THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 9, November, 1973
Printed in U.S.A.

A Geometric Theory of Intersymbol Interference

Part I: Zero-Forcing and Decision-Feedback Equalization

By D. G. MESSERSCHMITT

(Manuscript received May 14, 1973)

A linear-space geometric theory of intersymbol interference is introduced in this paper. An equivalence between the structure of intersymbol interference and a wide-sense stationary discrete random process is demonstrated and exploited to demonstrate the equivalence of zero-forcing (decision-feedback) equalization to minimum mean-square error linear interpolation (prediction) of a random process. This equivalence is used to quickly derive the properties of these equalizers and give them additional geometric interpretation. Results from prediction theory are used to develop practical computational methods of determining the tap-gains of the infinite equalizers for both rational and nonrational channel power spectra. Finally, the theory of reproducing kernel Hilbert spaces is used to develop a theory of equalization for nonstationary channels with nonstationary noise.

I. INTRODUCTION

The analysis of digital communication systems from a geometrical viewpoint—the viewing of waveforms as points in a signal space and the identification of cross-correlation with the formation of an inner product—is by now well established. To a large extent, this approach has been popularized by the book of Wozencraft and Jacobs.[1] However, when it comes to analyzing systems with intersymbol interference, frequency-domain techniques have almost exclusively been relied upon. The purpose of this paper is to consider pulse-amplitude modulation (PAM) systems with intersymbol interference from a geometric standpoint, and more specifically to develop a geometric theory of equalization.

Consideration of the geometric structure of intersymbol interference leads immediately to the observation of a striking correspondence to the theory of minimum mean-square error (MMSE) linear estimation of a wide-sense stationary discrete-parameter random process. The fact that the latter subject is almost exclusively treated by geometric methods[2,3] is further impetus for this approach to equalization.

The theories of linear zero-forcing equalization and decision-feedback equalization are well established. The properties of linear equalization

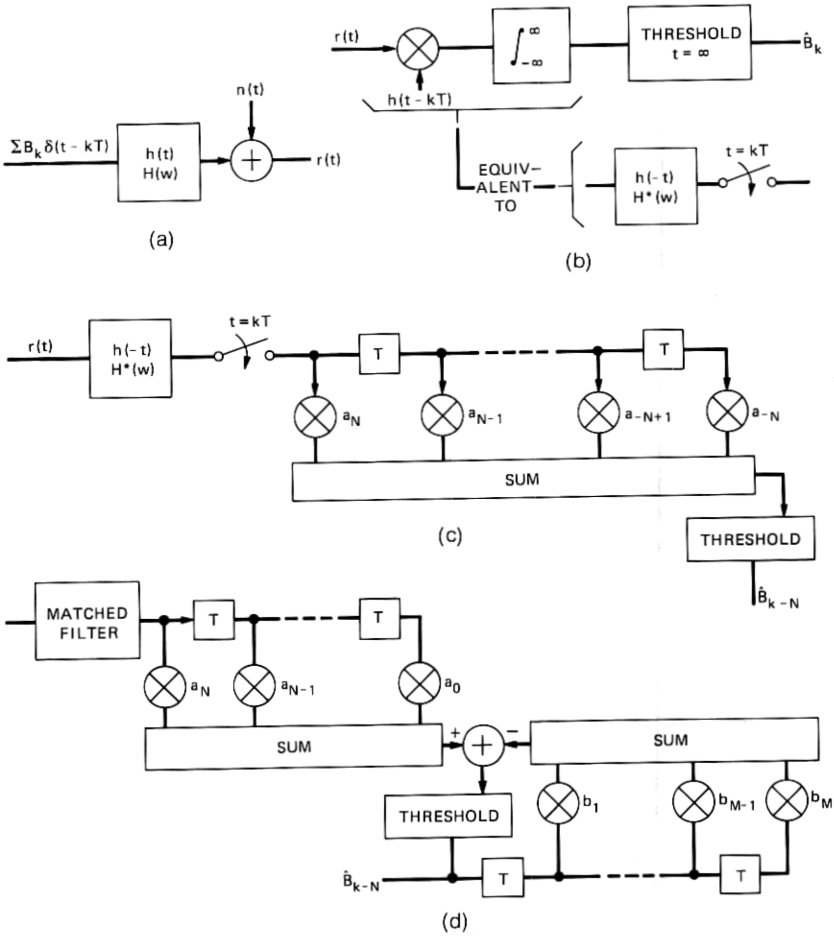

Fig. 1—(a) Communication system model. (b) Matched-filter receiver. (c) Zero-forcing equalizer. (d) Decision-feedback equalizer.

are summarized by Lucky, et al.,[4] while the present state of knowledge of decision-feedback equalization is summarized by Monsen[5] and Price.[6] The primary analysis tools which have been used are the calculus of variations in the case of linear equalization and Toeplitz forms in the case of decision-feedback equalization.

In this paper, the geometric approach enables us to treat the two types of equalization simultaneously using the same mathematical framework, in which the relationship between them becomes very clear and many of their known properties are given an additional geometric interpretation. Many of the results follow directly from the theory of MMSE estimation. In addition to the unification and reinterpretation of previously known results, the geometric approach leads to extensions of the theory in several directions. Among these are the derivation of an orthogonal expansion in Section 2.4 which is useful in many problems involving intersymbol interference, the development of practical iterative techniques for determining equalizer tap-gains (the infinite case) in Section 3.4, the extension of the theory of equalization to nonstationary noise and a time-varying channel in Section IV, and numerous results on the minimum distance problem associated with the performance analysis of the Viterbi algorithm maximum likelihood detector in a companion paper.[7]

This paper together with a companion one[7] expand upon an earlier talk.[8] Readers desiring a limited and short treatment of this subject may wish to refer there. The geometrical approach to intersymbol interference was also employed to a limited extent in the author's thesis.[9]

1.1 *Problem Statement*

We will consider the detection of a sequence of digital data digits, B_k, each assuming one of a finite and predetermined number of levels, from the reception

$$r(t) = \sum_{k=N_1}^{N_2} B_k h(t - kT) + n(t) \tag{1}$$

as determined from the communication system model of Fig. 1a. It will be assumed initially that $n(t)$ is white Gaussian noise (this assumption will be relaxed in Section IV).† A simple matched-filter receiver for the reception of $r(t)$ is shown in Fig. 1b. In the first of two equivalent formulations of this receiver, the reception is cross-correlated

† The assumption of Gaussian noise is not necessary for the majority of results to follow, and in particular those which involve only second-order statistics of the noise.

with $h(t - kT)$ and the decision on B_k made by applying a series of thresholds to the result; in the second formulation the cross-correlator is realized as a filter with impulse response $h(-t)$ (commonly called a matched filter) whose output is sampled at $t = kT$. The matched-filter receiver is optimum when there is no intersymbol interference, but in the presence of intersymbol interference the matched filter will respond to more than a single data digit and the performance of the receiver will be degraded.

When there is intersymbol interference, a common approach is to build a linear filter, called a zero-forcing equalizer (ZFE), which responds to only a single time-translate of $h(t)$ (this can only be approximated in practice). The most common form of this equalizer, shown in Fig. 1c, is a matched filter followed by transversal filter (MFTF). As $N \rightarrow \infty$ the tap-gains of the transversal filter can be chosen such that the threshold input is a function of only a single data digit. It is important to note for future reference that the MFTF can also be modeled in the manner of Fig. 1b as a cross-correlation of $r(t)$ with a linear sum of time translates of $h(t)$,

$$\sum_{m=-N}^{N} a_m h(t - mT).$$

The decision-feedback equalizer (DFE) embodies a slightly different philosophy in which the DFE forward filter is allowed to respond to past (but not future) translates at $h(t)$; the residual interference from past data digits is then subtracted out prior to the decision threshold using past decisions. A realization of the DFE using again the MFTF approach is shown in Fig. 1d. The tap coefficients are now chosen to null the response to future data digits; this can be accomplished as $N \rightarrow \infty$.

The shortcoming of both the ZFE and DFE is that their linear filters remove intersymbol interference without regard to the effect on the noise; the result is that in eliminating the intersymbol interference (or a portion thereof) they necessarily enhance the noise.† It seems clear intuitively that since the DFE eliminates interference from only future data digits, it has more degrees of freedom than the ZFE and should therefore be capable of less noise enhancement. A proof that this is always the case has been given by Price;[6] his method was to determine an explicit formula for the DFE S/N ratio using

† In addition, the DFE is susceptible to decision errors. The effect of errors will not receive consideration here.

Toeplitz form theory and compare it with the known S/N ratio of the ZFE.[4] Additional interpretation of this result will be given in Section 3.1.

A review of some requisite material on linear spaces and MMSE linear estimation is given in Sections 2.1 and 2.2. Readers familiar with this material are nevertheless urged to scan these sections for notation to be employed in the remainder of the paper. The ZFE and DFE are reformulated in Section 2.3. In Section 2.4 the relationship between intersymbol interference and MMSE estimation is discussed, and a useful orthogonal expansion arising out of this relationship is derived in Section 2.5.

Section III develops a geometric theory of the ZFE and DFE. Conditions necessary and sufficient for the existence of these equalizers are given in Section 3.1, their performance is discussed in Section 3.2, a useful property of the DFE with regard to its output noise sequence is interpreted in Section 3.3, methods of calculating the tap-gains are derived in Section 3.4, and the relationship between finite and infinite transversal filter equalizers receives consideration in Section 3.5.

Sections II and III are concerned with additive white noise exclusively. Section IV extends the theory to colored Gaussian noise, nonstationary Gaussian noise, and a time-varying channel using the theory of reproducing kernel Hilbert spaces (RKHS).

II. AN EQUIVALENCE TO DISCRETE RANDOM PROCESSES

The structure of the intersymbol interference in (1) will now be shown to have an equivalence to a wide-sense stationary random process. The starting point will be a quick review of linear spaces and of linear mean-square error (MMSE) estimation of a random process.

2.1 *Hilbert Space Notation*[10]

An inner product space $\mathfrak{L}$ consists of a linear space together with a defined inner product $\langle x, y\rangle$ between two elements x and y. All spaces in this paper are Hilbert spaces, which consist of an inner product space satisfying an additional closure property (specifically, the limits of Cauchy sequences must be in the space). The inner product induces a norm, or "length" of a vector,

$$\|x\| \triangleq \langle x,x\rangle \tag{2}$$

and the notion of the distance between two vectors, $\|x - y\|$. The geometrical interpretation of these quantities is illustrated in Fig. 2.

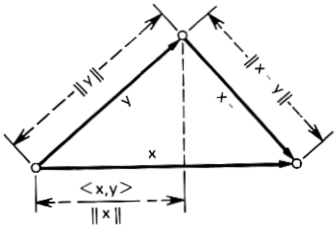

Fig. 2—Interpretation of inner product, norm, and distance.

A subspace of $\mathfrak{L}$ is any set of vectors which itself constitutes a linear space. If x_k, $k \in I$ is a countable or finite sequence of vectors, then we denote by $M(x_k, k \in I)$ the closure of the subspace consisting of all finite linear combinations of elements of the set $\{x_k, k \in I\}$ and call this the subspace spanned by the x_k's. It is convenient to think of elem nts of $M(x_k, k \in I)$ as convergent (possibly) infinite sums of the form

$$\sum_{k \in I} a_k x_k$$

even th h in some obscure cases not all elements can be expressed in this

In r ninimization problems it is desired to find the element of som d subspace M which is closest to a vector y; the resulting element called the projection of y on M, is denoted by $P(y; M)$, and satisfies the orthogonality property

$$\langle y - P(y; M), x \rangle = 0 \tag{3}$$

for all $x \in M$.† The geometric interpretation of (3) is shown in Fig. 3 for a one-dimensional subspace spanned by x; for this case the projection must be a scalar times x and the validity of (3) is apparent.

2.2 *Review of Linear Mean-Square Interpolation and Prediction*[2,3]

We will now quickly review the theory of linear mean-square estimation of a random variable.

The set of random variables with zero mean and finite variance is a linear space, since the sum of any two such random variables itself has these properties. This set is also a Hilbert space with inner product

$$\langle X, Y \rangle = E(XY), \tag{4}$$

† When, as in (3), a vector is orthogonal to every vector in M, it is said to be orthogonal to M.

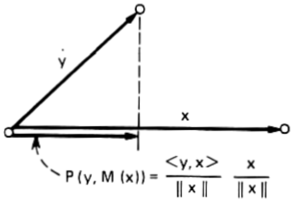

Fig. 3—Projection on subspace spanned by x.

where $E(\cdot)$ denotes expected value. It is standard to suppress the sample space dependence of a random variable as has been done in (4) because the geometric properties (inner product and norm) are determined by the value of the random variable on the whole sample space; that is, by its statistics in their entirety.

Consider now the following interpolation problem: Suppose that a sequence of zero-mean random variables X_k, $-\infty < k < \infty$, with finite variances are given and it is desired to estimate X_0 based on the observation of X_k, $k \neq 0$. If the estimate is further stipulated to be linear, it is the same as requiring that it be an element of $M(X_k, k \neq 0)$. Suppose that the estimate $\hat{X}_0$ is to be chosen in such a way that the mean-square error between X_0 and the estimate is minimized:

$$\min_{\hat{X}_0 \in M(X_k, k \neq 0)} E(X_0 - \hat{X}_0)^2. \tag{5}$$

From (4) and the previous section, the MMSE linear interpolator is

$$\hat{X}_0 = P[X_0, M(X_k, k \neq 0)], \tag{6}$$

the projection of X_0 on $M(X_k, k \neq 0)$.

A second estimation problem which will be of interest is the prediction of X_0 based only on X_k, $k > 0$ (an anticausal prediction). The MMSE linear predictor is the projection of X_0 on the subspace spanned by X_k, $k = 1, 2, \cdots$, denoted by $P[X_0, M(X_k, k > 0)]$.

2.3 *Zero-Forcing and Decision-Feedback Equalization*

We are now prepared to restate the problem of determining the ZFE and DFE filters in a linear space context. It will be assumed that the basic pulse $h(t)$ in (1) has finite energy (i.e., is square integrable),

$$\int_{-\infty}^{\infty} h^2(t)dt < \infty \cdot \tag{7}$$

The set of waveforms which satisfies (7) is a linear space, which we denote by L_2. L_2 is also a Hilbert space with inner product

$$\langle x, y\rangle = \int_{-\infty}^{\infty} x(t)y(t)dt \tag{8}$$

for any two L_2 waveforms $x(t)$ and $y(t)$. For the same reason that the sample space dependence of a random variable was suppressed in (4), the time dependence of the waveforms $x(t)$ and $y(t)$ has been suppressed on the left side of (8): it is the entire time waveform which determines the geometric properties.

The class of filters† which will be considered will be limited to those which can be modeled as an inner product (or cross-correlation) of the reception $r(t)$ with some L_2 waveform. A ZFE is a filter corresponding to a waveform $g_k(t)$ which does not respond to any translate of $h(t)$ except $h(t - kT)$,

$$\int_{-\infty}^{\infty} h(t - mT)g_k(t)dt = 0\,, \qquad m \neq k, \tag{9}$$

but does respond to $h(t - kT)$,

$$\int_{-\infty}^{\infty} h(t - kT)g_k(t)dt \neq 0\,, \tag{10}$$

in order that there be a signal on which to base the decision. It is evident that if $g_0(t)$ satisfies (9) and (10) for $k = 0$, then they are also satisfied by $g_k(t) = g_0(t - kT)$ for $k \neq 0$. Written in inner product notation, (9) and (10) become

$$\langle h_k, g_0\rangle = 0, \qquad k \neq 0, \tag{11}$$

$$\langle h_0, g_0\rangle \neq 0, \tag{12}$$

where we have written h_k for $h(t - kT)$. The analogous condition for a DFE forward filter is

$$\langle h_k, g_0\rangle = 0, \qquad k > 0, \tag{13}$$

$$\langle h_0, g_0\rangle \neq 0. \tag{14}$$

The forms of the ZFE and DFE in this symbolic notation are shown in Figs. 4a and b. The output of the linear filter is a function of B_k (a single data digit) for a ZFE and B_{k-m}, $m > 0$ (all past data digits) for a DFE. The tap-gains of the feedback transversal filter storing past decisions for the DFE are equal to the responses of g_0 to previous pulses, $\langle g_0, h_{-m}\rangle$, $m > 1$.

† In the case of the DFE, we refer only to the forward filter.

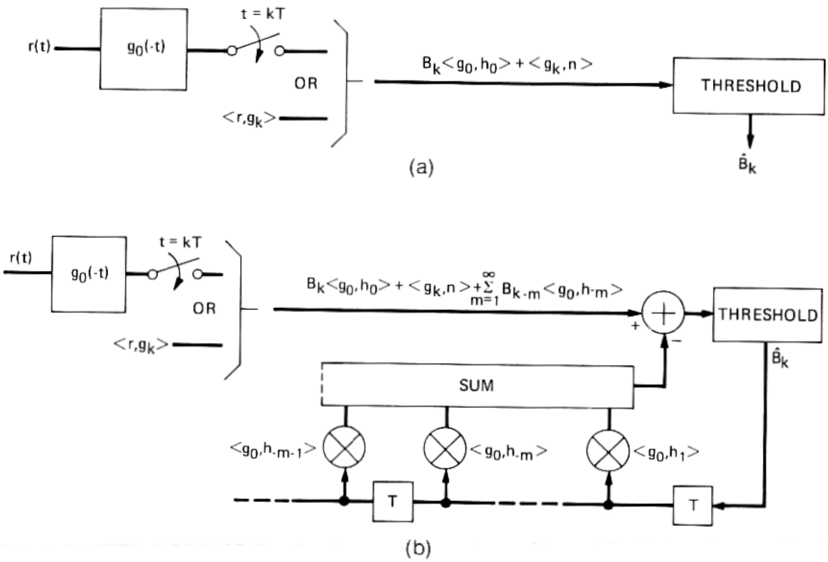

Fig. 4—Symbolic representations of the two equalizers: (a) zero-forcing equalizer; (b) decision-feedback equalizer.

2.4 *A Congruence Relationship*

Two Hilbert spaces which display an identical geometrical structure are said to be congruent[11] or unitarily equivalent.[10] Specifically, in order for two Hilbert spaces to be congruent, there must exist between them a one-to-one and onto linear mapping which preserves norms and inner products. Although the elements of two such spaces may be quite different entities, when considered as elements of their respective Hilbert spaces they have the same geometrical structure.

Define the autocorrelation function of the pulse sequence,

$$R_k = \langle h_m, h_{m+k} \rangle. \tag{15}$$

It follows from the inequality

$$0 \leq \left\| \sum_{m=0}^{N} \alpha_m h_{k_m} \right\|^2 = \sum_{m=0}^{N} \sum_{n=0}^{N} \alpha_m \alpha_n R_{k_m - k_n}$$

that $\{R_k\}$ is a nonnegative definite function. Therefore, there exists a second-order discrete random process $\{X_k\}$ which has autocorrelation R_k,

$$\begin{aligned} \langle X_m, X_{m+k} \rangle &= E(X_m, X_{m+k}) \\ &= R_k. \end{aligned} \tag{16}$$

For the random process defined in (16), $M(h_k, k \in I)$ and $M(X_k, k \in I)$ are congruent through the obvious mapping

$$\phi\left[\sum_{m=1}^{N} \alpha_m h_{k_m}\right] = \sum_{m=1}^{N} \alpha_m X_{k_m} \tag{17}$$

which is a unitary linear transformation. To verify this, observe that the mapping is linear, preserves norms,

$$\begin{aligned}\left\|\phi\left(\sum_{m=1}^{N} \alpha_m h_{k_m}\right)\right\|^2 &= \left\|\sum_{m=1}^{N} \alpha_m X_{k_m}\right\|^2 \\ &= \sum_{m=1}^{N}\sum_{n=1}^{N} \alpha_m \alpha_n R_{k_m - k_n} \\ &= \left\|\sum_{m=1}^{N} \alpha_m h_{k_m}\right\|^2,\end{aligned} \tag{18}$$

and preserves inner products by an equally simple derivation.

The mapping of (17) is only defined for finite sums. When I is an infinite set, ϕ can be extended to all of $M(h_k, k \in I)$ by taking limits in the mean. For any $f \in M(h_k, k \in I)$ there exists a sequence $\{f_k\}$, each consisting of a finite sum of the form of (17), such that $f_k \rightarrow f$. Since $\phi(f_k)$ is a Cauchy sequence from (18) ,we define $\phi(f)$ as the limit of $\phi(f_k)$, which is in $M(X_k, k \in I)$ by completeness.

There is an additional congruence which is useful. From the definition of R_k in (15), we see that

$$\begin{aligned}R_k &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(\omega)|^2 e^{j\omega kT} d\omega \\ &= \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} R(\omega) e^{j\omega kT},\end{aligned} \tag{19}$$

where

$$\begin{aligned}R(\omega) &\triangleq \sum_{m=-\infty}^{\infty} \left|H\left(\omega + m\frac{2\pi}{T}\right)\right|^2 \\ &= T\sum_{n=-\infty}^{\infty} R_n e^{jn\omega T},\end{aligned} \tag{20}$$

where $R(\omega)$ is an equivalent power spectrum of the channel. From (16), $R(\omega)/T$ is the power spectrum of the random process $\{X_k\}$. Let $L_2(-\pi/T, \pi/T; R)$ denote the Hilbert space of all complex-valued Lebesque measurable functions $f(\omega)$ with domain $|\omega| < \pi/T$ which satisfy

$$\|f(\omega)\|^2 = \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} |f(\omega)|^2 R(\omega) d\omega < \infty \tag{21}$$

with the obvious definition of the inner product. A frequently invoked congruence is between $M(X_k, -\infty < k < \infty)$ and $L_2(-\pi/T, \pi/T; R)$.[2] By implication, $L_2(-\pi/T, \pi/T; R)$ and $M(h_k, -\infty < k < \infty)$ are also congruent through the mapping

$$\psi\left(\sum_{m=1}^{N} \alpha_m h_{k_m}\right) = \sum_{m=1}^{N} \alpha_m e^{-j\omega k_m T} \tag{22}$$

as is readily verified.

In the remainder of this paper, the congruence demonstrated in this section will be exploited to demonstrate that many available results on MMSE interpolation and prediction theory are directly applicable to the equalization problems posed in Section 2.3.

2.5 *An Orthogonal Expansion*

The congruence relation of Section 2.4 will be used in this section to establish an orthogonal expansion in $M(h_k, -\infty < k < \infty)$ which will be particularly useful in the sequel.

Define the element

$$e_k^+ \triangleq h_k - P[h_k, M(h_m, m > k)] \tag{23}$$

which is the difference between a translate of $h(t)$, h_k, and its projection on the subspace of translates to its right. It will be shown later that this element is of particular significance to the DFE. For the moment, however, note that e_k^+ is equivalent to the MMSE prediction error of X_k based on X_m, $m > k$, since the projection is the optimum linear predictor. It is well known[3] that the successive prediction errors of a random process are uncorrelated random variables. The equivalent statement relating to e_k^+ is that

$$\langle e_m^+, e_n^+ \rangle = \|e_0^+\|^2 \delta_{m,n} \tag{24}$$

and it is an orthogonal sequence.† This is readily demonstrated directly by noting that e_m^+ is orthogonal to $M(h_k, k \geqq m)$, which contains e_n^+ for $n > m$. Hence, (24) follows for $n > m$ and by symmetry for $n < m$ also.

From (24) it follows that as long as

$$\|e_0^+\| > 0 \tag{25}$$

the sequence

$$w_n \triangleq e_n^+/\|e_0^+\|, \qquad -\infty < n < \infty, \tag{26}$$

is an orthonormal set in L_2. The significance of (25) is that the equiv-

† The norm of e_k^+ is independent of k since e_k^+ is a time translate of e_0^+.

alent random process must not be linearly predictable with vanishing mean-square error (in the language of Ref. 3, p. 564, X_k must be "regular," or "nondeterministic").

Expanding h_n in a Fourier series in w_n,

$$h_n = u_n + v_n$$

$$u_n = \sum_{m=-\infty}^{\infty} c_m w_{n+m}$$

$$c_m \triangleq \langle w_{n+m}, h_n \rangle = \langle w_m, h_0 \rangle$$

$$\langle v_n, w_m \rangle = 0, \quad -\infty < n < \infty, \quad -\infty < m < \infty, \tag{27}$$

where v_n is the remainder. Equation (27) can be simplified by observing that

$$\langle w_m, h_0 \rangle = 0, \qquad m < 0,$$

since $h_0 \in M(h_k,\, k \geqq 0)$ and w_m is orthogonal to $M(h_k,\, k \geqq m+1)$, which contains $M(h_k,\, k \geqq 0)$ when $m < 0$. In addition, it can be shown (Ref. 3, pp. 571–575) that $v_n = 0$, since the spectrum under consideration here is absolutely continuous.† Thus, (27) reduces to

$$h_n = \sum_{m=0}^{\infty} c_m w_{n+m}$$

$$c_m = \langle h_0, w_m \rangle. \tag{28}$$

The expansion of (28), which is used in the theory of linear prediction,[2,3] is similar in spirit to a straightforward Gram-Schmidt orthogonalization process, but is much more useful in that the coefficients of the expansion are independent of n. The main shortcoming of the expansion (28) is requirement (25).

The formula for c_m given in (27) is not very useful in explicitly evaluating the coefficients of (28). A more useful method of evaluation is to observe that it is a spectral factorization problem. Defining the bilateral z-transform‡ of the autocorrelation,

$$R^*(z) = \sum_{m=-\infty}^{\infty} R_m Z^m, \tag{29}$$

we claim that

$$R^*(z) = \sum_{n=0}^{\infty} c_n Z^n \sum_{n=0}^{\infty} c_n Z^{-n}, \tag{30}$$

† This is by virtue of the fact that integral (21) is in terms of $R(\omega)d\omega$; i.e., the underlying measure is presumed to be absolutely continuous with respect to Lebesque measure.

‡ Note that we define the z-transform in positive powers of z.

where the c_m are given by (28). To show (30), first calculate R_j from (15),

$$\begin{aligned} R_j &= \langle h_0, h_j \rangle \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} c_m c_n \langle w_m, w_{j+n} \rangle \\ &= \begin{cases} \sum_{n=0}^{\infty} c_n c_{n+j}, & j \geqq 0 \\ \sum_{n=-j}^{\infty} c_n c_{n+j}, & j < 0. \end{cases} \end{aligned} \tag{31}$$

Similarly, the right side of (30) can be manipulated,

$$\sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_n c_m Z^{n-m} = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} c_n c_{n+m} Z^m + \sum_{m=1}^{\infty} \sum_{n=-m}^{\infty} c_n c_{n-m} Z^{-m}, \tag{32}$$

and comparing (31) and (32), (30) is established. The representation of (30) is not unique. However, Doob (Ref. 3, p. 160) shows that the coefficients of (27) uniquely satisfy (30) when the additional conditions

$$\sum_{n=0}^{\infty} c_n Z^n \neq 0, \qquad |Z| < 1, \tag{33}$$

$$\sum_{n=0}^{\infty} c_n^2 < \infty \tag{34}$$

are required.† The necessity of (34) is obvious from (27), while the reason why (33) is needed is that otherwise (30) could be satisfied on the unit circle by another sequence with a larger zeroth term, contradicting the fact that

$$c_0 = \|e_0^+\|. \tag{35}$$

Equation (35) follows from the observation that $M(h_k, k \geqq n) = M(w_k, k \geqq n)$ and therefore $P[h_n, M(h_k, k \geqq n+1)] = \sum_{m=1}^{\infty} c_m w_{n+m}$ or

$$e_n^+ = c_0 w_n. \tag{36}$$

A simple example will serve to illustrate (30). Suppose $h(t)$ has an exponential autocorrelation with

$$R_k = A^{|k|}, \qquad 0 < A < 1. \tag{37}$$

† Of course, condition (25) is also required.

Direct calculation of (29) reveals that

$$R^*(z) = \frac{1 - A^2}{(1 - AZ)(1 - A/Z)} \tag{38}$$

which is in the form of (30) with

$$c_n = \sqrt{1 - A^2}\, A^n. \tag{39}$$

The validity of (39) can be demonstrated directly for this simple example by noting that

$$e_k^+ = h_k - Ah_{k+1} \tag{40}$$

(as can be verified by showing that e_k^+ is orthogonal to h_m, $m \geqq k + 1$) and thus

$$\begin{aligned} w_n &= \frac{h_m - Ah_{m+1}}{\|h_m - Ah_{m+1}\|} \\ &= \frac{h_m - Ah_{m+1}}{\sqrt{1 - A^2}}. \end{aligned} \tag{41}$$

From (28),

$$\begin{aligned} c_m &= \langle h_0, w_n \rangle \\ &= \sqrt{1 - A^2}\, A^m \end{aligned} \tag{42}$$

agreeing with (39).

The procedure for higher-order rational spectra is equally simple. From (29) and the fact that R_m is real and even ($R_{-m} = R_m$), it follows that

$$R^*(z) = R^*\left(\frac{1}{z}\right). \tag{43}$$

Thus, for every zero a_i and pole b_i of $R^*(z)$, a_i^{-1} and b_i^{-1} are also a zero and a pole respectively. Thus, $R^*(z)$ can be written in the form

$$R^*(z) = K \frac{\prod_{i=1}^{m} (1 - a_i z)\left(1 - \frac{a_i}{z}\right)}{\prod_{i=1}^{n} (1 - b_i z)\left(1 - \frac{b_i}{z}\right)} \tag{44}$$

$$|a_i|, \qquad |b_i| < 1$$

so that from (30)

$$C(z) = \sum_{n=0}^{\infty} c_n Z^n = \sqrt{K} \frac{\sum_{i=1}^{m} (1 - a_i z)}{\sum_{i=1}^{n} (1 - b_i z)}, \tag{45}$$

where (33) has been insured by the choice of zeros in (45).

When $R^*(z)$ is not rational, a more general method of determining the coefficients of (28) is required. For this purpose, we use the equivalent power spectrum of (20). The first form in (20) is the one required for analytically determining $C(z)$, whereas the second form is the one which would usually be used in numerical calculations. The relationship of $R(\omega)$ to $R^*(z)$ is, of course,

$$R(\omega) = TR^*(e^{j\omega T}), \tag{46}$$

the evaluation of $R^*(z)$ on the unit circle. The equivalent of (30) for $R(\omega)$ is

$$\frac{R(\omega)}{T} = \left| \sum_{k=0}^{\infty} c_k e^{j\omega kT} \right|^2. \tag{47}$$

Intuitively, (47) requires the expansion of $\sqrt{R(\omega)/T}$, with an arbitrary phase characteristic, in a complex Fourier series with only positive frequencies. Following Doob (Ref. 3, p. 161), expand $\log \sqrt{R(\omega)/T}$ in a Fourier series,

$$\frac{1}{2} \log \frac{R(\omega)}{T} = \sum_k r_k e^{j\omega kT}. \tag{48}$$

This is always possible because, as will be demonstrated later, in order for (25) to be satisfied, it is necessary and sufficient that $\log R(\omega)$ be integrable. Define

$$g(z) = r_0 + 2 \sum_{k=1}^{\infty} r_k z^k \tag{49}$$

and note that

$$\operatorname{Re} g(e^{j\omega T}) = \frac{1}{2} \log \frac{R(\omega)}{T}. \tag{50}$$

We claim that

$$C(z) = e^{g(z)} \tag{51}$$

satisfies (47), since

$$|C(e^{j\omega T})| = \exp[\operatorname{Re} g\,(e^{j\omega T})] = \sqrt{\frac{R(\omega)}{T}}.$$

Equation (33) is also satisfied since $g(z)$ is analytic for $|z| < 1$.

Equation (51) is an analytic solution to the problem initially posed, but a practical means of applying it numerically is required. It is shown in Appendix A that the Fourier coefficients of (48) can be calculated efficiently and accurately using the fast Fourier transform (FFT) algorithm. The second difficulty is in determining $C(z)$ from

$g(z)$ in (51). This is easily resolved by noting that

$$\begin{aligned} c_m &= \frac{1}{m!}\frac{d^m}{dz^m}C(z)\bigg|_{z=0} \qquad m \geqq 0 \\ r_m &= \frac{1}{2m!}\frac{d^m}{dz^m}g(z)\bigg|_{z=0} \qquad m \geqq 1 \\ c_0 &= e^{r_0} \end{aligned} \tag{52}$$

and applying Leibniz's differentiation rule

$$\frac{d^n}{dz^n}uv = \sum_{m=0}^{n}\binom{n}{m}\frac{d^{n-m}u}{dz^{n-m}}\frac{d^m v}{dz^m}$$

to the product

$$\begin{aligned} \frac{d^n}{dz^n}C(z) &= \frac{d^{n-1}}{dz^{n-1}}\left(e^{g(z)}\frac{dg(z)}{dz}\right) \\ &= \sum_{m=0}^{n-1}\binom{n-1}{m}\frac{d^{n-m}g(z)}{dz^{n-m}}\frac{d^m C(z)}{dz^m} \end{aligned}$$

and, setting $z = 0$,

$$c_n = \frac{2}{n}\sum_{m=0}^{n-1}(n-m)r_{n-m}c_m\,, \quad n \geqq 1. \tag{53}$$

Equations (52)–(53) give us a practical recursive method of determining the coefficients of (28) when the channel spectrum is not rational.

III. GEOMETRIC THEORY OF THE ZERO-FORCING AND DECISION-FEEDBACK EQUALIZERS

The zero-forcing equalizer (ZFE) and decision-feedback equalizer (DFE) have been introduced in Sections 1.1 and 2.3. In this section, we will describe fully the characteristics of these equalizers in the context of the geometric structure developed in Section II.

3.1 *Conditions for the Existence of the ZFE and DFE*

The existence of a ZFE and DFE will now be related to the interpolation and prediction of the equivalent random process defined in Section 2.2. This relationship will then be used to obtain directly the known conditions for their existence.

The first observation is that the subspaces $M(h_k,\ k \neq 0)$ and $M(X_k,\ k \neq 0)$ are identical, as are the subspaces $M(h_k,\ k > 0)$ and $M(X_k,\ k > 0)$. The element

$$e_0 = h_0 - P[h_0,\ M(h_k,\ k \neq 0)] \tag{54}$$

is the same as the interpolation error vector defined in Section 2.2, $(X_0 - \hat{X}_0)$, while the prediction error vector is the same as

$$e_0^+ = h_0 - P[h_0, M(h_k, k > 0)]. \tag{55}$$

These two vectors are likely candidates for a ZFE and a DFE because they are orthogonal to the subspaces $M(h_k, k \neq 0)$ and $M(h_k, k > 0)$ respectively [see Section 2.1 and eq. (3)]. Hence, they satisfy (11) and (13) respectively. To verify that they are indeed a ZFE and a DFE, conditions (12) and (14) must be checked. Noting that e_0 is orthogonal to $M(h_k, k \neq 0)$, we have

$$\begin{aligned} \langle e_0, h_0 \rangle &= \langle e_0, h_0 - P[h_0, M(h_k, k \neq 0)] \rangle \\ &= \|e_0\|^2 \end{aligned} \tag{56}$$

by definition (54). Similarly, it follows that

$$\langle e_0^+, h_0 \rangle = \|e_0^+\|^2. \tag{57}$$

Thus, we see that a necessary and sufficient condition for e_0 (e_0^+) to be a ZFE (DFE) is that $\|e_0\| > 0$ ($\|e_0^+\| > 0$). By definition, the projection of h_0 on a subspace is the element of that subspace which is at a minimum distance from h_0, and hence $\|e_0\|$ and $\|e_0^+\|$ are the minimum distances between h_0 and $M(h_k, k \neq 0)$ and $M(h_k, k > 0)$ respectively. Since $\|e_0\|$ can only vanish if $h_0 \in M(h_k, k \neq 0)$, and similarly for $\|e_0^+\|$, it follows that e_0 (e_0^+) is a ZFE (DFE) if and only if $h_0 \notin M(h_k, k \neq 0)$ [$h_0 \notin M(h_k, k > 0)$]. Physically, these conditions mean that $h(t)$ must not be representable as an infinite weighted sum of a subset of its own translates. Geometrically, it is evident in Fig. 5 that, as long as $\|e_0\| > 0$ (or $\|e_0^+\| > 0$), e_0 (or e_0^+) will have a component in the direction of h_0 and the equalizer will have a response to the desired signal.

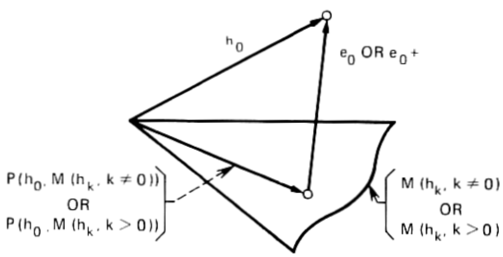

Fig. 5—Geometric interpretation of the zero-forcing equalizer and decision-feedback equalizer.

The weighting functions (54)–(55) can, under reasonable conditions,† be written in the form of a convergent linear sum of translates of h_0,

$$e_0 = h_0 - \sum_{k \neq 0} a_k h_k \tag{58}$$

$$e_0^+ = h_0 - \sum_{k>0} a_k^+ h_k \tag{59}$$

for some coefficients a_k^+. This demonstrates that these two elements are just the matched filter followed by transversal filter (MFTF) discussed in Section 1.1. It will be shown in the next section that the MFTF has particular significance, in that it maximizes the S/N ratio.

In general, there will be many ZFE's and DFE's other than (58)–(59). An example of a different ZFE is the element

$$h_0' - P[h_0', M(h_k, k \geqq 0)]$$

for any h_0' such that

$$\langle h_0', h_0 \rangle \neq 0$$
$$h_0' \notin M(h_k, k \neq 0).$$

An interesting question that arises is, then, whether there ever exists a ZFE and DFE when their corresponding MFTF's do not exist. To see that the answer is no for the ZFE (the proof for the DFE is identical), note that if $h_0 \in M(h_k, k \neq 0)$, then any g_0 orthogonal to $M(h_k, k \neq 0)$ is also necessarily orthogonal to h_0.‡ Thus, we have proven the following theorem:

Theorem 1: The following five statements are equivalent:

1. $h_0 \notin M(h_k, k \neq 0)$ $[h_0 \notin M(h_k, k > 0)]$.
2. $\|e_0\| > 0$ $[\|e_0^+\| > 0]$.
3. *There exists a ZFE [DFE].*
4. *There exists a ZFE [DFE] of the form of eq. (54) [eq. (55)], the MFTF.*
5. *The random process defined in (16) cannot be linearly interpolated [predicted] with vanishing mean-square error.*

The fifth condition of Theorem 1 follows from our earlier identification of e_0 and e_0^+ as the interpolation and prediction errors, respectively, of the equivalent random process. This observation also enables us to pull from the literature formulas for the norms of e_0 and e_0^+. The follow-

† This will be discussed fully in Section 3.5.
‡ We also make use of the trivial observation that any g_0 satisfying (11) is orthogonal to $M(h_k, k \neq 0)$.

ing corollary follows directly from the known formulas for the interpolation and prediction errors of a random process,[2,3]

$$\|e_0\|^2 = \left[\frac{T^2}{2\pi}\int_{-\pi/T}^{\pi/T} R^{-1}(\omega)d\omega\right]^{-1} \tag{60}$$

$$\|e_0^+\|^2 = \frac{1}{T}\exp\left[\frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} \log R(\omega)d\omega\right]. \tag{61}$$

Corollary 1: A ZFE [DFE] exists if and only if $R^{-1}(\omega)$ [$\log R(\omega)$] is integrable.

Both conditions relate to the fashion in which $R(\omega)$ vanishes. In particular, both require that $R(\omega)$ vanish on at most a set of measure zero. The relationship of (60) and (61) will be discussed more fully in the sequel.

It should be noted also that (61) follows directly from the orthogonal expansion of Section 2.5. From (35) we know that $\|e_0^+\|^2$ equals c_0^2, while (52) gives a relation for c_0. When the Fourier series of (48) is inverted and r_0 is substituted into (52), (61) results.

3.2 *Performance of the Equalizers*

It will now be shown that the MFTF among all ZFE's and DFE's maximizes the S/N ratio and minimizes the error probability in white Gaussian noise. The derivation will be a simple application of the Schwarz inequality.

Assume that the additive noise in (1) is white and Gaussian. Then the decision axis which is applied to a threshold is, for the ZFE,

$$\langle g_0, r\rangle = B_0\langle g_0, h_0\rangle + \langle g_0, n\rangle, \tag{62}$$

where $\langle g_0, n\rangle = n_0$ is a Gaussian random variable with mean zero and variance

$$En_0^2 = \frac{N_0}{2}\|g_0\|^2 \tag{63}$$

and $N_0/2$ is the two-sided spectral density of the noise. The minimum probability of error decision strategy is then to apply $\langle g_0, r\rangle$ to a series of $M - 1$ thresholds, with the specific thresholds depending on the probability law on B_k. For any such law and series of thresholds the probability of error will be a monotone decreasing function of the S/N ratio, which is proportional to

$$\text{S/N} \propto \frac{\langle g_0, h_0\rangle^2}{\|g_0\|^2}, \tag{64}$$

since $\langle g_0, n\rangle$ is a zero-mean Gaussian random variable with variance proportional to $\|g_0\|^2$. Noting from (11) that g_0 is orthogonal to $P[h_0, M(h_k, k \neq 0)]$ whenever g_0 is a ZFE, (64) can be rewritten

$$\mathrm{S/N} \propto \frac{\langle g_0, e_0\rangle^2}{\|g_0\|^2} \leqq \|e_0\|^2 \tag{65}$$

by the Schwarz inequality, with equality if and only if g_0 equals e_0 (the MFTF) within a multiplicative constant. Thus, the MFTF, among all ZFE's, maximizes the S/N ratio. By the same method an identical result can be demonstrated for the DFE, if it is assumed that the decision-feedback mechanism correctly cancels the tails of earlier pulses.

The preceding derivation, which is a generalization of the Schwarz inequality derivation of the matched filter, has the geometric interpretation of Fig. 6. In writing (65), the maximization of (64) is restricted to those g_0 which lie in the hyperplane orthogonal to $P[h_0, M(h_k, k \neq 0)]$. Since every ZFE is also orthogonal to this vector, it follows that the hyperplane so described contains the set of all ZFE's. However, the maximization over elements of the hyperplane does not guarantee a result which is a ZFE. The vector in the hyperplane which has the greatest component in the direction of h_0 per unit length is evidently the one which lines up with e_0, as verified by (65). Fortunately, this vector also turns out to be a ZFE, so that the maximization is complete.

An additional observation relative to (65) is that the maximum S/N ratio is proportional to $\|e_0\|^2$ for the ZFE and $\|e_0^+\|^2$ for the DFE. The maximum S/N ratio is therefore directly proportional to the mean-square interpolation and prediction errors of the equivalent random process. Thus, the maximum S/N ratios of the ZFE and DFE are given by (60) and (61) respectively, while the factor by which the

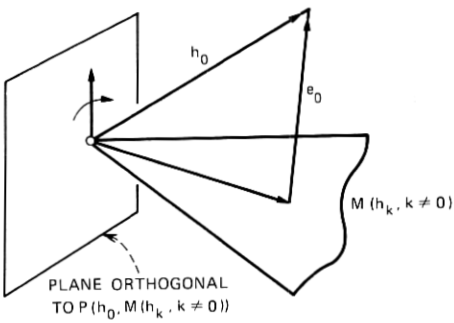

Fig. 6—S/N ratio maximized by the MFTF.

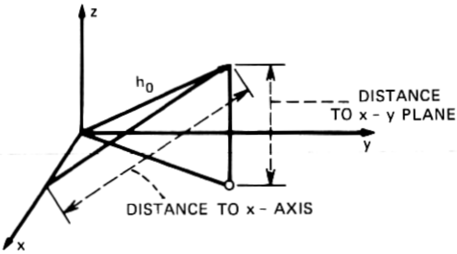

Fig. 7—Geometric interpretation of eq. (66).

S/N ratio is reduced relative to an isolated pulse with matched filter detection is obtained by dividing by R_0, the isolated pulse energy.

Price[6] derived (61) by a different method and used the geometric mean inequality for integrals to show from (60) and (61) that

$$\|e_0\|^2 \leqq \|e_0^+\|^2. \tag{66}$$

This important result implies that (*i*) the S/N ratio of the DFE MFTF always exceeds that of the ZFE MFTF,† and (*ii*) a DFE exists whenever a ZFE exists [the contrary is not true, as demonstrated by the important example of algebraic zeros in $R(\omega)$[6]]. Using the geometric method we have developed, two interpretations of (66) can be given. First, it is intuitively apparent that the mean-square interpolation error of a random process will be smaller than the mean-square prediction error, because an interpolation is based on more information; similarly, there will be some processes for which interpolation, but not prediction, with zero mean-square error is possible. Second, since $M(h_k,\ k \neq 0)$ contains $M(h_k,\ k > 0)$, the distance between h_0 and $M(h_k,\ k \neq 0)$ (equal to $\|e_0\|^2$) must be smaller than the distance between h_0 and $M(h_k,\ k > 0)$ (equal to $\|e_0^+\|^2$). This second interpretation is a rigorous way of establishing (66) by a method more direct than the integral inequality. It has the geometric interpretation of Fig. 7, where the distance between a vector h_0 and the larger subspace (the x-y plane) is less than between h_0 and the subspace it contains (the x axis).

The performance of the ZFE and DFE can be evaluated for any particular channel spectrum using (60)–(61). In particular, (60)–(61) can be evaluated in closed form for rational spectra. A different approach, which allows us to evaluate the tap-gains of the equalizers as well, will be pursued in Section 3.4.

† This result neglects the effect of decision errors on the DFE.

3.3 *On the DFE White Output Noise Property*

As observed by Price,[6] the DFE forward filter is identical to the "whitened matched filter" employed by Forney[12] as the first element of his maximum likelihood detector. The property of this filter which is essential to Forney's application is that the noise sequence at the filter output is uncorrelated. As with the other properties of this filter, this one has a simple explanation in terms of the relationship to linear prediction.

Identifying e_k^+ as $e_0^+(t - kT)$, the noise sequence at the DFE forward filter output is $\langle e_k^+, n \rangle$. Since $n(t)$ is white noise, this sequence will be uncorrelated if and only if

$$\langle e_m^+, e_n^+ \rangle = 0, \qquad m \neq n. \tag{67}$$

The validity of (67) and an interpretation of this result in terms of the uncorrelated nature of the successive prediction errors of a random process has already been given in Section 2.5.

3.4 *Determination of Tap-Gains*

In this section, we will use the orthogonal expansion of Section 2.5 to derive methods of determining the tap-gains of the forward and feedback filters of the MFTF DFE. For comparison purposes the well-known relation for the tap-gains of the ZFE will also be briefly developed.

If we write the weighting response of the MFTF ZFE as

$$a_0 e_0 = \sum_{k=-\infty}^{\infty} a_k h_k, \tag{68}$$

where the tap-gains of the transversal filter are a_k, $-\infty < k < \infty$, condition (11)–(12) becomes

$$\begin{aligned} \langle e_0, h_m \rangle &= \|e_0\|^2 \delta_{m,0} \\ &= \frac{1}{a_0} \sum_k a_k R_{m-k}. \end{aligned} \tag{69}$$

Taking the bilateral z-transform of (69),

$$a_0 \|e_0\|^2 = A(z) R^*(z), \tag{70}$$

where $A(z)$ is the z-transform of the tap-gains

$$A(z) \triangleq \sum_k a_k z^k. \tag{71}$$

Thus, from (70),

$$A(z) = \frac{a_0 \|r_0\|^2}{R^*(z)}. \tag{72}$$

This filter is illustrated in Fig. 8a. When $h(t)$ is applied to the input of a matched filter and the output sampled at a rate of $1/T$, the output has z-transform $R^*(z)$. The transversal filter weighting response has a z-transform proportional to $R^*(z)^{-1}$, so that the output is consistent with (69).

The S/N ratio of the ZFE, given by (60), is readily derived from (72). Writing the relation for tap-gain zero,

$$a_0 = \frac{1}{2\pi j} \oint \frac{A(z)}{z} dz = \frac{a_0 \|e_0\|^2}{2\pi j} \oint \frac{dz}{zR^*(z)}, \tag{73}$$

and solving for $\|e_0\|^2$, we immediately get (60) using (46).

As an example, for the exponential autocorrelation of (37), (72) becomes

$$A(z) = a_0 \|e_0\|^2 \left[-\frac{A}{1 - A^2} z^{-1} + \frac{1 + A^2}{1 - A^2} - \frac{A}{1 - A^2} z \right] \tag{74}$$

from which we get

$$\begin{aligned} \|e_0\|^2 &= \frac{1 - A^2}{1 + A^2} \\ a_{-1} = a_1 &= -\frac{a_0 A}{1 + A^2} \\ a_k &= 0, \qquad |k| > 1, \end{aligned} \tag{75}$$

a result derived by Tufts[13] by another method. This example points out that it is not ever necessary to actually evaluate (60) when the channel spectrum is rational, but rather the performance can be obtained by equating the zero-order tap-gains of (72) in the manner of (73).

The situation with the DFE is only slightly more complicated. In this case the DFE filter is

$$a_0^+ e_0^+ = \sum_{k=0}^{\infty} a_k h_k, \tag{76}$$

where only taps on one side are involved. Substituting from (28) and (36),

$$\begin{aligned} a_0^+ c_0 w_0 &= \sum_{k=0}^{\infty} a_k^+ \sum_{m=0}^{\infty} c_m w_{k+m} \\ &= \sum_{m=0}^{\infty} w_m \sum_{k=0}^{m} a_k^+ c_{m-k}, \end{aligned} \tag{77}$$

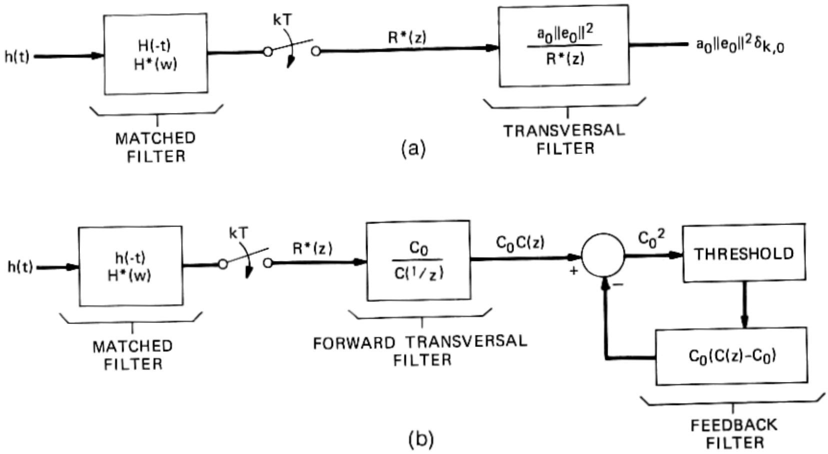

Fig. 8—Spectral representations of the MFTF zero-forcing equalizer (a) and decision-feedback equalizer (b).

and equating coefficients,

$$\sum_{k=0}^{m} a_k^+ c_{m-k} = \begin{cases} a_0^+ c_0, & m = 0 \\ 0, & m > 0. \end{cases} \tag{78}$$

From (78) we get a recursion relation for the tap coefficients which is useful for nonrational spectra,

$$a_m^+ = -\frac{1}{c_0} \sum_{k=0}^{m-1} a_k^+ c_{m-k}\,, \tag{79}$$

and a z-transform relation which is useful for rational spectra,

$$A^+(z) = \frac{a_0^+ c_0}{C(z)}\,, \tag{80}$$

where $A^+(z)$ is the z-transform of the tap-gains of (76). Performing (80) again for the autocorrelation of (37),

$$\begin{aligned} A^+(z) &= a_0^+(1 - Az) \\ \|e_0^+\|^2 &= c_0^2 = 1 - A^2 \end{aligned} \tag{81}$$

which is consistent with (40) and is larger than $\|e_0\|^2$ by a factor of $(1 + A^2)$. As with the ZFE, the performance of the DFE can be determined for rational spectra without the explicit evaluation of (61).

The comparison of (80) with (72) is interesting, in that they are identical except for the fact that in (80) $C(z)$ is substituted for $R^*(z)$

in (72). The annulus of convergence of $A(z)$ will always include the unit circle, since $R^*(z)$ converges in an annulus containing the unit circle. Similarly, $C(z)$ is analytic and nonzero in a region containing the unit disk, and hence $A^+(z)$ will have only positive powers of z and converge in a region containing the unit disk. Note that these properties of $A^+(z)$ are critically dependent on (33) being satisfied.

The spectral factorization method of determining the tap-gains of the DFE was given by Monsen[5] for rational spectra. Price[6] gave a formula valid for arbitrary spectra, but it is difficult to evaluate numerically. Since (79) is valid for arbitrary spectra, the method presented here represents a synthesis of the appeal and computational simplicity of the spectra factorization method with the generality of Price's Toeplitz form result.

We also need the tap-gains of the feedback filter for the DFE. From Fig. 4, the required feedback tap-gains are given by $\langle e_0^+, h_{-n}\rangle$, $1 \leqq n < \infty$. From (36) and (28),

$$b_n = \langle e_0^+, h_{-n}\rangle = c_0 \sum_{m=0}^{\infty} c_m \langle w_0, w_{m-n}\rangle$$
$$= c_0\, c_n. \tag{82}$$

Thus, the frequency response of the feedback filter is given by

$$\sum_{m=1}^{\infty} b_m z^m = c_0[C(z) - c_0]. \tag{83}$$

The z-transform representation of the DFE just derived is illustrated in Fig. 8b. When an isolated pulse $h(t)$ is applied to the matched filter, the sampled output has z-transform $R^*(z)$. The transversal filter multiplies by $A^+(1/z) = c_0/C(1/z)$, as can be verified from (76).† The z-transform of the forward transversal filter output is $c_0C(z)$ because of (30), which verifies the causal response which is characteristic of the DFE. The output of the feedback filter of (83) is then subtracted, to yield (hopefully) a delta function response c_0^2. The reader can verify that when the threshold is replaced by a gain of $1/c_0^2$ (the noise-free case) the response is as represented.

3.5 *Finite Transversal Filter Equalizers*

The previous sections have considered the rather idealized case of infinite transversal filter equalizers. Since only finite equalizers can

† This is because (76) is not in the form of a convolution sum. This distinction was not relevant to the ZFE due to the symmetry of that filter.

actually be implemented, the important question arises as to when and in what sense the infinite equalizer can be approximated by a finite one.

We have already seen in the example of the exponential autocorrelation that the infinite equalizer can degenerate into a finite transversal filter for some channel spectra. This will happen whenever $A(z)$ and $A^+(z)$ are finite polynomials in z. From (72) and (80) we see that this will occur whenever $R^*(z)$ is a rational function which has no zeros (only poles). When the spectrum is not rational, or is rational with zeros, it will be necessary to approximate the infinite MFTF.

It is straightforward to generalize the results of Sections 3.1 and 3.2 to subspaces spanned by a finite number of translates of h_0. In particular, if we replace the criteria of (11) and (13) by

$$\langle h_k, g_0 \rangle = 0 \qquad -N \leqq k \leqq N, \qquad k \neq 0 \tag{84}$$

for the ZFE and

$$\langle h_k, g_0 \rangle = 0 \qquad 1 \leqq k \leqq N \tag{85}$$

for the DFE, we are left with the consideration of the finite dimensional subspaces $M(h_k, -N \leqq k \leqq N, k \neq 0)$ and $M(h_k, 1 \leqq k \leqq N)$, which we will write as M_N and M_N^+ respectively. Then the MFTF equalizers which satisfy (84) and (85) are similar to (54) and (55),

$$e_0(N) \triangleq h_0 - P(h_0, M_N) \tag{86}$$

$$e_0^+(N) \triangleq h_0 - P(h_0, M_N^+). \tag{87}$$

It is straightforward to see that Theorem 1 can be replaced by the following version:

Theorem 2: The following four statements are equivalent:

1. $h_0 \notin M_N$ $[h_0 \notin M_N^+]$.
2. $\|e_0(N)\| > 0$ $[\|e_0^+(N)\| > 0]$.
3. *There exists a ZFE [DFE] in the restricted sense of (84) [(85)].*
4. *There exists an MFTF ZFE [DFE] in this restricted sense.*

The question of when it can be asserted that $\|e_0(N)\| > 0$ and $\|e_0^+(N)\| > 0$ deserves consideration. The condition that $h_0 \in M_N^+$ requires that coefficients $\{\alpha_m, 1 \leqq m \leqq N\}$ exist which satisfy

$$h_0 = \sum_{m=1}^{N} \alpha_m h_m. \tag{88}$$

This occurrence will be precluded if the set $\{h_m, -\infty < m < \infty\}$ is linearly independent. Similarly, linear independence is sufficient for a ZFE to exist in the sense of (84). The following lemma, which is proven in Appendix B, establishes sufficient conditions for the linear independence of $\{h_m, -\infty < m < \infty\}$:

Lemma 1: The following two conditions are sufficient for the linear independence of $\{h_m, -\infty < m < \infty\}$:

1. $\|e_0\| > 0$ *or* $\|e_0^+\| > 0$.
2. *There exists an interval* $[a, b]$, $a < b$, *such that* $R(\omega) > 0$, $\omega \in [a, b]$.

The first condition of Lemma 1 satisfies our intuition that if an infinite MFTF ZFE or DFE exists then the finite MFTF version should also exist. The second condition assures us that the finite equalizers also exist under much weaker conditions.

The following theorem establishes a relationship between the finite and infinite equalizers, and is proven in Appendix B:

Theorem 3: As $N \rightarrow \infty$, $\|e_0(N)\|^2$ *is monotonically decreasing and approaches* $\|e_0\|^2$, *and likewise for* $e_0^+(N)$. *Furthermore,* $\|e_0(N) - e_0\|^2 \rightarrow 0$ *and* $\|e_0^+(N) - e_0^+\|^2 \rightarrow 0$.

The primary conclusion of Theorem 3 is that the infinite equalizer can be approximated with arbitrary accuracy (in the sense of L_2 convergence) by a finite equalizer. In addition, it asserts that the S/N ratio of this finite equalizer is greater than that of the infinite equalizer; however, this desirable property may be entirely or partially offset by any residual intersymbol interference.

Each member of the sequence of equalizers guaranteed by Theorem 3 has different tap-gains, because the projection on a different subspace is being taken with each N. A more aesthetically pleasing approximation results when (58) and (59) are valid, for then

$$\left\| h_0 - \sum_{k=-N}^{N} a_k h_k - e_0 \right\| \rightarrow 0, \tag{89}$$

$$\left\| h_0 - \sum_{k=1}^{N} a_k^+ h_k - e_0^+ \right\| \rightarrow 0, \tag{90}$$

by the definition of convergence of the infinite sums in (58)–(59). Each succeeding equalizer defined by (89)–(90) is obtained by adding an additional tap, without changing the other tap-gains. As observed by Doob (Ref. 3, p. 564), a convergent sum of the form of (58)–(59) does not always exist; the following theorem gives sufficient conditions for the validity of (58)–(59) which are generally satisfied in practical problems:

Theorem 4†*: If there exist constants K_1 and K_2, $0 < K_1 \leqq K_2$, such that $K_1 \leqq R(\omega) \leqq K_2$, $|\omega| < \pi/T$, then convergent expansions of e_0 and e_0^+ of the form of (58)–(59) exist. Furthermore, the coefficients of the expansions are unique.*

This theorem is proven in Appendix B. The question of uniqueness of the tap-gains of the DFE is one which was not answered by Price.[6]

Finally, the white output noise property of the MFTF DFE also extends to a finite MFTF DFE in the following sense: If the reception of (1) extends from N_1 to N_2, where N_2 (but not necessarily N_1) is finite, then the DFE defined by

$$e_k^+ = h_k - P[h_k, M(h_m, k+1 \leqq m \leqq N_2)]$$

will have white output noise samples. This fact is easily verified from the same containment of subspaces that was used in the proof for the infinite case.

IV. EXTENSION TO NONSTATIONARY NOISE AND CHANNEL

The previous sections have considered only the case where the additive noise is white. The extension to colored Gaussian noise can be handled in a straightforward fashion with the addition of a whitening filter. In this section we will generalize the ZFE and DFE to the case of arbitrary nonstationary second-order Gaussian noise (which includes colored Gaussian noise as a special case) using the techniques of reproducing kernel Hilbert space (RKHS).[11] Although the cases for which the corresponding RKHS can be characterized explicitly correspond generally to those cases which can be handled by other techniques, the RKHS approach does allow us to treat all cases simultaneously and concisely. In addition, it enables us to generalize simultaneously to an arbitrary nonstationary channel (to be precise, a channel which is changing in time in a deterministic and known fashion) with no additional complications. Perhaps the most interesting outcome of this effort will be the observation that the DFE white output noise property (discussed in Section 3.3) remains valid in this general case. The result is an interesting generalization of Forney's whitened matched filter.[12]

---

† Theorem 4 remains valid under the weaker hypothesis that $0 < \text{ess inf } R(\omega)$ and $\text{ess sup } R(\omega) < \infty$.

To this end, modify (1) to

$$r(t) = \sum_{m=N_1}^{N_2} B_m h_m(t) + n(t), \tag{91}$$

where, as before, N_1 and N_2 can be infinite. The noise will be assumed to be Gaussian with arbitrary autocorrelation

$$K(t, s) = E[n(t)n(s)]. \tag{92}$$

The subscript m on $h_m(t)$ indicates that the received pulses need not be translates of the same elementary waveform. The reception will be termed *channel stationary* when

$$h_m(t) = h(t - mT)$$

and *noise stationary* when

$$K(t, s) = K(t - s).$$

We denote by $L_2(n)$ the subspace of the Hilbert space of square integrable random variables spanned by $n(t)$, $-\infty < t < \infty$. This subspace is entirely analogous to $M(X_k, -\infty < k < \infty)$ defined earlier, except that the underlying parameter t is continuous. The following lemma is applicable:[11]

Lemma 2: Let $H(K)$ consist of all functions $g(\cdot)$ of the form

$$g(\cdot) = E[n(\cdot)U] \tag{93}$$

for some $U \in L_2(n)$. Then $H(K)$ is a Hilbert space with inner product

$$\langle g, g \rangle_{H(K)} = E|U|^2. \tag{94}$$

The mapping ψ: $L_2(n) \rightarrow H(K)$ defined by (93) is a congruence which maps $n(t)$ into $K(\cdot, t)$.

The Hilbert space $H(K)$ defined by Lemma 2 is known as the reproducing kernel Hilbert space with reproducing kernel K. It is straightforward to show from (93) and (94) that $H(K)$ has the properties

$$K(\cdot, t) \in H(K), \qquad -\infty < t < \infty, \tag{95}$$

$$\langle g(\cdot), K(\cdot, t) \rangle_{H(K)} = g(t), \qquad g \in H(K). \tag{96}$$

It can be shown[11] that for any symmetric positive-definite kernel K there exists a unique Hilbert space satisfying (95)–(96).

The inverse of $g(\cdot)$ under ψ is usually given the suggestive notation

$$\langle g, n \rangle_{H(K)} \triangleq \psi^{-1}(g) \tag{97}$$

even though $n \notin H(K)$ with probability one and therefore (97) cannot be given an interpretation as an inner product.

It will be assumed that $h_m(t) \in H(K)$, since otherwise the detection problem is singular.† [11] In nonstationary noise the space $H(K)$ takes the place of L_2 in the earlier white noise problem. Accordingly, we restrict the class of filters under consideration to $H(K)$ inner products with elements of $H(K)$. Thus, a filter can be written in the form

$$\langle g, r\rangle_{H(K)} = \sum_{m=-N_1}^{N_2} B_m \langle g, h_m\rangle_{H(K)} + \langle g, n\rangle_{H(K)}, \tag{98}$$

where the noise term in (98) assumes the special meaning of (97). Analogously to (15), we define the pulse autocorrelation

$$R(m, n) = \langle h_m, h_n\rangle_{H(K)}. \tag{99}$$

When the reception is noise and channel stationary, $R(m, n)$ is a function of the difference of its arguments, as in (15). In general, however, it is an arbitrary symmetric positive definite function defined for $N_1 \leqq m, n \leqq N_2$.‡

In the white noise case, we saw that the subspace of L_2 spanned by translates of $h(t)$ was congruent to the subspace of second-order random variables spanned by a wide-sense stationary random process. In the nonstationary noise case, the subspace of $H(K)$ spanned by h_m, $N_1 \leqq m \leqq N_2$, is congruent to the subspace of the second-order random variables spanned by a possibly nonstationary second-order random process. In the white noise case the theory of minimum mean-square error estimation of a wide-sense stationary random process was relevant; in the present case the random process becomes nonstationary. As before, the ZFE and DFE have interpretations as interpolation and prediction errors of the corresponding random process with autocorrelation $R(m, n)$. However, rather than pursue these correspondences further (in view of our results for the white noise case they are obvious), we will directly pursue the theory of the ZFE and DFE for the detection of B_m, $N_1 \leqq m \leqq N_2$, from $r(t)$ in (91).

† A singular detection problem is one in which a decision can be made which is correct with probability one.

‡ The positive definite property follows from the inequality

$$0 \leqq \left\| \sum_{m=1}^{N} \alpha_m h_{k_m} \right\|_{H(K)}^2 = \sum_{m=1}^{N} \sum_{n=1}^{N} \alpha_m \alpha_n R(k_m, k_n).$$

The theory of Section 3.1 remains valid if the subspaces $M(h_m, m \in I)$ are considered as subspaces of $H(K)$ rather than L_2.† As before, the condition which is necessary and sufficient for the existence of a ZFE or DFE is that

$$h_k \notin M(h_m, m \in I).$$

The analogs of the MFTF versions of the DFE and ZFE are the elements given by (54) and (55), except that now we must work with e_k and e_k^+ instead of e_0 and e_0^+ (e_k is no longer necessarily simply a time translate of e_0, etc.). A derivation similar to that given in Section 3.3 establishes that e_k and e_k^+ maximize the S/N ratio as before. In particular, when the filter of (91) is restricted to be a ZFE, (91) becomes

$$\langle g, r\rangle_{H(K)} = B_k\langle g, h_k\rangle_{H(K)} + \langle g, n\rangle_{H(K)} \tag{100}$$

and the S/N ratio is proportional to

$$S/N \propto \frac{\langle g, h_k\rangle^2_{H(K)}}{\langle g, g\rangle_{H(K)}} \leqq \langle e_0, e_0\rangle_{H(K)} \tag{101}$$

since the variance of the noise term in (100) is, from (97),

$$\begin{aligned} E\,|\langle g, n\rangle_{H(K)}|^2 &\triangleq E\,|\psi^{-1}(g)|^2 \\ &= \langle g, g\rangle_{H(K)} \end{aligned}$$

through the congruence established in Lemma 2. Equation (101) demonstrates that the MFTF ZFE maximizes the S/N ratio, and the same result follows for the DFE by the same method.

A general equation can be given for the projection element required for the MFTF. This equation is entirely analogous to a result of Parzen[11] for stochastic estimation. To this end we require a lemma which is a restatement of Lemma 2:

Lemma 3: Let $H(R)$ consist of all functions $f(m)$, $m \in I$, of the form

$$f(m) = \langle h_m, F\rangle_{H(K)} m \in I \tag{102}$$

for some $F \in M(h_m, m \in I)$. Then $H(R)$ is the RKHS with reproducing kernel $R(m, n)$, $m,n \in I$, and has inner product

$$\langle f, f\rangle_{H(R)} = \langle F, F\rangle_{H(K)}. \tag{103}$$

The mapping ϕ: $M(h_m, m \in I) \rightarrow H(R)$ defined by (102) is a congruence which maps h_m into $R(\cdot, m)$.

---

† We use I as a set of indices to avoid repeating the equations twice. For the ZFE, $I = [N_1, k - 1]U[k + 1, N_2]$ and for the DFE $I = [k + 1, N_2]$. For the infinite case, $N_2 = - N_1 = \infty$. The digit B_k is being detected.

The reader might find it instructive to verify from (102)–(103) that the RKHS properties hold for $H(R)$,

$$R(\cdot, n) \in H(R), \tag{104}$$

$$\langle f(\cdot), R(\cdot, n)\rangle_{H(R)} = f(n), \tag{105}$$

where $f(\cdot) \in H(R)$.

The problem we want to attack is finding the projection P of some vector Q on $M(h_m, m \in I)$ (later we will let $Q = h_k$). From (3) we have

$$\langle Q - P, h_m\rangle_{H(K)} = 0, \qquad m \in I \tag{106}$$

or

$$\langle P, h_m\rangle_{H(K)} = \rho_Q(m), \qquad m \in I, \tag{107}$$

where

$$\rho_Q(m) \triangleq \langle Q, h_m\rangle_{H(K)}, \qquad m \in I. \tag{108}$$

In (107), $\rho_Q(m)$ is a known function and P is to be determined. Assuming for the moment that $\rho_Q \in H(R)$, from Lemma 3 we see that ρ_Q is the image of P under the congruence ϕ, and hence

$$P = \phi^{-1}(\rho_Q), \tag{109}$$

which is the solution we desire. Using the congruence properties of ϕ, the length of $Q - P$ is

$$\begin{aligned} \|Q - \phi^{-1}(\rho_Q)\|^2_{H(K)} &= \|Q\|^2_{H(K)} - 2\langle Q, \phi^{-1}(\rho_Q)\rangle_{H(K)} + \|\phi^{-1}(\rho_Q)\|^2_{H(K)} \\ &= \|Q\|^2_{H(K)} - \|\rho_Q\|^2_{H(R)}. \end{aligned} \tag{110}$$

Establishing that in fact $\rho_Q \in H(R)$ is straightforward. Note that

$$\begin{aligned} \rho_Q(m) &= \langle Q, h_m\rangle_{H(K)} \\ &= \langle Q - P, h_m\rangle_{H(K)} + \langle P, h_m\rangle_{H(K)} \\ &= \langle P, h_m\rangle_{H(K)}, \end{aligned} \tag{111}$$

which implies that $\rho_Q \in H(R)$ by Lemma 3 since $P \in M(h_m, m \in I)$.

Replacing Q by h_k in (109), we get the desired projection

$$P[h_k, M(h_m, m \in I)] = \phi^{-1}[R(k, \cdot)] \tag{112}$$

The ZFE and DFE are obtained by letting I equal the appropriate set. The S/N ratios of the receivers are proportional to, from (101) and (110),

$$\text{S/N} \propto \|h_k\|^2_{H(K)} - \|R(k, \cdot)\|^2_{H(R)}. \tag{113}$$

The RKHS approach has reduced the problem to that of finding

RKHS inner products. In some cases these inner products can be explicitly characterized, while in all others they can be determined by convergent iterative techniques.[11]

We can also quickly show that the DFE white output noise property discussed in Section 3.3 generalizes. From (98), the noise samples at the filter output are

$$\begin{aligned} n_k &= \langle e_k^+, n \rangle_{H(K)} \\ &= \psi^{-1}(e_k^+) \end{aligned} \tag{114}$$

by definition. From (114) and Lemma 2,

$$\begin{aligned} E(n_j n_k) &= E[\psi^{-1}(e_j^+)\psi^{-1}(e_k^+)] \\ &= \langle e_j^+, e_k^+ \rangle_{H(K)} \\ &= 0, \qquad j \neq k \end{aligned} \tag{115}$$

by the same reasoning as before.

Finally, it is instructive to demonstrate that this RKHS formulation reduces to the whitening filter approach when the reception is noise and channel stationary. Assume that

$$K(t, s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega(t-s)} N(\omega) d\omega\,, \tag{116}$$

where $N(\omega)$ is uniformly bounded and never vanishes. Under these conditions we claim that $H(K)$ consists of all integrable $g(t)$ with Fourier transforms $G(\omega)$ which satisfy

$$\|g\|_{H(K)}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |G(\omega)|^2 \frac{1}{N(\omega)}\, d\omega \cdot \tag{117}$$

To verify this, properties (95)–(96) must be checked. Equation (95) is valid since $N(\omega)$ is integrable, while (96) follows from

$$\begin{aligned} \langle g(\cdot), K(\cdot, t) \rangle_{H(K)} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega)[e^{-j\omega t} N(\omega)]^* \frac{1}{N(\omega)}\, d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) e^{j\omega t}\, d\omega \\ &= g(t), \end{aligned} \tag{118}$$

where (*) denotes complex conjugation. From (117), the $H(K)$ inner product consists of a filter with frequency response $N^{-1}(\omega)$ (which is the whitening filter) followed by an ordinary L_2 inner product, and is therefore consistent with the whitening filter formulation.

V. CONCLUSIONS

This paper has presented a unified and rather thorough treatment of the ZFE and DFE. In a companion paper,[7] the geometric model of intersymbol interference developed here will be used to study the minimum distance problem encountered in the performance analysis of the maximum likelihood detector[12] and in evaluating a lower bound on the performance of any receiver.[14] It is shown there that a canonical relationship exists between the minimum distance and the performance and tap-gains of the MFTF DFE.

No performance example comparing the DFE and ZFE on a channel of practical interest has been given in this paper in order that the maximum likelihood detector may enter into the comparison. In Ref. 7 the performance of three receivers is calculated for a channel whose loss in dB increases as the square-root of frequency. This channel is an excellent model of coaxial cable and some types of wire-pairs.

VI. ACKNOWLEDGMENTS

The author is indebted to R. Price for many valuable comments. In particular, it was he who suggested the extension to nonstationary noise using RKHS theory. The author also appreciates many valuable discussions with D. L. Duttweiler.

APPENDIX A

The purpose of this appendix is to derive an approximation to the Fourier coefficients of (48) in terms of discrete Fourier transform (DFT), which can be efficiently evaluated using the FFT algorithm.

Define a normalized function

$$F(\lambda) = \log \frac{R\left(\frac{2\pi}{T}\lambda\right)}{T} \tag{119}$$

so that

$$r_n = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-jn2\pi\lambda} F(\lambda) d\lambda \cdot \tag{120}$$

Approximating the integral by a summation,

$$\begin{aligned}\hat{r}_n &\cong \frac{1}{2N}\sum_{k=0}^{N-1} F\left(\lambda_0 + \frac{k}{N} - \frac{1}{2}\right) e^{-jn2\pi(\lambda_0 + k/N - \frac{1}{2})} \\ &= \frac{1}{2} e^{-jn2\pi(\lambda_0 - \frac{1}{2})} \frac{1}{N}\sum_{k=0}^{N-1} F\left(\lambda_0 + \frac{k}{N} - \frac{1}{2}\right) e^{-j2\pi(kn/N)},\end{aligned} \tag{121}$$

where the sum on the right is a discrete Fourier transform.

In order to determine the effect of this approximation, substitute

$$\tfrac{1}{2}F(\lambda) = \sum_k r_k e^{jk2\pi\lambda} \tag{122}$$

into the approximation equation (121) to yield

$$\begin{aligned}\hat{r}_n &= \sum_{m=-\infty}^{\infty} r_m \frac{1}{N} \sum_{k=0}^{N-1} e^{j(m-n)2\pi(\lambda_0+k/N-\frac{1}{2})} \\ &= r_n + \sum_{l\neq 0} e^{j2\pi lN\lambda_0}(-1)^l r_{n+lN}.\end{aligned} \tag{123}$$

Thus, the approximation of (121) yields the desired Fourier coefficient plus the sum of alias terms. N must be larger than the number of coefficients to be evaluated and large enough that the alias terms r_{n+lN} are small. In practice, $N \cong 5{,}000$ can be achieved with modest amounts of computer time using the FFT algorithm.

APPENDIX B

Proofs of Theorems

Proof of Lemma 1: Since $\|e_0^+\|^2 \geqq \|e_0\|^2$ it suffices to show that $\|e_0^+\| > 0$ implies that $\{h_m, -\infty < m < \infty\}$ is linearly independent set. To this end, assume that

$$\left\|\sum_{m=1}^{N} \alpha_m h_{k_m}\right\|^2 = 0, \qquad k_1 < k_2 < \cdots < k_N. \tag{124}$$

To show that $\alpha_1 = 0$, assume to the contrary that $\alpha_1 \neq 0$ and note that

$$0 = |\alpha_1|^2 \left\| h_{k_1} + \sum_{m=2}^{N} \frac{\alpha_m}{\alpha_1} h_{k_m} \right\|^2 \geqq |\alpha_1|^2 \|e_0^+\| > 0. \tag{125}$$

This contradiction establishes that $\alpha_1 = 0$. Continuing by induction in the same fashion, it can be shown that $\alpha_m = 0$, $1 \leqq m \leqq N$.

To show that the second condition of Lemma 1 implies linear independence, we use a proof similar to Tuft's.[13] By the congruence of (22), (124) is equivalent to

$$\int_{-\pi/T}^{\pi/T} \left| \sum_{m=1}^{N} \alpha_m e^{-j\omega k_m T} \right|^2 R(\omega) d\omega = 0,$$

which implies that the integrand is zero almost everywhere on $[a, b]$. This is impossible unless $\alpha_m = 0$, $1 \leqq m \leqq N$, since otherwise

$$\left| \sum_{m=1}^{N} \alpha_m e^{-j\omega k_m T} \right|^2$$

has at most a finite number of algebraic zeros on $[a, b]$ and $R(\omega)$ is strictly positive.

Proof of Theorem 3: We will prove the result for the ZFE; the proof for the DFE is identical. Since for $N \leqq M$

$$M(h_k,\ |k| \leqq N, k \neq 0) \subset M(h_k,\ |k| \leqq M, k \neq 0) \subset M(h_k,\ k \neq 0),$$

the inequality

$$\|e_0\| \leqq \|e_0(M)\| \leqq \|e_0(N)\|$$

follows. Hence $\|e_0(M)\|^2$ must approach a limit,

$$\lim_{N\to\infty} \|e_0(N)\| \geqq \|e_0\|.$$

Denote by the shortened notation P the projection of h_0 on $M(h_k, k \neq 0)$ (so that $e_0 \triangleq h_0 - P$). Since $P \in M(h_k,\ k \neq 0)$, there exists a sequence $\gamma_n \in M(h_k,\ |k| \leqq n,\ k \neq 0)$ such that $\gamma_n \to P$ and we have

$$\|h_0 - \gamma_n\|^2 = \|e_0\|^2 + \|P - \gamma_n\|^2.$$

For any $\epsilon > 0$, there exists an $N(\epsilon)$ such that

$$\|h_0 - \gamma_n\|^2 \leqq \|e_0\|^2 + \epsilon$$

for $n \geqq N(\epsilon)$, and since $\|e_0(n)\|^2 \leqq \|h_0 - \gamma_n\|^2$ we have

$$\|e_0\|^2 \leqq \|e_0(n)\|^2 \leqq \|e_0\|^2 + \epsilon,$$

which establishes that $\|e_0(n)\| \to \|e_0\|$. The remainder of the proof follows that of the projection theorem. By the parallelogram law,

$$\|e_0(N) - e_0\|^2 = 2\|e_0(N)\|^2 + 2\|e_0\|^2 - \|e_0(N) + e_0\|^2,$$

but defining $P(N) = P[h_0, M(h_k,\ |k| \leqq N,\ k \neq 0)]$

$$\begin{aligned}\|e_0(N) + e_0\|^2 &= \|h_0 - P(N) + h_0 - P\|^2 \\ &= 4\left\|h_0 - \frac{P(N) + P}{2}\right\|^2 \geqq 4\|e_0\|^2,\end{aligned}$$

we have

$$\|e_0(N) - e_0\|^2 \leqq 2[\|e_0(N)\|^2 - \|e_0\|^2] \to 0.$$

Proof of Theorem 4: From (22) we have

$$\left\| \sum_{m=1}^{N} \beta_m h_{k_m} \right\|^2 = \frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} \left| \sum_{m=1}^{N} \beta_m e^{-j\omega k_m T} \right|^2 R(\omega) d\omega \cdot$$

A standard result of Toeplitz theory asserts that

$$\frac{1}{T}\left\{\sum_{m=1}^{N} |\beta_m|^2\right\} \text{ess inf } R(\omega) \leqq \left|\left| \sum_{m=1}^{N} \beta_m h_k \right|\right|^2 \leqq \frac{1}{T}\left\{\sum_{m=1}^{N} |\beta_m|^2\right\} \text{ess sup } R(\omega)\cdot$$

The conclusions of the theorem then follow from Theorem 5.17.18 of Ref. 10.

REFERENCES

1. Wozencraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering*, New York: Wiley, 1967.
2. Grenander, U., and Rosenblatt, M., *Statistical Analysis of Stationary Time Series*, New York: Wiley, 1957.
3. Doob, J. L., *Stochastic Processes*, New York: Wiley, 1953.
4. Lucky, R. W., Salz, J., and Weldon, E. J., *Principles of Data Communication*, New York: McGraw-Hill, 1968.
5. Monsen, P., "Linear Equalization for Digital Transmission over Noisy Dispersive Channels," Ph.D. Dissertation, Columbia University, June 1970.
6. Price, R., "Nonlinearly Feedback-Equalized PAM vs Capacity for Noisy Filter Channels," 1972, Int. Conf. Commun., Philadelphia, June 1972.
7. Messerschmitt, D. G., "A Geometric Theory of Intersymbol Interference. Part II: Performance of the Maximum Likelihood Detector," B.S.T.J., this issue, pp. 1521–1539.
8. Messerschmitt, D. G., "A Unified Geometric Theory of Zero-Forcing and Decision-Feedback Equalization," 1973 Int. Conf. Commun., Seattle, June 1973.
9. Messerschmitt, D. G., "Digital Communications: Detectors and Estimators for the Time-Varying Channel with Intersymbol Interference," Ph.D. Dissertation, University of Michigan, December 1971.
10. Naylor, A. W., and Sell, G. R., *Linear Operator Theory in Science and Engineering*, New York: Holt, Rinehart and Winston, 1971.
11. Parzen, E., *Time Series Analysis Papers*, San Francisco: Holden-Day, 1967.
12. Forney, G. D., Jr., "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," IEEE Trans. Inform. Theory, *IT-18*, May 1972, p. 363.
13. Tufts, D. W., "Nyquist's Problem—The Joint Optimization of Transmitter and Receiver in Pulse Amplitude Modulation," Proc. IEEE, *53*, March 1965.
14. Forney, G. D., Jr., "Lower Bounds on Error Probability in the Presence of Large Intersymbol Interference," IEEE Trans. Commun., *COM-20*, February 1972, p. 76.