Binary Data Transmission by FM over a Real Channel

By W. R. BENNETT and J. SALZ

(Manuscript received April 26, 1963)

Formulas are derived for probability of error in the detection of binary FM signals received from a channel characterized by arbitrary amplitude- and phase-vs-frequency distortion as well as additive Gaussian noise. The results depend on the signal sequence and can be presented in terms of averages over all signal sequences or as bounds for the most and least vulnerable ones. Illustrative examples evaluated include Sunde's method of suppressing intersymbol interference in band-limited FM. The effects of various representative channel filters are also analyzed. A solution is given for the problem of optimizing the receiving bandpass filter to minimize error probability at constant transmitted signal power. It is found that a performance from 3 to 4 db poorer than that theoretically attainable from binary PM is realizable over a variety of filtering situations.

I. INTRODUCTION

This paper undertakes to refine and extend the state of knowledge concerning performance of FM systems for binary data transmission over real-life channels. The particular aim is application to facilities such as exist in the telephone plant. Efficient use of the available channels constrains the bandwidth allowed for a given signaling speed. The luxury of a bandwidth sufficient to permit frequency transitions without amplitude variations and without dependence of present waveform on past signal history would in general imply an unjustifiably low information rate for the frequency range occupied. We therefore concentrate our attention on the band-limited channel with its inherent distortion of the FM data wave.

We assume a linear time-invariant transmission medium specified by its amplitude- and phase-vs-frequency functions and the statistics of its additive noise sources. The limiting noise environment in the telephone plant is typically nongaussian and not well defined even in a statistical sense. Nevertheless, with the usual apology, we shall perform our analysis in terms of additive Gaussian noise. Justification of the relevancy is based on the following considerations:

(a) Laboratory tests on data transmission systems are made at present by adding Gaussian noise and counting errors. Good performance in terms of low error rate as a function of signal-to-noise ratio under such test conditions is found to be indicative of good performance on actual channels.

(b) Identification and removal of nongaussian disturbances is a feasible and continuing process which should eventually lead to a more nearly Gaussian description of the residue.

Our measure of performance is expressed in terms of error probability vs the ratio of average transmitted signal power to average Gaussian noise power. In most of the work we assume white Gaussian noise is added at the receiver input. A convenient reference is then the average noise power in a band of frequencies having width equal to the transmitted information rate in bits per second.

II. STATEMENT OF PROBLEM

A block diagram of the transmission system under study is shown in Fig. 1. The data source emits a sequence of binary symbols which for full information rate are independent of each other and have equal probability. The analysis can be generalized without analytical inconvenience to assign a probability m_1 to one of the two binary symbols and $1 - m_1$ to the other. In conventional binary notation the symbols are 1 and 0. It is convenient to express binary frequency modulation of

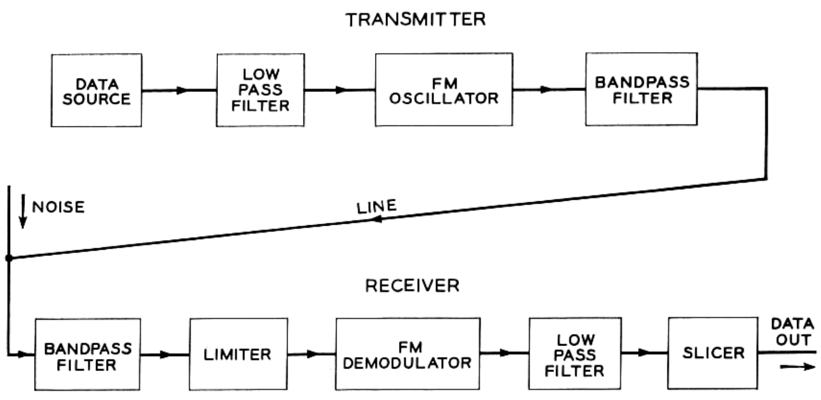

Fig. 1 — Binary FM transmission system.

an oscillator in terms of positive and negative frequency deviations. The combination of data source and low-pass filter is accordingly defined by the shaped baseband data wave train

$$s(t) = \sum_{n=-\infty}^{\infty} b_n g(t - nT) \tag{1}$$

where

$$b_n = 2a_n - 1. \tag{2}$$

The values of a_n represent the data sequence in binary notation. The probability is m_1 that the typical a_n is unity, and $1 - m_1$ that it is zero. The value of b_n is $+1$ if a_n is unity, and -1 if a_n is zero. The function $g(t)$ represents a standard pulse emitted by the low-pass filter for a signal element centered at $t = 0$.

Ideally, the oscillator frequency follows the baseband signal wave $s(t)$. This would imply an output voltage from the FM oscillator specified by

$$V(t) = A \cos \left[\omega_c t + \theta_0 + \mu \int_{t_0}^{t} s(\lambda) \, d\lambda \right]. \tag{3}$$

Here, A is the carrier amplitude, ω_c is the frequency of the oscillator with no modulating signal applied, t_0 is an arbitrary reference time, θ_0 is the phase at $t = t_0$, and μ is a conversion factor relating frequency displacement to baseband signal voltage. The instantaneous frequency of the wave (3) is defined as the derivative of the argument of the cosine function. It can be written in the form $\omega_c + \omega_i$, where ω_i, the deviation from midband, is ideally expressed by

$$\omega_i = \mu s(t). \tag{4}$$

In the practical case, the transmitting bandpass filter restricts the frequency-modulated wave to the range of frequencies passed by the channel. The purpose of this filter is to prevent both waste of transmitted power in components which will not reach the receiver and contamination of the line at frequencies assigned to other channels. The result is a transformation of the voltage wave (3) to a band-limited form, which must depart in more or less degree from the ideal conditions of constant amplitude and linear relationship between frequency and baseband signal. The line also inserts variations in amplitude- and phase-vs-frequency which cause further departures from the ideal. For our purposes it is sufficient to combine the line characteristics with those of the transmitting filter into a single composite network function determining the wave presented to the receiving bandpass filter.

The receiving bandpass filter is necessary to exclude out-of-band noise and interference from the detector input. It also shapes the signal waveform and can include compensation for linear in-band distortion suffered in transmission. Two contradictory attributes are sought in the filter — a narrow band to reject noise and a wide band to supply a good signal wave to the detector. An opportunity for an optimum design thus exists and will be explored in this paper.

The frequency detector is assumed to differentiate the phase with respect to time. The post-detection filter can do further noise rejection and shaping in the baseband range, but its only function in our present analysis is to separate the wave representing the frequency variation from the higher-frequency detection products. The slicer delivers positive voltage when the detected frequency is above midband and negative voltage when the detected frequency is below midband. The slicer output is sampled at appropriate instants to recover the binary data sequence.

The noise-free input to the detector will be written in the form

$$V_r(t) = P(t) \cos (\omega_c t + \theta) - Q(t) \sin (\omega_c t + \theta). \tag{5}$$

$P(t)$ and $Q(t)$ represent in-phase and quadrature signal modulation components respectively, which are associated with a carrier wave at the midband frequency ω_c with specified phase θ. Such a resolution can always be made, even though the details in actual examples may be burdensome. The added noise wave at the detector input is assumed to be Gaussian with zero mean and can likewise be written as

$$v(t) = x(t) \cos (\omega_c t + \theta) - y(t) \sin (\omega_c t + \theta). \tag{6}$$

If $v(t)$ represents Gaussian noise band-limited to $\pm 2\omega_c$, $x(t)$ and $y(t)$ are also Gaussian and are band-limited to $\pm\omega_c$. If the spectral density of $v(t)$ is $w_v(\omega)$, the spectral densities of $x(t)$ and $y(t)$ are given by[1]

$$w_x(\omega) = w_y(\omega) = w_v(\omega_c + \omega) + w_v(\omega_c - \omega), \qquad |\omega| < \omega_c \tag{7}$$

In general, $x(t)$ and $y(t)$ are dependent, with cross-spectral density

$$w_{xy}(\omega) = j[w_v(\omega_c - \omega) - w_v(\omega_c + \omega)] \tag{8}$$

and cross-correlation function expressed in terms of $R_v(\tau)$, the autocorrelation function of $v(t)$, by

$$R_{xy}(\tau) = -2R_v(\tau) \sin \omega_c \tau. \tag{9}$$

The cross correlation vanishes at $\tau = 0$, and hence the joint distribution of $x(t)$, $y(t)$ at any specified t is that of two independent Gaussian variables.

We shall also require the joint distribution of x and y with their time derivatives $\dot{x}$ and $\dot{y}$. The latter are Gaussian with spectral densities

$$w_{\dot{x}}(\omega) = w_{\dot{y}}(\omega) = \omega^2 w_x(\omega). \tag{10}$$

The cross-spectral densities are

$$w_{x\dot{x}}(\omega) = w_{y\dot{y}}(\omega) = j\omega w_x(\omega) \tag{11}$$

$$w_{x\dot{y}}(\omega) = j\omega_{xy}(\omega) = \omega[w_v(\omega_c + \omega) - w_v(\omega_c - \omega)] = -w_{\dot{x}y} . \tag{12}$$

The cross correlations are

$$\begin{aligned} R_{x\dot{x}}(\tau) &= \int_{-\infty}^{\infty} w_{x\dot{x}}(\omega)e^{j\tau\omega}\, d\omega = -\int_{-\infty}^{\infty} \omega w_x(\omega) \sin \tau\omega\, d\omega \\ &= R_{y\dot{y}}(\tau) \end{aligned} \tag{13}$$

$$\begin{aligned} R_{x\dot{y}}(\tau) &= -R_{\dot{x}y}(\tau) = \int_{-\infty}^{\infty} w_{x\dot{y}}(\omega)e^{j\tau\omega}\, d\omega \\ &= \int^{\infty} \omega[w_v(\omega_c + \omega) - w_v(\omega_c - \omega)] \cos \tau\omega\, d\omega. \end{aligned} \tag{14}$$

The cross correlation of x and $\dot{x}$ as well as of y and $\dot{y}$ vanish at $\tau = 0$, and hence at any instant $\dot{x}$ is independent of x, and $\dot{y}$ is independent of y. The cross correlations of x and $\dot{y}$, and of $\dot{x}$ and y, do not vanish in general, but do vanish in the special case in which

$$w_v(\omega_c + \omega) = w_v(\omega_c - \omega). \tag{15}$$

This is the case of a noise spectrum which is symmetrical with respect to the midband and represents a reasonable objective in system design. Since the simplification in computational details is quite considerable when the condition of symmetry is imposed, and since the departures caused by lack of symmetry are not of primary interest, we shall assume henceforth that (15) is satisfied. The four variables x, $\dot{x}$, y, and $\dot{y}$ are then independent and have the joint Gaussian probability density function

$$p(x, y, \dot{x}, \dot{y}) = \frac{1}{4\pi^2\sigma_0^2\sigma_1^2} \exp\left[-\frac{x^2 + y^2}{2\sigma_0^2} - \frac{\dot{x}^2 + \dot{y}^2}{2\sigma_1^2}\right] \tag{16}$$

$$\sigma_0^2 = \int_{-\infty}^{\infty} w_x(\omega)\, d\omega = 2\int_{-\infty}^{\infty} w_v(\omega_c + \omega)\, d\omega \tag{17}$$

$$\sigma_1^2 = \int_{-\infty}^{\infty} w_{\dot{x}}(\omega)\, d\omega = 2\int_{-\infty}^{\infty} \omega^2 w_v(\omega_c + \omega)\, d\omega. \tag{18}$$

The noise-free detector input wave (5) can be written in the equivalent form

$$V_r(t) = R(t) \cos\left[\omega_c t + \phi(t)\right] \tag{19}$$

where

$$R^2(t) = P^2(t) + Q^2(t) \tag{20}$$

$$\tan \phi(t) = Q(t)/P(t). \tag{21}$$

The frequency detector and post-detection filter combine to deliver a wave proportional to the instantaneous frequency deviation from mid-band. Taking the constant of proportionality as unity, we write for the output wave

$$\phi'(t) = \frac{d}{dt} \arctan \frac{Q(t)}{P(t)} = \frac{P(t)Q'(t) - Q(t)P'(t)}{P^2(t) + Q^2(t)}. \tag{22}$$

With the functional dependence on t understood, we write this equation in the form

$$\phi'(t) = \dot{\phi} = (P\dot{Q} - Q\dot{P})/R^2. \tag{23}$$

When the noise is added, the detected frequency is changed to

$$\psi'(t) = \dot{\psi} = \frac{(P + x)(\dot{Q} + \dot{y}) - (Q + y)(\dot{P} + \dot{x})}{(P + x)^2 + (Q + y)^2}. \tag{24}$$

Assuming that the system does not make errors in the absence of noise, we can express the probability of error in a given sample of instantaneous frequency taken at the time $t = nT$ as the probability that $\psi'(nT)$ is negative if $\phi'(nT)$ is positive or the probability that $\psi'(nT)$ is positive if $\phi'(nT)$ is negative. Since the system has memory, the values of P, Q, $\dot{P}$, and $\dot{Q}$ at any sampling instant depend on the entire signal sequence. Our procedure is first to show how the error probability can be evaluated at any sampling instant for any sequence. We then calculate error rates for specific sequences and establish bounds for most and least vulnerable sequences.

Since the denominators of (23) and (24) are inherently positive, the decisions are made entirely on the basis of the signs of the numerators. Therefore, we do not require the distribution function of the instantaneous frequency itself. In fact if we let

$$\begin{aligned} x + P &= x_1, & \dot{x} + \dot{P} &= \dot{x}_1 \\ y + Q &= y_1, & \dot{y} + \dot{Q} &= \dot{y}_1 \end{aligned} \tag{25}$$

we require only one value of the distribution function of the variable z defined by

$$z = x_1\dot{y}_1 - y_1\dot{x}_1 . \tag{26}$$

The error probability is fully determined in any specific case either by the probability that z is negative or by the probability that z is positive. That is, if $F(z)$ is the distribution function of z, we only require the value of $F(0)$.

We shall derive a general expression for $F(0)$ in terms of a single definite integral. From this integral we shall then obtain definite integrals representing bounds for the error probability when arbitrary binary data sequences are transmitted. No restrictions on range of signal-to-noise ratios are made. The results will be applied to special cases of practical interest. One is Sunde's binary FM system which avoids intersymbol interference in a finite band in the absence of noise. When noise is added in this system, the detected samples become dependent on past signal history. It has been found possible to give a complete treatment of the Sunde method, including optimization of the receiving filter for minimum probability of error with fixed average transmitted signal power. The other cases analyzed in detail are based on design parameters actually in use on FM data transmission terminals.

III. GENERAL SOLUTION

Our first observation is that when x_1 and y_1 are fixed, the variable z of (26) is defined by a linear operation on the two independent Gaussian variables $\dot{x}_1$ and $\dot{y}_1$. Hence the conditional probability density function $p(z \mid x_1, y_1)$ of z when x_1 and y_1 are given is Gaussian with readily determined parameters. We accordingly write

$$p(z \mid x_1, y_1) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(z - z_0)^2}{2\sigma^2}\right]. \tag{27}$$

The mean z_0 is the sum of the means of $x_1\dot{y}_1$ and $-y_1\dot{x}_1$, that is,

$$z_0 = x_1 \operatorname{av} \dot{y}_1 - y_1 \operatorname{av} \dot{x}_1 = x_1\dot{Q} - y_1\dot{P}. \tag{28}$$

The variance σ^2 is the sum of the variances of $x_1\dot{y}_1$ and $y_1\dot{x}_1$; hence

$$\sigma^2 = (x_1^2 + y_1^2)\sigma_1^2. \tag{29}$$

The complete probability density function $p(z)$ for z is obtained by averaging the conditional probability density function over x_1 and y_1 . This is done by multiplying (27) by the joint probability density function of x_1 and y_1 and then integrating over all x_1 and y_1 . Calling the latter function $q(x_1, y_1)$, we can express its value by substituting the values of x and y from (25) in (16) and integrating out the $\dot{x}$ and $\dot{y}$ terms. The result is

$$q(x_1, y_1) = \frac{1}{2\pi\sigma_0^2} \exp\left[-\frac{(x_1 - P)^2 + (y_1 - Q)^2}{2\sigma_0^2}\right]. \tag{30}$$

Then

$$p(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(z \mid x_1, y_1) q(x_1, y_1)\, dx_1\, dy_1. \tag{31}$$

The probability of error when the noise-free sample of frequency deviation is positive is

$$P_+ = \int_{-\infty}^{0} p(z)\, dz = \int_0^{\infty} p(-z)\, dz. \tag{32}$$

Likewise, when the noise-free sample is negative, we obtain a probability of error

$$P_- = \int_0^{\infty} p(z)\, dz. \tag{33}$$

The problem is thus reduced to the evaluation of the triple integral obtained by combining (27), (30), and (31) with either (32) or (33). It is shown in Appendix A that the result of these operations can be expressed in the following form

$$P_+ = \frac{1}{2}\operatorname{erfc}\frac{R}{\sqrt{2}\sigma_0} + \frac{R}{2\sigma_0\sqrt{2\pi}}\int_{-1}^{1} \exp\left(-\frac{R^2x^2}{2\sigma_0^2}\right) \operatorname{erfc}\frac{R\phi(1 - x^2)^{\frac{1}{2}} - \dot{R}x}{\sqrt{2}\sigma_1}\, dx. \tag{34}$$

The value of P_- is obtained by subtracting the right-hand member of (34) from unity. We note that ϕ is positive for P_+ and negative for P_-. The symbol $\dot{R}$ is used for dR/dt where R is given by (20). In a pure FM wave, $\dot{R} = 0$, but this condition cannot be maintained in a finite bandwidth.

Differentiating partially with respect to $\dot{R}$ and rearranging, we obtain

$$\frac{\partial P_+}{\partial \dot{R}} = \frac{R}{\pi\sigma_0\sigma_1}\int_0^1 x \exp\left[-\frac{R^2x^2}{2\sigma_0^2} - \frac{R^2\phi^2(1 - x^2) + \dot{R}^2x^2}{2\sigma_1^2}\right] \sinh\frac{R\dot{R}\phi x(1 - x^2)^{\frac{1}{2}}}{\sigma_1^2}\, dx. \tag{35}$$

We note that $\partial P_+/\partial \dot{R}$ vanishes when $\dot{R} = 0$ and at no other value of $\dot{R}$. The latter follows from the fact that the integrand of (35) cannot change sign in the interval of integration. We also find that $\partial^2 P_+/\partial \dot{R}^2$ is positive when $\dot{R} = 0$. We conclude that P_+ is minimum with respect to $\dot{R}$ when and only when $\dot{R} = 0$. A lower bound on the probability of error for any fixed R and ϕ is therefore obtained by setting $\dot{R} = 0$, giving

$$P_l = \frac{1}{2} \operatorname{erfc} \frac{R}{\sqrt{2}\sigma_0} + \frac{R}{2\sigma_0\sqrt{2\pi}} \int_{-1}^{1} \exp\left(-\frac{R^2x^2}{2\sigma_0^2}\right) \operatorname{erfc} \frac{R\phi(1-x^2)^{\frac{1}{2}}}{\sqrt{2}\sigma_1}\, dx. \tag{36}$$

Also, since P_+ must be monotonic increasing with $|\dot{R}|$, the largest probability of error for any fixed R and ϕ occurs when $\dot{R}$ has its largest possible absolute value. These deductions are of aid in selecting the data sequences which have most and least probabilities of error.

It is shown in Appendix A that P_l can be written in the equivalent form

$$P_l = \frac{1}{\pi} \int_0^{\pi/2} \exp\left[-\frac{R^2\phi^2/(2\sigma_1^2)}{1 + \left(\frac{\sigma_0^2\phi^2}{\sigma_1^2} - 1\right)\cos^2\theta}\right] d\theta. \tag{37}$$

It is also shown that when $\phi < (\sigma_1/\sigma_0)$, the limiting form for large signal-to-noise ratio — i.e., R large compared with σ_0 — is given by

$$P_l \sim \frac{\sigma_1}{R\phi\sqrt{2\pi}} \left(\frac{\sigma_0^2\phi^2}{\sigma_1^2} - 1\right)^{-\frac{1}{2}} \exp\left(-\frac{R^2}{2\sigma_0^2}\right). \tag{38}$$

When $\phi > (\sigma_1/\sigma_0)$, the limiting form becomes

$$P_l \sim \frac{\sigma_1}{R\phi\sqrt{2\pi}} \left(1 - \frac{\sigma_0^2\phi^2}{\sigma_1^2}\right)^{-\frac{1}{2}} \exp\left(-\frac{R^2\phi^2}{2\sigma_1^2}\right). \tag{39}$$

When $\phi = \sigma_1/\sigma_0$, we have the exact result

$$P_l = \frac{1}{2} \exp\left(-\frac{R^2}{2\sigma_0^2}\right). \tag{40}$$

The general equation for error probability (34) can conveniently be expressed in terms of the following three parameters

$$\rho^2 = \frac{R^2}{2\sigma_0^2} \tag{41}$$

$$a^2 = \frac{\sigma_0^2 \dot{\phi}^2}{\sigma_1^2} \tag{42}$$

$$b^2 = \frac{\dot{R}^2}{2\sigma_1^2}. \tag{43}$$

Equation (34) then becomes

$$P_+ = \frac{1}{2} \operatorname{erfc} \rho + \frac{\rho}{2\sqrt{\pi}} \int_{-1}^{1} e^{-\rho^2 x^2} \operatorname{erfc} [a\rho(1 - x^2)^{\frac{1}{2}} - bx]\, dx. \tag{44}$$

Evaluation of this equation in terms of the three parameters ρ, a, and b gives the error probability for any of the FM systems considered.

IV. ERROR PROBABILITY VS SIGNAL-TO-NOISE RATIO

In analog systems the performance is often expressed in terms of signal-to-noise ratio in the receiver output. In the case of audio and video signals, where subjective judgments determine the requirements, the signal-to-noise ratio furnishes a good criterion. In the case of data signals, however, performance is judged in terms of errors made, and the errors cannot be predicted from the signal-to-noise ratio alone. The error rate depends in general on the distribution of the noise values. Furthermore, in good systems the errors are rare and hence are associated with infrequent noise conditions. The central part of the noise distribution is of less importance than the tails.

We illustrate the difference between a straight signal-to-noise ratio analysis and a direct error probability calculation in FM by a simple example. Consider the case of a long sequence of mark signals leading to a constant signal frequency $\omega_c + \omega_d$. The signal wave can then be written in the form

$$\begin{aligned} V(t) &= A \cos (\omega_c + \omega_d)t \\ &= A \cos \omega_d t \cos \omega_c t - A \sin \omega_d t \sin \omega_c t. \end{aligned} \tag{45}$$

Comparing with (5) and noting that we are omitting the arbitrary phase angle θ, which is of trivial interest, we make the identifications

$$P(t) = A \cos \omega_d t \qquad Q(t) = A \sin \omega_d t. \tag{46}$$

Then, by differentiation

$$P'(t) = -\omega_d A \sin \omega_d t \qquad Q'(t) = \omega_d A \cos \omega_d t. \tag{47}$$

If a sample is taken at $t = 0$

$$P = A \qquad \dot{P} = 0 \qquad Q = 0 \qquad \dot{Q} = \omega_d A. \tag{48}$$

Then from (24) the error $\dot{\psi} - \omega_d$ in the detected frequency deviation because of additive Gaussian noise is

$$\nu = \dot{\psi} - \omega_d = \frac{(A + x)(\omega_d A + \dot{y}) - y\dot{x}}{(A + x)^2 + y^2} - \omega_d . \tag{49}$$

In a signal-to-noise ratio calculation for the case in which the signal amplitude is usually much larger than the noise on the line, (49) would be written in the form

$$\nu = \frac{\omega_d(1 + x/A) + \dot{y}/A + (x\dot{y} - y\dot{x})/A^2}{(1 + x/A)^2 + (y/A)^2} - \omega_d . \tag{50}$$

If we then assume that A is large compared with x, y, $\dot{x}$, and $\dot{y}$, we retain only first-order terms in small quantities and construct the following approximate result, valid most of the time

$$\begin{aligned} \nu &\approx \omega_d(1 + x/A) + \dot{y}/A - \omega_d(1 + 2x/A) \\ &= (\dot{y} - \omega_d x)/A. \end{aligned} \tag{51}$$

The approximate spectral density of the frequency deviation error is then

$$\begin{aligned} w_\nu(\omega) &\approx [w_{\dot{y}} + \omega_d^2 w_x(\omega)]/A^2 \\ &= 2(\omega^2 + \omega_d^2)w_v(\omega_c + \omega)/A^2. \end{aligned} \tag{52}$$

The approximate mean-square value of error can now be found by integrating the spectral density function $w_\nu(\omega)$ over all frequencies. However, we cannot obtain the probability of error from this value because we do not know the distribution function. A nonlinear operation has been performed on a Gaussian process, and the result must be non-gaussian. In this case Rice[2] has shown that the central part of the frequency error distribution is approximately Gaussian. His argument does not apply to the tail. When the signal exceeds the noise most of the time, it is only the tails of the distribution which are important in determining the probability that an error is made in distinguishing between mark and space frequencies.

Since there is no intersymbol interference in our example the exact expression for probability of error is given by (37) with $R = A$ and $\phi = \omega_d$. It can be seen from the limiting forms for large signal-to-noise ratio, (38) through (40), that the Gaussian approximation from (52) cannot approach the correct result. The result obtained from (52) must contain both the original and differentiated noise spectra in the argument of the exponential part of the approximation at large signal-to-noise ratios. In (38) and (39) the exponential depends on either σ_0 or σ_1 but not both.

As another example of the difference between inferences from signal-to-noise ratio and error probability, it is interesting to consider the case of differentially detected binary phase modulation. In this system the polarity of the present carrier wave is compared with the polarity one bit ago. The binary message is read as 1 for a phase reversal and 0 for no phase change. By intuitive reasoning one could easily conclude that there would be a 3-db penalty relative to synchronous detection with a noise-free period. Certainly, in the differential case noise is added to both waves under comparison, and the bit interval is usually long enough to make the two noise samples substantially independent of each other. Signal-to-noise ratio analysis supports the intuitive argument when the average noise power is small relative to the average signal power. A direct calculation of error probability, however, exposes the fallacy and reminds us sharply that the noise is not small compared with the signal when errors occur. If we focus attention on the large noise peaks which cause error, we can see that the simultaneous combination of disturbances on both waves does not imply the same probability of disaster as would follow from concentration of all the noise on one wave.

The differential binary PM problem can in fact be solved as a simple special case of the analysis we have developed for FM. The input wave to the detector can be written as

$$V_r(t) = [P(t) + x(t)] \cos \omega_c t - y(t) \sin \omega_c t. \tag{53}$$

The detector operates by multiplying $V_r(t)$ and $V_r(t - T)$, selecting the low-frequency components of the product, and sampling the output at intervals T apart. If we assume $\omega_c T$ is a multiple of 2π and identify quantities evaluated at $t - T$ by the subscript d, the binary decisions are based on the sign of the wave

$$V_a(t) = (P + x)(P_d + x_d) + y y_d\,. \tag{54}$$

When the correct binary decision is 0, the signs of P and P_d are the same, and an error occurs if the sampled value V_a is negative. When the correct binary decision is 1, the signs of P and P_d are opposite, and an error occurs if the sampled value of V_a is positive. The two cases are symmetric and an analysis of either suffices. For the case of the symbol 0, $P = P_d$, while for the case of 1, $P = -P_d$.

In calculating the signal-to-noise ratio for the case of a symbol 0, we would write

$$V_a = P\left(P + x + x_d + \frac{x x_d + y y_d}{P}\right). \tag{55}$$

Then if P is large compared with x, x_d, y, and y_d, we approach a condition in which the decisions are based on the sign of $P + x + x_d$. If x and x_d are independent, the sum $x + x_d$ represents samples from random noise with twice as much average power as the samples of either x or x_d alone. This tempting argument leads to the 3-db rule.

In a direct calculation of error probability, we recognize that the influence of xx_d and yy_d cannot be ignored at the tails of the noise distribution where the errors occur. In particular, if x and x_d are both very negative, tending to cause an error in a symbol 0, the value of xx_d is large and positive, tending to prevent the threatened damage.

To find the error probability, we compare (54) with (26), and note that we have a special case of the previous solution if we make the following identification

$$z \equiv V_a \qquad x_1 \equiv P + x \qquad \dot{y}_1 \equiv P_d + x_d$$
$$y_1 \equiv y \qquad \dot{x}_1 \equiv -y_d . \tag{56}$$

The remainder of the solution proceeds as before if x, y, x_d, and y_d are independent Gaussian variables. The independence is guaranteed if the second-order correlation functions vanish at lag time T. One difference between this case and the earlier one is that the variables x, y, x_d, and y_d all have the same variance. This specialization can be made in the earlier work by setting $\sigma_0 = \sigma_1 = \sigma$. By comparing with (25), we further note that we can now set $Q = \dot{P} = 0$, $\dot{Q} = P_d = P$. Hence we also have $R = P$ and $\dot{R} = 0$. Corresponding to ϕ we insert the value which V_a/R^2 assumes in the absence of noise, namely $\phi \equiv P^2/P^2 = 1$. In terms of (41), (42), and (43) we then have

$$\rho^2 = \frac{P^2}{2\sigma^2} \qquad a^2 = 1 \qquad b^2 = 0. \tag{57}$$

Hence the answer is given by (40), namely

$$P_+ = P_- = \tfrac{1}{2}e^{-\rho^2}. \tag{58}$$

In the ideal case, a bandwidth f_0 is sufficient to send signals by binary PM at a rate f_0 bits per second without intersymbol interference. This allows for upper and lower sidebands with widths $f_0/2$. If the spectral density of the noise is ν_0 watts/cps, it follows that $\sigma^2 = \nu_0 f_0$. Then M, the ratio of average signal power to the average noise power in a band of width equal to the bit rate, is equal to the ratio of $P^2/2$ to $\nu_0 f_0$ and hence $M = \rho^2$. The formula for error probability is thus found to agree with the one given by Lawton.[3] Average signal power 0.9 db greater than the coherent case is required for an error probability of 10^{-4}. The difference in performance between the differential and purely coherent cases approaches zero at very high signal-to-noise ratios.

V. SUNDE'S BAND-LIMITED FM SYSTEM WITHOUT INTERSYMBOL INTERFERENCE

E. D. Sunde[4] has described a binary FM system in which the intersymbol interference in the absence of noise can be made to vanish at the sampling instants, even when the bandwidth is limited to an extent comparable with that used in AM transmission. The method is remarkable in that a type of result similar to that given by Nyquist[5] for AM systems is obtained for all sequences in spite of the nonlinear FM detection process which invalidates the principle of superposition. The performance falls a little short of the corresponding AM case, in that some dependence on the message appears when noise is added.

Fig. 2 shows a diagram of Sunde's method. The binary message is sent by switching between two oscillators. The difference between the oscillator frequencies must be locked to the bit rate, and the oscillators must be so phased that the frequency transitions are accomplished with continuous phase. The combination of sending filter, line, and receiving filter modify the switched output to produce a spectrum at the input to the frequency detector with even symmetry about the midband and with Nyquist's vestigial symmetry about the marking and spacing fre-

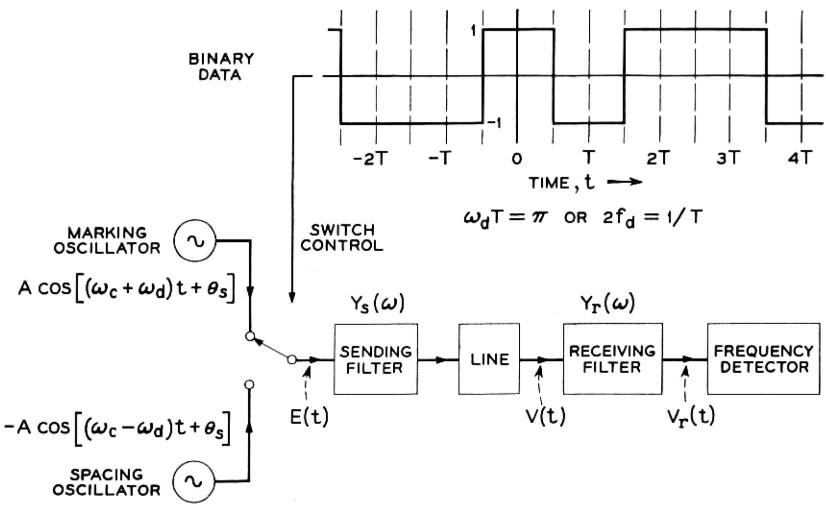

Fig. 2 — Sunde's band-limited binary FM system.

quencies. The latter must be high enough relative to the bit rate to prevent appreciable lower sideband foldover.

The output of the switch is represented by

$$E(t) = \frac{A}{2}[1 - s(t)] \cos[(\omega_c - \omega_d)t + \theta_s] + \frac{A}{2}[1 + s(t)] \cos[(\omega_c + \omega_d)t + \theta_m]. \tag{59}$$

In (59) A represents the amplitude of the output and must be the same for each oscillator. The switching function $s(t)$ represents the baseband data wave of (1). When $s(t) = -1$, the first term has amplitude A and the second term vanishes. When $s(t) = +1$, the first term vanishes and the second has amplitude A. The center of the band is the frequency ω_c and the total frequency shift is $2\omega_d$. For minimum bandwidth the angular signaling frequency $\omega_0 = 2\pi/T$ must be equal to $2\omega_d$. One of the two phase angles θ_s and θ_m can be arbitrary, but the two angles must differ by 180 degrees. Under these restrictions, the value of $E(t)$ can be written as

$$E(t) = A \sin \omega_d t \sin(\omega_c t + \theta_s) - As(t) \cos \omega_d t \cos(\omega_c t + \theta_s). \tag{60}$$

Sunde requires that the input wave to the frequency detector can be written in the form

$$V_r(t) = A \sin \omega_d t \sin(\omega_c t + \theta_r) - As_1(t) \cos(\omega_c t + \theta_r) \tag{61}$$

where $s_1(t)$ represents the data sequence with $g(t)$ replaced by $g_1(t)$. The latter must be a pulse which gives no intersymbol interference when the data rate is $1/T$. That is,

$$s_1(t) = \sum_{n=-\infty}^{\infty} (-)^n b_n g_1(t - nT) \tag{62}$$

and $g_1(t)$ assumes the value unity at $t = 0$ and has nulls at all instants differing from $t = 0$ by multiples of T. In mathematical notation

$$g_1[(m - n)T] = \delta_{mn} \tag{63}$$

and

$$s_1(mT) = (-)^m b_m. \tag{64}$$

The requirement as actually stated by Sunde differs from (61) in that his analysis is based on a switching function which assumes the values 1 and 0 at the sampling instant rather than 1 and -1. The two expressions for the requirement can be shown to be equivalent. Equation

(61) has the advantage that the function $s_1(t)$ has the average value zero for a random data sequence with equal probability of the two binary symbols. This fact enables an easy separation of the spectral density of $V_r(t)$ into line spectra contributed by the first term of (61) and a continuous spectral density function for the second part.

Incidentally, it is clear from (61) that all the signal information is contained in the second term, and that the first term can be regarded as a pair of pilot tones at the marking and spacing frequencies $\omega_c \pm \omega_d$. The sole function of these pilot tones is to enable an FM detector to recover the message. The information carrying part of $V_r(t)$ can equally well be regarded as double-sideband suppressed-carrier binary AM or binary phase modulation, with the carrier frequency placed at ω_c. The ideal way of detecting such signals is by multiplication with a coherent carrier wave, which must be transmitted as part of the data wave in some way. Detection of $V_r(t)$ as FM has a practical advantage in that there is no carrier recovery problem; the wave is ready for the frequency detector with no further processing. The penalty for transmitting pure sine waves is a waste of signal power. As will be shown quantitatively later, such waste results in an unfavorable comparison with more nearly ideal systems.

To show that the stipulated conditions are sufficient to suppress intersymbol interference in the detected frequency of $V_r(t)$, we identify $P(t)$ and $Q(t)$ of (5) with the applicable terms of (61) as follows

$$P(t) = -As_1(t) \tag{65}$$

$$Q(t) = -A \sin \omega_d t. \tag{66}$$

We then calculate

$$P'(t) = -As_1'(t) \tag{67}$$

$$Q'(t) = -\omega_d A \cos \omega_d t. \tag{68}$$

If we take frequency samples at $t = mT$ we find that since $\omega_d T = \pi$

$$\begin{aligned} P(mT) &= (-)^{m+1} b_m \\ P'(mT) &= -As_1'(mT) \\ Q(mT) &= 0 \\ Q'(mT) &= (-)^{m+1} \omega_d A. \end{aligned} \tag{69}$$

Hence in (23), evaluated at $t = mT$

$$\dot{\phi} = \dot{Q}/P = \omega_d/b_m = b_m \omega_d. \tag{70}$$

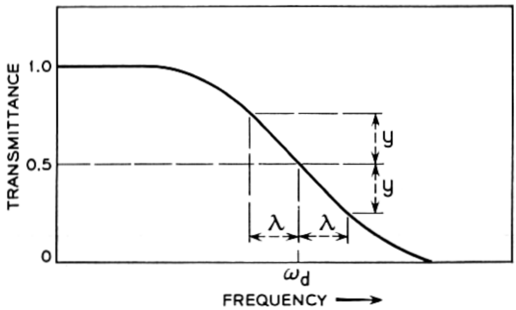

Fig. 3 — Nyquist's condition of vestigial symmetry.

The value of the instantaneous frequency deviation at the mth sampling point is, therefore, equal to ω_d if $s(mT) = 1$ and equal to $-\omega_d$ if $s(mT) = -1$. Freedom from intersymbol interference is thus obtained if (64) is satisfied.

As shown by Nyquist, a sufficient condition for obtaining (64) is that the standard pulse $g_1(t)$ is the impulse response of a network with transmittance $G_1(\omega)$ of the form shown in Fig. 3, described mathematically by

$$G_1(\pm\omega_d - \lambda) + G_1(\pm\omega_d + \lambda) = 2G_1(\omega_d) = T \qquad 0 < \lambda < \omega_d . \quad (71)$$

We say that a function satisfying (71) has vestigial symmetry about frequency ω_d because it has the type of symmetry called for in a vestigial sideband filter with the carrier at ω_d. We can think of the response at a frequency exceeding ω_d by an amount λ as exactly compensating the deficiency in the response at the frequency less than ω_d by the same amount λ. The ideal low-pass filter is a limiting special case occurring when the transmittance vanishes for $|\omega| > \omega_d$. The amplitude can be associated with linear phase shift, which changes only the origin of time. Unnecessary complication is avoided by carrying through the calculations with zero phase shift.

The conditions imposed on the filters and line to transform (60) to (61) can be expressed in terms of the Fourier transforms of $g(t) \cos \omega_d t$ and $g_1(t)$, which we represent respectively by $C(\omega)$ and $G_1(\omega)$. Both $C(\omega)$ and $G_1(\omega)$ are purely real and are given by

$$\begin{aligned} C(\omega) &= \int_{-\infty}^{\infty} g(t) \cos \omega_d t \cos \omega t \, dt \\ &= [G(\omega - \omega_d) + G(\omega + \omega_d)]/2 \end{aligned} \quad (72)$$

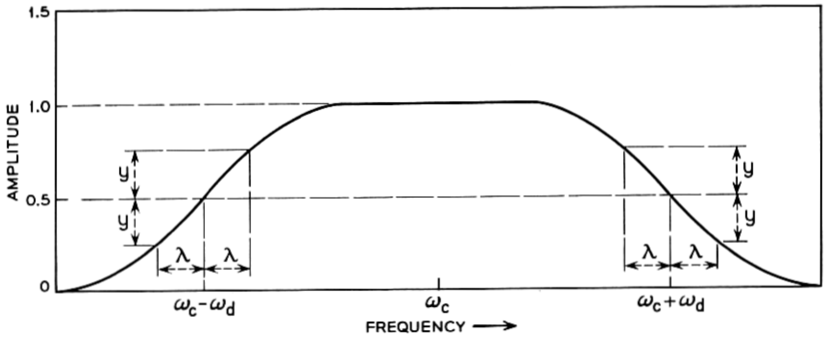

Fig. 4 — Spectrum at input to detector in Sunde's FM system.

$$G_1(\omega) = \int_{-\infty}^{\infty} g_1(t) \cos \omega t \, dt. \tag{73}$$

The result, obtained by multiplying $\cos(\omega_c t + \theta_s)$ by $g(t) \cos \omega_d t$ or $g_1(t)$, is to place upper and lower sidebands on the frequencies $\pm\omega_c$, as shown in Fig. 4, with spectra equal to $C(\omega - \omega_c)/2$ and $G_1(\omega - \omega_c)/2$ respectively on ω_c. The required transmittance function for the combination of sending filter, line, and receiving filter is then

$$Y(\omega) = \frac{G_1(\omega - \omega_c)}{C(\omega - \omega_c)}. \tag{74}$$

This function transforms the second term of (60) to the second term of (61). It is also necessary for the first term of (60) to remain unchanged. The first term can be written as the difference of sine waves of frequencies $\omega_c - \omega_d$ and $\omega_c + \omega_d$. These components will be unchanged by the operation $Y(\omega)$ if

$$C(\pm\omega_d) = G_1(\pm\omega_d) \quad \text{or} \quad Y(\omega_c \pm \omega_d) = 1. \tag{75}$$

It can readily be seen that the condition (71) required on $G_1(\omega)$ translates to the same condition for $G_1(u)$ where $u = \omega - \omega_c$.

The relations can be made clearer by working out an example. Suppose the switching is rectangular and there is no lost time between contacts. The function $g(t)$ is then defined by

$$g(t) = \begin{cases} 1, & -T/2 < t < T/2 \\ 0, & |t| > T/2. \end{cases} \tag{76}$$

Let the received signal $V_r(t)$ have a full raised cosine spectrum centered at ω_c, with vestigial symmetry about $\omega_c + \omega_d$ and $\omega_c - \omega_d$. We then write

$$G_1(u) = \begin{cases} T\left(1 + \cos \dfrac{\pi u}{2\omega_d}\right) \Big/ 2 & |u| \leqq 2\omega_d \\ 0 & |u| > 2\omega_d . \end{cases} \tag{77}$$

We calculate

$$C(\omega) = 2 \int_0^{T/2} \cos \omega_d t \cos \omega t \, dt = \frac{2\omega_d \cos (\omega T/2)}{\omega_d^2 - \omega^2} \tag{78}$$

$$Y(\omega) = \frac{\pi(\omega_d^2 - u^2)\left(1 + \cos \dfrac{\pi u}{2\omega_d}\right)}{4\omega_d^2 \cos \dfrac{\pi u}{2\omega_d}} \qquad u = \omega_c - \omega. \tag{79}$$

This function satisfies the required condition that $Y(\omega_c \pm \omega_d) = 1$.

In practice it is difficult to control two oscillators with the necessary precision to meet Sunde's requirements. One method of realizing the system approximately is to begin with two high-frequency crystal-controlled oscillators of frequencies $n(\omega_c - \omega_d)$ and $n(\omega_c + \omega_d)$, where n is a large integer. The phases of the two oscillators are not under control and are assumed to be θ_1 and θ_2, respectively. Frequency step-down circuits are introduced after each oscillator to give outputs of frequency $\omega_c - \omega_d$ and $\omega_c + \omega_d$ with respective phases θ_1/n and θ_2/n. By multiplying these two outputs and selecting the low-frequency component as shown in Fig. 5, we obtain a wave of frequency $2\omega_d$ and phase $(\theta_2 - \theta_1)/n$. This wave can be used to control the timing of the binary input symbols. For the switched marking and spacing frequency sources we use the stepped-down component of frequency $\omega_c - \omega_d$ directly and the component of frequency $\omega_c + \omega_d$ with reversed polarity. The required frequency and phase relations are then satisfied

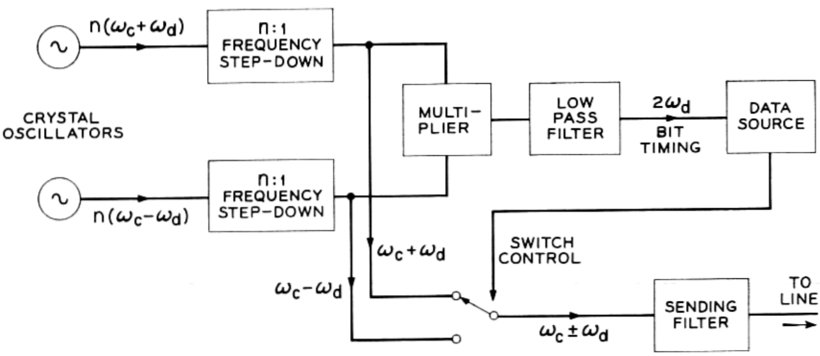

Fig. 5 — Practical realization of Sunde's system.

except for a slow drift in the time scale caused by the lack of perfect stability in the original oscillators.

To calculate the probability of error when Gaussian noise is added to Sunde's FM signal, we identify the values of $P(mT)$, $P'(mT)$, $Q(mT)$, and $Q'(mT)$ of (69) with P, $\dot{P}$, Q, and $\dot{Q}$ respectively. The general expression for the probability of error, (34), is expressed in terms of R and $\dot{R}$. We calculate

$$R = (P^2 + Q^2)^{\frac{1}{2}} = A \tag{80}$$

$$\begin{aligned} R'(t) &= \frac{d}{dt}\,[P^2(t) + Q^2(t)]^{\frac{1}{2}} \\ &= [P(t)P'(t) + Q(t)Q'(t)]/R(t) \end{aligned} \tag{81}$$

$$\dot{R} = R'(mT) = (P\dot{P} + Q\dot{Q})/R = (-)^{m+1}Ab_m s_1'(mT). \tag{82}$$

From (62)

$$s_1'(mT) = \sum_{n=-\infty}^{\infty} (-)^n b_n g_1'[(m-n)T]. \tag{83}$$

From (73) we verify

$$\begin{aligned} g_1(rT) &= \frac{1}{\pi}\int_0^{2\omega_d} G_1(\omega)\cos(\omega rT)\,d\omega \\ &= \frac{1}{\pi}\int_0^{\omega_d} G_1(\omega_d - \omega)\cos[rT(\omega_d - \omega)]\,d\omega \\ &\qquad + \frac{1}{\pi}\int_0^{\omega_d} G_1(\omega_d + \omega)\cos[rT(\omega_d + \omega)]\,d\omega \\ &= \frac{2}{\pi}\,G_1(\omega_d)\cos r\pi \int_0^{\omega_d} \cos r\omega T\,d\omega = \delta_{r0}\,. \end{aligned} \tag{84}$$

This checks our previous requirements expressed by (63) and (64). By differentiating (73) and substituting $t = rT$, we find

$$g_1'(rT) = -\frac{1}{\pi}\int_0^{2\omega_d} \omega G_1(\omega)\sin\omega rT\,d\omega. \tag{85}$$

The value of this integral in general is not zero except when $r = 0$. It appears, therefore, that at any sampling instant $t = mT$ the value of $\dot{R}$ depends on all the values of b_n in the sequence except b_m.

For further progress we take a specific example, namely the full raised cosine spectrum for $G_1(\omega)$. We set

$$G_1(\omega) = T\left(1 + \cos\frac{\pi\omega}{2\omega_d}\right)\Big/2 \qquad |\omega| \leqq 2\omega_d. \tag{86}$$

Then

$$\begin{aligned} g_1'(rT) &= -\frac{T}{2\pi}\int_0^{2\omega_d} \omega\left(1 + \cos\frac{\pi\omega}{2\omega_d}\right)\sin\omega rT\, d\omega \\ &= \frac{f_0}{r(1-4r^2)} \qquad r \neq 0. \end{aligned} \tag{87}$$

From (85), we noted that $g_1'(0) = 0$. The value of $\dot{R}$ can now be found from (82), thus

$$\begin{aligned} \dot{R} &= (-)^{m+1}\, b_m f_0 A\left[\sum_{n=-\infty}^{m-1} + \sum_{n=m+1}^{\infty}\right]\frac{(-)^n b_n}{(m-n)[1-4(m-n)^2]} \\ &= -b_m f_0 A \sum_{n=1}^{\infty} (-)^n \frac{b_{m+n} - b_{m-n}}{n(4n^2-1)}. \end{aligned} \tag{88}$$

We observe from our previous study of the integral defining the probability of error that for fixed R the most vulnerable sequence is the one which has the largest absolute value of $\dot{R}$. The least vulnerable sequence is the one for which $\dot{R} = 0$, and this can be obtained by setting $b_{m+n} = b_{m-n}$ for all n. The maximum absolute value of $\dot{R}$ occurs when b_{m+n} and b_{m-n} have opposite signs and the signs are reversed when n changes by unity. The resulting value of $|\dot{R}|$ is[6]

$$\begin{aligned} \dot{R}_m &= 2f_0 A \sum_{n=1}^{\infty} \frac{1}{n(4n^2-1)} \\ &= 2f_0 A\,(\log_e 4 - 1) = 0.7726\, f_0 A. \end{aligned} \tag{89}$$

The upper and lower bounds for the error probability are found by substituting $\dot{R}_m$ and 0 respectively for $\dot{R}$ in (34). By (80) the value of R is constant and equal to A. From (70), $\phi = b_m\omega_d$. It is important to note that while the intersymbol interference is suppressed in the absence of noise the error probability with noise present does depend on the signal sequence. This occurs because frequency detection is a nonlinear process, and the effect of noise cannot be found by merely adding a noise wave to the detected frequency output.

The actual spectral density of the noise facing the frequency detector is under the control of the system designer, since the selectivity of the receiving bandpass filter is not determined by the requirements thus far discussed. We have stated what the received signal spectrum at the detector input should be, but this is a resultant of signal shaping at the transmitter, the transmitting filter selectivity, and the transmittance of the line, as well as receiving filter selectivity. The latter can be varied within reasonable limits if the others are adjusted in a complementary fashion to obtain the desired output response. In evaluating the merit of different receiving filter designs it is reasonable to compare them with the same average signal power on the line. We shall also assume that the line has been equalized for unity gain and linear phase over the band so that it can be considered as a transparent link in the system.

The average signal power on the line can be computed in terms of (a) the transmittance function $Y_r(\omega)$ of the receiving filter, (b) the required function $G_1(\omega)$ representing the spectrum of the modified switching function $g_1(t)$ at the detector input, and (c) the statistics of the data sequence. Details of the calculation are given in Appendix B. An interesting consequence of the assumptions that the FM wave has continuous phase and that the frequency shift is equal to the signaling rate is the appearance of discrete components on the line at the marking and spacing frequencies even when the data sequence is random. This means there are transmitted sine waves which consume power but carry no information. An optimization procedure aimed at conserving power would very nearly suppress these components at the transmitter by balance or by sharp antiresonances and restore them to their proper relative amplitudes by complementary narrow-band resonance peaks in the response of the receiving bandpass filter. The bandwidth used to augment these frequencies at the receiver could in theory be made so small that no appreciable effect on the accepted noise would result. The system would then only have to deliver the average power associated with the continuous part of the FM spectrum.

Actually, even a partial suppression of the steady-state components on the line would destroy much of the advantage of signaling by FM. The system would become more sensitive to gain changes and overload distortion. Accurate tracking of the suppression and recovery circuits for the marking and spacing frequencies would be difficult at best and would be practically impossible over a channel with carrier frequency offset. The narrow-band recovery circuits would contribute to a sluggish start-up time. In fact, about the only remaining resemblance to FM would be the use of an FM detector. If low-level tones can actually be recovered successfully from a received wave, it would be better to use them for synchronous PM detection, which is a linear method capable of attaining ideal performance in the presence of additive Gaussian noise. It appears that Sunde's system should carry the power in the steady-state components in order to deserve the name of FM.

Standard variational procedures can be applied to find the shape of receiving filter selectivity which minimizes probability of error when the average signal power and the spectral density of added Gaussian noise on the line are specified. The solution of the optimization problem is given in Appendix B, and means are shown for completing the computation of the corresponding probabilities of error for the most and least vulnerable data sequences. In the case of white Gaussian noise on the line, the optimum receiving filter has very nearly the same cosine characteristic found by Sunde for optimum binary AM transmission. The bounds for error probability are plotted in Fig. 6 for both FM proper with no suppression of steady-state tones and the abnormal FM with marking and spacing frequencies suppressed. Also shown is the ideal curve representing what can be proved to give the best possible binary performance. The ideal curve can theoretically be obtained for example by coherent detection of binary phase modulation. Differentially detected phase modulation requires about 1 db more signal power than ideal at an error probability of 10^{-4}.

It is seen from Fig. 6 that when the suppression bands are inserted in Sunde's binary FM system, the theoretical performance is only about a half db poorer than ideal, but, as previously pointed out, this does not represent a true FM system. The more legitimate FM has error bounds from 3 to 3.5 db poorer than ideal. However, a penalty of this order of magnitude could be a fair trade in many cases for the advantages of a much simplified receiver relatively immune to many channel faults.

VI. APPLICATION TO DATA TERMINALS FOR USE ON TELEPHONE CHANNELS

We now apply our formulas to calculate error probabilities in binary FM transmission with terminals more closely resembling those actually in use on telephone channels. In the design of real-life terminals, the emphasis is placed on ruggedness and simplicity. The bit rate is not locked to the frequency deviation. The filters do not meet elaborate optimization requirements. The significant conclusion from our evaluation of error probabilities for the practical systems is that the degradation of performance compared with the ideal is actually very slight.

The probability of error as given in (44) is generally applicable to FM systems. There are three parameters, ρ, a, and b, given in (41) to (43). The first parameter ρ is a signal-to-noise ratio. It depends on the

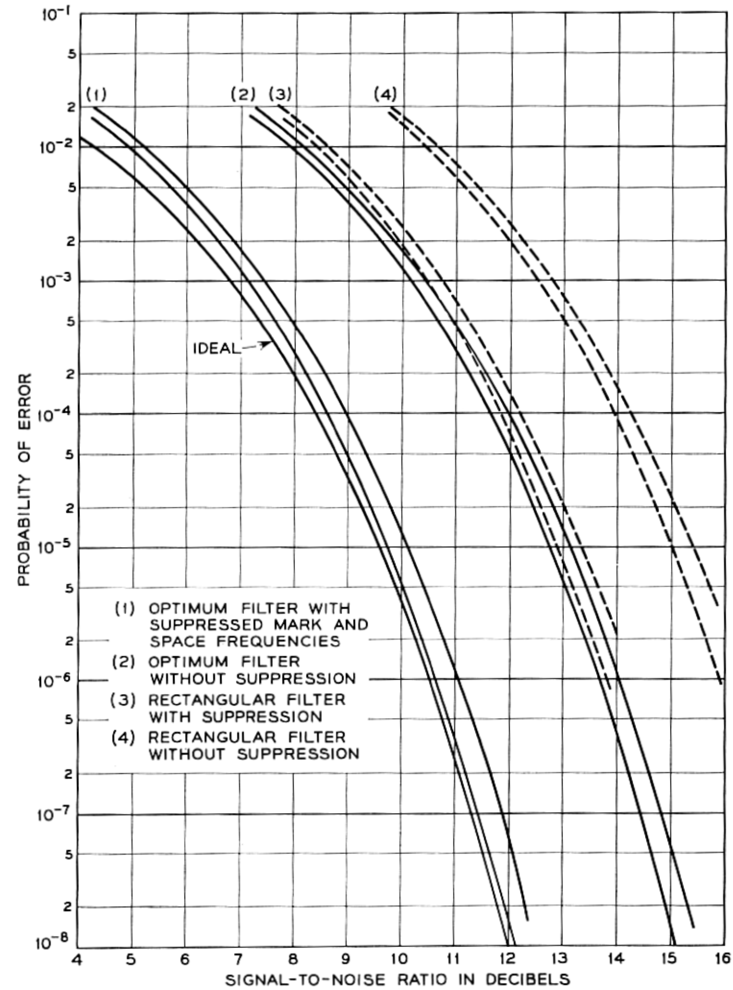

Fig. 6 — Error probabilities for Sunde's binary FM system with additive Gaussian noise. Bounds are for most and least vulnerable sequences. Noise reference is mean noise power in bandwidth equal to bit rate.

ratio of instantaneous envelope of the received signal to the rms noise voltage at the detector input. For any given front-end filter, this parameter can be expressed in terms of average signal-to-noise ratio at the input of the receiver. The parameter a depends on the ratio of instantaneous frequency displacement at the sampling time to the Gabor noise bandwidth, σ_1/σ_0, of the receiver. The third parameter b depends on the derivative of the instantaneous envelope at the sampling time. For a given channel these parameters can be computed for any particular signaling sequence. The true probability of error could conceivably be obtained by averaging over all possible sequences, but this would be a formidable task. Instead we will give bounds on the probability of error for the most and least vulnerable sequences over a finite representative set of signaling intervals.

We first consider the system in Fig. 7, which has amplitude-vs-frequency "raised cosine" type roll-off but no phase distortion. Equal filtering takes place at the transmitter and receiver. The modulator applies a pure FM wave of constant envelope to the transmitting filter. In other words, the modulator and the demodulator are ideal. The data source is composed of rectangular pulses. The frequency deviation in cps is equal to half the bit rate. These rates and deviations are characteristic of practical systems.

With the aid of a digital computer, S. Habib has calculated the parameters given in (41) to (43) for 2^{10} sequences. From these calculations we have computed an upper and a lower bound on the probability of error. These results are shown in Fig. 8. The probability of error for all other sequences will fall between the two curves labeled "best" and "worst." Superimposed on the same graph is the ideal curve, which can only be achieved with ideal phase systems and coherent detection. The FM detection is, of course, incoherent.

Our next example applies the theory to a real bandpass filter used in an operational data set. Fig. 9 shows the system considered. The curve

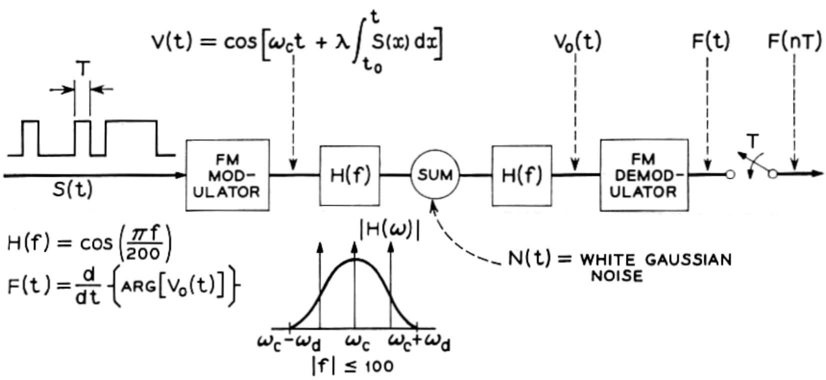

Fig. 7 — Ideal FM modulator and demodulator with transmitted and received signals equally shaped by "raised cosine" type roll-off amplitude characteristics and no phase distortion.

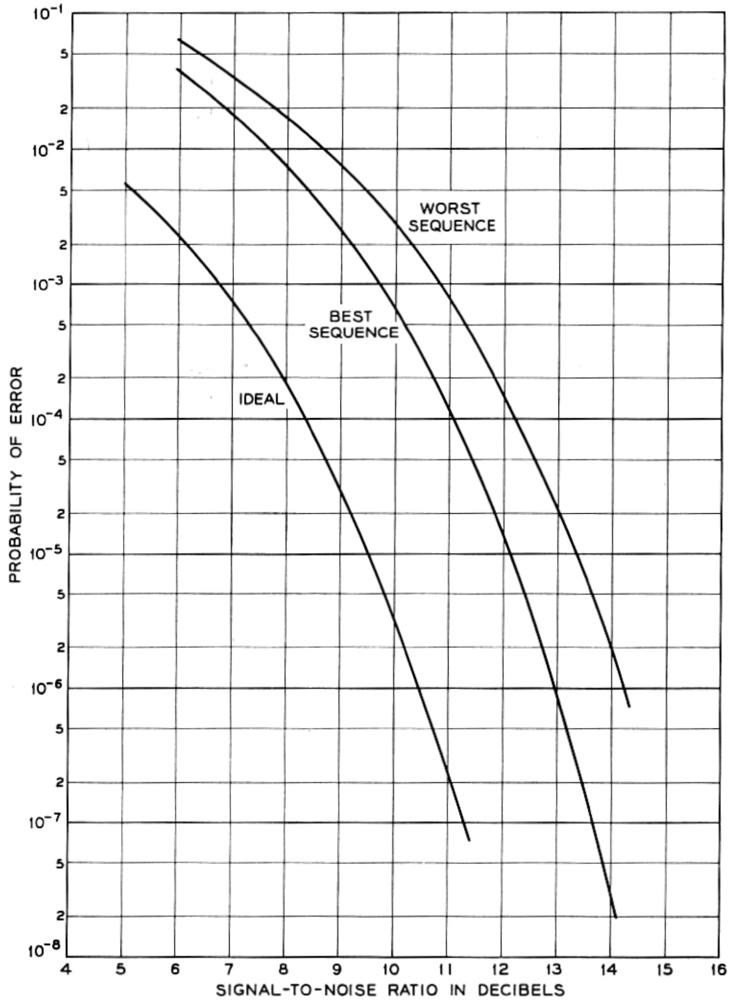

Fig. 8 — Probability of error for system depicted in Fig. 7. Noise reference is mean power in bandwidth equal to bit rate.

of loss vs frequency for the filter used is given in Fig. 10. The curve departs from the condition of symmetry about midband, and also the separation between the signal and carrier bands is not sufficient to make overlapping effects negligible. The marking and spacing frequencies were assumed to be 1200 and 2200 cps, respectively, and the signaling rate 1200 bits per second. As shown in Fig. 11, the calculated results are

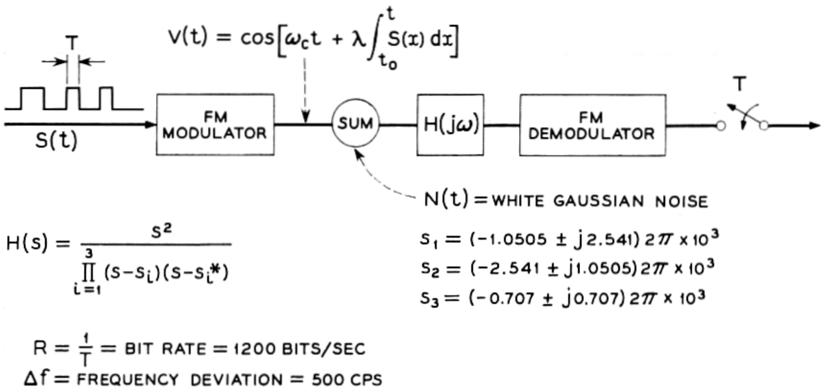

Fig. 9 — Ideal modulator and demodulator with received signal shaped by filter characteristics used in FM data set and shown in Fig. 10.

about 1 db better than the experimental results obtained with a random word generator, random noise generator, and error counter. The experimental system included an axis-crossing detector and post-detection low-pass filter, which do not correspond precisely with the theoretical model. In view of the differences cited, the agreement between calculated and experimental curves is good. The penalty suffered by the actual back-to-back channel compared with the best theoretical FM performance is between 2 and 3 db. Somewhat more optimistic estimates have been given in other published studies.[7,8] The effects of amplitude and delay-versus-frequency variation in the channel are calculable by use of the computer programs we have established.

It was shown in the previous sections that a lower bound on the probability of error occurs when the parameter b is set equal to zero. For

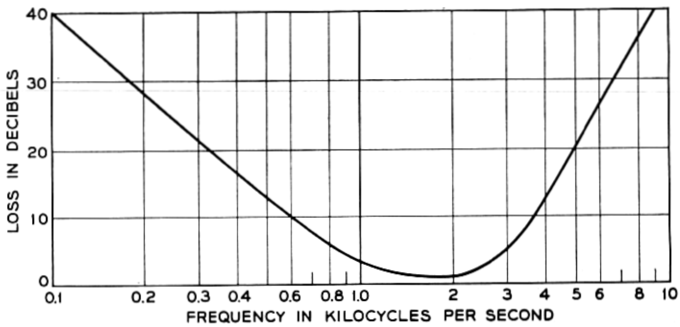

Fig. 10 — Receiver bandpass filter loss vs frequency characteristic.

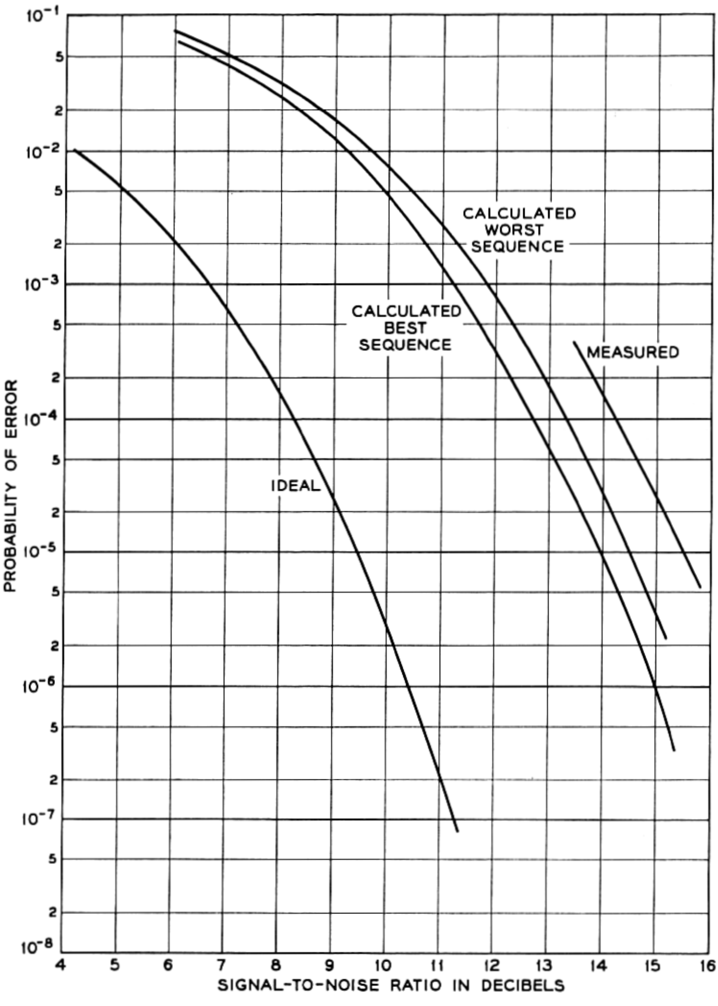

Fig. 11 — Calculated and measured error rates for system depicted in Fig. 9. Noise reference is mean noise power in bandwidth equal to bit rate.

this reason we include Fig. 12, showing a set of universal curves relating the corresponding minimum probability of error to ρ and a.

APPENDIX A

Evaluation of Integral for Error Probability

We evaluate the integral

$$P_+ = \int_0^\infty dz \int_{-\infty}^\infty \int_{-\infty}^\infty p(-z \mid x,y) q(x,y) \, dx \, dy \tag{90}$$

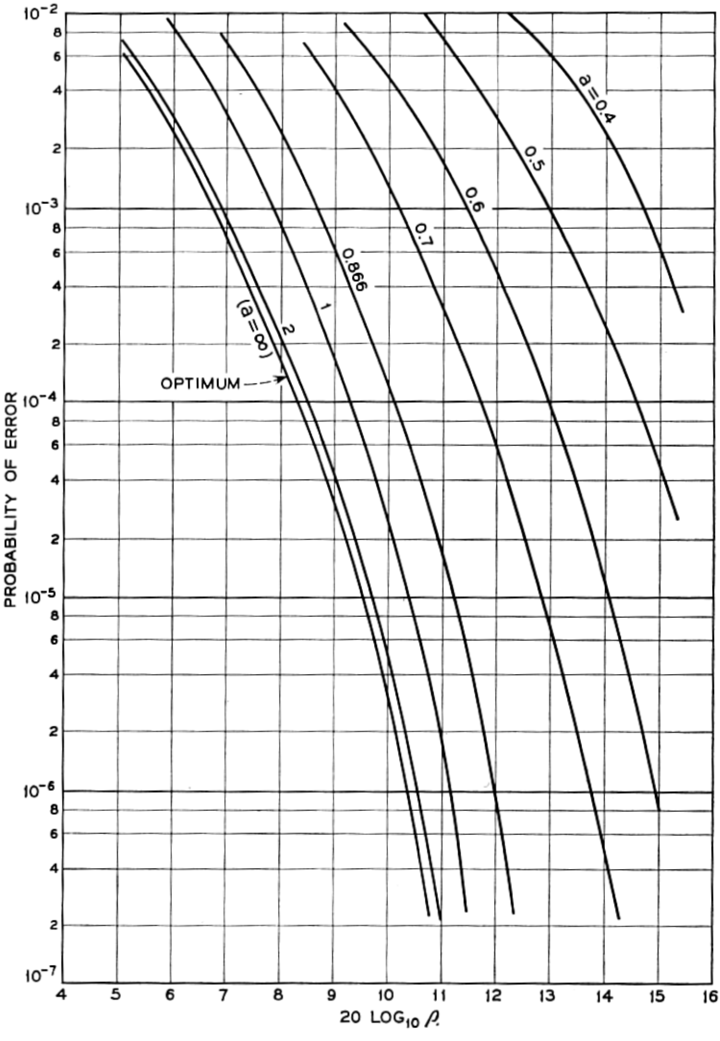

Fig. 12 — General curves of minimum probability of error vs ρ for different values of a in the range of interest.

where

$$p(-z \mid x,y) = \frac{1}{\sigma_1[2\pi(x^2 + y^2)]^{\frac{1}{2}}} \exp\left[-\frac{(z + \dot{Q}x - \dot{P}y)^2}{2\sigma_1^2(x^2 + y^2)}\right] \quad (91)$$

$$q(x,y) = \frac{1}{2\pi\sigma_0^2} \exp\left[-\frac{(x - P)^2 + (y - Q)^2}{2\sigma_0^2}\right]. \quad (92)$$

The integration with respect to z can be performed at once in terms of the error function by substituting a new variable u defined by

$$(z + \dot{Q}x - \dot{P}y)^2 = 2\sigma_1^2(x^2 + y^2)u^2. \tag{93}$$

The result is:

$$P_+ = \frac{1}{2}\frac{1}{4\pi\sigma_0^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \operatorname{erf}\frac{\dot{Q}x - \dot{P}y}{\sigma_1[2(x^2+y^2)]^{\frac{1}{2}}} \cdot\exp\left[-\frac{(x-P)^2 + (y-Q)^2}{2\sigma_0^2}\right] dx\, dy. \tag{94}$$

We now transform to polar coordinates, setting

$$x = r\cos\theta \qquad y = r\sin\theta \qquad dx\,dy = r\,dr\,d\theta \tag{95}$$

We also let

$$P\cos\theta + Q\sin\theta = R\cos(\theta - \alpha) = R\cos\psi$$
$$\dot{Q}\cos\theta - \dot{P}\sin\theta = D\cos(\theta + \beta) = D\cos(\psi + \gamma)$$

where

$$\begin{aligned} R^2 &= P^2 + Q^2 = 2\sigma_0^2\rho^2 & \tan\alpha &= Q/P \\ D^2 &= \dot{P}^2 + \dot{Q}^2 & \tan\beta &= \dot{P}/\dot{Q} \\ \psi &= \theta - \alpha & \gamma &= \alpha + \beta. \end{aligned} \tag{96}$$

The result of the transformation is

$$1 - 2P_+ = \frac{e^{-\rho^2}}{2\pi\sigma_0^2}\int_{-\pi}^{\pi} \operatorname{erf}\frac{D\cos(\psi+\gamma)}{\sqrt{2}\,\sigma_1}\,d\psi \int_0^{\infty}\exp\left[-\frac{r^2 - 2rR\cos\psi}{2\sigma_0^2}\right] r\,dr. \tag{97}$$

The integration with respect to r can be performed by subtracting and adding the term $R\cos\psi$ to r. This enables separation of the integrand into a perfect differential and a term which can be expressed as an error function. We thereby obtain

$$1 - 2P_+ = \frac{1}{2\pi}\int_{-\pi}^{\pi}\operatorname{erf}\frac{D\cos(\psi+\gamma)}{\sqrt{2}\,\sigma_1} \left[1 + \sqrt{2\pi}\,\frac{R}{2\sigma_0}\exp\left(-\frac{R^2}{2\sigma_0^2}\sin^2\psi\right)\cos\psi\left(1 - \operatorname{erf}\frac{R\cos\psi}{\sqrt{2}\,\sigma_0}\right)\right] d\psi. \tag{98}$$

We note that both $\cos\psi$ and $\cos(\psi + \gamma)$ change sign when ψ is increased by π and that $\sin^2(\psi + \pi) = \sin^2\psi$. Furthermore, the integration in (98) is over one full period in ψ, and for every value of ψ in the left half of the period there is a corresponding value in the right half at $\psi + \pi$. Since the error function, erf z, is an odd function of z, a change in the sign of $\cos \psi$ or $\cos (\psi + \gamma)$ changes the sign of the corresponding error function in the integrand. If we multiply the first term under the integral sign by the terms within the bracket following, we see that there is only one product which does not change sign at points π apart. The integral of the other products must vanish. The integral of the one which does not change sign is twice the integral over a half period of ψ. Hence

$$1 - 2P_+ = \frac{R}{\sigma_0\sqrt{2\pi}} \int_{-\pi/2}^{\pi/2} \exp\left(-\frac{R^2}{2\sigma_0^2}\sin^2\psi\right) \cdot \cos\psi \operatorname{erf} \frac{D\cos(\psi+\gamma)}{\sqrt{2}\,\sigma_1}\, d\psi. \tag{99}$$

From (96) and (23)

$$\begin{aligned} D\cos\gamma &= D(\cos\alpha\cos\beta - \sin\alpha\sin\beta) \\ &= D\left(\frac{P}{R}\frac{\dot{Q}}{D} - \frac{Q}{R}\frac{\dot{P}}{D}\right) = \frac{P\dot{Q} - Q\dot{P}}{R} = R\phi \end{aligned} \tag{100}$$

$$\begin{aligned} D\sin\gamma &= D(\sin\alpha\cos\beta + \cos\alpha\sin\beta \\ &= D\left(\frac{Q}{R}\frac{\dot{Q}}{D} + \frac{P}{R}\frac{\dot{P}}{D}\right) = \frac{Q\dot{Q} + P\dot{P}}{R} \\ &= \frac{1}{2R}\frac{d}{dt}(R^2) = \frac{dR}{dt} = \dot{R}. \end{aligned} \tag{101}$$

Therefore

$$\begin{aligned} D\cos(\psi+\gamma) &= D\cos\gamma\cos\psi - D\sin\gamma\sin\psi \\ &= R\phi\cos\psi - \dot{R}\sin\psi. \end{aligned} \tag{102}$$

Now substituting $x = \sin\psi$ in (99) we rearrange to obtain

$$P_+ = \frac{1}{2} - \frac{\rho}{2\sqrt{\pi}} \int_{-1}^{1} e^{-\rho^2 x^2} \operatorname{erf} \frac{R\phi(1-x^2)^{\frac{1}{2}} - \dot{R}x}{\sqrt{2}\,\sigma_1}\, dx. \tag{103}$$

Equation (34) of the main text is obtained from (103) by substituting the complementary function erfc $z = 1 - \operatorname{erf} z$.

The lower bound P_l on the probability of error for any fixed R and ϕ was shown in the text to be obtained by setting $\dot{R} = 0$. When this substitution is made in (103) and the definition of the error function in terms of an integral is inserted, we obtain

$$P_l = \frac{1}{2} - \frac{2\rho}{\pi} \int_0^1 e^{-\rho^2 x^2}\, dx \int_0^{a\rho(1-x^2)^{\frac{1}{2}}} e^{-z^2}\, dz. \tag{104}$$

The parameters a and ρ are defined by (41) and (42). If we substitute $\rho x = y$ the expression becomes

$$P_l = \frac{1}{2} - \frac{2}{\pi} \int_0^{\rho} \int_0^{a(\rho^2-y^2)^{\frac{1}{2}}} e^{-y^2-z^2}\, dy\, dz. \tag{105}$$

The region of integration in the double integral consists of the first quadrant of the ellipse

$$z^2/(a\rho)^2 + y^2/\rho^2 = 1. \tag{106}$$

After transforming to polar coordinates by setting $y = r \cos \theta$ and $z = r \sin \theta$, we can perform the integration with respect to r. The result is

$$P_l = \frac{1}{\pi} \int_0^{\pi/2} \exp\left[-\frac{a^2\rho^2}{\sin^2 \theta + a^2 \cos^2 \theta}\right] d\theta. \tag{107}$$

This is equivalent to (37) of the main text.

The integral has a simple value when $a = 1$, which is equivalent to $\phi = \sigma_1/\sigma_0$. For this case the integrand is seen to become a constant and (40) results. This coincides with a result given for a special case by Montgomery.[9] By a change in the meaning of the parameters it also gives the error probability for differential binary phase detection as discussed in Section IV. In the general case, the limiting form of P_l for large signal-to-noise ratio can be calculated by the method of steepest descents. Saddle points occur at $\theta = 0$ and $\theta = \pi/2$. When $a > 1$, the saddle point at $\theta = 0$ determines the asymptotic form of the integral for large ρ and (38) is obtained. When $a < 1$, the saddle point at $\theta = \pi/2$ is dominant and we obtain (39).

APPENDIX B

Optimization of Receiving Filter for Sunde's FM System

Our problem is to find the receiving filter characteristic which minimizes the probability of error in Sunde's FM system when the average transmitted signal power and the spectral density of the noise on the line are specified. In terms of Fig. 13 the transmittance function for the filter is $Y_r(\omega)$ and the output of the filter is $V_r(t)$ as defined by (61), (62), (71), and (73), namely

$$V_r(t) = A \sin \omega_d t \sin (\omega_c t + \theta_r) - A s_1(t) \cos (\omega_c t + \theta_r) \tag{108}$$

$$s_1(t) = \sum_{n=-\infty}^{\infty} (-)^n b_n g_1(t - nT) \tag{109}$$

$$G_1(\pm \omega_d - \lambda) + G_1(\pm \omega_d + \lambda) = 2G_1(\omega_d) = T \qquad 0 < \lambda < \omega_d \tag{110}$$

$$G_1(\omega) = \int_{-\infty}^{\infty} g_1(t) \cos \omega t \, dt. \tag{111}$$

The input to the filter is the sum of the signal wave $V(t)$ plus the Gaussian noise wave $v_0(t)$. The wave $V(t)$ is defined as that function of time which when operated on by $Y_r(\omega)$ produces $V_r(t)$. The noise wave $v(t)$ at the input to the frequency detector has a spectral density equal to $|Y_r(\omega)|^2$ times that of $v_0(t)$.

We shall simplify our treatment by assuming a random sequence of data in which the two binary symbols are selected with equal probability. The probability is then equal to 0.5 that any particular b_n has the value $+1$ and also 0.5 that the value is -1. We regard $V_r(t)$ as a member of an ensemble of random functions with a distribution in the infinite number of independent random parameters b_n. The randomness appears entirely in the function $s_1(t)$. We can calculate the ensemble average of $s_1(t)$ at fixed t by adding the individual averages of the terms in the infinite series defining $s_1(t)$. When we do this we find that the only random variable in each term is b_n, which assumes the values ± 1 with equal likelihood and therefore has the average value zero. Hence the ensemble average of $s_1(t)$, which we shall designate by $\langle s_1(t) \rangle$, is zero for any fixed value of t. It follows that $s_1(t)$ can contain no periodic components, for the presence of any such components would give a non-zero average at some values of t. Therefore, the spectral density function of the second term in $V_r(t)$ must be a continuous function of frequency.

To calculate the average square of $s_1(t)$ over the ensemble, we note that $s_1(t)$ is the sum of an infinite number of independent random variables of form

$$z_n = (-)^n b_n g_1(t - nT). \tag{112}$$

The average value of each z_n is zero and the variance, or mean square minus the square of the mean, is equal to the square of $g_1(t - nT)$.

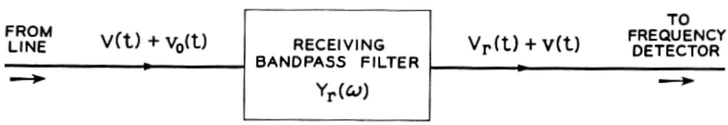

Fig. 13 — Function of receiving filter in Sunde's system.

Since the variance of the sum of independent variables is equal to the sum of the variances of the individual variables, we can write

$$\langle s_1^2(t) \rangle = \sum_{n=-\infty}^{\infty} g_1^2(t - nT). \tag{113}$$

The average in (113) is an ensemble average at fixed t. We can show that this average is periodic in t with period T by noting that

$$\begin{aligned} \langle s_1^2(t + T) \rangle &= \sum_{n=-\infty}^{\infty} g_1^2(t + T - nT) \\ &= \sum_{m=-\infty}^{\infty} g_1^2(t - mT) \\ &= \langle s_1^2(t) \rangle. \end{aligned} \tag{114}$$

Therefore the average over t can be computed by averaging over a single period from $t = 0$ to $t = T$. Hence the average over time which we shall designate by av is

$$\begin{aligned} \text{av } s_1^2(t) &= \frac{1}{T} \int_0^T \langle s_1^2(t) \rangle \, dt = \frac{1}{T} \sum_{n=-\infty}^{\infty} \int_0^T g_1^2(t - nT) \, dt \\ &= \frac{1}{T} \sum_{n=-\infty}^{\infty} \int_{-nT}^{-(n-1)T} g_1^2(\lambda) \, d\lambda = \frac{1}{T} \int_{-\infty}^{\infty} g_1^2(\lambda) \, d\lambda. \end{aligned} \tag{115}$$

By application of Parseval's theorem

$$\text{av } s_1^2(t) = \frac{1}{2\pi T} \int_{-\infty}^{\infty} G_1^2(\omega) \, d\omega. \tag{116}$$

From (116) we deduce that the spectral density of $s_1(t)$ is given by

$$w_1(\omega) = \frac{G_1^2(\omega)}{2\pi T} = \frac{\omega_d G_1^2(\omega)}{2\pi^2}. \tag{117}$$

The spectral density of $V_r(t)$ can now be easily calculated. The first term can be expressed as the sum of sine waves of amplitude $A/2$ and frequencies $\omega_c + \omega_d$ and $\omega_c - \omega_d$. The first term therefore contributes line spectra with mean square $A^2/8$ at the marking and spacing frequencies. The average square of the second term can be written

$$\text{av } [A^2 s_1^2(t) \cos^2 (\omega_c t + \theta_r)] = \frac{A^2}{2} \text{av } s_1^2(t). \tag{118}$$

The spectral components comprising $s_1(t) \cos (\omega_c t + \theta_r)$ are those of $s_1(t)$ shifted from their original positions to appear as sidebands around the frequencies $\pm\omega_c$. Hence $w_r(\omega)$, the spectral density of $V_r(t)$ with all power assigned to positive frequencies, is given by

$$w_r(\omega) = \frac{A^2}{8}\,\delta(\omega - \omega_c + \omega_d) + \frac{A^2}{8}\,\delta(\omega - \omega_c - \omega_d) + \frac{\omega_d A^2 G_1^2(\omega - \omega_c)}{4\pi^2} \qquad \omega \geqq 0. \tag{119}$$

It is convenient to let $\omega - \omega_c = u$ and write for the transmittance of the filter

$$U(u) = Y_r(\omega - \omega_c). \tag{120}$$

We shall also designate the spectral density of $V(t)$ as $w(u)$. Since the linear operator $U(u)$ can be applied individually to the components which make up (119) we must have

$$w(u) = \frac{A^2\delta(u + \omega_d)}{8\,|\,U(-\omega_d)\,|^2} + \frac{A^2\delta(u - \omega_d)}{8\,|\,U(\omega_d)\,|^2} + \frac{\omega_d A^2 G_1^2(u)}{4\pi^2\,|\,U(u)\,|^2}. \tag{121}$$

The average power on the line is proportional to W_0, the average square of $V(t)$, which is given by

$$W_0 = \int_{-2\omega_d}^{2\omega_d} w(u)\,du = \frac{A^2}{8\,|\,U(-\omega_d)\,|^2} + \frac{A^2}{8\,|\,U(\omega_d)\,|^2} + \frac{\omega_d A^2}{4\pi^2}\int_{-2\omega_d}^{2\omega_d} \frac{G_1^2(u)}{|\,U(u)\,|^2}\,du. \tag{122}$$

We make the reasonable assumption that $|\,U(u)\,|$ is an even function of u. Combined with the further assumption that the spectral density of the noise on the line is symmetrical about ω_c, this furnishes a convenient assurance of a symmetrical spectral density for the noise in the output of the receiving filter. Since $G_1(u)$ is also an even function of u, we can write (122) in the equivalent form

$$W_0 = \frac{A^2}{4X(\omega_d)} + \frac{\omega_d A^2}{2\pi^2}\int_0^{2\omega_d} \frac{G_1^2(u)}{X(u)}\,du \tag{123}$$

where

$$X(u) = |\,U(u)\,|^2. \tag{124}$$

The function $X(u)$ is to be chosen to minimize the probability of error under the constraint that W_0 is held constant. In calculating the optimum function, the signal power represented by the steady-state components can be ignored, since this power could be reduced to an arbitrarily small value by the use of narrow-band suppression techniques. The constraint on the signal power is therefore that the integral in (123) is to be held constant.

Let $N(u)$ represent the spectral density of the Gaussian noise wave $v_0(t)$ on the line. Then the spectral density of $v(t)$, the noise in the output of the receiving filter, is $X(u)N(u)$. In terms of the spectral density $w_v(\omega)$ previously defined for $v(t)$ with values symmetrically distributed between positive and negative frequencies, we have

$$X(u)N(u) = 2w_v(u + \omega_c). \tag{125}$$

The values of σ_0 and σ_1 necessary to complete the calculation of the probability of error by (34) can now be found by substituting (125) in (17) and (18) giving the results

$$\sigma_0^2 = 2 \int_0^{2\omega_d} X(u)N(u)\, du \tag{126}$$

$$\sigma_1^2 = 2 \int_0^{2\omega_d} u^2 X(u)N(u)\, du. \tag{127}$$

If we substitute (126) and (127) into the general expression for error probability, (34), and attempt to formulate a variational problem, the expressions become unmanageable. Instead, we concentrate attention on the lower bound for error probability obtained by setting $\dot{R} = 0$, (36), in which it is evident that to make the error probability as small as possible both σ_0 and σ_1 should be made as small as possible. As shown by (126) and (127), σ_0 and σ_1 are not independent. The effect of the dependence can be taken into account by performing the minimization problem in two steps. First we minimize σ_0 with both σ_1 and W_0 held constant. After this solution is obtained, we find by trial the value of σ_1 which yields the lowest minimum probability of error.

Omitting inconsequential multiplying factors, we set the variational problem as

$$\delta\left[\int_0^{2\omega_d} X(u)N(u)\, du + \lambda \int_0^{2\omega_d} u^2 X(u)N(u)\, du + \mu \int_0^{2\omega_d} \frac{|G_1(u)|^2}{X(u)}\, du\right] = 0 \tag{128}$$

where λ and μ are Lagrange multipliers and the function under variation is $X(u)$. The solution is

$$X(u) = \frac{\mu\, |G_1(u)|}{(1 + \lambda u^2)^{\frac{1}{2}}\, N^{\frac{1}{2}}(u)}. \tag{129}$$

It is straightforward to verify that this stationary value of $X(u)$ actually gives a minimum value of σ_0 and hence minimum probability of error for fixed values of σ_1 and W_0.

Substituting our partially optimized solution in (123), (126), and (127), we obtain

$$\sigma_0^2 = 2\mu \int_0^{2\omega_d} \frac{|G_1(u)| N^{\frac{1}{2}}(u)}{(1 + \lambda u^2)^{\frac{1}{2}}} du \tag{130}$$

$$\sigma_1^2 = 2\mu \int_0^{2\omega_d} \frac{u^2 |G_1(u)| N^{\frac{1}{2}}(u)}{(1 + \lambda u^2)^{\frac{1}{2}}} du. \tag{131}$$

$$W_0 - W_s = \frac{\omega_d A^2}{2\pi^2 \mu} \int_0^{2\omega_d} |G_1(u)| N^{\frac{1}{2}}(u)(1 + \lambda u^2)^{\frac{1}{2}} du \tag{132}$$

$$W_s = \frac{A^2 N^{\frac{1}{2}}(\omega_d)(1 + \lambda \omega_d^2)^{\frac{1}{2}}}{4\mu |G_1(\omega_d)|} \tag{133}$$

$$\frac{1}{\rho^2} = \frac{2\sigma_0^2}{A^2} = \frac{2\omega_d I_1 I_2}{\pi^2(W_0 - W_s)} \tag{134}$$

$$\frac{1}{a^2\rho^2} = \frac{2\sigma_1^2}{A^2\omega_d^2} = \frac{2 I_2 I_3}{\pi^2 \omega_d (W_0 - W_s)} \tag{135}$$

where

$$I_1 = \int_0^{2\omega_d} \frac{|G_1(u)| N^{\frac{1}{2}}(u)}{(1 + \lambda u^2)^{\frac{1}{2}}} du \tag{136}$$

$$I_2 = \int_0^{2\omega_d} |G_1(u)| N^{\frac{1}{2}}(u)(1 + \lambda u^2)^{\frac{1}{2}} du \tag{137}$$

and

$$I_3 = \int_0^{2\omega_d} \frac{u^2 |G_1(u)| N^{\frac{1}{2}}(u)}{(1 + \lambda u^2)^{\frac{1}{2}}} du. \tag{138}$$

These equations furnish a straightforward procedure for calculating the optimum filter characteristic. Each assumed value of λ determines a pair of values ρ and $a\rho$ from which the corresponding upper and lower bounds for the error probability can be evaluated by computer techniques. By successive trials the best value of λ can be approximated to any desired degree and substituted in (129) to obtain the best filter selectivity function. In actual examples tried, this procedure could be shortened because the error probability turned out to be very much more sensitive to the value of ρ than to the value of $a\rho$. If this were known beforehand, we would place no constraint on σ_1 in the minimization of σ_0. This is equivalent to setting $\lambda = 0$, leading to the simpler formulas

$$\frac{1}{\rho^2} = \frac{2\omega_d}{\pi^2(W_0 - W_s)} \left[\int_0^{2\omega_d} |G_1(u)| N^{\frac{1}{2}}(u) \, du \right]^2 \tag{139}$$

$$\frac{1}{a^2\rho^2} = \frac{2}{\pi^2\omega_d(W_0 - W_s)} \int_0^{2\omega_d} u^2 |G_1(u)| N^{\frac{1}{2}}(u) \, du \cdot \int_0^{2\omega_d} |G_1(u)| N^{\frac{1}{2}}(u) \, du. \tag{140}$$

By applying Schwarz' inequality to the products of integrals in (134) and (135), we verify that the case of $\lambda = 0$ gives the maximum value of ρ, but that the maximum value of $a\rho$ occurs when $\lambda = \infty$. It seems therefore that an intermediate nonzero value of λ would be best, but in the cases computed the improvement obtainable in this way turned out to be negligibly small.

As an example, consider the raised cosine signal spectrum in which $G_1(u)$ is given by (77). We also assume a white noise spectrum in which $N(u)$ is equal to a constant N_0. It is convenient to introduce as a signal-to-noise ratio the quantity M defined by

$$M = \frac{W_0}{N_0\omega_0} = \frac{W_0}{2N_0\omega_d}. \tag{141}$$

This is the ratio of average transmitted signal power to the average noise power in a band of frequencies of width equal to the bit rate. Computer results show that the case of $\lambda = 0$ is practically indistinguishable from the optimum λ. Hence we set $\lambda = 0$ and calculate for the optimum filter

$$U(u) = X^{\frac{1}{2}}(u) = \left(\frac{\mu\pi}{\omega_d N_0}\right)^{\frac{1}{2}} \cos \frac{\pi u}{4\omega_d} \qquad |u| < 2\omega_d. \tag{142}$$

This is the same cosine filter characteristic found by Sunde to be optimum for binary AM with synchronous detection. From (132) and (133) we find that with $\lambda = 0$

$$W_0 - W_s = \frac{A^2\omega_d N_0^{\frac{1}{2}}}{2\pi\mu} = W_s. \tag{143}$$

Hence

$$W_s = W_0/2 \quad \text{and} \quad W_0 - W_s = W_0/2. \tag{144}$$

From (139), (140), and (141) we then calculate

$$\rho^2 = \frac{W_0}{4\omega_d N_0} = \frac{M}{2} \tag{145}$$

$$\begin{aligned} a^2\rho^2 &= \frac{3\pi^2 W_0}{16\omega_d N_0(\pi^2 - 6)} \\ &= \frac{3\pi^2 M}{8(\pi^2 - 6)} = 0.956M. \end{aligned} \tag{146}$$

If the steady-state components were suppressed, we would set $W_s = 0$ and would then obtain $\rho^2 = M$, $a^2\rho^2 = 1.913M$. This would correspond to a 3-db shift in the direction of lower signal-to-noise ratio when the error probability curves are plotted against $10 \log_{10} M$.

The curves of Fig. 6, showing the upper and lower bounds for error probability when Sunde's FM system is optimized, were calculated by S. Habib on the digital computer. The case of a nonoptimum receiving filter is illustrated by the corresponding curves for a rectangular band defined by

$$X(u) = X_0 \qquad |u| < 2\omega_d . \tag{147}$$

For this case we compute from (126) and (127)

$$\sigma_0^2 = 2\int_0^{2\omega_d} X_0 N_0 \, du = 4\omega_d X_0 N_0 \tag{148}$$

$$\sigma_1^2 = 2\int_0^{2\omega_d} u^2 X_0 N_0 \, du = \frac{16\omega_d^3 X_0 N_0}{3}. \tag{149}$$

From (123)

$$W_0 = \frac{A^2}{4X_0} + \frac{A^2}{8\omega_d X_0}\int_0^{2\omega_d}\left(1 + \cos\frac{\pi u}{2\omega_d}\right)^2 du = \frac{5A^2}{8X_0}. \tag{150}$$

We then calculate

$$\rho^2 = 2M/5 \qquad a^2\rho^2 = 3M/10. \tag{151}$$

If the steady-state components are suppressed, the average transmitted power could be reduced to $(\frac{5}{8} - \frac{1}{4})/(\frac{5}{8}) = \frac{3}{5}$ of the previously determined value, which is a saving of 2.2 db.

REFERENCES

1. Bennett, W. R., *Electrical Noise*, McGraw-Hill Book Co., Inc., New York, 1960, pp. 234–238.
2. Rice, S. O., Statistical Properties of a Sine Wave Plus Random Noise, B.S.T.J., **27,** Jan., 1948, pp. 109–157.
3. Lawton, J. G., Theoretical Error Rates of "Differentially Coherent" Binary

and "Kineplex" Data Transmission Systems, Proc. I.R.E., **47,** Feb., 1959, pp. 333–334.
4. Sunde, E. D., Ideal Pulses Transmitted by AM and FM, B.S.T.J., **38,** Nov., 1959, pp. 1357–1426.
5. Nyquist, H., Certain Topics in Telegraph Transmission Theory, Trans. A.I.E.E., **47,** April, 1928, pp. 617–644.
6. Knopp, K., *Theory and Applications of Infinite Series*, Blackie and Son, Ltd., London, 1928, p. 269.
7. Meyerhoff, A. A. and Mazer, W. M., Optimum Binary FM Reception Using Discriminator Detection and I-F Shaping, RCA Review, **22,** Dec., 1961, pp. 698–728.
8. Smith, E. F., Attainable Error Probabilities in Demodulation of Random Binary PCM/FM Waveforms, I.R.E. Trans. on Space Elec. and Telemetry, **SET-8,** Dec., 1962, pp. 290–297.
9. Montgomery, G. F., A Comparison of Amplitude and Angle Modulation for Narrow-Band Communication of Binary-Coded Messages in Fluctuation Noise, Proc. I.R.E., **42,** Feb., 1954, pp. 447–454.