CHAPTER 12.2
FREQUENCY AND PHASE MODULATION

Nobel R. Powell

ANGLE MODULATION

The representation of angle modulation is conveniently made in terms of the notion of the analytic function. For real continuous functions of time $x(t)$, consider

$$m(t) = x(t) + jHx(t)$$

where $x(t)$ = real continuous function
$j = (-1)^{1/2}$
Hx = Hilbert transformation of x

$$Hx(t) = \pi^{-1} \int_{-z}^{z} x(\tau)(t-\tau)^{-1}\, d\tau$$

The angle of $m(t)$ is said to be the angle θ, defined by

$$\theta = \tan^{-1}(Hx/x)$$

Angle modulation may be considered as the change in θ with time or as that portion of the total change in θ which can be associated with the phenomenon of interest. If the relationship between the changes in θ and the effect of interest is direct, the modulation is called *phase modulation* (PM) and the devices producing this relationship *phase modulators*. If the relationship between the changes in the derivative $d\theta/dt$ and the effect of interest is direct, the modulation is called *frequency modulation* (FM), and the devices producing this relationship are called *frequency modulators*.

The derivative $d\theta/dt = \theta$ is related to the components of the analytical function $m(t)$ by

$$\dot{\theta} = \frac{\begin{vmatrix} x & Hx \\ \dot{x} & H\dot{x} \end{vmatrix}}{x^2 + (Hx)^2} \tag{1}$$

for functions $x(t)$ for which the differential and Hilbert operators commute.

Thus in angle modulators, whether implemented functionally in the general form (as with a general-purpose computer) suggested by the representation of θ and $\dot{\theta}$, or as some special combination of electronic networks,

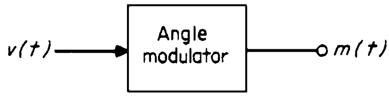

FIGURE 12.2.1 Linear angle modulator.

a direct or proportional relationship is established within the device between either of these two functions and an effect, call it the input $v(t)$, of interest. Diagrammatically, a linear angle modulator can be considered to be a device that transforms $v(t)$, as shown in Fig. 12.2.1.

Demodulators simply perform the inverse of this operation, providing an output function proportional to $\theta(t)$ from a function proportional to the angle of $m(t)$ as an input.

ANGLE-MODULATION SPECTRA

The spectral distribution of power for angle-modulated waveforms varies widely with the nature of the input function $v(t)$.

Random Modulation

For a random process having sample functions

$$x(t) = b \sin(2\pi f t + \phi) \quad b\text{-const}$$

where f and ϕ are independent random variables, ϕ being uniformly distributed over $-\pi \le \phi \le \pi$ and f having a symmetric probability density $p(f)$, this stationary random process has a spectral density function

$$S_{xx}(f) = (b^2/2)p(f) \tag{2}$$

Note that if $p(f)$ is not a discrete distribution, the process is not in general periodic and $S_{xx}(f)$ must be considered continuous.

Periodic Modulation

For a random process having sample functions

$$x(t) = b \cos[\omega_c t - \phi(t) + \theta] \quad b, \omega_c\text{-const}$$

θ uniformly distributed over $-\pi \le \theta \le \pi$. $\phi(t)$ is a stationary process independent of θ; that is,

$$\phi(t) = d \cos(\omega_m t + \theta') \quad d, \omega_m\text{-const}$$

For θ' uniformly distributed over $-\pi \le \theta' \le \pi$, the spectral density function is

$$S_{xx}(f) = (b^2/4)\Bigg(J_0^2(d)[\delta(f - f_c) + \delta(f + f_c)] + n\sum_{n=1}^{\infty} J_n^2(d)\{\delta[f - (f_c \pm nf_m)] + \delta[f + (f_c \pm nf_m)]\}\Bigg) \tag{3}$$

where $J_n(d)$ = nth-order Bessel function of first kind evaluated at d
δ = Kronecker delta function
$S_{xx}(f)$ = Fourier transformation of autocorrelation function of x

Deterministic Modulation

For a function of a completely specified type, e.g.,

$$x(t) = b \sin(\omega_c t + d \sin \omega_m t) \tag{4}$$

frequently it is possible to reexpress $x(t)$ in terms of $J_n(d)$ as

$$x(t) = \sum_{n=-\infty}^{\infty} J_n^2(d) \sin(\omega_c t + n\omega_m t) \tag{5}$$

Examination of Eqs. (1) and (3) and tables of Bessel functions permits the construction of Fig. 12.2.2 showing the implied increase in bandwidth versus modulation index for FM. Modulation index for sinusoidal modulation can be defined as

$$d = |\Delta f|/f_m$$

where $(\Delta f) = df_m$ = amount of instantaneous frequency change $\dot{\theta}$ to be associated with the input $v(t)$.

Such a set of curves can be readily constructed for most deterministic modulation waveforms, since $J_n(d)$ decreases monotonically and rapidly with n for $n > d > 1$. Using the criteria for n indicated for curves A, B, and C, the frequency range (or bandwidth) centered about ω_c, which contains all such spectral components, is indicated. These are the components that are found to be below the value of nf_m for which $J_n(d)$ is monotonically decreasing and equal to the value for each case. Thus the bandwidth (BW) required at the frequency ω_c is

$$\mathrm{BW} = 2df_m(1 + I) \tag{6}$$

Such criteria should be used with care, since the relationship which these measures bear to distortion of the input $v(t)$ when carried at frequency ω_c through linear-tuned circuits as angle modulation is rather indirect.

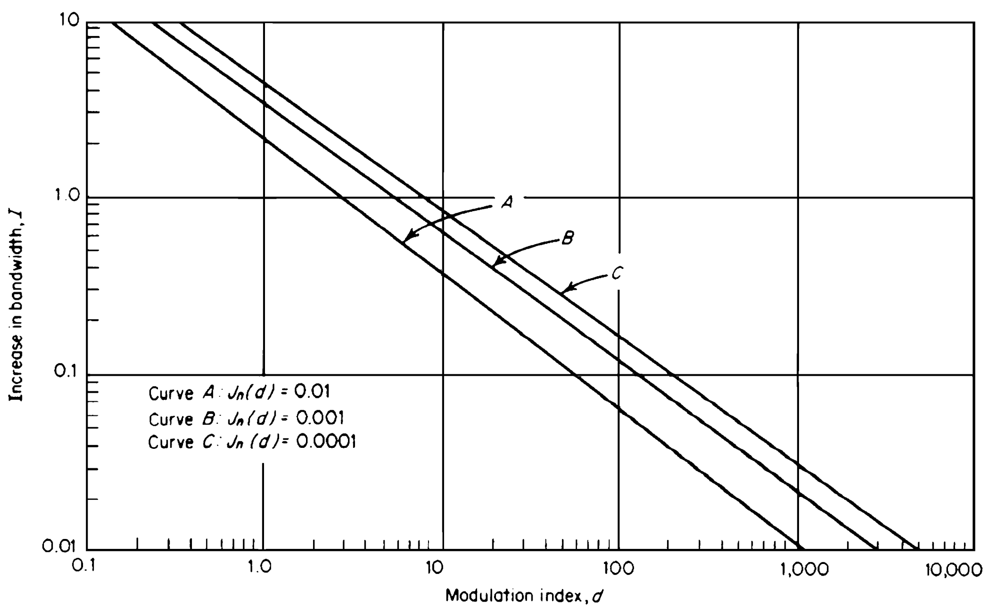

FIGURE 12.2.2 Bandwidth increase vs. modulation index.

ANGLE-MODULATION SIGNAL-TO-NOISE IMPROVEMENT

One of the principal reasons for using frequency modulation in communications and telemetry systems is that it provides a convenient and power-efficient method of trading power for bandwidth while providing high-quality transmission of the input. This is expressed in phase modulation by the relationship

$$\left.\frac{\mathrm{S}}{\mathrm{N}}\right|_{fm} = ad^2 \left.\frac{\mathrm{C}}{\mathrm{N}}\right|_{fm} \tag{7}$$

where $(\mathrm{S/N})|_{fm}$ = demodulated output signal-to-noise ratio measured in a bandwidth f_m
$(\mathrm{C/N})|_{fm}$ = demodulator input signal-to-noise ratio measured in a bandwidth f_m
d = modulation index
a = a constant of proportionality

The constant of proportionality a is unity for sinusoidal phase modulation of constant-amplitude sinusoidal carrier. The constant varies between 0.5 and 3.0 with class of modulation waveforms, type of network compensation, and noise spectrum; however, Eq. (7) can be conservatively applied to FM single-channel voice systems for $a = {}^3/_2$ and preemphasis and deemphasis networks that preshape the spectrum of the modulation to match the sloped noise spectrum and to restore the original spectrum after demodulation.

The trade-off between the signal-to-noise improvement and the required bandwidth corresponding to curve A in Fig. 12.2.2 is shown for FM in Fig. 12.2.3. These curves have been prepared for a conventional demodulator operating at an input carrier-to-noise ratio 1 dB above the threshold (13 dB) measured in the input noise bandwidth to the demodulator, with a constant of proportionality a equal to 0.5. Output signal-to-noise ratio referred to the output information bandwidth $(\mathrm{S/N})_{fm}$ is

$$(\mathrm{S/N})|_{fm} = 10 + G \text{ (dB)} \tag{8}$$

where G is obtained from the figure along with r, the ratio of premodulator bandwidth to output information bandwidth. The carrier-to-noise ratio in a noise bandwidth equal to the output information bandwidth is

$$(\mathrm{C/N})|_{fm} = 13 + 10 \log r \text{ (dB)} \tag{9}$$

NOISE-THRESHOLD PROPERTIES OF ANGLE MODULATION

The signal-to-noise improvement represented by Eq. (7) is achievable only when the input carrier-to-noise ratio is above certain minimum levels. These levels depend on the type of modulating waveforms, the type of noise interference prevalent, and the type of demodulator. As the foregoing discussion indicates, whenever a demodulator without phase or frequency feedback is used, an input carrier-to-noise ratio (measured in the premodulator bandwidth) of roughly 12 dB is required. Unless this condition is met, a small decrease in carrier-to-noise ratio will result in a sharp decrease in output signal-to-noise ratio, accompanied by undesirable noise effects, such as load clicking sounds in the case of voice modulation.

ANGLE MODULATORS

Angle modulators for communications and telemetry purposes generally fall into the category of what may be termed “hard” oscillators having relatively high-Q frequency-determining networks; or they fall into the category of “soft” oscillators having supply and bias sources as the frequency-determining networks. Examples of each are shown in Fig. 12.2.4.

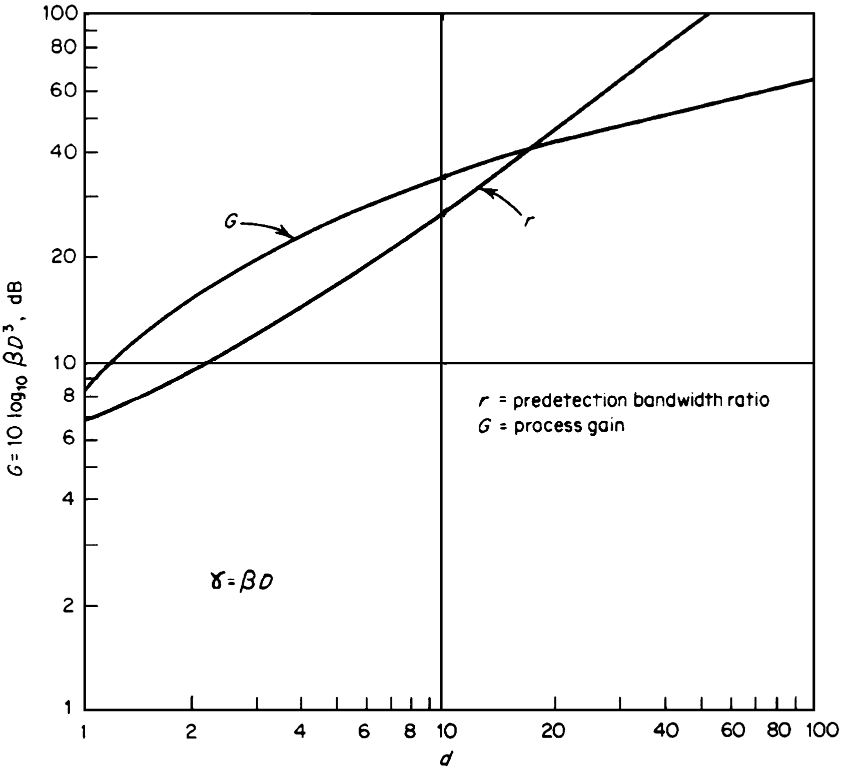

FIGURE 12.2.3 Process gain and rf bandwidth vs. deviation ratio.

Control of the hard oscillator is executed by symmetrical incremental variation of the reactive components. For the case of a Hartley oscillator, Z_1 and Z_2 are inductors and Z_3 is a capacitor, allowing the use of varicaps paralleling, Z_3 as the voltage-controllable reactance. In such a case

$$f_o = (2\pi)^{-1}[C_3(L_1 + L_2)]^{-1/2}$$

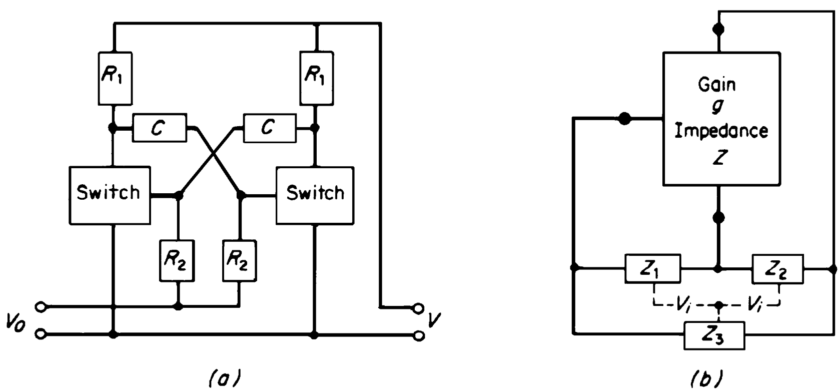

FIGURE 12.2.4 Voltage-controlled oscillators: (*a*) soft oscillator; (*b*) hard oscillator.

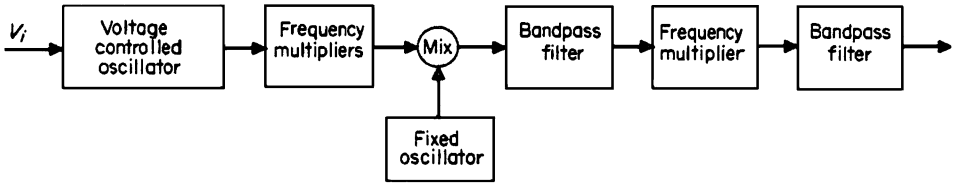

FIGURE 12.2.5 Frequency-modulator configuration.

where C_3 = total capacitance of varicaps and fixed capacitor of Z_3. Note that the bandwidth of this modulator is determined by the frequency-determining impedances of the network, i.e., the overall Q and center frequency. In view of the need for certain minimum bandwidth requirements from Fig. 12.2.2 and the need for good oscillator stability, the total frequency deviation required is sometimes obtained by following the oscillator with a series of frequency multipliers and frequency translators, as shown in Fig. 12.2.5. This configuration permits the attainment of good oscillator stability, constant proportionality between output frequency change and input voltage change, and the necessary modulator bandwidth to achieve wide-band FM.

Control for the soft oscillator is introduced as a change in the switching level of the active-device switches. The frequency of a transistor version of the oscillator is, roughly,

$$f = [2R_1C \ln (1 + V/V_i)]^{-1} \tag{10}$$

Since this type of oscillator is a relaxation oscillator, the rate at which the frequency of oscillation can be changed is limited only by the rate at which the switching points can be altered by voltage control. Modulators of this kind can be designed with bandwidths greater than the frequency of oscillation. The disadvantage of such networks is the relatively poor frequency stability compared with the high-Q hard oscillators.

DISCRIMINATORS

Basic angle demodulators can be designed using balanced tuned networks with suitably connected nonlinearities (such as diode switches); or angle demodulators can be designed using voltage-controlled oscillators (VCO), multipliers, and appropriate feedback networks.

The former are simpler to implement than the latter but require much higher input signal-to-noise ratios for operation above the noise threshold at which the angle modulation produces signal process gain. Two popular versions of the discriminator type of frequency demodulator are shown in Fig. 12.2.6. The diode discriminator is designed so that one diode conducts more with increases, the other with decreases, in frequency. For greatest linearity, the mutual coupling between the tuned circuits is generally greater than unity. The other basic type of conventional demodulator simply implements a version of the definition [Eq. (1)] of FM for the case of a constant-amplitude sinusoidal waveform. The mixing operation can be replaced by a simple phase shifter that provides a 90° phase relationship between x and Hx. Frequently, the reference oscillator will be phase-locked to the average value of the $J_0(d)$ component indicated by Eqs. (3) and (5).

FM FEEDBACK (FMFB) DEMODULATORS

These angle demodulators have the advantage of lower distortion, lower noise threshold, and little or no drift in center frequency of operation and cab be designed to be less sensitive to interference. No limiter is required for good performance, and there is no requirement to maintain a minimum predemodulator input carrier-to-noise ratio to avoid noise threshold.

A block diagram of a basic synchronous-filtering demodulator is shown in Fig. 12.2.7. The synchronous filter is indicated by the dashed lines that enclose a phase-locked loop designed to follow the instantaneous

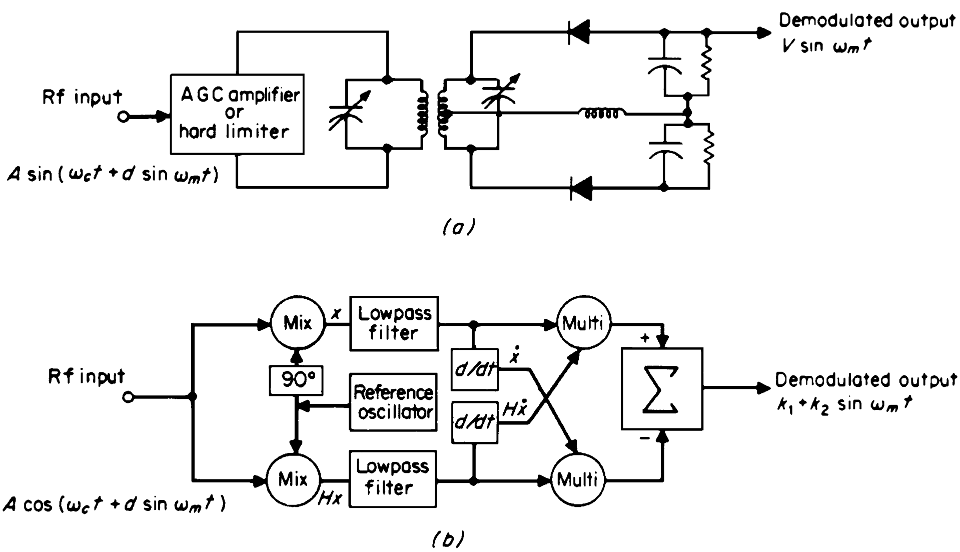

FIGURE 12.2.6 Conventional demodulators: (*a*) diode discriminator; (*b*) phase-shift discriminator.

excursions in the phase of the angle-modulated signal ϕ_2. The mixer, i.f. amplifier, discriminator, filter 2, and the VCO-2 form a frequency feedback loop for the demodulator.

Aside from the synchronous filter, the basic configuration can be considered that of a conventional FMFB demodulator which compresses the wide-band FM input signal so that it can be passed through a relatively narrow bandwidth fixed-tuned filter and to a discriminator for detection. It should be noted that the configuration shown here can also be considered simply as a phase-locked loop (PLL) with frequency feedback around it.

The significant advantages of each of these techniques (FMFB and PLL) can be combined. It is important to consider the design from both the synchronous and frequency-feedback viewpoints. Note in this regard Figs. 12.2.8 and 12.2.9 in which $F_1(S)$ and $F_2(S)$ have been assigned. Observe that Fig. 12.2.10 is an

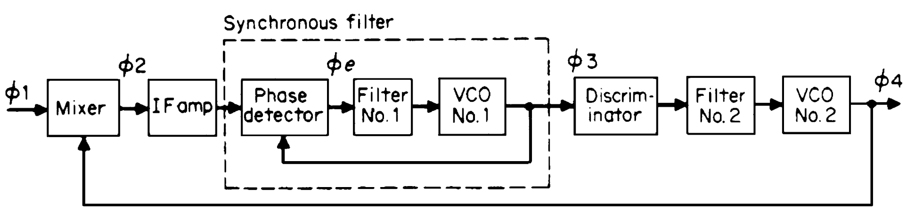

FIGURE 12.2.7 Block diagram of basic feedback demodulator.

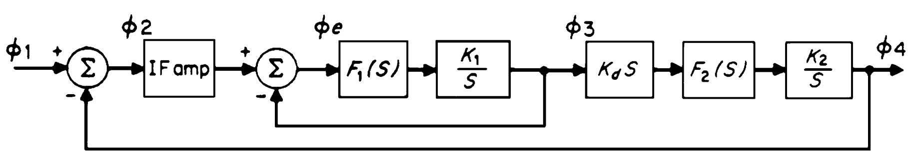

FIGURE 12.2.8 Equivalent network of basic demodulator shown in Fig. 12.2.7.

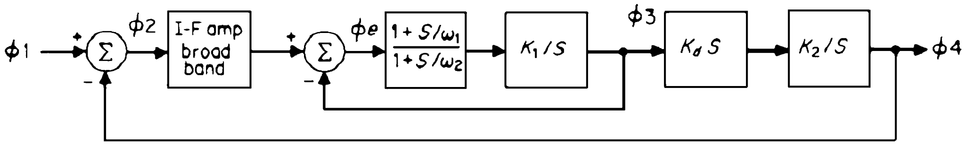

FIGURE 12.2.9 Direct linear equivalent form of basic demodulator.

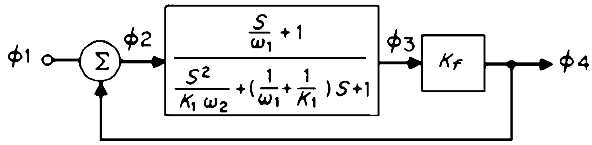

FIGURE 12.2.10 FMFB demodulator.

FMFB equivalent linear form obtained by substituting the closed-loop transfer function for the synchronous filter; however, by retaining the inner loop and combining phase comparators, we obtain the synchronous phase-locked-loop form, as shown in Fig. 12.2.11. To the extent that linear analysis and quasi-linear substitutions can be made in these block diagrams, the remarks that follow are thus relevant to both PLL and FMFB forms implemented with synchronous filters or with broadband amplifiers in cascade with single-tuned filters.

PHASE-LOCKED-LOOP DEMODULATORS

The introduction of the phase-locked loop between the i.f. amplifier and discriminator may be viewed simply as a means by which the i.f. signal can be tracked, limited, and filtered regardless of Doppler shift or oscillator drift. The synchronous filter is employed in conjunction with the relatively wide-bandwidth i.f. amplifier shown on the block diagram to perform this critical filtering function, as well as to take advantage of the phase coherence between the FM signal sidebands.

The operation of the demodulator can be understood from an examination of the equivalent network shown in Fig. 12.2.8, in which the transfer functions relate to phase as the input and output variables. If the bandwidth of the i.f. amplifier is broad by comparison with both the synchronous-filter bandwidth and significant modulation sidebands, it can be ignored in the equivalent linear representation of the demodulator. Since the synchronous-filter function is

$$\frac{\phi_3(s)}{\phi_2(s)} = \frac{(K_1/s)F_1(s)}{1+(K_1/s)F_1(s)} \tag{11}$$

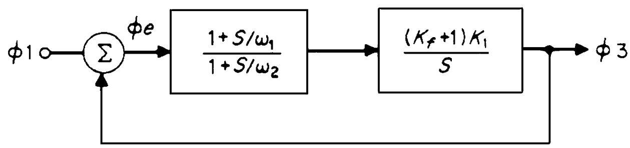

FIGURE 12.2.11 PLL demodulator.

the signal ϕ_2 is related to the input by

$$\frac{\phi_2(s)}{\phi_1(s)} = \frac{1}{1 + K_1 K_2 K_d F_1(s) F_2(s) / \{s[1 + (K_1/s) F_1(s)]} \tag{12}$$

For frequency components of $\phi_1(s)$ lying well within the bandwidths of $F_2(s)$ and the synchronous filter, this transfer function reduces to the familiar form of a type-zero feedback network: i.e.,

$$\frac{\phi_2(s)}{\phi_1(s)} = \frac{1}{1 + K_2 K_d} = \frac{s\phi_2(s)}{s\phi_1(s)} \tag{13}$$

Since this is also the transfer function with respect to frequency variations, the compression of frequency excursions is evident. If

$$\phi_1(t) = (\Delta\omega/\omega_m)\sin\omega_m t = D\sin\omega_m t \tag{14}$$

where $\phi(t)$ represents the instantaneous variation of the phase of the input signal relative to some reference carrier phase, and if the synchronous filter follows this instantaneous variation, the effective phase excursion to the discriminator is reduced to

$$\phi_3(t) = \frac{D}{1 + K_f}\sin\omega_m t \qquad K_t = K_2 K_d \tag{15}$$

if ω_m is well within the passband of $F_2(s)$ and the phase-locked loop. This reduction in deviation ratio by the use of frequency feedback gain K_f suggests that an optimum gain and synchronous-filter bandwidth combination should be sought for given modulation and noise characteristics, just as the proper i.f. amplifier and K_f must be chosen in a conventional FMFB demodulator.

Figure 12.2.10 shows the closed-loop transfer function of the synchronous filter along with the frequency-feedback loop. It has been assumed that the i.f. amplifier bandwidth is very broad compared with the bandwidth occupied by the significant portions of the signal spectrum as it appears at i.f. frequencies. This amounts to assuming that the dispersive effect produced by the i.f. amplifier is negligible. If the phase-locked-loop gain can be considered large compared with the filter zero ($K_1 >> \omega_1$ is usually satisfied in practice), the following relationships between the synchronous-filter and the complete feedback-demodulator parameters can be written

$$\zeta_f^2 = (K_f + 1)\zeta_\phi^2 \qquad \omega_{nf}^2 = (K_f + 1)\omega_{n\phi}^2 \qquad B_{nf} = B_{n\phi}\left(1 + K_f \frac{1}{1 + 1/4\zeta_\phi^2}\right) \tag{16}$$

where ζ_f = damping of demodulator
ω_{nf} = natural frequency of demodulator
ζ_ϕ = damping of synchronous filter
$\omega_{n\phi}$ = natural frequency of synchronous filter
B_{nf} = noise bandwidth of demodulator
$B_{n\phi}$ = noise bandwidth of synchronous filter

and in terms of the actual network parameters

$$\zeta_\phi^2 = \frac{K_1\omega_2}{4\omega_1^2} \qquad \omega_{n\phi}^2 = K_1\omega_2 \qquad B_{n\phi} = \frac{1}{2}\left(\frac{K_1\omega_2}{\omega_1} + \omega_1\right) \quad \text{(Hz)} \tag{17}$$

Note that all the demodulator response variables can be controlled by simple *RC* adjustments.

FM FEEDBACK-DEMODULATOR DESIGN FORMULAS

Acquisition. The rate at which the VCO in a phase-locked demodulator can be swept through the frequency and phase at which pull-in and phase lock will occur is indicated by the representative curves of Fig. 12.2.12.

Pull-In

The frequency difference $\Delta\omega$ (between an unmodulated carrier and the VCO) within which a phase-locked loop will pull into phase synchronism is, roughly,

$$|\Delta\omega|_p = (2pK\omega_n)^{1/2} \quad \text{for } K/\omega_n \gg 1$$

where K = total open-loop gain (rad/s)
ω_n = loop natural frequency (rad/s)
ρ = dimensionless constant ≈ 1

The time required for pull-in is given by

$$T = 4(\Delta f)^2/B_n^3 \quad \text{(seconds)}$$

for loop damping of 0.5 and $|\Delta f| < 0.8|\Delta f|_p$, where Δf = difference frequency (Hz) and B_n = closed-loop noise bandwidth (Hz).

Stability

The condition of sustained beat-note stability without the loop capacity to be swept into lock, which is exhibited by high-gain narrow-bandwidth phase-locked demodulators, is a condition of loop oscillation arising from the presence of the phase detector nonlinearity and extraneous memory, such as that in the tuned circuits in

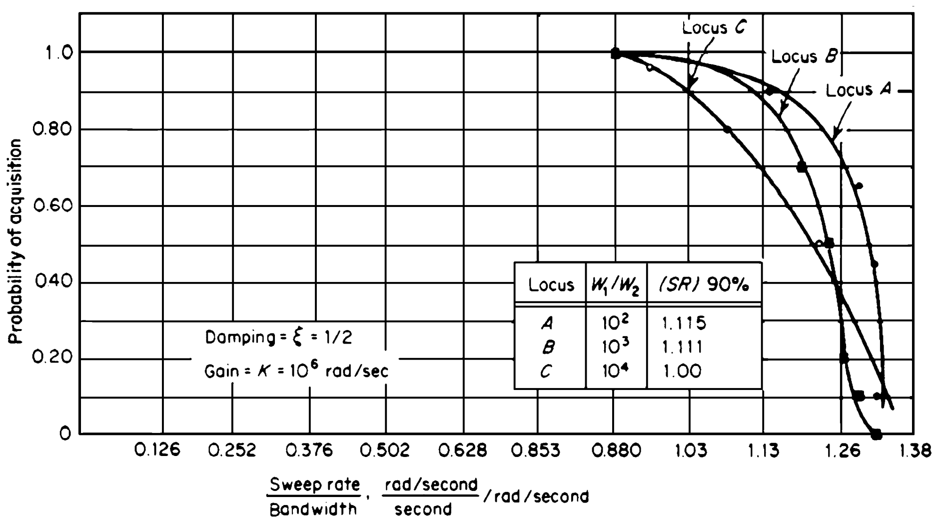

FIGURE 12.2.12 Probability of acquisition vs. sweep-rate per unit bandwidth.

phase detectors and in voltage-controlled oscillators. This condition can be predicted from the approximate condition

$$\frac{K^2 G|\omega_0|}{2|\omega_0|^2}[\cos\phi(\omega_0)]+1=0 \tag{18}$$

where K = total open-loop gain
G = complete transfer function between phase detector and VCO
ϕ = angle of G

This condition can be used to predict to nonlinear network oscillations preventing lockup.

Parameter Variations

The rate at which second-order demodulators damped for minimum-noise bandwidth can have the loop natural frequency changed at constant damping is given by

$$\frac{d\omega_n/dt}{\omega_n^2} \le 0.1 \tag{19}$$

for peak phase errors less than 5°, where ω_n is the natural frequency of the loop.

Minimum-Threshold Parameters

The parametric relationships for minimum-noise threshold for the demodulator in Fig. 12.2.9 are shown in Fig. 12.2.13.

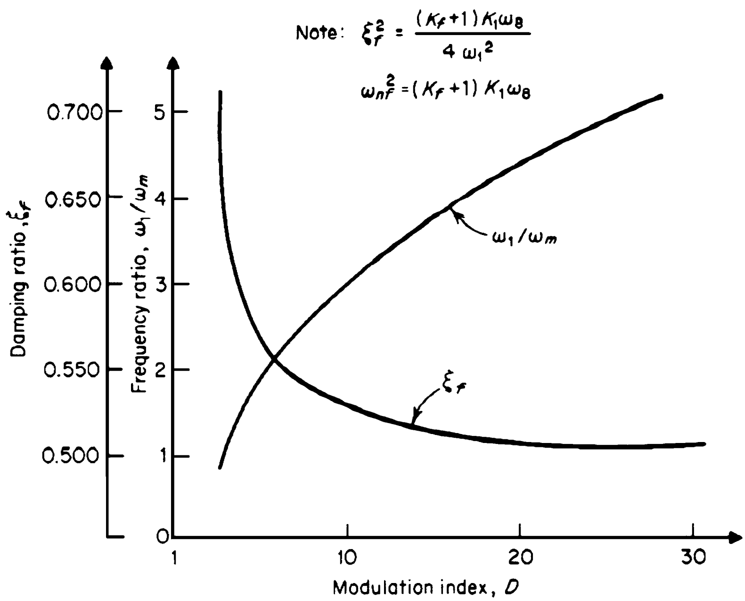

FIGURE 12.2.13 Parametric relationships for minimum threshold (voice).