CHAPTER 8

NOISE IN MODULATION SYSTEMS

In Chapters 6 and 7 the subjects of probability and random processes were studied. These concepts led to a representation for bandlimited noise, which will now be used for the analysis of basic analog communication systems and for introductory considerations of digital communication systems operating in the presence of noise. The remaining chapters of this book will focus on digital communication systems in more detail. This chapter is essentially a large number of example problems, most of which focus on different systems and modulation techniques.

Noise is present in varying degrees in all electrical systems. This noise is often low level and can often be neglected in those portions of a system where the signal level is high. However, in many communications applications the receiver input signal level is very small, and the effects of noise significantly degrade system performance. Noise can take several different forms, depending upon the source, but the most common form is due to the random motion of charge carriers. As discussed in more detail in Appendix A, whenever the temperature of a conductor is above 0 K, the random motion of charge carriers results in *thermal noise*. The variance of thermal noise, generated by a resistive element, such as a cable, and measured in a bandwidth B, is given by

$$\sigma_n^2 = 4kTRB \tag{8.1}$$

where k is Boltzman's constant (1.38×10^{-23} J/K), T is the temperature of the element in degrees kelvin, and R is the resistance in ohms. Note that the noise variance is directly proportional to temperature, which illustrates the reason for using supercooled amplifiers in low-signal environments, such as for radio astronomy. Note also that the noise variance is independent of frequency, which implies that the noise power spectral density is assumed constant or white. The range of B over which the thermal noise can be assumed white is a function of temperature. However, for temperatures greater than approximately 3 K, the white-noise assumption holds for bandwidths less than approximately 10 GHz. As the temperature increases, the bandwidth over which the white-noise assumption is valid increases. At standard temperature (290 K) the white-noise assumption is valid to bandwidths exceeding 1000 GHz. At very high frequencies other noise sources, such as quantum noise, become significant, and the white-noise assumption is no longer valid. These ideas are discussed in more detail in Appendix A.

We also assume that thermal noise is Gaussian (has a Gaussian amplitude pdf). Since thermal noise results from the random motion of a large number of charge carriers, with each charge carrier making a small contribution to the total noise, the Gaussian assumption is justified through the central-limit theorem. Thus, if we assume that the noise of interest is thermal noise, and the

bandwidth is smaller than 10 to 1000 GHz (depending on temperature), the additive white Gaussian noise (AWGN) model is a valid and useful noise model. We will make this assumption throughout this chapter.

As pointed out in Chapter 1, system noise results from sources external to the system as well as from sources internal to the system. Since noise is unavoidable in any practical system, techniques for minimizing the impact of noise on system performance must often be used if high-performance communications are desired. In the present chapter, appropriate performance criteria for system performance evaluation will be developed. After this, a number of systems will be analyzed to determine the impact of noise on system operation. It is especially important to note the differences between linear and nonlinear systems. We will find that the use of nonlinear modulation, such as FM, allows *improved performance* to be obtained at the expense of *increased transmission bandwidth*. Such trade-offs do not exist when linear modulation is used.

■ 8.1 SIGNAL-TO-NOISE RATIOS

In Chapter 3, systems that involve the operations of modulation and demodulation were studied. In this section we extend that study to the performance of linear demodulators in the presence of noise. We concentrate our efforts on the calculation of signal-to-noise ratios since the signal-to-noise ratio is often a useful and easily determined figure of merit of system performance.

8.1.1 Baseband Systems

In order to have a basis for comparing system performance, we determine the signal-to-noise ratio at the output of a baseband system. Recall that a baseband system involves no modulation or demodulation. Consider Figure 8.1(a). Assume that the signal power is finite at P_T W and that the additive noise has the double-sided power spectral density $\frac{1}{2}N_0$ W/Hz

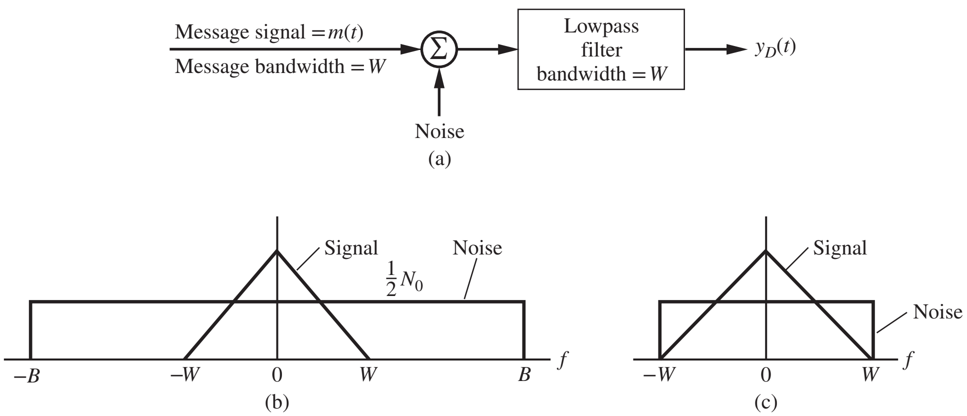

Figure 8.1
Baseband system. (a) Block diagram. (b) Spectra at filter input. (c) Spectra at filter output.

over a bandwidth B, which is assumed to exceed W, as illustrated in Figure 8.1(b). The total noise power in the bandwidth B is

$$\int_{-B}^{B} \frac{1}{2} N_0 df = N_0 B \tag{8.2}$$

and, therefore, the signal-to-noise ratio (SNR) at the filter input is

$$(\text{SNR})_i = \frac{P_T}{N_0 B} \tag{8.3}$$

Since the message signal $m(t)$ is assumed to be bandlimited with bandwidth W, a simple lowpass filter can be used to enhance the SNR. This filter is assumed to pass the signal component without distortion but removes the out-of-band noise as illustrated in Figure 7.1(c). Assuming an ideal filter with bandwidth W, the signal is passed without distortion. Thus, the signal power at the lowpass filter output is P_T, which is the signal power at the filter input. The noise at the filter output is

$$\int_{-W}^{W} \frac{1}{2} N_0 df = N_0 W \tag{8.4}$$

which is less than $N_0 B$ since $W < B$. Thus, the SNR at the filter output is

$$(\text{SNR})_o = \frac{P_T}{N_0 W} \tag{8.5}$$

The filter therefore enhances the SNR by the factor

$$\frac{(\text{SNR})_o}{(\text{SNR})_i} = \frac{P_T}{N_0 W} \frac{N_0 B}{P_T} = \frac{B}{W} \tag{8.6}$$

Since (8.5) describes the SNR achieved with a simple baseband system in which all out-of-band noise is removed by filtering, it is a reasonable standard for making comparisons of system performance. This reference, $P_T / N_0 W$, will be used extensively in the work to follow, in which the output SNR is determined for a variety of basic systems.

8.1.2 Double-Sideband Systems

As a first example, we compute the noise performance of the coherent DSB demodulator first considered in Chapter 3. Consider the block diagram in Figure 8.2, which illustrates a coherent demodulator preceded by a predetection filter. Typically, the predetection filter is the IF filter as discussed in Chapter 3. The input to this filter is the modulated signal plus white Gaussian noise of double-sided power spectral density $\frac{1}{2} N_0$ W/Hz. Since the transmitted signal $x_c(t)$ is

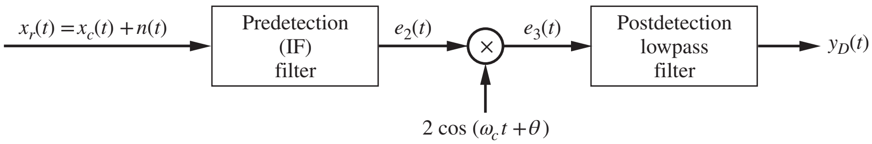

Figure 8.2
Double-sideband demoulator.

assumed to be a DSB signal, the received signal $x_r(t)$ can be written as

$$x_r(t) = A_c m(t) \cos(2\pi f_c t + \theta) + n(t) \tag{8.7}$$

where $m(t)$ is the message and θ is used to denote our uncertainty of the carrier phase or, equivalently, the time origin. Note that, using this model, the SNR at the input to the predetection filter is zero since the power in white noise is infinite. If the predetection filter bandwidth is (ideally) $2W$, the DSB signal is completely passed by the filter. Using the technique developed in Chapter 7, the noise at the predetection filter output can be expanded into its direct and quadrature components. This gives

$$\begin{aligned} e_2(t) = {} & A_c m(t) \cos(2\pi f_c t + \theta) \\ & + n_c(t) \cos(2\pi f_c t + \theta) - n_s(t) \sin(2\pi f_c t + \theta) \end{aligned} \tag{8.8}$$

where the total noise power is $\overline{n_0^2(t)} = \frac{1}{2}\overline{n_c^2(t)} = \frac{1}{2}\overline{n_s^2(t)}$ and is equal to $2N_0W$.

The predetection SNR, measured at the input to the multiplier, is easily determined. The signal power is $\frac{1}{2}A_c^2\overline{m^2}$, where m is understood to be a function of t and the noise power is $2N_0W$ as shown in Figure 8.3(a). This yields the predetection SNR,

$$(\text{SNR})_T = \frac{A_c^2\overline{m^2}}{4WN_0} \tag{8.9}$$

In order to compute the postdetection SNR, $e_3(t)$ is first computed. This gives

$$\begin{aligned} e_3(t) = {} & A_c m(t) + n_c(t) + A_c m(t) \cos[2(2\pi f_c t + \theta)] \\ & + n_c(t) \cos[2(2\pi f_c t + \theta)] - n_s(t) \sin[2(2\pi f_c t + \theta)] \end{aligned} \tag{8.10}$$

The double-frequency terms about $2f_c$ are removed by the postdetection filter to produce the baseband (demodulated) signal

$$y_D(t) = A_c m(t) + n_c(t) \tag{8.11}$$

Note that additive noise on the demodulator input gives rise to additive noise at the demodulator output. This is a property of linearity.

The postdetection signal power is $A_c^2\overline{m^2}$, and the postdetection noise power is $\overline{n_c^2}$ or $2N_0W$, as shown on Figure 8.3(b). This gives the postdetection SNR as

$$(\text{SNR})_D = \frac{A_c^2\overline{m^2}}{2N_0W} \tag{8.12}$$

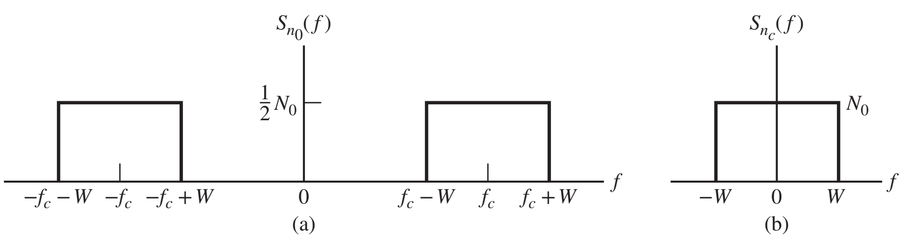

Figure 8.3
(a) Predetection and (b) postdetection filter output noise spectra for DSB demodulation.

Since the signal power is $\frac{1}{2}A_c^2\overline{m^2} = P_T$, we can write the postdetection SNR as

$$(\text{SNR})_D = \frac{P_T}{N_0 W} \tag{8.13}$$

which is equivalent to the ideal baseband system.

The ratio of $(\text{SNR})_D$ to $(\text{SNR})_T$ is referred to as *detection gain* and is often used as a figure of merit for a demodulator. Thus, for the coherent DSB demodulator, the detection gain is

$$\frac{(\text{SNR})_D}{(\text{SNR})_T} = \frac{A_c^2\overline{m^2}}{2N_0 W}\frac{4N_0 W}{A_c^2\overline{m^2}} = 2 \tag{8.14}$$

At first sight, this result is somewhat misleading, for it appears that we have gained 3 dB. This is true for the demodulator because it suppresses the quadrature noise component. However, a comparison with the baseband system reveals that nothing is gained, insofar as the SNR at the system output is concerned. The predetection filter bandwidth must be $2W$ if DSB modulation is used. This results in double the noise bandwidth at the output of the predetection filter and, consequently, double the noise power. The 3-dB detection gain is exactly sufficient to overcome this effect and give an overall performance equivalent to the baseband reference given by (8.5). Note that this ideal performance is only achieved if all out-of-band noise is removed and if the demodulation carrier is perfectly phase coherent with the original carrier used for modulation.

In practice, PLLs, as we studied in Chapter 4, are used to establish carrier recovery at the demodulator. If noise is present in the loop bandwidth, phase jitter will result. We will consider the effect on performance resulting from a combination of additive noise and demodulation phase errors in a later section.

8.1.3 Single-Sideband Systems

Similar calculations are easily carried out for SSB systems. For SSB, the predetection filter input can be written as

$$x_r(t) = A_c[m(t)\cos(2\pi f_c t + \theta) \pm \hat{m}(t)\sin(2\pi f_c t + \theta)] + n(t) \tag{8.15}$$

where $\hat{m}(t)$ denotes the Hilbert transform of $m(t)$. Recall from Chapter 3 that the plus sign is used for LSB SSB and the minus sign is used for USB SSB. Since the minimum bandwidth of the predetection bandpass filter is W for SSB, the center frequency of the predetection filter is $f_x = f_c \pm \frac{1}{2}W$, where the sign depends on the choice of sideband.

We could expand the noise about the center frequency $f_x = f_c \pm \frac{1}{2}W$, since, as we saw in Chapter 7, we are free to expand the noise about any frequency we choose. It is slightly more convenient, however, to expand the noise about the carrier frequency f_c. For this case, the predetection filter output can be written as

$$\begin{aligned} e_2(t) = {} & A_c[m(t)\cos(2\pi f_c t + \theta) \pm \hat{m}(t)\sin(2\pi f_c t + \theta)] \\ & + n_c(t)\cos(2\pi f_c t + \theta) - n_s(t)\sin(2\pi f_c t + \theta) \end{aligned} \tag{8.16}$$

where, as can be seen from Figure 8.4(a),

$$N_T = \overline{n^2} = \overline{n_c^2} = \overline{n_s^2} = N_0 W \tag{8.17}$$

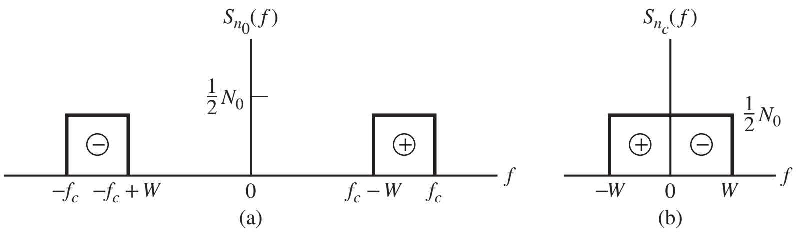

Figure 8.4
(a) Predetection and (b) postdection filter output spectra for lower-sideband SSB (+) and (−) signs denote spectral translation of positive and negative portions of $S_{n_0}(f)$ due to demodulation, respectively.

Equation (8.16) can be written

$$\begin{aligned} e_2(t) &= [A_c m(t) + n_c(t)]\cos(2\pi f_c t + \theta) \\ &+ [A_c \hat{m}(t) \mp n_s(t)]\sin(2\pi f_c t + \theta) \end{aligned} \tag{8.18}$$

As discussed in Chapter 3, demodulation is accomplished by multiplying $e_2(t)$ by the demodulation carrier $2\cos(2\pi f_c t + \theta)$ and lowpass filtering. Thus, the coherent demodulator illustrated in Figure 8.2 also accomplishes demodulation of SSB. It follows that

$$y_D(t) = A_c m(t) + n_c(t) \tag{8.19}$$

We see that coherent demodulation removes $\hat{m}(t)$ as well as the quadrature noise component $n_s(t)$. The power spectral density of $n_c(t)$ is illustrated in Figure 8.4(b) for the case of LSB SSB. Since the postdetection filter passes only $n_c(t)$, the postdetection noise power is

$$N_D = \overline{n_c^2} = N_0 W \tag{8.20}$$

From (8.19) it follows that the postdetection signal power is

$$S_D = A_c^2 \overline{m^2} \tag{8.21}$$

We now turn our attention to the predetection terms.

The predetection signal power is

$$S_T = \overline{\{A_c[m(t)\cos(2\pi f_c t + \theta) \pm \hat{m}(t)\sin(2\pi f_c t + \theta)]\}^2} \tag{8.22}$$

In Chapter 2 we pointed out that a function and its Hilbert transform are orthogonal. If $\overline{m(t)} = 0$, it follows that $\overline{m(t)\hat{m}(t)} = E\{m(t)\}E\{\hat{m}(t)\} = 0$. Thus, the preceding expression becomes

$$S_T = A_c^2\left[\frac{1}{2}\overline{m^2(t)} + \frac{1}{2}\overline{\hat{m}^2(t)}\right] \tag{8.23}$$

It was also shown in Chapter 2 that a function and its Hilbert transform have equal power. Applying this to (8.23) yields

$$S_T = A_c^2 \overline{m^2} \tag{8.24}$$

Since both the predetection and postdetection bandwidths are W, it follows that they have equal power. Therefore,

$$N_T = N_D = N_0 W \tag{8.25}$$

and the detection gain is

$$\frac{(\mathrm{SNR})_D}{(\mathrm{SNR})_T} = \frac{A_c^2\overline{m^2}}{N_0W}\,\frac{N_0W}{A_c^2\overline{m^2}} = 1 \tag{8.26}$$

The SSB system lacks the 3-dB detection gain of the DSB system. However, the predetection noise power of the SSB system is 3 dB less than that for the DSB system if the predetection filters have minimum bandwidth. This results in equal performance, given by

$$(\mathrm{SNR})_D = \frac{A_c^2\overline{m^2}}{N_0W} = \frac{P_T}{N_0W} \tag{8.27}$$

Thus, coherent demodulation of both DSB and SSB results in performance equivalent to baseband.

8.1.4 Amplitude Modulation Systems

The main reason for using AM is that simple envelope demodulation (or detection) can be used at the receiver. In many applications the receiver simplicity more than makes up for the loss in efficiency that we observed in Chapter 3. Therefore, coherent demodulation is not often used in AM. Despite this fact, we consider coherent demodulation briefly since it provides a useful insight into performance in the presence of noise.

Coherent Demodulation of AM Signals

We saw in Chapter 3 that an AM signal is defined by

$$x_c(t) = A_c[1 + am_n(t)]\cos(2\pi f_c t + \theta) \tag{8.28}$$

where $m_n(t)$ is the modulation signal normalized so that the maximum value of $|m_n(t)|$ is unity [assuming $m(t)$ has a symmetrical pdf about zero] and a is the modulation index. Assuming coherent demodulation, it is easily shown, by using a development parallel to that used for DSB systems, that the demodulated output in the presence of noise is

$$y_D(t) = A_c am_n(t) + n_c(t) \tag{8.29}$$

The DC term resulting from multiplication of $x_c(t)$ by the demodulation carrier is not included in (8.29) for two reasons. First, this term is not considered part of the signal since it contains no information. [Recall that we have assumed $\overline{m(t)} = 0$.] Second, most practical AM demodulators are not DC-coupled, so a DC term does not appear on the output of a practical system. In addition, the DC term is frequently used for automatic gain control (AGC) and is therefore held constant at the transmitter.

From (8.29) it follows that the signal power in $y_D(t)$ is

$$S_D = A_c^2a^2\overline{m_n^2} \tag{8.30}$$

and, since the bandwidth of the transmitted signal is $2W$, the noise power is

$$N_D = \overline{n_c^2} = 2N_0W \tag{8.31}$$

For the predetection case, the signal power is

$$S_T = P_T = \frac{1}{2}A_c^2(1 + a^2\overline{m_n^2}) \tag{8.32}$$

and the predetection noise power is

$$N_T = 2N_0W \tag{8.33}$$

Thus, the detection gain is

$$\frac{(\text{SNR})_D}{(\text{SNR})_T} = \frac{A_c^2 a^2 \overline{m_n^2}/2N_0W}{(A_c^2 + A_c^2 a^2 \overline{m_n^2})/4N_0W} = \frac{2a^2\overline{m_n^2}}{1 + a^2\overline{m_n^2}} \tag{8.34}$$

which is dependent on the modulation index.

Recall that when we studied AM in Chapter 3 the efficiency of an AM transmission system was defined as the ratio of sideband power to total power in the transmitted signal $x_c(t)$. This resulted in the efficiency E_{ff} being expressed as

$$E_{ff} = \frac{a^2\overline{m_n^2}}{1 + a^2\overline{m_n^2}} \tag{8.35}$$

where the overbar, denoting a statistical average, has been substituted for the time-average notation $\langle \cdot \rangle$ used in Chapter 3. It follows from (8.34) and (8.35) that the detection gain can be expressed as

$$\frac{(\text{SNR})_D}{(\text{SNR})_T} = 2E_{ff} \tag{8.36}$$

Since the predetection SNR can be written as

$$(\text{SNR})_T = \frac{S_T}{2N_0W} = \frac{P_T}{2N_0W} \tag{8.37}$$

it follows that the SNR at the demodulator output can be written as

$$(\text{SNR})_D = E_{ff}\frac{P_T}{N_0W} \tag{8.38}$$

Recall that in Chapter 3 we defined the efficiency of an AM system as the ratio of sideband power to the total power in an AM signal. The preceding expression gives another, and perhaps better, way to view efficiency.

If the efficiency *could* be 1, AM would have the same postdetection SNR as the ideal DSB and SSB systems. Of course, as we saw in Chapter 3, the efficiency of AM is typically much less than 1 and the postdetection SNR is correspondingly lower. Note that an efficiency of 1 requires that the modulation index $a \to \infty$ so that the power in the unmodulated carrier is negligible compared to the total transmitted power. However, for $a > 1$ envelope demodulation cannot be used and AM loses its advantage.

EXAMPLE 8.1

An AM system operates with a modulation index of 0.5, and the power in the normalized message signal is 0.1 W. The efficiency is

$$E_{ff} = \frac{(0.5)^2(0.1)}{1 + (0.5)^2(0.1)} = 0.0244 \tag{8.39}$$

and the postdetection SNR is

$$(\text{SNR})_D = 0.0244 \frac{P_T}{N_0 W} \tag{8.40}$$

The detection gain is

$$\frac{(\text{SNR})_D}{(\text{SNR})_T} = 2E_{ff} = 0.0488 \tag{8.41}$$

This is more than 16 dB inferior to the ideal system requiring the same bandwidth. It should be remembered, however, that the motivation for using AM is not noise performance but rather that AM allows the use of simple envelope detectors for demodulation. The reason, of course, for the poor efficiency of AM is that a large fraction of the total transmitted power lies in the carrier component, which conveys no information since it is not a function of the message signal.

■

Envelope Demodulation of AM Signals

Since envelope detection is the usual method of demodulating an AM signal, it is important to understand how envelope demodulation differs from coherent demodulation in the presence of noise. The received signal at the input to the envelope demodulator is assumed to be $x_c(t)$ plus narrowband noise. Thus,

$$\begin{aligned} x_r(t) = {} & A_c[1 + am_n(t)]\cos(2\pi f_c t + \theta) \\ & + n_c(t)\cos(2\pi f_c t + \theta) - n_s(t)\sin(2\pi f_c t + \theta) \end{aligned} \tag{8.42}$$

where, as before, $\overline{n_c^2} = \overline{n_s^2} = 2N_0 W$. The signal $x_r(t)$ can be written in terms of envelope and phase as

$$x_r(t) = r(t)\cos[2\pi f_c t + \theta + \phi(t)] \tag{8.43}$$

where

$$r(t) = \sqrt{\{A_c[1 + am_n(t)] + n_c(t)\}^2 + n_s^2(t)} \tag{8.44}$$

and

$$\phi(t) = \tan^{-1}\left(\frac{n_s(t)}{A_c[1 + am_n(t)] + n_c(t)}\right) \tag{8.45}$$

Since the output of an ideal envelope detector is independent of phase variations of the input, the expression for $\phi(t)$ is of no interest, and we will concentrate on $r(t)$. The envelope detector is assumed to be AC coupled so that

$$y_D(t) = r(t) - \overline{r(t)} \tag{8.46}$$

where $\overline{r(t)}$ is the average value of the envelope amplitude. Equation (8.46) will be evaluated for two cases. First, we consider the case in which $(\text{SNR})_T$ is large, and then we briefly consider the case in which the $(\text{SNR})_T$ is small.

Envelope Demodulation: Large $(\text{SNR})_T$ For $(\text{SNR})_T$ sufficiently large, the solution is simple. From (8.44), we see that if

$$|A_c[1 + am_n(t)] + n_c(t)| \gg |n_s(t)| \tag{8.47}$$

then *most of the time*

$$r(t) \cong A_c[1 + am_n(t)] + n_c(t) \tag{8.48}$$

yielding, after removal of the DC component,

$$y_D(t) \cong A_c am_n(t) + n_c(t) \tag{8.49}$$

This is the final result for the case in which the SNR is large.

Comparing (8.49) and (8.29) illustrates that the output of the envelope detector is equivalent to the output of the coherent detector if $(\mathrm{SNR})_T$ is large. The detection gain for this case is therefore given by (8.34).

Envelope Demodulation: Small $(\mathrm{SNR})_T$ For the case in which $(\mathrm{SNR})_T$ is small, the analysis is somewhat more complex. In order to analyze this case, we recall from Chapter 7 that $n_c(t)\cos(2\pi f_c t + \theta) - n_s(t)\sin(2\pi f_c t + \theta)$ can be written in terms of envelope and phase, so that the envelope detector input can be written as

$$\begin{aligned} e(t) = {} & A_c[1 + am_n(t)]\cos(2\pi f_c t + \theta) \\ & + r_n(t)\cos[2\pi f_c t + \theta + \phi_n(t)] \end{aligned} \tag{8.50}$$

For $(\mathrm{SNR})_T \ll 1$, the amplitude of $A_c[1 + am_n(t)]$ will usually be much smaller than $r_n(t)$. Consider the phasor diagram illustrated in Figure 8.5, which is drawn for $r_n(t)$ greater than $A_c[1 + am_n(t)]$. It can be seen that $r(t)$ is approximated by

$$r(t) \cong r_n(t) + A_c[1 + am_n(t)]\cos[\phi_n(t)] \tag{8.51}$$

yielding

$$y_D(t) \cong r_n(t) + A_c[1 + am_n(t)]\cos[\phi_n(t)] - \overline{r(t)} \tag{8.52}$$

The principal component of $y_D(t)$ is the Rayleigh-distributed noise envelope, and no component of $y_D(t)$ is proportional to the signal. Note that since $n_c(t)$ and $n_s(t)$ are random, $\cos[\phi_n(t)]$ is also random. Thus, the signal $m_n(t)$ is multiplied by a random quantity. This multiplication of the signal by a function of the noise has a significantly worse degrading effect than does additive noise.

This severe loss of signal at low-input SNR is known as the *threshold effect* and results from the nonlinear action of the envelope detector. In coherent detectors, which are linear, the signal and noise are additive at the detector output *if* they are additive at the detector input. The result is that the signal retains its identity even when the input SNR is low. We have seen

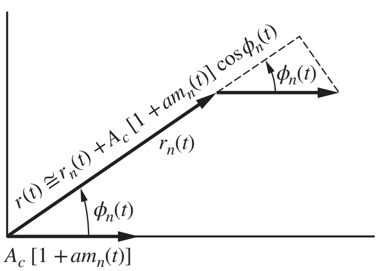

Figure 8.5
Phasor diagram for AM with $(\mathrm{SNR})_T \ll 1$ (drawn for $\theta = 0$).

that this is not true for nonlinear demodulators. For this reason, coherent detection is often desirable when the noise is large.

Square-Law Demodulation of AM Signals

The determination of the SNR at the output of a nonlinear system is often a very difficult task. The *square-law detector*, however, is one system for which this is not the case. In this section, we conduct a simplified analysis to illustrate the phenomenon of thresholding, which is characteristic of nonlinear systems.

In the analysis to follow, the postdetection bandwidth will be assumed twice the message bandwidth W. This is not a necessary assumption, but it does result in a simplification of the analysis without impacting the threshold effect. We will also see that harmonic and/or intermodulation distortion is a problem with square-law detectors, an effect that may preclude their use.

Square-law demodulators are implemented as a squaring device followed by a lowpass filter. The response of a square-law demodulator to an AM signal is $r^2(t)$, where $r(t)$ is defined by (8.44). Thus, the output of the square-law device can be written as

$$r^2(t) = \{A_c[l + am_n(t)] + n_c(t)\}^2 + n_s^2(t) \tag{8.53}$$

We now determine the output SNR. Carrying out the indicated squaring operation gives

$$\begin{aligned} r^2(t) = {} & A_c^2 + 2A_c^2 am_n(t) + A_c^2 a^2 m_n^2(t) \\ & +2A_c n_c(t) + 2A_c an_c(t)m_n(t) + n_c^2(t) + n_s^2(t) \end{aligned} \tag{8.54}$$

First, consider the first line of the preceding equation. The first term, A_c^2, is a DC term and is neglected. It is not a function of the signal and is not a function of noise. In addition, in most practical cases, the detector output is assumed AC coupled, so that DC terms are blocked. The second term is proportional to the message signal and represents the desired output. The third term is signal-induced distortion (harmonic and intermodulation) and will be considered separately. All four terms on the second line of (8.54) represent noise. We now consider the calculation of $(\text{SNR})_D$.

The signal and noise components of the output are written as

$$s_D(t) = 2A_c^2 am_n(t) \tag{8.55}$$

and

$$n_D(t) = 2A_c n_c(t) + 2A_c an_c(t)m_n(t) + n_c^2(t) + n_s^2(t) \tag{8.56}$$

respectively. The power in the signal component is

$$S_D = 4A_c^4 a^2 \overline{m_n^2} \tag{8.57}$$

and the noise power is

$$N_D = 4A_c^2 \overline{n_c^2} + 4A_c^2 a^2 \overline{n_c^2 m_n^2} + \sigma^2_{n_c^2+n_s^2} \tag{8.58}$$

The last term is given by

$$\sigma^2_{n_c^2+n_s^2} = E\left\{[n_c^2(t) + n_s^2(t)]^2\right\} - E^2[n_c^2(t) + n_s^2(t)] = 4\sigma_n^2 \tag{8.59}$$

where, as always, $\sigma_n^2 = \overline{n_c^2} = \overline{n_s^2}$. Thus,

$$N_D = 4A_c^2\sigma_n^2 + 4A_c^2 a^2 \overline{m_n^2(t)}\sigma_n^2 + 4\sigma_n^4 \tag{8.60}$$

This gives

$$(\text{SNR})_D = \frac{a^2\overline{m_n^2}(A_c^2/\sigma_n^2)}{\left(1 + a^2\overline{m_n^2}\right) + (\sigma_n^2/A_c^2)} \tag{8.61}$$

Recognizing that $P_T = \frac{1}{2}A_c^2(1 + a^2\overline{m_n^2})$ and $\sigma_n^2 = 2N_0W$, A_c^2/σ_n^2 can be written

$$\frac{A_c^2}{\sigma_n^2} = \frac{P_T}{\left[1 + a^2\overline{m_n^2(t)}\right] N_0W} \tag{8.62}$$

Substitution into (8.61) gives

$$(\text{SNR})_D = \frac{a^2\overline{m_n^2}}{\left(1 + a^2\overline{m_n^2}\right)^2} \frac{P_T/N_0W}{1 + N_0W/P_T} \tag{8.63}$$

For high SNR operation $P_T \gg N_0W$ and the second term in the denominator is negligible. For this case,

$$(\text{SNR})_D = \frac{a^2\overline{m_n^2}}{\left(1 + a^2\overline{m_n^2}\right)^2} \frac{P_T}{N_0W}, \quad P_T \gg N_0W \tag{8.64}$$

while for low SNR operation $N_0W \gg P_T$ and

$$(\text{SNR})_D = \frac{a^2\overline{m_n^2}}{\left(1 + a^2\overline{m_n^2}\right)^2} \left(\frac{P_T}{N_0W}\right)^2, \quad N_0W \gg P_T \tag{8.65}$$

Figure 8.6 illustrates (8.63) for several values of the modulation index a assuming sinusoidal modulation. We see that, on a log (decibel) scale, the slope of the detection gain characteristic below threshold is double the slope above threshold. The threshold effect is therefore obvious.

Recall that in deriving (8.63), from which (8.64) and (8.65) followed, we neglected the third term in (8.54), which represents signal-induced distortion. From (8.54) and (8.57) the distortion-to-signal-power ratio, denoted D_D/S_D, is

$$\frac{D_D}{S_D} = \frac{A_c^4 a^4 \overline{m_n^4}}{4A_c^4 a^2 \overline{m_n^2}} = \frac{a^2}{4}\frac{\overline{m_n^4}}{\overline{m_n^2}} \tag{8.66}$$

If the message signal is Gaussian with variance σ_m^2, the preceding becomes

$$\frac{D_D}{S_D} = \frac{3}{4}a^2\sigma_m^2 \tag{8.67}$$

We see that signal-induced distortion can be reduced by decreasing the modulation index. However, as illustrated in Figure 8.6, a reduction of the modulation index also results in a decrease in the output SNR. Is the distortion signal or noise? This question will be discussed in the following section.

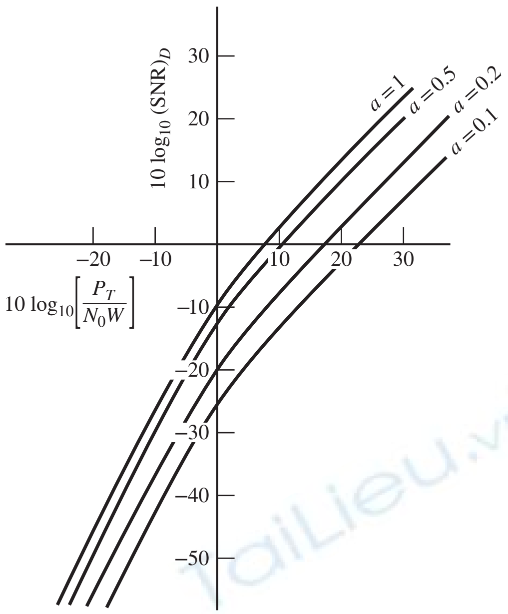

Figure 8.6
Performance of a square-law detector (sinusoidal modulation assumed).

The *linear envelope detector* defined by (8.44) is much more difficult to analyze over a wide range of SNRs because of the square root. However, to a first approximation, the performance of a linear envelope detector and a square-law envelope detector are the same. Harmonic distortion is also present in linear envelope detectors, but the amplitude of the distortion component is significantly less than that observed for square-law detectors. In addition, it can be shown that for high SNRs and a modulation index of unity, the performance of a linear envelope detector is better by approximately 1.8 dB than the performance of a square-law detector. (See Problem 8.13.)

8.1.5 An Estimator for Signal-to-Noise Ratios

Note that in the preceding section, the output consisted of a signal term, a noise term, and a distortion term. An important question arises. Is the distortion term part of the signal or is it part of the noise? It is clearly signal-generated distortion. The answer lies in the nature of the distortion term. A reasonable way of viewing this issue is to decompose the distortion term into a component orthogonal to the signal. This component is treated as noise. The other component, in phase with the signal, is treated as signal.

Assume that a signal, $x(t)$ is the input to a system and that $x(t)$ gives rise to an output, $y(t)$. We say that $y(t)$ is a ''perfect'' version of $x(t)$ if the waveform for $y(t)$ only differs from $x(t)$ by an amplitude scaling and a time delay. In Chapter 2 we defined a system having this property as a distortionless system We also require that $y(t)$ is noiseless. For such a system

$$y(t) = Ax(t - \tau) \tag{8.68}$$

where A is the system gain and τ is the system time delay. The SNR of $y(t)$, referred to $x(t)$, is infinite. Now let's assume that $y(t)$ contains both noise and distortion. Now

$$y(t) \neq Ax(t - \tau) \tag{8.69}$$

It is reasonable to assume that the noise power is the mean-square error

$$\overline{\epsilon^2(A,\tau)} = E\left\{[y(t) - Ax(t-\tau)]^2\right\} \tag{8.70}$$

where $E\{\bullet\}$ denotes statistical expection. Carrying out the obvious multipication, the preceding equation can be written

$$\overline{\epsilon^2(A,\tau)} = E\left\{y^2(t)\right\} + A^2 E\left\{x^2(t-\tau)\right\} - 2AE\{y(t)x(t-\tau)\} \tag{8.71}$$

The first term in the preceding expression is, by definition, the power in the observed (measurement) signal $y(t)$, which we denote P_y. Since A is a constant system parameter, the second term is A^2P_x, where P_x is the power in $x(t)$. We note that shifting a signal in time does not change the signal power. The last term is

$$2AE\{y(t)x(t-\tau)\} = 2AE\{x(t)y(t+\tau)\} = 2AR_{XY}(\tau) \tag{8.72}$$

Note that we have assumed stationarity. The final expression for the mean-square error is

$$\overline{\epsilon^2(A,\tau)} = P_y + A^2P_x - 2AR_{XY}(\tau) \tag{8.73}$$

We now find the values of A and τ that minimize the mean-square error.

Since A, the system gain, is a fixed but yet unknown positive constant, we wish to find the value of τ that maximizes the cross-correlation $R_{XY}(\tau)$. This is denoted $R_{XY}\left(\tau_m\right)$. Note that $R_{XY}\left(\tau_m\right)$ is the standard definition of system delay.

The value of A that minimizes the mean-square error, denoted A_m, is determined from

$$\frac{d}{dA}\left\{P_y + A^2P_x - 2AR_{XY}(\tau_m)\right\} = 0 \tag{8.74}$$

which gives

$$A_m = \frac{R_{XY}(\tau_m)}{P_x} \tag{8.75}$$

Substitution into (8.73) gives

$$\overline{\epsilon^2(A_m,\tau_m)} = P_y - \frac{R_{XY}^2(\tau_m)}{P_x} \tag{8.76}$$

which is the noise power. The signal power is A^2P_x. Therefore, the SNR is

$$\text{SNR} = \frac{\left[\frac{R_{XY}(\tau_m)}{P_x}\right]^2 P_x}{P_y - \frac{R_{XY}^2(\tau_m)}{P_x}} \tag{8.77}$$

Multiplying by P_x gives the SNR

$$\text{SNR} = \frac{R_{XY}^2(\tau_m)}{P_xP_y - R_{XY}^2(\tau_m)} \tag{8.78}$$

The MATLAB code for the signal-to-noise ratio, along with other important parameters such as gain (A), delay (τ_m), P_x, P_y and $R_{XY}\left(\tau_m\right)$, is as follows:

```
% File: snrest.m
function [gain,delay,px,py,rxy,rho,snrdb] = snrest(x,y)
ln = length(x);            % Set length of the reference (x) vector
fx = fft(x,ln);            % FFT the reference (x) vector
```

```
fy = fft(y,ln);              % FFT the measurement (y) vector
fxconj = conj(fx);           % Conjugate the FFT of the reference vector
sxy = fy .* fxconj;          % Determine the cross PSD
rxy = ifft(sxy,ln);          % Determine the cross-correlation function
rxy = real(rxy)/ln;          % Take the real part and scale
px = x*x'/ln;                % Determine power in reference vector
py = y*y'/ln;                % Determine power in measurement vector
[rxymax,j] = max(rxy);       % Find the max of the cross correlation
gain = rxymax/px;            % Estimate of the Gain
delay = j-1;                 % Estimate of the Delay
rxy2 = rxymax*rxymax;        % Square rxymax for later use
rho = rxymax/sqrt(px*py);    % Estimate of the correlation coefficient
snr = rxy2/(px*py-rxy2);     % Estimate of the SNR
snrdb = 10*log10(snr);       % SNR estimate in db
% End of script file.
```

The following three examples illustrate the technique.

COMPUTER EXAMPLE 8.1

In this example we consider a simple interference problem. The signal is assumed to be

$$x(t) = 2\sin(2\pi f t) \tag{8.79}$$

and the measurement signal is

$$y(t) = 10\sin(2\pi f t + \pi) + 0.1\sin(20\pi f t) \tag{8.80}$$

In order to determine the gain, delay, signal powers, and the SNR, we execute the following program:

```
% File: c8ce1.m
t = 1:6400;
fs = 1/32;
x = 2*sin(2*pi*fs*t);
y = 10*sin(2*pi*fs*t+pi)+0.1*sin(2*pi*fs*10*t);
[gain,delay,px,py,rxymax,rho,snr,snrdb] = snrest(x,y);
format long e
a = ['The gain estimate is ',num2str(gain),'.'];
b = ['The delay estimate is ',num2str(delay),' samples.'];
c = ['The estimate of px is ',num2str(px),'.'];
d = ['The estimate of py is ',num2str(py),'.'];
e = ['The snr estimate is ',num2str(snr),'.'];
f = ['The snr estimate is ',num2str(snrdb),' db.'];
disp(a); disp(b); disp(c); disp(d); disp(e); disp(f)
% End of script file.
```

Executing the program gives the following results:

```
The gain estimate is 5.
The delay estimate is 16 samples.
The estimate of px is 2.
The estimate of py is 50.005.
The snr estimate is 10000.
The snr estimate is 40 db.
```

Are these results reasonable? Examining $x(t)$ and $y(t)$, we see that the signal gain is $\frac{10}{2} = 5$. Note that there are 32 samples per period of the signal. Since the delay is π, or one-half period, it follows that the signal delay is 16 sample periods. The power in the signal $x(t)$ is clearly 2. The power in the measurement signal is

$$P_y = \frac{1}{2}\left[(10)^2 + (0.1)^2\right] = 50.005$$

and so P_y is correct. The SNR is

$$\text{SNR} = \frac{(10)^2/2}{(0.1)^2/2} = 10000 \tag{8.81}$$

and we see that all parameters have been correctly estimated.

■

COMPUTER EXAMPLE 8.2

In this example, we consider a combination of interference and noise. The signal is assumed to be

$$x(t) = 2\sin(2\pi f t) \tag{8.82}$$

and the measurement signal is

$$y(t) = 10\sin(2\pi f t + \pi) + 0.1\sin(20\pi f t) + n(t) \tag{8.83}$$

In order to determine the gain, delay, signal powers, and the SNR, the following MATLAB script is written:

```
% File: c8ce2.m
t = 1:6400;
fs = 1/32;
x = 2*sin(2*pi*fs*t);
y = 10*sin(2*pi*fs*t+pi)+0.1*sin(2*pi*fs*10*t);
A = 0.1/sqrt(2);
y = y+A*randn(1,6400);
[gain,delay,px,py,rxymax,rho,snr,snrdb] = snrest(x,y);
format long e
a = ['The gain estimate is ',num2str(gain),'.'];
b = ['The delay estimate is ',num2str(delay),' samples.'];
c = ['The estimate of px is ',num2str(px),'.'];
d = ['The estimate of py is ',num2str(py),'.'];
e = ['The snr estimate is ',num2str(snr),'.'];
f = ['The snr estimate is ',num2str(snrdb),' db.'];
disp(a); disp(b); disp(c); disp(d); disp(e); disp(f)
%End of script file.
```

Executing this program gives:

```
The gain estimate is 5.0001.
The delay estimate is 16 samples.
The estimate of px is 2.
The estimate of py is 50.0113.
The snr estimate is 5063.4892.
The snr estimate is 37.0445 db.
```

Are these correct? Comparing this result to the previous computer example we see that the SNR has been reduced by approximately a factor of two. That this is reasonable follows from the fact that the parameter A is the standard deviation of the noise. The noise variance is therefore given by

$$\sigma_n^2 = A^2 = \left(\frac{0.1}{\sqrt{2}}\right)^2 = \frac{(0.1)^2}{2} \tag{8.84}$$

We note from the previous computer example that the ''programmed'' noise variance is exactly equal to the interference power. The SNR should therefore be reduced by 3 dB since the power of the interference plus noise is twice the power of the interference acting alone. Comparing this to the previous computer example, we note, however, that the SNR is reduced by slightly less than 3 dB. The reason for this should be clear from Chapter 7. When the program is run, the noise generated is a finite-length sample function from the noise process. Therefore, the variance of the finite-length sample function is a random variable. The noise is assumed ergodic so that the estimate of the noise variance is consistent, which means that as the number of noise samples N increases, the variance of the estimate decreases. ■

COMPUTER EXAMPLE 8.3

In this example we consider the effect of a nonlinearity on the SNR, which will provide insight into the allocation of the distortion term in the previous section to signal and noise. For this example we define the signal as

$$x(t) = 2\cos(2\pi f t) \tag{8.85}$$

and assume the measurement signal

$$y(t) = 1 - \cos^3(2\pi f t + \pi) \tag{8.86}$$

In order to determine the SNR we execute the following MATLAB script:

```
% File: c8ce3.m
t = 1:6400;
fs = 1/32;
x = 2*cos(2*pi*fs*t);
y = 10*((cos(2*pi*fs*t+pi)).^3);
[gain,delay,px,py,rxymax,rho,snr,snrdb] = snrest(x,y);
format long e
a = ['The gain estimate is ',num2str(gain),'.'];
b = ['The delay estimate is ',num2str(delay),' samples.'];
c = ['The estimate of px is ',num2str(px),'.'];
d = ['The estimate of py is ',num2str(py),'.'];
e = ['The snr estimate is ',num2str(snr),'.'];
f = ['The snr estimate is ',num2str(snrdb),' db.'];
disp(a); disp(b); disp(c); disp(d); disp(e); disp(f)
%End of script file.
```

Executing the program gives the following results:

```
The gain estimate is 3.75.
The delay estimate is 16 samples.
The estimate of px is 2.
The estimate of py is 31.25.
The snr estimate is 9.
The snr estimate is 9.5424 db.
```

Since we take 32 samples per period and the delay is one-half of a period, the delay estimate of 16 samples is clearly correct as is the power in the reference signal P_x. Verifying the other results is not so obvious. Note that the measurement signal can be written

$$y(t) = 10\left(\frac{1}{2}\right)[1 + \cos(4\pi f t)][\cos(2\pi f t)] \tag{8.87}$$

which becomes

$$y(t) = 7.5\cos(2\pi f t) + 2.5\cos(6\pi f t) \tag{8.88}$$

Therefore, the power in the measurement signal is

$$P_y = \frac{1}{2}\left[(7.5)^2 + (2.5)^2\right] = 31.25 \tag{8.89}$$

The SNR is

$$\text{SNR} = \frac{(7.5)^2/2}{(2.5)^2/2} = 9 \tag{8.90}$$

This result provides insight into the allocation of the distortion power in the preceding section into signal and noise powers. That portion of the noise power that is orthogonal to the signal is classified as noise. The portion of the distortion that is correlated or ''in phase'' with the signal is classified as signal. ■

8.2 NOISE AND PHASE ERRORS IN COHERENT SYSTEMS

In the preceding section we investigated the performance of various types of demodulators. Our main interests were detection gain and calculation of the demodulated output SNR. Where coherent demodulation was used, the demodulation carrier was assumed to have *perfect* phase coherence with the carrier used for modulation. In a practical system, as we briefly discussed, the presence of noise in the carrier recovery system prevents perfect estimation of the carrier phase. Thus, system performance in the presence of both additive noise and demodulation phase errors is of interest.

The demodulator model is illustrated in Figure 8.7. The signal portion of $e(t)$ is assumed to be the quadrature double-sideband (QDSB) signal

$$m_1(t)\cos(2\pi f_c t + \theta) + m_2(t)\sin(2\pi f_c t + \theta)$$

where any constant A_c is absorbed into $m_1(t)$ and $m_2(t)$ for notational simplicity. Using this model, a general representation for the error in the demodulated signal $y_D(t)$ is obtained. After the analysis is complete, the DSB result is obtained by letting $m_1(t) = m(t)$ and $m_2(t) = 0$. The SSB result is obtained by letting $m_1(t) = m(t)$ and $m_2(t) = \pm\hat{m}(t)$, depending upon the

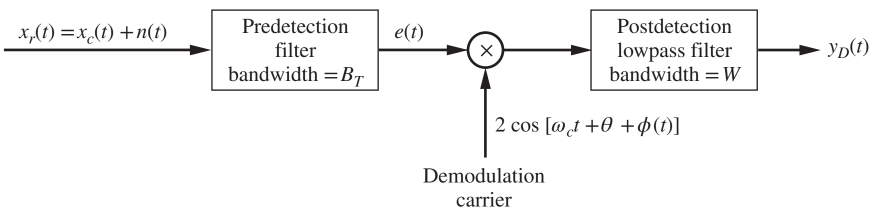

Figure 8.7
Coherent demodulation with phase error.

sideband of interest. For the QDSB system, $y_D(t)$ is the demodulated output for the direct channel. The quadrature channel can be demodulated using a demodulation carrier of the form $2\sin[2\pi f_c t + \theta + \phi(t)]$.

The noise portion of $e(t)$ is represented using the narrowband model

$$n_c(t)\cos(2\pi f_c t + \theta) - n_s(t)\sin(2\pi f_c t + \theta)$$

in which

$$\overline{n_c^2} = \overline{n_s^2} = N_0 B_T = \overline{n_2} = \sigma_n^2 \tag{8.91}$$

where B_T is the bandwidth of the predetection filter, $\frac{1}{2}N_0$ is the double-sided power spectral density of the noise at the filter input, and σ_n^2 is the noise variance (power) at the output of the predetection filter. The phase error of the demodulation carrier is assumed to be a sample function of a zero-mean Gaussian process of known variance σ_ϕ^2. As before, the message signals are assumed to have zero mean.

With the preliminaries of defining the model and stating the assumptions disposed of, we now proceed with the analysis. The assumed performance criterion is mean-square error in the demodulated output $y_D(t)$. Therefore, we will compute

$$\overline{\epsilon^2} = \overline{\{m_1(t) - y_D(t)\}^2} \tag{8.92}$$

for DSB, SSB, and QDSB. The multiplier input signal $e(t)$ in Figure 8.7 is

$$\begin{aligned} e(t) &= m_1(t)\cos(2\pi f_c t + \theta) + m_2(t)\sin(2\pi f_c t + \theta) \\ &\quad + n_c(t)\cos(2\pi f_c t + \theta) - n_s(t)\sin(2\pi f_c t + \theta) \end{aligned} \tag{8.93}$$

Multiplying by $2\cos(2\pi f_c t + \theta + \phi(t))$ and lowpass filtering gives us the output

$$y_D(t) = [m_1(t) + n_c(t)]\cos\phi(t) - [m_2(t) - n_s(t)]\sin\phi(t) \tag{8.94}$$

The error $m_1(t) - y_D(t)$ can be written as

$$\epsilon = m_1 - (m_1 + n_c)\cos\phi + (m_2 - n_s)\sin\phi \tag{8.95}$$

where it is understood that $\epsilon, m_1, m_2, n_c, n_s$, and ϕ are all functions of time. The mean-square error can be written as

$$\begin{aligned} \overline{\epsilon^2} = \overline{m_1^2} &- \overline{2m_1(m_1 + n_c)\cos\phi} \\ &+ 2\overline{m_1(m_2 + n_s)\sin\phi} \\ &+ \overline{(m_1 + n_c)^2\cos^2\phi} \\ &- 2\overline{(m_1 + n_c)(m_2 - n_s)\sin\phi\cos\phi} \\ &+ \overline{(m_2 - n_s)^2\sin^2\phi} \end{aligned} \tag{8.96}$$

The variables m_1, m_2, n_c, n_s, and ϕ are all assumed to be uncorrelated. It should be pointed out that for the SSB case, the power spectra of $n_c(t)$ and $n_s(t)$ will not be symmetrical about f_c. However, as pointed out in Section 7.5, $n_c(t)$ and $n_s(t)$ are uncorrelated, since there is no time displacement. Thus, the mean-square error can be written as

$$\overline{\epsilon^2} = \overline{m_1^2} - \overline{2m_1^2\cos\phi} + \overline{m_1^2\cos^2\phi} + \overline{n^2} \tag{8.97}$$

and we are in a position to consider specific cases.

First, let us assume the system of interest is QDSB with equal power in each modulating signal. Under this assumption, $\overline{m_1^2} = \overline{m_2^2} = \sigma_m^2$, and the mean-square error is

$$\overline{\epsilon_Q^2} = 2\sigma_m^2 - 2\sigma_m^2\overline{\cos\phi} + 2\sigma_n^2 \tag{8.98}$$

This expression can be easily evaluated for the case in which the maximum value of $|\phi(t)| \ll 1$ so that $\phi(t)$ can be represented by the first two terms in a power series expansion. Using the approximation

$$\overline{\cos\phi} \cong \overline{1 - \frac{1}{2}\phi^2} = 1 - \frac{1}{2}\sigma_\phi^2 \tag{8.99}$$

gives

$$\overline{\epsilon_Q^2} = \sigma_m^2\sigma_\phi^2 + \sigma_n^2 \tag{8.100}$$

In order to have an easily interpreted measure of system performance, the mean-square error is normalized by σ_m^2. This yields

$$\overline{\epsilon_{NQ}^2} = \sigma_\phi^2 + \frac{\sigma_n^2}{\sigma_m^2} \tag{8.101}$$

Note that the first term is the phase-error variance and the second term is simply the reciprocal of the SNR. Note that for high SNR the important error source is the phase error.

The preceding expression is also valid for the SSB case, since an SSB signal is a QDSB signal with equal power in the direct and quadrature components. However, σ_n^2 may be different for the SSB and QDSB cases, since the SSB predetection filter bandwidth need only be half the bandwidth of the predetection filter for the QDSB case. Equation (8.101) is of such general interest that it is illustrated in Figure 8.8.

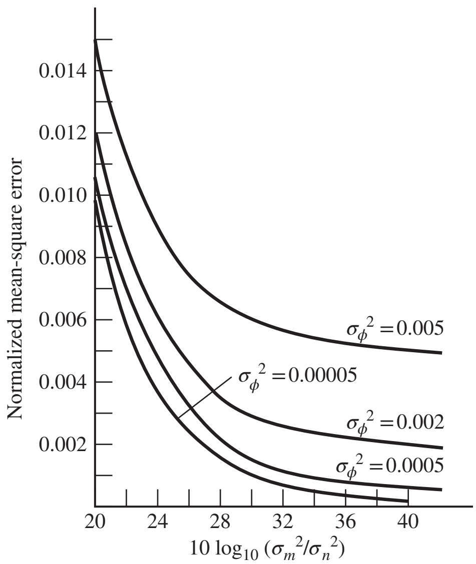

Figure 8.8
Mean-square error versus SNR for QDSB system.