Optimum Pulse Amplitude Modulation Part I: Transmitter-Receiver Design and Bounds from Information Theory

TOBY BERGER, MEMBER, IEEE, AND DONALD W. TUFTS, MEMBER, IEEE

Abstract—Intersymbol interference and additive noise are two common sources of distortion in data transmission systems. For pulse amplitude modulation (PAM) communication links, the combination of transmitter waveform and linear receiver that minimizes the mean-squared error arising from these sources is determined. An extension to include the effects of timing jitter is performed in a companion paper.

Performance characteristics of the optimal PAM systems, showing the mean-squared error versus the signal-to-noise ratio, are determined explicitly for several examples. These characteristics are compared both with those of certain suboptimal systems and with the optimal performance theoretically attainable (OPTA), derived by combining Shannon's concepts of the capacity of a channel and the rate distortion function of a source.

The optimal PAM systems are seen to perform very close to the OPTA for low signal-to-noise ratios. For high signal-to-noise ratios, however, the mean-squared error of optimal PAM systems decreases as the reciprocal of the signal-to-noise ratio, but the OPTA decreases more rapidly, except for band-limited channels. The performance of PAM systems can be improved at high signal-to-noise ratios by coding techniques. One such technique, called Shannon-Cantor coding, is discussed briefly.

I. Introduction

WE DISCUSS the joint optimization of transmitted pulse shape and linear receiver in pulse amplitude modulation (PAM) using a mean-square distortion measure. The receiver is required to estimate sequentially the elements of a random transmitted data sequence. The sources of error in these estimates considered are independent additive noise, intersymbol interference, and timing jitter, i.e., inexact synchronization. The intersymbol interference stems from overlapping of received pulses when data are transmitted at a high rate through a linear transmission medium. In addition to its intrinsic utility, the mean-squared-error criterion also has been shown to lead to nearly optimal digital PAM data transmission links, judged on the basis of error probability.[1]

Mean-squared error versus signal-to-noise ratio performance characteristics of optimal systems are determined for several examples. Certain subtle and challenging difficulties arise in solving our necessary equations. First, these equations are nonlinear, and second, the equations and an associated power constraint can be satisfied in an infinite number of ways. These difficulties are completely resolved in the case of deterministic sampling. For the timing jitter case, we present in a companion paper[2] the best solution in certain classes of examples, but a complete, general, and rigorous development has eluded us.

An optimal PAM system can be interpreted as the ideal combination of linear encoder and decoder for the given noisy channel with memory. It is fruitful to compare its performance with the optimal performance theoretically attainable (OPTA) in order to determine the additional improvement that could be realized with more complex coding techniques. The OPTA is computed using the rate distortion function results of Shannon[3] and, in particular, the extensions of Pinsker[4] and Dobrushin.[5] We also compare the OPTA and optimal PAM with optimal PAM subject to the additional, practically useful constraint that transmitted pulses are to be confined to a single time slot. (The received pulses will still overlap in general.) The latter results are taken from the work of Shnidman and Tufts.[6],[7]

The joint optimization of transmitter and receiver in PAM, subject only to a constraint on average transmitted power or average pulse energy, has been considered previously by Tufts,[8],[9] Berger,[10] Berger and Tufts,[11] Tsybakov,[12] and Smith.[13] The transmitter and receiver conditions [cf. (11) and (12)] were first derived in Tufts[8] and later considered in more detail in Tufts[9] and Berger.[10] The results presented below constitute a refinement of our technical report.[11] Tsybakov restricts attention to time-discrete channels in which the question of timing jitter does not arise. However, he investigates this case in considerable depth. Tsybakov's results are closely related to our continuous-channel, deterministic-sampling results. Smith recently considered the deterministic-sampling problem, obtaining results which partially overlap those of Tufts and Berger[9]–[11] and parts of Sections VI and VII. Smith's approach differs markedly from ours, an important consequence of the difference being that we are able to prove that the proposed system design minimizes the mean-squared error, whereas Smith conjectures that it will be minimized "to a close approximation" (Smith,[13] p. 2374).

The optimal PAM analysis performed below also is related to the work of Shnidman and Tufts.[6],[7] However,

Manuscript received January 4, 1966; revised August 1, 1966.
T. Berger is with the Advanced Development Laboratory, Raytheon Company, Wayland, Mass.
D. W. Tufts is with the Division of Engineering and Applied Physics, Harvard University, Cambridge, Mass.

their results are obtained under the constraints that a) the duration of the transmitted pulse is limited, and b) the transmitted data sequence has finite length. The character of the results is substantially different when these constraints are removed.

Our results can be used directly to evaluate mean-squared-error performance of PAM telemetry systems. Performance limits for such telemetry links are obtained by optimizing the PAM systems, and these limits are compared with the OPTA bounds. Indirectly, our results can be used to design digital PAM data transmission links according to an error-probability criterion. The theoretical and practical attractions of such links for data transmission and the development of the relation between mean-squared-error and probability-of-error design methods are beyond the scope of this paper. The reader is referred to the recent publications of Aaron and Tufts,[1] Bennett and Davey,[14] DiToro,[15] George and Coll,[21] Franks,[22] and Lucky.[16]

Sections II through V are devoted to the model, the notation, a discussion of the problem, and the basic equations. We present a detailed derivation of the jitter-free results in Sections VI and VII, and Appendix II. In Section VIII, the OPTA is derived and examples are presented. The merits of a particular nonlinear coding technique, called Shannon-Cantor coding, are discussed in Section IX.

II. Model and Notation

Our model for a PAM link is depicted in Fig. 1.

The message to be transmitted is assumed to be a random sequence of real numbers, $\{a_i\}$, which has the known discrete stationary autocorrelation function $\{m_i\}$, $i = 0, \pm 1, \pm 2, \cdots$. Such a message might arise, for example, from the sampling, quantizing, and coding of a continuous signal as is done in pulse-code modulation, or simply from the sampling of a random process.

The message sequence is used to amplitude modulate a train of identically-shaped waveforms in such a way that the overall transmitter output can be represented by the series

$$\sum_{k=-\infty}^{\infty} a_k s(t - kT), \tag{1}$$

where $s(t)$ is the transmitter waveform we wish to specify optimally, and $1/T$ is the constant data rate in message elements per second.

The channel is represented as a linear time-invariant network with a known impulse response $h(t)$ followed by an additive noise generator. The noise $n_g(t)$ has a zero mean and known autocorrelation function $n(\tau)$, and it is uncorrelated with the message. Thus, the channel output, i.e., the received waveform, can be expressed as

$$n_g(t) + \sum_{k=-\infty}^{\infty} a_k r(t - kT), \tag{2}$$

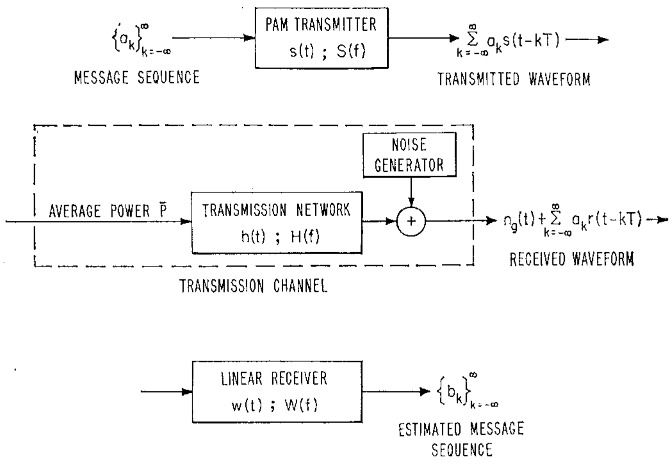


Fig. 1. PAM data transmission system model.

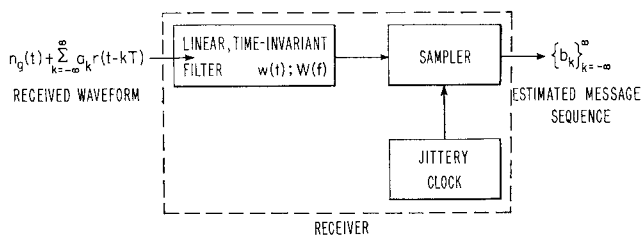


Fig. 2. Receiver block diagram.

where $r(t)$ is the response of the channel network to an individual input waveform $s(t)$. That is, $r(t)$ is defined by the convolution equation

$$r(t) = \int_{-\infty}^{\infty} s(u)h(t - u)\, du. \tag{3}$$

The linear receiver sequentially calculates an approximation to each element in the transmitted message sequence by performing a weighted integration of the received waveform. For example, b_k, the estimate of message element a_k, is calculated as follows:

$$b_k = \int_{-\infty}^{\infty} w(kT + \delta_k - t)\left[n_g(t) + \sum_{n=-\infty}^{\infty} a_n r(t - nT)\right] dt. \tag{4}$$

Here $w(t)$ is the impulse response of the linear, time-invariant receiver filter we wish to specify optimally, and $\{\delta_k\}$ is a sequence of random variables used to represent inexact synchronization, or timing jitter, between transmitter and receiver. For modeling purposes, this timing jitter has been ascribed to an imperfect clock that controls the sampler. In order to stay within the framework of time-invariant receivers, it is assumed that the elements of $\{\delta_k\}$ are statistically independent and identically distributed according to the known probability density $p(t)$. The physical significance of this assumption is discussed in Tufts and Berger.[2] Figure 2 shows a block diagram of the receiver.

Fourier transform pairs are denoted by upper and lower case letters. For example,

$$P(f) = \int_{-\infty}^{\infty} p(t)e^{-i2\pi ft}\,dt \quad \text{and} \quad p(t) = \int_{-\infty}^{\infty} P(f)e^{i2\pi ft}\,df.$$

A spectrum $M(f)$ is associated with the discrete message autocorrelation function in the following way:

$$M(f) = \sum_{k=-\infty}^{\infty} m_k e^{-i2\pi kTf} = m_0 + 2\sum_{k=1}^{\infty} m_k \cos 2\pi kTf. \quad (5a)$$

Thus,

$$m_k = T\int_{-1/2T}^{1/2T} M(f)e^{i2\pi kTf}\,df. \quad (5b)$$

Since the message autocorrelation matrix with entries $\{m_i\}$ is non-negative definite, the theory of Toeplitz forms[17] guarantees that the even, periodic function $M(f)$ is also non-negative.

III. The Performance Criterion

Our criterion for comparing system performances is that the average of $(b_k - a_k)^2$, with respect to the message, noise, and jitter distributions, should be minimized. This average value is called the interference, or distortion, and is denoted by D. In Appendix I, it is shown that D can be expressed in the following ways:

$$\begin{aligned} D &= m_0 \\ &+ \int_{-\infty}^{\infty} \sum_{i=-\infty}^{\infty} m_i v(t-iT)\left[v(t)\sum_{k=-\infty}^{\infty} p(t-kT) - 2p(t)\right] dt \\ &+ \int_{-\infty}^{\infty} w(t)\left[\int_{-\infty}^{\infty} w(\tau)n(t-\tau)\,d\tau\right] dt \end{aligned} \quad (6)$$

and

$$\begin{aligned} D &= m_0 \\ &+ \int_{-\infty}^{\infty} M(f)V^*(f)\left[\frac{1}{T}\sum_{k=-\infty}^{\infty} P(k/T)V\left(f-\frac{k}{T}\right) - 2P(f)\right] df \\ &+ \int_{-\infty}^{\infty} N(f)\,|W(f)|^2\,df, \end{aligned} \quad (7)$$

where "*" denotes a complex conjugate and $v(t)$ is defined by either of the following two equations:

$$v(t) = \int_{-\infty}^{\infty} r(x)w(t-x)\,dx \quad (8a)$$

or

$$V(f) = R(f)W(f) = S(f)H(f)W(f). \quad (8b)$$

The following useful facts can be deduced from (7). First, the distortion depends on the phase angle of $S(f)W(f)$, but is independent of the way in which phase is distributed between $S(f)$ and $W(f)$. This is because the first and third terms of (7) are independent of phase angles, while the second term depends only on the phase angle of $V(f)$. Accordingly, the system designer always may take advantage of a certain degree of flexibility in the assignment of phase to the transmitter waveform and the receiver filter. Second, if $p(t)$ is an even function, the distortion can only be reduced if $v(t)$ is replaced by its even part, or equivalently, if $V(f)$ is replaced by its real part. It is shown in Berger[10] that each term in (7) either decreases or remains constant upon this substitution.

Inspection of either (6) or (7) leads to the conclusion that unless some limitation is placed on the transmitter waveform, D can be made arbitrarily small. In this regard, a constraint is imposed on the average transmitted power. This requirement is represented as a restriction on the ensemble average of the time average of the squared channel input, namely

$$E\left\{(1/T)\int_{-1/2T}^{1/2T}\left[\sum_{k=-\infty}^{\infty} a_k s(t-kT)\right]^2 dt\right\} = \bar{P} \quad (9)$$

where E denotes the expectation operator of the message ensemble. The constraint of (9) may be expressed more compactly in the equivalent form

$$(1/T)\int_{-\infty}^{\infty} M(f)\,|S(f)|^2\,df = \bar{P}. \quad (10)$$

IV. Conditions for Minimum Distortion

Using a well-known variational argument[9],[18] it can be shown that

$$\begin{aligned} &M(f)H^*(f)W^*(f)(1/T) \\ &\quad\cdot\sum_{k=-\infty}^{\infty} P(k/T)S\left(f-\frac{k}{T}\right)H\left(f-\frac{k}{T}\right)W\left(f-\frac{k}{T}\right) \\ &\quad + M(f)S(f)/\lambda = M(f)H^*(f)W^*(f)P(f) \end{aligned} \quad (11)$$

is a necessary and sufficient condition for a transmitter waveform $S(f)$ that minimizes D for a fixed receiver filter, $W(f)$, subject to the average power constraint. The constant λ is a Lagrange multiplier that must be chosen so that the solution of (11) for $S(f)$ satisfies (10). Since there will be occasion to cancel the common factor $M(f)$ from both sides of (11), it will be assumed that $M(f)$ is positive almost everywhere; this assumption is satisfied by most message-power spectral densities of practical interest.

Similarly, it can be shown that

$$\begin{aligned} &M(f)H^*(f)S^*(f)(1/T) \\ &\quad\cdot\sum_{k=-\infty}^{\infty} P(k/T)S(f-k/T)H(f-k/T)W(f-k/T) \\ &\quad + N(f)W(f) = M(f)H^*(f)S^*(f)P(f) \end{aligned} \quad (12)$$

is a necessary and sufficient condition for a receiver filter $W(f)$ which minimizes D for a fixed transmitter waveform $S(f)$. In both (11) and (12), the functions $M(f)$, $H(f)$, $N(f)$, and $P(f)$ are regarded as fixed and known. In general, the receiver filter specified by (12) will not be causal, and practical approximations to the optimal PAM system will exhibit a coding delay.

It is of interest to calculate the value of distortion that results when $W(f)$ is a solution of (12). Upon multiplying (12) by $W^*(f)$, substituting the result into (7), and recalling the definition of $V(f)$, one finds that

$$\hat{D} = m_0 - \int_{-\infty}^{\infty} M(f)V^*(f)P(f)\,df. \tag{13}$$

The symbol $\hat{D}$ has been introduced in order to emphasize that (13) holds only when $W(f)$ is optimal for $S(f)$ in the sense of (12).

V. The Joint Optimization Problem

The sections that follow are devoted to the important design problem of simultaneously optimizing both the transmitter and the receiver. A transmitter-receiver combination (S, W), henceforth called a system, will be deemed optimal for a specified value $\bar{P}$ of average transmitted power if the distortion D which it achieves does not exceed that of any other system using this value of average transmitted power. The curve of minimum D vs. $\bar{P}$ is clearly monotonic nonincreasing by definition. In most practical cases, this curve will be strictly monotonic decreasing. Accordingly, an alternative characterization of optimality, leading to the same optimal systems, would be the achievement of a specified distortion level with the least possible expenditure of average transmitter power.

It is clear that (11) and (12) constitute a set of necessary conditions that must be satisfied by any system (S, W) that is to be a solution to the joint optimization problem. It is essential to stress, however, that simultaneous satisfaction of (11) and (12) is not a sufficient condition for an optimal system. In fact, it generally is possible to simultaneously satisfy both the necessary equations and the average power constraint in an infinite number of ways.[1] The problem then becomes that of determining which of these solutions yields the least distortion. It is this lack of a unique solution of the necessary equations for a specified value of average power that necessitates the intricacy of certain arguments employed below.

Note that (11) and (12) together imply that the optimal system must be such that, at any frequency f_0, $S(f_0) = 0$ if and only if $W(f_0) = 0$. Accordingly, if (11) is multiplied by $S^*(f)$ and (12) by $W^*(f)$, the resulting equations will have the same joint solutions as do (11) and (12) themselves. Moreover, when one of the two equations which results is subtracted from the other it follows that

$$M(f)\,|S(f)|^2 = \lambda N(f)\,|W(f)|^2. \tag{14}$$

Integration of (14) over all frequencies and reference to (10) demonstrate that, for any system satisfying (11) and (12), the quantity λ/T is a non-negative real number representing the ratio of the average transmitted power to the average output noise power.

[1] In this regard, suppose that (S_1, W_1) satisfies the necessary equations when the Lagrange multiplier is assigned the value λ_1, and let the resulting average transmitted power and distortion be denoted by $\bar{P}_1$ and D_1, respectively. A system (S_2, W_2), differing from (S_1, W_1) only in that for all f in some arbitrary set F the relation $S(f - k/T) = W(f - k/T) = 0$ holds for all integers k, is another solution of the necessary equations for $\lambda = \lambda_1$. It follows from (10) that $\bar{P}_2 < \bar{P}_1$ and from (13) that $D_2 > D_1$; hence, (S_1, W_1) and (S_2, W_2) cannot be ranked on the basis of optimality. However, (S_1, W_1) can be compared meaningfully with still other solutions of the necessary equations obtained for values of $\lambda \neq \lambda_1$ and adjusted in bandwidth, in accordance with the procedure described above for obtaining (S_2, W_2) from (S_1, W_1), until their average power equals $\bar{P}_1$. For further details see Section VII and Appendix II.

VI. The Jitter-Free Case

In this paper we concentrate on the case of deterministic sampling, i.e., the jitter-free case.

In the absence of jitter, $p(t) = \delta(t)$ and $P(f) = 1$. Therefore, (11) and (12), respectively, simplify to

$$M(f)H^*(f)W^*(f)(1/T) \cdot \sum_{k=-\infty}^{\infty} S(f - k/T)H(f - k/T)W(f - k/T) + M(f)S(f)/\lambda = M(f)H^*(f)W^*(f) \tag{15}$$

and

$$M(f)H^*(f)S^*(f)(1/T) \cdot \sum_{k=-\infty}^{\infty} S(f - k/T)H(f - k/T)W(f - k/T) + N(f)W(f) = M(f)H^*(f)S^*(f). \tag{16}$$

It is now demonstrated that the receiver equation (16) always has a solution $W(f)$, composed of a generalized matched filter $R^*(f)/N(f)$ in cascade with a linear system that has a periodic transfer function $Y(f)$, the period of which is equal to the data rate $1/T$. Substitution of the assumed solution

$$W(f) = R^*(f)Y(f)/N(f) \tag{17}$$

into (16) and use of the fact that $Y(f - k/T) = Y(f)$ yields

$$M(f)R^*(f)Y(f)\left[(1/T)\sum_{k=-\infty}^{\infty} |R(f - k/T)|^2/N(f - k/T)\right] + R^*(f)Y(f) = M(f)R^*(f). \tag{18}$$

It is clear that (16) has been satisfied at all frequencies for which the given function $R(f)$ vanishes, regardless of the value of $Y(f)$. For $R(f) \neq 0$, (18) is easily solved for the required $Y(f)$, namely

$$Y(f) = \frac{M(f)}{M(f)L(f) + 1}, \tag{19}$$

where

$$L(f) = \frac{1}{T}\sum_{k=-\infty}^{\infty} |R(f - k/T)|^2/N(f - k/T). \tag{20}$$

Substitution of $f + l/T$ for f in (20) and change of summation index to $n = k - l$ show that $L(f)$ is periodic $(1/T)$. Since $M(f)$ is periodic $(1/T)$ by definition, it follows from (19) that the function $Y(f)$ is indeed periodic $(1/T)$ as was assumed, which completes the demonstration. A block diagram of the optimal jitter-free receiver specified by (17), in which $Y(f)$ is shown realized as a feedback system, is depicted in Fig. 3.

Equation (13) may be used to evaluate the performance of the optimal receiver. Setting $P(f) = 1$, because we

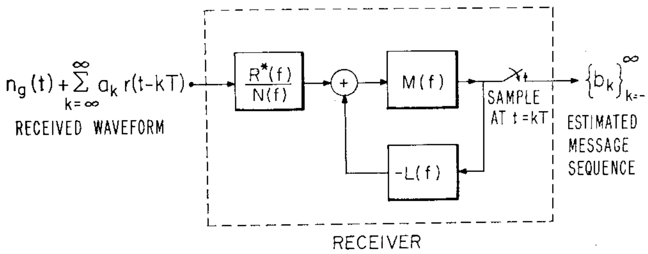


Fig. 3. Block diagram of the optimal jitter-free receiver.

are examining the jitter-free case, using (17) and noting that $Y(f)$ is real so that $V^*(f) = V(f) = |R(f)|^2\, Y(f)/N(f)$, we may cast (13) in the form

$$\hat{D} = m_0 - \int_{-\infty}^{\infty} M(f)\, |R(f)|^2\, [Y(f)/N(f)]\, df. \tag{21}$$

If the integral in (21) is represented as an infinite sum of integrals over successive intervals of length $1/T$, and the variables of integration are changed so that each integral in the sum is taken over the range $[-\frac{1}{2}T, \frac{1}{2}T]$, there results

$$\hat{D} = m_0 - \sum_{k=-\infty}^{\infty} \int_{-1/2T}^{1/2T} M(f+k/T)\, |R(f+k/T)|^2 \cdot [Y(f+k/T)/N(f+k/T)]\, df. \tag{22}$$

Commuting the operations of summation and integration, employing the periodicity of both $M(f)$ and $Y(f)$, and recalling the definition of $L(f)$, we obtain

$$\hat{D} = m_0 - T \int_{-1/2T}^{1/2T} M(f) L(f) Y(f)\, df. \tag{23}$$

If (5b) is used to represent m_0 and (19) to represent $Y(f)$, some algebraic manipulations then yield

$$\hat{D} = T \int_{-1/2T}^{1/2T} \frac{M(f)}{M(f)L(f)+1}\, df, \tag{24}$$

or simply

$$\hat{D} = T \int_{-1/2T}^{1/2T} Y(f)\, df. \tag{25}$$

It has not been proved that (17) is the unique solution of (16). However, the fact that (16) is a sufficient as well as a necessary condition for minimizing D for a given $S(f)$ assures us that no loss of generality results from assuming a solution of the form (17). In particular, whatever the optimal $S(f)$ may be, the $W(f)$ that corresponds to it may be assumed without loss of generality to be of the form (17). When (17) is substituted into (14), which must be satisfied for joint optimization, the result is

$$M(f)\, |S(f)|^2 = \lambda\, |S(f)|^2\, |H(f)|^2\, Y^2(f)/N(f). \tag{26}$$

It may be concluded directly from (26) that at every frequency f, a jointly optimal system must satisfy either

$$Y^2(f) = M(f)N(f)/\lambda\, |H(f)|^2 \tag{27a}$$

or

$$S(f) = W(f) = 0. \tag{27b}$$

The function $Y^2(f)$ must be periodic, but the given function $M(f)N(f)/\lambda\, |H(f)|^2$ is usually aperiodic. Accordingly, if (27a) is satisfied at some frequency f_0, it generally cannot be satisfied at any other frequency of the form $f = f_0 + k/T$, $k \neq 0$. Therefore, one must resort to (27b) and set $S(f_0 + k/T) = W(f_0 + k/T) = 0$ for all $k \neq 0$.

The principal conclusion to be drawn from the above analysis is that a jointly optimal jitter-free system generally is band-limited to a frequency set of total measure $1/T$, of which no two points coincide under translation by k/T for any $k \neq 0$. Such a collection of frequencies henceforth will be referred to as a Nyquist set. It is clear that there are infinitely many ways to choose a particular value of λ and a particular Nyquist set or part thereof over which to satisfy (27a), so that the resulting solution has the prescribed average transmitted power $\bar{P}$. The task is to determine which of these solutions yields the smallest distortion. Toward this end a further restriction on the frequencies at which a nonzero solution may be employed is now derived.

If (27a) is satisfied at the frequency f, and, hence, (27b) is satisfied at $f + k/T$ for all $k \neq 0$, then (15) and (16) reduce to

$$(MH^*W^*)(1/T)(SHW) + MS/\lambda = MH^*W^* \tag{28}$$

and

$$(MH^*S^*)(1/T)(SHW) + NW = MH^*S^*, \tag{29}$$

where it is to be understood that all functions involved are evaluated at the frequency f under consideration. Solving (28) for the transmitter waveform yields

$$S = \lambda T H^* W^* (T + \lambda\, |HW|^2)^{-1}. \tag{30}$$

When (30) is substituted into (29), a quadratic equation in $|W|^2$ results, the solutions of which are

$$|W|^2 = -\frac{T}{\lambda\, |H|^2} \pm \frac{T}{|H|} \sqrt{\frac{M}{\lambda N}}. \tag{31}$$

The choice of the minus sign would yield a negative $|W|^2$; the choice of the plus sign yields a positive $|W|^2$ if, and only if, the following inequality is satisfied at the frequency f:

$$1 > N(f)/\lambda M(f)\, |H(f)|^2. \tag{32}$$

Thus, for any f at which (32) is not satisfied, the optimal system must have $S(f) = W(f) = 0$. Also, note in passing that multiplication of (30) by S^* leads to the conclusion that the phases of S and W must be assigned so that V is real and positive.

VII. The Optimal Solution of the Necessary Conditions in the Jitter-Free Case

Since the value of distortion associated with any joint solution of (15) and (16) is given by (25), it is clear that the best solution is that which minimizes $Y(f)$ for each f in the basic interval $(-1/2T, 1/2T)$. For each such f, there are two alternatives: 1) satisfy (27a) at the frequency $f + k/T$ for some value of k and satisfy

(27b) at $f + j/T$ for all $j \neq k$; or 2) satisfy (27b) for all frequencies of the form $f + k/T$. Clearly, it is better to satisfy (27a) for some value of k than for none at all. This is because $S(f + k/T) = 0$ for all k implies that $L(f) = 0$ and, therefore, that $Y(f) = M(f)$; but, if $S(f + k/T) \neq 0$, then

$$L(f) = |S(f + k/T)H(f + k/T)|^2/TN(f + k/T) > 0$$

and

$$Y(f) = \frac{M(f)}{M(f)L(f) + 1} < M(f).$$

In order to see which value of k to select for a particular f, one can rewrite (27a) as follows, using the periodicity of $Y(f)$ and $M(f)$:

$$\begin{aligned} Y^2(f) &= Y^2(f + k/T) \\ &= M(f)N(f + k/T)/\lambda\ |H(f + k/T)|^2, \\ -1/2T &\leqq f \leqq 1/2T. \end{aligned} \tag{33}$$

It follows that $Y(f)$ will be minimized for a given value of λ if (33) is satisfied at the value of k for which $N(f + k/T)/|H(f + k/T)|^2$ is smallest.

In summary, the procedure recommended for determining an optimal system consists of the following steps:

1) For every f in the basic Nyquist set $I_0 = [-1/2T, 1/2T)$, determine $\min_k N(f+k/T)/|H(f+k/T)|^2$, and let k_f denote the minimizing value of k.[2]

Define the Nyquist set I_1 according to the prescription

$$I_1 = \left\{f : f = g + \frac{k_g}{T} \quad \text{for some} \quad g \,\varepsilon\, I_0\right\}. \tag{34}$$

2) Choose a positive number λ, and define the frequency set I_λ according to the prescription

$$I_\lambda = \{f : f \,\varepsilon\, I_1 \quad \text{and} \quad 1 > N(f)/\lambda M(f)\ |H(f)|^2\}. \tag{35}$$

3) For every $f \,\varepsilon\, I_\lambda$, set $S(f) = W(f) = 0$. Then, for every $f \,\varepsilon\, I_\lambda$, solve (15) and (16) for $S(f)$ and $W(f)$. Since there is only one nonzero term in the summation in (15) and (16), they simplify to (28) and (29). Accordingly

$$|S(f)|^2 = -\frac{TN(f)}{M(f)\ |H(f)|^2} + \frac{T}{|H(f)|}\sqrt{\frac{\lambda N(f)}{M(f)}} : f \,\varepsilon\, I_\lambda \tag{36a}$$

and

$$|W(f)|^2 = -\frac{T}{\lambda\ |H|^2} + \frac{T}{|H(f)|}\sqrt{\frac{M(f)}{\lambda N(f)}} : f \,\varepsilon\, I_\lambda. \tag{36b}$$

Assign phases to $S(f)$ and $W(f)$ so that $V(f) = S(f)H(f)W(f)$ is real and positive, in accordance with the remarks at the end of Section VI.

[2] The existence of $\min_k N(f + k/T)/|H(f + k/T)|^2$ presupposes that the function $N(f)/|H(f)|^2$ is well-behaved in some sense; an assumption of piecewise continuity, usually fulfilled in practice, suffices for the purpose. Furthermore, k_f generally will be unique for almost all $f \,\varepsilon\, I_0$; an important case in which this is not so, that of the ideal bandpass channel with white noise, will be discussed in the examples.

4) Evaluate the resulting average channel input power.

$$\begin{aligned} \bar{P}(\lambda) &= \frac{1}{T}\int_{I_\lambda} M(f)\ |S(f)|^2\, df \\ &= \int_{I_\lambda} [\sqrt{\lambda M(f)N(f)/|H(f)|^2} - N(f)/|H(f)|^2]\, df, \end{aligned} \tag{37}$$

and the resulting minimum distortion

$$\begin{aligned} D_{\min}(\lambda) &= T\int_{I_1} Y(f)\, df \\ &= m_0 - T\int_{I_\lambda} [M(f) - \sqrt{M(f)N(f)/\lambda\ |H(f)|^2}]\, df. \end{aligned} \tag{38}$$

5) Repeat steps 2) through 4) using different values of λ, thereby obtaining the curves $\bar{P}(\lambda)$ and $D_{\min}(\lambda)$ from which the optimal performance characteristic of $D_{\min}$ vs. $\bar{P}$ may be generated parametrically.

Actually, it has not been demonstrated yet that the systems obtained by the above procedure are truly optimal. From among all solutions of the necessary equations for the value of λ chosen in step 2), the recommended procedure determines the one that achieves the least distortion. However, an optimal system has been defined as one that minimizes D for fixed $\bar{P}$, not for fixed λ. Accordingly, it remains to be shown that no solution of the necessary equations, regardless of the value of the Lagrange multiplier used in obtaining it, possesses both an average power equal to $\bar{P}(\lambda)$ of (37) and a distortion less than $D_{\min}(\lambda)$ of (38). This is established in Appendix II.

VIII. OPTA and Examples

The performance characteristics of optimal jitter-free PAM systems are presented below for several practical examples. These characteristics have been drawn using the dimensionless coordinates D/m_0 and $\bar{E}/N_0$. The former is the mean-squared error incurred per unit variance of an individual message element; the latter is the ratio of the average transmitted energy available per message element, namely $\bar{E} \equiv \bar{P}T$, to the spectral density of the noise, which happens to be white in the particular examples treated. Also displayed on the same coordinates for purposes of comparison are both the best performance that can be achieved by any PAM system employing an $s(t)$ time limited to an interval of length T or less,[6],[7] and the OPTA.

The OPTA is derived by appealing to Shannon's powerful concepts of the information capacity of a communication channel[19] and the rate distortion function of an information source with respect to a fidelity criterion.[3] Specifically, the OPTA is the greatest lower bound to the performance characteristic of any practical communication system whatsoever (not necessarily PAM) which could be designed for transmitting the given source over the given channel, and reproducing it faithfully with respect to the given fidelity criterion. In determining the OPTA, it has been assumed that both the message and the channel noise are Gaussian with the prescribed second-order statistics $M(f)$ and $N(f)$; this has been shown to be the "least favorable distribution" for the higher-order statistics in the sense that the resulting source is the most difficult to reproduce with respect to the criterion under consideration[4],[19] and the resulting channel capacity is the smallest. Since the optimal PAM systems, which were designed solely on the basis of second-order statistics, perform in this least favorable case equally as well as for any other choice of the higher-order statistics, it seems only fair to make the comparison in this manner.

The OPTA is determined as follows. First, the channel capacity is found as a function of $\bar{P}$ by appealing to the recent treatment by Holsinger[20] of certain general formulas given originally by Shannon.[23] For example, it can be shown[10] that the capacity of both the RC channel and the RLC channel shown in Fig. 4 is

$$C(\bar{P}) = \alpha/\pi[\sqrt[3]{3\pi\bar{P}/2\alpha N_0} - \arctan(\sqrt[3]{3\pi\bar{P}/2\alpha N_0})] \quad \text{nats per second} \tag{39}$$

for white Gaussian noise, or equivalently,

$$C(\bar{E}/N_0) = \alpha T/\pi[\sqrt[3]{3\pi\bar{E}/2\alpha TN_0} - \arctan(\sqrt[3]{3\pi\bar{E}/2\alpha TN_0})] \text{ nats per message element.} \tag{40}$$

The quantity α is defined for each channel in Fig. 4. Note that the capacity of the RLC channel does not depend on the resonance frequency.

Next, the rate-distortion function $R(D)$ of the Gaussian message sequence with respect to the prescribed fidelity criterion is determined. The principal significance of Shannon's rate-distortion function resides in his proof[3] that $R(D^*)$ is the amount of channel capacity that is both necessary and sufficient for reproducing the source in question with distortion D^* or less as measured by the fidelity criterion. For an uncorrelated Gaussian message sequence and a mean-squared-error criterion, Shannon[3] has shown that

$$R(D) = -\tfrac{1}{2}\ln(D/m_0) \quad \text{nats per message element;} \quad 0 < D/m_0 \leqq 1. \tag{41}$$

The extension to Gaussian message sequences possessing arbitrary stationary correlation was first performed by Pinsker[4] and later generalized still further by Dobrushin and Tsybakov.[5]

It follows that the least value D of distortion that can possibly be achieved with a signal-to-noise ratio of $\bar{E}/N_0$ is that which satisfies $R(D) = C(\bar{E}/N_0)$. Since the functions $C(\bar{E}/N_0)$ and $R(D)$ are monotonic increasing and decreasing, respectively, the resulting OPTA curve will be monotonic decreasing in $\bar{E}/N_0$. For example, with an RC or RLC white-noise channel and uncorrelated data, (40) and (41) may be combined to yield

$$\text{OPTA} = \exp\{-2\alpha T/\pi[\sqrt[3]{3\pi\bar{E}/2\alpha TN_0} - \arctan(\sqrt[3]{3\pi\bar{E}/2\alpha TN_0})]\} \tag{42}$$

as depicted in Figs. 5(a) and 5(c). For more complicated examples, it usually is not possible to determine an explicit expression for the OPTA. In general, both $C(\bar{P})$ and $R(D)$ themselves can be determined only parametrically; the OPTA then may be generated from them by a procedure analogous to steps 4) and 5) of Section VII. The details of this procedure for the examples considered below appear in Berger.[10]

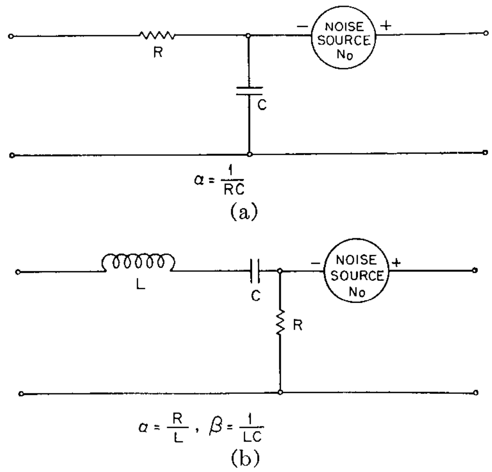


Fig. 4. (a) The RC channel; (b) The RLC channel.

In order to approach the OPTA in practice, it generally is necessary to use rather complicated coding procedures both for the source and for the channel. The optimal PAM systems, which may be construed as the ideal combinations of infinite memory linear encoders and decoders, are seen to perform very close to the OPTA for low signal-to-noise ratios. For high signal-to-noise ratios, however, the mean-squared error of optimal PAM systems decreases as the reciprocal of $\bar{E}/N_0$, whereas the OPTA decreases exponentially in most instances. Accordingly, there generally exists a potential for realizing significant improvement at high signal-to-noise ratios by resorting to more complex coding techniques.

Example 1: RC Channel, White Noise, Uncorrelated Data

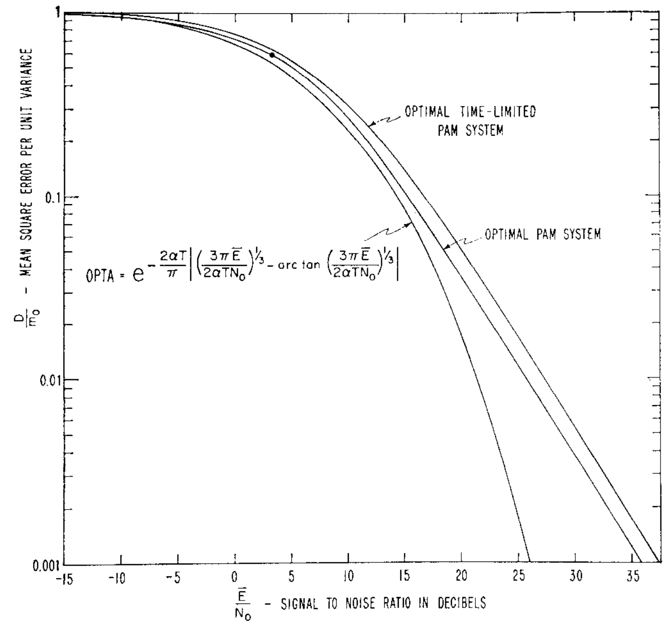
The channel block diagram is given in Fig. 4(a). Since

$$N(f)/|H(f)|^2 = N_0[1 + (2\pi f/\alpha)^2] \tag{43}$$

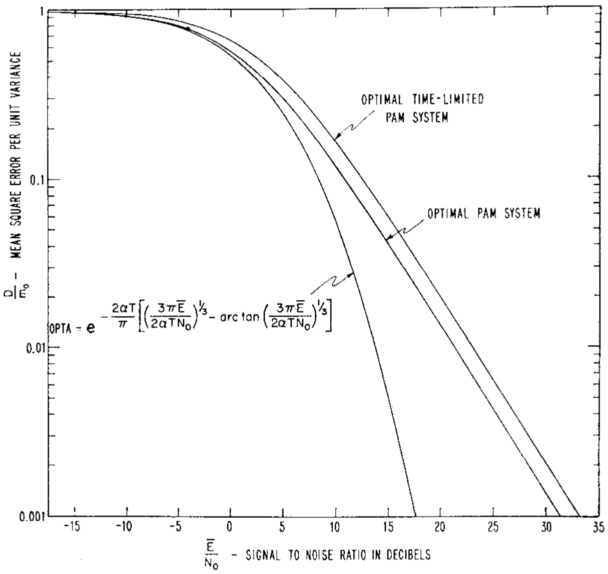
is a monotonic increasing function of $|f|$, it follows that k_f, as defined in step 1) of Section VII, is zero for all $f \in I_0$. The associated performance characteristics are shown in Fig. 5(a) for $\alpha T = 1$. In this and future examples, the dot on the optimal performance characteristic corresponds to the lowest signal-to-noise ratio for which (32) is satisfied over an entire Nyquist set. As is generally the case, the optimal $S(f)$ and $W(f)$ prescribed by (36) depend on λ, and hence on $\bar{P}$, in a more complex manner than simply through an amplitude scale factor.

Example 2: RC Channel, White Noise, Markov Correlated Data

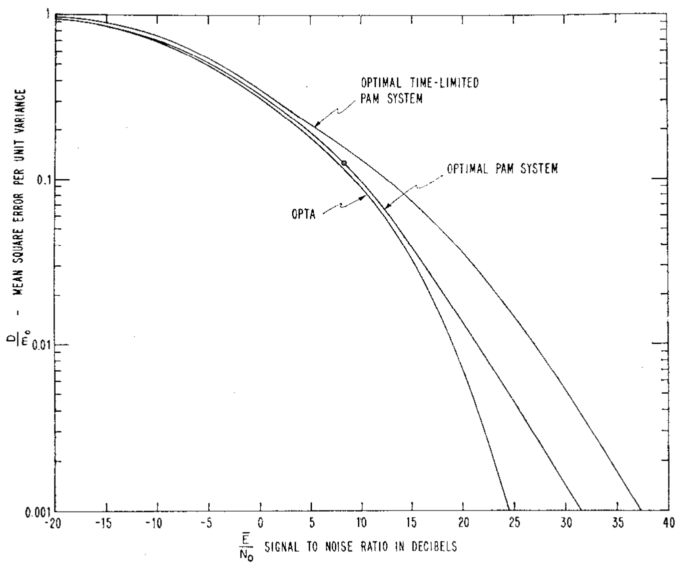
Again, the channel block diagram is given in Fig. 4(a) and $k_f = 0$ for all $f \in I_0$. By Markov correlation, it is meant that $m_k = m_0\rho^{|k|}$ for with $\rho \in (0, 1)$. The associated



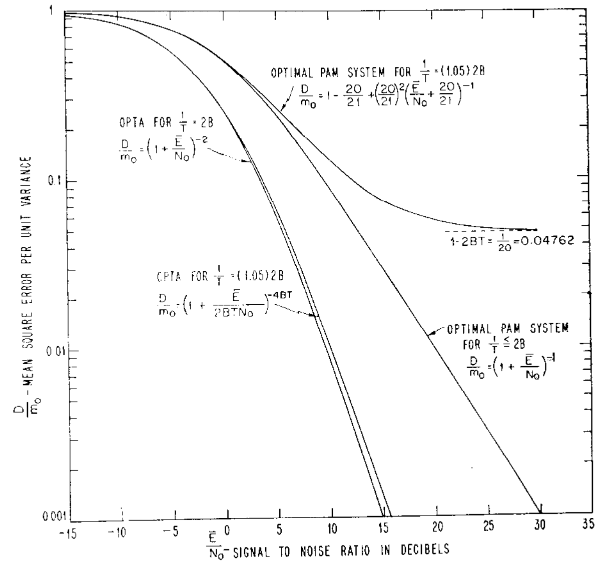
(a)



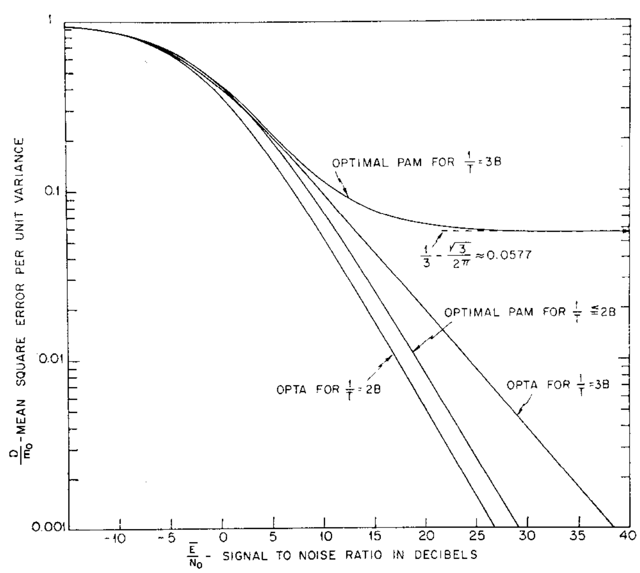
(c)



(b)



(d)



(e)

Fig. 5. (a) Comparison of D/m_0 vs. $\bar{E}/N_0$ performance characteristics for an RC channel with $\alpha T = 1$: uncorrelated data, white noise. (b) Comparison of D/m_0 vs. $\bar{E}/N_0$ performance characteristics for an RC channel with $\alpha T = 1$; Markov correlated data, $\rho = 0.78$. (c) Comparison of D/m_0 vs. $\bar{E}/N_0$ performance characteristics for an RLC channel with $\alpha T = RT/L = 4$ and $(\omega_0 T)^2 = T^2/LC = 100$: uncorrelated data, white noise. (d) Comparison of D/m_0 vs. $\bar{E}/N_0$ performance characteristics for an ideal bandpass channel: white noise: uncorrelated data: signaling rates both above and below the Nyquist rate. (e) Comparison of D/m_0 vs. $\bar{E}/N_0$ performance characteristics for an ideal bandpass channel: white noise: nearest neighbor correlated data, $\theta = \frac{1}{2}$: signaling rates both above and below the Nyquist rate.

periodic message spectrum is

$$M(f) = \sum_{k=-\infty}^{\infty} m_0 \rho^{|k|} e^{i2\pi kTf} = \frac{m_0(1-\rho^2)}{1 - 2\rho\cos 2\pi fT + \rho^2}. \quad (44)$$

The performance characteristics are shown in Fig. 5(b) for $\alpha T = 1$ and $\rho = 0.78$. Comparison with the previous example reveals that it is possible, of course, to capitalize upon the correlation in the data in order to improve system performance. However, an analysis of the subtle trade-offs between correlation, information rate, and system performance will not be undertaken here.

Example 3: RLC Channel, White Noise, Uncorrelated Data

Figure 4(b) shows the block diagram of the channel in question, which is characterized by the function

$$N(f)/|H(f)|^2 = N_0\{1 + [(2\pi f/\alpha) - (\beta/2\pi f\alpha)]^2\}. \quad (45)$$

Naturally, the value of k_f is assigned so that optimal system response is concentrated about the resonance frequency. The parameter values indicated in conjunction with the performance characteristics of Fig. 5(c), for instance, yield $k_f = +2$ for $-1/2T \leqq f < 0$ and $k_f = -2$ for $0 \leqq f < 1/2T$.

Example 4: Ideal Band-Limited Channel, White Noise, Uncorrelated Data

Without loss of generality, the passband may be assumed to consist of a single interval of width $2B$ centered about zero frequency, i.e.,

$$N(f)/|H(f)|^2 = \begin{cases} N_0; & |f| < B \\ \infty; & |f| \geqq B \end{cases}. \quad (46)$$

The nature of the solution depends crucially upon whether or not the signaling rate $1/T$ exceeds the Nyquist rate $2B$.

For $1/T \geqq 2B$, the entire passband is contained in the basic Nyquist set $I_0 = [-1/2T, 1/2T)$. For all significant values of λ, namely all $\lambda \geqq N_0/m_0$, the set I_λ of (35) is the whole passband $(-B, B)$. The functions $|S(f)|^2$ and $|W(f)|^2$ simply are constants over the passband, the values of which are determined by the amount of average power prescribed. Because I_λ does not vary with λ, the parametric dependence of both $\bar{E}/N_0$ and $D_{\min}/m_0$ upon λ easily may be eliminated. The resulting performance characteristic for the optimal system, plotted in Fig. 5(d), is given by

$$D_{\min}/m_0 = 1 - 2BT + (2BT)^2\left(2BT + \frac{\bar{E}}{N_0}\right)^{-1}; \qquad 2BT \leqq 1. \quad (47)$$

An irreducible mean-squared error per unit variance of $1 - 2BT$ in the limit of infinite signal-to-noise ratio is the price that must be paid for signaling faster than the Nyquist rate.

For $1/T < 2B$ the optimal system is not unique. That is, the optimal performance characteristic can be achieved with a variety of system designs. This is because the Nyquist set of I_1 of (34) is not defined uniquely within a set of measure zero as it is for most other channels. Indeed, it even is possible to construct optimal systems with a total bandwidth greater than $1/T$. For instance, suppose that signaling is at less than half the Nyquist rate, i.e., $1/T < B$. The reader may easily verify that, in addition to innumerable solutions of (15) and (16) limited to Nyquist sets, the following is a solution that is nonzero over a band of measure $2/T$:

$$S(f)/K_s = W(f)/K_w = (1 - |f|\,T)^{1/2}; \quad |f| \leqq 1/T, \quad (48)$$

where

$$\begin{aligned} K_s &= K_w\sqrt{\lambda N_0/m_0} \\ &= [T\sqrt{\lambda N_0/m_0}(1 - \sqrt{N_0/\lambda m_0})]^{1/2}. \end{aligned} \quad (49)$$

Straightforward algebraic manipulations eliminate the parameter λ, and the optimal performance characteristic assumes the simple form

$$D/m_0 = \left(1 + \frac{\bar{E}}{N_0}\right)^{-1} = \left(1 + \frac{\bar{P}T}{N_0}\right)^{-1}. \quad (50)$$

Note that if $\bar{E}$ is held constant, (50) is independent of the signaling rate. On the other hand, the second part of (50) shows that, if the average power $\bar{P}$ is constrained, one can, of course, reduce the distortion by reducing the signaling rate. The reader is reminded that (50) is valid only for signaling rates not exceeding the Nyquist rate.

It is shown in Tufts[9] and Berger[10] that the performance characteristic of (50) is indeed common to all solutions of the necessary equations for any signaling rate less than or equal to the Nyquist rate, provided that the solutions are nonzero at least over one Nyquist set. In this regard, note that lowering the signaling rate to the Nyquist rate in (47) yields (50). In Fig. 5(d) both the optimal performance characteristic of a system signaling slightly above the Nyquist rate and the universal characteristic of (50) for rates not exceeding the Nyquist rate are presented. The associated OPTA curves may be computed by appealing to Shannon's[19] classic formula

$$C = W \ln\left(\frac{P + N}{N}\right) \text{ nats per second} \quad (51)$$

for the capacity of a channel of bandwidth W, constrained average input power P, and additive thermal noise power $N = 2N_0W$. In the present notation, $W = B$ and $P = \bar{P} = \bar{E}/T$. Since T seconds are allotted to each message element, (51) may be written

$$C = BT \ln\left(1 + \frac{\bar{E}}{2BTN_0}\right) \text{ nats per message element.} \quad (52)$$

Combining (52) with (41) yields

$$\text{OPTA} = \left(1 + \frac{\bar{E}}{2BTN_0}\right)^{-2BT} = \left(1 + \frac{\bar{P}}{2BN_0}\right)^{-2BT}. \quad (53)$$

It may be observed from (53) that, in contrast to the exponential decrease of the OPTA encountered in most channels, the OPTA decreases only as the power $2BT$ of the signal-to-noise ratio because of the finite passband of the channel under consideration.

If signaling is at the Nyquist rate, (53) reduces to

$$\text{OPTA} = \left(1 + \frac{\bar{E}}{N_0}\right)^{-1} = \left(1 + \frac{\bar{P}}{2BN_0}\right)^{-1}, \tag{54}$$

which is identical to the optimal PAM performance characteristic of (50). This means that the optimal PAM system for signaling at the Nyquist rate through an ideal band-limited channel with an average input power constraint yields the smallest mean-squared error that can possibly be achieved by any communication system whatsoever when both the uncorrelated message elements and the additive white channel noise are zero mean and Gaussian. This, in turn, implies that the optimal PAM system codes ideally both for the source and for the channel in this situation.

On the other hand, when the signaling rate is either above or below the Nyquist rate, the optimal PAM performance is no longer ideal. At rates exceeding the Nyquist rate, intersymbol interference precludes ideal source encoding and even results in the irreducible error of $1 - 2BT$. For rates below the Nyquist rate, the optimal PAM system is not unique. However, since the optimal performance characteristic is unique by definition, the optimal PAM system may be taken to be band-limited to $|f| \leqq 1/2T$ without loss of generality. Such a system utilizes only part of the available bandwidth and, therefore, does not code ideally for the channel. A nonlinear coding technique that results in nearly ideal performance for data rates unequal to the Nyquist rate is discussed in Section IX.

Example 5: Ideal Band-Limited Channel, White Noise, Nearest-Neighbor Correlation

The $N(f)/|H(f)|^2$ function is given by (46). By nearest-neighbor correlation, we mean that each message element is correlated only with the ones immediately preceding and following it; i.e., $m_{\pm 1} = \theta m_0$ and $m_k = 0$ for $|k| \geqq 2$. If this is to be a permissible autocorrelation structure, the parameter θ, which controls the extent of the correlation, must satisfy the inequality $|\theta| \leqq \frac{1}{2}$. Otherwise, the message spectrum

$$M(f) = m_0(1 + 2\theta \cos 2\pi f T) \tag{55}$$

would assume negative values, violating the requirement that the infinite order correlation matrix be non-negative definite. Performance characteristics are shown in Fig. 5(e).

As in the uncorrelated case an irreducible error results for $1/T \geqq 2B$, the expression for which is

$$D_{\min}(\infty)/m_0 = 1 - 2BT - \frac{2\,|\theta|}{\pi} \sin 2\pi BT. \tag{56}$$

Another carry-over from the uncorrelated, or $\theta = 0$, case is the existence of a universal D/m_0 vs. $\bar{E}/N_0$ performance characteristic for $1/T < 2B$. For signal-to-noise ratios large enough so that inequality (32) is satisfied over an entire Nyquist set, this characteristic assumes the form

$$D/m_0 = \left(1 + \frac{\bar{E}}{N_0}\right)^{-1} \left(\frac{1}{\pi} \int_0^\pi \sqrt{1 + 2\theta \cos x}\, dx\right)^2. \tag{57}$$

This is a special case of the more general result

$$D/m_0 = \left(1 + \frac{\bar{E}}{N_0}\right)^{-1} \left(\frac{1}{\pi} \int_0^\pi \sqrt{\frac{M(x/2\pi T)}{m_0}}\, dx\right)^2, \tag{58}$$

which holds for arbitrary message spectra $M(f)$ when $\bar{E}/N_0$ is large and $1/T$ is less than $2B$. An application of the Schwarz inequality shows that the numerator in (58) is always less than 1. Accordingly, optimal performance characteristics for correlated data always lie below that specified by (50) for the uncorrelated case, as would be expected.

IX. Shannon-Cantor Coding

If appropriate nonlinear coding devices are appended to the transmitter and receiver of the optimal PAM system for the band-limited white-noise channel, the overall performance can be made to approach the OPTA even for data rates unequal to the Nyquist rate. In this regard, Shannon[23] has noted that efficient coding for cases in which the data rate exceeds the channel bandwidth requires the mapping of a space of high dimensionality onto one of lower dimensionality. In particular, he cites Cantor's technique for mapping the square $0 \leqq x,\, y < 1$ onto the line $0 \leqq z < 1$ by expanding the coordinates of each point (x, y) in the square in decimal notation, namely

$$\begin{aligned} x &= 0.\, x_1 x_2 x_3 \cdots \\ y &= 0.\, y_1 y_2 y_3 \cdots, \end{aligned} \tag{59}$$

and then obtaining the corresponding point z on the line by alternately interweaving the digits in their decimal expansions, i.e.,

$$z = 0.\, x_1 y_1 x_2 y_2 x_3 y_3 \cdots. \tag{60}$$

Since knowledge of z determines both x and y, and every number has a unique decimal expansion, the mapping in question is clearly one-to-one and onto.

Now consider, for example, the case in which the data rate is twice the Nyquist rate. In theory, we could attach the nonlinear "Shannon-Cantor" encoder specified by (59) and (60) to the front end of our optimal PAM system for signaling at the Nyquist rate. Naturally, we would then append a nonlinear decoder to act upon the received samples in order to undo the interwoven expansions. The advantage gained is that the encoder output actually transmitted by the PAM system results in efficient utilization of the channel, since it is being sent at the

Nyquist rate. As the signal-to-noise ratio increases, the interwoven digits are reproduced more and more accurately. This, in turn, implies that the actual source data can be reproduced at the data rate within any desired degree of accuracy by providing enough signal power, thereby overcoming the phenomenon of irreducible error.

The following heuristic argument indicates that the use of Shannon-Cantor coding can lead to considerable improvement over straight PAM at finite signal-to-noise ratios as well. Assume that the source produces data at twice the Nyquist rate of the band-limited channel and that the specified value of $\bar{P}/N_0$ is such that acceptable two-decimal-digit accuracy is obtained in the interleaved samples transmitted at the Nyquist rate by PAM. That is, the mean-squared error in the received samples, as given by (50) is

$$\left(1 + \frac{\bar{P}}{2BN_0}\right) \approx 10^{-4}. \tag{61}$$

Each of the two numbers formed by separating the expansion digits of a received sample then will have a mean-squared error of approximately 10^{-2}.

To complete the argument note that substitution of (61) into (53) shows that the OPTA has the value 10^{-2} for the assumed signal-to-noise ratio. In performing this substitution the reader is cautioned that the parameter $1/T$ in the OPTA formulas is the source data rate, but in PAM performance calculations $1/T$ is the signaling rate. The data rate and signaling rate were always equal in the previous sections.

The conclusion suggested by this argument is that one is able to approach the OPTA using suitably modified optimal PAM at finite signal-to-noise ratios and at source rates different from the Nyquist rate. Since Cantor's technique readily extends to the one-to-one mapping of an n-dimensional space onto an m-dimensional one, data rates either greater than or less than the Nyquist rate can be handled efficiently by Shannon-Cantor encoders mapping n message elements into m PAM system inputs.

X. Summary

Pulse amplitude modulation links minimizing mean-squared error caused by intersymbol interference and additive noise have been determined. These optimal PAM systems are band-limited to a Nyquist set and exhibit mean-squared error performance characteristics that ultimately are proportional to the reciprocal of the signal-to-noise ratio. The additional improvement possible with general coding techniques has been determined by comparison with performance characteristics derived for systems that are ideal from the standpoint of information theory. In the special case of a band-limited channel with white Gaussian noise, the two performances are identical for uncorrelated Gaussian data elements transmitted at the Nyquist rate. At other transmission rates Shannon-Cantor coding can be used to improve PAM performance.

Appendix I

Derivation of Distortion Expressions

Equations (6) and (7) are derived below. Using the definition of b_i in (4), one can write

$$\begin{aligned}(b_i - a_i)^2 &= \int_{-\infty}^{\infty} w(iT + \delta_i - t)\left[n_g(t) + \sum_{k=-\infty}^{\infty} a_k r(t - kT)\right] dt \\ &\quad \cdot \int_{-\infty}^{\infty} w(iT + \delta_i - x)\left[n_g(x) + \sum_{l=-\infty}^{\infty} a_l r(x - lT)\right] dx - 2a_i \\ &\quad \cdot \int_{-\infty}^{\infty} w(iT + \delta_i - t)\left[n_g(t) + \sum_{k=-\infty}^{\infty} a_k r(t - kT)\right] dt + a_i^2.\end{aligned} \tag{62}$$

Introducing $v(t)$ as defined by (8), interchanging the orders of summation and integration, and applying the expectation operators E_n and E_m of the zero-mean noise and stationary message ensembles yields

$$\begin{aligned}E_nE_m[(b_i - a_i)^2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(iT + \delta_i - t)w(iT + \delta_i - x)n(t - x)\, dt\, dx \\ &\quad + \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} m_{k-l}v(iT + \delta_i - kT)v(iT + \delta_i - lT) \\ &\quad - 2\sum_{k=-\infty}^{\infty} m_{k-i}v(iT + \delta_i - kT) + m_0.\end{aligned} \tag{63}$$

By making changes of integration variables and summation indexes, this can be transformed to

$$\begin{aligned}E_nE_m[(b_i - a_i)^2] &= \int_{-\infty}^{\infty} w(u)\left[\int_{-\infty}^{\infty} w(\tau)n(u - \tau)\, d\tau\right] du \\ &\quad + \sum_{l=-\infty}^{\infty}\sum_{z=-\infty}^{\infty} m_z v(iT + \delta_i - lT - zT)v(iT + \delta_i - lT) \\ &\quad - 2\sum_{z=-\infty}^{\infty} m_z v(\delta_i - zT) + m_0.\end{aligned} \tag{64}$$

The overall distortion, obtained by averaging with respect to the timing jitter as well, therefore, can be expressed as

$$\begin{aligned}D &= m_0 + \sum_{l=-\infty}^{\infty}\sum_{z=-\infty}^{\infty} m_z \\ &\quad \cdot \int_{-\infty}^{\infty} p(\delta_i)v(iT + \delta_i - lT - zT)v(iT + \delta_i - lT)\, d\delta_i \\ &\quad - 2\int_{-\infty}^{\infty}\sum_{z=-\infty}^{\infty} m_z v(\delta_i - zT)p(\delta_i)\, d\delta_i \\ &\quad + \int_{-\infty}^{\infty} w(u)\int_{-\infty}^{\infty} w(\tau)n(u - \tau)\, d\tau\, du.\end{aligned} \tag{65}$$

A change of integration variable to $u = \delta_i + iT - lT$, another interchange of summation and integration and, finally, a change of summation index transforms the second term on the right of (65) to

$$\int_{-\infty}^{\infty} \sum_{z=-\infty}^{\infty} m_z v(u - zT) \sum_{k=-\infty}^{\infty} v(u)p(u - kT)\, du. \qquad (66)$$

When this is substituted in (65) and combined with the third term, the result is (6). Equation (7) then follows from (6) by repeated application of Parseval's theorem.

Appendix II

Resolution of Difficulties Concerning the Optimal Jitter-Free Solution

An optimal PAM system was defined in Section V as one which minimizes D for a specified value of $\bar{P}$. However, the procedure recommended in Section VII determines that solution of the jitter-free necessary equations which yield the least possible D for a fixed value λ^* of the Lagrange multiplier rather than for a fixed value $\bar{P}$ of the average transmitted power. Thus, it remains to be shown that there is no other solution of the necessary equations using some $\lambda \neq \lambda^*$ that has an associated average power $\bar{P}$ which equals $\bar{P}(\lambda^*)$ as given by (37), and an associated distortion D less than $D_{\min}(\lambda^*)$ as given by (38). This is established below. For brevity we omit the "df" symbol in each integral and the argument f in the functions $M(f)$, $N(f)$, and $H(f)$.

Any solution of the jitter-free necessary equations (15) and (16) may be completely described by a pair (λ, I). The first entry λ is the value of the Lagrange multiplier for which the equations have been solved, and the second entry I is the Nyquist set or part thereof over which the solution is nonzero. The associated average transmitted power $\bar{P}(\lambda, I)$ and distortion $D(\lambda, I)$ are

$$\bar{P}(\lambda, I) = \int_I \left(\sqrt{\frac{\lambda MN}{|H|^2}} - \frac{N}{|H|^2}\right) = \int_I \sqrt{\frac{\lambda MN}{|H|^2}}\left(1 - \sqrt{\frac{N}{\lambda M\, |H|^2}}\right) \qquad (67)$$

and

$$D(\lambda, I) = m_0 - T \int_I \left(M - \sqrt{\frac{MN}{\lambda\, |H|^2}}\right) = m_0 - T\int_I M\left(1 - \sqrt{\frac{N}{\lambda M\, |H|^2}}\right). \qquad (68)$$

[When $I = I_\lambda$ of (35), then $\bar{P}(\lambda, I_\lambda) = \bar{P}(\lambda)$ of (37), and $D(\lambda, I_\lambda) = D_{\min}(\lambda)$ of (38).]

Because inequality (32) must be satisfied over the entire set I, we see that the integrands in (67) and (68) are always positive. This implies that $\bar{P}(\lambda, I)$ and $D(\lambda, I)$ respectively increase and decrease with increasing λ for fixed I, and also that, if $I_A \subset I_B$, than $\bar{P}(\lambda, I_A) \leqq \bar{P}(\lambda, I_B)$ and $D(\lambda, I_A) \geqq D(\lambda, I_B)$.

Consider the recommended solution $(\lambda^*, I_{\lambda^*})$ for the value $\bar{P}(\lambda^*)$ of the average transmitted power. It will now be shown that no solution (λ, I) exists such that both

1) $\bar{P}(\lambda, I) = \bar{P}(\lambda^*)$

and

2) $D(\lambda, I) < D_{\min}(\lambda^*)$.

Let (λ, I) be a solution reputed to have properties 1) and 2). Then $\lambda > \lambda^*$, since $\lambda \leqq \lambda^*$ would imply upon recalling (35) that $I_\lambda \subset I_{\lambda^*}$ and, therefore, that $D(\lambda, I) \geqq D(\lambda, I_\lambda) \geqq D(\lambda^*, I_\lambda) \geqq D(\lambda^*, I_{\lambda^*}) = D_{\min}(\lambda^*)$, which violates property 2). Hence, in the following discussion, it is assumed without loss of generality that λ is greater than λ^*.

Now suppose that I is composed entirely of translates which minimize $N/|H|^2$; i.e., suppose $I \subset I_1$ of (34). Then, since inequality (32) must be satisfied, it is necessary that $I \subset I_\lambda$ (cf. definition of I_λ given in (35)). In fact, it is necessary to delete a set of positive measure from I_λ in order to form I, for otherwise $\bar{P}(\lambda, I) = \bar{P}(\lambda, I_\lambda)$, whereas $\lambda > \lambda^*$ implies that $I_{\lambda^*} \subset I_\lambda$ and hence that $\bar{P}(\lambda, I) = \bar{P}(\lambda, I_\lambda) \geqq \bar{P}(\lambda, I_{\lambda^*}) > \bar{P}(\lambda^*, I_{\lambda^*}) = \bar{P}(\lambda^*)$, which violates 1). Of course, as frequencies are deleted in order to make $\bar{P}(\lambda, I)$ decrease, the value of $D(\lambda, I)$ will increase. It is best to delete first those frequencies for which the resultant increase in D per unit decrease in $\bar{P}$ is least. Therefore, the first frequencies deleted should be those for which the ratio of the integrand in (68) to that in (67) is smallest. This ratio is

$$\frac{\Delta D}{-(\Delta P)} = \frac{TM\left(1 - \sqrt{\frac{N}{\lambda M\, |H|^2}}\right)}{\sqrt{\frac{\lambda MN}{|H|^2}}\left(1 - \sqrt{\frac{N}{\lambda M\, |H|^2}}\right)} = T\sqrt{\frac{M\, |H|^2}{\lambda N}}. \qquad (69)$$

Since λ is being held constant during the deleting process, the first frequencies in I_λ to be deleted are those for which $\sqrt{N/M\, |H|^2}$ is largest. Accordingly, the best way to delete frequencies is to take away first those which were added to I_{λ^*} in order to form I_λ. Even after all these have been deleted, the resulting $\bar{P}$ is still too large, since it is $\bar{P}(\lambda, I_{\lambda^*}) > P(\lambda^*, I_{\lambda^*}) = P(\lambda^*)$. Hence, further deletion is required, which means that the I obtained when the optimal deletion process is finally completed is the result of removing a set of positive measure from I_{λ^*}.

Continuing to assume that $I \subset I_1$, we will now show that the resulting $D(\lambda, I)$ does not satisfy 2). In fact, the following argument establishes that, in general, if $(\lambda + \Delta\lambda, I - \Delta I)$ is a solution formed from any solution (λ, I) by adding an increment $\Delta\lambda$ to the Lagrange multiplier, and then compensating by deleting the set ΔI in such a manner that $\bar{P}$ does not change, then $D(\lambda + \Delta\lambda, I - \Delta I) \geqq D(\lambda, I)$. In this regard, the new average transmitted power expression is

$$P(\lambda + \Delta\lambda, I - \Delta I) = \int_I \left(\sqrt{\lambda + \Delta\lambda} \sqrt{\frac{MN}{|H|^2}} - \frac{N}{|H|^2} \right) - \int_{\Delta I} \left(\sqrt{\lambda + \Delta\lambda} \sqrt{\frac{MN}{|H|^2}} - \frac{N}{|H|^2} \right). \tag{70}$$

For convenience, let $\gamma = \sqrt{\lambda}$ and $\gamma + \Delta\gamma = \sqrt{\lambda + \Delta\lambda}$. Then the requirement that $\bar{P}$ is being held constant becomes

$$\begin{aligned} \frac{\Delta \bar{P}}{\gamma} &= \frac{\bar{P}(\lambda + \Delta\lambda, I - \Delta I) - \bar{P}(\lambda - I)}{\gamma} \\ &= \left(\frac{\Delta\gamma}{\gamma}\right) \int_I \sqrt{\frac{MN}{|H|^2}} - \int_{\Delta I} \left[\left(\frac{\gamma + \Delta\gamma}{\gamma}\right) \sqrt{\frac{MN}{|H|^2}} - \frac{N}{\gamma\,|H|^2} \right] \\ &= 0. \end{aligned} \tag{71}$$

The distortion increment, normalized by T, is

$$\begin{aligned} \frac{\Delta D}{T} &= \frac{D(\lambda + \Delta\lambda, I - \Delta I) - D(\lambda, I)}{T} \\ &= \int_{\Delta I} M + \left(\frac{1}{\gamma + \Delta\gamma}\right) \int_I \sqrt{\frac{MN}{|H|^2}} - \left(\frac{1}{\gamma + \Delta\gamma}\right) \\ &\quad \cdot \int_{\Delta I} \sqrt{\frac{MN}{|H|^2}} - \frac{1}{\gamma} \int_I \sqrt{\frac{MN}{|H|^2}} \\ \frac{\Delta D}{T} &= \int_{\Delta I} M - \left(\frac{1}{\gamma + \Delta\gamma}\right)\left(\frac{\Delta\gamma}{\gamma}\right) \\ &\quad \cdot \int_I \sqrt{\frac{MN}{|H|^2}} - \left(\frac{1}{\gamma + \Delta\gamma}\right) \int_{\Delta I} \sqrt{\frac{MN}{|H|^2}}. \end{aligned} \tag{72}$$

Substituting (71) into (72) yields

$$\frac{\Delta D}{T} = \int_{\Delta I} \left[M - \frac{1}{\gamma} \sqrt{\frac{MN}{|H|^2}} + \frac{N}{\gamma(\gamma + \Delta\gamma)\,|H|^2} - \frac{1}{\gamma + \Delta\gamma} \sqrt{\frac{MN}{|H|^2}} \right]. \tag{73}$$

Some algebraic manipulations cast (73) in the form

$$\frac{\Delta D}{T} = \left(\frac{\gamma}{\gamma + \Delta\gamma}\right) \int_{\Delta I} M\left(1 - \frac{1}{\gamma}\sqrt{\frac{N}{M\,|H|^2}}\right)^2 + \left(\frac{\Delta\gamma}{\gamma + \Delta\gamma}\right) \int_{\Delta I} M\left(1 - \frac{1}{\gamma}\sqrt{\frac{N}{M\,|H|^2}}\right). \tag{74}$$

Both terms on the right of (74) have positive integrands, the first because of the perfect square and the second because inequality (32) must be satisfied. Accordingly, $\Delta D > 0$ as claimed.

It follows from the above argument that no (λ, I) having $I \subset I_1$ satisfies both 1) and 2). It should be intuitively clear both physically and mathematically that using non-optimal translates instead of the frequencies in I_1 can only make matters worse. In this regard, note from (69) that, since M is invariant under such translations, the decrease in D obtained per unit $\bar{P}$ expended in this manner is reduced because $|H|^2/N$ is smaller outside I_1. Accordingly, the solution (λ, I_λ) recommended in Section VII is indeed the optimal jitter-free system for the average transmitted power value $\bar{P}(\lambda)$ which results.

References

[1] M. R. Aaron and D. W. Tufts, "Intersymbol interference and error probability," *IEEE Trans. on Information Theory*, vol. IT-12, pp. 26–34, January 1966.

[2] D. W. Tufts and T. Berger, "Optimum pulse amplitude modulation—part II: inclusion of timing jitter," *IEEE Trans. on Information Theory*, this issue, page 209.

[3] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *Information and Decision Processes*, R. E. Machol, Ed. New York: McGraw-Hill, 1960, pp. 93–126.

[4] M. S. Pinsker, "Calculation of the speed of information creation by means of stationary random processes and of the stationary channel capacity," *Proc. Acad. Sci. (USSR)*, vol. 113, pp. 753–756, 1965.

[5] R. L. Dobrushin and B. S. Tsybakov, "Information transmission with additional noise," *IEEE Trans. on Information Theory*, vol. IT-8, pp. s-293–304, September 1962.

[6] D. W. Tufts, "Nyquist's problem—the joint optimization of transmitter and receiver in pulse amplitude modulation," *Proc. IEEE*, vol. 53, pp. 248–259, March 1965.

[7] D. A. Shnidman and D. W. Tufts, "Effect of correlated data and colored noise on error probability and mean square error in PAM data transmission," *IEEE Internat'l Conv. Rec.*, pt. 7, vol 13, pp. 227–237, 1965.

[8] D. W. Tufts, "Design problems in pulse transmission," Sc.D. dissertation, M.I.T., Cambridge, Mass., May 1960; also M.I.T. Research Lab. of Electronics, Tech. Rept. 368, July 28, 1960.

[9] ——, "Certain results in pulse transmission theory," Cruft Lab., Harvard University, Cambridge, Mass., Tech. Rept. 355, February 5, 1962.

[10] T. Berger, "Nyquist's problem in data transmission," Ph.D. dissertation, Harvard University, Cambridge, Mass., December 1965; also Raytheon Co., Wayland, Mass., Tech. Rept. ESRN-68, February 1966.

[11] T. Berger and D. W. Tufts, "Waveforms and receivers for pulse amplitude modulation," Cruft Lab., Harvard University, Cambridge, Mass., Tech. Rept. 488, September 1965.

[12] B. S. Tsybakov, "Linear information coding," *Radio Engrg. Electronic Phys.*, vol. 7, pp. 22–33, January 1962.

[13] J. W. Smith, "The joint optimization of transmitted signal and receiving filter for data transmission systems," *Bell Sys. Tech. J.*, vol. 44, pp. 2363–2393, December 1965.

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

[15] M. J. DiToro, "A new method of high-speed adaptive serial communication through any time-variable and dispersive transmission link," *Conf. Rec. 1st IEEE Ann. Commun. Conv.*, pp. 763–767, June 1965.

[16] R. W. Lucky, "Automatic equalization for digital communication," *Bell Sys. Tech. J.*, vol. 44, pp. 547–588, April 1965.

[17] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. Berkeley: University of California Press, 1958.

[18] W. B. Davenport, Jr. and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*. New York: McGraw-Hill, 1958, pp. 183–184.

[19] C. E. Shannon, *The Mathematical Theory of Communication*. Urbana, Ill.: University of Illinois Press, 1949.

[20] J. L. Holsinger, "Digital communication over fixed time-continuous channels with memory and special applications to telephone channels," Ph.D. dissertation, M.I.T., Cambridge, Mass., 1964.

[21] D. A. George and D. C. Coll, "The reception of time-dispersed pulses," *Conf. Rec. 1st IEEE Ann. Commun. Conv.*, pp. 749–752, June 1965.

[22] L. E. Franks, "A method of optimizing performance of pulse transmission systems," *IEEE Internat'l Conv. Rec.*, pt. 5, vol. 12, pp. 279–290, 1964.

[23] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, January 1949.