Copyright © 1971 American Telephone and Telegraph Company
THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 6, July-August, 1971
Printed in U.S.A.

The Computation of Error Probability for Digital Transmission

By F. S. HILL, JR.

(Manuscript received October 19, 1970)

A method is presented for calculating the probability of error for a digital signal contaminated by intersymbol interference and additive Gaussian noise. The method constructs a close approximation to the probability density function of the intersymbol interference, circumventing the heretofore formidable computational problems by decomposing the calculation into a sequence of simple calculations. This method is applicable to binary transmissions, as well as 4-, 8-, 16-, . . . level transmission. The method is rapid enough for use on time-sharing facilities. As part of the method, a new and rather simple scheme is presented for including the effects of partial response source coding. Several interesting examples which use and give insight into the method are included.

I. INTRODUCTION

In a large class of digital transmission systems, a succession of amplitude modulated pulses is sent over a channel to the receiver. The signal suffers two main types of distortion: additive noise and intersymbol interference (ISI). The latter arises from dispersion in the channel, and is a noise-like process due to the overlapping of many neighboring pulses.[1] The number of these neighboring pulses, or "interferers," depends on the system impulse response, and can be rather large for systems with restricted bandwidth.

To properly analyze a system one must compute its probability of error due to the noise and intersymbol interference. The effect of the additive noise is simple to analyze, but that of the ISI is extremely difficult to find due to the complicated nature of its probability density function (pdf). The complexity of the pdf grows exponentially with the number of interferers, and for more than about 12 interferers the computation of error probability has long seemed impossible.

Until recently researchers have analyzed such systems by one of three main routes: (*i*) they have used more tractable—although less meaningful—performance measures such as peak distortion[2] or mean squared error,[1,3] (*ii*) they have found upper bounds on the probability of error,[4] or (*iii*) they have found the probability of error using a truncated impulse response, for which only a few interferers were assumed to be significant.[5,6] Unfortunately, in many restricted-bandwidth systems as many as 50 or 60 interferers can be important, so the use of a truncated impulse response is inadequate for a large class of systems. The goal is therefore to find methods which will treat large numbers of interferers while avoiding the exponential growth of computational complexity. This paper describes such a method.

Recently two other schemes have been presented which also provide accurate estimates of the error probability for large numbers of interferers.[7,8] These methods do not provide the pdf of ISI, but calculate instead the terms in a series expansion of the expression for the error probability. They were devised primarily for the binary transmission case, and the extension to multilevel transmission would be rather awkward.

The new technique presented here constructs a close approximation to the pdf of intersymbol interference, and uses it to find the error probability. Knowledge of the pdf can provide useful information about the intersymbol interference process, illuminating for the system designer the relationship between certain types of channel characteristics and error performance. The technique is applicable to binary transmission, as well as to 4-, 8-, 16-, $\cdots$ level transmission. The method also allows simple inclusion of certain kinds of source coding such as partial response.[9] The ability to treat the multilevel case and different kinds of source coding is very important. In feasibility studies a system designer might know the general type of channel characteristics the signal will encounter, but he does not know how sensitive the signal will be to various combinations of numbers of source levels and source coding. The method given here is envisioned as providing a valuable tool in such studies.

In Section II the probability of error is related to a single random variable, the error voltage z, which then characterizes system performance. It is this random variable whose pdf is sought. Examples taken from the applicable class of source coders are presented and it is shown that the effects of coding can be transferred from the source statistics to the channel description. Alternative performance measures are noted for later use. In Section III the details of the method are described with the computer programmer in mind, and various empirical rules-of-thumb are suggested. In Section IV several simple examples are presented which illuminate some of the potential uses of the method. In Section V the accuracy of the method is considered, and guidelines for choosing some parameter values in the method are given. The method presented here is an extended and refined version of a technique developed by W. E. Norris.[10]

II. ANALYSIS

Figure 1 indicates the system under consideration. The source emits a symbol every T seconds. The symbols are statistically independent, and form the sequence $\{a_k\}$ where each a_k takes on one of the N equally likely values $m(i)$

$$m(i) = (2i - N - 1)V/(N - 1), \quad i = 1, 2, \cdots, N. \tag{1}$$

The outermost levels are $\pm V$ volts. The sequence is then passed through a coder to form the sequence $\{c_k\}$. Only two kinds of coding are treated in the main body of the paper, although the method is extended to more general coders in the Appendix. The two special types treated here are the uncoded case and the class IV partial response case,[9] where

$$\begin{aligned} \text{uncoded: } c_k &= a_k , \\ \text{partial response IV: } c_k &= a_k - a_{k-2} . \end{aligned} \tag{2}$$

Partial response IV coding provides useful shaping of the signal spectrum, and precoding of the source sequence can be used to prevent error propagation.[9] It is assumed that the sequence $\{a_k\}$ has already been precoded appropriately.

The sequence $\{c_k\}$ modulates the impulse response $r(t)$ of the channel. White Gaussian noise is added, and the received signal $y(t)$ is sampled every T seconds. The decision device (typically an A/D converter), determines which of the possible transmitted levels each

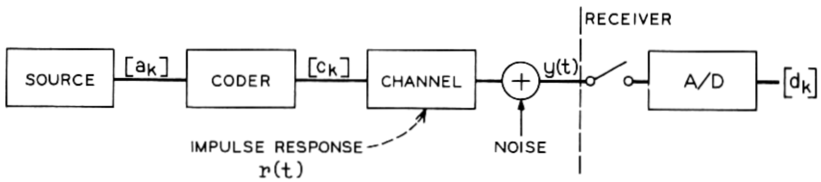

Fig. 1—System configuration.

sample is nearest, and outputs the sequence $\{d_k\}$. (The receiver can base its decision simply on the present sample because of the precoding.[9]) Ideally, each $d_k = c_k$, but because of the noise and ISI they will occasionally differ, and an error is made. The method discussed here calculates the probability of occurence of such an event.

The sample at $t = 0$ has the form

$$y(0) = \sum_{k=-\infty}^{\infty} c_k r(-kT) + n, \tag{3}$$

where n is a realization of a zero mean Gaussian random variable with variance σ^2. Because the sequence $\{a_k\}$ and the noise are stationary in a statistical sense, studying the error probability for the single sample $y(0)$ is equivalent to studying it for the entire ensemble of samples $y(mT)$.

Ideally $y(0)$ will equal c_0. The error voltage z is therefore given by

$$z = \sum_{k=-\infty}^{\infty} c_k x_k + n, \tag{4}$$

where

$$x_k = r(-kT) - \delta_k \tag{5*}$$

are the "error samples" of the system. It is very convenient to incorporate the effects of any source coding in the error samples. In the case of partial response coding:

$$\begin{aligned} z &= \sum_{k=-\infty}^{\infty} (a_k - a_{k-2})x_k + n \\ &= \sum_{k=-\infty}^{\infty} a_k e_k + n, \end{aligned} \tag{6}$$

where

$$e_k = x_k - x_{(k+2)} \tag{7}$$

are the "coded error samples." Thus z of (6) depends on the simple statistics of the uncoded source symbols a_k. (If no coding is used, then $e_k = x_k$.)

The error voltage z is the sum of a large number of independent random variables. Each a_k has the same discrete probability density function: A set of N equidistant Dirac delta-functions with weights

---

* δ_k = Kronecker delta function: $\delta_0 = 1$; $\delta_k = 0$, $k \neq 0$.

$1/N$. Each a_k is scaled by the coded error sample e_k. The n is Gaussian, and assumed independent of the a_k's. The pdf $p_z(\cdot)$ of z is found by convolving the individual density functions of the constituent random variables. Because each of these constituents has a symmetrical pdf about $z = 0$, $p_z(\cdot)$ is also symmetrical. Once $p_z(\cdot)$ has been found, the error probability for the sample $y(0)$ is simply the probability that $y(0)$ is further from c_0 than from some other level. When c_0 is an inner level this is just the probability that $|z| > V/(N-1)$. When c_0 is an outer level only one polarity of z causes an error, and an error occurs with the probability that $z > V/(N-1)$. For uncoded symbols outer levels occur with probability $1/N$, while for partial response IV coding they occur with probability $1/N^2$. Hence the error probability is

$$P_e = 2(1 - N^{-h}) \int_{V/(N-1)}^{\infty} p_z(e)\, de, \tag{8}$$

where $h = 1$ for an uncoded source, and $h = 2$ for a partial response IV coded source.

Clearly, the real task in finding P_e lies in computing $p_z(\cdot)$, since so many random variables are involved. The method used to achieve this is described in the next section, after two simpler measures of system distortion have been introduced. These measures have been widely used in the past, and will provide useful definitions in the discussion to follow. They are both defined in the noise-free case (i.e., $n = 0$) here in order to isolate the effect of the intersymbol interference.

2.1 *Mean Squared Error*[1,3]

The variance of the error voltage z is given by

$$E[y(0) - c_0]^2 = E[\sum_k a_k e_k]^2 = E[a^2] \sum_k e_k^2, \tag{9}$$

where $E[a^2]$ is the variance of the source. A simple calculation based on equation (1) shows that $E[a^2] = \frac{1}{3}V^2(N+1)/N-1)$. A normalized version of the mean squared error then depends only on the coded error samples, and will be called MSE:

$$\text{MSE} = \sum_k e_k^2. \tag{10}$$

2.2 *Peak Distortion*[2]

The maximum value that z of equation (6) can have when $n = 0$ is

$$\max_{a_k} z = V \sum_k | e_k |, \tag{11}$$

which occurs when all of the a_k symbols have maximum amplitude V and such signs that the interfering symbols add constructively. Such a sequence occurs with very low probability. Peak distortion here is defined as D:

$$D = \sum_k | e_k | . \tag{12}$$

III. THE COMPUTATIONAL METHOD

The starting point for computing $p_z(\cdot)$ is the set of coded error samples $\{e_k\}$. The only other relevant parameters are V, N and σ^2. Because each random variable a_k has the same pdf, and this function is symmetrical, one need only consider the magnitudes $|e_k|$, and these can be relabelled to rank order the samples for convenience.

3.1 *Obtaining the Error Samples*

Channel characteristics are typically measured or calculated in the frequency domain, although in some simple cases an analytical expression for the channel impulse response is available. If the impulse response is indeed available, it is simply sampled as in equation (5) to obtain the error samples. Otherwise the error samples are computed from $R(f)$ as follows.* Because by inspection

$$r(kT) = \int_{-\infty}^{\infty} R(f) e^{i2\pi fkT}\, df = \int_{-1/2T}^{1/2T} e^{i2\pi fkT} \sum_{m=-\infty}^{\infty} R(f - m/T)\, df \tag{13}$$

the Fourier coefficients of $\hat{R}(f) = \sum_m R(f - m/T)$ are desired. A Discrete Fourier Transform (DFT)—perhaps using a Fast Fourier Transform algorithm—of N_s samples of $\hat{R}(f)/N_sT$, $f_i = (i - 1)/N_sT$, $i = 1, 2, \cdots, N_s$, yields N_s samples of the aliased version $\hat{r}(t) = \sum_m r(t - mT)$, $t_k = (k - 1)T$, $k = 1, 2, \cdots, N_s$.[11] Although the samples of $\hat{r}(t)$ are not identical to those of $r(t)$, for well-behaved channels and sufficiently large N_s the two are indistinguishable.

The error samples of equation (5) are formed by removing unity from r_o. The coded error samples (for the partial response IV case) are then formed by subtracting from each sample the value of the error sample two places to the right in accordance with (7). The last two error samples, e_{N_s-1} and e_{N_s} are formed making use of the periodic nature of $r(t)$:[11] e.g., because $x_{N_s+2} = x_2$ we have $e_{N_s} = x_{N_s} - x_2$. Finally, only the magnitudes are retained, and these are rank ordered.

* Upper and lower case letters indicate Fourier Transform pairs.

The result is the vector $\mathbf{e}$:

$$\mathbf{e} = (e_1, e_2, \cdots, e_{N_s}), \; e_1 \geqq e_2 \geqq \cdots \geqq e_{N_s} \geqq 0. \qquad (14)$$

(Any $e_k = 0$ could be discarded.)

3.2 *Formation of the Pdf for Binary Transmission*

The number of levels N is normally a power of two. In such cases the pdf $p_z(\cdot)$ for N levels can be built up from the pdf for the binary case as discussed below. Thus the binary case is of fundamental importance. Ignoring the Gaussian noise for the moment, the error voltage z of equation (6) is

$$z = \sum_{k=1}^{N_s} a_k e_k, \qquad (15)$$

where the e_k are elements of $\mathbf{e}$ and for the binary case each a_k takes on the values $\pm V$ with equal probability. There are 2^{N_s} possible realizations of the sequence $(a_1, a_2, \cdots, a_{N_s})$, each occuring with probability 2^{-N_s}. Each realization yields a value for z, (not necessarily all different), which contributes a delta function of weight 2^{-N_s} to the pdf of z. Since N_s can be 100 or more, an attempt to form z for each possible sequence of a_k would be futile ($2^{100} = 10^{30}$). Instead, decomposition and quantization are used to reduce this formidable task to a sequence of rather simple ones.

3.2.1 *Decomposition*

The solution used here to circumvent the overwhelming computational problem is to decompose the N_s samples into blocks of K samples, and to find the pdf of ISI due to the samples for each block.* The resulting pdf's are convolved together to form the final binary pdf of z. Using specific numbers for concreteness, if $N_s = 54$ and one chooses $K = 9$, then z can be written as a sum of six random variables $A_1, \cdots, A_6$, where

$$A_i = \sum_{k=9i-8}^{9i} a_k e_k, \qquad i = 1, \cdots, 6. \qquad (16)$$

Each A_i is a discrete random variable, being the sum of nine discrete random variables. As the pdf of each A_i is formed, it is convolved with previous ones until all 54 samples have been included.

* Assume for the moment that N_s is an integer multiple of K. A rule-of-thumb for discarding some of the small error samples in order to reduce the computation time is described in the following text.

Consider the formation of A_1. Since a_k can take on values $\pm V$, there are 2^9 possible values for A_1. The largest value, denoted as A_1^*, occurs when all $a_k = V$, and has value $A_1^* = V(e_1 + e_2 + \cdots + e_9)$. The 2^9 delta functions that make up the exact pdf of A_1 will therefore lie in the interval $[-A_1^*, A_1^*]$. Since the pdf is symmetrical it is sufficient to compute it for nonnegative values of A_1 only. Every combination of $\pm V$ values for $a_1, \cdots, a_9$ is tried. For each the corresponding value of A_1 is found according to equation (16), and is discarded if $A_1 < 0$.

The following simple technique insures that all combinations are included. A 9-digit binary number B is initialized to zero, and then in turn incremented by one until all digits are one (2^9 iterations in all). For each value of B, the value of a_k is V if the kth digit of B is 1, and $-V$ otherwise.

3.2.2 *Quantization*

Decomposition alone would not significantly reduce the size of the computation, since the total number of delta functions in the final binary pdf of z would still be on the order of 2^{54}. Quantization of the voltage axis is used to keep the number of possible levels of z within reasonable bounds.[10]

The interval $[0, A_1^*]$ is marked off with $2N_1 + 1$ "location" points t_i, $i = 0, 1, \cdots, 2N_1$. Associated with each t_i is a probability p_i. Instead of recording the exact value of A_1, it is tested to see which t_i it is nearest, and the corresponding p_i is incremented by 2^{-9}. There are many roughly equivalent ways of assigning the values t_i. The following scheme was adopted for its computational simplicity. The interval $[0, A_1^*]$ is broken into two parts at some level kA_1^*, and then each part is uniformly quantized into N_1 levels. Hence, the t_i are given by

$$t_i = \begin{cases} kA_1^* i/N_1, & i = 0, 1, \cdots, N_1, \\ kA_1^* + (1 - k)A_1^*(i/N_1 - 1), & i = N_1 + 1, \cdots, 2N_1. \end{cases} \tag{17}$$

For $k = 1/2$, the quantizing is therefore uniform over the whole interval. For $k > 1/2$ the quantizing is finer for larger values of A_1, so that the most important levels of error voltage are most accurately represented. An efficient choice for k is 0.6, as shown in Section V. (However, uniform quantizing may have other advantages since convolution could be performed using a Fast Fourier Transform algorithm.[11]

The pdf's of $A_2, \cdots, A_9$ are computed in the same way. After each has been found, it is convolved with the resultant pdf for the previous ones. Quantization is also used in the convolution process.

3.3 *Formation of the N-Level Pdf*

Once the binary pdf for z has been obtained, it is a straightforward matter to find the pdf for 4-, 8-, 16-, $\cdots$ level sources. The key to this is to envision N level transmission as the sum of $\log_2 N$ independent binary transmission systems.[10] For instance, an 8-level source with voltage range from $-V$ to V can be decomposed into three binary sources with symbols f_k, g_k, and h_k having only values $\pm V$. Each 8-level symbol a_k can be written as:

$$a_k = \frac{4f_k + 2g_h + h_k}{7} \tag{18}$$

and the error voltage for eight levels can be written in the noise-free case as

$$\begin{aligned} z &= \tfrac{4}{7} \sum f_k e_k + \tfrac{2}{7} \sum g_k e_k + \tfrac{1}{7} \sum h_k e_k \\ &= \tfrac{4}{7} E_1 + \tfrac{2}{7} E_2 + \tfrac{1}{7} E_1, \end{aligned} \tag{19}$$

where E_1, E_2, and E_3 are independent random variables, all having the same pdf as z in the binary case. (The generalization to any power of 2 is straightforward.) Since z is the sum of three random variables, and the pdf of each is known, it is a simple matter to convolve scaled versions of the binary pdf to obtain the 8-level pdf. Quantization is used here also, employing the scheme described above. The final pdf is thus approximated over the interval $[0, D \cdot V]$ by a set of $2N_1 + 1$ delta-functions having locations t_i similar to those in equation (17), and probabilities p_i. By symmetry there is the same pattern of delta-functions at location $-t_i$ with probabilities p_i.

3.4 *Calculation of the Error Probability*

The Gaussian random variable n of equation (6) is reinserted by convolving its pdf with that found above. The result is simply

$$P_z(x) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=0}^{2N_1} p_i \{ e^{-(x-t_i)^2/2\sigma^2} + e^{-(x+t_i)^2/2\sigma^2} \}. \tag{20}$$

From equation (8) the error probability is then

$$P_e = 2(1 - N^{-h}) \sum_{i=0}^{2N_1} p_i \left[Q\left(\frac{V - t_i}{\sigma(N-1)}\right) + Q\left(\frac{V + t_i}{\sigma(N-1)}\right) \right] \tag{21}$$

where $Q(x)$ is the normal probability integral

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt. \tag{22}$$

The accuracy of the method is examined in Section V, following a set of examples of the method applied to typical channel distortions.

3.5 *Reducing the Computation Time by Discarding Samples*

In the preceding discussion all N_s error samples of equation (14) were involved in the computation of the pdf of ISI. In most practical cases, however, a sizable number of the error samples are very small, and consequently their effect on the nature of the pdf as calculated is negligible. In order to save some computation time it behooves one to truncate the rank-ordered vector **e** of (14) at some index M, and to lump the other samples together in some fashion. This is an optional procedure, since the computation time grows only approximately linearly with the number of blocks of K samples processed, but it can indeed improve the efficiency of the method.

One way to approximate the effect of the remaining samples is to treat the random variable

$$A_r = \sum_{K=M+1}^{N_s} a_k e_k \tag{23}$$

as a Gaussian random variable with variance[4]

$$\sigma_r^2 = E[a^2] \sum_{K=M+1}^{N_s} e_k^2. \tag{24}$$

The effect of these samples is thus a contribution to the Gaussian noise n, replacing the previous variance σ^2 in equation (21) with $\sigma_1^2 = \sigma^2 + \sigma_r^2$.

The question of choosing M still remains. One rule of thumb is that the sum of the samples lumped together should not exceed 1 percent of the total sum D. This was found to be very useful empirically: for all choices of channel studied P_e was the same whether all N_s samples were used, or only M of them, while if only M were used the saving in computation time was significant.

There are some pathological cases of mainly academic interest where D is theoretically infinite. This is true, for instance, for an ideal band-limited signal with a timing error. Saltzberg has shown, using an upper bound method, that the error probability can still be small in such cases.[4] The present method will not work, however, since the value of D is a crucial landmark in the computational procedure, and it must be finite. Such a restriction does not significantly reduce the power of the method.

IV. EXAMPLES OF ERROR RATE CALCULATION

In each of the examples the following definitions are used:

$$\text{Error Rate} = \frac{P_e}{\log_2 N}, \qquad \text{SNR} = \frac{E[c^2]}{\sigma^2} = \frac{h_s V^2(N+1)}{3\sigma^2(N-1)} \tag{25}$$

P_e is normalized by the number of information bits in each symbol so that comparisons between different values of N will be more meaningful. Signal-to-noise ratio is given by the ratio of the variance $E[c^2]$ of transmitted levels to average noise power. Since twice as much power is transmitted (if V is unchanged) when partial response IV coding is used, $h_s = 1$ for uncoded sources, and $h_s = 2$ for partial response IV sources.

(*i*) *Error in Sampling Time*:

A convenient example of an ideal signal shaping characteristic $S(f)$ is shown in Fig. 2. $S(f)$ satisfies the Nyquist criterion and consequently if $R(f)$ of Fig. 1 were equal to $S(f)$ no intersymbol interference would be present at the receiver. Instead it is assumed that the sampling time is in error by ρT seconds, so that $R(f)$ is given by

$$R(f) = S(f)e^{-j2\pi f\rho T}. \tag{26}$$

Figure 3 shows the vector of error samples $\mathbf{e}$ of equation (14)

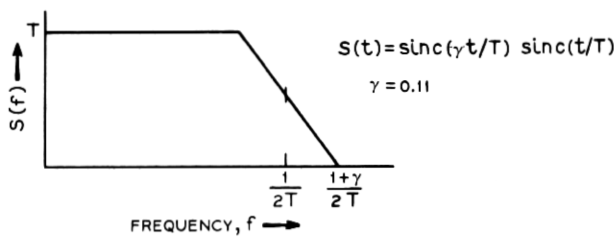

Fig. 2—Ideal signal shaping characteristic.

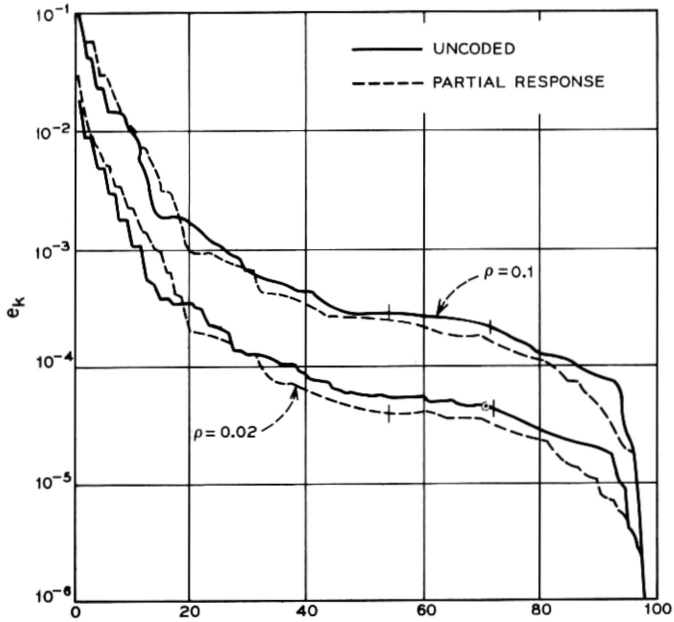

Fig. 3—Error samples e_k for timing error.

for $\rho = 0.02$ and $\rho = 0.1$, and for coded and uncoded sources. The samples fall off rapidly in each case, and the significant samples are somewhat larger for partial response IV coding. In order to include 99 percent of D, six blocks of nine samples must be used in the partial response IV case, and eight blocks of nine in the uncoded case. Figure 4 shows the resulting probability density functions for an uncoded source with timing error $\rho = 0.1$. (In the quantization process $N_1 = 40$ so that 81 locations were used in the interval $[0, V \cdot D]$.) The binary and 8-level pdf's are shown along with a Gaussian pdf having the same value of MSE. The Gaussian function is a reasonably good approximation to the binary case near the origin, but is of course very poor near $z = D \cdot V$, since the pdf's of intersymbol interference are strictly zero beyond this point. Figures 5 and 6 show curves of error-rate versus SNR for several values of timing error. An 8-level source is of course much more sensitive to timing error than is a binary source. In the absence of intersymbol interference a partial response IV signal requires about 3 dB more signal-to-noise ratio than an uncoded signal to achieve the same error-rate, under a constraint of equal average signal power. It is clear from the figures that this "SNR cost" need not be maintained when intersymbol interference is present. The partial response IV signal is more sensitive to timing error than is the uncoded signal, such that at $\rho = 0.15$ the cost is around 9 dB. However, for types of distortion other than timing error, this situation can be markedly reversed, since a partial response IV signal is rather impervious to distortions near dc or the Nyquist frequency.[1]

(*ii*) *Single Echo in the Channel:*

Echoes frequently occur in transmission channels, generated for instance by mismatched terminations. A sizable amount of analysis has been done on echo theory in relation to system distortion.[1,12] Here a single echo of amplitude g and relative delay βT is assumed to occur in the ideal channel of Fig. 2, so that

$$R(f) = S(f)(1 + ge^{-i2\pi f\beta T}). \tag{27}$$

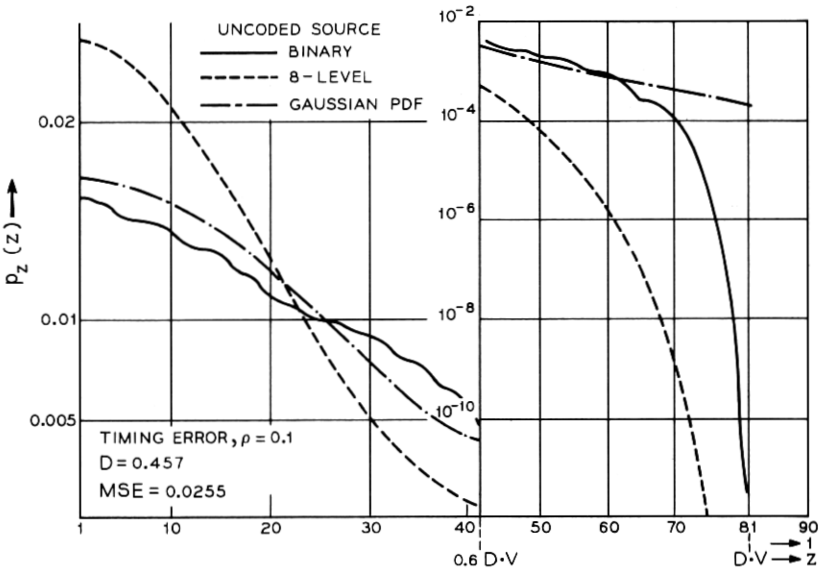

Fig. 4—PDF of intersymbol interference.

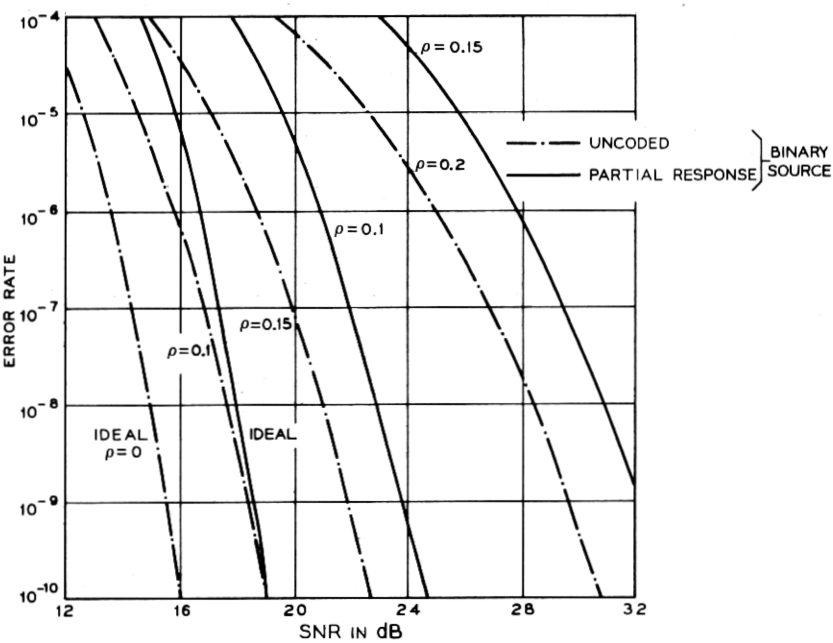

Fig. 5—Error rate vs SNR for timing error.

Because $s(t)$ is symmetrical and introduces no intersymbol interference, an echo at βT is identical in its effect to an echo at $(\pm\ \beta\ \pm\ n)T$ for any integer n. For instance, echoes at $\beta =$ $\pm$ 0.4, $\pm$ 3.4 $\pm$ 0.6 all have identical error sample vectors. Consequently one need only consider values of β between 0 and 0.5.

Figure 7 shows curves of error rate versus SNR for echoes of various strengths, and for an 8-level partial response IV coded source. Direct calculation shows that, because of the narrow excess bandwidth γ of $S(f)$ in Fig. 2, the sample error variance MSE of equation (10) is essentially independent of echo position. Echo strengths were chosen to yield convenient values of MSE = 0, 0.001, 0.003, and 0.005. It is clear that echo location has a strong effect on the error rate curves. Echoes at β = .5 yield the largest error rates. The error rate does not necessarily vanish as SNR $\rightarrow \infty$, as indicated by the asymptote labelled "isi alone."

(*iii*) *Equal Error Samples*:

Consider the error sample vector **e** consisting of n identical elements, $e_k = u$, $i = 1, \cdots, n$. A considerable amount of insight is gained by examining this case, although it is unlikely that any physical system would give rise to such a vector of error samples. In addition, one can derive the pdf analytically for the binary case providing a useful check on the computer method.

The error voltage z of equation (15) clearly has the value $uV(n - 2r)$ if and only if r of the symbols a_k equal $-V$, and the rest equal V. Such an event occurs with probability $\binom{n}{r} 2^{-n}$ so that z has the pdf

$$p_z(e) = \sum_{r=0}^{n} \binom{n}{r} 2^{-n} \delta(e - uV(n - 2r)), \tag{28}$$

and the error probability follows immediately using the method

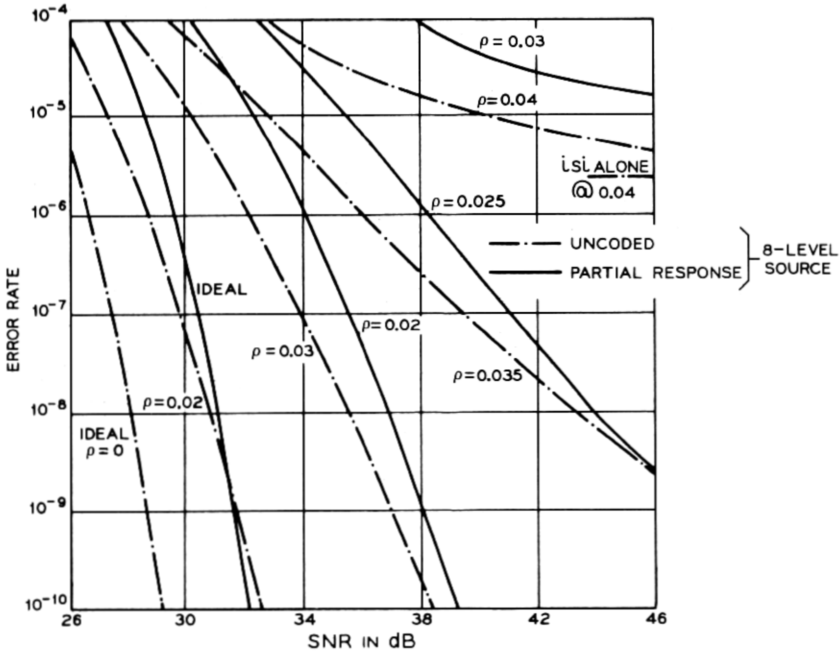

Fig. 6—Error rate vs SNR for timing error.

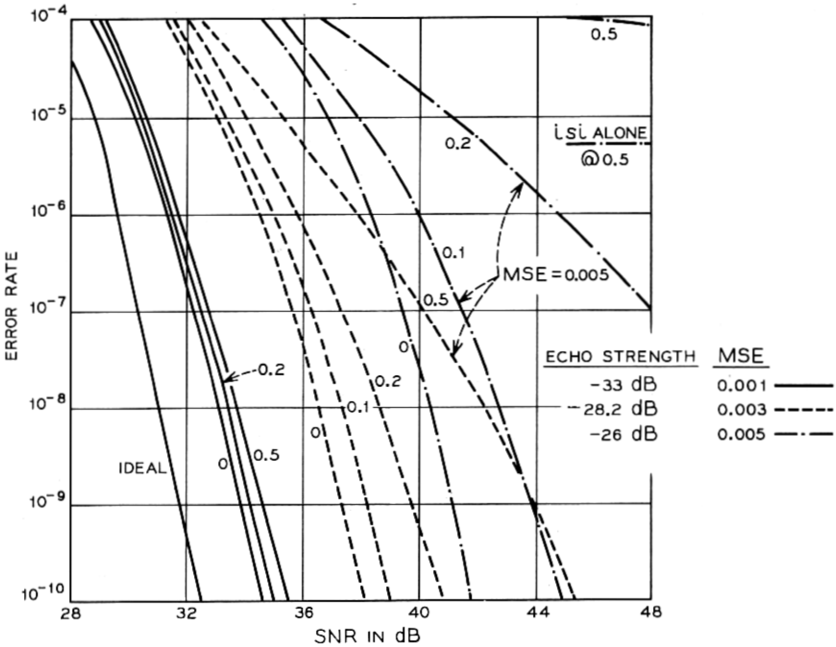

Fig. 7—Error rate vs SNR for single echo.

of Section 3.4:

$$P_e = \sum_{r=0}^{n} \binom{n}{r} 2^{-n} Q[\sqrt{\text{SNR}}\,[1 - u(n - 2r)]]. \tag{29}$$

The case $n = 1$ is the same as an echo positioned at $\beta T = 0$. Figure 8 shows curves of error rate versus SNR as a function of n. (Curves formed using equation (29) and the method just given were indistinguishable.) The samples have size $u = 0.5/n$ so that $d = 0.5$ for all n, while MSE $= 0.25/n$. From the curves it is clear that a few large error samples are more degrading than many small ones for the same value of D. This is intuitively reasonable since only very special source sequences $\{a_k\}$ will cause the error voltage z to be large when $\mathbf{e}$ consists of many small error samples, and such sequences occur with very small probability. Figure 9 shows similar curves, where now the samples have sizes $0.1/\sqrt{n}$ or $0.2/\sqrt{n}$, so that MSE is constant at 0.01 or 0.04. For MSE $= 0.01$ error rate is somewhat insensitive to the number of error samples, because the increasing value of $D = 0.1\sqrt{n}$ with n is offset by the decreasing probability that sequences $\{a_k\}$ yielding large errors will occur. However, as n increases for MSE $= 0.04$, D approaches 1, the value at which errors occur due to intersymbol interference alone, ($D = 1$ for $n = 25$). The offset in probability is not strong enough here, and error rate is very sensitive to n.

V. ACCURACY OF THE METHOD AND THE SALTZBERG BOUND

The applications of quantization at various steps in the computation shift the locations of some of the delta-functions slightly in one direction or the other, introducing some error in the t_i locations of equation (17). It would be extremely difficult to estimate analytically the error resulting in the value of P_e. Instead an empirical check was made over several channel characteristics. Figure 10 shows how error rate converges with increasing N_1 for a typical example of channel distortion. Convergence was typically most rapid for $k = 0.6$. For high values of SNR uniform quantization ($k = 0.5$) was noticeably

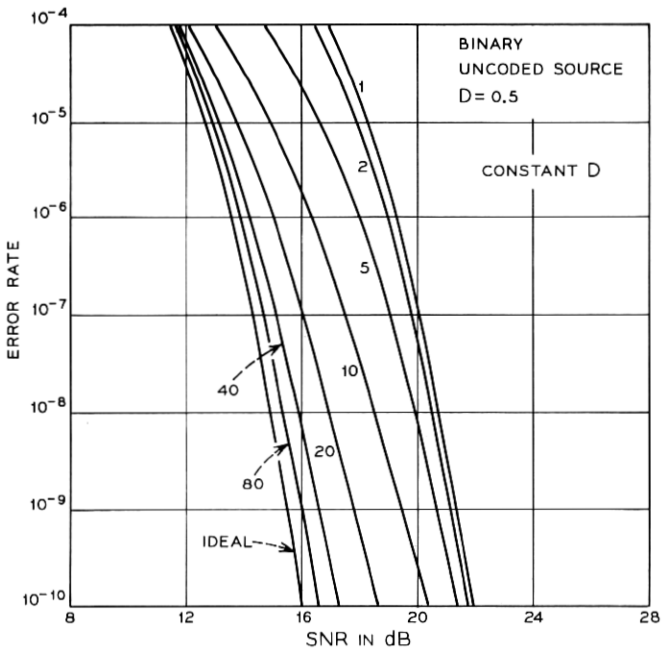

Fig. 8—Error rate for equal error samples.

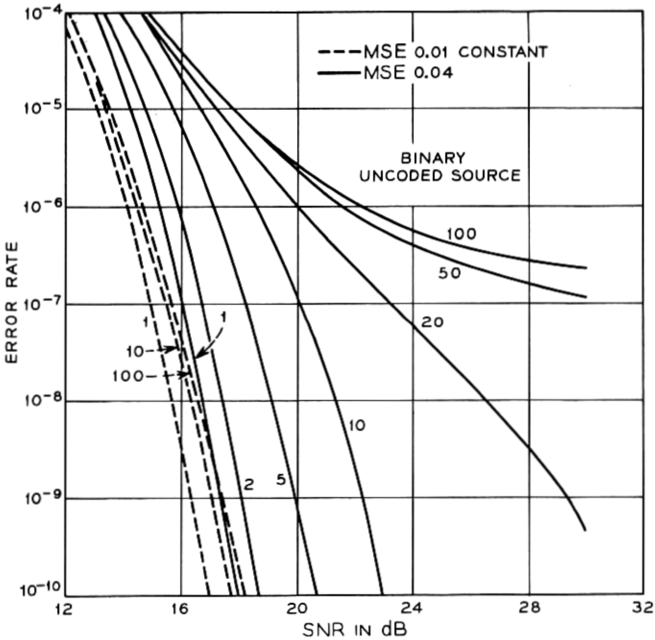

Fig. 9—Error rate for equal error samples.

inferior. Because computation time grows rapidly with N_1, values of $N_1 = 40$ and $k = 0.6$ were used in the examples of Section IV. With these values the algorithm is acceptably rapid, even on time-sharing facilities.

A much faster and simpler scheme was proposed by Saltzberg.[4] This technique found an upper bound for the error probability, but it was not clear from the theory how tight a bound was being computed. Figures 11 and 12 compare the error rate using the method given above with the upper bound found using this technique. For binary transmission the bound is between one and three orders of magnitude above the true value. For 8-level transmission the bound is less tight. The discontinuities in the bound versus signal-to-noise ratio occur due to the optimum partitioning of the error samples into two sets in the upper-bound method.

(*i*) *Choice of the Block Size in the Method*:

Given M error samples (the M "significant" samples from the original set of N_s), it was desired to find the binary pdf of z in equation (15). To do this, the M samples were partitioned into M/K blocks of K samples each. Consequently one must compute M/K pdf's based on sets of K samples, and perform $M/K - 1$ convolutions of these pdf's to obtain the binary pdf of z. The speed of the convolution process depends on $2N_1 + 1$, the number of quantization levels. By direct examination of the algorithm one finds: (*i*) that to form a pdf trying all combinations of K samples requires 2^K multiplications; (*ii*) that each convolution requires $5(2N_1 + 1)^2$ multiplications. The total number of multiplications is therefore

$$N_{\text{mult.}} = 2^K M/K + (M/K - 1)5(2N_1 + 1)^2. \tag{30}$$

For example, if $N_1 = 40$ and $M = 50$, a search shows that $N_{\text{mult.}}$ has a minimum with respect to K of 1.2×10^5 at $K = 12$. The saving in computation effort is dramatic: if no decomposition were used ($K = M$), $N_{\text{mult.}}$ would be 1.1×10^{15}. More generally, examination of equation (30) reveals that if $M/K \gg 1$ then $N_{\text{mult.}}$ varies linearly with M. Therefore $N_{\text{mult.}}$ has a minimum

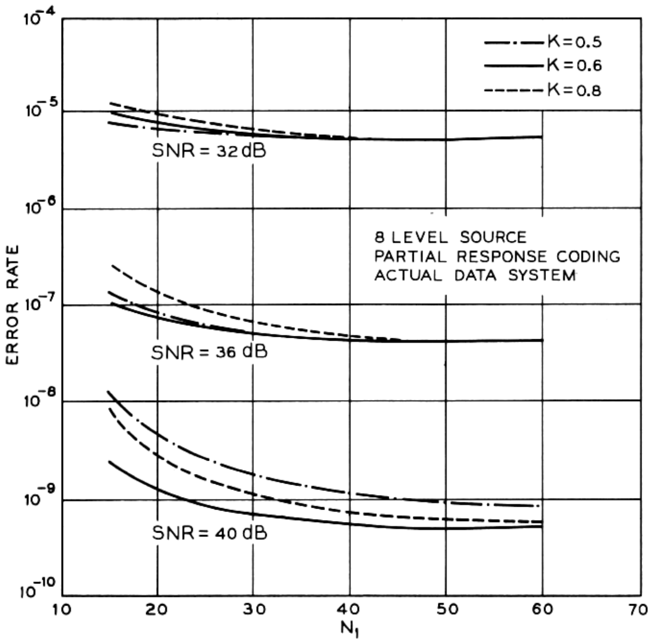

Fig. 10—Convergence of error-rate values.

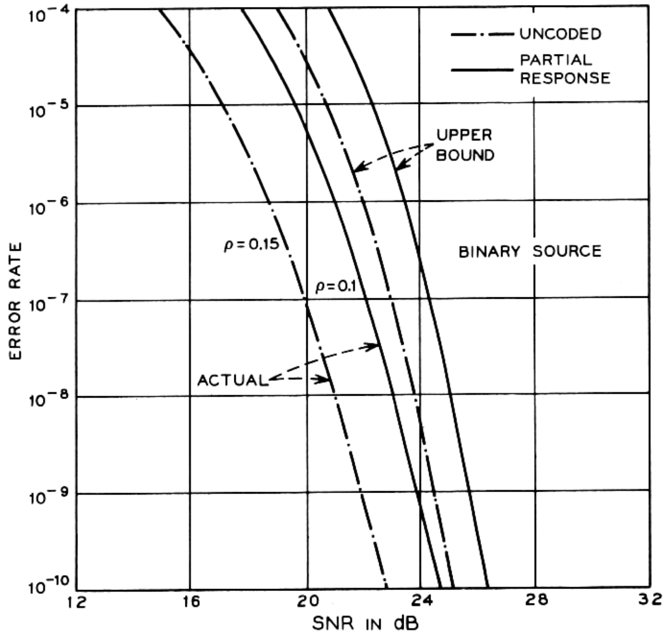

Fig. 11—Comparison with upper bound.

with respect to K that depends only on N_1 . This minimum is $K = 12$ for $N_1 = 40$, and $K = 13$ for $N_1 = 60$, although $N_{\text{mult.}}$ is rather insensitive to K for $K\epsilon$ (8, 15).

VI. CONCLUSIONS

A method has been developed that accurately computes the probability of error for a multilevel digital signal contaminated by intersymbol interference and noise. The method constructs a close approximation to the probability density function for the intersymbol interference, which can provide useful insight into the nature of the interference. The method is rapid enough for use on time-sharing facilities, and can provide a useful tool to aid the analysis and design of digital communication systems.

Starting with the transfer function or impulse response of the channel, a set of error samples is found, and the method operates on these to form the probability density function of intersymbol interference. Partial response coding of the source is easily included if desired. The formidable size of the computation that has heretofore blocked direct calculation of this probability density function is circumvented by partitioning the problem into a sequence of easily handled computations. Quantization keeps the total number of elements in the probability density function within reasonable limits, and the density function for binary transmission is used repeatedly to form the density function for 4-, 8-, 16-, ··· level transmission.

VII. ACKNOWLEDGMENT

The author wishes to acknowledge the large contribution of W. E. Norris who developed several of the key ideas in the method described here. Also, the aid and encouragement of John F. Gunn is greatly appreciated.

APPENDIX

Extension to Other Coding Types

The method described above considers only two coding types: the uncoded and partial response IV cases. With slight adjustments in the method one can also treat other interesting cases. The class of coders

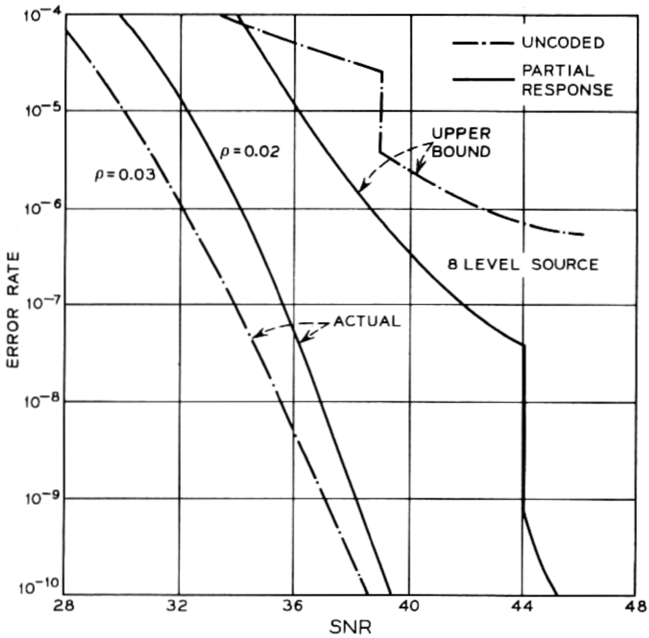

Fig. 12—Comparison with upper bound.

considered is that of Kretzmer[9,13]: each symbol c_k is related to the input symbols $\{a_k\}$ by

$$c_k = b_0 a_k + b_1 a_{k-1} + b_2 a_{k-2} + \cdots + b_v a_{k-v} , \tag{31}$$

where the elements b_i of the vector **b** are integers. (The uncoded case is then $b_0 = 1$, all other $b_i = 0$). Kretzmer tabulates five partial response classes which have proven most interesting: (I) $\mathbf{b} = (1, 1)$; (II) $\mathbf{b} = (1, 2, 1)$; (III) $\mathbf{b} = (2, 1, -1)$; (IV) $\mathbf{b} = (1, 0, -1)$; and (V) $\mathbf{b} = (-1, 0, 2, 0, -1)$. Other simple types exist as well, such as $\mathbf{b} = (1, 0, 0, 0, 0, 0, 0, 0, 0, -1)$, a generalization of class IV partial response that has five "lobes" in the signal spectrum.[9]

To see how codes of the type in equation (31) could be included in the method, note that there are only three places where the coding scheme used makes a difference: (*i*) in forming the coded error samples, as in equation (7), (*ii*) in finding P_e from the pdf, as in equation (8), and (*iii*) in evaluating signal-to-noise ratio, as in equation (25).

(*i*) The coded error samples are easily found for any vector **b**. Using equation (30) in equation (4), and manipulating

$$\begin{aligned} z &= \sum_k \sum_{m=0}^{v} b_m a_{k-m} x_k + n \\ &= \sum_{m=0}^{v} b_m \sum_j a_j x_{j+m} + n \\ &= \sum_j a_j e_j + n, \end{aligned}$$

where

$$e_j = \sum_{m=0}^{v} b_m x_{j+m} \tag{32}$$

are the coded error samples.

(*ii*) Once the pdf of isi has been found from the coded error samples, the error probability follows as in equation (8). To use signal power efficiently a coder will be chosen so that the transmitted voltage levels are equally spaced. Thus one need only evaluate the probability of an outer level in order to adapt equation (8) to this larger class of coders. An outer level of c_k occurs only if all relevant a_k have their maximum size and correct polarity. For N-level a_k symbols then, if M of the coefficients b_i in equation (31) are nonzero, each outer level has probability N^{-M}. Thus M replaces h in equation (8).

(*iii*) Finally, since the symbols a_k are statistically independent and have mean square value $E[a^2]$ as in equation (25), it is a simple matter to show that the symbols c_k have mean square value $E[c^2] = h_s \, E[a^2]$, where

$$h_s = \sum_{m=0}^{v} b_m^2 \, . \tag{32}$$

Therefore h_s replaces h_s in equation (25) in the evaluation of signal-to-noise ratio.

REFERENCES

1. Lucky, R. W., Salz, J., Weldon, E. J., Jr., "Principles of Data Communications," New York: McGraw Hill Book Co., 1968.
2. Lucky, R. W., "Automatic Equalization for Digital Communication," B.S.T.J., *44,* No. 4, part 1 (April 1965), pp. 547–588.
3. Gersho, A., "Adaptive Equalization of Highly Dispersive Channels for Data Transmission", B.S.T.J. *48,* No. 1, part 1 (January 1969), pp. 55–70.
4. Saltzberg, B. R., "Intersymbol Interference Error Bounds with Application to Ideal Bandlimited Signaling", IEEE Trans. Inform. Theory *IT-14* (July 1968), pp. 563–568.
5. Aein, J. M., and Hancock, J. C., "Reducing the Effects of Intersymbol Interference with Correlation Receivers," IEEE Trans. Inform. Theory, *IT-9,* No. 3 (July 1963), pp. 167–175.
6. Aaron, M. R., and Tufts, D. W., "Intersymbol Interference and Error Probability," IEEE Trans. Inform. Theory, *IT-12,* No. 1 (January 1966), pp. 26–34.
7. Ho, E. Y., and Yeh, Y. S., "A New Approach for Evaluating the Error Probability in the Presence of Intersymbol Interference and Additive Gaussian Noise," B.S.T.J., *49,* No. 9, part 3 (November 1970), pp. 2249–2265.
8. Celebiler, M. I., and Shimbo, O., "Intersymbol Interference Considerations in Digital Communications," Int. Conf. on Commun., *ICC 1970,* San Francisco (June 8, 1970), pp. 8.1–8.10.
9. Kretzmer, E. R., "Generalization of a Technique for Binary Data Communication," IEEE Trans. on Commun. Technology, *COM-14,* No. 1 (February 1966), pp. 67–68.
10. Norris, W. E., unpublished work, Bell Telephone Laboratories.
11. Cooley, J. W., Lewis, P. A., Welch, P. D., "Application of the Fast Fourier Transform to Computation of Fourier Integrals, Fourier Series, and Convolution Integrals," IEEE Trans. Audio and Electroactoustics, *AV-15* (June 1967), pp. 79–84.
12. Sunde, E. D., "Theoretical Fundamentals of Pulse Transmission," I and II, B.S.T.J., *33,* Nos. 3 and 4, parts 1 and 2 (May, July, 1954), pp. 721–88, 987–1010.
13. Gerrish, A. M., Howson, R. D., "Multilevel Partial-Response Signaling," IEEE Int. Conf. on Commun., Minneapolis, Minnesota, June 12, 1967, p. 186.