THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 8, October 1975
Printed in U.S.A.

Faster-Than-Nyquist Signaling

By J. E. MAZO

(Manuscript received March 27, 1975)

The degradation suffered when pulses satisfying the Nyquist criterion are used to transmit binary data in noise at supraconventional rates is studied. Optimum processing of the received waveforms is assumed, and attention is focused on the minimum distance between signal points as a performance criterion. An upper bound on this distance is given as a function of signaling speed. In particular, the pulse energy seems to be the minimum distance up to rates of transmission 25 percent faster than the Nyquist rate, but not beyond.

Some mathematical aspects related to the above problem are also considered. In particular, the minimum distance is rigorously shown to be nonzero for all transmission rates. This is tantamount to showing that, in the singular case of linear prediction, perfect prediction cannot be approached with bounded prediction coefficients.

I. INTRODUCTION

The use of Nyquist pulses

$$g(t) = \frac{\sin(\pi t/T)}{(\pi t/T)}$$

to send binary (or multilevel) data without intersymbol interference over a channel of bandwidth $W = (1/2T)Hz$ is classic. If we assume that one receives the pulse train

$$u(t) = \sum_{n=N_1}^{N_2} a_n g(t - nT), \quad a_n = \pm 1, \text{ independently,} \tag{1}$$

in additive white gaussian noise of two-sided spectral density $N_0/2$, then the optimum detector has a bit-error rate P_e given by

$$P_e = Q\left(\frac{2\sqrt{E}}{\sqrt{2N_0}}\right), \tag{2}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \equiv \tfrac{1}{2} \operatorname{erfc} \frac{x}{\sqrt{2}}, \tag{3}$$

erfc $(\cdot)$ denoting the co-error function, and E being the energy in the pulse $g(t)$. In our case, $E = T$. Asymptotically, for large signal-to-noise ratios, (2) becomes

$$P_e \sim \frac{1}{2}\sqrt{\frac{N_0}{\pi E}} \exp\left(-\frac{E}{N_0}\right). \tag{4}$$

We now address the following question: Suppose that in transmitting (1) we obtain a performance from (2) that is more than satisfactory. Thus, we may have a P_e of 10^{-6} or 10^{-7} when 10^{-5} would be adequate. To what extent can we trade this "excess performance" for speed by replacing T by $T' < T$ in (1), while keeping transmitted power constant? In other words, we still use pulses

$$g(t) = B \frac{\sin(\pi t/T)}{(\pi t/T)}, \tag{5}$$

but send them at intervals $T' < T$. We call this faster-than-Nyquist transmission and shall characterize T' by writing $T' = \rho T$, $0 < \rho < 1$. A particular motivation for this problem is to mathematically model, in a simple way, what would happen if voice-band telephone channels are "pushed" to their limits with more rapid transmission of pulses than has been conventional.

While simple detectors that match filter and sample can still be used for faster-than-Nyquist transmission, their performance is suboptimum.[1] We are concerned here with optimum detectors. Since exact analysis of nonlinear detectors is not presently feasible, we choose to give our detectors the benefit of the doubt and work rather with lower bounds to P_e. Nevertheless, interesting results can be obtained regarding the trade-off considered here. To see why degradation in error rate is inevitable, note that (2) is the well-known matched filter bound for antipodal pulses, each of energy E, which must bound performance for bit detection with a sequence of (perhaps interfering) pulses. On the other hand, as T' decreases, pulses are sent faster and the energy E in each pulse must be decreased in direct proportion so that the power E/T' is kept constant. This is an immediate unavoidable element in performance degradation, and may be regarded as a "fair" trade-off. Another cause of degradation is the degree to which the optimum detector can cope with the interference among pulses, i.e., the fact that the performance will drop below that of (2). Here, bounds other than (2) are useful, and in fact are the first item taken up in the next section.

II. DISCUSSION OF LOWER BOUND FOR ERROR RATE

Assuming (1) is received in white noise and an optimum detector is used for detecting the kth bit, a lower bound on the chance of making an error on this kth bit will now be derived. Since the data a_n are independent, this bound also serves for any sequence (1) starting at $n = N_1' \leqq N_1$, and ending at $n = N_2' \geqq N_2$. We begin with the fact that, for a binary hypothesis problem with equal *a priori* probabilities and having $p_+(x)$ or $p_-(x)$ as the two probability densities of the received signal x under the two respective hypotheses, one way[2] to write the probability of error is

$$P_e = \frac{1}{2} \int \min \left[p_+(x), p_-(x)\right] dx. \tag{6}$$

If we let $u_i^{\pm}(t)$ be a particular one of the equiprobable 2^N signals in (1), $N = N_2 - N_1$, which have ± 1 in the kth position, then formally

$$p_{\pm}(x) = \frac{1}{2^N} \sum_{i=1}^{2^N} p_{\pm}^i(x), \tag{7}$$

where $p_{\pm}^i(x)$ is the density of the observations conditioned on the entire sequence. Thus,

$$\begin{aligned} P_e &= \frac{1}{2} \cdot \frac{1}{2^N} \cdot \int \min \left(\sum_{i=1}^{2^N} p_+^i(x), \sum_{j=1}^{2^N} p_-^j(x) \right) dx \\ &\geqq \frac{1}{2} \frac{1}{2^N} \sum_{i=1}^{2^N} \int \min \left[p_+^i(x), p_-^{j(i)}(x)\right] dx. \end{aligned} \tag{8}$$

In writing (8), we have made use of the fact that the minimum of two sums with an equal number of terms is at least as large as the sum of the minimum of the two ith terms of each series. Of course, each series can be arranged in any permuted order before the pair-wise minimum is taken and, thus, the pairings i with $j(i)$ are indicated in (8) to allow for this permutation. Now

$$\frac{1}{2} \int \min \left[p_+^i(x), p_-^{j(i)}(x)\right] dx \tag{9}$$

is the probability of error with two fixed signals and has the well-known evaluation

$$Q\left(\frac{d[i, j(i)]}{\sqrt{2N_0}} \right), \tag{10}$$

where

$$d^2(i, j) = \int_{-\infty}^{\infty} \left[u_+^i(x) - u_-^j(t)\right]^2 dt \tag{11}$$

is the "distance" between two sequences (1) which differ in the kth position. Equation (8) then reads

$$P_e \geqq \frac{1}{2^N} \sum_{i=1}^{2^N} Q\left(\frac{d[i, j(i)]}{\sqrt{2N_0}} \right) \tag{12}$$

for any set of pairings $[i, j(i)]$. The bound (12) is intimately related to Forney's lower bound,[3] although our derivation is quite different. Forney's bound in the present situation reads

$$P_e \geqq p_m Q\left(\frac{d_{\min}}{\sqrt{2N_0}} \right), \tag{13}$$

where $d_{\min}$ is the minimum distance between signals (1) which differ in the kth position, and p_m is the probability that a sequence chosen at random has a sequence with opposite polarity in the kth position at distance $d_{\min}$. Equation (12) can be made to yield something like (13). Thus, in (12) discard all terms except for those pairings $[i, j(i)]$ such that $d[i, j(i)] = d_0$. Then (12) implies

$$P_e \geqq \frac{\text{no. of pairings}}{2^N} Q\left(\frac{d_0}{\sqrt{2N_0}} \right). \tag{14}$$

The coefficient in front of the Q function corresponds to the probability coefficient in (13). Choosing $d_0 = d_{\min}$ yields (13), but when we will not be able to find $d_{\min}$, eq. (14) will serve our purpose.

III. ESTIMATING THE MINIMUM DISTANCE

Clearly, in (14) we should like to find the smallest d_0 to maximize the lower bound, provided the coefficient is not too small. In our problem, $d^2_{\min}$ is given by

$$\frac{d^2_{\min}}{4E} = \inf_{N;\ \{a_l = \pm 1, 0\}} \frac{1}{2\pi\rho} \int_{-\rho\pi}^{\rho\pi} \left| 1 - \sum_{l=1}^{N} a_l e^{il\theta} \right|^2 d\theta, \tag{15}$$

where we have normalized by dividing by the pulse energy E. The expression (15) comes from taking the Fourier transform of (11) and manipulating the resulting expression slightly. We note particularly that in (15) only positive values of l need be considered, since

$$\left| e^{iK\theta} \left(1 - \sum_{\substack{l=-K \\ l \neq 0}}^{M} a_l e^{il\theta} \right) \right|^2 = \left| 1 - \sum_{l=1}^{M+K} b_l e^{il\theta} \right|^2$$

if $a_{-K} \neq 0$. We have set $b_l = -a_{-K} a_{s-K}$ if $l \neq K$ and $b_l = -a_{-K}$ if $l = K$.

We cannot claim to have found the minimum value of (15). However, a simple numerical effort has yielded the results for $d_0^2/4E$ shown in Fig. 1, where d_0 refers to the smallest distance we have found. We note in particular that d_0 is the pulse energy for ρ decreasing from 1

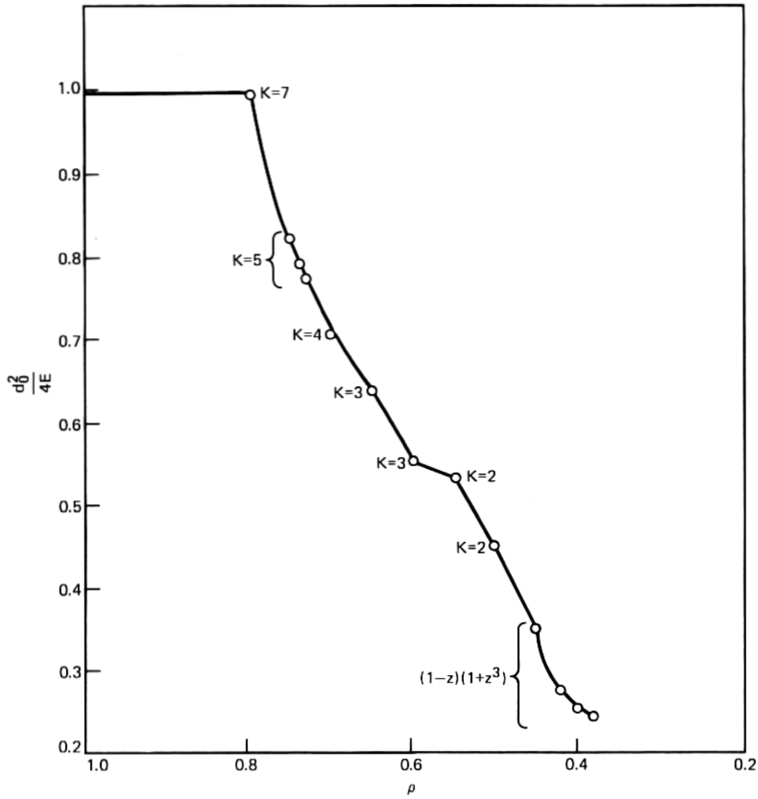

Fig. 1—The smallest distances between signal sequences that we have found are shown here for different values of signaling rate. Labeling a point by K indicates that the polynomial is

$$p(z) = 1 + \sum_{j=1}^{K} (-1)^j z^j.$$

to 0.8, or, in other words, for rates exceeding the Nyquist rate by 25 percent [percentage of excess $= 100(1/\rho - 1)$]. Thus, $d_{\min}^2/4$ cannot be the pulse energy for $\rho < 0.8$ for this problem. By the time ρ has decreased to 0.5, $d_0^2/4E$ has dropped to 0.465. (G. J. Foschini has informed the author that the use of the polynomial $p(z) = 1 - z + z^3 - z^4 + z^6 - z^7$, $z = \exp(i\theta)$, results in the value 0.410 for $d_0^2/4E$ at $\rho = 0.5$.) Except for some points in the neighborhood of $\rho = 0.4$, the values for d_0^2 have been obtained by considering numerically the best value of K which minimizes, for not too large K,

$$\frac{1}{2\pi\rho} \int_{-\rho\pi}^{\rho\pi} \left| 1 + \sum_{l=1}^{K} (-1)^l e^{il\theta} \right|^2 d\theta. \tag{16}$$

These points are labeled with the appropriate value of K in Fig. 1.

Somewhat surprisingly, the larger values of K are responsible for decreasing d_0 initially ($K = 7$ at $\rho = 0.8$), and then K gradually becomes smaller ($K = 2$ at $\rho = 0.5$). The value obtained with $K = 1$ always was suboptimum, as was the limiting value of (16) when $K \rightarrow \infty$, which is easily shown to be

$$\frac{1}{\pi\rho} \tan \frac{\rho\pi}{2}. \tag{17}$$

Why were the sequences given in (16) deemed to be of interest in the first place? The most interesting reason stems from the following argument. If one considers the Fourier transform of a doubly infinite pulse sequence like (1) when pulses are being sent faster than Nyquist and when the special case of the alternating sequence $a_n = (-1)^n$ is being sent, one finds that the Fourier transform consists of delta functions spaced at all odd multiples of π/T', that is, the Fourier transform is out-of-band, which suggests zero received energy. Actually, the doubly infinite model and its δ-function Fourier transforms are idealizations representing limiting behavior for signals consisting of pulses extending from $(-N, N)$ and N becoming large. We are really concerned with limiting behavior of the energy contained in the frequency interval $(-\pi/T, \pi/T)$, with $T > T'$, and evidently for the present case, if $S_N(\omega)$ is the Fourier transform of the truncated pulse sequence,

$$\lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} |S_N(\omega)|^2 d\omega \neq \frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} |\lim S_N(\omega)|^2 d\omega = 0. \tag{18}$$

In spite of the above subtlety, however, sequences which are alternating at least over part of their range are interesting and one might expect difficulty distinguishing between one such sequence and its negative.

In addition to the normalized distances given in Fig. 1, Fig. 2 plots the numerical values of lower bounds computed from expression (14), as well as the matched filter bound. These curves all assume constant power. Curves with initial ($\rho = 1$) error rates with 10^{-5} and 10^{-7} are chosen as examples in Fig. 2. In both cases, an order of magnitude of degradation in error rate is seen for a 25-percent increase in bit rate ($\rho = 0.8$) using only the matched filter bound. Decreasing ρ further on the 10^{-7} curve illustrates further degradations using (14) with an appropriate value of K. These bounds do not show a departure from the matched filter bound for as small a value of ρ as Fig. 1 would suggest, because the coefficient $1/2^K$ to be used in (14) swamps the effect of the decreasing "minimum" distance. For the 10^{-5} curve, this effect extends to even smaller ρ and no lower bound other than the matched filter one is shown for that case.

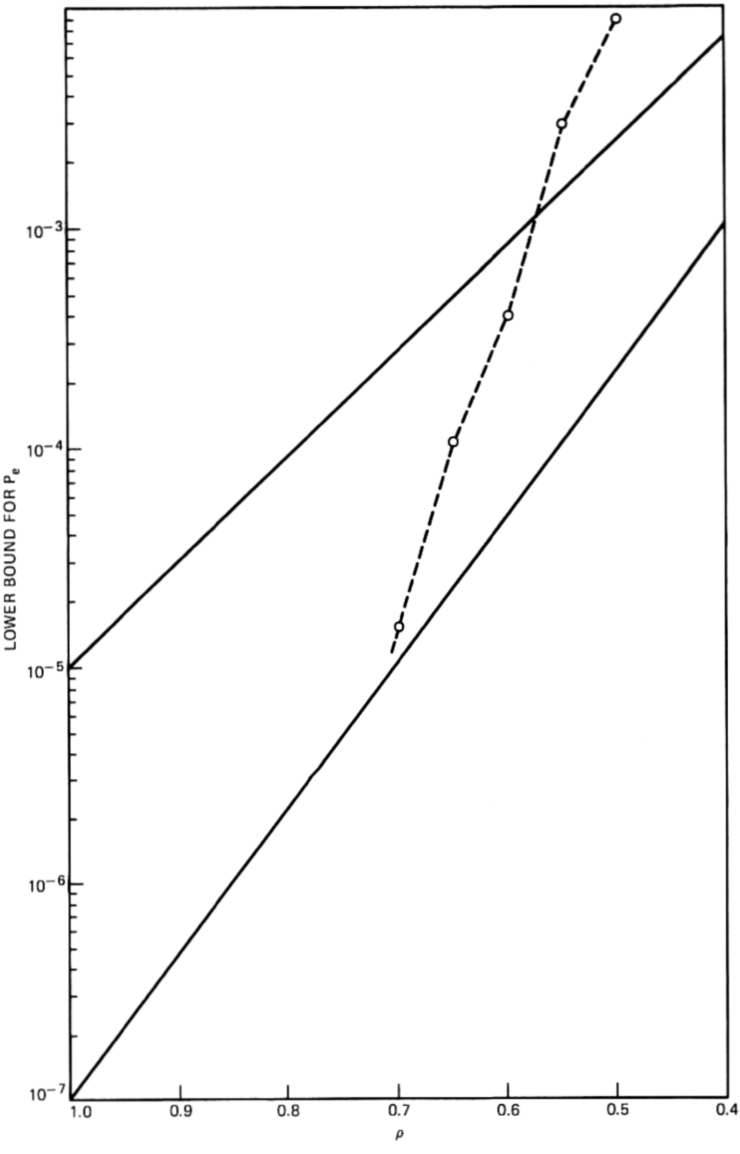

Fig. 2—Lower bounds on error rate vs signaling speed for two initial ($\rho = 1$) cases. The solid curves are both matched filter bounds. The dashed curve is based on minimum-distance considerations and applies to the 10^{-7} case. All curves are drawn for constant power.

IV. TWO MATHEMATICAL QUESTIONS

As we have already emphasized, the infimum of the right member of (15) over all the indicated trigonometric polynomials with ± 1, 0 coefficients is not displayed in Fig. 1. Figure 1 simply shows the smallest values we have found. Next, we want rigorously to establish here that $d^2_{\min} \neq 0$ if $\rho \neq 0$. Note that this would not be the case if the coefficients a_l in (15) were allowed to be any real numbers. In fact, for any nonnegative function $f(\theta)$ with $\ln f(\theta) \in L_1(-\pi, \pi)$, we have the Szego theorem[4] which states

$$\inf_{N;\, a_l \text{ real}} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) \left| 1 - \sum_1^N a_l e^{il\theta} \right|^2 d\theta = \exp \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln f(\theta) d\theta. \quad (19)$$

Expressions such as (19) occur, in particular, in linear prediction theory.

In our case, $f(\theta) = 0$ if $|\theta| > \rho\pi$ and $\ln f(\theta)$ is not L_1, but the appropriate limit of (19) indicates zero to be the infimum, which is the correct answer.[4] Thus, there is some cause to wonder if $d^2_{\min}$ as defined in (15) is zero as well. We shall in fact show it is slightly more.

Theorem 1: Let β be any positive (finite) real number and require $|a_l| \leqq \beta$, $l = 1, 2, \cdots$. Then

$$\inf_{N;\, \{a_l\}} \frac{1}{2\pi} \int_{-\rho\pi}^{\rho\pi} \left| 1 - \sum_1^N a_l e^{il\theta} \right|^2 d\theta > 0, \quad \rho \neq 0. \quad (20)$$

Proof: We first note that if there exists a sequence $\{p_n(\theta)\}_{n=1}^{\infty}$ of trigonometric polynomials of the form

$$p_n(\theta) = \sum_{l=1}^{n} a_l(n) e^{il\theta}, \quad |a_l| \leqq \beta < \infty \quad (21)$$

such that

$$\frac{1}{2\pi} \int_{-\rho\pi}^{\rho\pi} |1 - p_n(\theta)|^2 d\theta \to 0, \quad (22)$$

then, for any $G(\theta) \in L_2(-\rho\pi, \rho\pi)$,†

$$\int_{-\rho\pi}^{\rho\pi} G(\theta) p_n(\theta) d\theta \to \int_{-\rho\pi}^{\rho\pi} G(\theta) d\theta. \quad (23)$$

This is simply a statement of the fact that if $p_n(\theta)$ converges strongly to unity, it also converges weakly to unity. Now it is easy to see from (23) and the form of $p_n(\theta)$ that

$$\beta \sum_1^{\infty} \left| \int_{-\rho\pi}^{\rho\pi} d\theta\, e^{in\theta} G(\theta) \right| \geqq \left| \int_{-\rho\pi}^{\rho\pi} G(\theta) d\theta \right| \cdot \quad (24)$$

Or, in other words, if

$$\beta < \sup_{G(\theta) \in L_2(-\rho\pi, \rho\pi)} \frac{\left| \int_{-\rho\pi}^{\rho\pi} G(\theta) d\theta \right|}{\sum_1^{\infty} \left| \int_{-\rho\pi}^{\rho\pi} e^{in\theta} G(\theta) d\theta \right|}, \quad (25)$$

† In addition to $G(\theta) \in L_2(-\rho\pi, \rho\pi)$ it will sometimes be convenient to regard $G(\theta) \in L_2(-\pi, \pi)$ but having support confined to $(-\rho\pi, \rho\pi)$.

then (22) cannot be true. In particular, if (25) holds with $\beta \geqq 1$, then $d_{\min}^2$ is strictly positive. Regarding $G(\theta) \in L_2(-\pi, \pi)$ but supported on $[-\rho\pi, \rho\pi]$, and calling

$$g(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\theta t}G(\theta)d\theta, \tag{26}$$

$$g_n \equiv g(n),$$

the right member of (25) contains the quantity

$$\frac{|g_0|}{\sum_1^\infty |g_n|}. \tag{27}$$

Clearly, we have a question concerning the sample values g_n at the nonnegative integers of a function whose bandwidth is strictly less than π. Normalizing (27) with $g_0 = 1$, (25) prompts the question: How small can $\sum_1^\infty |g_n|$ be? If it can be zero, then (25) would be true for any finite β. In fact, by Carlson's lemma,[5] which states that a band-limited function having a bandwidth less than π is uniquely determined by its sample values taken at integers along a half line, it follows that if $g_0 = 1$, then $\sum_1^\infty |g_n| \neq 0$. But Carlson's lemma does not say that $\sum_1^\infty |g_n|$ cannot be made arbitrarily small under these conditions. Lemma 1 (see below) shows that $\sum_1^\infty |g_n|$ can be arbitrarily small. Thus, the right member of (25) is infinity, implying the truth of Theorem 1.

An immediate corollary of Theorem 1 is that for the singular case of Szego's theorem [$f(\theta)$ vanishing on an interval] the infimum value of zero cannot be approached without using *unbounded* coefficients.

Lemma 1: Let $g(t)$ [not identically zero and $\in L_2(-\infty, \infty)$] have Fourier transform $G(\theta)$ supported on $(-\rho\pi, \rho\pi)$ for some fixed ρ, $0 < \rho < 1$. Denote the samples of $g(t)$ at the integers by g_n [as in eq. (26)], and fix the normalization of $g(t)$ by setting $|g_0| = 1$. Then

$$\inf \sum_1^\infty |g_n| = 0, \tag{28}$$

where the infimum is taken over all $g(t)$ having the indicated properties.

Proof: We begin with the simple, but crucial, remark that it is sufficient that there be, for any ρ, a function $h(t;\rho) \in L^2(-\infty, \infty)$ whose Fourier transform is supported on $(-\rho\pi, \rho\pi)$, such that $h(0, \rho) = 1$ and such that $\sum_1^\infty |h_n(\rho)|^2$ can be arbitrarily small.† This is sufficient,

† We are grateful to H. J. Landau for pointing this out. Landau has also supplied an independent proof of the above refinement to Carlson's lemma, which we give in the appendix.

because to make (27) large (for some fixed value of ρ) we would just need to take

$$g(t) = h^2\left(t, \frac{\rho}{2}\right) \tag{29}$$

for an appropriate $h(t, \rho/2)$. Clearly, $g(t)$ is band-limited to ρ and is $L^2(-\infty, \infty)$ because $h(t, \rho/2)$ is bounded:

$$h\left(t, \frac{\rho}{2}\right) = \frac{1}{2\pi} \int_{-\rho\pi/2}^{\rho\pi/2} H(\theta) d\theta \leqq \frac{1}{2\pi}\left(\rho\pi \cdot \int_{-\rho\pi/2}^{\rho\pi/2} |H(\theta)|^2 dt\right). \tag{30}$$

But can we really find an appropriate $h(t)$ such that

$$h_0 = 1, \quad \sum_1^\infty |h_n|^2 < \epsilon, \tag{31}$$

or, equivalently, can we find a real $h(t)$, band-limited to $(-\rho\pi, \rho\pi)$, such that

$$(h_0 - 1)^2 + \sum_1^\infty h_n^2 < \epsilon? \tag{32}$$

Indeed we can, and in fact the answer may be extracted from an article by Salz[6] which discusses mean-square decision feedback equalization. Salz, in Section V of his paper, considered the equalization problem for faster-than-Nyquist signaling. His minimization problem was of the form in (32) plus an added term for the noise variance; $h(t)$ corresponds to the output of the equalizer when one pulse of the form $\sin \rho\pi t/\rho\pi t$ is the input. He remarks, in the last sentence on page 1354 of his paper, that the quantity that corresponds to (32) plus added output noise variance goes to zero as the input noise variance decreases. Hence, if we choose $h(t)$ to be the output pulse of a decision-feedback equalizer whose taps have been optimized for the case of sufficiently small input noise, then (32) will be sufficiently small. Thus, Lemma 1 is proven.

The second question we discuss in this section is the rapidity with which the minimum distance decreases as ρ approaches zero. We develop this in Theorem 2.

Theorem 2:

$$\lim_{\rho \to 0} \frac{d_{\min}^2(\rho)}{\rho^k} = 0 \quad \textit{for any } k > 0. \tag{33}$$

Proof: The proof is a simple construction. Consider the polynomials

$$P_L(z) = \prod_{l=0}^{L} (1 - z^{2^l}). \tag{34}$$

Clearly, $P_L(z)$ has a zero of order $(L+1)$ at $z = +1$, and has ± 1 coefficients, with $P_L(0) = +1$. Now, for small ρ, the $(L+1)$st order zero at $z = 1$ implies

$$\frac{1}{2\pi} \int_{-\rho\pi}^{\rho\pi} |P_L(e^{i\theta})|^2 d\theta = 0(\rho^{2L+3}) \tag{35}$$

for all integer L. Equation (33) follows immediately.

Short of finding $d_{\min}^2$ exactly, there are a few mathematical questions that suggest themselves and that may be less difficult than the full problem. Thus, Fig. 1 prompts one to ask if there is a neighborhood of $\rho = 1$, where $d_{\min}^2/4$ is the pulse energy? Another question has to do with pulse design. Given that $G(\theta)$ is symmetric, positive, L_2, and supported on $(-\rho\pi, \rho\pi)$, is $G(\theta) =$ constant the best choice to maximize the minimum distance (subject to fixed pulse energy)?

V. ACKNOWLEDGMENTS

It is a pleasure to acknowledge many helpful discussions with M. A. Kaplan, H. J. Landau, B. F. Logan, and H. O. Pollak during this work. We acknowledge the contribution of B. F. Logan who supplied an early proof that $d_{\min}^2 \neq 0$ if $\rho > \frac{1}{2}$, approximately.

APPENDIX

Landau's Proof

In Section IV we present another proof that

$$\sup \frac{g_0^2}{\sum_{n=1}^{\infty} |g_n|^2} = \infty, \tag{36}$$

where the sup is taken over all $g(t) \in L^2(-\infty, \infty)$, which are band-limited to $(-\rho\pi, \rho\pi)$. Our proof in the text relied on the published results of work by Salz.[6] Here we give a self-contained, but more mathematical, proof of (36) which was developed by H. J. Landau.

Suppose (36) is not true, i.e., suppose that

$$\sum_{1}^{\infty} |g_n|^2 \geqq \frac{g_0{}^2}{k} > 0 \quad \text{for all } g(t) \text{ of BW} = \rho\pi. \tag{37}$$

Then,

$$|g_0|^2 \leqq k \sum_{1}^{\infty} |g_n|^2. \tag{38}$$

From Carlson's lemma, g_0 is a linear functional on the l_2 sequence $\{g_1, g_2, \cdots, g_k, \cdots\}$ and, from (38), this linear functional is bounded. Therefore, by the standard Riesz representation† for bounded linear

† Not all l_2 sequences $\{g_i\}$ give rise to an appropriate $g(t)$, and hence, the linear functional g_0 is not defined on all of l_2. Therefore, before using the Riesz theorem, the Hahn-Banach theorem should be invoked to extend g_0 to a bounded linear functional on all of l_2.

functionals, we may write

$$g_0 = \sum_1^\infty b_\nu g_\nu, \quad \sum_1^\infty b_\nu^2 < \infty, \tag{39}$$

where the b_ν do not depend on $g(t)$. We now consider the function

$$p(z) = 1 - \sum_1^\infty b_n z^n, \tag{40}$$

which is analytic for $|z| < 1$. For any $G(\theta) \in L_2(-\rho\pi, \rho\pi)$, we may write, using (39),

$$\begin{aligned} \int_{-\rho\pi}^{\rho\pi} G(\theta)d\theta &= \sum_1^\infty b_n \int_{-\rho\pi}^{\rho\pi} e^{in\theta}G(\theta)d\theta \\ &= \int_{-\rho\pi}^{\rho\pi} \left(\sum_1^\infty b_n e^{in\theta}\right) G(\theta)d\theta. \end{aligned} \tag{41}$$

Therefore,

$$\lim_{|z|\to 1} \int_{-\rho\pi}^{\rho\pi} \left(1 - \sum_1^\infty b_\nu z^\nu\right) G(\theta)d\theta = 0 \tag{42}$$

for all $G(\theta) \in L_2(-\rho\pi, \rho\pi)$.† By the completeness of L_2, we must have $1 - \sum_1^\infty b_n e^{in\theta} = 0$ a.e. on $(-\rho\pi, \rho\pi)$. Since the radial limit of the H_2 function $p(z)$ vanishes on a set of positive measure, $p(z)$ itself must vanish for $|z| < 1$. (See Ref. 7, p. 373, Theorem 17.18.) However, $p(0) = 1$, and, hence, we have a contradiction, denying the validity of (37).

REFERENCES

1. B. R. Saltzberg, "Intersymbol Interference Error Bounds with Application to Ideal Bandlimited Signaling," IEEE Trans. Inform. Theory, *IT-14*, No. 4 (July 1968), pp. 563–568.
2. T. T. Kadota and L. A. Shepp, "On the Best Finite Set of Linear Observables for Discriminating Two Gaussian Signals," IEEE Trans. Inform. Theory, *IT-13*, No. 2 (April 1967), Appendix A.
3. G. D. Forney, "Lower Bounds on Error Probability in the Presence of Large Intersymbol Interference," IEEE Trans. Com., *COM-20*, No. 1 (February 1972), pp. 76–77.
4. Ulf Grenander and Gabor Szego, *Toeplitz Forms and Their Applications*, Berkeley: University of California Press, 1958, Chapter 3.
5. E. C. Titchmarsh, *Theory of Functions*, 2nd ed., London: Oxford University Press, 1952, p. 186, Section 5.81.
6. J. Salz, "Optimum Mean-Square Decision Feedback Equalization," B.S.T.J., *52*, No. 8 (October 1973), pp. 1341–1373.
7. W. Rudin, *Real and Complex Analysis*, 2nd ed., New York: McGraw-Hill, 1974.

† This is a simple application of Ref. 7, page 366, Theorem 17.10 supported by the fact that strong convergence in L_2 implies weak convergence.