Downloaded From : www.EasyEngineering.net



Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

Downloaded From : www.EasyEngineering.net

# PRINCIPLES OF COMMUNICATION SYSTEMS

Second Edition

**Herbert Taub**

**Donald L. Schilling**

*Professors of Electrical Engineering*
*The City College of New York*

**McGraw-Hill Publishing Company**

New York St. Louis San Francisco Auckland Bogotá
Caracas Hamburg Lisbon London Madrid Mexico Milan
Montreal New Delhi Oklahoma City Paris San Juan
São Paulo Singapore Sydney Tokyo Toronto



Downloaded From : www.EasyEngineering.net

CHAPTER

# ONE

# SPECTRAL ANALYSIS

## INTRODUCTION

Suppose that two people, separated by a considerable distance, wish to communicate with one another. If there is a pair of conducting wires extending from one location to another, and if each place is equipped with a microphone and earpiece, the communication problem may be solved. The microphone, at one end of the wire communications channel, impresses an electric signal voltage on the line, which voltage is then received at the other end. The received signal, however, will have associated with it an erratic, random, unpredictable voltage waveform which is described by the term *noise*. The origin of this noise will be discussed more fully in Chaps. 7 and 14. Here we need but to note that at the atomic level the universe is in a constant state of agitation, and that this agitation is the source of a very great deal of this noise. Because of the length of the wire link, the received message signal voltage will be greatly attenuated in comparison with its level at the transmitting end of the link. As a result, the message signal voltage may not be very large in comparison with the noise voltage, and the message will be perceived with difficulty or possibly not at all. An amplifier at the receiving end will not solve the problem, since the amplifier will amplify signal and noise alike. As a matter of fact, as we shall see, the amplifier itself may well be a source of additional noise.

A principal concern of communication theory and a matter which we discuss extensively in this book is precisely the study of methods to suppress, as far as possible, the effect of noise. We shall see that, for this purpose, it may be better not to transmit directly the original signal (the microphone output in our

example). Instead, the original signal is used to generate a different signal waveform, which new signal waveform is then impressed on the line. This processing of the original signal to generate the transmitted signal is called *encoding* or *modulation*. At the receiving end an inverse process called *decoding* or *demodulation* is required to recover the original signal.

It may well be that there is a considerable expense in providing the wire communication link. We are, therefore, naturally led to inquire whether we may use the link more effectively by arranging for the simultaneous transmission over the link of more than just a single waveform. It turns out that such multiple transmission is indeed possible and may be accomplished in a number of ways. Such simultaneous multiple transmission is called *multiplexing* and is again a principal area of concern of communication theory and of this book. It is to be noted that when wire communications links are employed, then, at least in principle, separate links may be used for individual messages. When, however, the communications medium is free space, as in radio communication from antenna to antenna, multiplexing is essential.

In summary then, *communication theory* addresses itself to the following questions: Given a communication channel, how do we arrange to transmit as many simultaneous signals as possible, and how do we devise to suppress the effect of noise to the maximum extent possible? In this book, after a few mathematical preliminaries, we shall address ourselves precisely to these questions, first to the matter of multiplexing, and thereafter to the discussion of noise in communications systems.

A branch of mathematics which is of inestimable value in the study of communications systems is *spectral analysis*. Spectral analysis concerns itself with the description of waveforms in the *frequency domain* and with the correspondence between the frequency-domain description and the time-domain description. It is assumed that the reader has some familiarity with spectral analysis. The presentation in this chapter is intended as a review, and will serve further to allow a compilation of results which we shall have occasion to use throughout the remainder of this text.

## 1.1 FOURIER SERIES[1]

A periodic function of time $v(t)$ having a fundamental period $T_0$ can be represented as an infinite sum of sinusoidal waveforms. This summation, called a *Fourier series*, may be written in several forms. One such form is the following:

$$v(t) = A_0 + \sum_{n=1}^{\infty} A_n \cos \frac{2\pi n t}{T_0} + \sum_{n=1}^{\infty} B_n \sin \frac{2\pi n t}{T_0} \tag{1.1-1}$$

The constant $A_0$ is the average value of $v(t)$ given by

$$A_0 = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} v(t)\, dt \tag{1.1-2}$$



while the coefficients $A_n$ and $B_n$ are given by

$$A_n = \frac{2}{T_0} \int_{-T_0/2}^{T_0/2} v(t) \cos \frac{2\pi n t}{T_0}\, dt \tag{1.1-3}$$

and

$$B_n = \frac{2}{T_0} \int_{-T_0/2}^{T_0/2} v(t) \sin \frac{2\pi n t}{T_0}\, dt \tag{1.1-4}$$

An alternative form for the Fourier series is

$$v(t) = C_0 + \sum_{n=1}^{\infty} C_n \cos \left( \frac{2\pi n t}{T_0} - \phi_n \right) \tag{1.1-5}$$

where $C_0$, $C_n$, and $\phi_n$ are related to $A_0$, $A_n$, and $B_n$ by the equations

$$C_0 = A_0 \tag{1.1-6a}$$

$$C_n = \sqrt{A_n^2 + B_n^2} \tag{1.1-6b}$$

and

$$\phi_n = \tan^{-1} \frac{B_n}{A_n} \tag{1.1-6c}$$

The Fourier series of a periodic function is thus seen to consist of a summation of harmonics of a fundamental frequency $f_0 = 1/T_0$. The coefficients $C_n$ are called *spectral amplitudes*; that is, $C_n$ is the amplitude of the *spectral component* $C_n \cos (2\pi n f_0 t - \phi_n)$ at frequency $nf_0$. A typical *amplitude spectrum* of a periodic waveform is shown in Fig. 1.1-1*a*. Here, at each harmonic frequency, a vertical line has been drawn having a length equal to the spectral amplitude associated with each harmonic frequency. Of course, such an amplitude spectrum, lacking the phase information, does not specify the waveform $v(t)$.

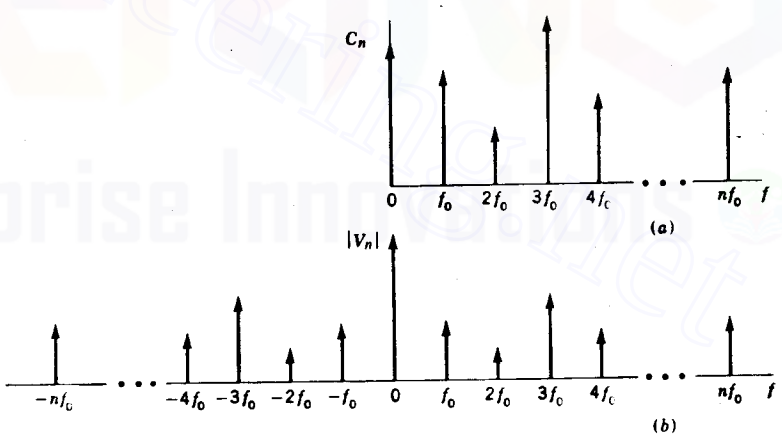



**Figure 1.1-1** (*a*) A one-sided plot of spectral amplitude of a periodic waveform. (*b*) The corresponding two-sided plot.

## 1.2 EXPONENTIAL FORM OF THE FOURIER SERIES

The exponential form of the Fourier series finds extensive application in communication theory. This form is given by

$$v(t) = \sum_{n=-\infty}^{\infty} V_n e^{j2\pi nt/T_0} \tag{1.2-1}$$

where $V_n$ is given by

$$V_n = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} v(t)e^{-j2\pi nt/T_0}\,dt \tag{1.2-2}$$

The coefficients $V_n$ have the property that $V_n$ and $V_{-n}$ are complex conjugates of one another, that is, $V_n = V^*_{-n}$. These coefficients are related to the $C_n$'s in Eq. (1.1-5) by

$$V_0 = C_0 \tag{1.2-3a}$$

$$V_n = \frac{C_n}{2}e^{-j\Phi_n} \tag{1.2-3b}$$

The $V_n$'s are the *spectral amplitudes* of the *spectral components* $V_n e^{j2\pi nf_0 t}$. The amplitude spectrum of the $V_n$'s shown in Fig. 1.1-1*b* corresponds to the amplitude spectrum of the $C_n$'s shown in Fig. 1.1-1*a*. Observe that while $V_0 = C_0$, otherwise each spectral line in 1.1-1*a* at frequency $f$ is replaced by the 2 spectral lines in 1.1-1*b*, each of half amplitude, one at frequency $f$ and one at frequency $-f$. The amplitude spectrum in 1.1-1*a* is called a *single-sided* spectrum, while the spectrum in 1.1-1*b* is called a *two-sided* spectrum. We shall find it more convenient to use the two-sided amplitude spectrum and shall consistently do so from this point on.

## 1.3 EXAMPLES OF FOURIER SERIES

A waveform in which we shall have occasion to have some special interest is shown in Fig. 1.3-1*a*. The waveform consists of a periodic sequence of impulses of strength $I$. As a matter of convenience we have selected the time scale so that an impulse occurs at $t = 0$. The impulse at $t = 0$ is written as $I\,\delta(t)$. Here $\delta(t)$ is the delta function which has the property that $\delta(t) = 0$ except when $t = 0$ and further

$$\int_{-\infty}^{\infty} \delta(t)\,dt = 1 \tag{1.3-1}$$

The strength of an impulse is equal to the area under the impulse. Thus the strength of $\delta(t)$ is 1, and the strength of $I\,\delta(t)$ is $I$.

The periodic impulse train is written

$$v(t) = I\sum_{k=-\infty}^{\infty} \delta(t - kT_0) \tag{1.3-2}$$



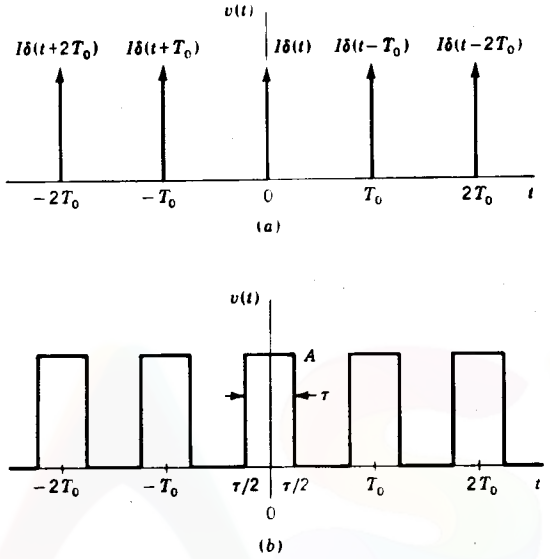



**Figure 1.3-1** Examples of periodic functions. (*a*) A periodic train of impulses. (*b*) A periodic train of pulses of duration $\tau$.

We then find, using Eqs. (1.1-2) and (1.1-3),

$$A_0 = \frac{I}{T_0}\int_{-T_0/2}^{T_0/2} \delta(t)\,dt = \frac{I}{T_0} \tag{1.3-3}$$

$$A_n = \frac{2I}{T_0}\int_{-T_0/2}^{T_0/2} \delta(t)\cos\frac{2\pi nt}{T_0}\,dt = \frac{2I}{T_0} \tag{1.3-4}$$

and, using Eq. (1.1-4),

$$B_n = \frac{2I}{T_0}\int_{-T_0/2}^{T_0/2} \delta(t)\sin\frac{2\pi nt}{T_0}\,dt = 0 \tag{1.3-5}$$

Further we have, using Eq. (1.1-6),

$$C_0 = \frac{I}{T_0} \qquad C_n = \frac{2I}{T_0} \qquad \phi_n = 0 \tag{1.3-6}$$

and from Eq. (1.2-3)

$$V_0 = V_n = \frac{I}{T_0} \tag{1.3-7}$$

Hence $v(t)$ may be written in the forms

$$v(t) = I \sum_{k=-\infty}^{\infty} \delta(t - kT_0) = \frac{I}{T_0} + \frac{2I}{T_0} \sum_{n=1}^{\infty} \cos \frac{2\pi nt}{T_0}$$

$$= \frac{I}{T_0} \sum_{n=-\infty}^{\infty} e^{j2\pi nt/T_0} \tag{1.3-8}$$

As a second example, let us find the Fourier series for the periodic train of pulses of amplitude $A$ and duration $\tau$ as shown in Fig. 1.3-1*b*. We find

$$A_0 = C_0 = V_0 = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} v(t)\, dt = \frac{A\tau}{T_0} \tag{1.3-9}$$

$$A_n = C_n = 2V_n = \frac{2}{T_0} \int_{-T_0/2}^{T_0/2} v(t) \cos \frac{2\pi nt}{T_0}\, dt$$

$$= \frac{2A\tau}{T_0} \frac{\sin (n\pi\tau/T_0)}{n\pi\tau/T_0} \tag{1.3-10}$$

and

$$B_n = 0 \qquad \phi_n = 0 \tag{1.3-11}$$

Thus

$$v(t) = \frac{A\tau}{T_0} + \frac{2A\tau}{T_0} \sum_{n=1}^{\infty} \frac{\sin (n\pi\tau/T_0)}{n\pi\tau/T_0} \cos \frac{2\pi nt}{T_0} \tag{1.3-12a}$$

$$= \frac{A\tau}{T_0} \sum_{n=-\infty}^{\infty} \frac{\sin (n\pi\tau/T_0)}{n\pi\tau/T_0} e^{j2\pi nt/T_0} \tag{1.3-12b}$$

Suppose that in the waveform of Fig. 1.3-1*b* we reduce $\tau$ while adjusting $A$ so that $A\tau$ is a constant, say $A\tau = I$. We would expect that in the limit, as $\tau \to 0$, the Fourier series for the pulse train in Eq. (1.3-12) should reduce to the series for the impulse train in Eq. (1.3-8). It is readily verified that such is indeed the case since as $\tau \to 0$

$$\frac{\sin (n\pi\tau/T_0)}{n\pi\tau/T_0} \to 1 \tag{1.3-13}$$

## 1.4 THE SAMPLING FUNCTION

A function frequently encountered in spectral analysis is the sampling function $Sa(x)$ defined by

$$Sa(x) \equiv \frac{\sin x}{x} \tag{1.4-1}$$

[A closely related function is sinc $x$ defined by sinc $x = (\sin \pi x)/\pi x$.] The function $Sa(x)$ is plotted in Fig. 1.4-1. It is symmetrical about $x = 0$, and at $x = 0$ has



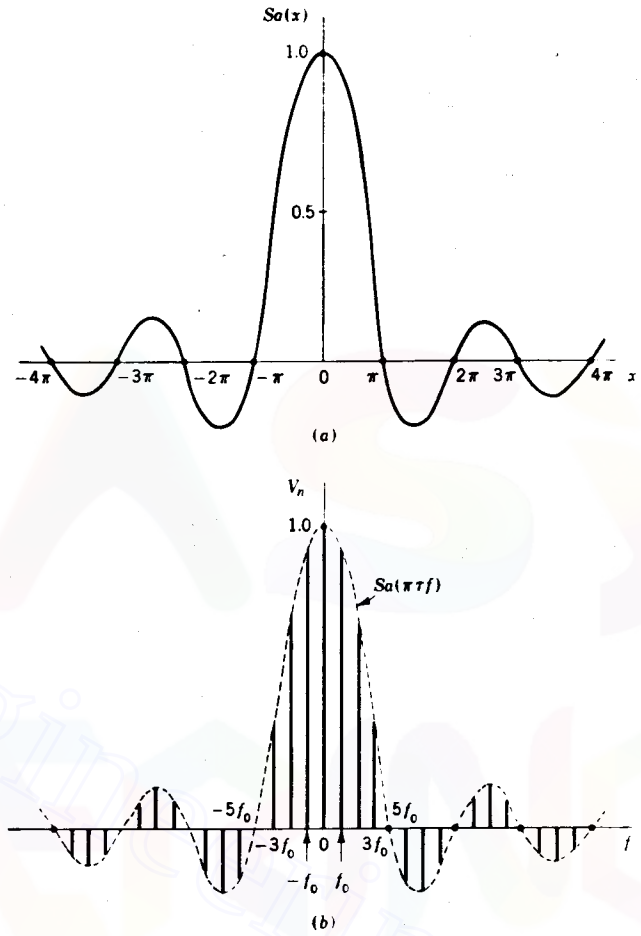



**Figure 1.4-1** (*a*) The function $Sa(x)$. (*b*) The spectral amplitudes $V_n$ of the two-sided Fourier representation of the pulse train of Fig. 1.3-1*b* for $A = 4$ and $\tau/T_0 = \frac{1}{4}$.

the value $Sa(0) = 1$. It oscillates with an amplitude that decreases with increasing $x$. The function passes through zero at equally spaced intervals at values of $x = \pm n\pi$, where $n$ is an integer other than zero. Aside from the peak at $x = 0$, the maxima and minima occur *approximately* midway between the zeros, i.e., at $x = \pm(n + \frac{1}{2})\pi$, where $|\sin x| = 1$. This approximation is poorest for the minima closest to $x = 0$ but improves as $x$ becomes larger. Correspondingly, the approximate value of $Sa(x)$ at these extremal points is

$$Sa[\pm(n + \tfrac{1}{2})\pi] = \frac{2(-1)^n}{(2n + 1)\pi} \tag{1.4-2}$$

We encountered the sampling function in the preceding section in Eq. (1.3-12) which expresses the spectrum of a periodic sequence of rectangular pulses. In that equation we have $x = n\pi\tau/T_0$. The spectral amplitudes of Eq. (1.3-12*b*) are plotted in Fig. 1.4-1*b* for the case $A = 4$ and $\tau/T_0 = \frac{1}{4}$. The spectral components appear at frequencies which are multiples of the fundamental frequency $f_0 = 1/T_0$, that is, at frequencies $f = nf_0 = n/T_0$. The envelope of the spectral components of $Sa(\pi\tau f)$ is also shown in the figure. Here we have replaced $x$ by

$$x = n\pi\tau/T_0 = \pi\tau nf_0 = \pi\tau f \tag{1.4-3}$$

## 1.5 RESPONSE OF A LINEAR SYSTEM

The Fourier trigonometric series is by no means the only way in which a periodic function may be expanded in terms of other functions.[2] As a matter of fact, the number of such possible alternative expansions is limitless. However, what makes the Fourier trigonometric expansion especially useful is the distinctive and unique characteristic of the sinusoidal waveform, this characteristic being that when a sinusoidal excitation is applied to a linear system, the response everyplace in the system is similarly sinusoidal and has the same frequency as the excitation. That is, the sinusoidal waveform preserves its waveshape. And since the waveshape is preserved, then, in order to characterize the relationship of the response to the excitation, we need but to specify how the response amplitude is related to the excitation amplitude and how the response phase is related to the excitation phase. Therefore with sinusoidal excitation, two numbers (amplitude ratio and phase difference) are all that are required to deduce a response. It turns out to be possible and very convenient to incorporate these two numbers into a single complex number.

Let the input to a linear system be the spectral component

$$v_i(t, \omega_n) = V_n e^{j2\pi nt/T_0} = V_n e^{j\omega_n t} \tag{1.5-1}$$

The waveform $v_i(t, \omega_n)$ may be, say, the voltage applied to the input of an electrical filter as in Fig. 1.5-1. Then the filter output $v_o(t, \omega_n)$ is related to the input by a complex transfer function

$$H(\omega_n) = |H(\omega_n)| e^{-j\theta(\omega_n)} \tag{1.5-2}$$

that is, the output is

$$\begin{aligned} v_o(t, \omega_n) = H(\omega_n)v_i(t, \omega_n) &= |H(\omega_n)| e^{-j\theta(\omega_n)} V_n e^{j\omega_n t} \\ &= |H(\omega_n)| V_n e^{j[\omega_n t - \theta(\omega_n)]} \end{aligned} \tag{1.5-3}$$

Actually, the spectral component in Eq. (1.5-1) is not a physical voltage. Rather, the physical input voltage $v_{ip}(t)$ which gives rise to this spectral component is the sum of this spectral component and its complex conjugate, that is,

$$v_{ip}(t, \omega_n) = V_n e^{j\omega_n t} + V_{-n} e^{-j\omega_n t} = V_n e^{j\omega_n t} + V_n^* e^{-j\omega_n t} = 2Re(V_n e^{j\omega_n t}) \tag{1.5-4}$$







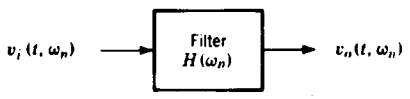



**Figure 1.5-1** A sinusoidal waveform $v_i(t, \omega_n)$ of angular frequency $\omega_n$ is applied at the input to a network (filter) whose transfer characteristic at $\omega_n$ is $H(\omega_n)$. The output $v_o(t, \omega_n)$ differs from the input only in amplitude and phase.

The corresponding *physical* output voltage is $v_{op}(t, \omega_n)$ given by

$$v_{op}(t, \omega_n) = H(\omega_n)V_n e^{j\omega_n t} + H(-\omega_n)V_n^* e^{-j\omega_n t} \tag{1.5-5}$$

Since $v_{op}(t, \omega_n)$ must be real, the two terms in Eq. (1.5-5) must be complex conjugates, and hence we must have that $H(\omega_n) = H^*(-\omega_n)$. Therefore, since $H(\omega_n) = |H(\omega_n)| e^{-j\theta(\omega_n)}$, we must have that

$$|H(\omega_n)| = |H(-\omega_n)| \tag{1.5-6}$$

and

$$\theta(\omega_n) = -\theta(-\omega_n) \tag{1.5-7}$$

that is, $|H(\omega_n)|$ must be an even function and $\theta(\omega_n)$ an odd function of $\omega_n$.

If, then, an excitation is expressed as a Fourier series in exponential form as in Eq. (1.2-1), the response is

$$v_o(t) = \sum_{n=-\infty}^{\infty} H(\omega_n)V_n e^{j2\pi nt/T_0} \tag{1.5-8}$$

If the form of Eq. (1.1-5) is used, the response is

$$v_o(t) = H(0)C_0 + \sum_{n=1}^{\infty} |H(\omega_n)| C_n \cos\left[\frac{2\pi nt}{T_0} - \phi_n - \theta(\omega_n)\right] \tag{1.5-9}$$

Given a periodic waveform, the coefficients in the Fourier series may be evaluated. Thereafter, if the transfer function $H(\omega_n)$ of a system is known, the response may be written out formally as in, say, Eq. (1.5-8) or Eq. (1.5-9). Actually these equations are generally of small value from a *computational* point of view. For, except in rather special and infrequent cases, we should be hard-pressed indeed to recognize the waveform of a response which is expressed as the sum of an infinite (or even a large) number of sinusoidal terms. On the other hand, the *concept* that a response may be written in the form of a linear superposition of responses to individual spectral components, as, say, in Eq. (1.5-8), is of inestimable value.

## 1.6 NORMALIZED POWER

In the analysis of communication systems, we shall often find that, given a waveform $v(t)$, we shall be interested in the quantity $\overline{v^2(t)}$ where the bar indicates the time-average value. In the case of periodic waveforms, the time averaging is done over one cycle. If, in our mind's eye, we were to imagine that the waveform $v(t)$ appear across a 1-ohm resistor, then the power dissipated in that resistor would

be $\overline{v^2(t)}$ volts$^2$/1 ohm = $W$ watts, where the number $W$ would be numerically equal to the numerical value of $\overline{v^2(t)}$, the mean-square value of $v(t)$. For this reason $\overline{v^2(t)}$ is generally referred to as the *normalized power* of $v(t)$. It is to be kept in mind, however, that the dimension of normalized power is volts$^2$ and not watts. When, however, no confusion results from so doing, we shall often follow the generally accepted practice of dropping the word "normalized" and refer instead simply to "power." We shall often have occasion also to calculate *ratios* of normalized powers. In such cases, even if the dimension "watts" is applied to normalized power, no harm will have been done, for the dimensional error in the numerator and the denominator of the ratio will cancel out.

Suppose that in some system we encounter at one point or another normalized powers $S_1$ and $S_2$. If the ratio of these powers is of interest, we need but to evaluate, say, $S_2/S_1$. It frequently turns out to be more convenient not to specify this ratio directly but instead to specify the quantity $K$ defined by

$$K \equiv 10 \log \frac{S_2}{S_1} \tag{1.6-1}$$

Like the ratio $S_2/S_1$, the quantity $K$ is dimensionless. However, in order that one may know whether, in specifying a ratio we are stating the number $S_2/S_1$ or the number $K$, the term *decibel* (abbreviated dB) is attached to the number $K$. Thus, for example, suppose $S_2/S_1 = 100$, then $\log S_2/S_1 = 2$ and $K = 20$ dB. The advantages of the use of the decibel are twofold. First, a very large power ratio may be expressed in decibels by a much smaller and therefore often more convenient number. Second, if power ratios are to be multiplied, such multiplication may be accomplished by the simpler arithmetic operation of addition if the ratios are first expressed in decibels. Suppose that $S_2$ and $S_1$ are, respectively, the normalized power associated with sinusoidal signals of amplitudes $V_2$ and $V_1$. Then $S_2 = V_2^2/2$, $S_1 = V_1^2/2$, and

$$K = 10 \log \frac{V_2^2/2}{V_1^2/2} = 20 \log \frac{V_2}{V_1} \tag{1.6-2}$$

The use of the decibel was introduced in the early days of communications systems in connection with the transmission of signals over telephone lines. (The "bel" in decibel comes from the name Alexander Graham Bell.) In those days the decibel was used for the purpose of specifying ratios of *real* powers, not normalized powers. Because of this early history, occasionally some confusion occurs in the meaning of a ratio expressed in decibels. To point out the source of this confusion and, we hope, thereby to avoid it, let us consider the situation represented in Fig. 1.6-1. Here a waveform $v_i(t) = V_i \cos \omega t$ is applied to the input of a linear amplifier of input impedance $R_i$. An output signal $v_o(t) = V_o \cos(\omega t + \theta)$ then appears across the load resistor $R_o$. A real power $P_i = V_i^2/2R_i$ is supplied to the input, and the real power delivered to the load is $P_o = V_o^2/2R_o$. The real power gain $P_o/P_i$ of the amplifier expressed in decibels is

$$K_{\text{real}} = 10 \log \frac{V_o^2/2R_o}{V_i^2/2R_i} \tag{1.6-3}$$





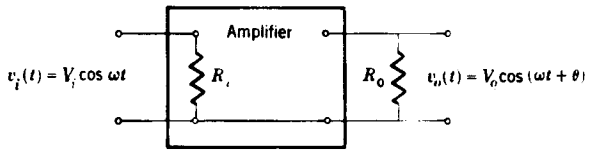



**Figure 1.6-1** An amplifier of input impedance $R_i$ with load $R_o$.

If it should happen that $R_i = R_o$, then $K_{\text{real}}$ may be written as in Eq. (1.6-2).

$$K_{\text{real}} = 20 \log \frac{V_o}{V_i} \tag{1.6-4}$$

But if $R_i \neq R_o$, then Eq. (1.6-4) does not apply. On the other hand, if we calculate the normalized power gain, then we have

$$K_{\text{norm}} = 10 \log \frac{V_o^2/2}{V_i^2/2} = 20 \log \frac{V_o}{V_i} \tag{1.6-5}$$

So far as the normalized power gain is concerned, the impedances $R_i$ and $R_o$ in Fig. 1.6-1 are *absolutely irrelevant*. If it should happen that $R_i = R_o$, then $K_{\text{real}} = K_{\text{norm}}$, but otherwise they would be different.

## 1.7 NORMALIZED POWER IN A FOURIER EXPANSION

Let us consider two typical terms of the Fourier expansion of Eq. (1.1-5). If we take, say, the fundamental and first harmonic, we have

$$v'(t) = C_1 \cos\left(\frac{2\pi t}{T_0} - \phi_1\right) + C_2 \cos\left(\frac{4\pi t}{T_0} - \phi_2\right) \tag{1.7-1}$$

To calculate the normalized power $S'$ of $v'(t)$, we must square $v'(t)$ and evaluate

$$S' = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} [v'(t)]^2 \, dt \tag{1.7-2}$$

When we square $v'(t)$, we get the square of the first term, the square of the second term, and then the cross-product term. However, the two cosine functions in Eq. (1.7-1) are *orthogonal*. That is, when their product is integrated over a complete period, the result is zero. Hence in evaluating the normalized power, we find no term corresponding to this cross product. We find actually that $S'$ is given by

$$S' = \frac{C_1^2}{2} + \frac{C_2^2}{2} \tag{1.7-3}$$

By extension it is apparent that the normalized power associated with the entire Fourier series is

$$S = C_0^2 + \sum_{n=1}^{\infty} \frac{C_n^2}{2} \tag{1.7-4}$$

Hence we observe that because of the orthogonality of the sinusoids used in a Fourier expansion, the total normalized power is the sum of the normalized power due to each term in the series separately. If we write a waveform as a sum of terms which are not orthogonal, this very simple and useful result will not apply. We may note here also that, in terms of the $A$'s and $B$'s of the Fourier representation of Eq. (1.1-1), the normalized power is

$$S = A_0^2 + \sum_{n=1}^{\infty} \frac{A_n^2}{2} + \sum_{n=1}^{\infty} \frac{B_n^2}{2} \tag{1.7-5}$$

It is to be observed that power and normalized power are to be associated with a *real* waveform and not with a *complex* waveform. Thus suppose we have a term $A_n \cos(2\pi nt/T_0)$ in a Fourier series. Then the normalized power contributed by this term is $A_n^2/2$ quite independently of all other terms. And this normalized power comes from averaging, over time, the product of the term $A_n \cos(2\pi nt/T)$ by *itself*. On the other hand, in the complex Fourier representation of Eq. (1.2-1) we have terms of the form $V_n e^{j2\pi nt/T_0}$. The average value of the square of such a term is zero. We find as a matter of fact that the contributions to normalized power come from product terms

$$V_n e^{j2\pi nt/T_0} V_{-n} e^{-j2\pi nt/T_0} = V_n V_{-n} = V_n V_n^* \tag{1.7-6}$$

The total normalized power is

$$S = \sum_{n=-\infty}^{n=+\infty} V_n V_n^* \tag{1.7-7}$$

Thus, in the complex representation, the power associated with a particular *real* frequency $n/T_0 = nf_0$ ($f_0$ is the fundamental frequency) is associated neither with the spectral component at $nf_0$ nor with the component at $-nf_0$, but rather with the *combination* of spectral components, one in the positive-frequency range and one in the negative-frequency range. This power is

$$V_n V_n^* + V_{-n} V_{-n}^* = 2V_n V_n^* \tag{1.7-8}$$

It is nonetheless a procedure of great convenience to associate one-half of the power in this combination of spectral components (that is, $V_n V_n^*$) with the frequency $nf_0$ and the other half with the frequency $-nf_0$. Such a procedure will always be valid provided that we are careful to use the procedure to calculate only the *total* power associated with frequencies $nf_0$ and $-nf_0$. Thus we may say that the power associated with the spectral component at $nf_0$ is $V_n V_n^*$ and the power associated with the spectral component at $-nf_0$ is similarly $V_n V_n^*$ ($= V_{-n} V_{-n}^*$). If we use these associations only to arrive at the result that the total power is $2V_n V_n^*$, we shall make no error.

In correspondence with the one-sided and two-sided spectral amplitude pattern of Fig. 1.1-1 we may construct one-sided and two-sided spectral (normalized) power diagrams. A two-sided power spectral diagram is shown in Fig. 1.7-1. The vertical axis is labeled $S_n$, the power associated with each spectral



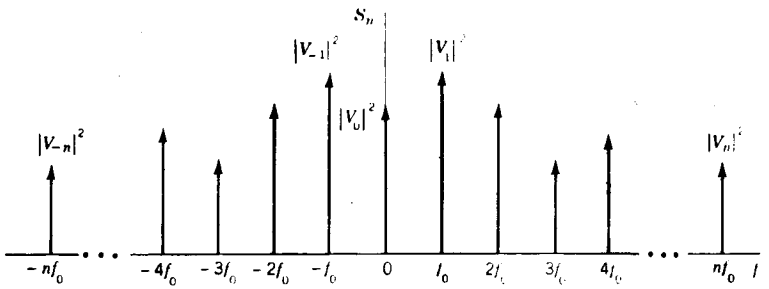



**Figure 1.7-1** A two-sided power spectrum.

component. The height of each vertical line is $|V_n|^2$. Because of its greater convenience and because it lends a measure of systemization to very many calculations, we shall use the two-sided amplitude and power spectral pattern exclusively throughout this text.

We shall similarly use a two-sided representation to specify the transmission characteristics of filters. Thus, suppose we have a low-pass filter which transmits without attenuation all spectral components up to a frequency $f_M$ and transmits nothing at a higher frequency. Then the magnitude of the transfer function will be given as in Fig. 1.7-2. The transfer characteristic of a bandpass filter will be given as in Fig. 1.7-3.

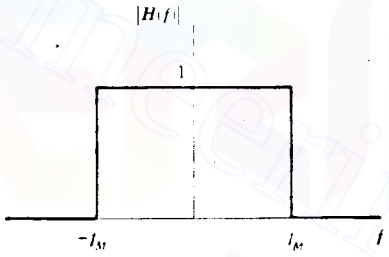



**Figure 1.7-2** The transfer characteristic of an idealized low-pass filter.

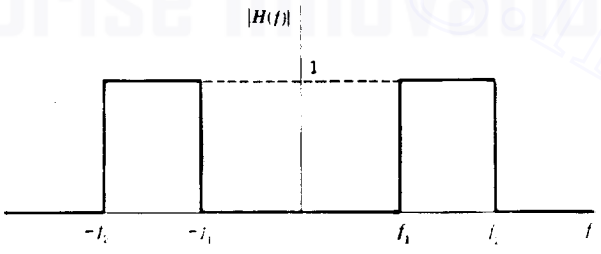



**Figure 1.7-3** The transfer characteristic of an idealized bandpass filter with passband from $f_1$ to $f_2$.

## 1.8 POWER SPECTRAL DENSITY

Suppose that, in Fig. 1.7-1 where $S_n$ is given for each spectral component, we start at $f = -\infty$ and then, moving in the positive-frequency direction, we add the normalized powers contributed by each power spectral line up to the frequency $f$. This sum is $S(f)$, a function of frequency. $S(f)$ typically will have the appearance shown in Fig. 1.8-1. It does not change as $f$ goes from one spectral line to another, but jumps abruptly as the normalized power of each spectral line is added. Now let us inquire about the normalized power at the frequency $f$ in a range $df$. This quantity of normalized power $dS(f)$ would be written

$$dS(f) = \frac{dS(f)}{df}\,df \tag{1.8-1}$$

The quantity $dS(f)/df$ is called the (normalized) *power spectral density* $G(f)$; thus

$$G(f) \equiv \frac{dS(f)}{df} \tag{1.8-2}$$

The power in the range $df$ at $f$ is $G(f)\,df$. The power in the positive-frequency range $f_1$ to $f_2$ is

$$S(f_1 \le f \le f_2) = \int_{f_1}^{f_2} G(f)\,df \tag{1.8-3}$$

The power in the negative-frequency range $-f_2$ to $-f_1$ is

$$S(-f_2 \le f \le -f_1) = \int_{-f_2}^{-f_1} G(f)\,df \tag{1.8-4}$$

The quantities in Eqs. (1.8-3) and (1.8-4) do not have physical significance. However, the total power in the *real* frequency range $f_1$ to $f_2$ does have physical significance, and this power $S(f_1 \le |f| \le f_2)$ is given by

$$S(f_1 \le |f| \le f_2) = \int_{-f_2}^{-f_1} G(f)\,df + \int_{f_1}^{f_2} G(f)\,df \tag{1.8-5}$$

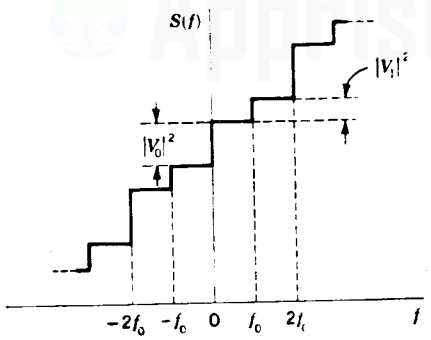



**Figure 1.8-1** The sum $S(f)$ of the normalized power in all spectral components from $f = -\infty$ to $f$.



To find the power spectral density, we must differentiate $S(f)$ in Fig. 1.8-1. Between harmonic frequencies we would have $G(f) = 0$. At a harmonic frequency, $G(f)$ would yield an impulse of strength equal to the size of the jump in $S(f)$. Thus we would find

$$G(f) = \sum_{n=-\infty}^{\infty} |V_n|^2\,\delta(f - nf_0) \tag{1.8-6}$$

If, in plotting $G(f)$, we were to represent an impulse by a vertical arrow of height proportional to the impulse strength, then a plot of $G(f)$ versus $f$ as given by Eq. (1.8-6) would have exactly the same appearance as the plot of $S_n$ shown in Fig. 1.7-1.

## 1.9 EFFECT OF TRANSFER FUNCTION ON POWER SPECTRAL DENSITY

Let the input signal $v_i(t)$ to a filter have a power spectral density $G_i(f)$. If $V_{in}$ are the spectral amplitudes of this input signal, then, using Eq. (1.8-6)

$$G_i(f) = \sum_{n=-\infty}^{\infty} |V_{in}|^2\,\delta(f - nf_0) \tag{1.9-1}$$

where, from Eq. (1.2-2),

$$V_{in} = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} v_i(t)e^{-j2\pi nt/T_0}\,dt \tag{1.9-2}$$

Let the output signal of the filter be $v_o(t)$. If $V_{on}$ are the spectral amplitudes of this output signal, then the corresponding power spectral density is

$$G_0(f) = \sum_{n=-\infty}^{\infty} |V_{on}|^2\,\delta(f - nf_0) \tag{1.9-3}$$

where

$$V_{on} = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} v_o(t)e^{-j2\pi nt/T_0}\,dt \tag{1.9-4}$$

As discussed in Sec. 1.5, if the transfer function of the filter is $H(f)$, then the output coefficient $V_{on}$ is related to the input coefficient by

$$V_{on} = V_{in}H(f = nf_0) \tag{1.9-5}$$

Hence,

$$|V_{on}|^2 = |V_{in}|^2\,|H(f = nf_0)|^2 \tag{1.9-6}$$

Substituting Eq. (1.9-6) into Eq. (1.9-3) and comparing the result with Eq. (1.9-1) yields the important result

$$G_0(f) = G_i(f)\,|H(f)|^2 \tag{1.9-7}$$

Equation (1.9-7) is of the greatest importance since it relates the power spectral density, and hence the power, at one point in a system to the power spectral density at another point in the system. Equation (1.9-7) was derived for the special case of periodic signals; however, it applies to nonperiodic signals and signals represented by random processes (Sec. 2.23) as well.

As a special application of interest of the result given in Eq. (1.9-7), assume that an input signal $v_i(t)$ with power spectral density $G_i(f)$ is passed through a differentiator. The differentiator output $v_o(t)$ is related to the input by

$$v_o(t) = \tau \frac{d}{dt} v_i(t) \tag{1.9-8}$$

where $\tau$ is a constant. The operation indicated in Eq. (1.9-8) multiplies each spectral component of $v_i(t)$ by $j2\pi f\tau = j\omega\tau$. Hence $H(f) = j\omega\tau$, and $|H(f)|^2 = \omega^2\tau^2$. Thus from Eq. (1.9-7) the spectral density of the output is

$$G_o(f) = \omega^2\tau^2 G_i(f) \tag{1.9-9}$$

## 1.10 THE FOURIER TRANSFORM

A periodic waveform may be expressed, as we have seen, as a sum of spectral components. These components have finite amplitudes and are separated by finite frequency intervals $f_0 = 1/T_0$. The normalized power of the waveform is finite, as is also the normalized energy of the signal in an interval $T_0$. Now suppose we increase without limit the period $T_0$ of the waveform. Thus, say, in Fig. 1.3-1*b* the pulse centered around $t = 0$ remains in place, but all other pulses move outward away from $t = 0$ as $T_0 \to \infty$. Then eventually we would be left with a single-pulse nonperiodic waveform.

As $T_0 \to \infty$, the spacing between spectral components becomes infinitesimal. The frequency of the spectral components, which in the Fourier series was a discontinuous variable with a one-to-one correspondence with the integers, becomes instead a continuous variable. The normalized energy of the nonperiodic waveform remains finite, but, since the waveform is not repeated, its normalized power becomes infinitesimal. The spectral amplitudes similarly become infinitesimal. The Fourier series for the periodic waveform

$$v(t) = \sum_{n=-\infty}^{\infty} V_n e^{j2\pi n f_0 t} \tag{1.10-1}$$

becomes (see Prob. 1.10-1)

$$v(t) = \int_{-\infty}^{\infty} V(f) e^{j2\pi f t}\, df \tag{1.10-2}$$

The finite spectral amplitudes $V_n$ are analogous to the infinitesimal spectral



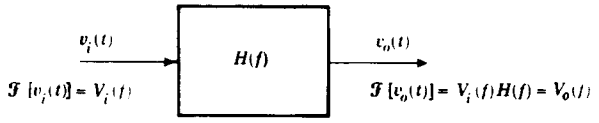



**Figure 1.10-1** A waveform $v_i(t)$ of transform $V_i(f)$ is transmitted through a network of transfer function $H(f)$. The output waveform $v_o(t)$ has a transform $V_o(f) = V_i(f)H(f)$.

amplitudes $V(f)\,df$. The quantity $V(f)$ is called the *amplitude spectral density* or more generally the *Fourier transform* of $v(t)$. The Fourier transform is given by

$$V(f) = \int_{-\infty}^{\infty} v(t) e^{-j2\pi f t}\, dt \tag{1.10-3}$$

in correspondence with $V_n$, which is given by

$$V_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} v(t) e^{-j2\pi n f_0 t}\, dt \tag{1.10-4}$$

Again, in correspondence with Eq. (1.5-8), let $H(f)$ be the transfer function of a network. If the input signal is $v_i(t)$, then the output signal will be $v_o(t)$ given by

$$v_o(t) = \int_{-\infty}^{\infty} H(f) V(f) e^{j2\pi f t}\, df \tag{1.10-5}$$

Comparing Eq. (1.10-5) with Eq. (1.10-2), we see that the Fourier transform $V_o(f) \equiv \mathcal{F}[v_o(t)]$ is related to the transform $V_i(f)$ of $v_i(t)$ by

$$\mathcal{F}[v_o(t)] = H(f)\mathcal{F}[v_i(t)] \tag{1.10-6}$$

or

$$V_o(f) = H(f)V_i(f) \tag{1.10-7}$$

as indicated in Fig. 1.10-1.

## 1.11 EXAMPLES OF FOURIER TRANSFORMS

In this section we shall evaluate the transforms of a number of functions both for the sake of review and also because we shall have occasion to refer to the results.

**Example 1.11-1** If $v(t) = \cos \omega_0 t$, find $V(f)$.

SOLUTION The function $v(t) = \cos \omega_0 t$ is periodic, and therefore has a Fourier series representation as well as a Fourier transform.

The exponential Fourier series representation of $v(t)$ is

$$v(t) = \tfrac{1}{2}e^{+j\omega_0 t} + \tfrac{1}{2}e^{-j\omega_0 t} \qquad \omega_0 = \frac{2\pi}{T_0} \tag{1.11-1}$$

Thus

$$V_1 = V_{-1} = \tfrac{1}{2} \tag{1.11-2}$$

and

$$V_n = 0 \qquad n \neq \pm 1 \tag{1.11-3}$$

The Fourier transform $V(f)$ is found using Eq. (1.10-3):

$$V(f) = \int_{-\infty}^{\infty} \cos \omega_0 t\, e^{-j2\pi f t}\, dt = \frac{1}{2}\int_{-\infty}^{\infty} e^{-j2\pi(f-f_0)t}\, dt + \frac{1}{2}\int_{-\infty}^{\infty} e^{-j2\pi(f+f_0)t}\, dt \tag{1.11-4a}$$

$$= \tfrac{1}{2}\delta(f-f_0) + \tfrac{1}{2}\delta(f+f_0) \tag{1.11-4b}$$

[See Eq. (1.11-22) below.]

From Eqs. (1.11-1) and (1.11-3) we draw the following conclusion: The Fourier transform of a sinusoidal signal (or other periodic signal) consists of impulses located at each harmonic frequency of the signal, i.e., at $f_n = n/T_0 = nf_0$. The strength of each impulse is equal to the amplitude of the Fourier coefficient of the exponential series.

**Example 1.11-2** A signal $m(t)$ is multiplied by a sinusoidal waveform of frequency $f_c$. The product signal is

$$v(t) = m(t) \cos 2\pi f_c t \tag{1.11-5}$$

If the Fourier transform of $m(t)$ is $M(f)$, that is,

$$M(f) = \int_{-\infty}^{\infty} m(t) e^{-j2\pi f t}\, dt \tag{1.11-6}$$

find the Fourier transform of $v(t)$.

SOLUTION Since

$$m(t) \cos 2\pi f_c t = \tfrac{1}{2} m(t) e^{j2\pi f_c t} + \tfrac{1}{2} m(t) e^{-j2\pi f_c t} \tag{1.11-7}$$

then the Fourier transform $V(f)$ is given by

$$V(f) = \frac{1}{2}\int_{-\infty}^{\infty} m(t) e^{-j2\pi(f+f_c)t}\, dt + \frac{1}{2}\int_{-\infty}^{\infty} m(t) e^{-j2\pi(f-f_c)t}\, dt \tag{1.11-8}$$

Comparing Eq. (1.11-8) with Eq. (1.11-6), we have the result that

$$V(f) = \tfrac{1}{2} M(f+f_c) + \tfrac{1}{2} M(f-f_c) \tag{1.11-9}$$

The relationship of the transform $M(f)$ of $m(t)$ to the transform $V(f)$ of $m(t) \cos 2\pi f_c t$ is illustrated in Fig. 1.11-1*a*. In Fig. 1.11-1*b* we see the spectral pattern of $M(f)$ replaced by two patterns of the same form. One is shifted to the right and one to the left, each by amount $f_c$. Further, the amplitudes of each of these two spectral patterns is one-half the amplitude of the spectral pattern $M(f)$.



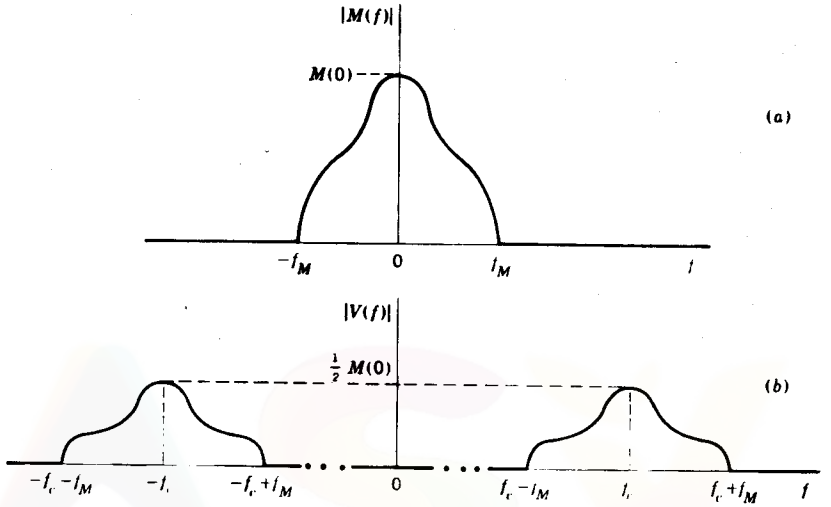



**Figure 1.11-1** (*a*) The amplitude spectrum of a waveform with no spectral component beyond $f_M$. (*b*) The amplitude spectrum of the waveform in (*a*) multiplied by $\cos 2\pi f_c t$.

A case of special interest arises when the waveform $m(t)$ is itself sinusoidal. Thus assume

$$m(t) = m \cos 2\pi f_m t \tag{1.11-10}$$

where $m$ is a constant. We then find that $V(f)$ is given by

$$V(f) = \frac{m}{4}\delta(f+f_c+f_m) + \frac{m}{4}\delta(f+f_c-f_m) + \frac{m}{4}\delta(f-f_c+f_m) + \frac{m}{4}\delta(f-f_c-f_m) \tag{1.11-11}$$

This spectral pattern is shown in Fig. 1.11-2. Observe that the pattern has four spectral lines corresponding to two real frequencies $f_c + f_m$ and $f_c - f_m$.

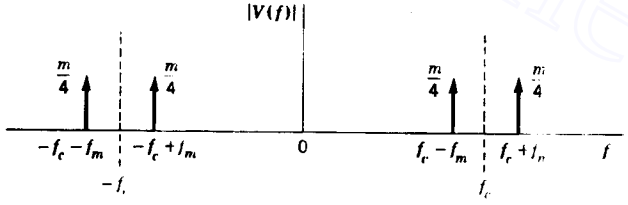



**Figure 1.11-2** The two-sided amplitude spectrum of the product waveform $v(t) = m \cos 2\pi f_c t$.

The waveform itself is given by

$$v(t) = \frac{m}{4}\left[e^{j2\pi(f_c+f_m)t} + e^{-j2\pi(f_c+f_m)t}\right]$$

$$+ \frac{m}{4}\left[e^{j2\pi(f_c-f_m)t} + e^{-j2\pi(f_c-f_m)t}\right] \qquad (1.11\text{-}12a)$$

$$= \frac{m}{4}\left[\cos 2\pi(f_c+f_m)t + \cos 2\pi(f_c-f_m)t\right] \qquad (1.11\text{-}12b)$$

**Example 1.11-3** A pulse of amplitude $A$ extends from $t = -\tau/2$ to $t = +\tau/2$. Find its Fourier transform $V(f)$. Consider also the Fourier series for a periodic sequence of such pulses separated by intervals $T_0$. Compare the Fourier series coefficients $V_n$ with the transform in the limit as $T_0 \to \infty$.

SOLUTION We have directly that

$$V(f) = \int_{-\tau/2}^{\tau/2} Ae^{-j2\pi ft}\,dt = A\tau\,\frac{\sin \pi f\tau}{\pi f\tau} \qquad (1.11\text{-}13)$$

The Fourier series coefficients of the periodic pulse train are given by Eq. (1.3-12*b*) as

$$V_n = \frac{A\tau}{T_0}\,\frac{\sin (n\pi\tau/T_0)}{n\pi\tau/T_0} \qquad (1.11\text{-}14)$$

The fundamental frequency in the Fourier series is $f_0 = 1/T_0$. We shall set $f_0 \equiv \Delta f$ in order to emphasize that $f_0 \equiv \Delta f$ is the frequency interval between spectral lines in the Fourier series. Hence, since $1/T_0 = \Delta f$, we may rewrite Eq. (1.11-14) as

$$V_n = A\tau\,\frac{\sin (\pi n\,\Delta f\tau)}{\pi n\,\Delta f\tau}\,\Delta f \qquad (1.11\text{-}15)$$

In the limit, as $T_0 \to \infty$, $\Delta f \to 0$. We then may replace $\Delta f$ by $df$ and replace $n\,\Delta f$ by a continuous variable $f$. Equation (1.11-15) then becomes

$$\lim_{\Delta f \to 0} V_n = A\tau\,\frac{\sin \pi f t}{\pi f t}\,df \qquad (1.11\text{-}16)$$

Comparing this result with Eq. (1.11-13), we do indeed note that

$$V(f) = \lim_{\Delta f \to 0} \frac{V_n}{\Delta f} \qquad (1.11\text{-}17)$$

Thus we confirm our earlier interpretation of $V(f)$ as an *amplitude spectral density*.



**Example 1.11-4**

(*a*) Find the Fourier transform of $\delta(t)$, an impulse of unit strength.

(*b*) Given a network whose transfer function is $H(f)$. An impulse $\delta(t)$ is applied at the input. Show that the response $v_o(t) \equiv h(t)$ at the output is the inverse transform of $H(f)$, that is, show that $h(t) = \mathscr{F}^{-1}[H(f)]$.

SOLUTION

(*a*) The impulse $\delta(t) = 0$ except at $t = 0$ and, further, has the property that

$$\int_{-\infty}^{\infty} \delta(t)\,dt = 1 \qquad (1.11\text{-}18)$$

Hence

$$V(f) = \int_{-\infty}^{\infty} \delta(t)e^{-j2\pi ft}\,dt = 1 \qquad (1.11\text{-}19)$$

Thus the spectral components of $\delta(t)$ extend with uniform amplitude and phase over the entire frequency domain.

(*b*) Using the result given in Eq. (1.10-7), we find that the transform of the output $v_o(t) \equiv h(t)$ is $V_o(f)$ given by

$$V_o(f) = 1 \times H(f) \qquad (1.11\text{-}20)$$

since the transform of $\delta(t)$, $\mathscr{F}[\delta(t)] = 1$. Hence the inverse transform of $V_o(f)$, which is the function $h(t)$, is also the inverse transform of $H(f)$. Specifically, for an impulse input, the output is

$$h(t) = \int_{-\infty}^{\infty} H(f)e^{j2\pi ft}\,df \qquad (1.11\text{-}21)$$

We may use the result given in Eq. (1.11-21) to arrive at a useful representation of $\delta(t)$ itself. If $H(f) = 1$, then the response $h(t)$ to an inpulse $\delta(t)$ is the impulse itself. Hence, setting $H(f) = 1$ in Eq. (1.11-21), we find

$$\delta(t) = \int_{-\infty}^{\infty} e^{j2\pi ft}\,df = \int_{-\infty}^{\infty} e^{-j2\pi ft}\,df \qquad (1.11\text{-}22)$$

## 1.12 CONVOLUTION

Suppose that $v_1(t)$ has the Fourier transform $V_1(f)$ and $v_2(t)$ has the transform $V_2(f)$. What then is the waveform $v(t)$ whose transform is the product $V_1(f)V_2(f)$? This question arises frequently in spectral analysis and is answered by the *convolution theorem*, which says that

$$v(t) = \int_{-\infty}^{\infty} v_1(\tau)v_2(t-\tau)\,d\tau \qquad (1.12\text{-}1)$$

or equivalently

$$v(t) = \int_{-\infty}^{\infty} v_2(\tau)v_1(t-\tau)\,d\tau \qquad (1.12\text{-}2)$$

The integrals in Eq. (1.12-1) or (1.12-2) are called *convolution integrals*, and the process of evaluating $v(t)$ through these integrals is referred to as *taking the convolution* of the functions $v_1(t)$ and $v_2(t)$.

To prove the theorem, we begin by writing

$$v(t) = \mathcal{F}^{-1}[V_1(f)V_2(f)] \tag{1.12-3a}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} V_1(f)V_2(f)e^{j\omega t}\,d\omega \tag{1.12-3b}$$

By definition we have

$$V_1(f) = \int_{-\infty}^{\infty} v_1(\tau)e^{-j\omega\tau}\,d\tau \tag{1.12-4}$$

Substituting $V_1(f)$ as given by Eq. (1.12-4) into the integrand of Eq. (1.12-3*b*), we have

$$v(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_1(\tau)e^{-j\omega\tau}\,d\tau V_2(f)e^{j\omega t}\,d\omega \tag{1.12-5}$$

Interchanging the order of integration, we find

$$v(t) = \int_{-\infty}^{\infty} v_1(\tau)\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} V_2(f)e^{j\omega(t-\tau)}\,d\omega\right]d\tau \tag{1.12-6}$$

We recognize that the expression in brackets in Eq. (1.12-6) is $v_2(t-\tau)$, so that finally

$$v(t) = \int_{-\infty}^{\infty} v_1(\tau)v_2(t-\tau)\,d\tau \tag{1.12-7}$$

Examples of the use of the convolution integral are given in Probs. 1.12-2 and 1.12-3. These problems will also serve to recall to the reader the relevance of the term *convolution*.

A special case of the convolution theorem and one of very great utility is arrived at through the following considerations. Suppose a waveform $v_i(t)$ whose transform is $V_i(f)$ is applied to a linear network with transfer function $H(f)$. The transform of the output waveform is $V_i(f)H(f)$. What then is the waveform of $v_o(t)$?

In Eq. (1.12-7) we identify $v_1(\tau)$ with $v_i(\tau)$ and $v_2(t)$ with the inverse transform of $H(f)$. But we have seen [in Eq. (1.11-21)] that the inverse transform of $H(f)$ is $h(t)$, the impulse response of the network. Hence Eq. (1.12-7) becomes

$$v_o(t) = \int_{-\infty}^{\infty} v_i(\tau)h(t-\tau)\,d\tau \tag{1.12-8}$$

in which the output $v_o(t)$ is expressed in terms of the input $v_i(t)$ and the inpulse response of the network.



## 1.13 PARSEVAL'S THEOREM

We saw that for periodic waveforms we may express the normalized power as a summation of powers due to individual spectral components. We also found for periodic signals that it was appropriate to introduce the concept of *power spectral density*. Let us keep in mind that we may make the transition from the periodic to the nonperiodic waveform by allowing the period of the periodic waveform to approach infinity. A nonperiodic waveform, so generated, has a finite *normalized energy*, while the normalized power approaches zero. We may therefore expect that for a nonperiodic waveform the energy may be written as a continuous summation (integral) of energies due to individual spectral components in a continuous distribution. Similarly we should expect that with such nonperiodic waveforms it should be possible to introduce an *energy spectral density*.

The normalized energy of a periodic waveform $v(t)$ in a period $T_0$ is

$$E = \int_{-T_0/2}^{T_0/2} [v(t)]^2\,dt \tag{1.13-1}$$

From Eq. (1.7-7) we may write $E$ as

$$E = T_0 S = T_0 \sum_{n=-\infty}^{n=+\infty} V_n V_n^* \tag{1.13-2}$$

Again, as in the illustrative Example 1.11-3, we let $\Delta f \equiv 1/T_0 = f_0$, where $f_0$ is the fundamental frequency, that is, $\Delta f$ is the spacing between harmonics. Then we have

$$V_n V_n^* = \frac{V_n}{\Delta f}\frac{V_n^*}{\Delta f}(\Delta f)^2 \tag{1.13-3}$$

and Eq. (1.13-2) may be written

$$E = \sum_{n=-\infty}^{\infty} \frac{V_n}{\Delta f}\frac{V_n^*}{\Delta f}\Delta f \tag{1.13-4}$$

In the limit, as $\Delta f \to 0$, we replace $\Delta f$ by $df$, we replace $V_n/\Delta f$ by the transform $V(f)$ [see Eq. (1.11-17)], and the summation by an integral. Equation (1.13-4) then becomes

$$E = \int_{-\infty}^{\infty} V(f)V^*(f)\,df = \int_{-\infty}^{\infty} |V(f)|^2\,df = \int_{-\infty}^{\infty} [v(t)]^2\,dt \tag{1.13-5}$$

This equation expresses *Parseval's theorem*. Parseval's theorem is the extension to the nonperiodic case of Eq. (1.7-7) which applies for periodic waveforms. Both results simply express the fact that the power (periodic case) or the energy (nonperiodic case) may be written as the superposition of power or energy due to individual spectral components separately. The validity of these results depends on the fact that the spectral components are orthogonal. In the periodic case the spectral components are orthogonal over the interval $T_0$. In the nonperiodic case

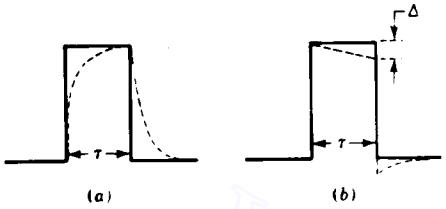



**Figure 1.15-2** (*a*) A rectangular pulse (solid) and the response (dashed) of a low-pass *RC* circuit. The rise time $t_r \approx 0.35\tau$. (*b*) The response of a high-pass *RC* circuit for $f_1\tau = 0.02$.

The essential distortion introduced by the frequency discrimination of this network is that the output does not sustain a constant voltage level when the input is constant. Instead, the output begins immediately to decay toward zero, which is its asymptotic limit. From Eq. (1.15-7) we have

$$\frac{1}{v_o}\frac{dv_o}{dt} = -\frac{1}{RC} = -2\pi f_1 \tag{1.15-8}$$

Thus the low-frequency cutoff $f_1$ alone determines the percentage drop in voltage per unit time. Again the importance of Eq. (1.15-8) is that, at least as a reasonable approximation, it applies to high-pass networks quite generally, even when the network is very much more complicated than the simple *RC* circuit.

If the input to the *RC* high-pass circuit is a pulse of duration $\tau$, then the output has the waveshape shown in Fig. 1.15-2*b*. The output exhibits a tilt and an undershoot. As a rule of thumb, we may assume that the pulse is reasonably faithfully reproduced if the tilt $\Delta$ is no more than $0.1V_0$. Correspondingly, this condition requires that $f_1$ be no higher than given by the condition

$$f_1\tau \approx 0.02 \tag{1.15-9}$$

## 1.16 CORRELATION BETWEEN WAVEFORMS

The correlation between waveforms is a measure of the similarity or relatedness between the waveforms. Suppose that we have waveforms $v_1(t)$ and $v_2(t)$, not necessarily periodic nor confined to a finite time interval. Then the correlation between them, or more precisely the *average cross correlation* between $v_1(t)$ and $v_2(t)$, is $R_{12}(\tau)$ defined as

$$R_{12}(\tau) \equiv \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} v_1(t)v_2(t+\tau)\,dt \tag{1.16-1}$$

If $v_1(t)$ and $v_2(t)$ are periodic with the same fundamental period $T_0$, then the average cross correlation is

$$R_{12}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} v_1(t)v_2(t+\tau)\,dt \tag{1.16-2}$$







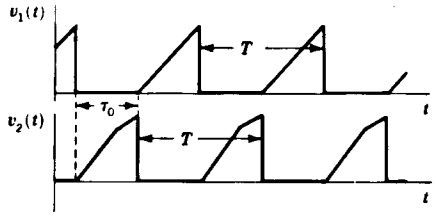



**Figure 1.16-1** Two related waveforms. The timing is such that the product $v_1(t)v_2(t) = 0$.

If $v_1(t)$ and $v_2(t)$ are waveforms of finite energy (for example, nonperiodic pulse-type waveforms), then the cross correlation is defined as

$$R_{12}(\tau) = \int_{-\infty}^{\infty} v_1(t)v_2(t+\tau)\,dt \tag{1.16-3}$$

The need for introducing the parameter $\tau$ in the definition of cross correlation may be seen by the example illustrated in Fig. 1.16-1. Here the two waveforms, while different, are obviously related. They have the same period and nearly the same form. However, the integral of the product $v_1(t)v_2(t)$ is zero since at all times one or the other function is zero. The function $v_2(t+\tau)$ is the function $v_2(t)$ shifted to the left by amount $\tau$. It is clear from the figure that, while $R_{12}(0) = 0$, $R_{12}(\tau)$ will increase as $\tau$ increases from zero, becoming a maximum when $\tau = \tau_0$. Thus $\tau$ is a "searching" or "scanning" parameter which may be adjusted to a proper time shift to reveal, to the maximum extent possible, the relatedness or correlation between the functions. The term *coherence* is sometimes used as a synonym for correlation. Functions for which $R_{12}(\tau) = 0$ for all $\tau$ are described as being *uncorrelated* or *noncoherent*.

In scanning to see the extent of the correlation between functions, it is necessary to specify which function is being shifted. In general, $R_{12}(\tau)$ is not equal to $R_{21}(\tau)$. It is readily verified (Prob. 1.16-2) that

$$R_{21}(\tau) \equiv \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} v_1(t+\tau)v_2(t)\,dt = R_{12}(-\tau) \tag{1.16-4}$$

with identical results for periodic waveforms or waveforms of finite energy.

## 1.17 POWER AND CROSS CORRELATION

Let $v_1(t)$ and $v_2(t)$ be waveforms which are not periodic nor confined to a finite time interval. Suppose that the normalized power of $v_1(t)$ is $S_1$ and the normalized power of $v_2(t)$ is $S_2$. What, then, is the normalized power of $v_1(t) + v_2(t)$? Or,

more generally, what is the normalized power $S_{12}$ of $v_1(t) + v_2(t+\tau)$? We have

$$S_{12} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} [v_1(t) + v_2(t+\tau)]^2 \, dt \tag{1.17-1a}$$

$$= \lim_{T\to\infty} \frac{1}{T} \left\{ \int_{-T/2}^{T/2} v_1^2(t)\, dt + \int_{-T/2}^{T/2} [v_2(t+\tau)]^2 \, dt + 2\int_{-T/2}^{T/2} v_1(t) v_2(t+\tau)\, dt \right\} \tag{1.17-1b}$$

$$= S_1 + S_2 + 2R_{12}(\tau) \tag{1.17-1c}$$

In writing Eq. (1.17-1c), we have taken account of the fact that the normalized power of $v_2(t+\tau)$ is the same as the normalized power of $v_2(t)$. For, since the integration in Eq. (1.17-1b) extends eventually over the entire time axis, a time shift in $v_2$ will clearly not affect the value of the integral.

From Eq. (1.17-1c) we have the important result that if two waveforms are uncorrelated, that is, $R_{12}(\tau) = 0$ for all $\tau$, then no matter how these waveforms are time-shifted with respect to one another, the normalized power due to the superposition of the waveforms is the sum of the powers due to the waveforms individually. Similarly if a waveform is the sum of any number of mutually uncorrelated waveforms, the normalized power is the sum of the individual powers. It is readily verified that the same result applies for periodic waveforms. For finite energy waveforms, the result applies to the normalized energy.

Suppose that two waveforms $v_1'(t)$ and $v_2'(t)$ are uncorrelated. If dc components $V_1$ and $V_2$ are added to the waveforms, then the waveforms $v_1(t) = v_1'(t) + V_1$ and $v_2(t) = v_2'(t) + V_2$ will be correlated with correlation $R_{12}(\tau) = V_1 V_2$. In most applications where the correlations between waveforms is of concern, there is rarely any interest in the dc component. It is customary, then, to continue to refer to waveforms as being uncorrelated if the only source of the correlation is the dc components.

## 1.18 AUTOCORRELATION

The correlation of a function with itself is called the *autocorrelation.* Thus with $v_1(t) = v_2(t)$, $R_{12}(\tau)$ becomes $R(\tau)$ given, in the general case, by

$$R(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} v(t) v(t+\tau)\, dt \tag{1.18-1}$$

A number of the properties of $R(\tau)$ are listed in the following:

(*a*)
$$R(0) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} [v(t)]^2 \, dt = S \tag{1.18-2}$$



That is, the autocorrelation for $\tau = 0$ is the average power $S$ of the waveform.

(*b*)
$$R(0) \geq R(\tau) \tag{1.18-3}$$

This result is rather intuitively obvious since we would surely expect that similarity between $v(t)$ and $v(t+\tau)$ be a maximum when $\tau = 0$. The student is guided through a more formal proof in Prob. 1.18-1.

(*c*)
$$R(\tau) = R(-\tau) \tag{1.18-4}$$

Thus the autocorrelation function is an even function of $\tau$. To prove Eq. (1.18-4), assume that the axis $t = 0$ is moved in the negative $t$ direction by amount $\tau$. Then the integrand in Eq. (1.18-1) would become $v(t-\tau)v(t)$, and $R(\tau)$ would become $R(-\tau)$. Since, however, the integration eventually extends from $-\infty$ to $\infty$, such a shift in time axis can have no effect on the value of the integral. Thus $R(\tau) = R(-\tau)$.

The three characteristics given in Eqs. (1.18-2) to (1.18-4) are features not only of $R(\tau)$ defined by Eq. (1.18-1) but also for $R(\tau)$ as defined by Eqs. (1.16-2) and (1.16-3) for the periodic case and the non-periodic case of finite energy. In the latter case, of course, $R(0) = E$, the energy rather than the power.

## 1.19 AUTOCORRELATION OF A PERIODIC WAVEFORM

When the function is periodic, we may write

$$v(t) = \sum_{n=-\infty}^{\infty} V_n e^{j2\pi nt/T_0} \tag{1.19-1}$$

and, using the correlation integral in the form of Eq. (1.16-2), we have

$$R(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} \left( \sum_{m=-\infty}^{\infty} V_m e^{j2\pi mt/T_0} \right) \left( \sum_{n=-\infty}^{\infty} V_n e^{j2\pi n(t+\tau)/T_0} \right) dt \tag{1.19-2}$$

The order of integration and summation may be interchanged in Eq. (1.19-2). If we do so, we shall be left with a double summation over $m$ and $n$ of terms $I_{m,n}$ given by

$$I_{m,n} = \frac{1}{T_0} e^{j2\pi n\tau/T_0} \int_{-T_0/2}^{T_0/2} V_m V_n e^{j2\pi(m+n)t/T_0}\, dt \tag{1.19-3a}$$

$$= V_m V_n e^{j2\pi n\tau/T_0} \frac{[\sin \pi(m+n)]}{\pi(m+n)} \tag{1.19-3b}$$

Since $m$ and $n$ are integers, we see from Eq. (1.19-3b) that $I_{m,n} = 0$ except when $m = -n$ or $m + n = 0$. To evaluate $I_{m,n}$ in this latter case, we return to Eq. (1.19-3a) and find

$$I_{m,n} = I_{-n,n} = V_n V_{-n} e^{j2\pi n\tau/T_0} \tag{1.19-4}$$

Finally, then.

$$R(\tau) = \sum_{n=-\infty}^{\infty} V_n V_{-n} e^{j2\pi nt/T_0} = \sum_{n=-\infty}^{\infty} |V_n|^2 e^{j2\pi nt/T_0} \tag{1.19-5a}$$

$$= |V_0|^2 + 2\sum_{n=1}^{\infty} |V_n|^2 \cos 2\pi n \frac{\tau}{T_0} \tag{1.19-5b}$$

We note from Eq. (1.19-5*b*) that $R(\tau) = R(-\tau)$ as anticipated, and we note as well that for this case of a periodic waveform the correlation $R(\tau)$ is also periodic with the same fundamental period $T_0$.

We shall now relate the correlation function $R(\tau)$ of a periodic waveform to its power spectral density. For this purpose we compute the Fourier transform of $R(\tau)$. We find, using $R(\tau)$ as in Eq. (1.19-5*a*), that

$$\mathcal{F}[R(\tau)] = \int_{-\infty}^{\infty} \left( \sum_{n=-\infty}^{\infty} |V_n|^2 e^{j2\pi nt/T_0} \right) e^{-j2\pi f\tau}\, d\tau \tag{1.19-6}$$

Interchanging the order of integration and summation yields

$$\mathcal{F}[R(\tau)] = \sum_{n=-\infty}^{\infty} |V_n|^2 \int_{-\infty}^{\infty} e^{-j2\pi(f-n/T_0)\tau}\, d\tau \tag{1.19-7}$$

Using Eq. (1.11-22), we may write Eq. (1.19-7) as

$$\mathcal{F}[R(\tau)] = \sum_{n=-\infty}^{\infty} |V_n|^2\, \delta\left(f - \frac{n}{T_0}\right) \tag{1.19-8}$$

Comparing Eq. (1.19-8) with Eq. (1.8-6), we have the interesting result that for a periodic waveform

$$G(f) = \mathcal{F}[R(\tau)] \tag{1.19-9}$$

and, of course, conversely

$$R(\tau) = \mathcal{F}^{-1}[G(f)] \tag{1.19-10}$$

Expressed in words, we have the following: *The power spectral density and the correlation function of a periodic waveform are a Fourier transform pair.*

## 1.20 AUTOCORRELATION OF NONPERIODIC WAVEFORM OF FINITE ENERGY

For pulse-type waveforms of finite energy there is a relationship between the correlation function of Eq. (1.18-1) and the energy spectral density which corresponds to the relationship given in Eq. (1.19-9) for the periodic waveform. This relationship is that the correlation function $R(\tau)$ and the *energy* spectral density are a Fourier transform pair. This result is established as follows.



We use the convolution theorem. We combine Eqs. (1.12-1) and (1.12-3) for the case where the waveforms $v_1(t)$ and $v_2(t)$ are the same waveforms, that is, $v_1(t) = v_2(t) = v(t)$, and get

$$\mathcal{F}^{-1}[V(f)V(f)] = \int_{-\infty}^{\infty} v(\tau)v(t-\tau)\, d\tau \tag{1.20-1}$$

Since $V(-f) = V^*(f) = \mathcal{F}[v(-t)]$, Eq. (1.20-1) may be written

$$\mathcal{F}^{-1}[V(f)V^*(f)] = \mathcal{F}^{-1}[|V(f)|^2] = \int_{-\infty}^{\infty} v(\tau)v(\tau - t)\, d\tau \tag{1.20-2}$$

The integral in Eq. (1.20-2) is a function of $t$, and hence this equation expresses $\mathcal{F}^{-1}[V(f)V^*(f)]$ as a function of $t$. If we want to express $\mathcal{F}^{-1}[V(f)V^*(f)]$ as a function of $\tau$ without changing the form of the function, we need but to interchange $t$ and $\tau$. We then have

$$\mathcal{F}^{-1}[V(f)V^*(f)] = \int_{-\infty}^{\infty} v(t)v(t-\tau)\, dt \tag{1.20-3}$$

The integral in Eq. (1.20-3) is precisely $R(\tau)$, and thus

$$\mathcal{F}[R(\tau)] = V(f)V^*(f) = |V(f)|^2 \tag{1.20-4}$$

which verifies that $R(\tau)$ and the energy spectral density $|V(f)|^2$ are Fourier transform pairs.

## 1.21 AUTOCORRELATION OF OTHER WAVEFORMS

In the preceding sections we discussed the relationship between the autocorrelation function and power or energy spectral density of *deterministic* waveforms. We use the term "deterministic" to indicate that at least, in principle, it is possible to write a function which specifies the value of the function at all times. For such deterministic waveforms, the availability of an autocorrelation function is of no particularly great value. The autocorrelation function does not include within itself complete information about the function. Thus we note that the autocorrelation function is related only to the amplitudes and not to the phases of the spectral components of the waveform. The waveform cannot be reconstructed from a knowledge of the autocorrelation functions. Any characteristic of a deterministic waveform which may be calculated with the aid of the autocorrelation function may be calculated by direct means at least as conveniently.

On the other hand, in the study of communication systems we encounter waveforms which are not deterministic but are instead random and unpredictable in nature. Such waveforms are discussed in Chap. 2. There we shall find that for such random waveforms no explicit function of time can be written. The waveforms must be described in statistical and probabilistic terms. It is in connection with such waveforms that the concepts of correlation and autocorrelation find their true usefulness. Specifically, it turns out that even for such random wave-

forms, the autocorrelation function and power spectral density are a Fourier transform pair. The proof[3] that such is the case is formidable and will not be undertaken here.

## 1.22 EXPANSIONS IN ORTHOGONAL FUNCTIONS

Let us consider a set of functions $g_1(x), g_2(x), \ldots, g_n(x) \cdots$, defined over the interval $x_1 \leq x < x_2$ and which are related to one another in the very special way that any two different ones of the set satisfy the condition

$$\int_{x_1}^{x_2} g_i(x) g_j(x)\, dx = 0 \tag{1.22-1}$$

That is, when we multiply two different functions and then integrate over the interval from $x_1$ to $x_2$ the result is zero. A set of functions which has this property is described as being *orthogonal over the interval from* $x_1$ *to* $x_2$. The term "orthogonal" is employed here in correspondence to a similar situation which is encountered in dealing with vectors. The *scalar* product of two vectors $\mathbf{V}_i$ and $\mathbf{V}_j$ (also referred to as the *dot* product or as the *inner* product) is a scalar quantity $V_{ij}$ defined as

$$V_{ij} = |\mathbf{V}_i||\mathbf{V}_j| \cos(\mathbf{V}_i, \mathbf{V}_j) = V_{ji} \tag{1.22-2}$$

In Eq. (1.22-2) $|\mathbf{V}_i|$ and $|\mathbf{V}_j|$ are the magnitudes of the respective vectors and $\cos(\mathbf{V}_i, \mathbf{V}_j)$ is the cosine of the angle between the vectors. If it should turn out that $V_{ij} = 0$ then (ignoring the trivial cases in which $\mathbf{V}_i = 0$ or $\mathbf{V}_j = 0$) $\cos(\mathbf{V}_i, \mathbf{V}_j)$ must be zero and correspondingly it means that the vectors $\mathbf{V}_i$ and $\mathbf{V}_j$ are perpendicular (i.e., orthogonal) to one another. Thus vectors whose scalar product is zero are physically orthogonal to one another and, in correspondence, functions whose integrated product, as in Eq. (1.22-1) is zero are also orthogonal to one another.

Now consider that we have some arbitrary function $f(x)$ and that we are interested in $f(x)$ only in the range from $x_1$ to $x_2$, i.e., in the interval over which the set of functions $g(x)$ are orthogonal. Suppose further that we undertake to write $f(x)$ as a linear sum of the functions $g_n(x)$. That is, we write

$$f(x) = C_1 g_1(x) + C_2 g_2(x) + \cdots + C_n g_n(x) + \cdots \tag{1.22-3}$$

in which the $C$'s are numerical coefficients. Assuming that such an expansion is indeed possible, the orthogonality of the $g$'s makes it very easy to compute the coefficients $C_n$. Thus to evaluate $C_n$ we multiply both sides of Eq. (1.22-3) by $g_n(x)$ and integrate over the interval of orthogonality. We have

$$\int_{x_1}^{x_2} f(x) g_n(x)\, dx = C_1 \int_{x_1}^{x_2} g_1(x) g_n(x)\, dx + C_2 \int_{x_1}^{x_2} g_2(x) g_n(x)\, dx + \cdots + C_n \int_{x_1}^{x_2} g_n^2(x)\, dx + \cdots \tag{1.22-4}$$



Because of the orthogonality, *all of the terms on the right-hand side of Eq. (1.22-4) become zero with a single exception* and we are left with

$$\int_{x_1}^{x_2} f(x) g_n(x)\, dx = C_n \int_{x_1}^{x_2} g_n^2(x)\, dx \tag{1.22-5}$$

so that the coefficient that we are evaluating becomes

$$C_n = \frac{\displaystyle\int_{x_1}^{x_2} f(x) g_n(x)\, dx}{\displaystyle\int_{x_1}^{x_2} g_n^2(x)\, dx} \tag{1.22-6}$$

The mechanism by which we use the orthogonality of the functions to "drain" away all the terms except the term that involves the coefficient we are evaluating is often called the "orthogonality sieve."

Next suppose that each $g_n(x)$ is selected so that the denominator of the right-hand member of Eq. (1.22-6) (which is a numerical constant) has the value

$$\int_{x_1}^{x_2} g_n^2(x)\, dx = 1 \tag{1.22-7}$$

In this case

$$C_n = \int_{x_1}^{x_2} f(x) g_n(x)\, dx \tag{1.22-8}$$

When the orthogonal functions $g_n(x)$ are selected as in (1.22-7) they are described as being *normalized*. The use of normalized functions has the merit that the $C_n$'s can then be calculated from Eq. (1.22-8) and thereby avoids the need to evaluate $\int_{x_1}^{x_2} g_n^2(x)\, dx$ in each case as called for in Eq. (1.22-6). A set of functions which is both orthogonal and normalized is called an *orthonormal* set.

## 1.23 COMPLETENESS OF AN ORTHOGONAL SET: THE FOURIER SERIES

Suppose on the one hand we expand a function $f(x)$ in terms of orthogonal functions as

$$f(x) = C_1 s_1(x) + C_2 s_2(x) + C_3 s_3(x) + \cdots \tag{1.23-1}$$

and on the other hand, we expand it as

$$f(x) = C_1 s_1(x) + C_3 s_3(x) + \cdots \tag{1.23-2}$$

That is, in the second case, we have deliberately omitted one term. A moment's review of the procedure, described in the previous section, for evaluating coefficients makes it apparent that all the coefficients $C_1$, $C_3$, etc., that appear in both expansions will turn out to be *the same*. Hence if one expansion is correct the other is in error. We might be suspicious of the expansion of Eq. (1.23-2) on the

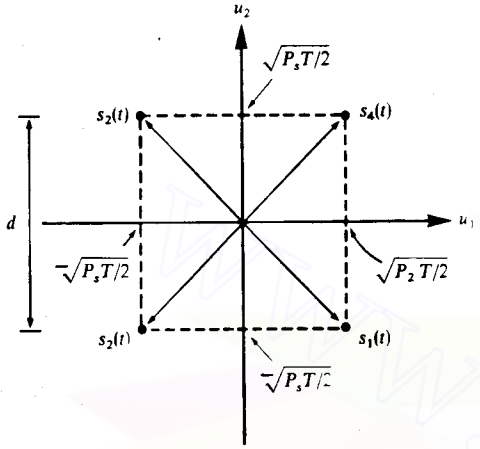



**Figure 1.26-2** Signal vectors representing the four signals of Eq. (1.26-2).

It is important to note that the magnitude of the signal vectors shown in Fig. 1.26-2 is equal to the square root of the signal energy. Thus, in Eq. (1.26-2) each signal $s_i(t)$ has a power $P_s$ and a duration $T$. Thus the signal energy is $P_s T$, which is equal to the magnitude squared of the signal vectors as shown in Fig. 1.26-2. This result simplifies the drawing of the signal vectors in signal space.

For example, consider the signal

$$s_i(t) = \sqrt{2P_s}\cos\left[\omega_0 t + (2i-1)\frac{\pi}{8}\right] \qquad \begin{cases} i = 1, 2, \ldots, 8 \\ 0 \le t \le T \end{cases} \qquad (1.26\text{-}6)$$

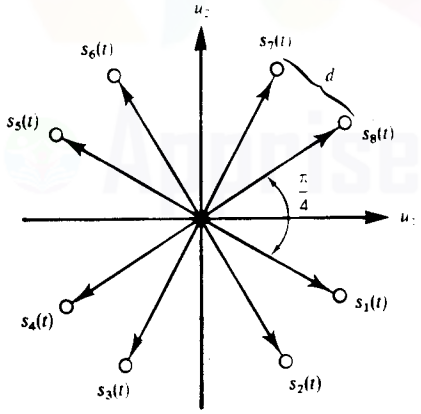



**Figure 1.26-3** Signal vectors representing the four signals of Eq. (1.26-6).



Then each signal has the power $P_s$ and duration $T$. Thus the magnitude of each signal vector is $\sqrt{P_s T}$. Furthermore, each vector is rotated $\pi/4$ radians from the previous vector as shown in Fig. 1.26-3. The reader should verify (Prob. 1.26-1) that Fig. 1.26-3 is correct by finding the two orthonormal components $u_1(t)$ and $u_2(t)$ as in Eqs. (1.26-3) and (1.26-4), expanding Eq. (1.26-6) in a form similar to Eq. (1.26-5) and then plotting the result.

## REFERENCES

1. Javid, M., and E. Brenner: "Analysis of Electric Circuits," 2d ed., McGraw-Hill Book Company, New York, 1967.
   Papoulis, A.: "The Fourier Integral and Applications," McGraw-Hill Book Company, New York, 1963.
2. Churchil, R. V.: "Fourier Series," McGraw-Hill Book Company, New York, 1941.
3. Papoulis, A.: "Probability, Random Variables, and Stochastic Processes," McGraw-Hill Book Company, New York, 1965.

## PROBLEMS

**1.1-1.** Verify the relationship of $C_n$ and $\phi_n$ to $A_n$ and $B_n$ as given by Eq. (1.1-6).

**1.1-2.** Calculate $A_n$, $B_n$, $C_n$, and $\phi_n$ for a waveform $v(t)$ which is a symmetrical square wave and which makes peak excursions to $+\frac{3}{4}$ volt and $-\frac{1}{4}$ volt, and has a period $T = 1$ sec. A positive going transition occurs at $t = 0$.

**1.2-1.** Verify the relationship of the complex number $V_n$ to $C_n$ and $\phi_n$ as given by Eq. (1.2-3).

**1.2-2.** The function

$$p(t) = \begin{cases} e^{-t} & 0 \le t \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

is repeated every $T = 1$ sec. Thus, with $u(t)$ the unit step function

$$v(t) = \sum_{n=-\infty}^{\infty} p(t-n)u(t-n)$$

Find $V_n$ and the exponential Fourier series for $v(t)$.

**1.3-1.** The impulse function can be defined as

$$\delta(t) = \lim_{a \to \infty} \frac{a}{2} e^{-a|t|}$$

Discuss.

**1.3-2.** In Eq. (1.3-8) we see that

$$v(t) = I \sum_{k=-\infty}^{\infty} \delta(t - kT_0) = \frac{I}{T_0} + \frac{2I}{T_0} \sum_{n=1}^{\infty} \cos \frac{2\pi n t}{T_0}$$

Set $I = 1$, $T_0 = 1$. Show by plotting

$$v(t) = 1 + 2 \sum_{n=1}^{3} \cos 2\pi n t$$

that the Fourier series does approximate the train of impulses.

**1.4-1.** $Sa(x) \equiv (\sin x)/x$. Determine the maxima and minima of $Sa(x)$ and compare your result with the approximate maxima and minima obtained by letting $x = (2n + 1)\pi/2$, $n = 1, 2, \ldots$.

**1.4-2.** A train of rectangular pulses, making excursions from zero to 1 volt, have a duration of 2 $\mu$s and are separated by intervals of 10 $\mu$s.

(*a*) Assume that the center of one pulse is located at $t = 0$. Write the exponential Fourier series for this pulse train and plot the spectral amplitude as a function of frequency. Include at least 10 spectral components on each side of $f = 0$, and draw also the envelope of these spectral amplitudes.

(*b*) Assume that the left edge of a pulse rather than the center is located at $t = 0$. Correct the Fourier expansion accordingly. Does this change affect the plot of spectral amplitudes? Why not?

**1.5-1.** (*a*) A periodic waveform $v_i(t)$ is applied to the input of a network. The output $v_o(t)$ of the network is $v_o(t) = \tau[dv_i(t)/dt]$, where $\tau$ is a constant. What is the transfer function $H(\omega)$ of this network?

(*b*) A periodic waveform $v_i(t)$ is applied to the $RC$ network shown, whose time constant is $\tau \equiv RC$. Assume that the highest frequency spectral component of $v_i(t)$ is of a frequency $f \ll 1/\tau$. Show that, under these circumstances, the output $v_o(t)$ is approximately $v_o(t) \simeq \tau[dv_i(t)/dt]$.

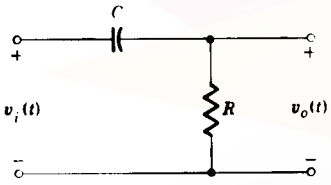



**Figure P1.5-1**

**1.5.2.** A voltage represented by an impulse train of strength $I$ and period $T$ is filtered by a low-pass $RC$ filter having a 3-dB frequency $f_c$.

(*a*) Find the Fourier series of the output voltage across the capacitor.

(*b*) If the third harmonic of the output is to be attenuated by 1000, find $f_c$ $T$.

**1.6-1.** Measurements on a voltage amplifier indicate a gain of 20 dB.

(*a*) If the input voltage is 1 volt, calculate the output voltage.

(*b*) If the input power is 1 mw, calculate the output power.

**1.6-2.** A voltage gain of 0.1 is produced by an attenuator.

(*a*) What is the gain in decibels?

(*b*) What is the power gain (not in decibels)?

**1.7-1.** A periodic triangular waveform $v(t)$ is defined by

$$v(t) = \frac{2t}{T} \qquad \text{for} \qquad -\frac{T}{2} < t < \frac{T}{2}$$

and $$v(t \pm T) = v(t)$$

and has the Fourier expansion

$$v(t) = \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin 2\pi n \frac{t}{T}$$

Calculate the fraction of the normalized power of this waveform which is contained in its first three harmonics.

**1.7-2.** The complex spectral amplitudes of a periodic waveform are given by

$$V_n = \frac{1}{|n|} e^{-j \arctan(n/2)} \qquad n = \pm 1, \pm 2, \ldots$$

Find the ratio of the normalized power in the second harmonic to the normalized power in the first harmonic.







**1.8-1.** Find $G(f)$ for the following voltages:

(*a*) An impulse train of strength $I$ and period $T$.

(*b*) A pulse train of amplitude $A$, duration $\tau = I/A$, and period $T$.

**1.8-2.** Plot $G(f)$ for a voltage source represented by an impulse train of strength $I$ and period $nT$ for $n = 1, 2, 10$, infinity. Comment on this limiting result.

**1.9-1.** $G_i(f)$ is the power spectral density of a *square-wave* voltage of peak-to-peak amplitude 1 and period 1. The square wave is filtered by a low-pass $RC$ filter with a 3-dB frequency 1. The output is taken across the capacitor.

(*a*) Calculate $G_i(f)$.

(*b*) Find $G_0(f)$.

**1.9-2.** (*a*) A symmetrical square wave of zero mean value, peak-to-peak voltage 1 volt, and period 1 sec is applied to an ideal low-pass filter. The filter has a transfer function $|H(f)| = \frac{1}{2}$ in the frequency range $-3.5 \leq f \leq 3.5$ Hz, and $H(f) = 0$ elsewhere. Plot the power spectral density of the filter output.

(*b*) What is the normalized power of the input square wave? What is the normalized power of the filter output?

**1.10-1.** In Eqs. (1.2-1) and (1.2-2) write $f_0 \equiv 1/T_0$. Replace $f_0$ by $\Delta f$, i.e., $\Delta f$ is the frequency interval between harmonics. Replace $n\,\Delta f$ by $f$, i.e., as $\Delta f \to 0$, $f$ becomes a continuous variable ranging from 0 to $\infty$ as $n$ ranges from 0 to $\infty$. Show that in the limit as $\Delta f \to 0$, so that $\Delta f$ may be replaced by the differential $df$, Eq. (1.2-1) becomes

$$v(t) = \int_{-\infty}^{\infty} V(f) e^{j2\pi ft}\, df$$

in which $V(f)$ is

$$V(f) = \lim_{\substack{f_0 = \Delta f \to 0 \\ nf_0 \to f}} \int_{-1/2f_0}^{1/2f_0} v(t) e^{-j2\pi nf_0 t}\, dt = \int_{-\infty}^{\infty} v(t) e^{-j2\pi ft}\, dt$$

**1.10-2.** Find the Fourier transform of $\sin \omega_0 t$. Compare with the transform of $\cos \omega_0 t$. Plot and compare the power spectral densities of $\cos \omega_0 t$ and $\sin \omega_0 t$.

**1.10-3.** The waveform $v(t)$ has the Fourier transform $V(f)$. Show that the waveform delayed by time $t_d$, i.e., $v(t - t_d)$ has the transform $V(f)e^{-j\omega t_d}$.

**1.10-4.** (*a*) The waveform $v(t)$ has the Fourier transform $V(f)$. Show that the time derivative $(d/dt)v(t)$ has the transform $(j2\pi f)V(f)$.

(*b*) Show that the transform of the integral of $v(t)$ is given by

$$\mathcal{F}\left[\int_{-\infty}^{t} v(\lambda)\, d\lambda\right] = \frac{V(f)}{j2\pi f}$$

**1.12-1.** Derive the convolution formula in the frequency domain. That is, let $V_1(f) = \mathcal{F}[v_1(t)]$ and $V_2(f) = \mathcal{F}[v_2(t)]$. Show that if $V(f) = \mathcal{F}[v_1(t)v_2(t)]$, then

$$V(f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} V_1(\lambda) V_2(f - \lambda)\, d\lambda$$

or

$$V(f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} V_2(\lambda) V_1(f - \lambda)\, d\lambda$$

**1.12-2.** (*a*) A waveform $v(t)$ has a Fourier transform which extends over the range from $-f_M$ to $+f_M$. Show that the waveform $v^2(t)$ has a Fourier transform which extends over the range from $-2f_M$ to $+2f_M$. (*Hint:* Use the result of Prob. 1.12-1.)

(*b*) A waveform $v(t)$ has a Fourier transform $V(f) = 1$ in the range $-f_M$ to $+f_M$ and $V(f) = 0$ elsewhere. Make a plot of the transform of $v^2(t)$.

**1.12-3.** A filter has an impulse response $h(t)$ as shown. The input to the network is a pulse of unit amplitude extending from $t = 0$ to $t = 2$. By graphical means, determine the output of the filter.

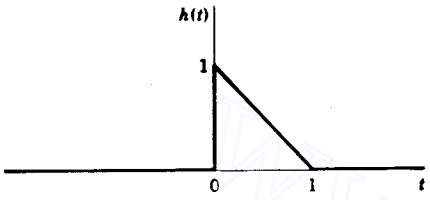



Figure P1.12-3

**1.13-1.** The energy of a nonperiodic waveform $v(t)$ is

$$E = \int_{-\infty}^{\infty} v^2(t)\, dt$$

(*a*) Show that this can be written as

$$E = \int_{-\infty}^{\infty} dt v(t) \int_{-\infty}^{\infty} V(f) e^{j2\pi ft}\, df$$

(*b*) Show that by interchanging the order of integration we have

$$E = \int_{-\infty}^{\infty} V(f)V^*(f)\, df = \int_{-\infty}^{\infty} |V(f)|^2\, df$$

which proves Eq. (1.13-5). This is an alternate proof of Parseval's theorem.

**1.13-2.** If $V(f) = AT \sin 2\pi fT/2\pi fT$, find the energy $E$ contained in $v(t)$.

**1.13-3.** A waveform $m(t)$ has a Fourier transform $M(f)$ whose magnitude is as shown.

(*a*) Find the normalized energy content of the waveform.

(*b*) Calculate the frequency $f_1$ such that one-half of the normalized energy is in the frequency range $-f_1$ to $f_1$.

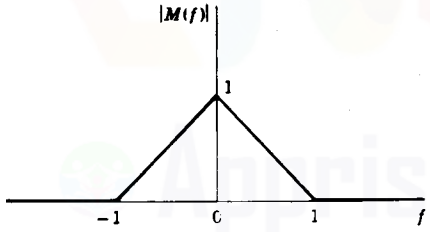



Figure P1.13-3

**1.14-1.** The signal $v(t) = \cos \omega_0 t + 2 \sin 3\omega_0 t + 0.5 \sin 4\omega_0 t$ is filtered by an $RC$ low-pass filter with a 3-dB frequency $f_c = 2f_0$.

(*a*) Find $G_i(f)$.

(*b*) Find $G_0(f)$.

(*c*) Find $S_0$.

**1.14-2.** The waveform $v(t) = e^{-t/t}u(t)$ is passed through a high-pass $RC$ circuit having a time constant equal to $t$.

(*a*) Find the energy spectral density at the output of the circuit.

(*b*) Show that the total output energy is one-half the input energy.



**1.15-1.** (*a*) An impulse of strength $I$ is applied to a low-pass $RC$ circuit of 3-dB frequency $f_2$. Calculate the output waveform.

(*b*) A pulse of amplitude $A$ and duration $\tau$ is applied to the low-pass $RC$ circuit. Show that, if $\tau \ll 1/f_2$, the response at the output is approximately the response that would result from the application to the circuit of an impulse of strength $I' = A\tau$. Generalize this result by considering any voltage waveform of area $I'$ and duration $\tau$.

**1.15-2.** A pulse extending from 0 to A volts and having a duration $\tau$ is applied to a high-pass $RC$ circuit. Show that the area under the response waveform is zero.

**1.16-1.** Find the cross correlation of the functions $\sin \omega t$ and $\cos \omega t$.

**1.16-2.** Prove that $R_{21}(\tau) = R_{12}(-\tau)$.

**1.17-1.** Find the cross-correlation function $R_{12}(\tau)$ of the two periodic waveforms shown.

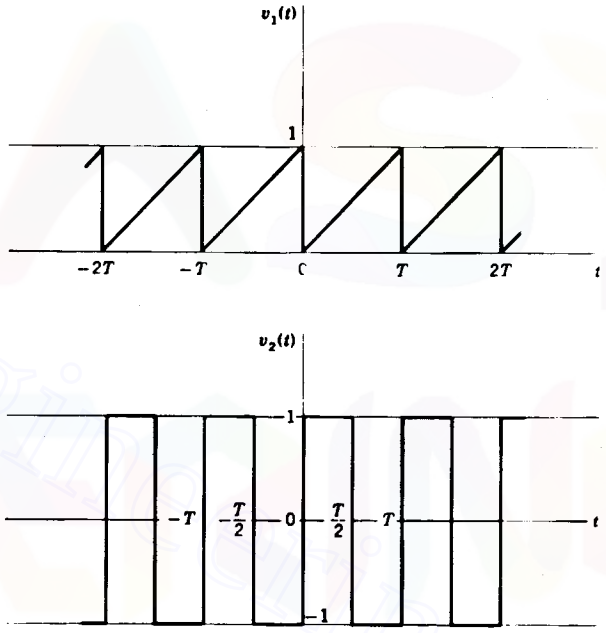



Figure P1.17-1

**1.18-1.** $$R(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} v(t)v(t+\tau)\, dt$$

Prove that $R(0) \geq R(\tau)$. *Hint:* Consider

$$I = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} [v(t) - v(t+\tau)]^2\, dt$$

Is $I \geq 0$? Expand and integrate term by term. Show that

$$I = 2[R(0) - R(\tau)] \geq 0$$

**1.18-2.** Determine an expression for the correlation function of a square wave having the values 1 or 0 and a period $T$.

**1.19-1.** (*a*) Find the power spectral density of a square-wave voltage by Fourier transforming the correlation function. Use the results of Prob. 1.18-2.

(*b*) Compare the answer to (*a*) with the spectral density obtained from the Fourier series (Prob. 1.9-1*a*) itself.

**1.19-2.** If $v(t) = \sin \omega_0 t$.

(*a*) Find $R(\tau)$.

(*b*) If $G(f) = \mathcal{F}[R(\tau)]$, find $G(f)$ directly and compare.

**1.20-1.** A waveform consists of a single pulse of amplitude $A$ extending from $t = -\tau/2$ to $t = \tau/2$.

(*a*) Find the autocorrelation function $R(\tau)$ of this waveform.

(*b*) Calculate the energy spectral density of this pulse by evaluating $G_E(f) = \mathcal{F}[R(\tau)]$.

(*c*) Calculate $G_E(f)$ directly by Parseval's theorem and compare.

**1.24-1.** Interchange the labels $s_1(t)$ and $s_2(t)$ on the waveforms of Fig. 1.24-1*a* and with this new labeling apply the Gram–Schmidt procedure to find expansions of the waveforms in terms of orthonormal functions.

**1.24-2.** Use the Gram–Schmidt procedure to express the functions in Fig. P1.24-2 in terms of orthonormal components.

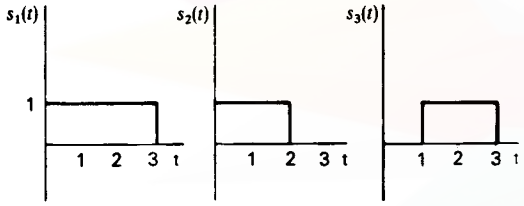



**Figure P1.24-2**

**1.24-3.** A set of signals ($k = 1, 2, 3, 4$) is given by

$$s_k(t) = \cos\left(\omega_0 t + k\frac{\pi}{2}\right) \qquad 0 \le t \le k\frac{2\pi}{\omega_0} = 0$$

$$= 0 \qquad \text{otherwise}$$

Use the Gram–Schmidt procedure to find an orthonormal set of functions in which the functions $s_k(t)$ can be expanded.

**1.25-1.** Verify Eq. (1.25-10*c*).

**1.25-2.** (*a*) Show that the pythagorean theorem applies to signal functions. That is, show that if $s_1(t)$ and $s_2(t)$ are orthogonal then the square of the length of $s_1(t)$ (defined as in Eq. (1.25-10*b*)) plus the square of the length of $s_2(t)$ is equal to the square of the length of the sum $s_1(t) + s_2(t)$.

(*b*) Let $s_1(t)$ and $s_2(t)$ be the signals shown in Fig. P1.25-2. Draw the signal $s_1(t) + s_2(t)$. Show that $s_1(t)$ and $s_2(t)$ are orthogonal. Show that $|s_1(t) + s_2(t)|^2 = |s_1(t)|^2 + |s_2(t)|^2$.

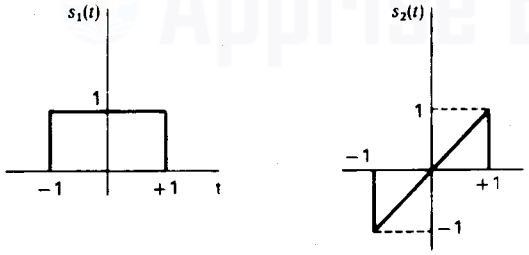



**Figure P1.25-2**

**1.25-3.** Verify Eq. (1.25-13).



**1.25-4.** (*a*) Refer to Fig. 1.25-2 and Eq. (1.25-13). Verify that the parameter $\alpha$ in Eq. (1.25-13) is the tangent of the angle between the axes of the $u_1$, $u_2$ coordinate system and the $u_1'u_2'$ coordinate system.

(*b*) Expand the functions $s_1(t)$ and $s_2(t)$ of Fig. 1.24-1 in terms of orthonormal functions $u_1'(t)$ and $u_2'(t)$ which are the axes of a coordinate system which is rotated 60° counterclockwise from the coordinate system whose axes are $u_1(t)$ and $u_2(t)$ of the same figure.

**1.26-1.** Apply the Gram–Schmidt procedure to the eight signals of Eq. (1.26-6). Verify that two orthonormal components suffice to allow a representation of any of the signals. Show that in a coordinate system of these two orthonormal signals, the geometric representation of the eight signals is as shown in Fig. 1.26-3.

**1.26-2.** A set of signals is $s_{a1}(t) = \sqrt{2P_a}\sin\omega_0 t$, $s_{a2} = -\sqrt{2P_a}\sin\omega_0 t$. A second set of signals is $s_{b2} = \sqrt{2P_b}\sin\omega_0 t$, $s_{b2} = \sqrt{2P_b}\sin(\omega_0 t + \pi/2)$. If the two sets of signals are to have the same distinguishability, what is the ratio $P_b/P_a$?

**1.26-3.** Given two signals

$$s_1(t) = f(t) \qquad 0 \le t \le T; \qquad s_2(t) = -f(t) \qquad 0 \le t \le T$$

Show that independently of the form of $f(t)$ the distinguishability of the signals is given by $\sqrt{2E}$ where $E$ is the normalized energy of the waveform $f(t)$.

**1.26-4.** (*a*) Use the Gram–Schmidt procedure to express the functions in Fig. P1.26-4 in terms of orthonormal components. In applying the procedure, involve the functions in the order $s_1(t)$, $s_2(t)$, $s_3(t)$, and $s_4(t)$. Plot the functions as points in a coordinate system in which the coordinate axes are measured in units of the orthonormal functions.

(*b*) Repeat part (*a*) except that the functions are to be involved in the order $s_1(t)$, $s_4(t)$, $s_3(t)$, and $s_2(t)$. Plot the functions.

(*c*) Show that the procedures of parts (*a*) and (*b*) yield the same distances between function points.

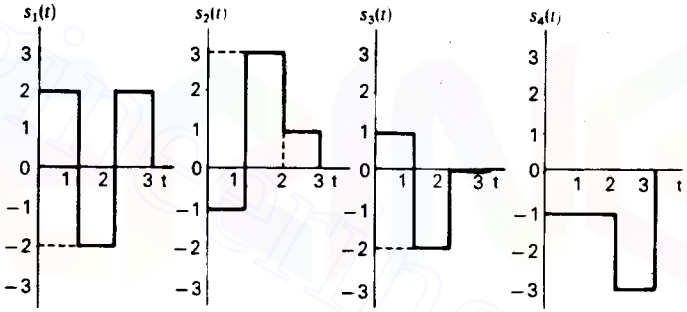



**Figure P1.26-4**

CHAPTER

# TWO

# RANDOM VARIABLES AND PROCESSES

A waveform which can be expressed, at least in principle, as an explicit function of time $v(t)$ is called a *deterministic* waveform. Such a waveform is determined for all times in that, if we select an arbitrary time $t = t_1$, there need be no *uncertainty* about the value of $v(t)$ at that time. The waveforms encountered in communication systems, on the other hand, are in many instances unpredictable. Consider, say, the very waveform itself which is transmitted for the purpose of communication. This waveform, which is called the *signal*, must, at least in part, be unpredictable. If such were not the case, i.e., if the signal were predictable, then its transmission would be unnecessary, and the entire communications system would serve no purpose. This point concerning the unpredictability of the signal is explored further in Chap. 13. Further, as noted briefly in Chap. 1, transmitted signals are invariably accompanied by *noise* which results from the ever-present agitation of the universe at the atomic level. These noise waveforms are also not predictable. Unpredictable waveforms such as a signal voltage $s(t)$ or a noise voltage $n(t)$ are examples of *random processes*. [Note that in writing symbols like $s(t)$ and $n(t)$ we do not imply that we can write explicit functions for these time functions.]

While random processes are not predictable, neither are they completely unpredictable. It is generally possible to predict the future performance of a random process with a certain *probability* of being correct. Accordingly, in this chapter we shall present some elemental ideas of probability theory and apply them to the description of random processes. We shall rather generally limit our discussion to the development of only those aspects of the subject which we shall have occasion to employ in this text.









## 2.1 PROBABILITY[1]

The concept of probability occurs naturally when we contemplate the possible outcomes of an experiment whose outcome is not always the same. Suppose that one of the possible outcomes is called $A$ and that when the experiment is repeated $N$ times the outcome $A$ occurs $N_A$ times. The relative frequency of occurrence of $A$ is $N_A/N$, and this ratio $N_A/N$ is not predictable unless $N$ is very large. For example, let the experiment consist of the tossing of a die and let the outcome $A$ correspond to the appearance of, say, the number 3 on the die. Then in 6 tosses the number 3 may not appear at all, or it may appear 6 times, or any number of times in between. Thus with $N = 6$, $N_A/N$ may be 0 or 1/6, etc., up to $N_A/N = 1$. On the other hand, we know from experience that when an experiment, whose outcomes are determined by chance, is repeated *very many times*, the relative frequency of a particular outcome approaches a fixed limit. Thus, if we were to toss a die very many times we would expect that $N_A/N$ would turn out to be very close to 1/6. This limiting value of the relative frequency of occurrence is called the probability of outcome $A$, written $P(A)$, so that

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N} \tag{2.1-1}$$

In many cases the experiments needed to determine the probability of an event are done more in thought than in practice. Suppose that we have 10 balls in a container, the balls being identical in every respect except that 8 are white and 2 are black. Let us ask about the probability that, in a single draw, we shall select a black ball. If we draw blindly, so that the color has no influence on the outcome, we would surely judge that the probability of drawing the black ball is 2/10. We arrive at this conclusion on the basis that we have postulated that there is absolutely nothing which favors one ball over another. There are 10 possible outcomes of the experiment, that is, any of the 10 balls may be drawn; of these 10 outcomes, 2 are favorable to our interest. The only reasonable outcome we can imagine is that, in very many drawings, 2 out of 10 will be black. Any other outcome would immediately suggest that either the black or the white balls had been favored. These considerations lead to an alternative definition of the probability of occurrence of an event $A$, that is

$$P(A) = \frac{\text{number of possible favorable outcomes}}{\text{total number of possible equally likely outcomes}} \tag{2.1-2}$$

It is apparent from either definition, Eq. (2.1-1) or (2.1-2), that the probability of occurrence of an event $P$ is a positive number and that $0 \leq P \leq 1$. If an event is not possible, then $P = 0$, while if an event is certain, $P = 1$.

## 2.2 MUTUALLY EXCLUSIVE EVENTS

Two possible outcomes of an experiment are defined as being *mutually exclusive* if the occurrence of one outcome precludes the occurrence of the other. In this case, if the events are $A_1$ and $A_2$ with probabilities $P(A_1)$ and $P(A_2)$, then the

the shaded area in Fig. 2.22-2 measures the probability that $m_2$ was transmitted and read as $m_1$.

In Fig. 2.22-3 we represent the situation in which one of four messages $m_1$, $m_2$, $m_3$, and $m_4$ are sent yielding receiver responses $s_1$, $s_2$, $s_3$, and $s_4$. The probability density function $f_N(n)$ of the noise (multiplied by $\frac{1}{4}$) is shown in Fig. 2.22-3*a*. Also shown are the functions $P(m_k)f_N(r - s_k)$ for $k = 1, 2, 3$, and 4. We have taken all the probabilities $P(m_k)$ to be equal at $P(m_k) = \frac{1}{4}$. The boundaries of the received response $R$ at which decisions change are also shown and are marked $\rho_{12}$, $\rho_{23}$, and $\rho_{34}$.

The sum of the two shaded areas shown, each of area $\Delta$, equals the probability $P(E, m_2)$ that $m_2$ is transmitted *and* that the message is erroneously read. Thus

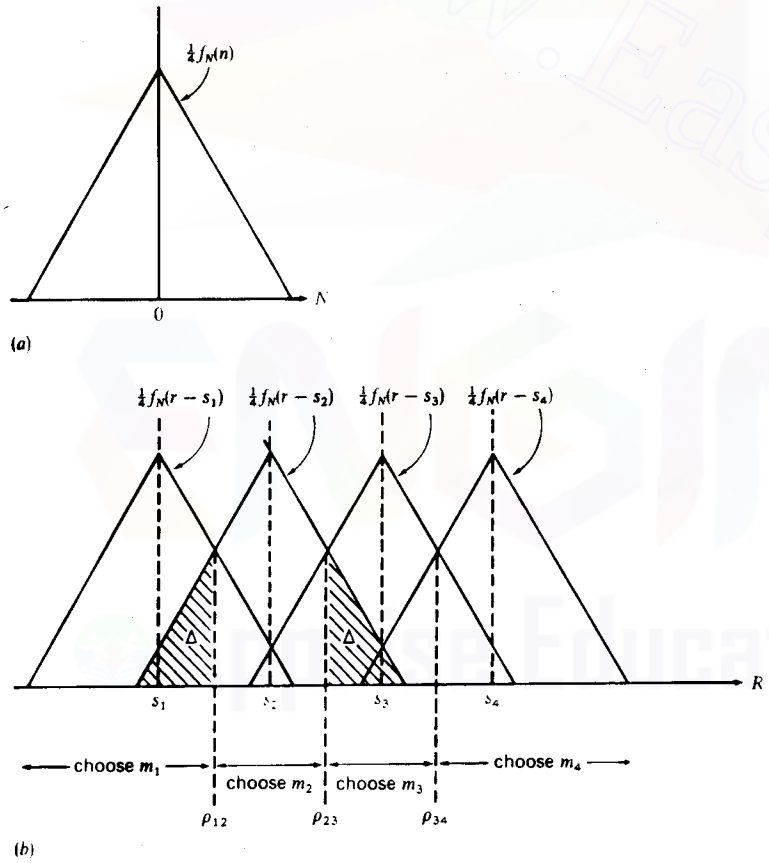



**Figure 2.22-3** Probability density when four signals are transmitted. (*a*) Probability density of $f_2(r)$ when noise alone is received. (*b*) Probability density when one of four signals is transmitted and received in noise.



$P(E, m_2) = 2\Delta$. The remaining area under $\frac{1}{4}f_N(r - s_2)$ is the probability that $m_2$ is transmitted and is correctly read. This probability is $P(C, m_2) = \frac{1}{4} - 2\Delta$. We find of course that $P(E, m_3) = P(E, m_2)$. But it is most interesting to note that when we evaluate $P(E, m_1)$ or $P(E, m_4)$ there is only a single area $\Delta$ involved. Hence $P(E, m_1) = P(E, m_4) = \frac{1}{2}P(E, m_2) = \frac{1}{2}P(E, m_3)$. Thus, we arrive at the most interesting result that given four equally likely messages, when we make boundary decisions which assure minimum average likelihood of error, the probability of making an error is not the same for all messages. Of course, this result applies for any number of equally likely messages greater than two.

The probability that a message is read correctly is

$$\begin{aligned} P(C) &= P(C, m_1) + P(C, m_2) + P(C, m_3) + P(C, m_4) \\ &= (\tfrac{1}{4} - \Delta) + (\tfrac{1}{4} - 2\Delta) + (\tfrac{1}{4} - 2\Delta) + (\tfrac{1}{4} - \Delta) \\ &= 1 - 6\Delta \end{aligned} \tag{2.22-4}$$

Thus the probability of an error is

$$P(E) = 1 - P(C) = 6\Delta \tag{2.22-5}$$

## 2.23 RANDOM PROCESSES

To determine the probabilities of the various possible outcomes of an experiment, it is necessary to repeat the experiment many times. Suppose then that we are interested in establishing the statistics associated with the tossing of a die. We might proceed in either of two ways. On one hand, we might use a single die and toss it repeatedly. Alternatively, we might toss simultaneously a very large number of dice. Intuitively, we would expect that both methods would give the same results. Thus, we would expect that a single die would yield a particular outcome, on the average, of 1 time out of 6. Similarly, with many dice we would expect that 1/6 of the dice tossed would yield a particular outcome.

Analogously, let us consider a random process such as a noise waveform $n(t)$ mentioned at the beginning of this chapter. To determine the statistics of the noise, we might make repeated measurements of the noise voltage output of a single noise source, or we might, at least conceptually, make simultaneous measurements of the output of a very large collection of statistically identical noise sources. Such a collection of sources is called an *ensemble*, and the individual noise waveforms are called *sample functions*. A statistical average may be determined from measurements made at some fixed time $t = t_1$ on all the sample functions of the ensemble. Thus to determine, say, $\overline{n^2(t)}$, we would, at $t = t_1$, measure the voltages $n(t_1)$ of each noise source, square and add the voltages, and divide by the (large) number of sources in the ensemble. The average so determined is the *ensemble average* of $n^2(t_1)$.

Now $n(t_1)$ is a random variable and will have associated with it a probability density function. The ensemble averages will be identical with the statistical averages computed earlier in Secs. 2.11 and 2.12 and may be represented by the same

symbols. Thus the statistical or ensemble average of $n^2(t_1)$ may be written $E[n^2(t_1)] = \overline{n^2(t_1)}$. The averages determined by measurements on a single sample function at successive times will yield a *time average*, which we represent as $\langle n^2(t) \rangle$.

In general, ensemble averages and time averages are not the same. Suppose, for example, that the statistical characteristics of the sample functions in the ensemble were changing with time. Such a variation could not be reflected in measurements made at a fixed time, and the ensemble averages would be different at different times. When the statistical characteristics of the sample functions do not change with time, the random process is described as being *stationary*. However, even the property of being stationary does not ensure that ensemble and time averages are the same. For it may happen that while each sample function is stationary the individual sample functions may differ statistically from one another. In this case, the time average will depend on the particular sample function which is used to form the average. When the nature of a random process is such that ensemble and time averages are identical, the process is referred to as *ergodic*. An ergodic process is stationary, but, of course, a stationary process is not necessarily ergodic.

Throughout this text we shall assume that the random processes with which we shall have occasion to deal are ergodic. Hence the ensemble average $E\{n(t)\}$ is the same as the time average $\langle n(t) \rangle$, the ensemble average $E\{n^2(t)\}$ is the same as the time average $\langle n^2(t) \rangle$, etc.

**Example 2.23-1** Consider the random process

$$V(t) = \cos(\omega_0 t + \Theta) \tag{2.23-1}$$

where $\Theta$ is a random variable with a probability density

$$f(\theta) = \frac{1}{2\pi} \qquad -\pi \le \theta \le \pi$$
$$= 0 \qquad \text{elsewhere} \tag{2.23-2}$$

(*a*) Show that the first and second moments of $V(t)$ are independent of time.
(*b*) If the random variable $\Theta$ in Eq. (2.23-1) is replaced by a fixed angle $\theta_0$, will the ensemble mean of $V(t)$ be time independent?

SOLUTION (*a*) Choose a fixed time $t = t_1$. Then

$$E\{V(t_1)\} = \int_{-\pi}^{\pi} \frac{1}{2\pi} \cos(\omega_0 t_1 + \theta)\, d\theta = 0 \tag{2.23-3}$$

and

$$E\{V^2(t_1)\} = \int_{-\pi}^{\pi} \frac{1}{2\pi} \cos^2(\omega_0 t_1 + \theta)\, d\theta = \frac{1}{2} \tag{2.23-4}$$

We note that these moments are independent of $t_1$ and hence independent of time.



In a similar manner it can be established that all of the moments and all other statistical characteristics of $V(t)$ are independent of time. *Hence $V(t)$ is a stationary process.*
(*b*) Since $\theta$ is known, $V(t)$ is deterministic. For example, if $\theta = 30$

$$E\{V(t) = V(t) = \cos(\omega_0 t + 30°) \neq \text{constant} \tag{2.23-5}$$

Thus, $V(t)$ is not stationary.

**Example 2.23-2** A voltage $V(t)$, which is a gaussian ergodic random process with a mean of zero and a variance of 4 volt$^2$, is measured by a dc meter, a true rms meter, and a meter which first squares $V(t)$ and then reads its dc component.

Find the output of each meter.

SOLUTION (*a*) The dc meter reads

$$\langle V(t) \rangle = E\{V(t)\}$$

since $V(t)$ is ergodic. Since $E\{V(t)\} = 0$, the dc meter reads zero.
(*b*) The true rms meter reads

$$\sqrt{\langle V^2(t) \rangle} = \sqrt{E\{V^2(t)\}}$$

since $V(t)$ is ergodic. Since $V(t)$ has a zero mean, the true rms meter reads $\sigma = 2$ volts.
(*c*) The square and average meter (a full-wave rectifier meter) yields a deflection proportional to

$$\langle v^2(t) \rangle = E\{V^2(t)\} = \sigma^2 = 4$$

## 2.24 AUTOCORRELATION

A random process $n(t)$, being neither periodic nor of finite energy has an autocorrelation function defined by Eq. (1.18-1). Thus

$$R(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} n(t) n(t + \tau)\, dt \tag{2.24-1}$$

In connection with deterministic waveforms we were able to give a physical significance to the concept of a power spectral density $G(f)$ and to show that $G(f)$ and $R(\tau)$ constitute a Fourier transform pair. As an extension of that result we shall *define* the power spectral density of a random process in the same way. Thus for a random process we take $G(f)$ to be

$$G(f) = \mathcal{F}[R(\tau)] = \int_{-\infty}^{\infty} R(\tau) e^{-j\omega\tau}\, d\tau \tag{2.24-2}$$

It is of interest to inquire whether $G(f)$ defined in Eq. (2.24-2) for a random process has a physical significance which corresponds to the physical significance of $G(f)$ for deterministic waveforms.

For this purpose consider a *deterministic* waveform $v(t)$ which extends from $-\infty$ to $\infty$. Let us select a section of this waveform which extends from $-T/2$ to $T/2$. This waveform $v_T(t) = v(t)$ in this range, and otherwise $v_T(t) = 0$. The waveform $v_T(t)$ has a Fourier transform $V_T(f)$. We recall that $|V_T(f)|^2$ is the energy spectral density; that is, $|V_T(f)|^2\, df$ is the normalized energy in the spectral range $df$. Hence, over the interval $T$ the normalized power density is $|V_T(f)|^2/T$. As $T \to \infty$, $v_T(t) \to v(t)$, and we then have the result that the physical significance of the power spectral density $G(f)$, at least for a deterministic waveform, is that

$$G(f) = \lim_{T\to\infty} \frac{1}{T} |V_T(f)|^2 \tag{2.24-3}$$

Correspondingly, we state, without proof, that when $G(f)$ is defined for a random process, as in Eq. (2.24-2), as the transform of $R(\tau)$, then $G(f)$ has the significance that

$$G(f) = \lim_{T\to\infty} E\left\{\frac{1}{T} |N_T(f)|^2\right\} \tag{2.24-4}$$

where $E\{\ \}$ represents the ensemble average or expectation and $N_T(f)$ represents the Fourier transform of a truncated section of a sample function of the random process $n(t)$.

The autocorrelation function $R(\tau)$ is, as indicated in Eq. (2.24-1), a *time average* of the product $n(t)$ and $n(t + \tau)$. Since we have assumed an ergodic process, we are at liberty to perform the averaging over any sample function of the ensemble, since every sample function will yield the same result. However, again because the noise process is ergodic, we may replace the time average by an ensemble average and write, instead of Eq. (2.24-1),

$$R(\tau) = E\{n(t)n(t + \tau)\} \tag{2.24-5}$$

The averaging indicated in Eq. (2.24-5) has the following significance: At some *fixed* time $t$, $n(t)$ is a random variable, the possible values for which are the values $n(t)$ assumed at time $t$ by the individual sample functions of the ensemble. Similarly, at the *fixed* time $t + \tau$, $n(t + \tau)$ is also a random variable. It then appears that $R(\tau)$ as expressed in Eq. (2.24-5) is the covariance between these two random variables.

Suppose then that we should find that for some $\tau$, $R(\tau) = 0$. Then the random variables $n(t)$ and $n(t + \tau)$ are uncorrelated, and for the gaussian process of interest to us, $n(t)$ and $n(t + \tau)$ are independent. Hence, if we should select some sample function, a knowledge of the value of $n(t)$ at time $t$ would be of no assistance in improving our ability to predict the value attained by that same sample function at time $t + \tau$.

The physical fact about the noise, which is of principal concern in connection with communications systems, is that such noise has a power spectral density







$G(f)$ which is uniform over all frequencies. Such noise is referred to as "white" noise in analogy with the consideration that white light is a combination of all colors, that is, colors of all frequencies. Actually as is pointed out in Sec. 14.5, there is an upper-frequency limit beyond which the spectral density falls off sharply. However, this upper-frequency limit is so high that we may ignore it for our purposes.

Now, since the autocorrelation $R(\tau)$ and the power spectral density $G(f)$ are a Fourier transform pair, they have the properties of such pairs. Thus when $G(f)$ extends over a wide frequency range, $R(\tau)$ is restricted to a narrow range of $\tau$. In the limit, if $G(f) = I$ (a constant) for all frequencies from $-\infty \le f \le +\infty$, then $R(\tau)$ becomes $R(\tau) = I\,\delta(\tau)$, where $\delta(\tau)$ is the delta function with $\delta(\tau) = 0$ except for $\tau = 0$. Since, then, for white noise, $R(\tau) = 0$ except for $\tau = 0$, Eq. (2.24-5) says that $n(t)$ and $n(t + \tau)$ are uncorrelated and hence independent, no matter how small $\tau$.

## 2.25 POWER SPECTRAL DENSITY OF A SEQUENCE OF RANDOM PULSES

We shall occasionally need to have information about the power spectral density of a sequence of random pulses such as is indicated in Fig. 2.25-1. The pulses are of the same form but have random amplitudes and statistically independent random times of occurrence. The waveform (the random process) is stationary so that the statistical features of the waveforms are time invariant. Correspondingly, there is an invariant *average* time of separation $T_s$ between pulses. We further assume that there is no overlap between pulses.

If the Fourier transform of a single sample pulse $P_1(t)$ is $P_1(f)$ then Parseval's theorem [Eq. (1.13-5)] states that the normalized energy of the pulse is

$$E_1 = \int_{-\infty}^{\infty} P_1(f)P_1^*(f)\, df = \int_{-\infty}^{\infty} |P_1(f)|^2\, df \tag{2.25-1}$$

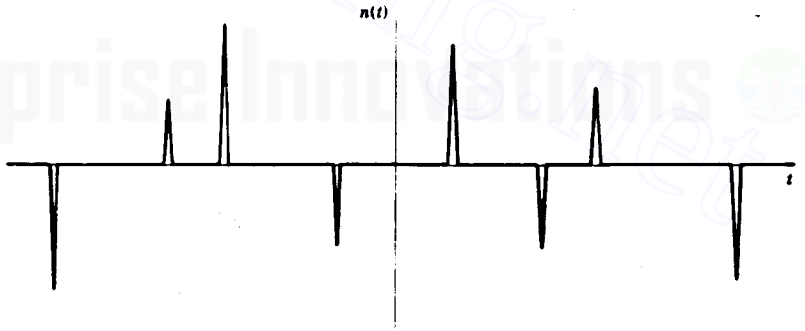



**Figure 2.25-1** Pulses of random amplitude and time of occurrence.

## PROBLEMS

**2.1-1.** Six dice are thrown simultaneously. What is the probability that at least 1 die shows a 3?

**2.1-2.** A card is drawn from a deck of 52 cards.
(*a*) What is the probability that a 2 is drawn?
(*b*) What is the probability that a 2 of clubs is drawn?
(*c*) What is the probability that a spade is drawn?

**2.2-1.** A card is picked from each of four 52-card decks of cards.
(*a*) What is the probability of selecting at least one 6 of spades?
(*b*) What is the probability of selecting at least 1 card larger than an 8?

**2.3-1.** A card is drawn from a 52-card deck, and without replacing the first card a second card is drawn. The first and second cards are not replaced and a third card is drawn.
(*a*) If the first card is a heart, what is the probability of the second card being a heart?
(*b*) If the first and second cards are hearts, what is the probability that the third card is the king of clubs?

**2.3-2.** Two factories produce identical clocks. The production of the first factory consists of 10,000 clocks of which 100 are defective. The second factory produces 20,000 clocks of which 300 are defective. What is the probability that a particular defective clock was produced in the first factory?

**2.3-3.** One box contains two black balls. A second box contains one black and one white ball. We are told that a ball was withdrawn from one of the boxes and that it turned out to be black. What is the probability that this withdrawal was made from the box that held the two black balls?

**2.4-1.** Two dice are tossed.
(*a*) Find the probability of a 3 and a 4 appearing.
(*b*) Find the probability of a 7 being rolled.

**2.4-2.** A card is drawn from a deck of 52 cards, then replaced, and a second card drawn.
(*a*) What is the probability of drawing the same card twice?
(*b*) What is the probability of first drawing a 3 of hearts and then drawing a 4 of spades?

**2.6-1.** A die is tossed and the number appearing is $n_i$. Let $N$ be the random variable identifying the outcome of the toss defined by the specifications $N = n_i$ when $n_i$ appears. Make a plot of the probability $P(N \le n_i)$ as a function of $n_i$.

**2.6-2.** A coin is tossed four times. Let $H$ be the random variable which identifies the number of heads which occur in these four tosses. It is defined by $H = h$, where $h$ is the number of heads which appear. Make a plot of the probability $P(H \le h)$ as a function of $h$.

**2.6-3.** A coin is tossed until a head appears. Let $T$ be the random variable which identifies the number of tosses $t$ required for the appearance of this first head. Make a plot of the probability $P(T \le t)$ as a function of $t$ up to $t = 5$.

**2.7-1.** An important probability density function is the Rayleigh density

$$f(x) = \begin{cases} xe^{-x^2/2} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

(*a*) Prove that $f(x)$ satisfies Eqs. (2.7-2) and (2.7-3).
(*b*) Find the distribution function $F(x)$.

**2.8-1.** Refer to Fig. 2.6-1.
(*a*) Find $P(2 < n \le 11)$.
(*b*) Find $P(2 \le n < 11)$.
(*c*) Find $P(2 \le n \le 11)$.
(*d*) Find $F(9)$.

**2.8-2.** Refer to the Rayleigh density function given in Prob. 2.7-1. Find the probability $P(x_1 < x \le x_2)$, where $x_2 - x_1 = 1$, so that $P(x_1 < x \le x_2)$ is a maximum. *Hint:* Find $P(x_1 < x \le x_2)$, replace $x_2$ by $1 + x_1$, and maximize $P$ with respect to $x_1$.







**2.8-3.** Refer to the Rayleigh density function given in Prob. 2.7-1. Find
(*a*) $P(0.5 < x \le 2)$.
(*b*) $P(0.5 \le x < 2)$.

**2.9-1.** The joint probability density of the random variables $X$ and $Y$ is $f(x, y) = ke^{-(x+y)}$ in the range $0 \le x \le \infty, 0 \le y \le \infty$, and $f(x, y) = 0$ otherwise.
(*a*) Find the value of the constant $k$.
(*b*) Find the probability density $f(x)$, the probability density of $X$ independently of $Y$.
(*c*) Find the probability $P(0 \le X \le 2; 2 \le Y \le 3)$.
(*d*) Are the random variables dependent or independent?

**2.9-2.** $X$ is a random variable having a gaussian density. $E(X) = 0$, $\sigma_x^2 = 1$. $V$ is a random variable having the values 1 or $-1$, each with probability $\frac{1}{2}$.
(*a*) Find the joint density $f_{X,V}(x, v)$.
(*b*) Show that $f_V(v) = \int_{-\infty}^{\infty} f_{XV}(x, v)\, dx$.

**2.9-3.** The joint probability density of the random variables $X$ and $Y$ is $f(x, y) = xe^{-x(y+1)}$ in the range $0 \le x \le \infty, 0 \le y \le \infty$, and $f(x, y) = 0$ otherwise.
(*a*) Find $f(x)$ and $f(y)$, the probability density of $X$ independently of $Y$ and $Y$ independently of $X$.
(*b*) Are the random variables dependent or independent?

**2.10-1.** (*a*) In a communication channel as represented in Fig. 2.10-1 $P(r_0 \mid m_0) = 0.9$, $P(r_1 \mid m_0) = 0.1$, $P(r_0 \mid m_1) = 0.4$, and $P(r_1 \mid m_1) = 0.6$. On a single set of coordinate axes make plots of $P(m_0 \mid r_0)$, $P(m_1 \mid r_0)$, $P(m_0 \mid r_1)$ and $P(m_1 \mid r_1)$ as a function of $P(m_0)$. Mark the range of $m_0$ for which the algorithm of Eq. (2.10-1) prescribes that we choose $m_0$ if $r_0$ is received and $m_1$ if $r_1$ is received, the range for which we choose $m_0$ no matter what is received and the range for which we choose $m_1$ no matter what is received.
(*b*) Calculate and plot the probability of error as a function of $P(m_0)$.

**2.10-2.** (*a*) For the channel and message probabilities given in Fig. P2.10-2 determine the best decisions about the transmitted message for each possible received response.
(*b*) With decisions made as in part (*a*) calculate the probability of error.
(*c*) Suppose the decision-making apparatus at the receiver were inoperative so that at the receiver nothing could be determined except that a message had been received. What would be the best strategy for determining what message had been transmitted and what would be the corresponding error probability?

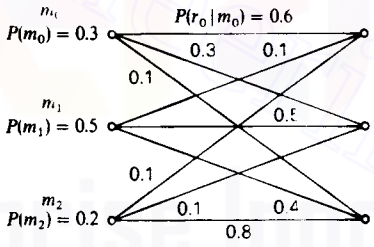



Figure P2.10-2

**2.11-1.** If $f_X(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for all $x$, show that
(*a*) $E(X^{2n}) = 1 \cdot 3 \cdot 5 \cdots (n-1)$, $n = 1, 2, \ldots$.
(*b*) $E(X^{2n-1}) = 0$, $n = 1, 2, \ldots$.

**2.11-2.** Compare the most probable [$f(x)$ is a maximum] and the average value of $X$ when

(*a*) $f_{X_1}(x) = \dfrac{1}{\sqrt{2\pi}} e^{-(x-m)^2/2}$ for all $x$

(*b*) $f_{X_2}(x) = \begin{cases} xe^{-x^2/2} & \text{for } x \ge 0 \\ 0 & \text{elsewhere} \end{cases}$

**2.12-1.** Calculate the variance of the random variables having densities:

(*a*) The gaussian density $f_{X_1}(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-m)^2/2}$, all $x$.

(*b*) The Rayleigh density $f_{X_2}(x) = xe^{-x^2/2}$, $x \geq 0$.

(*c*) The uniform density $f_{X_3}(x) = 1/a$, $-a/2 \leq x \leq a/2$.

**2.12-2.** Consider the Cauchy density function

$$f(x) = \frac{K}{1 + x^2} \qquad -\infty \leq x \leq \infty$$

(*a*) Find $K$ so that $f(x)$ is a density function.

(*b*) Find $E(X)$.

(*c*) Find the variance of $X$. Comment on the significance of this result.

**2.12-3.** The random variable $X$ has a variance $\sigma^2$ and a mean $m$. The random variable $Y$ is related to $X$ by $Y = aX + b$, where $a$ and $b$ are constants. Find the mean and variance of $Y$.

**2.13-1.** A probability density function $f(x)$ is uniform over the range from $-L$ to $+L$.

(*a*) Calculate and plot the probability $P[|x| \leq \varepsilon]$ as a function of $\varepsilon$.

(*b*) Calculate and plot the quantity $\sigma^2/\varepsilon^2$ and verify that Tchebycheff's rule is valid in this case.

**2.14-1.** Refer to the gaussian density given in Eq. (2.14-1).

(*a*) Show that $E((X - m)^{2n-1}) = 0$.

(*b*) Show that $E((X - m)^{2n}) = 1 \cdot 3 \cdot 5 \cdots (n - 1)\sigma^{2n}$.

**2.14-2.** Given a gaussian probability function $f(x)$ of mean value zero and variance $\sigma^2$.

(*a*) As a function of $\sigma$, plot the probability $P[|x| \geq \varepsilon]$.

(*b*) Calculate and plot the quantity $\sigma^2/\varepsilon^2$ and verify that Tchebycheff's rule is valid in this case.

**2.14-3.** A random variable $V = b + X$, where $X$ is a gaussian distributed random variable with mean 0 and variance $\sigma^2$, and $b$ is a constant. Show that $V$ is a gaussian distributed random variable with mean $b$ and variance $\sigma^2$.

**2.14-4.** The joint density function of two dependent variables $X$ and $Y$ is

$$f(x, y) = \frac{1}{\pi\sqrt{3}} e^{-2(x^2 - xy + y^2)/3}$$

(*a*) Show that, when $X$ and $Y$ are each considered without reference to the other, each is a gaussian variable, i.e., $f(x)$ and $f(y)$ are gaussian density functions.

(*b*) Find $\sigma_x^2$ and $\sigma_y^2$.

**2.15-1.** Obtain values for and plot erf $u$ versus $u$.

**2.15-2.** On the same set of axes used in Prob. 2.15-1 plot $e^{-u^2}$ and erfc $u$ versus $u$. Compare your results.

**2.15-3.** The probability

$$P_{\pm k\sigma} \equiv P(m - k\sigma \leq X \leq m + k\sigma) = \int_{m-k\sigma}^{m+k\sigma} \frac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma}\, dx$$

(*a*) Change variables by letting $u = x - m/\sqrt{2}\,\sigma$.

(*b*) Show that $P_{\pm k\sigma} = \text{erf}\,(k/\sqrt{2})$.

**2.16-1.** Show that a random variable with a Rayleigh density as in Eq. (2.16-1) has a mean value $R = \sqrt{(\pi/2)}\,\alpha$, a mean square value $R^2 = 2\alpha^2$, and a variance $\sigma^2 = (2 - \pi/2)\alpha^2$.

**2.16-2.** (*a*) A voltage $V$ is a function of time $t$ and is given by

$$V(t) = X \cos \omega t + Y \sin \omega t$$

in which $\omega$ is a constant angular frequency and $X$ and $Y$ are independent gaussian variables each with zero mean and variance $\sigma^2$. Show that $V(t)$ may be written

$$V(t) = R \cos(\omega t + \Theta)$$



in which $R$ is a random variable with a Rayleigh probability density and $\Theta$ is a random variable with uniform density.

(*b*) If $\sigma^2 = 1$, what is the probability that $R \geq 1$?

**2.17-1.** Derive Eq. (2.17-6) directly from the definition $\sigma_z^2 = E\{(Z - m_z)^2\}$.

**2.17-2.** $Z = X_1 + X_2 + \cdots + X_N$, $E(X_i) = m$.

(*a*) Find $E(Z)$.

(*b*) If $E(X_i X_j) = \begin{cases} 1 & j = i \\ \rho & j = i \pm 1 \\ 0 & \text{otherwise} \end{cases}$

find (1) $E(Z^2)$ and (2) $\sigma_z^2$.

**2.18-1.** The independent random variables $X$ and $Y$ are added to form $Z$. If

$$f_X(x) = xe^{-x^2/2} \qquad 0 \leq x \leq \infty \qquad \text{and} \qquad f_Y(y) = \tfrac{1}{2}e^{-|y|} \qquad |y| < \infty$$

find $f_Z(z)$.

**2.18-2.** The independent random variables $X$ and $Y$ have the probability densities

$$f(x) = e^{-x} \qquad 0 \leq x \leq \infty$$
$$f(y) = 2e^{-2y} \qquad 0 \leq y \leq \infty$$

Find and plot the probability density of the variable $Z = X + Y$.

**2.18-3.** The random variable $X$ has a probability density uniform in the range $0 \leq x \leq 1$ and zero elsewhere. The independent variable $Y$ has a density uniform in the range $0 \leq x \leq 2$ and zero elsewhere. Find and plot the density of $Z = X + Y$.

**2.18-4.** The $N$ independent gaussian random variables $X_1, \ldots, X_N$ are added to form $Z$. If the mean of $X_i$ is 1 and its variance is 1, find $f_Z(z)$.

**2.19-1.** Two gaussian random variables $X$ and $Y$, each with mean zero and variance $\sigma^2$, between which there is a correlation coefficient $\rho$, have a joint probability density given by

$$f(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1 - \rho^2}} \exp -\left[\frac{x^2 - 2\rho xy + y^2}{2\sigma^2(1 - \rho^2)}\right]$$

(*a*) Verify that the symbol $\rho$ in the expression for $f(x, y)$ is indeed the correlation coefficient. That is, evaluate $E\{XY\}/\sigma^2$ and show that the result is $\rho$ as required by Eq. (2.19-5).

(*b*) Show that the case $\rho = 0$ corresponds to the circumstance where $X$ and $Y$ are independent.

**2.19-2.** The random variables $X$ and $Y$ are related to the random variable $\Theta$ by $X = \sin \Theta$ and $Y = \cos \Theta$. The variable $\Theta$ has a uniform probability density in the range from 0 to $2\pi$. Show that $X$ and $Y$ are not independent but that, nonetheless, $E(XY) = 0$ so that they are uncorrelated.

**2.19-3.** The random variables $X_1, X_2, X_3, \ldots$ are dependent but uncorrelated. $Z = X_1 + X_2 + X_3 + \cdots$. Show that $\sigma_Z^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \cdots$.

**2.20-1.** The random variables $X_1, X_2, X_3$ are independent and each has a uniform probability density in the range $0 \leq x \leq 1$. Find and plot the probability density of $X_1 + X_2$ and of $X_1 + X_2 + X_3$.

**2.21-1.** Verify Eq. (2.21-16).

**2.21-2.** In a communication system used to transmit a sequence of messages, it is known that the error probability is $4 \times 10^{-5}$. A sample survey of $N$ messages is made. It is required that, with a probability not to exceed 0.05, the error rate in the sample is not to be larger than $5 \times 10^{-5}$. How many messages must be included in the sample?

**2.22-1.** A communication channel transmits, in random order, two messages $m_1$ and $m_2$. The message $m_1$ occurs three times more frequently than $m_2$. Message $m_1$ generates a receiver response $r_1 = -1V$ while $m_2$ generates $r_2 = +1V$. The channel is corrupted by noise $n$ with a uniform probability density which extends from $n = -1.5V$ to $n = +1.5V$.

(*a*) Find the probability that $m_1$ is mistaken for $m_2$ and the probability that $m_2$ is mistaken for $m_1$.

(*b*) What is the probability that the receiver will determine a message correctly?


Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

**2.22-2.** A communication channel transmits, in random order, two messages $m_1$ and $m_2$ with equal likelihood. Message $m_1$ generates response $r_1 = -1V$ at the receiver and message $m_2$ generates $r_2 = +1V$. The channel is corrupted with gaussian noise with variance $\sigma^2 = 1$ volt$^2$. Find the probability that the receiver will determine a message correctly.

**2.22-3.** A communication channel transmits, in random order, two messages $m_1$ and $m_2$. The message $m_1$ occurs three times more frequently than $m_2$. Message $m_1$ generates a receiver response $r_1 = +1V$ while $m_2$ generates $r_2 = -1V$. The channel is corrupted by noise $n$ whose probability density has the triangular form shown in Fig. 2.22-2*a* with $f_N(n) = 0$ at $f_N(n) = -2V$ and at $f_N(n) = +2V$.

(*a*) What are the ranges of $r$ for which the decision is to be made that $m_1$ was transmitted and for which the decision is to be made that $m_2$ was transmitted?

(*b*) What is the probability that the message will be read correctly?

**2.23-1.** The function of time $Z(t) = X_1 \cos \omega_0 t - X_2 \sin \omega_0 t$ is a random process. If $X_1$ and $X_2$ are independent gaussian random variables each with zero mean and variance $\sigma^2$, find

(*a*) $E(Z)$, $E(Z^2)$, $\sigma_z^2$, and

(*b*) $f_Z(z)$.

**2.23-2.** $Z(t) = M(t) \cos(\omega_0 t + \Theta)$. $M(t)$ is a random process, with $E(M(t)) = 0$ and $E(M^2(t)) = M_0^2$.

(*a*) If $\Theta = 0$, find $E(Z^2)$. Is $Z(t)$ stationary?

(*b*) If $\Theta$ is an independent random variable such that $f_\Theta(\theta) = 1/2\pi$, $-\pi \le \theta \le \pi$, show that $E(Z^2(t)) = E(M^2(t))E(\cos^2(\omega_0 t + \Theta)) = M_0^2/2$. Is $Z(t)$ now stationary?

**2.24-1.** Refer to Prob. 2.23-1. Find $R_z(\tau)$.

**2.24-2.** A random process $n(t)$ has a power spectral density $G(f) = \eta/2$ for $-\infty \le f \le \infty$. The random process is passed through a low-pass filter which has a transfer function $H(f) = 2$ for $-f_M \le f \le f_M$ and $H(f) = 0$ otherwise. Find the power spectral density of the waveform at the output of the filter.

**2.24-3.** White noise $n(t)$ with $G(f) = \eta/2$ is passed through a low-pass $RC$ network with a 3-dB frequency $f_c$.

(*a*) Find the autocorrelation $R(\tau)$ of the output noise of the network.

(*b*) Sketch $\rho(\tau) = R(\tau)/R(0)$.

(*c*) Find $\omega_c \tau$ such that $\rho(\tau) \le 0.1$.

**2.25-1.** Consider a train of rectangular pulses. The $k$th pulse has a width $\tau$ and a height $A_k$. $A_k$ is a random variable which can have the values 1, 2, 3, ... 10 with equal probability. Assuming statistical independence between amplitudes and assuming that the average separation between pulses is $T_s$, find the power spectral density $G_n(f)$ of the pulse train.

**2.25-2.** A pulse train consists of rectangular pulses having an amplitude of 2 volts and widths which are either 1 $\mu$s or 2 $\mu$s with equal probability. The mean time between pulses is 5 $\mu$s. Find the power spectral density $G_n(f)$ of the pulse train.

**2.26-1.** Consider the power spectral density of an NRZ waveform as given by Eq. (2.26-4) and as shown in Fig. 2.26-3. By consulting a table of integrals, show that the power of the NRZ waveform is reduced by only 10 percent if the waveform is passed through an ideal low-pass filter with cutoff at $f = f_b$.

**2.26-2.** Consider the power spectral density of a biphase waveform as given by Eq. (2.26-7) and as shown in Fig. 2.26-4. By consulting a table of integrals, show that, if the waveform is passed through an ideal low-pass filter with cutoff at $2f_b$, 95 percent of the power will be passed. Show also that, if the filter cutoff is set at $f_b$, then only 70 percent of the power is transmitted.

**2.27-1.** An NRZ waveform consists of alternating 0's and 1's. The bit interval is 1 $\mu$s and the waveform makes excursions between $+1V$ and $-1V$. The waveform is transmitted through an $RC$ high-pass filter of time constant 1 $\mu$s. Draw the output waveform and calculate numerical values of all details of the waveform.

**2.27-2.** An NRZ waveform consists of alternating 0's and 1's. The bit interval is 1 $\mu$s and the waveform makes excursions between $+1V$ and $-1V$. The waveform is transmitted through an $RC$ low-pass filter of time constant 1 $\mu$s. Draw the output waveform and calculate numerical values of all details of the waveform.





CHAPTER

# THREE

# AMPLITUDE-MODULATION SYSTEMS

One of the basic problems of communication engineering is the design and analysis of systems which allow many individual messages to be transmitted simultaneously over a single communication channel. A method by which such multiple transmission, called *multiplexing*, may be achieved consists in translating each message to a different position in the *frequency* spectrum. Such multiplexing is called *frequency multiplexing*. The individual message can eventually be separated by filtering. Frequency multiplexing involves the use of an auxiliary waveform, usually sinusoidal, called a *carrier*. The operations performed on the signal to achieve frequency multiplexing result in the generation of a waveform which may be described as the carrier modified in that its amplitude, frequency, or phase, individually or in combination, varies with time. Such a modified carrier is called a *modulated* carrier. In some cases the modulation is related simply to the message; in other cases the relationship is quite complicated. In this chapter, we discuss the generation and characteristics of amplitude-modulated carrier waveforms.

## 3.1 FREQUENCY TRANSLATION

It is often advantageous and convenient, in processing a signal in a communications system, to translate the signal from one region in the frequency domain to another region. Suppose that a signal is bandlimited, or nearly so, to the frequency range extending from a frequency $f_1$ to a frequency $f_2$. The process of frequency translation is one in which the original signal is replaced with a new

2-2. A communication channel transmits, in random order, two messages $m_1$ and $m_2$ with equal elihood. Message $m_1$ generates response $r_1 = -1V$ at the receiver and message $m_2$ generates $r_2 = 1V$. The channel is corrupted with gaussian noise with variance $\sigma^2 = 1$ volt$^2$. Find the probability t the receiver will determine a message correctly.

2-3. A communication channel transmits, in random order, two messages $m_1$ and $m_2$. The message occurs three times more frequently than $m_2$. Message $m_1$ generates a receiver response $r_1 = +1V$ ile $m_2$ generates $r_2 = -1V$. The channel is corrupted by noise $n$ whose probability density has the ngular form shown in Fig. 2.22-2*a* with $f_N(n) = 0$ at $f_N(n) = -2V$ and at $f_N(n) = +2V$.

(*a*) What are the ranges of $r$ for which the decision is to be made that $m_1$ was transmitted and which the decision is to be made that $m_2$ was transmitted?

(*b*) What is the probability that the message will be read correctly?

3-1. The function of time $Z(t) = X_1 \cos \omega_0 t - X_2 \sin \omega_0 t$ is a random process. If $X_1$ and $X_2$ are ependent gaussian random variables each with zero mean and variance $\sigma^2$, find

(*a*) $E(Z)$, $E(Z^2)$, $\sigma_z^2$, and

(*b*) $f_z(z)$.

3-2. $Z(t) = M(t) \cos(\omega_0 t + \Theta)$. $M(t)$ is a random process, with $E(M(t)) = 0$ and $E(M^2(t)) = M_0^2$.

(*a*) If $\Theta = 0$, find $E(Z^2)$. Is $Z(t)$ stationary?

(*b*) If $\Theta$ is an independent random variable such that $f_\Theta(\theta) = 1/2\pi$, $-\pi \le \theta \le \pi$, show that $R_z(t) = E[M^2(t)]E[\cos^2(\omega_0 t + \Theta)] = M_0^2/2$. Is $Z(t)$ now stationary?

3-11. Refer to Prob. 2.23-1. Find $R_z(\tau)$.

3-12. A random process $n(t)$ has a power spectral density $G(f) = \eta/2$ for $-\infty \le f \le \infty$. The random cess is passed through a low-pass filter which has a transfer function $H(f) = 2$ for $-f_M \le f \le f_M$ $H(f) = 0$ otherwise. Find the power spectral density of the waveform at the output of the filter.

3-13. White noise $n(t)$ with $G(f) = \eta/2$ is passed through a low-pass $RC$ network with a 3-dB frequency $f_c$.

(*a*) Find the autocorrelation $R(\tau)$ of the output noise of the network.

(*b*) Sketch $\rho(\tau) = R(\tau)/R(0)$.

(*c*) Find $\omega_c \tau$ such that $\rho(\tau) \le 0.1$.

3-14. Consider a train of rectangular pulses. The $k$th pulse has a width $\tau$ and a height $A_k$. $A_k$ is a dom variable which can have the values 1, 2, 3, ... 10 with equal probability. Assuming statistical ependence between amplitudes and assuming that the average separation between pulses is $T_s$, find power spectral density $G_s(f)$ of the pulse train.

3-15. A pulse train consists of rectangular pulses having an amplitude of 2 volts and widths which either 1 $\mu$s or 2 $\mu$s with equal probability. The mean time between pulses is 5 $\mu$s. Find the power ctral density $G_s(f)$ of the pulse train.

3-16. Consider the power spectral density of an NRZ waveform as given by Eq. (2.26-4) and as wn in Fig. 2.26-3. By consulting a table of integrals, show that the power of the NRZ waveform is uced by only 10 percent if the waveform is passed through an ideal low-pass filter with cutoff at $f_b$.

3-17. Consider the power spectral density of a biphase waveform as given by Eq. (2.26-7) and as wn in Fig. 2.26-4. By consulting a table of integrals, show that, if the waveform is passed through deal low-pass filter with cutoff at $2f_b$, 95 percent of the power will be passed. Show also that, if the r cutoff is set at $f_b$, then only 70 percent of the power is transmitted.

3-1. An NRZ waveform consists of alternating 0's and 1's. The bit interval is 1 $\mu$s and the wave- n makes excursions between $+1V$ and $-1V$. The waveform is transmitted through an $RC$ high- s filter of time constant 1 $\mu$s. Draw the output waveform and calculate numerical values of all ils of the waveform.

3-2. An NRZ waveform consists of alternating 0's and 1's. The bit interval is 1 $\mu$s and the wave- n makes excursions between $+1V$ and $-1V$. The waveform is transmitted through an $RC$ low- s filter of time constant 1 $\mu$s. Draw the output waveform and calculate numerical values of all ils of the waveform.





CHAPTER

# THREE

# AMPLITUDE-MODULATION SYSTEMS

One of the basic problems of communication engineering is the design and analysis of systems which allow many individual messages to be transmitted simultaneously over a single communication channel. A method by which such multiple transmission, called *multiplexing*, may be achieved consists in translating each message to a different position in the *frequency* spectrum. Such multiplexing is called *frequency multiplexing*. The individual message can eventually be separated by filtering. Frequency multiplexing involves the use of an auxiliary waveform, usually sinusoidal, called a *carrier*. The operations performed on the signal to achieve frequency multiplexing result in the generation of a waveform which may be described as the carrier modified in that its amplitude, frequency, or phase, individually or in combination, varies with time. Such a modified carrier is called a *modulated* carrier. In some cases the modulation is related simply to the message; in other cases the relationship is quite complicated. In this chapter, we discuss the generation and characteristics of amplitude-modulated carrier waveforms.[1]

## 3.1 FREQUENCY TRANSLATION

It is often advantageous and convenient, in processing a signal in a communications system, to translate the signal from one region in the frequency domain to another region. Suppose that a signal is bandlimited, or nearly so, to the frequency range extending from a frequency $f_1$ to a frequency $f_2$. The process of frequency translation is one in which the original signal is replaced with a new

signal whose spectral range extends from $f'_1$ to $f'_2$ and which *new* signal bears, in recoverable form, the same *information* as was borne by the original signal. We discuss now a number of useful purposes which may be served by frequency translation.

### Frequency Multiplexing

Suppose that we have several different signals, all of which encompass the same spectral range. Let it be required that all these signals be transmitted along a single communications channel in such a manner that, at the receiving end, the signals be separately recoverable and distinguishable from each other. The single channel may be a single pair of wires or the free space that separates one radio antenna from another. Such multiple transmissions, i.e., multiplexing, may be achieved by translating each one of the original signals to a different frequency range. Suppose, say, that one signal is translated to the frequency range $f'_1$ to $f'_2$, the second to the range $f''_1$ to $f''_2$, and so on. If these new frequency ranges do not overlap, then the signal may be separated at the receiving end by appropriate bandpass filters, and the outputs of the filters processed to recover the original signals.

### Practicability of Antennas

When free space is the communications channel, antennas radiate and receive the signal. It turns out that antennas operate effectively only when their dimensions are of the order of magnitude of the wavelength of the signal being transmitted. A signal of frequency 1 kHz (an audio tone) corresponds to a wavelength of 300,000 m, an entirely impractical length. The required length may be reduced to the point of practicability by translating the audio tone to a higher frequency.

### Narrowbanding

Returning to the matter of the antenna, just discussed, suppose that we wanted to transmit an audio signal directly from the antenna, and that the inordinate length of the antenna were no problem. We would still be left with a problem of another type. Let us assume that the audio range extends from, say, 50 to $10^4$ Hz. The ratio of the highest audio frequency to the lowest is 200. Therefore, an antenna suitable for use at one end of the range would be entirely too short or too long for the other end. Suppose, however, that the audio spectrum were translated so that it occupied the range, say, from $(10^6 + 50)$ to $(10^6 + 10^4)$ Hz. Then the ratio of highest to lowest frequency would be only 1.01. Thus the processes of frequency translation may be used to change a "wideband" signal into a "narrowband" signal which may well be more conveniently processed. The terms "wideband" and "narrowband" are being used here to refer not to an absolute range of frequencies but rather to the fractional change in frequency from one band edge to the other.







### Common Processing

It may happen that we may have to process, in turn, a number of signals similar in general character but occupying different spectral ranges. It will then be necessary, as we go from signal to signal, to adjust the frequency range of our processing apparatus to correspond to the frequency range of the signal to be processed. If the processing apparatus is rather elaborate, it may well be wiser to leave the processing apparatus to operate in some fixed frequency range and instead to translate the frequency range of each signal in turn to correspond to this fixed frequency.

## 3.2 A METHOD OF FREQUENCY TRANSLATION

A signal may be translated to a new spectral range by *multiplying* the signal with an auxiliary sinusoidal signal. To illustrate the process, let us consider initially that the signal is sinusoidal in waveform and given by

$$v_m(t) = A_m \cos \omega_m t = A_m \cos 2\pi f_m t \tag{3.2-1a}$$

$$= \frac{A_m}{2}(e^{j\omega_m t} + e^{-j\omega_m t}) = \frac{A_m}{2}(e^{j2\pi f_m t} + e^{-j2\pi f_m t}) \tag{3.2-1b}$$

in which $A_m$ is the constant amplitude and $f_m = \omega_m/2\pi$ is the frequency. The two-sided spectral amplitude pattern of this signal is shown in Fig. 3.2-1*a*. The pattern consists of two lines, each of amplitude $A_m/2$, located at $f = f_m$ and at $f = -f_m$. Consider next the result of the multiplication of $v_m(t)$ with an auxiliary sinusoidal signal

$$v_c(t) = A_c \cos \omega_c t = A_c \cos 2\pi f_c t \tag{3.2-2a}$$

$$= \frac{A_c}{2}(e^{j\omega_c t} + e^{-j\omega_c t}) = \frac{A_c}{2}(e^{j2\pi f_c t} + e^{-j2\pi f_c t}) \tag{3.2-2b}$$

in which $A_c$ is the constant amplitude and $f_c$ is the frequency. Using the trigonometric identity $\cos \alpha \cos \beta = \frac{1}{2}\cos(\alpha + \beta) + \frac{1}{2}\cos(\alpha - \beta)$, we have for the product $v_m(t)v_c(t)$

$$v_m(t)v_c(t) = \frac{A_m A_c}{2}[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t] \tag{3.2-3a}$$

$$= \frac{A_m A_c}{4}(e^{j(\omega_c + \omega_m)t} + e^{-j(\omega_c + \omega_m)t} + e^{j(\omega_c - \omega_m)t} + e^{-j(\omega_c - \omega_m)t}) \tag{3.2-3b}$$

The new spectral amplitude pattern is shown in Fig. 3.2-1*b*. Observe that the two original spectral lines have been *translated*, both in the positive-frequency direction by amount $f_c$ and also in the negative-frequency direction by the same amount. There are now four spectral components resulting in two sinusoidal waveforms, one of frequency $f_c + f_m$ and the other of frequency $f_c - f_m$. Note that

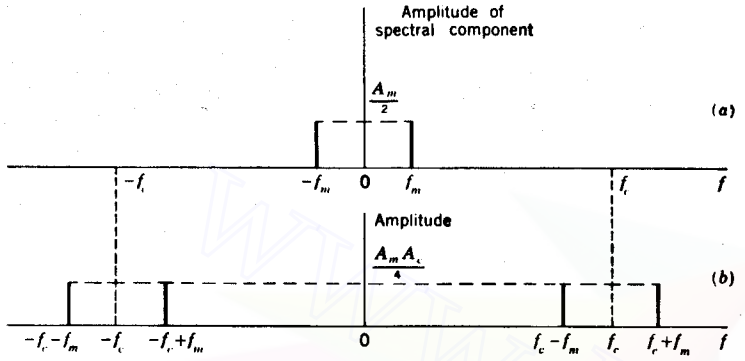



**Figure 3.2-1** (*a*) Spectral pattern of the waveform $A_m \cos \omega_m t$. (*b*) Spectral pattern of the product waveform $A_m A_c \cos \omega_m t \cos \omega_c t$.

while the product signal has four spectral components each of amplitude $A_m A_c/4$, there are only two frequencies, and the amplitude of each sinusoidal component is $A_m A_c/2$.

A generalization of Fig. 3.2-1 is shown in Fig. 3.2-2. Here a signal is chosen which consists of a superposition of four sinusoidal signals, the highest in frequency having the frequency $f_M$. Before translation by multiplication, the two-sided spectral pattern displays eight components centered around zero frequency. After multiplication, we find this spectral pattern translated both in the positive- and the negative-frequency directions. The 16 spectral components in this two-sided spectral pattern give rise to eight sinusoidal waveforms. While the original signal extends in range up to a frequency $f_M$, the signal which results from multiplication has sinusoidal components covering a range $2f_M$, from $f_c - f_M$ to $f_c + f_M$.

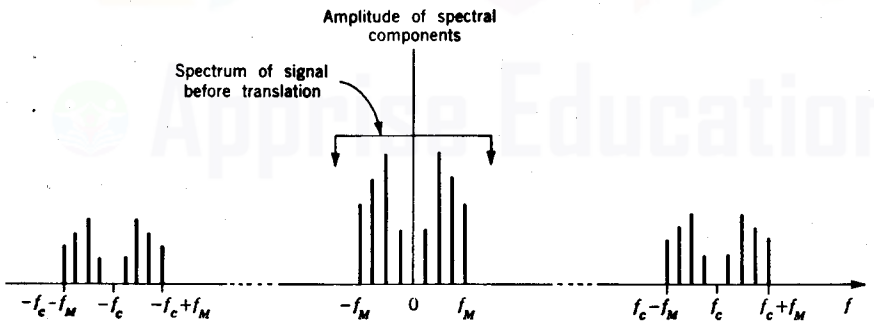



**Figure 3.2-2.** An original signal consisting of four sinusoids of differing frequencies is translated through multiplication and becomes a signal containing eight frequencies symmetrically arranged about $f_c$.



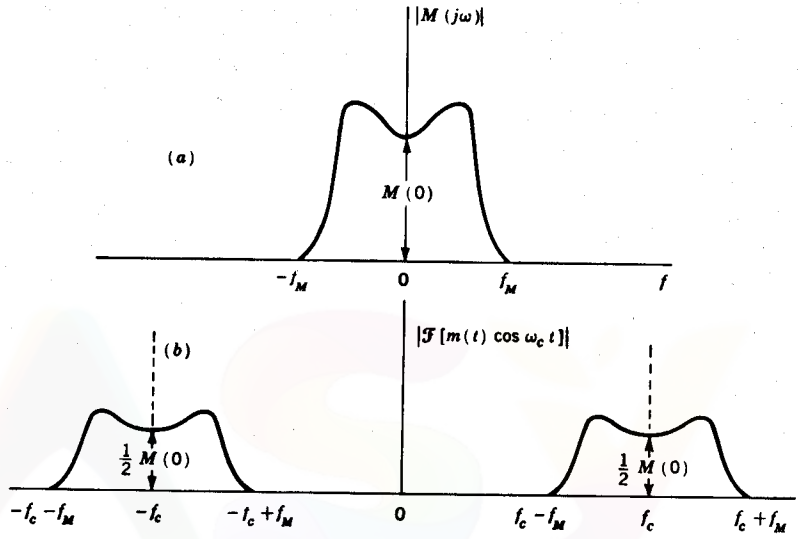



**Figure 3.2-3** (*a*) The spectral density $|M(j\omega)|$ of a nonperiodic signal $m(t)$. (*b*) The spectral density of $m(t) \cos 2\pi f_c t$.

Finally, we consider in Fig. 3.2-3 the situation in which the signal to be translated may not be represented as a superposition of a number of sinusoidal components at sharply defined frequencies. Such would be the case if the signal were of finite energy and nonperiodic. In this case the signal is represented in the frequency domain in terms of its Fourier transform, that is, in terms of its spectral density. Thus let the signal $m(t)$ be bandlimited to the frequency range 0 to $f_M$. Its Fourier transform is $M(j\omega) = \mathcal{F}[m(t)]$. The magnitude $|M(j\omega)|$ is shown in Fig. 3.2-3*a*. The transform $M(j\omega)$ is symmetrical about $f = 0$ since we assume that $m(t)$ is a real signal. The spectral density of the signal which results when $m(t)$ is multiplied by $\cos \omega_c t$ is shown in Fig. 3.2-3*b*. This spectral pattern is deduced as an extension of the results shown in Figs. 3.2-1 and 3.2-2. Alternatively, we may easily verify (Prob. 3.2-2) that if $M(j\omega) = \mathcal{F}[m(t)]$, then

$$\mathcal{F}[m(t) \cos \omega_c t] = \tfrac{1}{2}[M(j\omega + j\omega_c) + M(j\omega - j\omega_c)] \qquad (3.2\text{-}4)$$

The spectral range occupied by the original signal is called the *baseband frequency range* or simply the *baseband*. On this basis, the original signal itself is referred to as the baseband signal. The operation of multiplying a signal with an auxiliary sinusoidal signal is called *mixing* or *heterodyning*. In the translated signal, the part of the signal which consists of spectral components *above* the auxiliary signal, in the range $f_c$ to $f_c + f_M$, is called the *upper-sideband* signal. The part of the signal which consists of spectral components below the auxiliary signal, in the range $f_c - f_M$ to $f_c$, is called the *lower-sideband* signal. The two sideband signals are also referred to as the *sum* and the *difference* frequencies, respec-

tively. The auxiliary signal of frequency $f_c$ is variously referred to as the *local oscillator signal*, the *mixing signal*, the *heterodyning signal*, or as the *carrier signal*, depending on the application. The student will note, as the discussion proceeds, the various contexts is which the different terms are appropriate.

We may note that the process of translation by multiplication actually gives us something somewhat different from what was intended. Given a signal occupying a baseband, say, from zero to $f_M$, and an auxiliary signal $f_c$, it would often be entirely adequate to achieve a simple translation, giving us a signal occupying the range $f_c$ to $f_c + f_M$, that is, the upper sideband. We note, however, that translation by multiplication results in a signal that occupies the range $f_c - f_M$ to $f_c + f_M$. This feature of the process of translation by multiplication may, depending on the application, be a nuisance, a matter of indifference, or even an advantage. Hence, this feature of the process is, of itself, neither an advantage nor a disadvantage. It is, however, to be noted that there is no other operation so simple which will accomplish translation.

## 3.3 RECOVERY OF THE BASEBAND SIGNAL

Suppose a signal $m(t)$ has been translated out of its baseband through multiplication with $\cos \omega_c t$. How is the signal to be recovered? The recovery may be achieved by a reverse translation, which is accomplished simply by multiplying the translated signal with $\cos \omega_c t$. That such is the case may be seen by drawing spectral plots as in Fig. 3.2-2 or 3.2-3 and noting that the difference-frequency signal obtained by multiplying $m(t) \cos \omega_c t$ by $\cos \omega_c t$ is a signal whose spectral range is back at baseband. Alternatively, we may simply note that

$$[m(t) \cos \omega_c t] \cos \omega_c t = m(t) \cos^2 \omega_c t = m(t)(\tfrac{1}{2} + \tfrac{1}{2} \cos 2\omega_c t) \quad (3.3\text{-}1a)$$

$$= \frac{m(t)}{2} + \frac{m(t)}{2} \cos 2\omega_c t \quad (3.3\text{-}1b)$$

Thus, the baseband signal $m(t)$ reappears. We note, of course, that in addition to the recovered baseband signal there is a signal whose spectral range extends from $2f_c - f_M$ to $2f_c + f_M$. As a matter of practice, this latter signal need cause no difficulty. For most commonly $f_c \gg f_M$, and consequently the spectral range of this double-frequency signal and the baseband signal are widely separated. Therefore the double-frequency signal is easily removed by a low-pass filter.

This method of signal recovery, for all its simplicity, is beset by an important inconvenience when applied in a physical communication system. Suppose that the auxiliary signal used for recovery differs in phase from the auxiliary signal used in the initial translation. If this phase angle is $\theta$, then, as may be verified Prob. 3.3-1), the recovered baseband waveform will be proportional to $m(t) \cos \theta$. Therefore, unless it is possible to maintain $\theta = 0$, the signal strength at recovery will suffer. If it should happen that $\theta = \pi/2$, the signal will be lost entirely. Or consider, for example, that $\theta$ drifts back and forth with time. Then in this case the signal strength will wax and wane, in addition, possibly, to disappearing entirely from time to time.



Alternatively, suppose that the recovery auxiliary signal is not precisely at frequency $f_c$ but is instead at $f_c + \Delta f$. In this case we may verify (Prob. 3.3-2) that the recovered baseband signal will be proportional to $m(t) \cos 2\pi \Delta f t$, resulting in a signal which will wax and wane or even be entirely unacceptable if $\Delta f$ is comparable to, or larger than, the frequencies present in the baseband signal. This latter contingency is a distinct possibility in many an instance, since usually $f_c \gg f_M$ so that a small percentage change in $f_c$ will cause a $\Delta f$ which may be comparable or larger than $f_M$. In telephone or radio systems, an offset $\Delta f \leq 30$ Hz is deemed acceptable.

We note, therefore, that signal recovery using a second multiplication requires that there be available at the recovery point a signal which is precisely *synchronous* with the corresponding auxiliary signal at the point of the first multiplication. In such a *synchronous* or *coherent* system a *fixed* initial phase discrepancy is of no consequence since a simple phase shifter will correct the matter. Similarly it is not essential that the recovery auxiliary signal be sinusoidal (see Prob. 3.3-3). What is essential is that, in any time interval, the number of cycles executed by the two auxiliary-signal sources be the same. Of course, in a physical system, where some signal distortion is tolerable, some lack of synchronism may be allowed.

When the use of a common auxiliary signal is not feasible, it is necessary to resort to rather complicated means to provide a synchronous auxiliary signal at the location of the receiver. One commonly employed scheme is indicated in Fig. 3.3-1. To illustrate the operation of the synchronizer, we assume that the baseband signal is a sinusoidal $\cos \omega_m t$. The received signal is $s_i(t) = A \cos \omega_m t \cos \omega_c t$, with $A$ a constant amplitude. This signal $s_i(t)$ does not have a spectral component at the angular frequency $\omega_c$. The output of the squaring circuit is

$$s_i^2(t) = A^2 \cos^2 \omega_m t \cos^2 \omega_c t \quad (3.3\text{-}2a)$$

$$= A^2(\tfrac{1}{2} + \tfrac{1}{2} \cos 2\omega_m t)(\tfrac{1}{2} + \tfrac{1}{2} \cos 2\omega_c t) \quad (3.3\text{-}2b)$$

$$= \frac{A^2}{4}[1 + \tfrac{1}{2} \cos 2(\omega_c + \omega_m)t + \tfrac{1}{2} \cos 2(\omega_c - \omega_m)t + \cos 2\omega_m t + \cos 2\omega_c t] \quad (3.3\text{-}2c)$$

The filter selects the spectral component $(A^2/4) \cos 2\omega_c t$, which is then applied to a circuit which divides the frequency by a factor of 2. (See Prob. 3.3-4.) This fre-

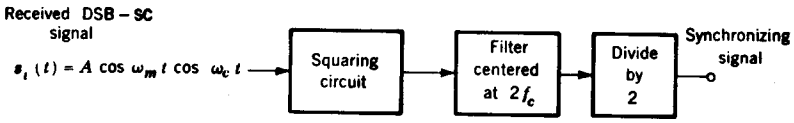



**Figure 3.3-1** A simple squaring synchronizer.

quency division may be accomplished by using, for example, a bistable multivibrator. The output of the divider is used to demodulate (multiply) the incoming signal and thereby recover the baseband signal cos $\omega_m t$.

We turn our attention now to a modification of the method of frequency translation, which has the great merit of allowing recovery of the baseband signal by an extremely simple means. This technique is called *amplitude modulation.*

## 3.4 AMPLITUDE MODULATION

A frequency-translated signal from which the baseband signal is easily recoverable is generated by adding, to the product of baseband and carrier, the carrier signal itself. Such a signal is shown in Fig. 3.4-1. Figure 3.4-1*a* shows the carrier signal with amplitude $A_c$, in Fig. 3.4-1*b* we see the baseband signal. The translated signal (Fig. 3.4-1*c*) is given by

$$v(t) = A_c[1 + m(t)] \cos \omega_c t \qquad (3.4\text{-}1)$$

We observe, from Eq. (3.4-1) as well as from Fig. 3.4-1*c*, that the resultant waveform is one in which the carrier $A_c \cos \omega_c t$ is *modulated in amplitude.* The process of generating such a waveform is called *amplitude modulation*, and a communication system which employs such a method of frequency translation is called an *amplitude-modulation* system, or *AM* for short. The designation "carrier" for the auxiliary signal $A_c \cos \omega_c t$ seems especially appropriate in the present connection since this signal now "carries" the baseband signal as its envelope. The term "carrier" probably originated, however, in the early days of radio when this relatively high-frequency signal was viewed as the *messenger* which actually "carried" the baseband signal from one antenna to another.

The very great merit of the amplitude-modulated carrier signal is the ease with which the baseband signal can be recovered. The recovery of the baseband signal, a process which is referred to as *demodulation* or *detection*, is accomplished with the simple circuit of Fig. 3.4-2*a*, which consists of a diode *D* and the resistor-capacitor *RC* combination. We now discuss the operation of this circuit briefly and qualitatively. For simplicity, we assume that the amplitude-modulated carrier which is applied at the input terminals is supplied by a voltage source of zero internal impedance. We assume further that the diode is ideal, i.e., of zero or infinite resistance, depending on whether the diode current is positive or the diode voltage negative.

Let us initially assume that the input is of fixed amplitude and that the resistor $R$ is not present. In this case, the capacitor charges to the peak positive voltage of the carrier. The capacitor holds this peak voltage, and the diode would not again conduct. Suppose now that the input-carrier amplitude is increased. The diode again conducts, and the capacitor charges to the new higher carrier peak. In order to allow the capacitor voltage to follow the carrier peaks when the carrier amplitude is decreasing, it is necessary to include the resistor $R$, so that the capacitor may discharge. In this case the capacitor voltage $v_c$ has the form



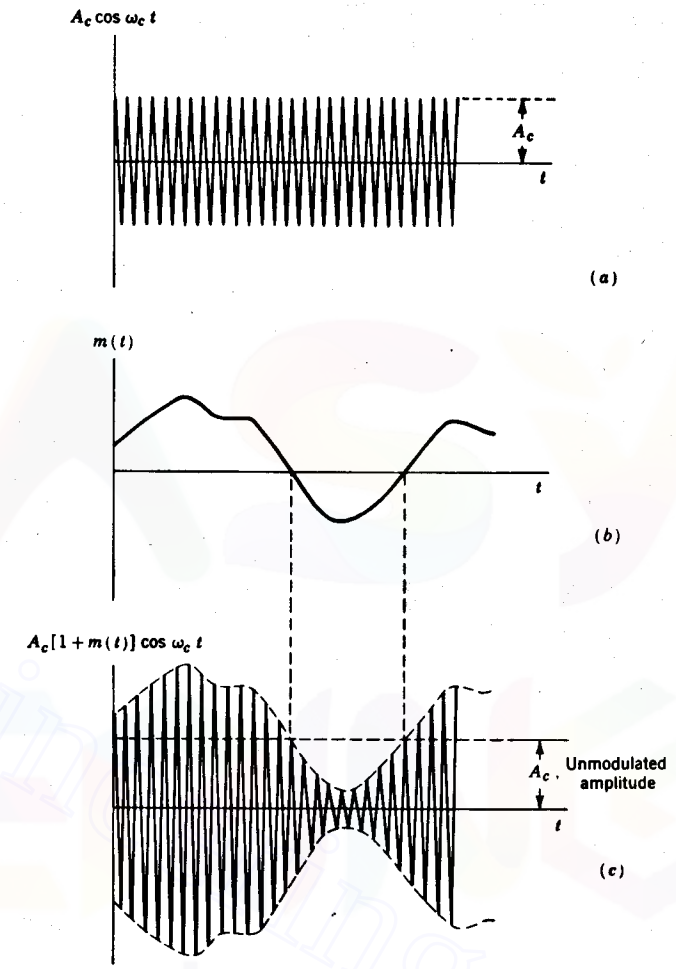



**Figure 3.4-1** (*a*) A sinusoidal carrier. (*b*) A modulating waveform. (*c*) The sinusoidal carrier in (*a*) modulated by the waveform in (*b*).

shown in Fig. 3.4-2*b*. The capacitor charges to the peak of each carrier cycle and decays slightly between cycles. The time constant $RC$ is selected so that the change in $v_c$ between cycles is at least equal to the decrease in carrier amplitude between cycles. This constraint on the time constant $RC$ is explored in Probs. 3.4-1 and 3.4-2.

It is seen that the voltage $v_c$ follows the carrier envelope except that $v_c$ also has superimposed on it a sawtooth waveform of the carrier frequency. In Fig. 3.4-2*b* the discrepancy between $v_c$ and the envelope is greatly exaggerated. In

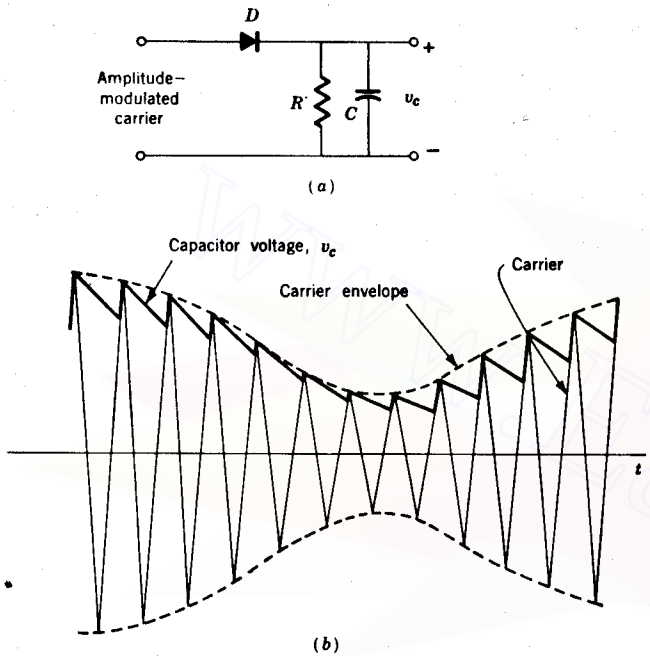



**Figure 3.4-2** (*a*) A demodulator for an AM signal. (*b*) Input waveform and output voltage $v_c$ across capacitor.

practice, the normal situation is one in which the time interval between carrier cycles is extremely small in comparison with the time required for the envelope to make a sizeable change. Hence $v_c$ follows the envelope much more closely than is suggested in the figure. Further, again because the carrier frequency is ordinarily much higher than the highest frequency of the modulating signal, the sawtooth distortion of the envelope waveform is very easily removed by a filter.

## 3.5 MAXIMUM ALLOWABLE MODULATION

If we are to avail ourselves of the convenience of demodulation by the use of the simple diode circuit of Fig. 3.4-2*a*, we must limit the extent of the modulation of the carrier. That such is the case may be seen from Fig. 3.5-1. In Fig. 3.5-1*a* is shown a carrier modulated by a sinusoidal signal. It is apparent that the envelope of the carrier has the waveshape of the modulating signal. The modulating signal is sinusoidal; hence $m(t) = m \cos \omega_m t$, where $m$ is a constant. Equation (3.4-1) becomes

$$v(t) = A_c(1 + m \cos \omega_m t) \cos \omega_c t \qquad (3.5\text{-}1)$$



In Fig. 3.5-1*b* we have shown the situation which results when, in Eq. (3.5-1), we adjust $m > 1$. Observe now that the diode demodulator which yields as an output the positive envelope (a negative envelope if the diode is reversed) will not reproduce the sinusoidal modulating waveform. In this latter case, where $m > 1$, we may recover the modulating waveform but not with the diode modulator. Recovery would require the use of a coherent demodulation scheme such as was employed in connection with the signal furnished by a multiplier.

It is therefore necessary to restrict the excursion of the modulating signal in the direction of decreasing carrier amplitude to the point where the carrier amplitude is just reduced to zero. No such similar restriction applies when the modulation is increasing the carrier amplitude. With sinusoidal modulation, as in Eq. (3.5-1), we require that $|m| \leq 1$. More generally in Eq. (3.4-1) we require that the maximum negative excursion of $m(t)$ be $-1$.

The extent to which a carrier has been amplitude-modulated is expressed in terms of a *percentage modulation*. Let $A_c$, $A_c(\max)$, and $A_c(\min)$, respectively, be

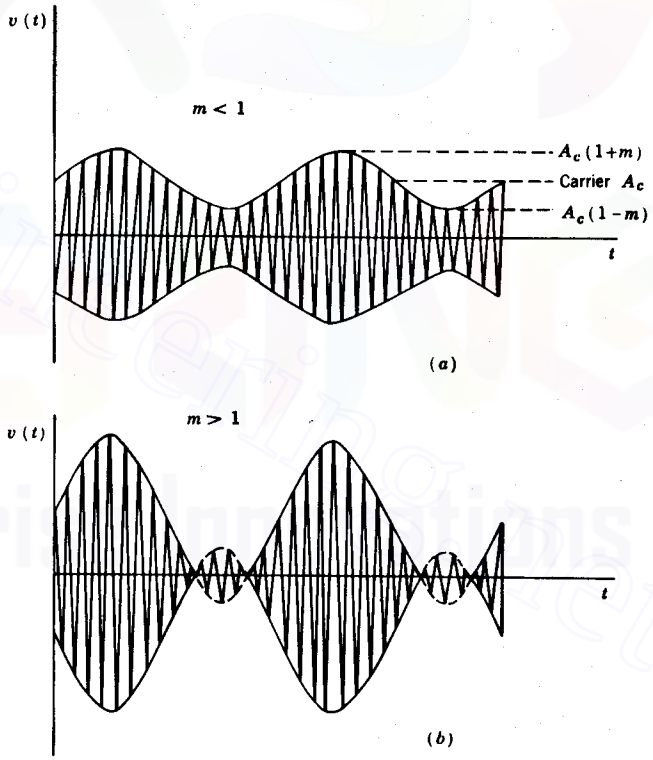



**Figure 3.5-1** (*a*) A sinusoidally modulated carrier ($m < 1$). (*b*) A carrier overmodulated ($m > 1$) by a sinusoidal modulating waveform.

the unmodulated carrier amplitude and the maximum and minimum carrier levels. Then if the modulation is symmetrical, the percentage modulation is defined as $P$, given by

$$\frac{P}{100\%} = \frac{A_c(\max) - A_c}{A_c} = \frac{A_c - A_c(\min)}{A_c} = \frac{A_c(\max) - A_c(\min)}{2A_c} \qquad (3.5\text{-}2)$$

In the case of sinusoidal modulation, given by Eq. (3.5-1) and shown in Fig. 3.5-1*a*, $P = m \times 100$ percent.

Having observed that the signal $m(t)$ may be recovered from the waveform $A_c[1 + m(t)] \cos \omega_c t$ by the simple circuit of Fig. 3.4-2*a*, it is of interest to note that a similar easy recovery of $m(t)$ is not possible from the waveform $m(t) \cos \omega_c t$. That such is the case is to be seen from Fig. 3.5-2. Figure 3.5-2*a* shows the carrier signal. The modulation or baseband signal $m(t)$ is shown in Fig.

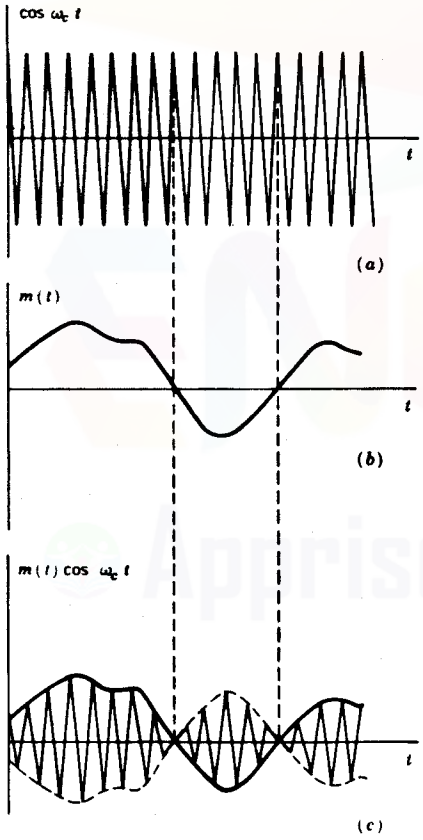



**Figure 3.5-2** (*a*) A carrier $\cos \omega_c t$. (*b*) A baseband signal $m(t)$. (*c*) The product $m(t) \cos \omega_c t$ and its envelope.



3.5-2*b*, and the product $m(t) \cos \omega_c t$ is shown in Fig. 3.5-2*c*. We note that the envelope in Fig. 3.5-2*c* has the waveform not of $m(t)$ but rather of $|m(t)|$, the *absolute value* of $m(t)$. Observe the reversal of phase of the carrier in Fig. 3.5-2*c* whenever $m(t)$ passes through zero.

## 3.6 THE SQUARE-LAW DEMODULATOR

An alternative method of recovering the baseband signal which has been superimposed as an amplitude modulation on a carrier is to pass the AM signal through a nonlinear device. Such demodulation is illustrated in Fig. 3.6-1. We assume here for simplicity that the device has a square-law relationship between input signal $x$ (current or voltage) and output signal $y$ (current or voltage). Thus $y = kx^2$, with $k$ a constant. Because of the nonlinearity of the transfer characteristic of the device, the output response is different for positive and for negative excursions of the carrier away from the quiescent operating point $O$ of the device.

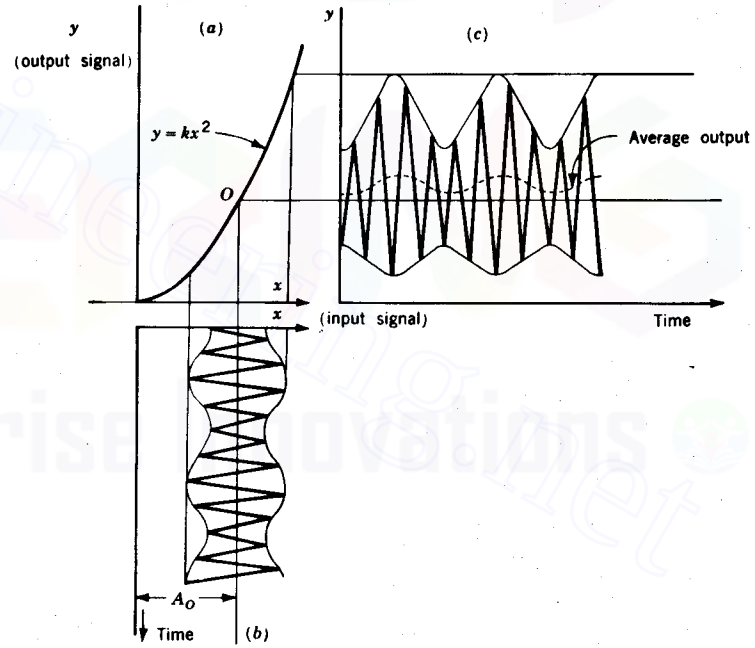



**Figure 3.6-1** Illustrating the operation of a square-law demodulator. The output is the value of $y$ averaged over many carrier cycles.

As a result, and as is shown in Fig. 3.6-1*c*, the output, when averaged over a time which encompasses many carrier cycles but only a very small part of the modulation cycle, has the waveshape of the envelope.

The applied signal is

$$x = A_o + A_c[1 + m(t)] \cos \omega_c t \tag{3.6-1}$$

Thus the output of the squaring circuit is

$$y = k\{A_o + A_c[1 + m(t)] \cos \omega_c t\}^2 \tag{3.6-2}$$

Squaring, and dropping dc terms as well as terms whose spectral components are located near $\omega_c$ and $2\omega_c$, we find that the output signal $s_o(t)$, that is, the signal output of a low-pass filter located after the squaring circuit, is

$$s_o(t) = kA_c^2[m(t) + \tfrac{1}{2}m^2(t)] \tag{3.6-3}$$

Observe that the modulation $m(t)$ is indeed recovered but that $m^2(t)$ appears as well. Thus the total recovered signal is a distorted version of the original modulation. The distortion is small, however, if $\frac{1}{2}m^2(t) \ll |m(t)|$ or if $|m(t)| \ll 2$.

There are two points of interest to be noted in connection with the type of demodulation described here; the first is that the demodulation does not depend on the nonlinearity being square-law. Any type of nonlinearity which does not have odd-function symmetry with respect to the initial operating point will similarly accomplish demodulation. The second point is that even when demodulation is not intended, such demodulation may appear incidentally when the modulated signal is passed through a system, say, an amplifier, which exhibits some nonlinearity.

## 3.7 SPECTRUM OF AN AMPLITUDE-MODULATED SIGNAL

The spectrum of an amplitude-modulated signal is similar to the spectrum of a signal which results from multiplication except, of course, that in the former case a carrier of frequency $f_c$ is present. If in Eq. (3.4-1) $m(t)$ is the superposition of three sinusoidal components $m(t) = m_1 \cos \omega_1 t + m_2 \cos \omega_2 t + m_3 \cos \omega_3 t$, then the (one-sided) spectrum of this baseband signal appears as at the left in Fig. 3.7-1*a*. The spectrum of the modulated carrier is shown at the right. The spectral lines at the sum frequencies $f_c + f_1$, $f_c + f_2$, and $f_c + f_3$ constitute the *upper-sideband* frequencies. The spectral lines at the difference frequencies constitute the *lower sideband*.

The spectrum of the baseband signal and modulated carrier are shown in Fig. 3.7-1*b* for the case of a bandlimited nonperiodic signal of finite energy. In this figure the ordinate is the spectral density, i.e., the magnitude of the Fourier transform rather than the spectral amplitude, and consequently the carrier is represented by an impulse.







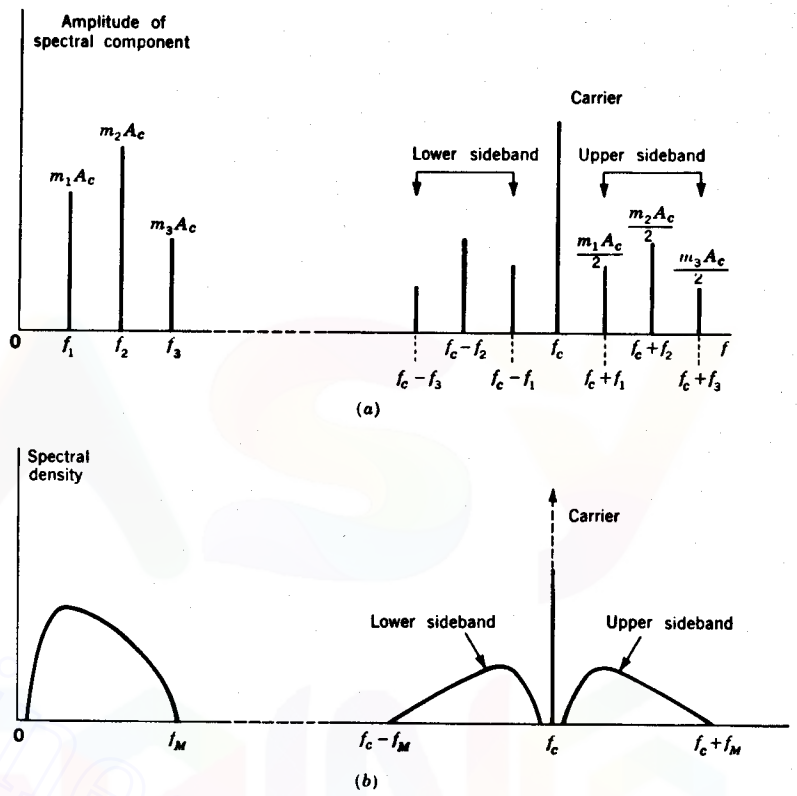



**Figure 3.7-1** (*a*) At left the one-sided spectrum of $m(t)A_c$, where $m(t)$ has three spectral components. At right the spectrum of $A_c[1 + m(t)] \cos 2\pi f_c t$. (*b*) Same as in (*a*) except $m(t)$ is a nonperiodic signal and the vertical axis is spectral density rather than spectral amplitude.

## 3.8 MODULATORS AND BALANCED MODULATORS

We have described a "multiplier" as a device that yields as an output a signal which is the product of two input signals. Actually no simple physical device now exists which yields the product alone. On the contrary, all such devices yield, at a minimum, not only the product but the input signals themselves. Suppose, then, that such a device has as inputs a carrier $\cos \omega_c t$ and a modulating baseband signal $m(t)$. The device output will then contain the product $m(t) \cos \omega_c t$ and also the signals $m(t)$ and $\cos \omega_c t$. Ordinarily, the baseband signal will be bandlimited to a frequency range very much smaller than $f_c = \omega_c/2\pi$. Suppose, for example, that the baseband signal extends from zero frequency to 1000 Hz, while $f_c = 1$ MHz. In this case, the carrier and its sidebands extend from 999,000 to 1,001,000 Hz, and the baseband signal is easily removed by a filter.

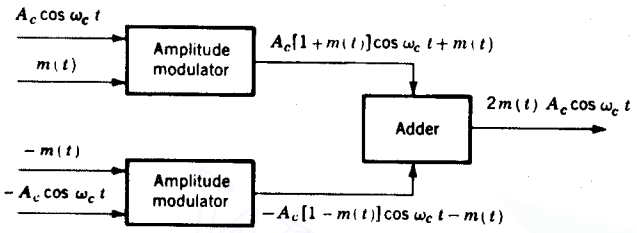



**Figure 3.8-1** Showing how the outputs of two amplitude modulators are combined to produce a double-sideband suppressed-carrier output.

The overall result is that the devices available for multiplication yield an output carrier as well as the lower- and upper-sideband signals. The output is therefore an amplitude-modulated signal. If we require the product signal alone, we must take steps to cancel or *suppress* the carrier. Such a suppression may be achieved by adding, to the amplitude-modulated signal, a signal of carrier frequency equal in amplitude but opposite in phase to the carrier of the amplitude-modulated signal. Under these circumstances only the sideband signals will remain. For this reason, a product signal is very commonly referred to as a *double-sideband suppressed-carrier* signal, abbreviated DSB-SC.

An alternative arrangement for carrier suppression is shown in Fig. 3.8-1. Here two *physical* multipliers are used which are labeled in the diagram as *amplitude modulators*. The carrier inputs to the two modulators are of reverse polarity, as are the modulating signals. The modulator outputs are added with consequent suppression of the carrier. We observe a cancellation not only of the carrier but of the baseband signal $m(t)$ as well. This last feature is not of great import, since, as noted previously, the baseband signal is easily eliminated by a filter. We note that the product terms of the two modulators reinforce. The arrangement of Fig. 3.8-1 is called a *balanced modulator*.

## 3.9 SINGLE-SIDEBAND MODULATION

We have seen that the baseband signal may be recovered from a double-sideband suppressed-carrier signal by multiplying a second time, with the same carrier. It can also be shown that the baseband signal can be recovered in a similar manner even if only one sideband is available. For suppose a spectral component of the baseband signal is multiplied by a carrier $\cos \omega_c t$, giving rise to an upper sideband at $\omega_c + \omega$ and a lower sideband at $\omega_c - \omega$. Now let us assume that we have filtered out one sideband and are left with, say, only the upper sideband at $\omega_c + \omega$. If now this sideband signal is again multiplied by $\cos \omega_c t$, we shall generate a signal at $2\omega_c + \omega$ and the original baseband spectral component at $\omega$. If we had used the lower sideband at $\omega_c - \omega$, the second multiplication would have yielded a signal at $2\omega_c - \omega$ and again restored the baseband spectral component.



Since it is possible to recover the baseband signal from a single sideband, there is an obvious advantage in doing so, since spectral space is used more economically. In principle, two single-sideband (abbreviated SSB) communications systems can now occupy the spectral range previously occupied by a single amplitude-modulation system or a double-sideband suppressed-carrier system.

The baseband signal may *not* be recovered from a single-sideband signal by the use of a diode modulator. That such is the case is easily seen by considering, for example, that the modulating signal is a sinusoid of frequency $f$. In this case the single-sideband signal will consist also of a single sinusoid of frequency, say, $f_c + f$, and there is no amplitude variation at all at the baseband frequency to which the diode modulator can respond.

Baseband recovery is achieved at the receiving end of the single-sideband communications channel by heterodyning the received signal with a local carrier signal which is synchronous (coherent) with the carrier used at the transmitting end to generate the sideband. As in the double-sideband case it is necessary, in principle, that the synchronism be exact and, in practice, that synchronism be maintained to a high order of precision. The effect of a lack of synchronism is different in a double-sideband system and in a single-sideband system. Suppose that the carrier received is $\cos \omega_c t$ and that the local carrier is $\cos (\omega_c t + \theta)$. Then with DSB-SC, as noted in Sec. 3.3, the spectral component $\cos \omega t$ will, upon demodulation, reappear as $\cos \omega t \cos \theta$. In SSB, on the other hand, the spectral component, $\cos \omega t$ will reappear (Prob. 3.9-2) in the form $\cos (\omega t - \theta)$. Thus, in one case a phase offset in carriers affects the amplitude of the recovered signal and, for $\phi = \pi/2$, may result in a total loss of the signal. In the other case the offset produces a phase change but not an amplitude change.

Alternatively, let the local oscillator carrier have an angular frequency offset $\Delta\omega$ and so be of the form $\cos (\omega_c + \Delta\omega)t$. Then as already noted, in DSB-SC, the recovered signal has the form $\cos \omega t \cos \Delta\omega t$. In SSB, however, the recovered signal will have the form $\cos (\omega + \Delta\omega)t$. Thus, in one case the recovered spectral component $\cos \omega t$ reappears with a "warble," that is, an amplitude fluctuation at the rate $\Delta\omega$. In the other case the amplitude remains fixed, but the frequency of the recovered signal is in error by amount $\Delta\omega$.

A phase offset between the received carrier and the local oscillator will cause distortion in the recovered baseband signal. In such a case each spectral component in the baseband signal will, upon recovery, have undergone the *same* phase shift. Fortunately, when SSB is used to transmit voice or music, such *phase distortion* does not appear to be of major consequence, because the human ear seems to be insensitive to the phase distortion.

A frequency offset between carriers in amount of $\Delta f$ will cause each recovered spectral component of the baseband signal to be in error by the same amount $\Delta f$. Now, if it had turned out that the frequency error were proportional to the frequency of the spectral component itself, then the recovered signal would sound like the original signal except that it would be at a higher or lower pitch. Such, however, is not the case, since the frequency error is fixed. Thus frequencies in the original signal which were harmonically related will no longer be so related after

recovery. The overall result is that a frequency offset between carriers adversely affects the intelligibility of spoken communication and is not well tolerated in connection with music. As a matter of experience, it turns out that an error $\Delta f$ of less than 30 Hz is acceptable to the ear.

The need to keep the frequency offset $\Delta f$ between carriers small normally imposes severe restrictions on the frequency stabilities of the carrier signal generators at both ends of the communications system. For suppose that we require to keep $\Delta f$ to 10 Hz or less, and that our system uses a carrier frequency of 10 MHz. Then the sum of the frequency drift in the two carrier generators may not exceed 1 part in $10^6$. The required equality in carrier frequency may be maintained through the use of quartz crystal oscillators using crystals cut for the same frequency at transmitter and receiver. The receiver must use as many crystals (or equivalent signals derived from crystals) as there are channels in the communications system.

It is also possible to tune an SSB receiver manually and thereby reduce the frequency offset. To do this, the operator manually adjusts the frequency of the receiver carrier generator until the received signal sounds "normal." Experienced operators are able to tune correctly to within 10 or 20 Hz. However, because of oscillator drift, such tuning must be readjusted periodically.

When the carrier frequency is very high, even quartz crystal oscillators may be hard pressed to maintain adequate stability. In such cases it is necessary to transmit the carrier itself along with the sideband signal. At the receiver the carrier may be separated by filtering and used to synchronize a local carrier generator. When used for such synchronization, the carrier is referred to as a "pilot carrier" and may be transmitted at a substantially reduced power level.

It is interesting to note that the squaring circuit used to recover the frequency and phase information of the DSB-SC system cannot be used here. In any event, it is clear that a principal complication in the way of more widespread use of single sideband is the need for supplying an accurate carrier frequency at the receiver.

## 3.10 METHODS OF GENERATING AN SSB SIGNAL

### Filter Method

A straightforward method of generating an SSB signal is illustrated in Fig. 3.10-1. Here the baseband signal and a carrier are applied to a balanced modulator. The output of the balanced modulator bears both the upper- and lower-sideband signals. One or the other of these signals is then selected by a filter. The filter is a bandpass filter whose passband encompasses the frequency range of the sideband selected. The filter must have a cutoff sharp enough to separate the selected sideband from the other sideband. The frequency separation of the sidebands is twice the frequency of the lowest frequency spectral components of the baseband signal. Human speech contains spectral components as low as about 70 Hz.





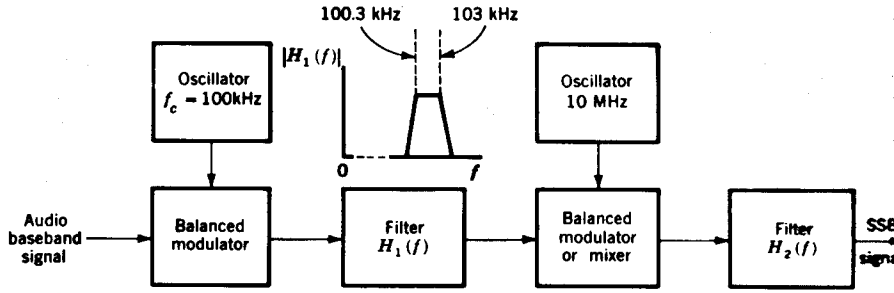



**Figure 3.10-1** Block diagram of the filter method of generating a single-sideband signal.

However, to alleviate the sideband filter selectivity requirements in an SSB system, it is common to limit the lower spectral limit of speech to about 300 Hz. It is found that such restriction does not materially affect the intelligibility of speech. Similarly, it is found that no serious distortion results if the upper limit of the speech spectrum is cut off at about 3000 Hz. Such restriction is advantageous for the purpose of conserving bandwidth. Altogether, then, a typical sideband filter has a passband which, measured from $f_c$, extends from about 300 to 3000 Hz and in which range its response is quite flat. Outside this passband the response falls off sharply, being down about 40 dB at 4000 Hz and rejecting the unwanted sideband also be at least 40 dB. The filter may also serve, further, to suppress the carrier itself. Of course, in principle, no carrier should appear at the output of a balanced modulator. In practice, however, the modulator may not balance exactly, and the precision of its balance may be subject to some variation with time. Therefore, even if a pilot carrier is to be transmitted, it is well to suppress it at the output of the modulator and to add it to the signal at a later point in a controllable manner.

Now consider that we desire to generate an SSB signal with a carrier of, say, 10 MHz. Then we require a passband filter with a selectivity that provides 40 dB of attenuation within 600 Hz at a frequency of 10 MHz, a percentage frequency change of 0.006 percent. Filters with such sharp selectivity are very elaborate and difficult to construct. For this reason, it is customary to perform the translation of the baseband signal to the final carrier frequency in several stages. Two such stages of translation are shown in Fig. 3.10-1. Here we have selected the first carrier to be of frequency 100 kHz. The upper sideband, say, of the output of the balanced modulator ranges from 100.3 to 103 kHz. The filter following the balanced modulator which selects this upper sideband need now exhibit a selectivity of only a hundredth of the selectivity (40 dB in 0.6 percent frequency change) required in the case of a 10-MHz carrier. Now let the filter output be applied to a second balanced modulator, supplied this time with a 10-MHz carrier. Let us again select the upper sideband. Then the second filter must provide 40 dB of attenuation in a frequency range of 200.6 kHz, which is nominally 2 percent of the carrier frequency.

We have already noted that the simplest physical frequency-translating device is a multiplier or mixer, while a balanced modulator is a balanced arrangement of two mixers. A mixer, however, has the disadvantage that it presents at its output not only sum and difference frequencies but the input frequencies as well. Still, when it is feasible to discriminate against these input signals, there is a merit of simplicity in using a mixer rather than a balanced modulator. In the present case, if the second frequency-translating device in Fig. 3.10-1 were a mixer rather than a multiplier, then in addition to the upper and lower sidebands, the output would contain a component encompassing the range 100.3 to 103 kHz as well as the 10-MHz carrier. The range 100.3 to 103 kHz is well out of the range of the second filter intended to pass the range 10,100,300 to 10,103,000 Hz. And it is realistic to design a filter which will suppress the 10-MHz carrier, since the carrier frequency is separated from the lower edge of the upper sideband (10,100,300) by nominally a 1-percent frequency change.

Altogether, then, we note in summary that when a single-sideband signal is to be generated which has a carrier in the megahertz or tens-of-megahertz range, the frequency translation is to be done in more than one stage—frequently two but not uncommonly three. If the baseband signal has spectral components in the range of hundreds of hertz or lower (as in an audio signal), the first stage invariably employs a balanced modulator, while succeeding stages may use mixers.

## Phasing Method

An alternative scheme for generating a single-sideband signal is shown in Fig. 3.10-2. Here two balanced modulators are employed. The carrier signals of angular frequency $\omega_c$ which are applied to the modulators differ in phase by 90°.

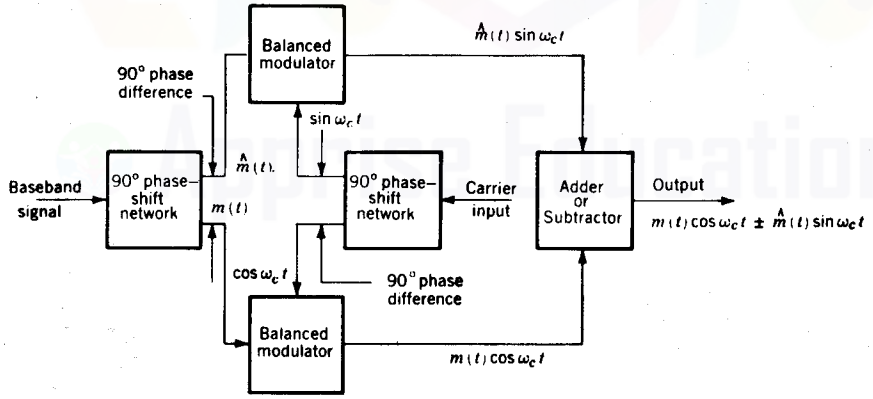



**Figure 3.10-2** A method of generating a single-sideband signal using balanced modulators and phase shifters.



Similarly the baseband signal, before application to the modulators, is passed through a 90° phase-shifting network so that there is a 90° phase shift between any spectral component of the baseband signal applied to one modulator and the like-frequency component applied to the other modulator.

To see most simply how the arrangement of Fig. 3.10-2 operates, let us assume that the baseband signal is sinusoidal and appears at the input to one modulator as $\cos \omega_m t$ and hence as $\sin \omega_m t$ at the other. Also, let the carrier be $\cos \omega_c t$ at one modulator and $\sin \omega_c t$ at the other. Then the outputs of the balanced modulators (multipliers) are

$$\cos \omega_m t \cos \omega_c t = \tfrac{1}{2}[\cos (\omega_c - \omega_m)t + \cos (\omega_c + \omega_m)t] \tag{3.10-1}$$

$$\sin \omega_m t \sin \omega_c t = \tfrac{1}{2}[\cos (\omega_c - \omega_m)t - \cos (\omega_c + \omega_m)t] \tag{3.10-2}$$

If these waveforms are added, the lower sideband results; if subtracted, the upper sideband appears at the output. In general, if the modulation $m(t)$ is given by

$$m(t) = \sum_{i=1}^{m} A_i \cos (\omega_i t + \theta_i) \tag{3.10-3}$$

then, using Fig. 3.10-2, we see that the output of the SSB modulator is in general

$$m(t) \cos \omega_c t \pm \hat{m}(t) \sin \omega_c t \tag{3.10-4}$$

where

$$\hat{m}(t) \equiv \sum_{i=1}^{m} A_i \sin (\omega_i t + \theta_i) \tag{3.10-5}$$

The single-sideband generating system of Fig. 3.10-2 generally enjoys less popularity than does the filter method. The reason for this lack of favor is that the present phasing method requires, for satisfactory operation, that a number of constraints be rather precisely met if the carrier and one sideband are adequately to be suppressed. It is required that each modulator be rather carefully balanced to suppress the carrier. It requires also that the baseband signal phase-shifting network provide to the modulators signals in which equal frequency spectral components are of exactly equal amplitude and differ in phase by precisely 90°. Such a network is difficult to construct for a baseband signal which extends over many octaves. It is also required that each modulator display equal sensitivity to the baseband signal. Finally, the carrier phase-shift network must provide exactly 90° of phase shift. If any of these constraints is not satisfied, the suppression of the rejected sideband and of the carrier will suffer. The effect on carrier and sideband suppression due to a failure precisely to meet these constraints is explored in Probs. 3.10-3 and 3.10-4. Of course, in any physical system a certain level of carrier and rejected sideband is tolerable. Still, there seems to be a general inclination to achieve a single sideband by the use of passive filters rather than by a method which requires many exactly maintained balances in passive and active circuits. There is an alternative single-sideband generating scheme[2] which avoids the need for a wideband phase-shifting network but which uses four balanced modulators.

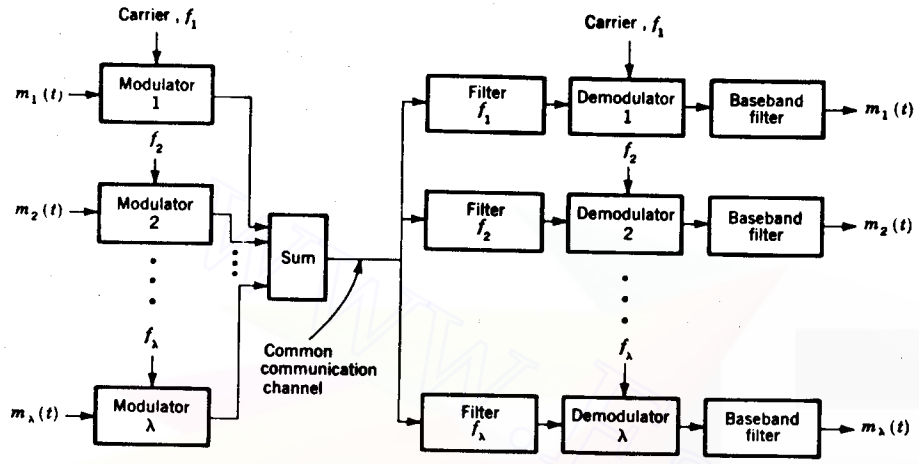



**Figure 3.13-1** Multiplexing many baseband signals over a single communications channel.

*multiplexing*. As a matter of fact, to facilitate this separation of the individual signals, the carrier frequencies are selected to leave a comfortable margin (guard band) between the limit of one frequency range and the beginning of the next. The combined output of all the modulators, i.e., the *composite* signal, is applied to a common communications channel. In the case of radio transmission, the channel is free space, and coupling to the channel is made by means of an antenna. In other cases wires are used.

At the receiving end the composite signal is applied to each of a group of bandpass filters whose passbands are in the neighborhood $f_1, f_2, \ldots, f_\lambda$. The filter $f_1$ is a bandpass filter which passes only the spectral range of the output of modulator 1 and similarly for the other bandpass filters. The signals have thus been separated. They are then applied to individual demodulators which extract the baseband signals from the carrier. The carrier inputs to the demodulators are required only for synchronous demodulation and are not used otherwise.

The final operation indicated in Fig. 3.13-1 consists in passing the demodulator output through a baseband filter. The baseband filter is a low-pass filter with cutoff at the frequency $f_M$ to which the baseband signal is limited. This baseband filter will pass, without modification, the baseband signal output of the modulator and in this sense serves no function in the system as described up to the present point. We shall, however, see in Chaps. 8 and 9 that such baseband filters are essential to suppress the noise which invariably accompanies the signal.

## REFERENCES

1. Bell Telephone Laboratories: "Transmission Systems for Communications," Western Electric Company, Tech. Pub., Winston-Salem, N.C., 1964.







2. Norgaard, D. E.: A Third Method of Generation and Detection of Single-sideband Signals, *Proc. IRE*, December, 1956.
3. Voelcker, H.: Demodulation of Single-sideband Signals Via Envelope Detection, *IEEE Trans. on Communication Technology*, pp. 22–30, February, 1966.

## PROBLEMS

**3.2-1.** A signal $v_m(t)$ is bandlimited to the frequency range 0 to $f_M$. It is frequency-translated by multiplying it by the signal $v_c(t) = \cos 2\pi f_c t$. Find $f_c$ so that the bandwidth of the translated signal is 1 percent of the frequency $f_c$.

**3.2-2.** The Fourier transform of $m(t)$ is $\mathcal{F}[m(t)] \equiv M(f)$. Show that

$$\mathcal{F}[m(t) \cos 2\pi f_c t] = \tfrac{1}{2}[M(f + f_c) + M(f - f_c)]$$

**3.2-3.** The signals

$$v_1(t) = 2 \cos \omega_1 t + \cos 2\omega_1 t$$

and

$$v_2(t) = \cos \omega_2 t + 2 \cos 2\omega_2 t$$

are multiplied. Plot the resultant amplitude-frequency characteristic, assuming that $\omega_2 > 2\omega_1$, but is not a harmonic of $\omega_1$. Repeat for $\omega_2 = 2\omega_1$.

**3.3-1.** The baseband signal $m(t)$ in the frequency-translated signal $v(t) = m(t) \cos 2\pi f_c t$ is recovered by multiplying $v(t)$ by the waveform $\cos (2\pi f_c t + \theta)$.

(*a*) The product waveform is transmitted through a low-pass filter which rejects the double-frequency signal. What is the output signal of the filter?

(*b*) What is the maximum allowable value for the phase $\theta$ if the recovered signal is to be 90 percent of the maximum possible value?

(*c*) If the baseband signal $m(t)$ is bandlimited to 10 kHz, what is the minimum value of $f_c$ for which it is possible to recover $m(t)$ by filtering the product waveform $v(t) \cos (2\pi f_c t + \theta)$?

**3.3-2.** The baseband signal $m(t)$ in the frequency-translated signal $v(t) = m(t) \cos 2\pi f_c t$ is recovered by multiplying $v(t)$ by the waveform $\cos 2\pi(f_c + \Delta f)t$. The product waveform is transmitted through a low-pass filter which rejects the double-frequency signal. Find the output signal of the filter.

**3.3-3.** (*a*) The baseband signal $m(t)$ in the frequency-translated signal $v(t) = m(t) \cos 2\pi f_c t$ is to be recovered. There is available a waveform $p(t)$ which is periodic with period $1/f_c$. Show that $m(t)$ may be recovered by appropriately filtering the product waveform $p(t)v(t)$.

(*b*) Show that $m(t)$ may be recovered as well if the periodic waveform has a period $n/f_c$, where $n$ is an integer. Assume $m(t)$ bandlimited to the frequency range from 0 to 5 kHz and let $f_c = 1$ MHz. Find the largest $n$ which will allow $m(t)$ to be recovered. Are all periodic waveforms acceptable?

**3.3-4.** The signal $m(t)$ in the DSB-SC signal $v(t) = m(t) \cos (\omega_c t + \theta)$ is to be reconstructed by multiplying $v(t)$ by a signal derived from $v^2(t)$.

(*a*) Show that $v^2(t)$ has a component at the frequency $2f_c$. Find its amplitude.

(*b*) If $m(t)$ is bandlimited to $f_M$ and has a probability density

$$f(m) = \frac{1}{\sqrt{2\pi}} e^{-m^2/2} \qquad -\infty \le m \le \infty$$

find the expected value of the amplitude of the component of $v^2(t)$ at $2f_c$.

**3.4-1.** The envelope detector shown in Fig. 3.4-2*a* is used to recover the signal $m(t)$ from the AM signal $v(t) = [1 + m(t)] \cos \omega_c t$, where $m(t)$ is a square wave taking on the values 0 and $-0.5$ volt and having a period $T \gg 1/f_c$. Sketch the *recovered* signal if $RC = T/20$ and $4T$.

**3.4-2.** (*a*) The waveform $v(t) = (1 + m \cos \omega_m t) \cos \omega_c t$, with $m$ a constant ($m \leq 1$), is applied to the diode demodulator of Fig. 3.4-2*a*. Show that, if the demodulator output is to follow the envelope of $v(t)$, it is required that at any time $t_0$:

$$\frac{1}{RC} \geq \omega_m \left( \frac{m \sin \omega_m t_0}{1 + m \cos \omega_m t_0} \right)$$

(*b*) Using the result of part (*a*), show that if the demodulator is to follow the envelope at all times then $m$ must be less than or equal to the value of $m_0$, determined from the equation

$$RC = \frac{1}{\omega_m} \frac{\sqrt{1 - m_0^2}}{m_0}$$

(*c*) Draw, qualitatively, the form of the demodulator output when the condition specified in part (*b*) is not satisfied.

**3.5-1.** The signal $v(t) = (1 + m \cos \omega_m t) \cos \omega_c t$ is detected using a diode envelope detector. Sketch the detector output when $m = 2$.

**3.6-1.** The signal $v(t) = [1 + 0.2 \cos (\omega_M/3)t] \cos \omega_c t$ is demodulated using a square-law demodulator having the characteristic $v_o = (v + 2)^2$. The output $v_o(t)$ is then filtered by an ideal low-pass filter having a cutoff frequency at $f_M$ Hz. Sketch the amplitude-frequency characteristics of the output waveform in the frequency range $0 \leq f \leq f_M$.

**3.6-2.** Repeat Prob. 3.6-1 if the square-law demodulator is centered at the origin so that $v_o = v^2$.

**3.6-3.** The signal $v(t) = [1 + m(t)] \cos \omega_c t$ is square-law detected by a detector having the characteristic $v_o = v^2$. If the Fourier transform of $m(t)$ is a constant $M_0$ extending from $-f_M$ to $+f_M$, sketch the Fourier transform of $v_o(t)$ in the frequency range $-f_M < f < f_M$. *Hint:* Convolution in the frequency domain is needed to find the Fourier transform of $m^2(t)$. See Prob. 1.12-2.

**3.6-4.** The signal $v(t) = (1 + 0.1 \cos \omega_1 t + 0.1 \cos 2\omega_1 t) \cos \omega_c t$ is detected by a square-law detector, $v_o = 2v^2$. Plot the amplitude-frequency characteristic of $v_o(t)$.

**3.9-1.** (*a*) Show that the signal

$$v(t) = \sum_{i=1}^{N} [\cos \omega_c t \cos (\omega_i t + \theta_i) - \sin \omega_c t \sin (\omega_i t + \theta_i)]$$

is an SSB-SC signal ($\omega_c \gg \omega_N$). Is it the upper or lower sideband?

(*b*) Write an expression for the missing sideband.

(*c*) Obtain an expression for the total DSB-SC signal.

**3.9-2.** The SSB signal in Prob. 3.9-1 is multiplied by $\cos \omega_c t$ and then low-pass filtered to recover the modulation.

(*a*) Show that the modulation is completely recovered if the cutoff frequency of the low-pass filter $f_0$ is $f_M < f_0 < 2f_c$.

(*b*) If the multiplying signal were $\cos (\omega_c t + \theta)$, find the recovered signal.

(*c*) If the multiplying signal were $\cos (\omega_c + \Delta\omega)t$, find the recovered signal. Assume that $\Delta\omega \ll \omega_1$.

**3.9-3.** Show that the squaring circuit shown in Fig. 3.3-1 will not permit the generation of a local oscillator signal capable of demodulating an SSB-SC signal.

**3.10-1.** A baseband signal, bandlimited to the frequency range 300 to 3000 Hz, is to be superimposed on a carrier of frequency of 40 MHz as a single-sideband modulation using the filter method. Assume that bandpass filters are available which will provide 40 dB of attenuation in a frequency interval which is about 1 percent of the filter center frequency. Draw a block diagram of a suitable system. At each point in the system draw plots indicating the spectral range occupied by the signal present there.

**3.10-2.** The system shown in Fig. 3.10-2 is used to generate a single-sideband signal. However, an ideal 90° phase-shifting network which is independent of a frequency is unattainable. The 90° phase shift is approximated by a lattice network having the transfer function

$$H(f) = e^{-j \arctan (f/30)}$$







The input to this network is $m(t)$, given by Eq. (3.10-3). If $f_1 = 300$ Hz and $f_M = 3000$ Hz show that $H(f) \cong e^{-j\pi/2} e^{j30/f}$, for $f_1 \leq f \leq f_M$.

**3.10-3.** In the SSB generating system of Fig. 3.10-2, the carrier phase-shift network produces a phase shift which differs from 90° by a *small* angle $\alpha$. Calculate the output waveform and point out the respects in which the output no longer meets the requirements for an SSB waveform. Assume that the input is a single spectral component $\cos \omega_m t$.

**3.10-4.** Repeat Prob. 3.10-3 except, assume instead, that the baseband phase-shift network produces a phase shift differing from 90° by a *small* angle $\alpha$.

**3.11-1.** A received SSB signal in which the modulation is a single spectral component has a normalized power of 0.5 volt$^2$. A carrier is added to the signal, and the carrier plus signal are applied to a diode demodulator. The carrier amplitude is to be adjusted so that at the demodulator output 90 percent of the normalized power is in the recovered modulating waveform. Neglect dc components. Find the carrier amplitude required.

CHAPTER

# FOUR

# FREQUENCY-MODULATION SYSTEMS

In the amplitude-modulation systems described in Chap. 3, the modulator output consisted of a carrier which displayed variations in its amplitude. In the present chapter we discuss modulation systems in which the modulator output is of constant amplitude and in which the signal information is superimposed on the carrier through variations of the carrier frequency.

## 4.1 ANGLE MODULATION[1]

All the modulation schemes considered up to the present point have two principal features in common. In the first place, each spectral component of the baseband signal gives rise to one or two spectral components in the modulated signal. These components are separated from the carrier by a frequency difference equal to the frequency of the baseband component. Most importantly, the nature of the modulators is such that the spectral components which they produce depend only on the carrier frequency and the baseband frequencies. The amplitudes of the spectral components of the modulator output may depend on the amplitude of the input signals; however, the frequencies of the spectral components do not. In the second place, all the operations performed on the signal (addition, subtraction, and multiplication) are linear operations so that superposition applies. Thus, if a baseband signal $m_1(t)$ introduces one spectrum of components into the modulated signal and a second signal $m_2(t)$ introduces a second spectrum, the application of the sum $m_1(t) + m_2(t)$ will introduce a spectrum which is the sum of the spectra separately introduced. All these systems are referred to under the designation "amplitude or linear modulation." This terminology must be taken







with some reservation, for we have noted that, at least in the special case of single sideband using modulation with a single sinusoid, there is no amplitude variation at all. And even more generally, when the amplitude of the modulated signal does vary, the carrier envelope need not have the waveform of the baseband signal.

We now turn our attention to a new type of modulation which is not characterized by the features referred to above. The spectral components in the modulated waveform depend on the amplitude as well as the frequency of the spectral components in the baseband signal. Furthermore, the modulation system is *not* linear and superposition does *not* apply. Such a system results when, in connection with a carrier of constant amplitude, the phase angle is made to respond in some way to a baseband signal. Such a signal has the form

$$v(t) = A \cos [\omega_c t + \phi(t)] \tag{4.1-1}$$

in which $A$ and $\omega_c$ are constant but in which the phase angle $\phi(t)$ is a function of the baseband signal. Modulation of this type is called *angle modulation* for obvious reasons. It is also referred to as *phase modulation* since $\phi(t)$ is the phase angle of the argument of the cosine function. Still another designation is *frequency modulation* for reasons to be discussed in the next section.

## 4.2 PHASE AND FREQUENCY MODULATION

To review some elementary ideas in connection with sinusoidal waveforms, let us recall that the function $A \cos \omega_c t$ can be written as

$$A \cos \omega_c t = \text{real part } (Ae^{j\omega_c t}) \tag{4.2-1}$$

The function $Ae^{j\theta}$ is represented in the complex plane by a phasor of length $A$ and an angle $\theta$ measured counterclockwise from the real axis. If $\theta = \omega_c t$, then the phasor rotates in the counterclockwise direction with an angular velocity $\omega_c$. With respect to a coordinate system which also rotates in the counterclockwise direction with angular velocity $\omega_c$, the phasor will be stationary. If in Eq. (4.1-1) $\phi$ is actually not time-dependent but is a constant, then $v(t)$ is to be represented precisely in the manner just described. But suppose $\phi = \phi(t)$ does change with time and makes positive and negative excursions. Then $v(t)$ would be represented by a phasor of amplitude $A$ which runs ahead of and falls behind the phasor representing $A \cos \omega_c t$. We may, therefore, consider that the angle $\omega_c t + \phi(t)$, of $v(t)$, undergoes a *modulation* around the angle $\theta = \omega_c t$. The waveform of $v(t)$ is, therefore, a representation of a signal which is *modulated in phase.*

If the phasor of angle $\theta + \phi(t) = \omega_c t + \phi(t)$ alternately runs ahead of and falls behind the phasor $\theta = \omega_c t$, then the first phasor must alternately be rotating more, or less, rapidly than the second phasor. Therefore we may consider that the angular velocity of the phasor of $v(t)$ undergoes a modulation around the nominal angular velocity $\omega_c$. The signal $v(t)$ is, therefore, an angular-velocity-modulated waveform. The angular velocity associated with the argument of a sinusoidal function is equal to the time rate of change of the argument (i.e., the

angle) of the function. Thus we have that the instantaneous radial frequency $\omega = d(\theta + \phi)/dt$, and the corresponding frequency $f = \omega/2\pi$ is

$$f = \frac{1}{2\pi}\frac{d}{dt}[\omega_c t + \phi(t)] = \frac{\omega_c}{2\pi} + \frac{1}{2\pi}\frac{d}{dt}\phi(t) \qquad (4.2\text{-}2)$$

The waveform $v(t)$ is, therefore, *modulated in frequency*.

In initial discussions of the sinusoidal waveform it is customary to consider such a waveform as having a fixed frequency and phase. In the present discussion we have generalized these concepts somewhat. To acknowledge this generalization, it is not uncommon to refer to the frequency $f$ in Eq. (4.2-2) as the *instantaneous frequency* and $\phi(t)$ as the *instantaneous phase*. If the frequency variation about the nominal frequency $\omega_c$ is small, that is, if $d\phi(t)/dt \ll \omega_c$, then the resultant waveform will have an appearance which is readily recognizable as a "sine wave," albeit with a period which changes somewhat from cycle to cycle. Such a waveform is represented in Fig. 4.2-1. In this figure the modulating signal is a square wave. The frequency-modulated signal changes frequency whenever the modulation changes level.

Among the possibilities which suggest themselves for the design of a modulator are the following. We might arrange that the phase $\phi(t)$ in Eq. (4.1-1) be directly proportional to the modulating signal, or we might arrange a direct proportionality between the modulating signal and the derivative, $d\phi(t)/dt$. From Eq. (4.2-2), with $f_c = \omega_c/2\pi$

$$\frac{d\phi(t)}{dt} = 2\pi(f - f_c) \qquad (4.2\text{-}3)$$

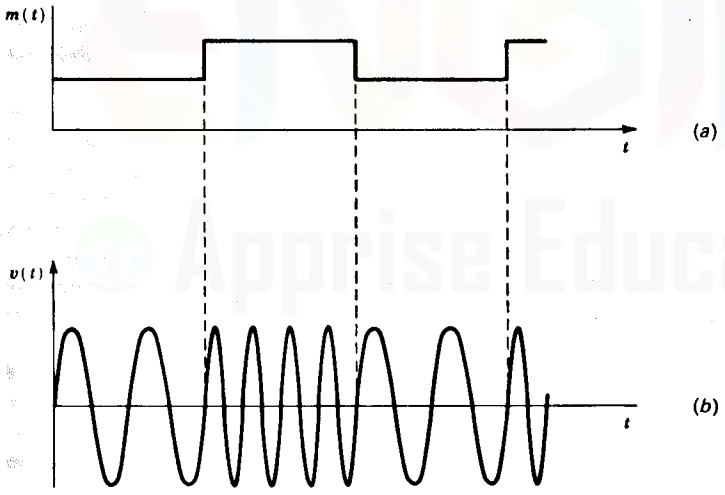



**Figure 4.2-1** An angle-modulated waveform. (*a*) Modulating signal. (*b*) Frequently-modulated sinusoidal carrier signal.





where $f$ is the instantaneous frequency. Hence in this latter case the proportionality is between modulating signal and the departure of the instantaneous frequency from the carrier frequency. Using standard terminology, we refer to the modulation of the first type as *phase* modulation, and the term *frequency modulation* refers only to the second type. On the basis of these definitions it is, of course, not possible to determine which type of modulation is involved simply from a visual examination of the waveform or from an analytical expression for the waveform. We would also have to be given the waveform of the modulating signal. This information is, however, provided in any practical communication system.

## 4.3 RELATIONSHIP BETWEEN PHASE AND FREQUENCY MODULATION

The relationship between phase and frequency modulation may be visualized further by a consideration of the diagrams of Fig. 4.3-1. In Fig. 4.3-1*a* the phase-modulator block represents a device which furnishes an output $v(t)$ which is a carrier, phase-modulated by the input signal $m_i(t)$. Thus

$$v(t) = A\cos[\omega_c t + k' m_i(t)] \qquad (4.3\text{-}1)$$

$k'$ being a constant. Let the waveform $m_i(t)$ be derived as the integral of the modulating signal $m(t)$ so that

$$m_i(t) = k'' \int_{-\infty}^{t} m(t)\,dt \qquad (4.3\text{-}2)$$

in which $k''$ is also a constant. Then with $k = k'k''$ we have

$$v(t) = A\cos\left[\omega_c t + k\int_{-\infty}^{t} m(t)\,dt\right] \qquad (4.3\text{-}3)$$

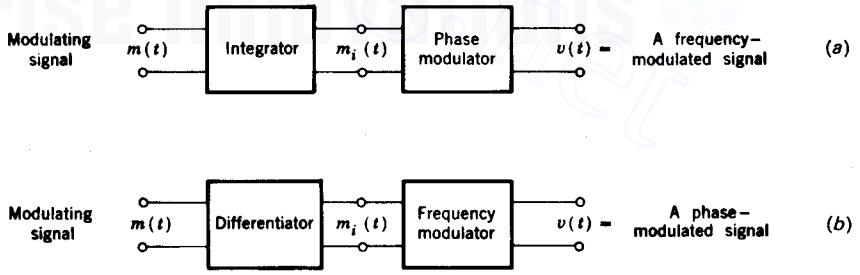



**Figure 4.3-1** Illustrating the relationship between phase and frequency modulation.

The instantaneous angular frequency is

$$\omega = \frac{d}{dt}\left[\omega_c t + k\int_{-\infty}^{t} m(t)\,dt\right] = \omega_c + km(t) \tag{4.3-4}$$

The deviation of the instantaneous frequency from the carrier frequency $\omega_c/2\pi$ is

$$v \equiv f - f_c = \frac{k}{2\pi}\,m(t) \tag{4.3-5}$$

Since the deviation of the instantaneous frequency is directly proportional to the modulating signal, the combination of *integrator* and *phase modulator* of Fig. 4.3-1*a* constitutes a device for producing a *frequency-modulated* output. Similarly, the combination in Fig. 4.3-1*b* of the differentiator and frequency modulator generates a *phase-modulated output*, i.e., a signal whose phase departure from the carrier is proportional to the modulating signal.

In summary, we have referred generally to the waveform given by Eq. (4.1-1) as an *angle-modulated* waveform, an appropriate designation when we have no interest in, or information about, the modulating signal. When $\phi(t)$ is proportional to the modulating signal $m(t)$, we use the designation *phase modulation* or PM. When the time derivative of $\phi(t)$ is proportional to $m(t)$, we use the term *frequency modulation* or FM. In an FM waveform, the form of Eq. (4.3-3) is of special interest, since here the instantaneous frequency deviation is directly proportional to the signal $m(t)$ which appears explicitly in the expression. In general usage, however, we find that such precision of language is not common. Very frequently the terms angle modulation, phase modulation, and frequency modulation are used rather interchangeably and without reference to, or even interest in, the modulating signal.

## 4.4 PHASE AND FREQUENCY DEVIATION

In the waveform of Eq. (4.1-1) the maximum value attained by $\phi(t)$, that is, the maximum phase deviation of the total angle from the carrier angle $\omega_c t$, is called the *phase deviation*. Similarly the maximum departure of the instantaneous frequency from the carrier frequency is called the *frequency deviation*.

When the angular (and consequently the frequency) variation is sinusoidal with frequency $f_m$, we have, with $\omega_m = 2\pi f_m$

$$v(t) = A\cos(\omega_c t + \beta\sin\omega_m t) \tag{4.4-1}$$

where $\beta$ is the peak amplitude of $\phi(t)$. In this case $\beta$ which is the maximum phase deviation, is usually referred to as the *modulation index*. The instantaneous frequency is

$$f = \frac{\omega_c}{2\pi} + \frac{\beta\omega_m}{2\pi}\cos\omega_m t \tag{4.4-2a}$$

$$= f_c + \beta f_m \cos\omega_m t \tag{4.4-2b}$$





The maximum frequency deviation is defined as $\Delta f$ and is given by

$$\Delta f = \beta f_m \tag{4.4-3}$$

Equation (4.4-1) can, therefore, be written

$$v(t) = A\cos\left(\omega_c t + \frac{\Delta f}{f_m}\sin\omega_m t\right) \tag{4.4-4}$$

While the instantaneous frequency $f$ lies in the range $f_c \pm \Delta f$, it should not be concluded that all spectral components of such a signal lie in this range. We consider next the spectral pattern of such an angle-modulated waveform.

## 4.5 SPECTRUM OF AN FM SIGNAL: SINUSOIDAL MODULATION

In this section we shall look into the frequency spectrum of the signal

$$v(t) = \cos(\omega_c t + \beta\sin\omega_m t) \tag{4.5-1}$$

which is the signal of Eq. (4.4-1) with the amplitude arbitrarily set at unity as a matter of convenience. We have

$$\cos(\omega_c t + \beta\sin\omega_m t) = \cos\omega_c t\cos(\beta\sin\omega_m t) - \sin\omega_c t\sin(\beta\sin\omega_m t) \tag{4.5-2}$$

Consider now the expression $\cos(\beta\sin\omega_m t)$ which appears as a factor on the right-hand side of Eq. (4.5-2). It is an *even*, periodic function having an angular frequency $\omega_m$. Therefore it is possible to expand this expression in a Fourier series in which $\omega_m/2\pi$ is the fundamental frequency. We shall not undertake the evaluation of the coefficients in the Fourier expansion of $\cos(\beta\sin\omega_m t)$ but shall instead simply write out the results. The coefficients are, of course, functions of $\beta$, and, since the function is *even*, the coefficients of the odd harmonics are zero. The result is

$$\cos(\beta\sin\omega_m t) = J_0(\beta) + 2J_2(\beta)\cos 2\omega_m t + 2J_4(\beta)\cos 4\omega_m t + \cdots + 2J_{2n}(\beta)\cos 2n\omega_m t + \cdots \tag{4.5-3}$$

while for $\sin(\beta\sin\omega_m t)$, which is an *odd* function, we find the expansion contains only odd harmonics and is given by

$$\sin(\beta\sin\omega_m t) = 2J_1(\beta)\sin\omega_m t + 2J_3(\beta)\sin 3\omega_m t + \cdots + 2J_{2n-1}(\beta)\sin(2n-1)\omega_m t + \cdots \tag{4.5-4}$$

The functions $J_n(\beta)$ occur often in the solution of engineering problems. They are known as Bessel functions of the first kind and of order $n$. The numerical values of $J_n(\beta)$ are tabulated in texts of mathematical tables.[2]

Putting the results given in Eqs. (4.5-3) and (4.5-4) back into Eq. (4.5-2) and using the identities

$$\cos A \cos B = \tfrac{1}{2}\cos(A - B) + \tfrac{1}{2}\cos(A + B) \tag{4.5-5}$$

$$\sin A \sin B = \tfrac{1}{2}\cos(A - B) - \tfrac{1}{2}\cos(A + B) \tag{4.5-6}$$

we find that $v(t)$ in Eq. (4.5-1) becomes

$$\begin{aligned} v(t) = {} & J_0(\beta)\cos\omega_c t - J_1(\beta)[\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t] \\ & + J_2(\beta)[\cos(\omega_c - 2\omega_m)t + \cos(\omega_c + 2\omega_m)t] \\ & - J_3(\beta)[\cos(\omega_c - 3\omega_m)t - \cos(\omega_c + 3\omega_m)t] \\ & + \cdots \end{aligned} \tag{4.5-7}$$

Observe that the spectrum is composed of a carrier with an amplitude $J_0(\beta)$ and a set of sidebands spaced symmetrically on either side of the carrier at frequency separations of $\omega_m$, $2\omega_m$, $3\omega_m$, etc. In this respected the result is unlike that which prevails in the amplitude-modulation systems discussed earlier, since in AM a sinusoidal modulating signal gives rise to only one sideband or one pair of sidebands. A second difference, which is left for verification by the student (Prob. 4.5-1), is that the present modulation system is nonlinear, as anticipated from the discussion of Sec. 4.1.

## 4.6 SOME FEATURES OF THE BESSEL COEFFICIENTS

Several of the Bessel functions which determine the amplitudes of the spectral components in the Fourier expansion are plotted in Fig. 4.6-1. We note that, at $\beta = 0$, $J_0(0) = 1$, while all other $J_n$'s are zero. Thus, as expected when there is no modulation, only the carrier, of normalized amplitude unity, is present, while all sidebands have zero amplitude. When $\beta$ departs slightly from zero, $J_1(\beta)$ acquires a magnitude which is significant in comparison with unity, while all higher-order $J$'s are negligible in comparison. That such is the case may be seen either from Fig. 4.6-1 or from the approximations[3] which apply when $\beta \ll 1$, that is,

$$J_0(\beta) \cong 1 - \left(\frac{\beta}{2}\right)^2 \tag{4.6-1}$$

$$J_n(\beta) \cong \frac{1}{n!}\left(\frac{\beta}{2}\right)^n \qquad n \neq 0 \tag{4.6-2}$$

Accordingly, for $\beta$ very small, the FM signal is composed of a carrier and a single pair of sidebands with frequencies $\omega_c \pm \omega_m$. An FM signal which is so constituted, that is, a signal where $\beta$ is small enough so that only a single sideband pair is of significant magnitude, is called a *narrowband* FM signal. We see further, in Fig. 4.6-1, as $\beta$ becomes somewhat larger, that the amplitude $J_1$ of the first sideband pair increases and that also the amplitude $J_2$ of the second sideband pair

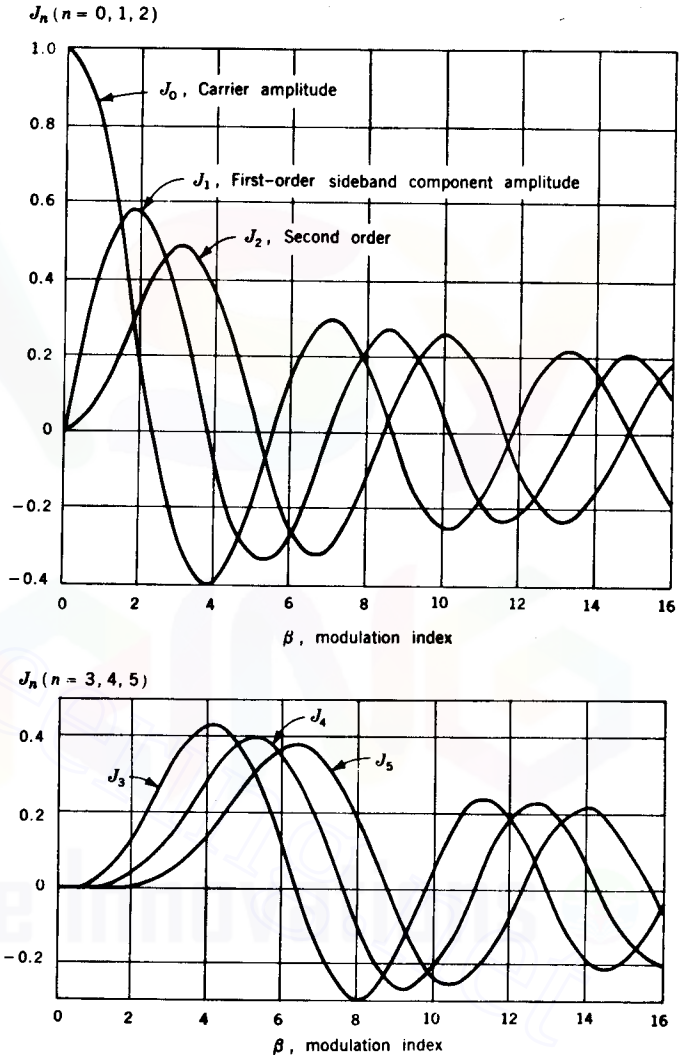



**Figure 4.6-1** The Bessel functions $J_n(\beta)$ plotted as a function of $\beta$ for $n = 0, 1, 2, \ldots, 5$.


Downloaded From : www.EasyEngineering.net

becomes significant. Further, as $\beta$ continues to increase, $J_3$, $J_4$, etc. begin to acquire significant magnitude, giving rise to sideband pairs at frequencies $\omega_c \pm 2\omega_m$, $\omega_c \pm 3\omega_m$, etc.

Another respect in which FM is unlike the linear-modulation schemes described earlier is that in an FM signal the amplitude of the spectral component at the carrier frequency is not constant independent of $\beta$. It is to be expected that such should be the case on the basis of the following considerations. The envelope of an FM signal has a constant amplitude. Therefore the power of such a signal is a constant independent of the modulation, since the power of a periodic waveform depends only on the square of its amplitude and not on its frequency. The power of a unit amplitude signal, as in Eq. (4.5-1), is $P_v = \frac{1}{2}$ and is independent of $\beta$. When the carrier is modulated to generate an FM signal, the power in the sidebands may appear only at the expense of the power originally in the carrier. Another way of arriving at the same conclusion is to make use of the identity[3] $J_0^2 + 2J_1^2 + 2J_2^2 + 2J_3^2 + \cdots = 1$. We calculate the power $P_v$ by squaring $v(t)$ in Eq. (4.5-7) and then averaging $v^2(t)$. Keeping in mind that cross-product terms average to zero, we find, independently of $\beta$, that

$$P_v = \frac{1}{2}\left(J_0^2 + 2\sum_{n=1}^{\infty} J_n^2\right) = \tfrac{1}{2} \tag{4.6-3}$$

as expected. We observe in Fig. 4.6-1 that, at various values of $\beta$, $J_0(\beta) = 0$. At these values of $\beta$ all the power is in the sidebands and none in the carrier.

## 4.7 BANDWIDTH OF A SINUSOIDALLY MODULATED FM SIGNAL

In principle, when an FM signal is modulated, the number of sidebands is infinite and the bandwidth required to encompass such a signal is similarly infinite in extent. As a matter of practice, it turns out that for any $\beta$, so large a fraction of the total power is confined to the sidebands which lie within some finite bandwidth that no serious distortion of the signal results if the sidebands outside this bandwidth are lost. We see in Fig. 4.6-1 that, except for $J_0(\beta)$, each $J_n(\beta)$ hugs the zero axis initially and that as $n$ increases, the corresponding $J_n$ remains very close to the zero axis up to a larger value of $\beta$. For any value of $\beta$ only those $J_n$ need be considered which have succeeded in making a significant departure from the zero axis. How many such sideband components need to be considered may be seen from an examination of Table 4.7-1 where $J_n(\beta)$ is tabulated for various values of $n$ and of $\beta$.

It is found experimentally that the distortion resulting from bandlimiting an FM signal is tolerable as long as 98 percent or more of the power is passed by the bandlimiting filter. This definition of the bandwidth of a filter is, admittedly, somewhat vague, especially since the term "tolerable" means different things in different applications. However, using this definition for bandwidth, one can proceed with an initial tentative design of a system. When the system is built, the

**Table 4.7-1 Values of the Bessel functions $J_n(\beta)$ for various orders $n$ and integral values of $\beta$**

| $n$ \ $\beta$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.7652 | 0.2239 | −0.2601 | −0.3971 | −0.1776 | 0.1506 | 0.3001 | 0.1717 | −0.09033 | −0.2459 |
| 1 | 0.4401 | 0.5767 | 0.3391 | −0.06604 | −0.3276 | −0.2767 | −0.004683 | 0.2346 | 0.2453 | 0.04347 |
| 2 | 0.1149 | 0.3528 | 0.4861 | 0.3641 | 0.04657 | −0.2429 | −0.3014 | −0.1130 | 0.1448 | 0.2546 |
| 3 | 0.01956 | 0.1289 | 0.3091 | 0.4302 | 0.3648 | 0.1148 | −0.1676 | −0.2911 | −0.1809 | 0.05838 |
| 4 | 0.002477 | 0.03400 | 0.1320 | 0.2811 | 0.3912 | 0.3576 | 0.1578 | −0.1054 | −0.2655 | −0.2196 |
| 5 | | 0.007040 | 0.04303 | 0.1321 | 0.2611 | 0.3621 | 0.3479 | 0.1858 | −0.05504 | −0.2341 |
| 6 | | 0.001202 | 0.01139 | 0.04909 | 0.1310 | 0.2458 | 0.3392 | 0.3376 | 0.2043 | −0.01446 |
| 7 | | | 0.002547 | 0.01518 | 0.05338 | 0.1296 | 0.2336 | 0.3206 | 0.3275 | 0.2167 |
| 8 | | | | 0.004029 | 0.01841 | 0.05653 | 0.1280 | 0.2235 | 0.3051 | 0.3179 |
| 9 | | | | | 0.005520 | 0.02117 | 0.05892 | 0.1263 | 0.2149 | 0.2919 |
| 10 | | | | | 0.001468 | 0.006964 | 0.02354 | 0.06077 | 0.1247 | 0.2075 |
| 11 | | | | | | 0.002048 | 0.008335 | 0.02560 | 0.06222 | 0.1231 |
| 12 | | | | | | | 0.002656 | 0.009624 | 0.02739 | 0.06337 |
| 13 | | | | | | | | 0.003275 | 0.01083 | 0.02897 |
| 14 | | | | | | | | 0.001019 | 0.003895 | 0.01196 |
| 15 | | | | | | | | | 0.001286 | 0.004508 |
| 16 | | | | | | | | | | 0.001567 |

bandwidth may thereafter be readjusted, if necessary. In each column of Table 4.7-1, a line has been drawn after the entries which account for at least 98 percent of the power. To illustrate this point, consider $\beta = 1$. Then the power contained in the terms $n = 0$, 1, and 2 is

$$P = \tfrac{1}{2}J_0^2(1) + J_1^2(1) + J_2^2(1)$$
$$= 0.289 + 0.193 + 0.013 = 0.495 \qquad (4.7\text{-}1)$$

The sum 0.495 is 99 percent of the power in the FM signal, which is $\frac{1}{2}$.

We note that the horizontal lines in Table 4.7-1, which indicate the value of $n$ for 98 percent power transmission, always occur just after $n = \beta + 1$. Thus, for sinusoidal modulation the bandwidth required to transmit or receive the FM signal is

$$B = 2(\beta + 1)f_m \qquad (4.7\text{-}2)$$

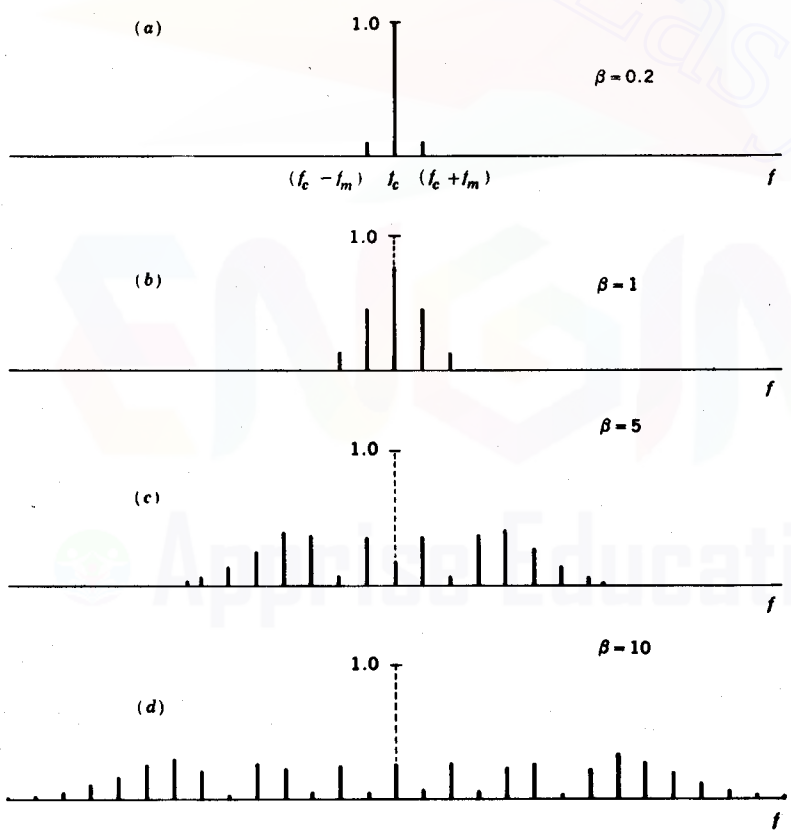



**Figure 4.7-1** The spectra of sinusoidally modulated FM signals for various values of $\beta$.





By way of example, when $\beta = 5$, the sideband components furthest from the carrier which have adequate amplitude to require consideration are those which occur at frequencies $f_c \pm 6f_m$. From the table of Bessel functions published in Jahnke and Emde[2] it may be verified on a numerical basis that the rule given in Eq. (4.7-2) holds without exception up to $\beta = 29$, which is the largest value of $\beta$ for which $J_n$ is tabulated there.

Using Eq. (4.4-3), we may put Eq. (4.7-2) in a form which is more immediately significant. We find

$$B = 2(\Delta f + f_m) \qquad (4.7\text{-}3)$$

Expressed in words, *the bandwidth is twice the sum of the maximum frequency deviation and the modulating frequency.* This rule for bandwidth is called *Carson's rule.*

We deduced Eqs. (4.7-2) and (4.7-3) as a generalization from Table 4.7-1, which begins with $\beta = 1$. We may note, however, that the bandwidth approximation applies quite well even when $\beta \ll 1$. For, in that case, we find that Eq. (4.7-2) gives $B = 2f_m$, which we know to be correct from our earlier discussion of narrowband FM.

The spectra of several FM signals with sinusoidal modulation are shown in Fig. 4.7-1 for various values of $\beta$. These spectra are constructed directly from the entries in Table 4.7-1 except that the signs of the terms have been ignored. The spectral lines have, in every case, been drawn upward even when the corresponding entry is negative. Hence, the lines represent the *magnitudes* only of the spectral components. Not all spectral components have been drawn. Those, far removed from the carrier, which are too small to be drawn conveniently to scale, have been omitted.

## 4.8 EFFECT OF THE MODULATION INDEX $\beta$ ON BANDWIDTH

The modulation index $\beta$ plays a role in FM which is not unlike the role played by the parameter $m$ in connection with AM. In the AM case, and for sinusoidal modulation, we established that to avoid distortion we must observe $m = 1$ as an upper limit. It was also apparent that when it is feasible to do so, it is advantageous to adjust $m$ to be close to unity, that is, 100 percent modulation; by so doing, we keep the magnitude of the recovered baseband signal at a maximum. On this same basis we expect the advantage to lie with keeping $\beta$ as large as possible. For, again, the larger is $\beta$, the stronger will be the recovered signal. While in AM the constraint that $m \leq 1$ is imposed by the necessity to avoid distortion, there is no similar absolute constraint on $\beta$.

There is, however, a constraint which needs to be imposed on $\beta$ for a different reason. From Eq. (4.7-2) for $\beta \gg 1$ we have $B \cong 2\beta f_m$. Therefore the maximum value we may allow for $\beta$ is determined by the maximum allowable bandwidth and the modulation frequency. In comparing AM with FM, we may

then note, in review, that in AM the recovered modulating signal may be made progressively larger subject to the onset of distortion in a manner which keeps the occupied bandwidth constant. In FM there is no similar limit on the modulation, but increasing the magnitude of the recovered signal is achieved at the expense of bandwidth. A more complete comparison is deferred to Chaps. 8 and 9, where we shall take account of the presence of noise and also of the relative power required for transmission.

## 4.9 SPECTRUM OF "CONSTANT BANDWIDTH" FM

Let us consider that we are dealing with a modulating signal voltage $v_m \cos 2\pi f_m t$ with $v_m$ the peak voltage. In a phase-modulating system the phase angle $\phi(t)$ would be proportional to this modulating signal so that $\phi(t) = k' v_m \cos 2\pi f_m t$, with $k'$ a constant. The phase deviation is $\beta = k' v_m$, and, for constant $v_m$, the bandwidth occupied increases linearly with modulating frequency since $B \cong 2\beta f_m = 2k' v_m f_m$. We may avoid this variability of bandwidth with modulating frequency by arranging that $\phi(t) = (k/2\pi f_m) v_m \sin 2\pi f_m t$ ($k$ a constant). For, in this case

$$\beta = \frac{k v_m}{2\pi f_m} \tag{4.9-1}$$

and the bandwidth is $B \cong (2k/2\pi) v_m$, independently of $f_m$. In this latter case, however, the instantaneous frequency is $\omega = \omega_c + k v_m \cos 2\pi f_m t$. Since the instantaneous frequency is proportional to the modulating signal, the initially angle-modulated signal has become a frequency-modulated signal. Thus a signal intended to occupy a nominally constant bandwidth is a frequently-modulated rather than an angle-modulated signal.

In Fig. 4.9-1 we have drawn the spectrum for three values of $\beta$ for the condition that $\beta f_m$ is kept constant. The nominal bandwidth $B \cong 2\,\Delta f = 2\beta f_m$ is consequently constant. The amplitude of the unmodulated carrier at $f_c$ is shown by a dashed line. Note that the extent to which the actual bandwidth extends beyond the nominal bandwidth is greatest for small $\beta$ and large $f_m$ and is least for large $\beta$ and small $f_m$.

In commercial FM broadcasting, the Federal Communications Commission allows a frequency deviation $\Delta f = 75$ kHz. If we assume that the highest audio frequency to be transmitted is 15 kHz, then at this frequency $\beta = \Delta f / f_m = 75/15 = 5$. For all other modulation frequencies $\beta$ is larger than 5. When $\beta = 5$, there are $\beta + 1 = 6$ significant sideband pairs so that at $f_m = 15$ kHz the bandwidth required is $B = 2 \times 6 \times 15 = 180$ kHz, which is to be compared with $2\,\Delta f = 150$ kHz. When $\beta = 20$, there are 21 significant sideband pairs, and $B = 2 \times 21 \times 15/4 = 157.5$ kHz. In the limiting case of very large $\beta$ and correspondingly very small $f_m$, the actual bandwidth becomes equal to the nominal bandwidth $2\,\Delta f$.

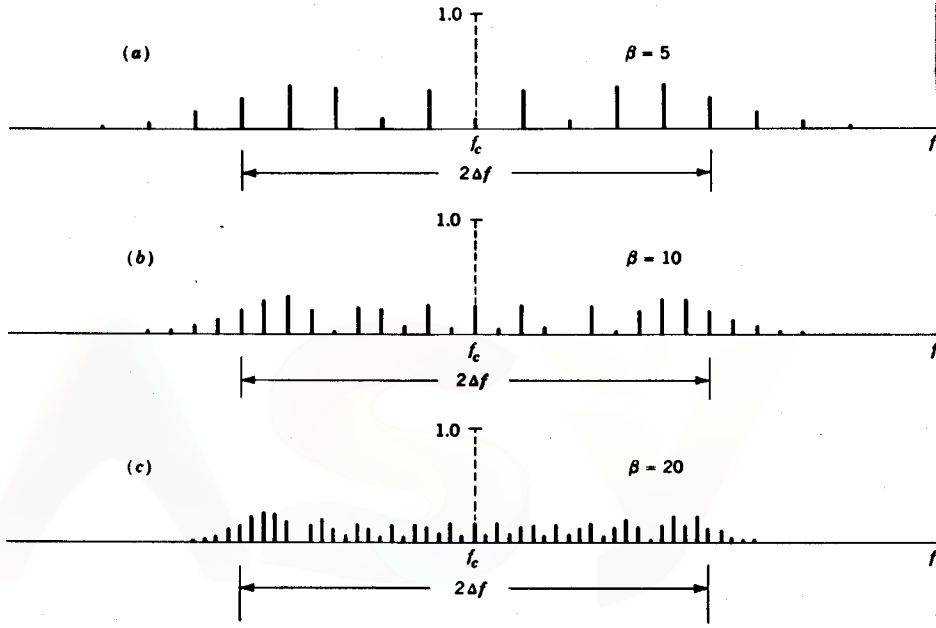



**Figure 4.9-1** Spectra of sinusoidally modulated FM signals. The nominal bandwidth $B \approx 2\beta f_m = 2\,\Delta f$ is kept fixed.

## 4.10 PHASOR DIAGRAM FOR FM SIGNALS

With the aid of a phasor diagram we shall be able to arrive at a rather physically intuitive understanding of how so odd an assortment of sidebands as in Eq. (4.5-7) yields an FM signal of constant amplitude. The diagram will also make clear the difference between AM and narrowband FM (NBFM). In both of these cases there is only a single pair of sideband components.

Let us consider first the case of narrowband FM. From Eqs. (4.4-1), (4.6-1), and (4.6-2) we have for $\beta \ll 1$ that

$$v(t) = \cos(\omega_c t + \beta \sin \omega_m t) \tag{4.10-1a}$$

$$\cong \cos \omega_c t - \frac{\beta}{2} \cos(\omega_c - \omega_m)t + \frac{\beta}{2} \cos(\omega_c + \omega_m)t \tag{4.10-1b}$$

Refer to Fig. 4.10-1*a*. Assuming a coordinate system which rotates counterclockwise at an angular velocity $\omega_c$, the phasor for the carrier-frequency term in Eq. (4.10-1) is fixed and oriented in the horizontal direction. In the same coordinate system, the phasor for the term $(\beta/2) \cos(\omega_c + \omega_m)t$ rotates in a counterclockwise direction at an angular velocity $\omega_m$, while the phasor for the term

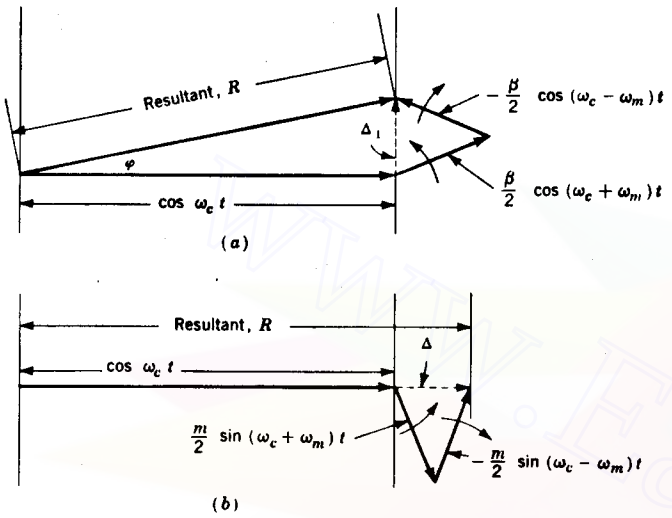



**Figure 4.10-1** (*a*) Phasor diagram for a narrowband FM signal. (*b*) Phasor diagram for an AM signal.

$-(\beta/2)\cos(\omega_c - \omega_m)t$ rotates in a clockwise direction, also at the angular velocity $\omega_m$. At the time $t = 0$, both phasors, which represent the sideband components, have maximum projections in the horizontal direction. At this time one is parallel to, and one is antiparallel to, the phasor representing the carrier, so that the two cancel. The situation depicted in Fig. 4.10-1*a* corresponds to a time shortly after $t = 0$. At this time, the rotation of the sideband phasors which are in opposite directions, as indicated by the curved arrows, have given rise to a sum phasor $\Delta_1$. In the coordinate system in which the carrier phasor is stationary, the phasor $\Delta_1$ always stands perpendicularly to the carrier phasor and has the magnitude

$$\Delta_1 = \beta \sin \omega_m t \tag{4.10-2}$$

The carrier, now slightly reduced in amplitude, and $\Delta_1$ combine to give rise to a resultant $R$. The angular departure of $R$ from the carrier phasor is $\phi$. It is readily seen from Fig. 4.10-1*a* that since $\beta \ll 1$, the maximum value of $\phi \simeq \tan\phi = \beta$, as is to be expected. The small variation in the amplitude of the resultant which appears in Fig. 4.10-1*a* is only the result of the fact that we have neglected higher-order sidebands.

Now let us consider the phasor diagram for AM. The AM signal is

$$(1 + m \sin \omega_m t)\cos \omega_c t = \cos \omega_c t + \frac{m}{2}\sin(\omega_c + \omega_m)t - \frac{m}{2}\sin(\omega_c - \omega_m)t \tag{4.10-3}$$

and the individual terms are represented as phasors in Fig. 4.10-1*b*. Comparing Eqs. (4.10-1) and (4.10-3), we see that there is a 90° phase shift in the phases of the sidebands between the FM and AM cases. In Fig. 4.10-1*b* the sum $\Delta$ of the sideband phasors is given by

$$\Delta = m \sin \omega_m t \tag{4.10-4}$$

The important difference between the FM and AM cases is that in the former the sum $\Delta_1$ is always perpendicular to the carrier phasor, while in the latter the sum $\Delta$ is always parallel to the carrier phasor. Hence in the AM case, the resultant $R$ does not rotate with respect to the carrier phasor but instead varies in amplitude between $1 + m$ and $1 - m$.

Another way of looking at the difference between AM and NBFM is to note that in NBFM where $\beta \ll 1$

$$v(t) \approx \cos \omega_c t - \beta \sin \omega_m t \sin \omega_c t \tag{4.10-5}$$

while in AM

$$v(t) = \cos \omega_c t + m \sin \omega_m t \cos \omega_c t \tag{4.10-6}$$

Note that in NBFM the first term is $\cos \omega_c t$, while the second term involves $\sin \omega_c t$, a *quadrature* relationship. In AM both first and second terms involve $\cos \omega_c t$, an *in-phase* relationship.

To return now to the FM case and to Fig. 4.10-1*a*, the following point is worth noting. When the angle $\phi$ completes a full cycle, that is, $\Delta_1$ varies from $+\beta$ to $-\beta$ and back again to $+\beta$, the magnitude of the resultant $R$ will have executed *two* full cycles. For $R$ is a maximum at $\Delta_1 = \beta$, a minimum at $\Delta_1 = 0$, a maximum again when $\Delta_1 = -\beta$, and so on. On this basis, it may well be expected that if an additional sideband pair is to be added to the first to make $R$ more nearly constant, this new pair must give rise to a resultant $\Delta_2$ which varies at the frequency $2\omega_m$. Thus, we are not surprised to find that as the phase deviation $\beta$ increases, a sideband pair comes into existence at the frequencies $\omega_c \pm 2\omega_m$.

As long as we depend on the first-order sideband pair only, we see from Fig. 4.10-1*a* that $\phi$ cannot exceed 90°. A deviation of such magnitude is hardly adequate. For consider, as above, that $\Delta f = 75$ kHz and that $f_m = 50$ Hz. Then $\omega_m = 75{,}000/50 = 1500$ rad, and the resultant $R$ must, in this case, spin completely about $1500/2\pi$ or about 240 times. Such wild whirling is made possible through the effect of the higher-order sidebands. As noted, the first-order sideband pair gives rise to a phasor $\Delta_1 = J_1(\beta)\sin\omega_m t$, which phasor is perpendicular to the carrier phasor. It may also be established by inspection of Eq. (4.5-7) that the second-order sideband pair gives rise to a phasor $\Delta_2 = J_2(\beta)\cos 2\omega_m t$ and that this phasor is *parallel* to the carrier phasor. Continuing, we easily establish that all odd-numbered sideband pairs give rise to phasors

$$\Delta_n = J_n(\beta)\sin n\omega_m t \qquad n \text{ odd} \tag{4.10-7}$$


Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

which are perpendicular to the carrier phasor, while all even-numbered sideband pairs give rise to phasors

$$\Delta_n = J_n(\beta) \cos n\omega_m t \qquad n \text{ even} \tag{4.10-8}$$

which are parallel to the carrier phasor. Thus, phasors $\Delta_1$, $\Delta_2$, $\Delta_3$, etc., alternately perpendicular and parallel to the carrier phasor, are added to carry the end point of the resultant phasor $R$ completely around as many times as may be required, while maintaining $R$ at constant magnitude. It is left as an exercise for the student to show by typical examples how the superposition of a carrier and sidebands may swing a constant-amplitude resultant around through an arbitrary angle (Prob. 4.10-2).

## 4.11 SPECTRUM OF NARROWBAND ANGLE MODULATION: ARBITRARY MODULATION

Previously we considered the spectrum, in NBFM, which is produced by sinusoidal modulation. We found that, just as in AM, such modulation gives rise to two sidebands at frequencies $\omega_c + \omega_m$ and $\omega_c - \omega_m$. We extend the result now to an arbitrary modulating waveform.

We may readily verify (Prob. 4.11-1) that superposition applies in narrowband angle modulation just as it does to AM. That is, if $\beta_1 \sin \omega_1 t + \beta_2 \sin \omega_2 t$ is substituted in Eq. (4.10-1*a*) in place of $\beta \sin \omega_m t$, the sidebands which result are the sum of the sidebands that would be yielded by either modulation alone. Hence even if a modulating signal of waveform $m(t)$, with a continuous distribution of spectral components, is used in either AM or narrowband angle modulation the forms of the sideband spectra will be the same in the two cases.

More formally, we have in AM, when the modulating waveform is $m(t)$, the signal is

$$v_{AM}(t) = A[1 + m(t)] \cos \omega_c t = A \cos \omega_c t + Am(t) \cos \omega_c t \tag{4.11-1}$$

Let us assume, for simplicity, that $m(t)$ is a finite energy waveform with a Fourier transform $M(j\omega)$. We use the theorem that if the Fourier transform $\mathcal{F}[m(t)] = M(j\omega)$, then $\mathcal{F}[m(t) \cos \omega_c t]$ is as given in Eq. (3.2-4). We then find that

$$\mathcal{F}[v_{AM}(t)] = \frac{A}{2}[\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] + \frac{A}{2}[M(j\omega + j\omega_c) + M(j\omega - j\omega_c)] \tag{4.11-2}$$

The narrowband angle-modulation signal of Eq. (4.10-1), except of amplitude $A$ and with phase modulation $m(t)$, may be written, for $|m(t)| \ll 1$,

$$v_{PM}(t) \cong A \cos \omega_c t - Am(t) \sin \omega_c t \tag{4.11-3}$$

so that

$$\mathcal{F}[v_{PM}(t)] = \frac{A}{2}[\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] + \frac{A}{2} e^{-j\pi/2}[M(j\omega + j\omega_c) - M(j\omega - j\omega_c)] \tag{4.11-4}$$

Comparing Eq. (4.11-2) with Eq. (4.11-4), we observe that

$$|\mathcal{F}[v_{AM}]|^2 = |\mathcal{F}[v_{PM}]|^2 \tag{4.11-5}$$

Thus, if we were to make plots of the energy spectral densities of $v_{AM}(t)$ and of $v_{PM}(t)$, we would find them identical. Similarly, if $m(t)$ were a signal of finite power, we would find that plots of power spectral density would be the same.

## 4.12 SPECTRUM OF WIDEBAND FM (WBFM): ARBITRARY MODULATION[4]

In this section we engage in a heuristic discussion of the spectrum of a wideband FM signal. We shall not be able to deduce the spectrum with the precision that is possible in the NBFM case described in the previous section. As a matter of fact, we shall be able to do no more than to deduce a means of expressing approximately the power spectral density of a WBFM signal. But this result is important and useful.

Previously, to characterize an FM signal as being narrowband or wideband, we had used the parameter $\beta \equiv \Delta f/f_m$, where $\Delta f$ is the frequency deviation and $f_m$ the frequency of the sinusoidal modulating signal. The signal was then NBFM or WBFM depending on whether $\beta \ll 1$ or $\beta \gg 1$. Alternatively we distinguished one from the other on the basis of whether one or very many sidebands were produced by each spectral component of the modulating signal, and on the basis of whether or not superposition applies. We consider now still another alternative.

Let the symbol $\nu \equiv f - f_c$ represent the frequency difference between the instantaneous frequency $f$ and the carrier frequency $f_c$; that is, $\nu(t) = (k/2\pi)m(t)$ [see Eq. (4.3-5)]. The period corresponding to $\nu$ is $T = 1/\nu$. As $f$ varies, so also will $\nu$ and $T$. The frequency $\nu$ is the frequency with which the resultant phasor $R$ in Fig. 4.10-1 rotates in the coordinate system in which the carrier phasor is fixed. In WBFM this resultant phasor rotates through many complete revolutions, and its speed of rotation does not change radically from revolution to revolution. Since, the resultant $R$ is constant, then if we were to examine the plot as a function of time of the projection of $R$ in, say, the horizontal direction, we would recognize it as a sinusoidal waveform because its frequency would be changing very slowly. No appreciable change in frequency would take place during the course of a cycle. Even a long succession of cycles would give the appearance of being of rather constant frequency. In NBFM, on the other hand, the phasor $R$ simply oscillates about the position of the carrier phasor. Even though, in this case, we may still formally calculate a frequency $\nu$, there is no corresponding time interval

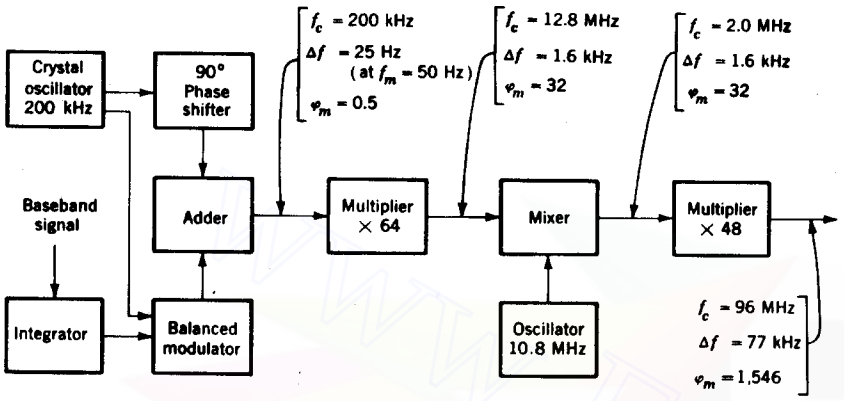



**Figure 4.19-1** Block diagram of an Armstrong system of generating an FM signal using multipliers to increase the frequency deviation.

before multiplication, to the extent of $\phi_m = 0.5$. Thus, at $f_m = 50$ Hz, we have $\Delta f = 25$ Hz. Note that at higher modulating frequencies, $\phi_m$ is less than 0.5 rad. The carrier frequency before multiplication has been selected at 200 kHz, a frequency at which very stable crystal oscillators and balanced modulators are readily constructed.

As already noted, if we require that $\Delta f \approx 75$ kHz, then a multiplication by a factor of 3000 is required. In Fig. 4.19-1 the multiplication is actually $3072 (= 64 \times 48)$. The values were selected so that the multiplication may be done by factors of 2 and 3, that is, $64 = 2^6$, $48 = 3 \times 2^4$. Direct multiplication would yield a signal of carrier frequency

$$200 \text{ kHz} \times 3072 = 614.4 \text{ MHz}$$

This signal might then be heterodyned with a signal of frequency, say, $614.4 - 96.0 = 518.4$ MHz. The difference signal output of such a mixer would be a signal of carrier frequency 96 MHz. Note particularly that a mixer, since it yields sum and difference frequencies, will translate the frequency spectrum of an FM signal but will have no effect on its frequency deviation. In the system of Fig. 4.19-1, in order to avoid the inconvenience of heterodyning at a frequency in the range of hundreds of megahertz, the frequency translation has been accomplished at a point in the chain of multipliers where the frequency is only in the neighborhood of approximately 10 MHz.

A feature, not indicated in Fig. 4.19-1 but which may be incorporated, is to derive the 10.8 MHz mixing signal not from a separate oscillator but rather through multipliers from the 0.2 MHz crystal oscillator. The multiplication required is $10.8/0.2 = 54 = 2 \times 3^3$. Such a derivation of the 10.8-MHz signal will suppress the effect of any drift in the frequency of this signal (see Prob. 4.19-1).





## 4.20 FM DEMODULATORS

With a view toward describing how we can recover the modulating signal from a frequency-modulated carrier we consider the situation represented in Fig. 4.20-1. Here a waveform of frequency $f_0$ and input amplitude $A_i$ is applied to a frequency selective network which then yields an output of amplitude $A_o$. The ratio of amplitudes $A_o/A_i$ is the absolute value of the transfer function of the network, that is, $|H(j\omega)|$. This output waveform is then applied to a diode AM demodulator (see Fig. 3.4-2). The diode demodulator generates an output which is equal to the peak value of the sinusoidal input so that the diode demodulator output is equal to $A_o$. Suppose now that the input waveform, instead of being of fixed frequency $f_0$, is actually a frequency modulated waveform. Then even for a fixed input amplitude $A_i$, the output amplitude $A_o$ will not remain fixed but instead be modulated because of the frequency selectivity of the transmission network. Correspondingly the diode demodulator output will follow the variation of $A_o$.

In short, a fixed amplitude, frequency-modulated input will generate, at the output of the frequency selective network, a waveform which is not only frequency modulated but also amplitude modulated. The diode demodulator will ignore the frequency modulation but will respond to the amplitude modulation. (In general, the frequency-selective network will not only give rise to an amplitude change but will also generate a frequency-dependent phase change, as noted in Fig. 4.20-1. But such a phase change is simply additional angle modulation which the diode demodulator will ignore.)

What we require of an FM demodulator is that the instantaneous output signal $A_o$ be proportional to the instantaneous frequency deviation of the received signal from the carrier frequency. If the carrier frequency is $f_0$ then we require a linear relationship between $A_o$ and $(f - f_0)$ where $f$ is the instantaneous frequency. Such a linear relationship is indicated in the plot of Fig. 4.20-1*b*. We

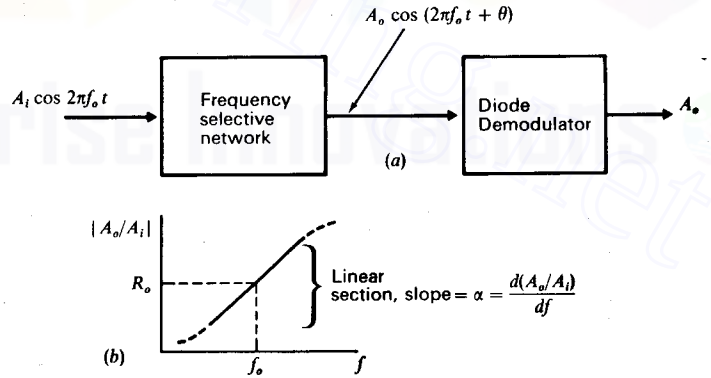



**Figure 4.20-1** (*a*) FM demodulation. (*b*) Frequency selective network, typically an LC circuit.

require actually that the linearity extend only as far as is necessary to accommodate the maximum frequency deviation to which the carrier is subject.

As indicated in Fig. 4.20-1*b* let us consider that the frequency selective network has a linear transfer characteristic of slope $\alpha$ over an adequate range in the neighborhood of the carrier frequency $f_0$, and that, at $f = f_0$, $|A_o/A_i| = R_0 = |H(f = f_0)|$. Then we shall have, as required

$$A_o = R_0 A_i + \alpha A_i(f - f_0) \tag{4.20-1}$$

the term $R_0 A_i$ in Eq. (4.20-1) is a term of fixed value which displays no response to frequency deviation and the second term $\alpha A_i(f - f_0)$ provides the required response to the instantaneous frequency deviation of the input frequency-modulated signal.

We observe however from Eq. (4.20-1) that if the amplitude $A_i$ of the input signal is not fixed then the demodulator output will respond to the input amplitude variations as well as to frequency deviations. Ordinarily in a frequency-modulated communication system the amplitude of the transmitted signal will not be modulated deliberately so that any such modulation which does appear will be due to noise. Hence it is of advantage, for the purpose of suppressing the noise, to reduce the dependence of $A_o$ on $A_i$ in Eq. (4.20-1). This is accomplished by passing the incoming signal through a hard limiter as shown in Fig. 4.20-2. The purpose of the hard limiter (or comparator) is to insure that variations of $A_i(t)$ are removed. Figure 4.20-2 shows the original FM waveform with amplitude variations at the output of the hard limiter. Since the waveform has been reduced to a frequency modulated "square wave" a bandpass filter is inserted to extract to first harmonic frequency $f_0$. The resulting FM waveform is now applied to the FM demodulator.

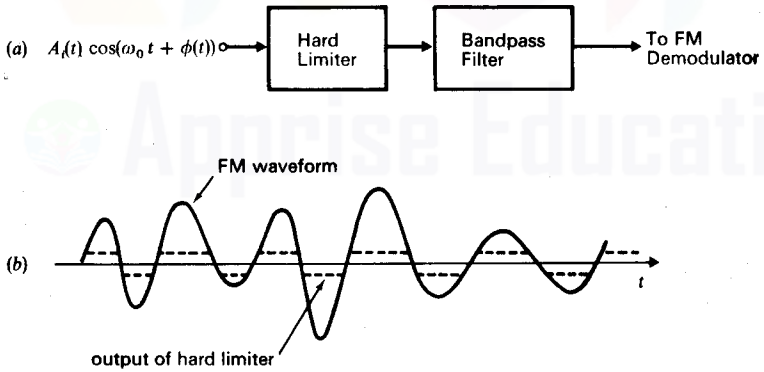



**Figure 4.20-2** (*a*) Hard limiter (or comparator) input to FM demodulator. (*b*) FM waveform at input and output of hard limiter.





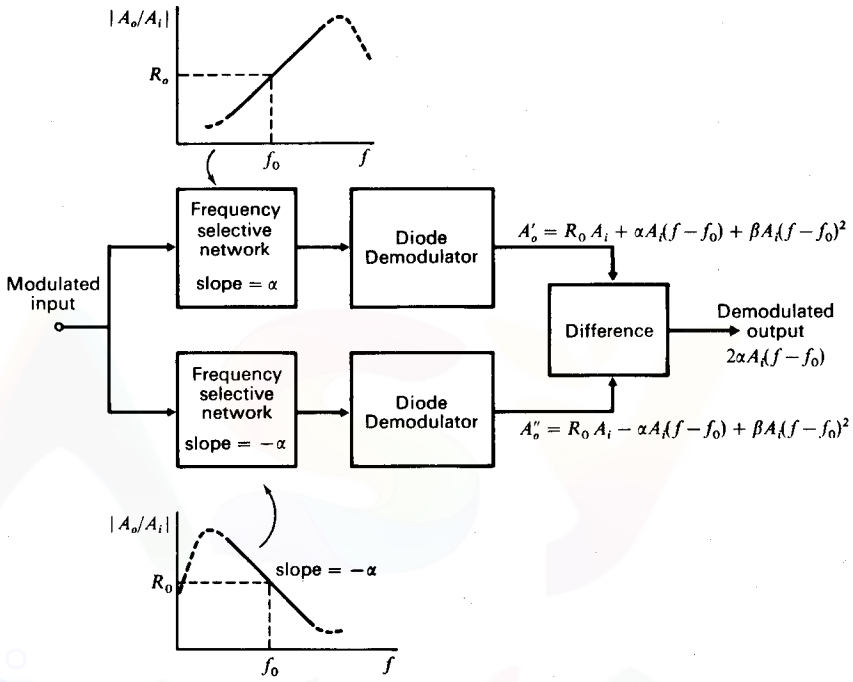



**Figure 4.20-3** A balanced FM demodulator.

Unfortunately, one cannot construct a network having the precisely linear transfer characteristic shown in Eq. (4.20-1). Indeed, in practical networks the output amplitude appears as

$$A_o = R_0 A_i + \alpha A_i(f - f_0) + \beta A_i(f - f_0)^2 + \cdots \tag{4.20-2}$$

The balanced FM demodulator shown in Fig. 4.20-3 can be used to remove the constant term $R_0 A_i$ and all even harmonics, thereby reducing the distortion produced by the nonlinearity of the bandpass filters. Here two demodulators are employed, differing only in that in one case the frequency-selective network has a slope $\alpha$ and in the other the slope is $-\alpha$. The output provided by this balanced modulator is the difference between the two individual demodulators. These individual outputs are (assuming a second order nonlinearity):

$$A'_o = R_0 A_i + \alpha A_i(f - f_0) + \beta A_i(f - f_0)^2 \tag{4.20-3}$$

and

$$A''_o = R_0 A_i - \alpha A_i(f - f_0) + \beta A_i(f - f_0)^2 \tag{4.20-4}$$

The difference output is then

$$A_o = 2\alpha A_i(f - f_0) \tag{4.20-5}$$

and we see that the linearity of the output of the FM demodulator has been improved.

In practice an LC tuned circuit is used as the frequency selective network. The relationship between the center frequency of each of the two tuned circuits and their bandwidths on the linearity of the FM demodulator is explored in Prob. 4.20-4.

It has also been shown in the literature that a hard limiter is not required when using a balanced discriminator and that any overdriven amplifier (i.e., an amplifier which is driven betweeen cutoff and saturation) can be used.

## 4.21 APPROXIMATELY COMPATIBLE SSB SYSTEMS

### SSB-AM

We noted earlier in Sec. 3.12 the advantages that would accrue from the availability of compatible single-sideband systems. While precisely compatible systems are presently impractical, approximately compatible systems are feasible, such systems are in use, and commercial equipment for such systems is available. In a strictly compatible AM system a waveform would be generated whose envelope exactly reproduces the baseband signal. Additionally if the highest baseband frequency component is $f_M$, the frequency range encompassed by the compatible signal would extend from $f_c$, the carrier frequency, to, say, $f_c + f_M$. An approximately compatible system is one in which there is some relaxation of these specifications concerning envelope shape or spectral range. On the basis of our discussion of FM-type waveforms we are now able briefly and qualitatively to discuss the principle of operation of one type of approximately compatible AM system.

We saw in Sec. 3.11 that if we suppressed one sideband component of an amplitude-modulated waveform, the envelope of the resultant waveform would still have the form of the modulating signal, provided the percentage modulation was kept small. Consider now the phasor diagram of the AM signal shown in Fig. 4.10-1*b*. From this phasor diagram it is apparent that when one of the sidebands is suppressed, the resultant is a waveform which is modulated in *both amplitude and phase*. The amplitude will vary between $1 + m/2$ and $1 - m/2$. The resultant $R$ will no longer always be parallel to the carrier phasor. Instead it will rotate clockwise by an angle $\phi$ such that $\tan \phi = m/2$ and rotate counterclockwise by a similar angle.

Thus we find that a carrier of fixed frequency (or phase), when amplitude-modulated, gives rise to two sidebands. But an angle-modulated carrier, when amplitude-modulated, may give rise to a single sideband. Suppose then that we arrange to amplitude-modulate a carrier in such manner that its envelope faithfully reproduces the modulating signal. Is it then possible to also angle-modulate that carrier so that a single sideband results? It turns out that, to a good approximation, the answer is yes!

In the system described by Kahn[5] the modulating signal modulates not only the amplitude but the carrier phase as well. The relationship between phase and modulating signal is *nonlinear* and has been determined, at least in part experimentally, on the basis of system performance. An analysis of the system is very involved because of the two nonlinearities involved: the inherent nonlinearity of FM and the additional nonlinearity between phase and modulating signal. The system is not strictly single sideband in the sense that a modulating tone gives rise not to a single side tone but to a spectrum of side tones. However, all the side tones are on the *same side* of the carrier. The predominant side tone is separated from the carrier by the tone frequency, and the others are separated by multiples of the modulating-tone frequency.

The system is able to operate, in effect, as a single-sideband system because of the characteristics of speech or music for which it is intended, and because the side tones, other than the predominant one, fall off sharply in power content with increasing harmonic number. Most of the power in sound is in the lower-frequency ranges. High-power low-frequency spectral components in sound may give rise to numerous harmonic side tones, but because of the low frequency of the fundamental, the harmonics will still fall in the audio spectrum. On the other hand, high-frequency tones, which may give rise to harmonic side tones that may fall outside the allowed spectral range, are of very small energy. The overall result is that there may well be spectral components that fall outside the spectral range allowable in a spectrum-conserving single-sideband system. However, it has been shown experimentally that such components are small enough to cause no interference with the signal in an adjacent single-sideband channel.

### SSB-FM

A compatible SSB-FM signal has frequency components extending either above or below the carrier frequency. In addition it can be demodulated using a standard limiter-discriminator. The compatible FM signal is constructed by adding amplitude modulation to the frequency-modulated signal. Since the limiter removes any and all amplitude modulation, the addition of the AM does not affect the recovery of the modulation by the discriminator. By properly adjusting the amplitude modulation, either the upper or lower sideband can be removed.

## 4.22 STEREOPHONIC FM BROADCASTING

In *monophonic* broadcasting of sound, a single audio baseband signal is transmitted from broadcasting studio to a home receiver. At the receiver, the audio signal is applied to a loudspeaker which then reproduces the original sound. The orig-

Turning now to the composite signal $M(t)$ in Eq. (4.22-1), we note that $\cos 2\pi f_{sc} t$ oscillates rapidly between $+1$ and $-1$, and, ignoring temporarily the pilot carrier, we have that $M(t)$ oscillates rapidly between $M(t) = 2L(t)$ and $M(t) = 2R(t)$. The maximum attained by $M(t)$ is then $M_m = 2L_m$ or $M_m = 2R_m$. From Eq. (4.22-2) $M_m = V_{sm}$. Hence, in summary, we find that the addition of the difference signal $V_d$ to the sum signal $V_s$ does not increase the peak signal excursion.

### Effect of the Pilot Carrier

Unlike the DSB-SC signal, the pilot carrier, when added to the other components of the composite modulating signal, does produce an increase in peak excursion. Hence the addition of the pilot carrier calls for a reduction in the sound signal modulation level. A low-level pilot carrier allows greater sound signal modulation, while a high-level pilot carrier eases the burden of extracting the pilot carrier at the receiver. As an engineering compromise, the FCC standards call for a pilot carrier of such level that the peak sound modulation amplitude has to be reduced to about 90 percent of what would be allowed in the absence of a carrier. This 10 percent reduction corresponds to a loss in signal level of less than 1 dB.

## REFERENCES


1. Bell Telephone Laboratories: "Transmission Systems for Communications," Western Electric Company, Tech. Pub., Winston-Salem, N.C., 1964.
2. Jahnke, E., and F. Emde: "Tables of Functions," Dover Publications Inc., New York, 1945.
3. Pipes, L. A.: "Applied Mathematics for Engineers and Physicists," McGraw-Hill Book Company, New York, 1958.
4. Blachman, N.: Calculation of the Spectrum of an FM Signal Using Woodwards Theorem, *IEEE Trans. Communication Technology*, August, 1969.
5. Kahn, L. R.: Compatible Single Sideband. *Proc. IRE*, vol. 49, pp. 1503–1527, October, 1961.


## PROBLEMS

**4.2-1.** Consider the signal $\cos [\omega_c t + \phi(t)]$ where $\phi(t)$ is a square wave taking on the values $\pm\pi/3$ every $2/f_c$ sec.

(*a*) Sketch $\cos [\omega_c t + \phi(t)]$.

(*b*) Plot the phase as a function of time.

(*c*) Plot the frequency as a function of time.

**4.2-2.** If the waveform $\cos (\omega_c t + k \sin \omega_m t)$ is a phase-modulated carrier, sketch the waveform of the modulating signal. Sketch the waveform of the modulating signal if the carrier is frequency-modulated.

**4.3-1.** What are the dimensions of the constants $k'$, $k''$, and $k$ that appear in Eqs. (4.3-1), (4.3-2), and (4.3-3)?

**4.4-1.** An FM signal is given by

$$v(t) = \cos \left[ \omega_c t + \sum_{k=1}^{K} \beta_k \cos (k\omega_0 t + \theta_k) \right]$$

(*a*) If $\theta_k = 0$ and $K = 1, 2$, find the maximum frequency deviations.

(*b*) If each $\theta_k$ is an independent random variable, uniformly distributed between $-\pi$ and $\pi$, find the rms frequency deviation.

(*c*) Under the condition of (*b*) calculate the rms phase deviation.

**4.4-2.** If $v(t) = \cos \left[ \omega_c t + k \int_{-\infty}^{t} m(\lambda)\, d\lambda \right]$, where $m(t)$ has a probability density

$$f(m) = \frac{1}{\sqrt{2\pi}} e^{-m^2/2}$$

calculate the rms frequency deviation.

**4.4-3.** A carrier which attains a peak voltage of 5 volts has a frequency of 100 MHz. This carrier is frequency-modulated by a sinusoidal waveform of frequency 2 kHz to such extent that the frequency deviation from the carrier frequency is 75 kHz. The modulated waveform passes through zero and is increasing at time $t = 0$. Write an expression for the modulated carrier waveform.

**4.4-4.** A carrier of frequency $10^6$ Hz and amplitude 3 volts is frequency-modulated by a sinusoidal modulating waveform of frequency 500 Hz and of peak amplitude 1 volt. As a consequence, the frequency deviation is 1 kHz. The level of the modulating waveform is changed to 5 volts peak, and the modulating frequency is changed to 2 kHz. Write the expression for the new modulated waveform.

**4.5-1.** A carrier is angle-modulated by two sinusoidal modulating waveforms simultaneously so that

$$v(t) = A \cos (\omega_c t + \beta_1 \sin \omega_1 t + \beta_2 \sin \omega_2 t)$$

Show that this waveform has sidebands separated from the carrier not only at multiples of $\omega_1$ and of $\omega_2$ but also has sidebands as well at separations of multiples of $\omega_1 + \omega_2$ and of $\omega_1 - \omega_2$.

**4.6-1.** Bessel functions are said to be *almost periodic* with a period of almost $2\pi$. Demonstrate this by recording the values of $\beta$, for $J_0(\beta)$ and $J_1(\beta)$, required to make these functions equal to zero.

**4.6-2.** The primary difference between the Bessel functions and the sine wave is that the envelope of the Bessel function decreases.

(*a*) Tabulate the magnitude of all peak values of $J_0(\beta)$, positive and negative peaks, as a function of $\beta$.

(*b*) Plot the magnitude of the peak values obtained in part (*a*) versus $\beta$ and draw a smooth curve through the points.

(*c*) Show that the magnitude decreases as $\frac{1}{\sqrt{\beta}}$.

**4.6-3.** An FM carrier is sinusoidally modulated. For what values of $\beta$ does all the power lie in the sidebands (i.e., no power in the carrier)?

**4.7-1.** A bandwidth rule sometimes used for space communication systems is $B = (2\beta + 1)f_M$. What fraction of the signal power is included in that frequency band. Consider $\beta = 1$ and 10.

**4.7-2.** A carrier is frequency-modulated by a sinusoidal modulating signal of frequency 2 kHz, resulting in a frequency deviation of 5 kHz. What is the bandwidth occupied by the modulated waveform? The amplitude of the modulating sinusoid is increased by a factor of 3 and its frequency lowered to 1 kHz. What is the new bandwidth?

**4.7-3.** Plot the spectrum of $\cos (2\pi \times 4t + 5 \sin 2\pi t)$. Note that the spectrum indicates the presence of a dc component. Plot the waveform as a function of time to indicate that the dc component is to have been expected.

**4.10-1.** $v(t) = \cos \omega_c t + 0.2 \cos \omega_m t \sin \omega_c t$.

(*a*) Show that $v(t)$ is a combination AM-FM signal.

(*b*) Sketch the phasor diagram at $t = 0$.


Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

**4.10-2.** Consider the angle-modulated waveform $\cos(\omega_c t + 2\sin\omega_m t)$, i.e., $\beta = 2$, so that the waveform may be approximated by a carrier and three pairs of sidebands. In a coordinate system in which the carrier phasor $\Delta_0$ is at rest, determine the phasors $\Delta_1$, $\Delta_2$, and $\Delta_3$, representing respectively the first, second, and third sideband pairs. Draw diagrams combining the four phasors for the cases $\omega_m t = 0, \pi/4, \pi/2, 3\pi/4$, and $\pi$. For each case calculate the magnitude of the resultant phasor.

**4.10-3.** (*a*) Show with a phasor diagram that $v(t)$ given by

$$v(t) = \cos(2\pi \times 10^6 t) + 0.02\cos[2\pi \times (10^6 + 10^3)t]$$

represents a carrier which is modulated both in amplitude and frequency.

(*b*) Show that, on the basis of the relative magnitudes of the two terms in $v(t)$, the amplitude and the frequency variations both vary approximately sinusoidally with time with frequency $10^3$ Hz.

(*c*) Express $v(t)$ in the form

$$v(t) \approx (1 + m\cos 2\pi \times 10^3 t)\cos(2\pi \times 10^6 t + \beta \sin 2\pi \times 10^3 t)$$

Find $m$ and $\beta$. Write an expression for the instantaneous frequency as a function of time.

**4.10-4.** Consider the angle-modulated waveform $\cos(\omega_c t + 6\sin\omega_m t)$, i.e., $\beta = 6$, so that the waveform may be approximated by a carrier and seven pairs of sidebands. In a coordinate system in which the carrier phasor $\Delta_0$ is at rest, determine the phasor $\Delta_1$, $\Delta_2$, etc., representing the first, second, etc., sideband pairs at a time when $\sin\omega_m t = 1$. For this time draw a phasor diagram showing each phasor $\Delta_i$ and the resultant phasor.

**4.11-1.** Verify the comment made in Sec. 4.11 that superposition applies in NBFM. To do this consider $v(t) = \cos[\omega_c t + \phi(t)]$ where $\phi(t) = \beta_1 \sin\omega_1 t + \beta_2 \sin\omega_2 t$. Let $\beta_1$ and $\beta_2$ be sufficiently small so that $|\phi(t)| \ll \pi/2$. Show that $v(t) \simeq \cos\omega_c t - (\beta_1 \sin\omega_1 t + \beta_2 \sin\omega_2 t)\sin\omega_c t$.

**4.12-1.** The frequency of a laboratory oscillator is varied back and forth extremely slowly and at a uniform rate between the frequencies of 99 and 101 kHz. The amplitude of the oscillator output is constant at 2 volts. Make a plot of the two-sided power spectral density of the oscillator output waveform.

**4.12-2.** If the probability density of the amplitude of $m(t)$ is Rayleigh:

$$f(m) = \begin{cases} me^{-m^2/2} & m \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find the power spectral density $G_v(f)$ of the FM signal

$$v(t) = \cos\left[\omega_c t + k\int_{-\infty}^{t} m(\lambda)\,d\lambda\right]$$

**4.12-3.** Repeat Prob. 4.12-2 if the probability density of $m(t)$ is $f(m) = \frac{1}{2}e^{-|m|}$.

**4.12-4.** The frequency of a laboratory oscillator is varied back and forth extremely slowly and in such a manner that the instantaneous frequency of the oscillator varies sinusoidally with time between the limits of 99 and 101 kHz. The amplitude of the oscillator output is constant at 2 volts.

(*a*) Find a function of frequency $g(f)$ such that $g(f)\,df$ is the fraction of the time that the instantaneous frequency is in the range between $f$ and $f + df$.

(*b*) Make a plot of the two-sided power spectral density of the oscillator output waveform.

**4.13-1.** Consider that the WBFM signal having the power spectral density of Eq. (4.13-1) is filtered by a gaussian filter having the bandpass characteristic

$$|H(f)|^2 = e^{-(f-f_c)^2/2B^2} + e^{-(f+f_c)^2/2B^2}$$

Assume $f_c \gg B$:

(*a*) Sketch $|H(f)|^2$ as a function of $f$.

(*b*) Calculate the 3-dB bandwidth of the filter in terms of $B$.

(*c*) Find $B$ so that 98 percent of the signal power of the WBFM signal is passed.





**4.13-2.** The two independent modulating signals $m_1(t)$ and $m_2(t)$ are both gaussian and both of zero mean and variance 1 volt$^2$. The modulating signal $m_1(t)$ is connected to a source which can be frequency-modulated in such manner that, when $m_1(t) = 1$ volt (constant), the source frequency, initially 1 MHz, increases by 3 kHz. The modulating signal $m_2(t)$ is connected in such manner that, when $m_2(t) = 1$ volt (constant), the source frequency decreases by 4 kHz. The carrier amplitude is 2 volts. The two modulating signals are applied simultaneously. Write an expression for the power spectral density of the output of the frequency-modulated source.

**4.14-1.** Consider the FM signal

$$v(t) = \cos\left[\omega_c t + \sum_{k=1}^{K} \beta_k \cos(\omega_k t + \theta_k)\right]$$

Let $\beta_k \omega_k = 1$ for each $k$.

(*a*) Find $B$ if $B \equiv 2[(\Delta f)_1 + (\Delta f)_2 + \cdots + (\Delta f)_K]$.

(*b*) Find $B$ if $B \equiv 2[(\Delta f)_{1_{rms}} + (\Delta f)_{2_{rms}} + \cdots + (\Delta f)_{K_{rms}}]$.

**4.14-2.** If $G(f)$ is gaussian and is given by Eq. (4.13-1), find the rms bandwidth $B$. Compare your result with the value $B = 4.6\,\Delta f_{rms}$ given in Eq. (4.13-5).

**4.15-1.** In Fig. 4.15-1 the voltage-variable capacitor is a reversed-biased *pn* junction diode whose capacitance is related to the reverse-biasing voltage $v$ by $C_v = (100/\sqrt{1 + 2v})$ pF. The capacitance $C_0 = 200$ pF and $L$ is adjusted for resonance at 5 MHz when a fixed reverse voltage $v = 4$ volts is applied to the capacitor $C_v$. The modulating voltage is $m(t) = 4 + 0.045\sin 2\pi \times 10^3 t$. If the oscillator amplitude is 1 volt, write an expression for the angle-modulated output waveform which appears across the tank circuit.

**4.17-1.** (*a*) In the multiplier circuit of Fig. 4.17-1 assume that the transistor acts as a current source and is so biased and so driven that the collector current consists of alternate half-cycles of a sinusoidal waveform with a peak value of 50 mA. The input frequency of the driving signal is 1 MHz, and the multiplication by a factor of 3 is to be accomplished. If $C = 200$ pF and the inductor $Q = 30$, find the inductance of the inductor and calculate the amplitude of the third harmonic voltage across the tank.

(*b*) If multiplication by 10 is to be accomplished, calculate the amplitude of the tank voltage. Assume that the resonant impedance of the tank remains the same as in part (*a*).

**4.18-1.** (*a*) Consider the narrowband waveform $v(t) = \cos(\omega_c t + \beta\sin\omega_m t)$, with $\beta \ll 1$ and with $\omega_m \ll \omega_c$. Show that $v(t)$, which has a frequency deviation $\Delta f = \beta f_m$, may be written approximately as

$$v(t) = \cos\omega_c t - \beta/2\cos(\omega_c - \omega_m)t + \beta/2\cos(\omega_c + \omega_m)t$$

and that this approximation is consistent with the general expansion for an angle-modulated waveform as given by Eq. (4.5-7). Use the approximations of Eqs. (4.6-1) and (4.6-2).

(*b*) Let $v(t)$ be applied as the input to a device whose output is $v^2(t)$ (i.e., the device is nonlinear and is to be used for frequency multiplication by a factor of 2). Square the approximate expression for $v(t)$ as given in part (*a*). Compare the spectrum of $v^2(t)$ so calculated with the exact spectrum for an angle-modulated waveform with frequency deviation $2\beta f_m$.

**4.19-1.** Assume that the 10.8-MHz signal in Fig. 4.19-1 is derived from the 200-kHz oscillator by multiplying by 54 and that the 200-kHz oscillator drifts by 0.1 Hz.

(*a*) Find the drift, in hertz, in the 10.8-MHz signal.

(*b*) Find the drift in the carrier of the resulting FM signal.

**4.19-2.** In an Armstrong modulator, as shown in Fig. 4.19-1, the crystal-oscillator frequency is 200 kHz. It is desired, in order to avoid distortion, to limit the maximum angular deviation to $\phi_m = 0.2$. The system is to accommodate modulation frequencies down to 40 Hz. At the output of the modulator the carrier frequency is to be 108 MHz and the frequency deviation 80 kHz. Select multiplier and mixer oscillator frequencies to accomplish this end.

**4.20-1.** The narrowband phase modulator of Fig. 4.16-1 is converted to a frequency modulator by preceding the balanced modulator with an integrator. The input signal is a sinusoid of angular frequency $\omega_m$.

(*a*) Show that, unless the frequency deviation is kept small, the modulator output, when demodulated, will yield not only the input signal but also its odd harmonics.

(*b*) If the modulation frequency is 50 Hz, find the allowable frequency deviation if the normalized power associated with the third harmonic is to be no more than 1 percent of the fundamental power.

**4.20-2.** (*a*) Consider the FM demodulator of Fig. 4.20-1. Let the frequency selective network be an *RC* integrating network. The 3-dB frequency of the network is $f_2 (= 1/2\pi RC)$. If the carrier frequency of the FM waveform is $f_c$, how should $f_2$ be selected so that the demodulator has the greatest sensitivity (i.e., greatest change in output per change in input frequency)?

(*b*) With $f_2$ selected for maximum sensitivity and with $f_c = 1$ MHz, find the change in demodulator output for a 1-Hz change in input frequency.

**4.20-3.** A "zero-crossing" FM discriminator operates in the following manner. The modulated waveform

$$v(t) = A \cos\left[2\pi f_c t + k \int_{-\infty}^{t} m(\lambda)\, d\lambda\right]$$

is applied to an electronic circuit which generates a narrow pulse on each occasion when $v(t)$ passes through zero. The pulses are of fixed polarity, amplitude, and duration. This pulse train is applied to a low-pass filter, say an *RC* low-pass network of 3-dB frequency $f_2$. Assume that the bandwidth of the baseband waveform $m(t)$ is $f_M$. Discuss the operation of this discriminator. Show that if $f_c \gg f_2 \gg f_M$ the output of the low-pass network is indeed proportional to the instantaneous frequency of $v(t)$.

**4.20-4.** (*a*) A frequency selective network is shown in Fig. P4.20-4. Calculate the ratio $|V_o(f)/V_i(f)|$, i.e., the ratio of the amplitude of the output to the amplitude of the input, as a function of frequency. Verify that

$$\left|\frac{V_o(f)}{V_i(f)}\right| = \left\{1 + \frac{Q_0^2}{(f/f_0)^2}\,[(f/f_0)^2 - 1]^2\right\}^{1/2}$$

in which $f_0 = 1/2\pi\sqrt{1/LC}$ is the resonant frequency and $Q_0 = R/2\pi f_0 L$ is the energy storage factor of the resonant network at $f_0$.

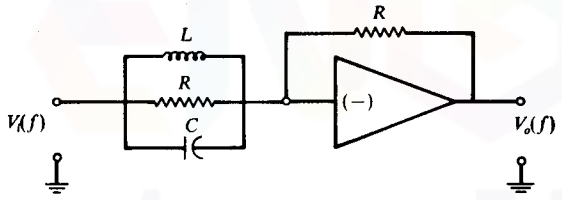



**Figure P4.20-4**

(*b*) Find the frequencies at which $|V_o/V_i|$ is higher by 3 dB than its value at the resonant frequency. Show that, for large $Q_0$, these frequencies are $f = f_0(1 \pm 1/Q_0)$. Calculate the slope of $|V_o/V_i|$ at the frequency $f_H = f_0(1 + 1/Q_0)$. A frequency modulated waveform with carrier frequency $f_H$ is applied as $V_i(f)$. What is the ratio of the change in amplitude $\Delta V_o(f)$ to the change, $\Delta f$, of the instantaneous frequency of the input?

(*c*) In the discriminator of Fig. 4.20-1 let the two frequency selective networks be the network shown in Fig. P4.20-4. The resonant frequencies of the two networks are to be $f_H = f_c(1 + 1/Q_0)$ and $f_L = f_c(1 - 1/Q_0)$, $f_c$ being the carrier frequency of an input frequency modulated waveform. Make a plot of the discriminator output as a function of the instantaneous frequency of the input. Use $Q_0 = 50$.





CHAPTER

# FIVE

# ANALOG-TO-DIGITAL CONVERSION

## PULSE-MODULATION SYSTEMS

In Chaps. 3 and 4 we described systems with which we can transmit many signals simultaneously over a single communications channel. We found that at the transmitting end the individual signals were translated in frequency so that each occupied a separate and distinct frequency band. It was then possible at the receiving end to separate the individual signals by the use of filters.

In the present chapter we shall discuss a second method of multiplexing. This second method depends on the fact that a bandlimited signal, even if it is a continuously varying function of time, may be specified exactly by samples taken sufficiently frequently. Multiplexing of several signals is then achieved by interleaving the samples of the individual signals. This process is called time-division multiplexing. Since the sample is a pulse, the systems to be discussed are called pulse-amplitude modulation systems.

Pulse-amplitude modulation (PAM) systems are analog systems and share a common problem with the AM and FM modulation systems studied earlier. That is, each of these analog modulation systems are extremely sensitive to the noise present in the receiver. In this chapter we consider how to convert the analog signal into a digital signal prior to modulation. Such a conversion is called *source encoding*. We shall see in Chap. 12 that when a digital signal is modulated and transmitted, the received signal is far less sensitive to receiver noise than is analog modulation.

### Noisy Communications Channels

We consider a basic problem associated with the transmission of a signal over a noisy communication channel. For the sake of being specific, suppose we require that a telephone conversation be transmitted from New York to Los Angeles. If

(*a*) Show that, unless the frequency deviation is kept small, the modulator output, when demodulated, will yield not only the input signal but also its odd harmonics.

(*b*) If the modulation frequency is 50 Hz, find the allowable frequency deviation if the normalized power associated with the third harmonic is to be no more than 1 percent of the fundamental power.

**4.20-2.** (*a*) Consider the FM demodulator of Fig. 4.20-1. Let the frequency selective network be an *RC* integrating network. The 3-dB frequency of the network is $f_2 (= 1/2\pi RC)$. If the carrier frequency of the FM waveform is $f_c$, how should $f_2$ be selected so that the demodulator has the greatest sensitivity (i.e., greatest change in output per change in input frequency)?

(*b*) With $f_2$ selected for maximum sensitivity and with $f_c = 1$ MHz, find the change in demodulator output for a 1-Hz change in input frequency.

**4.20-3.** A "zero-crossing" FM discriminator operates in the following manner. The modulated waveform

$$v(t) = A \cos\left[2\pi f_c t + k \int_{-\infty}^{t} m(\lambda)\, d\lambda\right]$$

is applied to an electronic circuit which generates a narrow pulse on each occasion when $v(t)$ passes through zero. The pulses are of fixed polarity, amplitude, and duration. This pulse train is applied to a low-pass filter, say an *RC* low-pass network of 3-dB frequency $f_2$. Assume that the bandwidth of the baseband waveform $m(t)$ is $f_M$. Discuss the operation of this discriminator. Show that if $f_c \gg f_2 \gg f_M$ the output of the low-pass network is indeed proportional to the instantaneous frequency of $v(t)$.

**4.20-4.** (*a*) A frequency selective network is shown in Fig. P4.20-4. Calculate the ratio $|V_o(f)/V_i(f)|$, i.e., the ratio of the amplitude of the output to the amplitude of the input, as a function of frequency. Verify that

$$\left|\frac{V_o(f)}{V_i(f)}\right| = \left\{1 + \frac{Q_0^2}{(f/f_0)^2}\,[(f/f_0)^2 - 1]^2\right\}^{1/2}$$

in which $f_0 = 1/2\pi\sqrt{1/LC}$ is the resonant frequency and $Q_0 = R/2\pi f_0 L$ is the energy storage factor of the resonant network at $f_0$.

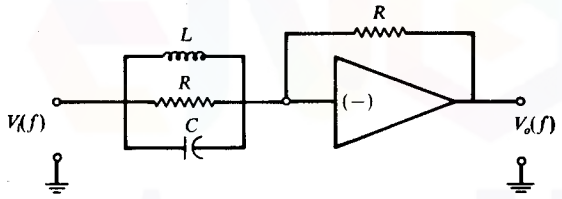



**Figure P4.20-4**

(*b*) Find the frequencies at which $|V_o/V_i|$ is higher by 3 dB than its value at the resonant frequency. Show that, for large $Q_0$, these frequencies are $f = f_0(1 \pm 1/Q_0)$. Calculate the slope of $|V_o/V_i|$ at the frequency $f_H = f_0(1 + 1/Q_0)$. A frequency modulated waveform with carrier frequency $f_H$ is applied as $V_i(f)$. What is the ratio of the change in amplitude $\Delta V_o(f)$ to the change, $\Delta f$, of the instantaneous frequency of the input?

(*c*) In the discriminator of Fig. 4.20-1 let the two frequency selective networks be the network shown in Fig. P4.20-4. The resonant frequencies of the two networks are to be $f_H = f_c(1 + 1/Q_0)$ and $f_L = f_c(1 - 1/Q_0)$, $f_c$ being the carrier frequency of an input frequency modulated waveform. Make a plot of the discriminator output as a function of the instantaneous frequency of the input. Use $Q_0 = 50$.





CHAPTER

# FIVE

# ANALOG-TO-DIGITAL CONVERSION

## PULSE-MODULATION SYSTEMS

In Chaps. 3 and 4 we described systems with which we can transmit many signals simultaneously over a single communications channel. We found that at the transmitting end the individual signals were translated in frequency so that each occupied a separate and distinct frequency band. It was then possible at the receiving end to separate the individual signals by the use of filters.

In the present chapter we shall discuss a second method of multiplexing. This second method depends on the fact that a bandlimited signal, even if it is a continuously varying function of time, may be specified exactly by samples taken sufficiently frequently. Multiplexing of several signals is then achieved by interleaving the samples of the individual signals. This process is called time-division multiplexing. Since the sample is a pulse, the systems to be discussed are called pulse-amplitude modulation systems.

Pulse-amplitude modulation (PAM) systems are analog systems and share a common problem with the AM and FM modulation systems studied earlier. That is, each of these analog modulation systems are extremely sensitive to the noise present in the receiver. In this chapter we consider how to convert the analog signal into a digital signal prior to modulation. Such a conversion is called *source encoding*. We shall see in Chap. 12 that when a digital signal is modulated and transmitted, the received signal is far less sensitive to receiver noise than is analog modulation.

### Noisy Communications Channels

We consider a basic problem associated with the transmission of a signal over a noisy communication channel. For the sake of being specific, suppose we require that a telephone conversation be transmitted from New York to Los Angeles. If

the signal is transmitted by radio, then, when the signal arrives as its destination, it will be greatly attenuated and also combined with noise due to thermal noise present in all receivers (Chap. 14), and to all manner of random electrical disturbances which are added to the radio signal during its propagation across country. (We neglect as irrelevant, for the present discussion, whether such direct radio communication is reliable over such long channel distances.) As a result, the received signal may not be distinguishable against its background of noise. The situation is not fundamentally different if the signal is transmitted over wires. Any physical wire transmission path will both attenuate and distort a signal by an amount which increases with path length. Unless the wire path is completely and perfectly shielded, as in the case of a perfect coaxial cable, electrical noise and crosstalk disturbances from neighboring wire paths will also be picked up in amounts increasing with the path length. In this connection it is of interest to note that even coaxial cable does not provide complete freedom from crosstalk. External low-frequency magnetic fields will penetrate the outer conductor of the coaxial cable and thereby induce signals on the cable. In telephone cable, where coaxial cables are combined with parallel wire signal paths, it is common practice to wrap the coax in Permalloy for the sake of magnetic shielding. Even the use of fiber optic cables which are relatively immune to such interference, does not significantly alter the problem since receiver noise is often the noise source of largest power.

One attempt to resolve this problem is simply to raise the signal level at the transmitting end to so high a level that, in spite of the attenuation, the received signal substantially overrides the noise. (Signal distortion may be corrected separately by equalization.) Such a solution is hardly feasible on the grounds that the signal power and consequent voltage levels at the transmitter would be simply astronomical and beyond the range of amplifiers to generate, and cables to handle. For example, at 1 kHz, a telephone cable may be expected to produce an attenuation of the order of 1 dB per mile. For a 3000-mile run, even if we were satisfied with a received signal of 1 mV, the voltage at the transmitting end would have to be $10^{147}$ volts.

An amplifier at the receiver will not help the above situation, since at this point both signal and noise levels will be increased together. But suppose that a repeater (repeater is the term used for an amplifier in a communications channel) is located at the midpoint of the long communications path. This repeater will raise the signal level; in addition, it will raise the level of only the noise introduced in the first half of the communications path. Hence, such a midway repeater, as contrasted with an amplifier at the receiver, has the advantage of improving the received signal-to-noise ratio. This midway repeater will relieve the burden imposed on transmitter and cable due to higher power requirements when the repeater is not used.

The next step is, of course, to use additional repeaters, say initially at the one-quarter and three-quarter points, and thereafter at points in between. Each added repeater serves to lower the maximum power level encountered on the communications link, and each repeater improves the signal-to-noise ratio over what would result if the corresponding gain were introduced at the receiver.

In the limit we might, conceptually at least, use an infinite number of repeaters. We could even adjust the gain of each repeater to be infinitesimally greater than unity by just the amount to overcome the attenuation in the infinitesimal section between repeaters. In the end we would thereby have constructed a channel which had no attenuation. The signal at the receiving terminal of the channel would then be the unattenuated transmitted signal. We would then, in addition, have at the receiving end all the noise introduced at all points of the channel. This noise is also received without attenuation, no matter how far away from the receiving end the noise was introduced. If now, with this finite array of repeaters, the signal-to-noise ratio is not adequate, there is nothing to be done but to raise the signal level or to make the channel quieter.

The situation is actually somewhat more dismal than has just been intimated, since each repeater (transistor amplifier) introduces some noise on its own accord. Hence, as more repeaters are cascaded, each repeater must be designed to more exacting standards with respect to noise figure (see Sec. 14.10). The limitation of the system we have been describing for communicating over long channels is that once noise has been introduced any place along the channel, we are "stuck" with it.

If we now were to transmit a digital signal over the same channel we would find that significantly less signal power would be needed in order to obtain the same performance at the receiver. The reason for this is that the significant parameter is now not the signal-to-noise ratio but the probability of mistaking a digital signal for a different digital signal. In practice we find that signal-to-noise ratios of 40–60 dB are required for analog signals while 10–12 dB are required for digital signals. This reason and others, to be discussed subsequently, have resulted in a large commercial and military switch to digital communications.

## 5.1 THE SAMPLING THEOREM. LOW-PASS SIGNALS

We consider at the outset the fundamental principle of digital communications; the sampling theorem:

Let $m(t)$ be a signal which is bandlimited such that its highest frequency spectral component is $f_M$. Let the values of $m(t)$ be determined at regular intervals separated by times $T_s \leq 1/2f_M$, that is, the signal is periodically *sampled* every $T_s$ seconds. Then these samples $m(nT_s)$, where $n$ is an integer, uniquely determine the signal, and the signal may be reconstructed from these samples with no distortion.

The time $T_s$ is called the *sampling time*. Note that the theorem requires that the *sampling rate* be rapid enough so that at least two samples are taken during the course of the period corresponding to the highest-frequency spectral com-

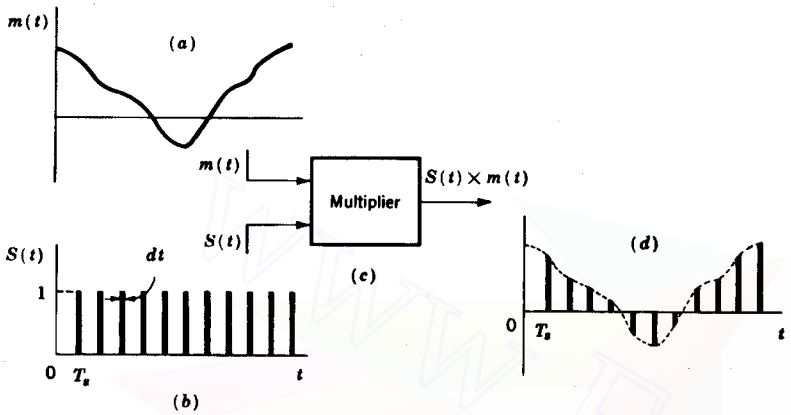



**Figure 5.1-1** (*a*) A signal $m(t)$ which is to be sampled. (*b*) The sampling function $S(t)$ consists of a train of very narrow unit amplitude pulses. (*c*) The sampling operation is performed in a multiplier. (*d*) The samples of the signal $m(t)$.

ponent. We shall now prove the theorem by showing how the signal may be reconstructed from its samples.

The baseband signal $m(t)$ which is to be sampled is shown in Fig. 5.1-1*a*. A periodic train of pulses $S(t)$ of unit amplitude and of period $T_s$ is shown in Fig. 5.1-1*b*. The pulses are arbitrarily narrow, having a width $dt$. The two signals $m(t)$ and $S(t)$ are applied to a multiplier as shown in Fig. 5.1-1*c*, which then yields as an output the product $S(t)m(t)$. This product is seen in Fig. 5.1-1*d* to be the signal $m(t)$ *sampled* at the occurrence of each pulse. That is, when a pulse occurs, the multiplier output has the same value as does $m(t)$, and at all other times the multiplier output is zero.

The signal $S(t)$ is periodic, with period $T_s$, and has the Fourier expansion [see Eq. (1.3-8) with $I = dt$ and $T_0 = T_s$]

$$S(t) = \frac{dt}{T_s} + \frac{2\,dt}{T_s}\left(\cos 2\pi\frac{t}{T_s} + \cos 2\times 2\pi\frac{t}{T_s} + \cdots\right) \tag{5.1-1}$$

For the case $T_s = 1/2f_M$, the product $S(t)m(t)$ is

$$S(t)m(t) = \frac{dt}{T_s}m(t) + \frac{dt}{T_s}[2m(t)\cos 2\pi(2f_M)t + 2m(t)\cos 2\pi(4f_M)t + \cdots] \tag{5.1-2}$$

We now observe that the first term in the series is, aside from a constant factor, the signal $m(t)$ itself. Again, aside from a multiplying factor, the second term is the product of $m(t)$ and a sinusoid of frequency $2f_M$. This product then, as discussed in Sec. 3.2, gives rise to a double-sideband suppressed-carrier signal with carrier frequency $2f_M$. Similarly, succeeding terms yield DSB-SC signals with carrier frequencies $4f_M$, $6f_M$, etc.

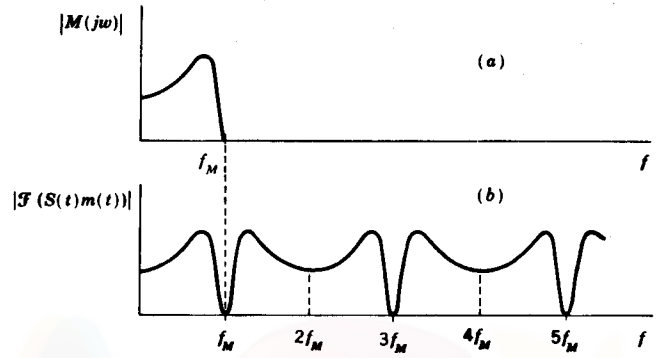



**Figure 5.1-2** (*a*) The magnitude plot of the spectral density of a signal bandlimited to $f_M$. (*b*) Plot of amplitude of spectrum of sampled signal.

Let the signal $m(t)$ have a spectral density $M(j\omega) = \mathcal{F}[m(t)]$ which is as shown in Fig. 5.1-2*a*. Then $m(t)$ is bandlimited to the frequency range below $f_M$. The spectrum of the first term in Eq. (5.1-2) extends from 0 to $f_M$. The spectrum of the second term is symmetrical about the frequency $2f_M$ and extends from $2f_M - f_M = f_M$ to $2f_M + f_M = 3f_M$. Altogether the spectrum of the sampled signal has the appearance shown in Fig. 5.1-2*b*. Suppose then that the sampled signal is passed through an ideal low-pass filter with cutoff frequency at $f_M$. If the filter transmission were constant in the passband and if the cutoff were infinitely sharp at $f_M$, the filter would pass the signal $m(t)$ and nothing else.

The spectral pattern corresponding to Fig. 5.1-2b is shown in Fig. 5.1-3*a* for the case in which the sampling rate $f_s = 1/T_s$ is larger than $2f_M$. In this case there

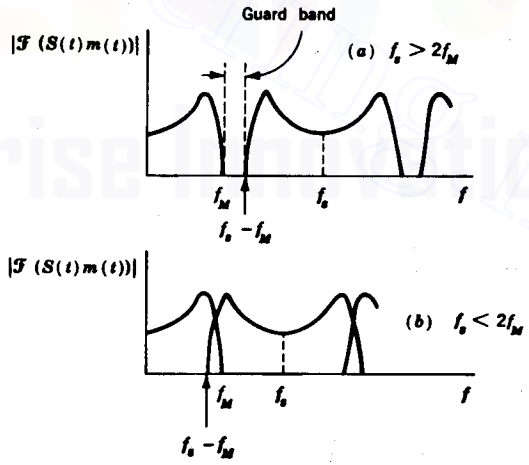



**Figure 5.1-3** (*a*) A guard band appears when $f_s > 2f_M$. (*b*) Overlapping of spectra when $f_s < 2f_M$.


Downloaded From : www.EasyEngineering.net

is a gap between the upper limit $f_M$ of the spectrum of the baseband signal and the lower limit of the DSB-SC spectrum centered around the carrier frequency $f_s > 2f_M$. For this reason the low-pass filter used to select the signal $m(t)$ need not have an infinitely sharp cutoff. Instead, the filter attenuation may begin at $f_M$ but need not attain a high value until the frequency $f_s - f_M$. This range from $f_M$ to $f_s - f_M$ is called a *guard band* and is always required in practice, since a filter with infinitely sharp cutoff is, of course, not realizable. Typically, when sampling is used in connection with voice messages on telephone lines, the voice signal is limited to $f_M = 3.3$ kHz, while $f_s$ is selected at 8.0 kHz. The guard band is then $8.0 - 2 \times 3.3 = 1.4$ kHz.

The situation depicted in Fig. 5.1-3*b* corresponds to the case where $f_s < 2f_M$. Here we find an overlap between the spectrum of $m(t)$ itself and the spectrum of the DSB-SC signal centered around $f_s$. Accordingly, no filtering operation will allow an exact recovery of $m(t)$.

We have just proved the *sampling theorem* since we have shown that, in principle, the sampled signal can be recovered exactly when $T_s \leq 1/2f_M$. It has also been shown why the minimum allowable sampling rate is $2f_M$. This minimum sampling rate is known as the *Nyquist rate*. An increase in sampling rate above the Nyquist rate increases the width of the guard band, thereby easing the problem of filtering. On the other hand, we shall see that an increase in rate extends the bandwidth required for transmitting the sampled signal. Accordingly an engineering compromise is called for.

An interesting special case is the sampling of a sinusoidal signal having the frequency $f_M$. Here, *all* the signal power is concentrated precisely at the cutoff frequency of the low-pass filter, and there is consequently some ambiguity about whether the signal frequency is inside or outside the filter passband. To remove this ambiguity, we require that $f_s > 2f_M$ rather than that $f_s \geq 2f_M$. To see that this condition is necessary, assume that $f_s = 2f_M$ but that an initial sample is taken at the moment the sinusoid passes through zero. Then all successive samples will also be zero. This situation is avoided by requiring $f_s > 2f_M$.

## Bandpass Signals

For a signal $m(t)$ whose highest-frequency spectral component is $f_M$, the sampling frequency $f_s$ must be no less than $f_s = 2f_M$ only if the lowest-frequency spectral component of $m(t)$ is $f_L = 0$. In the more general case, where $f_L \neq 0$, it may be that the sampling frequency need be no larger than $f_s = 2(f_M - f_L)$. For example, if the spectral range of a signal extends from 10.0 to 10.1 MHz, the signal may be recovered from samples taken at a frequency $f_s = 2(10.1 - 10.0) = 0.2$ MHz.

To establish the sampling theorem for such bandpass signals, let us select a sampling frequency $f_s = 2(f_M - f_L)$ and let us initially assume that it happens that the frequency $f_L$ turns out to be an integral multiple of $f_s$, that is, $f_L = nf_s$ with $n$ an integer. Such a situation is represented in Fig. 5.1-4. In part *a* is shown the two-sided spectral pattern of a signal $m(t)$ with Fourier transform $M(j\omega)$. Here it has been arranged that $n = 2$; that is, $f_L$ coincides with the second harmonic of

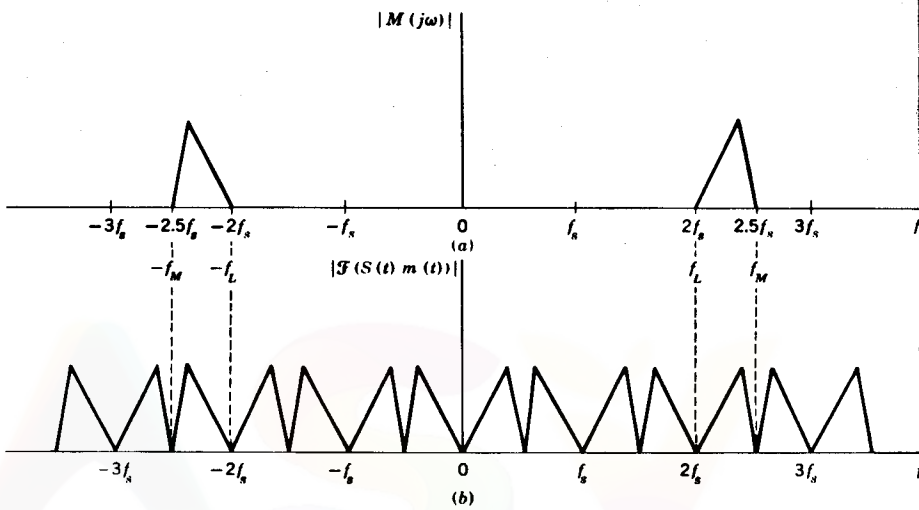



**Figure 5.1-4** (*a*) The spectrum of a bandpass signal. (*b*) The spectrum of the sampled bandpass signal.

the sampling frequency, while the sampling frequency is exactly $f_s = 2(f_M - f_L)$. In part *b* is shown the spectral pattern of the sampled signal $S(t)m(t)$. The product of $m(t)$ and the dc term of $S(t)$ [Eq. (5.1-1)] duplicates in part *b* the form of the spectral pattern in part *a* and leaves it in the same frequency range from $f_L$ to $f_M$. The product of $m(t)$ and the spectral component in $S(t)$ of frequency $f_s(= 1/T_s)$ gives rise in part *b* to a spectral pattern derived from part *a* by shifting the pattern in part *a* to the right and also to the left by amount $f_s$. Similarly, the higher harmonics of $f_s$ in $S(t)$ give rise to corresponding shifts, right and left, of the spectral pattern in part *a*. We now note that if the sampled signal $S(t)m(t)$ is passed through a bandpass filter with arbitrarily sharp cutoffs and with passband from $f_L$ to $f_M$, the signal $m(t)$ will be recovered exactly.

In Fig. 5.1-4 the spectrum of $m(t)$ extends over the first half of the frequency interval between harmonics of the sampling frequency, that is, from $2.0f_s$ to $2.5f_s$. As a result, there is no spectrum overlap, and signal recovery is possible. It may also be seen from the figure that if the spectral range of $m(t)$ extended over the second half of the interval from $2.5f_s$ to $3.0f_s$, there would similarly be no overlap. Suppose, however, that the spectrum of $m(t)$ were confined neither to the first half nor to the second half of the interval between sampling-frequency harmonics. In such a case, there would be overlap between the spectrum patterns, and signal recovery would not be possible. Hence the minimum sampling frequency allowable is $f_s = 2(f_M - f_L)$ provided that either $f_M$ or $f_L$ is a harmonic of $f_s$.

If neither $f_M$ nor $f_L$ is a harmonic of $f_s$, a more general analysis is required. In Fig. 5.1-5*a* we have reproduced the spectral pattern of Fig. 5.1-4. The positive-frequency part and the negative-frequency part of the spectrum are called PS and NS respectively. Let us, for simplicity, consider separately PS and NS and the

selection of a sampling frequency $f_s = 2B = 2$ kHz brings us to a point in an overlap region. As $f_s$ is increased there is a small range of $f_s$, corresponding to $N = 3$, where there is no overlap. Further increase in $f_s$ again takes us to an overlap region, while still further increase in $f_s$ provides a nonoverlap range, corresponding to $N = 2$ (from $f_s = 3.5B$ to $f_s = 5B$). Increasing $f_s$ further we again enter an overlap region while at $f_s = 7B$ we enter the nonoverlap region for $N = 1$. When $f_s \geq 7B$ we do not again enter an overlap region. (This is the region where $f_s \geq 2f_M$; that is, we assume we have a lowpass rather than a bandpass signal.)

### The Discrete Fourier Transform

There are occasions when the only information we have available about a signal is a set of sample values, $N$ in number, taken at regularly spaced intervals $T_s$ over a period of time $T_0$. From this sampled data we should often like to be able to arrive at some reasonable approximation of the spectral content of the signal. If the sample period and the number of samples is adequate to give us some confidence that what has been observed of the signal is representative of the signal generally, we may indeed estimate its spectral content.

We pretend that the signal is periodic with period $T_0$ and we pretend, as well, that the sampling rate is adequate to satisfy the Nyquist criterion. A typical set of sample values is shown in Fig. 5.1-7. Here, for simplicity we have assumed an even number of samples so that we can place them symmetrically in the period interval $T_0$ and symmetrically about the origin of the coordinate system. The sample values are located at $\pm T_s/2$, $\pm 3T_s/2$, etc. If there are $N$ samples then the samples most distant from the origin are at $\pm[(N-1)/2]T_s$.

If the waveform to be sampled is $m(t)$, then after sampling, the waveform we have available is $m(t)S(t)$, where $S(t)$ is the sampling function, i.e., $S(t) = 1$ during the sampling duration $dt$, as shown in Fig. 5.1-7, and $S(t) = 0$ elsewhere. We note from Figs. 5.1-2 and 5.1-3*a* that the spectrum of $m(t)$ and the part of the spectrum of $m(t)S(t)$ up to $f_M$ are identical in form. Hence, to find the spectrum of $m(t)$ we may evaluate instead the spectrum of $m(t)S(t)$. The spectral amplitudes of $m(t)S(t)$ are:

$$M_n = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} m(t)e^{-j2\pi nt/T_0}\,dt \tag{5.1-9}$$

If there are $N$ samples in all, then the sample times are:

$$-\left(\frac{N-1}{2}\right)T_s,\ -\left(\frac{N-1}{2}\right)T_s + T_s,\ -\left(\frac{N-1}{2}\right)T_s + 2T_s, \ldots, \left(\frac{N-1}{2}\right)T_s + \cdots$$

Using these values for $t$ in the integrand, we find that Eq. (5.1-9) becomes:

$$M_n = \frac{dt}{T_0}\sum_{k=-(N-1)/2}^{k=(N-1)/2} m(kT_s)e^{-j2\pi nkT_s/T_0} \tag{5.1-10}$$





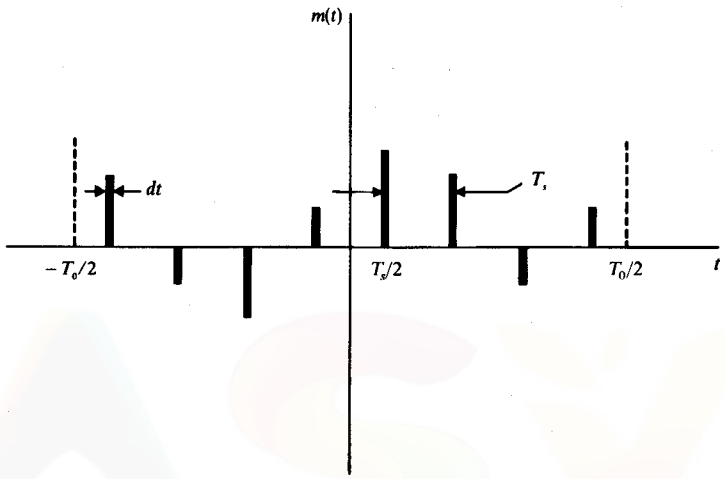



**Figure 5.1-7** A possible set of sample values of a waveform $m(t)$, taken every $T_s$ over a time interval $T_0$.

As a purely mathematical exercise we could use Eq. (5.1-10) to calculate spectral components $v_n$ for any value of $n$. But consistently with our assumptions, the largest value of $n$ allowable is determined by the Nyquist criterion. The highest frequency component should have a period which is $2T_s$ so that:

$$f_n(\max) = \frac{1}{2T_s} \tag{5.1-11}$$

The fundamental period is $T_0$ so that the fundamental frequency is therefore $f_0 = 1/T_0$. Since $T_0 = NT_s$, we have:

$$f_n(\max) = \frac{N}{2}\frac{1}{T_0} = \frac{N}{2}f_0 \tag{5.1-12}$$

Hence, the highest value of $n$ for which Eq. (5.1-10) should be used is $n = N/2$.

## 5.2 PULSE-AMPLITUDE MODULATION

A technique by which we may take advantage of the sampling principle for the purpose of time-division multiplexing is illustrated in the idealized representation of Fig. 5.2-1. At the transmitting end on the left, a number of bandlimited signals

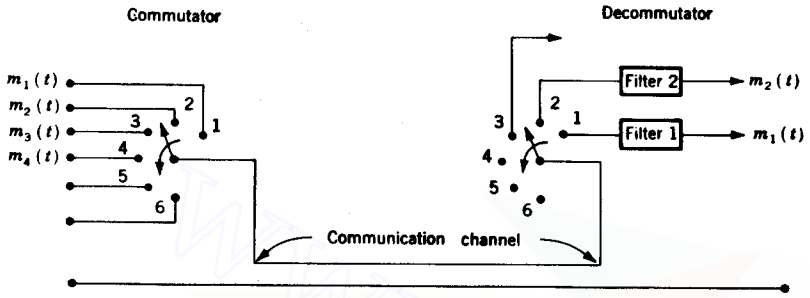



**Figure 5.2-1** Illustrating how the sampling principle may be used to transmit a number of bandlimited signals over a single communications channel.

are connected to the contact point of a rotary switch. We assume that the signals are similarly bandlimited. For example, they may all be voice signals, limited to 3.3 kHz. As the rotary arm of the switch swings around, it samples each signal sequentially. The rotary switch at the receiving end is in synchronism with the switch at the sending end. The two switches make contact simultaneously at similarly numbered contacts. With each revolution of the switch, one sample is taken of each input signal and presented to the correspondingly numbered contact of the receiving-end switch. The train of samples at, say, terminal 1 in the receiver, pass through low-pass filter 1, and, at the filter output, the original signal $m_1(t)$ appears reconstructed. Of course, if $f_M$ is the highest-frequency spectral component present in any of the input signals, the switches must make at least $2f_M$ revolutions per second.

When the signals to be multiplexed vary slowly with time, so that the sampling rate is correspondingly slow, mechanical switches, indicated in Fig. 5.2-1, may be employed. When the switching speed required is outside the range of

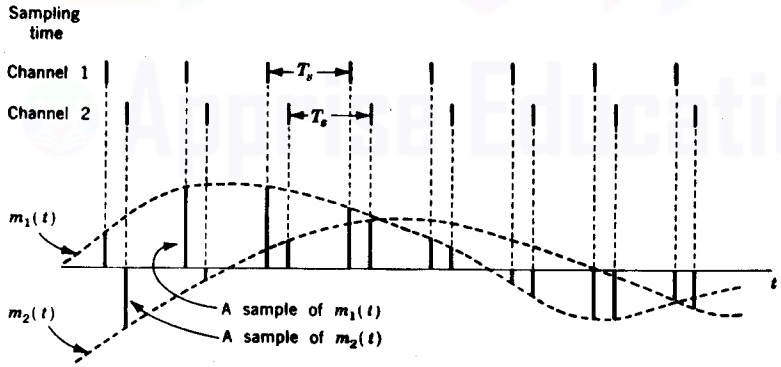



**Figure 5.2-2** The interlacing of two baseband signals.

mechanical switches, electronic switching systems may be employed. In either event, the switching mechanism, corresponding to the switch at the left in Fig. 5.2-1, which samples the signals, is called the *commutator*. The switching mechanism which performs the function of the switch at the right in Fig. 5.2-1 is called the *decommutator*. The commutator samples and combines samples, while the decommutator separates samples belonging to individual signals so that these signals may be reconstructed.

The interlacing of the samples that allows multiplexing is shown in Fig. 5.2-2. Here, for simplicity, we have considered the case of the multiplexing of just two signals $m_1(t)$ and $m_2(t)$. The signal $m_1(t)$ is sampled regularly at intervals of $T_s$ and at the times indicated in the figure. The sampling of $m_2(t)$ is similarly regular, but the samples are taken at a time different from the sampling time of $m_1(t)$. The input waveform to the filter numbered 1 in Fig. 5.2-1 is the train of samples of $m_1(t)$, and the input to the filter numbered 2 is the train of samples of $m_2(t)$. The timing in Fig. 5.2-2 has been deliberately drawn to suggest that there is room to multiplex more than two signals. We shall see shortly, in principle, how many signals may be multiplexed.

We observe that the train of pulses corresponding to the samples of each signal are *modulated in amplitude* in accordance with the signal itself. Accordingly, the scheme of sampling is called *pulse-amplitude modulation* and abbreviated PAM.

Multiplexing of several PAM signals is possible because the various signals are kept distinct and are separately recoverable by virtue of the fact that they are sampled at different times. Hence this system is an example of a *time-division multiplex* (TDM) system. Such systems are the counterparts in the time domain of the systems of Chap. 3. There, the signals were kept separable by virtue of their translation to different portions of the frequency domain, and those systems are called frequency-division multiplex (FDM) systems.

If the multiplexed signals are to be transmitted directly, say, over a pair of wires, no further signal processing need be undertaken. Suppose, however, we require to transmit the TDM-PAM signal from one antenna to another. It would then be necessary to amplitude-modulate or frequency-modulate a high-frequency carrier with the TDM-PAM signal; in such a case the overall system would be referred to, respectively, as PAM-AM or PAM-FM. Note that the same terminology is used whether a single signal or many signals (TDM) are transmitted.

## 5.3 CHANNEL BANDWIDTH FOR A PAM SIGNAL

Suppose that we have $N$ independent baseband signals $m_1(t)$, $m_2(t)$, etc., each of which is bandlimited to $f_M$. What must be the bandwidth of the communications channel which will allow all $N$ signals to be transmitted simultaneously using PAM time-division multiplexing? We shall now show that, in principle at least, the channel need not have a bandwidth larger than $Nf_M$.


Downloaded From : www.EasyEngineering.net

The baseband signal, say $m_1(t)$, must be sampled at intervals not longer than $T_s = 1/2f_M$. Between successive samples of $m_1(t)$ will appear samples of the other $N - 1$ signals. Therefore the interval of separation between successive samples of different baseband signals is $1/2f_M N$. The composite signal, then, which is presented to the transmitting end of the communications channel, consists of a sequence of samples, that is, a sequence of *impulses*. If the bandwidth of the channel were arbitrarily great, the waveform at the receiving end would be the same as at the sending end and demultiplexing could be achieved in a straightforward manner.

If, however, the bandwidth of the channel is restricted, the channel response to an instantaneous sample will be a waveform which may well persist with significant amplitude long after the time of selection of the sample. In such a case, the signal at the receiving end at any particular sampling time may well have significant contributions resulting from previous samples of other signals. Consequently the signal which appears at any of the output terminals in Fig. 5.2-1 will not be a single baseband signal but will be instead a combination of many or even all the baseband signals. Such combining of baseband signals at a communication system output is called *crosstalk* and is to be avoided as far as possible.

Let us assume that our channel has the characteristics of an ideal low-pass filter with angular cutoff frequency $\omega_c = 2\pi f_c$, unity gain, and no delay. Let a sample be taken, say, of $m_1(t)$, at $t = 0$. Then at $t = 0$ there is presented at the transmitting end of the channel an impulse of strength $I_1 = m_1(0)\,dt$. The response at the receiving end is $s_{R1}(t)$ given by (see Prob. 5.3-1)

$$s_{R1}(t) = \frac{I_1 \omega_c}{\pi} \frac{\sin \omega_c t}{\omega_c t} \tag{5.3-1}$$

The normalized response $\pi s_R(t)/\omega_c$ is shown in Fig. 5.3-1 by the solid plot. At $t = 0$ the reponse attains a peak value proportional to the strength of the impulse

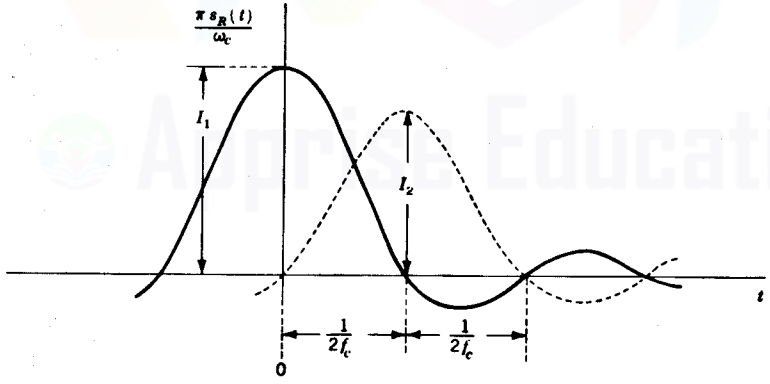



**Figure 5.3-1** The response of an ideal low-pass filter to an instantaneous sample at $t = 0$ (solid plot). The response to a sample at $t = 1/2f_c$ (dashed plot).

$I_1 = m_1(0)\,dt$, which is in turn proportional to the value of the sample $m_1(0)$. This response persists indefinitely. Observe, however, that the response passes through zero at intervals which are multiples of $\pi/\omega_c = 1/2f_c$. Suppose, then, that a sample of $m_2(t)$ is taken and transmitted at $t = 1/2f_c$. If $I_2 = m_2(t = 1/2f_c)\,dt$,

$$s_{R2}(t) = \frac{I_2 \omega_c}{\pi} \frac{\sin \omega_c(t - 1/2f_c)}{\omega_c(t - 1/2f_c)} \tag{5.3-2}$$

This response is shown by the dashed plot. Suppose, finally, that the demultiplexing is done also by instantaneous sampling at the receiving end of the channel, for $m_1(t)$ at $t = 0$ and for $m_2(t)$ at $t = 1/2f_c$. Then, in spite of the persistence of the channel response, there will be no crosstalk, and the signals $m_1(t)$ and $m_2(t)$ may be completely separated and individually recovered. Similarly, additional signals may be sampled and multiplexed, provided that each new sample is taken synchronously, every $1/2f_c$ s. The sequence must, of course, be continually repeated every $1/2f_M$ s, so that each signal is properly sampled.

We have then the result that with a channel of bandwidth $f_c$ we need to separate samples by intervals $1/2f_c$. The sampling theorem requires that the samples of an individual baseband signal be separated by intervals not longer than $1/2f_M$. Hence the total number of signals which may be multiplexed is $N = f_c/f_M$, or $f_c = Nf_M$ as indicated earlier.

In principle then, multiplexing a number of signals by PAM time division requires no more bandwidth than would be required to multiplex these signals by frequency-division multiplexing using single-sideband transmission.

## 5.4 NATURAL SAMPLING

It was convenient, for the purpose of introducing some basic ideas, to begin our discussion of time multiplexing by assuming instantaneous commutation and decommutation. Such instantaneous sampling, however, is hardly feasible. Even if it were possible to construct switches which could operate in an arbitrarily short time, we would be disinclined to use them. The reason is that instantaneous samples at the transmitting end of the channel have infinitesimal energy, and when transmitted through a bandlimited channel give rise to signals having a peak value which is infinitesimally small. We recall that in Fig. 5.3-1 $I_1 = m_1(0)\,dt$. Such infinitesimal signals will inevitably be lost in background noise.

A much more reasonable manner of sampling, referred to as *natural sampling*, is shown in Fig. 5.4-1. Here the sampling waveform $S(t)$ consists of a train of pulses having duration $\tau$ and separated by the sampling time $T_s$. The baseband signal is $m(t)$, and the sampled signal $S(t)m(t)$ is shown in Fig. 5.4-1*c*. Observe that the sampled signal consists of a sequence of pulses of varying amplitude whose tops are not flat but follow the waveform of the signal $m(t)$.

With natural sampling, as with instantaneous sampling, a signal sampled at the Nyquist rate may be reconstructed exactly by passing the samples through an ideal low-pass filter with cutoff at the frequency $f_M$, where $f_M$ is the highest-

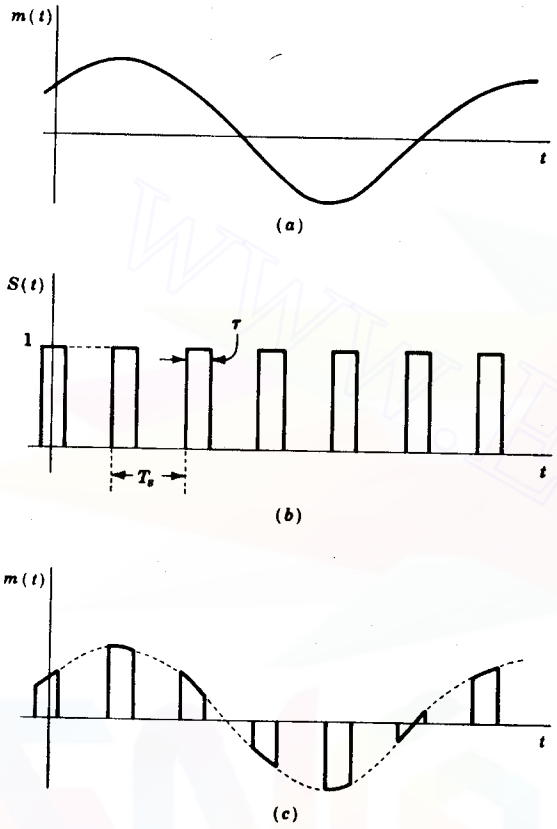



**Figure 5.4-1** (*a*) A baseband signal $m(t)$. (*b*) A sampling signal $S(t)$ with pulses of finite duration. (*c*) The *naturally* sampled signal $S(t)m(t)$.

frequency spectral component of the signal. To prove this, we note that the sampling waveform $S(t)$ shown in Fig. 5.4-1 is given by [see Eq. (1.3-12) with $A = 1$ and $T_0 = T_s$]

$$S(t) = \frac{\tau}{T_s} + \frac{2\tau}{T_s}\left(C_1 \cos 2\pi \frac{t}{T_s} + C_2 \cos 2 \times 2\pi \frac{t}{T_s} + \cdots\right) \tag{5.4-1}$$

with the constant $C_n$ given by

$$C_n = \frac{\sin (n\pi\tau/T_s)}{n\pi\tau/T_s} \tag{5.4-2}$$

This sampling waveform differs from the sampling waveform of Eq. (5.1-1) for instantaneous sampling only in that $dt$ is replaced by $\tau$ and by the fact that the amplitudes of the various harmonics are not the same. The sampled baseband signal $S(t)m(t)$ is, for $T_s = 1/2f_M$,

$$S(t)m(t) = \frac{\tau}{T_s} m(t) + \frac{2\tau}{T_s}\left[m(t)C_1 \cos 2\pi(2f_M)t + m(t)C_2 \cos 2\pi(4f_M)t + \cdots\right] \tag{5.4-3}$$

Therefore, as in instantaneous sampling, a low-pass filter with cutoff at $f_M$ will deliver an output signal $s_o(t)$ given by

$$s_o(t) = \frac{\tau}{T_s} m(t) \tag{5.4-4}$$

which is the same as is given by the first term of Eq. (5.1-2) except with $dt$ replaced by $\tau$.

With samples of finite duration, it is not possible to completely eliminate the crosstalk generated in a channel, sharply bandlimited to a bandwidth $f_c$. If $N$ signals are to be multiplexed, then the maximum sample duration is $\tau = T_s/N$. It is advantageous, for the purpose of increasing the level of the output signal, to make $\tau$ as large as possible. For, as is seen in Eq. (5.4-4), $s_o(t)$ increases with $\tau$. However, to help suppress crosstalk, it is ordinarily required that the samples be limited to a duration much less than $T_s/N$. The result is a large *guard time* between the end of one sample and the beginning of the next.

## 5.5 FLAT-TOP SAMPLING

Pulses of the type shown in Fig. 5.4-1*c*, with tops contoured to follow the waveform of the signal, are actually not frequently employed. Instead *flat-topped* pulses are customarily used, as shown in Fig. 5.5-1*a*. A flat-topped pulse has a

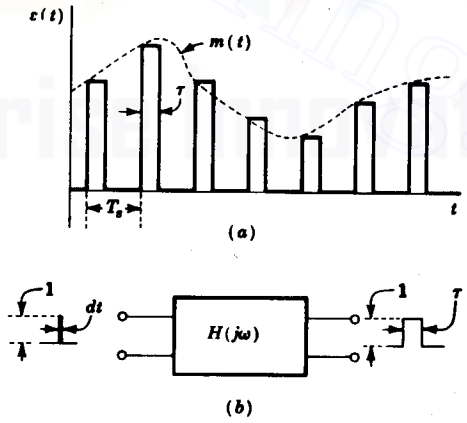



**Figure 5.5-1** (*a*) Flat-topped sampling. (*b*) A network with transform $H(j\omega)$ which converts a pulse of width $dt$ into a rectangular pulse of like amplitude but of duration $\tau$.

constant amplitude established by the sample value of the signal at some point within the pulse interval. In Fig. 5.5-1*a* we have arbitrarily sampled the signal at the beginning of the pulse. In sampling of this type the baseband signal $m(t)$ cannot be recovered exactly by simply passing the samples through an ideal low-pass filter. However, the distortion need not be large. Flat-top sampling has the merit that it simplifies the design of the electronic circuitry used to perform the sampling operation.

To show the extent of the distortion, consider the signal $m(t)$ having a Fourier transform $M(j\omega)$. We have seen (see Figs. 5.1-2 and 5.1-3) how to deduce the transform of the sampled signal, when the sampling is instantaneous. The transform of the sampled signal for flat-top sampling is determined by considering that the flat-top pulse can be generated by passing the instantaneously sampled signal through a network which broadens a pulse of duration $dt$ (an impulse) into a pulse of duration $\tau$. The transform of a pulse of unit amplitude and width $dt$ is

$$\mathcal{F}[\text{impulse of strength } dt \text{ at } t = 0] = dt \tag{5.5-1}$$

The transform of a pulse of unit amplitude and width $\tau$ is [see Eq. (1.10-20)]

$$\mathcal{F}\left[\text{pulse, amplitude} = 1, \text{ extending from } t = -\frac{\tau}{2} \text{ to } t = \frac{\tau}{2}\right] = \tau \frac{\sin(\omega\tau/2)}{\omega\tau/2} \tag{5.5-2}$$

Hence, the transfer function of the network shown in Fig. 5.5-1*b*, is required to be

$$H(j\omega) = \frac{\tau}{dt} \frac{\sin(\omega\tau/2)}{\omega\tau/2} \tag{5.5-3}$$

Let the signal $m(t)$, with transform $M(j\omega)$, be bandlimited to $f_M$ and be sampled at the Nyquist rate or faster. Then in the range 0 to $f_M$ the transform of the flat-topped sampled signal is given by the product $H(j\omega)M(j\omega)$ or, from Eqs. (5.5-1), (5.5-2), and (5.5-3)

$$\mathcal{F}[\text{flat-topped sampled } m(t)] = \frac{\tau}{T_s} \frac{\sin(\omega\tau/2)}{\omega\tau/2} M(j\omega) \qquad 0 \le f \le f_M \tag{5.5-4}$$

To illustrate the effect of flat-top sampling, we consider for simplicity that the signal $m(t)$ has a flat spectral density equal to $M_0$ over its entire range from 0 to $f_M$, as is shown in Fig. 5.5-2*a*. The form of the transform of the instantaneously sampled signal is shown in Fig. 5.5-2*b*. The sampling frequency $f_s = 1/T_s$ is assumed large enough to allow for a guard band between the spectrum of the baseband signal and the DSB-SC signal with carrier $f_s$. The spectrum of the flat-topped sampled signal is shown in Fig. 5.5-2*d*. We are, of course, interested only in the part of the spectrum in the range 0 to $f_M$. If, in this range, the spectra of the sampled signal and the original signal are identical, then the original signal may be recovered by a low-pass filter as has already been discussed. We observe,





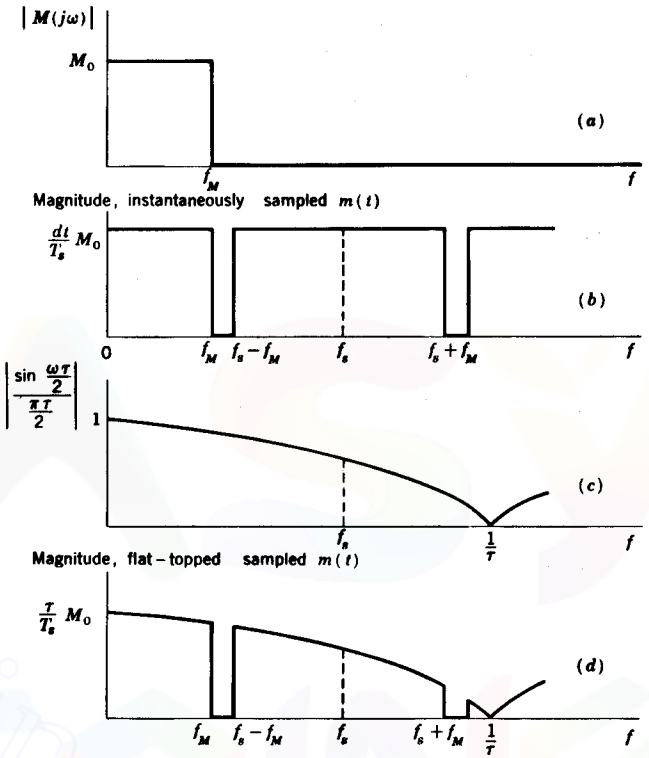



**Figure 5.5-2** (*a*) An idealized spectrum of a baseband signal. (*b*) The spectrum of the signal with instantaneous sampling. (*c*) The form $[(\sin x)/x$, with $x = \omega\tau/2]$ of the distortion factor (aperture effect) introduced by flat-topped sampling. (*d*) The spectrum of the signal with flat-topped sampling.

however, that such is not the case and that, as a result, distortion will result. This distortion results from the fact that the original signal was "observed" through a finite rather than an infinitesimal time "aperture" and is hence referred to as *aperture effect* distortion.

The distortion results from the fact that the spectrum is multiplied by the sampling function $Sa(x) \equiv (\sin x)/x$ (with $x = \omega\tau/2$). The magnitude of the sampling function (see Sec. 1.4) falls off slowly with increasing $x$ in the neighborhood of $x = 0$ and does not fall off sharply until we approach $x = \pi$, at which point $Sa(x) = 0$. To minimize the distortion due to the aperture effect, it is advantageous to arrange that $x = \pi$ correspond to a frequency very large in comparison with $f_M$. Since $x = \pi f\tau$, the frequency $f_0$ corresponding to $x = \pi$ is $f_0 = 1/\tau$. If $f_0 \gg f_M$, or, correspondingly, if $\tau \ll 1/f_M$, the aperture distortion will be small. The distortion becomes progressively smaller with decreasing $\tau$. And, of course, as $\tau \to 0$ (instantaneous sampling), the distortion similarly approaches zero.

### Equalization[1,2]

As in the case of natural sampling, so also in the present case of flat-top sampling, it is advantageous to make $\tau$ as large as practicable for the sake of increasing the amplitude of the output signal. If, in a particular case, it should happen that the consequent distortion is not acceptable, it may be corrected by including an *equalizer* in cascade with the output low-pass filter. An equalizer, in the present instance, is a passive network whose transfer function has a frequency dependence of the form $x/\sin x$, that is, a form inverse to the form of $H(j\omega)$ given in Eq. (5.5-3). The equalizer in combination with the aperture effect will then yield a flat overall transfer characteristic between the original baseband signal and the output at the receiving end of the system. The equalizer $x/\sin x$ cannot be exactly synthesized, but can be approximated.

If $N$ signals are multiplexed, $\tau \leq 1/2f_M N$, and hence for large $N$ $\tau \ll 1/f_M$ and $x/\sin x \simeq 1$. In this case the equalizer is not needed as negligible distortion results.

## 5.6 SIGNAL RECOVERY THROUGH HOLDING

We have already noted that the maximum ratio $\tau/T_s$, of the sample duration to the sampling interval, is $\tau/T_s = 1/N$, $N$ being the number of signals to be multiplexed. As $N$ increases, $\tau/T_s$ becomes progressively smaller, and, as is to be seen from Eq. (5.4-4), so correspondingly does the output signal. We discuss now an alternative method of recovery of the baseband signal which raises the level of the output signal (without the use of amplifiers which may introduce noise). The method has the additional advantage that rather rudimentary filtering is often quite adequate, but has the disadvantage that some distortion must be accepted.

The method is illustrated in Fig. 5.6-1, where the baseband signal $m(t)$ and its flat-topped samples are shown. At the receiving end, and after *demultiplexing*, the sample pulses are extended; that is, the sample value of each individual baseband signal is *held* until the occurrence of the next sample of that same baseband signal. This operation is shown in Fig. 5.6-1 as the dashed extension of the sample pulses. The output waveform consists then, as shown, of an up and down staircase waveform with no blank intervals.

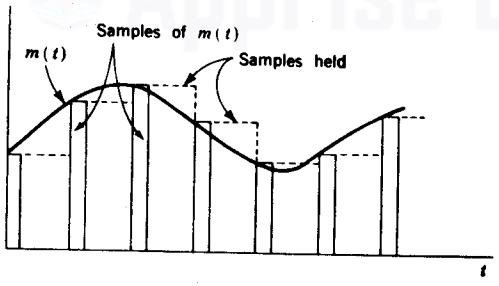



**Figure 5.6-1** Illustrating the operation of holding.

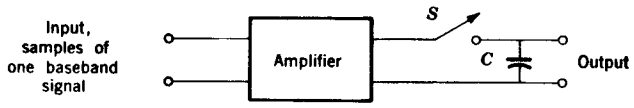



**Figure 5.6-2** Illustrating a method of performing the operation of holding.

A method, in principle, by which this holding operation may be performed is shown in Fig. 5.6-2. The switch $S$ operates in synchronism with the occurrence of input samples. This switch, ordinarily open, closes somewhat after the occurrence of the leading edge of a sample pulse and opens somewhat before the occurrence of the trailing edge. The amplifier, whose gain, if any, is incidental to the present discussion, has a low-output impedance. Hence, at the closing of the switch, the capacitor $C$ charges abruptly to a voltage proportional to the sample value, and the capacitor holds this voltage until the operation is repeated for the next sample. In Fig. 5.6-1 we have idealized the situation somewhat by showing the output waveform maintaining a perfectly constant level throughout the sample pulse interval and its following holding interval. We have also indicated abrupt transitions in voltage level from one sample to the next. In practice, these voltage transitions will be somewhat rounded as the capacitor charges and discharges exponentially. Further, if the received sample pulses are natural samples rather than flat-topped samples, there will be some departure from a constant voltage level during the sample interval itself. As a matter of practice however, the sample interval is very small in comparison with the interval between samples, and the voltage variation of the baseband signal during the sampling interval is small enough to be neglected.

If the baseband signal is $m(t)$ with spectral density $M(j\omega) = \mathcal{F}[m(t)]$, we may deduce the spectral density of the sampled and held waveform in the manner of Sec. 5.5 and in connection with flat-topped sampling. We need but to consider that the flat tops have been stretched to encompass the entire interval between instantaneous samples. Hence the spectral density is given as in Eq. (5.5-4) except with $\tau$ replaced by the time interval between samples. We have, then,

$$\mathcal{F}[m(t), \text{ sampled and held}] = \frac{\sin(\omega T_s/2)}{\omega T_s/2} M(j\omega) \qquad 0 \leq f \leq f_M \qquad (5.6\text{-}1)$$

In Fig. 5.6-3 we have again assumed for simplicity that the band-limited signal $m(t)$ has a flat spectral density of magnitude $M_0$. In Fig. 5.6-3*a* is shown the spectrum of the instantaneously sampled signal. In Fig. 5.6-3*b* has been drawn the magnitude of the aperture factor $(\sin x)/x$ (with $x = \omega T_s/2$), while in Fig. 5.6-3*c* is shown the magnitude of the spectrum of the sampled-and-held signal. These plots differ from the plots of Fig. 5.5-2 only in the location of the nulls of the factor $(\sin x)/x$. In Fig. 5.6-3 the first null occurs at the sampling frequency $f_s$. We observe that, as a consequence, the aperture effect, which is responsible for the $(\sin x)/x$ term, has accomplished most of the filtering which is required to suppress the part of the spectrum of the output signal above the

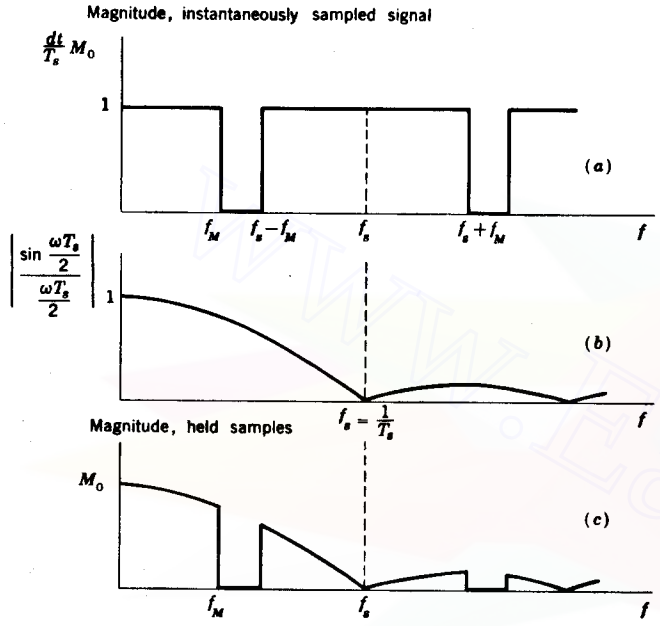



**Figure 5.6-3** (*a*) Spectrum of instantaneously sampled signal $m(t)$ with $m(t)$ having idealized spectrum shown in Fig. 5.5-2*a*. (*b*) The magnitude of the aperture effect factor. (*c*) Spectrum of sampled-and-held signal.

bandlimit $f_M$. Of course, the filtering is not perfect, and some additional filtering may be required. We also note that, as in the case of flat-top sampling, there will be some distortion introduced by the unequal transmission of spectral components in the range 0 to $f_M$. If the distortion is not acceptable, then, as before, it may be corrected by an $x/\sin x$ equalizer.

Most importantly we note in comparing Eq. (5.6-1) with Eq. (5.5-4) that, aside from the relatively small effect of the $(\sin x)/x$ terms in the two cases, the sampled-and-held signal has a magnitude larger by the factor $T_s/\tau$ than the signal of sample duration $\tau$. This increase in amplitude is, of course, intuitively to have been anticipated.

## 5.7 QUANTIZATION OF SIGNALS

The limitation of the system we have been describing for communicating over long channels is that once noise has been introduced any place along the channel, we are "stuck" with it. We now describe how the situation is modified by subjecting a signal to the operation of quantization. When quantizing a signal $m(t)$, we create a new signal $m_q(t)$ which is an approximation to $m(t)$. However, the quantized signal $m_q(t)$ has the great merit that it is, in large measure, separable from the additive noise.

The operation of quantization is represented in Fig. 5.7-1. Here we contemplate a signal $m(t)$ whose excursion is confined to the range from $V_L$ to $V_H$. We have divided this total range into $M$ equal intervals each of size $S$. Accordingly $S$, called the *step size*, is $S = (V_H - V_L)/M$. In Fig. 5.7-1 we show the specific example in which $M = 8$. In the center of each of these steps we locate *quantization levels* $m_0, m_1, \ldots, m_7$. The quantized signal $m_q(t)$ is generated in the following way: Whenever $m(t)$ is in the range $\Delta_0$, the signal $m_q(t)$ maintains the constant level $m_0$; whenever $m(t)$ is in the range $\Delta_1$, $m_q(t)$ maintains the constant level $m_1$; and so on. Thus the signal $m_q(t)$ will at all times be found at one of the levels $m_0, m_1, \ldots, m_7$. The transition in $m_q(t)$ from $m_q(t) = m_0$ to $m_q(t) = m_1$ is made abruptly when $m(t)$ passes the transition level $L_{01}$ which is midway between $m_0$ and $m_1$ and so on. To state the matter in an alternative fashion, we say that, at every instant of time, $m_q(t)$ has the value of the quantization level to which $m(t)$ is closest. Thus the signal $m_q(t)$ does not change at all with time or it makes a "quantum" jump of step size $S$. Note the disposition of the quantization levels in the range from $V_L$ to $V_H$. These levels are each separated by an amount $S$, but the separation of the extremes $V_L$ and $V_H$ each from its nearest quantization level is only $S/2$. Also, at every instant of time, the quantization error $m(t) - m_q(t)$ has a magnitude which is equal to or less than $S/2$.

We see, therefore, that the quantized signal is an approximation to the original signal. The quality of the approximation may be improved by reducing the size of the steps, thereby increasing the number of allowable levels. Eventually, with small enough steps, the human ear or the eye will not be able to distinguish the original from the quantized signal. To give the reader an idea of the number of quantization levels required in a practical system, we note that 256 levels can be used to obtain the quality of commercial color TV, while 64 levels gives only fairly good color TV performance. These results are also found to be valid when quantizing voice.

Now let us consider that our quantized signal has arrived at a repeater somewhat attenuated and corrupted by noise. This time our repeater consists of a quantizer and an amplifier. There is noise superimposed on the quantized levels of $m_q(t)$. But suppose that we have placed the repeater at a point on the communications channel where the instantaneous noise voltage is almost always less than half the separation between quantized levels. Then the output of the quantizer will consist of a succession of levels duplicating the original quantized signal and *with the noise removed*. In rare instances the noise results in an error in quantization level. A noisy quantized signal is shown in Fig. 5.7-2*a*. The allowable quantizer output levels are indicated by the dashed lines separated by amount $S$. The output of the quantizer is shown in Fig. 5.7-2*b*. The quantizer output is the level to which the input is closest. Therefore, as long as the noise has an instantaneous amplitude less than $S/2$, the noise will not appear at the output. One instance in which the noise does exceed $S/2$ is indicated in the figure, and, corre-


Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

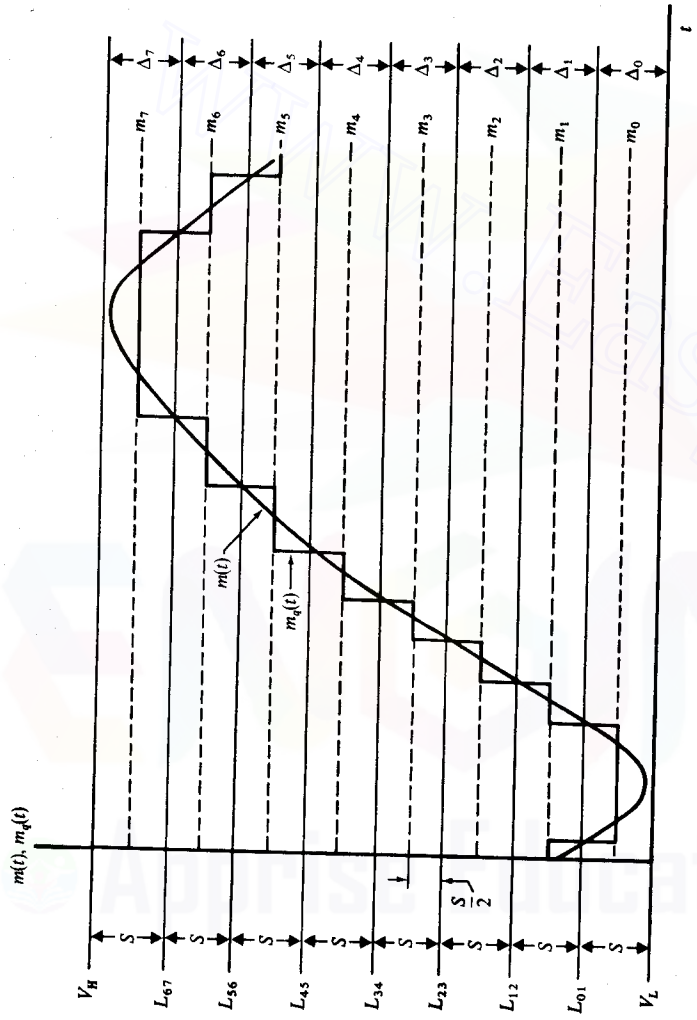



Figure 5.7-1 The operation of quantization.







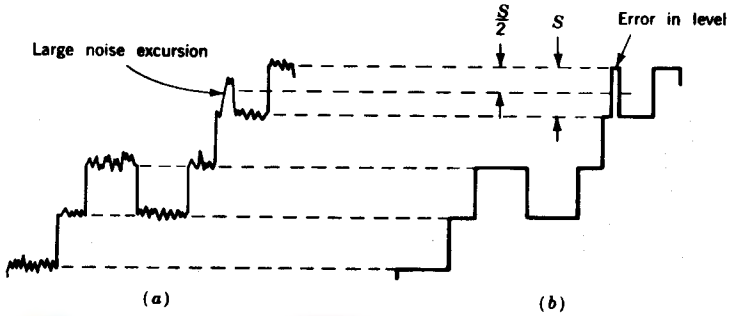



**Figure 5.7-2** (*a*) A quantized signal with added noise. (*b*) The signal after requantization. One instance is recorded in which the noise level is so large that an error results.

spondingly, an error in level does occur. The statistical nature of noise is such that even if the average noise magnitude is much less than $S/2$, there is always a finite probability that from time to time, the noise magnitude will exceed $S/2$. Note that it is never possible to suppress completely level errors such as the one indicated in Fig. 5.7-2.

We have shown that through the method of signal quantization, the effect of additive noise can be significantly reduced. By decreasing the spacing of the repeaters, we decrease the attenuation suffered by $m_q(t)$. This effectively decreases the relative noise power and hence decreases the probability $P_q$ of an error in level. $P_q$ can also be reduced by increasing the step size $S$. However, increasing $S$ results in an increased discrepancy between the true signal $m(t)$ and the quantized signal $m_q(t)$. This difference $m(t) - m_q(t)$ can be regarded as noise and is called *quantization noise*. Hence, the received signal is not a perfect replica of the transmitted signal $m(t)$. The difference between them is due to errors caused by additive noise and quantization noise. These noises are discussed further in Chap. 12.

## 5.8 QUANTIZATION ERROR

It has been pointed out that the quantized signal and the original signal from which it was derived differ from one another in a random manner. This difference or error may be viewed as a noise due to the quantization process and is called *quantization error*. We now calculate the mean-square quantization error $\overline{e^2}$, where $e$ is the difference between the original and quantized signal voltages.

Let us divide total peak-to-peak range of the message signal $m(t)$ into $M$ equal voltage intervals, each of magnitude $S$ volts. At the center of each voltage interval we locate a quantization level $m_1, m_2, \ldots, m_M$ as shown in Fig. 5.8-1*a*. The dashed level represents the instantaneous value of the message signal $m(t)$ at a time $t$. Since, in this figure, $m(t)$ happens to be closest to the level $m_k$, the quantizer output will be $m_k$, the voltage corresponding to that level. The error is $e = m(t) - m_k$.

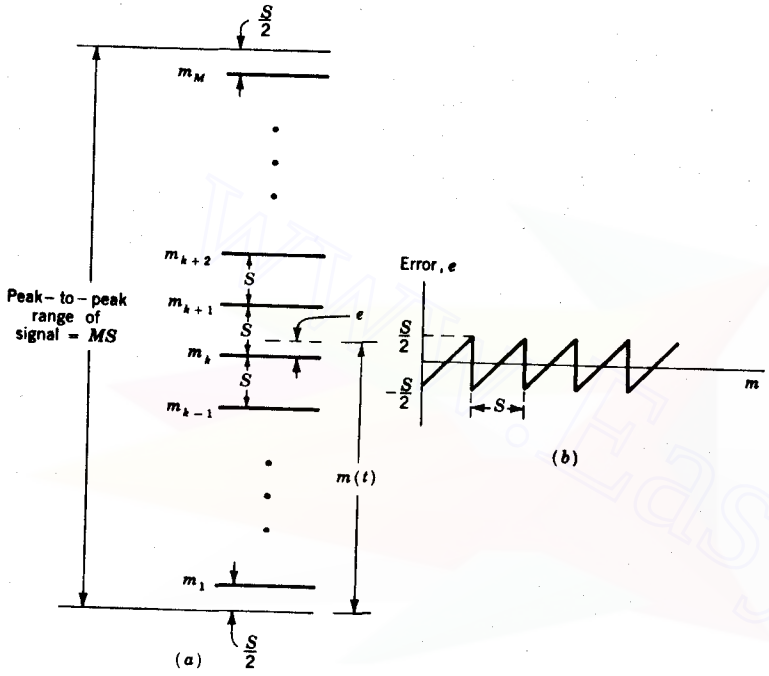



**Figure 5.8-1** (*a*) A range of voltage over which a signal $m(t)$ makes excursions is divided into $M$ quantization ranges each of size $S$. The quantization levels are located at the center of the range. (*b*) The error voltage $e(t)$ as a function of the instantaneous value of the signal $m(t)$.

Let $f(m)\,dm$ be the probability that $m(t)$ lies in the voltage range $m - dm/2$ to $m + dm/2$. Then the mean-square *quantization error* is

$$\overline{e^2} = \int_{m_1 - S/2}^{m_1 + S/2} f(m)(m - m_1)^2\,dm + \int_{m_2 - S/2}^{m_2 + S/2} f(m)(m - m_2)^2\,dm + \cdots \quad (5.8\text{-}1)$$

Now, ordinarily the probability density function $f(m)$ of the message signal $m(t)$ will certainly not be constant. However, suppose that the number $M$ of quantization is large, so that the step size $S$ is small in comparison with the peak-to-peak range of the message signal. In this case, it is certainly reasonable to make the approximation that $f(m)$ is constant within each quantization range. Then in the first term of Eq. (5.8-1) we set $f(m) = f^{(1)}$, a constant. In the second term $f(m) = f^{(2)}$, etc. We may now remove $f^{(1)}, f^{(2)}$, etc., from inside the integral sign. If we make the substitution $x \equiv m - m_k$, Eq. (5.8-1) becomes

$$\overline{e^2} = (f^{(1)} + f^{(2)} + \cdots) \int_{-S/2}^{S/2} x^2\,dx = (f^{(1)} + f^{(2)} + \cdots)\frac{S^3}{12} \quad (5.8\text{-}2a)$$

$$= (f^{(1)}S + f^{(2)}S + \cdots)\frac{S^2}{12} \quad (5.8\text{-}2b)$$





Now $f^{(1)}S$ is the probability that the signal voltage $m(t)$ will be in the first quantization range, $f^{(2)}S$ is the probability that $m$ is in the second quantization range, etc. Hence the sum of terms in the parentheses in Eq. (5.8-2*b*) has a total value of unity. Therefore, the mean-square quantization error is

$$\overline{e^2} = \frac{S^2}{12} \quad (5.8\text{-}3)$$

## 5.9 PULSE-CODE MODULATION

A signal which is to be quantized prior to transmission is usually sampled as well. The quantization is used to reduce the effects of noise, and the sampling allows us to time-division multiplex a number of messages if we choose to do so. The combined operations of sampling and quantizing generate a quantized PAM waveform, that is, a train of pulses whose amplitudes are restricted to a number of discrete magnitudes.

We may, if we choose, transmit these quantized sample values directly. Alternatively we may represent each quantized level by a code number and transmit the code number rather than the sample value itself. The merit of so doing will be developed in the subsequent discussion. Most frequently the code number is converted, before transmission, into its representation in binary arithmetic, i.e., base-2 arithmetic. The digits of the binary representation of the code number are transmitted as pulses. Hence the system of transmission is called (binary) *pulse-code modulation* (PCM).

We review briefly some elementary points about binary arithmetic. The binary system uses only two digits, 0 and 1. An arbitrary number $N$ is represented by the sequence $\ldots k_2 k_1 k_0$, in which the $k$'s are determined from the equation

$$N = \cdots + k_2 2^2 + k_1 2^1 + k_0 2^0 \quad (5.9\text{-}1)$$

with the added constraint that each $k$ has the value 0 or 1. The binary representations of the decimal numbers 0 to 15 are given in Table 5.9-1. Observe that to represent the four (decimal) numbers 0 to 3, we need only two binary digits $k_1$ and $k_0$. For the eight (decimal) numbers from 0 to 7 we require only three binary places, and so on. In general, if $M$ numbers $0, 1, \ldots, M-1$ are to be represented, then an $N$ binary digit sequence $k_{N-1} \cdots k_0$ is required, where $M = 2^N$.

The essential features of binary PCM are shown in Fig. 5.9-1. We assume that the analog message signal $m(t)$ is limited in its excursions to the range from $-4$ to $+4$ volts. We have set the step size between quantization levels at 1 volt. Eight quantization levels are employed, and these are located at $-3.5, -2.5, \ldots, +3.5$ volts. We assign the code number 0 to the level at $-3.5$ volts, the code number 1 to the level at $-2.5$ volts, etc., until the level at $+3.5$ volts, which is assigned the code number 7. Each code number has its representation in binary arithmetic ranging from 000 for code number 0 to 111 for code number 7.

In Fig. 5.9-1, in correspondence with each sample, we specify the sample value, the nearest quantization level, and the code number and its binary rep-

**Table 5.9-1 Equivalent numbers in decimal and binary representation**

| Binary | | | | Decimal |
|---|---|---|---|---|
| $k_3$ | $k_2$ | $k_1$ | $k_0$ | |
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 2 |
| 0 | 0 | 1 | 1 | 3 |
| 0 | 1 | 0 | 0 | 4 |
| 0 | 1 | 0 | 1 | 5 |
| 0 | 1 | 1 | 0 | 6 |
| 0 | 1 | 1 | 1 | 7 |
| 1 | 0 | 0 | 0 | 8 |
| 1 | 0 | 0 | 1 | 9 |
| 1 | 0 | 1 | 0 | 10 |
| 1 | 0 | 1 | 1 | 11 |
| 1 | 1 | 0 | 0 | 12 |
| 1 | 1 | 0 | 1 | 13 |
| 1 | 1 | 1 | 0 | 14 |
| 1 | 1 | 1 | 1 | 15 |

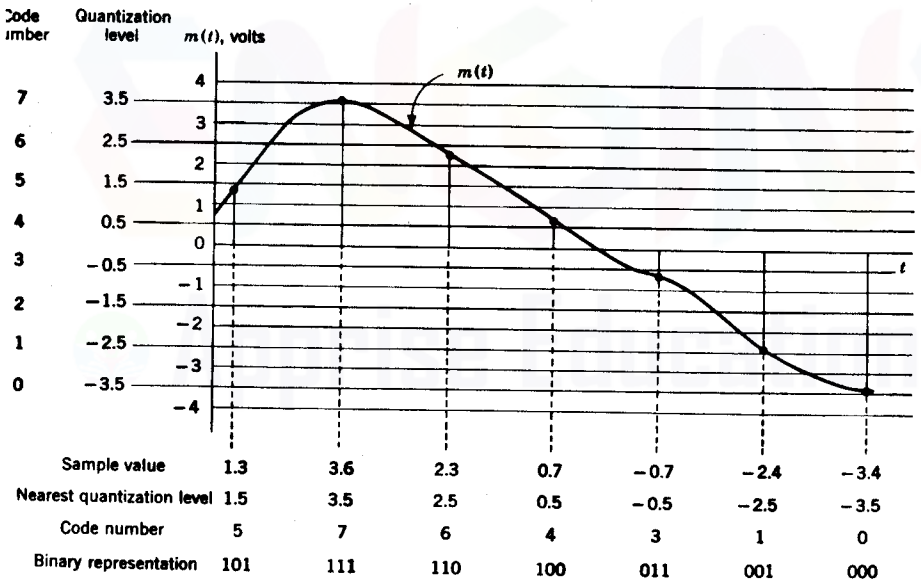



**Figure 5.9-1** A message signal is regularly sampled. Quantization levels are indicated. For each sample the quantized value is given and its binary representation is indicated.





resentation. If we were transmitting the analog signal, we would transmit the sample values 1.3, 3.6, 2.3, etc. If we were transmitting the quantized signal, we would transmit the quantized sample values 1.5, 3.5, 2.5, etc. In binary PCM we transmit the binary representations 101, 111, 110, etc.

## 5.10 ELECTRICAL REPRESENTATIONS OF BINARY DIGITS

As intimated in the previous section, we may represent the binary digits by electrical pulses in order to transmit the code representations of each quantized level over a communication channel. Such a representation is shown in Fig. 5.10-1. Pulse time slots are indicated at the top of the figure, and, as shown in Fig. 5.10-1*a*, the binary digit 1 is represented by a pulse, while the binary digit 0 is represented by the absence of a pulse. The row of three-digit binary numbers given in Fig. 5.10-1 is the binary representation of the sequence of quantized samples in Fig. 5.9-1. Hence the pulse pattern in Fig. 5.10-1*a* is the (binary) PCM waveform that would be transmitted to convey to the receiver the sequence of quantized samples of the message signal $m(t)$ in Fig. 5.9-1. Each three-digit binary number that specifies a quantized sample value is called a *word*. The spaces between words allow for the multiplexing of other messages.

At the receiver, in order to reconstruct the quantized signal, all that is required is that a determination be made, within each pulse time slot, about whether a pulse is present or absent. The exact amplitude of the pulse is not important. There is an advantage in making the pulse width as wide as possible since the pulse energy is thereby increased and it becomes easier to recognize a pulse against the background noise. Suppose then that we eliminate the guard time $\tau_g$ between pulses. We would then have the waveform shown in Fig. 5.10-1*b*. We would be rather hard put to describe this waveform as either a sequence of positive pulses or of negative pulses. The waveform consists now of a sequence of transitions between two levels. When the waveform occupies the lower level in a

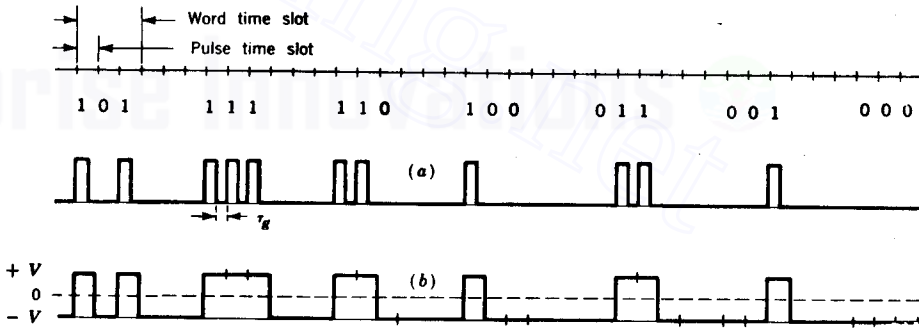



**Figure 5.10-1** (*a*) Pulse representation of the binary numbers used to code the samples in Fig. 5.9-1. (*b*) Representation by voltage levels rather than pulses.

particular time slot, a binary 0 is represented, while the upper voltage level represents a binary 1.

Suppose that the voltage difference of $2V$ volts between the levels of the waveform of Fig. 5.10-1*b* is adequate to allow reliable determination at the receiver of which digit is being transmitted. We might then arrange, say, that the waveform make excursions between 0 and $2V$ volts or between $-V$ volts and $+V$ volts. The former waveform will have a dc component, the latter waveform will not. Since the dc component wastes power and contributes nothing to the reliability of transmission, the latter alternative is preferred and is indicated in Fig. 5.10-1*b*.

## 5.11 THE PCM SYSTEM

### The Encoder

A PCM communication system is represented in Fig. 5.11-1. The analog signal $m(t)$ is sampled, and these samples are subjected to the operation of quantization. The quantized samples are applied to an *encoder*. The encoder responds to each such sample by the generation of a unique and identifiable binary pulse (or binary level) pattern. In the example of Figs. 5.9-1 and 5.10-1 the pulse pattern happens to have a numerical significance which is the same as the order assigned to the quantized levels. However, this feature is not essential. We could have assigned any pulse pattern to any level. At the receiver, however, we must be able to identify the level from the pulse pattern. Hence it is clear that not only does the encoder number the level, it also assigns to it an identification code.

The combination of the quantizer and encoder in the dashed box of Fig. 5.11-1 is called an *analog-to-digital converter*, usually abbreviated A/D converter. In commercially available A/D converters there is normally no sharp distinction between that portion of the electronic circuitry used to do the quantizing and that portion used to accomplish the encoding. In summary, then, the A/D converter accepts an analog signal and replaces it with a succession of *code* symbols, each symbol consisting of a train of pulses in which each pulse may be interpreted as the representation of a *digit* in an arithmetic system. Thus the signal transmitted over the communications channel in a PCM system is referred to as a digitally encoded signal.

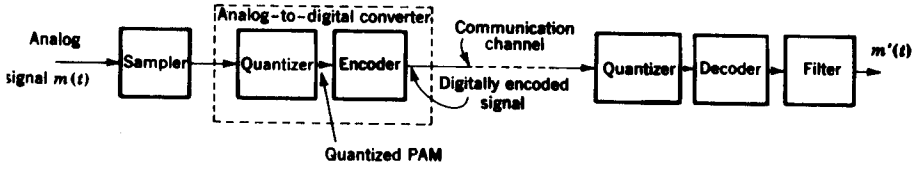



**Figure 5.11-1** A PCM communication system.

### The Decoder

When the digitally encoded signal arrives at the receiver (or repeater), the first operation to be performed is the separation of the signal from the noise which has been added during the transmission along the channel. As noted previously, separation of the signal from the noise is possible because of the quantization of the signal. Such an operation is again an operation of *requantization;* hence the first block in the receiver in Fig. 5.11-1 is termed a quantizer. A feature which eases the burden on this quantizer is that for each pulse interval it has only to make the relatively simple decision of whether a pulse has or has not been received or which of two voltage levels has occurred. Suppose the quantized sample pulses had been transmitted instead, rather than the binary-encoded codes for such samples. Then this quantizer would have had to have yielded, in each pulse interval, not a simple yes or no decision, but rather a more complicated determination about which of the many possible levels had been received. In the example of Fig. 5.10-1, if a quantized PAM signal had been transmitted, the receiver quantizer would have to decide which of the levels 0 to 7 was transmitted, while with a binary PCM signal the quantizer need only distinguish between two possible levels. The relative reliability of the yes or no decision in PCM over the multivalued decision required for quantized PAM constitutes an important advantage for PCM.

The receiver quantizer then, in each pulse slot, makes an educated and sophisticated estimate and then decides whether a positive pulse or a negative pulse was received and transmits its decisions, in the form of a reconstituted or regenerated pulse train, to the decoder. (If repeater operation is intended, the regenerated pulse train is simply raised in level and sent along the next section of the transmission channel.) The decoder, also called a *digital-to-analog* (*D/A*) *converter*, performs the inverse operation of the encoder. The decoder output is the sequence of quantized multilevel sample pulses. The quantized PAM signal is now reconstituted. It is then filtered to reject any frequency components lying outside of the baseband. The final output signal $m'(t)$ is identical with the input $m(t)$ except for quantization noise and the occasional error in yes-no decision making at the receiver due to the presence of channel noise.

## 5.12 COMPANDING

Referring again to Figs. 5.7-1 and 5.8-1 let us consider that we have established a quantization process employing $M$ levels with step size $S$, the levels being established at voltages to accommodate a signal $m(t)$ which ranges from a low voltage $V_L$ to a high voltage $V_H$. We can readily see that if the signal $m(t)$ should make excursions beyond the bounds $V_H$ and $V_L$ the system will operate at a disadvantage. For, within these bounds, the instantaneous quantization error never exceeds $\pm S/2$ while outside these bounds the error is larger.

Further, whenever $m(t)$ does not swing through the full available range the system is equally at a disadvantage. For, in order that $m_q(t)$ be a good approx-


Downloaded From : www.EasyEngineering.net

imation to $m(t)$ it is necessary that the step size $S$ be small in comparison to the *range over which* $m(t)$ *swings*. As a very pointed example of this consideration consider a case in which $m(t)$ has a peak-to-peak voltage which is less than $S$ and never crosses one of the transition levels in Fig. 5.7-1. In such a case $m_q(t)$ will be a fixed (dc) voltage and will bear no relationship to $m(t)$.

To explore this latter point somewhat more quantitatively let us consider that $m(t)$ is a signal, such as the sound signal output of a microphone, in which $V_H = -V_L = V$, i.e., a signal without dc components, and with (at least approximately) equal positive and negative peaks. Further, for simplicity, let us assume that in the range $\pm V$ the signal $m(t)$ is characterized by a uniform probability density. The probability density is then equal to $1/2V$ and the normalized average signal power of the applied input signal is

$$S_i = \overline{m^2(t)} = \int_{-V}^{+V} m^2(t) \frac{1}{2V}\, dm = \frac{V^2}{3} \tag{5.12-1}$$

The quantization noise, as given by Eq. (5.8-3) is

$$N_Q = \frac{S^2}{12} \tag{5.12-2}$$

If the number of quantization levels is $M$, then $MS = 2V$ so that

$$V = \frac{MS}{2} \tag{5.12-3}$$

Combining Eqs. (5.12-1), (5.12-2), and (5.12-3) we have that the input signal-to-quantization noise power ratio is

$$\frac{S_i}{N_Q} = M^2 \tag{5.12-4}$$

Eventually the received quantized signal will be smoothed out to generate an output signal with power $S_o$. If we have a useful communication system then presumably the effect of the quantization noise is not such as to cause an easily perceived difference between input and output signal. In such a case the output power $S_o$ may be taken to be the same as the input power i.e., $S_o \simeq S_i$ so that finally we may replace Eq. (5.12-4) by

$$\frac{S_o}{N_Q} = M^2 \tag{5.12-5}$$

If there are $M$ quantization levels then the code which singles out the closest quantization to the signal $m(t)$ will have to have $N$ bits with $M = 2^N$. Hence Eq. (5.12-5) becomes

$$\frac{S_o}{N_Q} = (2^N)^2 = 2^{2N} \tag{5.12-6}$$





In decibels we have

$$\left[\frac{S_o}{N_Q}\right]_{\text{dB}} = 10 \log_{10} \left[\frac{S_o}{N_Q}\right] = 10 \log_{10} 2^{2N} = 6N \tag{5.12-7}$$

Equation (5.12-7) has the interpretation that, in a system where the signal is quantized using an $N$-bit code (i.e., the number of quantization levels is $2^N$), and where the signal amplitude is capable of swinging through all available quantization regions without extending beyond the outermost ranges, the output *signal-to-quantization noise ratio* is $6N$ dB. In voice communication we use $N = 8$ corresponding to 256 quantization regions and $S_o/N_Q = (6)(8) = 48$ dB. If the signal is reduced in amplitude so that not all quantization ranges are used then $S_o/N_Q$ becomes smaller, since $N_Q$, depending as it does only on step $S$ is not affected by the amplitude reduction, while $S_o$ is reduced. For example, if the amplitude were reduced by a factor of 2, the power is then reduced by a factor of 4 reducing the signal-to-noise ratio by 6 dB. It is interesting to observe that the effective number of quantization levels is also reduced by a factor of 2. Correspondingly, the number $N$ of code bits is reduced by 1. In summary, the dependence of $S_o/N_Q$ on the input signal power $S_i$ is such that as the number of code bits needed decreases, $S_o/N_Q$ decreases by 6 dB/bit.

It is generally required, for acceptable voice transmission, that the received signal have a ratio $S_o/N_Q$ not less than 30 dB, and that this minimum 30 dB figure hold even though the signal power itself may vary by 40 dB. (The signal, in this case, is described as having a 40 dB *dynamic* range.) In an 8-bit system, at maximum signal level we have $S_o/N_Q \cong S_i/N_Q = 48$ dB. Now $N_Q$ is fixed, and depends only on step size. If we are to allow $S_i/N_Q$ to drop to no lower than 30 dB then the dynamic range would be restricted to $48 - 30 = 18$ dB.

The dynamic range can be materially improved by a process called *companding* (a word formed by combining the words *compressing* and *expanding*). As we have seen, to keep the signal-to-quantization noise ratio high we must use a signal which swings through a range which is large in comparison with the step size. This requirement is not satisfied when the signal is small. Accordingly before applying the signal to the quantizer we pass it through a network which has an input-output characteristic as shown in Fig. 5.12-1. Note that at low amplitudes the slope is larger than at large amplitudes. A signal transmitted through such a network will have the extremities of its waveform *compressed*. The peak signal which the system is intended to accommodate will, as before, range through all available quantization regions. But now, a small amplitude signal will range through *more* quantization regions than would be the case in the absence of compression. Of course, the compression produces signal distortion. To undo the distortion, at the receiver we pass the recovered signal through an *expander* network. An expander network has an input-output characteristic which is the inverse of the characteristic of the compressor. The inverse distortions of compressor and expander generate a final output signal without distortion.

The determination of the form of the compression plot of Fig. 5.12-1 is a somewhat subjective matter. In the United States, Canada, and Japan a *μ-law*

Downloaded From : www.EasyEngineering.net

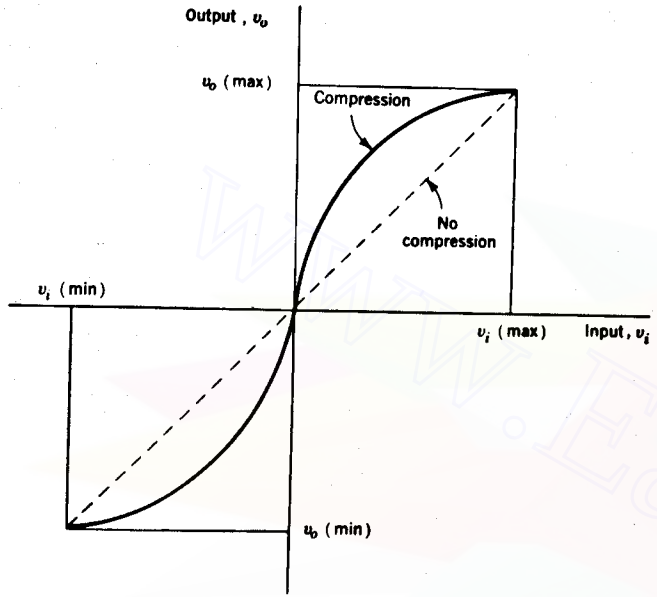



**Figure 5.12-1** An input–output characteristic which provides compression.

compandor is used (see Prob. 5.12-1) and it differs somewhat from the *A-law* compander (see Prob. 5.12-2) used by the rest of the world. The "$\mu$" and the "$A$" refer to parameters which appear in the equations for the compression and expansion characteristic.

In implementing the compression characteristic, the analog signal $m(t)$ is left unmodified and, instead, the step size is tapered so that the quantization levels are close together at low signal amplitudes and progressively further apart in regions attained as the signal increases in amplitude. To see one way in which the step sizes are altered consider again an 8-bit PCM system employing $2^8 = 256$ quantization levels. In such a system, the A/D converter shown in Fig. 5.11-1 would ordinarily be an 8-bit converter. Each input sample would generate an 8-bit output code identifying the closest quantization level. If the total range of the input was $\pm V$ then the step size would be $S = 2V/2^8$. Let us start however with a 12-bit A/D converter so that the step size is $2V/2^{12}$. This 12-bit PCM signal is not applied to the communication channel directly but is instead applied to the address pins of a read-only memory (ROM) whose content is to be described. The ROM has 12 address pins and 8 output data pins. The signal transmitted is the 8-bit data output of the ROM.

The content of the ROM is as follows: As shown in Fig. 5.12-2, in each successive memory location in the region where the step size is to be smallest (i.e., step size $= 2V/2^{12} = \Delta$) there are written successive 8-bit code words. Hence in this region a change in analog signal of step size $\Delta$ will change the code word presented for transmission by the amount $\Delta$. This one-to-one correspondence between addresses and transmitted code applies as shown for the middle 64 addresses. For the next 64 addresses (32 on one side and 32 on the other side of the original 64) we arrange that at the memory locations specified by pairs of addresses, i.e., for *two* adjacent addresses the *same* code word is written into the memory. Hence the transmitted code word will change at every *second* addresses change and correspondingly the step size is $2\Delta$. Next we arrange that for the next

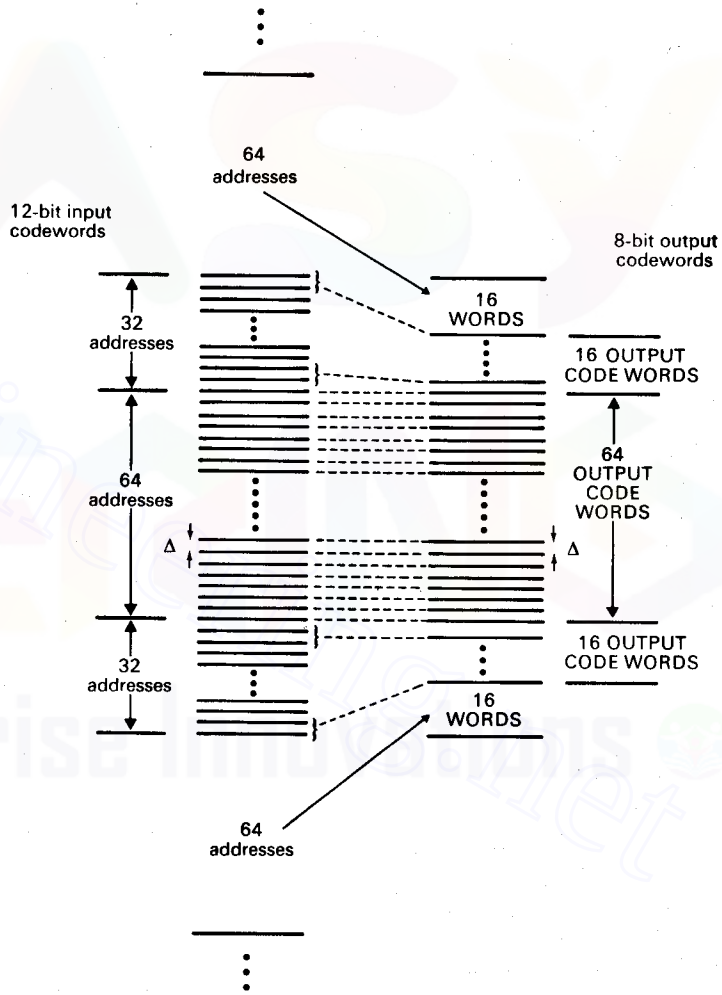



**Figure 5.12-2** ROM characteristic for compression.

Downloaded From : www.EasyEngineering.net

Downloaded From : www.EasyEngineering.net

128 addresses (64 above + 64 below) the *same* code word is written into the locations of *four* successive memory locations so that the step size is 4Δ. We proceed in this manner, each time doubling the number of successive memory locations which contain the same code word. As can be verified, when we have used all $2^{12} = 4096$ addresses we shall have generated $2^8 = 256$ code words. At the receiver we have a mechanism for generating the quantization corresponding to each code word.

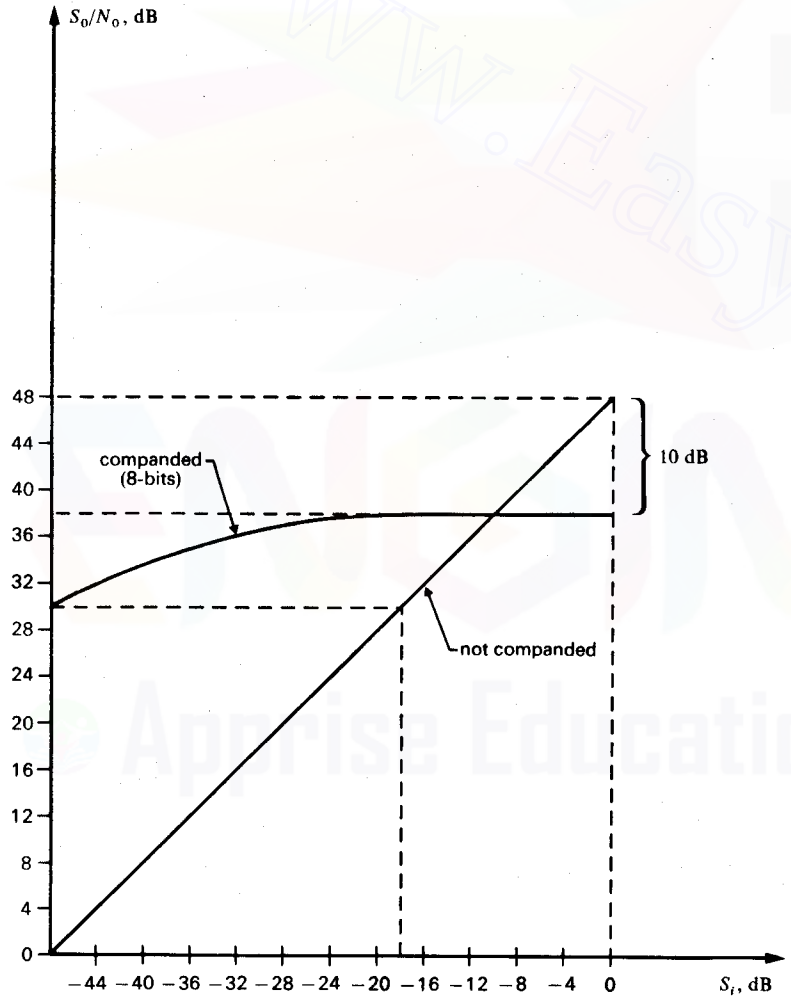



**Figure 5.12-3** Comparison of companded and uncompanded systems.



Downloaded From : www.EasyEngineering.net

It is interesting to observe that for the smallest 64 levels, the input and output signals are the same, i.e., small signals are *not* companded. One reason for this is that signals other than voice are often digitized using the same PCM system employed for voice. If a non-voice signal is subjected to the compression algorithm the result is usually a degradation of performance. Therefore, to avoid this possibility, such non-voice signals are kept 40 dB below the peak level of a voice signal.

In conclusion, we refer to Fig. 5.12-3 which compares the variation of output signal-to-noise ratio as a function of input signal power when companding is used, to the case of an uncompanded system. Note that the companded system has a far greater dynamic range than the uncompanded system and that theoretically the companded system has an output signal-to-noise ratio which exceeds 30 dB over a dynamic range of input signal power of 48 dB, while the uncompanded system has a dynamic range of 18 dB for the same conditions. It should be noted however, that the penalty paid using companding is approximately 10 dB. Thus an 8-bit uncompanded system, when operating at maximum amplitude, produces a signal-to-noise ratio of 48 dB. The same 8-bit system, using a compandor, yields a 38 dB SNR.

## 5.13 MULTIPLEXING PCM SIGNALS[3]

We have already noted the advantage of converting an analog signal into a PCM waveform when the signal must be transmitted over a noisy channel. When a large number of such PCM signals must be transmitted over a common channel, multiplexing of these PCM signals is required. In this section we discuss the multiplexing methods for PCM waveforms used in the United States by the common carriers (AT&T, GTE, etc.)

### 5.13-1 The *T*1 Digital System

Figure 5.13-1 shows the basic time division multiplexing scheme, called the *T*1 digital system, which is used to convey multiple signals over telephone lines using wideband coaxial cable. It accommodates 24 analog signals which we shall refer to as $s_1$ through $s_{24}$. Each signal is bandlimited to approximately 3.3 kHz and is sampled at the rate 8 kHz which exceeds, by a comfortable margin, the Nyquist rate of $2 \times 3.3 = 6.6$ kHz. Each of the time division multiplexed signals (still analog) is next A/D converted and companded as described in Sec. 5.12. The resulting digital waveform is transmitted over a coaxial cable, the cable serving to minimize signal distortion and serving also to suppress signal corruption due to noises from external sources. Periodically, at approximately 6000 ft intervals, the signal is regenerated by amplifiers called repeaters and then sent on toward its eventual destination. The repeater eliminates from each bit the effect of the distortion introduced by the channel. Also, the repeater removes from each bit any superimposed noise and thus, even having received a distorted and noisy signal, it

Downloaded From : www.EasyEngineering.net

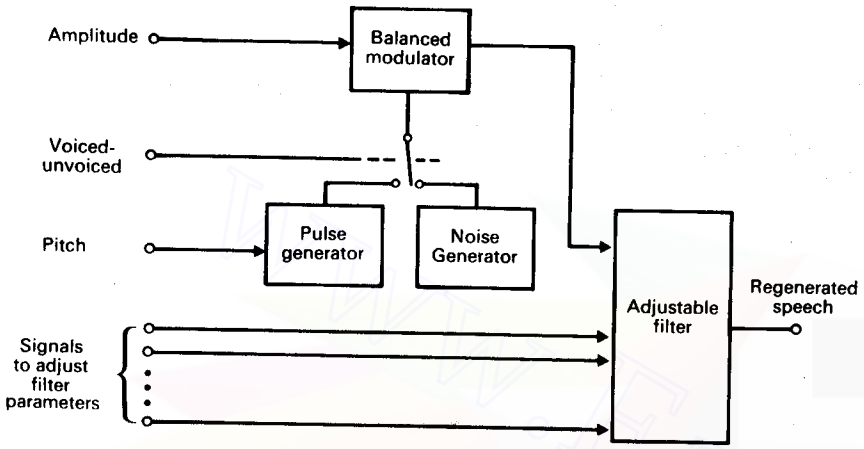



**Figure 5.19-2** A simplified linear predictive decoder.

pitch signals. These are used in connection with the modulator, and pulse and noise generators, as at the encoder, to provide an input to the adjustable filter. The filter-parameter adjusting signals are also received and are used, as at the encoder, to adjust the filter characteristics for optimum voice regeneration.

Not explicitly shown in Figs. 5.19-1 and 5.19-2, but nonetheless to be understood is that, as in the simpler vocoder of Fig. 5.18-1, the transmitted signal must be the time-division multiplex of the individual signals to be transmitted. Typically, 18 filter-adjusting signals $a_i$ are employed. If, as before we sample at the rate 40 samples/s and encode each sample value into three bits, the bit rate $R$ is

$$R = (18 + 3)\,\text{signals} \times 40\,\frac{\text{samples/s}}{\text{signal}} \times 3\ \text{bits/sample}$$

$$= 2.52\ \text{kb/s} \qquad (5.19\text{-}1)$$

## REFERENCES


. Lucky, R. W., J. Salz, and E. J. Weldon, Jr.: "Principles of Data Communication," pp. 61–87, McGraw-Hill Book Company, New York, 1968.

. Lucky, R. W., J. Salz, and E. J. Weldon, Jr.: "Principles of Data Communication," pp. 128–165, McGraw-Hill Book Company, New York, 1968.

. Bell Telephone Laboratories: "Telecommunications Transmission Engineering," Vol II, AT&T, Western Electric Company, Tech. Pub., Winston-Salem, N.C., 1977.

. de Jager, F.: Deltamodulation: A Method of PCM Transmission Using a 1-unit Code, *Philips Res. Rept.* 7, pp. 442–446, 1952.

. Jayant, N. S., and P. Noll: "Digital Coding of Waveforms," Prentice-Hall Inc., Englewood Cliffs, N.J., 1984.




Downloaded From : www.EasyEngineering.net

## PROBLEMS

**5.1-1.** It is required to transmit telephone messages across the United States, a 3000 mile run. The signal level is not to be allowed to drop below 1 millivolt before amplification and the signal is not to be allowed to be larger than 15 volts in order to avoid amplifier overload. Assuming that repeaters are to be located with equal spacings, how many repeaters will be required?

**5.1-2.** A bandpass signal has a spectral range that extends from 20 to 82 kHz. Find the acceptable range of the sampling frequency $f_s$.

**5.1-3.** A bandpass signal has a center frequency $f_0$ and extends from $f_0 - 5$ kHz to $f_0 + 5$ kHz. The signal is sampled at a rate $f_s = 25$ kHz. As the center frequency $f_0$ varies from $f_0 = 5$ kHz to $f_0 = 50$ kHz find the ranges of $f_0$ for which the sampling rate is adequate.

**5.1-4.** The signal $v(t) = \cos 5\pi t + 0.5 \cos 10\pi t$ is instantaneously sampled. The interval between samples is $T_s$.

(*a*) Find the maximum allowable value for $T_s$.

(*b*) If the sampling signal is $S(t) = 5 \sum_{k=-\infty}^{\infty} \delta(t - 0.1k)$, the sampled signal $v_s(t) = v(t)S(t)$ consists of a train of impulses, each with a different strength $v_s(t) = \sum_{k=-\infty}^{\infty} I_k\, \delta(t - 0.1k)$. Find $I_0$, $I_1$, and $I_2$, and show that $I_k = I_{4+k}$.

(*c*) To reconstruct the signal $v_s(t)$ is passed through a rectangular low-pass filter. Find the minimum filter bandwidth to reconstruct the signal without distortion.

**5.1-5.** We have the signal $v(t) = \cos 2\pi f_0 t + \cos 2 \times 2\pi f_0 t + \cos 3 \times 2\pi f_0 t$. Our interest extends, however, only to spectral components up to and including $2f_0$. We therefore sample at the rate $5f_0$ which is adequate for the $2f_0$ component of the signal.

(*a*) If sampling is asccomplished by multiplying $v(t)$ by an impulse train in which the impulses are of unit strength, write an expression for the sampled signal.

(*b*) To recover the part of the signal of interest, the sampled signal is passed through a rectangular low-pass filter with passband extending from 0 to slightly beyond $2f_0$. Write an expression for the filter output. Is the part of the signal of interest recovered exactly? If we want to reproduce the first two terms of $v(t)$ without distortion, what operation must be performed at the very outset?

**5.1-6.** The bandpass signal $v(t) = \cos 10\omega_0 t + \cos 11\omega_0 t + \cos 12\omega_0 t$ is sampled by an impulse train $S(t) = I \sum_{k=-\infty}^{\infty} \delta(t - kT_s)$.

(*a*) Find the maximum time between samples, $T_s$, to ensure reproduction without error.

(*b*) Using the result obtained in (*a*), obtain an expression for $v_s(t) = S(t)v(t)$.

(*c*) The sampled signal $v_s(t)$ is filtered by a rectangular low-pass filter with a bandwidth $B = 2f_0$. Obtain an expression for the filter output.

(*d*) The sampled signal $v_s(t)$ is filtered by a rectangular bandpass filter extending from $2f_0$ to $4f_0$. Obtain an expression for the filter output.

**5.1-7.** The bandpass signal $v(t) = \cos 10\omega_0 t + \cos 11\omega_0 t + \cos 12\omega_0 t$ is sampled by an impulse train, $S(t) = I \sum_{k=-\infty}^{\infty} \delta(t - k/8f_0)$. The sampled signal $v_s(t) = S(t)v(t)$ is then filtered by a rectangular low-pass filter having a bandwidth $B = 2f_0$. Obtain an expression for the filter output.

**5.1-8.** Let us view the waveform $v(t) = \cos \omega_0 t$ as a bandpass signal occupying an arbitrarily narrow frequency band. On this basis we find that the required sampling rate is $f_s = 0$. Discuss.

**5.2-1.** The TDM system shown in Fig. 5.2-1 is used to multiplex the four signals $m_1(t) = \cos \omega_0 t$, $m_2(t) = 0.5 \cos \omega_0 t$, $m_3(t) = 2 \cos 2\omega_0 t$, and $m_4(t) = \cos 4\omega_0 t$.

(*a*) If each signal is sampled at the same sampling rate, calculate the minimum sampling rate $f_s$.

(*b*) What is the commutator speed in revolutions per second.

(*c*) Design a commutator which will allow each of the four signals to be sampled at a rate faster than is required to satisfy the Nyquist criterion for the individual signal.

**5.2-2.** Three signals $m_1$, $m_2$, and $m_3$ are to be multiplexed. $m_1$ and $m_2$ have a 5-kHz bandwidth, and $m_3$ has a 10-kHz bandwidth. Design a commutator switching system so that each signal is sampled at its Nyquist rate.

Downloaded From : www.EasyEngineering.net

**5.3-1.** Show that the response of a rectangular low-pass filter, with a bandwidth $f_c$, to the impulse function $I\,\delta(t - k/2f_c)$ is

$$S_R(t) = \frac{I\omega_c}{\pi}\,\frac{\sin\,\omega_c(t - k/2f_c)}{\omega_c(t - k/2f_c)}$$

Assume that in its passband the filter has $H(f) = 1$.

**5.3-2.** Four signals, $m_1(t) = 1\cos\omega_0 t$, $m_2(t) = 1\sin\omega_0 t$, $m_3(t) = -1\cos\omega_0 t$, and $m_4(t) = -1\sin\omega_0 t$ are sampled every $1/2f_0$ sec by the sampling function

$$S(t) = 1\sum_{k=-\infty}^{\infty}\delta\left(t - \frac{k}{2f_0}\right)$$

The signals are then time-division multiplexed. The TDM signal is filtered by a rectangular low-pass filter having a bandwidth $f_c = 4f_0$ and then decommutated.

(*a*) Sketch the four outputs of the decommutator.

(*b*) Each of the four output signals is filtered by a rectangular low-pass filter having a bandwidth $f_0$. Show that the four signals are reconstructed without error.

**5.3-3.** The four signals of Prob. 5.3-2 are sampled, as indicated in that problem, and time-division multiplexed. The TDM signal is filtered by a rectangular low-pass filter having a bandwidth $f_c = 2f_0$ and then decommutated. Sketch the output at the decommutator switch segment where the samples of $m_1(t)$ should appear and show that $m_1(t)$ cannot be recovered.

**5.4-1.** The signal $v(t) = \cos\omega_0 t + \cos 8\omega_0 t$ is sampled by using *natural sampling.*

(*a*) Determine the minimum sampling rate $f_s$.

(*b*) Sketch $v_s(t) = S(t)v(t)$ if $S(t)$ is a train of pulses having unit height, occurring at the rate $f_s$; and $S(t) = 1$ for $nT - \tau/2 \le t \le nT + \tau/2$. The pulse duration is $\tau = 1/32f_0$.

(*c*) Repeat (*b*) if $\tau = 1/320f_0$.

**5.5-1.** Show that an impulse function $I\,\delta(t)$ can be stretched to have a width $\tau$ by passing the impulse function through a filter $(1 - e^{-j\omega\tau})/j\omega$. Show that this operation is identical with integrating the impulse for $\tau$ sec, that is, that the output $v_o(t)$ is given by

$$v_o(t) = \int_0^t I\,\delta(t)\,dt \qquad 0 \le t \le \tau$$
$$= 0 \qquad \text{otherwise}$$

**5.8-1.** For the quantizer characteristic shown in Fig. 5.8-1*b*.

(*a*) Plot the error characteristic $e = v_o - v_i$ versus $v_i$. Assume that $S = 1$, that is, $m_{k+1} - m_k = 1$ volt.

(*b*) Since the error $e = v_o - v_i$ is periodic, it can be expanded in a Fourier series. Write the Fourier series for the error $e = e(v_i)$.

(*c*) If $v_i = S\sin\omega_0 t$, find the component of the error $e$ at the angular frequency $\omega_0$.

**5.8-2.** Show that if the signal is uniformly distributed, Eq. (5.8-3) results even if $M$ is not large.

**5.8-3.** Consider a signal having a probability density

$$f(v) = \begin{cases} Ke^{-|v|} & -4 < v < 4 \\ 0 & \text{elsewhere} \end{cases}$$

(*a*) Find $K$.

(*b*) Determine the step size $S$ if there are four quantization levels.

(*c*) Calculate the variance of the quantization error when there are four quantization levels. Do not assume that $f(v)$ is constant over each level. Compare your result with Eq. (5.8-3).

**5.8-4.** Consider a signal having a probability density

$$f(v) = K(1 - |v|) \qquad -1 \le v \le 1$$

Calculate (*a*) to (*c*) of Prob. 5.8-3.



Downloaded From : www.EasyEngineering.net

**5.9-1.** Show that the numbers 0 to 7 can be written using 3 binary digits (bits). How many bits are required to write the numbers 0 to 5?

**5.10-1.** Consider that the signal $\cos 2\pi t$ is quantized into 16 levels. The sampling rate is 4 Hz. Assume that the sampling signal consists of pulses each having a unit height and duration $dt$. The pulses occur every $t = k/4$ sec, $-\infty < k < \infty$.

(*a*) Sketch the binary signal representing each sample voltage.

(*b*) How many bits are required per sample?

**5.11-1.** A D/A converter is shown in Fig. P5.11-1. Using the set-reset flip-flops shown explain the operation of the device.

**5.11-2.** An A/D converter is shown in Fig. P5.11-2. Using *trigger* flip-flops as indicated explain the operation of the device.

**5.12-1.** A $\mu$-law compander uses a compressor which relates output to input by the relation

$$y = \pm\frac{\log(1 + \mu|x|)}{\log(1 + \mu)}$$

Here the $+$ sign applies when $x$ is positive and the $-$ sign applies when $x$ is negative. Also $x \equiv v_i/V$ and $y = v_o/V$ where $v_i$ and $v_o$ are the input and output voltages and the range of allowable voltage is $-V$ to $+V$. The parameter $\mu$ determines the degree of compression.

(*a*) A commonly used value is $\mu = 255$. For this value make a plot of $y$ versus $x$ from $x = -1$ to $x = +1$.

(*b*) If $V = 40$ volts and 256 quantization levels are employed what is the voltage interval between levels when there is no compression? For $\mu = 255$ what is the minimum and what is the maximum effective separation between levels?

**5.12-2.** An $A$-law compander uses a compressor which relates output to input by the relations

$$y = \pm\frac{A|x|}{1 + \log A} \qquad \text{for} \qquad |x| \le \frac{1}{A}$$

$$y = \pm\frac{1 + \log A|x|}{1 + \log A} \qquad \text{for} \qquad \frac{1}{A} \le |x| \le 1$$

Here the $+$ sign applies when $x$ is positive and the $-$ sign when $x$ is negative. Also $x = v_i/V$ and $y = v_o/V$ where $v_i$ and $v_o$ are the input and output voltages and the range of allowable voltage is $-V$ to $+V$. The parameter $A$ determines the degree of compression.

(*a*) A commonly used value is $A = 87.6$. For this value make a plot of $y$ versus $x$ from $x = -1$ to $x = +1$.

(*b*) If $\pm V = \pm 10$ volts and 256 quantization levels are employed, what is the voltage interval between levels when there is no compression? For $A = 87.6$ what is the minimum and the maximum effective separation between levels?

**5.12-3.** An analog signal $m(t)$ has a probability density function

$$f(x) = 1 - |x| \qquad -1 \le x \le +1$$

This signal is quantized by a quantizer that has eight quantization levels.

(*a*) Determine the voltages at which the quantization levels should be located so that there shall be equal probabilities that $m(t)$ is between any two adjacent levels or between the extreme levels and the $+1$ and $-1$ extreme voltages.

(*b*) If a quantizer is used with equal spacings between quantization levels, make an input–output plot of the compander that must precede the quantizer to satisfy the requirement of part (*a*).

**5.12-4.** An analog signal $m(t)$ has a probability density function

$$f(x) = 1 - |x| \qquad \text{for} \qquad -1 \le x \le -1$$

The signal is applied to a quantizer with two quantization levels at $\pm 0.5$ volts. Calculate the mean square quantization error and compare with the result that would be given if Eq. (5.12-2) were applied.

Downloaded From : www.EasyEngineering.net

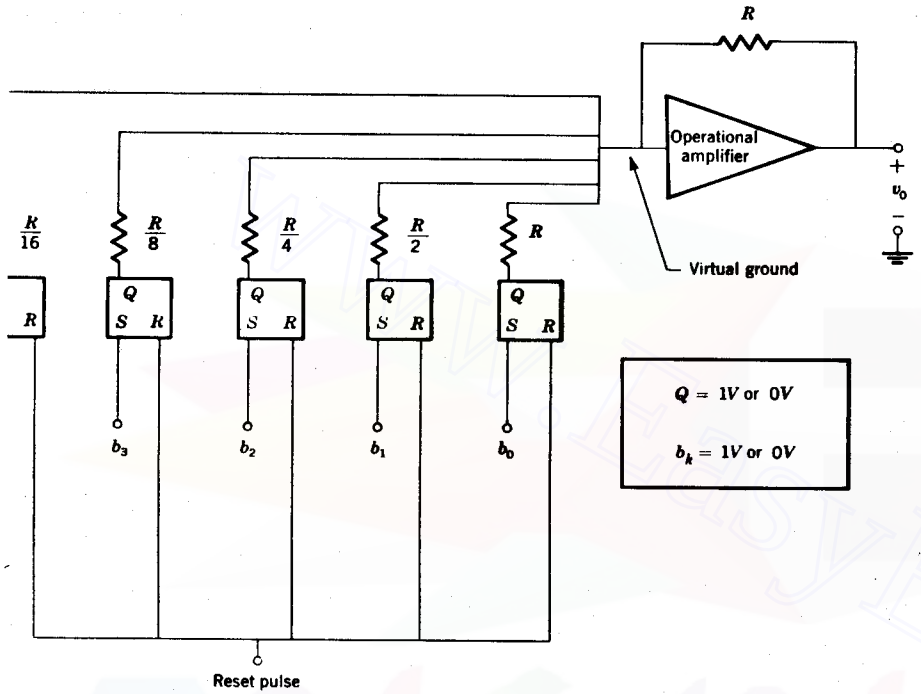



**Figure P5.11-1**

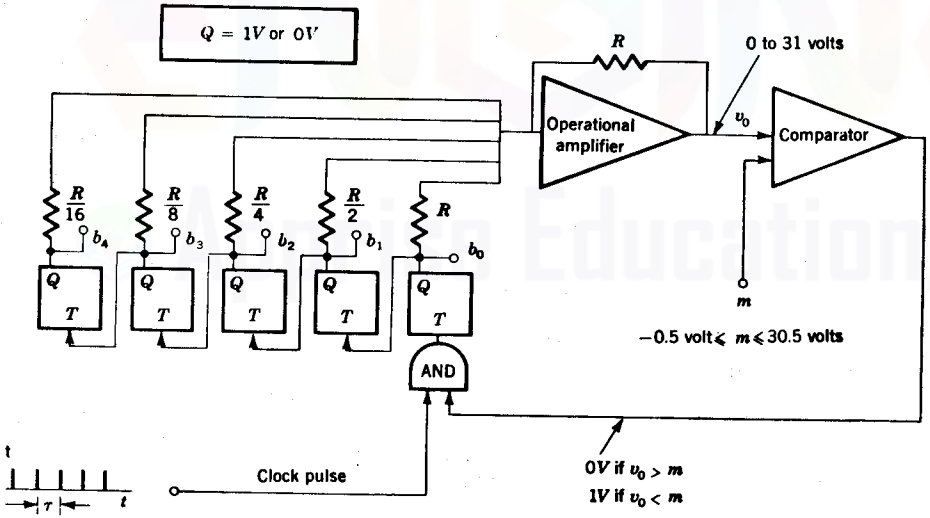



**Figure P5.11-2**

**5.12-5.** Verify that the procedure described in the text and represented in Fig. 5.12-2 for setting the content of the ROM yields $2^8 = 256$ code words for $2^{12} = 4096$ addresses. What is the ratio between the smallest quantization $\Delta$ and the largest quantization level?

**5.13-1.** (*a*) An NRZ waveform as in Fig. 5.13-3 consists of alternating 1's and 0's. The waveform swings between $+V$ and $-V$ and the bit duration is $T_b$. The waveform is passed through a $RC$ network as in Fig. 5.13-3*b* whose time constant is $RC = T_b$. Calculate all the voltage levels of the output waveform for the circumstances that the NRZ waveform has persisted for a long time.

(*b*) For the circumstances of part (*a*) calculate the area under the waveform during the bit time $T_b$. Now consider that a sequence of 10 successive 1's appears. Calculate the area under the waveform during the 10th 1 bit and compare the area calculated for the case when 1's and 0's alternate.

**5.13-2.** Draw an AMI waveform corresponding to a binary bit stream 100110010101. The polarity of the first pulse in the waveform may be set arbitrarily.

**5.13-3.** Two of the $T1$ inputs to the $M12$ multiplexer of Fig. 5.13-5 are generated in systems whose clocks have frequencies which are different by 50 parts/million. In how long a time will the faster clock have generated one more time slot than the slower clock. In this time interval, how many frames will have been generated?

**5.14-1.** Consider the delta PCM system shown in Fig. P5.14-1.

(*a*) Explain its operation.

(*b*) Sketch the receiver.

(*c*) If $m(t) = 0.05 \sin 2\pi t$, find $\tilde{m}(t)$ and $\Delta(t)$ graphically.

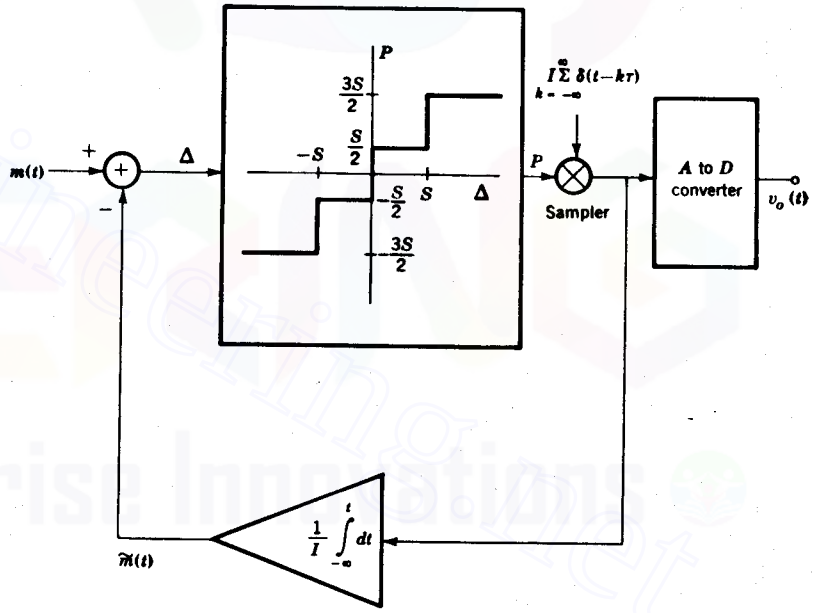



**Figure P5.14-1**

**5.15-1.** (*a*) Write the nonlinear difference equation of the delta modulator shown in Fig. 5.15-1. Let $m(t) = 18 \times 10^{-3} \sin 2\pi t$ and let the samples occur every 0.05 sec starting at $t = 0.01$ sec. The step size is 5 mV.

(*b*) Write and run a computer program to solve for $\Delta(t)$.

(*c*) Repeat (*b*) if the samples occur every 0.1 sec.

(*d*) Compare the results of (*b*) and (*c*).

**5.15-2.** The input to a DM is $m(t) = 0.01t$. The DM operates at a sampling frequency of 20 Hz and has a step size of 2 mV. Sketch the delta modulator output, $\Delta(t)$ and $\tilde{m}(t)$.

**5.15-3.** The input to a DM is $m(t) = kt$. Prove, by graphically determining $\tilde{m}(t)$, that slope overload occurs when $k$ exceeds a specified value. What is this value in terms of the step size $S$ and the sampling frequency $f_s$?

**5.15-4.** If the step size is $S$, the sampling frequency is $f_s^{(\Delta)}$, and $m(t) = M \sin \omega t$, explain what happens to $\tilde{m}(t)$ if $2M < S$. This is called *step-size limiting*.

**5.15-5.** The signal $m(t) = M \sin \omega_0 t$ is to be encoded by using a delta modulator. If the step size $S$ and sampling frequency $f_s^{(\Delta)}$ are selected so as to ensure that neither slope overloads [Eq. (5.15-1)] nor step-size limiting (Prob. 5.15-4) occurs, show that $f_s^{(\Delta)} > 3f_0$.

**5.16-1.** The adaptive delta-modulation system described in Sec. 5.16 has an input which is $m(t) = 0$ until time $t = 0$ and thereafter $m(t) = 1250 \sin 2\pi t$. An inactive edge of the clock occurs at $t = 0$ and the clock period is 0.05 sec. On a single set of coordinate axes draw the clock waveform, the input $m(t)$ and the approximation $\hat{m}(t)$. Extend the plot through a full cycle of $m(t)$.

**5.16-2.** An adaptive delta modulator is shown in Fig. P5.16-2. The gain $K$ is variable and is adjusted using the following logic. If $p_0(t)$ alternates between $+1$ and $-1$, $K = 1$; if a sequence of $N$ positive or $N$ negative pulses occurs, $K$ increases by $N$; if after a sequence of $N$ pulses the polarity changes, $|K|$ decreases by 2. Thus, consider the sequence 1, 1, 1. Then $K = 1, 2, 3, 1, 2$.

Find $\tilde{m}(t)$ if $m(t) = \sin 2\pi t$. Perform the analysis graphically. Consider a sampling time of 0.05 sec, and at $t = 0.01$ sec a sample occurs. At this time the step size $S = 1$ volt when $K = 1$.

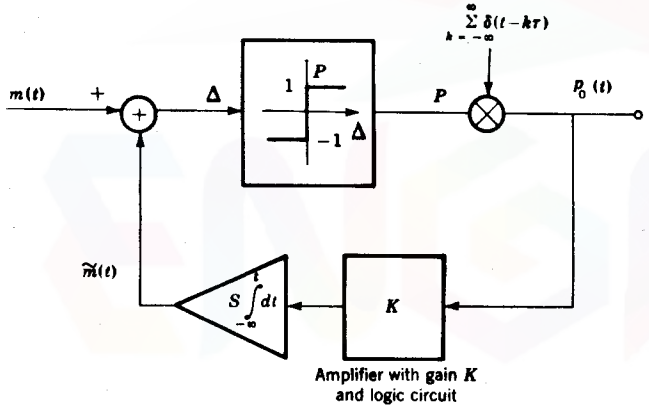



**Figure P5.16-2**





CHAPTER

# SIX

# DIGITAL MODULATION TECHNIQUES

## 6.1 INTRODUCTION

When it becomes necessary, for the purpose of transmission, to superimpose a binary waveform on a carrier, amplitude modulation (AM), phase modulation (PM), or frequency modulation (FM) may be used. Combination AM-PM systems such as quadrature amplitude modulation (QAM) are also commonly employed.

The selection of the particular modulation method used is determined by the application intended as well as by the channel characteristics such as available bandwidth and the susceptibility of the channel to fading. When radio communication is intended we must take account of antenna characteristics. As we have already noted (Sec. 3.1) an antenna is a narrow-band device whose operating frequency is related to its physical dimensions. When data transmission using a telephone channel is intended, we need to take account of the fact that often the channel may not transmit dc and low frequencies because of transformers which may be included in the transmission path. A fading channel is one in which the received signal amplitude varies with time because of variabilities in the transmission medium. When we must contend with such channels it is useful to use FM which is relatively insensitive to amplitude fluctuations.

In each of these situations we need a modulator at the transmitter and, at the receiver, a demodulator to recover the baseband signal. Such a modulator-demodulator combination is called a *MODEM*. In this chapter we present a description of many of the available modulation/demodulation techniques and compare them on the basis of their spectral occupancy. In Chap. 11 we shall complete our comparison and determine in each case the probability of error for each system as a function of signal-to-noise ratio and bandwidth available for transmission.

## 6.2 BINARY PHASE-SHIFT KEYING

In binary phase-shift keying (BPSK) the transmitted signal is a sinusoid of fixed amplitude. It has one fixed phase when the data is at one level and when the data is at the other level the *phase* is different by 180°. If the sinusoid is of amplitude $A$ it has a power $P_s = \frac{1}{2}A^2$ so that $A = \sqrt{2P_s}$. Thus the transmitted signal is either

$$v_{\text{BPSK}}(t) = \sqrt{2P_s}\cos(\omega_0 t) \tag{6.2-1}$$

or

$$v_{\text{BPSK}}(t) = \sqrt{2P_s}\cos(\omega_0 t + \pi) \tag{6.2-2a}$$

$$= -\sqrt{2P_s}\cos(\omega_0 t) \tag{6.2-2b}$$

In BPSK the data $b(t)$ is a stream of binary digits with voltage levels which, as a matter of convenience, we take to be at $+1V$ and $-1V$. When $b(t) = 1V$ we say it is at logic level 1 and when $b(t) = -1V$ we say it is at logic level 0. Hence $v_{\text{BPSK}}(t)$ can be written, with no loss of generality, as

$$v_{\text{BPSK}}(t) = b(t)\sqrt{2P_s}\cos\omega_0 t \tag{6.2-3}$$

In practice, a BPSK signal is generated by applying the waveform $\cos\omega_0 t$, as a carrier, to a *balanced* modulator and applying the baseband signal $b(t)$ as the modulating waveform. In this sense BPSK can be thought of as an AM signal.

### Reception of BPSK

The received signal has the form

$$v_{\text{BPSK}}(t) = b(t)\sqrt{2P_s}\cos(\omega_0 t + \theta) = b(t)\sqrt{2P_s}\cos\omega_0(t + \theta/\omega_0) \tag{6.2-4}$$

Here $\theta$ is a nominally fixed phase shift corresponding to the time delay $\theta/\omega_0$ which depends on the length of the path from transmitter to receiver and the phase shift produced by the amplifiers in the "front-end" of the receiver preceding the demodulator. The original data $b(t)$ is recovered in the demodulator. The demodulation technique usually employed is called synchronous demodulation and requires that there be available at the demodulator the waveform $\cos(\omega_0 t + \theta)$. A scheme for generating the carrier at the demodulator and for recovering the baseband signal is shown in Fig. 6.2-1.

The received signal is squared to generate the signal

$$\cos^2(\omega_0 t + \theta) = \tfrac{1}{2} + \tfrac{1}{2}\cos 2(\omega_0 t + \theta) \tag{6.2-5}$$

The dc component is removed by the bandpass filter whose passband is centered around $2f_0$ and we then have the signal whose waveform is that of $\cos 2(\omega_0 t + \theta)$. A frequency divider (composed of a flip-flop and narrow-band filter tuned to $f_0$) is used to regenerate the waveform $\cos(\omega_0 t + \theta)$. Only the waveforms of the signals at the outputs of the squarer, filter and divider are relevant to our discussion and not their amplitudes. Accordingly in Fig. 6.2-1 we have arbitrarily taken each amplitude to be unity. In practice, the amplitudes will

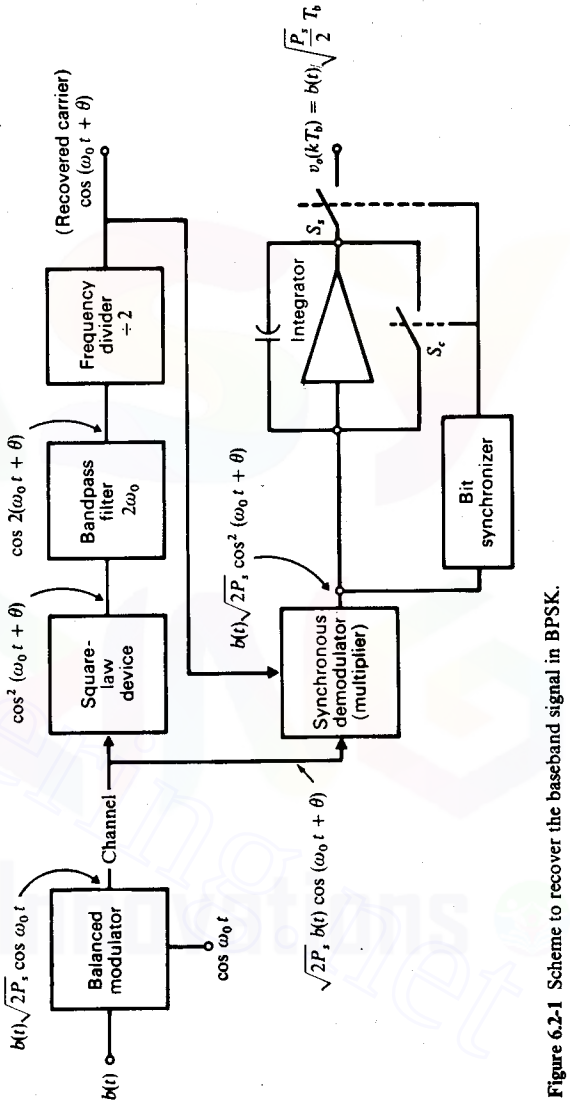



**Figure 6.2-1** Scheme to recover the baseband signal in BPSK.

be determined by features of these devices which are of no present concern. In any event, the carrier having been recovered, it is multiplied with the received signal to generate

$$b(t)\sqrt{2P_s}\cos^2(\omega_0 t+\theta) = b(t)\sqrt{2P_s}\left[\tfrac{1}{2}+\tfrac{1}{2}\cos 2(\omega_0 t+\theta)\right] \qquad (6.2\text{-}6)$$

which is then applied to an integrator as shown in Fig. 6.2-1.

We have included in the system a *bit synchronizer*. This device, whose operation is described in Sec. 10.15, is able to recognize precisely the moment which corresponds to the end of the time interval allocated to one bit and the beginning of the next. At that moment, it closes switch $S_c$ very briefly to discharge (*dump*) the integrator capacitor and leaves the switch $S_c$ open during the entire course of the ensuing bit interval, closing switch $S_c$ again very briefly at the end of the next bit time, etc. (This circuit is called an "integrate-and-dump" circuit.) The output signal of interest to us is the integrator output at the end of a bit interval but immediately before the closing of switch $S_c$. This output signal is made available by switch $S_s$ which samples the output voltage just prior to dumping the capacitor. Let us assume for simplicity that the bit interval $T_b$ is equal to the duration of an integral number $n$ of cycles of the carrier of frequency $f_0$, that is, $n \cdot 2\pi = \omega_0 T_b$. In this case the output voltage $v_o(kT_b)$ at the end of a bit interval extending from time $(k-1)T_b$ to $kT_b$ is, using Eq. (6.2-6)

$$v_o(kT_b) = b(kT_b)\sqrt{2P_s}\int_{(k-1)T_b}^{kT_b}\tfrac{1}{2}dt + b(kT_b)\sqrt{2P_s}\int_{(k-1)T_b}^{kT_b}\tfrac{1}{2}\cos 2(\omega_0 t+\theta)dt \qquad (6.2\text{-}7a)$$

$$= b(kT_b)\sqrt{\frac{P_s}{2}}\,T_b \qquad (6.2\text{-}7b)$$

since the integral of a sinusoid over a whole number of cycles has the value zero. Thus we see that our system reproduces at the demodulator output the transmitted bit stream $b(t)$. The operation of the bit synchronizer allows us to sense each bit independently of every other bit. The brief closing of both switches, after each bit has been determined, wipes clean all influence of a preceding bit and allows the receiver to deal exclusively with the present bit.

Our discussion has been rather naive since it has ignored the effects of thermal noise, frequency jitter in the carrier and random fluctuations in propagation delay. When these perturbing influences need to be taken into account a phase-locked synchronization system is called for as discussed in Sec. 10.16.

### Spectrum of BPSK

The waveform $b(t)$ is a NRZ (non-return-to-zero) binary waveform whose power spectral density is given in Eq. (2.25-4) for a waveform which makes excursions between $+\sqrt{P_s}$ and $-\sqrt{P_s}$. We have

$$G_b(f) = P_s T_b\left(\frac{\sin \pi f T_b}{\pi f T_b}\right)^2 \qquad (6.2\text{-}8)$$





The BPSK waveform is the NRZ waveform multiplied by $\sqrt{2}\cos\omega_0 t$. Thus following the analysis of Sec. 3.2 we find that the power spectral density of the BPSK signal is

$$G_{\text{BPSK}}(f) = \frac{P_s T_b}{2}\left\{\left[\frac{\sin \pi(f-f_0)T_b}{\pi(f-f_0)T_b}\right]^2 + \left[\frac{\sin \pi(f+f_0)T_b}{\pi(f+f_0)T_b}\right]^2\right\} \qquad (6.2\text{-}9)$$

Equations (6.2-8) and (6.2-9) are plotted in Fig. 6.2-2.

Note that, in principle at least, the spectrum of $G_b(f)$ extends over all frequencies and correspondingly so does $G_{\text{BPSK}}(f)$. Suppose then that we tried to multiplex signals using BPSK, using different carrier frequencies for different baseband signals. There would inevitably be overlap in the spectra of the various signals and correspondingly a receiver tuned to one carrier would also receive, albeit at a lower level, a signal in a different channel. This overlapping of spectra causes *interchannel* interference.

Since efficient spectrum utilization is extremely important in order to maximize the number of simultaneous users in a multi-user communication system, the FCC and CCITT require that the side-lobes produced in BPSK be reduced below certain specified levels. To accomplish this we employ a filter to restrict the bandwidth allowed to the NRZ baseband signal. For example, before modulation we might pass the bit stream $b(t)$ through a low-pass filter which suppresses (but does not completely eliminate) all the spectrum except the principal lobe. Since 90 percent of the power of the waveform is associated with this lobe the suggestion is not unreasonable. There is, however, the difficulty that such spectrum suppression distorts the signal and as a result, as we shall see, there is a partial overlap of a bit (symbol) and its adjacent bits in a single channel. This overlap is called *intersymbol* interference (ISI). Intersymbol interference can be somewhat alleviated by the use of *equalizers* at the receiver. Equalizers are filter-type structures used to undo the adverse effects of filters introduced, intentionally or unavoidably, at other places in a communications channel.

### Geometrical Representation of BPSK Signals

Referring to Sec. 1.9 we see that a BPSK signal can be represented, in terms of one orthonormal signal $u_1(t) = \sqrt{(2/T_b)}\cos\omega_0 t$ as [see Eq. (6.2-1)]

$$v_{\text{BPSK}}(t) = \left[\sqrt{P_s T_b}\,b(t)\right]\sqrt{\frac{2}{T_b}}\cos\omega_0 t = \left[\sqrt{P_s T_b}\,b(t)\right]u_1(t) \qquad (6.2\text{-}10)$$

The binary PSK signal can then be drawn as shown in Fig. 6.2-3. Note that the distance $d$ between signals is

$$d = 2\sqrt{P_s T_b} = 2\sqrt{E_b} \qquad (6.2\text{-}11)$$

where $E_b = P_s T_b$ is the energy contained in a bit duration. We show in Sec. 11.13

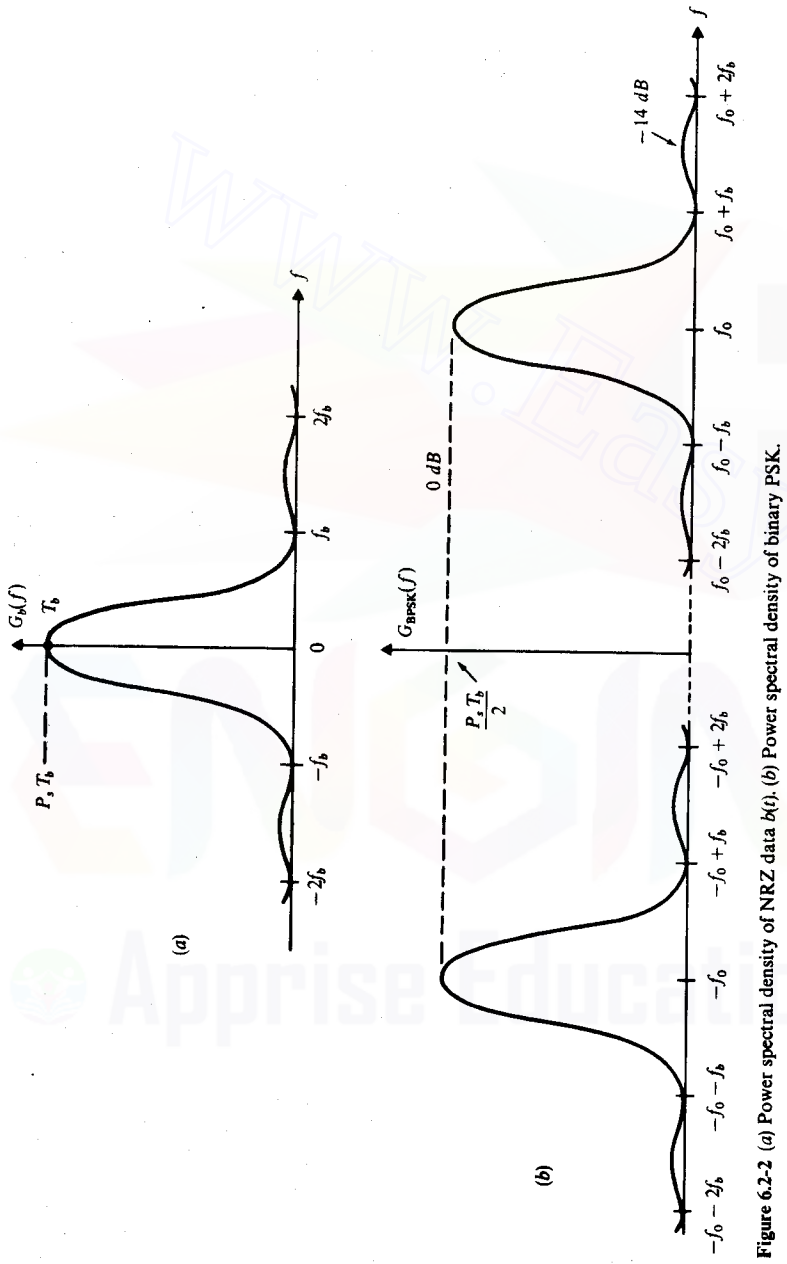



**Figure 6.2-2** (*a*) Power spectral density of NRZ data $b(t)$. (*b*) Power spectral density of binary PSK.



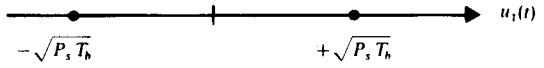



**Figure 6.2-3** Geometrical representation of BPSK signals.

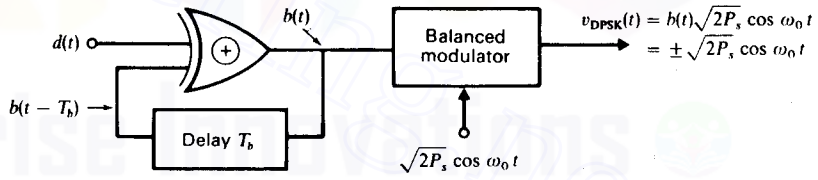
that the distance $d$ is inversely proportional to the probability that we make an error when, in the presence of noise, we try to determine which of the levels of $b(t)$ is being received.

## 6.3 DIFFERENTIAL PHASE-SHIFT KEYING

We observed in Fig. 6.2-1 that, in BPSK, to regenerate the carrier we start by squaring $b(t)\sqrt{2P_s}\cos\omega_0 t$. Accordingly, if the received signal were instead $-b(t)\sqrt{2P_s}\cos\omega_0 t$, the recovered carrier would remain as before. Therefore we shall not be able to determine whether the received baseband signal is the transmitted signal $b(t)$ or its negative $-b(t)$.

Differential phase-shift keying (DPSK) and differential encoded PSK (DEPSK) which is discussed in Sec. 6.4 are modifications of BPSK which have the merit that they eliminate the ambiguity about whether the demodulated data is or is not inverted. In addition DPSK avoids the need to provide the synchronous carrier required at the demodulator for detecting a BPSK signal.

A means for generating a DPSK signal is shown in Fig. 6.3-1. The data stream to be transmitted, $d(t)$, is applied to one input of an exclusive-OR logic gate. To the other gate input is applied the output of the exclusive or gate $b(t)$ delayed by the time $T_b$ allocated to one bit. This second input is then $b(t - T_b)$. In Fig. 6.3-2 we have drawn logic waveforms to illustrate the response $b(t)$ to an input $d(t)$. The upper level of the waveforms corresponds to logic 1, the lower level to logic 0. The truth table for the exclusive-OR gate is given in Fig. 6.3-1




| $d(t)$ | | $b(t - T_b)$ | | $b(t)$ | |
|---|---|---|---|---|---|
| logic level | voltage | logic level | voltage | logic level | voltage |
| 0 | −1 | 0 | −1 | 0 | −1 |
| 0 | −1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | −1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 0 | −1 |

**Figure 6.3-1** Means of generating a DPSK signal.

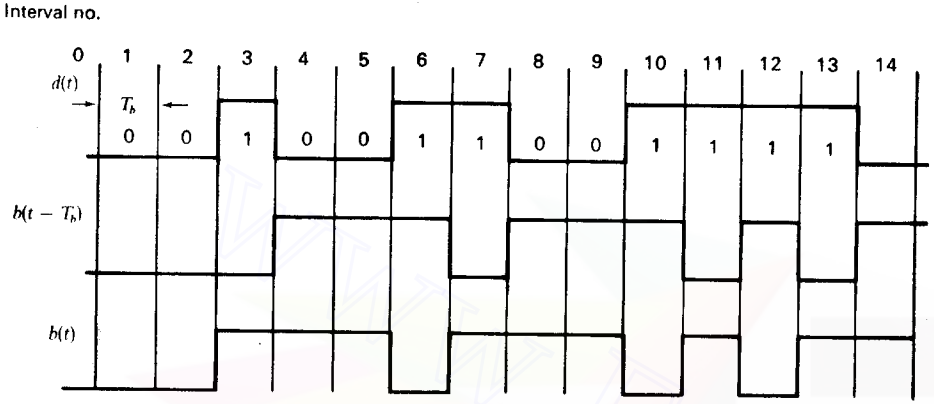



**Figure 6.3-2** Logic waveforms to illustrate the response $b(t)$ to an input $d(t)$.

and with this table we can easily verify that the waveforms for $d(t)$, $b(t - T_b)$, and $b(t)$ are consistent with one another. We observe that, as required, $b(t - T_b)$ is indeed $b(t)$ delayed by one bit time and that in any bit interval the bit $b(t)$ is given by $b(t) = d(t) \oplus b(t - T_b)$. In the ensuing discussion we shall use the symbolism $d(k)$ and $b(k)$ to represent the logic levels of $d(t)$ and $b(t)$ during the $k$th interval.

Because of the feedback involved in the system of Fig. 6.3-2 there is a difficulty in determining the logic levels in the interval in which we start to draw the waveforms (interval 1 in Fig. 6.3-2). We cannot determine $b(t)$ in this first interval of our waveform unless we know $b(k = 0)$. But we cannot determine $b(0)$ unless we know both $d(0)$ and $b(-1)$, etc. Thus, to justify any set of logic levels in an initial bit interval we need to know the logic levels in the preceding interval. But such a determination requires information about the interval two bit times earlier and so on. In the waveforms of Fig. 6.3-2 we have circumvented the problem by *arbitrarily assuming* that in the first interval $b(0) = 0$. It is shown below that in the demodulator, the data will be correctly determined regardless of our assumption concerning $b(0)$.

We now observe that the response of $b(t)$ to $d(t)$ is that $b(t)$ *changes* level at the beginning of each interval in which $d(t) = 1$ and $b(t)$ does not change level when $d(t) = 0$. Thus during interval 3, $d(3) = 1$, and correspondingly $b(3)$ *changes* at the beginning at that interval. During intervals 6 and 7, $d(6) = d(7) = 1$ and there are *changes* in $b(t)$ at the beginnings of both intervals. During bits 10, 11, 12, and 13 $d(t) = 1$ and there are *changes* in $b(t)$ at the beginnings of each of these intervals. This behavior is to be anticipated from the truth table of the exclusive-OR gate. For we note that when $d(t) = 0$, $b(t) = b(t - T_b)$ so that, whatever the initial value of $b(t - T_b)$, it reproduces itself. On the other hand when $d(t) = 1$ then $b(t) = \overline{b(t - T_b)}$. Thus, in each successive bit interval $b(t)$ changes from its value in the previous interval. Note that in some intervals where $d(t) = 0$ we have $b(t) = 0$ and in other intervals when $d(t) = 0$ we have $b(t) = 1$. Similarly, when $d(t) = 1$ sometimes $b(t) = 1$ and sometimes $b(t) = 0$. Thus there is no correspondence between the levels of $d(t)$ and $b(t)$, and the only invariant feature of the system is that a *change* (sometimes up and sometimes down) in $b(t)$ occurs whenever $d(t) = 1$, and that no change in $b(t)$ will occur whenever $d(t) = 0$.

Finally, we note that the waveforms of Fig. 6.3-2 are drawn on the assumption that, in interval 1, $b(0) = 0$. As is easily verified, if not intuitively apparent, if we had assumed $b(0) = 1$, the invariant feature by which we have characterized the system would continue to apply. Since $b(0)$ must be either $b(0) = 0$ or $b(0) = 1$, there being no other possibilities, our result is valid quite generally. If, however, we had started with $b(0) = 1$ the levels $b(1)$ and $b(0)$ would have been inverted.

As is seen in Fig. 6.3-1 $b(t)$ is applied to a balanced modulator to which is also applied the carrier $\sqrt{2P_s}\cos\omega_0 t$. The modulator output, which is the transmitted signal is

$$\begin{aligned} v_{\text{DPSK}}(t) &= b(t)\sqrt{2P_s}\cos\omega_0 t \\ &= \pm\sqrt{2P_s}\cos\omega_0 t \end{aligned} \tag{6.3-1}$$

Thus altogether when $d(t) = 0$ the phase of the carrier does *not* change at the beginning of the bit interval, while when $d(t) = 1$ there is a phase change of magnitude $\pi$.

A method of recovering the data bit stream from the DPSK signal is shown in Fig. 6.3-3. Here the received signal and the received signal delayed by the bit time $T_b$ are applied to a multiplier. The multiplier output is

$$\begin{aligned} &b(t)b(t - T_b)(2P_s)\cos(\omega_0 t + \theta)\cos[\omega_0(t - T_b) + \theta] \\ &\quad = b(t)b(t - T_b)P_s\left\{\cos\omega_0 T_b + \cos\left[2\omega_0\left(t - \frac{T_b}{2}\right) + 2\theta\right]\right\} \end{aligned} \tag{6.3-2}$$

and is applied to a bit synchronizer and integrator as shown in Fig. 6.2-1 for the BPSK demodulator. The first term on the right-hand side of Eq. (6.3-2) is, aside from a multiplicative constant, the waveform $b(t)b(t - T_b)$ which, as we shall see is precisely the signal we require. As noted previously in connection with BPSK, and so here, the output integrator will suppress the double frequency term. We should select $\omega_0 T_b$ so that $\omega_0 T_b = 2n\pi$ with $n$ an integer. For, in this case we shall have $\cos\omega_0 T_b = +1$ and the signal output will be as large as possible.

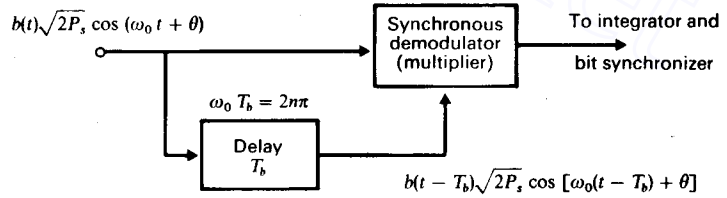



**Figure 6.3-3** Method of recovering data from the DPSK signal.

Further, with this selection, the bit duration encompasses an integral number of clock cycles and the integral of the double-frequency term is exactly zero.

The transmitted data bit $d(t)$ can readily be determined from the product $b(t)b(t - T_b)$. If $d(t) = 0$ then there was no phase change and $b(t) = b(t - T_b)$ both being $+1V$ or both being $-1V$. In this case $b(t)b(t - T_b) = 1$. If however, $d(t) = 1$ then there was a phase change and either $b(t) = 1V$ with $b(t - T_b) = -1V$ or vice versa. In either case $b(t)b(t - T_b) = -1$.

The differentially coherent system, DPSK, which we have been describing has a clear advantage over the coherent BPSK system in that the former avoids the need for complicated circuitry used to generate a local carrier at the receiver. To see the relative disadvantage of DPSK in comparison with PSK, consider that during some bit interval the received signal is so contaminated by noise that in a PSK system an error would be made in the determination of whether the transmitted bit was a 1 or a 0. In DPSK a bit determination is made on the basis of the signal received in two successive bit intervals. Hence noise in one bit interval may cause errors to two bit determinations. The error rate in DPSK is therefore greater than in PSK, and, as a matter of fact, there is a tendency for bit errors to occur in pairs. It is not inevitable however that errors occur in pairs. Single errors are still possible. For consider a case in which the received signals in $k$th and $(k + 1)$st bit intervals are both somewhat noisy but that the signals in the $(k - 1)$st and $(k + 2)$nd intervals are noise free. Assume further that the $k$th interval signal is *not* so noisy that an error results from the comparison with the $(k - 1)$st interval signal and assume a similar situation prevails in connection with the $(k + 1)$st and the $(k + 2)$nd interval signals. Then it may be that only a single error will be generated, that error being the result of the comparison of the $k$th and $(k + 1)$st interval signals both of which are noisy.

## 6.4 DIFFERENTIALLY-ENCODED PSK (DEPSK)

As is noted in Fig. 6.3-3 the DPSK demodulator requires a device which operates at the carrier frequency and provides a delay of $T_b$. Differentially-encoded PSK eliminates the need for such a piece of hardware. In this system, synchronous demodulation recovers the signal $b(t)$, and the decoding of $b(t)$ to generate $d(t)$ is done at baseband.

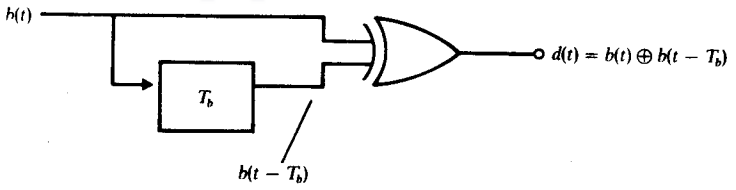



**Figure 6.4-1** Baseband decoder to obtain $d(t)$ from $b(t)$.

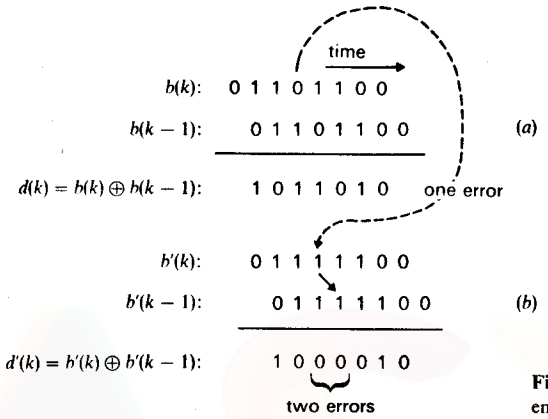



**Figure 6.4-2** Errors in differentially-encoded PSK occur in pairs.

The transmitter of the DEPSK system is identical to the transmitter of the DPSK system shown in Fig. 6.3-1. The signal $b(t)$ is recovered in exactly the manner shown in Fig. 6.2-1 for a BPSK system. The recovered signal is then applied directly to one input of an exclusive-OR logic gate and to the other input is applied $b(t - T_b)$ (see Fig. 6.4-1). The gate output will be at one or the other of its levels depending on whether $b(t) = b(t - T_b)$ or $b(t) = \overline{b(t - T_b)}$. In the first case $b(t)$ did not change level and therefore the transmitted bit is $d(t) = 0$. In the second case $d(t) = 1$.

We have seen that in DPSK there is a tendency for bit errors to occur in pairs but that single bit errors are possible. In DEPSK errors *always* occur in pairs. The reason for the difference is that in DPSK we do not make a hard decision, in each bit interval about the phase of the received signal. We simply allow the received signal in one interval to compare itself with the signal in an adjoining interval and, as we have seen, a single error is not precluded. In DEPSK, a firm definite hard decision is made in each interval about the value of $b(t)$. If we make a mistake, then errors must result from a comparison with the preceding and succeeding bit. This result is illustrated in Fig. 6.4-2. In Fig. 6.4-2*a* is shown the error-free signals $b(k)$, $b(k - 1)$ and $d(k) = b(k) \oplus b(k - 1)$. In Fig. 6.4-2*b* we have assumed that $b'(k)$ has a single error. Then $b'(k - 1)$ must also have a single error. We note that the reconstructed waveform $d'(k)$ now has two errors.

## 6.5 QUADRATURE PHASE-SHIFT KEYING (QPSK)

We have seen that when a data stream whose bit duration is $T_b$ is to be transmitted by BPSK the channel bandwidth must be nominally $2f_b$ where $f_b = 1/T_b$. Quadrature phase-shift keying, as we shall explain, allows bits to be transmitted using half the bandwidth. In describing the QPSK system we shall have occasion to use the type-$D$ flip-flop as a one bit storage device. We therefore digress, very briefly, to remind the reader of the essential characteristics of this flip-flop.

### Type-*D* flip-flop

The type-*D* flip-flop represented in Fig. 6.5-1*a* has a single data input terminal (*D*) to which a data stream $d(t)$ is applied. The operation of the flip-flop is such that at the "active" edge of the clock waveform the logic level at *D* is transferred to the output *Q*. Representative waveforms are shown in Fig. 6.5-1(*b*). We assume arbitrarily that the negative-going edge of the clock waveform is the active edge. At the active edge numbered 1 we find that $d(t) = 0$. Hence, after a short delay (the delay is not shown) from the time of occurrence of this edge we shall find that $Q = 0$. The delay results from the fact that some time is required for the input data to propagate through the flip-flop to the output *Q*. (On the basis of the waveform $d(t)$ shown, we have no basis on which to determine *Q* at an earlier time.) At active edge 2 it appears that the clock edge occurs precisely at the time when $d(t)$ is changing. If such were indeed the case, the response of the flip-flop would be ambiguous. As a matter of practice, the change in $d(t)$ will occur slightly after the active edge. Such is the case because, rather inevitably, the change in $d(t)$ is itself the response of some digital component (gate, flip-flop, etc.) to the very same clock waveform which is driving our type-*D* flip-flop. (The delays referred to are normally not indicated on a waveform diagram as in Fig. 6.5-1 because these delays are ordinarily very small in comparison with the bit time $T_b$.) In any event, the fact is that at active edge 2 we have $d(t) = 0$ and *Q* remains at $Q = 0$. The remainder of the waveforms are easily verified on the same basis. Observe that the *Q* waveform is the $d(t)$ waveform delayed by one bit interval $T_b$. The relevant point about the flip-flop in the matter of our present concern is the following: Once the flip-flop, in response to an active clock edge, has registered a data bit, it will hold that bit until updated by the occurrence of the next succeeding active edge.

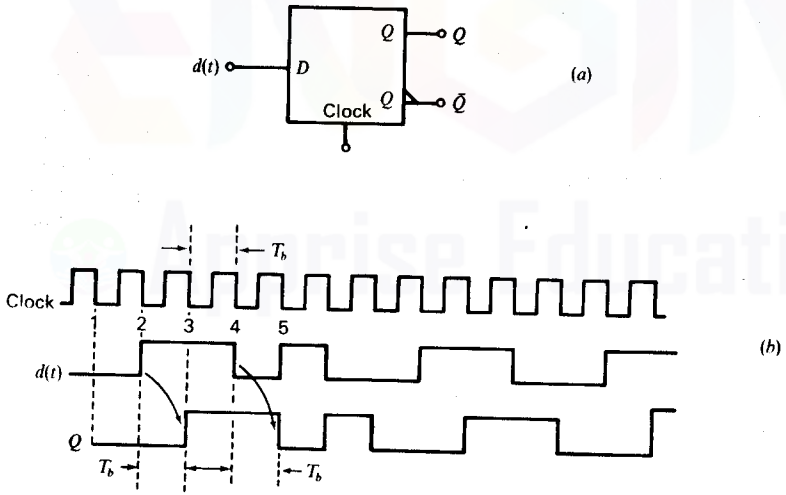



**Figure 6.5-1** (*a*) Type-*D* flip-flop symbol. (*b*) Waveforms showing flip-flop characteristics.

### QPSK Transmitter

The mechanism by which a bit stream $b(t)$ generates a QPSK signal for transmission is shown in Fig. 6.5-2 and relevant waveforms are shown in Fig. 6.5-3. In these waveforms we have arbitrarily assumed that in every case the active edge of the clock waveforms is the downward edge. The toggle flip-flop is driven by a clock waveform whose period is the bit time $T_b$. The toggle flip-flop generates an odd clock waveform and an even waveform. These clocks have periods $2T_b$. The active edge of one of the clocks and the active edge of the other are separated by the bit time $T_b$. The bit stream $b(t)$ is applied as the data input to both type-*D* flip-flops, one driven by the odd and one driven by the even clock waveform. The flip-flops register alternate bits in the stream $b(t)$ and hold each such registered bit for two bit intervals, that is for a time $2T_b$. In Fig. 6.5-3 we have numbered the bits in $b(t)$. Note that the bit stream $b_o(t)$ (which is the output of the flip-flop driven by the odd clock) registers bit 1 and holds that bit for time $2T_b$, then registers bit 3 for time $2T_b$, then bit 5 for $2T_b$, etc. The even bit stream $b_e(t)$ holds, for times $2T_b$ each, the alternate bits numbered 2, 4, 6, etc.

The bit stream $b_e(t)$ (which, as usual, we take to be $b_e(t) = \pm 1$ volt) is superimposed on a carrier $\sqrt{P_s}\cos\omega_0 t$ and the bit stream $b_o(t)$ (also $\pm 1$ volt) is

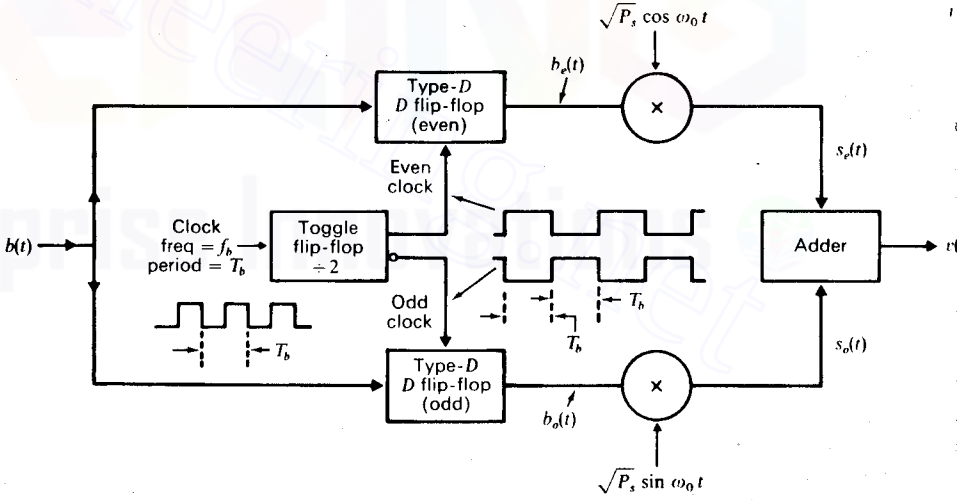



**Figure 6.5-2** An offset QPSK transmitter.

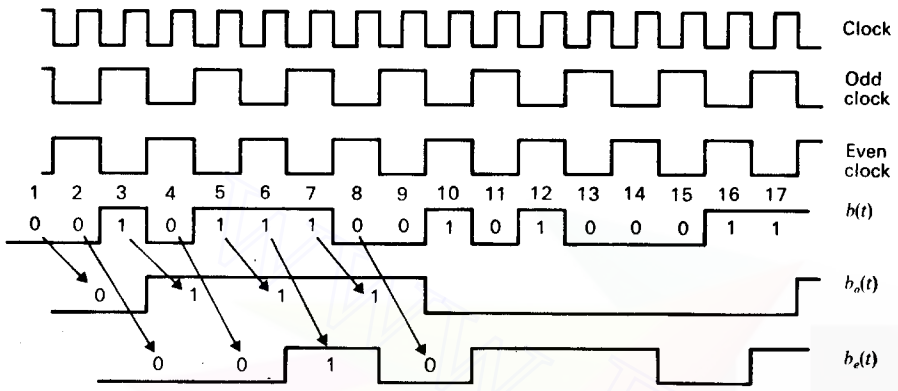



**Figure 6.5-3** Waveforms for the QPSK transmitter of Fig. 6.5-2.

superimposed on a carrier $\sqrt{P_s}\sin\omega_0 t$ by the use of two multipliers (i.e., balanced modulators) as shown, to generate two signals $s_e(t)$ and $s_o(t)$. These signals are then added to generate the transmitted output signal $v_m(t)$ which is

$$v_m(t) = \sqrt{P_s}\, b_o(t)\sin\omega_0 t + \sqrt{P_s}\, b_e(t)\cos\omega_0 t \qquad (6.5\text{-}1)$$

As may be verified, the total normalized power of $v_m(t)$ is $P_s$.

As we have noted in BPSK, a bit stream with bit time $T_b$ multiplies a carrier, the generated signal has a nominal bandwidth $2 \times 1/T_b$. In the waveforms $b_o(t)$ and $b_e(t)$ the bit times are each $1/2T_b$, hence both $s_e(t)$ and $s_o(t)$ have nominal bandwidths which are half the bandwidth in BPSK. Both $s_e(t)$ and $s_o(t)$ occupy the same spectral range but they are nonetheless individually identifiable because of the phase quadrature of their carriers.

When $b_o = 1$ the signal $s_o(t) = \sqrt{P_s}\sin\omega_0 t$, and $s_o(t) = -\sqrt{P_s}\sin\omega_0 t$ when $b_o = -1$. Correspondingly, for $b_e(t) = \pm 1$, $s_e(t) = \pm\sqrt{P_s}\cos\omega_0 t$. These four signals have been represented as phasors in Fig. 6.5-4. They are in mutual phase

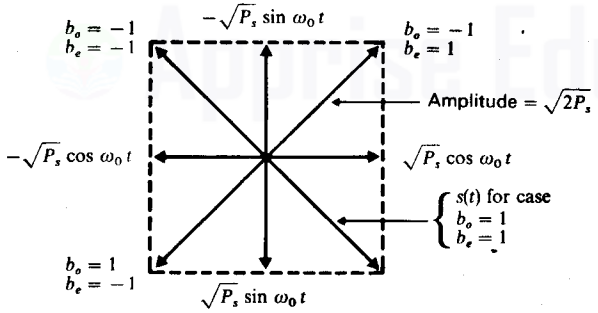



**Figure 6.5-4** Phasor diagram for sinusoids in Fig. 6.5-2.

quadrature. Also drawn are the phasors representing the four possible output signals $v_m(t) = s_o(t) + s_e(t)$. These four possible output signals have equal amplitude $\sqrt{2P_s}$ and are in phase quadrature; they have been identified by their corresponding values of $b_o$ and $b_e$. At the end of each bit interval (i.e., after each time $T_b$) either $b_o$ or $b_e$ can change, but both cannot change at the same time. Consequently, the QPSK system shown in Fig. 6.5-2 is called *offset* or *staggered* QPSK and abbreviated OQPSK. After each time $T_b$, the transmitted signal, if it changes, changes phase by 90° rather than by 180° as in BPSK.

## Non-offset QPSK

Suppose that in Fig. 6.5-2 we introduce an additional flip-flop before either the odd or even flip-flop. Let this added flip-flop be driven by the clock which runs at the rate $f_b$. Then one or the other bit streams, odd or even, will be delayed by one bit interval. As a result, we shall find that two bits which occur in time sequence (i.e., serially) in the input bit stream $b(t)$ will appear at the same time (i.e., in parallel) at the outputs of the odd and even flip-flops. In this case $b_e(t)$ and $b_o(t)$ can change at the same time, after each time $2T_b$, and there can be a phase change of 180° in the output signal. There is no difference, in principle, between a staggered and non-staggered system.

In practice, there is often a significant difference between QPSK and OQPSK. At each transition time, $T_b$ for OQPSK and $2T_b$ for QPSK, one bit for OQPSK and perhaps two bits for QPSK change from 1V to −1V or −1V to 1V. Now the bits $b_e(t)$ and $b_o(t)$ can, of course, not change instantaneously and, in changing, must pass through zero and dwell in that neighborhood at least briefly. Hence there will be brief variations in the *amplitude* of the transmitted waveform. These variations will be more pronounced in QPSK than in OQPSK since in the first case both $b_e(t)$ and $b_o(t)$ may be zero simultaneously so that the signal amplitude may actually be reduced to zero temporarily. There is a second mechanism through which amplitude variations are caused at the transmitter. In QPSK as in BPSK a filter is used to suppress sidebands. It turns out that when waveforms which exhibit abrupt phase changes, are filtered, the effect of the filter, at the time of the abrupt phase changes, is to cause substantial changes again in the *amplitude* of the waveform. Here too, we expect larger changes in QPSK where phase changes of 180° are possible than in OQPSK where the maximum phase change is 90°.

The amplitude variations can cause difficulty in QPSK communication systems which employ repeaters, i.e., stations which receive and rebroadcast signals, such as earth satellites. For such stations generally employ output power stages which operate nonlinearly, the nonlinearity being deliberately introduced because such nonlinear stages can operate with improved efficiency. However, precisely because of their nonlinearity, when presented with amplitude variations, they generate spectral components outside the range of the main lobe, thereby undoing the effect of the band limiting filter and causing interchannel inter-


Downloaded From : www.EasyEngineering.net

ference. Further filtering to suppress the effect of amplitude variation has an effect on the phase of the signal and it is, of course, precisely the phase which conveys the signal message.

### Symbol Versus Bit Transmission

In BPSK we deal individually with each bit of duration $T_b$. In QPSK we lump two bits together to form what is termed a *symbol*. The symbol can have any one of four possible values corresponding to the two-bit sequences 00, 01, 10, and 11. We therefore arrange to make available for transmission four distinct signals. At the receiver each signal represents *one symbol* and, correspondingly, *two bits*. When bits are transmitted, as in BPSK, the signal changes occur at the bit rate. When symbols are transmitted the changes occur at the symbol rate which is one-half the bit rate. Thus the symbol time is $T_s = 2T_b$.

### The QPSK Receiver

A receiver for the QPSK signal is shown in Fig. 6.5-5. Synchronous detection is required and hence it is necessary to locally regenerate the carriers $\cos \omega_0 t$ and $\sin \omega_0 t$. The scheme for carrier regeneration is similar to that employed in BPSK. In that earlier case we squared the incoming signal, extracted a waveform

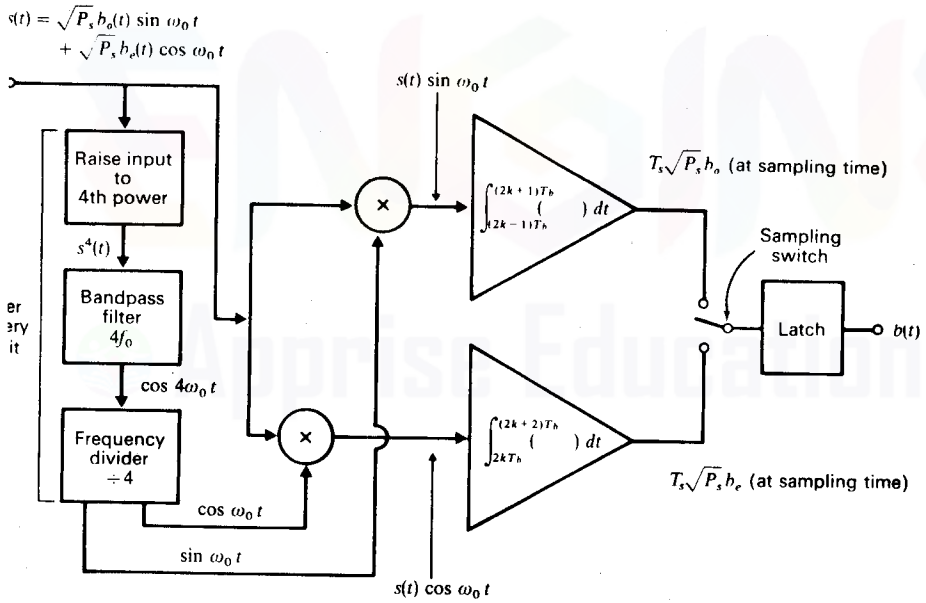



**Figure 6.5-5** A QPSK receiver.

at twice the carrier frequency by filtering, and recovered the carrier by frequency dividing by two. In the present case, it is required that the incoming signal be raised to the fourth power after which filtering recovers a waveform at four times the carrier frequency and finally frequency division by four regenerates the carrier. In the present case, also, we require both $\sin \omega_0 t$ and $\cos \omega_0 t$. It is left as a problem (see Prob. 6.5-3) to verify that the scheme indicated in Fig. 6.5-5 does indeed yield the required waveforms $\sin \omega_0 t$ and $\cos \omega_0 t$.

The incoming signal is also applied to two synchronous demodulators consisting, as usual, of a multiplier (balanced modulator) followed by an integrator. The integrator integrates over a two-bit interval of duration $T_s = 2T_b$ and then dumps its accumulation. As noted previously, ideally the interval $2T_b = T_s$ should encompass an integral number of carrier cycles. One demodulator uses the carrier $\cos \omega_0 t$ and the other the carrier $\sin \omega_0 t$. We recall that when sinusoids in phase quadrature are multiplied, and the product is integrated over an integral number of cycles, the result is zero. Hence the demodulators will selectively respond to the parts of the incoming signal involving respectively $b_e(t)$ or $b_o(t)$.

Of course, as usual, a bit synchronizer is required to establish the beginnings and ends of the bit intervals of each bit stream so that the times of integration can be established. The bit synchronizer is needed as well to operate the sampling switch. At the end of each integration time for each individual integrator, and just before the accumulation is dumped, the integrator output is sampled. Samples are taken alternately from one and the other integrator output at the end of each bit time $T_b$ and these samples are held in the latch for the bit time $T_b$. Each individual integrator output is sampled at intervals $2T_b$. The latch output is the recovered bit stream $b(t)$.

The voltages marked on Fig. 6.5-5 are intended to represent the waveforms of the signals only and not their amplitudes. Thus the actual value of the sample voltages at the integrator outputs depends on the amplitude of the local carrier, the gain, if any, in the modulators and the gain in the integrators. We have however indicated that the sample values depend on the normalized power $P_s$ of the received signal and on the duration $T_s$ of the symbol.

The mechanism used in Fig. 6.5-5 to regenerate the local carriers is a source of phase ambiguity of the type described in Sec. 6.4. That is, the carrier may be 180° out of phase with the carriers at the transmitter and as a result the demodulated signals may be complementary to the transmitted signal. This situation can be corrected, as before, by using differential encoding and decoding as in Figs. 6.4-1 and 6.4-2.

### Signal Space Representation

In Sec. 1.26 we investigated four quadrature signals. Equation (1.26-2), repeated here, is

$$v_m(t) = \sqrt{2P_s} \cos\left[\omega_0 t + (2m+1)\frac{\pi}{4}\right] \qquad m = 0, 1, 2, 3 \qquad (6.5\text{-}2)$$

These signals were then represented in terms of the two orthonormal signals $u_1(t) = \sqrt{(2/T)} \cos \omega_0 t$ and $u_2(t) = \sqrt{(2/T)} \sin \omega_0 t$. The result in Eq. (1.9-22), repeated here, is

$$v_m(t) = \left[\sqrt{P_s T} \cos (2m+1)\frac{\pi}{4}\right]\sqrt{\frac{2}{T}} \cos \omega_0 t - \left[\sqrt{P_s T} \sin (2m+1)\frac{\pi}{4}\right]\sqrt{\frac{2}{T}} \sin \omega_0 t \qquad (6.5\text{-}3)$$

The QPSK signal $v_m(t)$ in Eq. (6.5-1) can be put in the form of Eq. (6.5-3) by setting

$$b_e = \sqrt{2} \cos (2m+1)\frac{\pi}{4} \qquad (6.5\text{-}4a)$$

and

$$b_o = -\sqrt{2} \sin (2m+1)\frac{\pi}{4} \qquad (6.5\text{-}4b)$$

Thus

$$v_m(t) = \sqrt{E_b}\, b_e(t) u_1(t) - \sqrt{E_b}\, b_o(t) u_2(t) \qquad (6.5\text{-}5)$$

where $T = 2T_b = T_s$. Now Fig. 1.9-9 can be redrawn as shown in Fig. 6.5-6, to show the geometrical representation of QPSK. The points in signal space corresponding to each of the four possible transmitted signals is indicated by dots. From each such signal we can recover two bits rather than one. The distance of a signal point from the origin is $\sqrt{E_s}$ which is the square root of the signal energy associated with the symbol, that is $E_s = P_s T_s = P_s(2T_b)$. As we have noted earlier

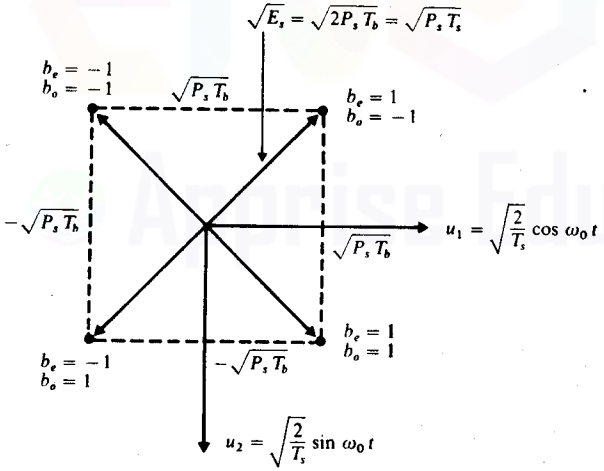



**Figure 6.5-6** The four QPSK signals drawn in signal space.

and will verify in Sec. 11.14, our ability to determine a bit without error is measured by the distance in signal space between points corresponding to the different values of the bit. We note in Fig. 6.5-6 that points which differ in a single bit are separated by the distance

$$d = 2\sqrt{P_s T_b} = 2\sqrt{E_b} \qquad (6.5\text{-}6)$$

where $E_b$ is the energy contained in a bit transmitted for a time $T_b$. This distance for QPSK is the same as for BPSK (see Eq. (6.2-11)). Hence, altogether, we have the important result that, in spite of the reduction by a factor of two in the bandwidth required by QPSK in comparison with BPSK, the noise immunities of the two systems are the same.

## 6.6 *M*-ARY PSK

In BPSK we transmit each bit individually. Depending on whether $b(t)$ is logic 0 or logic 1, we transmit one or another of a sinusoid for the bit time $T_b$, the sinusoids differing in phase by $2\pi/2 = 180°$. In QPSK we lump together two bits. Depending on which of the four two-bit words develops, we transmit one or another of four sinusoids of duration $2T_b$, the sinusoids differing in phase by amount $2\pi/4 = 90°$. The scheme can be extended. Let us lump together $N$ bits so that in this $N$-bit *symbol*, extending over the time $NT_b$, there are $2^N = M$ possible symbols. Now let us represent the symbols by sinusoids of duration $NT_b = T_s$ which differ from one another by the phase $2\pi/M$. Hardware to accomplish such $M$-ary communication is available.

Thus in $M$-ary PSK the waveforms used to identify the symbols are

$$v_m(t) = \sqrt{2P_s} \cos (\omega_0 t + \phi_m) \qquad (m = 0, 1, \ldots, M-1) \qquad (6.6\text{-}1)$$

with the symbol phase angle given by

$$\phi_m = (2m+1)\frac{\pi}{M} \qquad (6.6\text{-}2)$$

The waveforms of Eq. (6.6-1) are represented by the dots in Fig. 6.6-1 in a signal space in which the coordinate axes are the orthonormal waveforms $u_1(t) = \sqrt{(2/T_s)} \cos \omega_0 t$ and $u_2(t) = \sqrt{(2/T_s)} \sin \omega_0 t$. The distance of each dot from the origin is $\sqrt{E_s} = \sqrt{P_s T_s}$.

From Eq. (6.6-1) we have

$$v_m(t) = (\sqrt{2P_s} \cos \phi_m) \cos \omega_0 t - (\sqrt{2P_s} \sin \phi_m) \sin \omega_0 t \qquad (6.6\text{-}3)$$

Defining $p_e$ and $p_o$ by

$$p_e = \sqrt{2P_s} \cos \phi_m \qquad (6.6\text{-}4a)$$

$$p_o = \sqrt{2P_s} \sin \phi_m \qquad (6.6\text{-}4b)$$


Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

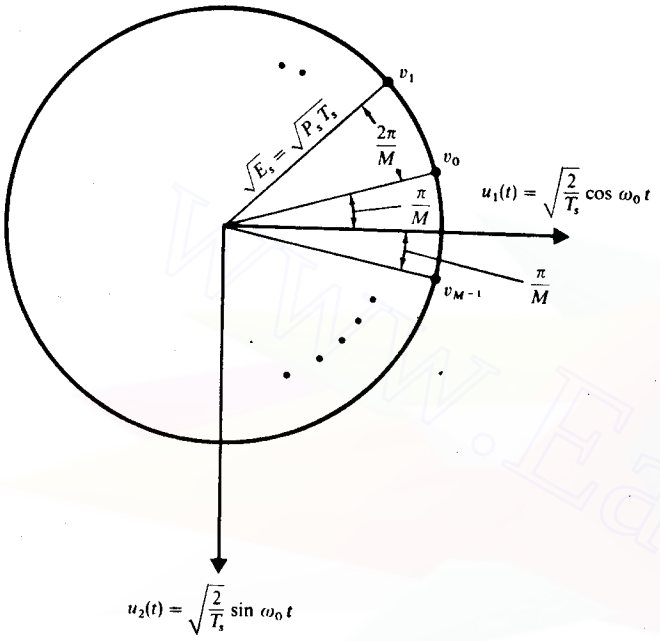



**Figure 6.6-1** Geometrical representation of $M$-ary PSK signals.

Eq. (6.6-3) becomes

$$v_m(t) = p_e \cos \omega_0 t - p_o \sin \omega_0 t \tag{6.6-5}$$

Both $p_e$ and $p_o$ can change every $T_s = NT_b$ and can assume any of $M$ possible values. The quantities $p_e$, $p_o$, and $\phi_m$ are random processes. The power spectral densities of $p_e$ and $p_o$ are given by Eq. (2.26-4) as

$$G_e(f) = \frac{\overline{|P_e(f)|^2}}{T_s} = 2P_s T_s \overline{\cos^2 \phi_m} \left(\frac{\sin \pi f T_s}{\pi f T_s}\right)^2 \tag{6.6-6}$$

and

$$G_o(f) = \frac{\overline{|P_o(f)|^2}}{T_s} = 2P_s T_s \overline{\sin^2 \phi_m} \left(\frac{\sin \pi f T_s}{\pi f T_s}\right)^2 \tag{6.6-7}$$

However, since $\phi_m$ is uniformly distributed

$$\overline{\cos^2 \phi_m} = \overline{\sin^2 \phi_m} = \tfrac{1}{2} \tag{6.6-8}$$

so that

$$G_e(f) = G_o(f) = P_s T_s \left(\frac{\sin \pi f T_s}{\pi f T_s}\right)^2 \tag{6.6-9}$$

As we have already noted, when signals with spectral density given by Eq. (6.6-9) are multiplied by a carrier, the resultant spectrum is centered at the carrier frequency and extends nominally over a bandwidth

$$B = \frac{2}{T_s} = 2f_s = 2\frac{f_b}{N} \tag{6.6-10}$$

We thus note that as we increase the number of bits $N$ per symbol the bandwidth becomes progressively smaller. On the other hand as we can see from Fig. 6.6-1 the distance between symbol signal points becomes smaller. We readily calculate (using the law of cosines) that this distance is

$$d = \sqrt{4E_s \sin^2 (\pi/M)} = \sqrt{4NE_b \sin^2 (\pi/2^N)} \tag{6.6-11}$$

where $E_s$ is the symbol energy $P_s \times (NT_b) = P_s T_s = NE_b$ and $E_b = P_s T_b$ is the energy associated with one bit. Thus as we increase $N$, i.e., as we increase the duration of the symbol, the bandwidth decreases, the distance $d$ decreases and, as we shall see, the probability of error becomes higher. Such is the case for all increases in $N$ except for the increase from $N = 1$ (BPSK) to $N = 2$ (QPSK).

## $M$-ary Transmitter and Receiver

The physical implementation of an $M$-ary PSK transmission system is moderately elaborate. Such hardware is only of incidental concern to us in this text so we shall describe the $M$-ary transmitter-receiver somewhat superficially.

As shown in Fig. 6.6-2, at the transmitter, the bit stream $b(t)$ is applied to a *serial-to-parallel converter*. This converter has facility for storing the $N$ bits of a symbol. The $N$ bits have been presented serially, that is, in time sequence, one after another. These $N$ bits, having been assembled, are then presented all at once on $N$ output lines of the converter, that is they are presented in *parallel*. The converter output remains unchanging for the duration $NT_b$ of a symbol during which time the converter is assembling a new group of $N$ bits. Each symbol time the converter output is updated.

The converter output is applied to a D/A converter. This D/A converter generates an output voltage which assumes one of $2^N = M$ different values in a one-to-one correspondence to the $M$ possible symbols applied to its input. That is,

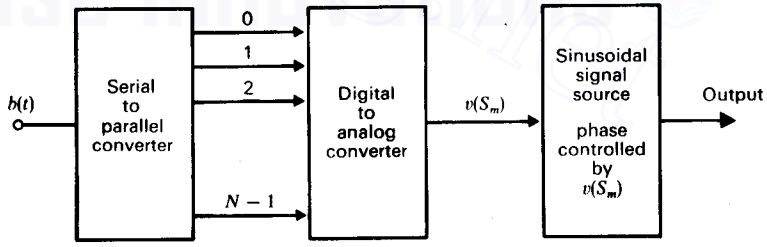



**Figure 6.6-2** $M$-ary PSK transmitter.


Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

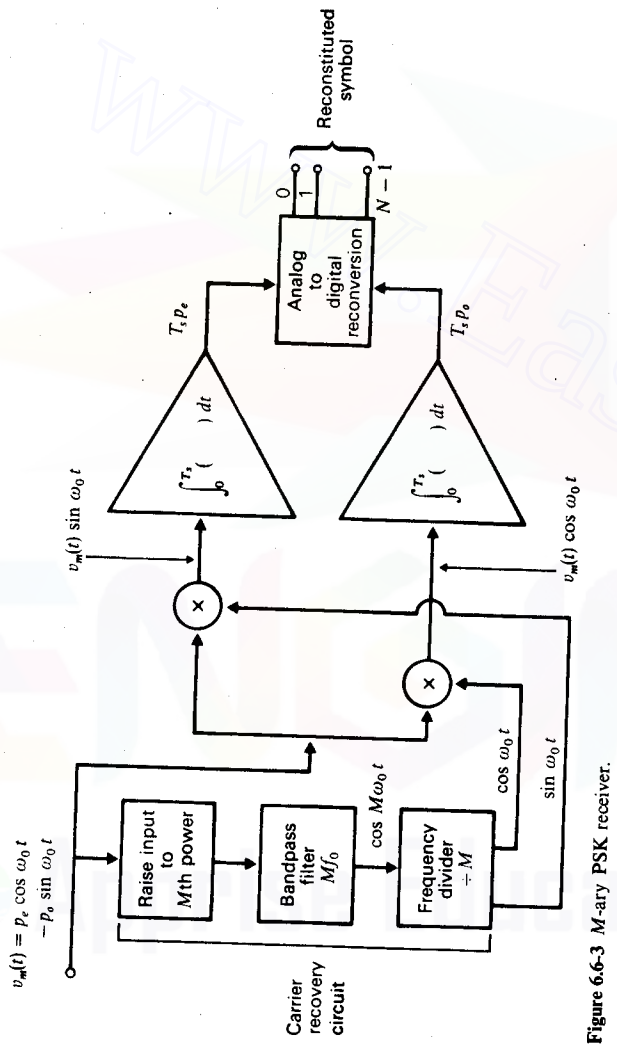



Figure 6.6-3 $M$-ary PSK receiver.



the D/A output is a voltage $v(S_m)$ which depends on the symbol $S_m$ ($m = 0, 1, \ldots, M-1$). Finally $v(S_m)$ is applied as a control input to a special type of constant-amplitude sinusoidal signal source whose phase $\phi_m$ is determined by $v(S_m)$. Altogether, then, the output is a fixed amplitude, sinusoidal waveform, whose phase has a one-to-one correspondence to the assembled $N$-bit symbol. The phase can change once per symbol time.

The receiver, shown in Fig. 6.6-3 is similar to the nonoffset QPSK receiver. The carrier recovery system requires, in the present case a device to raise the received signal to the $M$th power, filter to extract the $Mf_0$ component and then divide by $M$. As was the case in Fig. 6.5-5 the signals indicated in Fig. 6.5-5 are intended to represent the waveforms only and not the amplitude. Since there is no staggering of parts of the symbol, the integrators extend their integration over the same time interval. Of course, again, a bit synchronizer is needed. The integrator outputs are voltages whose amplitudes are proportional to $T_s p_e$ and $T_s p_o$ respectively and change at the symbol rate. These voltages measure the components of the received signal in the directions of the quadrature phasors $\sin \omega_0 t$ and $\cos \omega_0 t$. Finally the signals $T_s p_e$ and $T_s p_o$ are applied to a device which reconstructs the digital $N$-bit signal which constitutes the transmitted signal.

There may or may not be a need to regenerate the bit stream. As a matter of fact, the idea of transmitting information one bit at a time by a bit stream $b(t)$ arises when we have a system, like BPSK which can handle only one bit at a time. If, on the other hand, our system handles $M$-bit symbols, then the data may originate as $M$-bit words. In such a case the serial to parallel converter shown in Fig. 6.6-2 is not needed.

Current operating systems are common in which $M = 16$. In this case the bandwidth is $B = 2f_b/4 = f_b/2$ in comparison to $B = f_b$ for QPSK. PSK systems transmit information through signal phase and not through signal amplitude. Hence such systems have great merit in situations where, on account of the vagaries of the transmission medium, the received signal varies in amplitude (i.e., fading channels).

## 6.7 QUADRATURE AMPLITUDE SHIFT KEYING (QASK)

In BPSK, QPSK, and $M$-ary PSK we transmit, in any symbol interval, one signal or another which are distinguished from one another in phase but are all of the *same amplitude*. In each of these individual systems the end points of the signal vectors in signal space falls on the circumference of a circle. Now we have noted that our ability to distinguish one signal vector from another in the presence of noise will depend on the distance between the vector end points. It is hence rather apparent that we shall be able to improve the noise immunity of a system by allowing the signal vectors to differ, not only in their phase but also in amplitude. We now describe such an *amplitude and phase shift keying* system. Like QPSK it involves direct (balanced) modulation of carriers in *quadrature* (i.e.,

integrators have an integration time equal to the symbol time $T_s$ and, of course, as usual, symbol time synchronizers (not shown) are required. Finally, the original input bits are recovered by using A/D converters.

## 6.8 BINARY FREQUENCY-SHIFT KEYING

In binary frequency-shift keying (BFSK) the binary data waveform $d(t)$ generates a binary signal

$$v_{\text{BFSK}}(t) = \sqrt{2P_s}\cos\left[\omega_0 t + d(t)\Omega t\right] \tag{6.8-1}$$

Here $d(t) = +1$ or $-1$ corresponding to the logic levels 1 and 0 of the data waveform. The transmitted signal is of amplitude $\sqrt{2P_s}$ and is either

$$v_{\text{BFSK}}(t) = s_H(t) = \sqrt{2P_s}\cos(\omega_0 + \Omega)t \tag{6.8-2}$$

or

$$v_{\text{BFSK}}(t) = s_L(t) = \sqrt{2P_s}\cos(\omega_0 - \Omega)t \tag{6.8-3}$$

and thus has an angular frequency $\omega_0 + \Omega$ or $\omega_0 - \Omega$ with $\Omega$ a constant offset from the nominal carrier frequency $\omega_0$. We shall call the higher frequency $\omega_H (= \omega_0 + \Omega)$ and the lower frequency $\omega_L (= \omega_0 - \Omega)$. We may conceive that the BFSK signal is generated in the manner indicated in Fig. 6.8-1. Two balanced modulators are used, one with carrier $\omega_H$ and one with carrier $\omega_L$. The voltage values of $p_H(t)$ and of $p_L(t)$ are related to the voltage values of $d(t)$ in the following manner

| $d(t)$ | $p_H(t)$ | $p_L(t)$ |
|---|---|---|
| +1V | +1V | 0V |
| −1V | 0V | +1V |

Thus when $d(t)$ changes from $+1$ to $-1$ $p_H$ changes from 1 to 0 and $p_L$ from 0 to 1. At any time either $p_H$ or $p_L$ is 1 but not both so that the generated signal is either at angular frequency $\omega_H$ or at $\omega_L$.

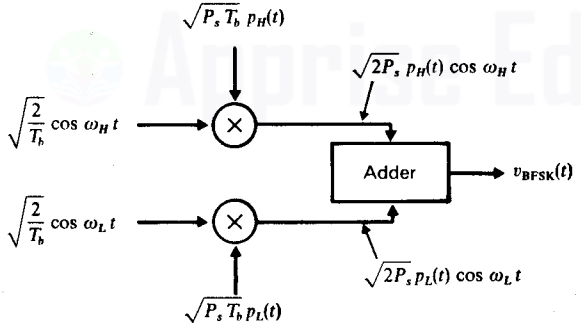



**Figure 6.8-1** A representation of a manner in which a BFSK signal can be generated.

### Spectrum of BFSK

In terms of the variables $p_H$ and $p_L$ the BFSK signal is

$$v_{\text{BFSK}}(t) = \sqrt{2P_s}\, p_H \cos(\omega_H t + \theta_H) + \sqrt{2P_s}\, p_L \cos(\omega_L t + \theta_L) \tag{6.8-4}$$

where we have assumed that each of the two signals are of independent and random, uniformly distributed phase. Each of the terms in Eq. (6.8-4) looks like the signal $\sqrt{2P_s}\, b(t)\cos\omega_0 t$ which we encountered in BPSK [see Eq. (6.2-3)] and for which we have already deduced the spectrum, but there is an important difference. In the BPSK case, $b(t)$ is bipolar, i.e., it alternates between $+1$ and $-1$ while in the present case $p_H$ and $p_L$ are unipolar, alternating between $+1$ and 0. We may, however, rewrite $p_H$ and $p_L$ as the sums of a constant and a bipolar variable, that is

$$p_H(t) = \tfrac{1}{2} + \tfrac{1}{2}\, p'_H(t) \tag{6.8-5a}$$

$$p_L(t) = \tfrac{1}{2} + \tfrac{1}{2}\, p'_L(t) \tag{6.8-5b}$$

In Eq. (6.8-5) $p'_H$ and $p'_L$ are bipolar, alternating between $+1$ and $-1$ and are complementary. When $p'_H$ is $+1$, $p'_L = -1$ and vice versa. We have then

$$v_{\text{BFSK}}(t) = \sqrt{\frac{P_s}{2}}\cos(\omega_H t + \theta_H) + \sqrt{\frac{P_s}{2}}\cos(\omega_L t + \theta_L) + \sqrt{\frac{P_s}{2}}\, p'_H \cos(\omega_H t + \theta_H) + \sqrt{\frac{P_s}{2}}\, p'_L \cos(\omega_L t + \theta_L) \tag{6.8-6}$$

The first two terms in Eq. (6.8-6) produce a power spectral density which consists of two impulses, one at $f_H$ and one at $f_L$. The last two terms produce the spectrum of two binary PSK signals (see Fig. 6.2-2*a*) one centered about $f_H$ and one about $f_L$. The individual power spectral density patterns of the last two terms in Eq. (6.8-6) are shown in Fig. 6.8-2 for the case $f_H - f_L = 2f_b$. For this separation between $f_H$ and $f_L$ we observe that the overlapping between the two parts of the spectra is not large and we may expect to be able, without excessive difficulty, to distinguish the levels of the binary waveform $d(t)$. In any event, with this separation the bandwidth of BFSK is

$$BW(\text{BFSK}) = 4f_b \tag{6.8-7}$$

which is twice the bandwidth of BPSK.

### Receiver for BFSK Signal

A BFSK signal is typically demodulated by a receiver system as in Fig. 6.8-3. The signal is applied to two bandpass filters one with center frequency at $f_H$ the other at $f_L$. Here we have assumed, as above, that $f_H - f_L = 2(\Omega/2\pi) = 2f_b$. The filter frequency ranges selected do not overlap and each filter has a passband wide enough to encompass a main lobe in the spectrum of Fig. 6.8-2. Hence one filter


Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

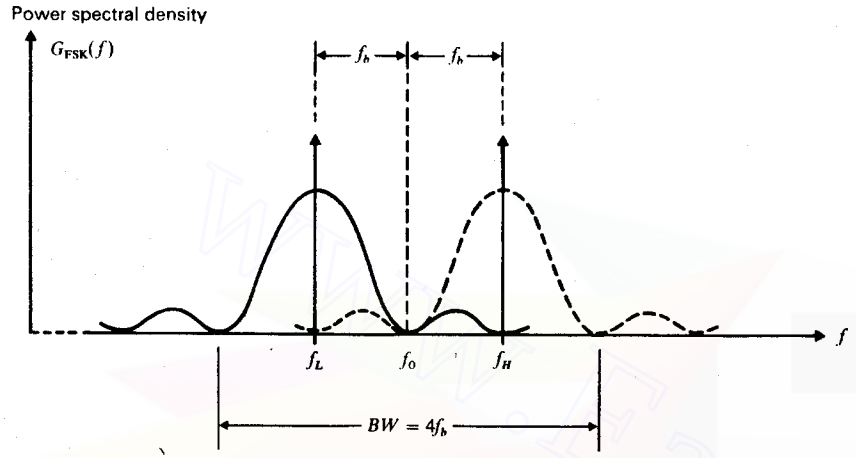



**Figure 6.8-2** The power spectral densities of the individual terms in Eq. (6.8-6).

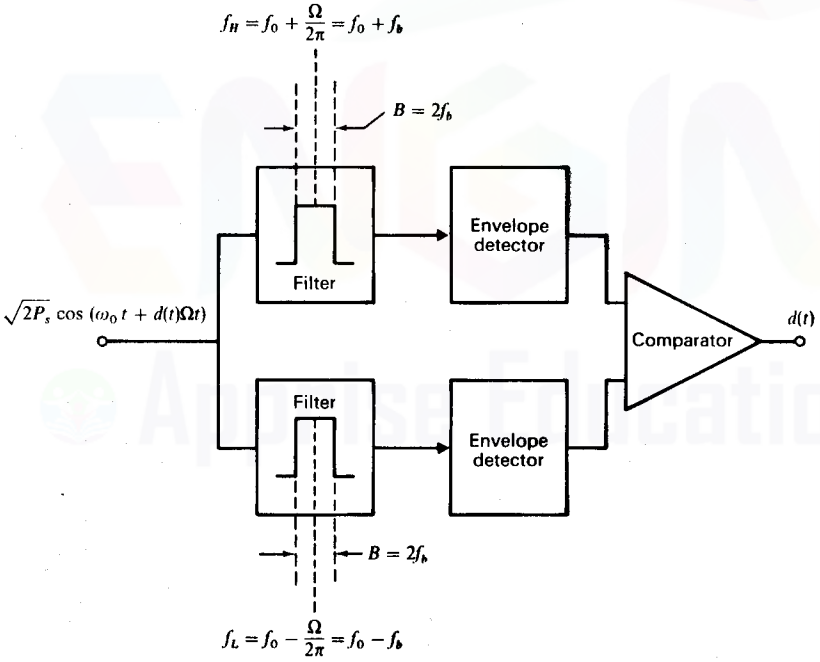



**Figure 6.8-3** A receiver for a BFSK signal.

will pass nearly all the energy in the transmission at $f_H$ the other will perform similarly for the transmission at $f_L$. The filter outputs are applied to envelope detectors and finally the envelope detector outputs are compared by a comparator. A comparator is a circuit that accepts two input signals. It generates a binary output which is at one level or the other depending on which input is larger. Thus at the comparator output the data $d(t)$ will be reproduced.

When noise is present, the output of the comparator may vary due to the systems response to the signal and noise. Thus, practical systems use a bit synchronizer and an integrator and sample the comparator output only once at the end of each time interval $T_b$.

## Geometrical Representation of Orthogonal BFSK

We noted, in $M$-ary phase-shift keying and in quadrature-amplitude shift keying, that any signal could be represented as $C_1 u_1(t) + C_2 u_2(t)$. There $u_1(t)$ and $u_2(t)$ are the orthonormal vectors in signal space, that is, $u_1(t) = \sqrt{2/T_s}\cos\omega_0 t$ and $u_2(t) = \sqrt{2/T_s}\sin\omega_0 t$. The functions $u_1$ and $u_2$ are orthonormal over the symbol interval $T_s$ and, if the symbol is a single bit, $T_s = T_b$. The coefficients $C_1$ and $C_2$ are constants. The normalized energies associated with $C_1 u_1(t)$ and with $C_2 u_2(t)$ are respectively $C_1^2$ and $C_2^2$ and the total signal energy is $C_1^2 + C_2^2$. In $M$-ary PSK and QASK the orthogonality of the vectors $u_1$ and $u_2$ results from their *phase quadrature*. In the present case of BFSK it is appropriate that the orthogonality should result from a special *selection of the frequencies* of the unit vectors. Accordingly, with $m$ and $n$ integers, let us establish unit vectors

$$u_1(t) = \sqrt{\frac{2}{T_b}}\cos 2\pi m f_b t \tag{6.8-8}$$

and

$$u_2(t) = \sqrt{\frac{2}{T_b}}\cos 2\pi n f_b t \tag{6.8-9}$$

in which, as usual, $f_b = 1/T_b$. The vectors $u_1$ and $u_2$ are the $m$th and $n$th harmonics of the (fundamental) frequency $f_b$. As we are aware, from the principles of Fourier analysis, different harmonics ($m \pm n$) are orthogonal over the interval of the fundamental period $T_b = 1/f_b$.

If now the frequencies $f_H$ and $f_L$ in a BFSK system are selected to be (assuming $m > n$)

$$f_H = m f_b \tag{6.8-10a}$$

$$f_L = n f_b \tag{6.8-10b}$$

then the corresponding signal vectors are

$$s_H(t) = \sqrt{E_b}\, u_1(t) \tag{6.8-11a}$$

$$s_L(t) = \sqrt{E_b}\, u_2(t) \tag{6.8-11b}$$


Downloaded From : www.EasyEngineering.net

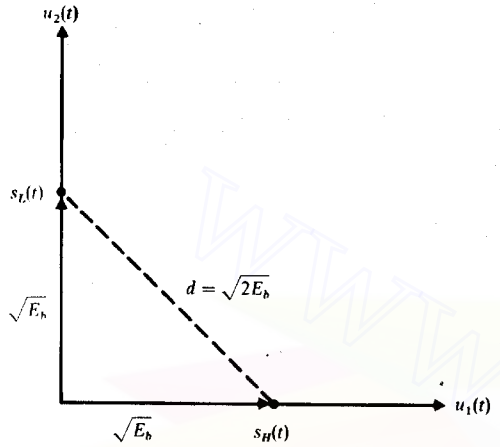



**Figure 6.8-4** Signal space representation of orthogonal BFSK.

The signal space representation of these signals is shown in Fig. 6.8-4. The signals, like the unit vectors are orthogonal. The distance between signal end points is therefore

$$d = \sqrt{2E_b} \tag{6.8-12}$$

Note that this distance is considerably smaller than the distance separating end points of BPSK signals, which are *antipodal.*

### Geometrical Representation of Non-Orthogonal BFSK

When the two FSK signals $s_H(t)$ and $s_L(t)$ are not orthogonal, the Gram–Schmidt procedure can still be used to represent the signals of Eqs. (6.8-2) and (6.8-3).

Let us represent the higher frequency signal $s_H(t)$ as:

$$s_H(t) = \sqrt{2P_s}\cos\omega_H t = s_{11}u_1(t) \qquad 0 \le t \le T_b \tag{6.8-13a}$$

and the lower frequency signal $s_L(t)$ as:

$$s_L(t) = \sqrt{2P_s}\cos\omega_L t = s_{12}u_1(t) + s_{22}u_2(t) \qquad 0 \le t \le T_b \tag{6.8-13b}$$

The representation of these two signals in signal space is shown in Fig. 6.8-5. Referring to this figure we see that the distance separating $s_H$ and $s_L$ is:

$$d^2_{\text{BFSK}} = (s_{11} - s_{12})^2 + s_{22}^2 = s_{11}^2 - 2s_{11}s_{12} + s_{12}^2 + s_{22}^2 \tag{6.8-14}$$

In order to determine $d^2_{\text{BFSK}}$ when the two signals are not orthogonal we must evaluate $s_{11}$, $s_{12}$, and $s_{22}$ using Eqs. (6.8-13). From Eq. (6.8-13*a*) we have:

$$s_{11}^2 = 2P_s\int_0^{T_b}\cos^2\omega_H t\,dt = E_b\left[1 + \frac{\sin 2\omega_H T_b}{2\omega_H T_b}\right] \tag{6-8-15}$$





Using Eq. (6.8-13*b*) we first determine $s_{12}$ by multiplying both sides of the equation by $u_1(t)$ and integrating from $0 \le t \le T_b$. The result is:

$$s_{12} = \sqrt{2P_s}\int_0^{T_b} u_1(t)\cos\omega_L t\,dt$$

$$= \frac{E_b}{s_{11}}\left[\frac{\sin(\omega_H - \omega_L)T_b}{(\omega_H - \omega_L)T_b} + \frac{\sin(\omega_H + \omega_L)T_b}{(\omega_H + \omega_L)T_b}\right] \tag{6.8-16a}$$

To arrive at this result we have used Eq. (6.8-13*a*) where

$$u_1(t) = (\sqrt{2P_s}/s_{11})\cos\omega_H t \tag{6.8-16b}$$

Finally, $s_{22}$ is found from Eq. (6.8-13*b*) by squaring both sides of the equation and then integrating from 0 to $T_b$. Since $u_1$ and $u_2$ are orthogonal, the result is:

$$\int_0^{T_b} s_L^2(t)\,dt = 2P_s\int_0^{T_b}\cos^2\omega_L t\,dt = s_{12}^2 + s_{22}^2 \tag{6.8-17a}$$

Hence

$$s_{12}^2 + s_{22}^2 = E_b\left[1 + \frac{\sin 2\omega_L T_b}{2\omega_L T_b}\right] \tag{6.8-17b}$$

The distance $d$ between $s_H$ and $s_L$ given in Eq. (6.8-14) can now be determined by substituting Eqs. (6.8-15), (6.8-16*a*), and (6.8-17*b*) into Eq. (6.8-14). The result is:

$$d^2 = E_b\left[1 + \frac{\sin 2\omega_H T_b}{2\omega_H T_b}\right] - 2E_b\left[\frac{\sin(\omega_H - \omega_L)T_b}{(\omega_H - \omega_L)T_b} + \frac{\sin(\omega_H + \omega_L)T_b}{(\omega_H + \omega_L)T_b}\right]$$
$$+ E_b\left[1 + \frac{\sin 2\omega_L T_b}{2\omega_L T_b}\right] \tag{6.8-18}$$

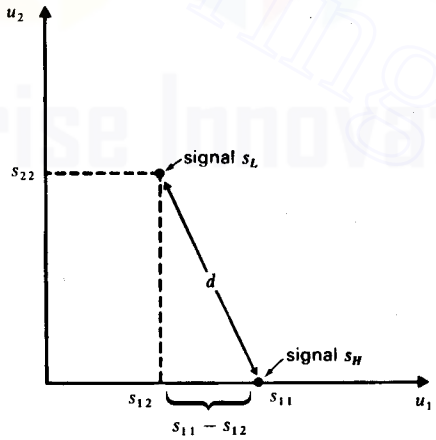



**Figure 6.8-5** Signal space representation of BFSK when $s_H(t)$ and $s_L(t)$ are not orthogonal.

Equation (6.8-18) can be simplified by recognizing that:

$$\left|\frac{\sin 2\omega_H T_b}{2\omega_H T_b}\right| \ll 1$$

$$\left|\frac{\sin 2\omega_L T_b}{2\omega_L T_b}\right| \ll 1$$

and

$$\left|\frac{\sin(\omega_H + \omega_L)T_b}{(\omega_H + \omega_L)T_b}\right| \ll \left|\frac{\sin(\omega_H - \omega_L)T_b}{(\omega_H - \omega_L)T_b}\right|$$

The final result is then

$$d^2 \cong 2E_b\left[1 - \frac{\sin(\omega_H - \omega_L)T_b}{(\omega_H - \omega_L)T_b}\right] \tag{6.8-19}$$

Note that when $s_H(t)$ and $s_L(t)$ are orthogonal $(\omega_H - \omega_L)T_b = 2\pi(m-n)f_bT_b = 2\pi(m-n)$ and Eq. (6.8-19) gives $d = \sqrt{2E_b}$ as expected. Note also that if $(\omega_H - \omega_L)T_b = 3\pi/2$, the distance $d$ is increased and becomes

$$d_{\text{opt}} = \left[2E_b\left(1 + \frac{2}{3\pi}\right)\right]^{1/2} \simeq \sqrt{2.4E_b} \tag{6.8-20}$$

an increase in $d^2$ by 20 percent.

## 6.9 COMPARISON OF BFSK AND BPSK

Let us start with the BFSK signal of Eq. (6.8-1). Using the trigonometric identity for the cosine of the sum of two angles and recalling that $\cos\theta = \cos(-\theta)$ while $\sin\theta = -\sin(-\theta)$ we are led to the alternate equivalent expression

$$v_{\text{BFSK}}(t) = \sqrt{2P_s}\cos\Omega t\cos\omega_0 t - \sqrt{2P_s}\,d(t)\sin\Omega t\sin\omega_0 t \tag{6.9-1}$$

Note that the second term in Eq. (6.9-1) looks like the signal encountered in BPSK i.e., a carrier $\sin\omega_0 t$ multiplied by a data bit $d(t)$ which changes the carrier phase. In the present case however, the carrier is not of fixed amplitude but rather the amplitude is shaped by the factor $\sin\Omega t$. We note further the presence of a quadrature reference term $\cos\Omega t\cos\omega_0 t$ which contains no information. Since this quadrature term carries energy, the energy in the information bearing term is thereby diminished. Hence we may expect that BFSK will not be as effective as BPSK in the presence of noise. For orthogonal BFSK, each term has the same energy, hence the information bearing term contains only one-half of the total transmitted energy.

## 6.10 *M*-ARY FSK

An $M$-ary FSK communications system is shown in Fig. 6.10-1. It is an obvious extension of a binary FSK system. At the transmitter an $N$-bit symbol is presented each $T_s$ to an $N$-bit D/A converter. The converter output is applied to a fre-

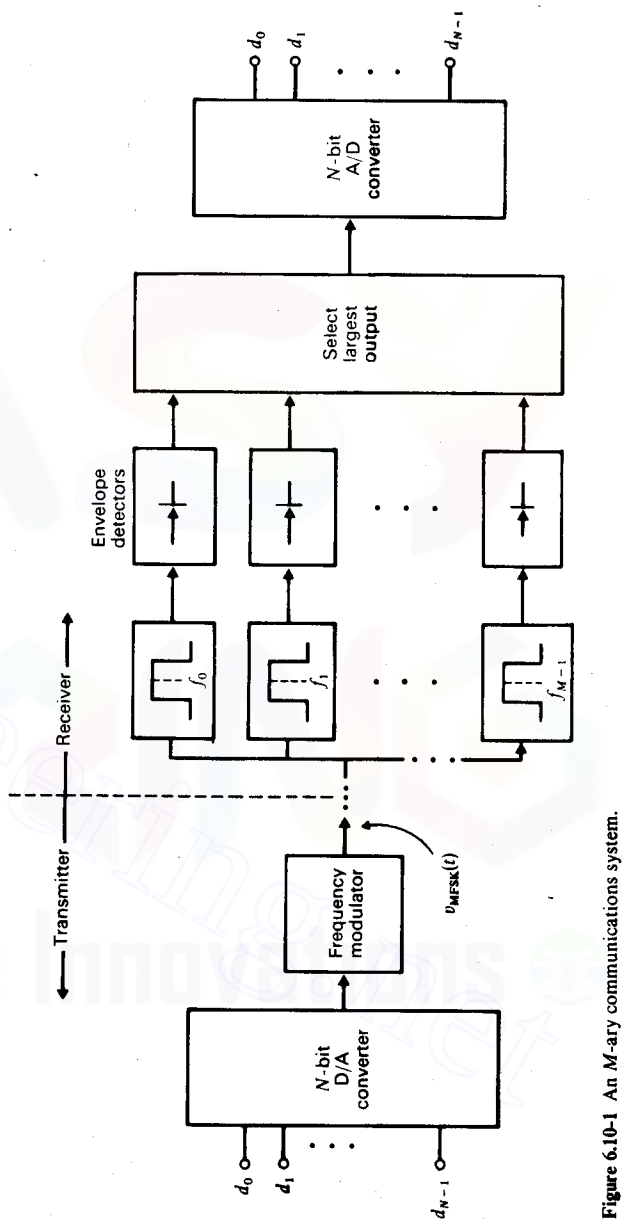



**Figure 6.10-1** An $M$-ary communications system.

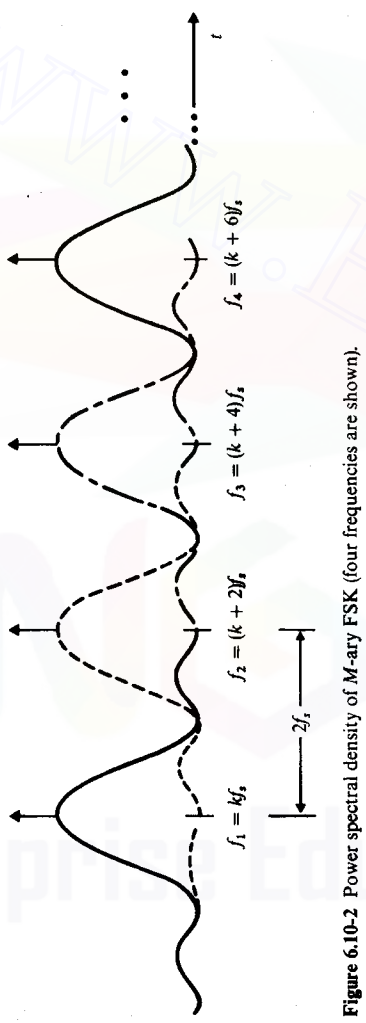



**Figure 6.10-2** Power spectral density of $M$-ary FSK (four frequencies are shown).



quency modulator, i.e., a piece of hardware which generates a carrier waveform whose frequency is determined by the modulating waveform. The transmitted signal, for the duration of the symbol interval, is of frequency $f_0$ or $f_1$ ... or $f_{M-1}$ with $M = 2^N$. At the receiver, the incoming signal is applied to $M$ paralleled bandpass filters each followed by an envelope detector. The bandpass filters have center frequencies $f_0, f_1, \ldots, f_{M-1}$. The envelope detectors apply their outputs to a device which determines which of the detector indications is the largest and transmits that envelope output to an $N$-bit A/D converter.

As we shall see (Sec. 11.15) the probability of error is minimized by selecting frequencies $f_0, f_1, \ldots, f_{M-1}$ so that the $M$ signals are mutually orthogonal. One commonly employed arrangement simply provides that the carrier frequency be successive even harmonics of the symbol frequency $f_s = 1/T_s$. Thus the lowest frequency, say $f_0$, is $f_0 = kf_s$, while $f_1 = (k+2)f_s, f_2 = (k+4)f_s$, etc. In this case, the spectral density patterns of the individual possible transmitted signals overlap in the manner shown in Fig. 6.10-2, which is an extension to $M$-ary FSK of the pattern of Fig. 6.8-2 which applies to binary FSK. We observe that to pass $M$-ary FSK the required spectral range is

$$B = 2Mf_s \tag{6.10-1}$$

Since $f_s = f_b/N$ and $M = 2^N$ we have

$$B = 2^{N+1} f_b/N \tag{6.10-2}$$

Note that $M$-ary FSK requires a considerably increased bandwidth in comparison with $M$-ary PSK. However, as we shall see, the probability of error for $M$-ary FSK *decreases* as $M$ increases, while for $M$-ary PSK, the probability of error increases with $M$.

### Geometrical Representation of $M$-ary FSK

In Fig. 6.8-4, we provided a signal space representation for the case of orthogonal binary FSK. The case of $M$-ary orthogonal FSK signals is clearly an extension of this figure. We simply conceive of a coordinate system with $M$ mutually orthogonal coordinate axes. The signal vectors are then parallel to these axes. The best we can do pictorially is the three-dimensional case shown in Fig. 6.10-3. As usual, and as is indicated in the figure, the square of the length of the signal vector is the normalized signal energy. Note that, as in Fig. 6.8-4, the distance between signal points is

$$d = \sqrt{2E_s} = \sqrt{2NE_b} \tag{6.10-3}$$

Note that this value of $d$ is greater than the values of $d$ calculated for $M$-ary PSK with the exception of the cases $M = 2$ and $M = 4$. It is also greater than $d$ in the case of 16-QASK.

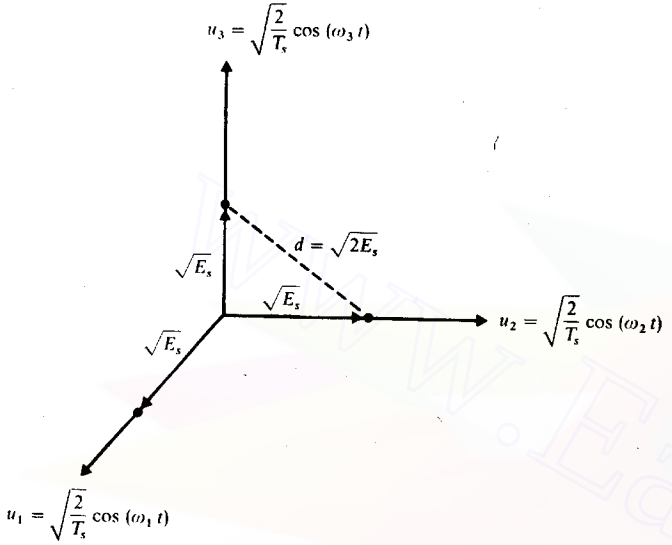



**Figure 6.10-3** Geometrical representation of orthogonal $M$-ary FSK ($M = 3$) when the frequencies are selected to generate orthogonal signals.

## 6.11 MINIMUM SHIFT KEYING (MSK)

In discussing *minimum shift keying*, we shall want to make a number of comparisons between MSK and QPSK. One of these comparisons will concern the spectra of the two systems. For this reason we review briefly some matters concerning the spectrum of QPSK. In offset QPSK the transmitted signal is [see Eq. (6.5-1)]

$$v_{OQPSK}(t) = \sqrt{P_s}\, b_e(t) \cos \omega_0 t + \sqrt{P_s}\, b_o(t) \sin \omega_0 t \qquad (6.11\text{-}1)$$

To find the power spectral density of this signal we start with the power spectral density of the baseband waveform $b_e(t)$. The waveform $p(t) \equiv \sqrt{P_s}\, b_e(t)$ is a random sequence of rectangular waveforms, flat topped for symbol duration $T_s = 2T_b$, and having an amplitude $\pm\sqrt{P_s}$. Its power spectral density $G(f)$ is given by the general formula (Eq. (2.24-14))

$$G_p(f) = \frac{\overline{|P(f)|^2}}{T_s} \qquad (6.11\text{-}2)$$

where $P(f)$ is the Fourier transform $p(t)$. We readily calculate that the two-sided power spectral density of $p(t)$ is, with $f_s = 1/T_s = f_b/2$ and $P_s = E_b/T_b$

$$G_p(f) = 2E_b \left(\frac{\sin 2\pi f/f_b}{2\pi f/f_b}\right)^2 \qquad (6.11\text{-}3)$$

The spectral density of $\sqrt{P_s}\, b_e(t) \cos \omega_0 t$ is generated by translating the two-sided pattern of Eq. (6.11-3) to $+f_0$ and also to $-f_0$, each such translated pattern being reduced in magnitude by 4 because the multiplication of $\sqrt{P_s}\, b_e(t)$ by $\cos \omega_0 t$ translates one-half of the voltage spectrum to $f_0$ and the other half to $-f_0$. (See Sec. 3.2.) Thus the voltage of each term is reduced by 2 and the power by 4. The power spectral density of the second term in Eq. (6.11-1), that is the term $\sqrt{P_s}\, b_o(t) \sin \omega_0 t$ is identical to the density of the first. Finally we note that the two terms are not correlated since $b_e(t)$ and $b_o(t)$ are quite independent of one another. Hence the total power density is twice the density generated by either term. Altogether then we find

$$G_{OQPSK}(f) = E_b \left\{ \left[\frac{\sin 2\pi(f - f_0)/f_b}{2\pi(f - f_0)/f_b}\right]^2 + \left[\frac{\sin 2\pi(f + f_0)f_b}{2\pi(f + f_0)/f_b}\right]^2 \right\} \qquad (6.11\text{-}4)$$

Earlier, upon examining the pattern for the power spectral density (see Fig. 6.2-2) we took the attitude that the bandwidth of QPSK (QPSK and OQPSK yield the same result) is $B = f_b$ because such a bandwidth is adequate to encompass the main lobe. The main lobe contains 90 percent of the signal energy. Still, the not inconsiderable power outside the main lobe is a source of trouble when QPSK is to be used for multichannel communication on adjacent carriers. If, say, we establish additional channels at carrier frequencies $f_0' = f_0 \pm f_b$, then the side lobe associated with the first channel, having a peak value at frequency $f_0 + 3f_b/4$, will be a source of serious interchannel interference. These side lobes, as is to be noted in Fig. 6.2-2, are smaller than the main lobe by only 14 dB.

This difficulty, i.e., the wide spectrum of QPSK, is due to the character of the baseband signal. This signal consists of *abrupt changes*, and abrupt changes give rise to *spectral components at high frequencies*. In short, the baseband spectral range is very large and multiplication by a carrier translates the spectral pattern without changing its form. We might try to alleviate the difficulty by passing the baseband signal through a low-pass filter to suppress the many side lobes. Such filtering will cause intersymbol interference. The problem of interchannel interference in QPSK is so serious that regulatory and standardization agencies such as the FCC and CCIR will not permit these systems to be used except with bandpass filtering at the carrier frequency (i.e., at the transmitter output) to suppress the side lobes.

The filtering which we have just described does not, in certain situations, necessarily resolve the problem of interchannel interference. We shall discuss the matter qualitatively: We recall that QPSK (staggered or not) is a system in which the signal is of constant amplitude, the information content being borne by phase changes. In both QPSK and OQPSK there are abrupt phase changes in the signal. In QPSK these changes can occur at the symbol rate $1/T_s = 1/2T_b$ and can be as large as 180°. In OQPSK phase changes of 90° can occur at the bit rate. Now it turns out that when such waveforms with abrupt phase changes, are filtered to suppress sidebands, the effect of the filter, at the times of the abrupt phase changes, is to cause substantial changes in the *amplitude* of the waveform. Such amplitude variations can cause problems in QPSK communication systems


Downloaded From : www.EasyEngineering.net

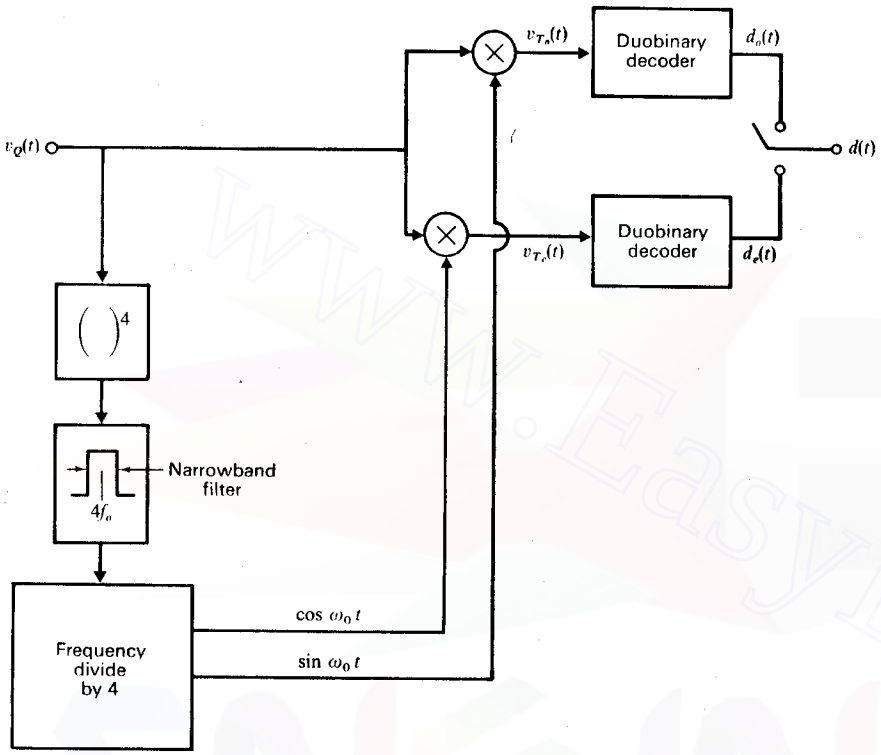



**Figure 6.15-2** QPR decoder.

takes place; each duobinary decoder being similar to the decoder shown in Fig. 6.14-3*b* except that they operate at $f_b/2$ rather than at $f_b$. The reconstructed data $d_o(t)$ and $d_e(t)$ is then combined to yield the data $d(t)$.

## PROBLEMS

**6.2-1** The data $b(t)$ consists of the bit stream 001010011010. Assume that the bit rate $f_b$ is equal to the carrier frequency $f_0$ and sketch $v_{\text{BPSK}}(t)$.

**6.2-2.** Calculate and plot the power $P_f$ contained in the power spectral density $G_b(f)$ in Eq. (6.2-8) as a function of the frequency range 0 to $f$. Note that when $f = 0$, $P_f = 0$ and when $f$ is infinite $P_f = P_s$.

**6.3-1.** The bit stream $d(t)$ is to be transmitted using DPSK. If $d(t)$ is 001010011010, determine $b(t)$. Show that $b(t)b(t - T_b)$ yields the original data.

**6.4-1.** The bit stream $d(t)$ is to be transmitted using DEPSK. If $d(t)$ is 001010011010, determine $b(t)$. Show that after decoding, using the circuit of Fig. 6.4-1, the data $d(t)$ is recovered. Show that if the fourth bit in $b(t)$ is in error, then the fourth and fifth data bits, $d(t)$, will also be in error.

**6.5-1.** The bit stream $b(t)$ is to be transmitted using OQPSK. If $b(t)$ is 001010011010 sketch $v_m(t)$ (see Eq. 6.5-1). Assume $f_s = f_b/2 = f_0$.

**6.5-2.** Repeat Prob. 6.5-1 if the odd bit stream is delayed by $T_b$ so that QPSK is generated.

**6.5-3.** Verify that the circuit shown in Fig. 6.5-5 yields sin $\omega_0 t$ and cos $\omega_0 t$, each with the same phase as the received signal.

**6.5-4.** Verify Eq. (6.5-5). Show that the minimum distance between signals is given by Eq. (6.5-6).

**6.6-1.** 8-ary PSK is used to modulate the bit stream 001010011010. Sketch the transmitted waveform $v_{\text{8PSK}}(t)$. Assume $f_s = f_b/3 = f_0$.

**6.6-2.** Repeat Prob. 6.6-1 if 16-ary PSK is used. Assume $f_s = f_b/4 = f_0$.

**6.6-3.** Verify Eq. (6.6-9) taking note of the fact that $\phi_m$ has the discrete values given by Eq. (6.6-2) and does not take on all values from $-\pi$ to $\pi$.

**6.6-4.** Verify Eq. (6.6-11).

**6.6-5.** If $M$ becomes large, in Eq. (6.6-11), show that $d$ approaches zero.

**6.7-1.** Verify Eqs. (6.7-2) and (6.7-3).

**6.7-2.** The bit stream 001010011010 is to be transmitted using 16-QASK. Sketch the transmitted waveform. Assume $f_s = f_b/4 = f_0$.

**6.7-3.** Verify Eq. (6.7-5).

**6.7-4.** Verify Eq. (6.7-11).

**6.7-5.** Verify Eq. (6.7-12).

**6.7-6.** Verify Eq. (6.7-15). Calculate the average power contained in the signal $v^4_{\text{QASK}}(t)$ at the frequency $4f_0$.

**6.8-1.** The bit stream 001010011010 is to be transmitted using BFSK. Sketch the transmitted waveform. Assume $f_L = f_b$ and $f_H = 2f_b$.

**6.8-2.** One way to obtain the power spectral density of a waveform is to first obtain the autocorrelation function $R(\tau)$. The power spectral density is the Fourier transform of $R(\tau)$. Since $R(\tau) = E[v(t)v(t + \tau)]$,

(*a*) Find $R(\tau)$ for the BFSK signal of Eq. (6.8-6) and verify Fig. (6.8-2).

(*b*) If $\theta_H = \theta_L$ find $R(\tau)$ and the power spectral density.

**6.8-3.** Verify Eq. (6.8-12).

**6.8-4.** Verify Eqs. (6.8-15).

**6.8-5.** Verify Eqs. (6.8-16*a*) and (6.8-17*b*).

**6.8-6.** Verify Eq. (6.8-19). What is the relationship between $f_H$, $f_L$ and $f_b$ in order that Eq. (6.8-19) reduce to Eq. (6.8-12).

**6.10-1.** Consider using 4-ary FSK.

(*a*) Show that if the frequencies are separated by $f_s$, they are each orthogonal.

(*b*) Calculate the bandwidth $B$ under the condition of (*a*). Show that the bandwidth is five-eighths of the value required by Eq. (6.10-1). Explain the difference.

(*c*) Determine the ratio of bandwidths of 2-ary FSK and 4-ary FSK when the frequencies are separated by $f_s$. Repeat for $2f_s$.

**6.11-1.** Verify Eq. (6.11-6*a*).

**6.11-2.** Verify Eq. (6.11-8) using Eq. (6.11-12).

**6.11-3.** If $b(t)$ is 001010011010 sketch $v_{\text{MSK}}(t)$. Assume in Eq. (6.11-12) that $m = 5$.

**6.11-4.** Repeat Prob. 6.11-3 if $m = 3$.

**6.11-5.** Verify Eq. (6.11-18).

**6.11-6.** Verify that in Fig. 6.11-5*a* the waveforms $x(t)$ and $y(t)$ are as shown and that the output is $v_{\text{MSK}}(t)$.

**6.12-1.** Verify Table 6.12-1.

**6.12-2.** Verify Eqs. (6.12-7) and (6.12-8).


Downloaded From : www.EasyEngineering.net

**6.12-3.** Verify Eqs. (6.12-10) and (6.12-11).

**6.12-4.** Plot $G_D(f)$ and $G_T(f)$ as a function of frequency using Eqs. (6.12-10) and (6.12-11).

**6.14-1.** Verify Eq. (6.14-5).

**6.15-1.** Show that if two duobinary signals $v_{T_e} = b_e(k) \cdot b_e(k-1)$ and $v_{T_o} = b_o(k)b_o(k-1)$ are modulated using carrier signals which are in time quadrature so that:

$$v_Q(t) = \sqrt{P_s}\, V_{T_e}(t) \cos \omega_0 t + \sqrt{P_s}\, V_{T_o}(t) \sin \omega_0 t$$

then the carrier signals $\cos \omega_0 t$ and $\sin \omega_0 t$ can be recovered in the receiver by filtering the signal $V_Q^4(t)$.

**6.15-2.** Repeat 6.15-1 if $V_{T_e}(t)$ and $V_{T_o}(t)$ are each filtered by a "brick wall" low-pass-filter prior to being amplitude modulated.





CHAPTER

# SEVEN

# MATHEMATICAL REPRESENTATION OF NOISE

All the communication systems discussed in the preceding chapters accomplish the same end. They allow us to reproduce the signal, impressed on the communication channel at the transmitter, at the demodulator output. Our only basis for comparison between systems, up to this point, has been bandwidth occupancy, convenience of multiplexing, and ease of implementation of the physical hardware. We have neglected, however, in our preceding discussions, the very important and fundamental fact that, in any real physical system, when the signal voltage arrives at the demodulator, it will be accompanied by a voltage waveform which varies with time in an entirely *unpredictable* manner. This unpredictable voltage waveform is a random process called *noise*. A signal accompanied by such a waveform is described as being *contaminated* or *corrupted* by noise. We now find that we have a new basis for system comparison, that is, the extent to which a communication system is able to distinguish the signal from the noise and thereby yield a low-*distortion* and low-*error* reproduction of the original signal.

In the present chapter we shall make only a brief reference to the sources of noise which are discussed more extensively in Chap. 14. Here we shall be concerned principally with a discussion of the mathematical representation and statistical characterizations of noise.

## 7.1 SOME SOURCES OF NOISE

One source of noise is the constant agitation which prevails throughout the universe at the molecular level. Thus, a piece of solid metal may appear to our gross view to be completely at rest. We know, however, that the individual molecules

CHAPTER

# EIGHT

# NOISE IN AMPLITUDE-MODULATION SYSTEMS

In Chap. 3 we described a number of different amplitude-modulation communication systems. In the present chapter we shall compare the performance of these systems under the circumstances that the received signal is corrupted by noise.

## 8.1 AMPLITUDE-MODULATION RECEIVER

A system for processing an amplitude-modulated carrier and recovering the baseband modulating system is shown in Fig. 8.1-1. We assume that the signal has suffered great attenuation during the course of its transmission over the communication channel and hence is in need of amplification. The input to the system might be a signal furnished by a receiving antenna which receives its signal from a transmitting antenna. The carrier of the received signal is called a *radio-frequency* (RF) carrier, and its frequency is the *radio* frequency $f_{rf}$. The input signal is amplified in an RF amplifier and then passed on to a *mixer*. In the mixer the modulated RF carrier is mixed (i.e., multiplied) with a sinusoidal waveform generated by a local oscillator which operates at a frequency $f_{osc}$. The process of mixing is also called *heterodyning*, and since, as is to be explained, the heterodyning local-oscillator frequency $f_{osc}$ is selected to be *above* the radio frequency $f_{rf}$, the system is often referred to as a *superheterodyne* system.

The process of mixing generates sum and difference frequencies. Thus the mixer output consists of a carrier of frequency $f_{osc} + f_{rf}$ and a carrier $f_{osc} - f_{rf}$. Each carrier is modulated by the baseband signal to the same extent as was the input RF carrier. The sum frequency is rejected by a filter. This filter is not shown









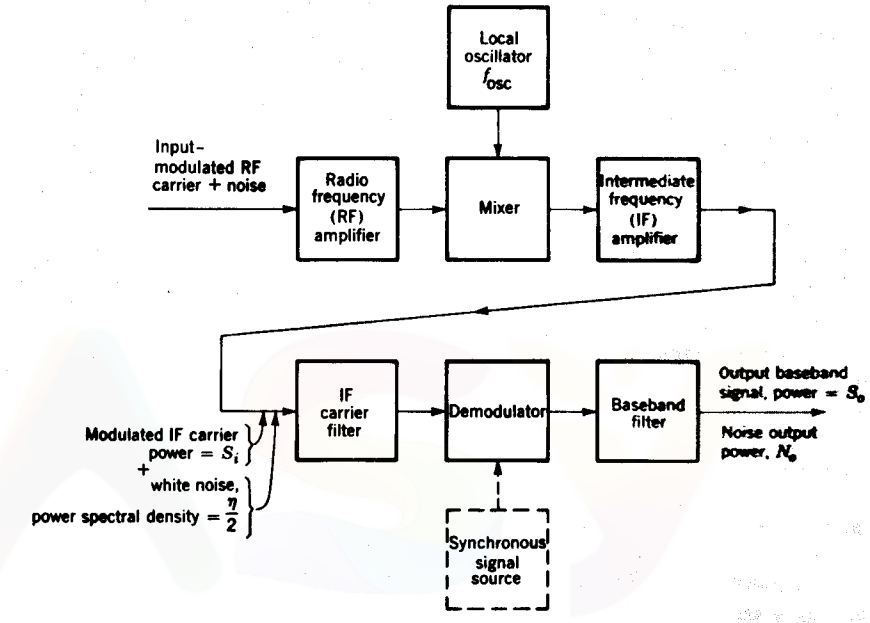



**Figure 8.1-1** A receiving system for an amplitude-modulated signal.

explicitly in Fig. 8.1-1 and may be considered to be part of the mixer. The difference-frequency carrier is called the *intermediate frequency* (IF) carrier, that is, $f_{if} = f_{osc} - f_{rf}$. The modulated IF carrier is applied to an IF amplifier. The process just described, in which a modulated RF carrier is replaced by a modulated IF carrier, is called *conversion*. The combination of the mixer and local oscillator is called a *converter*.

The IF amplifier output is passed, through an IF carrier filter, to the demodulator in which the baseband signal is recovered, and finally through a baseband filter. The baseband filter may include an amplifier, not explicitly indicated in Fig. 8.1-1. If synchronous demodulation is used, a synchronous signal source will be required.

The only absolutely essential operation performed by the receiver is the process of frequency translation back to baseband. This process is, of course, the inverse of the operation of modulation in which the baseband signal is frequency-translated to a carrier frequency. The process of frequency translation is performed in the system of Fig. 8.1-1 in part by the converter and in part by the demodulator. For this reason the converter is sometimes referred to as the *first detector*, while the demodulator is then called the *second detector*. The only other components of the system are the linear amplifiers and filters, none of which would be essential if the signal were strong enough and there were no need for multiplexing.

It is apparent that there is no essential need for an initial conversion before demodulation. The modulated RF carrier may be applied directly to the demodulator. However, the superheterodyne principle, which is rather universally incorporated into receivers, has great merit, as is discussed in the next section.

## 8.2 ADVANTAGE OF THE SUPERHETERODYNE PRINCIPLE: SINGLE CHANNEL

A signal furnished by an antenna to a receiver may have a power as low as some tens of picowatts, while the required output signal may be of the order of tens of watts. Thus the magnitude of the required gain is very large. In addition, to minimize the noise power presented to the demodulator, filters are used which are no wider than is necessary to accommodate the baseband signal. Such filters should be rather flat-topped and have sharp skirts. It is more convenient to provide gain and sharp flat-topped filters at low frequencies than at high. By way of example, in commercial FM broadcasting, the RF carrier frequency is in the range of 100 MHz, while at the FM receiver the IF frequency is 10.7 MHz.

Thus, in Fig. 8.1-1 the largest part, by far, of the required gain is provided by the IF amplifier, and the critical filtering done by the IF filter. While Fig. 8.1-1 suggests a separate amplifier and filter, actually in physical receivers these two usually form an integral unit. For example, the IF amplifier may consist of a number of amplifier stages, each one contributing to the filtering. Some filtering will also be incorporated in the RF amplifier. But this filtering is not critical. It serves principally to limit the total noise power input to the mixer and thereby avoids *overloading* the mixer with a noise waveform of excessive amplitude.

RF amplification is employed whenever the incoming signal is very small. This is because of the fact that RF amplifiers, such as masers, are *low-noise* devices; i.e., an RF amplifier can be designed to provide relatively high gain while generating relatively little noise. When RF amplification is not employed, the signal is applied directly to the mixer. The mixer provides relatively little gain and generates a relatively large noise power. Calculations showing typical values of gain and noise power generation in RF, mixer, and IF amplifiers are presented in Sec. 14.14.

### Multiplexing

An even greater merit of the superheterodyne principle becomes apparent when we consider that we shall want to tune the receiver to one or another of a number of different signals, each using a different RF carrier. If we were not to take advantage of the superheterodyne principle, we would require a receiver in which many stages of RF amplification were employed, each stage requiring tuning. Such tuned-radio-frequency (TRF) receivers were, as a matter of fact, commonly employed during the early days of radio communication. It is difficult enough to operate at the higher radio frequencies; it is even more difficult to







*gang-tune* the individual stages over a wide band, maintaining at the same time a reasonably sharp flat-topped filter characteristic of constant bandwidth.

In a *superhet* receiver, however, we need but change the frequency of the local oscillator to go from one RF carrier frequency to another. Whenever $f_{osc}$ is set so that $f_{osc} - f_{rf} = f_{if}$, the mixer will convert the input modulated RF carrier to a modulated carrier at the IF frequency, and the signal will proceed through the demodulator to the output. Of course, it is necessary to gang the tuning of the RF amplifier to the frequency control of the local oscillator. But again this ganging is not critical, since only one or two RF amplifiers and filters are employed.

Finally, we may note the reason for selecting $f_{osc}$ higher than $f_{rf}$. With this higher selection the fractional change in $f_{osc}$ required to accommodate a given range of RF frequencies is smaller than would be the case for the alternative selection.

## 8.3 SINGLE-SIDEBAND SUPPRESSED CARRIER (SSB-SC)

The receiver of Fig. 8.1-1 is suitable for the reception and demodulation of all types of amplitude-modulated signals, single sideband or double sideband, with and without carrier. The only essential changes required to accommodate one type of signal or another are in the demodulator and in the bandwidth of the IF carrier filter. Hence, from this point, our interest will focus on the section of receiver beginning with the IF filter and through to the output.

The signal input to the IF filter is an amplitude-modulated IF carrier. The normalized power (power dissipated in a 1-ohm resistor) of this signal is $S_i$. The signal arrives with noise. Added, is the noise generated in the RF amplifier and amplified in the RF amplifier and IF amplifiers. The IF amplifiers and mixer are also sources of noise, i.e., thermal noise, shot noise, etc., but this noise, lacking the gain of the RF amplifier, represents a second-order effect. (See Sec. 14.11.) We shall assume that the noise is gaussian, white, and of two-sided power spectral density $\eta/2$. The IF filter is assumed rectangular and of bandwidth no wider than is necessary to accommodate the signal. The output baseband signal has a power $S_o$ and is accompanied by noise of total power $N_o$.

### Calculation of Signal Power

With a single-sideband suppressed-carrier signal, the demodulator is a multiplier as shown in Fig. 8.3-1*a*. The carrier is $A \cos 2\pi f_c t$. For synchronous demodulation the demodulator must be furnished with a synchronous locally generated carrier $\cos 2\pi f_c t$. We assume that the upper sideband is being used; hence the carrier filter has a bandpass, as shown in Fig. 8.3-1*b*, that extends from $f_c$ to $f_c + f_M$, where $f_M$ is the baseband bandwidth. The bandwidth of the baseband filter extends from zero to $f_M$ as shown in Fig. 8.3-1*c*.

Let us assume that the baseband signal is a sinusoid of angular frequency

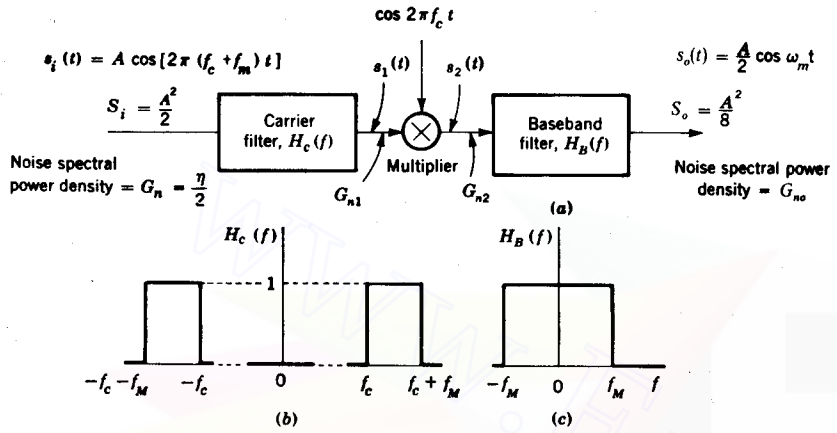



**Figure 8.3-1** (*a*) A synchronous demodulator operating on a single-sideband single-tone signal. (*b*) The bandpass range of the carrier filter. (*c*) The passband of the lowpass baseband filter.

$f_m (f_m \leq f_M)$. The carrier frequency is $f_c$, and, since we have assumed that the upper sideband is being used, the received signal is

$$s_i(t) = A \cos [2\pi(f_c + f_m)t] \tag{8.3-1}$$

The output of the multiplier is

$$s_2(t) = s_1(t) \cos \omega_c t = \frac{A}{2} \cos [2\pi(2f_c + f_m)t] + \frac{A}{2} \cos 2\pi f_m t \tag{8.3-2}$$

Only the difference-frequency term will pass through the baseband filter. Therefore the output signal is

$$s_o(t) = \frac{A}{2} \cos 2\pi f_m t \tag{8.3-3}$$

which is the modulating signal amplified by $\frac{1}{2}$.

The input signal power is

$$S_i = \frac{A^2}{2} \tag{8.3-4}$$

while the output signal power is

$$S_o = \frac{1}{2}\left(\frac{A}{2}\right)^2 = \frac{A^2}{8} = \frac{S_i}{4} \tag{8.3-5}$$

Thus

$$\frac{S_o}{S_i} = \frac{1}{4} \tag{8.3-6}$$



We may readily see that, even though Eq. (8.3-6) was deduced on the assumption of a sinusoidal baseband signal, the result is entirely general. For suppose that the baseband signal were quite arbitrary in waveshape. Then the single-sideband signal generated by this baseband signal may be resolved into a series of harmonically related spectral components. The input power is the sum of the powers in these individual components. Next, we note, as was discussed in Chap. 3, that superposition applies to the multiplication process being used for demodulation. Therefore, the output signal power generated by the simultaneous application at the input of many spectral components is simply the sum of the output powers that would result from each spectral component individually. Hence $S_i$ and $S_o$ in Eqs. (8.3-4) and (8.3-5) are properly the *total* powers, independently of whether a single or many spectral components are involved.

## Calculation of Noise Power

We now calculate the output noise $N_o$. For this purpose, we recall from Sec. 7.8 that when a noise spectral component at a frequency $f$ is multiplied by $\cos 2\pi f_c t$, the original noise component is replaced by two components, one at frequency $f_c + f$ and one at frequency $f_c - f$, each new component having one-fourth the power of the original.

The input noise is white and of spectral density $\eta/2$. The noise input to the multiplier has a spectral density $G_{n1}$ as shown in Fig. 8.3-2*a*. The density of the noise after multiplication by $\cos 2\pi f_c t$ is $G_{n2}$ as is shown in Fig. 8.3-2*b*. Finally the noise transmitted by the baseband filter is of density $G_{no}$ as in Fig. 8.3-2*c*. The total noise output is the area under the plot in Fig. 8.3-2*c*. We have, then, that

$$N_o = 2f_M \frac{\eta}{8} = \frac{\eta f_M}{4} \tag{8.3-7}$$

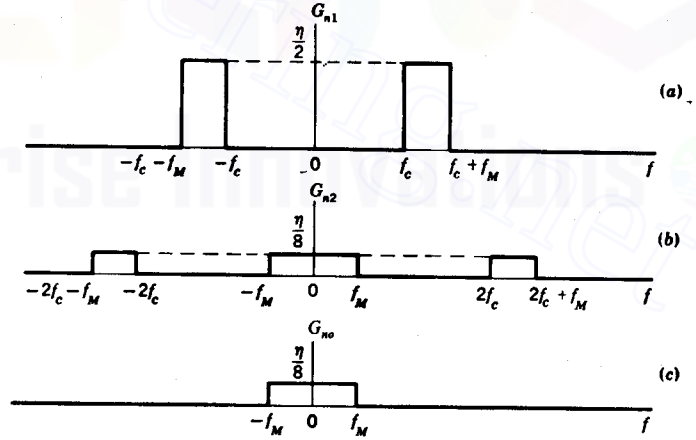



**Figure 8.3-2** Spectral densities of noises in SSB demodulator. (*a*) Density $G_{n1}$ of noise input to multiplier. (*b*) Density $G_{n2}$ of noise output of multiplier. (*c*) Density $G_{no}$ of noise output of baseband filter.

### Use of Quadrature Noise Components

It is of interest to calculate the output noise power $N_o$ in an alternative manner using the transformation of Eq. (7.11-2):

$$n(t) = n_c(t) \cos 2\pi f_c t - n_s(t) \sin 2\pi f_c t \tag{8.3-8}$$

We now apply Eq. (8.3-8) to the noise output of the IF filter so that $n(t)$ has the spectral density $G_{n1}$ as in Fig. 8.3-2*a*. The spectral densities of $n_c(t)$ and $n_s(t)$ are [see Eqs. (7.12-7) and (7.12-8)]:

$$G_{n_c}(f) = G_{n1}(f) = G_{n1}(f_c - f) + G_{n1}(f_c + f) \tag{8.3-9}$$

We observe that for $0 \le f \le f_M$, $G_{n1}(f_c + f) = \eta/2$, while $G_{n1}(f_c - f) = 0$, so that $G_{n_c}(f)$ and $G_{n_s}(f)$ are as shown in Fig. 8.3-3.

Multiplying $n(t)$ by $\cos 2\pi f_c t$ yields

$$\begin{aligned} n(t) \cos 2\pi f_c t &= n_c(t) \cos^2 2\pi f_c t - n_s(t) \sin 2\pi f_c t \cos 2\pi f_c t \\ &= \tfrac{1}{2} n_c(t) + \tfrac{1}{2} n_c(t) \cos 4\pi f_c t - \tfrac{1}{2} n_s(t) \sin 4\pi f_c t \end{aligned} \tag{8.3-10}$$

The spectra of the second and third terms in Eq. (8.3-10) extend over the range $2f_c - f_M$ to $2f_c + f_M$ and are outside the baseband filter. The output noise is, therefore,

$$n_o(t) = \tfrac{1}{2} n_c(t) \tag{8.3-11}$$

The spectral density of $n_o(t)$ is then $G_{no} = \frac{1}{4} G_{n_c} = \frac{1}{4}(\eta/2) = \eta/8$. Hence as before, as shown in Fig. 8.3-2*c*, the spectral density $G_{no}$ is $\eta/8$ over the range $-f_M$ to $f_M$, and the total noise is again $N_o = \eta f_M/4$.

### Calculation of Signal-to-Noise Ratio (SNR)

Finally we may calculate, using Eqs. (8.3-6) and (8.3-7), the *signal-to-noise ratio* at the output, $S_o/N_o$. We have

$$\frac{S_o}{N_o} = \frac{S_i/4}{\eta f_M/4} = \frac{S_i}{\eta f_M} \tag{8.3-12}$$

The importance of $S_o/N_o$ is that it serves as a *figure of merit* of the performance of a communication system. Certainly, as $S_o/N_o$ increases, it becomes easier to distinguish and to reproduce the modulating signal without error or confusion. If a

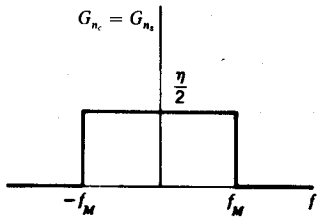



**Figure 8.3-3** Power spectral densities of $G_{n_c}$ and $G_{n_s}$.



system of communication allows the use of more than a single type of demodulator (say, synchronous or nonsynchronous), that ratio $S_o/N_o$ will serve as a figure of merit with which to compare demodulators.

We observe from Eq. (8.3-12) that to increase the output signal-to-noise power ratio, we can increase the transmitted signal power, restrict the baseband frequency range, or make the receiver quieter.

## 8.4 DOUBLE-SIDEBAND SUPPRESSED CARRIER (DSB-SC)

When a baseband signal of frequency range $f_M$ is transmitted over a DSB-SC system, the bandwidth of the carrier filter must be $2f_M$ rather than $f_M$. Thus, input noise in the frequency range $f_c - f_M$ to $f_c + f_M$ will contribute to the output noise, rather than only in the range $f_c$ to $f_c + f_M$ as in the SSB case.

### Calculation of Noise Power

This situation is illustrated in Fig. 8.4-1*a*, which shows the spectral density $G_{n1}(f)$ of the white input noise after the IF filter. This noise is multiplied by $\cos \omega_c t$. The multiplication results in a frequency shift by $\pm f_c$ and a reduction of power in the power spectral density of the noise by a factor of 4. Thus, the noise in region *d* of Fig. 8.4-1*a* shifts to regions *d* shown in Fig. 8.4-1*b*. Similarly regions *a*, *b*, and *c* of Fig. 8.4-1*a* are translated by $\pm f_c$ and are also attenuated by 4 as shown in

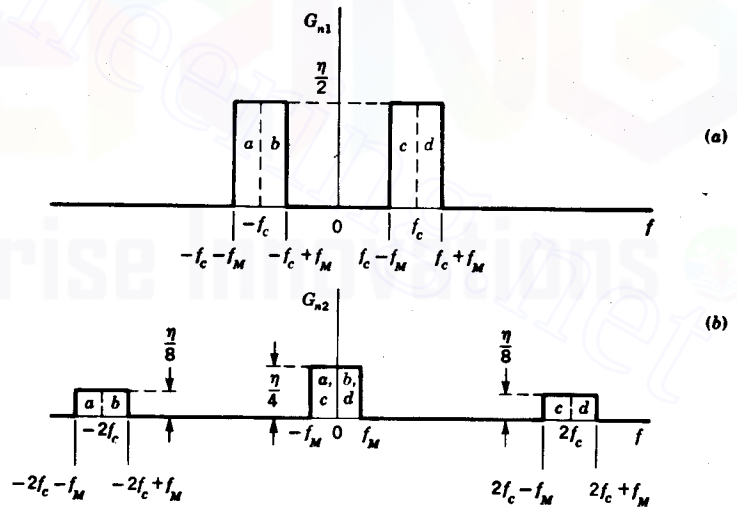



**Figure 8.4-1** Spectral densities of noise in DSB demodulation. (*a*) Density $G_{n1}$ of noise at output of IF filter. (*b*) Density $G_{n2}$ of noise output of baseband filter.

Fig. 8.4-1*b*. Note that the noise-power spectral density in the region between $-f_M$ and $+f_M$ is $\eta/4$, while the noise density in the SSB case, as shown in Fig. 8.3-2*c* is only $\eta/8$. Hence the output noise power in twice as large as the output noise power for SSB given in Eq. (8.3-5). The output noise for DSB after baseband filtering is therefore

$$N_o = \frac{\eta}{4}(2f_M) = \frac{\eta f_M}{2} \tag{8.4-1}$$

### Calculation of Signal Power

We might imagine that for equal received powers, the ratio $S_o/N_o$ for DSB would be only half the corresponding ratio for SSB. We shall now see that such is not the case, and that the ratio $S_o/N_o$ is the *same* in the two cases. Again, without loss in generality, let us assume a sinusoidal baseband signal of frequency $f_m \leq f_M$. To keep the received power the same as in the SSB case, that is, $S_i = A^2/2$, we write

$$\begin{aligned} s_i(t) &= \sqrt{2}\, A \cos 2\pi f_m t \cos 2\pi f_c t \\ &= \frac{A}{\sqrt{2}} \cos [2\pi(f_c + f_m)t] + \frac{A}{\sqrt{2}} \cos [2\pi(f_c - f_m)t] \end{aligned} \tag{8.4-2}$$

The received power is then

$$S_i = \frac{1}{2}\left(\frac{A}{\sqrt{2}}\right)^2 + \frac{1}{2}\left(\frac{A}{\sqrt{2}}\right)^2 = \frac{A^2}{2} \tag{8.4-3}$$

as in Eq. (8.3-4).

In the demodulator (multiplier), $s_i(t)$ is multiplied by $\cos \omega_c t$. The upper-sideband term in Eq. (8.4-2) yields a signal within the passband of the baseband filter given by

$$s_o'(t) = \frac{A}{2\sqrt{2}} \cos 2\pi f_m t \tag{8.4-4}$$

The lower-sideband term of Eq. (8.4-2) yields

$$s_o''(t) = \frac{A}{2\sqrt{2}} \cos 2\pi f_m t \tag{8.4-5}$$

Observe, most particularly in Eqs. (8.4-4) and (8.4-5), that $s_o'(t)$ and $s_o''(t)$ are in phase and that hence the output signal is

$$s_o(t) = s'(t) + s''(t) = \frac{A}{\sqrt{2}} \cos 2\pi f_m t \tag{8.4-6}$$

which has a power

$$S_o = \frac{A^2}{4} = \frac{S_i}{2} \tag{8.4-7}$$



rather than $S_o = A^2/8 = S_i/4$ as in Eq. (8.3-5) for the SSB case. Thus we see that when a received signal of fixed power is split into two sideband components each of half power, as in DSB, rather than being left in a single sideband, the output signal power increases by a factor of 2. This increase results from the fact that the contributions from each sideband yield output signals which are in phase. A doubling in amplitude causes a fourfold increase in power. This fourfold increase, due to the inphase addition of $s_o'$ and $s_o''$, is in part undone by the need to split the input power into two half-power sidebands. Thus the overall improvement in output signal power is by a factor of 2.

On the other hand, the noise outputs due to noise spectral components symmetrically located with respect to the carrier are uncorrelated with one another. The two resultant noise spectral components in the output, although of the same frequency, are uncorrelated. Hence the combination of the two yields a power which is the sum of the two powers individually, not larger than the sum, as is the case with the signal.

### Calculation of Signal-to-Noise Ratio

Returning now to the calculation of signal-to-noise ratio for DSB-SC, we find from Eqs. (8.4-1) and (8.4-7) that

$$\frac{S_o}{N_o} = \frac{S_i}{\eta f_M} \tag{8.4-8}$$

exactly as for SSB-SC.

### Arbitrary Modulating Signal

In the discussion in the present section concerning DSB we have assumed that the baseband signal waveform is sinusoidal. As pointed out in Sec. 8.3, this assumption causes no loss of generality because of the linearity of the demodulation. Nonetheless it is often convenient to have an expression for the power of a DSB signal in terms of the arbitrary waveform $m(t)$ of the baseband modulating signal. Hence let the received signal be

$$s_i(t) = m(t) \cos 2\pi f_c t \tag{8.4-9}$$

The power of $s_i(t)$ is

$$S_i \equiv \overline{s_i^2(t)} = \overline{m^2(t) \cos^2 2\pi f_c t} = \tfrac{1}{2}\overline{m^2(t)} + \tfrac{1}{2}\overline{m^2(t) \cos (4\pi f_c t)} \tag{8.4-10}$$

Now $m(t)$ can always be represented as a sum of sinusoidal spectral components. [Of interest, albeit of no special relevance in the present discussion, is the fact that if $m(t)$ is bandlimited to $f_M$, $m^2(t)$ is bandlimited to $2f_M$. See Prob. 8.4-1.] Hence $m^2(t) \cos 4\pi f_c t$ consists of a sum of sinusoidal waveforms in the frequency range $2f_c \pm 2f_M$. The average value of such a sum is zero, and we therefore have

$$S_i \equiv \overline{s_i^2(t)} = \tfrac{1}{2}\overline{m^2(t)} \tag{8.4-11}$$

When the signal $s_i(t)$ in Eq. (8.4-9) is demodulated by multiplication by $\cos 2\pi f_c t$, and the product passed through the baseband filter, the output is $s_o(t) = m(t)/2$. The output signal power is

$$S_o = \frac{\overline{m^2(t)}}{4} \tag{8.4-12}$$

so that, from Eqs. (8.4-11) and (8.4-12),

$$S_o = \frac{S_i}{2} \tag{8.4-13}$$

that is, the same result as given in Eq. (8.4-7) for an assumed single sinusoidal modulating signal.

### Use of Quadrature Noise Components to Calculate $N_o$

It is again interesting to calculate $N_o$ using the transformation of Eq. (7.11-2):

$$n(t) = n_c(t) \cos 2\pi f_c t - n_s(t) \sin 2\pi f_c t \tag{8.4-14}$$

The power spectral density of $n_c(t)$ and $n_s(t)$ are [see Eqs. (7.12-7) and (7.12-8)]

$$G_{n_c}(f) = G_{n_s}(f) = G_{n1}(f_c + f) + G_{n1}(f_c - f) \tag{8.4-15}$$

In the frequency range $|f| \leq f_M$, $G_{n1}(f_c + f) = G_{n1}(f_c - f) = \eta/2$. Thus

$$G_{n_c}(f) = G_{n_s}(f) = \eta \qquad |f| \leq f_M \tag{8.4-16}$$

(This result was also derived in Example 7.12-1.)

The result of multiplying $n(t)$ by $\cos 2\pi f_c t$ yields

$$n(t) \cos 2\pi f_c t = \tfrac{1}{2} n_c(t) + \tfrac{1}{2} n_c(t) \cos 4\pi f_c t - \tfrac{1}{2} n_s(t) \sin 4\pi f_c t \tag{8.4-17}$$

Baseband filtering eliminates the second and third terms, leaving

$$n_o(t) = \tfrac{1}{2} n_c(t) \tag{8.4-18}$$

The power spectral density of $n_o(t)$ is then

$$G_{no}(f) = \frac{1}{4} G_{n_c}(f) = \frac{\eta}{4} \qquad -f_M \leq f \leq f_M \tag{8.4-19}$$

The output noise power $N_o$ is, therefore,

$$N_o = \frac{\eta}{4} 2f_M = \frac{\eta f_M}{2} \tag{8.4-20}$$

This result is, of course, identical with Eq. (8.4-1), which was obtained by considering directly the effect on a noise spectral component of a multiplication by $\cos 2\pi f_c t$.







## 8.5 DOUBLE SIDEBAND WITH CARRIER

Let us now consider the case where a carrier accompanies the double-sideband signal. Demodulation is achieved synchronously as in SSB-SC and DSB-SC. The carrier is used as a *transmitted reference* to obtain the reference signal $\cos \omega_c t$ (see Prob. 8.5-1). We note that the carrier increases the total input-signal power but makes no contribution to the output-signal power. Equation (8.4-8) applies directly to this case, provided only that we replace $S_i$ by $S_i^{(SB)}$, where $S_i^{(SB)}$ is the power in the sidebands alone. Then

$$\frac{S_o}{N_o} = \frac{S_i^{(SB)}}{\eta f_M} \tag{8.5-1}$$

Suppose that the received signal is

$$\begin{aligned} s_i(t) &= A[1 + m(t)] \cos 2\pi f_c t \\ &= A \cos 2\pi f_c t + Am(t) \cos 2\pi f_c t \end{aligned} \tag{8.5-2}$$

where $m(t)$ is the baseband signal which amplitude-modulates the carrier $A \cos 2\pi f_c t$. The carrier power is $A^2/2$. The sidebands are contained in the term $Am(t) \cos 2\pi f_c t$. The power associated with this term is $(A^2/2)\overline{m^2(t)}$, where $\overline{m^2(t)}$ is the time average of the square of the modulating waveform. We then have that the total input power $S_i$ is given by

$$S_i = \frac{A^2}{2} + S_i^{(SB)} = \frac{A^2}{2}[1 + \overline{m^2(t)}] \tag{8.5-3}$$

Eliminating $A^2$, we have

$$S_i^{(SB)} = \frac{\overline{m^2(t)}}{1 + \overline{m^2(t)}} S_i \tag{8.5-4}$$

or, with Eq. (8.5-1),

$$\frac{S_o}{N_o} = \frac{\overline{m^2(t)}}{1 + \overline{m^2(t)}} \frac{S_i}{\eta f_M} \tag{8.5-5}$$

In terms of the carrier power $P_c \equiv A^2/2$, we get, from Eqs. (8.5-3) and (8.5-5), that

$$\frac{S_o}{N_o} = \overline{m^2(t)} \frac{P_c}{\eta f_M} \tag{8.5-6}$$

If the modulation is sinusoidal, with $m(t) = m \cos 2\pi f_m t$ ($m$ a constant), then

$$s_i(t) = A(1 + m \cos 2\pi f_m t) \cos 2\pi f_c t \tag{8.5-7}$$

In this case $\overline{m^2(t)} = m^2/2$ and

$$\frac{S_o}{N_o} = \frac{m^2}{2 + m^2} \frac{S_i}{\eta f_M} \tag{8.5-8}$$

When the carrier is transmitted only to synchronize the local demodulator waveform $\cos 2\pi f_c t$, relatively little carrier power need be transmitted. In this case $m \gg 1$, $m^2/(2 + m^2) \cong 1$, and the signal-to-noise ratio is not greatly reduced by the presence of the carrier. On the other hand, when envelope demodulation is used (Sec. 3.4), it is required that $m \leq 1$. When $m = 1$, the carrier is 100 percent modulated. In this case $m^2/(2 + m^2) = \frac{1}{3}$, so that of the power transmitted, only one-third is in the sidebands which contribute to signal power output.

### A Figure of Merit

We observe that in each demodulation system considered so far, the ratio $S_i/\eta f_M$ appeared in the expression for output SNR [see Eqs. (8.3-12), (8.4-8), and (8.5-5)]. This ratio is the output signal power $S_i$ divided by the product $\eta f_M$. To give the product $\eta f_M$ some physical significance, we consider it to be the noise power $N_M$ at the input, measured in a frequency band equal to the *baseband frequency*. Thus

$$N_M \equiv \frac{\eta}{2} 2f_M = \eta f_M \tag{8.5-9}$$

The ratio $S_i/\eta f_M$ is, therefore, often referred to as the *input signal-to-noise ratio* $S_i/N_M$. It needs to be kept in mind that $N_M$ is the noise power transmitted through the IF filter only when the IF filter bandwidth is $f_M$. Thus $N_M$ is the true input noise power only in the case of single sideband.

For the purpose of comparing systems, we introduce the *figure of merit* $\gamma$, defined by

$$\gamma \equiv \frac{S_o/N_o}{S_i/N_M} \tag{8.5-10}$$

The results given above may now be summarized as follows:

$$\gamma = \begin{cases} 1 & \text{SSB-SC} \quad (8.5\text{-}11) \\ 1 & \text{DSB-SC} \quad (8.5\text{-}12) \\ \dfrac{\overline{m^2(t)}}{1 + \overline{m^2(t)}} & \text{DSB} \quad (8.5\text{-}13) \\ \dfrac{m^2}{2 + m^2} & \text{DSB with sinusoidal modulation} \quad (8.5\text{-}14) \end{cases}$$

A point of interest in connection with *double-sideband* synchronous demodulation is that, for the purpose of suppressing output-noise power, the carrier filter of Fig. 8.3-1 is not necessary. A noise spectral component at the input which lies outside the range $f_c \pm f_M$ will, after multiplication in the demodulator, lie outside the passband of the baseband filter. On the other hand, if the carrier filter is eliminated, the magnitude of the noise signal which reaches the modulator may be large enough to overload the active devices used in the demodulator. Hence,







such carrier filters are normally included, but the purpose is *overload suppression* rather than *noise suppression*. In single sideband, of course, the situation is different, and the carrier filter does indeed suppress noise.

## 8.6 SQUARE-LAW DEMODULATOR

We saw in Sec. 3.6 that a double-sideband signal with carrier may be demodulated by passing the signal through a network whose input-output characteristic is not linear. Such nonlinear demodulation has the advantage, over the linear synchronous demodulation methods, that a synchronous local carrier need not be obtained. This eliminates the rather costly synchronizing circuits. In this section we discuss the performance and determine the output SNR of a nonlinear demodulator which uses a network whose output signal $y$ (voltage or current) is related to the input signal $x$ (voltage or current) by

$$y = \lambda x^2 \tag{8.6-1}$$

in which $\lambda$ is a constant. As shown in Fig. 8.6-1, this nonlinear network, which constitutes the demodulator, is preceded by a bandpass IF filter of bandwidth $2f_M$ and is followed by a baseband low-pass filter of bandwidth $f_M$.

We discuss this square-law demodulator in part for its own intrinsic interest, but also because it exhibits an important characteristic which is not displayed by the linear synchronous demodulators. We have previously adopted the quantity $\gamma \equiv (S_o/N_o)/(S_i/N_M)$ as a figure of merit for the performance of demodulators in the presence of noise. We observed [Eqs. (8.5-11) to (8.5-14)] that this figure of merit is not a function of the input signal-to-noise ratio $S_i/N_M$. Therefore, if the input $S_i/N_M$ decreases, say, by a factor $\alpha$, the output $S_o/N_o$ will also decrease by $\alpha$. The nonlinear demodulator also has a range where the figure of merit $\gamma$ is inde-

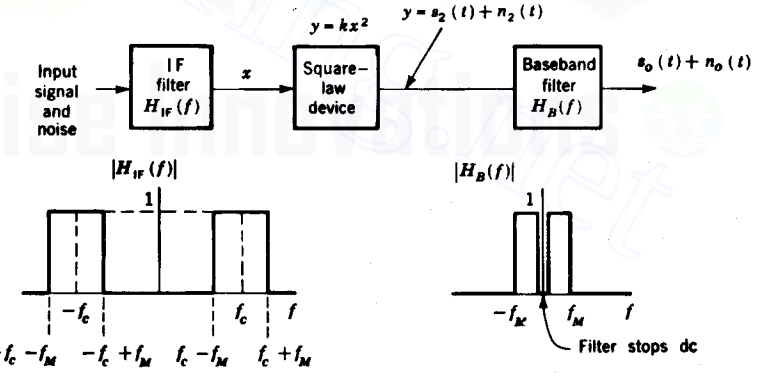



**Figure 8.6-1** The square-law AM demodulator.

Equation (8.6-22) is plotted (solid plot) in Fig. 8.6-4 with the variables expressed in terms of their decibel (dB) equivalents, i.e., the abscissa is marked off in units of 10 log $(P_c/N_M)$. Above threshold, when $P_c/N_M$ is very large, Eq. (8.6-22) becomes

$$\frac{S_o}{N_o} = \overline{m^2(t)}\,\frac{P_c}{N_M} \tag{8.6-23}$$

Below threshold, when $P_c/N_M \ll 1$, Eq. (8.6-22) becomes

$$\frac{S_o}{N_o} = \frac{4}{3}\,\overline{m^2(t)}\left(\frac{P_c}{N_M}\right)^2 \tag{8.6-24}$$

For comparison, Eqs. (8.6-23) and (8.6-24) have also been plotted in Fig. 8.6-4. We observe the occurrence of a threshold in that, as $P_c/N_M$ decreases, the demodulator performance curve falls progressively further away from the straight-line plot corresponding to $P_c/N_M$ very large. The *threshold* point is

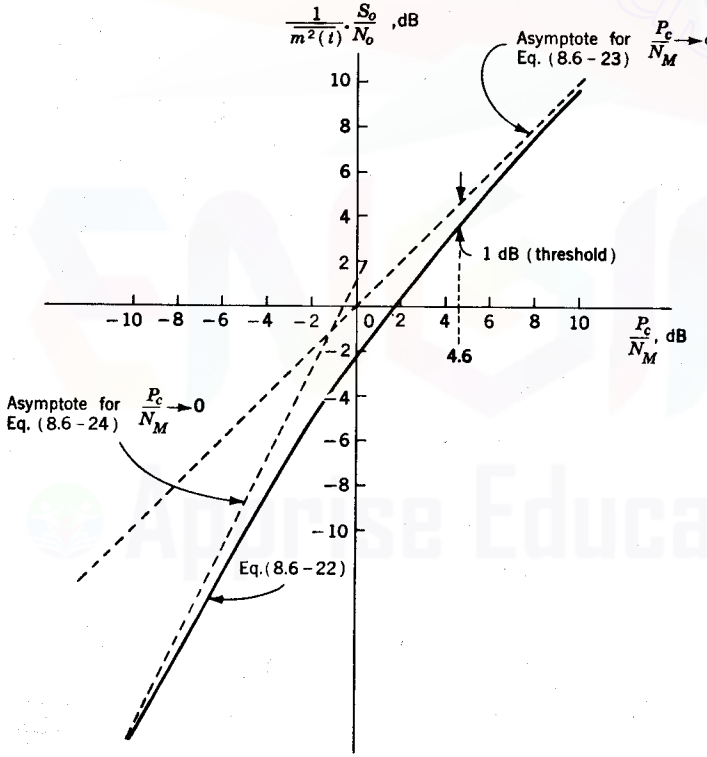



**Figure 8.6-4** Performance of a square-law demodulator illustrating the phenomenon of threshold.



chosen arbitrarily to be the point at which the performance curve falls away by 1 dB as shown. On this basis it turns out that the threshold occurs when $P_c/N_M = 4.6$ dB or when $P_c = 2.9N_M$.

## 8.7 THE ENVELOPE DEMODULATOR

We again consider an AM signal with modulation $|m(t)| < 1$. To demodulate this DSB signal we shall use a network which accepts the modulated carrier and provides an output which follows the waveform of the *envelope* of the carrier. The diode demodulator of Sec. 3.6 is a physical circuit which performs the required operation to a good approximation. As usual, as in Fig. 8.3-1, the demodulator is preceded by a bandpass filter with center frequency $f_c$ and bandwidth $2f_M$, and is followed by a low-pass baseband filter of bandwidth $f_M$.

It is convenient in the present discussion to use the noise representation given in Eq. (7.11-2):

$$n(t) = n_c(t)\cos\omega_c t - n_s(t)\sin\omega_c t \tag{8.7-1}$$

If the noise $n(t)$ has a power spectral density $\eta/2$ in the range $|f - f_c| \le f_M$ and is zero elsewhere as shown in Fig. 8.4-1, then, as explained in Sec. 7.12, both $n_c(t)$ and $n_s(t)$ have the spectral density $\eta$ in the frequency range $-f_M$ to $f_M$.

At the demodulator input, the input signal plus noise is

$$s_1(t) + n_1(t) = A[1 + m(t)]\cos\omega_c t + n_c(t)\cos\omega_c t - n_s(t)\sin\omega_c t \tag{8.7-2a}$$

$$= \{A[1 + m(t)] + n_c(t)\}\cos\omega_c t - n_s(t)\sin\omega_c t \tag{8.7-2b}$$

where $A$ is the carrier amplitude and $m(t)$ the modulation. In a phasor diagram, the first term of Eq. (8.7-2*b*) would be represented by a phasor of amplitude $A[1 + m(t)] + n_c(t)$, while the second term would be represented by a phasor perpendicular to the first and of amplitude $n_s(t)$. The phasor sum of the two terms is then represented by a phasor of amplitude equal to the square root of the sum of the squares of the amplitudes of the two terms. Thus, the output signal plus noise just prior to baseband filtering is the envelope (phasor sum):

$$s_2(t) + n_2(t) = \{(A[1 + m(t)] + n_c(t))^2 + n_s^2(t)\}^{1/2} \tag{8.7-3a}$$

$$= \{A^2[1 + m(t)]^2 + 2A[1 + m(t)]n_c(t) + n_c^2(t) + n_s^2(t)\}^{1/2} \tag{8.7-3b}$$

We should now like to make the simplification in Eq. (8.7-3*b*) that would be allowed if we might assume that both $|n_c(t)|$ and $|n_s(t)|$ were much smaller than the carrier amplitude $A$. The difficulty is that $n_c$ and $n_s$ are noise "waveforms" for which an explicit time function may not be written and which are described only in terms of the statistical distributions of their instantaneous amplitudes. No matter how large $A$ and how small the values of the standard deviation of $n_c(t)$ or $n_s(t)$, there is always a finite probability that $|n_c(t)|$, $|n_s(t)|$, or both, will be com-

parable to, or even larger than, $A$. On the other hand, if the standard deviations of $n_c(t)$ and $n_s(t)$ are much smaller than $A$, the *likelihood* that $n_c$ or $n_s$ will approach or exceed $A$ is rather small. For example, since $n_c$ and $n_s$ have gaussian distributions, the probability that $n_c(t)$ is greater than twice the standard deviation is only 0.045, and only 0.00006 that it exceeds 4 times its standard deviation. Hence, if $\sqrt{\overline{n_c^2(t)}} \ll A$ and we assume that $|n_c(t)| \ll A$, the assumption is *usually* valid.

Assuming then that $|n_c(t)| \ll A$ and $|n_s(t)| \ll A$, the "noise-noise" terms $n_c^2(t)$ and $n_s^2(t)$ may be dropped, leaving us with the approximation

$$s_2(t) + n_2(t) \approx \{A^2[1 + m(t)]^2 + 2A[1 + m(t)]n_c(t)\}^{1/2} \tag{8.7-4a}$$

$$= A[1 + m(t)]\left\{1 + \frac{2n_c(t)}{A[1 + m(t)]}\right\}^{1/2} \tag{8.7-4b}$$

Using now the further approximation that $(1 + x)^{1/2} \approx 1 + x/2$ for small $x$ we have finally that

$$s_2(t) + n_2(t) \approx A[1 + m(t)] + n_c(t) \tag{8.7-5}$$

The output-signal power measured after the baseband filter, and neglecting dc terms, is $S_o = A^2\overline{m^2(t)}$. Since the spectral density of $n_c(t) = \eta$, the output-noise power after baseband filtering is $N_o = 2\eta f_M$. Again using the symbol $N_M (\equiv \eta f_M)$ to stand for the noise power at the input in the baseband range $f_M$, and using Eq. (8.5-3), we find that

$$\gamma \equiv \frac{S_o/N_o}{S_i/N_M} = \frac{\overline{m^2(t)}}{1 + \overline{m^2(t)}} \tag{8.7-6}$$

The result is the same as given in Eq. (8.5-13) for synchronous demodulation. To make a comparison with the square-law demodulator, we assume $\overline{m^2(t)} \ll 1$. In this case, as before, $S_i \cong P_c$, and Eq. (8.7-6) reduces to Eq. (8.5-6). Hence we have the important result that *above threshold* the synchronous demodulator, the square-law demodulator, and the envelope demodulator all perform equally well, provided $\overline{m^2(t)} \ll 1$.

### Threshold

Like the square-law demodulator, the envelope demodulator exhibits a threshold. As the input signal-to-noise ratio decreases, a point is reached where the signal-to-noise ratio at the output decreases more rapidly than at the input. The calculation of signal-to-noise ratio is quite complex, and we shall therefore be content to simply state the result[1] that for $S_i/N_M \ll 1$, and $\overline{m^2(t)} \ll 1$

$$\frac{S_o}{N_o} = \frac{\overline{m^2(t)}}{1.1}\left(\frac{S_i}{N_M}\right)^2 \tag{8.7-7}$$

Equation (8.7-7) obviously indicates a poorer performance than indicated by Eq. (8.7-6), which applies above threshold.



Since both square-law demodulation and envelope demodulation exhibit a threshold, a comparison is of interest. We had assumed in square-law demodulation that $\overline{m^2(t)} \ll 1$. Then, as noted above, $S_i \cong A^2/2 = P_c$ the carrier power, and Eq. (8.7-7) becomes

$$\frac{S_o}{N_o} = \frac{\overline{m^2(t)}}{1.1}\left(\frac{P_c}{N_M}\right)^2 \tag{8.7-8}$$

which is to be compared with Eq. (8.6-24) giving $S_o/N_o$ below threshold for the square-law demodulator.

The comparison indicates that, below threshold, the square-law demodulator performs better than the envelope detector. Actually this advantage of the square-law demodulator is of dubious value. Generally, when a demodulator is operated below threshold to any appreciable extent, the performance may be so poor as to be nearly useless. What is of greater significance is that the comparison suggests that the threshold in square-law demodulation is lower than the threshold in envelope demodulation. Therefore a square-law demodulator will operate above threshold on a weaker signal than will an envelope demodulator.

In summary, on strong signals all demodulators work equally well except that the square-law demodulator requires that $\overline{m^2(t)} \ll 1$ to avoid baseband-signal distortion. On weak signals, synchronous demodulation does best since it exhibits no threshold. When synchronous demodulation is not feasible, square-law demodulation does better than envelope demodulation. It is also interesting to note that voice signals require a 40-dB output signal-to-noise ratio for high quality. In this case both the linear-envelope detector and the square-law detector operate above threshold.

## REFERENCE

1. Davenport, W., and W. Root: "Random Signals and Noise," McGraw-Hill Book Company, New York, 1958.

## PROBLEMS

**8.1-1.** (*a*) A superheterodyne receiver using an IF frequency of 455 kHz is tuned to 650 kHz. It is found that the receiver picks up a transmission from a transmitter whose carrier frequency is 1560 kHz. Suggest a reason for this undesired reception and suggest a remedy. (These frequencies, 650 kHz and 1560 kHz, are referred to as *image frequencies*. Why?)

**8.3-1.** Let $g(t)$ be a waveform characterized by a power spectral density $G(f)$. Assume $G(f) = 0$ for $|f| \geq f_1$. Show that the time-average value of $g(t) \cos 2\pi f_c t$ is zero if $f_c > f_1$.

**8.3-2.** As noted in Sec. 3.10, if $m(t)$ is an arbitrary baseband waveform, a received SSB signal may be written $s_i(t) = m(t) \cos 2\pi f_c t + \hat{m}(t) \sin 2\pi f_c t$. Here $\hat{m}(t)$ is derived from $m(t)$ by shifting by 90° the phase of every spectral component in $m(t)$.

(*a*) Show that $m(t)$ and $\hat{m}(t)$ have the same power spectral densities and that $\overline{m^2(t)} = \overline{\hat{m}^2(t)}$.

(*b*) Show that if $m(t)$ has a spectrum which extends from zero frequency to a maximum frequency $f_M$, then $m^2(t)$, $\hat{m}^2(t)$, and $m(t)\hat{m}(t)$ all have spectra which extend from zero frequency to $2f_M$.

(*c*) Show that the normalized power $S_i$ of $s_i(t)$ is $\overline{m^2(t)} = \overline{\hat{m}^2(t)}$. (*Hint:* Use the results of Prob. 8.3-1.)

(*d*) Calculate the normalized power $S_o$ of the demodulated SSB signal, i.e., the signal $s_i(t)$ multiplied by $\cos 2\pi f_c t$ and then passed through a baseband filter. Show that $S_o = \overline{m^2(t)}/4$ and hence that $S_o/S_i = \frac{1}{4}$. [*Note:* This problem establishes more generally the result given in Eq. (8.3-6) which was derived on the basis of the assumption that the modulating waveform is a sinusoid.]

**8.3-3.** Prove that Eq. (8.3-11) is correct by sketching the power spectral density of Eq. (8.3-10).

**8.3-4.** A baseband signal $m(t)$ is transmitted using SSB as in Prob. 8.3-2. Assume that the power spectral density of $m(t)$ is

$$G_m(f) = \begin{cases} \frac{\eta_m}{2}\frac{|f|}{f_M} & |f| < f_M \\ 0 & |f| > f_M \end{cases}$$

Find:

(*a*) The input signal power.

(*b*) The output signal power.

(*c*) If white gaussian noise with power spectral density $\eta/2$ is added to the SSB signal, find the output SNR. The baseband filter cuts off at $f = f_M$.

**8.3-5.** A received SSB signal has a power spectrum which extends over the frequency range from $f_c = 1$ MHz to $f_c + f_M = 1.003$ MHz. The signal is accompanied by noise with uniform power spectral density $10^{-9}$ watt/Hz.

(*a*) The noise $n(t)$ is expressed as $n(t) = n_c(t)\cos 2\pi f_c t - n_s(t)\sin 2\pi f_c t$. Find the power spectral densities of the quadrature components $n_c(t)$ and $n_s(t)$ of the noise in the spectral range specified as $f_c \le f \le f_c + f_M$.

(*b*) The signal plus its accompanying noise is multiplied by a local carrier $\cos 2\pi f_c t$. Plot the power spectral density of the noise at the output of the multiplier.

(*c*) The signal plus noise, after multiplication, is passed through a baseband filter and an amplifier which provides a voltage gain of 10. Plot the power spectral density of the noise at the amplifier output, and calculate the total noise output power.

**8.4-1.** Show that $m^2(t)$ in Eq. (8.4-10) is bandlimited to $2f_M$.

**8.4-2.** Repeat Prob. 8.3-4 if DSB rather than SSB modulation is employed.

**8.4-3.** A carrier of amplitude 10 mV at $f_c$ is 50 percent amplitude-modulated by a sinusoidal waveform of frequency 750 Hz. It is accompanied by thermal noise of two-sided power spectral density

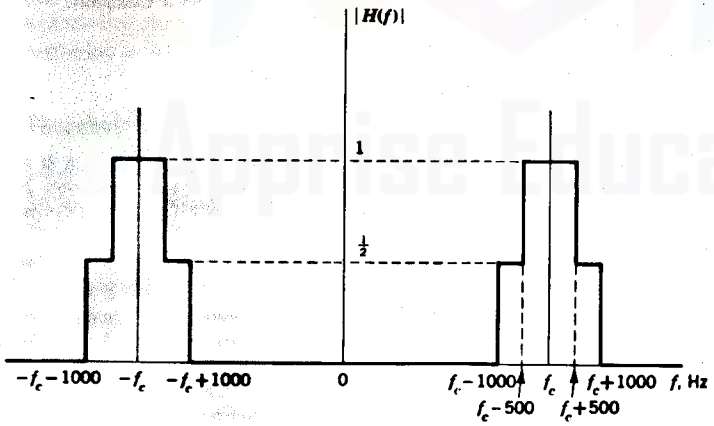



**Figure P8.4-3**



$\eta/2 = 10^{-3}$ watt/Hz. The signal plus noise is passed through the filter shown. The signal is demodulated by multiplication with a local carrier of amplitude 1 volt.

(*a*) Find the output signal power.

(*b*) Find the output noise power.

**8.5-1.** The signal $[\varepsilon + m(t)]\cos \omega_c t$ is synchronously detected. The reference signal $\cos \omega_c t$ used to multiply the incoming signal, is obtained by passing the input signal through a narrowband filter of bandwidth $B$, as shown in Fig. P8.5-1.

(*a*) Calculate $S_i$.

(*b*) Calculate $v_R(t)$ due to the input signal alone. Calculate the noise power accompanying $v_R(t)$.

(*c*) Comment on the effect of the noisy reference on the output SNR.

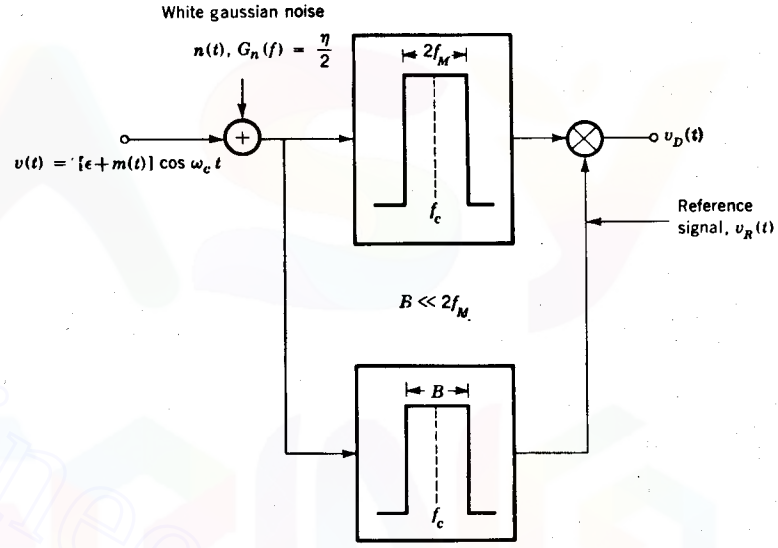



**Figure P8.5-1**

**8.5-2.** Verify Eqs. (8.5-5) and (8.5-8).

**8.5-3.** (*a*) Show that the output SNR of a DSB-SC signal which is synchronously detected is independent of the IF bandwidth; i.e., Eq. (8.4-8) is independent of the IF bandwidth.

(*b*) Show that the output SNR of an SSB signal which is synchronously detected is dependent on the IF bandwidth. To do this, consider an IF bandwidth which extends from $f_0 - B$ to $f_0 + f_M$, where $f_M > B > 0$. Calculate the output SNR using this bandwidth.

**8.5-4.** In the received amplitude-modulated signal $s_i(t) = A[1 + m(t)]\cos 2\pi f_c t$, $m(t)$ has the power spectral density $G_m(f)$ specified in Prob. 8.3-4. The received signal is accompanied by noise of power spectral density $\eta/2$. Calculate the output signal-to-noise ratio.

**8.6-1.** Verify Eq. (8.6-4) by showing graphically that all the neglected terms have spectra falling outside the range $|f| \le f_M$.

**8.6-2.** Given $2K + 1$ spectral components of a noise waveform spaced by intervals $\Delta f$. Show that the number of pairs of components separated by a frequency $\rho\,\Delta f$ is $p = 2K - 1 - \rho$ [i.e., verify the discussion leading to Eq. (8.6-15)].

**8.6-3.** In a DSB transmission, a carrier of frequency 1 MHz and of amplitude 2 volts is amplitude-modulated to the extent of 10 percent by a sinusoidal baseband signal of frequency 5 kHz. The signal is corrupted by white noise of two-sided power spectral density $10^{-6}$ watt/Hz. The demodulator is a

device whose input/output characteristic is given by $v_o = 3v_i^2$, where $v_o$ and $v_i$ are respectively the output and input voltage. The IF filter, before the demodulator, has a rectangular transfer characteristic of unity gain and 10-kHz bandwidth. By error, the IF filter is tuned so that its center frequency is at 1.002 MHz.

(*a*) Calculate the signal waveform at the demodulator output and calculate its normalized power.

(*b*) Calculate the noise power at the demodulator output and the signal-to-noise ratio.

**8.6-4.** Plot $(1/\overline{m^2})\,(S_o/N_o)$ versus $P_c/N_M$ in Eq. (8.6-22) and show that the 1-dB threshold occurs when $P_c/N_M = 4.6$ dB.

**8.6-5.** Assume that $\overline{m^2(t)} = 0.1$ and that $S_o/N_o = 30$ dB. Find $P_c/N_M$ in Eq. (8.6-22). Are we above threshold?

**8.7-1.** A baseband signal $m(t)$ is superimposed as an amplitude modulation on a carrier in a DSB transmission from a transmitting station. The instantaneous amplitude of $m(t)$ has a probability density which falls off linearly from a maximum at $m = 0$ to zero at $m = 0.1$ volt. The spectrum of $m(t)$ extends over a frequency range from zero to 10 kHz. The level of the modulating waveform applied to the modulator is adjusted to provide for maximum allowable modulation. It is required that under this condition the total power supplied by the transmitter to its antenna be 10 kW. Assume that the antenna appears to the transmitter as a resistive load of 72 ohms.

(*a*) Write an expression for the voltage at the input to the antenna.

(*b*) At a receiver, tuned to pick up the transmission, the level of the carrier at the input to the diode demodulator is 3 volts. What is the maximum allowable power spectral density at the demodulator input if the signal-to-noise ratio at the receiver output is to be at least 20 dB?

**8.7-2.** Plot $(1/\overline{m^2})\,(S_o/N_o)$ versus $S_i/N_M$ for the envelope demodulator. Assume that $\overline{m^2} \ll 1$, and find the intersection of the above-threshold and the below-threshold asymptotes.





CHAPTER

# NINE

# NOISE IN FREQUENCY-MODULATION SYSTEMS

In this chapter we discuss the performance of frequency-modulation systems in the presence of additive noise. We shall show how, in an FM system, an improvement in output signal-to-noise power ratio can be made through the sacrifice of bandwidth.

## 9.1 AN FM DEMODULATOR

The receiving system of Fig. 8.1-1 may be used with an AM or FM signal. When used to recover a frequency-modulated signal, the AM demodulator is replaced by an FM demodulator such as the limiter-discriminator shown in Fig. 9.1-1.

### The Limiter

In an FM system, the baseband signal varies only the frequency of the carrier. Hence any amplitude variation of the carrier must be due to noise alone. The *limiter* is used to suppress such amplitude-variation noise. In a limiter, diodes, transistors, or other devices are used to construct a circuit in which the output voltage $v_1$ is related to the input voltage $v_i$ in the manner shown in Fig. 9.1-2*a*. The output follows the input only over a limited range. A cycle of the carrier is shown in Fig. 9.1-2*b* and the output waveform in Fig. 9.1-2*c*. In limiter operation

**10.11-1.** Find the differential equation of the piecewise-linear second-order PLL in the region in which $\pi/2 \leq \psi \leq 3\pi/2$.

**10.11-2.** Find the differential equation of the second-order PLL if the phase comparator has a sinusoidal characteristic.

**10.12-1.** Show that Fig. 10.12-1 represents the first-order PLL.

**10.12-2.** Plot $N/f_M$ versus $S_i/\eta f_M$ at threshold. This is called the *threshold hyperbola*. If $S_i/\eta f_M = 20$ dB and $f_M = 5$ kHz, find $N$ to produce threshold.

**10.14-1.** Show that the ratio $S_i/\eta f_M$ is unchanged when measured after the carrier filter or the bandpass filter shown in Fig. 10.13-1.

**10.14-2.** Show that for the FMFB, $\delta N \gg N_c$. Use Eq. (10.14-6) and plot $N_c/\delta N$ as a function of $S_i/\eta B_p$ with $[(1 + \alpha G_0)B_p]/\Delta f = 4$ (this corresponds to choosing $1 + \alpha G_0 = \beta$, which represents reasonably good design).

**10.14-3.** Find the threshold extension possible for $\beta = 3$ and 5, if $1 + \alpha G_0 = \beta$.

**10.15-1.** Refer to Fig. 10.15-3. Show that if, due to noise, $\tau_Q$ is momentarily displaced to $\tau = 3T_h/4$ or $T_h/4$, then when the disturbance subsides, $\tau_Q$ will return to $\tau_Q = T_h/2$.

**10.15-2.** Design a *time-to-voltage* converter having the characteristics that 20 past samples are used to compute $v_o$ and that all samples receive equal weight.

**10.16-1.** Show that if $H(s) = \alpha(1 + 1/s)$ and the PLL is adjusted to have a bandwidth $2Mf_d$, the PLL will be able to follow jitter occurring at the rate $f_d$ with peak deviations less than $\Delta f$.

**10.16-2.** Repeat Prob. 10.16-1 for the Costas loop shown in Fig. 10.16-2.



Downloaded From : www.EasyEngineering.net

CHAPTER

# ELEVEN

# DATA TRANSMISSION

A data transmission system using binary encoding transmits a sequence of binary digits, that is, 1's and 0's. These digits may be represented in a number of ways. For example, a 1 may be represented by a voltage $V$ held for a time $T$, while a zero is represented by a voltage $-V$ held for an equal time. In general the binary digits are encoded so that a 1 is represented by a signal $s_1(t)$ and a 0 by a signal $s_2(t)$, where $s_1(t)$ and $s_2(t)$ each have a duration $T$. The resulting signal may be transmitted directly or, as is more usually the case, used to modulate a carrier. The received signal is corrupted by noise, and hence there is a finite probability that the receiver will make an error in determining, within each time interval, whether a 1 or a 0 was transmitted.

In this chapter we make calculations of such error probabilities and discuss methods to minimize them. The discussion will lead us to the concept of the *matched filter* and *correlator*.

## 11.1 A BASEBAND SIGNAL RECEIVER

Consider that a binary-encoded signal consists of a time sequence of voltage levels $+V$ or $-V$. If there is a guard interval between the bits, the signal forms a sequence of positive and negative pulses. In either case there is no particular interest in preserving the waveform of the signal after reception. We are interested only in knowing within each bit interval whether the transmitted voltage was $+V$ or $-V$. With noise present, the received signal and noise together will yield sample values generally different from $\pm V$. In this case, what deduction shall we make from the sample value concerning the transmitted bit?



Downloaded From : www.EasyEngineering.net

Downloaded From : www.EasyEngineering.net

Suppose that the noise is gaussian and therefore the noise voltage has a probability density which is entirely symmetrical with respect to zero volts. Then the probability that the noise has increased the sample value is the same as the probability that the noise has decreased the sample value. It then seems entirely reasonable that we can do no better than to assume that if the sample value is positive the transmitted level was $+V$, and if the sample value is negative the transmitted level was $-V$. It is, of course, possible that at the sampling time the noise voltage may be of magnitude larger than $V$ and of a polarity opposite to the polarity assigned to the transmitted bit. In this case an error will be made as indicated in Fig. 11.1-1. Here the transmitted bit is represented by the voltage $+V$ which is sustained over an interval $T$ from $t_1$ to $t_2$. Noise has been superimposed on the level $+V$ so that the voltage $v$ represents the received signal and noise. If now the sampling should happen to take place at a time $t = t_1 + \Delta t$, an error will have been made.

We can reduce the probability of error by processing the received signal plus noise in such a manner that we are then able to find a sample time where the sample voltage due to the signal is emphasized relative to the sample voltage due to the noise. Such a processer (receiver) is shown in Fig. 11.1-2. The signal input during a bit interval is indicated. As a matter of convenience we have set $t = 0$ at the beginning of the interval. The waveform of the signal $s(t)$ before $t = 0$ and after $t = T$ has not been indicated since, as will appear, the operation of the receiver during each bit interval is independent of the waveform during past and future bit intervals.

The signal $s(t)$ with added white gaussian noise $n(t)$ of power spectral density $\eta/2$ is presented to an integrator. At time $t = 0+$ we require that capacitor $C$ be uncharged. Such a discharged condition may be ensured by a brief closing of switch $SW_1$ at time $t = 0-$, thus relieving $C$ of any charge it may have acquired during the previous interval. The sample is taken at the output of the integrator by closing this sampling switch $SW_2$. This sample is taken at the end of the bit interval, at $t = T$. The signal processing indicated in Fig. 11.1-2 is described by the phrase *integrate and dump*, the term *dump* referring to the abrupt discharge of the capacitor after each sampling.

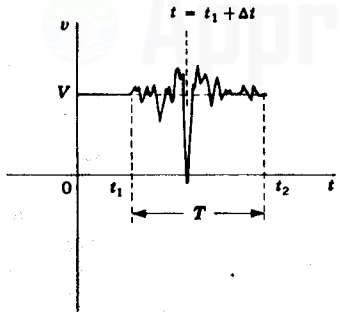



**Figure 11.1-1** Illustration that noise may cause an error in the determination of a transmitted voltage level.

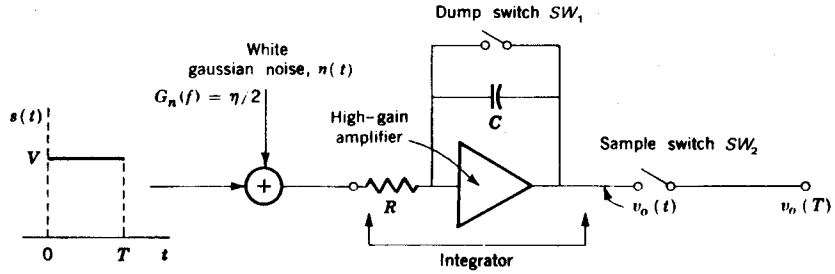



**Figure 11.1-2** A receiver for a binary coded signal.

## Peak Signal to RMS Noise Output Voltage Ratio

The integrator yields an output which is the integral of its input multiplied by $1/RC$. Using $\tau = RC$, we have

$$v_o(T) = \frac{1}{\tau}\int_0^T [s(t) + n(t)]\,dt = \frac{1}{\tau}\int_0^T s(t)\,dt + \frac{1}{\tau}\int_0^T n(t)\,dt \qquad (11.1\text{-}1)$$

The sample voltage due to the signal is

$$s_o(T) = \frac{1}{\tau}\int_0^T V\,dt = \frac{VT}{\tau} \qquad (11.1\text{-}2)$$

The sample voltage due to the noise is

$$n_o(T) = \frac{1}{\tau}\int_0^T n(t)\,dt \qquad (11.1\text{-}3)$$

This noise-sampling voltage $n_o(T)$ is a gaussian random variable in contrast with $n(t)$, which is a gaussian random process.

The variance of $n_o(T)$ was found in Sec. 7.9 [see Eq. (7.9-17)] to be

$$\sigma_o^2 = \overline{n_o^2(T)} = \frac{\eta T}{2\tau^2} \qquad (11.1\text{-}4)$$

and, as noted in Sec. 7.3, $n_o(T)$ has a gaussian probability density.

The output of the integrator, before the sampling switch, is $v_o(t) = s_o(t) + n_o(t)$. As shown in Fig. 11.1-3*a*, the signal output $s_o(t)$ is a ramp, in each bit interval, of duration $T$. At the end of the interval the ramp attains the voltage $s_o(T)$ which is $+VT/\tau$ or $-VT/\tau$, depending on whether the bit is a 1 or a 0. At the end of each interval the switch $SW_1$ in Fig. 11.1-2 closes momentarily to discharge the capacitor so that $s_o(t)$ drops to zero. The noise $n_o(t)$, shown in Fig. 11.1-3*b*, also starts each interval with $n_o(0) = 0$ and has the random value $n_o(T)$ at the end of each interval. The sampling switch $SW_2$ closes briefly just before the closing of $SW_1$ and hence reads the voltage

$$v_o(T) = s_o(T) + n_o(T) \qquad (11.1\text{-}5)$$


Downloaded From : www.EasyEngineering.net

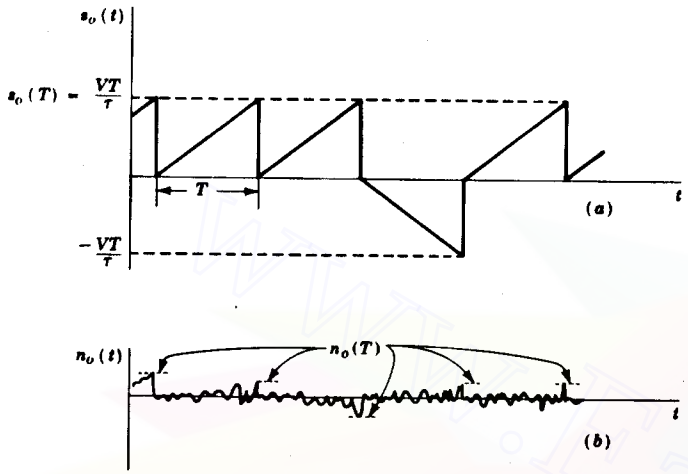



**Figure 11.1-3** ($a$) The signal output and ($b$) the noise output of the integrator of Fig. 11.1-2.

We would naturally like the output signal voltage to be as large as possible in comparison with the noise voltage. Hence a figure of merit of interest is the signal-to-noise ratio

$$\frac{[s_o(T)]^2}{\overline{[n_o(T)]^2}} = \frac{2}{\eta} V^2 T \tag{11.1-6}$$

This result is calculated from Eqs. (11.1-2) and (11.1-4). Note that the signal-to-noise ratio increases with increasing bit duration $T$ and that it depends on $V^2T$ which is the normalized energy of the bit signal. Therefore, a bit represented by a narrow, high amplitude signal and one by a wide, low amplitude signal are equally effective, provided $V^2T$ is kept constant.

It is instructive to note that the integrator filters the signal and the noise such that the signal voltage increases linearly with time, while the standard deviation (rms value) of the noise increases more slowly, as $\sqrt{T}$. Thus, the integrator enhances the signal relative to the noise, and this enhancement increases with time as shown in Eq. (11.1-6).

## 11.2 PROBABILITY OF ERROR

Since the function of a receiver of a data transmission is to distinguish the bit 1 from the bit 0 in the presence of noise, a most important characteristic is the probability that an error will be made in such a determination. We now calculate this error probability $P_e$ for the integrate-and-dump receiver of Fig. 11.1-2.

We have seen that the probability density of the noise sample $n_o(T)$ is gaussian and hence appears as in Fig. 11.2-1. The density is therefore given by

$$f[n_o(T)] = \frac{e^{-n_o^2(T)/2\sigma_o^2}}{\sqrt{2\pi\sigma_o^2}} \tag{11.2-1}$$

where $\sigma_o^2$, the variance, is $\sigma_o^2 \equiv \overline{n_o^2(T)}$ given by Eq. (11.1-4). Suppose, then, that during some bit interval the input-signal voltage is held at, say, $-V$. Then, at the sample time, the signal sample voltage is $s_o(T) = -VT/\tau$, while the noise sample is $n_o(T)$. If $n_o(T)$ is positive and larger in magnitude than $VT/\tau$, the total sample voltage $v_o(T) = s_o(T) + n_o(T)$ will be positive. Such a positive sample voltage will result in an error, since as noted earlier, we have instructed the receiver to interpret such a positive sample voltage to mean that the signal voltage was $+V$ during the bit interval. The probability of such a misinterpretation, that is, the probability that $n_o(T) > VT/\tau$, is given by the area of the shaded region in Fig. 11.2-1. The probability of error is, using Eq. (11.2-1).

$$P_e = \int_{VT/\tau}^{\infty} f[n_o(T)]\, dn_o(T) = \int_{VT/\tau}^{\infty} \frac{e^{-n_o^2(T)/2\sigma_o^2}}{\sqrt{2\pi\sigma_o^2}}\, dn_o(T) \tag{11.2-2}$$

Defining $x \equiv n_o(T)/\sqrt{2}\sigma_o$, and using Eq. (11.1-4), Eq. (11.2-2) may be rewritten as

$$P_e = \frac{1}{2}\frac{2}{\sqrt{\pi}} \int_{x=V\sqrt{T/\eta}}^{\infty} e^{-x^2}\, dx$$

$$= \frac{1}{2}\operatorname{erfc}\left(V\sqrt{\frac{T}{\eta}}\right) = \frac{1}{2}\operatorname{erfc}\left(\frac{V^2T}{\eta}\right)^{1/2} = \frac{1}{2}\operatorname{erfc}\left(\frac{E_s}{\eta}\right)^{1/2} \tag{11.2-3}$$

in which $E_s = V^2T$ is the signal energy of a bit.

If the signal voltage were held instead at $+V$ during some bit interval, then it is clear from the symmetry of the situation that the probability of error would again be given by $P_e$ in Eq. (11.2-3). Hence Eq. (11.2-3) gives $P_e$ quite generally.

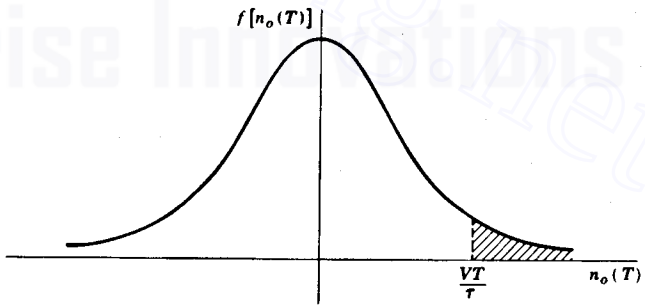



**Figure 11.2-1** The gaussian probability density of the noise sample $n_o(T)$.


Downloaded From : www.EasyEngineering.net

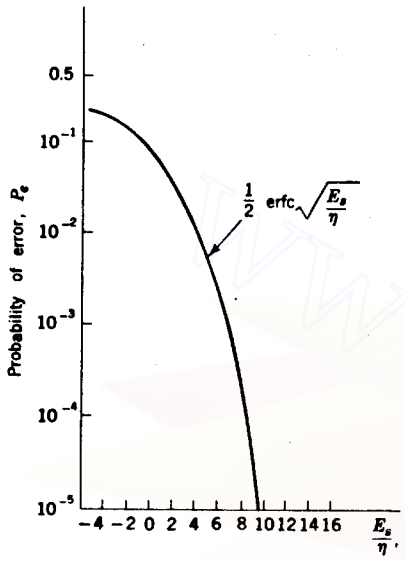



**Figure 11.2-2** Variation of $P_e$ versus $E_s/\eta$.

The probability of error $P_e$, as given in Eq. (11.2-3), is plotted in Fig. 11.2-2. Note that $P_e$ decreases rapidly as $E_s/\eta$ increases. The maximum value of $P_e$ is $\frac{1}{2}$. Thus, even if the signal is entirely lost in the noise so that any determination of the receiver is a sheer guess, the receiver cannot be wrong more than half the time on the average.

## 11.3 THE OPTIMUM FILTER

In the receiver system of Fig. 11.1-2, the signal was passed through a filter (i.e., the integrator), so that at the sampling time the signal voltage might be emphasized in comparison with the noise voltage. We are naturally led to ask whether the integrator is the optimum filter for the purpose of minimizing the probability of error. We shall find that for the received signal contemplated in the system of Fig. 11.1-2 the integrator is indeed the optimum filter. However, before returning specifically to the integrator receiver, we shall discuss optimum filters more generally.

We assume that the received signal is a binary waveform. One binary digit (bit) is represented by a signal waveform $s_1(t)$ which persists for time $T$, while the other bit is represented by the waveform $s_2(t)$ which also lasts for an interval $T$. For example, in the case of transmission at baseband, as shown in Fig. 11.1-2, $s_1(t) = +V$, while $s_2(t) = -V$; for other modulation systems, different waveforms are transmitted. For example, for PSK signalling, $s_1(t) = A \cos \omega_0 t$ and $s_2(t) = -A \cos \omega_0 t$; while for FSK, $s_1(t) = A \cos (\omega_0 + \Omega)t$ and $s_2(t) = A \cos (\omega_0 - \Omega)t$.

As shown in Fig. 11.3-1 the input, which is $s_1(t)$ or $s_2(t)$, is corrupted by the addition of noise $n(t)$. The noise is gaussian and has a spectral density $G(f)$. [In most cases of interest the noise is white, so that $G(f) = \eta/2$. However, we shall assume the more general possibility, since it introduces no complication to do so.] The signal and noise are filtered and then sampled at the end of each bit interval. The output sample is either $v_o(T) = s_{o1}(T) + n_o(T)$ or $v_o(T) = s_{o2}(T) + n_o(T)$. We assume that immediately after each sample, every energy-storing element in the filter has been discharged.

We have already considered in Sec. 2.22, the matter of signal determination in the presence of noise. Thus, we note that in the absence of noise the output sample would be $v_o(T) = s_{o1}(T)$ or $s_{o2}(T)$. When noise is present we have shown that to minimize the probability of error one should assume that $s_1(t)$ has been transmitted if $v_o(T)$ is *closer* to $s_{o1}(T)$ than to $s_{o2}(T)$. Similarly, we assume $s_2(t)$ has been transmitted if $v_o(T)$ is *closer* to $s_{o2}(T)$. The decision boundary is therefore midway between $s_{o1}(T)$ and $s_{o2}(T)$. For example, in the baseband system of Fig. 11.1-2, where $s_{o1}(T) = VT/\tau$ and $s_{o2}(T) = -VT/\tau$, the decision boundary is $v_o(T) = 0$. In general, we shall take the decision boundary to be

$$v_o(T) = \frac{s_{o1}(T) + s_{o2}(T)}{2} \tag{11.3-1}$$

The probability of error for this general case may be deduced as an extension of the considerations used in the baseband case. Suppose that $s_{o1}(T) > s_{o2}(T)$ and that $s_2(t)$ was transmitted. If, at the sampling time, the noise $n_o(T)$ is positive and larger in magnitude than the voltage difference $\frac{1}{2}[s_{o1}(T) + s_{o2}(T)] - s_{o2}(T)$, an error will have been made. That is, an error [we decide that $s_1(t)$ is transmitted rather than $s_2(t)$] will result if

$$n_o(T) \geq \frac{s_{o1}(T) - s_{o2}(T)}{2} \tag{11.3-2}$$

Hence the probability of error is

$$P_e = \int_{[s_{o1}(T) - s_{o2}(T)]/2}^{\infty} \frac{e^{-n_o^2(T)/2\sigma_o^2}}{\sqrt{2\pi\sigma_o^2}}\, dn_o(T) \tag{11.3-3}$$

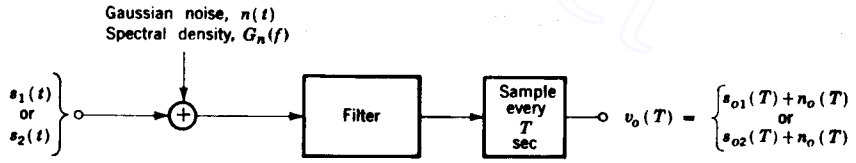



**Figure 11.3-1** A receiver for binary coded signalling.


Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

If we make the substitution $x \equiv n_o(T)/\sqrt{2}\sigma_o$, Eq. (11.3-3) becomes

$$P_e = \frac{1}{2}\frac{2}{\sqrt{\pi}}\int_{[s_{o1}(T)-s_{o2}(T)]/2\sqrt{2}\sigma_o}^{\infty} e^{-x^2}\,dx \tag{11.3-4a}$$

$$P_e = \frac{1}{2}\operatorname{erfc}\left[\frac{s_{o1}(T)-s_{o2}(T)}{2\sqrt{2}\sigma_o}\right] \tag{11.3-4b}$$

Note that for the case $s_{o1}(T) = VT/\tau$ and $s_{o2}(T) = -VT/\tau$, and, using Eq. (11.1-4), Eq. (11.3-4*b*) reduces to Eq. (11.2-3) as expected.

The complementary error function is a monotonically decreasing function of its argument. (See Fig. 11.2-2.) Hence, as is to be anticipated, $P_e$ decreases as the difference $s_{o1}(T) - s_{o2}(T)$ becomes larger and as the rms noise voltage $\sigma_o$ becomes smaller. The optimum filter, then, is the filter which maximizes the ratio

$$\gamma = \frac{s_{o1}(T)-s_{o2}(T)}{\sigma_o} \tag{11.3-5}$$

We now calculate the transfer function $H(f)$ of this optimum filter. As a matter of mathematical convenience we shall actually maximize $\gamma^2$ rather than $\gamma$.

### Calculation of the Optimum-Filter Transfer Function $H(f)$

The fundamental requirement we make of a binary encoded data receiver is that it distinguishes the voltages $s_1(t) + n(t)$ and $s_2(t) + n(t)$. We have seen that the ability of the receiver to do so depends on how large a particular receiver can make $\gamma$. It is important to note that $\gamma$ is proportional not to $s_1(t)$ nor to $s_2(t)$, but rather to the *difference* between them. For example, in the baseband system we represented the signals by voltage levels $+V$ and $-V$. But clearly, if our only interest was in distinguishing levels, we would do just as well to use $+2$ volts and 0 volt, or $+8$ volts and $+6$ volts, etc. (The $+V$ and $-V$ levels, however, have the advantage of requiring the least average power to be transmitted.) Hence, while $s_1(t)$ or $s_2(t)$ is the received signal, the signal which is to be compared with the noise, i.e., the signal which is relevant in all our error-probability calculations, is the difference signal

$$p(t) \equiv s_1(t) - s_2(t) \tag{11.3-6}$$

Thus, for the purpose of calculating the minimum error probability, we shall assume that the input signal to the optimum filter is $p(t)$. The corresponding *output signal* of the filter is then

$$p_o(t) \equiv s_{o1}(t) - s_{o2}(t) \tag{11.3-7}$$

We shall let $P(f)$ and $P_o(f)$ be the Fourier transforms, respectively, of $p(t)$ and $p_o(t)$.

If $H(f)$ is the transfer function of the filter,

$$P_o(f) = H(f)P(f) \tag{11.3-8}$$

and

$$p_o(T) = \int_{-\infty}^{\infty} P_o(f)e^{j2\pi fT}\,df = \int_{-\infty}^{\infty} H(f)P(f)e^{j2\pi fT}\,df \tag{11.3-9}$$

The input noise to the optimum filter is $n(t)$. The output noise is $n_o(t)$ which has a power spectral density $G_{n_o}(f)$ and is related to the power spectral density of the input noise $G_n(f)$ by

$$G_{n_o}(f) = |H(f)|^2 G_n(f) \tag{11.3-10}$$

Using Parseval's theorem (Eq. 1.13-5), we find that the normalized output noise power, i.e., the noise variance $\sigma_o^2$, is

$$\sigma_o^2 = \int_{-\infty}^{\infty} G_{n_o}(f)\,df = \int_{-\infty}^{\infty} |H(f)|^2 G_n(f)\,df \tag{11.3-11}$$

From Eqs. (11.3-9) and (11.3-11) we now find that

$$\gamma^2 = \frac{p_o^2(T)}{\sigma_o^2} = \frac{\left|\int_{-\infty}^{\infty} H(f)P(f)e^{j2\pi Tf}\,df\right|^2}{\int_{-\infty}^{\infty}|H(f)|^2 G_n(f)\,df} \tag{11.3-12}$$

Equation (11.3-12) is unaltered by the inclusion or deletion of the absolute value sign in the numerator since the quantity within the magnitude sign $p_o(T)$ is a positive real number. The sign has been included, however, in order to allow further development of the equation through the use of the *Schwarz inequality*.

The *Schwarz inequality* states that given arbitrary complex functions $X(f)$ and $Y(f)$ of a common variable $f$, then

$$\left|\int_{-\infty}^{\infty} X(f)Y(f)\,df\right|^2 \le \int_{-\infty}^{\infty}|X(f)|^2\,df\int_{-\infty}^{\infty}|Y(f)|^2\,df \tag{11.3-13}$$

The equal sign applies when

$$X(f) = KY^*(f) \tag{11.3-14}$$

where $K$ is an arbitrary constant and $Y^*(f)$ is the complex conjugate of $Y(f)$.

We now apply the Schwarz inequality to Eq. (11.3-12) by making the identification

$$X(f) \equiv \sqrt{G_n(f)}\,H(f) \tag{11.3-15}$$

and

$$Y(f) \equiv \frac{1}{\sqrt{G_n(f)}}P(f)e^{j2\pi Tf} \tag{11.3-16}$$

Using Eqs. (11.3-15) and (11.3-16) and using the Schwarz inequality, Eq. (11.3-13), we may rewrite Eq. (11.3-12) as

$$\frac{p_o^2(T)}{\sigma_o^2} = \frac{\left|\int_{-\infty}^{\infty} X(f)Y(f)\,df\right|^2}{\int_{-\infty}^{\infty}|X(f)|^2\,df} \le \int_{-\infty}^{\infty}|Y(f)|^2\,df \tag{11.3-17}$$


Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

or, using Eq. (11.3-16),

$$\frac{p_o^2(T)}{\sigma_o^2} \leq \int_{-\infty}^{\infty} |Y(f)|^2\, df = \int_{-\infty}^{\infty} \frac{|P(f)|^2}{G_n(f)}\, df \tag{11.3-18}$$

The ratio $p_o^2(T)/\sigma_o^2$ will attain its maximum value when the equal sign in Eq. (11.3-18) may be employed as is the case when $X(f) = KY^*(f)$. We then find from Eqs. (11.3-15) and (11.3-16) that the optimum filter which yields such a maximum ratio $p_o^2(T)/\sigma_o^2$ has a transfer function

$$H(f) = K \frac{P^*(f)}{G_n(f)} e^{-j2\pi fT} \tag{11.3-19}$$

Correspondingly, the maximum ratio is, from Eq. (11.3-18),

$$\left[\frac{p_o^2(T)}{\sigma_o^2}\right]_{\max} = \int_{-\infty}^{\infty} \frac{|P(f)|^2}{G_n(f)}\, df \tag{11.3-20}$$

In succeeding sections we shall have occasion to apply Eqs. (11.13-19) and (11.13-20) to a number of cases of interest.

## 11.4 WHITE NOISE: THE MATCHED FILTER

An optimum filter which yields a maximum ratio $p_o^2(T)/\sigma_o^2$ is called a *matched filter* when the input noise is *white*. In this case $G_n(f) = \eta/2$, and Eq. (11.3-19) becomes

$$H(f) = K \frac{P^*(f)}{\eta/2} e^{-j2\pi fT} \tag{11.4-1}$$

The impulsive response of this filter, i.e., the response of the filter to a unit strength impulse applied at $t = 0$, is

$$h(t) = \mathscr{F}^{-1}[H(f)] = \frac{2K}{\eta} \int_{-\infty}^{\infty} P^*(f) e^{-j2\pi fT} e^{j2\pi ft}\, df \tag{11.4-2a}$$

$$= \frac{2K}{\eta} \int_{-\infty}^{\infty} P^*(f) e^{j2\pi f(t-T)}\, df \tag{11.4-2b}$$

A physically realizable filter will have an impulse response which is real, i.e., not complex. Therefore $h(t) = h^*(t)$. Replacing the right-hand member of Eq. (11.4-2*b*) by its complex conjugate, an operation which leaves the equation unaltered, we have

$$h(t) = \frac{2K}{\eta} \int_{-\infty}^{\infty} P(f) e^{j2\pi f(T-t)}\, df \tag{11.4-3a}$$

$$= \frac{2K}{\eta} p(T-t) \tag{11.4-3b}$$

Finally, since $p(t) \equiv s_1(t) - s_2(t)$ [see Eq. (11.3-6)], we have

$$h(t) = \frac{2K}{\eta} [s_1(T-t) - s_2(T-t)] \tag{11.4-4}$$

The significance of these results for the matched filter may be more readily appreciated by applying them to a specific example. Consider then, as in Fig. 11.4-1*a*, that $s_1(t)$ is a triangular waveform of duration $T$, while $s_2(t)$, as shown in Fig. 11.4-1*b*, is of identical form except of reversed polarity. Then $p(t)$ is as shown in Fig. 11.4-1*c*, and $p(-t)$ appears in Fig. 11.4-1*d*. The waveform $p(-t)$ is the waveform $p(t)$ rotated around the axis $t = 0$. Finally, the waveform $p(T-t)$ called for as the impulse response of the filter in Eq. (11.4-3*b*) is this rotated waveform $p(-t)$ translated in the positive $t$ direction by amount $T$. This last translation ensures that $h(t) = 0$ for $t < 0$ as is required for a *causal* filter.

In general, the impulsive response of the matched filter consists of $p(t)$ rotated about $t = 0$ and then delayed long enough (i.e., a time $T$) to make the filter realiz-

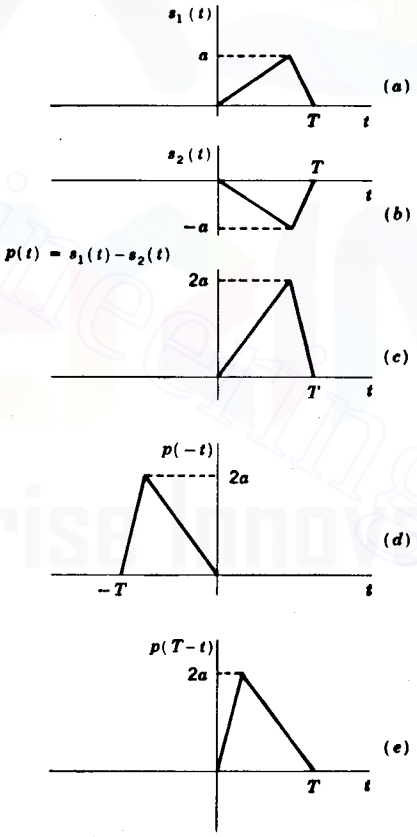



**Figure 11.4-1** The signals (*a*) $s_1(t)$, (*b*) $s_2(t)$, and (*c*) $p(t) = s_1(t) - s_2(t)$. (*d*) $p(t)$ rotated about the axis $t = 0$. (*e*) The waveform in (*d*) translated to the right by amount $T$.


Downloaded From : www.EasyEngineering.net

able. We may note in passing, that any additional delay that a filter might introduce would in no way interfere with the performance of the filter, for both signal and noise would be delayed by the same amount, and at the sampling time (which would need similarly to be delayed) the ratio of signal to noise would remain unaltered.

## 11.5 PROBABILITY OF ERROR OF THE MATCHED FILTER

The probability of error which results when employing a matched filter, may be found by evaluating the maximum signal-to-noise ratio $[p_o^2(T)/\sigma_o^2]_{max}$ given by Eq. (11.3-20). With $G_n(f) = \eta/2$, Eq. (11.3-20) becomes

$$\left[\frac{p_o^2(T)}{\sigma_o^2}\right]_{max} = \frac{2}{\eta}\int_{-\infty}^{\infty} |P(f)|^2\, df \tag{11.5-1}$$

From Parseval's theorem we have

$$\int_{-\infty}^{\infty} |P(f)|^2\, df = \int_{-\infty}^{\infty} p^2(t)\, dt = \int_0^T p^2(t)\, dt \tag{11.5-2}$$

In the last integral in Eq. (11.5-2), the limits take account of the fact that $p(t)$ persists for only a time $T$. With $p(t) = s_1(t) - s_2(t)$, and using Eq. (11.5-2), we may write Eq. (11.5-1) as

$$\left[\frac{p_o^2(T)}{\sigma_o^2}\right]_{max} = \frac{2}{\eta}\int_0^T [s_1(t) - s_2(t)]^2\, dt \tag{11.5-3a}$$

$$= \frac{2}{\eta}\left[\int_0^T s_1^2(t)\, dt + \int_0^T s_2^2(t)\, dt - 2\int_0^T s_1(t)s_2(t)\, dt\right] \tag{11.5-3b}$$

$$= \frac{2}{\eta}(E_{s1} + E_{s2} - 2E_{s12}) \tag{11.5-3c}$$

Here $E_{s1}$ and $E_{s2}$ are the energies, respectively, in $s_1(t)$ and $s_2(t)$, while $E_{s12}$ is the energy due to the correlation between $s_1(t)$ and $s_2(t)$.

Suppose that we have selected $s_1(t)$, and let $s_1(t)$ have an energy $E_{s1}$. Then it can be shown that if $s_2(t)$ is to have the *same energy*, the optimum choice of $s_2(t)$ is

$$s_2(t) = -s_1(t) \tag{11.5-4}$$

The choice is optimum in that it yields a maximum output signal $p_o^2(T)$ for a given signal energy. Letting $s_2(t) = -s_1(t)$, we find

$$E_{s1} = E_{s2} = -E_{s12} \equiv E_s$$

and Eq. (11.5-3c) becomes

$$\left[\frac{p_o^2(T)}{\sigma_o^2}\right]_{max} = \frac{8E_s}{\eta} \tag{11.5-5}$$





Rewriting Eq. (11.3-4b) using $p_o(T) = s_{o1}(T) - s_{o2}(T)$, we have

$$P_e = \frac{1}{2}\operatorname{erfc}\left[\frac{p_o(T)}{2\sqrt{2}\,\sigma_o}\right] = \frac{1}{2}\operatorname{erfc}\left[\frac{p_o^2(T)}{8\sigma_o^2}\right]^{1/2} \tag{11.5-6}$$

Combining Eq. (11.5-6) with (11.5-5), we find that the minimum error probability $(P_e)_{min}$ corresponding to a maximum value of $p_o^2(T)/\sigma_o^2$ is

$$(P_e)_{min} = \frac{1}{2}\operatorname{erfc}\left\{\frac{1}{8}\left[\frac{p_o^2(T)}{\sigma_o^2}\right]_{max}\right\}^{1/2} \tag{11.5-7}$$

$$= \frac{1}{2}\operatorname{erfc}\left(\frac{E_s}{\eta}\right)^{1/2} \tag{11.5-8}$$

We note that Eq. (11.5-8) establishes more generally the idea that the error probability depends only on the signal energy and not on the signal waveshape. Previously we had established this point only for signals which had constant voltage levels.

We note also that Eq. (11.5-8) gives $(P_e)_{min}$ for the case of the matched filter and when $s_1(t) = -s_2(t)$. In Sec. 11.2 we considered the case when $s_1(t) = +V$ and $s_2(t) = -V$ and the filter employed was an integrator. There we found [Eq. (11.2-3)] that the result for $P_e$ was identical with $(P_e)_{min}$ given in Eq. (11.5-8). This agreement leads us to suspect that for an input signal where $s_1(t) = +V$ and $s_2(t) = -V$, the integrator is the matched filter. Such is indeed the case. For when we have

$$s_1(t) = V \qquad 0 \le t \le T \tag{11.5-9a}$$

and

$$s_2(t) = -V \qquad 0 \le t \le T \tag{11.5-9b}$$

the impulse response of the matched filter is, from Eq. (11.4-4),

$$h(t) = \frac{2K}{\eta}[s_1(T - t) - s_2(T - t)] \tag{11.5-10}$$

The quantity $s_1(T - t) - s_2(T - t)$ is a pulse of amplitude $2V$ extending from $t = 0$ to $t = T$ and may be rewritten, with $u(t)$ the unit step,

$$h(t) = \frac{2K}{\eta}(2V)[u(t) - u(t - T)] \tag{11.5-11}$$

The constant factor of proportionality $4KV/\eta$ in the expression for $h(t)$ (that is, the gain of the filter) has no effect on the probability of error since the gain affects signal and noise alike. We may therefore select the coefficient $K$ in Eq. (11.5-11) so that $4KV/\eta = 1$. Then the inverse transform of $h(t)$, that is, the transfer function of the filter, becomes, with $s$ the Laplace transform variable,

$$H(s) = \frac{1}{s} - \frac{e^{-sT}}{s} \tag{11.5-12}$$

The first term in Eq. (11.5-12) represents an integration beginning at $t = 0$, while the second term represents an integration with reversed polarity beginning at $t = T$. The overall response of the matched filter is an integration from $t = 0$ to $t = T$ and a zero response thereafter. In a physical system, as already described, we achieve the effect of a zero response after $t = T$ by sampling at $t = T$, so that so far as the determination of one bit is concerned we ignore the response after $t = T$.

## 11.6 COHERENT RECEPTION: CORRELATION

We discuss now an alternative type of receiving system which, as we shall see, is identical in performance with the matched filter receiver. Again, as shown in Fig. 11.6-1, the input is a binary data waveform $s_1(t)$ or $s_2(t)$ corrupted by noise $n(t)$. The bit length is $T$. The received signal plus noise $v_i(t)$ is multiplied by a locally generated waveform $s_1(t) - s_2(t)$. The output of the multiplier is passed through an integrator whose output is sampled at $t = T$. As before, immediately after each sampling, at the beginning of each new bit interval, all energy-storing elements in the integrator are discharged. This type of receiver is called a *correlator*, since we are *correlating* the received signal and noise with the waveform $s_1(t) - s_2(t)$.

The output signal and noise of the correlator shown in Fig. 11.6-1 are

$$s_o(T) = \frac{1}{\tau}\int_0^T s_i(t)[s_1(t) - s_2(t)]\,dt \tag{11.6-1}$$

$$n_o(T) = \frac{1}{\tau}\int_0^T n(t)[s_1(t) - s_2(t)]\,dt \tag{11.6-2}$$

where $s_i(t)$ is either $s_1(t)$ or $s_2(t)$, and where $\tau$ is the constant of the integrator (i.e., the integrator output is $1/\tau$ times the integral of its input). We now compare these outputs with the matched filter outputs.

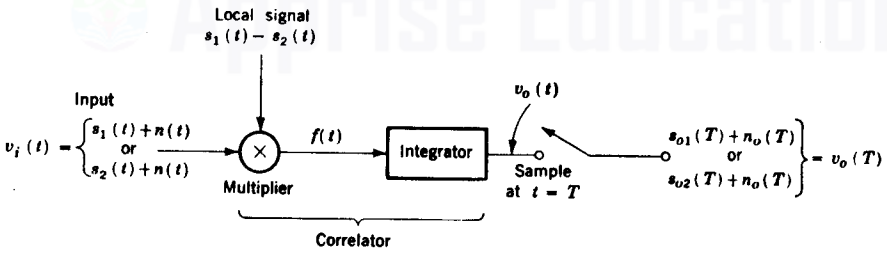



**Figure 11.6-1** A coherent system of signal reception.

If $h(t)$ is the impulsive response of the matched filter, then the output of the matched filter $v_o(t)$ can be found using the convolution integral (see Sec. 1.12). We have

$$v_o(t) = \int_{-\infty}^{\infty} v_i(\lambda)h(t - \lambda)\,d\lambda = \int_0^T v_i(\lambda)h(t - \lambda)\,d\lambda \tag{11.6-3}$$

The limits on the integral have been changed to 0 and $T$ since we are interested in the filter response to a bit which extends only over that interval. Using Eq. (11.4-4) which gives $h(t)$ for the matched filter, we have

$$h(t) = \frac{2K}{\eta}[s_1(T - t) - s_2(T - t)] \tag{11.6-4}$$

so that

$$h(t - \lambda) = \frac{2K}{\eta}[s_1(T - t + \lambda) - s_2(T - t + \lambda)] \tag{11.6-5}$$

Substituting Eq. (11.6-5) into (11.6-3), we have

$$v_o(t) = \frac{2K}{\eta}\int_0^T v_i(\lambda)[s_1(T - t + \lambda) - s_2(T - t + \lambda)]\,d\lambda \tag{11.6-6}$$

Since $v_i(\lambda) = s_i(\lambda) + n(\lambda)$, and $v_o(t) = s_o(t) + n_o(t)$, setting $t = T$ yields

$$s_o(T) = \frac{2K}{\eta}\int_0^T s_i(\lambda)[s_1(\lambda) - s_2(\lambda)]\,d\lambda \tag{11.6-7}$$

where $s_i(\lambda)$ is equal to $s_1(\lambda)$ or $s_2(\lambda)$. Similarly we find that

$$n_o(T) = \frac{2K}{\eta}\int_0^T n(\lambda)[s_1(\lambda) - s_2(\lambda)]\,d\lambda \tag{11.6-8}$$

Thus $s_o(T)$ and $n_o(T)$, as calculated from Eqs. (11.6-1) and (11.6-2) for the correlation receiver, and as calculated from Eqs. (11.6-7) and (11.6-8) for the matched filter receiver, are identical. Hence the performances of the two systems are identical.

The *matched filter* and the *correlator* are not simply two distinct, independent techniques which happen to yield the same result. In fact they are two techniques of synthesizing the optimum filter $h(t)$.

## 11.7 PHASE-SHIFT KEYING

An important application of the coherent reception system of Sec. 11.6 is its use in phase-shift keying (PSK). Here the input signal is

$$s_1(t) = A\cos\omega_0 t \tag{11.7-1}$$

or

$$s_2(t) = -A\cos\omega_0 t \tag{11.7-2}$$


Downloaded From : www.EasyEngineering.net
Downloaded From : www.EasyEngineering.net

modulating baseband signal. For SSB modulation $W$ is the bandwidth of both the modulating baseband signal and the modulated carrier. However, all of the other cases are doubled sideband systems and hence:

$$W = \frac{B}{2} \tag{11.20-1}$$

$B$ being the (two-sided) bandwidth of the modulated carrier. For the sake of having a uniform basis of comparison, we have in every case tabulated the error probability $P_e$.

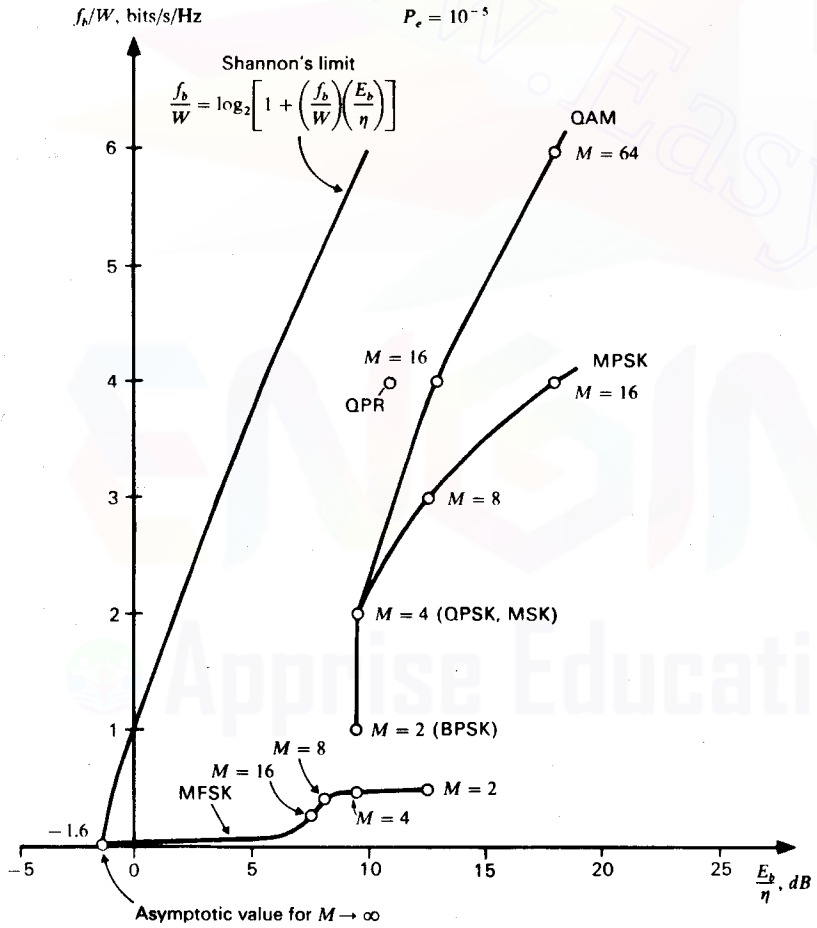



**Figure 11.20-1** Bits/s/Hz vs. $E_b/\eta$ for Probability of error = $10^{-5}$.

In connection with a communication system which is to be designed, it is common place that, in addition to such practical constraints as cost, volume, etc., there be a specification on allowable error rate. In such circumstances, plots as shown in Fig. 11.20-1 are very useful. Here $P_e$ is fixed at $P_e = 10^{-5}$. As we shall see in Sec. 13.7, there is an ultimate limit to the performance of a communications system. This limit, deduced by Shannon, is given by the inequality:

$$\frac{f_b}{W} \leq \log_2\left[1 + \left(\frac{f_b}{W}\right)\left(\frac{E_b}{\eta}\right)\right] \tag{11.20-2}$$

and is independant of the error probability. We have included a plot of Eq. (11.20-2) in Fig. 11.20-1. In Fig. 11.20-1, as well as for any given error probability, $P_e$, we shall always find that all modulation schemes yield plots which lie to the *right* of Shannon's limiting plot.

Figure 11.20-1 shows that QAM more closely approaches Shannon's curve than does MPSK, indicating that QAM is a more efficient system for operation in a white gaussian noise environment. Note that when we compare MPSK with MFSK (see Table 11.20-1 and Fig. 11.20-1) we find that as $M$ increases the bandwidth of MPSK decreases while the SNR required to obtain $P_e = 10^{-5}$ increases. However, with MFSK the bandwidth increases while the SNR decreases. In all cases however we note that the *slope* of each curve is positive, so that an increase in $f_b/W$ requires an increase in $E_b/\eta$ if a constant bit error rate is to be maintained, i.e., for a constant error rate

$$\frac{f_b}{W} \sim \frac{E_b}{\eta} \tag{11.20-3}$$

If the error rate were, for example, $P_e = 10^{-6}$ each curve would shift approximately 1 dB to the *right*, except of course, for Shannon's curve which is independent of $P_e$.

## REFERENCES

1. Stein, S., and J. Jones: "Modern Communication Principles," McGraw-Hill Book Company, New York, 1967.
2. Schwartz, M., W. R. Bennett, and S. Stein: "Communication Systems and Techniques," McGraw-Hill Book Company, New York, 1966.
3. Wozencraft, J., and I. Jacobs: "Communication Engineering," John Wiley and Sons, New York, 1966.

## PROBLEMS

**11.1-1.** (*a*) Find the power spectral density $G_n(f)$ of noise $n(t)$ which has an autocorrelation function

$$R_n(\tau) = \sigma^2 e^{-|\tau/\tau_0|}$$

(*b*) The noise in (*a*) is applied to an integrator at $t = 0$. Find the mean square value $\overline{[n_o(T)]^2}$ of the noise output of the integrator at $t = T$.


Downloaded From : www.EasyEngineering.net

(*c*) The noise in (*a*) accompanies a signal which consists of either the voltage $+V$ or the voltage $-V$ sustained for a time $T$. At time $t = T$ find the ratio of the integrator output due to the signal to the rms noise voltage.

**11.1-2.** A received signal $s(t) = \pm V$ is held for an interval $T$. The signal is accompanied by white gaussian noise of power spectral density $\eta/2$. The received signal is to be processed as in Fig. 11.1-2. However, as an approximation to the required integrator we use a low-pass $RC$ circuit of 3-dB bandwidth $f_c$. Calculate the value of $f_c$ for which the signal-to-noise voltage, at the sampling time, will be a maximum. For this value of $f_c$ calculate the signal-to-rms noise ratio and compare with Eq. (11.1-6) which applies when an integrator is used. Show that for the $RC$ network the signal-to-noise ratio is about 1 dB smaller than for the integrator.

**11.2-1.** A signal which can assume one of the voltage $+V$ or $-V$ is transmitted. Consider that the probability of transmitting $+V$ is $\frac{3}{4}$ while the probability of transmitting $-V$ is $\frac{1}{4}$. The signal is accompanied by white gaussian noise.

(*a*) Assume that the threshold voltage for decision between the two possible signals is $V_t$ rather than zero volts. Write an expression for the probability that an error in decision will be made. An integrate and dump receiver is used as in Fig. 11.1-2.

(*b*) Find $V_t$ such that the probability of error is a minimum and calculate the corresponding probability of error.

**11.2-2.** A signal, which can take on the voltages $+V$, 0, $-V$ with equal likelihood, is transmitted. When received, it is embedded in white gaussian noise. The receiver integrates the signal and noise for a time $T$ s.

Write an expression for the threshold voltages $\pm V_t$ so that the probability of error is independent of which signal is transmitted.

**11.2-3.** A received signal is either $+2V$ or $-2V$ held for a time $T$. The signal is corrupted by white gaussian noise of power spectral density $10^{-4}$ volt$^2$/Hz. If the signal is processed by an integrate and dump receiver, what is the minimum time $T$ during which a signal must be sustained if the probability of error is not to exceed $10^{-4}$?

**11.3-1.** A transmitter transmits the signals $\pm V$ with equal probability; the channel noise has the power spectral density $G_n(f) = G_0/[1 + (f/f_1)^2]$.

(*a*) Find the transfer function $H(f)$ of the *matched filter*, and comment on its realizability.

(*b*) Find the average probability of error when using the matched filter.

(*c*) An integrator is employed rather than the optimum filter. Find its $P_e$ and compare with (*b*).

**11.5-1.** A signal is either $s_1(t) = A \cos 2\pi f_0 t$ or $s_2(t) = 0$ for an interval $T = n/f_0$ with $n$ an integer. The signal is corrupted by white noise with $G_n(f) = \eta/2$. Find the transfer function of the matched filter for this signal. Write an expression for the probability of error $P_e$.

**11.5-2.** Repeat Prob. 11.5-1 if the signal is $s(t) = \pm A(1 - \cos 2\pi f_0 t)$.

**11.6-1.** Compare the outputs of the MF and the correlator, when the input signal is either $\pm V$, as a function of time $t$ for $0 \le t \le T$. Assume white gaussian noise. Are the outputs the same for all $t$, or just when $t = T$?

**11.6-2.** A signal is $s(t) = \pm 2(t/T)$ for $0 \le t \le T$. The signal is corrupted by white gaussian noise of power spectral density $10^{-6}$ volt$^2$/Hz.

(*a*) Draw the signal waveform at the output of a matched filter receiver.

(*b*) If the probability of error $P_e$ is to be no larger than $10^{-4}$, find the minimum allowable interval $T$.

**11.7-1.** Verify Eq. (11.7-5).

**11.8-1.** Plot Eq. (11.8-8) versus $E_s/\eta$.

**11.8-2.** If the frequency offset $\Omega$ in FSK satisfies $\Omega T = n\pi$, $s_1(t)$ and $s_2(t)$ are orthogonal.

(*a*) Prove this statement.

(*b*) Calculate $P_e$.

(*c*) Plot $P_e$ versus $E_s/\eta$ and compare with the results given in Eq. (11.8-8).

**11.8-3.** Plot the $P_e$ in binary FSK as a function of $\Omega T$. Select $E_s/\eta = 15$.



Downloaded From : www.EasyEngineering.net

**11.9-1.** Plot Eq. (11.9-1) versus $E_s/\eta$.

**11.11-1.** $M$-ary PSK involves choosing signals of the form $(\cos \omega_0 t + \theta_i)$ for $M$ values of $i$.

(*a*) Show how to choose the $\theta_i$ so that the probability of error of each $\theta_i$ is the same.

(*b*) Find the correlator detector.

(*c*) Obtain an expression for the probability of error.

**11.11-2** Verify the entries in Table 11.11-1.

**11.12-1.** Verify Eq. (11.12-2).

**11.13-1.** Refer to Eq. (11.13-5*a* and *b*). Determine $\omega_2 - \omega_1$ such that the unit vectors $u_1(t)$ and $u_2(t)$ are orthonormal.

**11.13-2.** (*a*) If $u_1(t) = \sqrt{2/T_b} \cos(\omega_1 t + \Theta)$ and $u_2(t) = \sqrt{2/T_b} \cos(\omega_2 t + \phi)$ where $\Theta$ and $\phi$ are independent random variables uniformly distributed between $-\pi \le \Theta, \phi < \pi$, find $\omega_2 - \omega_1$ so that $u_1$ and $u_2$ are orthonormal.

(*b*) Compare your answer to the answer obtained for Prob. 11.13-1. Why do they differ?

**11.13-3.** Plot a graph having an ordinate which is $f_b/W$ and an abscisssa which is $E_b/\eta$ for BPSK and BFSK for $P_e = 10^{-4}$, $10^{-5}$ and $10^{-6}$. Use Eqs. (11.13-4) and (11.13-10*a*) and the results,

$$B(\text{BPSK}) = 2f_b \quad \text{and} \quad B(\text{BFSK}) = 4f_b.$$

**11.14-1.** Verify that Eq. (11.14-3) reduces to Eq. (11.14-4).

**11.15-1.** Equations (11.15-5) and (11.15-7) assume that the *arc* of a circle is of the same length as the *chord*. (*a*) Derive an expression for $P_e$ to replace Eq. (11.15-7) if the above approximation is *not* made. (*b*) If the approximation is to be used but the resulting $P_e$ is to be correct to within 50 percent determine the minimum value of $M$.

**11.15-2.** If the probability of BPSK, BFSK and MPSK for $M = 4$, 8, 16, 32, and 64 are to be the same, what is the ratio of their $E_b/\eta$ compared to $(E_b/\eta)_{\text{BPSK}}$?

**11.15-3.** Calculate $P_e$ for 16 QAM using the Union Bound. Compare your answer with that of Eq. (11.15-11).

**11.15-4.** Calculate $P_e$ for 16 QAM without using the Union Bound approximation.

**11.15-5.** (*a*) Using the Union Bound approximation, calculate $P_e$ for 64 and 256 QAM.

(*b*) To obtain the same $P_e$ for 16, 64 and 256 QAM determine the ratio of each $E_b/\eta$ as compared to $(E_b/\eta)_{\text{BPSK}}$.

**11.15-6.** Verify Eq. (11.15-16).

**11.15-7.** A 9.6 kb/s NRZ data stream is to be transmitted over a 2.4 kHz bandwidth channel. What modulation system would you choose if an error rate of $10^{-4}$ is to be achieved with a minimum signal-to-noise ratio (minimum $E_b/\eta$)?

**11.15-8.** A 9.6 kb/s NRZ data stream is to be transmitted over a 2.4 kHz bandwidth channel. An error rate of $10^{-4}$ is desired. Due to channel nonlinearities and its phase characteristics, intersymbol interference (ISI) is produced which results in an "eye pattern" which closes by 3 dB compared to the response obtained when the transmitter is connected directly to the receiver (by-passing the channel) when $E_b/\eta$ is 12 dB.

Can a modulation system be found to achieve this error rate of $10^{-4}$? What is the value of $E_b/\eta$ that is needed?

**11.16-1.** The probability of a bit being in error is $10^{-3}$. (*a*) If a message consists of 10 bits, calculate the probability of the message being in error. (*b*) Repeat (*a*), if the message length is 100, 1000, 10,000. Note how quickly the probability of a message being in error increases toward unity.

**11.16-2.** The probability of a bit being in error is $10^{-3}$. (*a*) What is the probability of a 6-bit word containing an error? (*b*) The bits are grouped so that instead of transmitting 6 bits, bit-by-bit, a 64 phase, 64 PSK signal is sent. (The symbol rate is one sixth of the bit rate.) Calculate the probability of a 6-bit symbol being in error. (*c*) Repeat (*b*) using 64 QAM. (*d*) Repeat using 64 FSK. (*e*) Discuss your results.

**11.17-1.** Verify Eq. (11.17-2).

Downloaded From : www.EasyEngineering.net

**11.17-2.** Consider that bits are grouped so that there are 4 bits/symbol and then modulated using 16 QAM. Following Fig. 11.17-1 show how to arrange the Grey code so that an error in a symbol will with high probability correspond to an error in only 1 of the 4 bits.

**11.18-1.** Prove that $H_D(f)$ as defined by Eqs. (11.18-2) and (11.18-3) is the matched filter for an impulse transmitted through $H_E(f)$.

**11.18-2.** Verify Eq. (11.18-6).

**11.18-3.** Verify Eq. (11.18-9) and Eq. (11.18-10).

**11.18-4.** Verify Eq. (11.18-11).

**11.18-5.** $\Delta_0(kT_s)$ and $\Delta_e(kT_s)$ can each assume one of three values. Thus, a signal space sketch would indicate nine signal points. Determine the nine signal points in signal space to yield Eq. (11.18-15).

**11.18-6.** A probability of error of $10^{-5}$ is desired and a channel bandwidth of 20 kHz is available. If the bit rate is 80 kb/s, 16 PSK, 16 QAM, or QPR can be used. Calculate the value of $E_b/\eta$ required for each of these systems.

**11.19-1.** MSK can be viewed as an FM system with a peak frequency deviation of $f_b/4$. If coherent FM detection is employed as in MFSK (see Sec. 11.8) calculate the probability of error.

**11.19-2.** MSK can be detected in a noncoherent manner using an FM discriminator. Referring to Fig. 10.2-1, the IF filter bandwidth is set to $B = 1.5\,f_b$ (since the spectral density of MSK is zero at $f - f_c = 0.75\,f_b$, very little required signal power lies outside this region and hence ISI can be neglected). The baseband filter is often replaced by an integrate-and-dump filter, which integrates over each bit, i.e. for a time $T_b$.

(*a*) Using Eqs. (9.2-16), (9.2-21), and (10.6-1) where $|\delta f| = f_b/4$, calculate the probability of error. (*b*) Why is $f = f_b/4$? (*c*) How much degradation is produced when the FM discriminator is employed?

**11.20-1.** Using Table 11.20-1 and Shannon's limit, plot Fig. 11.20-1 for $P_e = 10^{-6}$.





CHAPTER

# TWELVE

# NOISE IN PULSE-CODE AND DELTA-MODULATION SYSTEMS

## 12.1 PCM TRANSMISSION

A binary PCM transmission system is shown in Fig. 12.1-1. The baseband signal $m(t)$ is quantized, giving rise to the quantized signal $m_q(t)$, where

$$m_q(t) = m(t) + e(t) \tag{12.1-1}$$

The term $e(t)$ is the error signal which results from the process of quantization. The quantized signal is next sampled. Sampling takes place at the Nyquist rate. The sampling interval is $T_s = 1/2f_M$, where $f_M$ is the frequency to which the signal $m(t)$ is bandlimited.

Sampling is accomplished by multiplying the signal $m_q(t)$ by a waveform which consists of a periodic train of pulses, the pulses being separated by the sampling interval $T_s$. We shall assume that the sampling pulses are narrow enough so that the sampling may be considered as instantaneous. It will be recalled (Sec. 5.1) that with such instantaneous sampling, the sampled signal may be reconstructed *exactly* by passing the sequence of samples through a low-pass filter with cut-off frequency at $f_M$. Now, as a matter of mathematical convenience, we shall represent each sampling pulse as an impulse. Such an impulse is infinitesimally narrow yet is characterized by having a finite area. The area of an impulse is called its *strength*, and an impulse of strength $I$ is written $I\delta(t)$. The sampling-impulse train is therefore $S(t)$, given by

$$S(t) = I \sum_{k=-\infty}^{\infty} \delta(t - kT_s) \qquad T_s = \frac{1}{2f_M} \tag{12.1-2}$$