times of $t = kT_s$, the samples are shifted into the register (as shown in Figure 4.10a) so that earlier samples are located to the right of later samples. Once the received signal has been sampled, the continuous time notation t is changed to kT_s or simply to k to allow use of the simple discrete notation

$$r(k) = s_i(k) + n(k) \qquad i = 1, 2 \qquad k = 0, 1, \ldots$$

where the index i identifies a particular symbol out of the M-ary set (binary in this example), and k is the sampling-time index. In Figure 4.10, the shift register, with its coefficients or weights $c_i(n)$ approximate an MF, where $n = 0, \ldots, N-1$ is the time index of the weights and register stages. In this example, $N = 4$ represents the number of stages in the register and the number of samples per symbol. Hence, the summation shown in the figure takes place over the time $n = 0$ to $n = 3$. Following the summer, a symbol decision will be made after 4 time samples have entered the registers. Note that for simplicity, the example in Figure 4.10b uses only 3-level amplitude values (0, ±1) for the $s_i(k)$ samples. In real systems, each sample (and weight) would be represented by 6–10 bits, depending on the density of the signal constellation. The set of filter weights $\{c_i(n)\}$ constitutes the filter impulse response; the weights are matched to signal samples according to the discrete form of Equation (4.26), as follows:

$$c_i(n) = s_i[(N-1) - n] = s_i(3 - n) \tag{4.27}$$

By using the discrete form of the *convolution integral* in Equation (A.44b), the output at a time corresponding to the kth sample can be expressed as

$$z_i(k) = \sum_{n=0}^{N-1} r(k-n)\, c_i(n) \qquad k = 0, 1, \ldots, \text{modulo-}N \tag{4.28}$$

where x modulo-y is defined as the remainder of dividing x by y, the index k represents time for both the received samples and the filter output, and n is a dummy time variable. In the expression $r(k-n)$ of Equation (4.28), it is helpful to think of n as the "age" of the sample (how long has it been sitting in the filter). In the expression $c_i(n)$, it is helpful to think of n as the address of the weight. We assume that because the system is synchronized, the symbol timing is known. Also, we assume the noise to have a zero mean, so that the expected value of a received sample is

$$\mathbf{E}\{r(k)\} = s_i(k) \qquad i = 1, 2$$

Thus, if $s_i(t)$ is transmitted, the expected matched filter output is

$$\mathbf{E}\{z_i(k)\} = \sum_{n=0}^{N-1} s_i(k-n)\, c_i(n) \qquad k = 0, 1, \ldots, \text{modulo-}N \tag{4.29}$$

In Figure 4.10b, where the prototype waveforms are plotted as a function of time, we see that the leftmost sample (amplitude equals + 1) of the $s_1(t)$ plot represents the earliest sample at time $k = 0$. Assuming that $s_1(t)$ was sent, and the noise is neglected for notational simplicity, we then denote the received samples of $r(k)$ as $s_1(k)$. As these samples fill the stages of the MF, at the end of each symbol time the

$k = 0$ sample is located in the rightmost stage of each register. In Equations (4.28) and (4.29), notice that the time indexes n of the reference weights are in reverse order compared with the time index $k - n$ of the samples, which is a key aspect of the convolution integral. The fact that the earliest time sample now corresponds to the rightmost weight will ensure a meaningful correlation. Even though we describe the mathematical operation of an MF to be *convolution* of a signal with the impulse response of the filter, the end result appears to be the *correlation* of a signal with a replica of that same signal. That is why it is valid to describe a correlator as an implementation of a matched filter.

In Figure 4.10b, detection will follow the MF in the usual way. For the binary decision, the $z_i(k)$ outputs are examined at each value of $k = N - 1$ corresponding to the end of a symbol. Under the condition that $s_1(t)$ had been transmitted and noise is neglected, we combine Equations (4.27) through (4.29) to express the correlator outputs at time $k = N - 1 = 3$ as

$$z_1(k{=}3) = \sum_{n=0}^{3} s_1(3 - n)\, c_1(n) = 2 \tag{4.30a}$$

and

$$z_2(k{=}3) = \sum_{n=0}^{3} s_1(3 - n)\, c_2(n) = -2 \tag{4.30b}$$

Since $z_1(k = 3)$ is greater than $z_2(k = 3)$, the detector chooses $s_1(t)$ as the transmitted symbol.

One might ask, "What is the difference between the MF in Figure 4.10b and the correlator in Figure 4.8?" In the case of the MF, a new output value is available in response to each new input sample; thus the output will be a time series such as the MF output seen in Figure 3.7b (a succession of increasing positive and negative correlations to an input sine wave). Such an MF output sequence can be equated to several correlators operating at different starting points of the input time series. Note that a correlator only computes an output once per symbol time, such as the value of the peak signal at time T in Figure 3.7b. If the timing of the MF and correlator are aligned, then their outputs at the end of a symbol time are identical. An important distinction between the MF and correlator is that since the correlator yields a single output value per symbol, it must have side information, such as the start and stop times over which the product integration should take place. If there are timing errors in the correlator, then the sampled output fed to the detector may be badly degraded. On the other hand, since the MF yields a *time series* of output values (reflecting time-shifted input samples multiplied by fixed weights), then with the use of additional circuitry, the best time for sampling the MF output can be learned.

Example 4.1 Sampled Matched Filter

Consider the waveform set

$$s_1(t) = At \qquad 0 \le t \le kT$$

and

$$s_2(t) = -At \qquad 0 \le t \le kT$$

where $k = 0, 1, 2, 3$.

Illustrate how a *sampled* matched filter as shown in Figure 4.10 can be used to detect a received signal, say $s_1(t)$, from this sawtooth waveform set in the absence of noise.

Solution

First, the waveform is sampled so that $s_1(t)$ is transformed into the set of samples $\{s_1(k)\}$. The sampled matched filter receiver will be shown with two branches, following the implementation in Figure 4.10b. The top branch is made up of shift registers and coefficients matched to the $\{s_1(k)\}$ sample points. The bottom branch is similarly matched to the $\{s_2(k)\}$ sample points. The four equally spaced sample points ($k = 0, 1, 2, 3$) for each of the $\{s_i(k)\}$ are as follows:

$$s_1(k=0) = 0 \qquad s_1(k=1) = \frac{A}{4} \qquad s_1(k=2) = \frac{A}{2} \qquad s_1(k=3) = \frac{3A}{4}$$

$$s_2(k=0) = 0 \qquad s_2(k=1) = -\frac{A}{4} \qquad s_2(k=2) = -\frac{A}{2} \qquad s_2(k=3) = -\frac{3A}{4}$$

The $c_i(n)$ coefficients represent the delayed mirror-image rotation of the signal to which the filter is matched. Therefore, $c_i(n) = s_i(N-1-n)$, where $n = 0, \ldots, N-1$, and we can write $c_i(0) = s_i(3)$, $c_i(1) = s_i(2)$, $c_i(2) = s_i(1)$, $c_i(3) = s_i(0)$.

Consider the top branch in Figure 4.10b. At the $k = 0$ clock time, the first sample $s_1(k=0) = 0$ enters the leftmost stage of each register. At the next clock time, the second sample $s_1(k=1) = A/4$ enters the leftmost stage of each register; at this same time the first sample has been shifted to the next right stage in each register, and so on. At the $k = 3$ clock time the sample $s_1(k=3) = 3A/4$ enters the leftmost stage; by this time the first sample has been shifted into the rightmost stage. The four signal samples are now located in the registers in mirror-image arrangement compared with how they would be plotted in time. Hence, the convolution operation is an appropriate expression for describing the alignment of the incoming waveform samples with the reference coefficients to maximize the correlation in the proper branch.

4.4.3 Coherent Detection of Multiple Phase-Shift Keying

Figure 4.11 illustrates the signal space for a multiple phase-shift keying (MPSK) signal set; the figure describes a four-level (4-ary) PSK or quadriphase shift keying (QPSK) example ($M = 4$). At the transmitter, binary digits are collected two at a time, and for each symbol interval, the two sequential digits instruct the modulator as to which of the four waveforms to produce. For typical coherent M-ary PSK (MPSK) systems, $s_i(t)$ can be expressed as

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left(\omega_0 t - \frac{2\pi i}{M}\right) \qquad \begin{array}{l} 0 \le t \le T \\ i = 1, \ldots, M \end{array} \tag{4.31}$$

where E is the received energy of such a waveform over each symbol duration T, and ω_0 is the carrier frequency. If an orthonormal signal space is assumed in Equations (3.10) and (3.11), we can choose a convenient set of axes, such as

$$\psi_1(t) = \sqrt{\frac{2}{T}} \cos \omega_0 t \tag{4.32a}$$

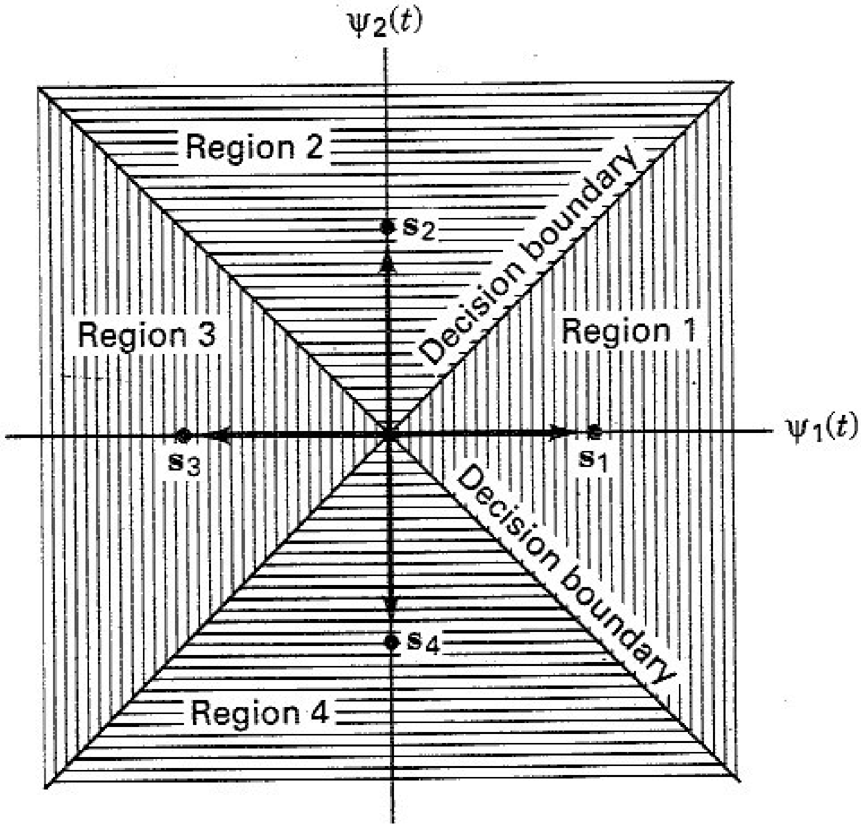


Figure 4.11 Signal space and decision regions for a QPSK system.

and

$$\psi_2(t) = \sqrt{\frac{2}{T}} \sin \omega_0 t \tag{4.32b}$$

where the amplitude $\sqrt{2/T}$ has been chosen to normalize the expected output of the detector, as was done in Section 4.4.1. Now $s_i(t)$ can be written in terms of these orthonormal coordinates, giving

$$s_i(t) = a_{i1}\psi_1(t) + a_{i2}\psi_2(t) \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix} \tag{4.33a}$$

$$= \sqrt{E} \cos\left(\frac{2\pi i}{M}\right) \psi_1(t) + \sqrt{E} \sin\left(\frac{2\pi i}{M}\right) \psi_2(t) \tag{4.33b}$$

Notice that Equation (4.33) describes a set of M multiple phase waveforms (intrinsically nonorthogonal) in terms of only two orthogonal carrier-wave components. The $M = 4$ (QPSK) case is unique among MPSK signal sets in the sense that the QPSK waveform set is represented by a combination of antipodal and orthogonal members. The decision boundaries partition the signal space into $M = 4$ regions; the construction is similar to the procedure outlined in Section 4.3.1 and Figure 4.6 for $M = 2$. The decision rule for the detector (see Figure 4.11) is to decide that $s_1(t)$ was transmitted if the received signal vector falls in region 1, that $s_2(t)$ was transmitted if the received signal vector falls in region 2, and so on. In other words, the decision rule is to choose the ith waveform if $z_i(T)$ is the largest of the correlator outputs (seen in Figure 4.7).

The form of the correlator shown in Figure 4.7a implies that there are always M product correlators used for the demodulation of MPSK signals. The figure infers that for each of the M branches, a reference signal with the appropriate phase shift is configured. In practice, the implementation of an MPSK demodulator follows Figure 4.7b, requiring only $N = 2$ product integrators regardless of the size of the signal set M. The savings in implementation is possible because any arbitrary integrable waveform set can be expressed as a linear combination of orthogonal waveforms, as shown in Section 3.1.3. Figure 4.12 illustrates such a demodulator. The received signal $r(t)$ can be expressed by combining Equations (4.32) and (4.33) as

$$r(t) = \sqrt{\frac{2E}{T}}(\cos \phi_i \cos \omega_0 t + \sin \phi_i \sin \omega_0 t) + n(t) \qquad \begin{array}{c} 0 \le t \le T \\ i = 1, \dots, M \end{array} \tag{4.34}$$

where $\phi_i = 2\pi i/M$, and $n(t)$ is a zero-mean white Gaussian noise process. Notice that in Figure 4.12, there are only two reference waveforms or basis functions, $\psi_1(t) = \sqrt{2/T}\cos \omega_0 t$ for the upper correlator and $\psi_2(t) = \sqrt{2/T}\sin \omega_0 t$ for the lower correlator. The upper correlator computes

$$X = \int_0^T r(t)\psi_1(t)\, dt \tag{4.35}$$

and the lower correlator computes

$$Y = \int_0^T r(t)\psi_2(t)\, dt \tag{4.36}$$

Figure 4.13 illustrates that the computation of the received phase angle $\hat{\phi}$ can be accomplished by computing the arctan of Y/X, where X can be thought of as the in-phase component of the received signal, Y is the quadrature component, and $\hat{\phi}$ is a

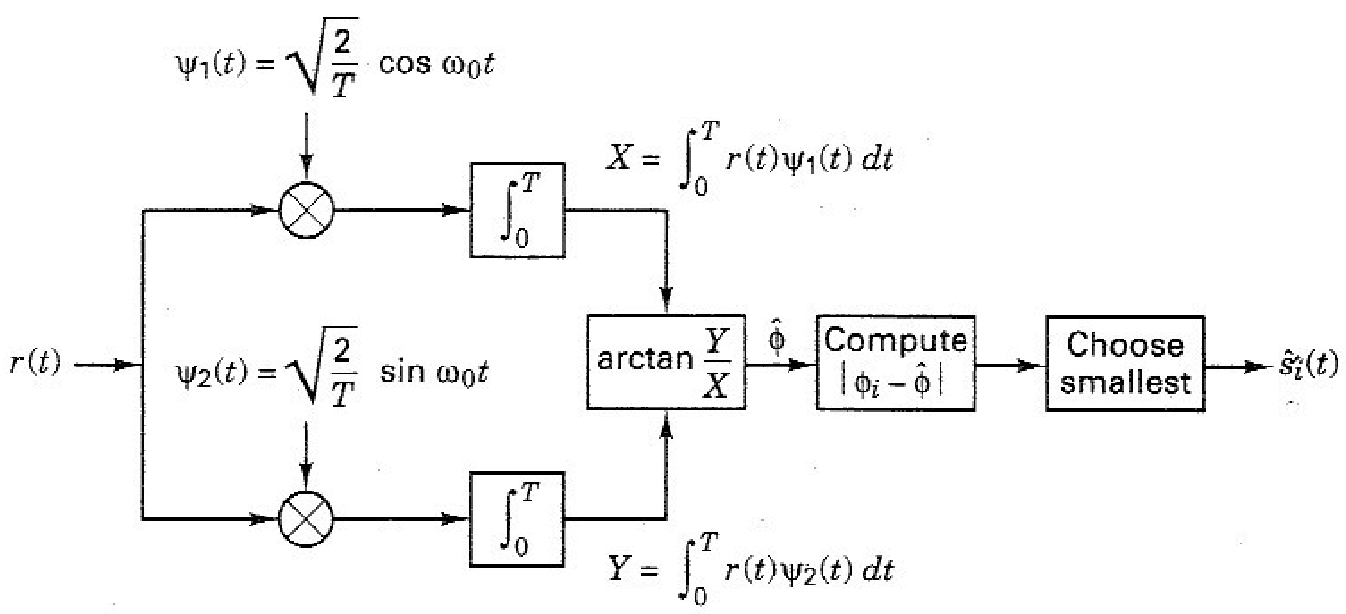


Figure 4.12 Demodulator for MPSK signals.

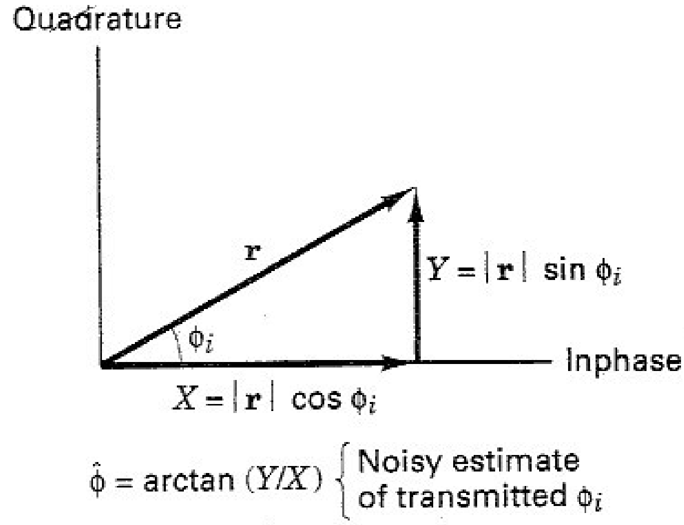


Figure 4.13 In-phase and quadrature components of the received signal vector **r**.

noisy estimate of the transmitted ϕ_i. In other words, the upper correlator of Figure 4.12 produces an output X, the magnitude of the in-phase projection of the vector $\mathbf{r}$, and the lower correlator produces an output Y, the magnitude of the quadrature projection of the vector $\mathbf{r}$, where $\mathbf{r}$ is the vector representation of $r(t)$. The X and Y outputs of the correlators feed into the block marked arctan (Y/X). The resulting value of the angle $\hat{\phi}$ is compared with each of the prototype phase angles, ϕ_i. The demodulator selects the ϕ_i that is closest to the angle $\hat{\phi}$. In other words, the demodulator computes $|\phi_i - \hat{\phi}|$ for each of the ϕ_i prototypes and chooses the ϕ_i yielding the smallest output.

4.4.4 Coherent Detection of FSK

FSK modulation is characterized by the information being contained in the frequency of the carrier. A typical set of FSK signal waveforms was described in Equation (4.8) as

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos(\omega_i t + \phi) \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix}$$

where E is the energy content of $s_i(t)$ over each symbol duration T, and $(\omega_{i+1} - \omega_i)$ is typically assumed to be an integral multiple of π/T. The phase term ϕ is an arbitrary constant and can be set equal to zero. Assuming that the basis functions $\psi_1(t)$, $\psi_2(t), \ldots, \psi_N(t)$ form an orthonormal set, the most useful form for $\{\psi_j(t)\}$ is

$$\psi_j(t) = \sqrt{\frac{2}{T}} \cos \omega_j t \qquad j = 1, \ldots, N \tag{4.37}$$

where, as before, the amplitude $\sqrt{2/T}$ normalizes the expected output of the matched filter. From Equation (3.11), we can write

$$a_{ij} = \int_0^T \sqrt{\frac{2E}{T}} \cos(\omega_i t) \sqrt{\frac{2}{T}} \cos \omega_j t \, dt \tag{4.38}$$

Therefore,

$$a_{ij} = \begin{cases} \sqrt{E} & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \tag{4.39}$$

In other words, the ith prototype signal vector is located on the ith coordinate axis at a displacement $\sqrt{E}$ from the origin of the signal space. In this scheme, for the general M-ary case and a given E, the distance between any two prototype signal vectors $\mathbf{s}_i$ and $\mathbf{s}_j$ is constant:

$$d(\mathbf{s}_i, \mathbf{s}_j) = \|\mathbf{s}_i - \mathbf{s}_j\| = \sqrt{2E} \qquad \text{for } i \neq j \tag{4.40}$$

Figure 4.14 illustrates the prototype signal vectors and the decision regions for a 3-ary ($M = 3$) coherently detected orthogonal FSK signaling scheme. A natural choice for the size M of a signaling set is any power-of-two value. However in this case, the reasons for the unorthodox selection of $M = 3$, are: We desire to examine a signaling set that is greater than binary, and the signaling space for orthogonal signaling is best visualized as having mutually perpendicular axes. It is not possible, beyond 3-axes, to convey the notion of mutual perpendicularity in some visually accurate way. As in the PSK case, the signal space is partitioned into M distinct regions, each containing one prototype signal vector; in this example, because the decision region is three-dimensional, the decision boundaries are planes instead of lines. The optimum decision rule is to decide that the transmitted signal belongs to the class whose index corresponds to the region where the received signal is

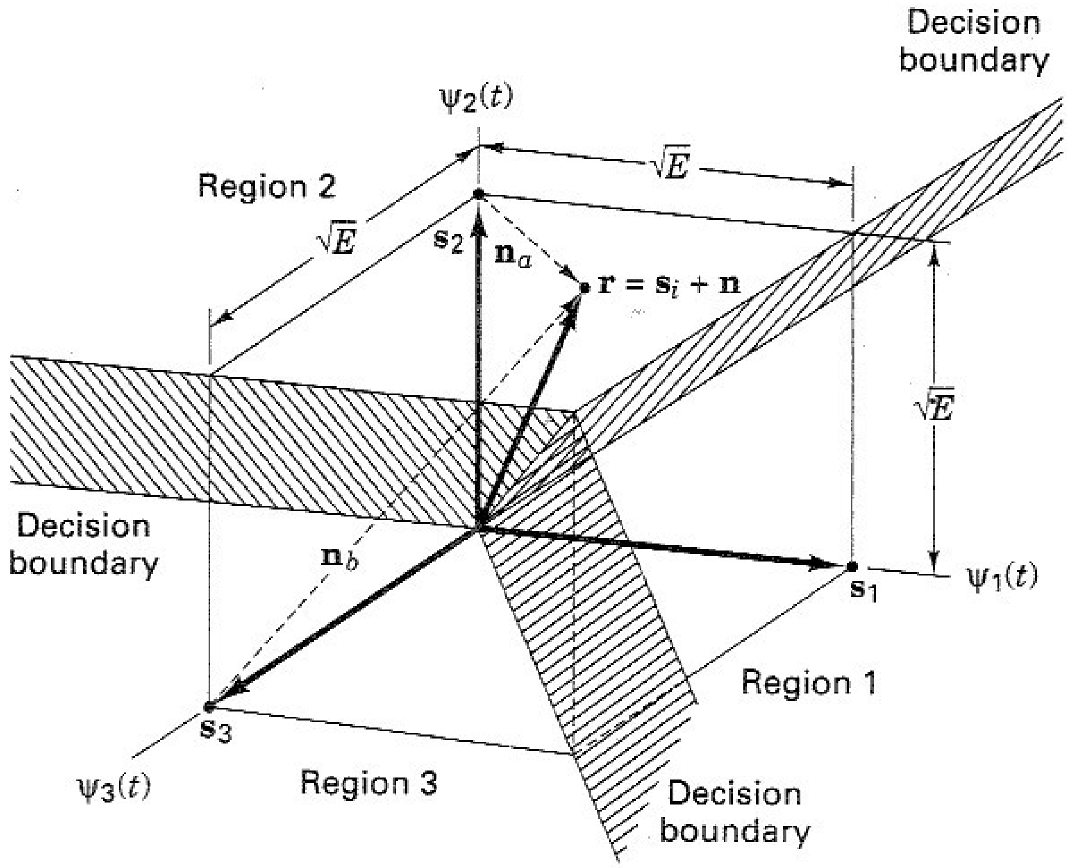


Figure 4.14 Partitioning the signal space for a 3-ary FSK signal.

found. In Figure 4.14, a received signal vector **r** is shown in region 2. Using the decision rule stated above, the detector classifies **r** as signal $\mathbf{s}_2$. Since the noise is a Gaussian random vector, there is a probability greater than zero that **r** could have been produced by some signal other than $\mathbf{s}_2$. For example, if the transmitter had sent $\mathbf{s}_2$, then **r** would be the sum of signal plus noise, $\mathbf{s}_2 + \mathbf{n}_a$, and the decision to choose $\mathbf{s}_2$ is correct; however, if the transmitter had actually sent $\mathbf{s}_3$, then **r** would be the sum of signal plus noise, $\mathbf{s}_3 + \mathbf{n}_b$, and the decision to select $\mathbf{s}_2$ is an error. The error performance of coherently detected FSK systems is treated in Section 4.7.3.

Example 4.2 Received Phase as a Function of Propagation Delay

(a) In Figure 4.8, the schematics fail to indicate where the correlator reference signals come from. One might think that they are known for all time and stored in memory until needed. Under some controlled circumstances, it is conceivable that the receiver might predict some expected value of the arriving signal's amplitude or its frequency within tolerable limits. But the one parameter that cannot be estimated without special help is the received signal phase. The most popular way to achieve phase estimatio is through the use of circuitry called a *phase-locked loop* (PLL). This carrier-recovery method locks on to the arriving carrier wave (or recreates it) and estimates its phase. To show how futile it would be to predict the phase without a PLL, consider the mobile radio link shown in Figure 4.15. In the figure, a mobile user is positioned at point *A*, at a distance *d* from the base station, such that the propagation delay is T_d. Using complex notation, the transmitted waveform emanating from the transmitter can be described as $s(t) = \exp(j2\pi f_0 t)$. Let us take f_0 equal to 1 GHz. Neglecting noise, the waveform received at the base station can be denoted as $r(t) = \exp[j2\pi f_0(t + T_d)]$. If the mobile user moves in-line away from the base station to point *B*, or in-line toward the base station to point *C*, what is the minimum distance of movement that will cause a 2π phase rotation of the received waveform?

(b) Do we really care about a 2π phase rotation? Of course not, because the received phasor would then be located at exactly the point at which it was initially postulated with the user located at point *A*. But let us ask, What is the minimum distance that will cause a $\pi/2$ phase rotation (say a lag of $\pi/2$)? Had the receiver predicted that the received phasor would correspond to $r(t)$, as denoted in part (a),

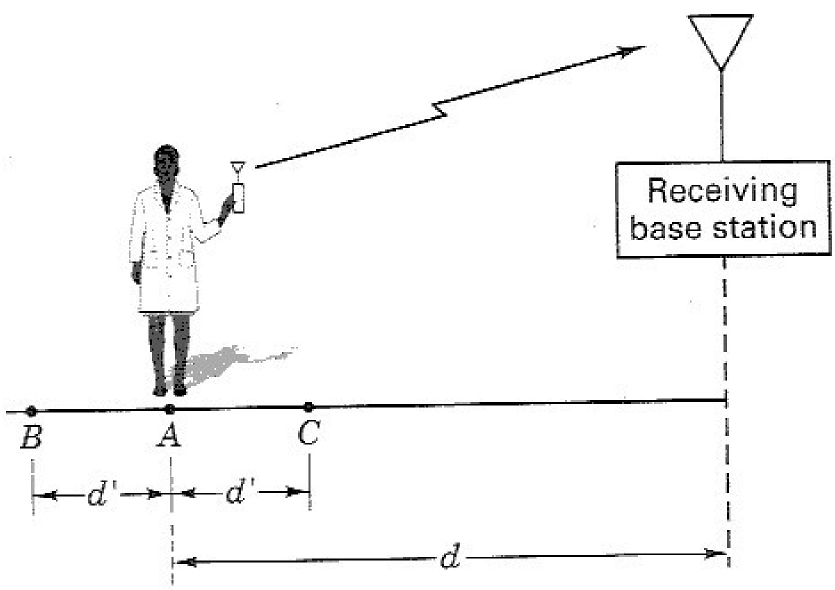


Figure 4.15 Mobile radio link.

but instead the lag caused the received phasor to appear as exp $[j2\pi f_0(t + T_d) - \pi/2]$, then the detecting correlator would yield a zero output. This is so because

$$\int_0^T \cos \omega_0 t \cos \left(\omega_0 t - \frac{\pi}{2}\right) dt = \int_0^T \cos \omega_0 t \sin \omega_0 t \, dt = 0$$

Find the user's minimum distance movement that will cause a $\pi/2$ phase rotation.

Solution

(a) Initially, let $t = 0$, so that when the mobile user is located at point A, the received phasor at the base station can be expressed as $r(t) = \exp(j2\pi f_0 T_d)$. Then, after the user's movement to point B, the received (further delayed) phasor $r_d(t = T_d + T_d')$, can be written as $r_d(t) = \exp[j2\pi f_0(T_d + T_d')]$. The minimum delay time $T_{d'}$ corresponding to a 2π (one wavelength) phasor rotation is $T_d' = 1/f_0 = 10^{-9}$ second. Therefore, the minimum distance for such a rotation (assuming ideal electromagnetic propagation at the speed of light) is

$$d' = \frac{c}{f_0} = 10^8 \text{ m/s} \times 10^{-9} \text{ s} = 0.3 \text{ m}$$

(b) Thus, for a $\pi/2$ phasor rotation, the minimum distance is

$$d'' = \frac{d'}{4} = \frac{0.3 \text{ m}}{4} = 7.5 \text{ cm}$$

It should be clear that even if a transmitter and receiver are located on fixed towers, a small amount of wind movement can bring about complete uncertainty regarding phase. If we scale our example from a frequency of 1 GHz to that of 10 GHz, the minimum distance scales from 7.5 cm to 0.75 cm. Very often we might want to avoid building receivers with PLLs for carrier recovery. The results of this example might then motivate us to ask, How will the error performance suffer if phase information is not used in the detection process? In other words, how will the system fare if the detection is performed noncoherently? We address this question in the sections that follow.

4.5 NONCOHERENT DETECTION

4.5.1 Detection of Differential PSK

The name *differential* PSK (DPSK) sometimes needs clarification because two separate aspects of the modulation/demodulation format are being referred to: the encoding procedure and the detection procedure. The term *differential encoding* refers to the procedure of encoding the data differentially; that is, the presence of a binary one or zero is manifested by the symbol's similarity or difference when compared with the preceding symbol. The term *differentially coherent detection* of differentially encoded PSK, the usual meaning of DPSK, refers to a detection scheme often classified as noncoherent because it does not require a reference in phase with the received carrier. Occasionally, differentially encoded PSK is *coherently* detected. This will be discussed in Section 4.7.2.

With noncoherent systems, no attempt is made to determine the actual value of the phase of the incoming signal. Therefore, if the transmitted waveform is

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left[\omega_0 t + \theta_i(t)\right] \qquad \begin{array}{l} 0 \le t \le T \\ i = 1, \ldots, M \end{array}$$

the received signal can be characterized by

$$r(t) = \sqrt{\frac{2E}{T}} \cos\left[\omega_0 t + \theta_i(t) + \alpha\right] + n(t) \qquad \begin{array}{l} 0 \le t \le T \\ i = 1, \ldots, M \end{array} \tag{4.41}$$

where α is an arbitrary constant and is typically assumed to be a random variable uniformly distributed between zero and 2π, and $n(t)$ is an AWGN process.

For coherent detection, matched filters (or their equivalents) are used; for noncoherent detection, this is not possible because the matched filter output is a function of the unknown angle α. However, if we assume that α varies slowly relative to two period times ($2T$), the phase difference between two successive waveforms $\theta_j(T_1)$ and $\theta_k(T_2)$ is independent of α; that is,

$$\left[\theta_k(T_2) + \alpha\right] - \left[\theta_j(T_1) + \alpha\right] = \theta_k(T_2) - \theta_j(T_1) = \phi_i(T_2) \tag{4.42}$$

The basis for *differentially coherent detection* of differentially encoded PSK (DPSK) is as follows. The carrier phase of the previous signaling interval can be used as a phase reference for demodulation. Its use requires *differential encoding* of the message sequence at the transmitter since the information is carried by the difference in phase between two successive waveforms. To send the ith message ($i = 1, 2, \ldots, M$), the present signal waveform must have its phase advanced by $\phi_i = 2\pi i/M$ radians over the previous waveform. The detector, in general, calculates the coordinates of the incoming signal by correlating it with locally generated waveforms, such as $\sqrt{2/T}\cos\omega_0 t$ and $\sqrt{2/T}\sin\omega_0 t$. the detector then measures the angle between the currently received signal vector and the previously received signal vector, as illustrated in Figure 4.16.

In general, DPSK signaling performs less efficiently than PSK, because the errors in DPSK tend to propagate (to adjacent symbol times) due to the correlation between signaling waveforms. One way of viewing the difference between PSK and

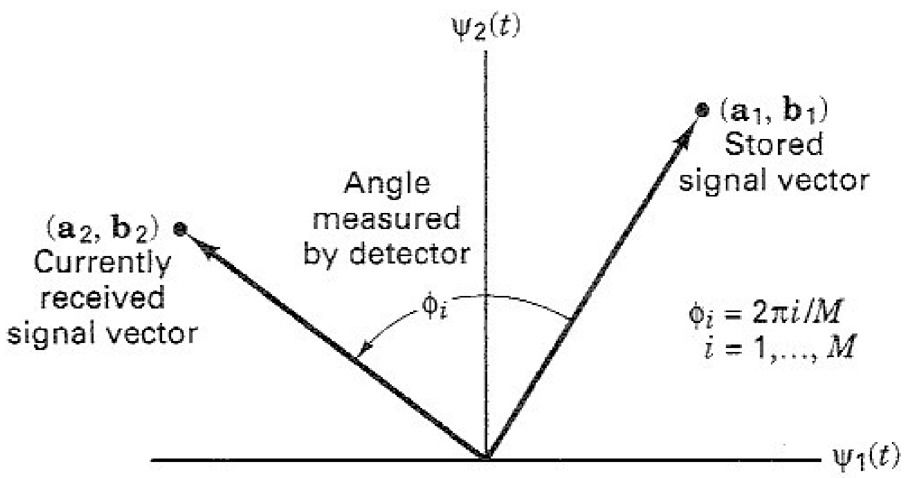


Figure 4.16 Signal space for DPSK.

DPSK is that the former compares the received signal with a clean reference; in the latter, however, two noisy signals are compared with each other. We might say that there is twice as much noise associated with DPSK signaling compared to PSK signaling. Consequently, as a first guess, we might estimate that the error probability for DPSK is approximately two times (3 dB) worse than PSK; this degradation decreases rapidly with increasing signal-to-noise ratio. The trade-off for this performance loss is reduced system complexity. The error performance for the detection of DPSK is treated in Section 4.7.5.

4.5.2 Binary Differential PSK Example

The essence of differentially coherent detection in DPSK is that the identity of the data is inferred from the changes in phase from symbol to symbol. Therefore, because the data are detected by differentially examining the waveform, the transmitted waveform must first be encoded in a differential fashion. Figure 4.17a illustrates a differential encoding of a binary message data stream $m(k)$, where k is the sample time index. The differential encoding starts (third row in the figure) with the first bit of the code-bit sequence $c(k = 0)$, chosen arbitrarily (here taken to be a one). Then the sequence of encoded bits $c(k)$ can, in general, be encoded in one of two ways:

$$c(k) = c(k-1) \oplus m(k) \tag{4.43}$$

or

$$c(k) = \overline{c(k-1) \oplus m(k)} \tag{4.44}$$

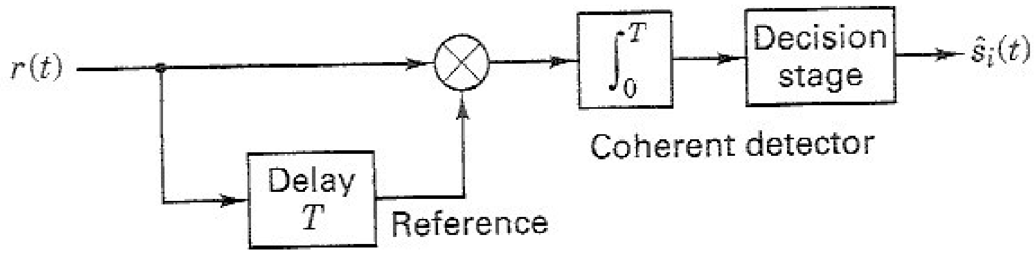
where the symbol $\oplus$ represents modulo-2 addition (defined in Section 2.9.3) and the overbar denotes complement. In Figure 4.17a the differentially encoded message was obtained by using Equation (4.44). In other words, the present code bit $c(k)$ is a one if the message bit $m(k)$ and the prior coded bit $c(k-1)$ are the same, otherwise, $c(k)$ is a zero. The fourth row translates the coded bit sequence $c(k)$ into the phase shift sequence $\theta(k)$, where a one is characterized by a 180° phase shift, and a zero is characterized by a 0° phase shift.

Figure 4.17b illustrates the binary DPSK detection scheme in block diagram form. Notice that the basic product integrator of Figure 4.7 is the essence of the demodulator; as with coherent PSK, we are still attempting to correlate a received signal with a reference. The interesting difference here is that the reference signal is simply a delayed version of the received signal. In other words, during each symbol time, we are matching a received symbol with the prior symbol and looking for a correlation or an anticorrelation (180° out of phase).

Consider the received signal with phase shift sequence $\theta(k)$ entering the correlator of Figure 4.17b, in the absence of noise. The phase $\theta(k = 1)$ is matched with $\theta(k = 0)$; they have the same value, π; hence the first bit of the detected output is $\hat{m}(k = 1) = 1$. Then $\theta(k = 2)$ is matched with $\theta(k = 1)$; again they have the same value, and $\hat{m}(k = 2) = 1$. Then $\theta(k = 3)$ is matched with $\theta(k = 2)$; they are different, so that $\hat{m}(k = 3) = 0$, and so on.

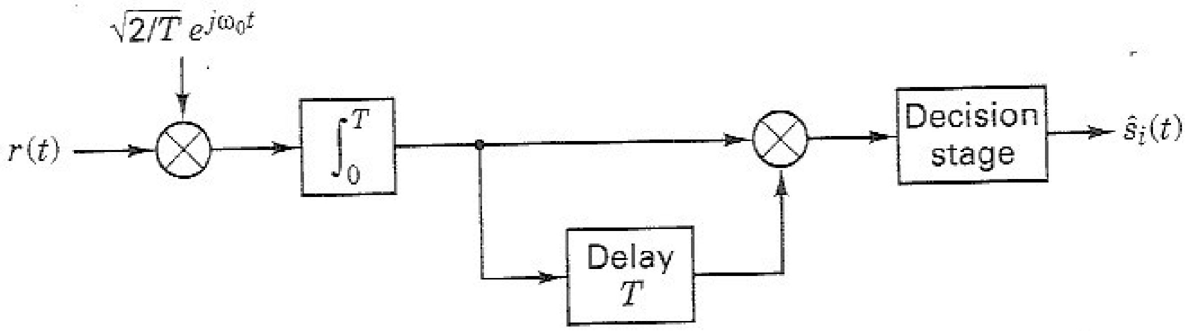
Sample index, k	0	1	2	3	4	5	6	7	8	9	10
Information message, $m(k)$		1	1	0	1	0	1	1	0	0	1
Differentially encoded message (first bit arbitrary), $c(k)$	1	1	1	0	0	1	1	1	0	1	1
Corresponding phase shift, $\theta(k)$	π	π	π	0	0	π	π	π	0	π	π

(a)



Detected message, $\hat{m}(k)$ 1 1 0 1 0 1 1 0 0 1

(b)



(c)

Figure 4.17 Differential PSK (DPSK). (a) Differential encoding. (b) Differentially coherent detection. (c) Optimum differentially coherent detection.

It must be pointed out that the detector in Figure 4.17b is suboptimum [3] in the sense of error performance. The optimum differential detector for DPSK requires a reference carrier in frequency but not necessarily in phase with the received carrier. Hence the optimum differential detector is shown in Figure 4.17c [4]. Its performance is treated in Section 4.7.5. In Figure 4.17c, the reference signal is shown in complex form ($\sqrt{2/T}\ e^{j\omega_0 t}$) to indicate that a quadrature implementation using both I and Q references are required (see Section 4.6.1).

4.5.3 Noncoherent Detection of FSK

A detector for the *noncoherent* detection of FSK waveforms described by Equation (4.8) can be implemented with correlators similar to those shown in Figure 4.7. However, the hardware must be configured as an *energy detector,* without exploiting phase measurements. For this reason, the noncoherent detector typically requires twice as many channel branches as the coherent detector. Figure 4.18 illustrates the in-phase (I) and quadrature (Q) channels used to detect a binary FSK (BFSK) signal set noncoherently. Notice that the upper two branches are configured to detect the signal with frequency ω_1; the reference signals are $\sqrt{2/T}\cos\omega_1 t$ for the I branch and $\sqrt{2/T}\sin\omega_1 t$ for the Q branch. Similarly, the lower two branches are configured to detect the signal with frequency ω_2; the reference signals are $\sqrt{2/T}\cos\omega_2 t$ for the I branch and $\sqrt{2/T}\sin\omega_2 t$ for the Q branch. Imagine that the received signal $r(t)$, by chance alone, is exactly of the form $\cos\omega_1 t + n(t)$; that is, the phase is exactly zero, and thus the signal component of the received signal exactly matches the top-branch reference signal with regard to frequency and phase. In that event, the product integrator of the top branch should yield the maximum output. The second branch should yield a near-zero output (integrated zero-mean noise), since its reference signal $\sqrt{2/T}\sin\omega_1 t$ is orthogonal to the sig-

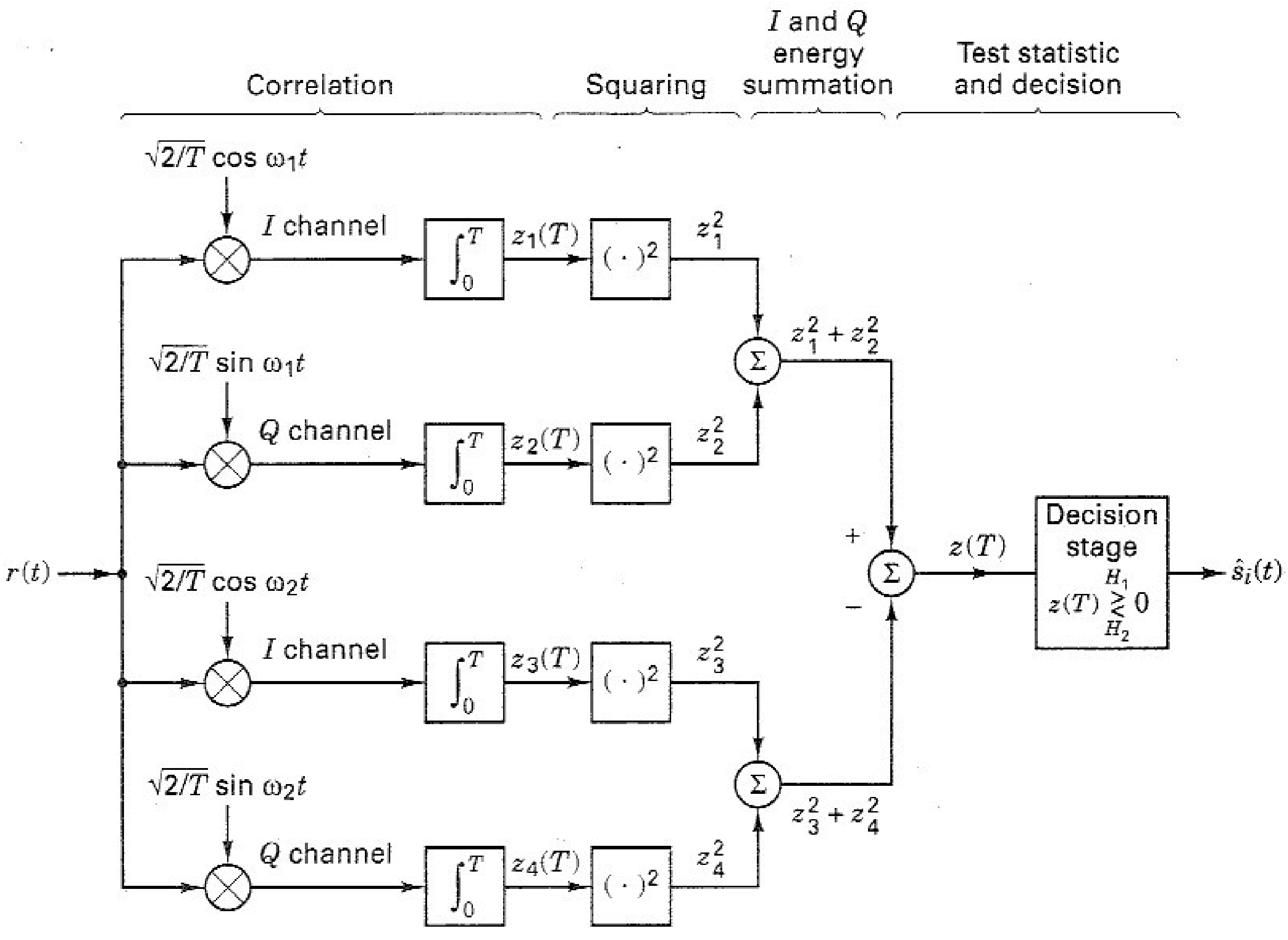


Figure 4.18 Quadrature receiver.

nal component of $r(t)$. For orthogonal signaling (see Section 4.5.4), the third and fourth branches should also yield near-zero outputs, since their ω_2 reference signals are also orthogonal to the signal component of $r(t)$.

Now imagine a different scenario: Suppose that by chance alone, the received signal $r(t)$ is of the form $\sin \omega_1 t + n(t)$. In that event, the second branch in Figure 4.18 should yield the maximum output, while the others should yield near-zero outputs. In actual practice, the most likely scenario is that $r(t)$ is of the form $\cos(\omega_1 t + \phi) + n(t)$; that is, the incoming signal will *partially* correlate with the $\cos \omega_1 t$ reference and *partially* correlate with the $\sin \omega_1 t$ reference. Hence a noncoherent quadrature receiver for orthogonal signals requires an I and Q branch for each candidate signal in the signaling set. In Figure 4.18, the blocks following the product integrators perform a squaring operation to prevent the appearance of any negative values. Then for each signal class in the set (two in this binary example), z_1^2 from the I channel and z_2^2 from the Q channel are added. The final stage forms the test statistic $z(T)$ and chooses the signal with frequency ω_1 or the signal with frequency ω_2, depending upon which pair of energy detectors yields the maximum output.

Another possible implementation for noncoherent FSK detection uses bandpass filters—centered at $f_i = \omega_i/2\pi$ with bandwidth $W_f = 1T$—followed by *envelope detectors,* as shown in Figure 4.19. An envelope detector consists of a rectifier and a low-pass filter. The detectors are matched to the *signal envelopes* and not to the signals themselves. The phase of the carrier is of no importance in defining the envelope; hence no phase information is used. In the case of binary FSK, the decision as to whether a one or a zero was transmitted is made on the basis of which of two envelope detectors has the largest amplitude at the moment of measurement. Similarly, for a multiple frequency shift-keying (MFSK) system, the decision as to which of M signals was transmitted is made on the basis of which of the M envelope detectors has the maximum output.

Even though the envelope detector block diagram of Figure 4.19 looks functionally more simple than the quadrature receiver shown in Figure 4.18, the use of

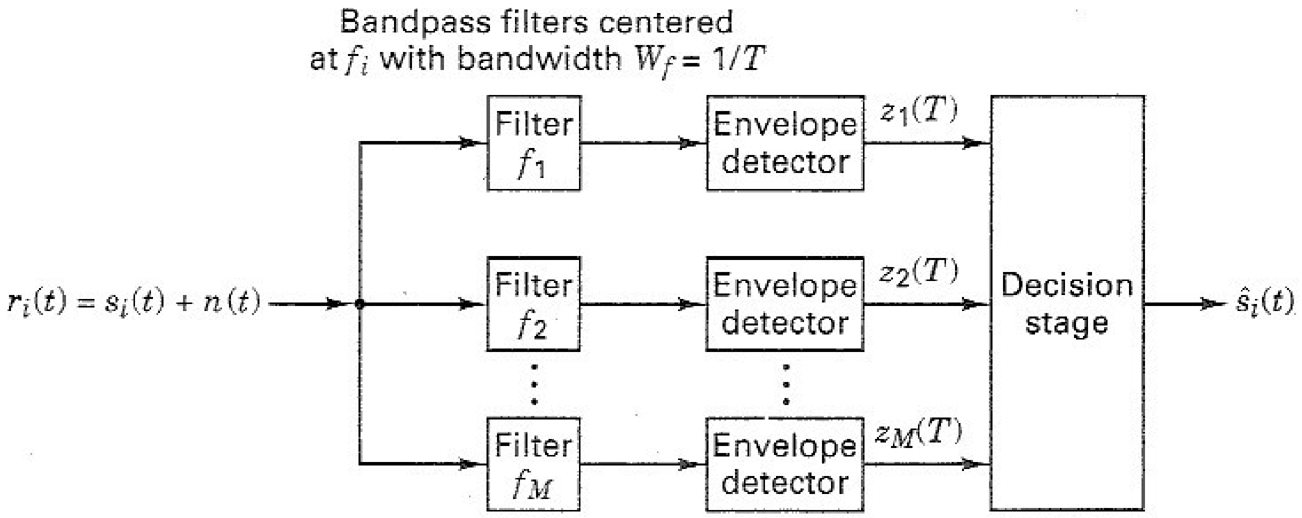


Figure 4.19 Noncoherent detection of FSK using envelope detectors.

(analog) filters usually results in the envelope detector design having greater weight and cost than the quadrature receiver. Quadrature receivers can be implemented digitally; thus, with the advent of large-scale integrated (LSI) circuits, they are often the preferred choice for noncoherent detectors. The detector shown in Figure 4.19 can also be implemented digitally by performing discrete Fourier transformations instead of using analog filters, but such a design is usually more complex than a digital implementation of the quadrature receiver.

4.5.4 Required Tone Spacing for Noncoherent Orthogonal FSK Signaling

Frequency shift keying (FSK) is usually implemented as orthogonal signaling, but not all FSK signaling is orthogonal. How can we tell if the tones in a signaling set form an orthogonal set? For example, suppose we have two tones $f_1 = 10{,}000$ Hz and $f_2 = 11{,}000$ Hz. Are they orthogonal to each other? In other words, do they fulfill the criterion (as set forth in Equation 3.69) that they are uncorrelated over a symbol time T? We do not have enough information to answer that question yet. Tones f_1 and f_2 manifest orthogonality if, for a transmitted tone at f_1, the sampled envelope of the receiver output filter tuned to f_2 is zero (i.e., no crosstalk). A property that insures such orthogonality between tones in an FSK signaling set states that any pair of tones in the set must have a frequency separation that is a multiple of $1/T$ hertz. (This will be proven below, in Example 4.3.) A tone with frequency f_i that is switched on for a symbol duration of T seconds and then switched off, such as the FSK tone described in Equation (4.8), can be analytically described by

$$s_i(t) = (\cos 2\pi f_i t) \operatorname{rect}(t/T)$$

$$\text{where } \operatorname{rect}(t/T) = \begin{cases} 1 & \text{for } -T/2 \le t \le T/2 \\ 0 & \text{for } |t| > T/2 \end{cases}$$

From Table A.1, the Fourier transform of $s_i(t)$ is

$$\mathscr{F}\{s_i(t)\} = T \operatorname{sinc}(f - f_i)T$$

where the sinc function is as defined in Equation (1.39). The spectra of two such adjacent tones—tone 1 with frequency f_1 and tone 2 with frequency f_2—are plotted in Figure 4.20.

4.5.4.1 Minimum Tone Spacing and Bandwidth

In order for a noncoherently detected tone to manifest a maximum output signal at its associated receiver filter and a zero-output signal at any neighboring filter (as implemented in Figure 4.19), the peak of the spectrum of tone 1 must coincide with one of the zero crossings of the spectrum of tone 2, and similarly, the peak of the spectrum of tone 2 must coincide with one of the zero crossings of the spectrum of tone 1. The frequency difference between the center of the spectral main lobe and the first zero crossing represents the *minimum required spacing.* With noncoherent detection, this corresponds to a minimum tone separation of $1/T$ hertz, as can be seen in Figure 4.20. Even though FSK signaling entails the trans-

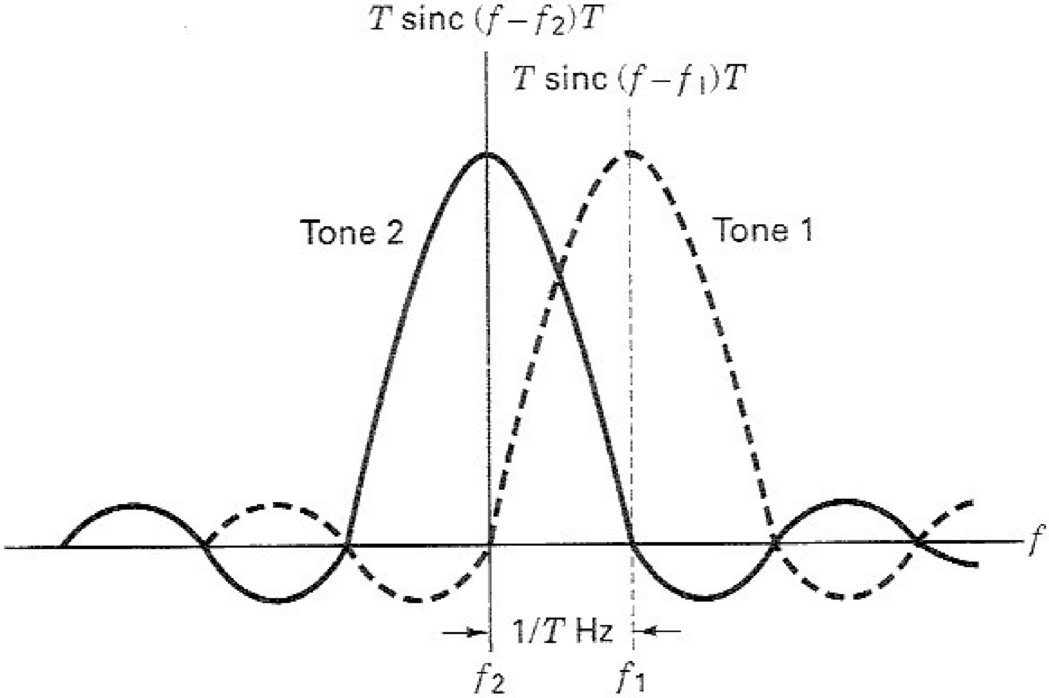


Figure 4.20 Minimum tone spacing for noncoherently detected orthogonal FSK signaling.

mission of just one single-sideband tone during each symbol time, when we speak of the signal bandwidth, we mean the amount of spectrum that needs to be available for the entire range of all the tones in the M-ary set. Thus for FSK, the bandwidth requirements are related to the spectral spacing between the tones. From a group of adjacent tones, the spectrum associated with just one tone can be considered to make up one-half of the tone spacing on each side of the tone. Thus, for the binary FSK case shown in Figure 4.20, the signaling bandwidth is equal to the spectrum that separates the tone centers plus one-half the tone spacing on each side of the set. This amounts to two times the tone spacing. Extrapolating this approach to the M-ary case, the bandwidth of noncoherently detected orthogonal MFSK is equal to M/T.

Thus far, we have only addressed noncoherently detected orthogonal FSK. Is the minimum tone-spacing criterion (and hence the bandwidth) different if the signaling is detected coherently? Indeed, it is. As we will see below, in Example 4.3, the minimum tone spacing is reduced to $1/2T$ when coherent detection is used.

4.5.4.2 Dual Relationships

The engineering concept of duality can be defined as follows: Two processes (functions, elements, or systems) are *dual* to each other if their mathematical relationships are identical, even though they are described with different variables (e.g., time and frequency). Consider the case of FSK signaling, where rectangular-shaped keying corresponds to sinc (fT) shaped tones, as seen in Figure 4.20. A given tone duration gives rise to the minimum frequency spacing between tones that would be needed to achieve orthogonality. This frequency-domain relationship has as its dual counterpart a time-domain relationship—the case of pusle signaling, where a rectangular-shaped bandwidth corresponds to sinc (t/T) shaped pulses, as seen in Figure 3.16b. A given bandwidth gives rise to the minimum time spacing between pulses that would be needed to achieve zero ISI.

Example 4.3 Minimum Tone Spacing for Orthogonal FSK

Consider two waveforms $\cos(2\pi f_1 t + \phi)$ and $\cos 2\pi f_2 t$ to be used for *noncoherent* FSK-signaling, where $f_1 > f_2$. The symbol rate is equal to $1/T$ symbols/s, where T is the symbol duration and ϕ is a constant arbitrary angle from 0 to 2π.

(a) Prove that the minimum tone spacing for *noncoherently detected* orthogonal FSK-signaling is $1/T$.

(b) What is the minimum tone spacing for *coherently detected* orthogonal FSK signaling?

Solution

(a) For the two waveforms to be orthogonal, they must fulfill the orthogonality constraint of Equation (3.69):

$$\int_0^T \cos(2\pi f_1 t + \phi)\cos 2\pi f_2 t \, dt = 0 \tag{4.45}$$

Using the basic trigonometric identities shown in Equations (D.6) and (D.1) to (D.3), we can write Equation (4.45) as

$$\cos\phi \int_0^T \cos 2\pi f_1 t \cos 2\pi f_2 t \, dt \tag{4.46}$$

$$- \sin\phi \int_0^T \sin 2\pi f_1 t \cos 2\pi f_2 t \, dt = 0$$

so that

$$\cos\phi \int_0^T [\cos 2\pi(f_1 + f_2)t + \cos 2\pi(f_1 - f_2)t] \, dt$$

$$- \sin\phi \int_0^T [\sin 2\pi(f_1 + f_2)t + \sin 2\pi(f_1 - f_2)t] \, dt = 0 \tag{4.47}$$

which, in turn, yields

$$\cos\phi \left[\frac{\sin 2\pi(f_1 + f_2)t}{2\pi(f_1 + f_2)} + \frac{\sin 2\pi(f_1 - f_2)t}{2\pi(f_1 - f_2)}\right]_0^T$$

$$+ \sin\phi \left[\frac{\cos 2\pi(f_1 + f_2)t}{2\pi(f_1 + f_2)} + \frac{\cos 2\pi(f_1 - f_2)t}{2\pi(f_1 - f_2)}\right]_0^T = 0 \tag{4.48}$$

or

$$\cos\phi \left[\frac{\sin 2\pi(f_1 + f_2)T}{2\pi(f_1 + f_2)} + \frac{\sin 2\pi(f_1 - f_2)T}{2\pi(f_1 - f_2)}\right]$$

$$+ \sin\phi \left[\frac{\cos 2\pi(f_1 + f_2)T - 1}{2\pi(f_1 + f_2)} + \frac{\cos 2\pi(f_1 - f_2)T - 1}{2\pi(f_1 - f_2)}\right] = 0 \tag{4.49}$$

We can assume that $f_1 + f_2 \gg 1$ and can thus make the following approximation:

$$\frac{\sin 2\pi(f_1 + f_2)T}{2\pi(f_1 + f_2)} \approx \frac{\cos 2\pi(f_1 + f_2)T}{2\pi(f_1 + f_2)} \approx 0 \tag{4.50}$$

Then, combining Equations (4.49) and (4.50), we can write

$$\cos \phi \sin 2\pi(f_1 - f_2)T + \sin \phi \left[\cos 2\pi(f_1 - f_2)T - 1\right] \approx 0 \tag{4.51}$$

Note that for arbitrary ϕ, the terms in Equation (4.51) can sum to zero only when $\sin 2\pi(f_1 - f_2)T = 0$, and simultaneously $\cos 2\pi(f_1 - f_2)T = 1$.

Since

$$\sin x = 0 \qquad \text{for } x = n\pi$$

and

$$\cos x = 1 \qquad \text{for } x = 2k\pi$$

where n and k are integers, then both $\sin x = 0$ and $\cos x = 1$ occur simultaneously when $n = 2k$. From Equation (4.51), for arbitrary ϕ, we can therefore write

$$2\pi(f_1 - f_2)T = 2k\pi$$

or (4.52)

$$f_1 - f_2 = \frac{k}{T}$$

Thus the minimum tone spacing for *noncoherent* FSK signaling occurs for $k = 1$, in which case we write

$$f_1 - f_2 = \frac{1}{T} \tag{4.53}$$

Recall the question posed earlier. Having two tones f_1 = 10,000 hertz and f_2 = 11,000 hertz, we asked, Are they orthogonal? Now we have sufficient information to answer that question. The answer depends on the speed of the FSK signaling. If the tones are being keyed (switched) at the rate of 1,000 symbols/s and noncoherently detected, then they are orthogonal. If they are being keyed faster—say, at the rate of 10,000 symbols/s—they are not orthogonal.

(b) For the noncoherent detection in part (a) of this example, the tone spacings that rendered the signals orthogonal was found by satisfying Equation (4.45) for any arbitrary phase. In this case, however, for coherent detection, the tone spacings needed for orthogonality are found by setting $\phi = 0$. This is because we know the phase of the received signal (from our phase-locked loop estimate). This received signal will be correlated with each reference signal using the same phase estimate for the reference signals. We can now rewrite Equation (4.51) with $\phi = 0$, which gives

$$\sin 2\pi(f_1 - f_2)T = 0 \tag{4.54}$$

or

$$f_1 - f_2 = \frac{n}{2T} \tag{4.55}$$