THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 9, November 1978
Printed in U.S.A.

Signal Design for PAM Data Transmission to Minimize Excess Bandwidth

By A. D. WYNER

(Manuscript received December 9, 1977)

In a conventional PAM data transmission system, the transmitted signal is $x(t) = \Sigma\, \alpha_n g(t - nT)$, where $\{\alpha_n\}$ is a 2^L-level data sequence, and $g(t)$ is a Nyquist pulse ($g(0) \neq 0, g(mT) = 0, m \neq 0$). Ideally, the bandwidth of the pulse $g(t)$ and, therefore, the bandwidth of $x(t)$ can be made equal to $1/2T = \rho/2L$, where ρ is the data rate. In practice, however, an "excess bandwidth" of at least 10 to 20 percent is required.

Using a class of real sequences called "discrete prolate spheroidal sequences," we show how to construct a modulated signal with bandwidth just slightly in excess of the optimal $\rho/2L$ (say, by 2 to 4 percent). The new signal is similar in many ways to a conventional PAM signal, and in particular an ad-hoc receiver structure is suggested for which the resulting error performance is about the same as for a conventional PAM system operating in the same environment.

I. INTRODUCTION

To fix ideas, consider the following conventional (baseband) PAM data-transmission scheme (see, for example, Ref. 1). The data to be transmitted is a sequence $\{\alpha_k\}_{-\infty}^{\infty}$. The α_k are independent identically distributed copies of the random variable α, which is uniformly distributed on the set $\{\pm 1, \pm 3, \cdots, \pm 2^L - 1\}$. Thus, α takes 2^L equally likely values, where $L = 1, 2, \cdots$, is a fixed parameter. The modulated signal is

$$x_0(t) = \sum_{k=-\infty}^{\infty} \alpha_k g_0(t - kT_0), \tag{1}$$

where $g_0(\cdot)$ is a real-valued "Nyquist pulse"—i.e.,

$$g_0(0) \neq 0,$$
$$g_0(kT_0) = 0, \qquad k \neq 0. \tag{2}$$

Since (2) implies that $x_0\,(kT_0) = \alpha_k g\,(0)$, $k = 0, \pm 1, \pm 2, \cdots$, the data sequence $\{\alpha_k\}$ can be obtained from $x_0(\cdot)$ simply by sampling. We assume that the Fourier transform

$$G_0(f) = \int_{-\infty}^{\infty} g_0(t) e^{-i2\pi ft} dt, \qquad -\infty < f < \infty, \tag{3}$$

of $g_0(\cdot)$ has support on the interval $[-F_0, \pm F_0]$, where $F_0 \leq 1/T_0$. Under this assumption, the Nyquist condition (2) is known[1] to be equivalent to

$$G_0(f) + G_0\left(f - \frac{1}{T_0}\right) = Tg_0(0), \qquad 0 \leq f \leq \frac{1}{T_0} \tag{4}$$

(except perhaps in a set of measure zero). Figure 1 is an example of a real $G_0(f)$ which satisfies (4). An often-used Nyquist pulse $G_0(f)$ is the so-called raised-cosine pulse. (See Ref. 1, pp. 50–51.) The bandwidth of $x_0(\cdot)$, which is the same as the bandwidth of $g_0(\cdot)$, is taken as F_0. The difference $F_0 - 1/2T_0$ is called the "excess bandwidth."

To conserve bandwidth, it is desirable to make F_0 as close to $1/2T_0$ as possible, but typically $(F_0 - 1/2T_0)/F_0 \geq 10$ to 20 percent in real systems. Further reduction in the excess bandwidth is difficult, since the very sharp cutoff filter used to generate $g_0(t)$ with F_0 close to $1/2T_0$ will introduce either phase distortion or ripples in the amplitude characteristic.

In a practical data transmission system for the voice-grade telephone channel, a reduction in bandwidth is also desirable, since the channel characteristics at the band edges are poor.

In this paper, we suggest another approach to the signal design problem which will allow a further reduction in the excess bandwidth, perhaps to as little as 2 to 4 percent. The technique involves a family of sequences called "discrete prolate spheroidal sequences" (DPSS) and is also intimately tied up with notions concerning the space of square summable sequences (l_2). Therefore, before presenting our scheme, we must digress to review some notions about the space l_2 and to introduce the DPSSs. We do this in Sections II and III, respectively. In Section IV we discuss our new scheme.

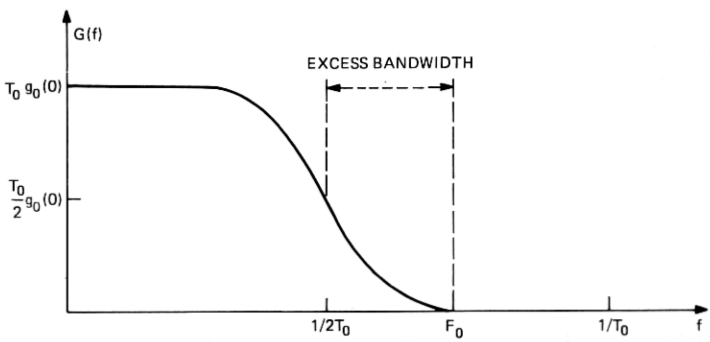

Fig. 1—Example of real $G(f)$ satisfying eq. (4).

At this point, I would like to acknowledge with thanks three of my colleagues without whose help this project could never have gotten off the ground. D. Slepian introduced me to DPSSs, and with much kindness and no small amount of work helped me to get numerical values for these sequences and their associated eigenvalues. J. Mazo taught me most of what I know about data communication, and J. Salz's interest and enthusiasm stimulated me to obtain a full understanding of the properties of the modulation scheme.

II. REVIEW OF THE SPACE l_2

The space l_2 of square-summable, real-valued sequences is the set of sequences $\{a(n)\}_{n=-\infty}^{\infty}$ (or $a(\cdot)$) such that

$$\sum_{n=-\infty}^{\infty} a^2(n) < \infty. \tag{5}$$

Let $a(\cdot)$, $b(\cdot) \in l_2$; then the *inner product* of $a(\cdot)$ and $b(\cdot)$ is

$$\langle a, b \rangle = \sum_{n=-\infty}^{\infty} a(n)b(n). \tag{6}$$

Also, the norm of $a(\cdot)$ is

$$\|a\| = \langle a, a \rangle^{1/2}. \tag{7}$$

We will need the following facts. For $a(\cdot)$, $b(\cdot)$, $c(\cdot) \in l_2$, and any real number γ,

$$\langle \gamma a, b + c \rangle = \gamma \langle a, b \rangle + \gamma \langle a, c \rangle, \tag{8}$$

which implies that, for $a_j \in l_2$, $j = 1, 2, \cdots$,

$$\| \sum_j a_j \|^2 = \sum_j \|a_j\|^2 + 2 \sum_{j<k} \langle a_j, a_k \rangle. \tag{9}$$

Further, the Schwarz inequality is, for a, $b \in l_2$,

$$|\langle a, b \rangle| \le \|a\| \, \|b\|. \tag{10}$$

For $a(\cdot) \in l_2$, the (sequence) *Fourier transform* is defined by

$$A_T(f) = \sum_{n=-\infty}^{\infty} a(n) e^{-i2\pi fTn}, \qquad -\infty < f < \infty, \tag{11}$$

where $T > 0$ is a fixed parameter. Of course, $A_T(f)$ is periodic with period $1/T$, and usually we will be concerned only with its values on the interval $[-(1/2T), (1/2T)]$. The sequence $\{a(n)\}$ can be recovered from $A_T(\cdot)$ by the formula

$$a(n) = T \int_{-1/2T}^{1/2T} A_T(f) e^{i2\pi fTn} df, \qquad -\infty < n < \infty. \tag{12}$$

The *convolution theorem* states that if

$$c(n) = \sum_{m=-\infty}^{\infty} a(m) b(n-m)$$

(which we denote $c = a*b$), then

$$C_T(f) = A_T(f)B_T(f), \qquad -\infty < f < \infty, \tag{13}$$

where A_T, B_T, and C_T are the transforms of a, b, and c, respectively.

The Parseval relation is, for $a, b \in l_2$,

$$\langle a, b \rangle = T \int_{-1/2T}^{+1/2T} A_T(f)B_T^*(f)df, \tag{14}$$

where "$*$" denotes complex conjugate. Thus, in particular,

$$\|a\|^2 = \langle a, a \rangle = T \int_{-1/2T}^{+1/2T} |A_T(f)|^2 df. \tag{15}$$

We say that a sequence $a(\cdot) \in l_2$ is *bandlimited* to $[0, F]$, $0 \le F \le 1/2T$, if its transform $A_T(f) = 0$, for $F \le |f| \le 1/2T$. Thus, a bandlimited $a(\cdot)$ can be written

$$a(n) = \int_{-F}^{F} A(f)e^{i2\pi fTn}df. \tag{16}$$

A sequence $a(\cdot)$ has *support* on the interval $[N_1, N_2]$, $-\infty \le N_1 \le N_2 \le \infty$, if $a(n) = 0$, for $n \notin [N_1, N_2]$. A sequence with support on $[N_1, N_2]$, where $|N_1|$, $|N_2| < \infty$, cannot be bandlimited to $[0, F]$ with $F < 1/2T$.

It is convenient to define the *bandlimiting* operator on l_2, $\mathcal{B} = \mathcal{B}_F$, $0 \le F \le 1/2T$, by (for $a \in l_2$)

$$\mathcal{B}a = b, \tag{17a}$$

where

$$b(n) = \int_{-F}^{F} A_T(f)e^{i2\pi fTn}df. \tag{17b}$$

In other words, the transform of $b(\cdot)$ is

$$B_T(f) = \begin{cases} A_T(f), & |f| \le F, \\ 0, & F \le |f| \le \dfrac{1}{2T}. \end{cases} \tag{17c}$$

A sequence $a \in l_2$ is bandlimited to $[0, F]$ iff $\mathcal{B}_F a = a$. Corresponding to the operator $\mathcal{B}_F$, we also define the complementary operator $\mathcal{B}' = \mathcal{B}_F' = I - \mathcal{B}_F$, where I is the identity operator.

We also define the *index-limiting* (or *time-limiting*) operator $\mathcal{D} = \mathcal{D}_N$ $(1 \le N < \infty)$, by (for $a \in l_2$)

$$\mathcal{D}a = b, \tag{18a}$$

where

$$b(n) = \begin{cases} a(n), & 1 \le n \le N, \\ 0, & \text{otherwise}. \end{cases} \tag{18b}$$

Thus $a \in l_2$ has support on $[1, N]$ iff $\mathcal{D}_N a = \dot{a}$. We will need the following easily established propositions.

Proposition 1: Let $x(t)$, $-\infty < t < \infty$, be a real-valued function with ordinary Fourier transform (as defined by (3)) $X(f)$, $-\infty < f < \infty$. Let the sequence $a(\cdot)$ be defined by $a(n) = x(nT)$. Then, the sequence Fourier transform of $a(\cdot)$ is

$$A_T(f) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X\left(f - \frac{k}{T}\right), \qquad -\infty < f < \infty.$$

In particular, if $x(t)$ is bandlimited to F Hz, and $1/T > 2F$, then

$$A_T(f) = \frac{1}{T} X(f), \qquad |f| \le \frac{1}{2T}.$$

Thus the sequence $a(\cdot)$ is bandlimited to $[0,F]$.

Proposition 2: Let $a(\cdot) \in l_2$, and let $g(t)$ be a real-valued function of the continuous variable t. Let

$$x(t) = \sum_{n=-\infty}^{\infty} a(n) g(t - nT), \qquad -\infty < t < \infty.$$

Then the ordinary Fourier transform of $x(t)$ is

$$X(f) = A_T(f) G(f), \qquad -\infty < f < \infty.$$

where $A_T(f)$ is the sequence Fourier transform of $a(\cdot)$ and $G(f)$ is the ordinary Fourier transform of $g(t)$.

III. DISCRETE PROLATE SPHEROIDAL SEQUENCES

Let $T, F > 0$ (where $W = FT \le 1/2$) and $N > 1$ be given; let the operators $\mathcal{B} = \mathcal{B}_F$, $\mathcal{D} = \mathcal{D}_N$ be as defined in Section II. The following theorem is proved in Appendix A.

Theorem 3: There exists a set of real sequences $\{\phi_j(\cdot)\}_{j=1}^{N}$, called "discrete prolate spheroidal sequences" (DPSS), with support on $[1, N]$ and a corresponding set of real numbers $\{\lambda_j\}_1^N$, called "eigenvalues," with the following properties.

(A) $1 \ge \lambda_1 \ge \lambda_2 \cdots \ge \lambda_N > 0$, *and* $\sum_{j=1}^{N} \lambda_j = 2FTN$.

(B) $\mathcal{D}\mathcal{B}\phi_j = \lambda_j \phi_j$, $1 \le j \le N$.

(C) $\langle \phi_j, \phi_k \rangle = \delta_{jk}$

(D) $\langle \mathcal{B}\phi_j, \mathcal{B}\phi_k \rangle = \lambda_j \delta_{jk}$

(E) *With $\delta > 0$, and F, T held fixed, as $N \to \infty$,*

$$\frac{1}{N} \left\{ \begin{matrix} \text{number of } j \text{ such that} \\ \delta < \lambda_j < 1 - \delta \end{matrix} \right\} \to 0.$$

(F) *With $\epsilon > 0$, and F, T held fixed, as $N \to \infty$,*

$$\lambda_{2FTN(1-\epsilon)} \to 1,$$

$$\lambda_{2FTN(1+\epsilon)} \to 0.$$

(G) *(Slepian [Ref. 4, eq. (63)]), with* $\epsilon > 0$ *and F, T held fixed, as* $N \to \infty$,

$$1 - \lambda_{2FTN(1-\epsilon)} = \exp\{-C(\epsilon)N + o(N)\},$$

where $C(\epsilon) > 0$.

(H) *The* $\phi_j(\cdot)$ *and* λ_j, $1 \le j \le N$, *depend on F, T only through their product* $W = FT$.

Remarks:

(*i*) In the course of giving the proof of Theorem 3, we will show explicitly how to find the DPSSs $\{\phi_j\}$ and the corresponding $\{\lambda_j\}$.

(*ii*) Theorem 3A, *F* implies that, with *N* large, about $2FTN$ of the λ_js are about 1, and that the remainder (about $(1 - 2FT)N$) of the λ_js are about 0. Theorem 3G indicates that the convergence as $N \to \infty$ is quite rapid. Since this fact is crucial to our modulation scheme, we list some of the λ_js for $FT = 1/4$, and various values of *N* in Table I. Here $2FTN = N/2$, so that about half of the λ_js are 1 and the remainder are about 0.

Table I — $\{\lambda_j\}$ for $W = 0.25$, for $N = 5, 10, 20, 50, 100$

	j	λ_j		j	λ_j
$N = 5$	1	0.9976686		15	0.0000212
	2	0.9244132		16	0.0000008
	3	0.5000000		17	0.0000000
	4	0.0755868		18	0.0000000
	5	0.0023143		19	0.0000000
				20	0.0000000
$N = 10$	1	0.9999994			
	2	0.9999490	$N = 50$	1–21	>0.9997
	3	0.9980787		22	0.998
	4	0.9650286		23	0.985
	5	0.7326630		24	0.914
	6	0.2673371		25	0.680
	7	0.0349714		26	0.320
	8	0.0019213		27	0.086
	9	0.0000510		28	0.015
	10	0.0000005		29	0.002
				30–50	<0.00023
$N = 20$	1	0.9999999			
	2	0.9999999	$N = 100$	1–45	>0.9998
	3	0.9999999		46	0.9993
	4	0.9999999		47	0.996
	5	0.9999992		48	0.976
	6	0.9999788		49	0.892
	7	0.9995798		50	0.664
	8	0.9940340		51	0.336
	9	0.9435514		52	0.108
	10	0.7070557		53	0.024
	11	0.2929445		54	0.004
	12	0.0564486		55	0.0007
	13	0.0059659		56–100	<0.0001
	14	0.0004201			

(*iii*) Theorem 3C implies that $\|\phi_j\|^2 = 1$, and Theorem 3D implies that $\|\mathcal{B}\phi_j\|^2 = \lambda_j$. Thus, the fraction of the energy of ϕ_j within the band $[0, F]$ is λ_j. Therefore, when λ_j is close to unity, ϕ_j is a sequence with support on $[1, N]$ with most of its energy in the band $[0, F]$. Theorem 3 implies that, with N large, there are about $2FTN$ orthogonal sequences, i.e., ϕ_j ($j = 1, 2, \cdots, 2FTN(1-\epsilon)$), with support on $[1, N]$ which are approximately bandlimited to $[0, F]$.

(*iv*) Slepian has made an exceptionally detailed study of DPSSs and their properties. Reference 4 contains most of his results, and Ref. 5 describes a Fortran program for computing the DPSSs and their eigenvalues.

IV. HEURISTIC DESCRIPTION OF THE MODULATION SCHEME

Let the data to be transmitted be as in Section I, the 2^L-level sequence $\{\alpha_j\}_{-\infty}^{\infty}$. We break this sequence into blocks of length ν, where the kth block is $\alpha_{k\nu+1}, \cdots, \alpha_{(k+1)\nu}$, $-\infty < k < \infty$, and where ν is an integer to be chosen later. Consider the 0th block $\alpha_1, \cdots, \alpha_\nu$. Let $N > \nu$ be another integer parameter, and let $F, T > 0$, with $FT < 1/2$, be given. Let $\phi_j, \lambda_j, 1 \le j \le N$, be the DPSSs and eigenvalues guaranteed by Theorem 3, with parameters N, F, T. Then define the sequence

$$a(n) = \sum_{j=1}^{\nu} \alpha_j \phi_j(n), \qquad -\infty < n < \infty. \tag{19}$$

Observe that $a(\cdot)$, like the $\phi_j(\cdot)$, has support on the interval $[1, N]$. Further, if we take $\nu = 2FTN(1-\epsilon)$, with N sufficiently large so that $\lambda_\nu \approx 1$, then from Theorem 3 (see remark iii), $a(\cdot)$ is approximately bandlimited to $[0, F]$.

Now the modulated waveform corresponding to the 0th data block is

$$x_0(t) = \sum_{n=1}^{N} a(n) g(t - nT), \qquad -\infty < t < \infty, \tag{20}$$

where the pulse $g(t)$ has Fourier transform $G(f)$ which satisfies

$$G(f) = \begin{cases} T, & |f| \le F, \\ 0, & |f| > \dfrac{1}{2T}, \end{cases} \tag{21a}$$

$$|G(f)| \le T, \qquad F \le |f| \le \frac{1}{2T}. \tag{21b}$$

Thus, we do not specify $G(f)$ in the interval $[F, 1/2T]$, except by (21b). Since $G(f)$ need not have sharp transitions, it is not difficult to implement in practice. For the kth data block ($-\infty < k < \infty$), $\alpha_{k\nu+1}, \cdots, \alpha_{(k+1)\nu}$, we set

$$a(n) = \sum_{j=1}^{\nu} \alpha_{k\nu+j} \phi_j(n - Nk), \qquad Nk + 1 \le n \le N(k+1), \tag{22}$$

and let the modulated waveform be

$$x_k(t) = \sum_{n=Nk+1}^{N(k+1)} a(n)g(t - nT). \tag{23}$$

The entire modulated signal is

$$x(t) = \sum_{k=-\infty}^{\infty} x_k(t) = \sum_{n=-\infty}^{\infty} a(n)g(t - nT). \tag{24}$$

Since the number of bits in each data block is $L\nu$ and each data block "occupies" NT seconds, the transmission rate is

$$\rho = \left(\frac{\nu}{N}\right)\frac{L}{T} \text{ bits/s}. \tag{25}$$

We now give an intuitive, though imprecise, explanation of the properties of the modulation scheme. Consider $x_0(t)$ given by (20). Its Fourier transform is, from Proposition 2,

$$X_0(f) = A_T(f)G(f), \tag{26a}$$

where

$$A_T(f) = \sum_{n=1}^{N} a(n)e^{-i2\pi fTn}$$

$$= \sum_{j=1}^{\nu} \alpha_j \Phi_{jT}(f), \tag{26b}$$

where Φ_{jT} is the sequence Fourier transform of $\phi_j(\cdot)$. In the light of remark iii following Theorem 3, the $\{\Phi_{jT}(f)\}_{j=1}^{\nu}$ and therefore $A_T(f)$ are approximately zero for $|f| \in [F, 1/2T]$ provided $\nu \le 2FN(1 - \epsilon)$. Since $G(f)$ is bounded in this interval and 0 for $|f| > 1/2T$, we see that $X_0(f)$ is approximately bandlimited to $|f| \le F$. Further, if we take $\nu = 2FTN(1 - \epsilon)$, we have from (25) that the transmission rate ρ is $2FL(1 - \epsilon)$. Thus in our scheme we can transmit $2F(1 - \epsilon)$, 2^L-level data symbols per second with bandwidth F. If $\epsilon = 0$, then we would have effectively constructed a PAM system with no excess bandwidth. Since, in practice, ϵ can be made very small, we can in fact come quite close to the ideal.

So far so good. But we still must show that the data symbols $\{\alpha_j\}_1^{\nu}$ can be recovered conveniently from $x_0(t)$. In fact, we claim that the samples $x_0(nT) \approx a(n)$, $1 \le n \le N$. The key observation here is that, since $A_T(f) \approx 0$, $|f| \in [F, 1/2T]$, then $X_0(f)$ is not appreciably changed when $G(f)$ is replaced by $G_I(f)$ ("I" for "ideal") where

$$G_I(f) = \begin{cases} T, & |f| \le 1/2T, \\ 0, & |f| > 1/2T. \end{cases} \tag{27}$$

The inverse transform of G_I is $g_I(t) = (\sin \pi t/T)/(\pi t/T)$. Let us therefore define $x_I(t)$ by replacing $g(t)$ by $g_I(t)$ in the definition of $x_0(t)$. We obtain

$$x_I(t) = \sum_{n=1}^{N} a(n)g_I(t - nT),$$

so that

$$x_I(nT) = a(n), \qquad 1 \le n \le N.$$

It follows that

$$x_0(nT) - a(n) = x_0(nT) - x_I(nT), \qquad 1 \le n \le N.$$

Now define the sequence $c(\cdot)$ by

$$c(n) = x_0(nT) - x_I(nT), \quad 1 \le n \le N.$$

Since $x_0(t) - x_I(t)$ is bandlimited to $1/2T$ Hz, we have from Proposition 1 that

$$C_T(f) = \frac{1}{T}[X_0(f) - X_I(f)], \qquad |f| \le \frac{1}{2T},$$

and from Proposition 2 that

$$\begin{aligned} C_T(f) &= \frac{1}{T}[X_0(f) - X_I(f)] \\ &= \frac{1}{T}A_T(f)[G(f) - G_I(f)]. \end{aligned}$$

From the Parseval relation (15),

$$\begin{aligned} \sum_{n=1}^{N} [x_0(nT) - a(n)]^2 &\le \|c\|^2 \\ &= T\int_{-1/2T}^{1/2T} |C_T(f)|^2 df \\ &= \frac{1}{T}\int_{F<|f|\le 1/2T} |A_T(f)|^2|G(f) - G_I(f)|^2 df \\ &\le 4T\int_{F<f\le 1/2T} |A_T(f)|^2 df = 4\|\mathcal{B}_F' a\|^2 \approx 0. \end{aligned}$$

The inequality follows from (21b), which implies that

$$|G(f) - G_I(f)| \le 2T.$$

Thus, $a(n)$, $1 \le n \le N$ can be recovered from $x_0(t)$ simply by sampling. From (19) and the orthonormality of the $\{\phi_j\}_1^{\nu}$,

$$\alpha_j = \sum_{n=1}^{N} a(n)\phi_j(n), \tag{28}$$

so that the $\{\alpha_j\}_1^\nu$ can be recovered from samples $x(nT)$, $1 \le n \le N$.

Aside from imprecision, the above arguments completely ignored the effects of the other data blocks ($k \neq 0$) and the effects of channel distortion and noise. In the next section, we give a precise definition of the modulation scheme and of a proposed receiver, and then state theorems that give bounds on the error introduced by linear channel distortion and a channel noise. We will also bound the instantaneous power $x^2(t)$.

V. PRECISE STATEMENT OF RESULTS AND DISCUSSION

Let $\{\alpha_j\}_{-\infty}^{\infty}$ be, as in Section IV, a sequence of independent, identically distributed copies of the 2^L-valued random variable α, where

$$\Pr\{\alpha = m\} = 2^{-L}, \qquad m = \pm 1, \pm 3, \cdots, \pm(2^L - 1). \tag{29}$$

The sequence $\{\alpha_j\}$ is the data sequence to be transmitted. Let $\nu, N, F, T > 0$ be parameters such that ν, N are integers, and

$$\nu < N, \tag{30a}$$

$$W \triangleq FT < 1/2. \tag{30b}$$

Partition the data sequence into blocks of length ν, such that the kth block is

$$\alpha_{\nu k+1}, \cdots, \alpha_{\nu(k+1)},$$

$k = 0, \pm 1, \pm 2, \cdots$. Let $\{\phi_j(\cdot), \lambda_j\}_{j=1}^N$ be the quantities (DPSSs and eigenvalues) whose existence is guaranteed by Theorem 3 with parameters N, W. Corresponding to the kth data block, define the sequence $a_k(\cdot)$ by

$$a_k(n) = \sum_{j=1}^{\nu} \alpha_{\nu k+j}\phi_j(n - Nk), \tag{31a}$$

and let

$$a(\cdot) = \sum_{k=-\infty}^{\infty} a_k(\cdot). \tag{31b}$$

Since $\phi_j(\cdot)$ has support on $[1, N]$, $a_k(\cdot)$ has support on $[Nk + 1, N(k + 1)]$. Finally, the modulated signal is

$$x(t) = \sum_{k=-\infty}^{\infty} x_k(t), \qquad -\infty < t < \infty, \tag{32a}$$

where

$$\begin{aligned} x_k(t) &= \sum_{n=-\infty}^{\infty} a_k(n)g(t - nT), \\ &= \sum_{n=Nk+1}^{N(k+1)} a_k(n)g(t - nT), \end{aligned} \tag{32b}$$

and the pulse $g(t)$ is the inverse Fourier transform of $G(f)$, which we leave unspecified for now.

A block diagram for the modulator described above is given in Fig. 2. The data symbols appear at a rate of ν/NT per second. Box A takes the data symbols ν at a time and calculates the numbers $\{a(n)\}$—producing N outputs for every ν inputs. Thus, the $a(n)$ appear at a rate of $1/T$ per second. Box B produces $x(t)$ by modulating the amplitude of a pulse train with the $\{a(n)\}$. Although we will not specify the ratio ν/N and the pulse $g(t)$ now, it will be useful to informally think of

$$\nu/N = 2FTN(1 - \epsilon) = 2WN(1 - \epsilon),$$

and $g(t) \leftrightarrow G(f)$ as in (21). We will allow the possibility of non-physically realizable pulses $g(t)$, with the usual understanding that a close approximation to $g(t)$ can be obtained with a finite delay (which we shall ignore).

The received signal is taken as

$$y(t) = w(t) + z(t), \tag{33a}$$

where

$$w(t) = \int_{-\infty}^{\infty} x(\tau)\, h_c(t - \tau) d\tau, \tag{33b}$$

and where $h_c(t)$ is the impulse response of the channel ($H_c(f)$, the channel transfer function, is the transform of $h_c(t)$), and $z(t)$ is noise with zero mean and power spectral density $N_Z(f)$.

We now turn to the receiver. We will postulate a simple receiver structure which, though not optimum, has the virtues of simplicity and amenability to analysis. Furthermore it is probably not very far from being optimal itself. Refer to Fig. 3. The received waveform $y(t)$ is first sampled at $t = nT$, to produce the sequence $\{y(nT)\}_{n=-\infty}^{\infty}$. These samples are the input to box C, a tapped delay line with $2M + 1$ taps. The output of box C is the sequence $\{\hat{a}(n)\}$ given by

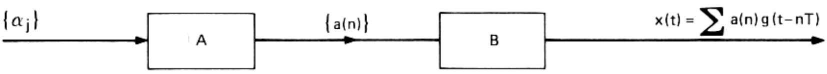

Fig. 2—The modulator.

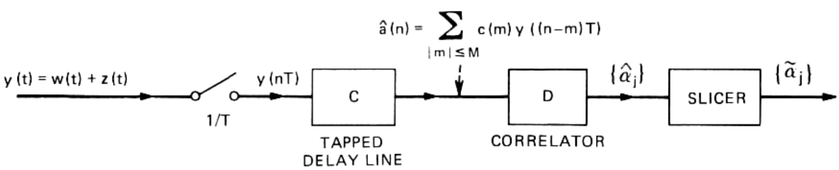

Fig. 3—The receiver.

$$\hat{a}(n) = \sum_{m=-M}^{+M} c(m)y((n-m)T), \qquad -\infty < n < \infty, \tag{34}$$

where $\{c(m)\}_{-M}^{+M}$ are the tap weight coefficients. As we shall see, the tap weights should be chosen so that the sequence $\hat{a}(\cdot)$ is the receiver's best guess of the sequence $a(\cdot)$.

Consider the 0th block of $\{\hat{a}(n)\}$, i.e., $\hat{a}(1), \cdots, \hat{a}(N)$. If $\hat{a}(n) \equiv a(n)$, then the 0th block of data symbols $\alpha_1, \cdots, \alpha_\nu$ could be recovered from $\{\hat{a}(n)\}$ using (28). Our approach will be to use these same formulas to obtain an estimate $\{\hat{\alpha}_j\}$ of the $\{\alpha_j\}$, i.e.,

$$\hat{\alpha}_j = \sum_{n=1}^{N} \hat{a}(n)\phi_j(n), \qquad 1 \le j \le \nu. \tag{35a}$$

For the remaining blocks ($k \neq 0$), we proceed analogously, viz.,

$$\hat{\alpha}_{k\nu+j} = \sum_{n=Nk+1}^{N(k+1)} \hat{a}(n)\phi_j(n-Nk), \qquad 1 \le j \le \nu. \tag{35b}$$

This is box D. The final step in the demodulation process is a "slicer," which examines $\hat{\alpha}_j(-\infty < j < \infty)$ and emits $\tilde{\alpha}_j$ where $\tilde{\alpha}_j$ equals a value of $m \in \{\pm 1, \pm 3, \cdots, \pm 2^L - 1\}$ which minimizes $|\tilde{\alpha}_j - m|$.

As in (25), the transmission rate is $\rho = (\nu/N) \cdot (L/T)$ bits/s.

We are now ready to state the properties of the modulation scheme in the form of theorems. The proofs of these theorems are given in Section VI. Theorem 4 gives an upper bound on the average power of the transmitted signal $x(t)$. Theorem 5 gives an upper bound on the expected instantaneous power $E\,x^2(t)$, as a function of t. Finally, Theorem 6 gives an upper bound on the mean-squared error. We state these results with no restrictions on $G(f)$, $H_c(f)$, and $N_Z(f)$. In the remarks which follow the statement of the theorems, we will look at some interesting special cases.

We begin with a bit of notation. Denote the variance of the data random variable α, defined in (29), by

$$\sigma_\alpha^2 = E\,\alpha^2 = 2^{-L} \sum_{\substack{|m| \le 2^L - 1 \\ m \text{ odd}}} m^2 = \frac{(2^L - 1)(2^L + 1)}{3}. \tag{36}$$

Also, for $\nu, N, W = FT$ satisfying (30), let

$$Q = Q(\nu, N, W) = \frac{1}{\nu} \sum_{j=1}^{\nu} (1 - \lambda_j). \tag{37}$$

Of course, if we set $\nu = 2WN(1-\epsilon)$, and let $N \to \infty$, with $W, \epsilon > 0$ held fixed, then $Q \to 0$. It will be helpful to think of Q as a small quantity. Here are the theorems. Although they may seem formidable at first glance, please stick with it! In the extensive discussion following the theorem statements, you will see that they can be easily applied and yield useful information.

Theorem 4: (*average power*)

$$P_{AV} \triangleq \frac{1}{NT} E \int_0^{NT} x^2(t)dt$$
$$\leq \frac{\sigma_\alpha^2}{T}\left[\int_{-F}^{F} \sum_{k=-\infty}^{\infty} \left|G\left(f-\frac{k}{T}\right)\right|^2 df + \frac{\nu}{N} A_1 Q\right], \tag{38a}$$

where

$$A_1 = \sup_{F \leq |f| \leq 1/2T} \frac{1}{T^2} \sum_{k=-\infty}^{\infty} \left|G\left(f-\frac{k}{T}\right)\right|^2. \tag{38b}$$

Theorem 5: For $-\infty < t_1 < \infty$,

$$E\, x^2(t_1) \leq \frac{\sigma_\alpha^2}{T} \int_{-1/2T}^{1/2T} \left|\sum_{k=-\infty}^{\infty} G\left(f-\frac{k}{T}\right) e^{i2\pi f t_1}\right|^2 df. \tag{39}$$

Theorem 6: (mean-squared error)

$$\epsilon^2 \triangleq \frac{1}{\nu} E \sum_{j=1}^{\nu} (\hat{\alpha}_j - \alpha_j)^2 = \epsilon_N^2 + \epsilon_I^2, \tag{40}$$

(N stands for noise, and I for intersymbol interference). The noise error ϵ_N^2 *is bounded by*

$$\epsilon_N^2 \leq \frac{N}{\nu} \int_{-F}^{F} |C_T(f)|^2 \left(\sum_{k=-\infty}^{\infty} N_Z\left(f-\frac{k}{T}\right)\right) df + A_2 Q, \tag{41a}$$

where

$$C_T(f) \triangleq \sum_{n=-M}^{M} c(n) e^{-i2\pi f T n}, \tag{41b}$$

$N_Z(f)$ *is the power spectral density of the noise, and*

$$A_2 \triangleq \sup_{F \leq |f| \leq 1/2T} \frac{|C_T(f)|^2}{T} \sum_{k=-\infty}^{\infty} N\left(f-\frac{k}{T}\right). \tag{41c}$$

The intersymbol interference error ϵ_I^2 *is bounded by*

$$\epsilon_I^2 \leq \sigma_\alpha^2 \left(\frac{N}{\nu}\right)\left[T \int_{-F}^{F} |C_T(f) B_T(f) - 1|^2 df + A_3 Q\right], \tag{42a}$$

where

$$B_T(f) \triangleq \frac{1}{T} \sum_{k=-\infty}^{\infty} G\left(f-\frac{k}{T}\right) H_c\left(f-\frac{k}{T}\right), \tag{42b}$$

$C_T(f)$ *is given by (41b), and*

$$A_3 \triangleq \sup_{F \leq |f| \leq 1/2T} |C_T(f) B_T(f) - 1|^2. \tag{42c}$$

Remarks:

(*i*) The reason that P_{AV} as defined by (38a) is the "average power" is that the random function $x(t)$ is cyclostationary with period NT. In other words, the shifted sequence $\tilde{x}(t) \triangleq x(t - kNT)$ has the same statistical properties as $x(t)$ itself (for $k = 0, \pm 1, \pm 2, \cdots$). Thus, it follows that

$$\lim_{\tau \to \infty} E \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} x^2(t)dt = \frac{1}{NT} E \int_0^{NT} x^2(t)dt = P_{AV}. \tag{43}$$

(*ii*) When Q is small, the upper bound of (38a) on P_{AV} depends essentially on the folded power spectrum

$$\sum_{k=-\infty}^{\infty} G\left(f - \frac{k}{T}\right), \qquad |f| \le F.$$

Furthermore, if $G(f) = 0$ for $|f| \le 1/2T$, then

$$\sum_k G\left(f - \frac{k}{T}\right) = G(f), \qquad |f| \le F,$$

so that Theorem 4 becomes

$$P_{AV} \le \frac{\sigma_\alpha}{T}\left[\int_{-F}^{F} |G(f)|^2 df + \frac{\nu}{N} A_1 Q\right], \tag{44a}$$

where

$$A_1 = \sup_{F \le |f| \le 1/2T} |G(f)|^2. \tag{44b}$$

(*iii*) If we assume, as in Remark *ii*, that $G(f) = 0$, $|f| > 1/2T$, then Theorem 5 becomes

$$E\,x^2(t) \le \frac{\sigma_\alpha^2}{T}\int_{-1/2T}^{1/2T} |G(f)|^2 df. \tag{45}$$

Thus the upper bound on $E\,x^2(t)$ depends on $G(f)$ for $|f| \le 1/2T$.

(*iv*) Saltzberg's[3] bound can be applied (see Appendix B) to our problem to show that the distribution for the instantaneous power satisfies

$$\Pr\{|x(t)|^2 > r^2\} \le 2 \exp\left\{-\frac{r^2}{2E\,x^2(t)}\right\}.$$

The bound of Theorem 5 can be applied here to further overbound this probability.

(*v*) We now explain the rationale for using the mean-squared error $\epsilon^2 = (1/\nu)\sum_{j=1}^{\nu} E(\alpha_j - \hat{\alpha}_j)^2$. First note that with $\epsilon_j^2 \triangleq E(\hat{\alpha}_j - \alpha_j)^2$, Saltzberg's bound (see Appendix B) can again be used to show that, if the noise is Gaussian, then

$$P_{ej} = \Pr\{\tilde{\alpha}_j \ne \alpha_j\} \le 2 \exp\left\{-\frac{1}{2\epsilon_j^2}\right\}.$$

Since the sequence of random pairs $\{\alpha_j, \tilde{\alpha}_j\}_{j=-\infty}^{\infty}$ is cyclostationary with period ν, the overall error probability is

$$P_e = \frac{1}{\nu}\sum_{j=1}^{\nu} P_{ej} \le \frac{2}{\nu}\sum_{j=1}^{\nu} \exp\left\{-\frac{1}{2\epsilon_j^2}\right\}. \tag{46}$$

Now let $\epsilon_{\max}^2 = \max_{1 \le j \le \nu} \epsilon_j^2$. Ineq. (46) yields

$$P_e \leq 2 \exp\left\{-\frac{1}{2\epsilon_{\max}^2}\right\}.$$

Thus it would appear that $\epsilon_{\max}^2$ rather than ϵ^2 is the appropriate criterion. Nevertheless the use of ϵ^2 (which was, of course, chosen for its mathematical tractability) can be justified by the following argument.

Let R be the $\nu \times \nu$ covariance matrix with (j, k)th entry $E(\hat{\alpha}_j - \alpha_j)\cdot(\hat{\alpha}_k - \alpha_k)$, $1 \leq j, k \leq \nu$. Let M_0 be a $\nu \times \nu$ orthogonal matrix such that the diagonal elements of $M_0^t R M_0$ are all identical.* One choice of M_0 is $\tilde{M}_0 \triangleq M_1 M_2$, where M_1 is a $\nu \times \nu$ orthogonal matrix which diagonalizes R (i.e., $M_1^t R M_1$ is a diagonal matrix), and M_2 is a normalized Hadamard matrix (i.e., a $\nu \times \nu$ orthogonal matrix with entries $\pm 1/\sqrt{\nu}$). A normalized Hadamard matrix is known to exist for all ν which are multiples of 4 up to 200.

Now modify the communication system as follows. Preceding box A in Fig. 2, insert a device which multiplies the data symbols $\{\alpha_j\}$, taken in blocks of ν, by the matrix M_0. Then, following box D in Fig. 3, insert a device which multiplies the input $\{\hat{\alpha}_j\}$, taken in corresponding blocks of length ν, by $M_0^{-1} = M_0^t$. Let the output of this device be $\{\alpha_j'\}$. It is easy to show that (i) the analysis which yields Theorems 4 to 6 is unchanged in the modified system and (ii) $E(\alpha_j' - \alpha_j)^2 \equiv \epsilon^2$, $1 \leq j \leq \nu$.

We must emphasize that the choice of the matrix M_0 depends on R which in turn depends on the channel which is usually unknown or variable. Although it is undoubtedly possible to find an adaptive procedure for finding a good matrix M_0, our conjecture is that, in most real situations, M_0 can be chosen to be any normalized Hadamard matrix with fairly good results.

(vi) To obtain more insight into our scheme, let us assume that $Q \approx 0$, and $G(f) = H_c(f) = N(f) = 0$, $|f| > 1/2T$. Then (38), (41), (42) become

$$P_{AV} \leq \frac{\sigma_\alpha^2}{T} \int_{-F}^{F} |G(f)|^2 df, \tag{47a}$$

$$\epsilon_N^2 \leq \frac{N}{\nu} \int_{-F}^{F} |C_T(f)|^2 N(f) df, \tag{47b}$$

$$\epsilon_I^2 \leq \sigma_\alpha^2 T \left(\frac{N}{\nu}\right) \int_{-F}^{F} \left|\frac{1}{T} C_T(f) G(f) H_c(f) - 1\right|^2 df. \tag{47c}$$

We see immediately that our bounds on the important quantities P_{AV}, ϵ_N^2, ϵ_I^2 depend on $G(f)$, $H_c(f)$, $N(f)$ only for $|f| \leq F$ and not for $F \leq |f| \leq 1/2T$. In particular, we need a channel whose bandwidth is F. Since the transmission rate

$$\rho = \frac{\nu}{N} \cdot \frac{L}{T} = \frac{\nu}{N} \cdot \frac{L}{2FT} \cdot 2F = \frac{\nu}{2WN} \cdot L \cdot 2F$$

---

* Witsenhausen (Ref. 6) has shown that there always exists an M_0 with the desired property for all $\nu = 1, 2, \cdots$.

we have that

$$F = \left(\frac{\rho}{2L}\right)\left(\frac{2WN}{\nu}\right). \tag{48}$$

Now the ideal bandwidth in a conventional 2^L-level PAM system with rate ρ is $(\rho/2L)$. We pointed out in Section I that the required channel bandwidth in real systems is usually no less than 10 to 20 percent in excess of this. For our system with $\nu = 2WN(1-\epsilon)$, the ratio of the required bandwidth given by (48) to $\rho/2L$ is $(1-\epsilon)^{-1}$, which can be made very small. See the numerical example in remark *vii*.

(*vii*) Roughly speaking, Theorem 6 tells us that there are three sources of error. The first is given by the integral in (41a) which is a bound on the error introduced by the noise in the band $[-F, F]$. The second is given by the integral in (42a), which is a bound on the error introduced by the imperfections of the channel, as compensated by $C_T(f)$, in the band $[-F, F]$. The third source of error is the fact that $Q > 0$. The first two of these sources appear in conventional PAM systems, but the last is unique to our system. The following numerical example shows that Q can in fact be made small.

Let $W = 2FT = 0.415$, $N = 80$, $\nu = 64$. Then $Q(\nu, N, W) = (1/\nu)\,\Sigma_{j=1}^{\nu}(1-\lambda_j) = 1.01 \times 10^{-4}$, which corresponds to -40.0 dB. The ratio $2WN/\nu = 1.0375$, so that the required bandwidth is 3.7 percent in excess of the ideal $\rho/2L$.

Continuing with this example, let us say that $1/T = 6 \times 10^3$/s, $F = W/T = 2490$ Hz, and $L = 2$. Then the transmission rate $\rho = (\nu/N)\cdot(L/T) = 9.6$ kb/s. Note that $1/2T = 3$ kHz, so that the system performance is essentially independent of the channel characteristics or noise in the band [2.49 kHz, 3 kHz]. Of course, we are assuming that $H_c(f) = N(f) = 0$, $|f| > 3$ kHz.

Finally, observe that the receiver-correlator (box D in Fig. 3) must perform $N \cdot \nu$ multiplications every $N \cdot T$ second, or ν/T multiplications per second. For $\nu = 64$, and $1/T = 6 \times 10^3$, this works out to one multiplication every 2.6 μs. To store the $N \times \nu$ $\phi_j(n)$, $1 \le n \le N$, $1 \le j \le \nu$, to say 10-bit accuracy, we need a ROM with capacity $10 \cdot N \cdot \nu = 51.2$ kb.

(*viii*) Continuing with the assumptions made in remark *vi*, let us further assume that $G(f) \equiv T$, $|f| \le F$, and $H_c(f) \equiv 1$, $|f| \le F$, i.e., a perfect channel response (in band). Also, let $N(f) = N_0/2$, $|f| \le 1/2T$. Then, from (47),

$$P_{AV} \le 2FT\,\sigma_\alpha^2.$$

Further, the upper bound on the total mean-squared error is minimized for $C_T(f) \equiv 1$ (i.e., $c(0) = 1$, $c(m) = 0$, $m \ne 0$). Then

$$\epsilon^2 \le \frac{N}{\nu}\int_{-F}^{F}\frac{N_0}{2}\,df = \frac{N}{\nu}(N_0F).$$

If $\nu/N \approx 2FT$, the total mean-squared error

$$\epsilon^2 \leq \frac{(N_0F)}{P_{AV}} \sigma_\alpha^2. \tag{49}$$

Note that (N_0F) is the noise power in the band $[-F, F]$. Observe that in a conventional PAM system with a perfect Nyquist equivalent channel and additive white noise, the mean-squared error is given the right member of (49) with F = the Nyquist bandwidth (see Ref. 1, Chap. 5).

(*ix*) Suppose that it is desired to transmit our data using a modulated signal $x(t)$ which is bandpass in the band $[F_1, F_2]$. Then, using quadrature amplitude modulation (QAM) in a straightforward manner, we can modify the present scheme to achieve a bandpass signal. We will now outline the procedure.

Let $0 < F_1 < F_2$ be given. Set $F = (F_2 - F_1)/2$, and choose $T < 1/2\,F$. With F, T so chosen, form two modulated signals (with independent data) according to our (baseband) prescription. Denote these baseband signals by $x^{(1)}(t)$, $x^{(2)}(t)$. Their rates are each $\nu L/NT$. Then form a bandpass signal

$$x(t) = x^{(1)}(t) \cos 2\pi F_c t + x^{(2)}(t) \sin 2\pi F_c t,$$

where $F_c = (F_2 + F_1)/2$. The signal $x(t)$ is essentially bandpass with lower frequency

$$F_c - F = \left(\frac{F_2 + F_1}{2}\right) - \left(\frac{F_2 - F_1}{2}\right) = F_1,$$

and upper frequency $F_c + F = F_2$. The transmission rate for $x(t)$ is

$$\rho = \frac{2\nu L}{NT} = 2 \frac{\nu}{N} \frac{L}{2FT} (2F) = \left(\frac{\nu}{N}\right) \frac{L}{W} (F_2 - F_1).$$

Thus the required channel bandwidth to pass $x(t)$ is

$$(F_2 - F_1) = \frac{\rho}{2L} \left(\frac{2WN}{\nu}\right),$$

exactly as in the baseband case (48). It is a fairly simple matter to analyze the QAM system and obtain results analogous to Theorems 4 to 6.

Another way of accomplishing the synthesis of a bandpass signal is to use "bandpass" DPSSs instead of the conventional DPSS characterized in Section III.

(*x*) Combining (31) and (32), we can rewrite the modulated signal as

$$x(t) = \sum_{j=-\infty}^{\infty} \alpha_j \tilde{g}(t, j),$$

where for $k\nu < j \leq (k + 1)\nu$,

$$\tilde{g}(t, j) = \sum_{n=Nk+1}^{N(k+1)} \phi_{j-k\nu}(n - Nk) g(t - nT).$$

Note that, for $s = 0, \pm 1, \pm 2, \cdots$,

$$\tilde{g}(t, j + s\nu) = \tilde{g}(t + sNT, j),$$

so that there are only ν possible shapes for $\tilde{g}(t, j)$.

We conclude from this that the present system is a kind of PAM, with the data $\{\alpha_j\}$ modulating the amplitude pulses $\{\tilde{g}(t, j)\}$.

(xi) Computation of the Tap Weights: Let us again make the assumptions of remark *vi*. Then, to minimize ϵ^2, it is a reasonable strategy to choose the coefficients $\{c(m)\}_{-M}^{M}$ so that $C_T(f)G(f)H_c(f)$ is as close as possible to unity for $|f| \leq F$. Of course, we must take care not to enhance the noise by making $C_T(f)$ too large. In fact, it is a simple matter to solve for the optimal set $\{c(m)\}_{-M}^{M}$ which minimizes our bound on the total mean-squared error (with $Q \approx 0$)

$$= \left(\frac{N}{\nu} T\right) \left[\int_{-F}^{F} |C_T(f)|^2 \frac{N(f)}{T} df + \sigma_\alpha^2 \int_{-F}^{F} \left| C_T(f) \frac{G(f)\, H_c(f)}{T} - 1 \right|^2 df \right]. \quad (50)$$

Let the sequences $\xi_0(\cdot)$, $\xi_1(\cdot)$, $\xi_2(\cdot)$ be the inverse transforms of

$$\frac{N^{1/2}(f)}{T^{1/2}} \Gamma(f), \qquad \frac{G(f)H_c(f)\Gamma(f)}{T} \Gamma(f), \qquad \Gamma(f),$$

respectively, where

$$\Gamma(f) = \begin{cases} 1, & |f| \leq F, \\ 0, & F < |f| \leq \dfrac{1}{2T}. \end{cases}$$

Then the bound of (50) is, from the Parseval relation (15) ("$*$" indicates convolution),

$$= \frac{N}{\nu} [\|c * \xi_0\|^2 + \sigma_\alpha^2 \|c * \xi_1 - \xi_2\|^2]$$

$$= \frac{N}{\nu} \sum_{n=-\infty}^{\infty} \left\{ \left[\sum_{m'=-M}^{+M} c(m')\xi_0(n - m') \right]^2 + \sigma_\alpha^2 \left[\sum_{m'=-M}^{M} c(m')\xi_1(n - m') - \xi_2(n) \right]^2 \right\}.$$

Differentiating with respect to $c(m)$, $-M \leq m \leq M$, and setting the result equal to zero, yields

$$\sum_{m'=-M}^{M} c(m') \left\{ \sum_{n=-\infty}^{\infty} \xi_0(n - m)\xi_0(n - m') + \sigma_\alpha^2 \sum_{n=-\infty}^{\infty} \xi_1(n - m)\xi_1(n - m') \right\} = \sum_{n=-\infty}^{\infty} \xi_1(n - m)\xi_2(n),$$

or

$$\boxed{\sum_{m=-M}^{M} c(m)\mu_0(m-m') = \mu_1(m)}, \qquad m = 0, \pm 1, \cdots, \pm M, \tag{51a}$$

where

$$\mu_0(m) = \sum_{n=-\infty}^{\infty} [\xi_0(n)\xi_0(n-m) + \sigma_\alpha^2 \xi_1(n)\xi_1(n-m)] \tag{51b}$$

$$\mu_1(m) = \sum_{n=-\infty}^{\infty} \xi_1(n-m)\xi_2(n). \tag{51c}$$

Clearly, $\mu_0(\cdot)$ is the inverse transform of

$$\left[\frac{N(f)}{T} + \sigma_\alpha^2 \frac{|G(f)|^2 |H_c(f)|^2}{T^2}\right] \Gamma(f),$$

and $\mu_1(\cdot)$ is the inverse transform of $(\sigma_\alpha^2/T)\ G^*(f)H_c^*(f)\Gamma(f)$. The tap weights $\{c(m)\}_{-M}^{M}$ are found by solving the linear equations (51a).

Of course, the above computation of the tap weight coefficients is possible only when the channel transfer function $H_c(f)$ and the noise spectrum $N(f)$ are known. In most real applications, these quantities are unknown or changing, so that an adaptive learning technique is required.

VI. PROOF OF THEOREMS

Proof of Theorem 4: Using (32), we have

$$\begin{aligned}
P_{AV} &\triangleq \frac{1}{NT} E \int_0^{NT} x^2(t)dt \\
&= \frac{1}{NT} E \int_0^{NT} \left(\sum_{k=-\infty}^{\infty} x_k(t)\right)^2 dt \\
&\overset{(1)}{=} \frac{1}{NT} \sum_{k=-\infty}^{\infty} E \int_0^{NT} x_k^2(t)dt \\
&\overset{(2)}{=} \frac{1}{NT} \sum_{k=-\infty}^{\infty} E \int_{-kNT}^{-kNT+NT} x_0^2(t)dt \\
&= \frac{1}{NT} E \int_{-\infty}^{\infty} x_0^2(t)dt = \frac{1}{NT} E \int_{-\infty}^{\infty} |X_0(f)|^2 df.
\end{aligned} \tag{52}$$

Step (1) follows from the independence of the $\{\alpha_j\}_{-\infty}^{\infty}$, which implies [see (31) and (32b)] that $E\, x_k(t)x_{k'}(t) = 0,\ k \neq k'$. Step (2) follows from (32b), and the fact that $\{a_k(n)\}_{n=Nk+1}^{N(k+1)}$ has the same statistics as $\{a_0(n)\}_{n=1}^{N}$. Thus, $x_k(t)$ has the same statistics as $x_0(t - kNT)$.

We next apply Proposition 2, which implies that

$$X_0(f) = A_T(f)G(f), \tag{53a}$$

where

$$A_T(f) = \sum_{n=1}^{N} a(n)e^{-i2\pi fTn}. \tag{53b}$$

Substituting (53a) into (52), we obtain

$$P_{AV} = \frac{1}{NT} E \int_{-\infty}^{\infty} |A_T(f)|^2 |G(f)|^2 df$$

$$= \frac{1}{NT} \sum_{k=-\infty}^{\infty} E \int_{k/T}^{(k+1)/T} |G(f)|^2 |A_T(f)|^2 df$$

$$= \frac{1}{NT} \sum_{k=-\infty}^{\infty} E \int_{0}^{1/T} \left|G\left(f - \frac{k}{T}\right)\right|^2 |A_T(f)|^2 df,$$

where we have used the fact that $A_T(f)$ is periodic with period $1/T$. Thus

$$P_{AV} = \frac{1}{NT} \int_{0}^{1/T} \left(\sum_{k=-\infty}^{\infty} \left|G\left(f - \frac{k}{T}\right)\right|^2\right) (E|A_T(f)|^2) df$$

$$= \frac{1}{NT} \int_{-1/2T}^{+1/2T} \left(\sum_{k=-\infty}^{\infty} \left|G\left(f - \frac{k}{T}\right)\right|^2\right) (E|A_T(f)|^2 df,$$

where the last step follows from the fact that the integrand is periodic with period $1/T$, so that we can change the interval of integration from $[0, 1/T]$ to $[-1/2T, 1/2T]$. Continuing, we have

$$P_{AV} = \frac{1}{NT} \int_{-F}^{F} \left(\sum_{k=-\infty}^{\infty} \left|G\left(f - \frac{k}{T}\right)\right|^2\right) (E|A_T(f)|^2) df$$

$$+ \frac{1}{NT} E \int_{F \le |f| \le 1/2T} \left(\sum_{k} \left|G\left(f - \frac{k}{T}\right)\right|^2\right) |A_T(f)|^2 df$$

$$= I_1 + I_2. \tag{54}$$

Now the second integral I_2 can be overbounded by

$$I_2 \le \frac{1}{NT} (A_1 T^2) E \int_{F \le |f| \le 1/2T} |A_T(f)|^2 df = \frac{1}{N} A_1 T E \frac{1}{T} \|\mathcal{B}' a_0\|^2,$$

where A_1 is defined by (38b). From (31), with $k = 0$, and $\langle \mathcal{B}'\phi_j, \mathcal{B}'\phi_{j'} \rangle = (1 - \lambda_j)\delta_{jj'}$, we have

$$I_2 \le \frac{1}{N} A_1 E \left\| \sum_{j=1}^{\nu} \alpha_j \mathcal{B}' \phi_j \right\|^2$$

$$= \frac{1}{N} A_1 E \sum_{j=1}^{\nu} \alpha_j^2 (1 - \lambda_j)$$

$$= \frac{\nu}{N} A_1 \sigma_\alpha^2 Q, \tag{55}$$

where Q is defined by (37).

To overbound I_1, the first term in (54), we again use (31) to obtain

$$E|A_T(f)|^2 = E \left| \sum_{j=1}^{\nu} \alpha_j \Phi_{jT}(f) \right|^2$$

where $\Phi_{jT}(f)$ is the transform of $\phi_j(\cdot)$. Since $E\,\alpha_j\alpha_{j'} = \sigma_\alpha^2\delta_{jj'}$, we have

$$E|A_T(f)|^2 = \sum_{j=1}^{\nu} \sigma_\alpha^2 |\Phi_{jT}(f)|^2$$
$$\leq \sigma_\alpha^2 N,$$

by Theorem 7 (proved in Appendix C). Thus

$$I_1 \leq \frac{\sigma_\alpha^2}{T} \int_{-F}^{F} \left(\sum_{k=-\infty}^{\infty} \left| G\left(f - \frac{k}{T}\right) \right|^2 \right) df. \tag{56}$$

Substituting (55) and (56) into (54) yields Theorem 4.

Proof of Theorem 5: Let t_1, $-\infty < t_1 < \infty$, be given. Then from (32) and (31),

$$x(t_1) = \sum_{k=-\infty}^{\infty} x_k(t_1) = \sum_{k=-\infty}^{\infty} \sum_{n=Nk+1}^{N(k+1)} \sum_{j=1}^{\nu} \alpha_{k\nu+j} \phi_j(n - Nk) g(t_1 - nT)$$
$$= \sum_k \sum_j \alpha_{k\nu+j} \left[\sum_{n=Nk+1}^{N(k+1)} \phi_j(n - Nk) g(t_1 - nT) \right].$$

Using $E\,\alpha_j\alpha_{j'} = \sigma_\alpha^2\delta_{jj'}$, we obtain

$$E\,x^2(t_1) = \sum_{k=-\infty}^{\infty} \sum_{j=1}^{\nu} \sigma_\alpha^2 \left[\sum_{n=Nk+1}^{N(k+1)} \phi_j(n - Nk) g(t_1 - nT) \right]^2. \tag{57}$$

Now with t_1 held fixed, define the sequence $c(\cdot)$ by

$$c(n) = g(t_1 - nT), \qquad -\infty < n < \infty.$$

Also for $-\infty < k < \infty$, $1 \leq j \leq \nu$, define the sequences $\phi_{kj}(\cdot)$ by

$$\phi_{kj}(n) = \phi_j(n - kN), \qquad -\infty < n < \infty.$$

Thus $\phi_{kj}(\cdot)$ has support in the interval $[Nk + 1, N(k + 1)]$. Of course, for $-\infty < k, k' < \infty$, $1 \leq j, j' \leq \nu$,

$$\langle \phi_{kj}, \phi_{k'j'} \rangle = \begin{cases} 1, & k = k', j = j', \\ 0, & \text{otherwise}, \end{cases}$$

so that $\{\phi_{kj}\}_{k,j}$ is a family of orthonormal sequences. Furthermore, the term in brackets in (57) is $\langle \phi_{kj}, c \rangle$, so that (58) can be written

$$E\,x^2(t_1) = \sigma_\alpha^2 \sum_{k=-\infty}^{\infty} \sum_{j=1}^{\nu} \langle \phi_{kj}, c \rangle^2. \tag{58}$$

Letting $\mathcal{S}$ be the subspace of l_2 spanned by the $\{\phi_{kj}\}$, and $P_{\mathcal{S}}c$ the projection of the sequence c into $\mathcal{S}$, (58) is

$$E\,x(t_1) = \sigma_\alpha^2 \|P_{\mathcal{S}}c\|^2 \leq \sigma_\alpha^2 \|c\|^2. \tag{59}$$

We will now bound $\|c\|^2$.

Define the function $w_1(t)$, $-\infty < t < \infty$, by

$$w_1(t) = g(t_1 - t),$$

with t_1 still held fixed. Then

$$c(n) = w_1(nT),$$

and Proposition 1 yields

$$C_T(f) = \frac{1}{T} \sum_{k=-\infty}^{\infty} W_1\left(f - \frac{k}{T}\right),$$

where $W_1(f)$ is the ordinary Fourier transform of $w(t)$. Since $W_1(f) = G^*(f)e^{-i2\pi f t_1}$, we have, from (15),

$$\|c\|^2 = T \int_{-1/2T}^{1/2T} |C_T(f)|^2 df$$

$$= \frac{1}{T} \int_{-1/2T}^{1/2T} \left| \sum_{k=-\infty}^{\infty} G^*\left(f - \frac{k}{T}\right) e^{-i2\pi f t_1} \right|^2 df.$$

Combining this with (59) yields Theorem 5.

Proof of Theorem 6: We begin by observing that the entire system described in Section V (up to the slicer in Fig. 3) is linear and the noise is additive and independent of the data. Thus, the error sequence $\{\hat{\alpha}_j - \alpha_j\}_{-\infty}^{\infty}$ can be written as the sum of two sequences $\{\beta_j\}_{-\infty}^{\infty}$ and $\{\gamma_j\}_{-\infty}^{\infty}$. The sequence $\{\beta_j\}$ is data dependent and is of the form

$$\beta_j = \sum_{j'} a_{jj'} \alpha_{j'}.$$

In fact, the sequence $\{\beta_j\}$ is the output of box D in Fig. 3 when the noise $z(t) \equiv 0$. The sequence $\{\gamma_j\}$ is due to the noise and is, in fact, equal to the output of box D in Fig. 3 when we set $w(t) \equiv 0$. Since the data and noise are uncorrelated, so are $\{\beta_j\}$ and $\{\gamma_j\}$. Thus the mean-squared error

$$\begin{aligned} \epsilon^2 &= \frac{1}{\nu} E \sum_{j=1}^{\nu} (\hat{\alpha}_j - \alpha_j)^2 \\ &= \frac{1}{\nu} E \sum_{j=1}^{\nu} (\beta_j + \gamma_j)^2 \\ &= \frac{1}{\nu} E \sum_{j=1}^{\nu} \beta_j^2 + \frac{1}{\nu} E \sum_{j=1}^{\nu} \gamma_j^2 \\ &\triangleq \epsilon_I^2 + \epsilon_N^2. \end{aligned} \tag{60}$$

We will overbound ϵ_I^2 and ϵ_N^2 separately.

We begin with ϵ_N^2, the error due to the noise. Thus we must overbound

$$\epsilon_N^2 = E \frac{1}{\nu} \sum_{j=1}^{\nu} \gamma_j^2,$$

where $\{\gamma_j\}$ is the output sequence of box D in Fig. 3 when $w(t) \equiv 0$. Let us define the sequence $b_0(\cdot)$ to be the output of box C when $w(t) \equiv 0$. Then

$$\gamma_j = \langle b_0, \phi_j \rangle = \sum_{n=1}^{N} b_0(n)\phi_j(n), \qquad 1 \le j \le N$$

and

$$\epsilon_N^2 = \frac{1}{\nu} E \sum_{j=1}^{\nu} \gamma_j^2 = \frac{1}{\nu} \sum_{j=1}^{\nu} \sum_{n=1}^{N} \sum_{m=1}^{N} \phi_j(n)\phi_j(m) E\, b_0(n) b_0(m). \qquad (61)$$

Next observe that $b_0(\cdot)$ is a stationary random sequence with $E\, b_0(n) = 0$, $E\, b_0(n)b_0(m) = R_{b_0}(n - m)$. The sequence $R_{b_0}(n)$, $-\infty < n < \infty$, is the inverse Fourier transform of its spectral density $S_{b_0}(f)$, which is, from Proposition 1,

$$S_{b_0}(f) = \frac{1}{T} |C_T(f)|^2 \sum_{k=-\infty}^{\infty} N_Z\left(f - \frac{k}{T}\right). \qquad (62)$$

Returning to (61), we write

$$\epsilon_N^2 = \frac{1}{\nu} \sum_{j=1}^{\nu} \sum_{n=1}^{N} \phi_j(n) \left[\sum_{m=1}^{N} R_{b_0}(n - m)\phi_j(m) \right]. \qquad (63)$$

The quantity in brackets in (63) is $d_j(n)$, where the sequence $d_j(\cdot)$ is the convolution of the sequences $R_{b_0}(\cdot)$ and $\phi_j(\cdot)$. Further, (63) can be written as

$$\epsilon_N^2 = \frac{1}{\nu} \sum_{j=1}^{N} \sum_{n=1}^{N} \phi_j(n) d_j(n) = \frac{1}{\nu} \sum_{j=1}^{N} \langle \phi_j, d_j \rangle.$$

From the Parseval relation (14), we have

$$\epsilon_N^2 = \frac{T}{\nu} \sum_{j=1}^{N} \int_{-1/2T}^{1/2T} D_{jT}(f) \Phi_{jT}^*(f) df, \qquad (64)$$

where $\Phi_{jT}(f)$, $D_{jT}(f)$ are the transforms of $\phi_j(\cdot)$, $d_j(\cdot)$, respectively. Furthermore, the convolution theorem (13) yields

$$D_{jT}(f) = \Phi_{jT}(f) S_{b_0}(f),$$

so that (64) becomes

$$\begin{aligned} \epsilon_N^2 &= \frac{T}{\nu} \sum_{j=1}^{\nu} \int_{-1/2T}^{1/2T} S_{b_0}(f) |\Phi_{jT}(f)|^2 df \\ &= \frac{T}{\nu} \int_{-F}^{F} S_{b_0}(f) \left(\sum_{j=1}^{\nu} |\Phi_j(f)|^2 \right) df \\ &+ \frac{T}{\nu} \sum_{j=1}^{\nu} \int_{F<|f|\le 1/2T} S_{b_0}(f) |\Phi_j(f)|^2 df\,. \end{aligned}$$

Using Theorem 7 (Appendix C), we obtain

$$\begin{aligned} &\le \frac{N}{\nu} \int_{-F}^{F} (T\, S_{b_0}(f)) df + \left[\sup_{F<|f|\le 1/2T} S_{b_0}(f) \right] \frac{1}{\nu} \sum_{j=1}^{\nu} \| \mathcal{B}' \phi_j \|^2 \\ &= \frac{N}{\nu} \int_{-F}^{F} |C_T(f)|^2 \sum_k N_Z\left(f - \frac{k}{T}\right) df + A_2 \frac{1}{\nu} \sum_{j=1}^{\nu} (1 - \lambda_j), \end{aligned}$$

which is (41).

It remains to verify (42), which is an upper bound on the data-dependent error or intersymbol interference error ϵ_I^2. Thus, set the noise $z(t) \equiv 0$, and

$$\epsilon_I^2 = E \frac{1}{\nu} \sum_{j=1}^{\nu} (\hat{\alpha}_j - \alpha_j)^2. \tag{65}$$

We begin by observing that, since the sequences $\phi_j(\cdot), j = 1, \cdots, \nu$, are orthonormal, any sequence $c_0(\cdot)$ can be written as

$$c_0(\cdot) = \sum_{j=1}^{\nu} \gamma_j \phi_j(\cdot) + r(\cdot)\,, \tag{66}$$

where $\langle r, \phi_j \rangle = 0$ and $\gamma_j = \langle c_0, \phi_j \rangle$, $1 \le j \le \nu$. Applying (66) to the sequence $\mathcal{D}(\hat{a} - a)$, where $\hat{a}(\cdot)$ and $a(\cdot)$ are as defined in Section V and $\mathcal{D} = \mathcal{D}_N$ is the index-limiting operator defined in Section II, we obtain

$$\begin{aligned}\mathcal{D}(\hat{a} - a) &= \sum_{j=1}^{\nu} \langle \mathcal{D}(\hat{a} - a), \phi_j \rangle \phi_j + r(\cdot) \\ &= \sum_{j=1}^{\nu} \langle \hat{a} - a, \phi_j \rangle \phi_j + r\,,\end{aligned}$$

where $\langle r, \phi_j \rangle = 0$, $1 \le j \le \nu$. Thus

$$\begin{aligned}\|\mathcal{D}(\hat{a} - a)\|^2 &= \sum_{j=1}^{\nu} \langle \hat{a} - a, \phi_j \rangle^2 + \|r\|^2 \\ &\ge \sum_{j=1}^{\nu} \langle \hat{a} - a, \phi_j \rangle^2.\end{aligned} \tag{67}$$

Now from (31),

$$\langle a, \phi_j \rangle = \alpha_j, \qquad 1 \le j \le \nu,$$

and from (35a),

$$\langle \hat{a}, \phi_j \rangle = \hat{\alpha}_j.$$

Thus (67) is

$$\|\mathcal{D}(\hat{a} - a)^2\|^2 \ge \sum_{j=1}^{\nu} (\hat{\alpha}_j - \alpha_j)^2,$$

so that the mean-squared error,

$$\epsilon_I^2 \triangleq \frac{1}{\nu} E \sum_{j=1}^{\nu} (\hat{\alpha}_j - \alpha_j) \le \frac{1}{\nu} E \|\mathcal{D}(\hat{a} - a)\|^2. \tag{68}$$

Ineq. (68) relates the error ϵ_I^2 to the error which the system makes in transmitting the sequence $a(\cdot)$.

We proceed to overbound $(1/\nu)E\|\mathcal{D}(\hat{a} - a)\|^2$. We now define

$$H_T(f) = C_T(f)B_T(f) = \frac{1}{T}C_T(f) \sum_{k=-\infty}^{\infty} G\left(f - \frac{k}{T}\right) H_C\left(f - \frac{k}{T}\right). \tag{69}$$

$B_T(f)$ is defined by (42b). It is easy to verify that $H_T(f)$ is the (sequence) transfer function of the overall system from the input to box B at the transmitter (Fig. 2), through the channel [defined by (33)], and through the sampler and box C at the receiver (Fig. 3). Thus with $h(\cdot)$, the inverse transform of $H_T(f)$,

$$\hat{a}(n) = \sum_{m=-\infty}^{\infty} h(n-m)a(m), \qquad -\infty < n < \infty. \tag{70}$$

Further, the error sequence

$$\begin{aligned} \hat{a}(n) - a(n) &= \sum_{m=-\infty}^{\infty} a(m)[h(n-m) - \delta_{n,m}] \\ &= \sum_{m=-\infty}^{\infty} a(m)u(n-m), \qquad -\infty < n < \infty, \end{aligned} \tag{71a}$$

where the sequence $u(\cdot)$ is defined by

$$u(n) = h(n) - \delta_{n,0}, \qquad -\infty < n < \infty. \tag{71b}$$

The transform of the sequence $u(\cdot)$ is

$$U_T(f) = H_T(f) - 1. \tag{72}$$

Now, with $a_k(\cdot)$ as defined by (31a), we define the convolution of $a_k(\cdot)$ and $u(\cdot)$ to be

$$\begin{aligned} v_k(n) = (a_k * u)(n) &= \sum_{m=-\infty}^{\infty} a_k(m)u(n-m) \\ &= \sum_{m=Nk+1}^{N(k+1)} a_k(m)u(n-m). \end{aligned}$$

Since $a(\cdot) = \sum_{k=-\infty}^{\infty} a_k(\cdot)$, (71) is

$$\hat{a} - a = a * u = \sum_{k=-\infty}^{\infty} v_k. \tag{73}$$

We next introduce the time-truncation operators $\mathcal{D}^{(k)}$, $-\infty < k < \infty$, defined by

$$(\mathcal{D}^{(k)}b_0)(n) = \begin{cases} b_0(n), & Nk+1 \le n \le N(k+1), \\ 0, & \text{otherwise}. \end{cases} \tag{74}$$

Of course, $\mathcal{D} = \mathcal{D}^{(0)}$. Then, from (73),

$$\begin{aligned} \frac{1}{\nu}\|\mathcal{D}(\hat{a} - a)\|^2 &= \frac{1}{\nu}\|\mathcal{D}^{(0)}(\hat{a} - a)\|^2 \\ &= \frac{1}{\nu}\left\|\sum_{k=-\infty}^{\infty} \mathcal{D}^{(0)}v_k\right\|^2 = \frac{1}{\nu}\sum_{k,k'=-\infty}^{\infty} \langle \mathcal{D}^{(0)}v_k, \mathcal{D}^{(0)}v_{k'}\rangle. \end{aligned} \tag{75}$$

Now let us observe that v_k depends (through a_k) on the data symbols in and only in the kth data block

$$\alpha_{\nu k+1}, \cdots, \alpha_{\nu(k+1)}.$$

Since all the data symbols are assumed to be statistically independent, we conclude that v_k and $v_k{}'$ $(k \neq k')$ are also statistically independent and uncorrelated. Thus

$$E\langle \mathcal{D}^{(0)}v_k, \mathcal{D}^{(0)}v_{k'}\rangle = 0, \qquad k \neq k',$$

and (68) and (75) are

$$\epsilon_I^2 \leq \frac{1}{\nu} E\|\mathcal{D}(\hat{a} - a)\|^2 = \frac{1}{\nu} \sum_{k=-\infty}^{\infty} E\|\mathcal{D}^{(0)}v_k\|^2. \tag{76}$$

We now make another observation about the sequence $v_k(\cdot)$. As we observed in the proof of Theorem 4, the N random variables $\{a_k(n)\}_{n=Nk+1}^{N(k+1)}$ have the same statistics as the N random variables $\{a_0(n)\}_{n=1}^{N}$. It follows that, for $-\infty < n < \infty$, $v_k(n)$ has the same statistics as $v_0(n - Nk)$, so that

$$\begin{aligned} E\|\mathcal{D}^{(0)}v_k\|^2 &= E \sum_{n=1}^{N} v_k^2(n) \\ &= E \sum_{n=1}^{N} v_0^2(n - Nk) = E \sum_{n=-Nk+1}^{N(-k+1)} v_0^2(n) \\ &= E\|\mathcal{D}^{(-k)}v_0\|^2. \end{aligned} \tag{77}$$

Substituting (77) into (76), we have

$$\epsilon_I^2 \leq \frac{1}{\nu} \sum_{k=-\infty}^{\infty} E\|\mathcal{D}^{(-k)}v_0\|^2 = \frac{1}{\nu} E\|v_0\|^2. \tag{77a}$$

Now from the convolution formula (13) and the Parseval relation (15),

$$\begin{aligned} E\|v_0\|^2 &= E\|a_0 * u\|^2 \\ &= TE\int_{-1/2T}^{1/2T} |A_T(f)|^2|U_T(f)|^2 df. \end{aligned} \tag{78}$$

Since

$$A_T(f) = \sum_{j=1}^{\nu} \alpha_j \Phi_{jT}(f)$$

and the $\{\alpha_j\}_1^{\nu}$ are uncorrelated,

$$E|A_T(f)|^2 = \sum_{j=1}^{\nu} \sigma_\alpha^2 |\Phi_{jT}(f)|^2. \tag{79}$$

Substituting (79) and (78) into (77), we have

$$\epsilon_I^2 \leq \frac{1}{\nu} T \int_{-1/2T}^{1/2T} |U_T(f)|^2 \sum_{j=1}^{\nu} \sigma_\alpha^2 |\Phi_{jT}(f)|^2 df$$

$$= \frac{\sigma_\alpha^2}{\nu} T \int_{-F}^{F} |U_T(f)|^2 \sum_{j=1}^{\nu} |\Phi_{jT}(f)|^2 df$$

$$+ \frac{\sigma_\alpha^2}{\nu} T \sum_{j=1}^{\nu} \int_{F \le |f| \le 1/2T} |U_T(f)|^2 |\Phi_{jT}(f)|^2 df$$

$$\le \frac{\sigma_\alpha^2}{\nu} T N \int_{-F}^{F} |U_T(f)|^2 df$$

$$+ \frac{\sigma_\alpha^2}{\nu} T \sup_{F \le |f| \le 1/2T} |U_T(f)|^2 \sum_{j=1}^{\nu} \int_{F \le |f| \le 1/2T} |\Phi_j(f)|^2 df.$$

Using (72) and (69), which define $U(f)$, and the definition of A_3 (42c), we obtain

$$\epsilon_I^2 \le \sigma_\alpha^2 \left(\frac{N}{\nu}\right) T \int_{-F}^{F} |C_T(f) B_T(f) - 1|^2 df$$

$$+ A_3 \frac{1}{\nu} \sum_{j=1}^{\nu} \|\mathcal{B}'\phi_j\|^2.$$

Since $\|\mathcal{B}'\phi_j\|^2 = 1 - \lambda_j$, we have established (42), completing the proof of Theorem 6.

APPENDIX A

Proof of Theorem 3

Let $T, F > 0$, with $W \triangleq FT < 1/2$ be given. Let $N = 1, 2, \cdots$, also be given. Define the sequence $\gamma(\cdot)$ by

$$\gamma(n) = \frac{\sin 2\pi W n}{\pi n}, \qquad -\infty < n < \infty. \tag{80}$$

The transform of $\gamma(\cdot)$ is easily seen to be

$$\Gamma_T(f) = \sum_{n=-\infty}^{\infty} \gamma(n) e^{-i2\pi f T n} = \begin{cases} 1, & |f| \le F \\ 0, & F < |f| \le \dfrac{1}{2T}. \end{cases} \tag{81}$$

The bandlimiting operator $\mathcal{B} = \mathcal{B}_F$ is therefore defined by $b = \mathcal{B}a$, where $B_T(f) = \Gamma_T(f) A_T(f)$.

Let K be the $N \times N$ matrix with (m, n)th entry $\gamma(n - m)$, $1 \le n, m \le N$. Consider the matrix eigenvalue equation

$$K\overline{\alpha} = \lambda \overline{\alpha}, \tag{82}$$

where $\overline{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_N)^t$. Equation (82) is equivalent to

$$\sum_{m=1}^{N} \gamma(n - m)\alpha_m = \lambda \alpha_n, \qquad 1 \le n \le N. \tag{82'}$$

Since K is a symmetric matrix, eq. (82) or (82′) has N (not necessarily distinct) real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ and a corresponding set of N real orthogonal eigenvectors $\overline{\alpha}_j = (\alpha_{j1}, \cdots, \alpha_{jN})^t$, $1 \leq j \leq N$. We assume that the eigenvectors are normalized so that $\overline{\alpha}_j^t \overline{\alpha}_{j'} = \delta_{jj'}$. Also we can write

$$\sum_{m=1}^{N} \gamma(n-m)\alpha_{jm} = \lambda_j \alpha_{jn}, \qquad 1 \leq n \leq N. \tag{83}$$

We can now define our sequences $\{\phi_j(\cdot)\}$. Let

$$\phi_j(n) = \begin{cases} \alpha_{jn}, & 1 \leq n \leq N, \\ 0, & \text{otherwise.} \end{cases} \tag{84}$$

Equation (83) is therefore

$$\sum_{n=1}^{N} \gamma(n-m)\phi_j(m) = \lambda_j \phi_j(n), \qquad 1 \leq n \leq N, \tag{85}$$

which is equivalent to

$$\mathcal{D}\mathcal{B}\phi_j = \lambda_j \phi_j, \qquad 1 \leq j \leq N, \tag{86}$$

where $\mathcal{D} = \mathcal{D}_N$, and $\mathcal{B} = \mathcal{B}_F$. This is Theorem 3B.

Now, for $1 \leq j \leq N$, let $c_j(\cdot) = \mathcal{B}\phi_j$. Then (86) implies that

$$c_j(n) = \lambda_j \phi_j(n), \qquad 1 \leq n \leq N. \tag{87}$$

Thus

$$\langle c_j, \phi_j \rangle = \sum_{n=1}^{N} c_j(n)\phi_j(n) = \sum_{n=1}^{N} \lambda_j \phi_j^2(n)$$
$$= \lambda_j \|\phi_j\|^2 = \lambda_j.$$

Further, since the transform of $c_j(\cdot)$ is $C_{jT}(f) = \Gamma_T(f)\Phi_{jT}(f)$, the Parseval relation (14) yields

$$\lambda_j = \langle c_j, \phi_j \rangle = T \int_{-1/2T}^{1/2T} \Gamma_T(f)\Phi_{jT}(f)\Phi_{jT}^*(f)df$$
$$= T \int_{-F}^{F} |\Phi_j(f)|^2 df. \tag{88}$$

From (88), $\lambda_j \leq \|\phi_j\|^2 = 1$ and $\lambda_j \geq 0$. In fact, if $\lambda_j = 0$, then $\Phi_j(f) = 0$ for $|f| \leq F$. But, since $\Phi_j(f)$ is a polynomial in $e^{-i2\pi fT}$, it vanishes on an interval only if it vanishes identically, which contradicts $\|\phi_j\| = 1$. Thus $\lambda_j > 0$ for $1 \leq j \leq N$. Since $\Sigma_1^N \lambda_j = \text{trace } K = 2WN$, the $\{\lambda_j\}_1^N$ can be labeled so that they satisfy Theorem 3A.

Now Theorem 3C follows from

$$\langle \phi_j, \phi_k \rangle = \overline{\alpha}_j^t \overline{\alpha}_k = \delta_{jk}.$$

Theorem 3D is established as follows:

$$\langle \mathcal{B}\phi_j, \mathcal{B}\phi_k \rangle = \langle c_j, c_k \rangle = T \int_{-1/2T}^{1/2T} C_{jT}(f) C_{kT}^*(f) df$$

$$= T \int_{-1/2T}^{1/2T} \Gamma_T(f)\Phi_{jT}(f)\Phi^*_{kT}(f)df$$

$$= \langle \mathcal{B}\phi_j, \phi_k \rangle \overset{(1)}{=} \langle \mathcal{D}\mathcal{B}\phi_j, \phi_k \rangle$$

$$\overset{(2)}{=} \lambda_j \langle \phi_j, \phi_k \rangle = \lambda_j \delta_{jk} .$$

Step (1) follows from the fact that ϕ_k has support on $[1, N]$ so that for any $a(\cdot)$, $\langle a, \phi_k \rangle = \langle \mathcal{D}a, \phi_k \rangle$. Step (2) follows from Theorem 3B.

To prove Theorem 3F, observe* that

$$\sum_{j=1}^{N} \lambda_j = \text{trace } K = 2WN \tag{89}$$

and

$$\sum_{j=1}^{N} \lambda_j^2 = \text{trace } (K^t K) = \sum_{n,m=1}^{N} \gamma^2(n-m).$$

Substitution of the formula for $\gamma(n)$ (80) and a simple computation yields

$$\sum_{1}^{N} \lambda_j^2 \geq 2WN - 0 \,(\log N), \tag{90}$$

as $N \to \infty$. Combining (89) and (90), we have

$$\frac{1}{N}\sum_{1}^{N} \lambda_j(1-\lambda_j) = \frac{1}{N}\sum_{1}^{N} \lambda_j - \lambda_j^2 \leq \frac{0\,(\log N)}{N} \xrightarrow{N} 0. \tag{91}$$

Let $S = \{j:\ \delta < \lambda_j < 1 - \delta\}$. Then (91) yields

$$\frac{1}{N}\,\delta^2\,(\text{card } S) \leq \frac{1}{N}\sum_{1}^{N} \lambda_j(1-\lambda_j) \xrightarrow{N} 0,$$

which is, on dividing by δ^2, Theorem 3E.

Theorem 3F follows directly from Theorems 3A and 3F. Theorem 3G is established in Ref. 4. Finally, Theorem 3H follows immediately from the definition of the $\phi_j(\cdot)$, λ_j, $1 \leq j \leq N$.

APPENDIX B

Saltzberg's Bound

Saltzberg's bound[2,3] states that if ξ is a random variable defined by

$$\xi_1 = \sum_{j=-\infty}^{\infty} \eta_j \alpha_j,$$

where $\{\alpha_j\}_{-\infty}^{\infty}$ are i.i.d. copies of the r.v. α, defined by (29), and $\{\eta_j\}$ are fixed coefficients, then the moment generating function of ξ_1

$$M_{\xi_1}(s) = E\, e^{s\xi_1} \geq \exp\left\{\frac{s^2\sigma_1^2}{2}\right\},\ s \geq 0, \tag{92a}$$

* This trick is used in Ref. 7.

where

$$\sigma_1^2 = \text{Var}\ \xi_1 = \sigma_\alpha^2 \sum_{j=-\infty}^{\infty} \eta_j^2. \tag{92b}$$

The right member of (92a) is the moment-generating function of a Gaussian r.v. with zero mean and variance σ_1^2.

Now let $\xi = \xi_1 + \xi_2$, where ξ_1 is as above and ξ_2 is a Gaussian r.v. with zero mean and variance σ_2^2. Let ξ_1 and ξ_2 be statistically independent. Then, from Saltzberg's bound (92), the moment-generating function of ξ,

$$M_\xi(s) = M_{\xi_1}(s) \cdot M_{\xi_2}(s) \leq \exp\left\{\frac{s^2}{2}(\sigma_1^2 + \sigma_2^2)\right\}, \qquad s \geq 0.$$

It follows from the Chernoff bounding technique that, for $r > 0$

$$\Pr\{\xi > \mathrm{r}\} \leq \exp\left\{-\frac{r^2}{2(\sigma_1^2 + \sigma_2^2)}\right\}. \tag{93}$$

Let us now apply (93) to establish the claims made in remarks *iv* and *v* in Section V. In remark *iv*, observe that $x(t)$ is a random variable of the form of ξ_1, i.e., a linear combination of the data symbols $\{\alpha_j\}$. If we apply (93) with $\xi_1 = x(t)$, $\xi_2 \equiv 0$, we obtain the inequality of remark *iv*.

Next consider Remark *v*. Observe that, due to the linearity of the system, the error $\hat{\alpha}_j - \alpha_j$ is of the form of ξ, with $\sigma_1^2 + \sigma_2^2 = \epsilon_j^2$. Thus (93) yields

$$\Pr\{(\hat{\alpha}_j - \alpha_j) > 1\} \leq \exp\left\{\frac{-1}{2\epsilon_j^2}\right\}.$$

Since $\tilde{\alpha}_j \neq \alpha_j$, only when $|\hat{\alpha}_j - \alpha_j| > 1$, we have

$$P_{ej} = \Pr\{\tilde{\alpha}_j \neq \alpha_j\} \leq 2 \exp\left\{\frac{-1}{2\epsilon_j^2}\right\}.$$

APPENDIX C

Theorem 7: Let $\{\phi_j(\cdot)\}_{j=1}^N$ be an orthonormal set of sequences (in l_2) with support on $[1, N]$. That is, $D_N\phi_j = \phi_j$ and $\langle \phi_j, \phi_k \rangle = \delta_{jk}$, $1 \leq j,k \leq N$. Let $\Phi_{jT}(f)$, $-\infty < f < \infty$, be the Fourier transform of $\phi_j(\cdot)$, $1 \leq j \leq N$. Then

$$\sum_{j=1}^{N} |\Phi_{jT}(f)|^2 = N, \qquad -\infty < f < \infty.$$

Proof: Let $v(n) = e^{-i2\pi fTn}$ or 0 according as $n \in [1, N]$ or $n \notin [1, N]$. From the orthonormality of the $\{\phi_j\}_1^N$, we conclude that they span the N-dimensional space of complex-valued sequences with support on $[1, N]$. Thus we can write

$$v(n) = \sum_{j=1}^{N} v_j \phi_j(n), \qquad -\infty < n < \infty,$$

where

$$v_j = \langle v, \phi_j \rangle = \sum_{n=1}^{N} v(n)\phi_j(n)$$

$$= \sum_{n=-\infty}^{\infty} e^{-i2\pi fTn}\phi_j(n) = \Phi_{jT}(f), \qquad 1 \le j \le N.$$

Furthermore, the orthonormality of the $\{\phi_j\}$ also implies

$$\sum_{n=1}^{N} |v(n)|^2 = \sum_{j=1}^{N} |v_j|^2 = \sum_{j=1}^{N} |\Phi_{jT}(f)|^2.$$

Since $|v(n)| \equiv 1$, we have established the theorem.

REFERENCES

1. R. W. Lucky, J. Salz, and E. J. Weldon, *Principles of Data Communication,* New York: McGraw-Hill, 1968.
2. J. E. Mazo, "Quantizing Noise and Data Transmission," B.S.T.J., *47,* No. 8 (October 1968), pp. 1737–1753 (Section IV).
3. B. R. Saltzberg, "Intersymbol Interference Error Bounds with Application to Ideal Bandlimited Signalling," IEEE Trans. Inform. Th., *14* (July 1968), pp. 563–568.
4. D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—V: The Discrete Case," B.S.T.J., *57,* No. 5 (May–June 1978), pp. 1371–1430.
5. D. Slepian, "Program to Compute Discrete Prolate Spheroidal Wave Functions, Sequences and Eigenvalues," unpublished work.
6. H. S. Witsenhausen, unpublished work.
7. H. J. Landau and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty—III," B.S.T.J., *41,* No. 4 (July 1962), pp. 1295–1336.