Technical Report No. 32-637

Optimum and Suboptimum Frequency Demodulation
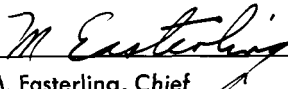
William C. Lindsey

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

June 15, 1964

Technical Report No. 32-637

Optimum and Suboptimum Frequency Demodulation

William C. Lindsey

M. Easterling, Chief
Communications Systems Research Section

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

June 15, 1964

CONTENTS

FIGURES

ABSTRACT

$\backslash 3300$
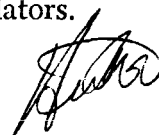
The design and mechanization of frequency demodulators for use in telemetry communications are discussed. The frequency demodulator consists of an ordinary phase-locked loop (PLL) followed by a linear-output filter. The linear and quasi-linear PLL demodulator models are optimized by choosing the loop filter transfer function that minimizes the mean-squared phase error, while the output filter transfer function is optimized by selecting the linear filter that minimizes the mean-squared frequency error. In general these filters are adaptive in that their pole–zero configuration depends on the available signal power, the noise spectral density, and the modulation index. For this reason, i.e., because the filters are difficult to mechanize, two alternate and more easily implemented suboptimum frequency demodulators are analyzed. These are obtained by selecting the loop and output filters that are optimum for a certain set of design levels and by preceding the loop either by a bandpass limiter or an automatic gain control (AGC) amplifier. For parameter levels that remain close to the design levels, either of these realizations performs almost optimally.

Synthesis procedures for the optimum-loop and output filters are given, and the transient behavior of the optimized system is determined. For various receiver threshold characteristics, the performance of the linear and quasi-linear demodulator models is graphically illustrated and compared for all three receiver realizations. A major conclusion from these results is that the PLL that is preceded by an AGC amplifier outperforms the same loop when it is preceded by a bandpass limiter. Heretofore, analysis of the behavior of the PLL as a frequency discriminator has avoided considering the effects on receiver performance when the quasi-linear or linear demodulator models are preceded by an AGC amplifier or a bandpass limiter.

Finally, an amplitude-modulated single-sideband or double-sideband system (which is "amplitude matched" to the same modulating spectra) is compared with that of the optimum and suboptimum linear demodulators. The results are useful when one is faced with the problem of designing optimum and near-optimum frequency demodulators.

I. INTRODUCTION

The purpose of this Report is to review, investigate, and determine design trends and to present an optimum theory of phase or frequency demodulation suitable for use in designing discriminators used in telemetry communication systems, e.g., ballistic missiles, space communications, etc. The theory and results presented are applicable to such practical problems as designing optimum and suboptimum frequency discriminators (1) for detecting video information relayed to the Earth concerning planetary surface characteristics, (2) for use in detecting audio signals reflected from one of the neighboring planets, or (3) for detecting scientific data transmitted from a particular spacecraft. The graphical results presented are valid for the demodulation of a radio frequency (RF) carrier that has been frequency-modulated by an RC–filtered white-noise spectrum and corrupted in the channel by additive white Gaussian noise. On the other hand, the theory may be applied to the demodulation or filtering of rather arbitrary types of stationary signal and noise processes.

Section II-A defines the problem and gives a brief history of the significant papers that have been written on frequency demodulation by means of phase-locked and frequency-feedback detectors. Section II-B is devoted to the problem of specifying equivalent receiver models for the PLL. In particular, three models are presented: the "Exact-Equivalent Model", the Quasi-linear Model, and the Linear Model. The phase-error spectral density is derived for the latter two models and related

to the mean-squared Wiener error. In Section II-C we discuss and derive two types of optimum Wiener filters. These are commonly referred to as the realizable (zero-lag) and nonrealizable (infinite-lag) filters. Section II-D has been used for specifying the optimum loop and output filter transfer functions and the ways of physically realizing and synthesizing these functions with electrical networks. Section II-E is devoted to specifying the performance of the demodulator when the loop and output filter pole–zero configurations are fixed in accordance with *a priori* given design levels and the PLL is preceded by an AGC amplifier. The receiver "threshold characteristic" is also defined in this Section. Section II-F presents the demodulator performance when the received signal power differs from the initial design value, while Section II-G is used for developing the performance of the demodulator when the PLL is preceded by a band-pass limiter. This Section succinctly presents the operating characteristics of the bandpass limiter and its role when used ahead of the PLL. Section II-H is concerned with the transient response of the optimum demodulator and compares this response with that obtained if the demodulator is used for tracking purposes. Finally, Section II-I is utilized for graphically illustrating and comparing the performance of the optimum demodulator with two more easily mechanized suboptimum demodulators. Graphical results are given (for various threshold characteristics) for both the linear and quasi-linear demodulator models. These results are compared with analogous results for balanced and single-sideband amplitude demodulation.

II. GENERAL THEORY AND RESULTS

A. Defining the Problem and Review of the Literature

The problem to be considered is illustrated in Fig. 1. At the transmitter a random signal $m(t)$ is used to control the instantaneous frequency, the phase, or some other parameter of a voltage-controlled oscillator (VCO) with a gain of K_t rad/sec/v. The resulting waveform $\xi(t)$ is transmitted into the channel where it is corrupted by additive white Gaussian noise $n(t)$ with single-sided
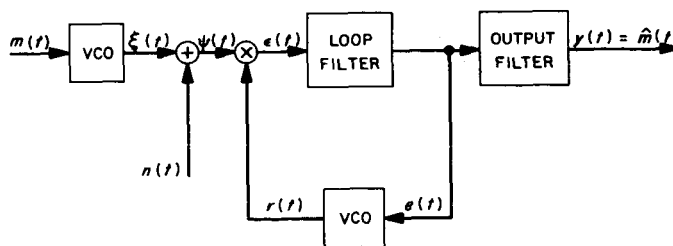

Fig. 1. Frequency-modulation communication system

spectral density of N_0 w/cps before being available as a waveform $\psi(t)$ to the receiver. The random processes $m(t)$ and $n(t)$ are presumed to be wide-sense stationary and statistically independent. We intend to specify the optimum demodulator and two alternate, more easily implemented suboptimum demodulators (under several restrictive assumptions) for a given transmitter power P, noise spectral density N_0, and VCO gain K_t, and to investigate the "output" signal-to-noise ratio relationship when the input power and the noise spectral density differ from those expected.

We realize, of course, that the communication system of Fig. 1 represents only a small fraction of a complete telemetering link. The major difference, however, is that at the transmitter, several data signals (not necessarily different) say $m_1(t), \ldots m_N(t)$, are used to modulate N subcarrier oscillators. The modulated outputs of all subcarrier oscillators are amplified and then multiplexed, thus providing a source of modulation energy that is used to modulate the master transmitter oscillator. For extraction of the transmitted information, we require (at the receiver) N detectors that are to be "matched" to the N data signals. This double modulation technique, which is well known to telemetry system design engineers, is analogous to that employed in diversity systems used in telemetry (and other types of communications) for offsetting failures in receiving and transmitting equipment, for example, (1) loss of the RF signal owing to cross-polarization, (2) multipath propagation and fading, (3) rocket flame attenuation, and (4) failures that are not considered to be important at the design time. The important thing here is that even though Fig. 1 does not depict an over-all multisignal telemetry link, the communication link shown does represent a system that ideally separates (zero crosstalk) the N subcarrier signals from the received signal; hence, it justifies the study to be carried out here.

The history of frequency demodulation (or phase modulation for that matter) employing the Foster-Seeley type of discriminator (or phase detector) rests in a hierarchical arrangement of technical papers that are too numerous to mention here. Middleton (Ref. 1), however, probably contains the most elaborate set of results and bibliography that is presently available. One major difficulty that is immediately encountered in employing ordinary discriminators of the Foster-Seeley type for use in long-range missiles or space flight telemetry communication systems is the discriminator threshold, i.e., that point in the input–output signal-to-noise ratio characteristic where the noise captures the detector (see Ref. 1).

The standard type of discriminator, i.e., a bandpass filter–limiter–discriminator followed by a low-pass filter, demonstrates an improvement over the amplitude-modulation (AM) system until the input signal-to-noise ratio reaches a low of about $+10$ db. This high threshold has caused frequency modulation (FM) systems to be criticized for use in telemetry because of the large input powers required to maintain this signal-to-noise ratio. In attempting to reduce this threshold, present emphasis seems to be placed on two types of discriminators: the PLL detector and the frequency-feedback detector. Analytical work relating standard FM discriminator performance to FM discriminators employing negative feedback may be found in Refs. 2–6. Recently, Cahn (Ref. 7) has shown (using his linearized frequency-feedback receiver based on a twin-threshold concept) that a design approach influenced by the open-loop threshold for FM feedback yields a receiver sensitivity inferior to the PLL at large modulation indices. On the other hand, Develet (Refs. 8 and 9) presents the derivation of a quasi-linear model for the frequency-feedback receiver and shows that (although the quasi-linear receiver models differ in detail) the threshold for a "maximum sensitivity" FM feedback or PLL receiver is identical.

Turning now to a brief history of the PLL, we find that the main idea seems to be due to Chaffee (Ref. 3). Among the first applications of the device appearing in the literature was its use as a synchronizing circuit (Ref. 10) while Richman (Ref. 11) applied the concept to color television systems. Jaffe and Rechtin (Ref. 12) and Viterbi (Ref. 13) presented the capabilities and behavior of the device when tracking doppler-shifted signals of known waveforms and where the receiver bandwidths were considerably narrower than the doppler shift (also see Ref. 14). Gilchriest (Ref. 15), Viterbi (Refs. 16, 17) and Choate (Ref. 18) presented results when the PLL was used as a telemetry discriminator and a tracking filter. The following work is related mainly to that of Viterbi (Refs. 16, 17) and Develet (Ref. 8).

B. Equivalent Receiver Models

It is well known that the PLL receiver of Fig. 1, i.e., the closed-loop portion of the receiver, is a nonlinear device. Furthermore, the "exact" performance of the loop in the absence of modulation and additive Gaussian noise is available only in special cases (Ref. 19). Consequently, we seek a workable model that will closely approximate the performance of the physical device when signal modulation and noise are present.

1. Exact-Equivalent Receiver

Denote the received waveform by $\psi(t)$, i.e.,

$$\psi(t) = \sqrt{2P} \sin [\omega_0 t + \theta_1(t)] + n(t) \qquad (1)$$

where

$$\theta_1(t) = K_t \int^t m(\tau)d\tau$$

and where

$$n(t) = n_1(t) \sin \omega_0 t + n_2(t) \cos \omega_0 t$$

represents the noise process of infinite duration (Ref. 20). Both n_1 and n_2 are assumed to be statistically independent, stationary, white Gaussian noise processes of one-sided spectral density N_0 w/cps. The reference signal $r(t)$ at the output of the receiver's VCO is presumed to be a sinusoid whose instantaneous frequency is related to the input control voltage $e(t)$ through the relationship (see Fig. 1)

$$r(t) = \sqrt{2} \cos [\omega_0 t + \theta_2(t)] \qquad (2)$$

where

$$\theta_2(t) = K_r \int^t e(\tau)d\tau$$

Thus, the product of the input $\psi(t)$ and $r(t)$ may be shown to be related, in operational form, to the phase error $\phi(t)$ by

$$\phi(t) = \theta_1(t) - \theta_2(t)$$

$$= \theta_1(t) - \frac{K_r}{s} F(s) \left\{ \sqrt{P} \sin \phi(t) + n'(t) \right\} \qquad (3)$$

where all double-frequency terms have been neglected because neither the loop filter $F(s)$ nor the VCO will respond significantly to them. Further, it may be shown (Ref. 19 and Ref. 8) that $n'(t)$, consisting of two noise terms, is white Gaussian phase noise with the same spectral density as that of the original additive process $n(t)$. Thus, to study the receiver structure from the input $\psi(t)$ to the output $e(t)$, Eq. (3) evidences the pertinent quantities, and the PLL may be conveniently represented by the block diagram of Fig. 2. For convenience, the output filter $F_0(s)$ has also been included. Note that the receiver structure under study is now placed into a form that is familiar to control engineers, and we shall refer to this model as the "exact-equivalent receiver."
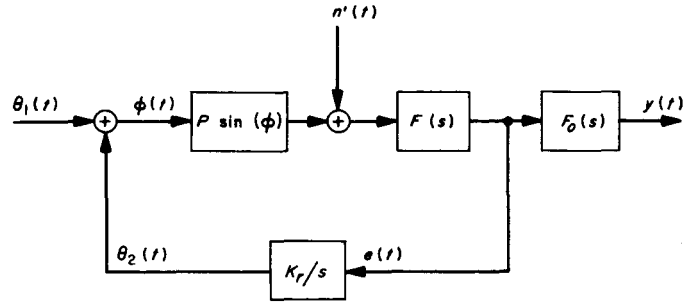
Fig. 2. Exact-equivalent receiver

Even though the loop of Fig. 1 has been replaced by that of Fig. 2, it is still impossible to specify the "optimum" demodulator. The heart of the problem lies in the sinusoidal nonlinearity. In the next Section we shall introduce a quasi-linear model that will approximate the operation of the loop when the phase error $\phi(t)$ is small.

2. Quasi-linear Receiver

The quasi-linear model, introduced by Develet (Ref. 8), is based on Booton's idea (Ref. 20) that any nonlinear device subjected to a Gaussian process may be replaced by an equivalent gain K. The concept is simple and may be briefly described as follows: denote the input and response of an amplitude-sensitive element by x_i and x_r respectively. Let the output x_r be written in the form

$$x_r = f(x_i)$$

where f is the *describing function* of the element. Then for a particular input process, we select that value of K for which the mean-squared error

$$\overline{[x_r - Kx_i]^2} = \overline{[f(x_i) - Kx_i]^2}$$

is a minimum. (The bar signifies an average taken in the statistical sense.) For the problem at hand $f = \sin [\quad]$. Assuming that the input $\phi(t)$ to the nonlinear element is a zero-mean Gaussian process, Develet (Ref. 8) shows that

$$K_\varphi = \exp \left[- \left(\sigma_\varphi^2 / 2 \right) \right] \qquad (4)$$

where σ_φ^2 is the mean-squared value of the phase error $\phi(t)$.

In passing, two comments are important. In order to use this gain, one further assumption is necessary, i.e., the process $\phi(t)$ must be first-order stationary. Otherwise the gain K_φ of the element, and hence the over-all

receiver, is time-varying. Second is the fact that the nonlinear element occurs as a part of the feedback system and even though the input to the PLL may be Gaussian, the operation of a nonlinear element on a Gaussian process is in general non-Gaussian. These remarks should be kept in mind when employing the quasi-linear model; however, if we are going to take the nonlinearity into consideration in the design at all, this approximation seems to be as simple and as mathematically tractable as any. As a matter of fact, Viterbi (Ref. 19) shows that, for the first-order loop and with no modulation present, Develet's approximation is good so long as $\sigma_\varphi^2 < 1/2$.

The equivalent quasi-linear model is depicted in Fig. 3. For reasons that will become obvious later, we define the following transfer functions, i.e.,

$$H(s) = \frac{y(s)}{\theta_1(s)} \; ; H_\varphi(s) = \frac{K_r e(s)}{s\theta_1(s)} = \frac{\theta_2(s)}{\theta_1(s)} \qquad (5)$$

and referring to Fig. 3, it is obvious that

$$F_o(s) = \frac{K_r H(s)}{s \, H_\varphi(s)} \qquad (6)$$

Further, the closed-loop transfer function of the quasi-linear PLL may be written by inspection from Fig. 3 as

$$H_\varphi(s) = \frac{K_\varphi K_r \sqrt{P} F(s)}{[s + K_\varphi K_r \sqrt{P} F(s)]} \qquad (7)$$

where

$$K_\varphi = \exp\left(-\sigma_\varphi^2/2\right)$$
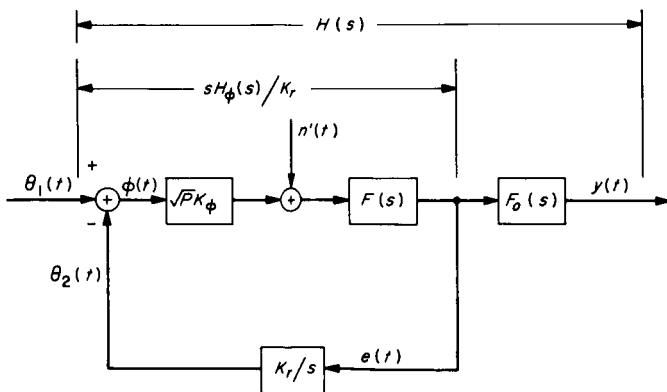

Fig. 3. Quasi-linear phase-locked receiver

and the loop filter $F(s)$, obtained from Eq. (7), is

$$F(s) = \frac{[K_r K_\varphi \sqrt{P}]^{-1} s H_\varphi(s)}{1 - H_\varphi(s)} \qquad (8)$$

From Eq. (5), we see that the output $\theta_2(t)$ of the PLL is proportional to the amplitude of the input $\theta_1(t)$. Thus the closed-loop response of the PLL is a function of the amplitude of the transmitted signal (see Eqs. 7 and 8). In a physical situation where the input power may deviate from the expected value, a bandpass limiter (when used ahead of the PLL) or an AGC amplifier helps remedy the situation. This phenomenon is due to an intrinsic characteristic of the limiter; i.e., it may be shown (Ref. 21) that the bandpass limiter is a constant-power device. In a later Section, we will determine the effects on the optimum design when no limiter or AGC amplifier is present and the power changes from the expected design value.

Before leaving this discussion pertaining to the quasi-linear model, we point out that the noise component $n'(t)$ of Fig. 3 may be moved to the input and designated as $n_1'(t)$, which is a wide-sense stationary, white Gaussian noise process with a single-sided spectral density of

$$N_0' = \frac{N_0}{P} \exp\left(\sigma_\varphi^2\right) = \frac{N_0 K_\varphi^{-2}}{P} \qquad (9)$$

Thus the noise that the equivalent receiver sees is now a function of the mean-squared phase error and the transmitter signal power.

3. Linear Receiver Model

To obtain the frequently used linear model, we utilize the assumption of Jaffe and Rechtin (Ref. 12) that $\phi(t)$ is small enough at all times that $\sin \phi \simeq \phi$ and $\cos \phi \simeq 1$, so that the quasi-linear model becomes the linear model and $K_\varphi = 1$. In the study to be carried out here, we will make use of the two models: the quasi-linear model and the linear model. At present these two models appear to be the only known forms that yield analytical results.

4. Error Spectral Density for the Linear and Quasi-linear Models

Through the use of the quasi-linear or linear approximation, Eq. (3) becomes, in operator form,

$$\phi(t) = \theta_1(t) - \frac{K_r}{s} F(s) \left[K_\varphi \sqrt{P} \phi(t) + n'(t)\right]$$

where K_φ is given by Eq. (4) for the quasi-linear model, and $K_\varphi = 1$ for the linear model. Taking the Laplace transform of each term in this equation gives

$$\phi(s) = \theta_1(s) - \frac{K_r}{s} F(s) \left[K_\varphi \sqrt{P} \, \phi(s) + n'(s) \right]$$

or

$$\phi(s) \left[1 + \frac{K_r K_\varphi \sqrt{P}}{s} F(s) \right] = \theta_1(s) - \frac{K_r F(s)}{s} n'(s)$$

whereas the absolute value of the phase error is $\phi(s)\phi(-s)$, and the average value $\overline{\phi(s)\phi(-s)}$ becomes

$$\overline{\phi(s)\phi(-s)} = \left| \frac{1}{1 + \dfrac{K_r K_\varphi \sqrt{P}}{s} F(s)} \right|^2 \overline{\theta_1(s)\theta_1(-s)}$$

$$\tag{10}$$

$$+ \left| \frac{K_r F(s)/s}{1 + \dfrac{K_r K_\varphi \sqrt{P}}{s} F(s)} \right|^2 \overline{n'(s)n'(-s)}$$

since $n'(t)$ and $\theta_1(t)$ are uncorrelated. Now $\overline{\phi(s)\phi(-s)}$, $\overline{\theta_1(s)\theta_1(-s)}$, and $\overline{n'(s)n'(-s)}$ are respectively the phase error spectral density $S_\varphi(s)$, the signal phase spectral density $S_{\theta_1}(s)$, and the noise spectral density $S_{n'}(s)$. Thus Eq. (10) becomes

$$S_\varphi(s) = \left| \frac{1}{1 + \dfrac{K_r K_\varphi \sqrt{P}}{s} F(s)} \right|^2 S_{\theta_1}(s)$$

$$+ \left| \frac{K_r F(s)/s}{1 + \dfrac{K_r K_\varphi \sqrt{P}}{s} F(s)} \right|^2 S_{n'}(s)$$

Notice that the coefficient multiplying $S_{\theta_1}(s)$ is $1 - H_\varphi(s)$, from Eq. (7), and that the coefficient multiplying $S_{n'}(s)$ is

$$\left[K_\varphi^2 P \right]^{-1} |H_\varphi(s)|^2$$

Upon referring the noise spectral density $S_{n'}(s)$ to the input and renaming, we have

$$S_\varphi(s) = |1 - H_\varphi(s)|^2 S_{\theta_1}(s) + |H_\varphi(s)|^2 S_{n'_1}(s) \tag{11}$$

where

$$S_{n'_1}(s) = \left[K_\varphi^2 P \right]^{-1} S_{n'}(s)$$

$$= \frac{N_0 \exp(\sigma_\varphi^2)}{2P} \tag{12}$$

The average phase error σ_φ^2 may be obtained from Eq. (11) by averaging over all frequencies, namely,

$$2\pi j \sigma_\varphi^2 = \int_{-j\infty}^{j\infty} S_\varphi(s)\,ds = \int_{-j\infty}^{j\infty} \left[S_{\theta_1}(s) \left| 1 - H_\varphi(s) \right|^2 ds \right.$$

$$\tag{13}$$

$$\left. + S_{n'_1}(s) \left| H_\varphi(s) \right|^2 ds \right]$$

The first term in Eq. (13) is, of course, that portion of the phase error due to $H_\varphi(s) \neq 1$, while the second term is that portion of the phase error introduced by the additive noise process. Unfortunately, when the quasi-linear model is used, Eq. (13) is a transcendental equation in σ_φ. For this reason the Wiener mean-squared-error filtering theory cannot be immediately invoked, although, as we shall see in the next Section, the filter (realizable or non-realizable) $H_\varphi(s)$ which minimizes σ_φ is nothing more than the Wiener filter.

C. Specification of the Optimum Filter

In the previous two Sections, we have discussed the problem of demodulation and presented receiver models to which the linear feedback type of analysis may be applied. The problem considered in this Section is that of following a signal (or some linear operation on a signal) corrupted by noise as "closely" as possible with a linear filter. The filter input, which is a function of time, is in general made up of both signal $s(t)$ and noise $n(t)$. It is desired to choose a linear filter such that the output $y(t)$ "matches" the input signal. Owing to the noise, it is clearly impossible for $y(t)$ to be identical with $s(t)$ or some linear operation on $s(t)$. The object then is to choose the linear filter that matches the two as "closely" as possible.

1. Derivation of the Filter Equation

Mathematically, the filtering problem may be stated as follows: Let the observed input function

$$f(t) = s(t) + n(t) \tag{14}$$

composed of a useful signal $s(t)$ and noise $n(t)$, be applied to the input of a filter $\mathcal{K}$. This filter $\mathcal{K}$ performs certain mathematical operations on the function $f(t)$, as a result of which we obtain $y(t)$ (the filter output) at the output of $\mathcal{K}$. The output function $y(t)$ may be regarded as the result of applying an operator $\mathcal{K}$ to the input function $f(t)$.

The problem now is to select the filter $\mathcal{K}$ in such a manner that it reproduces some desired function $g(t)$ with the "least" possible error. Here $g(t)$ may be considered to be functionally related to the signal and may be written in the form

$$g(t) = \mathcal{L}[s(t)] \tag{15}$$

where $\mathcal{L}$ is a known mathematical operator.

The difference

$$\sigma(t) = y(t) - g(t) = y(t) - \mathcal{L}[s(t)] \tag{16}$$

is the instantaneous error of reproduction and, if noise is present, it is clear that the value of $\sigma(t)$ will fluctuate randomly with time. The intensity of the fluctuations will be characterized by the mean-squared error, i.e., by the mathematical expectation of the square denoted by σ^2. Hence we look for that realizable filter $\mathcal{K} = H(s)$ (employing the well-known and accepted Laplace transform theory notation) which makes σ^2 a minimum. This is what we mean by selecting the filter that will follow the signal $m(t)$ as "closely" as possible. By "realizable" we mean filters that may be constructed with circuits containing linear elements [more specifically, $h(t) = L^{-1}[H(s)] = 0$ for $t < 0$, as per Davenport and Root (Ref. 22, p. 174)]. It is possible to discuss the filtering of stored or recorded processes; however, we are not interested in this type of filtering here.

The input signal and noise to the filter $\mathcal{K}$ are assumed to be stationary random processes; i.e., the probability distributions that describe the time functions are invariant with respect to a change of the time zero. Further, the signal and noise are presumed to be uncorrelated.

For such a system the mean-square difference between the filter output $y(t)$ and some linear operation $\mathcal{L}$ on the input signal is given by

$$\sigma^2 = \frac{1}{2\pi j} \left\{ \int\!\!\int_{-j\infty}^{j\infty} [S_s(s)\,|L(s) - H(s)|^2]\,ds \right.$$

$$\left. + \int_{-j\infty}^{j\infty} S_n(s)\,|H(s)|^2\,ds \right\} \tag{17}$$

where $S_s(s)$ is the spectral density of the signal, $L(s)$ is the desired linear operation, and $S_n(s)$ is the spectral density of the noise. From Eq. (17) we see that the mean-squared error is composed of two parts. The first part is the error obtained because the transfer function differs from $L(s)$. We shall refer to this portion of the error as the signal distortion

$$S_d = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} S_s(s)\,|L(s) - H(s)|^2\,ds \tag{18}$$

and the other portion of the error is due to the additive noise.

Eq. (17) is a standard equation, the derivation of which is immediate when the signal and noise are transformed into their Laplace transforms. The mean-squared error is then found by squaring the absolute difference between the actual output signal and the desired output signal and integrating over all frequencies.

Thus we desire to minimize σ^2 of Eq. (17) by the proper choice of a realizable transfer function $H(s)$. The transfer function $H(s)$ [for a given $S_s(s)$ and $S_n(s)$] may be found by standard variational methods. In fact as Eq. (17) now stands, the optimum filter is nothing more than the standard Wiener filter (Ref. 23); however, the problem that we face is a bit different, although (as we shall see) the Wiener results will still apply. The primary difference is the following: in the quasi-linear model of the PLL we found that the input noise spectral density was a function of the phase error σ_φ [see Eq. (13)]. In fact, if we remove this dependence in the term containing the noise spectral density by replacing it by $f(\sigma^2)\,S_n'(s)$ where $f(\sigma^2)$ is some function depending on σ^2

and where $S_n(s)$ depends only on the original additive noise process, Eq. (14) becomes

$$\sigma^2 = \frac{1}{2\pi j} \left[\int_{-j\infty}^{j\infty} S_s(s) |L(s) - H(s)|^2 \, ds + f(\sigma^2) \right.$$

$$\left. \int_{-j\infty}^{j\infty} S_n'(s) |H(s)|^2 \, ds \right] \tag{19}$$

The variational technique that may be used to determine $H(s)$ is to add to the optimum realizable filter $H(s)$ the function $\varepsilon(s)$, differentiate with respect to ε, and set the derivative $d\sigma_\varepsilon^2/d\varepsilon$ and ε equal to zero. Since $\varepsilon(s)$ is arbitrary, an equation for $H(s)$ is obtained. The result is easily shown to be the filter that satisfies the following equation

$$H(s) = \frac{1}{S_f^+(s)} \left[\frac{L(s) \, S_s(s)}{S_f^-(s)} \right]_{PR} \tag{20}$$

where the input spectra are rational functions and where the plus superscript signifies the left half-plane factors. The subscript *PR* denotes that only those terms in the partial fraction expansion of the bracketed quantity which have left half-plane poles are to be taken. The spectral density $S_f(s)$ is given by

$$S_f(s) = S_s(s) + f(\sigma^2) \, S_n'(s) \tag{21}$$

Thus the general results due to Wiener are applicable to optimizing the quasi-linear receiver, with the exception that the noise spectral density, for the white-noise case, is replaced by Eq. (9).

2. The Infinite-Lag Filter and Other Optimization Criteria

As is often the case, the input to the filter $H(s)$ is stored (recorded) for a certain interval of time (theoretically, for $-\infty t < \infty$) and is then processed. The advantages of the filters given by Eq. (16) are the simplicity with which they can be implemented and the rapidity with which the output $y(t)$ is obtained. On the other hand, the so-called "infinite lag" filters are more difficult to implement; however, their ability to suppress the noise $n(t)$ is more effective since they make complete use of the input message. If, instead of imposing the realizability constraint on $H(s)$, we solve for the

optimum filter (under a mean-squared-error criterion) that makes use of the input data for all t, we find from Eq. (19) that

$$H(s) = \frac{|L(s)|^2 \, S_s(s)}{S_f(s)}$$

and the corresponding mean-squared error is given by

$$\sigma_\varphi^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} |L(s)|^2 \left[\frac{S_s(s) S_n(s)}{S_f(s)} \right] ds$$

In retrospect, there are other criteria that are available for specifying an optimum filter. For example, Gilchriest (Ref. 15) defines an optimum filter as one that minimizes the sum of a fixed transient error (properly defined) plus the mean-squared noise error. The optimum filter equation [analogous to Eq. (20)] is identical with Eq. (20), with the exception that an arbitrary constant λ, the so-called Lagrangian multiplier, premultiplies the equation, and $S_s(s)$ is replaced by $|\theta_1(s)|^2$. The Jaffe and Rechtin criterion (Ref. 12) is a special case of the Gilchriest criterion, which is suitable for selecting filters when the input's signal process is deterministic (or non-stationary), as would be the case in tracking applications, e.g., in tracking signals of the form $\theta_1(t) = kt^2$, etc.

Thus far, the theory discussed above has been restricted to selecting linear filters, and we have not considered the possibility of using nonlinear filters. This idea is, however, a bit of a paradox in that the PLL is inherently a nonlinear filter whose "exact" performance characteristics still remain unknown in the presence of modulation. For linear filtering of Gaussian processes it is well known that a large class of criteria is equivalent to minimizing the mean-squared error; e.g., consider the detection of a sure signal in noise where the criterion of maximizing the output signal-to-noise ratio is equivalent to minimizing the mean-squared error or alternatively maximizing the likelihood function (see Refs. 1 and 28).

Finally, we note that this theory requires a statistical knowledge about the data signal and noise, i.e., the spectral densities. Consequently, the detector or demodulator that we select is highly dependent upon these quantities. For the noise process, available evidence does indeed indicate that "additive white Gaussian noise" is the best way to characterize the channel disturbances in space telemetry; however, the random process governing the behavior of $m(t)$ may not be clear at all. It has been fashionable to postulate the data signal spectral density (and we shall do so here) and thus determine the detector.

D. Specification of the Optimum Filters for FM Reception

In this Section we shall specify the optimum loop filter $F(s)$ and the optimum output filter $F_o(s)$. In order to apply the filter theory of Section C, however, we must specify the message spectral density $S_m(s)$. For purposes here, we shall assume that $m(t)$ is a unit-variance stationary time series whose spectrum is[1]

$$S_m(s) = \frac{2a}{a^2 - s^2} \tag{22}$$

Physically, Eq. (22) represents a modulation waveform that has the same power spectrum as a white-noise process passed through an RC filter possessing a 3-db frequency of $a/2\pi$ cps. Lehan and Parks (Ref. 24) were apparently the first to make the PLL analysis with this type of spectrum. Viterbi (Ref. 17) also uses this spectrum in making FM, PM, and AM calculations.

1. The Optimum Filters

It is clear from Fig. 3 that there are two types of errors that are made at the receiver. One is the phase error σ_φ^2, which is the error due to the inability of the receiver VCO to follow the incoming phase $\theta_1(t)$ exactly. To insure proper operation of the PLL, i.e., to make the loop operate as linearly as possible, it is necessary to minimize σ_φ^2 by properly selecting the loop filter $F(s)$. The other error is the frequency error σ_f^2, i.e., the total modulation error. In order that the receiver reproduce the modulation waveform $m(t)$ as faithfully as possible and that the PLL follow the observed data as closely as possible, it is necessary that we simultaneously minimize σ_f^2 and also σ_φ^2.

It may seem, at first, that it is impossible to minimize the phase and frequency errors simultaneously; however, since the system is linear, the filter $H_\varphi(s)$ and hence the loop filter $F(s)$ may be determined so as to minimize σ_φ^2, and the overall receiver filter $H(s)$ may be adjusted to minimize σ_f^2. Having determined the loop filter $F(s)$ that minimizes σ_φ^2, we may adjust the output filter $F_o(s)$ so as to minimize σ_f^2 when $H(s)$ is fixed. The ability to carry

out the optimization procedure in this manner is a result of the quasi-linear receiver model.

Since $\theta_1(t)$ is the phase variation of the signal portion of the received waveform $\psi(t)$ [see Eq. (1)] and since the desired output is $m(t)$, i.e., the time derivative of $K_t^{-1} \theta_1(t)$, the linear operation that must be performed on the input signal is one of differentiation; hence, $L(s) = K_t^{-1}s$ and

$$S_f(s) = S_{\theta_1}(s) + S_{n_1'}(s) \tag{23}$$

where, from Eq. (9), the double-sided noise spectral density is

$$S_{n_1'}(s) = \frac{N_0}{2P} \exp (\sigma_\varphi^2) \tag{24}$$

and the phase spectral density $S_{\theta_1}(s)$, obtained from Eq. (22), becomes

$$S_{\theta_1}(s) = -\frac{K_t^2}{s^2} S_m(s) = \frac{2a K_t^2}{s^2(s^2 - a^2)} \tag{25}$$

Direct substitution of Eqs. (24) and (25) into Eq. (20) produces the optimum linear filter that minimizes σ_f, namely,[2]

$$H(s) = \frac{a^2(\gamma - 1)^2 s}{2K_t(s^2 + a\gamma s + a^2\delta)} \tag{26}$$

where

$$\gamma = \sqrt{1 + 2\delta} \; ; \quad \delta = m_f K_\varphi \sqrt{R}$$

$$m_f = \frac{K_t}{a} \; ; \quad R = \frac{P}{aN_0} \tag{27}$$

Note the *dependence* of $H(s)$ on the phase error σ_φ. It seems appropriate to refer to m_f, the ratio of the modulation index to the 3-db frequency, as the *deviation ratio* and to R as the input signal-to-noise ratio in a bandwidth of 1 cps. Thus, the filter $H(s)$ produces the optimum replica of $K_t^{-1} \theta_1'(t)$ for each set of input signal and noise powers.

---

[1]From the physical standpoint this spectrum is convenient to use, in that the optimum closed-loop transfer function turns out to be a second-order system. Second-order loops have been mechanized in the past, and the second-order loop is the optimum filter for separating such a time series from white Gaussian noise.

[2]This result is a trivial generalization to Viterbi's result (Ref. 17). Apparently Lehan and Parks (Ref. 24) were the first to arrive at this equation with $K_\varphi = 1$.

To minimize σ_φ^2, we determine the quasi-linear filter $H_\varphi(s)$ using Eqs. (24) and (25) in Eq. (20) with $L(s) = 1$. Direct substitution yields the result[3]

$$H_\varphi(s) = \frac{a[(\gamma - 1)s + \delta a]}{s^2 + a\gamma s + a^2 \delta} \qquad (28)$$

where δ and γ are defined in Eq. (27).

The loop filter $F(s)$ may be determined using Eq. (28) in Eq. (8). Thus

$$F(s) = \left[\sqrt{P} K_\varphi K_r\right]^{-1} \left[\frac{a(\gamma - 1)s + a(\gamma + 1)}{s + a}\right] \qquad (29)$$

whereas the output filter is determined using Eq. (6). This filter function is

$$F_o(s) = \frac{a(\gamma - 1)^2 K_r}{2K_t[(\gamma - 1)s + a\delta]} \qquad (30)$$

Both filters $F(s)$ and $F_o(s)$ may be realized with RC networks.

2. Network Realization Methods

Consider first the realization of the loop filter $F(s)$ by means of a lag-lead network. Rewriting $F(s)$, we have

$$\sqrt{P} K_\varphi K_r F(s) = a\delta \left[\frac{1 + \dfrac{\gamma - 1}{a\delta} s}{1 + s/a}\right] \qquad (31)$$

$$= K_o \left[\frac{1 + \tau_1 s}{1 + \tau_2 s}\right]$$

where

$$K_o = a\delta = \text{a dc gain}$$

and

$$\tau_2 = \frac{1}{a}; \; \tau_1 = \frac{\gamma - 1}{a\delta} = \left(\frac{2}{\gamma + 1}\right)\tau_2$$

Thus the time constants τ_1 and τ_2 are directly related to the basic communication parameters P, N_0, m_f, and a.

Networks of this type are sometimes referred to as proportional-plus-integral-control networks (Refs. 10 and

25), and typical network configurations are shown in Fig. 4. Various methods of physically realizing the filter function $F(s)$ have been investigated by Martin (Ref. 26). The output filter $F_o(s)$ may be accomplished with network (b) of Fig. 4 by letting $C_1 = 0$. Rewriting $F_o(s)$ and relating this to the pertinent time constants, we have

$$F_o(s) = K_o \left[\frac{1}{1 + \tau s}\right] \qquad (32)$$

where

$$K_o = \frac{K_r}{K_t} \left[\frac{\gamma - 1}{\gamma + 1}\right] = \text{a dc gain}$$

and

$$\tau = R_1 C_2 = \text{time constant}$$

An important conclusion is that the design of the filter function $H(s)$ does not depend on K_r; however, design of the output filter $F_o(s)$ and the loop filter $F(s)$ does. Consequently, the performance of the receiver is not affected by variations in the receiver VCO gain, provided the gain of $F(s)$ and $F_o(s)$ is adjusted accordingly. Of course, $K_r \neq 0$. Further, we see that the optimum filter's pole-zero configuration depends, unfortunately, on both the input signal and noise powers. This means that in order to maintain optimum demodulation at the receiver in the presence of variations in the received signal and noise mean-squared strengths, the filter must be adaptive. Designing such a filter would be extremely difficult, if not impossible. Therefore, in Section E, we shall fix the filter structure, which is optimum only for a certain signal power P and noise power N_0, and compute the filter performance when the filter is subjected to new signal and noise levels.
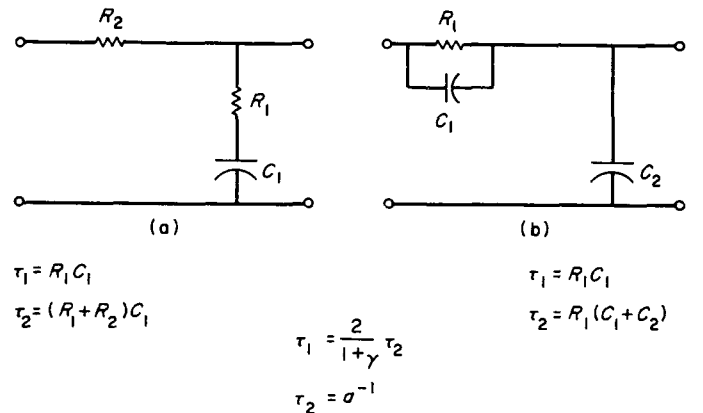

$$\tau_1 = R_1 C_1$$
$$\tau_2 = (R_1 + R_2)C_1$$

$$\tau_1 = R_1 C_1$$
$$\tau_2 = R_1(C_1 + C_2)$$

$$\tau_1 = \frac{2}{1 + \gamma}\tau_2$$

$$\tau_2 = a^{-1}$$

Fig. 4. Realization of the loop filter $F(s)$

[3]See Footnote 1.

E. Receiver Performance When the PLL Is Preceded by an AGC Amplifier

Having determined the various filter components that comprise the receiver structure, we shall now determine its performance. The problem here is to "compare" a fixed filter, which is optimum for only one value of input signal-to-noise ratio, say R_D, with the optimum receiver structure. Briefly, we shall fix the filter, designed to be optimum at a value of $R = R_D$, and then subject it to a new signal power P_1 and a new noise power N_1. Of course, even though the loop filter is fixed, the loop bandwidth changes as a function of P_1. Practically speaking, however, the loop would normally be preceded by an AGC amplifier; we may hence assume that, for this Section, $P_1 = P$. In Section I, we investigate the performance when the input power changes from the expected value. The loop bandwidth is defined in Section H.

1. The Receiver Performance Criterion

First, we must define what we mean by receiver performance. One performance criterion that has been used successfully for design purposes is that of output signal-to-noise ratio. There are several definitions that are available for describing the signal-to-noise ratio; for example, Choate (Ref. 18) defines the output signal-to-noise ratio as the ratio of the mean-squared output signal power to total mean-squared error. An obvious flaw (error) in this definition is the fact that Choate has included the signal distortion term [see Eq. (18)] in both the numerator and denominator. If we neglect the signal distortion term in the denominator, Choate's definition becomes the ratio of the mean-squared output signal power to the mean-squared output noise power, which has been used by many authors (see Ref. 1). This definition, however, has the disadvantage of approaching infinity as the additive noise approaches zero, and in any physical (nonadaptive) system this is not true because the system introduces a certain amount of signal distortion. Another definition is to consider the output signal-to-noise ratio as being the ratio of that portion of the output signal $y(t)$ which correlates with the desired output signal $m(t)$ to the total mean-squared error found by subtracting the output $y(t)$ from the desired output $m(t)$, squaring, and averaging, i.e.,

$$\text{SNR} = \frac{\overline{y(t)m(t)}}{\overline{[y(t) - m(t)]^2}} \qquad (33)$$

However, in this definition the numerator includes an output power component that does not appear in the output.

For our purposes here we define the "output" signal-to-noise ratio ρ_0 as the ratio of the mean-squared value of the desired signal $m(t)$ to the total mean-squared filter frequency error σ_f, i.e., the noise in the system is taken to be equal to σ_f. Thus

$$\rho_0 = \frac{\overline{m^2(t)}}{\overline{[y(t) - m(t)]^2}} = \frac{1}{\sigma_f^2} \qquad (34)$$

is our measure of signal fidelity, and we see that this definition is nothing more than the reciprocal of the Wiener error. [Note that $\overline{m^2(t)} = 1$.]

Justification for the use of Eq. (34) follows. Regardless of how the output signal $y(t)$ is written, it contains a certain amount of power, say σ_y^2. Furthermore, any definition of output signal-to-noise ratio that is reasonable must be consistent with σ_y^2, i.e., the sum of the numerator and the denominator must equal σ_y^2. Choate's definition (Ref. 18) is not consistent with a constant mean-squared output power, and, neglecting the signal distortion term in the numerator, it has the other disadvantage of approaching infinity when the system noise is zero. Although Eq. (33) is not consistent with σ_y^2, Eq. (34) is; moreover, Eq. (34) is nothing more than the reciprocal of the criterion chosen for determining the optimum filter.

2. The Receiver Threshold Criterion

Before the characteristics of the receiver are computed analytically, it remains to define the PLL threshold. Develet (Ref. 18) and Viterbi (Ref. 17) take the threshold condition of the PLL to exist when the total mean-squared phase error (signal distortion plus noise distortion) is one radian squared. This definition seems to be justifiable from an experimental viewpoint.

In the results to follow we shall assume that the PLL threshold exists when two conditions are satisfied. First, the threshold of the PLL is defined to be the locus of a point that moves so that the total mean-squared phase error σ_φ^2 is equal to a constant C. (We shall specify C in Section I.) The other condition assumes that the noise and available signal powers at threshold are those that are expected under the worst channel conditions, i.e., those resulting from communication at the greatest range expected.

3. Derivation of the Output Signal-to-Noise Ratio and the Receiver Threshold Characteristic

The output signal-to-noise ratio may be easily computed using Eq. (17) and the appropriate filter and signal density equations. First we compute the signal distortion from Eq. (18). This term is easily obtained by direct substitution of Eqs. (25) and (48) into Eq. (18) where $L(s) = K_t^{-1}s$. Equation (48) is derived in Section F under the assumption that the loop filter $F(s)$ is fixed and is determined by the initial design conditions. Obviously, when the loop is preceded by an AGC amplifier, the over-all filter function $H(s)$ changes only when the quasi-linear model is used, i.e., it depends on the total phase error. For the linear model the over-all fixed filter is determined by the design signal and noise levels.

Substituting Eqs. (25) and (48) into Eq. (18) and carrying out the integration in the complex plane give

$$S_d = \frac{4\delta\beta\,(\delta + \alpha) + 4\delta\beta\,(\alpha^2 - \kappa) + \kappa^2\,(\alpha + 1)}{4\delta\beta\,[(\alpha + 1)\,(\delta + \alpha) - \delta\beta]} \quad (35)$$

where

$$\alpha = 1 + \beta\,(\gamma - 1)$$

$$\kappa = \beta\,(\gamma^2 - 1) - (\gamma - 1)^2 \quad (36)$$

$$\beta = \exp\left(\frac{\sigma_\varphi^2 - \sigma_1^2}{2}\right)$$

Similarly, that portion of the total error due to the additive noise is easily found from the second term of Eq. (17). This turns out to be, for $S_{n'}(s) = N_0/K_\varphi^2 P$,

$$\sigma_{n'}^2 = \frac{(\gamma - 1)^2}{\alpha\,(\gamma + 1)^2} \quad (37)$$

while for a new noise spectral density of $S_{n_1'}(s) = N_1/K_1^2 P$,

$$\sigma_{n_1'}^2 = \frac{N_1}{N_0}\frac{(\gamma - 1)^2}{\alpha\,(\gamma + 1)^2} \quad (38)$$

If we substitute Eqs. (35) and (38) into Eq. (34) we find the "output" signal-to-noise ratio ρ_0, namely,

$$\rho_0 = \frac{1}{S_d + \dfrac{1}{\mu}\left[\dfrac{(\gamma - 1)^2}{\alpha\,(\gamma + 1)^2}\right]} \quad (39)$$

where

$$\mu = \frac{P}{aN_0} \times \frac{aN_1}{P} = \frac{R_D}{R} \quad (40)$$

and R_D is the input signal-to-noise ratio in a bandwidth of $a/2\pi$ cps for the greatest range expected, while R is the actual input signal-to-noise ratio (due to changing channel conditions) referred to a bandwidth of $a/2\pi$ cps. As a novel by-product of the analysis, the signal-to-noise ratio for the optimum (adaptive) receiver is obtained when $R = R_D$, i.e.,

$$\rho_{\mathrm{opt}} = \frac{(\gamma + 1)^2}{4\gamma} \quad (41)$$

This result was originally derived by Viterbi (Ref. 17).

One final computation is that of determining the threshold locus of the PLL. This locus is found from Eq. (17) with $L(s) = 1$, $H(s)$ replaced by $H_\varphi(s)$ of Eq. (28), and $S_x(s)$ replaced by Eq. (25). The noise spectral density $S_{n_1'}(s)$ is obtained from Eq. (24) by replacing N_0 by N_1.

Substituting these quantities into Eq. (17) and performing the integration,

$$\sigma_1^2 = \frac{2m_f^2}{\alpha\beta\,(\gamma^2 - 1)} \times \left[1 + \mu\,\frac{[2\beta\,(\gamma - 1)^2 + \gamma^2 - 1]}{(\gamma^2 - 1)}\right] \quad (42)$$

for the suboptimum or fixed-receiver structure. Letting $\mu = 1$, we obtain the desired threshold locus

$$\sigma_\varphi^2 = \frac{8m_f^2}{(\gamma - 1)\,(\gamma + 1)^2} = C \quad (43)$$

In Section I, we will have more to say about the constant C.

F. Receiver Performance for Reception of a Signal Whose Power Is Different From the Design Level

It was pointed out in Section B that the closed-loop transfer function of the PLL depends on the received signal power P. In this Section, we wish to determine the effects on receiver performance when the transmitter power, say P_1, differs from the design level P_D. We shall assume further that the noise spectral density is N_0 w/cps and that it remains fixed.

Since the closed-loop transfer function of the PLL varies with the received signal power, it is necessary to determine the new transfer function when the filter is subjected to the new signal power P_1. Contrary to the situation in Section E (where only the error due to noise affected the performance) we find that a change in signal power affects both the noise and the signal error. This is, of course, due to the changing closed-loop transfer function, which occurs in both of the error expressions.

From Eq. (7) we may rewrite, for an arbitrary signal power P_1, the quasi-linear PLL receiver transfer function as

$$H_\varphi(s) = \frac{K_r K_1 \sqrt{P_1}\, F(s)}{s + K_r K_1 \sqrt{P_1}\, F(s)} \qquad (44)$$

where

$$K_1 = \exp\left(-\sigma_1^2/2\right) \qquad (45)$$

and σ_1^2 is the mean-squared phase error for this new power level. Substituting the fixed-loop filter $F(s)$, Eq. (29), evaluated at design levels into Eq. (44), and rearranging,

$$H_\varphi'(s)$$
$$= \frac{\beta\,[a\,(\gamma - 1)\,s + a\,(\gamma + 1)]}{s^2 + as\,[1 + \beta\,(\gamma - 1)] + a\,[(\gamma + 1)\,\beta]} \qquad (46)$$

where

$$\beta = \sqrt{\frac{P_1}{P_D}}\,\exp\left(+\,\frac{\sigma_\varphi^2 - \sigma_1^2}{2}\right) \qquad (47)$$

At $\beta = 1$, $H_\varphi'(s)$ reduces to $H_\varphi(s)$ as it should. The overall fixed FM receiver filter is obtained from Eqs. (6) and (30), namely,

$$H'(s) = \frac{\beta a^2\,(\gamma - 1)\,s}{2K_r\,[s^2 + a\,\{1 + \beta\,(\gamma - 1)\}\,s + a^2\,\delta\beta]} \qquad (48)$$

The new frequency error σ_f^2 is easily obtained from Eq. (17). First we compute the new signal distortion term using Eqs. (25) and (48) in Eq. (18). Carrying out the necessary integration, we find

$$S_d' = \frac{4\delta\beta\,(\delta + \alpha) + 4\delta\beta\,(\alpha^2 - \kappa) + \kappa^2\,(\alpha + 1)}{4\delta\beta\,[(\alpha + 1)\,(\delta + \alpha) - \delta\beta]} \qquad (49)$$

where

$$\alpha = 1 + \beta\,(\gamma - 1)$$

$$\kappa = \beta\,(\gamma^2 - 1) - (\gamma - 1)^2$$

and β is defined in Eq. (47).

On the other hand, the variance of the new noise error is easily found by substituting Eq. (48) in the second term of Eq. (17) and integrating. The result is

$$\sigma_{n_1'}^2 = \frac{(\gamma - 1)^2}{\alpha\,(\gamma + 1)^2} \qquad (50)$$

Hence, the new "output" signal-to-noise ratio ρ_0' is, from Eqs. (49) and (50),

$$\rho_0' = \frac{1}{S_d' + \dfrac{(\gamma - 1)^2}{\alpha\,(\gamma + 1)^2}} \qquad (51)$$

which is the required result.

The new phase error σ_1^2 that results from a change in received signal power may be written as

$$\sigma_1^2 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} S_{\theta_1}(s)\,|1 - H_\varphi'(s)|^2\,ds + \frac{1}{2\pi j}$$
$$\int_{-j\infty}^{j\infty} S_{n'}(s)\,|H_\varphi'(s)|^2\,ds \qquad (52)$$

and becomes

$$\sigma_1^2 = \frac{2m_f^2}{\alpha\beta\,(\gamma^2 - 1)}\left[1 + \frac{2\beta\,(\gamma - 1)^2 + \gamma^2 - 1}{(\gamma^2 - 1)}\right] \qquad (53)$$

when Eq. (46) is substituted into Eq. (52) and the integration is performed. Thus, Eqs. (51) and (53) are the required results needed for specifying the receiver performance.

G. Receiver Performance When the PLL Is Preceded by a Bandpass Limiter

In Section E, we investigated the demodulator performance when the PLL is preceded by an AGC amplifier that holds the signal level constant and when the

loop filter is designed in accordance with the signal power P and noise power N_0 expected under worst channel conditions. In this Section we again fix the loop filter as was done in Section E and insert a bandpass limiter ahead of the PLL. Since the optimum demodulator requires the use of complex auxiliary servo loops that must be capable of continually adjusting the pole-zero configuration of the demodulator, it behooves the design engineer to seek ways and means of mechanizing the receiver so as to have it perform at a near-optimum level without the use of the exact equipment called for by the deceptively simple adaptive filter equations. It appears, from an experimental standpoint, that the band-pass limiter may be an excellent engineering approximation to the complex servo system that we require for optimum performance. A comparison of the results to follow with those of Section E provides the engineer with design data that will allow him to select the AGC amplifier or the bandpass limiter as a possible means for obtaining near-optimum performance.

1. Derivation of the Threshold Characteristic and Output Signal-to-Noise Ratio

The appropriate filter functions required in the derivation are given by Eqs. (46) and (48). As a matter of fact we may make use of the results obtained in Section F for the signal distortion term. That portion of the error due to the new noise N_1 is easily found by substituting Eq. (48) into the second term of Eq. (13) and integrating. The result is

$$\sigma_{n'_1}^2 = \frac{N_1}{N_0} \left[\frac{(\gamma - 1)^2}{\alpha (\gamma + 1)^2} \right] \tag{54}$$

The new phase error, which results from a change in received signal and noise power, may be written as

$$\sigma_1^2 = \frac{1}{2\pi j} \left[\int_{-j\infty}^{j\infty} S_{\theta_1}(s) | 1 - H'_\varphi (s) |^2 \, ds \right.$$

$$\left. + \int_{-j\infty}^{j\infty} S_{n'_1}(s) | H'_\varphi (s) |^2 \, ds \right] \tag{55}$$

and becomes

$$\sigma_1^2 = \frac{2m_f^2}{\alpha\beta (\gamma^2 - 1)} \left[1 + \frac{N_1}{N_0} \left(\frac{2\beta (\gamma - 1)^2 + \gamma^2 - 1}{\gamma^2 - 1} \right) \right]$$

when Eq. (46) is substituted into Eq. (55) and the integration is performed.

Thus to obtain the resulting "output" signal-to-noise ratio we need to relate the ratios N_1/N_0 and P_1/P_D to the input–output signal-to-noise ratio relationship of the limiter. This, however, requires analyzing the limiter, and we shall present this analysis next.

2. Signal-to-Noise Ratios in Bandpass Limiters

In Ref. 21, a general analysis has been made relating output signal and noise powers and input signal and noise powers for bandpass limiters having odd symmetry in their limiting characteristics.

Briefly, a bandpass limiter is an electronic device that consists of a limiter followed by a bandpass filter. The form of the limiter characteristic is

$$g[x(t)] = \begin{cases} \alpha_1 [x(t)]^{\frac{1}{n}} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases} \tag{56}$$

where α_1 is the gain of the limiter and x is a Gaussian time series. The case $n = 1$ corresponds to a linear amplifier, while the case $n = \infty$ corresponds to the ideal symmetrical limiter (or clipper). For $n = \infty$, Davenport (Ref. 21) has shown that the following results are true:

$$\left(\frac{S}{N} \right)_o \approx 2 \left(\frac{S}{N} \right)_i \qquad \text{for } \left(\frac{S}{N} \right)_i \to \infty$$

$$\left(\frac{S}{N} \right)_o \approx \frac{\pi}{4} \left(\frac{S}{N} \right)_i \qquad \text{for } \left(\frac{S}{N} \right)_i \to 0 \tag{57}$$

where $(S/N)_o$ is the signal-to-noise power ratio at the limiter output, $(S/N)_i$ is the signal-to-noise power ratio at the limiter input, and

$$S_o + N_o \sim \frac{2\alpha_1^2}{\pi} \left[\frac{4}{\pi} + \left(\frac{S}{N} \right)_i \right] \text{ for } \left(\frac{S}{N} \right)_i \to 0 \tag{58}$$

$$S_o + N_o \sim \frac{2\alpha_1^2}{\pi} \left[\frac{4}{\pi} + \left(\frac{N}{S} \right)_i \right] \text{ for } \left(\frac{S}{N} \right)_i \to \infty$$

Further, Davenport shows that

$$\frac{\pi}{4} \leq \left(\frac{S}{N}\right)_o \times \left(\frac{S}{N}\right)_i^{-1} \leq 2 \text{ for } 0.001 \leq \left(\frac{S}{N}\right)_i \leq 100$$

(59)

which is a variation of about 4.06 db. (See Fig. 5 for the exact results.) These results have been experimentally verified by Youla of this Laboratory (Ref. 12). Equations (58) and (59) suggest that the bandpass limiter possesses the property of being able to hold the output power constant. Consequently, if the noise input to the system is increased, the signal strength at the limiter output must decrease. However, the noise bandwidth B_L of the loop has been shown to be directly dependent on the signal power P (see Section H). Thus an increase in input noise, for a fixed signal power, causes the output signal power, and hence the loop bandwidth, to be reduced; as a result, the phase jitter at the VCO output is a smaller percentage of the input noise. Therefore, the over-all demodulator appears to be self-adaptive.

Using Eqs. (58) and (59) and assuming that the gain of the bandpass limiter is adjusted so that the output signal power plus noise power are those values for which the optimum loop was designed, it is easy to show that

$$\frac{N_1}{N_o} = \frac{1 + R_D}{1 + R}$$

and   (60)

$$\frac{P_1}{P_D} = \frac{R}{R_D}\left(\frac{1 + R_D}{1 + R}\right)$$

where

R = Input signal-to-noise ratio of the bandpass limiter

R_D = Input signal-to-noise ratio for which the loop was designed

P_1, N_1 = Bandpass limiter output noise and signal levels, respectively

and

P_D, N_o = Signal and noise levels for which the loop was designed.

Thus Eq. (60) may be substituted into Eqs. (49) and (54) to yield the output signal-to-noise ratio, i.e.,

$$\rho'_0 = \frac{1}{S'_d + \left(\dfrac{1 + R_D}{1 + R}\right)\left[\dfrac{(\gamma - 1)^2}{\alpha\,(\gamma + 1)}\right]}$$

(61)

where S'_d is given by Eq. (49) and

$$\alpha = 1 + \beta\,(\gamma - 1); \quad \kappa = \beta\,(\gamma^2 - 1) - (\gamma - 1)^2$$

(62)

$$\beta = \sqrt{\frac{R}{R_D}\left(\frac{1 + R_D}{1 + R}\right)}\,\exp\left(\sigma_\varphi^2 - \sigma_1^2\right)$$

which is the required result.

Before using the results obtained above, a word of caution should be issued. Recall that the optimum demodulator filters were selected on the basis that the perturbing noise was white and Gaussian. In fact, now that the bandpass limiter has been inserted, it is inconceivable that the noise that the PLL sees is white or even Gaussian. The fact is that under the original criterion, the PLL is now suboptimum since the disturbing noise is non-Gaussian and non-white; however, the over-all system is optimum when $R = R_D$, and it may actually perform optimally for $R > R_D$. This would indicate that even though the exact mechanization is not possible, the bandpass limiter, in combination with the PLL and the output filter, serves as the engineering device that performs almost identically with the exact mechanization. In Section I, we shall graphically illustrate its performance and compare it with the adaptive filter.

Furthermore, in deriving the "output" signal-to-noise ratio ρ'_0 we were forced to assume that the noise out of the bandpass limiter was white or at least contained a spectrum that was relatively flat over a band of frequencies that was wide in comparison with the bandwidth of the PLL. In computing the signal distortion no such assumption was made, nor was it necessary. If the bandwidth of the bandpass filter turns out to be wide when compared with the loop bandwidth B_L, we would expect, to a good approximation, that the output noise spectrum of the bandpass limiter would be relatively flat; thus, for all practical purposes, the PLL sees noise that has a relatively flat spectrum. Hence the assumption that the noise has a flat spectrum over a band of frequencies much greater than B_L cps seems to be justifiable, and the "output" signal-to-noise ratio ρ'_0 is valid.
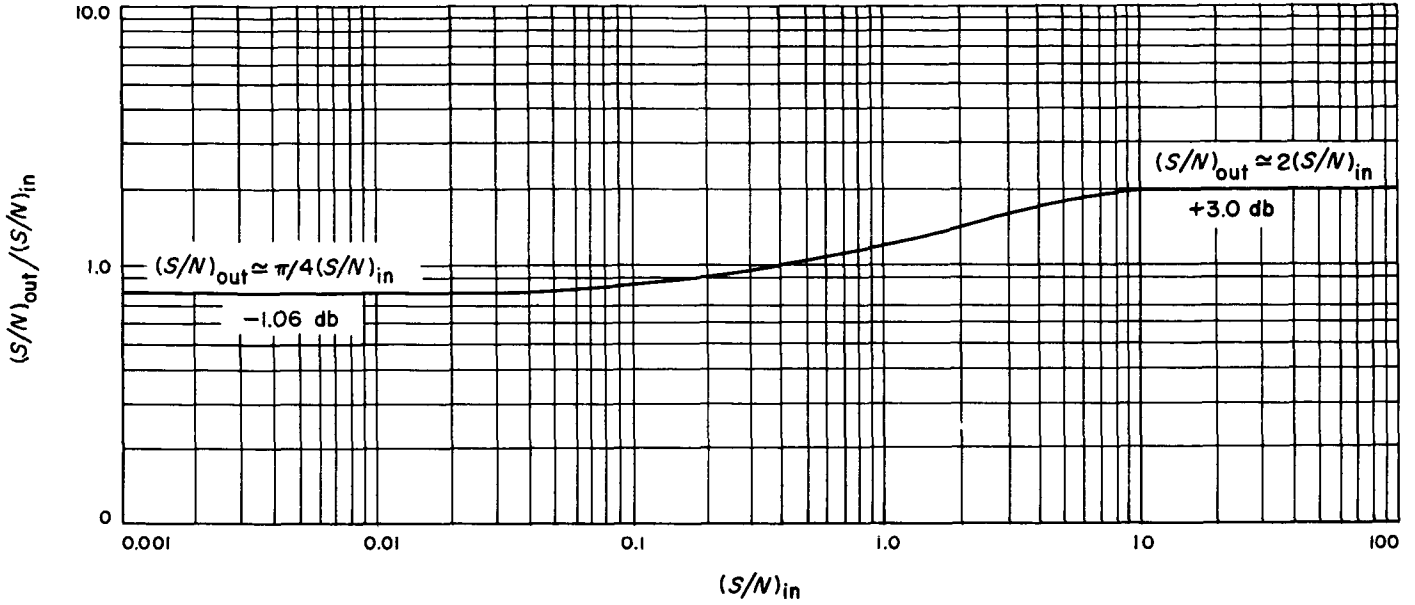
Fig. 5. Bandpass limiter characteristics

H. Transient Response of the Optimum Filter

In order to discuss the transient response of the optimum filter, it is necessary to relate the transfer function $H_\varphi(s)$ to a standard form given by Truxal (Ref. 27) for a second-order system. This standard form, using Truxal's notation, is

$$G_2(s) = \frac{\omega_n^2}{z_1} \frac{s + z_1}{s^2 + 2\xi\omega_n s + \omega_n^2} \qquad (63)$$

where ξ represents the system damping factor. A ξ of 1 corresponds to critical damping, $\xi < 1$ corresponds to an underdamped (oscillatory) system, and $\xi > 1$ corresponds to an overdamped (nonoscillatory) system. For purposes of comparison, we rewrite Eq. (28) as

$$H_\varphi(s) = \frac{a(\gamma - 1)\left[s + \dfrac{a\delta}{\gamma - 1}\right]}{s^2 + a\gamma s + a^2\delta} \qquad (64)$$

The quantity ω_n of Eq. (63) is taken to be the natural loop frequency, while z_1 is the closed-loop system zero. Thus, comparison of Eqs. (63) and (64) shows that

$$z_1 = \frac{a\delta}{\gamma - 1} = \frac{a}{2}(\gamma + 1)$$

$$\omega_n^2 = a^2\delta \qquad (65)$$

$$\xi = \frac{1}{\sqrt{2}}\left(1 + \frac{1}{2\delta}\right)^{1/2} \sim \frac{1}{\sqrt{2}}$$

with a total error of less than $25\delta^{-1}\%$. Experimental evidence shows that if the damping factor of the loop is made equal to $1/\sqrt{2}$, a satisfactory balance is achieved between the transient error and the output noise. Thus if $\delta = 25$, an equality that must be true if the optimum frequency demodulator is to produce the signal with any fidelity at all, $\xi \doteq 0.707$ with a total error less than, or equal to, 1%.

A parameter that has received considerable attention in the design of a PLL loop is the loop-noise bandwidth B_L, defined as

$$2B_L = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} |H_\varphi(s)|^2\, ds \qquad (66)$$

Substituting Eq. (64) into Eq. (66) and integrating give

$$\frac{B_L}{a} = \frac{(\gamma - 1)(3\gamma - 1)}{8\gamma} \qquad (67)$$

15

Solving Eq. (67) for γ, we find

$$\gamma = \frac{4}{3}\left(1 + \frac{3B_L}{a}\right)\left[1 + \sqrt{1 + \frac{1}{\left[\frac{2}{3}\left(1 + \frac{2B_L}{a}\right)\right]}}\right] \quad (68)$$

which is asymptotic to

$$\gamma \sim \frac{4}{3}\left(1 + \frac{3B_L}{a}\right) \quad (69)$$

with an error less than

$$\varepsilon_\gamma < \frac{8}{3}\left(1 + \frac{2B_L}{a}\right)^2$$

For large γ or a large normalized 3-db loop bandwidth B_L/a, we have

$$\gamma \sim \frac{8B_L}{3a} \quad (70)$$

Substituting Eq. (70) into Eq. (64) we obtain, for reasonable demodulator input conditions,

$$H_\varphi(s) \doteq \frac{1 + \frac{3}{4B_L}s}{1 + \frac{3}{4B_L}s + \frac{9}{32B_L^2}s^2} \quad (71)$$

which has a damping factor of $1/\sqrt{2}$. Interestingly enough, if one derives the optimum closed-loop linear filter that minimizes the peak error $|\theta_1 - \theta_2|_{\max} + n_{\max}$ when the filter is subjected to a frequency ramp and further demands that the damping factor be equal to $1/\sqrt{2}$, a filter that is identical with Eq. (71) results (Refs. 12 and 16). This is a striking result, for it requires that, given reasonable input demodulator conditions and proper construction of the loop filter $F(s)$, the frequency demodulator that minimizes the Wiener error for an RC–filtered white-noise modulating spectrum will also perform optimally, for all practical purposes, under a peak error criterion where the signal to be tracked or followed is a frequency ramp. This result should be very
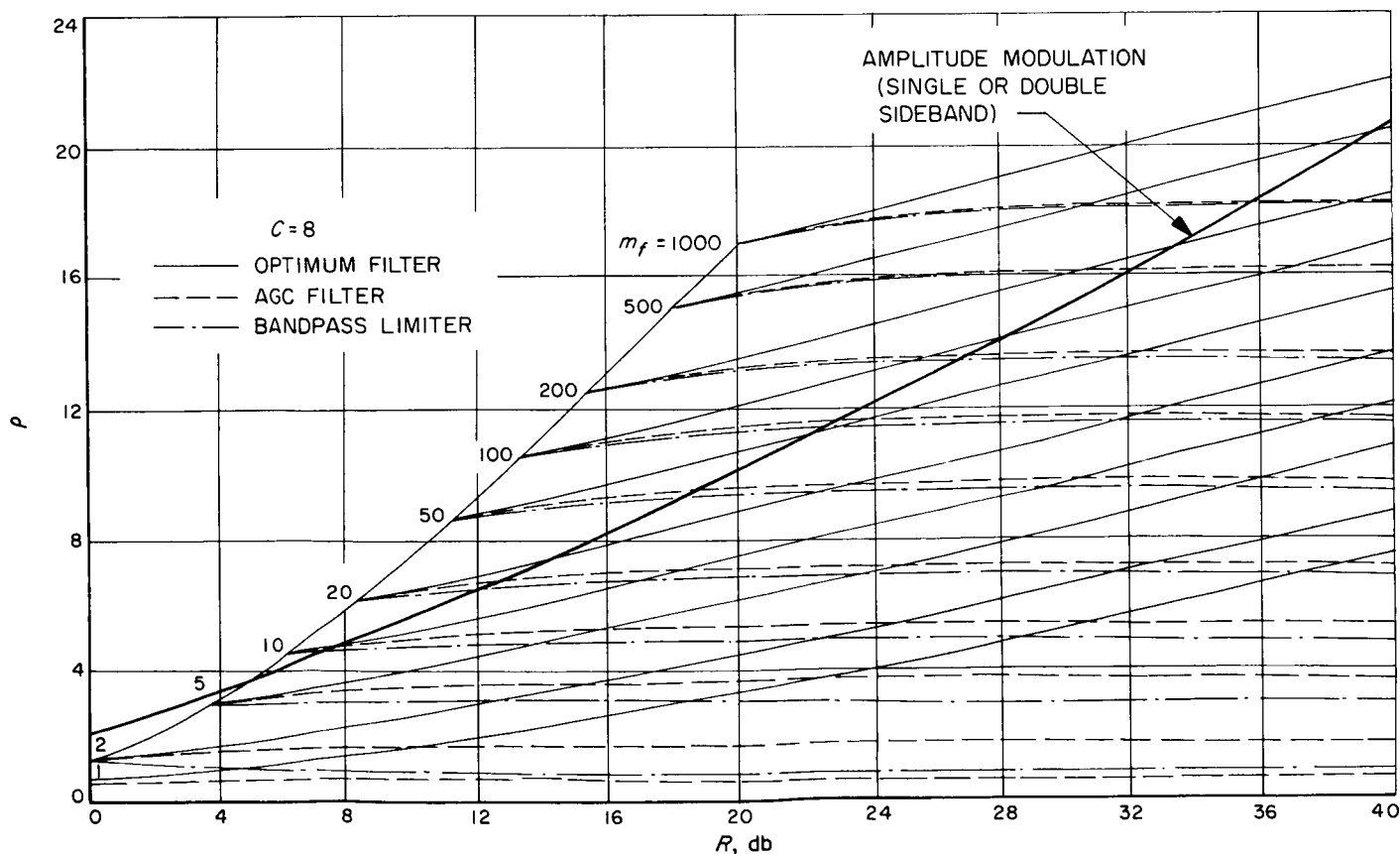

Fig. 6. Performance characteristics for the linear demodulator $\left(\sigma_\varphi^2 = 1\right)$

pleasing to design engineers who (for some time) have apparently been building phase-locked loops with transfer functions approaching that of Eq. (71).

I. Graphical Results for the Linear and Quasi-linear Demodulators

In Fig. 6 we have graphically illustrated the performance of the linear FM demodulator for three different receiver terminations: the optimum PLL demodulator and the phase-locked demodulator (which is designed to be optimum for the most deleterious channel conditions) preceded by a bandpass limiter or an AGC amplifier. From Fig. 6 it is evident that the optimum demodulator outperforms either of the other two realizations; however, the "fixed" PLL, when preceded by the AGC amplifier, performs better than the loop that is preceded by a bandpass limiter. The amount of superiority becomes smaller as the modulation index m_f increases. In either of the two suboptimum systems the signal-to-noise ratio ρ becomes asymptotic (for large R) to the

reciprocal of the signal distortion S_d initially designed into the system. For comparison purposes we indicate the performance (derived in Ref. 29) of an amplitude-modulated double-sideband or single-sideband suppressed carrier system that demodulates the noisy received data by coherent frequency translation and by smoothing the resulting waveform with a Wiener filter. The improvement obtained by using FM is clearly manifested.

In Figs. 7 and 8 we have illustrated the performance of the same three systems when the quasi-linear receiver models are used. Figure 7 has been prepared under the assumption that receiver threshold occurs when the total phase error equals one-half radian squared, i.e., $C = 1/2$, and Fig. 8 illustrates the performance of the three systems for $C = 1$. Notice the higher threshold characteristic for the situation where $C = 1/2$.

The "dipping" behavior for either of the suboptimum quasi-linear systems may be explained as follows. As before, the asymptotic behavior (for large R) is ultimately
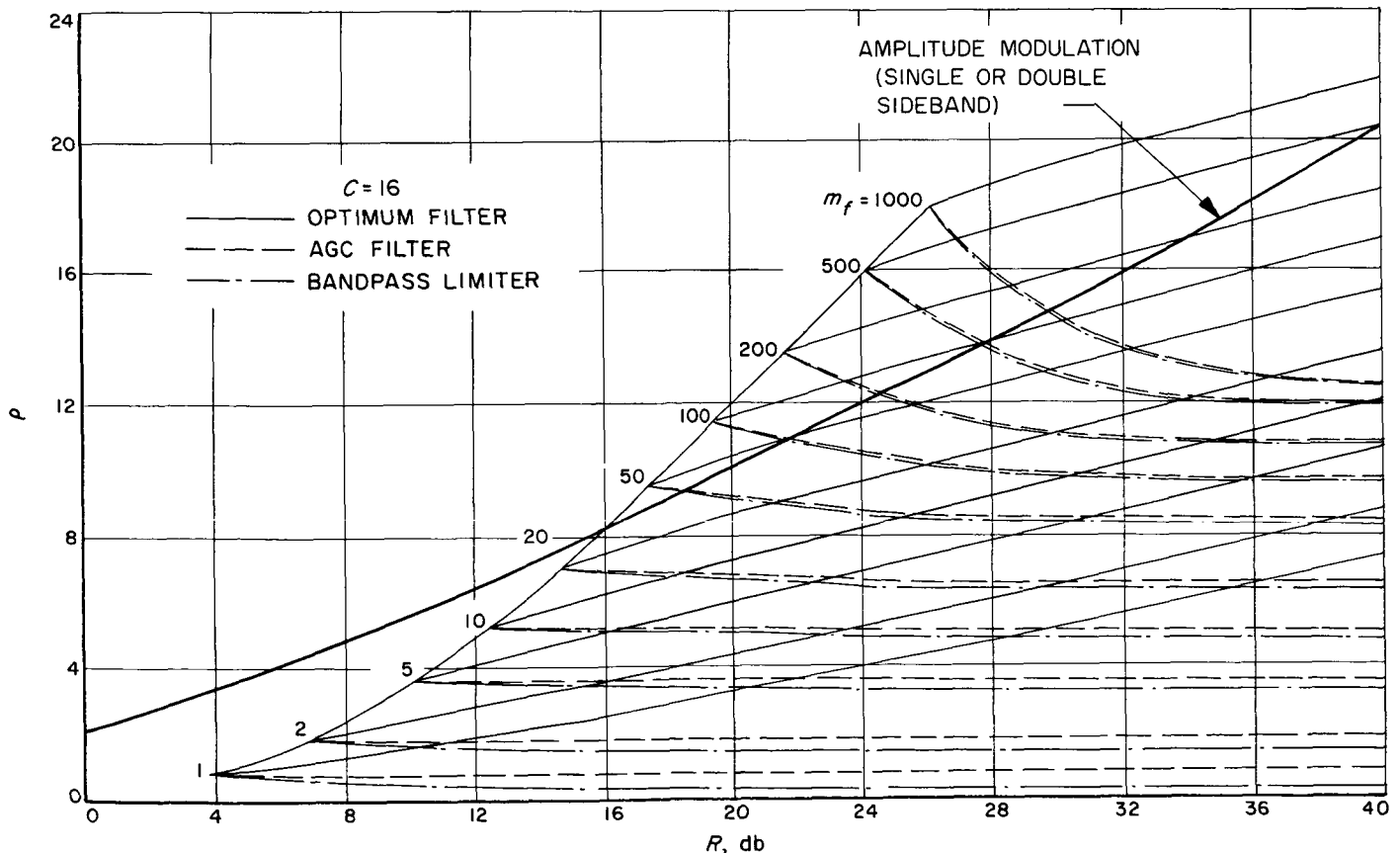

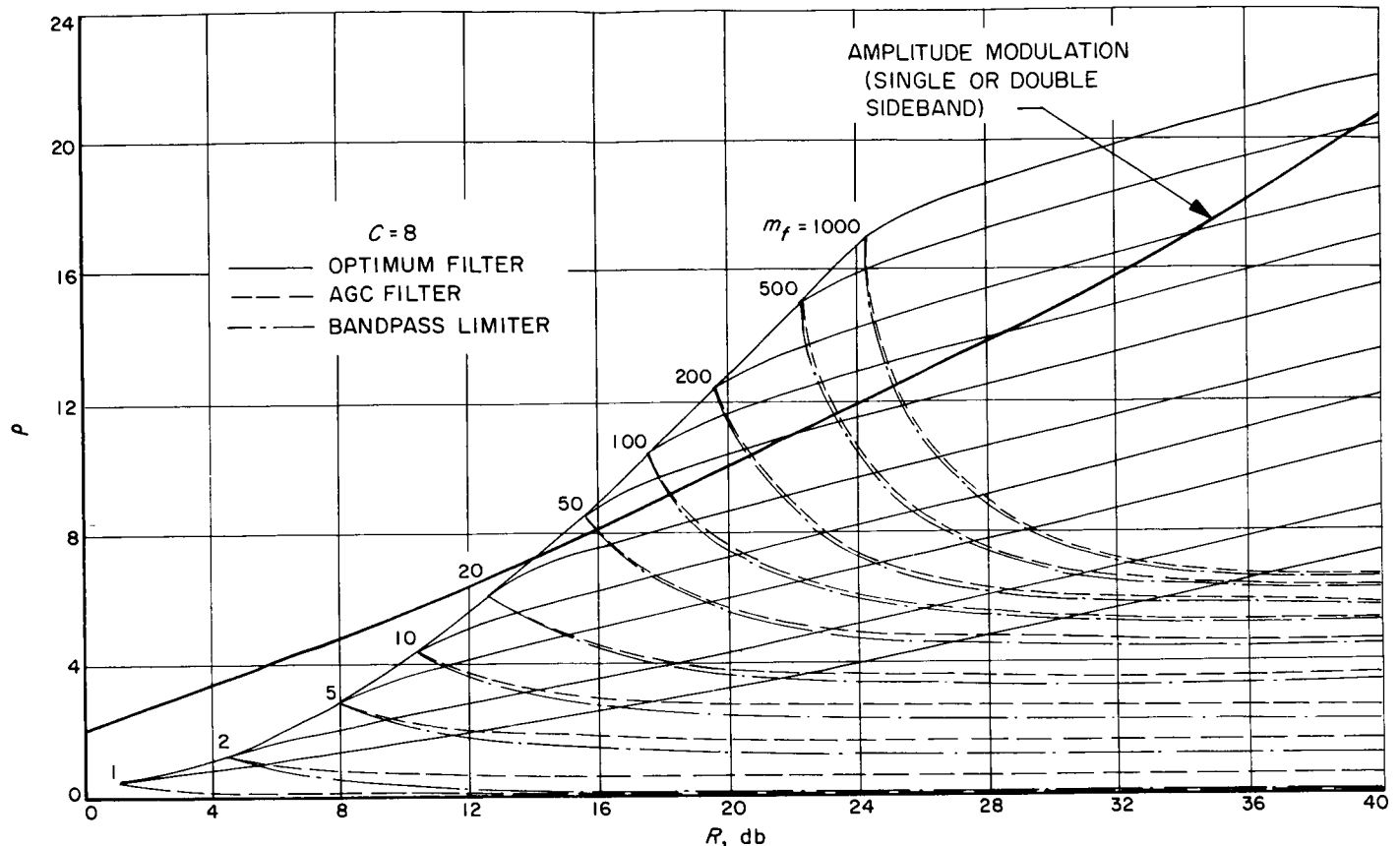Fig. 7. Performance characteristics for the quasi-linear demodulator $\left(\sigma_\varphi^2 = 1 \right)$

Fig. 8. Performance characteristics for the quasi-linear demodulator $\left(\sigma_\varphi^2 = \frac{1}{2} \right)$

determined by the signal distortion S_d designed into the system. For the linear receiver models, this value of S_d was constant for all $R > R_D$. In the quasi-linear receivers, however, the signal distortion S_d is a function of the initial design parameters and the phase error is a minimum at threshold (worst channel conditions expected). As R increases beyond R_D, the signal distortion increases until the phase error is so small that S_d is again determined by the same function of the signal-to-noise ratio existing in the channel. These curves show the importance of keeping the system operating optimally. In fact, for

$C < 1/2$, it has been shown that the quasi-linear model specifies fairly accurately the performance of a PLL. As before, the "fixed" loop, when preceded by an AGC amplifier, outperforms the "fixed" loop preceded by a bandpass limiter.

Comparison of Fig. 6 with Fig. 8 shows that the linear system threshold characteristic is slightly lower than that of the linear receiver. These results should prove beneficial to engineers faced with the problem of designing second-order PLL frequency demodulators.

NOMENCLATURE

$a/2\pi$ 3-db bandwidth of message spectrum

B_L loop bandwidth

C constant

C_1 capacitance

$e(s)$ Laplace transform of the input signal to the VCO

$e(t)$ loop filter output

$F(s)$ loop filter

$F_o(s)$ output filter

$H(s)$ transfer function of the receiver

$H_\varphi(s)$ closed-loop transfer function

$h(t)$ impulse response of the receiver

K_0 dc gain

K_1, K_φ gain of the nonlinear receiver element

K_r gain of the receiver VCO, rad/sec/v

K_t gain of the transmitter VCO, rad/sec/v

$\mathcal{K}$ arbitrary filter

L^{-1} inverse Laplace transform

$L(s)$ desired linear operation in transform notation

$\mathcal{L}$ known mathematical operator

m_f deviation ratio (ratio of modulation index to the 3-db frequency)

$m(t)$ random signal used to modulate the transmitter

N_0 single-sided spectral density, w/cps

N_1 bandpass limiter output noise level

N_o average noise power at the limiter output

$n'(s)$ Laplace transform of equivalent additive noise $n(t)$

$n(t)$ additive white Gaussian channel noise

$n'_1(t)$ wide-sense stationary, white Gaussian noise process

P average power of received signal

P_1 bandpass limiter output signal level

P_D average signal level for which loop was designed

R input signal-to-noise ratio

R_D input signal-to-noise ratio in a bandwidth of $a/2\pi$ cps for the greatest range expected, or the input signal-to-noise ratio for which the loop was designed

$r(t)$ reference signal at output of receiver VCO

$(S/N)_i$ signal-to-noise power at the limiter input

$(S/N)_o$ signal-to-noise power at the limiter output

SNR   signal-to-noise ratio

S_d signal distortion

$S_f(s)$ arbitrary spectral density

$S_m(s)$ message spectral density

$S_n(s)$ spectral density of the additive channel noise

$S_{n'}(s)$ spectral density of the noise $n'(t)$ that produces the same effects as $n(t)$

$S_{n'_1}(s)$ double-sided noise spectral density

S_o average signal power at the limiter output

$S_s(s)$ spectral density of an arbitrary signal $s(t)$

$S_{o_1}(s)$ signal-phase spectral density

$S_\varphi(s)$ phase-error spectral density

$s(t)$ arbitrary signal

x the value that the Gaussian time series $x(t)$ assumes at time t

x_i input of an amplitude-sensitive element

x_r response of amplitude-sensitive element

$y(t)$ linear filter output

z_1 closed-loop system zero

$\alpha, \beta, \gamma, \delta, \kappa, \mu$ relationships between basic system parameters

α_1 gain of the limiter

$\varepsilon, \varepsilon(s)$ arbitrary variation in $H(s)$

$\varepsilon(t)$ instantaneous error of reproduction

NOMENCLATURE (Cont'd)

$\theta_1(s)$ Laplace transform of phase modulation used at the transmitter

$\theta_2(s)$ Laplace transform of the estimate of $\theta_1(s)$

$\theta_1(t)$ phase modulation used at the transmitter

$\theta_2(t)$ estimate of the phase modulation used at the transmitter

λ Lagrangian multiplier

ξ system damping factor

$\xi(t)$ frequency-modulated wave

ρ_0 ratio of the mean-squared value of the desired signal to the total mean-squared filter frequency error

σ^2 mean-squared error of reproduction

σ_f^2 mean-squared value of the total frequency error

$\sigma_{n'_1}^2$ mean-squared error due to n'_1

σ_y^2 mean-squared value of $y(t)$

σ_ε^2 mean-squared error in resulting variation

σ_φ^2 mean-squared value of the total phase error

$\phi(t)$ phase error

$\phi(s)$ Laplace transform of the phase error $\phi(t)$

$\psi(t)$ observed data

ω_n natural loop frequency

REFERENCES

1. Middleton, D., *Introduction to Statistical Communication Theory*, McGraw-Hill Book Company, Inc., New York, 1960.

2. Chaffee, J. G., "The Application of Negative Feedback to Frequency Modulation Systems," *Proceedings of the IRE*, May 1939, pp. 317–331.

3. Carson, J. R., "Frequency Modulation: Theory of the Feedback Receiver Circuit," *Bell System Technical Journal*, Vol. 18, No. 3, July 1939, pp. 395–403.

4. Ruthroff, C. L., "FM Demodulators with Negative Feedback," *Bell System Technical Journal*, Vol. 40, No. 4, July 1961, pp. 1149–1156.

5. Sydnor, R., "Investigation of an FM Demodulator with Negative Feedback," *Research Summary No. 36-9*, Vol. I, Jet Propulsion Laboratory, Pasadena, Calif., July 1961.

6. Enloe, L. H., "Decreasing the Threshold in FM by Frequency Feedback," *Proceedings of the IRE*, January 1962, pp. 18–30.

7. Cahn, C. R., "Optimum Performance of Phase-Lock and Frequency-Feedback Demodulators for FM," Magnavox Corporation, Torrance, Calif. To be published.

8. Develet, J. A., "A Threshold Criterion for Phase-Lock Demodulation," *Proceedings of the IRE*, February 1963, pp. 349–356.

9. Develet, J. A., *Statistical Design and Performance of High-Sensitivity Frequency-Feedback Receivers*, Aerospace Corporation, El Segundo, Calif., February 1963.

REFERENCES (Cont'd)

10. Gruen, W. J., "Theory of AFC Synchronization," *Proceedings of the IRE*, August 1953, pp. 1043–1048.

11. Richman, D., "Color Carrier Reference Phase Synchronization Accuracy in NTSC Color TV," *Proceedings of the IRE*, January 1954, pp. 106–133.

12. Jaffe, R., and Rechtin, E., "Design and Performance of Phase-Lock Circuits Capable of Near-Optimum Performance Over a Wide Range of Input Signal and Noise Levels," *IRE Transactions on Information Theory*, Vol. IT-1, March 1955, pp. 66–72.

13. Viterbi, A. J., "Acquisition and Tracking Behavior of Phase-Locked Loops," *Proceedings of the Symposium on Active Networks and Feedback Systems*, Polytechnic Institute of Brooklyn, April 1960, pp. 583–620.

14. Margolis, S. G., "The Response of a Phase-Locked Loop to a Sinusoid Plus Noise," *IRE Transactions on Information Theory*, Vol. IT-3, June 1957, pp. 136–143.

15. Gilchriest, C. E., "Application of the Phase-Locked Loop to Telemetry as a Discriminator or Tracking Filter," *IRE Transactions on Telemetry and Remote Control*, Vol. TRC-4, June 1958, pp. 20–35.

16. Viterbi, A. J., "System Design Criteria for Space Television," *Journal of the British Institution of Radio Engineers*, No. 9, September 1959, pp. 561–570.

17. Viterbi, A. J., "Optimum Coherent Demodulation for Continuous Modulation Systems," *Proceedings of the National Electronics Conference*, Vol. 18, 1962, pp. 498–508.

18. Choate, R. L., *Design Techniques for Low-Power Telemetry*, Technical Report No. 32-153, Jet Propulsion Laboratory, Pasadena, Calif., March 1962.

19. Viterbi, A. J., *Phase-Locked Loop Dynamics in the Presence of Noise by Fokker-Planck Techniques*, Technical Report No. 32-427, Jet Propulsion Laboratory, Pasadena, Calif., March 1963.

20. Booton, R. C., Jr., "The Analysis of Nonlinear Control Systems with Random Inputs," *Proceedings on Nonlinear Circuit Analysis*, Polytechnic Institute of Brooklyn, New York, April 1953, Vol. II, pp. 369–391.

21. Davenport, W. B., Jr., "Signal-to-Noise Ratios in Band-Pass Limiters," *Journal of Applied Physics*, Vol. 24, No. 6, June 1953, pp. 720–727.

22. Davenport, W. B., Jr., and Root, W., *Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Company, Inc., New York, 1958.

23. Wiener, N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, The M.I.T. Press, Cambridge, Mass., 1960.

24. Lehan, F. W., and Parks, R. J., "Optimum Demodulation," *IRE Convention Record*, Part 8, PGIT, March 1953, pp. 101–103.

25. Campbell, D. P., and Brown, G. S., *Principles of Servomechanisms*, John Wiley and Sons, Inc., New York, 1948.

26. Martin, B. D., *A Coherent Minimum-Power Lunar Probe Telemetry System*, External Publication No. 610, Revised, Jet Propulsion Laboratory, Pasadena, Calif., August 1959, pp. 1–72.

REFERENCES (Cont'd)

27. Truxal, J. G., *Control System Synthesis*, McGraw-Hill Book Company, Inc., New York, 1955.

28. Helstrom, C. W., *Statistical Theory of Signal Detection*, Pergamon Press, New York, 1960.

29. Lindsey, W. C., "Optimum Coherent Amplitude Demodulation," *Proceedings of the National Electronics Conference*, Vol. 20, 1964, pp. 497–503.

ACKNOWLEDGMENTS