Optimum Detection and Signal Design for Channels With Non- but Near-Gaussian Additive Noise

ADISAI BODHARAMIK, JOHN B. MOORE, AND ROBERT W. NEWCOMB

Abstract—**The Gram–Charlier series representation of the noise-probability density function is used to determine an optimum detector for signals in non-Gaussian but near-Gaussian (NGNG) noise. Solutions are obtained for coherent and incoherent detection. Optimal detectors for several typical transmitting systems are determined. Generally these detectors consist of the standard detector for Gaussian noise with the addition of a few, not too sophisticated, nonlinear elements. The performance of a detector, specified by the upper bound on the probability of error, is assessed and is seen to depend on the signal shape, the time-bandwidth product, and the signal-to-noise ratio. The optimal signal to minimize the probability of error is determined and is seen to result as a solution to Duffing's second-order nonlinear differential equation.**

I. INTRODUCTION

AT THE present time there is likely to be an increasing interest in digital data transmission, which in some respects is more effective than analog transmission, especially for long distance communication. Considerable work has been done [1]-[3] on designing the optimum detector for digital data transmission systems in the presence of interference, but most of the effort has been devoted to the case of Gaussian interference. In this case the solutions for both coherent and incoherent reception schemes have been well analyzed and the optimal detector in the presence of white Gaussian noise is well known to be a matched filter [1]. Under the Gaussian assumption the receiver performance can be analytically evaluated [4] and in some situations it seems to agree rather well with the results for many physical systems [5, p. 286]. However, the performances of some practical communication channels are not in close agreement with the predicted values [6] based upon the Gaussian noise assumption. Thus there is an obvious need to consider the detection of signals in non-Gaussian noise, as a closer investigation [7], [8] reveals that for some communication circuits the major source of additive disturbance may be non-Gaussian.

Paper approved by the Communication Theory Committee of the IEEE Communications Society for publication without oral presentation. Manuscript received December 13, 1971; revised July 6, 1972. This work was supported in part by the U.S. Air Force under Grant 70-1910. It is based in part on a portion of a dissertation by A. Bodharamik in partial fulfillment of the requirements for the Ph.D. degree at the University of Maryland, College Park.

A. Bodharamik was with the Department of Electrical Engineering, University of Maryland, College Park, Md. He is now with the Telephone Organization of Thailand, Bangkok, Thailand.

J. B. Moore was on leave at the Department of Electrical Engineering, University of Maryland, College Park, Md. He is with the Department of Electrical Engineering, University of Newcastle, New South Wales, Australia.

R. W. Newcomb is with the Department of Electrical Engineering, University of Maryland, College Park, Md. 20742.

Several authors have evaluated the system performance for some digital systems in the presence of non-Gaussian noise utilizing the detector that is optimum under the Gaussian assumption [9]-[13]. However it should be noted that this detector is not optimum for non-Gaussian interference. A derivation for an optimal nonlinear detector for non-Gaussian noise is given in [14], but the treatment is limited to large variance noise specified by the Cauchy distribution. Although these results are quite interesting many other distributions than the Cauchy may occur in practice.

In this paper we introduce another model of non-Gaussian interference, which uses a series expansion, consisting of a sum with a Gaussian multiplier, to represent the probability density function. We assume that the first few terms of the series are sufficient to represent the noise probability density.

We shall discuss the detection problem for coherent and incoherent detection for various types of transmitting. As seen in Section III the detector consists of a standard detector, the matched filter of the Gaussian case with a few nonlinear elements added. We can observe (Section IV), from the upper bounds on the probability of error that the receiver performance depends upon the signal-to-noise ratio, the time-bandwidth product, and the particular signal used. Consequently, a solution for an optimal signal to achieve the minimum probability of error is derived in Section V, this resulting as the solution of the second order nonlinear equation (66). We begin in the next section with a development of the series expansion used to describe the probability density functions.

II. NOISE SPECIFICATIONS

We introduce the expression of a non-Gaussian noise-probability density function (pdf) by the Gram–Charlier series [15, p. 156], [16, p. 222]. This is a representation involving orthonormal functions and a normal reference function.

Let $f(n)$ be a pdf of non-Gaussian noise having zero mean and let

$$N(0, \sigma^2, n) = (1/\sqrt{2\pi}\,\sigma) \exp(-n^2/2\sigma^2)$$

be a Gaussian reference function.

If an expansion exists of the form

$$f(n) = \sum_{j=0}^{\infty} c_{2j} h_{2j}(n) N(0, \sigma^2, n) \tag{1}$$

having

$$h_{2j}(n) = (1/\sqrt{2j!})H_{2j}(n/\sigma) \qquad H_{2j} = \text{Hermite polynomial} \tag{2}$$

then this series (1) is called a modified *Gram-Charlier series* and we have

$$c_{2j} = \int_{-\infty}^{\infty} h_{2j}(n)f(n)\,dn \tag{3}$$

$$\int_{-\infty}^{\infty} h_{2ij}(n)h_2(n)N(0,\sigma^2,N)\,dn = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases} \tag{4}$$

We comment that our zero mean assumption allows us to consider only even order terms for (1). The conditions under which the Gram-Charlier expansion is valid have been discussed by Cramer [16] where it is pointed out that finite sums of the form of (1) are of most use. Indeed, for our purposes we will assume that the noise can be represented by a $P+1$ term truncated version of the Gram-Charlier series. Thus

$$f(n) = \sum_{j=0}^{P} c_{2j}h_{2j}(n)N(0,\sigma^2,n). \tag{5}$$

Then we define *non-Gaussian near-Gaussian* (NGNG) noise as that for which $P = 2$ and either c_2 or c_4 nonzero, i.e., NGNG noise is that which can be adequately represented by the first three terms of a modified Gram-Charlier series. For NGNG noise we will also choose the σ^2 of the normal multiplier equal to the second moment of the noise process.

Equation (5) can be written in an alternate form, which will be more useful for detection problems, as

$$f(n) = \sum_{j=0}^{P} \left(\frac{n}{\sigma}\right)^{2j}\alpha_{2j}N(0,\sigma^2,n). \tag{6}$$

Here α_{2j} is a function of various noise moments and the Hermite polynomials as follows

$$\alpha_{2j} = \sum_{k=0}^{P} a_{2k,2j}q_{2k}, \qquad j = 0,1,2,\cdots,P \tag{7}$$

with

$$q_{2k} = \frac{1}{(2k)!}\sum_{l=0}^{P}\frac{a_{2k,2l}\mu_{2l}}{\sigma^{2l}}, \qquad k = 0,1,2,\cdots,P \tag{8}$$

in which μ_{2l} is the $2l$th moment of the noise process. The coefficients $a_{2k,2l}$ can be generated from the recurrence formula [16, p. 156]

$$H_i(x) = xH_{i-1}(x) - (i-1)H_{i-2}(x), \qquad i = 2,3,4,\cdots \tag{9}$$

given $H_0(x) = 1$ and $H_1(x) = x$. Therefore

$$H_{2j}(x) = \sum_{k=0}^{P} a_{2j,2k}x^{2k}, \qquad j = 0,1,2,\cdots,P, \tag{10}$$

where $a_{2j,2k} = 0$ for all $k \rangle j$.

It is important to note that the sum of a finite number of terms of the series (6) may give a negative density function, particularly near the tails. Furthermore, the series may behave irregularly in the sense that the sum of j terms may give a worse fit than the sum of $(j-1)$ terms [15]. It is also true that for a fixed number of terms in the series many different normal functions can be chosen. However, there exists an optimum normal function such that the mean-square error is minimized [17].

In order to simplify the problem we shall assume from this point on, unless otherwise stated, that the noise is NGNG. We would comment that this representation has never given us a negative density function in the range of a satisfactory approximation.

This type of noise model includes the contribution of a Gaussian term that is always present in any physical system and results for the Gaussian case can be obtained by setting $c_0 = 1, c_2 = c_4 = 0$.

In general, it has been known that in order to optimize the detection process a knowledge of all the higher order probability density functions of the interference are required [2]. If the interfering noise is Gaussian, a second-order statistic implies all the higher order statistics, so that solutions for optimal detection are quite simple. This is not true for non-Gaussian noise. Moreover, in practice when one meets non-Gaussian noise it is in general nonstationary, for example its statistics may depend on many factors including geography and the time of day [18]. Here we make the assumption that the noise is considered to be quasi-stationary with statistics that remain unchanged over a period that is long compared to the signal interval. This allows us to use stationary results. Another assumption is that the correlation time of the noise is small compared to the duration of a signal to be detected. In other words the noise bandwidth is large compared to the signal bandwidth. Then successive noise samples are considered to be independent.

III. Detection Problems

We shall formulate the optimum receiver or detector under the optimality criteria of minimum average probability of error. We assume that data to be transmitted is presented to the transmitter in the form of a sequence of binary digits that can be denoted by zeros and ones appearing at a rate of one every T seconds. During the interval $mT < t < (m+1)T$, if the mth position of the sequence is 1, the system transmits a signal $s_1(t)$; if it is a 0, the transmitted signal is $s_2(t)$. Having observed the received waveform $y(t)$ during the signal duration T, the detector is to choose between two hypotheses,

$$H_j: y(t) = s_j(t) + n(t) \qquad j = 1,2: 0 < t < T, \tag{11}$$

where $n(t)$ denotes stationary NGNG noise.

The detector observes the received waveform by uniformly sampling the waveform with sampling interval δ. If B is the system bandwidth, then due to the sampling theorem [19] we choose the sampling interval δ as

$$\delta = (1/2B) \tag{12}$$

in which case the maximum number of independent samples during the duration T is

$$M = 2TB. \tag{13}$$

The optimum receiver for these M samples is specified by the log likelihood functional that is given by [2, p. 91]

$$\ln L(Y) = \ln F_1(Y) - \ln F_2(Y), \tag{14}$$

where

$$Y = [y_1, y_2, y_3, \cdots, y_m] \tag{15}$$

is used to denote a vector consisting of M samples received during the interval T and $F_k(Y)$ is used to denote the joint pdf of the sampled waveform assuming that hypothesis H_k of (11) is true.

A. Coherent Detection

The binary signals in digital communication systems are usually of the narrow-band type. The signal consists of a high-frequency sinusoidal carrier modulated in amplitude or phase by a slowly varying function of time. Such signals can be written as

$$s_k(t) = C_k(t) \cos(\omega t + \phi_k(t)), \quad k = 1, 2 \tag{16}$$

where $C_k(t)$ and $\phi_k(t)$ are called the amplitude modulation and the phase modulation, respectively.

By definition coherent detection means that the functions $C_k(t)$ and $\phi_k(t)$ of the received signal are precisely known, so that $s_k(t)$ is known exactly. From (6) it is clear that the pdf $\mathfrak{Z}_k(y_i)$ of a received sample under the hypothesis H_k is given by

$$\begin{aligned}\mathfrak{Z}_k(y_i) &= f(y_i - s_{ki}) \\ &= \sum_{j=0}^{P=2} \alpha_{2j} \left[\frac{y_i - s_{ki}}{\sigma}\right]^{2j} \exp\left(-\left[\frac{(y_i - s_{ki})^2}{2\sigma^2}\right] \Big/ \sqrt{2\pi\sigma^2}\right), \\ &\qquad i = 1, 2, 3, \cdots, M\end{aligned} \tag{17}$$

where σ^2 is the average noise energy and s_{ki} denotes the ith sample of signal $s_k(t)$.

Since the joint pdf $F_k(Y)$ is equal to the product of the pdf's of individual samples, (14) yields

$$\ln L[Y] = \sum_{i=1}^{M} \frac{y_i}{\sigma^2}(s_{1i} - s_{2i}) + \sum_{i=1}^{M} \frac{(s_{2i}^2 - s_{1i}^2)}{2\sigma^2} + \sum_{i=1}^{M} \ln \left[\frac{\sum_{j=0}^{P=2} \alpha_{2j} \left(\frac{y_i - s_{1i}}{\sigma}\right)^{2j}}{\sum_{j=0}^{P=2} \alpha_{2j} \left(\frac{y_i - s_{2i}}{\sigma}\right)^{2j}}\right]. \tag{18}$$

It is interesting to notice that the first two terms of (18) represent the optimum detector for the white Gaussian noise case (matched filter or correlator) [2], [20] and the third term adds a nonlinearity in which the nonlinear element is specified by the statistics of the NGNG noise. If we were to assume the additive noise is Gaussian then the third term in (18) is absent. Here we show that if the noise parameters could be adjusted during the detection process (adaptive receiver) the receiver performance would appear optimum for non-Gaussian as well as Gaussian interference.

When the sample size M is large enough, the sums of (18) can be approximately represented by integrals yielding the detector structure shown in Fig. 1 where $N_0/2$ denotes the power spectral density of the noise process. We are using (13) and the fact that the average noise energy is $\sigma^2 = N_0 B$. This structure consists of a linear matched filter in parallel with a nonlinear portion in which the nonlinearity is specified by the NGNG noise parameters. When the transmitting signals are specified, for instance as in ON-OFF (ASK), frequency shift keying (FSK), or phase-shift keying (PSK) systems, the receiver structure can be obtained from a reduction of Fig. 1. Here we do this for the simplest transmitting system, which is the ASK system. For this case the transmitting signals are $s_1(t) = s(t)$ and $s_2(t) = 0$. The receiver structure is readily obtained by substituting $s_1(t)$ and $s_2(t)$ into the integral form of (18) and is shown in Fig. 2.

For future reference we recall that the FSK system has

$$s_1(t) = C_1(t) \cos(\omega_1 t + \varphi)$$

and

$$s_2(t) = C_1(t) \cos(\omega_2 t + \varphi)$$

while the PSK system is defined by

$$s_1(t) = C_1(t) \cos(\omega_1 t + \varphi)$$

and

$$s_2(t) = C_1(t) \cos(\omega_1 t + \varphi + \pi).$$

B. Incoherent Detection

For this case it is presumed that the phase of the transmitted signal is distorted during transmission through the channel. We assume the carrier phase is totally unknown to the receiver, thus the received signal can be expressed in the following form

$$s_k(t, \theta) = C_k(t) \cos(\omega t + \phi_k(t) + \theta), \quad k = 1, 2 \tag{19}$$

where θ denotes the random phase assumed to be uniformly distributed between 0 to 2π.

Under the hypothesis $H_k (k = 1, 2)$ and fixed random phase θ the joint probability density function of the received waveform samples becomes

$$F_k(Y/\theta) = \prod_{i=1}^{M} A_i(y_i - s_{ki}), \tag{20}$$

where

$$A_i(y_i - s_{ki}) = \sum_{j=0}^{P=2} \frac{\alpha_{2j}}{\sigma^{2j}} [y_i - s_{ki}(\theta)]^{2j} \cdot \frac{\exp\left[\frac{-(y_i - s_{ki}(\theta))^2}{2\sigma^2}\right]}{\sqrt{2\pi\sigma^2}}. \tag{21}$$

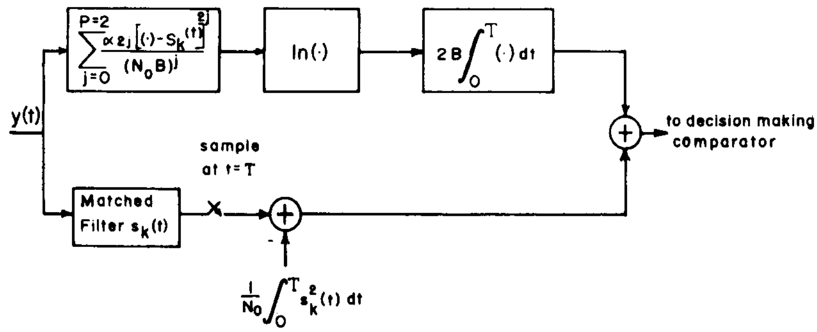

Fig. 1. kth branch of optimal nonlinear detector for deterministic signals in NGNG noise.

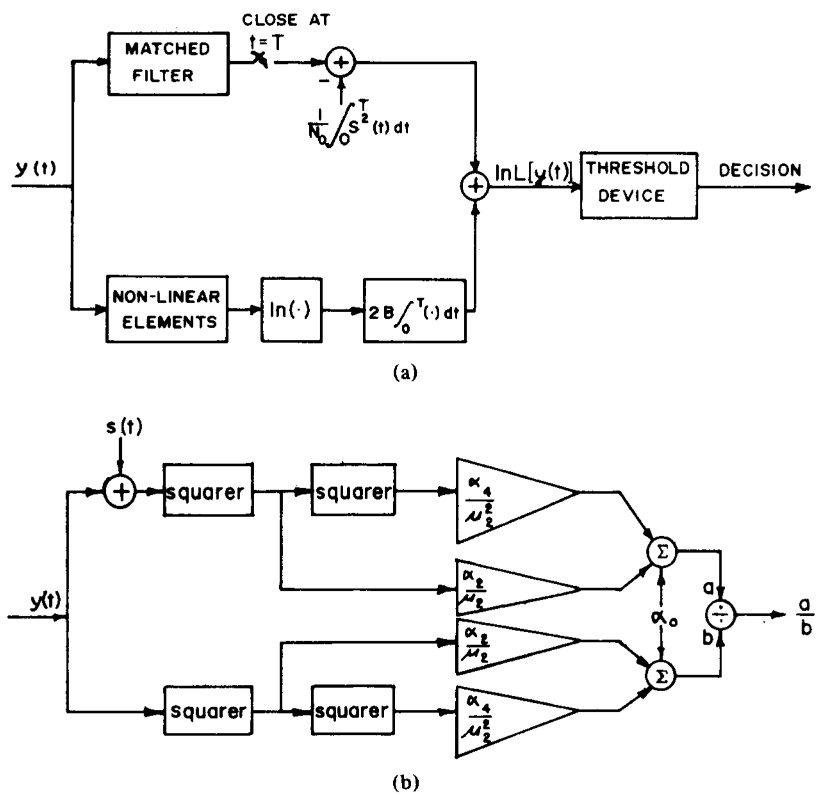

Fig. 2. (a) Optimal detector for ASK system (coherent reception). (b) Nonlinear elements in detection of Fig. 2(a).

Averaging (20) over θ results in

$$F_k(Y) = \frac{1}{2\pi} \int_0^{2\pi} \exp\left[\sum_{i=1}^{M} \ln A_i(y_i - s_{ki})\right] d\theta. \quad (22)$$

Equation (22) can now be applied to the log-likelihood functional, which from (14) is given by

$$\ln L[Y] = \ln\left\{\frac{\frac{1}{2\pi}\int_0^{2\pi} \exp\left[\sum_{i=1}^{M} \ln A_i(y_i - s_{1i})\right] d\theta}{\frac{1}{2\pi}\int_0^{2\pi} \exp\left[\sum_{i=1}^{M} \ln A_i(y_i - s_{2i})\right] d\theta}\right\}. \quad (23)$$

To evaluate (23) we make the assumption that the receiver will operate under not unreasonably large signal-to-noise ratios. Accordingly we can keep the first two terms using (21) of the expansion of $\ln A_i$ in (23). We then write (23) in integral form; on some computation after substituting $s_k(t, \theta)$ of (19) into the integral form of (23) we obtain

$$\ln L[y(t)] = 2B(K_1 - K_2) + \ln\left[\frac{I_0(\sqrt{R_1^2 + Q_1^2})}{I_0(\sqrt{R_2^2 + Q_2^2})}\right], \quad (24)$$

where $I_0(\cdot)$ denotes the modified Bessel function of the first kind and

$$K_k = \frac{\gamma_2}{N_0 B}\int_0^T y^2(t)\,dt + \frac{\gamma_2}{N_0 B} E_k + \frac{\gamma_4}{(N_0 B)^2}\int_0^T y^4(t)\,dt + \frac{3}{2}\frac{\gamma_4}{(N_0 B)^2} G_k + \frac{3\gamma_4}{(N_0 B)^2}\int_0^T y^2(t) C_k^2(t)\,dt \quad (25)$$

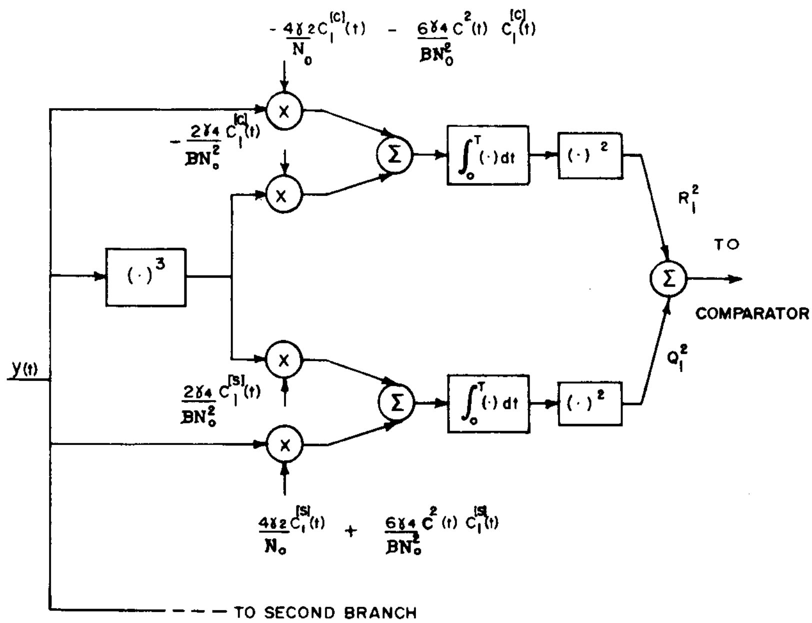

Fig. 3. First branch of the optimal detector for FSK system. (Incoherent reception.)

$$R_k = -2B\left\{\frac{2\gamma_2}{N_0 B}\int_0^T y(t)C_k^{[c]}(t)dt + \frac{3\gamma_4}{(N_0B)^2}\int_0^T y(t)C_k^2(t)C_k^{[c]}(t)dt + \frac{\gamma_4}{(N_0B)^2}\int_0^T y^3(t)C_k^{[c]}(t)dt\right\} \quad (26)$$

$$Q_k = 2B\left\{\frac{2\gamma_2}{N_0 B}\int_0^T y(t)C_k^{[s]}(t)dt + \frac{3\gamma_4}{(N_0B)^2}\int_0^T y(t)C_k^2(t)C_k^{[s]}(t)dt + \frac{\gamma_4}{(N_0B)^2}\int_0^T y^3(t)C_k^{[s]}(t)dt\right\} \quad (27)$$

and

$$\gamma_2 = \frac{\alpha_2}{\alpha_0} - \frac{1}{2} \qquad \gamma_4 = \frac{\alpha_4}{\alpha_0} - \frac{\alpha_2^2}{2\alpha_0^2}$$

$$E_k = \int_0^T s_k^2(t,\theta)dt = \frac{1}{2}\int_0^T C_k^2(t)dt \quad (28)$$

$$G_k = \int_0^T s_k^4(t,\theta)dt = \frac{1}{4}\int_0^T C_k^4(t)dt \quad (29)$$

$$C_k^{[s]}(t) = C_k(t)\cos(\omega_k t + \phi_k(t)) \quad (30)$$

$$C_k^{[c]}(t) = C_k(t)\sin(\omega_k t + \phi_k(t)). \quad (31)$$

In general (24) describes the structure of the optimum receiver for narrow-band random phase signals. It should be noticed that the receiver for this case is similar to the standard detector for Gaussian noise [21], [22, p. 217] except for some additional nonlinear elements. Just as with the Gaussian noise case, the optimum receiver correlates the sine and cosine demodulator outputs against each of the lowpass signals $C_k(t)$. For each received $s_k(t)$ the receiver forms the sum of the square of the cosine correlation and the sine correlation with the results obtained then fed to a comparison device.

At this point, we look at optimum receivers for the different signaling systems. For an ASK system the log-likelihood functional becomes, from (24)

$$\ln L[y(t)] = \frac{2\gamma_2 E}{N_0} + \frac{3\gamma_4 G}{N_0^2 B} + \frac{6\gamma_4}{N_0^2 B}\int_0^T y^2(t)C^2(t)dt + \ln I_0\left[\sqrt{R^2+Q^2}\right]. \quad (32)$$

For an FSK system with orthogonal signals of equal energy, the receiver consists of two similar branches each of which is to generate $R_k^2 + Q_k^2$ for $k = 1, 2$. One branch of the structure of the receiver is shown in Fig. 3.

IV. System Performance

In the previous section we have been concerned primarily with the structure of the optimal detection system. In this section we shall investigate the receiver performance that is completely specified by the probability of detection error P_e.

In many cases of interest, the test likelihood ratio can be derived but an exact performance calculation is sometimes impossible. For our noise model we encounter this difficulty. Therefore, it is useful to search for another measure that may be weaker than the probability of error but that is easier to evaluate.

We shall use the upper bound given by Chernoff [23]. Let $P_{e\max}$ designate the upper bound on P_e, and let the signals be equally likely, then

$$P_e \leqslant P_{e\max} = \frac{1}{2} \int \underset{(M)}{\cdots} \int [F_1(Y)]^m [F_2(Y)]^{1-m} \, dY, \qquad 0 < m < 1, \quad (33)$$

where F_k and Y are defined as at (15).

It can be shown that $P_{e\max}$ is a convex function of m for $0 < m < 1$ [24]. Therefore, a true minimum $P_{e\max}$ exists by a proper choice of m. For simplicity we choose m to be one-half; thus our bound is

$$P_{e\max} = \frac{1}{2} \int \underset{(M)}{\cdots} \int F_1^{1/2}(Y) F_2^{1/2}(Y) \, dY. \quad (34)$$

Substituting (17) into (34) and taking logarithms of both sides yields

$$\ln (2P_{e\max}) = \sum_{i=1}^{M} \ln \left[\int_{-\infty}^{\infty} \sqrt{\mathcal{Z}_1(y_i)\mathcal{Z}_2(y_i)} \, dy_i \right]. \quad (35)$$

Before proceeding further to evaluate (35) we must specify the transmitting system as well as the reception scheme. We first develop the performance bound for ASK systems with coherent detection. For this particular case we have, from (17) with $s_2 = 0$,

$$[\mathcal{Z}_1(y)\mathcal{Z}_2(y)]^{1/2} = [\alpha_0^2 + \textcircled{B}\,(y,s)]^{1/2} \frac{1}{2\pi\sigma^2} \exp\left[-\frac{s^2}{8\sigma^2} - \frac{1}{2\sigma^2}\left(y - \frac{s}{2}\right)^2 \right], \quad (36)$$

where $\textcircled{B}\,(y,s)$ is a polynomial in y and s.

Generally for NGNG noise the noise parameters satisfy $|\alpha_0| \gg |\alpha_2| \gg |\alpha_4|$, in which case $\textcircled{B}\,(y,s) \ll \alpha_0^2$, in a reasonable range of signal-to-noise ratio. Using the series expansions for the square root, keeping the first two terms, and substituting (36) into (35) yields the following result in integral form.

$$P_{e\max} \approx \frac{1}{2} \exp\left\{ -\frac{1}{4} \int_0^T \frac{s^2(t)}{N_0} dt + 2B \int_0^T \ln (g_0\alpha_0) dt + 2B \int_0^T \ln \left(1 + \frac{g_2}{g_0} \frac{s^2(t)}{N_0 B} + \frac{g_4}{g_0} \frac{s^4(t)}{(N_0 B)^2} + \frac{g_6}{g_0} \frac{s^6(t)}{(N_0 B)^3} + \frac{g_8}{g_0} \frac{s^8(t)}{(N_0 B)^4} \right) dt \right\}, \quad (37)$$

where g_0, g_2, g_4, etc., are functions of the α_i. It can be shown from (7), (8), (10) that

$$\alpha_0 = 1 + \frac{1}{8}\left(\frac{\mu_4}{\sigma^4} - 3\right) \qquad \alpha_2 = 2(1 - \alpha_0) \qquad 3\alpha_4 = \alpha_0 - 1. \quad (38a)$$

Defining also

$$\beta = (\alpha_0 - 1)/\alpha_0 \quad (38b)$$

gives

$$g_0 = 1 - \beta + \frac{11}{6}\beta^2 \qquad g_2 = -\frac{4}{3}\beta^2 \qquad g_4 = \frac{1}{48}(\beta + 11\beta^2). \quad (38c)$$

Using the fact that $g_0\alpha_0 \approx 1$ and $\ln(1 + x) \approx x - x^2/2$ for $x \ll 1$, (37) reduces to

$$P_{e\max} \approx \frac{1}{2} \exp\left\{ -\left(\frac{1}{4} - \frac{2g_2}{g_0^2}\right) \int_0^T \frac{s^2(t)}{N_0} dt + 4\left(\frac{g_4}{g_0} - \frac{g_2^2}{2g_0^2}\right) \frac{1}{2BN_0^2} \int_0^T s^4(t) dt \right\}. \quad (39)$$

It should be noticed that the higher order terms become negligible for the case of interest when P_e is in a reasonable range such that the signal-to-noise ratio, e.g., the integral of s^2 divided by N_0, is not very large. Equation (39) is the upper bound on P_e of an ASK system with coherent detection. Similarly, the upper bounds for FSK and PSK systems can be obtained. The results are shown in (40) and (41), respectively:

$$P_{e\max} \approx \frac{1}{2} \exp\left\{ -\left[\frac{1}{2} - 4h_2\right] \frac{1}{N_0} \int_0^T C^2(t) dt + \left[h_{22} + 3h_4 - \frac{5}{2}h_2^2\right] \frac{1}{2BN_0^2} \int_0^T C^4(t) dt \right\} \quad (40)$$

and

$$P_{e\max} \approx \frac{1}{2} \exp\left\{ -\left(1 - \frac{2b_2}{b_0}\right) \frac{1}{N_0} \int_0^T s^2(t) dt + 4\left(\frac{b_4}{b_0} - \frac{b_2^2}{2b_0^2}\right) \frac{1}{2BN_0^2} \int_0^T s^4(t) dt \right\}, \quad (41)$$

where the h_i and b_i are functions of noise parameters, i.e.,

$$h_2 = \frac{2 - g_2}{2g_0} \qquad h_4 = \frac{1 + g_4}{g_0} \qquad h_{22} = \left(1 + \frac{3}{2}\beta + \frac{93}{8}\beta^2\right) / g_0 \quad (42)$$

and

$$b_0 = g_0 \qquad b_2 = 4(\beta + g_2) \qquad b_4 = 16 g_4. \quad (43)$$

It is important to emphasize that the upper bound $P_{e\max}$ does not only depend on the signal-to-noise ratio as in the case of Gaussian noise but also depends on the particular signal used as well as the time bandwidth product $(2TB)$.

For the case of incoherent detection the upper bound on probability of error is given by

$$P_{e\max} = \frac{1}{2\pi} \int_0^{2\pi} P_{e\max/\theta} \, d\theta, \quad (44)$$

where $P_{e\max/\theta}$ denotes the upper bound on P_e given the random phase θ. We observe that for narrow-band and equal-energy signals the two integrals in (41) are independent of θ.

Therefore, for narrow-band signals $P_{e\max/\theta}$ is independent of θ and the upper bound for incoherent detection becomes the same as the one for coherent detection.

For the later purpose of choosing optimal signals we rewrite the equations of $P_{e\max}$, (39), (40), and (41) in a general form, i.e.,

$$P_{e\max} = \frac{1}{2} \exp\left\{-\frac{D_1}{2N_0}\int_0^T C^2(t)\,dt + \frac{D_2}{8BN_0^2}\int_0^T C^4(t)\,dt\right\}, \tag{45}$$

where D_1, D_2 depend on the transmitting system used as shown for the three cases in Table I.

As illustrations, some plots of bounds on the probability of error versus signal-to-noise ratio are shown in Fig. 4. This $P_{e\max}$ requires the knowledge of the optimal signal (to be shown in Section V). Upon examining the curves of Fig. 4 we see first of all that by increasing the time bandwidth product $2TB$ a better detector performance is obtained; secondly among the three transmitting systems the performance of the PSK system is the best, whereas the ASK system gives the highest probability of error. This result is the same as for the case when the noise in the system is Gaussian [4].

We have covered a wide range of detection problems including the implementation of optimum nonlinear detectors, as well as evaluation of the system performance for some typical transmitting systems. Perhaps one could say that the main difference between optimum signal detection in Gaussian noise and NGNG noise is that the performance of the latter case could be improved by the selection of a proper signal, whereas for the Gaussian noise case the performance depends only on signal energy regardless of its shape.

TABLE I

	ASK	FSK	PSK
D_1	$\frac{1}{4} - \frac{2g_2}{g_0^2}$	$\frac{1}{2} - 4h_2$	$1 - \frac{2b_2}{b_0}$
D_2	$4\left(\frac{g_4}{g_0} - \frac{1}{2}\frac{g_2^2}{g_0^2}\right)$	$4\left(h_{22} + 3h_4 - \frac{5}{2}h_2^2\right)$	$4\left(\frac{b_4}{b_0} - \frac{b_2^2}{2b_0^2}\right)$

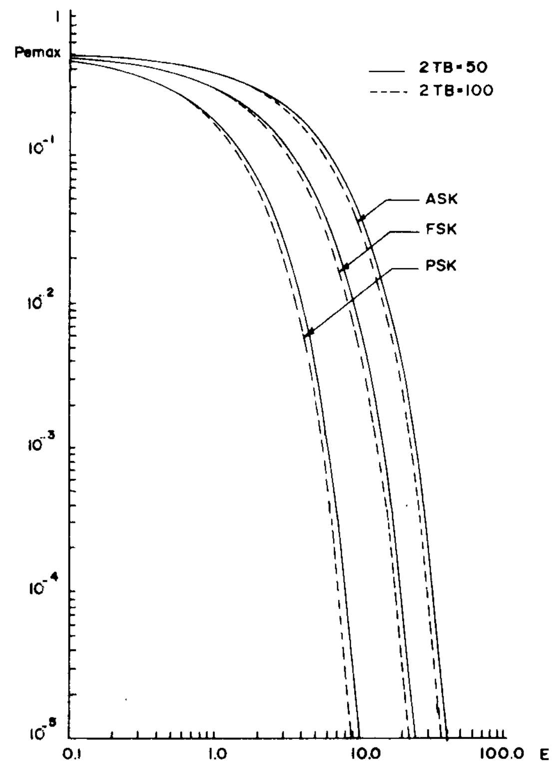

Fig. 4. Bounds on the probability of error (coherent and incoherent receptions).

V. Signal Design

We shall consider the problem of minimizing the upper bound on the probability of error by a choice of transmitting signal, since it is known that the signal that minimizes the upper bound is also the signal that minimizes the actual probability of error [25].

For convenience, we assume the observation interval is $[-T/2, T/2]$. We shall develop the optimal signal design (to minimize $P_{e\max}$) subject to some physically meaningful constraints. As a preliminary case to these, we minimize the $P_{e\max}$ with no constraint. The second case is to minimize $P_{e\max}$ subject to an energy constraint. The third case is the most meaningful since we will introduce both signal energy and signal bandwidth constraints in the minimization process.

Here we define the mean-square bandwidth of the envelope signal as

$$W^2 = \int_{-T/2}^{T/2} [\dot{C}(t)]^2\,dt, \tag{46}$$

where $\dot{C}(t)$ is the time derivative of $C(t)$ and W^2 is measured in H_z^2. As shown in [26], the mean-square bandwidth is the bandwidth that contains the major part of the signal energy.

Before proceeding to the solution, we define the index

$$I[C] = \int_{-T/2}^{T/2} [C^4(t) - \tau C^2(t)]\,dt \tag{47}$$

where

$$\tau = \tau_1/\tau_2 \tag{48}$$

and

$$\tau_1 = (D_1/2N_0) \tag{49}$$

$$\tau_2 = (D_2/8BN_0^2). \tag{50}$$

Therefore, $P_{e\max}$ in (45) can be written as

$$P_{e\max} = \frac{1}{2}\exp\{\tau_2 I[C]\}. \tag{51}$$

From (51) it is obvious that to minimize $P_{e\max}$ we need to minimize the performance index $I[C]$.

We first consider the simplest case when the mean-square bandwidth W^2 is allowed to take any value and the signal energy E is required to be finite.

To carry out the minimization, using the calculus of variations [27], [28, p. 659] we let

$$C(t) = C_0(t) + \epsilon C_A(t), \tag{52}$$

where $C_0(t)$ is the optimum signal envelope and $C_A(t)$ is an arbitrary function. We require that

$$\left.\frac{dI}{d\epsilon}\right|_{\epsilon=0} = 0. \tag{53}$$

Substituting (52) in (47) and carrying out the step indicated in (53) yields

$$\int_{-T/2}^{T/2} C_A(t)\,[4C_0^3(t) - 2\tau]\,dt = 0. \tag{54}$$

Since C_A is arbitrary, the terms in the brackets must be identically zero in which case

$$4C_0^2(t) - 2\tau = 0, \quad -T/2 \leqslant t \leqslant T/2 \tag{55}$$

or

$$C_0(t) = \sqrt{\frac{T}{2}}, \quad -T/2 \leqslant t \leqslant T/2. \tag{56}$$

Equation (56) indicates the optimal signal that is shown in Fig. 5. Observe that the optimum signal for this case turns out to be a rectangular signal with infinite mean-square bandwidth W^2 and finite energy determined by noise in the system. The result is obviously logical due to the fact that without any restriction on energy the best signal—in the sense of minimum P_e—should have energy proportional to the noise energy.

The second case of more practical interest is the signal design problem with specified energy

$$\mathcal{E} = \int_{-T/2}^{T/2} C^2(t)dt. \tag{57}$$

From (47) the index that we have to minimize becomes

$$I_1[C] = \int_{-T/2}^{T/2} C^4(t)dt. \tag{58}$$

Using a standard technique in constrained minimization theory [27], we define the function

$$J = I_1[C] + \lambda_1\left[\int_{-T/2}^{T/2} C^2(t)dt - \mathcal{E}\right], \tag{59}$$

where λ_1 is a Lagrange multiplier, and $\mathcal{E}$ is the energy. Then on substituting (52) in (59) and carrying out

$$(dJ/d\epsilon)|_{\epsilon=0} = 0 \tag{60}$$

the final result becomes

$$\int_{-T/2}^{T/2} C_A(t)[\lambda_1 C_0(t) + 2C_0^3(t)]\,dt = 0. \tag{61}$$

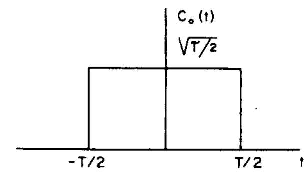

Fig. 5. Optimal signal with no restriction on energy and bandwidth.

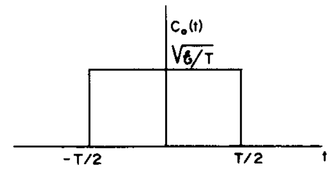

Fig. 6. Optimal signal with energy constraint.

By arguments similar to those of the first case this integral equation requires that

$$\lambda_1 C_0(t) + 2C_0^3(t) = 0, \quad -T/2 \leqslant t \leqslant T/2. \tag{62}$$

We use the given constraint on $\mathcal{E}$ of (57) to evaluate the constant λ_1 as $-2C_0^2(t)$. Finally the solution for optimal signal, found by integrating $\lambda_1 = -2C_0^2(t)$ and substituting for λ_1 into (62) becomes

$$C_0(t) = \sqrt{\frac{\mathcal{E}}{T}}, \quad -T/2 < t < T/2. \tag{63}$$

The optimal signal for this case, once again, is a rectangular signal with infinite mean-square bandwidth and given energy $\mathcal{E}$. It is shown in Fig. 6.

Although the optimal signal shapes of the first and the second case appear to be the same, actually the design concepts are different. For the first case, the signal amplitude depends upon the noise energy whereas the amplitude of the signal in the second case is constrained by the energy $\mathcal{E}$.

The problem of signal optimization is more meaningful if we constrain both the energy and bandwidth of $C(t)$. We define the bandwidth as in (46) and the energy as in (57) and then assume that both are bounded by fixed finite values. We require in addition the end point assumption

$$C(\pm T/2) = 0 \tag{64}$$

in order to avoid discontinuities at the end points.

Then for this case the index J for minimization is

$$J = I[C] + \lambda_1\left[\int_{-T/2}^{T/2} C^2(t)dt - \mathcal{E}\right] + \lambda_2\left[\int_{-T/2}^{T/2} \dot{C}^2(t)dt - W^2\right], \tag{65}$$

where λ_1 and λ_2 are Lagrange multipliers.

The details of the minimization process are similar to the first and the second cases. Using the calculus of variations with the boundary condition (64) we obtain the equation that specifies $C_0(t)$ in the following form

$$\ddot{C}_0(t) - \frac{\lambda_1}{\lambda_2} C_0(t) - \frac{2}{\lambda_2} C_0^3(t) = 0. \tag{66}$$

This is the nonlinear differential equation that is known as Duffing's equation [29, pp. 16-22]. From (66), separating the variables $\dot{C}_0(t)$ and $C_0(t)$, then integrating both sides of the equation yields

$$[\dot{C}_0(t)]^2 - \frac{\lambda_1}{\lambda_2} C_0^2(t) - \frac{1}{\lambda_2} C_0^4(t) = \Omega, \tag{67}$$

where Ω is the constant of integration.

For $\lambda_2 < 0$, a real solution of $C_0(t)$ exists only when $\Omega \geqslant 0$. Observe that at $\dot{C}_0(t) = 0$, there exists the real roots $C_0 = \pm r$ if λ_2 is negative, where the root is

$$r = \sqrt{\left(\frac{\lambda_1^2}{4} - \lambda_2 \Omega\right)^{1/2} - \frac{\lambda_1}{2}}. \tag{68}$$

By changing the variable such that

$$C_0(t) = r \cos \psi(t) \tag{69}$$

(67) can be expressed in the following form

$$dt = \frac{\pm d[\psi(t)]}{\mathcal{H}\sqrt{1 - k^2 \sin^2 \psi(t)}} \tag{70}$$

where

$$\mathcal{H} = \sqrt{\frac{2\Omega}{r^2} + \frac{\lambda_1}{\lambda_2}} \tag{71}$$

and

$$k^2 = \frac{\Omega + \frac{\lambda_1}{\lambda_2} r^2}{2\Omega + \frac{\lambda_1}{\lambda_2}} < 1. \tag{72}$$

Integrating both sides of (70) and applying the boundary condition given in (64), the final result is

$$t = \pm \frac{T}{2} \frac{\mathcal{F}(\psi, k)}{\mathcal{F}\left(\frac{\pi}{2}, k\right)}, \tag{73}$$

where

$$\mathcal{F}(\psi, k) = \int_0^{\psi} \frac{du}{\sqrt{1 - k^2 \sin^2 u}} \tag{74}$$

is an elliptic integral. Equation (73) is the solution for $C_0(t)$ in terms of λ_1 and λ_2. To complete the problem it is necessary to evaluate λ_1 and λ_2 by using the constraint equations (46) and (57). Unfortunately, (73) only gives an implicit solution for ψ. However, in practice, λ_1 and λ_2 numerically can be chosen such that C_0 of (60) meets the required energy and bandwidth constraints using trial and error. Actually, it is not difficult to evaluate limits within which λ_1 and λ_2 must lie [17]. To obtain the optimal signals, a family of curves can be tabulated for the whole range of λ_1 and λ_2. Some typical optimal signals with equal energy but differences in bandwidth and duration are shown in Fig. 7.

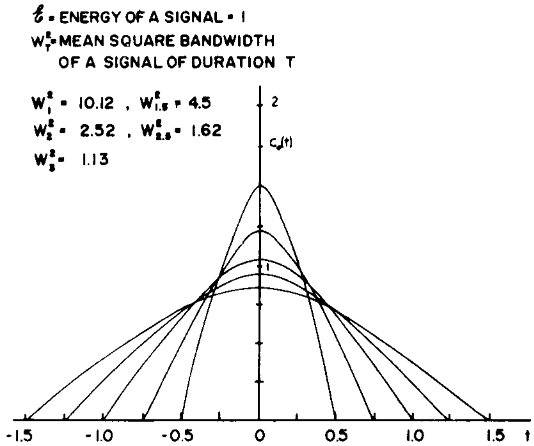

Fig. 7. Optimal signals of equal energy.

It might be interesting to look at the general shape of the optimal signal. The signal is time limited with duration T and symmetrical about $t = 0$. The slope of $C_0(t)$ takes its maximum in absolute value $\sqrt{\Omega}$ at the end points ($t = \pm T/2$) and monotonically decreases reaching the minimum value of zero at $t = 0$.

It should be noticed that the first and second cases of signal design are impractical. This is due to the fact that the optimal signals for both cases are rectangular types that require infinite mean-square bandwidth. Since practical systems have finite systems bandwidth B, the signal bandwidth W^2 in the physical system must be finite as well.

VI. Conclusions and Remarks

We have developed detection schemes using a three-term truncated Gram–Charlier series to represent the probability density function of NGNG noise. It should be understood that the theory can be developed for an arbitrary number of terms in the Gram–Charlier series; physically however, this is only meaningful if the noise is near Gaussian. In such cases, an optimal choice for σ^2 of the normal reference function also exists. Since the Gram–Charlier series contains a Gaussian term, the results obtained when additional additive Gaussian noise is present are contained in the results.

The development of detection theory has given results in the structure of the optimal detector, which, in general, consists of a standard detector for the Gaussian noise case (matched filter) in parallel with nonlinear elements. Since our detection model is restricted to the case of near-Gaussian noise, it may not be adequate for noise that considerably departs from the Gaussian form. Nonetheless, the near-Gaussian model may still provide useful guides to designing receivers operating in general non-Gaussian noise. The important result suggested by the model is that the performance of an optimal receiver in the presence of non-Gaussian interference is sensitive to the signal shape. Here we developed the case of signal design only when the signal-to-noise ratio is not very large, since, as is intuitively true, then the signal shape is relatively insensitive to the system performance when the signal energy is very large. On investigating the optimal signal design it seems that, in theory, "the rectangular signal" might be better than the optimal signal with both energy and bandwidth constraints in the sense of smaller probability of error. However, it is obvious that the rectangular signal never practically exists at the receiver, owing to the band limited nature of any physical communication channel. Actually, the transmitting signal appears distorted due to loss of high frequency components. For this reason, the results from a design subject to energy and bandwidth constraints remain practically optimum.

The system performance was evaluated by observing $P_{e\max}$, which can be calculated through (45). Fig. 4 shows $P_{e\max}$ as calculated for the various cases, and, as for the Gaussian case, the PSK performance is seen to be the best. The results are seen to be comparable to those for the Gaussian case [30, p. 359].

References

[1] W. W. Peterson, T. G. Birdsall, and W. C. Fox, "The theory of signal detectability," *IRE Trans. Inform. Theory*, vol. PGIT-4, pp. 171–212, Sept. 1954.

[2] C. W. Helstrom, *Statistical Theory of Signal Detection*. Hungary: Pergamon, 1968.

[3] D. Middleton, *An Introduction to Statistical Communication Theory*. New York: McGraw-Hill, 1960.

[4] A. J. Viterbi, *Principles of Coherent Communication*. New York: McGraw-Hill, 1966.

[5] W. R. Bennett and J. R. Davey, *Data Transmission*. New York: McGraw-Hill, 1965.

[6] A. B. Bodonyi, "Effects of impulse noise on digital data transmission," *IRE Trans. Commun. Syst.*, vol. CS-9, pp. 355–361, Dec. 1961.

[7] P. A. Bello and R. Esposito, "A new method for calculating probabilities of errors due to impulsive noise," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 368–379, June 1969.

[8] B. Shepelavey, "Non-Gaussian atmospheric noise in binary-data, phase-coherent communication systems," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 280–284, Sept. 1963.

[9] J. H. Halton and A. Spaulding, "Error rates in differentially coherent phase systems in non-Gaussian noise," *IEEE Trans. Commun. Technol.*, vol. COM-14, pp. 594–601, Oct. 1966.

[10] L. R. Halsted, "On binary data transmission error rates due to combinations of Gaussian and impulsive noise," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 428–435, Dec. 1963.

[11] P. A. Bello, "Bounds on the error probability of FSK and PSK receivers due to non-Gaussian noise in fading channels," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 315–326, July 1966.

[12] —, "Error probabilities due to atmospheric noise and flat fading in HF ionospheric communication systems," *IEEE Trans. Commun. Technol.*, vol. COM-13, pp. 266–279, Sept. 1965.

[13] R. E. Ziemer, "Character error probabilities for M-ary signaling in impulsive noise environments," *IEEE Trans. Commun. Technol.*, vol. COM-15, pp. 32–44, Feb. 1967.

[14] S. S. Rappaport and L. Kurz, "An optimal nonlinear detector for digital data transmission through non-Gaussian channels," *IEEE Trans. Commun. Technol.*, vol. COM-14, pp. 266–274, June 1966.

[15] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 1. New York: Hafner, 1958.

[16] H. Cramér, *Mathematical Methods of Statistics*. Princeton, N.J.: Princeton Univ. Press, 1946.

[17] A. Bodharamik, "Optimal signal detection for low frequency data transmission in the presence of Non-Gaussian noise," Ph.D. thesis, Univ. Maryland, College Park, 1971.

[18] "World distribution and characteristics of atmospheric radio noise," Nat. Bur. Stand. (Document submitted to CCIR), Feb. 7, 1964.

[19] R. S. Berkowitz, "Methods of sampling band-limited functions," *Proc. IRE*, vol. 44, pp. 231–235, Feb. 1956.

[20] R. Deutsch, *System Analysis Techniques*. Englewood Cliffs, N.J.: Prentice-Hall, 1969.

[21] J. C. Hancock and P. A. Wintz, *Signal Detection Theory*. New York: McGraw-Hill, 1966.

[22] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1967.

[23] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, vol. 23, pp. 493–507, 1962.

[24] M. E. Hellman and J. Raviv, "Probability of error, equivocation, and the Chernoff bound," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 368–372, July 1970.

[25] D. Blackwell, "Comparison of experiments," in *Proc. 2nd Berkeley Symp. Probability and Statistics*, vol. 1. Berkeley, Calif.: Univ. California Press, 1955, pp. 93–102.

[26] N. Abramson, "Bandwidth and spectra of phase-and-frequency-modulated waves," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 407–414, Dec. 1963.

[27] R. P. Weinstock, *Calculus of Variations*. New York: McGraw-Hill 1952.

[28] H. L. Van Trees, "Applications of state-variable techniques in detection theory," *Proc. IEEE*, vol. 58, pp. 653–669, May 1970.

[29] R. A. Struble, *Nonlinear Differential Equations*. New York: McGraw-Hill, 1962.

[30] H. L. Van Trees, *Detection, Estimation, and Modulation Theory–Part I*. New York: Wiley, 1968.

Adisai Bodharamik was born in Bangkok, Thailand, on April 23, 1940. He received the B.Eng. degree from Chulalongkorn University, Bangkok, Thailand, in 1962, the M.S. degree from the University of Hawaii, Honolulu, in 1967, and the Ph.D. degree from the University of Maryland, College Park, in 1971.

In 1962 he joined the Telephone Organization of Thailand where he engaged in the long-line communication systems. In 1966 he attended the University of Hawaii under an East-West Center Grant. From 1968 to 1971 he was a Graduate Assistant at the Department of Electrical Engineering, University of Maryland. At the present time he is working with the Telephone Organization of Thailand, Bangkok.

Dr. Bodharamik is a member of Sigma Xi.

R. W. Newcomb (S'52–M'56) was born in Glendale, Calif., in June 1933. He received the B.S.E.E. degree from Purdue University, Lafayette, Ind., in 1955, the M.S. degree from Stanford University, Stanford, Calif., in 1957, and the Ph.D. degree from the University of California, Berkeley, in 1960.

In 1957, he was a Research Intern at the Stanford Research Institute, Stanford, and in 1960 he was a Teaching Associate at the University of California, Berkeley. In 1960 he joined the professorial staff at Stanford University. He spent the academic year 1963-1964 as Visiting Professor at the University of Newcastle, New South Wales, Australia, and the first half of the academic year 1967–1968 as Professeur Invité at the Université Catholique de Louvain, Belgium. In 1970 he joined the Faculty of Electrical Engineering, University of Maryland, College Park. He presently acts as the External Examiner in Electrical Engineering for Lagos University, and is cochairman of several international meetings. He is the network series Editor for Prentice-Hall, and author of *Linear Multiport Synthesis*, *Concepts of Linear Systems and Controls*, *Active Integrated Circuit Synthesis*, and *Network Theory: the State Space Approach*. His research concentrates in the areas of microsystems and generalized networks.

John B. Moore (S'66–M'68) received the B.S. and M.S. degrees in electrical engineering from the University of Queensland, Brisbane, Australia, and the Ph.D. degree in electrical engineering from the University of Santa Clara, Santa Clara, Calif., in 1966.

Since 1967, he has been a faculty member of the Department of Electrical Engineering at the University of Newcastle, New South Wales, Australia, where he is currently an Associate Professor. He has also been a Visiting Faculty Member at the University of Santa Clara and the University of Maryland, College Park. His research interests are in control and communication systems.