CHAPTER

6 Angle Modulation

A continuous-wave (CW) sinusoidal signal can be varied by changing its amplitude and its phase angle. Recalling Eq. (5.2), we write

$$\phi(t) = a(t) \cos [\omega_c t + \gamma(t)] .$$

In Chapter 5 we kept $\gamma(t)$ constant and varied $a(t)$ proportional to $f(t)$. This introduced the concept of amplitude modulation. Now we shall investigate the case in which $a(t) = A$ (a constant) and the phase angle $\gamma(t)$ is varied in proportion to $f(t)$. This introduces the concept of angle modulation.

6.1 FM AND PM

The angle of a sinusoidal signal is described in terms of a frequency and/or a phase angle. Before proceeding here, however, we must decide precisely what we mean by the frequency of a sinusoid. If a sinusoid has a constant angular rate ω_0, then we say that the frequency of the sinusoid is ω_0 radians per second. But what happens if the angular rate is not constant? It is helpful at this point to return to a phasor representation.

The phasor representation of a constant-amplitude sinusoid is shown in Fig. 6.1. This phasor has a magnitude A and a phase angle $\theta(t)$. If $\theta(t)$ increases linearly with time [that is, $\theta(t) = \omega_0 t$], we say that the phasor has an angular rate, or "frequency," of ω_0 radians per second. If the angular rate is not constant, we can still write a relation between the instantaneous angular rate $\omega_i(t)$ and $\theta(t)$:

$$\theta(t) = \int_0^t \omega_i(\tau)\, d\tau + \theta_0 . \tag{6.1}$$

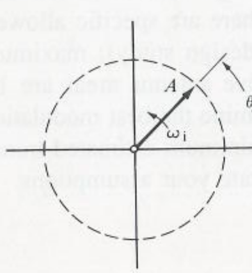

Figure 6.1 A general phasor representation.

Taking the derivative of both sides of Eq. (6.1), we have

$$\omega_i(t) = \frac{d\theta}{dt} . \tag{6.2}$$

Therefore we conclude that the instantaneous frequency of a sinusoidal signal is given by the time derivative of its phase. Note that this definition agrees with our usual concept of frequency when the phase is linear with time.

EXAMPLE 6.1.1

Determine the instantaneous frequency of the signal $\phi(t) = A \cos (10\pi t + \pi t^2)$.

Solution

$$\theta(t) = 10\pi t + \pi t^2$$

$$\omega_i(t) = \frac{d\theta}{dt} = 10\pi + 2\pi t = 2\pi(5 + t)$$

The frequency of $\phi(t)$ is 5 Hz at $t = 0$ and increases linearly at a rate of 1 Hz per second. Thus a quadratic phase shift gives a linear frequency dependence.

DRILL PROBLEM 6.1.1

Determine the instantaneous frequency of the following signal at $t = 0$: $\phi(t) = 5 \cos (10t + \sin 5t)$.

ANSWER: 15 rad/sec.

The concept of instantaneous frequency now permits us to describe two obvious possibilities for angle modulation (there are many more). If the phase angle $\theta(t)$ is varied linearly with the input signal $f(t)$, we can write

$$\theta(t) = \omega_c t + k_p f(t) + \theta_0 \tag{6.3}$$

where ω_c, k_p, θ_0 are constants. Because the phase is linearly related to $f(t)$, this type of angle modulation is called *phase modulation* (PM). The instantaneous frequency of this phase-modulated signal is

$$\omega_i = \frac{d\theta}{dt} = \omega_c + k_p \frac{df}{dt} . \tag{6.4}$$

Another possibility is to make the instantaneous *frequency* proportional to the input signal,

$$\omega_i = \omega_c + k_f f(t) , \tag{6.5}$$

where ω_c, k_f are constants. Because the frequency is linearly related to $f(t)$, this type of angle modulation is called *frequency modulation* (FM). The phase angle of this frequency-modulated signal is

$$\theta(t) = \int_0^t \omega_i(\tau)\,d\tau = \omega_c t + \int_0^t k_f f(\tau)\,d\tau + \theta_0 . \tag{6.6}$$

A comparison of Eqs. (6.3)–(6.6) shows that PM and FM are closely related. The phase angle of the PM carrier signal is proportional to the modulating signal. But in FM the instantaneous frequency of the carrier signal is proportional to the modulating signal, so that the phase angle of the carrier is proportional to the integral of the modulating signal. Therefore, if the modulating signal $f(t)$ is first integrated and then used to phase modulate a carrier, the result is a signal that is frequency-modulated. Figure 6.2 is an illustration of FM and PM waveforms for given $f(t)$.

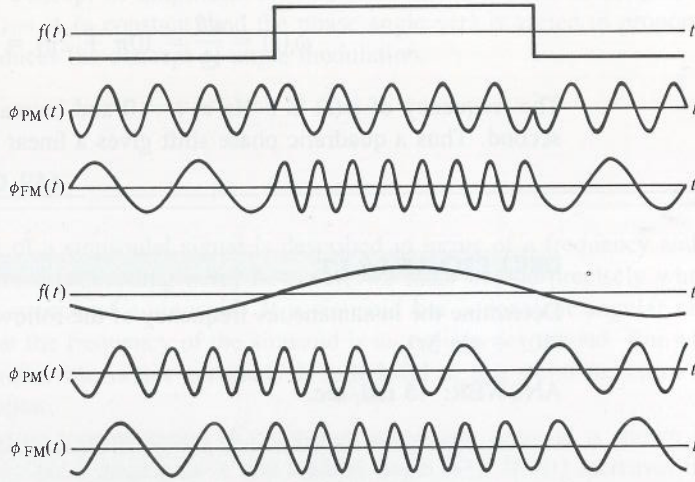

Figure 6.2 Examples of frequency and phase modulation.

Because frequency and phase modulation are so closely related, any variation in phase will necessarily result in a variation in frequency and vice versa. The essential difference between FM and PM is the nature of the dependency on the modulating signal. Although we shall discuss FM in more detail, our discussion is also valid for PM with only minor differences and these are pointed out in a later section.

In the case of AM signals, there was always a one-to-one correspondence between the modulated signal and the modulating signal. When this condition holds, the modulation is said to be *linear.*† For PM and FM this is not always true, however, as can be seen from the following reasoning.

† More formally, if $f(t)$ is the modulating signal and $\phi(t)$ is the modulated signal, the modulation is linear if $d\phi/df$ is independent of $f(t)$.

A general PM (or FM, with the appropriate modifications) signal can be represented by (note our return to complex notation)†

$$\phi_{PM}(t) = Ae^{j\theta(t)} = Ae^{j(\omega_c t + \theta_0)}e^{jk_p f(t)} . \tag{6.7}$$

Using a series expansion for the exponential modulation factor in Eq. (6.7), we have

$$\phi_{PM}(t) = Ae^{j(\omega_c t + \theta_0)}\left[1 + jk_p f(t) - \frac{1}{2!}k_p^2 f^2(t) - j\frac{1}{3!}k_p^3 f^3(t) + \cdots\right] . \tag{6.8}$$

From this result we conclude that, unless $|k_p f(t)| \ll 1$, angle modulation—in this case PM—is not linear. Therefore we can expect that, in general, the sidebands arising in angle modulation will not obey the principle of superposition. An analysis of spectra, etc., will have to be carried out choosing a particular waveform. When confronted by this choice, we shall use the sinusoidal waveform unless otherwise specified.

6.2 NARROWBAND FM

The linear condition in Eq. (6.8) maintains a linear modulation for FM and this appears to be a good place to begin. To lay the groundwork for the nonlinear modulation case we shall use a sinusoidal modulating signal. To be specific, let

$$f(t) = a \cos \omega_m t . \tag{6.9}$$

Because we are dealing with FM [cf. Eq. (6.5)],

$$\omega_i = \omega_c + k_f f(t)$$

$$= \omega_c + ak_f \cos \omega_m t , \tag{6.10}$$

where k_f is the frequency modulation constant; typical units are in radians per second per volt. Defining a new constant called the *peak frequency deviation,*

$$\Delta\omega = ak_f , \tag{6.11}$$

we can rewrite Eq. (6.10) as

$$\omega_i = \omega_c + \Delta\omega \cos \omega_m t . \tag{6.12}$$

The phase of this FM signal is [cf. Eq. (6.6)] (let $\theta_0 = 0$ for convenience)

$$\theta(t) = \omega_c t + \frac{\Delta\omega}{\omega_m} \sin \omega_m t = \omega_c t + \beta \sin \omega_m t , \tag{6.13}$$

† Although these are written in terms of PM signals, the conclusions are applicable to FM as well by substituting

$$k_f \int_0^t f(\tau)\,d\tau \quad \text{for} \quad k_p f(t) .$$

where

$$\beta = \Delta\omega/\omega_m \tag{6.14}$$

is a dimensionless ratio of the peak frequency deviation to the modulating frequency.

The resulting FM signal is

$$\phi_{FM}(t) = \mathcal{Re}\{Ae^{j\theta(t)}\}$$
$$= \mathcal{Re}\{Ae^{j\omega_c t}e^{j\beta \sin \omega_m t}\}. \tag{6.15}$$

Note that Eq. (6.15) can be rewritten as

$$\phi_{FM}(t) = A \cos (\omega_c t + \beta \sin \omega_m t). \tag{6.16}$$

Alternatively, an identity for the real part of a product (cf. Appendix A) may be used to rewrite Eq. (6.16) as

$$\phi_{FM}(t) = A \cos \omega_c t \cos (\beta \sin \omega_m t) - A \sin \omega_c t \sin (\beta \sin \omega_m t). \tag{6.17}$$

It is fairly obvious that either we will have to approximate this result, or seek some alternative approach. For small values of β we can write

$$\cos (\beta \sin \omega_m t) \approx 1, \tag{6.18}$$
$$\sin (\beta \sin \omega_m t) \approx \beta \sin \omega_m t. \tag{6.19}$$

The condition where β is small enough for these approximations is the condition for *narrowband FM* (NBFM). Usually a value of $\beta < 0.2$ is taken to be sufficient to satisfy this condition. Using Eqs. (6.18), (6.19) in Eq. (6.17), we obtain an approximate solution for small β:

$$\phi_{NBFM}(t) = A \cos \omega_c t - \beta A \sin \omega_m t \sin \omega_c t. \tag{6.20}$$

It is instructive to compare Eq. (6.20) with an equivalent expression for an AM signal:

$$\phi_{AM}(t) = A \cos \omega_c t + mA \cos \omega_m t \cos \omega_c t. \tag{6.21}$$

As suggested by a comparison of Eqs. (6.20), (6.21), the parameter β is called the *modulation index* of the FM signal.

Although the narrowband FM signal and the AM signal have similarities, they are distinctively different methods of modulation. The similarities and differences are portrayed by considering their phasor representations. Expanding Eq. (6.20) [or Eq. (6.15)] in phasor form, we have

$$\phi_{NBFM}(t) = \mathcal{Re}\{Ae^{j\omega_c t}(1 + j\beta \sin \omega_m t)\}$$
$$= \mathcal{Re}\{Ae^{j\omega_c t}(1 + \tfrac{1}{2}\beta e^{j\omega_m t} - \tfrac{1}{2}\beta e^{-j\omega_m t})\}. \tag{6.22}$$

Similarly, Eq. (6.21) can be written in phasor form as

$$\phi_{AM}(t) = \mathcal{Re}\{Ae^{j\omega_c t}(1 + m \cos \omega_m t)\}$$
$$= \mathcal{Re}\{Ae^{j\omega_c t}(1 + \tfrac{1}{2}me^{j\omega_m t} + \tfrac{1}{2}me^{-j\omega_m t})\}. \tag{6.23}$$

Taking the term $Ae^{j\omega_c t}$ as the reference (i.e., suppressing the continuous ω_c rotation), we show the phasor representation of each of these waveforms in Fig. 6.3. The resultant waveform can be found by rotating the entire phasor diagram at an angular rate of ω_c rad/sec and then taking the projection of the resultant on the real axis.

From Fig. 6.3, the differences between Eqs. (6.22) and (6.23) become quite evident. In the AM waveform, the modulation is added in phase with the carrier whereas in NBFM the modulation is added in quadrature with the carrier. The NBFM case gives rise to phase variations with very little amplitude change, whereas the AM case gives amplitude variations with no phase deviation.

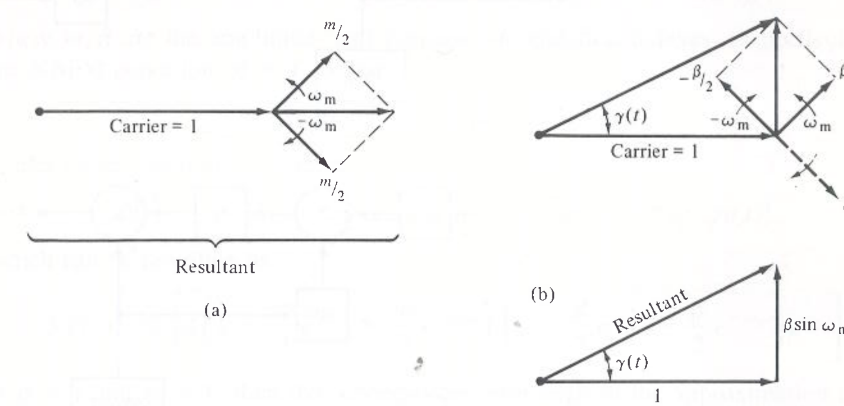

Figure 6.3 Phasor representation of (a) AM and (b) NBFM.

It is instructive to determine the limits on β from the results of the phasor diagrams in Fig. 6.3(b); the phase angle from the carrier is

$$\gamma(t) = \tan^{-1} (\beta \sin \omega_m t). \tag{6.24}$$

The instantaneous frequency deviation from the carrier frequency should be equal to $\Delta\omega \cos \omega_m t = \beta\omega_m \cos \omega_m t$ and is found by taking the derivative of this phase angle, or,

$$\frac{d\gamma}{dt} = \frac{\beta\omega_m \cos \omega_m t}{1 + \beta^2 \sin^2 \omega_m t} \approx \beta\omega_m \cos \omega_m t, \quad \text{if} \quad \beta^2 \sin^2 \omega_m t \ll 1. \tag{6.25}$$

The amplitude of the resultant phasor should be a constant (A); checking from the phasor diagram, we find

$$A\sqrt{1 + \beta^2 \sin^2 \omega_m t} \approx A \quad \text{if} \quad \beta^2 \sin^2 \omega_m t \ll 1. \tag{6.26}$$

Because $\sin^2 \omega_m t \le 1$, these approximations are valid if $\beta^2 \ll 1$. Choosing $\beta^2 < 0.1$, we find that $\beta < 1/\sqrt{10} = 0.316$ is a reasonable bound for the narrowband approximation. Values as high as 0.50 can be used in practice if the resulting amplitude modulation is removed by amplitude-limiting the angle-modulated waveform.

The addition of the modulation in quadrature with the carrier in NBFM, in contrast to that in phase in AM, suggests a method of generation for the NBFM or NBPM case using phase shifters and balanced modulators as shown in Fig. 6.4. This method is commonly used in the generation of NBFM and NBPM signals. Note that even though we have been discussing the FM case, the PM case follows in the same manner.

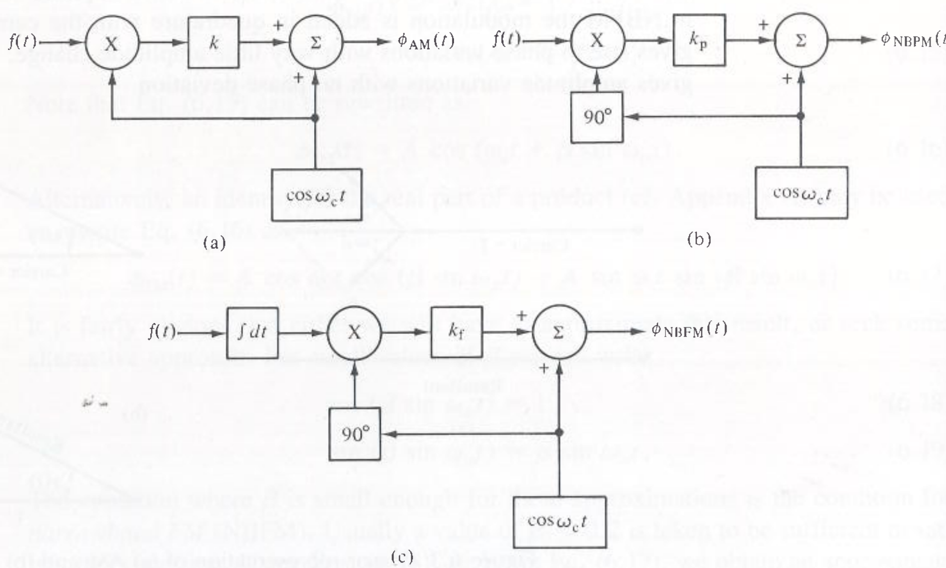

Figure 6.4 Generation of signals using balanced modulators: (a) AM; (b) NBPM; (c) NBFM.

Summarizing, narrowband FM (and PM), like AM, is an example of linear modulation. A major difference is that the modulation is added in phase with the carrier in AM, whereas it is added in quadrature with the carrier in NBFM. Both systems require a bandwidth of $W = 2\omega_m$ to transmit a signal of ω_m rad/sec in spectral width. The modulation index in FM is $\beta = \Delta\omega/\omega_m$ and a useful criterion for NBFM is $\beta < 0.2$.

DRILL PROBLEM 6.2.1

Calculate the maximum (peak)-percentage amplitude, phase, and frequency error incurred in using the phasor approximations to narrowband FM for the sinusoidal case when (a) $\beta = 0.20$; (b) $\beta = 0.50$.

ANSWER: (a) 2.0%, 1.3%, 3.8%; (b) 11.8%, 7.3%, 20.0%.

EXAMPLE 6.2.1

A comparison of the two phasor representations in Fig. 6.3 motivates us to investigate the simultaneous use of both amplitude modulation and narrowband frequency modulation for the possible elimination of one sideband. Investigate the possible use of this technique, for sinusoidal modulation, to generate SSB-LC signals.

Solution For simultaneous amplitude and frequency modulation we write

$$\phi(t) = \mathscr{Re}\{A(1 + m \cos \omega_m t) \exp [j(\omega_c t + \beta \sin \omega_m t)]\},$$

where m, β are the amplitude- and frequency-modulation indexes, respectively. For the NBFM condition, $\beta \ll 1$ so that

$$\exp (j\beta \sin \omega_m t) \approx 1 + j\beta \sin \omega_m t.$$

Under these conditions, we have

$$\phi(t) \approx \mathscr{Re}\{A(1 + m \cos \omega_m t)(1 + j\beta \sin \omega_m t) \exp (j\omega_c t)\},$$

which can be rewritten as

$$\phi(t) \approx \mathscr{Re}\left\{A\left(1 + \frac{m}{2}e^{j\omega_m t} + \frac{m}{2}e^{-j\omega_m t}\right)\left(1 + \frac{\beta}{2}e^{j\omega_m t} - \frac{\beta}{2}e^{-j\omega_m t}\right)e^{j\omega_c t}\right\}.$$

If $\beta \ll 1$ and $m \ll 1$, then the second-order term $(m\beta)$ in the approximation is very small. Neglecting these second-order effects, we get

$$\phi(t) \approx \mathscr{Re}\left\{A\left[1 + \left(\frac{m+\beta}{2}\right)e^{j\omega_m t} + \left(\frac{m-\beta}{2}\right)e^{-j\omega_m t}\right]e^{j\omega_c t}\right\}.$$

Note that, in contrast to AM, the two sidebands may have unequal magnitudes if some FM is also present. Now if we set $\beta = m$, we obtain the (approximate) SSB-LC signal

$$\phi(t) \approx \mathscr{Re}\{A[1 + me^{j\omega_m t}]e^{j\omega_c t}\},$$

or,

$$\phi(t) \approx A \cos \omega_c t + mA \cos (\omega_c + \omega_m)t.$$

DRILL PROBLEM 6.2.2

Sometimes when AM is the desired modulation, a combination of AM and FM can actually occur as a result of an imperfect modulator. Combined AM and NBFM is characterized on a spectrum-analyzer display by two sidebands of unequal amplitude. This arises because the AM sidebands are of the same phase but the NBFM sidebands are of opposite phase. Because the intended modulation is AM, the incidental FM introduced is assumed to be the smaller of the two effects.

As an example, suppose that a spectrum-analyzer measurement of the output of a modulator using sinusoidal modulation indicates a carrier line of unit magnitude, an upper sideband line of magnitude 0.45 and a lower sideband line of magnitude 0.35. Calculate the percentage AM and the percentage FM present using the result of Example 6.2.1.

ANSWER: 80%; 10%.

6.3 WIDEBAND FM

Up to this point we have relied heavily on the use of the Fourier transform of a general signal $f(t)$ to give us the spectral density $F(\omega)$. However, if the value of β is not small, the Fourier transform of a general angle-modulated waveform cannot be evaluated. For specific cases the integration can be performed numerically or in terms of tabulated values. Therefore we shall first try to establish some bounds on the spectral density before we are forced to restrict the analysis to a few given modulating signal waveforms.

A measure of the peak amplitude-to-frequency conversion is the peak frequency deviation, $\Delta\omega$. This represents the maximum amount that ω_i deviates from the "average" value of ω_c. This is demonstrated for two differing cases in Fig. 6.5.

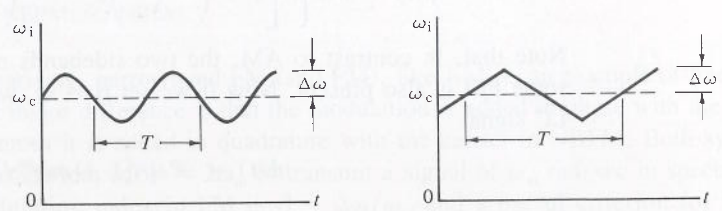

Figure 6.5 Definition of maximum (peak) frequency deviation.

There are two identifiable mechanisms in the description of the spectrum of an FM waveform. The first is attributable to the rate of change of the modulating signal; i.e., the frequency content of the modulating signal. The second effect, peculiar to FM, is the proportionality between the amplitude of the modulating signal and the instantaneous frequency of the FM signal. The instantaneous frequency follows the amplitude of the modulating signal, but this does not imply necessarily that the spectral density follows the same pattern. The concept of instantaneous frequency and the frequency used in the Fourier transform are not identical.

In the NBFM approximation, it is seen that the second effect was neglected in favor of the first since $\Delta\omega \ll \omega_m$. In fact, we now see that for the sinusoidal case the

modulation index $\beta = \Delta\omega/\omega_m$ gives us a relative measure as to the importance of these two effects in FM.

The idea of a modulation index can be extended to more general waveforms. For a general pulse waveform we can define a peak frequency deviation $\Delta\omega$ and a time duration T; if the waveform is periodic, then T is the period. The product $\beta_1 = (\Delta\omega/2\pi)T$ is a dimensionless number, called a *dispersion index*, which takes the place of the modulation index for more general modulation waveforms. It is easy to see that $\beta_1 \to \beta$ for sinusoidal modulation. For very low dispersion indexes the spectral content of a modulating signal largely controls the magnitude of the FM spectral density. For very high dispersion indexes, the amplitude-to-frequency conversion largely controls the magnitude spectral density. Phase effects are not as predictable because they depend on the relative phasing between signals. What happens for intermediate values must be examined on the basis of each given type of signal.

Returning to the sinusoidal case of FM, let $f(t) = a \cos \omega_m t$ and $\omega_i = \omega_c + \Delta\omega \cos \omega_m t$. The spectral content of the modulating signal is at ω_m rad/sec. The peak amplitude-to-frequency conversion is $\Delta\omega$ rad/sec. Then for very low values of $\beta = \Delta\omega/\omega_m$ (that is, $\Delta\omega \ll \omega_m$) the spectrum will be band-limited to $2\omega_m$. On the other hand, for very high values of $\beta = \Delta\omega/\omega_m$ (that is, $\Delta\omega \gg \omega_m$) the amplitude-to-frequency conversion will predominate and we would expect the bandwidth to be on the order of $2\Delta\omega$. We therefore have some rather intuitive bounds on the bandwidth at both extremes.

6.3.1 ★ General Approximations

Another general intuitive comment can be made here before restricting ourselves to specific waveforms. If we let $\beta_1 \to \infty$ ($\beta \to \infty$ for the sinusoidal case), we would expect the amplitude-to-frequency conversion to completely predominate. From the concept of a spectral density we would then expect the spectral magnitudes to be in proportion to the fractional time spent at each frequency.† For example, let $f(t) = a \cos \omega_m t$ so that the frequency deviation about the carrier, $\omega_i' = \omega_i - \omega_c$, is

$$\omega_i' = \Delta\omega \cos \omega_m t\,, \tag{6.27}$$

or

$$t = \frac{1}{\omega_m} \cos^{-1}\left(\frac{\omega_i'}{\Delta\omega}\right) \quad \text{for } |\omega_i'| \le \Delta\omega\,. \tag{6.28}$$

The fractional amount of time per unit of frequency is‡:

$$\frac{1}{T}\left|\frac{dt}{d\omega_i'}\right| = \frac{1/(2\pi)}{\Delta\omega\sqrt{1 - (\omega_i'/\Delta\omega)^2}} \quad \text{for } |\omega_i'| \le \Delta\omega\,. \tag{6.29}$$

† This is sometimes referred to as Woodward's theorem.

‡ The reader with some knowledge of probability will recognize this as the probability density function of the modulating waveform for uniform phase.

Therefore as $\beta_1 \to \infty$ (in this case $\beta \to \infty$) the magnitude weighting of the spectral density of the FM waveform will approach the shape shown in Fig. 6.6 over band limits $2\Delta\omega$ in width. Note that this is based on a signal T units long; for the periodic case the spectral density will be composed of impulses with weights determined from this curve. Effects of phase may cause the individual components to vary somewhat from this approximation. This result gives a relative distribution of power; the correct scaling factors can be found by setting the integral of this result equal to the average power in the modulated waveform.

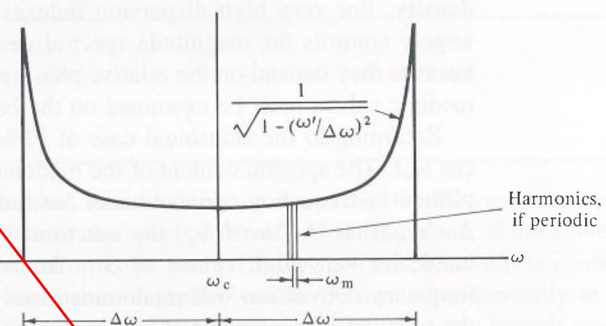

Figure 6.6 Approximation to the magnitude FM spectral density as $\beta \to \infty$, sinusoidal case.

EXAMPLE 6.3.1

A sinusoidal signal at a frequency of ω_c rad/sec is frequency-modulated by the saw-tooth waveform shown in Fig. 6.7(a). The peak frequency deviation on each side of the carrier is $\Delta\omega$ rad/sec, as shown in Fig. 6.7(b). Describe the approximate magnitude spectral density as the dispersion index of the system becomes very large.[†]

Solution As $\beta_1 \to \infty$, the bandwidth approaches $2\Delta\omega$ and the magnitude spectrum is approximated by

$$\omega_i' = \frac{\Delta\omega}{T/2}t, \qquad -T/2 < t < T/2,$$

$$t = \frac{T}{2\Delta\omega}\omega_i',$$

$$\frac{1}{T}\left|\frac{dt}{d\omega_i'}\right| = \frac{1}{2\Delta\omega}, \qquad -\Delta\omega < \omega_i' < \Delta\omega.$$

[†] This is a simplified version of the type of modulation that a bat uses (at ultrasonic frequencies) for navigation and target location. It is also used for radar purposes.

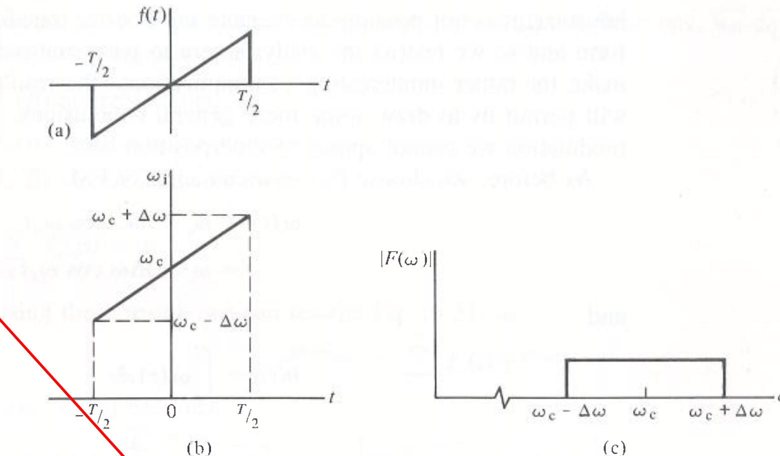

Figure 6.7 Example of FM spectral density as $\beta_1 \to \infty$.

This is shown in Fig. 6.7(c). If the modulating signal were repeated periodically, a series of impulses would be present spaced by ω_0 units. A numerical example of the computation of a magnitude spectrum for $\beta_1 = 50$ is shown in Fig. 6.8.

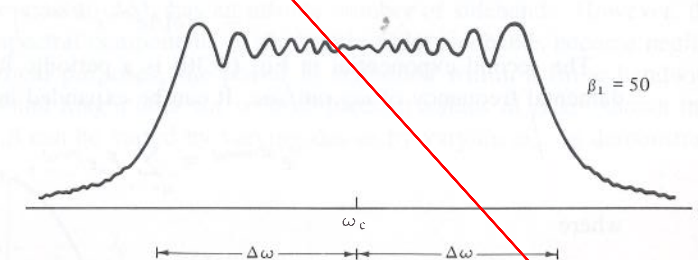

Figure 6.8 Computed magnitude spectrum for the FM discussed in Example 6.3.1.

DRILL PROBLEM 6.3.1

A sinusoidal signal at a frequency of ω_c rad/sec is frequency-modulated by a symmetrical square wave. The peak frequency deviation is $\Delta\omega$. Describe the approximate magnitude spectral density as the dispersion index becomes very large.

ANSWER: $\frac{1}{2}\delta(\omega - \omega_c + \Delta\omega) + \frac{1}{2}\delta(\omega - \omega_c - \Delta\omega)$.

6.3.2 Sinusoidal Case

Having obtained some intuitive insights into the mechanisms of FM, we now seek to extend our knowledge by using the Fourier transform. As pointed out earlier, however, it is not possible to evaluate the Fourier transform of a general FM waveform and so we restrict the analysis here to pure sinusoids. Although pure sinusoids make for rather uninteresting communications, the results of the analysis hopefully will permit us to draw some more general conclusions. Because FM is a nonlinear modulation we cannot appeal to superposition here.

As before, we choose $f(t) = a \cos \omega_m t$; for FM,[†]

$$\omega_i(t) = \omega_c + ak_f \cos \omega_m t$$
$$= \omega_c + \Delta\omega \cos \omega_m t\,,$$

and

$$\theta(t) = \int_0^t \omega_i(\tau)\, d\tau$$
$$= \omega_c t + \frac{\Delta\omega}{\omega_m} \sin \omega_m t$$
$$= \omega_c t + \beta \sin \omega_m t\,.$$

Using complex notation,

$$\phi_{FM}(t) = \mathcal{R}e\{Ae^{j\theta(t)}\}$$
$$= \mathcal{R}e\{Ae^{j\omega_c t}e^{j\beta \sin \omega_m t}\}\,. \tag{6.30}$$

The second exponential in Eq. (6.30) is a periodic function of time with a fundamental frequency of ω_m rad/sec. It can be expanded in a Fourier series,

$$e^{j\beta \sin \omega_m t} = \sum_{n=-\infty}^{\infty} F_n e^{jn\omega_m t}\,, \tag{6.31}$$

where

$$F_n = \frac{1}{T}\int_{-T/2}^{T/2} e^{j\beta \sin \omega_m t} e^{-jn\omega_m t}\, dt\,. \tag{6.32}$$

Making a change of variable $\xi = \omega_m t = (2\pi/T)t$, we get

$$F_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j(\beta \sin \xi - n\xi)}\, d\xi\,. \tag{6.33}$$

This integral can be evaluated numerically in terms of the parameters n and β, and because it occurs in many physical problems it has been tabulated extensively. It is a function of n and β, denoted by $J_n(\beta)$, and is called the Bessel function of the first kind (signified by the "J") of order n and argument β.[‡] Note that in our case n is an integer (negative and positive) and β is a continuous variable (positive values only). Some of these functions are plotted in Fig. 6.9. Though we do not wish to get involved in discussing the detailed characteristics of Bessel functions, we do have use for the following properties:

[†] A constant term is introduced if the lower limit in the integral is not zero; this does not change the analysis and will be omitted for convenience.
[‡] A table of Bessel functions is given in Appendix G.

1. $J_n(\beta)$ are real valued,
2. $J_n(\beta) = J_{-n}(\beta)$, for n even,
3. $J_n(\beta) = -J_{-n}(\beta)$, for n odd, (6.34)
4. $\sum_{n=-\infty}^{\infty} J_n^2(\beta) = 1$.

Using these results we can rewrite Eq. (6.31) as

$$e^{j\beta \sin \omega_m t} = \sum_{n=-\infty}^{\infty} J_n(\beta) e^{jn\omega_m t}\,, \tag{6.35}$$

and Eq. (6.30) becomes

$$\phi_{FM}(t) = \mathcal{R}e\left\{Ae^{j\omega_c t}\sum_{n=-\infty}^{\infty} J_n(\beta)e^{jn\omega_m t}\right\}, \tag{6.36}$$

$$\phi_{FM}(t) = A\sum_{n=-\infty}^{\infty} J_n(\beta)\cos(\omega_c + n\omega_m)t\,. \tag{6.37}$$

This can be expanded, if desired, using properties (2) and (3) of Eq. (6.34).

From these results, it is evident that an FM waveform with sinusoidal modulation, in contrast to AM, has an infinite number of sidebands. However, the magnitudes of the spectral components of the higher-order sidebands become negligible and, for all practical purposes, the power is contained within a finite bandwidth. Plots of the sideband magnitudes for several different values of β are shown in Fig. 6.10. Note that β can be varied by varying $\Delta\omega$ or by varying ω_m, as demonstrated in Fig. 6.10.

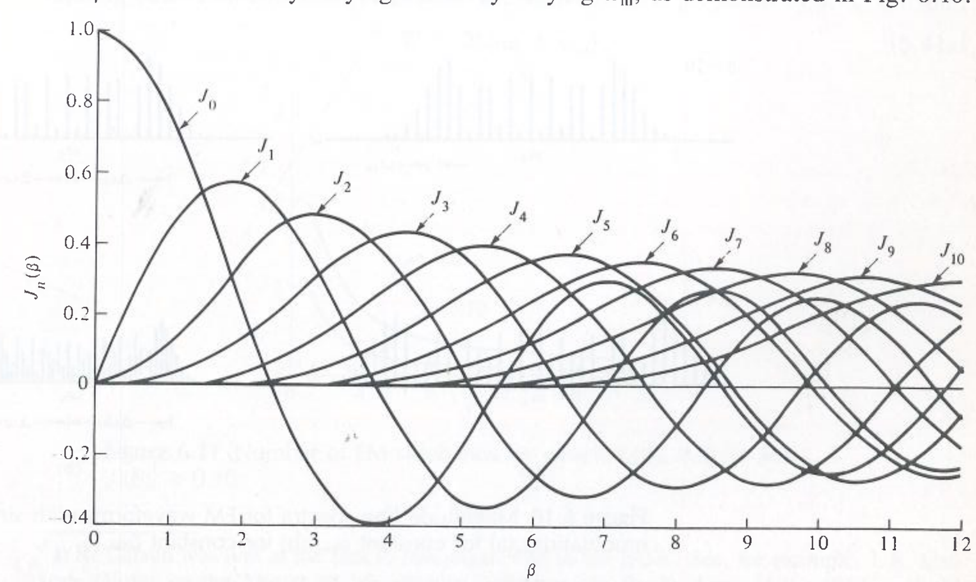

Figure 6.9 Plot of Bessel function of the first kind, $J_n(\beta)$.

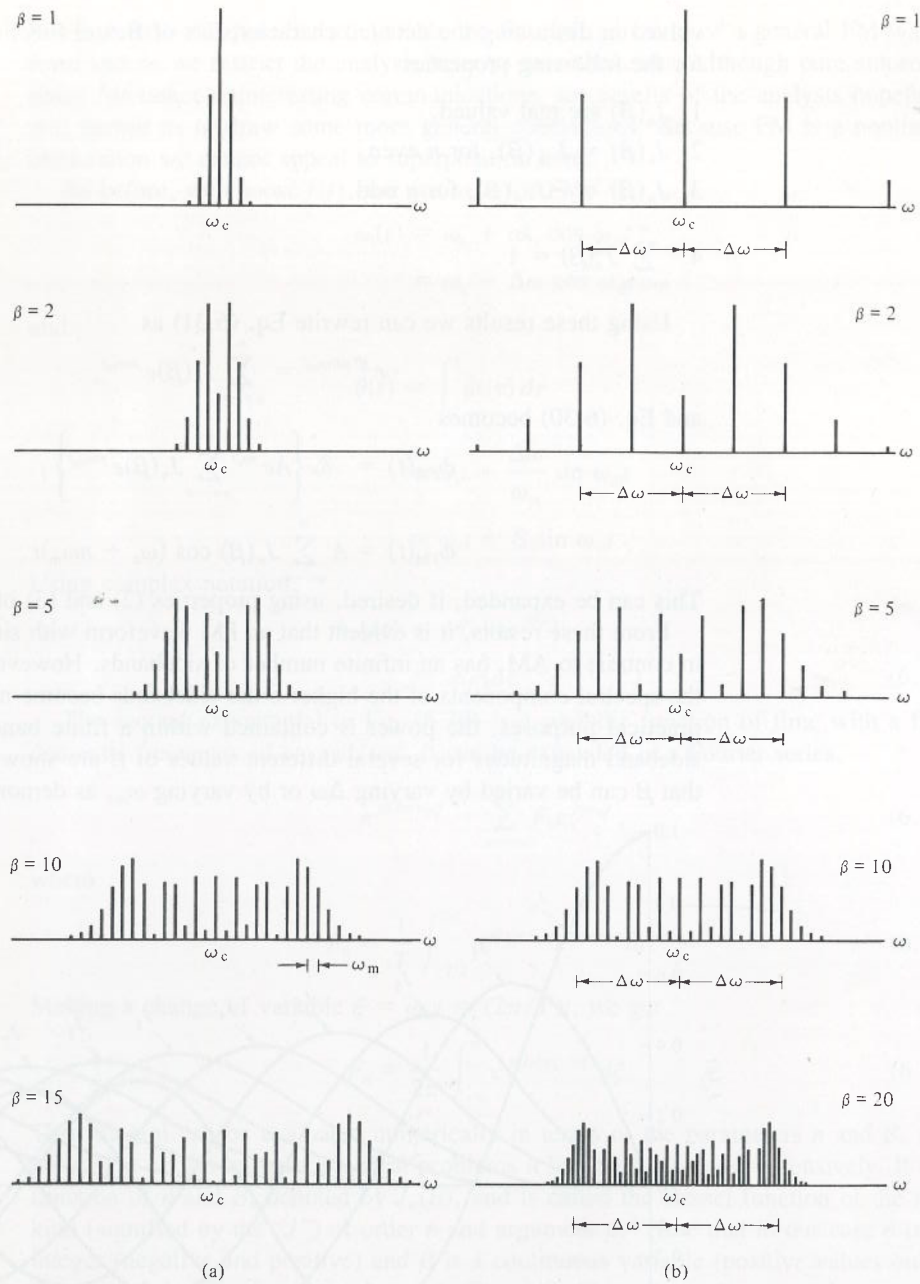

Figure 6.10 Magnitude line spectra for FM waveforms with sinusoidal modulation: (a) for constant ω_m; (b) for constant $\Delta\omega$.

How many sidebands are important to the FM transmission of a signal? This will depend on the intended application and the fidelity requirements. A rule commonly adopted is that a sideband is *significant* if its magnitude is equal to or exceeds 1% of the unmodulated carrier, i.e., if

$$|J_n(\beta)| \geq 0.01\,. \tag{6.38}$$

The actual number of significant sidebands for different values of β can be found from a plot or a table of Bessel functions. It can be seen from the plots of Fig. 6.10 that the $J_n(\beta)$ diminish rapidly for $n > \beta$, particularly as β becomes large. A graph of the ratio n/β for $|J_n(\beta)| \geq 0.01$ is shown in Fig. 6.11 and it is seen that the ratio approaches one as β becomes very large. The bandwidth for very large β can then be approximated by taking the last significant sideband at $n = \beta$ so that

$$W = 2n\omega_m \approx 2\beta\omega_m = 2\frac{\Delta\omega}{\omega_m}\omega_m\,,$$

or

$$W \approx 2\Delta\omega \quad \text{for large } \beta. \tag{6.39}$$

For very small values of β, the only Bessel functions of significant magnitude (see Fig. 6.9) are $J_0(\beta)$ and $J_1(\beta)$. Therefore the bandwidth for the narrowband case (verifying an earlier result) is

$$W \approx 2\omega_m \quad \text{for small } \beta. \tag{6.40}$$

We now have bounds on the limiting cases. It would be convenient to have a more general rule to take care of the intermediate cases and, if possible, also approach the limiting cases in a continuous manner. One such rule was proposed by J. R. Carson†:

$$W \approx 2(\Delta\omega + \omega_m)\,, \tag{6.41a}$$

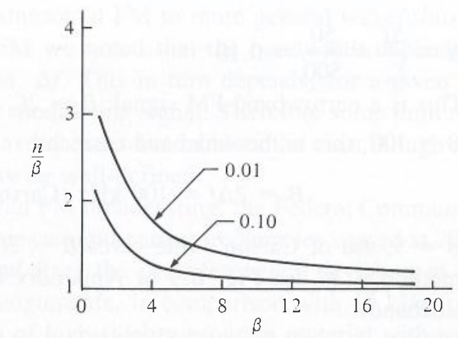

Figure 6.11 Number of FM sidebands for which $|J_n(\beta)| \geq 0.01$ and $|J_n(\beta)| \geq 0.10$.

† J. R. Carson was one of the first to investigate FM in the 1920s. See, for example, J. R. Carson, "Notes on the Theory of Modulation," reprinted in *Proceedings of the IEEE*, vol. 51 (1951): 893–896.

which can also be written as

$$W \approx 2\omega_m(1 + \beta)\,. \tag{6.41b}$$

Carson's rule approaches the correct limits for both very large and very small β; it is widely used in practice because it gives a very convenient approximation that is reasonably accurate. It always gives less bandwidth than our definition of significant sidebands, with a maximum bandwidth error in the neighborhood of $\beta = 1$. The average power in the sidebands neglected, however, is small and less than 1% of the total average power in the FM waveform. In fact, the approximation is good enough that we now release the restriction that the modulating signal be purely sinusoidal and make the wide generalization that Carson's rule holds for general modulating signals that are band-limited and have finite power. An intuitive justification for this is that the two terms in Carson's rule display the effects of the two mechanisms in the generation of FM, and that these effects on the bandwidth are additive. It works surprisingly well.

The analytical method used here in expanding in a Fourier series is a very powerful one, and can be used for more general periodic modulating signals. This is explored in problems at the end of this chapter. Because we must integrate the modulating signal in FM to obtain the phase, we require that the modulating signal for FM have zero mean value.

EXAMPLE 6.3.2

A 10 MHz carrier is frequency-modulated by a sinusoidal signal such that the peak frequency deviation is 50 kHz. Determine the approximate bandwidth of the FM signal if the frequency of the modulating sinusoid is (a) 500 kHz; (b) 500 Hz; (c) 10 kHz.

Solution

a) $\beta = \dfrac{\Delta f}{f_m} = \dfrac{50}{500} = 0.10$

This is a narrowband FM signal; $B \approx 2f_m = 1$ MHz.

b) $\beta = 100$; this is the wideband case and

$$B \approx 2\Delta f = 100 \text{ kHz (Carson's rule gives 101 kHz)}.$$

c) $\beta = 5$; use of Carson's rule gives $B \approx 2(\Delta f + f_m) = 120$ kHz. A more accurate method is to use Fig. 6.9 or Appendix G to find the number n of significant sidebands:

$$B = 2nf_m = 2(8)(10 \text{ kHz}) = 160 \text{ kHz}.$$

A magnitude line spectrum for $\beta = 5$ is shown in Fig. 6.10; in this case the spacing between lines would be 10 kHz.

DRILL PROBLEM 6.3.2

Repeat Example 6.3.2 if the peak frequency deviation were decreased to 20 kHz.

ANSWER: (a) 1 MHz; (b) 41 kHz; (c) 80 kHz (60 kHz if you use Carson's rule).

DRILL PROBLEM 6.3.3

A given FM signal is

$$\phi_{FM}(t) = 10 \cos\left[10^6\pi t + 8 \sin(10^3\pi t)\right].$$

Determine the following: (a) the carrier frequency, f_c; (b) the modulation index, β; (c) the peak frequency deviation, Δf.

ANSWER: (a) 500 kHz; (b) 8; (c) 4 kHz.

6.3.3 Commercial FM Transmissions

As noted earlier, narrowband FM is linear and therefore much of the analysis for AM applies. Advantages in using narrowband FM over AM include the possibility of a response to zero Hz (important in telemetry and recording) and the rejection of large noise pulses (as a result of clipping, or limiting, the amplitude of the waveform) which may tend to saturate the receiver. Narrowband FM is used primarily in telemetry and mobile communications.

Provided that we are content with only the bandwidth, we can apply our knowledge of purely sinusoidal FM to more general waveforms also in the wideband case. For wideband FM we noted that the bandwidth depended mainly on the peak frequency deviation, Δf. This in turn depends, for a given modulator constant, on the amplitude of the modulating signal. Therefore some limit must be placed on the modulating signal to avoid excessive bandwidths even though the bandwidth of the modulating signal may be well-defined.

For commercial FM broadcasting, the Federal Communications Commission (FCC) in the United States assigns carrier frequencies spaced at 200 kHz intervals in the range 88–108 MHz and fixes the peak frequency deviation at 75 kHz.[†] The 200 kHz between station assignments, in comparison with 10 kHz for AM broadcasting, allows the transmission of high-fidelity program material with room to spare, and wideband FM is used to fill the band. Suppose we take the modulating frequency f_m to be 15 kHz (typically the maximum audio frequency in FM transmissions). Use of Carson's rule then yields a bandwidth of $B \approx 2(\Delta f + f_m) = 180$ kHz, well within the required

[†] See Appendix C.

bandwidth. Our sinusoidal analysis indicates that $\beta = 5$ and the bandwidth occupied by significant sidebands is

$$2(8)(15\text{ kHz}) = 240\text{ kHz}$$

(see Fig. 6.9, 6.10, or 6.11). The discrepancy, of course, lies in the definition of bandwidth. However, we chose an extreme case as far as typical audio transmission is concerned, because we assumed that the 15 kHz tone was set at the maximum amplitude to produce a peak frequency deviation of 75 kHz. Typical program material does not contain as much at the higher frequencies. For lower audio frequencies the value of β increases and the bandwidth occupied by the significant sidebands approaches the wideband limit of $2\Delta f = 150$ kHz. (For audio signals with full maximum amplitude and frequencies below about 5 kHz all significant sidebands are within the 200 kHz bandwidth.) Note that if the amplitude weighting is uniform, it is the highest modulating frequency that governs the final bandwidth.

The transmission of one audio channel leaves room for additional program material within the bandwidth allocated. Stereo multiplexing and other auxiliary transmissions often occupy the higher frequency portions of the modulating spectrum. These were discussed in Chapter 5. To keep the bandwidth restricted, the maximum amplitude of these transmissions is reduced. The spectrum of a typical commercial transmission before the FM transmitter is shown in Fig. 5.12.

In the FM station, the left (L) and right (R) audio signals are derived from microphones, records, tapes, etc., and a preemphasis is applied to each channel (this is discussed later in this chapter). For stereo broadcasts, a pilot subcarrier at 19 kHz is permitted 10% of the total peak-frequency deviation (of 75 kHz). When there is a pause in program material (and no auxiliary transmissions), the modulation index is $\beta = (10\%)(75\text{ kHz})/(19\text{ kHz}) = 0.395$. This is approximately in the narrowband condition. Thus when there is a pause in program material, a stereo FM broadcast can be identified on a spectrum analyzer by a large carrier line plus two first-order sidebands each spaced 19 kHz from the carrier.

The Subsidiary Communications Authorization (SCA) system permits a commercial FM station to add another broadcasting channel in addition to the monophonic and stereo channels. The SCA transmissions carry no commercial messages and are intended for private subscribers who pay a fee for background music in stores, physicians' offices, etc. In contrast, the other FM transmissions are for general public use and are supported by commercial advertisements. The SCA channel uses narrowband FM. The subcarrier center frequency is usually set at 67 kHz, although this choice is not set by the FCC. A total peak-frequency deviation not exceeding 75 kHz is still required. For monophonic transmission only, this entire 75 kHz is available. If SCA is used with mono, the FCC limits the SCA portion to 30% of the maximum peak-frequency deviation, leaving 70% for the mono channel.

In stereo broadcasting with no SCA, 10% of the maximum peak deviation is used for the 19 kHz pilot subcarrier, leaving 90% to be divided between the (L + R) and (L − R) stereo channels. The average amplitudes of the L and R channels are normally kept equal. The maximum (L + R) amplitude is set to provide 90% modulation when the (L − R) amplitude is zero. Now if (L − R) is maximum, then (L + R) will go to zero, and if either L or R goes to zero, (L + R) and (L − R) will each take a maximum of 45% of the total peak frequency deviation. Thus there is a seesaw effect between the (L + R) and (L − R) channels such that the total does not exceed 90% of the peak frequency deviation capability allowed.

When used with stereo multiplexing, the SCA channel is limited to 10% of the maximum peak frequency deviation. This leaves 80% for the stereo channels (i.e., 10% for SCA plus 10% for the 19 kHz pilot subcarrier). The system performance is the same as in the preceding paragraph except that now the stereo channels are allowed 80% of the maximum peak frequency deviation instead of 90%. Also, with the relatively low peak frequency deviation allowed, the SCA transmission does not have a very good signal-to-noise ratio and is used for only local coverage. Station muting (e.g., dropping the subcarrier to actuate an audio silencer circuit in the receiver) is often employed to silence the noise between records or tapes in the SCA transmissions.

Public FM broadcasting is an example of the use of DSB-SC and NBFM methods to frequency-multiplex several channels before using wideband FM for the final transmission. Frequency modulation is also used for the audio in commercial television transmissions. The peak frequency deviation for this use is fixed at 25 kHz by the FCC. Assuming a maximum audio frequency of 15 kHz, use of Carson's rule gives a bandwidth of 80 kHz for the sound channel of a television receiver.

The relatively large bandwidth required for commercial FM, as compared with AM, is the penalty for obtaining substantial improvement in noise and interference rejection. This noise rejection increases with increasing Δf and, therefore, with increasing bandwidth. These topics will be discussed in a later section.

6.4 AVERAGE POWER IN ANGLE-MODULATED WAVEFORMS

For sinusoidal modulation, we can write [cf. Eq. (6.16)]

$$\phi_{FM}(t) = A\cos(\omega_c t + \beta\sin\omega_m t)\,.$$

The mean-square value of this expression is

$$\overline{\phi_{FM}^2(t)} = A^2/2\,, \tag{6.42}$$

showing that the total average power in an FM waveform is a constant regardless of the modulation index. This is in contrast to AM where the total average power was proportional to the modulation index. This conclusion can be extended to any arbitrary band-limited modulating waveform.

Equation (6.42) can be verified by writing $\phi_{FM}(t)$ in a series expansion [cf. Eq. (6.37)],

$$\phi_{FM}(t) = A\sum_{n=-\infty}^{\infty} J_n(\beta)\cos(\omega_c + n\omega_m)t\,.$$

As a result of the orthogonality of the cosine terms, the mean-square value of the sum is equal to the sum of the mean-square values and we get

$$\overline{\phi_{\mathrm{FM}}^2(t)} = \tfrac{1}{2}A^2 \sum_{n=-\infty}^{\infty} J_n^2(\beta) . \tag{6.43}$$

But from property (4) in Eq. (6.34),

$$\sum_{n=-\infty}^{\infty} J_n^2(\beta) = 1 ,$$

so that

$$\overline{\phi_{\mathrm{FM}}^2(t)} = A^2/2 .$$

The mean-square value of the unmodulated carrier is $A^2/2$. As the modulation index β is increased from zero and the sidebands are nonzero, the carrier component decreases. According to Eqs. (6.42) and (6.43), this takes place in such a manner as to always keep the total mean-square value constant. The mean-square value of each sideband is $\frac{1}{2}A^2J_n^2(\beta)$ (recall also that sidebands occur in pairs). The mean-square value, of course, is identical to the average power if the resistance is 1 ohm and is related to the average power by a constant (i.e., the resistance) for all other cases, so the conversion to units of power is straightforward.

It is possible to make any particular sideband, including the carrier, as small as desired by a proper choice of the modulation index β. From a table or graph of Bessel functions (e.g., Fig. 6.9), we see that the carrier term, $J_0(\beta)$, can be made zero for $\beta = 2.405, 5.52, \ldots$, and in these cases all of the average power is in the sidebands. These points are easy to read with a spectrum analyzer and serve as very convenient calibration points for β and Δf.

EXAMPLE 6.4.1

A given FM transmitter is modulated with a single sinusoid. The output for no modulation is 100 watts into a 50-ohm resistive load. The peak frequency deviation of the transmitter is carefully increased from zero until the first sideband amplitude in the output is zero. Under these conditions determine (a) the average power at the carrier frequency; (b) the average power in all the remaining sidebands; and (c) the average power in the second-order sidebands.

Solution a) Using Fig. 6.9 and Appendix G, we see that $J_1(\beta) = 0$ first occurs at $\beta \approx 3.8$ and that $J_0(3.8) \approx -0.40$. The average carrier power is then

$$P_c = \frac{J_0^2(3.8)}{J_0^2(0)}(100 \text{ W}) = 16 \text{ W}.$$

b) The average power in the sum of the remaining sidebands is

$$P_s = P_t - P_c = 100 \text{ W} - 16 \text{ W} = 84 \text{ W}.$$

c) $J_2(3.8) \approx 0.41$. The average power in the second-order sidebands is

$$P_c = 2\frac{J_2^2(3.8)}{J_0^2(0)}(100 \text{ W}) = 34 \text{ W}.$$

DRILL PROBLEM 6.4.1

Determine the peak amplitude of (a) the total waveform and (b) the upper second-order sideband in Example 6.4.1.

ANSWER: (a) 100 V; (b) 41 V.

DRILL PROBLEM 6.4.2

Show that the rms value of Eq. (6.16) can be written as

$$\sqrt{\overline{\phi^2(t)}} = \frac{A}{\sqrt{2}}\sqrt{J_0^2(\beta) + 2\sum_{n=1}^{\infty} J_n^2(\beta)} .$$

6.5 PHASE MODULATION

There is no basic difference between the mechanisms involved in the generation of phase modulation (PM) and frequency modulation (FM). In fact, the only difference is that the phase in the modulated waveform is proportional to the input signal amplitude in PM and to the integral of the input signal in FM. This introduces only a slight modification and we shall point that out here.

For an FM signal with the sinusoidal modulation $f(t) = a \cos \omega_m t$, the instantaneous frequency is

$$\omega_i(t) = \omega_c + ak_f \cos \omega_m t$$

$$= \omega_c + \Delta\omega \cos \omega_m t ,$$

where $\Delta\omega$ is the peak frequency deviation (in radians per second) and k_f is the frequency-modulator constant (in radians per second per volt). The modulation index, $\beta = \Delta\omega/\omega_m$, is a dimensionless number and serves as a guide to the behavior of the carrier and sidebands.

For PM with the same modulating signal we have

$$\theta(t) = \omega_c t + ak_p \cos \omega_m t + \theta_0$$
$$= \omega_c t + \Delta\theta \cos \omega_m t + \theta_0 ,$$

where $\Delta\theta$ is the peak phase deviation (in radians) and k_p is the phase-modulator constant (in radians per volt). The instantaneous frequency is

$$\omega_i(t) = \frac{d\theta}{dt}$$
$$= \omega_c - ak_p\omega_m \sin \omega_m t$$
$$= \omega_c - \Delta\omega \sin \omega_m t .$$

Thus we see that the peak frequency deviation in PM is proportional not only to the amplitude of the modulating waveform but also to its frequency; that is,

$$\Delta\omega = \begin{cases} ak_f & \text{for FM} \\ ak_p\omega_m = (\Delta\theta)\omega_m & \text{for PM} \end{cases} \tag{6.44}$$

This makes PM less desirable to transmit when $\Delta\omega$ is fixed (as in commercial FM). There are some advantages in the demodulation of PM, however, which make its use desirable. (These will become more evident later in this chapter.) The role of the modulation index β remains the same as in FM. Formally, then, we can compute $\Delta\omega = ak_p\omega_m = \Delta\theta\omega_m$ and then proceed as if the modulation were FM as far as bandwidth, sidebands, etc. are concerned. Note that the numerical value of β is the peak phase deviation, $\Delta\theta$, in the PM case.

EXAMPLE 6.5.1

A carrier is phase modulated by a sinusoidal signal of 5 kHz and unit amplitude and the peak phase deviation is one radian. Calculate the bandwidth of the PM signal (a) using Carson's rule; and (b) using the definition of significant sidebands.

Solution a) $\Delta f = (\Delta\theta)f_m = 5$ kHz and Carson's rule gives

$$B \approx 2(\Delta f + f_m) = 20 \text{ kHz} .$$

b) $\beta = \Delta\theta = 1$; using a table of Bessel functions,

$$B = 2nf_m = 2(3)(5 \text{ kHz}) = 30 \text{ kHz} .$$

DRILL PROBLEM 6.5.1

Here we consider a PM system in which the phase may take on only two possible values—known as *phase-shift-keying* (PSK). Assume that a phase modulator is modulated by a periodic symmetric square wave of unit amplitude. Determine the required value of the peak phase deviation, $\Delta\theta$ $(-\pi/2 \le \Delta\theta < \pi/2)$, such that the spectral component at the carrier frequency is zero in the output.

ANSWER: ±90°.

6.6 GENERATION OF WIDEBAND FM SIGNALS

One method of generating wideband FM signals is to produce a narrowband FM signal first and then to use frequency multiplication to increase the modulation index to the desired range of values. This is known as the indirect method of generating wideband FM signals. A second method—known as the direct method—is to vary the carrier frequency directly with the modulating signal. We shall now examine these two methods.

6.6.1 Indirect FM

Section 6.2 showed that the generation of narrowband PM is relatively easy and that narrowband FM can then be generated by first integrating the modulating signal. However, the modulation index obtainable by use of this method is restricted to very low values ($\beta < 0.2$ in theory; $\beta < 0.5$ in practice). To generate wideband FM, a method of increasing the modulation index must be used in this approach. The method used is that of the frequency multiplier.

A *frequency multiplier* is a nonlinear device designed to multiply the frequencies of the input signal by a given factor. For example, the input-output characteristic of an ideal square-law device is

$$e_o(t) = ae_i^2(t) . \tag{6.45}$$

If the input signal is the FM signal,

$$e_i(t) = A \cos (\omega_c t + \beta \sin \omega_m t) ,$$

the output is

$$e_o(t) = aA^2\cos^2 (\omega_c t + \beta \sin \omega_m t)$$
$$= (1/2)aA^2[1 + \cos (2\omega_c t + 2\beta \sin \omega_m t)] . \tag{6.46}$$

The first term in this result is simply a constant level and is easily removed with a filter. We conclude that both the carrier frequency and the modulation index have been doubled in this process. In a similar manner, use of an nth law device followed by a filter yields a carrier and a modulation index that have been increased by a factor of n. Equivalently, the peak frequency deviation $\Delta\omega$ has been increased by n in the multiplication (ω_m has remained unaltered).

In practice, very abrupt nonlinearities can be generated using special diodes (e.g., the varactor and the step-recovery diode) which yield many harmonic terms. It is possible to multiply by an order of magnitude or more in one step using these techniques. Limitations include the fact that losses incurred in the harmonic generation require additional amplification and small phase instabilities in the multiplication process accumulate and appear as noise in the output. With good design techniques, multiplication factors on the order of 10^3 are achievable with only a few degrees of phase noise.

Use of frequency multiplication increases the carrier of the FM waveform as well as the modulation index. This may result in very high carrier frequencies in order to achieve a given modulation index. To avoid this, frequency converters are often used to control the value of the carrier frequency. The frequency converter is essentially the same as discussed in connection with AM and translates the spectrum of a signal by a given amount but does not alter its spectral content. Block diagrams of the frequency multiplier and the frequency converter are shown in Fig. 6.12. Note carefully the differences between these two operations. In the frequency multiplier all spectral components of the input signal are multiplied by themselves (so that all cross-products are present), whereas in the frequency converter all spectral components of the input signal are multiplied by a sinusoid of a fixed frequency. The former operation spreads the spectral content (this can be verified using the frequency convolution property discussed in Chapter 3) and the latter translates the spectral content in frequency.

The method of obtaining a wideband FM waveform from a narrowband one using frequency multiplication is called the Armstrong indirect FM transmitter.[†] A block diagram of a typical Armstrong-type transmitter is shown in Fig. 6.13. As a result of the multiplication and heterodyne operations, it is difficult in this system to maintain the correct magnitude of carrier to sidebands, and thus it could not be used for an information signal with dc content. This drawback can be remedied with use of the phase-locked loop, a subject introduced in a subsequent section.

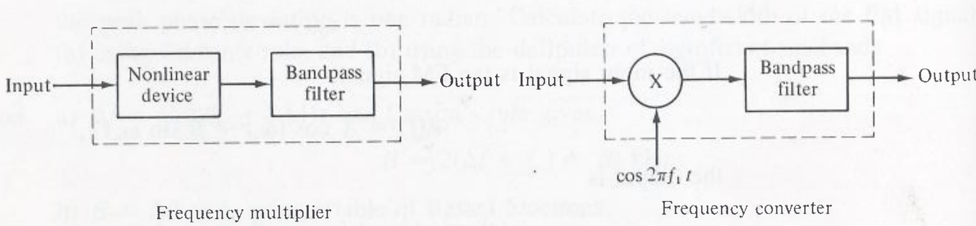

Figure 6.12 Block diagrams of frequency multiplication and frequency conversion.

[†] E. H. Armstrong was one of the first engineers to recognize the possible merits of FM broadcasting in the 1930s. See, for example, E. H. Armstrong, "A Method of Reducing Disturbance in Radio Signaling by a System of Frequency Modulation," *Proceedings of the IRE*, vol. 24 (1936); 689–740.

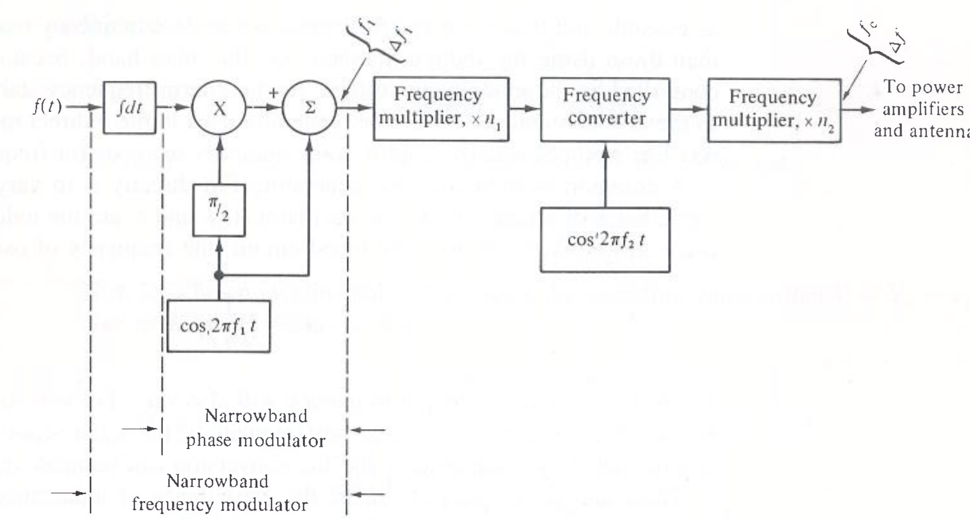

Figure 6.13 Block diagram of an indirect (Armstrong) FM transmitter.

EXAMPLE 6.6.1

A given angle-modulated signal has a peak frequency deviation of 20 Hz for an input sinusoid of unit amplitude and a frequency of 50 Hz. Determine the required frequency multiplication factor, n, to produce a peak frequency deviation of 20 kHz when the input sinusoid has unit amplitude and a frequency of 100 Hz, and the angle-modulation used is (a) FM; (b) PM.

Solution a) $\Delta f_2 = 20$ kHz; $\Delta f_1 = 20$ Hz; $n = \Delta f_2/\Delta f_1 = 1000$

b) $\Delta f_2 = 20$ kHz; $\Delta f_1 = (100/50)(20 \text{ Hz}) = 40$ Hz; $n = \Delta f_2/\Delta f_1 = 500$

DRILL PROBLEM 6.6.1

Compute the carrier frequency f_c and the peak frequency deviation Δf of the output of the FM transmitter shown in Fig. 6.13 if $f_1 = 200$ kHz; $f_2 = 10.8$ MHz; $\Delta f_1 = 25$ Hz; $n_1 = 64$; $n_2 = 48$.

ANSWER: 96.0 MHz or 1132.8 MHz; 76.8 kHz.

6.6.2 Direct FM

In the direct method of generating FM the modulating signal directly controls the carrier frequency. An attempt is usually made to generate as wide a frequency deviation as possible and thus these systems often require less frequency multiplication, if any, than those using the indirect method. On the other hand, because the frequency is controlled by the modulating voltage, the long-term frequency stability is not as good as the crystal-stabilized oscillators generally used in the indirect method. Thus the direct FM methods usually employ some auxiliary methods for frequency stabilization.

A common method used for generating FM directly is to vary the inductance or capacitance of a tuned electronic oscillator. If L and C are the inductance and capacitance, respectively, of a simple tuned circuit, the frequency of oscillation is

$$\omega_0 = \frac{1}{\sqrt{LC}}.$$

If L or C is varied, the output frequency will also vary. For very small variations (that is, $\Delta\omega \ll \omega_c$, where ω_c is the carrier frequency), the square-root relationship can be approximated by a linear term and the conversion can be made quite linear.

There are various ways to make the capacitance or inductance of a tuned circuit dependent on the input signal. One common method at medium and high frequencies is to use a reverse-biased semiconductor diode as a voltage-variable capacitance. Although any semiconductor diode exhibits some capacitance change with change in reverse bias, the type of diode frequently used for this application is the varactor diode. The percentage frequency deviation that can be attained in this manner is quite small. To increase the percentage frequency deviation, the frequency modulation is performed at a high frequency and then heterodyned down to a lower frequency.

Other methods that are used successfully at high frequencies include the reflex klystron and the reactance-tube modulator. The latter consists of a pentode that is operated in such a manner as to produce a capacitance which is proportional to the grid voltage over a wide range. At lower frequencies the control of RC oscillators with FET's and similar devices has been used. Any oscillator whose frequency is controlled by the modulating-signal voltage is called a *voltage-controlled oscillator*, or VCO.

EXAMPLE 6.6.2

A reverse-biased semiconductor diode can be used as a voltage-variable capacitance for frequency modulation. Assume that the capacitance of a given PN junction is given in terms of its reverse-bias voltage V by $C = C_0/\sqrt{1 + 2V}$. Such a diode is to be used as the capacitance in a parallel LC circuit tuned to a center frequency of 10 MHz when the reverse-bias voltage is 4 V.

a) Determine the modulation constant k_f (i.e., the frequency-voltage slope near center frequency).

b) Determine the peak frequency deviation permissible for a maximum error of 1% from a linear frequency-voltage characteristic.

Solution a) We can write the frequency f as

$$f = \frac{1}{2\pi\sqrt{LC}} = \frac{(1 + 2V)^{1/4}}{2\pi\sqrt{LC_0}}.$$

Letting $f = f_0$ when $V = V_0$ (i.e., at the operating point), we get

$$f = f_0\frac{(1 + 2V)^{1/4}}{(1 + 2V_0)^{1/4}}.$$

Now let v be an incremental voltage about the operating point so that $V = V_0 + v$; also let $K = (1 + 2V_0)$ so that

$$f = f_0[1 + (2v/K)]^{1/4}.$$

Assuming that $v \ll K$, we can use the binomial expansion to obtain

$$f \approx f_0\left[1 + \frac{1}{4}\left(\frac{2v}{K}\right) - \frac{3}{32}\left(\frac{2v}{K}\right)^2 + \cdots\right].$$

The modulation constant k_f is the slope of the linear frequency-voltage characteristic and is given by

$$k_f = \frac{1}{4}\left(\frac{2v}{K}\right)\frac{f_0}{v} = \frac{f_0}{2K} = \frac{f_0}{2(1 + 2V_0)}.$$

For the given operating point, $f_0 = 10$ MHz and $V_0 = 4$ volts so that

$$k_f = 0.56 \text{ MHz/V}.$$

b) Most of the error will arise from the second-order term in the series expansion so that we require

$$\frac{\frac{3}{32}\left(\frac{2v}{K}\right)^2}{\frac{1}{4}\left(\frac{2v}{K}\right)} \leq 0.01$$

which gives

$$v \leq \frac{4K}{300}.$$

The peak frequency deviation is then

$$\Delta f = k_f v_{max} = \left(\frac{f_0}{2K}\right)\left(\frac{4K}{300}\right) = \frac{f_0}{150}$$

$$= 66.7 \text{ kHz}.$$

In some cases it is not necessary that the output voltage be sinusoidal, or the output may be wave-shaped by nonlinear shaping circuits or by filtering. In these cases it is possible — and quite attractive — to generate wideband FM digitally. Perhaps the simplest way is to control the oscillation frequency of a relaxation oscillator or a multivibrator with the modulating-signal voltage. A more accurate and stable method is to generate the zero crossings of a PM waveform using digital techniques. Basically this involves sampling the input waveform and then using a precision ramp generator that resets at each sample point, followed by a voltage-variable threshold. The point in time at which the threshold is exceeded is used to generate a short pulse to signify a zero crossing. Applying this sequence of pulses to a bandpass filter results in a wideband PM signal. This method is stable and accurate and is capable of generating signals of very wide bandwidth. A simplified block diagram is shown in Fig. 6.14. These operations are discussed in more detail in Chapter 7.

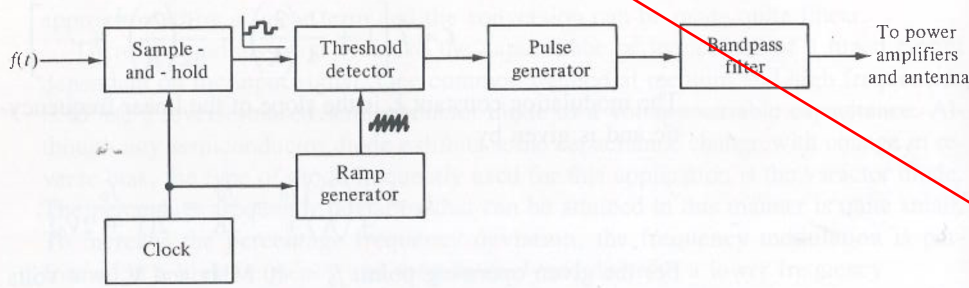

Figure 6.14 Digital generation of wideband PM.

6.6.3 FM Multiplexing

It is a common practice in data transmission to combine several channels of modulated signals using frequency multiplexing methods and then modulate a high-frequency carrier with the composite multiplexed signal. To do this, the individual data signals each modulate an assigned subcarrier. These subcarriers are arranged so that the channels occupy adjacent frequency bands with some frequency space between them, known as *guard bands*. The modulated subcarriers are used to angle-modulate a high-frequency carrier, as shown in Fig. 6.15.

If FM is used for the subcarrier modulation and for the main carrier modulation, the composite modulation is referred to as FM-FM; if AM is used, then it is referred to as AM-FM. The amplitude-modulation methods used for the subcarrier modulation are DSB-SC or SSB-SC. Large-carrier methods are avoided because too much of the peak frequency deviation would be used merely to send the AM carrier. Usually a pilot subcarrier is also sent for demodulation. Note that the stereo multiplexing used in commercial FM is an example of an AM-FM system.†

† Applications of both amplitude modulation and angle modulation to proposed stereo AM transmissions are discussed in Appendix H.

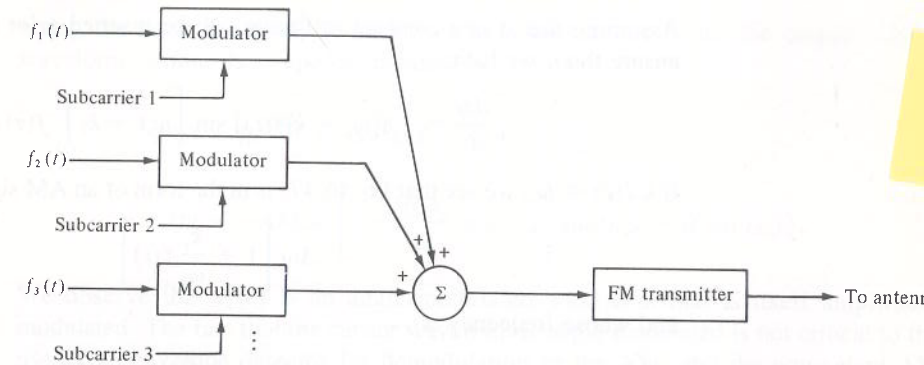

Figure 6.15 A simplified composite modulation system.

Angle modulation (both FM and PM) is widely used in multichannel data transmission and telemetry systems. Standards exist for the assignment of subcarriers and guard bands for the latter.† To allow for realizable filter designs to separate adjacent channels, it is common to allow some frequency separation between channels.

6.7 DEMODULATION OF FM SIGNALS

There are a number of ways to recover the modulating signal from the FM waveform and we shall discuss only some of them. The overall characteristic must be the same, however — to provide an output signal whose amplitude is linearly proportional to the instantaneous frequency of the input waveform.

6.7.1 Direct Method

One method is to use some system that has a linear frequency-to-voltage transfer characteristic. Such a system is called a frequency *discriminator*. In our search for a simple discriminator, we need something with a linear amplitude versus frequency characteristic. The simplest conceptually is that of the ideal differentiator, for we recall that its transfer function is given by $H(\omega) = j\omega$. (Certainly the magnitude characteristic is very linear!)

An expression for the general FM waveform is

$$\phi_{FM}(t) = A \cos\left[\omega_c t + k_f \int_0^t f(\tau)\, d\tau\right].$$

† See, for example, E. L. Gruenberg (ed.), *Handbook of Telemetry and Remote Control*, New York: McGraw-Hill, 1967.