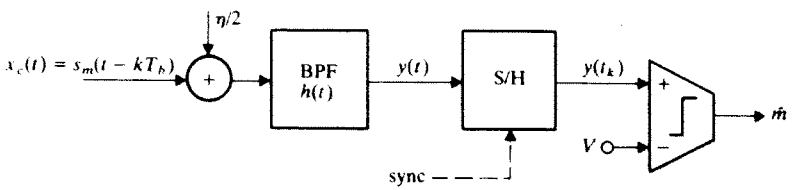


Figure 14.2-1 Bandpass binary receiver.

plus noise $y(t)$ is sampled at $t_k = (k+1)T_b$, the end of the bit interval, and compared with a threshold level to regenerate the most likely message bit $\hat{m}$. We seek the BPF impulse response $h(t)$ and threshold level V for optimum binary detection, resulting in the smallest average regeneration error probability.

As in Sect. 11.2, let H_1 and H_0 denote the hypotheses that $m = 1$ and $m = 0$, respectively. The receiver decides between H_1 and H_0 according to the observed value of the random variable

$$Y = y(t_k) = z_m + n$$

where

$$z_m \triangleq z_m(t_k) = [s_m(t - kT_b) * h(t)]\Big|_{t=t_k} = \int_{kT_b}^{(k+1)T_b} s_m(\lambda - kT_b) h(t_k - \lambda)\, d\lambda = \int_0^{T_b} s_m(\lambda) h(T_b - \lambda)\, d\lambda \tag{3}$$

The noise sample $n = n(t_k)$ is a gaussian r.v. with zero mean and variance σ^2, so the conditional PDFs of Y given H_1 or H_0 will be gaussian curves centered at z_1 or z_0, portrayed by Fig. 14.2-2. With the usual assumption of equally likely zeros and ones, the optimum threshold is at the intersection point, i.e.,

$$V_{\text{opt}} = \tfrac{1}{2}(z_1 + z_0)$$

Then, from the symmetry of the PDFs, $P_{e_1} = P_{e_0}$ and

$$P_e = Q(|z_1 - z_0|/2\sigma)$$

in which the absolute-value notation $|z_1 - z_0|$ includes the case of $z_1 < z_0$.

But what BPF impulse response $h(t)$ maximizes the ratio $|z_1 - z_0|/2\sigma$ or, equivalently, $|z_1 - z_0|^2/4\sigma^2$? To answer this question, we note from Eq. (3) that

$$|z_1 - z_0|^2 = \left| \int_{-\infty}^{\infty} [s_1(\lambda) - s_0(\lambda)] h(T_b - \lambda)\, d\lambda \right|^2 \tag{4a}$$

where the infinite limits are allowed since $s_m(t) = 0$ outside of $0 < t < T_b$. We also note that

$$\sigma^2 = \frac{\eta}{2} \int_{-\infty}^{\infty} |h(t)|^2\, dt = \frac{\eta}{2} \int_{-\infty}^{\infty} |h(T_b - \lambda)|^2\, d\lambda \tag{4b}$$

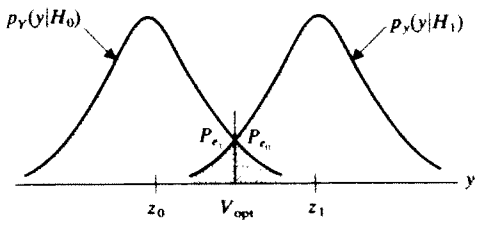


Figure 14.2-2 Conditional PDFs.

Application of Schwarz's inequality now yields

$$\frac{|z_1 - z_0|^2}{4\sigma^2} \le \frac{1}{2\eta}\int_{-\infty}^{\infty}[s_1(t) - s_0(t)]^2\,dt \tag{5}$$

and the ratio is maximum if $h(T_b - t) = K[s_1(t) - s_0(t)]$. Thus,

$$h_{opt}(t) = K[s_1(T_b - t) - s_0(T_b - t)] \tag{6}$$

with K being an arbitrary constant.

Equation (6) says that the filter for optimum binary detection should be matched to the *difference* between the two signaling waveforms. Alternatively, you could use two matched filters with $h_1(t) = Ks_1(T_b - t)$ and $h_0(t) = Ks_0(T_b - t)$ arranged in parallel per Fig. 14.2-3*a*; subtracting the output of the lower branch from the upper branch yields the same optimum response. In either case, any stored energy in the filters must be discharged after each sampling instant to prevent ISI in subsequent bit intervals.

Another alternative, with built-in discharge, is based on the observation that the sampled signal value from the upper branch in Fig. 14.2-3*a* is

$$z_{m1}(t_k) = \int_0^{T_b} s_m(\lambda)h_1(T_b - \lambda)\,d\lambda$$
$$= \int_{kT_b}^{(k+1)T_b} s_m(t - kT_b)Ks_1(t - kT_b)\,dt$$

and likewise for $z_{m0}(t_k)$. Hence, optimum filtering can be implemented by the system diagrammed in Fig. 14.2-3*b*, which requires two multipliers, two integrators, and stored copies of $s_0(t)$ and $s_1(t)$. This system is called a *correlation*

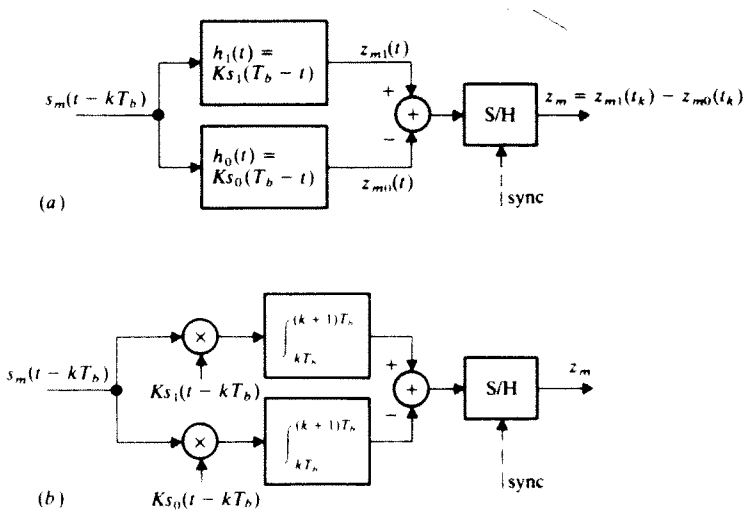


Figure 14.2-3 Optimum binary detection. (*a*) Parallel matched filters; (*b*) correlation detector.

detector because it correlates the received signal plus noise with noise-free copies of the signaling waveforms. Note that correlation detection is a generalization of the integrate-and-dump technique for matched filtering.

Regardless of the particular implementation method, the error probability with optimum binary detection depends upon the ratio maximized in Eq. (5). This ratio, in turn, depends on the signal energy per bit and on the similarity of the signaling waveforms. To pursue this point, consider the expansion

$$\int_0^{T_b}[s_1(t) - s_0(t)]^2\,dt = E_1 + E_0 - 2E_{10}$$

where

$$E_1 \triangleq \int_0^{T_b} s_1^2(t)\,dt \qquad E_0 \triangleq \int_0^{T_b} s_0^2(t)\,dt$$
$$E_{10} \triangleq \int_0^{T_b} s_1(t)s_0(t)\,dt \tag{7}$$

We identify E_1 and E_0 as the respective energies of $s_1(t)$ and $s_0(t)$, while E_{10} is proportional to their correlation coefficient. Since zeros and ones are equally likely, the average signal energy per bit is

$$E_b = \tfrac{1}{2}(E_1 + E_0)$$

Therefore,

$$\left(\frac{z_1 - z_0}{2\sigma}\right)^2_{max} = \frac{E_1 + E_0 - 2E_{10}}{2\eta} = \frac{E_b - E_{10}}{\eta} \tag{8a}$$

and

$$P_e = Q[\sqrt{(E_b - E_{10})/\eta}] \tag{8b}$$

Equation (8) brings out the importance of E_{10} relative to system performance when E_b and η are fixed.

Finally, substituting Eq. (6) into Eq. (3) yields $z_1 = K(E_1 - E_{10})$ and $z_0 = K(E_{10} - E_0)$, so

$$V_{opt} = \frac{1}{2}(z_1 + z_0) = \frac{K}{2}(E_1 - E_0) \tag{9}$$

Note that the optimum threshold does not involve E_{10}.

Exercise 14.2-1 Derive Eqs. (5) and (6) from Eqs. (4*a*) and (4*b*). Use Eq. (17), Sect. 3.5, written in the form

$$\frac{\left|\int_{-\infty}^{\infty} V(\lambda)W^*(\lambda)\,d\lambda\right|^2}{\int_{-\infty}^{\infty}|W(\lambda)|^2\,d\lambda} \le \int_{-\infty}^{\infty}|V(\lambda)|^2\,d\lambda$$

and recall that the equality holds when $V(\lambda)$ and $W(\lambda)$ are proportional functions.

Coherent OOK, PRK, and FSK

Although the crude nature of ASK hardly warrants sophisticated system design, a brief look at coherent on-off keying helps clarify optimum detection concepts. The OOK signaling waveforms are just

$$s_1(t) = A_c\, p_{T_b}(t) \cos \omega_c t \qquad s_0(t) = 0 \tag{10}$$

Our carrier-frequency condition $f_c = N_c/T_b$ means that $s_1(t - kT_b) = A_c \cos \omega_c t$ for any bit interval while, of course, $s_0(t - kT_b) = 0$. Thus, a receiver with correlation detection simplifies to the form of Fig. 14.2-4, in which a local oscillator synchronized with the carrier provides the stored copy of $s_1(t)$. The bit sync signal actuates the sample-and-hold unit and resets the integrator. Both sync signals may be derived from a single source, thanks to the harmonic relationship between f_c and r_b.

Now we use Eqs. (7) and (10) to obtain $E_0 = E_{10} = 0$ and

$$E_1 = A_c^2 \int_0^{T_b} \cos^2 \omega_c t\, dt = \frac{A_c^2\, T_b}{2}\left[1 + \operatorname{sinc} \frac{4f_c}{r_b}\right] = \frac{A_c^2\, T_b}{2}$$

so $E_b = E_1/2 = A_c^2\, T_b/4$. Setting the threshold at $V = K(E_1 - E_0)/2 = KE_b$ yields the minimum average error probability given by Eq. (8), namely

$$P_e = Q(\sqrt{E_b/\eta}) = Q(\sqrt{\gamma_b}) \tag{11}$$

Not surprisingly, the performance of coherent OOK is identical to unipolar baseband transmission.

Better performance is achieved by coherent phase-reversal keying. Let the two phase shifts be 0 and π radians, so

$$s_1(t) = A_c\, p_{T_b}(t) \cos \omega_c t \qquad s_0(t) = -s_1(t) \tag{12}$$

The relation $s_0(t) = -s_1(t)$ defines *antipodal* signaling, analogous to polar baseband transmission. It quickly follows that

$$E_b = E_1 = E_0 = A_c^2\, T_b/2 \qquad E_{10} = -E_b$$

so $E_b - E_{10} = 2E_b$ and

$$P_e = Q(\sqrt{2E_b/\eta}) = Q(\sqrt{2\gamma_b}) \tag{13}$$

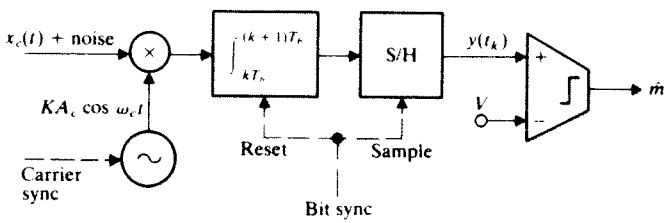


Figure 14.2-4 Correlation receiver for OOK or PRK.

PRK therefore gets by with 3 dB less signal energy than OOK, other factors being equal.

Since $s_0(t) = -s_1(t)$, a coherent PRK receiver requires only one matched filter or correlator, just like OOK. But now $V = 0$ since $E_1 = E_0$, so the PRK threshold level need not be readjusted if the received signal undergoes fading. Furthermore, the constant envelope of PRK makes it relatively invulnerable to nonlinear distortion. PRK is therefore superior to OOK on several counts, and has the same spectral efficiency. We'll see next that PRK is also superior to binary FSK.

Consider binary FSK with frequency shift $\pm f_d$ and signaling waveforms

$$\begin{aligned} s_1(t) &= A_c\, p_{T_b}(t) \cos 2\pi(f_c + f_d)t \\ s_0(t) &= A_c\, p_{T_b}(t) \cos 2\pi(f_c - f_d)t \end{aligned} \tag{14}$$

When $f_c \pm f_d \gg r_b$, $E_b \approx A_c^2\, T_b/2$, whereas

$$E_{10} = E_b \operatorname{sinc}(4f_d/r_b) \tag{15}$$

which depends on the frequency shift. If $f_d = r_b/2$, corresponding to Sunde's FSK, then $E_{10} = 0$ and the error probability is the same as OOK.

Some improvement is possible when phase discontinuities are allowed in $x_c(t)$, but $E_b - E_{10} \leq 1.22E_b$ for any choice of f_d. Hence, binary FSK does not provide any significant wideband noise reduction, and PRK has an energy advantage of at least 10 log (2/1.22) $\approx$ 2 dB. Additionally, an optimum FSK receiver is more complicated than Fig. 14.2-4.

Exercise 14.2-2 Suppose the optimum receiver for Sunde's FSK is implemented in the form of Fig. 14.2-3*a*. Find and sketch the amplitude response of the two filters.

Timing and Synchronization

Finally, we should give some attention to the timing and synchronization problems associated with optimum coherent detection. For this purpose, consider the bandpass signaling waveform and matched filter

$$\begin{aligned} s(t) &= A_c\, p_{T_b}(t) \cos \omega_c t \qquad f_c\, T_b = N_c \gg 1 \\ h(t) &= Ks(T_b - t) = KA_c\, p_{T_b}(t) \cos \omega_c t \end{aligned}$$

When $s(t)$ is applied to its matched filter, the resulting response is

$$z(t) = s(t) * h(t) \approx KE\Lambda\left(\frac{t - T_b}{T_b}\right) \cos \omega_c t \tag{16}$$

where $E = A_c^2\, T_b/2$. The sketch of $z(t)$ in Fig. 14.2-5 shows the expected maximum value $z(T_b) = KE$, and the response for $t > T_b$ would be eliminated by discharging the filter after the sampling instant.

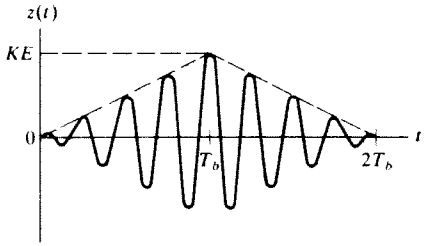


Figure 14.2-5 Response of bandpass matched filter.

But suppose there's a small timing error such that sampling actually occurs at $t_k = T_b(1 \pm \epsilon)$. Then

$$z(t_k) \approx KE \cos \theta_\epsilon \qquad \theta_\epsilon = \omega_c T_b \epsilon = 2\pi N_c \epsilon$$

so the timing error reduces the effective signal level by the factor $\cos \theta_\epsilon$. Since $|z_1 - z_0|^2$ will be reduced by $\cos^2 \theta_\epsilon$, while σ^2 remains unchanged, the error probability becomes

$$P_e = Q\left(\sqrt{\frac{E_b - E_{10}}{\eta} \cos^2 \theta_\epsilon}\right) \tag{17}$$

which follows from Eq. (8). As an example of the magnitude of this problem, take PRK with $\gamma_b = 8$, $r_b = 2$ kbps, and $f_c = 100$ kHz; perfect timing gives $P_e = Q(\sqrt{16}) \approx 3 \times 10^{-5}$, while an error of just 0.3% of the bit interval results in $\theta_\epsilon = 2\pi(100/2) \times 0.003 = 54°$ and $P_e = Q(\sqrt{16 \cos^2 54°}) \approx 10^{-2}$. These numbers illustrate why bandpass matched filter is not a practical method for coherent detection.

A correlation detector like Fig. 14.2-4 has much less sensitivity to timing error, since the integrated output does not oscillate at the carrier frequency. Correlation detection is therefore used in most coherent binary systems. However, the local oscillator must be synchronized accurately with the carrier, and a phase synchronization error θ_ϵ again reduces the effective signal level by the factor $\cos \theta_\epsilon$.

In the case of PRK, the carrier sync signal can be derived from $x_c(t)$ using techniques such as the Costas PLL system back in Fig. 8.3-4. Another approach known as *phase-comparison detection* is discussed in the next section, along with noncoherent detection of OOK and FSK.

14.3 NONCOHERENT BINARY SYSTEMS

Optimum coherent detection may not be essential if the signal is strong enough for adequate reliability with a less sophisticated receiver. A prime example of this situation is digital transmission over voice telephone channels, which have relatively large signal-to-noise ratios dictated by analog performance standards. There are also applications in which it would be very difficult and expensive to carry out coherent detection. For instance, the propagation delay on some radio channels changes too rapidly to permit accurate tracking of the carrier phase at the receiver, and unsynchronized or noncoherent detection becomes the only viable recourse.

Here we examine the suboptimum performance of noncoherent OOK and FSK systems that employ envelope detection to bypass the synchronization problems of coherent detection. We'll also look at *differentially coherent* PSK systems with phase-comparison detection. For all three cases we must first analyze the envelope of a sinusoid plus bandpass noise.

Envelope of a Sinusoid plus Bandpass Noise

Consider the sinusoid $A_c \cos(\omega_c t + \theta)$ plus gaussian bandpass noise $n(t)$ with zero mean and variance σ^2. Using the quadrature-carrier expression

$$n(t) = n_i(t) \cos(\omega_c t + \theta) - n_q(t) \sin(\omega_c t + \theta)$$

we write the sum as

$$A_c \cos(\omega_c t + \theta) + n(t) = A(t) \cos[\omega_c t + \theta + \phi(t)]$$

where, at any instant t,

$$A = \sqrt{(A_c + n_i)^2 + n_q^2} \qquad \phi = \arctan \frac{n_q}{A_c + n_i} \tag{1}$$

We recall from Sect. 9.2 that the i and q noise components are independent r.v.'s having the same distribution as $n(t)$. Now we seek the PDF of the envelope A.

Before plunging into the analysis, let's speculate on the nature of A under extreme conditions. If $A_c = 0$, then A reduces to the noise envelope A_n, with the Rayleigh distribution

$$p_{A_n}(A_n) = \frac{A_n}{\sigma^2} e^{-A_n^2/2\sigma^2} \qquad A_n \geq 0 \tag{2}$$

At the other extreme, if $A_c \gg \sigma$, then A_c will be large compared to the noise components most of the time, so

$$A = A_c \sqrt{1 + (2n_i/A_c) + (n_i^2 + n_q^2)/A_c^2} \approx A_c + n_i$$

which implies that A will be approximately gaussian.

For an arbitrary value of A_c, we must perform a rectangular-to-polar conversion following the procedure that led to Eq. (14), Sect. 4.4. The joint PDF of A and ϕ then becomes

$$p_{A\phi}(A, \phi) = \frac{A}{2\pi\sigma^2} \exp\left(-\frac{A^2 - 2A_c A \cos\phi + A_c^2}{2\sigma^2}\right) \tag{3}$$

for $A \geq 0$ and $|\phi| \leq \pi$. The term $A \cos\phi$ in the exponent prevents us from factoring Eq. (3) as a product of the form $p_A(A)p_\phi(\phi)$, meaning that A and ϕ are not

statistically independent. The envelope PDF must therefore be found by integrating the joint PDF over the range of ϕ, so

$$p_A(A) = \frac{A}{2\pi\sigma^2} \exp\left(-\frac{A^2 + A_c^2}{2\sigma^2}\right) \int_{-\pi}^{\pi} \exp\left(\frac{A_c A \cos\phi}{\sigma^2}\right) d\phi$$

Now we introduce the *modified Bessel function* of the first kind and order zero, defined by

$$I_0(v) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(v \cos\phi)\, d\phi \tag{4a}$$

with the properties

$$I_0(v) \approx \begin{cases} e^{v^2/4} & v \ll 1 \\ \dfrac{e^v}{\sqrt{2\pi v}} & v \gg 1 \end{cases} \tag{4b}$$

We then have

$$p_A(A) = \frac{A}{\sigma^2} e^{-(A^2 + A_c^2)/2\sigma^2} I_0\left(\frac{A_c A}{\sigma^2}\right) \qquad A \geq 0 \tag{5}$$

which is called the *Rician distribution.*

Although Eq. (5) has a formidable appearance, it easily simplifies under large-signal conditions to

$$p_A(A) \approx \sqrt{\frac{A}{2\pi A_c \sigma^2}}\, e^{-(A - A_c)^2/2\sigma^2} \qquad A_c \gg \sigma \tag{6}$$

obtained from the large-v approximation in Eq. (4*b*). Since the exponential term dominates in Eq. (6), we have confirmed that the envelope PDF is essentially a gaussian curve with variance σ^2 centered at $\bar{A} \approx A_c$. Figure 14.3-1 illustrates the transition of the envelope PDF from a Rayleigh curve to a gaussian curve as A_c becomes large compared to σ.

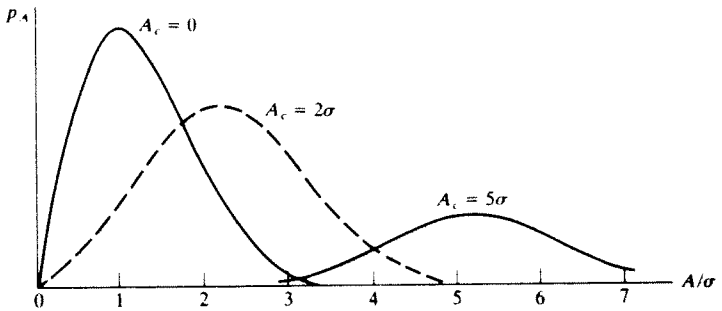


Figure 14.3-1 PDFs for the envelope of a sinusoid plus bandpass noise.

Noncoherent OOK

Noncoherent on-off keying is intended to be a simple system. Usually the carrier and data are unsynchronized so, for an arbitrary bit interval $kT_b < t < (k+1)T_b$, we write

$$x_c(t) = A_c\, a_k\, p_{T_b}(t - kT_b) \cos(\omega_c t + \theta) \qquad a_k = 0, 1 \tag{7}$$

The signaling energies are $E_0 = 0$ and

$$E_1 = \frac{A_c^2 T_b}{2}\left[1 + \frac{\sin(2\omega_c T_b + 2\theta) - \sin 2\theta}{2\omega_c T_b}\right] \approx \frac{A_c^2 T_b}{2}$$

where we've assumed that $f_c \gg r_b$. The average signal energy per bit is then $E_b = E_1/2 \approx A_c^2 T_b/4$ since we'll continue to assume that 1s and 0s are equally likely.

The OOK receiver diagrammed in Fig. 14.3-2 consists of a BPF followed by an envelope detector and regenerator. The BPF is a matched filter with

$$h(t) = K A_c\, p_{T_b}(t) \cos\omega_c t \tag{8}$$

which ignores the carrier phase θ. The envelope detector eliminates dependence on θ by tracing out the dashed line back in Fig. 14.2-5. Thus, when $a_k = 1$, the peak signal component of the envelope $y(t)$ is $A_1 = KE_1$. Let's take $K = A_c/E_1$ for convenience, so that $A_1 = A_c$. Then

$$A_c^2/\sigma^2 = 4E_b/\eta = 4\gamma_b \tag{9}$$

where σ^2 is the variance of the bandpass noise at the input to the envelope detector, calculated from $h(t)$ using Eq. (4*b*), Sect. 14.2.

Now consider the conditional PDFs of the random variable $Y = y(t_k)$. When $a_k = 0$, we have a sample value of the envelope of the noise alone; hence, $p_Y(y \mid H_0)$ is the Rayleigh function $p_{A_n}(y)$. When $a_k = 1$, we have a sample value of the envelope of a sinusoid plus noise; hence, $p_Y(y \mid H_1)$ is the Rician function $p_A(y)$. Figure 14.3-3 shows these two curves for the case of $\gamma_b \gg 1$, so the Rician PDF has a nearly gaussian shape. The intersection point defines the optimum threshold, which turns out to be

$$V_{\text{opt}} \approx \frac{A_c}{2}\sqrt{1 + \frac{2}{\gamma_b}} \approx \frac{A_c}{2} \qquad \gamma_b \gg 1$$

But we no longer have symmetry with respect to the threshold and, consequently, $P_{e1} \neq P_{e0}$ when P_e is minimum.

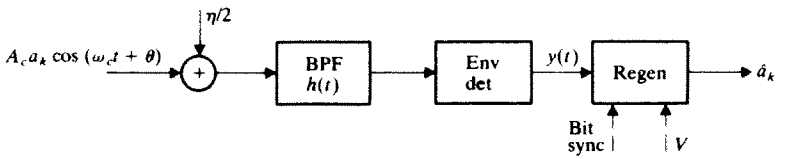


Figure 14.3-2 Noncoherent OOK receiver.

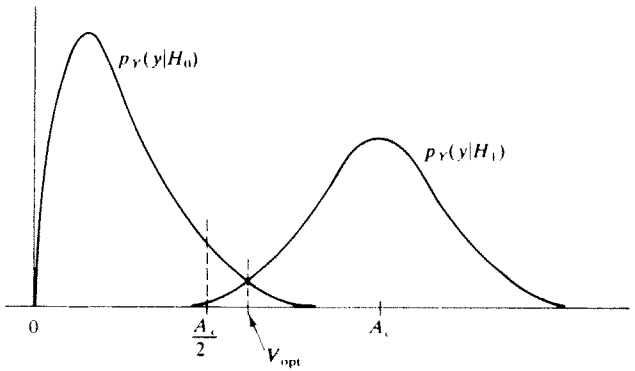


Figure 14.3-3 Conditional PDFs for noncoherent OOK.

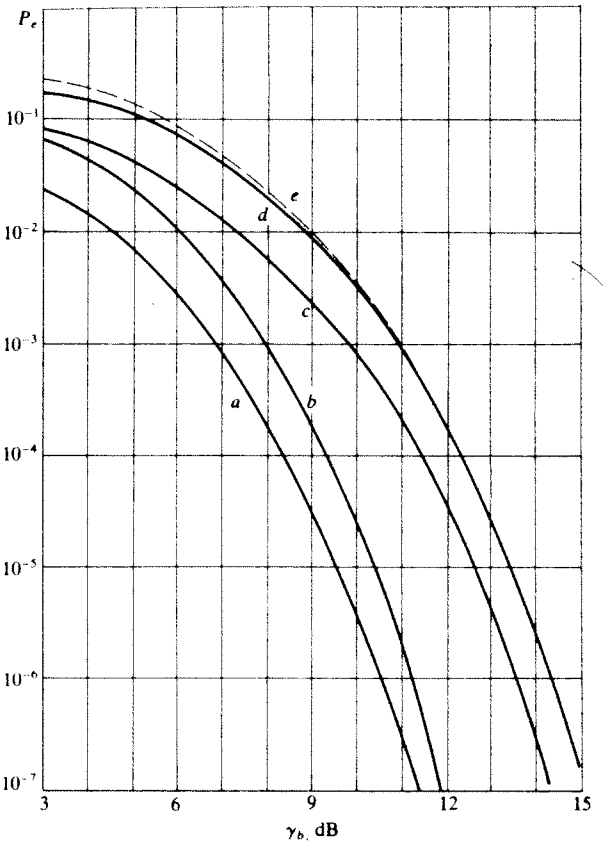


Figure 14.3-4 Binary error probability curves. (*a*) Coherent PRK; (*b*) DPSK; (*c*) coherent OOK or FSK; (*d*) noncoherent FSK; (*e*) noncoherent OOK.

Noncoherent OOK systems require $\gamma_b \gg 1$ for reasonable performance, and the threshold is normally set at $A_c/2$. The resulting error probabilities are

$$P_{e_0} = \int_{A_c/2}^{\infty} p_{A_n}(y)\, dy = e^{-A_c^2/8\sigma^2} = e^{-\gamma_b/2} \tag{10a}$$

$$P_{e_1} = \int_0^{A_c/2} p_A(y)\, dy \approx Q\left(\frac{A_c}{2\sigma}\right) = Q(\sqrt{\gamma_b}) \approx \frac{1}{\sqrt{2\pi\gamma_b}} e^{-\gamma_b/2} \qquad \gamma_b \gg 1 \tag{10b}$$

where we've introduced the asymptotic approximation for $Q(\sqrt{\gamma_b})$ to bring out the fact that $P_{e_1} \ll P_{e_0}$ when $\gamma_b \gg 1$. Finally,

$$P_e = \tfrac{1}{2}(P_{e_0} + P_{e_1}) = \tfrac{1}{2}[e^{-\gamma_b/2} + Q(\sqrt{\gamma_b})] \approx \tfrac{1}{2}e^{-\gamma_b/2} \qquad \gamma_b \gg 1 \tag{11}$$

which is plotted versus γ_b in Fig. 14.3-4 along with curves for other binary systems.

Exercise 14.3-1 Consider the BPF output $z(t) = x_c(t) * h(t)$ when $x_c(t) = A_c\, p_{T_b}(t) \cos(\omega_c t + \theta)$ and $K = 2/A_c T_b$. Show that, for $0 < t < T_b$,

$$z(t) = \frac{A_c t}{T_b}\left[\cos\theta \cos\omega_c t - \left(\sin\theta - \frac{\cos\theta}{\omega_c t}\right)\sin\omega_c t\right]$$

Then find and sketch the envelope of $z(t)$ assuming $f_c \gg r_b$.

Noncoherent FSK

Although envelope detection seems an unlikely method for FSK, a reexamination of the waveform back in Fig. 14.1-1*b* reveals that binary FSK consists of two interleaved OOK signals with the same amplitude A_c but different carrier frequencies, $f_1 = f_c + f_d$ and $f_0 = f_c - f_d$. Accordingly, noncoherent detection can be implemented with a pair of bandpass filters and envelope detectors, arranged per Fig. 14.3-5 where

$$h_1(t) = KA_c\, p_{T_b}(t) \cos\omega_1 t \qquad h_0(t) = KA_c\, p_{T_b}(t) \cos\omega_0 t \tag{12}$$

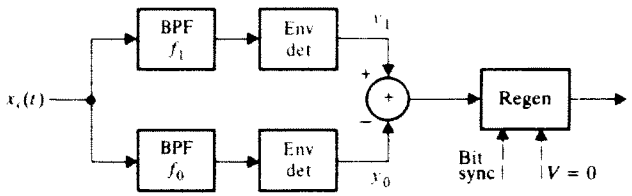


Figure 14.3-5 Noncoherent detection of binary FSK.

We'll take $K = A_c/E_b$, noting that $E_b = E_1 = E_0 \approx A_c^2 T_b/2$. Then

$$A_c^2/\sigma^2 = 2E_b/\eta = 2\gamma_b \tag{13}$$

where σ^2 is the noise variance at the output of either filter.

We'll also take the frequency spacing $f_1 - f_0 = 2f_d$ to be an integer multiple of r_b, as in Sunde's FSK. This condition ensures that the BPFs effectively separate the two frequencies, and that the two bandpass noise waveforms are uncorrelated at the sampling instants. Thus, when $a_k = 1$, the sampled output $y_1(t_k)$ at the upper branch has the signal component $A_1 = KE_1 = A_c$ and a Rician distribution, whereas $y_0(t_k)$ at the lower branch has a Rayleigh distribution — and vice versa when $a_k = 0$.

Regeneration is based on the envelope difference $Y_1 - Y_0 = y_1(t_k) - y_0(t_k)$. Without resorting to conditional PDFs, we conclude from the symmetry of the receiver that the threshold should be set at $V = 0$, regardless of A_c. It then follows that $P_{e_1} = P(Y_1 - Y_0 < 0 \mid H_1)$ and $P_{e_0} = P_{e_1} = P_e$. Therefore,

$$\begin{aligned} P_e &= P(Y_0 > Y_1 \mid H_1) \\ &= \int_0^\infty p_{Y_1}(y_1 \mid H_1)\left[\int_{y_1}^\infty p_{Y_0}(y_0 \mid H_1)\, dy_0\right] dy_1 \end{aligned}$$

where the inner integral is the probability of the event $Y_0 > Y_1$ for a *fixed* value of y_1. Inserting the PDFs $p_{Y_0}(y_0 \mid H_1) = p_{A_n}(y_0)$ and $p_{Y_1}(y_1 \mid H_1) = p_A(y_1)$ and performing the inner integration yields

$$P_e = \int_0^\infty \frac{y_1}{\sigma^2}\, e^{-(2y_1^2 + A_c^2)/2\sigma^2} I_0\left(\frac{A_c\, y_1}{\sigma^2}\right) dy_1$$

Rather amazingly, this integral can be evaluated in closed form by letting $\lambda = \sqrt{2}\, y_1$ and $\alpha = A_c/\sqrt{2}$ so that

$$P_e = \frac{1}{2}\, e^{-A_c^2/4\sigma^2} \int_0^\infty \frac{\lambda}{\sigma^2}\, e^{-(\lambda^2 + \alpha^2)/2\sigma^2} I_0\left(\frac{\alpha\lambda}{\sigma^2}\right) d\lambda$$

The integrand is now exactly the same function as the Rician PDF in Eq. (5), whose total area equals unity. Hence, our final result simply becomes

$$P_e = \tfrac{1}{2} e^{-A_c^2/4\sigma^2} = \tfrac{1}{2} e^{-\gamma_b/2} \tag{14}$$

having used Eq. (13).

A comparison of the performance curves for noncoherent FSK and OOK plotted in Fig. 14.3-4 reveals little difference except at small values of γ_b. However, FSK does have three advantages over OOK: constant modulated signal envelope, equal digit error probabilities, and fixed threshold level $V = 0$. These advantages usually justify the extra hardware needed for the FSK receiver.

Differentially Coherent PSK

Noncoherent detection of binary PSK would be impossible since the message information resides in the phase. Instead, the clever technique of *phase-comparison detection* gets around the phase synchronization problems associated with coherent PRK and provides much better performance than noncoherent OOK or FSK. The phase-comparison detector in Fig. 14.3-6 looks something like a correlation detector except that the local oscillator signal is replaced by the PRK signal itself after a delay of T_b. A BPF at the front end prevents excess noise from swamping the detector.

Successful operation requires f_c to be an integer multiple of r_b, as in coherent PRK. We therefore write

$$\begin{aligned} x_c(t) &= A_c\, p_{T_b}(t - kT_b) \cos(\omega_c t + \theta + a_k \pi) \\ a_k &= 0, 1 \qquad kT_b < t < (k+1)T_b \end{aligned} \tag{15}$$

In the absense of noise, the phase-comparison product for the kth bit interval is

$$\begin{aligned} x_c(t) \times 2x_c(t - T_b) &= 2A_c^2 \cos(\omega_c t + \theta + a_k \pi) \\ &\quad \times \cos[\omega_c(t - T_b) + \theta + a_{k-1}\pi] \\ &= A_c^2\{\cos[(a_k - a_{k-1})\pi] \\ &\quad + \cos[2\omega_c t + 2\theta + (a_k + a_{k-1})\pi]\} \end{aligned}$$

where we've used the fact that $\omega_c T_b = 2\pi N_c$. Lowpass filtering then yields

$$z(t_k) = \begin{cases} +A_c^2 & a_k = a_{k-1} \\ -A_c^2 & a_k \neq a_{k-1} \end{cases} \tag{16}$$

so we have polar symmetry and the threshold should be set at $V = 0$.

Since $z(t_k)$ only tells you whether a_k differs from a_{k-1}, a PRK system with phase-comparison detection is called *differentially coherent* PSK (DPSK). Such systems generally include *differential encoding* at the transmitter, which makes it possible to regenerate the message bits directly from $z(t_k)$. Differential encoding starts with an arbitrary initial bit, say $a_0 = 1$. Subsequent bits are determined by the message sequence m_k according to the rule: $a_k = a_{k-1}$ if $m_k = 1$, $a_k \neq a_{k-1}$ if

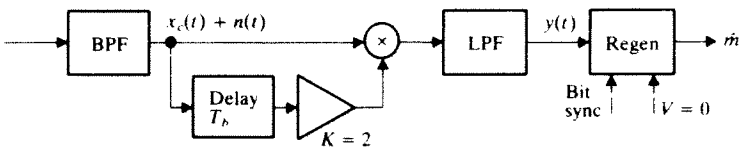


Figure 14.3-6 Differentially coherent receiver for binary PSK.

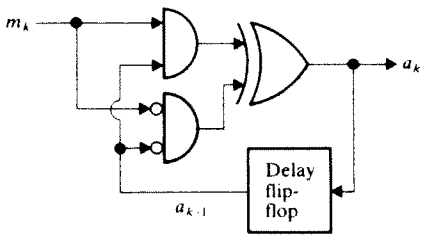


Figure 14.3-7 Logic circuit for differential encoding.

$m_k = 0$. Thus, $z(t_k) = +A_c^2$ means that $m_k = 1$ and $z(t_k) = -A_c^2$ means that $m_k = 0$. Figure 14.3-7 shows a logic circuit for differential encoding; this circuit implements the logic equation

$$a_k = a_{k-1}m_k \oplus \bar{a}_{k-1}\bar{m}_k \tag{17}$$

where the overbar stands for logical inversion. An example of differential encoding and phase-comparison detection (without noise) is given in Table 14.3-1.

To analyze the performance of DPSK with noise, we'll assume that the BPF performs most of the noise filtering, like the BPFs in an FSK receiver. Hence, the carrier amplitude and noise variance at the BPF output are related by

$$A_c^2/\sigma^2 = 2E_b/\eta = 2\gamma_b$$

We'll also exploit the symmetry and focus on the case when $a_k = a_{k-1} = 0$, so an error occurs if $y(t_k) < 0$.

Now let the delayed i and q noise components be denoted by $n_i'(t) = n_i(t - T_b)$ and $n_q'(t) = n_q(t - T_b)$. The inputs to the multiplier during the kth bit interval are $x_c(t) + n(t) = [A_c + n_i(t)] \cos(\omega_c t + \theta) - n_q(t) \sin(\omega_c t + \theta)$ and $2[x_c(t - T_b) + n(t) - T_b)] = 2[A_c + n_i'(t)] \cos(\omega_c t + \theta) - 2n_q'(t) \sin(\omega_c t + \theta)$. The LPF then removes the high-frequency terms from the product, leaving

$$Y = y(t_k) = (A_c + n_i)(A_c + n_i') + n_q n_q' \tag{18}$$

where all four noise components are independent gaussian r.v.'s with zero mean and variance σ^2.

Equation (18) has a *quadratic form* that can be simplified by a diagonalization process, resulting in

$$Y = \alpha^2 - \beta^2 \tag{19a}$$

with

$$\alpha^2 = (A_c + \alpha_i)^2 + \alpha_q^2 \qquad \beta^2 = \beta_i^2 + \beta_q^2 \tag{19b}$$

Table 14.3-1

Input message		1	0	1	1	0	1	0	0
Encoded message	1	1	0	0	0	1	1	0	1
Transmitted phase	π	π	0	0	0	π	π	0	π
Phase-comparison sign		+	−	+	+	−	+	−	−
Regenerated message		1	0	1	1	0	1	0	0

and

$$\begin{aligned} \alpha_i &\triangleq \tfrac{1}{2}(n_i + n_i') & \beta_i &\triangleq \tfrac{1}{2}(n_i - n_i') \\ \alpha_q &\triangleq \tfrac{1}{2}(n_q + n_q') & \beta_q &\triangleq \tfrac{1}{2}(n_q - n_q') \end{aligned} \tag{19c}$$

Note that α_i is a zero-mean gaussian r.v. with variance $\overline{\alpha_i^2} = (\overline{n_i^2} + \overline{n_i'^2})/4 = 2\sigma^2/4 = \sigma^2/2$; identical conclusions hold for the other i and q components of α and β. Therefore, α has a *Rician* PDF given by Eq. (5) with $\sigma^2/2$ in place of σ^2, while β has a *Rayleigh* PDF given by Eq. (2) with $\sigma^2/2$ in place of σ^2.

Lastly, since α and β are nonnegative, we can write the average error probability as

$$P_e = P(Y < 0 \mid a_k = a_{k-1}) = P(\alpha^2 < \beta^2) = P(\beta > \alpha)$$

and we've arrived at an expression equivalent to the one previously solved for noncoherent FSK. Substituting $\sigma^2/2$ for σ^2 in Eq. (14) now gives our DPSK result

$$P_e = \tfrac{1}{2}e^{-A_c^2/2\sigma^2} = \tfrac{1}{2}e^{-\gamma_b} \tag{20}$$

The performance curves in Fig. 14.3-4 now show that DPSK has a 3-dB energy advantage over noncoherent binary systems and a penalty of less than 1 dB compared to coherent PRK at $P_e \leq 10^{-4}$.

DPSK does not require the carrier phase synchronization essential for coherent PRK, but it does involve somewhat more hardware than noncoherent OOK or FSK — including differential encoding and carrier-frequency synchronization with r_b at the transmitter. A minor annoyance is that DPSK errors tend to occur in groups of two (why?).

Example 14.3-1 Binary data is to be sent at the rate $r_b = 100$ kbps over a channel with 60-dB transmission loss and noise density $\eta = 10^{-12}$ W/Hz at the receiver. What transmitted power S_T is needed to get $P_e = 10^{-3}$ for various types of modulation and detection?

To answer this question, we first write the received signal power as $S_R = E_b r_b = \eta\gamma_b r_b = S_T/L$ with $L = 10^6$. Thus,

$$S_T = L\eta\gamma_b r_b = 0.1\gamma_b$$

Next, using the curves in Fig. 14.3-4 or our previous formulas for P_e, we find the value of γ_b corresponding to the specified error probability and calculate S_T therefrom.

Table 14.3-2 summarizes the results. The systems have been listed here in order of increasing difficulty of implementation, bringing out the trade-off between signal power and hardware complexity.

Table 14.3-2

System	S_T, W
Noncoherent OOK or FSK	1.26
Differentially coherent PSK	0.62
Coherent PRK	0.48

Exercise 14.3-2 Suppose the system in the previous example has a limitation on the *peak envelope power*, such that $LA_c^2 \leq 2$ watts at the transmitter. Find the resulting minimum error probability for noncoherent OOK and FSK and for DPSK.

14.4 QUADRATURE-CARRIER AND *M*-ARY SYSTEMS

This section investigates the performance of *M*-ary modulation systems with coherent or phase-comparison detection, usually in a quadrature-carrier configuration. Our primary motivation here is the increased modulation speed afforded by QAM and related quadrature-carrier methods, and by *M*-ary PSK and APK modulation. These are the modulation types best suited to digital transmission on telephone lines and other bandwidth-limited channels. (Appendix C discusses wideband modulation, such as *M*-ary FSK, which is better suited to power-limited applications.)

As in previous sections, we continue to assume independent equiprobable symbols and AWGN contamination. We also assume that M is a power of two, consistent with binary to *M*-ary data conversion. This assumption allows a practical comparison of binary and *M*-ary systems.

Quadrature-Carrier Systems

We pointed out in Sect. 14.1 that both quadriphase PSK and keyed polar QAM are equivalent to the sum of two PRK signals impressed on quadrature carriers. Here we'll adopt that viewpoint to analyze the performance of QPSK/QAM with coherent detection. Accordingly, let the source information be grouped into *dibits* represented by $I_k Q_k$. Each dibit corresponds to one symbol from a quaternary ($M = 4$) source or two successive bits from a binary source. In the latter case, which occurs more often in practice, the dibit rate is $r = r_b/2$ and $D = 1/r = 2T_b$.

Coherent quadrature-carrier detection requires synchronized modulation, as discussed in Sect. 14.2. Thus, for the kth dibit interval $kD < t < (k+1)D$, we write

$$x_c(t) = s_i(t - kD) - s_q(t - kD) \tag{1a}$$

with

$$\begin{aligned} s_i(t) &= A_c I_k p_D(t) \cos \omega_c t \qquad I_k = \pm 1 \\ s_q(t) &= A_c Q_k p_D(t) \sin \omega_c t \qquad Q_k = \pm 1 \end{aligned} \tag{1b}$$

Since f_c must be harmonically related to $r = 1/D$, the signaling energy is

$$\int_{kD}^{(k+1)D} x_c^2(t)\, dt = \tfrac{1}{2}A_c^2(I_k^2 + Q_k^2)D = A_c^2 D$$

and we have

$$E = 2E_b \qquad E_b = A_c^2 D/2 \tag{2}$$

where E is the energy per dibit or quaternary symbol.

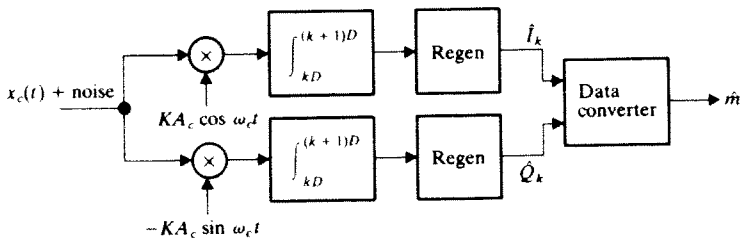


Figure 14.4-1 Quadrature-carrier receiver with correlation detectors.

From Eq. (1) and our prior study of coherent PRK, it follows that the optimum quadrature-carrier receiver can be implemented with two correlation detectors arranged as in Fig. 14.4-1. Each correlator performs coherent binary detection, independent of the other. Hence, the average error probability *per bit* is

$$P_{be} = Q(\sqrt{2E_b/\eta}) = Q(\sqrt{2\gamma_b}) \tag{3}$$

where the function $Q(\sqrt{2\gamma_b})$ denotes the area under the gaussian tail — not to be confused with Q symbolizing quadrature modulation.

We see from Eq. (3) that coherent QPSK/QAM achieves the same bit error probability as coherent PRK. But recall that the transmission bandwidth for QPSK/QAM is

$$B_T \approx r_b/2$$

whereas PRK requires $B_T \approx r_b$. This means that the additional quadrature-carrier hardware allows you to cut the transmission bandwidth in half for a given bit rate or to double the bit rate for a given transmission bandwidth. The error probability remains unchanged in either case.

Equation (3) and the bandwidth/hardware trade-off also hold for *minimum-shift keying*, whose i and q components illustrated back in Fig. 14.1-11*b* suggest quadrature-carrier detection. An MSK receiver has a structure like Fig. 14.4-1 modified in accordance with the pulse shaping and staggering of the i and q components. There are only two significant differences between MSK and QPSK: (1) the MSK spectrum has a broader main lobe but smaller side lobes than the spectrum of QPSK with the same bit rate; (2) MSK is inherently binary frequency modulation, whereas QPSK can be viewed as either binary or quaternary phase modulation.

When QPSK/QAM is used to transmit quaternary data, the output converter in Fig. 14.4-1 reconstructs quaternary symbols from the regenerated dibits. Since bit errors are independent, the probability of obtaining a *correct* symbol is

$$P_c = (1 - P_{be})^2$$

The average error probability *per symbol* thus becomes

$$\begin{aligned} P_e = 1 - P_c &= 2Q(\sqrt{E/\eta}) - Q^2(\sqrt{E/\eta}) \\ &\approx 2Q(\sqrt{E/\eta}) \qquad E/\eta \gg 1 \end{aligned} \tag{4}$$

where $E = 2E_b$ represents the average symbol energy.

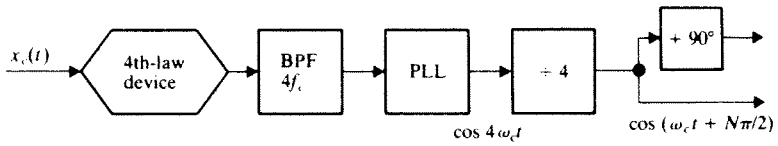


Figure 14.4-2 PLL system for carrier synchronization in a quadrature-carrier receiver.

Various methods have been devised to generate the carrier sync signals necessary for coherent detection in quadrature-carrier receivers. Figure 14.4-2 shows a simple PLL system based on the fact that the fourth power of $x_c(t)$ contains a discrete frequency component at $4f_c$. However, since $\cos 4\omega_c t = \cos(4\omega_c t + 2\pi N)$, fourfold frequency division produces $\cos(\omega_c t + N\pi/2)$ so the output has a fixed phase error of $N\pi/2$ with N being an integer whose value depends on the lock-in transient. A known preamble may be transmitted at the start of the message to permit phase adjustment, or differential encoding may be used to nullify the phase error effects. Another carrier sync system will be described in conjunction with M-ary PSK; additional methods are covered by Lindsey (1972).

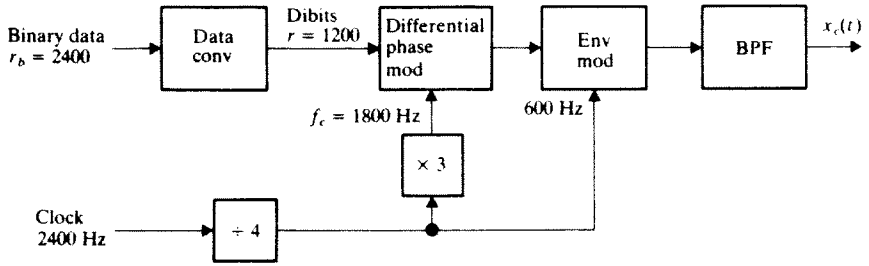
Phase-comparison detection is also possible in quadrature-carrier systems with differential encoding. From our study of DPSK in Sect. 14.3, you may correctly infer that differentially coherent QPSK (DQPSK) requires somewhat more signal energy than coherent QPSK to get a specified error probability. The difference turns out to be about 2.3 dB.

Example 14.4-1 DQPSK Telephone modem Digital communication over telephone lines has been the subject of intense effort for many years. One of the earliest and most successful systems is the AT & T series 201 modem, which employs DQPSK to achieve binary transmission at $r_b = 2400$ bps within the limitations of a voice channel. (See Table 12.4-2 for a listing of other modems.) This modem is described here to illustrate how basic design concepts are augmented in practical implementation.

Each modem consists of a modulator unit and a demodulator unit to allow transmission in either direction. The modulator is diagrammed in Fig. 14.4-3*a*, where incoming binary digits are converted to dibits that modulate the carrier at half the bit rate. The process of differential phase modulation incorporates both differential encoding and phase-shift keying, as indicated in Fig. 14.4-3*b*. The carrier frequency is $f_c = 1800$ Hz, derived from the 2400-Hz clock signal.

Although the nominal transmission bandwidth for DQPSK is $B_T \approx r_b/2 = 1200$ Hz, the spectrum has considerable spillover that would exceed the voice channel bandwidth. Synchronized envelope modulation at 600 Hz provides pulse shaping to reduce spillover outside $f_c \pm r_b/2$. Bandpass filtering finally yields an output spectrum confined to the range 600–3000 Hz.

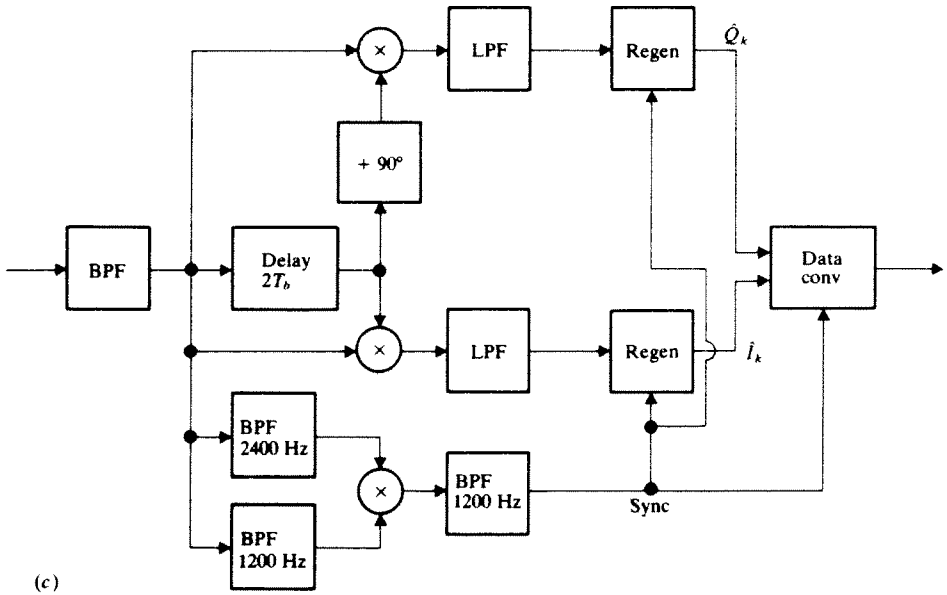
Observe in Fig. 14.4-3*b* that the differential phase shift is never zero, so ϕ_k always differs from ϕ_{k-1} and there will be a phase shift in the modulated



(*a*)

Dibit	$\phi_k - \phi_{k-1}$	Q_k	I_k
0 0	$+\pi/4$	+	+
0 1	$+3\pi/4$	+	−
1 0	$-3\pi/4$	−	−
1 1	$-\pi/4$	−	+

(*b*)



(*c*)

Figure 14.4-3 DQPSK telephone modem. (*a*) Modulator; (*b*) differential encoding table; (*c*) demodulator.

wave every $D = 1/1200$ seconds. When combined with the envelope modulation, these phase shifts produce discrete frequency components at $f_c + 600 = 2400$ Hz and $f_c - 600 = 1200$ Hz used for regeneration timing.

The demodulator unit in Fig. 14.4-3*c* has quadrature phase-comparison detection with outputs proportional to $I_k = \sqrt{2} \cos(\phi_k - \phi_{k-1})$ and $Q_k = \sqrt{2} \sin(\phi_k - \phi_{k-1})$, from which the message bits are regenerated. The dibit sync signal is obtained by mixing the 2400- and 1200-Hz components and selecting the 1200-Hz difference frequency. This roundabout method, rather than directly filtering the 1200-Hz component, takes better account of the delay distortion induced by a typical telephone channel.

Tests have shown that the resulting error probability is $P_{be} \approx 10^{-5}$ with $\gamma_b = 15$ dB. A fully optimized DQPSK system under ideal conditions would yield the same performance with about 3 dB less energy per bit. Of course a telephone channel or any other real channel does not provide ideal conditions—a fact of life that must be anticipated by the design engineer.

Exercise 14.4-1 Consider a QPSK signal like Eq. (1) written as $x_c(t) = A_c \cos(\omega_c t + \phi_k)$ with $\phi_k = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$. Show that $x_c^4(t)$ includes an unmodulated component at $4f_c$.

M-ary PSK Systems

Now let's extend our investigation of coherent quadrature-carrier detection to encompass M-ary PSK. The carrier is again synchronized with the modulation, and f_c is harmonically related to the symbol rate r. We write the modulated signal for a given symbol interval as

$$x_c(t) = s_i(t - kD) - s_q(t - kD) \tag{5a}$$

with

$$\begin{aligned} s_i(t) &= A_c \cos \phi_k \, p_D(t) \cos \omega_c t \\ s_q(t) &= A_c \sin \phi_k \, p_D(t) \sin \omega_c t \end{aligned} \tag{5b}$$

where

$$\phi_k = 2\pi a_k / M \qquad a_k = 0, 1, \ldots, M - 1$$

from Eq. (13), Sect. 14.1, taking $N = 0$. The signaling energy per symbol then becomes

$$E = \tfrac{1}{2} A_c^2 (\cos^2 \phi_k + \sin^2 \phi_k) D = \tfrac{1}{2} A_c^2 D \tag{6}$$

equivalent to $E_b = E/\log_2 M$ if each symbol represents $\log_2 M$ binary digits. The transmission bandwidth requirement is $B_T \approx r = r_b/\log_2 M$, from our spectral analysis in Sect. 14.1.

An optimum receiver for M-ary PSK can be modeled in the form of Fig. 14.4-4. We'll let $K = A_c/E$ so, in absence of noise, the quadrature correlators produce $z_i(t_k) = A_c \cos \phi_k$ and $z_q(t_k) = A_c \sin \phi_k$ from which $\phi_k = \arctan z_q/z_i$.

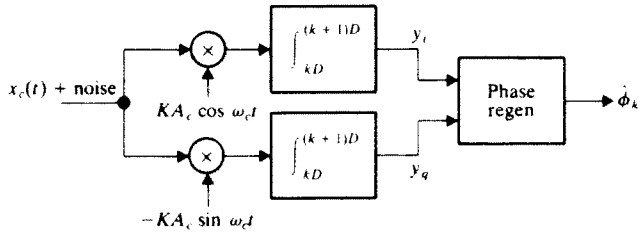


Figure 14.4-4 Coherent M-ary PSK receiver.

When $x_c(t)$ is contaminated by noise, message symbol regeneration is based on the noisy samples

$$y_i = A_c \cos \phi_k + n_i \qquad y_q = A_c \sin \phi_k + n_q$$

in which the i and q noise components are independent gaussian r.v.'s with zero mean and variance

$$\sigma^2 = K^2 E \eta/2 = A_c^2 \eta/2E = \eta r \tag{7}$$

The generator has M *angular* thresholds equispaced by $2\pi/M$, as illustrated in Fig. 14.4-5, and it selects the point from the signal constellation whose angle is closest to arctan y_q/y_i.

The circular symmetry of Fig. 14.4-5, together with the symmetry of the noise PDFs, means that all phase angles have the same error probability. We'll therefore focus on the case of $\phi_k = 0$, so

$$\arctan \frac{y_q}{y_i} = \arctan \frac{n_q}{A_c + n_i} = \phi$$

and we recognize ϕ as the phase of a sinusoid plus bandpass noise. Since no error results if $|\phi| < \pi/M$, the symbol error probability can be calculated using

$$P_e = P(|\phi| > \pi/M) = 1 - \int_{-\pi/M}^{\pi/M} p_\phi(\phi) \, d\phi \tag{8}$$

for which we need the PDF of the phase ϕ.

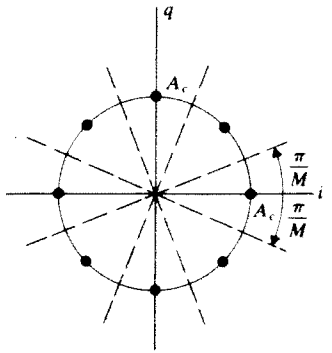


Figure 14.4-5 Decision thresholds for M-ary PSK.

The joint PDF for the envelope and phase of a sinusoid plus bandpass noise was given in Eq. (3), Sect. 14.3. The PDF of the phase alone is found by integrating the joint PDF over $0 \le A < \infty$. A few manipulations lead to the awesome-looking expression

$$p_\phi(\phi) = \frac{1}{2\pi} e^{-A_c^2/2\sigma^2} + \frac{A_c \cos\phi}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{A_c^2 \sin^2\phi}{2\sigma^2}\right)\left[1 - Q\left(\frac{A_c \cos\phi}{\sigma}\right)\right] \quad (9)$$

for $-\pi < \phi < \pi$. Under the large-signal condition $A_c \gg \sigma$, Eq. (9) simplifies to

$$p_\phi(\phi) \approx \frac{A_c \cos\phi}{\sqrt{2\pi\sigma^2}} e^{-(A_c \sin\phi)^2/2\sigma^2} \qquad |\phi| < \frac{\pi}{2} \quad (10)$$

which, for small values of ϕ, approximates a *gaussian* with $\bar{\phi} = 0$ and $\overline{\phi^2} = \sigma^2/A_c^2$. Equation (10) is invalid for $|\phi| > \pi/2$, but the probability of that event is small if $A_c \gg \sigma$. Figure 14.4-6 depicts the transition of $p_\phi(\phi)$ from a uniform distribution when $A_c = 0$ to a gaussian curve when A_c becomes large compared to σ. (See Fig. 14.3-1 for the corresponding transition of the envelope PDF.)

We'll assume that $A_c \gg \sigma$ so we can use Eq. (10) to obtain the error probability of coherent M-ary PSK with $M > 4$. (We already have the results for $M = 2$ and 4.) Inserting Eq. (10) with $A_c^2/\sigma^2 = 2E/\eta$ into Eq. (8) gives

$$P_e \approx 1 - \frac{1}{\sqrt{2\pi}} \int_{-\pi/M}^{\pi/M} \sqrt{\frac{2E}{\eta}} \cos\phi \, e^{-(2E/\eta)(\sin\phi)^2/2} \, d\phi$$

$$\approx 1 - \frac{2}{\sqrt{2\pi}} \int_0^L e^{-\lambda^2/2} \, d\lambda \quad (11)$$

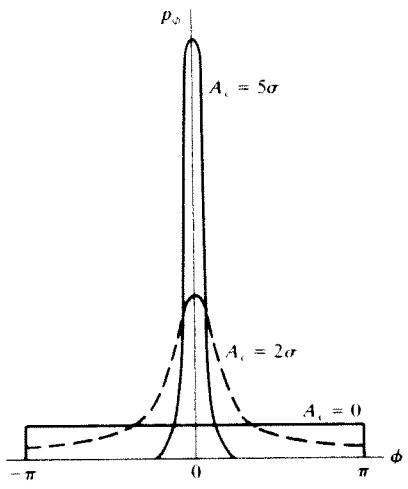


Figure 14.4-6 PDFs for the phase of a sinusoid plus bandpass noise.

where we've noted the even symmetry and made the change of variable $\lambda = \sqrt{2E/\eta} \sin\phi$ so $L = \sqrt{2E/\eta} \sin(\pi/M)$. But the integrand in Eq. (11) is a gaussian function, so $P_e \approx 1 - [1 - 2Q(L)] = 2Q(L)$. Hence,

$$P_e \approx 2Q\left(\sqrt{\frac{2E}{\eta} \sin^2\frac{\pi}{M}}\right) \quad (12)$$

which is our final result for the symbol error probability with $M > 4$. We'll discuss the equivalent bit error probability in our comparisons at the end of the chapter.

Returning to the receiver in Fig. 14.4-4, the carrier sync signals can be derived from the Mth power of $x_c(t)$ using a modified version of Fig. 14.4-2. The more sophisticated *decision-feedback* PLL system in Fig. 14.4-7 uses the estimated phase $\hat{\phi}_k$ to generate a control signal $v(t)$ that corrects any VCO phase error. The two delayors here simply account for the fact that $\hat{\phi}_k$ is obtained at the end of the kth symbol interval.

If accurate carrier synchronization proves to be impractical, then differentially coherent detection may be used instead. The noise analysis is quite complicated, but Lindsey and Simon (1973) have obtained the simple approximation

$$P_e \approx 2Q\left(\sqrt{\frac{4E}{\eta} \sin^2\frac{\pi}{2M}}\right) \quad (13)$$

which holds for $E/\eta \gg 1$ with $M \ge 4$. We see from Eqs. (12) and (13) that M-ary DPSK achieves the same error probability as coherent PSK when the energy is increased by the factor

$$\Gamma = \frac{\sin^2(\pi/M)}{2\sin^2(\pi/2M)}$$

This factor equals 2.3 dB for DQPSK ($M = 4$), as previously asserted, and it approaches 3 dB for $M \gg 1$.

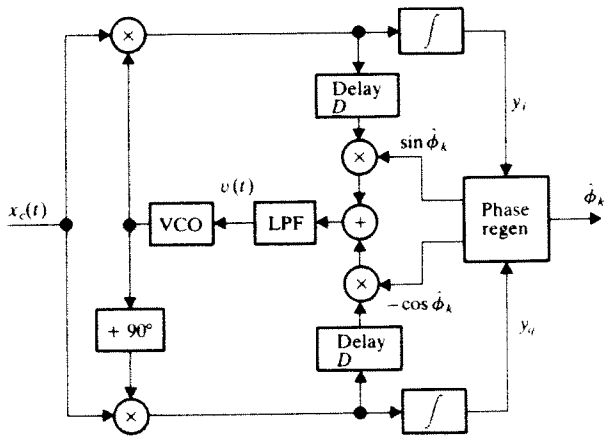


Figure 14.4-7 M-ary PSK receiver with decision-feedback system for carrier synchronization.

Exercise 14.4-2 Derive Eq. (7) by replacing one of the correlation detectors in Fig. 14.4-4 with an equivalent BPF, as in Fig. 14.2-3.

M-ary APK Systems

Combined amplitude-phase keying (APK) is an important new method for digital transmission on bandpass channels having limited bandwidth. Specifically, at the cost of more elaborate hardware, coherent APK provides lower error rates than other M-ary systems with keyed modulation operating at the same symbol rate. Here we'll study the class of APK systems defined by square signal constellations, after a preliminary treatment of suppressed-carrier M-ary ASK.

Consider M-ary ASK with synchronized modulation and suppressed carrier. Carrier-suppression is readily accomplished by applying a polar modulating signal. Thus, for the kth symbol interval, we write

$$x_c(t) = A_c I_k p_D(t - kD) \cos \omega_c t \tag{14a}$$

where

$$I_k = \pm 1, \pm 3, \ldots, \pm(M - 1) \tag{14b}$$

The transmission bandwidth is $B_T \approx r$, the same as M-ary PSK.

An optimum coherent receiver consists of just one correlation detector, since there's no quadrature component, and regeneration is based on the noisy samples

$$y_i = A_c I_k + n_i$$

The noise component is a zero-mean gaussian r.v. with variance $\sigma^2 = \eta r$, as in Eq. (7). Figure 14.4-8 shows the one-dimensional signal constellation and the corresponding $M - 1$ equispaced thresholds when $M = 4$. The symbol error probability for any even value of M is

$$P_e = 2\left(1 - \frac{1}{M}\right) Q\left(\frac{A_c}{\sqrt{\eta r}}\right) \tag{15}$$

obtained by the same analysis used for polar M-ary baseband transmission in Sect. 11.2.

Suppose that two of these ASK signals are transmitted on the same channel via quadrature-carrier multiplexing, which requires no more bandwidth than one signal. Let the information come from an M-ary source with $M = \mu^2$ so the message can be converted into two μ-ary digit streams, each having the same rate r. We could describe this process as μ-ary QAM, but the more common designation APK emphasizes the property that the source symbols are represented by

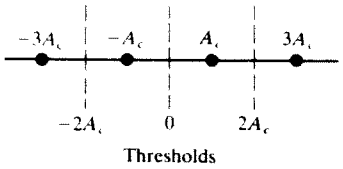


Figure 14.4-8 Decision thresholds for ASK with $M = 4$.

amplitude and phase modulation. Even so, the performance of APK fundamentally depends upon the μ-ary error rate and therefore will be superior to direct M-ary modulation with $M > \mu$.

Figure 14.4-9a diagrams the structure of our APK transmitter. The output signal for the kth symbol interval is

$$x_c(t) = s_I(t - kD) - s_q(t - kD) \tag{16a}$$

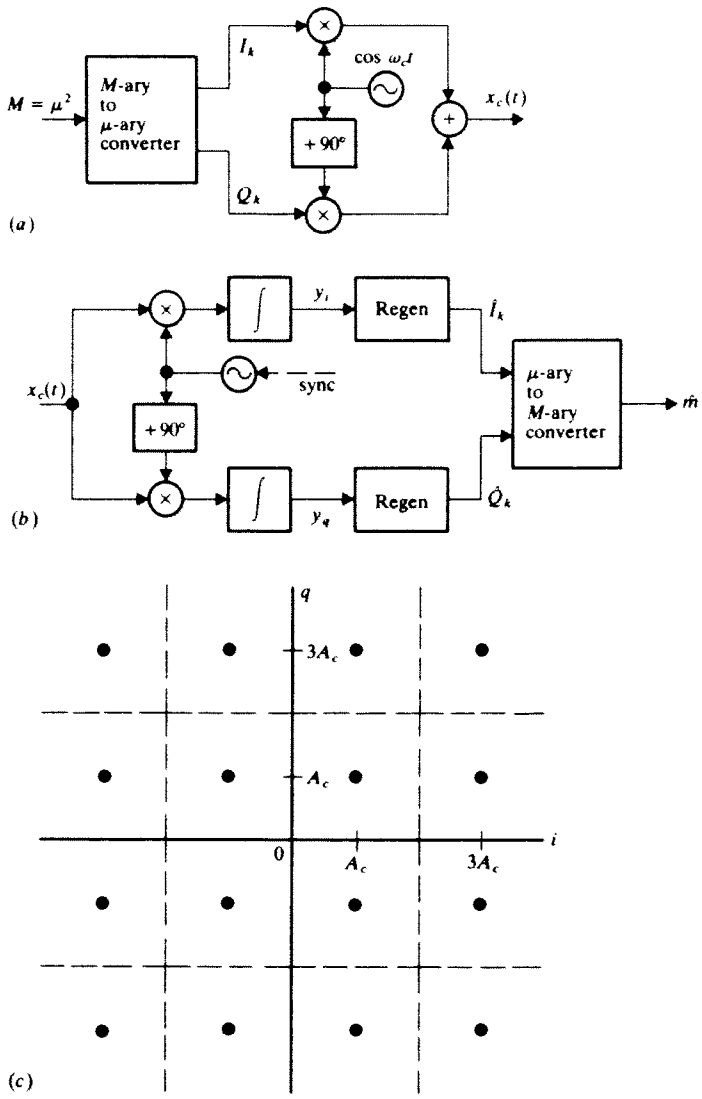


Figure 14.4-9 APK system. (a) Transmitter; (b) receiver; (c) square signal constellation and thresholds with $M = 16$.

with

$$s_i(t) = A_c I_k p_D(t) \cos \omega_c t \qquad I_k = \pm 1, \pm 3, \ldots, \pm(\mu - 1)$$
$$s_q(t) = A_c Q_k p_D(t) \sin \omega_c t \qquad Q_k = \pm 1, \pm 3, \ldots, \pm(\mu - 1) \tag{16b}$$

The average energy per M-ary symbol is

$$E = \tfrac{1}{2}A_c^2(\overline{I_k^2} + \overline{Q_k^2})D = \tfrac{1}{3}A_c^2(\mu^2 - 1)D \tag{17}$$

since $\overline{I_k^2} = \overline{Q_k^2} = (\mu^2 - 1)/3$.

Coherent APK detection is performed by the receiver in Fig. 14.4-9*b*, whose quadrature correlators produce the sample values

$$y_i = A_c I_k + n_i \qquad y_q = A_c Q_k + n_q$$

We then have a *square* signal constellation and threshold pattern, illustrated in Fig. 14.4-9*c* taking $M = 4^2 = 16$. Now let P denote the probability of error for I_k or Q_k, as given by Eq. (15) with M replaced by $\mu = \sqrt{M}$. The error probability per M-ary symbol is $P_e = 1 - (1 - P)^2$ and $P_e \approx 2P$ when $P \ll 1$. Therefore,

$$P_e \approx 4\left(1 - \frac{1}{\sqrt{M}}\right) Q\left[\sqrt{\frac{3E}{(M-1)\eta}}\right] \tag{18}$$

in which we've inserted the average symbol energy from Eq. (17).

Calculations using this result confirm the superior performance of APK. By way of example, if $M = 16$ and $E/\eta = 100$, then $P_e \approx 4 \times \frac{3}{4} \times Q(\sqrt{20}) = 1.2 \times 10^{-5}$, whereas an equivalent PSK system with $M = 16$ would have $P_e \approx 2Q(\sqrt{7.6}) = 6 \times 10^{-3}$.

Comparison of Digital Modulation Systems

A performance comparison of digital modulation systems should consider several factors, including: error probability, transmission bandwidth, spectral spillover, hardware requirements, and the differences between binary and M-ary signaling. To establish an equitable basis for comparison, we'll make the realistic assumption that the information comes from a binary source with bit rate r_b. This allows us to compare systems in terms of the modulation speed r_b/B_T and the energy-to-noise ratio γ_b needed to get a specified error probability per bit.

Our previous results for binary modulation systems apply directly to the comparison at hand, especially the error probability curves back in Fig. 14.3-4. Table 14.4-1 serves as a more abbreviated summary when γ_b is large enough to justify the applicable approximations. (Thus, in the case of noncoherent OOK, almost all the errors correspond to the carrier "off" state.) The table omits coherent OOK and FSK, which have little practical value, but it includes QAM and QPSK viewed as binary rather than quaternary modulation. This listing emphasizes the fact that doubled modulation speed goes hand-in-hand with coherent quadrature-carrier detection. Also recall that minimizing spectral spillover

Table 14.4-1 Summary of binary modulation systems

Modulation	Detection	r_b/B_T	P_{be}
OOK or FSK ($f_d = r_b/2$)	Envelope	1	$\frac{1}{2}e^{-\gamma_b/2}$
DPSK	Phase-comparison	1	$\frac{1}{2}e^{-\gamma_b}$
PRK	Coherent	1	$Q(\sqrt{2\gamma_b})$
MSK, QAM, or QPSK	Coherent quadrature	2	$Q(\sqrt{2\gamma_b})$

requires staggered keyed modulation (MSK or OQPSK) or additional pulse shaping.

Now consider M-ary transmission with symbol rate r and energy E per symbol. We'll take $M = 2^K$ and introduce the data-conversion factor

$$K = \log_2 M$$

which equals the number of bits per M-ary symbol. The equivalent bit rate and energy are $r_b = Kr$ and $E_b = E/K$, so

$$\gamma_b = E/K\eta$$

The modulation speed of M-ary PSK or APK is

$$r_b/B_T \approx K \tag{19}$$

since $B_T \approx r = r_b/K$. The error probability per bit is given by

$$P_{be} \approx P_e/K$$

providing that the data converter employs a Gray code, as discussed in conjunction with Eq. (24), Sect. 11.2. After incorporating these adjustments in our previous expressions, we get the comparative results listed in Table 14.4-2.

All of the quadrature-carrier and M-ary systems increase modulation speed at the expense of error probability or signal energy. Suppose, for example, that you want to keep the error probability fixed at $P_{be} \approx 10^{-4}$—a common standard for comparison purposes. The value of γ_b needed for different modulation systems with various modulation speeds then can be calculated from our tabulated

Table 14.4-2 Summary of M-ary modulation systems with $r_b/B_T = K = \log_2 M$

Modulation	Detection	P_{be}
DPSK ($M \geq 4$)	Phase-comparison quadrature	$\frac{2}{K} Q\left(\sqrt{4K\gamma_b \sin^2 \frac{\pi}{2M}}\right)$
PSK ($M \geq 8$)	Coherent quadrature	$\frac{2}{K} Q\left(\sqrt{2K\gamma_b \sin^2 \frac{\pi}{M}}\right)$
APK (K even)	Coherent quadrature	$\frac{4}{K}\left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3K}{M-1}\gamma_b}\right)$

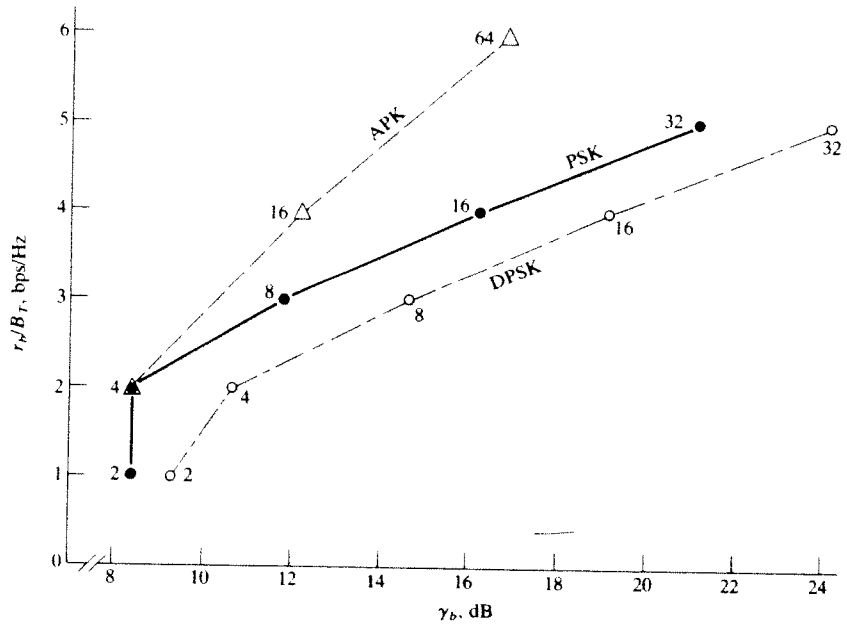


Figure 14.4-10 Performance comparison of M-ary modulation systems with $P_{be} = 10^{-4}$.

expressions. Figure 14.4-10 depicts the results as plots of r_b/B_T versus γ_b in dB, and each point is labeled with the corresponding value of M. Clearly, you would choose APK over PSK for $r_b/B_T \geq 4$ with coherent detection. M-ary DPSK eliminates the carrier-synchronization problems of coherent detection, but it requires at least 7 dB more energy than APK for $r_b/B_T \geq 4$.

As our final comparison, Table 14.4-3 combines M-ary data from Fig. 14.4-10 and calculated values for binary systems with the same error probability. The various systems are listed here in order of increasing complexity to bring out the trade-offs between modulation speed, signal energy, and hardware expense.

Table 14.4-3 Comparison of digital modulation systems with $P_{be} = 10^{-4}$

Modulation	Detection	r_b/B_T	γ_b, dB
OOK or FSK ($f_d = r_b/2$)	Envelope	1	12.3
DPSK ($M = 2$)	Phase-comparison	1	9.3
DQPSK	Phase-comparison quadrature	2	10.7
PRK	Coherent	1	8.4
MSK, QAM, or QPSK	Coherent quadrature	2	8.4
DPSK ($M = 8$)	Phase-comparison quadrature	3	14.6
PSK ($M = 8$)	Coherent quadrature	3	11.8
PSK ($M = 16$)	Coherent quadrature	4	16.2
APK ($M = 16$)	Coherent quadrature	4	12.2

You should keep two points in mind when you examine this table. First, the numerical values correspond to *ideal* systems. The modulation speed of an actual system is typically about 80% of the theoretical value, and the required energy is at least 1–2 dB higher. Second, the characteristics of specific transmission channels may impose additional considerations. In particular, rapidly changing transmission delay prohibits coherent detection, while transmission nonlinearities dictate against the envelope modulation of OOK and APK.

Other factors not covered here include the effects of interference, fading, and delay distortion. These are discussed in an excellent paper by Oetting (1979), which also contains an extensive list of references.

14.5 PROBLEMS

14.1-1 Find from Eq. (7) the average power $\overline{x_c^2}$ and the carrier-frequency power P_c of an M-ary ASK signal. Then form the ratio $P_c/\overline{x_c^2}$ and simplify for $M = 2$ and $M \gg 1$.

14.1-2 Suppose a binary ASK signal consists of RZ rectangular pulses with duration $T_b/2$, where $r_b = 1/T_b \ll f_c$. (a) Find the equivalent lowpass spectrum, and sketch and label $G_c(f)$ for $f > 0$. (b) Sketch the signal representing the sequence 010110. Then find the ratio of the carrier-frequency power P_c to the average power $\overline{x_c^2}$.

14.1-3 Consider a binary ASK signal with *raised-cosine pulse shaping* so, from Example 2.4-2

$$p(t) = \frac{1}{2}\left[1 + \cos\left(\frac{\pi t}{\tau}\right)\right]\Pi\left(\frac{t}{2\tau}\right) \qquad P(f) = \frac{\tau \operatorname{sinc} 2f\tau}{1 - (2f\tau)^2}$$

(a) Sketch the signal representing the sequence 010110 when $\tau = T_b/2$. Then find the equivalent lowpass spectrum, and sketch and label $G_c(f)$ for $f > 0$.

(b) Redo part a with $\tau = T_b$.

14.1-4 The *envelope* and *phase* variations of a QAM signal are

$$A(t) = A_c[x_i^2(t) + x_q^2(t)]^{1/2} \qquad \phi(t) = \arctan\,[x_q(t)/x_i(t)]$$

(a) By considering the time interval $kD < t < (k+1)D$, obtain expressions for $A(t)$ and $\phi(t)$ with a rectangular pulse shape $p_D(t)$.

(b) Redo part a with an arbitrary pulse shape $p(t)$ whose duration does not exceed D.

14.1-5 Let a polar M-ary VSB signal have Nyquist pulse shaping per Eq. (6), Sect. 11.3. Find the equivalent lowpass spectrum before VSB filtering. Then sketch and label $G_c(f)$ for $f > 0$ when the filter has $\beta_v \ll r$.

14.1-6 Before bandpass filtering, the i and q components of the OQPSK signal generated in Fig. 14.1-6 can be written as

$$x_i(t) = \sum_k a_{2k}\, p(t - 2kT_b) \qquad x_q(t) = \sum_k a_{2k+1} p(t - 2kT_b - T_b)$$

where $a_k = (2A_k - 1)$ is the polar sequence corresponding to the message bit sequence A_k, and $p(t) = \Pi(t/2T_b)$ for NRZ rectangular pulse shaping.

(a) Sketch $x_i(t)$ and $x_q(t)$ for the bit sequence 10011100. Use your sketch to draw the signal constellation and to confirm that the phase $\phi(t) = \arctan\,[x_q(t)/x_i(t)]$ never changes by more than $\pm\pi/2$ rad.

(b) Find the equivalent lowpass spectrum.

14.1-7‡ Let $A_k B_k C_k$ denote the Gray-code binary words for the eight-phase PSK constellation in Fig. 14.1-5*b*. Construct a table listing $A_k B_k C_k$ and the corresponding values of I_k and Q_k expressed in terms of $\alpha = \cos \pi/8$ and $\beta = \sin \pi/8$. Then write *algebraic* expressions for I_k and Q_k as functions of