[13] R. W. Lucky, J. Salz, and E. J. Weldon, *Principles of Data Communication.* New York: McGraw-Hill, 1968.
[14] F. S. Hill, Jr., "The computation of error probability for digital transmission," *Bell Syst. Tech. J.*, vol. 50, pp. 2055–2077, July–Aug. 1971.
[15] E. R. Kretzmer, "Generalization of a technique for binary data communication," *IEEE Trans. Commun. Technol.* (Concise Papers), vol. COM-14, pp. 67–68, Feb. 1966.
[16] E. Hansler, "An upper bound of error probability in data transmission systems," *Nachrichtentech. Z.*, no. 12, pp. 625–627, 1970.

Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference

G. DAVID FORNEY, JR., MEMBER, IEEE

Abstract—**A maximum-likelihood sequence estimator for a digital pulse-amplitude-modulated sequence in the presence of finite intersymbol interference and white Gaussian noise is developed. The structure comprises a sampled linear filter, called a whitened matched filter, and a recursive nonlinear processor, called the Viterbi algorithm. The outputs of the whitened matched filter, sampled once for each input symbol, are shown to form a set of sufficient statistics for estimation of the input sequence, a fact that makes obvious some earlier results on optimum linear processors. The Viterbi algorithm is easier to implement than earlier optimum nonlinear processors and its performance can be straightforwardly and accurately estimated. It is shown that performance (by whatever criterion) is effectively as good as could be attained by any receiver structure and in many cases is as good as if intersymbol interference were absent. Finally, a simplified but effectively optimum algorithm suitable for the most popular partial-response schemes is described.**

Introduction

INTERSYMBOL interference arises in pulse-modulation systems whenever the effects of one transmitted pulse are not allowed to die away completely before the transmission of the next. It is the primary impediment to reliable high-rate digital transmission over high signal-to-noise ratio narrow-bandwidth channels such as voice-grade telephone circuits. Intersymbol interference is also introduced deliberately for the purpose of spectral shaping in certain modulation schemes for narrow-band channels, called duobinary, partial-response, and the like [1]–[3].

The simplest model of a digital communication system subject to intersymbol interference occurs in pulse amplitude modulation (PAM), illustrated in Fig. 1. A sequence of real numbers x_k drawn from a discrete alphabet passes through a linear channel whose impulse response $h(t)$ is longer than the symbol separation T, and the filtered signal

$$s(t) = \sum_k x_k h(t - kT) \tag{1}$$

is corrupted by white Gaussian noise $n(t)$ to give a received signal

$$r(t) = s(t) + n(t). \tag{2}$$

In this paper we shall restrict ourselves to finite impulse responses $h(t)$.

This model dates back to Nyquist and is so simple that it would seem unlikely that at this late date anything new could be said about it. However, no serious attention seems to have been given to this problem until the last decade, when practical requirements for high-speed digital transmission over telephone circuits have begun to become important.

While lip service has long been paid to the idea that symbol decisions ought to be based on the entire received sequence, the fact that straightforward likelihood calculations grow exponentially with message length [4] has justified a retreat to simple symbol-by-symbol decisions in most theoretical and practical work. Early work analyzed and optimized linear transmitter and receiver filters subject to various criteria [5]–[11]. In this work the optimum receiver filter always turned out to be a combination of a matched filter and a transversal filter, the general reason for which is explained below.

More recently, nonlinear receivers have been investigated. Several authors [12]–[16] have developed "optimum" or approximately optimum nonlinear receiver structures, again subject to a variety of criteria. The intimidating complexity of these structures has led to interest in suboptimum nonlinear structures such as decision feedback [17], [18]. Invariably, the complaint is made that it is difficult to estimate the performance of nonlinear receivers analytically and resort is made to simulation.

In this paper we introduce a receiver structure (Fig. 2) consisting of a linear filter, called a whitened matched filter, a symbol-rate sampler, and a recursive nonlinear processor, called the Viterbi algorithm. This structure is a maximum-likelihood estimator of the entire transmitted

Manuscript received December 17, 1970; revised December 29, 1971. This work was supported in part by NSF Grant GK-31627.

The author is with the Codex Corporation, Newton, Mass. 02195. He is now on leave of absence at the Department of Electrical Engineering, Stanford University, Stanford, Calif. 94305.

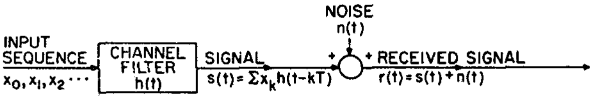

Fig. 1. PAM communications channel.

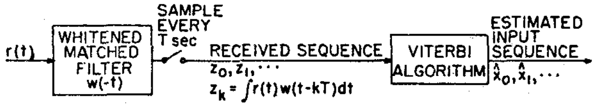

Fig. 2. Maximum-likelihood sequence estimator.

sequence; furthermore, it can be implemented and analyzed.

The whitened matched filter has the following properties.

1) Simplicity: The transition from continuous to discrete time does not require a bank of filters, or oversampling, but merely a single sample for each input symbol. There is some freedom in specifying the response $w(-t)$; when $h(t)$ is finite, $w(-t)$ can be chosen to be causal and hence realizable. (It can also be chosen to be anticausal, which is more natural in principle.)

2) Sufficiency: The filter is information lossless, in the sense that its sampled outputs are a set of sufficient statistics for estimation of the input sequence.

3) Whiteness: The noise components of the sampled outputs are independent identically distributed Gaussian random variables. (This and the sufficiency property follow from the fact that the set of waveforms $w(t - kT)$ is an orthonormal basis for the signal space.)

The Viterbi algorithm was originally invented to decode convolutional codes [19]–[21] and later shown to be a shortest-route algorithm of a type long known in operations research [20]–[27], which in turn can be expressed as a variant of dynamic programming [26]–[28]. Its applicability to partial-response systems was noticed by Omura and Kobayashi at U.C.L.A. [29]–[33] independently of and about the same time as the present work, and Kobayashi independently determined its performance for discrete-time responses of the form $1 \pm D^n$ [33]. It has the following properties.

1) Implementability: Like the best of the earlier "optimum" nonlinear processors, the Viterbi algorithm is a recursive structure that does not grow with the length of the message sequence, and that has complexity proportional to m^L, where m is the size of the input alphabet and L is the length of the impulse response $h(t)$ in units of T. In detail, it is superior in not requiring any multiplications, but only m^L additions and m^{L-1} m-ary comparisons per received symbol, which greatly simplifies hardware implementation; it also requires only m^{L-1} words of memory (of the order of 10–30 bits per word).

2) Analyzability: The ease with which performance can be analyzed is in marked contrast to all earlier work with nonlinear processors. We show that at moderate-to-high signal-to-noise ratios the symbol-error probability is accurately overbounded and estimated by

$$\Pr(e) \lesssim K_2 Q[d_{\min}/2\sigma], \tag{3}$$

where $d_{\min}^2$ is the minimum energy of any nonzero signal, σ^2 is the spectral density of the noise, K_2 is a small constant independent of σ^2, and $Q(\cdot)$ is the probability of error function

$$Q(x) \triangleq (2\pi)^{-1/2} \int_x^\infty dy\, e^{-y^2/2}. \tag{4}$$

We observe that for any estimator,

$$\Pr(e) \geq K_0 Q[d_{\min}/2\sigma], \tag{5}$$

where K_0 is another constant typically within an order of magnitude of K_2. When, as usually happens, $d_{\min}^2$ is equal to the energy $\|h\|^2$ in a single pulse, then the probability of error is approximately equal to $Q[\|h\|/2\sigma]$, which is what it would be for one-shot communication (no intersymbol interference). When $d_{\min}^2 < \|h\|^2$, if we can insert a partial-response-type preemphasis filter at the transmitter, we get a similar result (Appendix II).

3) Optimality: Our structure is optimum for maximum-likelihood estimation of the entire transmitted sequence, provided unbounded delay is allowed at the output, and is effectively optimum in the same sense for reasonable finite delays. This implies that it also minimizes the error-event probability $\Pr(E)$ (maximizes the mean time between error events), which is more significant than the symbol-error probability in many applications. As for the symbol-error probability $\Pr(e)$ itself, while examples can be constructed [34] in which the Viterbi algorithm does not minimize $\Pr(e)$, the bounds of (3) and (5) show that at moderate-to-high signal-to-noise ratios the estimator that minimizes $\Pr(e)$ cannot improve substantially on the performance of our structure.

Finally, in the last section we shall describe a practical embodiment of these ideas: a simple approximation to the preceding structure that has been implemented in a commercially available partial-response modem and that reduces error rates by one to two orders of magnitude on typical channels.

Definitions

We shall use the PAM model of Fig. 1. The inputs x_k are assumed to be equally spaced and scaled and biased to be integers in the range $0 \leq x_k \leq m - 1$; they are assumed to start at time $k = 0$ and to continue until $k = K$, where K may be infinite. With the input sequence we associate the formal power series in the delay operator D (D-transform)

$$x(D) \triangleq x_0 + x_1 D + x_2 D^2 + \cdots, \tag{6}$$

which will itself frequently be referred to as the input sequence.

The channel is characterized by a finite impulse response $h(t)$ of length L symbol intervals; i.e., L is the smallest integer such that $h(t) = 0$ for $t \geq LT$. The response $h(t)$ is assumed square-integrable,

$$\|h\|^2 \triangleq \int_{-\infty}^{\infty} h^2(t)\, dt < \infty. \tag{7}$$

We define

$$h^{(k)}(t) \triangleq h(t - kT) \tag{8}$$

and use inner-product notation in which

$$[a(t),b(t)] \triangleq \int_{-\infty}^{\infty} a(t)b(t)\, dt. \tag{9}$$

Then the pulse autocorrelation coefficients of $h(t)$ are

$$R_{k-k'} \triangleq [h^{(k)}(t),h^{(k')}(t)]$$
$$= \begin{cases} \int_{-\infty}^{\infty} h(t - kT)h(t - k'T)\, dt, & |k - k'| \leq L - 1 \\ 0, & |k - k'| \geq L. \end{cases} \tag{10}$$

We define the pulse autocorrelation function of $h(t)$ as

$$R(D) \triangleq \sum_{k=-\nu}^{\nu} R_k D^k, \tag{11}$$

where $\nu = L - 1$ is called the *span* of $h(t)$. We shall also write $R_{hh}(D)$ when it is necessary to distinguish different autocorrelation functions.

The response $h(t)$ may be regarded as a sequence of L *chips* $h_i(t)$, $0 \leq i \leq \nu$, where $h_i(t)$ is a function that is nonzero only over the interval $[0,T)$; i.e.,

$$h(t) = h_0(t) + h_1(t - T) + \cdots + h_\nu(t - \nu T). \tag{12}$$

Then it is natural to associate with $h(t)$ the *chip D-transform*

$$h(D,t) \triangleq \sum_{i=0}^{\nu} h_i(t)D^i, \tag{13}$$

which is a polynomial in D of degree ν with coefficients in the set of functions over $[0,T)$. It is easy to verify that the chip D-transform has the following properties.

1) The pulse autocorrelation function is given by

$$R(D) = [h(D,t),h(D^{-1},t)] = \int_0^T h(D,t)h(D^{-1},t)\, dt. \tag{14}$$

2) A transversal filter is a filter with response

$$g(t) = \sum_i g_i\, \delta(t - iT) \tag{15}$$

for some set of coefficients g_i. This response is not square integrable, but we assume the coefficients g_i are square summable, $\sum_i g_i^2 < \infty$. We say a transversal filter $g(t)$ is characterized by $g(D)$ if

$$g(D) = \sum_i g_i D^i. \tag{16}$$

The chip D-transform $g(D,t)$ of a transversal filter is

$$g(D,t) = g(D)\, \delta(t). \tag{17}$$

3) The cascade of a square-integrable filter with response $g(t)$ and a transversal filter characterized by a square-summable $f(D)$ is a filter with square-integrable response $h(t)$, chip D-transform

$$h(D,t) = f(D)g(D,t) \tag{18}$$

and pulse autocorrelation function

$$R_{hh}(D) = f(D)f(D^{-1})R_{gg}(D). \tag{19}$$

The Matched Filter

The signal $s(t)$ is defined as

$$s(t) \triangleq \sum_{k=0}^{K} x_k h(t - kT) = \sum_{k=0}^{K} x_k h^{(k)}(t). \tag{20}$$

The received signal $r(t)$ is $s(t)$ plus white Gaussian noise $n(t)$.

It is well known (see, for example, [35]) that in the detection of signals that are linear combinations of some finite set of square-integrable basis functions $h^{(k)}(t)$, the outputs of a bank of matched filters, one matched to each basis function, form a set of sufficient statistics for estimating the coefficients. Thus the $K + 1$ quantities

$$a_k \triangleq [r(t),h^{(k)}(t)]$$
$$= \int_{-\infty}^{\infty} r(t)h(t - kT)\, dt, \qquad 0 \leq k \leq K \tag{21}$$

form a set of sufficient statistics for estimation of the x_k, $0 \leq k \leq K$, when K is finite. But these are simply the sampled outputs of a filter $h(-t)$ matched to $h(t)$. Hence we have the following proposition.

Proposition 1: When $x(D)$ is finite, the sampled outputs a_k [defined by (21)] of a filter matched to $h(t)$ form a set of sufficient statistics for estimation of the input sequence $x(D)$.

It is obvious on physical grounds that this property does not depend on $x(D)$ being finite, but the corresponding result does not seem to be available in the literature. If $x(D)$ is infinite the signals have infinite duration, so that the Karhunen–Loève expansion is not applicable, and infinite energy, so that the generalization of Bharucha and Kadota [36] cannot be applied. We leave the proof to the reader, using his favorite definition of white Gaussian noise, which as usual is the only technical difficulty. (The Alexandrian method of dealing with this Gordian knot would be to *define* white Gaussian noise as any noise such that Proposition 1 is valid for infinite $x(D)$ and any square integrable $h(t)$.)

In view of the obviousness of Proposition 1, it is remarkable that it has not been much exploited previously in the intersymbol interference literature. (It does appear as a problem in Van Trees [37], attributed to Austin.) Some authors make the transition from continuous to discrete time by a sampler without a matched filter, with no explicit recognition that such a procedure is information lossy in general, or else gloss over the problem entirely. Others express the signal waveform in each symbol interval as a linear combination of chips and sample the outputs of a bank of filters matched to all the chips.

Furthermore, there is a whole series of papers (see [11] and the references therein) showing that the optimum linear receiving filter under various criteria can be expressed as the cascade of a matched filter and a transversal filter.

But since the matched filter is linear and its sampled outputs can be used without loss of optimality, any optimal linear receiver must be expressible as a linear combination of the sampled matched filter outputs a_k. Hence Proposition 1 has the following corollary.

Corollary: For any criterion of optimality, the optimum linear receiver is expressible as the cascade of a matched filter and a (possibly time-varying) transversal filter.

(If the criterion is the minimization of the ensemble average of some quantity per symbol and $x(D)$ is long enough so that end effects are unimportant, then it is easy to show that the optimum transversal filter is time invariant.)

The Whitened Matched Filter

Define the matched-filter output sequence as

$$a(D) \triangleq \sum_{k=0}^{K} a_k D^k. \tag{22}$$

Since

$$\begin{aligned} a_k &= [r(t),h^{(k)}(t)] \\ &= \sum_{k'} x_{k'}[h^{(k')}(t),h^{(k)}(t)] + [n(t),h^{(k)}(t)] \\ &= \sum_{k'} x_{k'}R_{k-k'} + n_k' \end{aligned} \tag{23}$$

we have

$$a(D) = x(D)R(D) + n'(D). \tag{24}$$

Here $n'(D)$ is zero-mean colored Gaussian noise with autocorrelation function $\sigma^2 R(D)$, since

$$\begin{aligned} \overline{n_k' n_{k'}'} &= \iint dt\, d\tau\, \overline{n(t)n(\tau)}\, h(t - kT)h(\tau - k'T) \\ &= \sigma^2 R_{k-k'}, \end{aligned} \tag{25}$$

where σ^2 is the spectral density of the noise $n(t)$, so that $\overline{n(t)n(\tau)} = \sigma^2\delta(t - \tau)$.

Since $R(D)$ is finite with $2\nu + 1$ nonzero terms, it has 2ν complex roots; further, since $R(D) = R(D^{-1})$, the inverse β^{-1} of any root β is also a root of $R(D)$, so the roots break up into ν pairs. Then if $f'(D)$ is any polynomial of degree ν whose roots consist of one root from each pair of roots of $R(D)$, $R(D)$ has the spectral factorization

$$R(D) = f'(D)f'(D^{-1}). \tag{26}$$

We can generalize (26) slightly by letting $f(D) = D^n f'(D)$ for any integer delay n; then

$$R(D) = f(D)f(D^{-1}). \tag{27}$$

Now let $n(D)$ be zero-mean white Gaussian noise with autocorrelation function σ^2; we can represent the colored noise $n'(D)$ by

$$n'(D) = n(D)f(D^{-1}) \tag{28}$$

since $n'(D)$ then has the autocorrelation function $\sigma^2 f(D^{-1})f(D) = \sigma^2 R(D)$ and zero-mean Gaussian noise is entirely specified by its autocorrelation function. Consequently we may write (24) as

$$a(D) = x(D)f(D)f(D^{-1}) + n(D)f(D^{-1}). \tag{29}$$

This suggests that we simply divide out the factor $f(D^{-1})$ formally to obtain a sequence

$$z(D) = a(D)/f(D^{-1}) = x(D)f(D) + n(D) \tag{30}$$

in which the noise is white.

When $f(D^{-1})$ has no roots on or inside the unit circle, the transversal filter characterized by $1/f(D^{-1})$ is actually realizable in the sense that its coefficients are square summable. Then the sequence $z(D)$ of (30) can actually be obtained by sampling the outputs of the cascade of a matched filter $h(-t)$ with a transversal filter characterized by $1/f(D^{-1})$ (with whatever delay is required to assure causality). We call such a cascade a whitened matched filter.

More generally, we shall now show that for any spectral factorization of the form (27), the filter $w(t)$ whose chip D-transform is

$$w(D,t) \triangleq \frac{1}{f(D)} h(D,t) \tag{31}$$

is well defined, and its time reversal $w(-t)$ can be used as a whitened matched filter in the sense that its sampled outputs

$$z_k = \int_{-\infty}^{\infty} r(t)w(t - kT)\, dt \tag{32}$$

satisfy (30) with $n(D)$ a white Gaussian noise sequence.

We write $f(D)$ as

$$f(D) = cD^n \prod_{i=1}^{\nu} (1 - \beta_i^{-1}D) \tag{33}$$

for some constant c, integer n, and complex roots β_i. Since realizability is not our main concern, we make the definitions

Definition 1: If $|\beta| > 1$,

$$(1 - \beta^{-1}D)^{-1} \triangleq 1 + \beta^{-1}D + \beta^{-2}D^2 + \cdots.$$

Definition 2: If $|\beta| < 1$,

$$\begin{aligned} (1 - \beta^{-1}D)^{-1} &= -\beta D^{-1}(1 - \beta D^{-1})^{-1} \\ &\triangleq -(\beta D^{-1} + \beta^2 D^{-2} + \cdots). \end{aligned}$$

Then if there are no roots β_i on the unit circle, $1/f(D)$ can be represented as a cascade of ν square-summable transversal filters and (31) makes sense.

To handle roots on the unit circle, we introduce the following useful lemma.

Lemma 1: If $h(t)$ is a finite square-integrable impulse response of span ν and the corresponding pulse autocorrelation function $R_{hh}(D)$ has a root β with $|\beta| = 1$, then $h(t)$ may be represented as the cascade of a transversal filter characterized by $(1 - \beta^{-1}D)$ and a filter with impulse response $g(t)$, where $g(t)$ has pulse autocorrelation function

$$R_{gg}(D) = \frac{R_{hh}(D)}{(1 - \beta^{-1}D)(1 - \beta^{-1}D^{-1})} \tag{34}$$

and is finite with span $\nu - 1$.

Proof: Let $h(D,t)$ and $g(D,t)$ be the corresponding chip D-transforms; the lemma asserts that

$$g(D,t)(1 - \beta^{-1}D) = h(D,t). \tag{35}$$

This suggests that we define $g(D,t)$ formally as

$$\begin{aligned} g(D,t) &= \frac{h(D,t)}{(1 - \beta^{-1}D)} \\ &= h(D,t) \sum_{k=0}^{\infty} \beta^{-k}D^k, \end{aligned} \tag{36}$$

where the chips $g_i(t)$ are defined in terms of the chips $h_j(t)$ and β as

$$g_i(t) \triangleq \sum_{j=0}^{i} \beta^{j-i}h_j(t), \qquad i \geq 0. \tag{37}$$

For $i \geq \nu$,

$$g_i(t) = \sum_{j=0}^{\nu} \beta^{j-i}h_j(t) = \beta^{-i}h(\beta,t), \tag{38}$$

where $h(\beta,t)$ is the chip

$$h(\beta,t) = \sum_{j=0}^{\nu} \beta^j h_j(t). \tag{39}$$

But, using an asterisk to represent complex conjugation,

$$\begin{aligned} \|h(\beta,t)\|^2 &= [h(\beta,t),h^*(\beta,t)] \\ &= [h(\beta,t),h(\beta^*,t)] \\ &= [h(\beta,t),h(\beta^{-1},t)] \\ &= R(\beta) = 0, \end{aligned} \tag{40}$$

where we have used (14) and the fact that $\beta^* = \beta^{-1}$ since $|\beta| = 1$. Consequently $h(\beta,t) = 0$ and $g_i(t) = 0$ for $i \geq \nu$. Hence the filter of span $\nu - 1$ defined by

$$g(D,t) \triangleq \sum_{i=0}^{\infty} g_i(t)D^i, \tag{41}$$

where $g_i(t)$ is given by (37), is the required filter. The expression for the autocorrelation function follows from combination of (35) with (19). Q.E.D.

Since the spectrum $S(\omega)$ of $h(t)$ in the Nyquist band is given by $S(\omega) = R[\exp(j2\pi\omega T)]$, $0 \leq \omega \leq 1/2T$, Lemma 1 has the interesting interpretation that any filter with nulls in its Nyquist spectrum can be regarded as the cascade of a null-free filter and a transversal filter that inserts the nulls at the appropriate places. For example, a filter with nulls at the upper and lower band edges is the cascade of a null-free filter and the transversal filter characterized by $1 - D^2$ (assuming the nulls are simple).

In view of Lemma 1, the following definition makes sense.

Definition 3: If $|\beta| = 1$ and $R_{hh}(\beta) = 0$, then $(1 - \beta^{-1}D)^{-1}h(D,t)$ is the filter $g(D,t)$ whose existence is implied by Lemma 1 [defined by (37) and (41)].

With this interpretation, (31) is well defined for any finite $h(t)$. Furthermore, from (19) and Lemma 1,

$$R_{ww}(D) = \frac{R_{hh}(D)}{f(D)f(D^{-1})} = 1 \tag{42}$$

so that the set of functions $w(t - kT)$ is orthonormal. Finally, the set is a basis for the signal space since the signals $s(t)$ have chip D-transforms

$$\begin{aligned} s(D,t) &= x(D)h(D,t) \\ &= x(D)f(D)w(D,t) \end{aligned} \tag{43}$$

so that

$$s(t) = \sum y_k w(t - kT), \tag{44}$$

where the signal sequence $y(D)$ is defined as

$$y(D) \triangleq x(D)f(D). \tag{45}$$

We note that only $K + \nu + 1$ of the y_k are nonzero. That the set of sampled outputs $z_k = [r(t),w(t - kT)]$ is a set of sufficient statistics for $y(D)$ and hence for $x(D)$ follows immediately from this observation, or alternately from the fact that the sufficient statistics $a(D)$ can be recovered by passing $z(D)$ through the finite filter $f(D^{-1})$.

We collect these results into the following theorem.

Theorem 1: Let $h(t)$ be finite with span ν and let $f(D)f(D^{-1})$ be any spectral factorization of $R_{hh}(D)$. Then the filter whose chip D-transform is $w(D,t) = h(D,t)/f(D)$ has square-integrable impulse response $w(t)$ under Definitions 1–3, and the sampled outputs z_k of its time reverse

$$z_k = \int_{-\infty}^{\infty} r(t)w(t - kT)\,dt \tag{46}$$

form a sequence

$$z(D) = x(D)f(D) + n(D) \tag{47}$$

in which $n(D)$ is a white Gaussian noise sequence with variance $\overline{n_k^2} = \sigma^2$, and which is a set of sufficient statistics for estimation of the input sequence $x(D)$.

A factorization $R(D) = f_c(D)f_c(D^{-1})$ is said to be canonical if $f_c(D)$ is a real polynomial of degree ν that contains all the roots of $R(D)$ outside the unit circle. Correspondingly there are two canonical choices for $w(t)$:

$$w_{c1}(D,t) = \frac{h(D,t)}{f_c(D)} \tag{48}$$

$$w_{c2}(D,t) = \frac{h(D,t)D^{-\nu}}{f_c(D^{-1})}. \tag{49}$$

The first choice seems more natural and yields a causal $w_{c1}(t)$; however, in the latter case $w_{c2}(t)$ is purely anticausal, so that the whitened matched filter response $w_{c2}(-t)$ is purely causal and thus corresponds to a realizable filter. The corresponding signal sequences $y(D)$ are

$$\begin{aligned} y_1(D) &= x(D)f_c(D) \\ y_2(D) &= x(D)D^\nu f_c(D^{-1}) \end{aligned} \tag{50}$$

so that one impulse response $f_c(D)$ is the time reversal of the other, $D^\nu f_c(D^{-1})$.

While we have developed these ideas only for finite $R(D)$, they extend practically without change to rational

$R(D)$ (except that there will in general be no purely causal whitened matched filter) and appear to apply in much more general situations whenever $R(D)$ has any kind of spectral factorization.

Discrete-Time Model

We have now seen that by use of a whitened matched filter we may confine our attention to the following discrete-time model, without loss of optimality. The signal sequence

$$y(D) = x(D)f(D) \tag{51}$$

is the convolution of the input sequence $x(D)$ with the finite impulse response $f(D)$, whose autocorrelation function is $R(D) = f(D)f(D^{-1})$. Without loss of generality we assume that $f(D)$ is a polynomial of degree ν with $f_0 \neq 0$. The received sequence $z(D)$ is the sum of the signal sequence $y(D)$ and a white Gaussian noise sequence $n(D)$ with autocorrelation function σ^2.

The output signal-to-noise ratio is defined to be

$$\begin{aligned} \text{SNR} &\triangleq \sigma_y^2/\sigma^2 \\ &= \sigma_x^2 \|f\|^2/\sigma^2, \end{aligned} \tag{52}$$

where σ_x^2 is the input variance $[(m^2 - 1)/12]$ and

$$\|f\|^2 \triangleq \sum_{i=0}^{\nu} f_i^2 = R_0 \tag{53}$$

is the energy in the impulse response $f(D)$. (If $f(D)$ is derived from a continuous-time response $h(t)$, then $\|f\|^2 = \|h\|^2 = R_0$.)

In some contexts the channel itself is discrete time rather than continuous time and such a model arises directly. For example, in a partial-response system the spectral shaping may be achieved by passing the input sequence $x(D)$ through a discrete-time filter such as $1 - D^2$ to give the signal sequence

$$y(D) = x(D)(1 - D^2) \tag{54}$$

whose Nyquist spectrum has nulls at the upper and lower band edges.

However the model arises, it is crucial to observe that the signal sequence $y(D)$ may be taken to be generated by a finite-state machine driven by the input sequence $x(D)$. We may imagine a shift register of ν m-state memory elements containing the ν most recent inputs, with y_k formed as the weighted sum of the shift register contents and the current input x_k as pictured in Fig. 3. Clearly the machine has m^ν states, the state at any time being given by the ν most recent inputs:

$$s_k \triangleq (x_{k-1}, x_{k-2}, \cdots, x_{k-\nu}), \tag{55}$$

where by convention $x_k = 0$ for $k < 0$. We define the state sequence $s(D)$ as

$$s(D) \triangleq s_0 + s_1 D + s_2 D^2 + \cdots, \tag{56}$$

where each state s_k takes on values from an alphabet of m^ν states S_j, $1 \leq j \leq m^\nu$. The maps from input sequences $x(D)$ to state sequences $s(D)$ and thence to signal sequences $y(D)$ are obviously one-to-one and hence invertible. In fact, two successive states uniquely determine an output

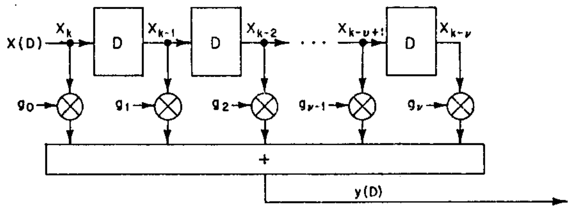

Fig. 3. Finite-state machine model.

$$y_k = y(s_k, s_{k+1}) \tag{57}$$

i.e., given a transition from s_k to s_{k+1}, the corresponding output y_k is determined.

An allowable state sequence $s(D)$ or signal sequence $y(D)$ is defined as one that could result from an allowable input sequence.

Maximum-Likelihood Sequence Estimation

Maximum-likelihood sequence estimation is defined as the choice of that $x(D)$ for which the probability density $p[z(D) \mid x(D)]$ is maximum. Since we have permitted sequences to be semi-infinite, so that $p[z(D) \mid x(D)]$ may be zero for all $x(D)$, some sort of limiting operation is implied; we shall see below that the estimate $\hat{x}(D)$ can be determined recursively in a way that gives sense to this definition.

Since the maps from $x(D)$ to $s(D)$ and to $y(D)$ are one to one, maximum-likelihood sequence estimation can equivalently be defined as choosing from the allowable $s(D)$ that which maximizes $p[z(D) \mid s(D)]$, or from the allowable $y(D)$ that which maximizes $p[z(D) \mid y(D)]$. We feel that it is most illuminating to consider the problem to be the estimation of the state sequence of a finite-state machine from noisy observations.

To construct the recursive estimation algorithm known as the Viterbi algorithm, we first use the fact that the noise terms n_k are independent. Then the log likelihood $\ln p[z(D) \mid s(D)]$ breaks up into a sum of independent increments:

$$\ln p[z(D) \mid s(D)] = \sum_k \ln p_n[z_k - y(s_k, s_{k+1})], \tag{58}$$

where $p_n(\cdot)$ is the probability density of each noise term n_k. For notational convenience we define the partial sums

$$\Gamma[s(D)]_{k_1}^{k_2} \triangleq \sum_{k=k_1}^{k_2-1} \ln p_n[z_k - y(s_k, s_{k+1})], \qquad 0 \leq k_1 < k_2. \tag{59}$$

Suppose for the moment that we knew that the state s_k at time k was S_j. Then for any allowable state sequence $s(D)$ that starts with the known initial state $s_0 = \mathbf{0}$ and passes through the state S_j at time k, the log likelihood would break up into two independent parts:

$$\Gamma[s(D)]_0^K = \Gamma[s(D)]_0^k + \Gamma[s(D)]_k^K. \tag{60}$$

Let $\hat{s}_j(D)$ be the allowable state sequence from time 0 to k that has minimum log likelihood $\Gamma[s(D)]_0{}^k$ among all allowable state sequences starting with $s_0 = \mathbf{0}$ and ending with $s_k = S_j$. We call $\hat{s}_j(D)$ the *survivor* at time k corresponding to state S_j. Then we assert that $\hat{s}_j(D)$ must be the initial segment of the maximum likelihood state sequence $s(D)$; for we can replace the initial segment $s'(D)$ of any allowable state sequence passing through S_j with the initial segment $\hat{s}_j(D)$ and obtain another allowable sequence with greater log likelihood $\Gamma[s(D)]_0{}^K$, unless $\Gamma[s'(D)]_0{}^k = \Gamma[\hat{s}_j(D)]_0{}^k$.

In fact, we do not know the state s_k at time k; but we do know that it must be one of the finite number of states S_j, $1 \leq j \leq m^\nu$, of the shift register of Fig. 3. Consequently, while we cannot make a final decision as to the identity of the initial segment of the maximum-likelihood state sequence at time k, we know the initial segment must be among the m^ν survivors $\hat{s}_j(D)$, $1 \leq j \leq m^\nu$, one for each state S_j. Thus we need only store m^ν sequences $\hat{s}_j(D)$ and their log likelihoods $\Gamma[\hat{s}_j(D)]_0{}^k$, regardless of how large k becomes. To update the memory at time $k + 1$, recursion proceeds as follows.

1) For each of the m allowable continuations $s_{j'}(D)$ to time $k + 1$ of each of the m^ν survivors $\hat{s}_j(D)$ at time k compute

$$\Gamma[s_{j'}(D)]_0^{k+1} = \Gamma[\hat{s}_j(D)]_0{}^k + \ln p_n[z_k - y_k(S_j,S_{j'})]. \quad (61)$$

This involves $m^{\nu+1} = m^L$ additions.

2) For each of the states $S_{j'}$, $1 \leq j' \leq m^\nu$, compare the log likelihoods $\Gamma[s_{j'}(D)]_0^{k+1}$ of the m continuations terminating in that state and select the largest as the corresponding survivor. This involves m^ν m-ary comparisons, or $(m - 1)$ m^ν binary comparisons.

In principle the Viterbi algorithm can make a final decision on the initial state segment up to time $k - \tau$ when and only when all survivors at time k have the same initial state sequence segment up to time $k - \tau$. The decoding delay τ is unbounded but is generally finite with probability 1 [34]. In implementation, one actually makes a final decision after some fixed delay δ, with δ chosen large enough that the degradation due to premature decisions is negligible. Although, as we shall see later, δ may have to be much larger than ν, it is typically of the order of 20 symbols or less. Parenthetically, our analysis following shows that the capability of deferring decisions is essential, in the sense that any receiver that does not have the capability of deferring decisions for the appropriate δ cannot approach optimum performance.

We further note that in the derivation of the algorithm we have used only the finite-state machine structure and the independence of the noise, so that the technique can be adapted to account for Markov context dependence in the input and other Markov-modelable statistics of the source and channel, as recounted by Hilborn [16], for example. Omura [30] has considered the situation in which the input sequence $x(D)$ is a code word from a convolutional code.

Error Events

We now begin our analysis of the probability of error in the estimated state sequence $\hat{s}(D)$ finally decided upon by the Viterbi algorithm. We let $\hat{x}(D)$ and $\hat{y}(D)$ stand for the corresponding estimated input sequence and signal sequence.

In the detection of a semi-infinite sequence there will generally occur an infinite number of errors. The idea of error events (see also [20] and [28]) makes precise our intuitive notion that these errors can be grouped into independent finite clumps. An *error event* $\mathscr{E}$ is said to extend from time k_1 to k_2 if the estimated state sequence $\hat{s}(D)$ is equal to the correct state sequence $s(D)$ at times k_1 and k_2, but nowhere in between ($s_{k_1} = \hat{s}_{k_1}$; $s_{k_2} = \hat{s}_{k_2}$; $s_k \neq \hat{s}_k$, $k_1 < k < k_2$). The length of the error event is defined as $n \triangleq k_2 - k_1 - 1$. Clearly $n \geq \nu$, with no upper bound; however, we shall find that n is finite with probability 1.

When the channel is linear, in the sense that $y(D) = x(D)f(D)$, we can say more about an error event. Since $s_{k_1} = \hat{s}_{k_1}$ and $s_{k_2} = \hat{s}_{k_2}$, we have

$$\begin{aligned} x_k = \hat{x}_k, \qquad & k_1 - \nu \leq k \leq k_1 - 1 \\ & k_2 - \nu \leq k \leq k_2 - 1 \end{aligned} \quad (62)$$

from the definition of s_k. However, $x_{k_1} \neq \hat{x}_{k_1}$ and $x_{k_2-\nu-1} \neq \hat{x}_{k_2-\nu-1}$ since $s_{k_1+1} \neq \hat{s}_{k_1+1}$ and $s_{k_2-1} \neq \hat{s}_{k_2-1}$. We define

$$\begin{aligned} \varepsilon_x(D) \triangleq (x_{k_1} - \hat{x}_{k_1}) + (x_{k_1+1} - \hat{x}_{k_1+1})D + \cdots \\ + (x_{k_2-\nu-1} - \hat{x}_{k_2-\nu-1})D^{n-\nu} \end{aligned} \quad (63)$$

as the input error sequence associated with the error event. It is a polynomial with nonzero constant term ε_{x0} and degree $n - \nu$ and contains no sequence of ν consecutive zero coefficients (since then s_k would equal $\hat{s}_k$ for some intermediate k and we would have two distinct error events). Furthermore, since the x_k are integers, the coefficients of $\varepsilon_x(D)$ are integral.

Correspondingly, we define the signal error sequence associated with the error event as

$$\begin{aligned} \varepsilon_y(D) \triangleq (y_{k_1} - \hat{y}_{k_1}) + (y_{k_1+1} - \hat{y}_{k_1+1})D + \cdots \\ + (y_{k_2-1} - \hat{y}_{k_2-1})D^n. \end{aligned} \quad (64)$$

Since $y(D) = x(D)f(D)$ and $\hat{y}(D) = \hat{x}(D)f(D)$, it follows that

$$\varepsilon_y(D) = \varepsilon_x(D)f(D). \quad (65)$$

Thus $\varepsilon_y(D)$ is a polynomial with nonzero constant term and degree n.

We define the Euclidean weight $d^2(\mathscr{E})$ of an error event as the energy in the associated signal-error sequence

$$\begin{aligned} d^2(\mathscr{E}) \triangleq \|\varepsilon_y\|^2 &= \sum_{i=0}^{n} \varepsilon_{yi}{}^2 \\ &= [\varepsilon_y(D)\varepsilon_y(D^{-1})]_0 \\ &= [\varepsilon_x(D)f(D)f(D^{-1})\varepsilon_x(D^{-1})]_0 \\ &= [\varepsilon_x(D)R(D)\varepsilon_x(D^{-1})]_0. \end{aligned} \quad (66)$$

We note that the energy depends only on $\varepsilon_x(D)$ and $R(D)$ and is therefore independent of the factorization $R(D) = f(D)f(D^{-1})$. In fact, when the received sequence is derived from a continuous-time received signal via a whitened matched filter, $d^2(\mathscr{E})$ is identifiable as the energy of the signal that results from passing the sequence $\varepsilon_x(D)$ through $h(t)$:

$$d^2(\mathscr{E}) = \|\varepsilon_x(D)h(D,t)\|^2 = \int \left[\sum_{i=0}^{n-v} \varepsilon_{xi} h(t - iT)\right]^2 dt. \qquad (67)$$

The number of errors in the associated input error sequence is defined as the Hamming weight $w_H(\mathscr{E})$ of the event; that is,

$$w_H(\mathscr{E}) \triangleq \text{number of nonzero coefficients in } \varepsilon_x(D). \qquad (68)$$

Probability of a Particular Error Event

We now calculate the probability that a particular error event identified by a starting time k_1 and an associated input error sequence $\varepsilon_x(D)$ of degree $n - v$ will actually occur. Three subevents must occur.

$\mathscr{E}_1$: At time k_1 we must have $s_{k_1} = \hat{s}_{k_1}$.

$\mathscr{E}_2$: Between k_1 and $k_1 + n - v$ the input sequence $x(D)$ must be such that $x(D) + \varepsilon_x(D)$ is an allowable sequence $\hat{x}(D)$. For example, if $\varepsilon_x(D) = 1$, then x_{k_1} must not equal $m - 1$, since then $\hat{x}_{k_1}$ would equal m, which is not an allowable level.

$\mathscr{E}_3$: The noise terms n_k, $k_1 \leq k \leq k_1 + n$, must be such that over this segment $\hat{x}(D)$ has greater likelihood than $x(D)$ or, in terms of the earlier notation of (59),

$$\Gamma[\hat{s}(D)]_{k_1}^{k_1+n+1} \geq \Gamma[s(D)]_{k_1}^{k_1+n+1}. \qquad (69)$$

When $n(D)$ is white and Gaussian with variance σ^2, we have

$$\ln p_n(z_k - y_k) = -\tfrac{1}{2} \ln 2\pi\sigma^2 - (z_k - y_k)^2/2\sigma^2 \qquad (70)$$

so that

$$[\hat{\Gamma} - \Gamma]_{k_1}^{k_1+n+1} = \frac{1}{2\sigma^2} \sum_{k=k_1}^{k_1+n} [(z_k - y_k)^2 - (z_k - \hat{y}_k)^2] = \frac{1}{2\sigma^2} [\|z(D) - y(D)\|^2 - \|z(D) - \hat{y}(D)\|^2]_{k_1}^{k_1+n} \qquad (71)$$

in obvious notation. In words, $\hat{y}(D)$ is more likely than $y(D)$ if $\hat{y}(D)$ is closer to $z(D)$ than is $y(D)$ in the $(n + 1)$-dimensional Euclidean space corresponding to times k_1 to $k_1 + n$. (The decision rule is thus independent of signal-to-noise ratio.) The three points

$$y(D) \mid_{k_1}^{k_1+n}, \qquad \hat{y}(D) \mid_{k_1}^{k_1+n}, \qquad z(D) \mid_{k_1}^{k_1+n}$$

define a two-dimensional subspace illustrated in Fig. 4. Since our Gaussian noise has equal variance in all dimensions, it is spherically symmetric and by coordinate rotation we can see that the probability of $\mathscr{E}_3$ is simply the probability that a single Gaussian variable of variance σ^2 exceeds half the Euclidean distance between $y(D)$ and $\hat{y}(D)$. But since

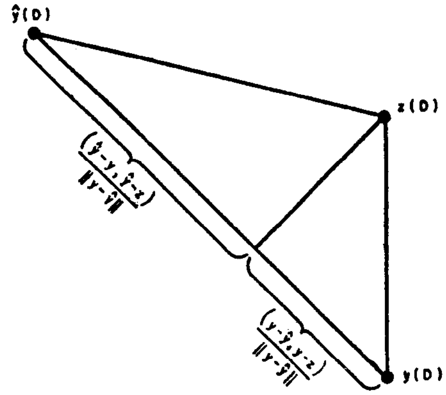

Fig. 4. Projection of $z(D)$ on $y(D) - \hat{y}(D)$.

$$\varepsilon_y(D) = [y(D) - \hat{y}(D)]_{k_1}^{k_1+n}$$

this distance squared is just the Euclidean weight $d^2(\mathscr{E})$ of $\varepsilon_y(D)$. Hence

$$\Pr(\mathscr{E}_3) = \int_{d(\mathscr{E})/2}^{\infty} dn (2\pi\sigma^2)^{-1/2} \exp - n^2/2\sigma^2 = Q[d(\mathscr{E})/2\sigma], \qquad (72)$$

where $Q(x)$ is defined in (4). We recognize this as the error probability when a binary signal of amplitude $\pm d(\mathscr{E})/2$ is sent through a Gaussian channel with noise variance σ^2. It depends only on $\varepsilon_y(D)$ and σ.

The subevent $\mathscr{E}_2$ is independent of $\mathscr{E}_1$ and $\mathscr{E}_3$, being dependent only on the message ensemble. Clearly when $|\varepsilon_{xi}| = j$ only $m - j$ values of x_{k+i} are permissible, so that

$$\Pr(\mathscr{E}_2) = \prod_{i=0}^{n-v} \frac{m - |\varepsilon_{xi}|}{m} \qquad (73)$$

assuming the inputs to be independent and equiprobable. Error events with any $|\varepsilon_{xi}| \geq m$ are impossible.

The subevent $\mathscr{E}_1$ is possibly dependent on the noise terms involved in $\mathscr{E}_3$ and the joint probability is not easily calculable. However, the probability that $\mathscr{E}_1$ is not true is of the order of the error probability, so that in the normal operating region $\Pr(\mathscr{E}_1 \mid \mathscr{E}_3)$ is closely approximated as well as overbounded by 1. In sum, therefore, the probability of the particular error event $\mathscr{E}$ is given by

$$\Pr(\mathscr{E}) = \Pr(\mathscr{E}_3) \Pr(\mathscr{E}_2) \Pr(\mathscr{E}_1 \mid \mathscr{E}_3) \lesssim Q[d(\mathscr{E})/2\sigma] \left[\prod_{i=0}^{n-v} \frac{m - |\varepsilon_{xi}|}{m}\right]. \qquad (74)$$

Probability of Error

From the probabilities of individual error events we obtain a simple and, for moderately low error probabilities like 10^{-3}, rather tight bound on overall probability of error through the union bound, which simply says that the probability of a union of events is not greater than the sum of their individual probabilities.

Let E be the set of all possible error events $\mathscr{E}$ starting at time k_1. Then the probability that any error event starts at time k_1 is bounded by

$$\Pr(E) \leq \sum_{\mathscr{E} \in E} \Pr(\mathscr{E}). \tag{75}$$

Let D be the set of all possible $d(\mathscr{E})$ and for each $d \in D$ let E_d be the subset of error events for which $d(\mathscr{E}) = d$. Then from (74)

$$\Pr(E) \leq \sum_{d \in D} Q[d/2\sigma] \sum_{\mathscr{E} \in E_d} \left[\prod_{i=0}^{n-\nu} \frac{m - |\varepsilon_{xi}|}{m}\right]. \tag{76}$$

Because of the exponential decrease of the Gaussian distribution function, this expression will be dominated at moderate signal-to-noise ratios by the term involving the minimum value $d_{\min}$ of $d(\mathscr{E})$:

$$\Pr(E) \simeq K_1 Q[d_{\min}/2\sigma], \tag{77}$$

where

$$K_1 \triangleq \sum_{\mathscr{E} \in E_{d_{\min}}} \left[\prod_{i=0}^{n-\nu} \frac{m - |\varepsilon_{xi}|}{m}\right] \tag{78}$$

is a constant independent of σ. Since this expression is independent of k_1, it may be read as the probability of an error event per unit time, and its reciprocal $1/\Pr(E)$ as the mean time between error events. The size of the signal-to-noise ratio at which this expression becomes a good estimate depends on the coefficients of $Q[d(\mathscr{E})/2\sigma]$ for larger $d(\mathscr{E})$, which in all cases we examine are of the order of magnitude of K_1.

A true bound with the same asymptotic behavior can be obtained by generating-function methods similar to those used by Viterbi [21]. Let N_d be the coefficient multiplying $Q[d/2\sigma]$ in (76) and let the generating function $g_N(z)$ be defined as

$$g_N(z) \triangleq \sum_{d \in D} N_d z^{d^2}. \tag{79}$$

As suggested by Viterbi in [21], we use the fact that $Q(x) \exp x^2/2$ is a monotonically decreasing function of x for $x \geq 0$ to obtain the bound

$$Q[d/2\sigma] \leq Q[d_{\min}/2\sigma] \exp (d_{\min}^2 - d^2)/8\sigma^2, \tag{80}$$

which can be substituted in (76) to obtain the upper bound

$$\begin{aligned} \Pr(E) &\leq \sum_{d \in D} N_d Q[d_{\min}/2\sigma] \exp (d_{\min}^2 - d^2)/8\sigma^2 \\ &= Q[d_{\min}/2\sigma]\{e^{d_{\min}{}^2/8\sigma^2} g_N(e^{-1/8\sigma^2})\}. \end{aligned} \tag{81}$$

As $\sigma \to 0$ the expression in brackets approaches $N_{d_{\min}} = K_1$. Evaluation of this bound involves finding the generating function, which in general can be done through flow-graph techniques [21], illustrated in Appendix I.

The symbol probability of error $\Pr(e)$ may be similarly computed by weighting each error event $\mathscr{E}$ by the number of decision errors $w_H(\mathscr{E})$ it entails:

$$\begin{aligned} \Pr(e) &\leq \sum_{\mathscr{E} \in E} w_H(\mathscr{E}) \Pr(\mathscr{E}) \\ &\leq \sum_{d \in D} Q[d/2\sigma] \sum_{\mathscr{E} \in E_d} w_H(\mathscr{E}) \left[\prod_{i=0}^{n-\nu} \frac{m - |\varepsilon_{xi}|}{m}\right] \\ &\simeq K_2 Q[d_{\min}/2\sigma], \end{aligned} \tag{82}$$

where

$$K_2 \triangleq \sum_{\mathscr{E} \in E_{d_{\min}}} \left[w_H(\mathscr{E}) \prod_{i=0}^{n-\nu} \frac{m - |\varepsilon_{xi}|}{m}\right] \tag{83}$$

is another constant independent of σ. The quantity K_2/K_1 may be interpreted as the average number of symbol errors per error event at high signal-to-noise ratios.

Obviously the average of any variable that is a finite function of error events (e.g., the average length of error events, the bit error probability, etc.) can be calculated in the same way. In each case we can approximate the result by a constant multiplied by $Q[d_{\min}/2\sigma]$ for sufficiently high signal-to-noise ratios. Strict upper bounds like (81) can also be obtained by the flow-graph techniques of [21] as we indicate in Appendix I.

The obvious lower bound shows that these bounds and estimates are very tight. Let $E_{d_{\min}}$ be the set of error events $\varepsilon_x(D)$ of Euclidean weight $d_{\min}$ and let $K_0 \leq 1$ be the probability that the input sequence $x(D)$ will be such that $\hat{x}(D) = x(D) + D^k\varepsilon_x(D)$ is an allowable input sequence for at least one $\varepsilon_x(D) \in E_{d_{\min}}$. When $d_{\min}^2 = \|f\|^2$, $E_{d_{\min}}$ contains $\varepsilon_x(D) = \pm 1$, so $K_0 = 1$. The probability that such an $\hat{x}(D)$ will be closer than $x(D)$ to the received sequence $z(D)$ is $Q[d_{\min}/2\sigma]$. Hence, with probability K_0, the probability of an error event starting at time k for any k is at least $Q[d_{\min}/2\sigma]$, so

$$\begin{aligned} \Pr(E) &\geq K_0 Q[d_{\min}/2\sigma] \\ \Pr(e) &\geq K_0 Q[d_{\min}/2\sigma]. \end{aligned} \tag{84}$$

Thus the upper estimate and lower bound differ only in their constant coefficient. Applications of this lower bound are given in [38].

If the channel were used for only one pulse, i.e., $x_k = 0$ for $k \neq 0$, then intersymbol interference would be absent, the signal sequence would be of the form

$$y(D) = x_0 f(D) \tag{85}$$

and the symbol-error probability would be very nearly

$$\Pr(e) = K_3 Q(\|f\|/2\sigma), \tag{86}$$

where $K_3 = 2(m-1)/m$. This gives a slightly tighter version of the lower bound above when $\|f\|^2 = d_{\min}^2$.

It also suggests that we define the effective signal-to-noise ratio as

$$\mathrm{SNR}_{\mathrm{eff}} \triangleq \sigma_x{}^2 d_{\min}^2/\sigma^2, \tag{87}$$

where again $\sigma_x{}^2 = (m^2 - 1)/12$, for (82) and (86) then show that the probability of error of an m-level PAM system with intersymbol interference differs at most by the ratio K_2/K_3 from that of an m-level system without intersymbol interference and output signal-to-noise ratio $\mathrm{SNR}_{\mathrm{eff}}$. In decibels such a difference is small and goes to zero as

$\mathrm{SNR}_{\mathrm{eff}}$ goes to infinity. But in the common case when $\|f\|^2 = d_{\min}^2$, SNR equals $\mathrm{SNR}_{\mathrm{eff}}$, so that the degradation due to intersymbol interference is negligible. This result, while conjectured in [15], seems effectively unanticipated in the intersymbol interference and partial-response literature. In particular, partial-response techniques had been thought to cost at least 3 dB in output SNR, whereas we see in the following (see also [33]) that for $f(D) = 1 \pm D^n$ the penalty in SNR with the Viterbi algorithm is a small fraction of a decibel.

Upon a little reflection, we are not surprised that when the only constraint is on the output signal-to-noise ratio[1] no degradation need be suffered because of intersymbol interference; for we could simply choose the output sequences $y(D)$ to suit our purposes within the constraint and let $x(D) = y(D)/f(D)$. That the inputs x_k would become very large if $f(D)$ had an unstable inverse might bother us physically, but not mathematically. The surprising result is that under the rigid constraint that the inputs x_k be m equally spaced amplitudes, we can do nearly as well when $\|f\|^2 = d_{\min}^2$.

When $\|f\|^2 > d_{\min}^2$, which tends to happen when intersymbol interference is severe, the ratio $\mathrm{SNR}_{\mathrm{eff}}/\mathrm{SNR}$ is $d_{\min}^2/\|f\|^2$ and measures the degradation due to intersymbol interference. We show in Appendix II that even in this case degradation can be avoided if it is permissible to insert a certain type of partial-response preemphasis filter at the transmitter.

Example: Partial Response

Let $f(D) = 1 - D$. (Everything that follows also holds with the obvious modifications for any partial response of the form $1 \pm D^n$.)

The finite-state machine realizing $f(D)$ has only one m-state memory element. Thus the maximum-likelihood detector needs to keep in mind only m survivors at any time. Fig. 5 shows the trellis representing the state transition diagram spread out in time when $m = 2$, with the associated output signal attached to each branch. Fig. 6 displays the progress of a maximum-likelihood sequence estimator through a typical received sequence. If all zeros were sent, we recognize an error event extending from time 1 through time 4.

It is easy to show that the Lee weight, defined as

$$w_L(\mathscr{E}) \triangleq \sum_{i=0}^{n} |\varepsilon_{yi}| \tag{88}$$

of all signal error sequences $\varepsilon_y(D)$ is even when $f(D) = 1 - D$, hence that $d_{\min}^2 = \|f\|^2 = 2$ and $d^2(\mathscr{E}) \geq 4$ if $d^2(\mathscr{E}) \neq 2$. Consequently only the signal error sequences of Euclidean weight 2 need be considered even for very modest signal-to-noise ratios. By inspection these sequences are the set of sequences of the form $\pm(1 - D^n)$, $n \geq 1$, which result from the input error sequences $\varepsilon_x(D) = \pm(1 + D + \cdots + D^{n-1})$. The error-event probability is then estimated by (77) and (78) as

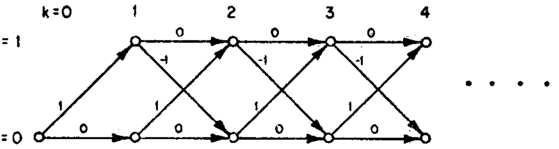

Fig. 5. State-diagram trellis for $m = 2, f(D) = 1 - D$.

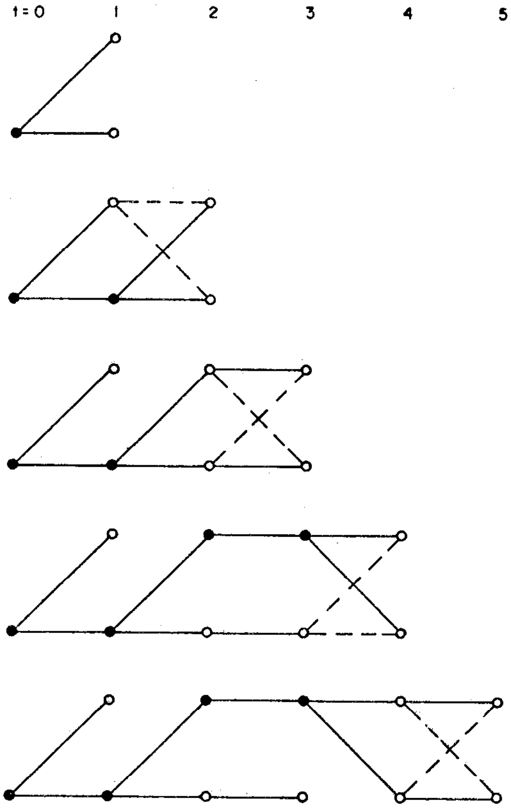

Fig. 6. Survivors of successive recursions of maximum-likelihood sequence estimator.

$$\begin{aligned}\Pr(E) &\simeq Q[1/\sigma\sqrt{2}] \sum_{n=1}^{\infty} 2\left(\frac{m-1}{m}\right)^n \\ &= 2(m-1)Q[1/\sigma\sqrt{2}]. \end{aligned}\tag{89}$$

The probability of symbol error in the estimation of $x(D)$ is, from (82) and (83),

$$\begin{aligned}\Pr(e) &\simeq Q[1/\sigma\sqrt{2}] \sum_{n=1}^{\infty} 2n\left(\frac{m-1}{m}\right)^n \\ &= 2m(m-1)Q[1/\sigma\sqrt{2}]. \end{aligned}\tag{90}$$

As an example of the generating-function approach, we have for $m = 2$

$$\begin{aligned} g_N(z) &= \sum_{d \in D} N_d z^{d^2} \\ &= \frac{2z^2}{1 - z^4} \\ &= 2(z^2 + z^6 + z^{10} + \cdots) \end{aligned}\tag{91}$$

[1] In filter optimization problems one usually constrains the input signal-to-noise ratio; this makes no sense here since the input waveform $x(t)$ is a train of delta functions.

as is shown by flow-graph techniques in Appendix I. This means that there are error events of Euclidean weights 2,6,10,$\cdots$, and that the coefficient of each weight is 2. We obtain the strict upper bound

$$\Pr(E) \leq \frac{2Q[1/\sigma\sqrt{2}]}{1 - e^{-1/2\sigma^2}}. \tag{92}$$

The accuracy of approximating the series (76) by its first term even for σ of the order of 1 is evident.

Precoding is a technique that has been used in partial response systems to prevent infinite error propagation in recovery of the transmitted sequence. As we have seen, with maximum-likelihood sequence estimation error events are quite finite; even so, precoding is useful in reducing symbol-error probability when $m > 2$. The idea [1], [3] is to take the original m-ary sequence, which we now call $d(D)$, and let the input sequence be another m-ary sequence defined by

$$x(D) \triangleq d(D)/f(D) \text{ modulo } m, \tag{93}$$

which is well-defined when $f(D) = 1 \pm D^n$ (or more generally when f_0 and m are relatively prime). Then if we could take the output sequence modulo m we would obtain

$$y(D) = x(D)f(D) \equiv d(D) \text{ modulo } m. \tag{94}$$

In fact, we obtain the estimated data sequence $\hat{d}(D)$ by the zero-memory modulo-m operation on $\hat{y}(D)$:

$$\hat{d}(D) \triangleq \hat{y}(D) \text{ modulo } m. \tag{95}$$

The number of symbol errors in $\hat{d}(D)$ is therefore the same as the number of nonzero coefficients in $\varepsilon_y(D)$. For $f(D) = 1 \pm D^n$, this number is 2 for all the $\varepsilon_y(D)$ of weight 2, which is to say that with precoding all the likely error events result in 2 symbol errors in the estimation of $d(D)$. Hence

$$\Pr(e) \simeq 2\Pr(E) \simeq 4(m-1)Q[1/\sigma\sqrt{2}]. \tag{96}$$

In Fig. 7, we plot the predicted symbol-error probability as a function of output signal-to-noise ratio with $m = 2$ and $m = 4$ for the three cases of 1) no intersymbol interference; 2) a partial response of the class $f(D) = 1 \pm D^n$ with precoding and maximum likelihood sequence estimation; and 3) the same partial response with precoding and with symbol-by-symbol decisions. (Bit-error probability is half the symbol-error probability for $m = 4$, assuming Gray coding.) Simulation results (10 000 symbols) for the maximum-likelihood sequence estimator at low signal-to-noise ratios are also given.

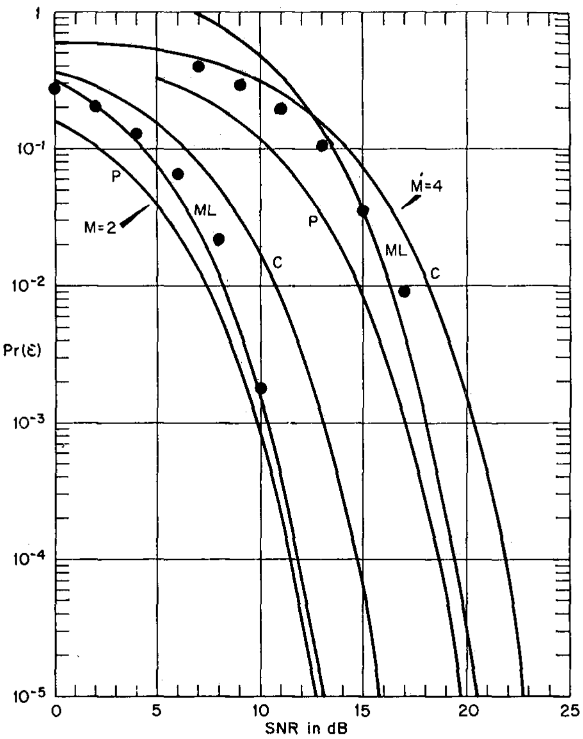

Fig. 7. Probability of error versus output signal-to-noise ratio for $m = 2$ and $m = 4$. P—perfect response; ML—partial response $f(D) = 1 \pm D^n$ with maximum-likelihood sequence estimation and precoding; C—partial response $f(D) = 1 \pm D^n$ with conventional sample-by-sample decisions and precoding; black dots—results of 10 000-sample simulations of ML.

A Practical Algorithm

In this section we show that simple suboptimal approximations to maximum-likelihood sequence estimation can perform nearly as well. We shall describe an algorithm suitable for the class of partial responses $f(D) = 1 \pm D^n$, with $f(D) = 1 - D$ again the particular illustrative example. We shall introduce the algorithm from a different, more concrete viewpoint than before, corresponding to the way it was actually invented; such an explanation may appeal to the more practical reader more than the previous abstractions. The scheme involves making tentative decisions that are then corrected by a single-error-detecting and correcting decoder. We shall show in the end that the two schemes perform nearly identically.

With m-level inputs x_k, the output signals $y_k = x_k - x_{k-1}$ take on $2m - 1$ different levels. Tentative decisions $\hat{y}_k$ may be made from the noisy outputs $z_k = y_k + n_k$ as to which output signal level is most probable. Since there are more $(2m-1)$-ary sequences than m-ary sequences, certain sequences of tentative decisions can be recognized as impossible, in the sense that no allowable input sequence could have caused them. That this redundancy can be used to detect errors in the tentative decisions has been recognized by previous authors [1], [2], [39].

A general method of determining whether the tentative decision sequence $\hat{y}(D)$ is allowable is to pass it through an inverse linear filter with impulse response $1/f(D)$ and see whether an allowable input sequence $\hat{x}(D)$ comes out. In this case such a filter is realized by the feedback circuit in which $\hat{x}_k = \hat{y}_k + \hat{x}_{k-1}$, illustrated in Fig. 8.

Whenever a single error is made, say $\hat{y}_k = y_k + 1$, the output $\hat{x}_k$ at that time will be one unit higher than the corresponding input x_k. Because of the feedback, the error continues to propagate in the circuit and raises all subsequent $\hat{x}_k$ by one unit as well. At the first time $k' \geq k$ for which $x_{k'} = m - 1$, we will observe the unallowable level $\hat{x}_{k'} = m$ and detect the error. At this time we can reset $\hat{x}_{k'}$ to $m - 1$ and error propagation ceases. Similarly, negative errors are detected when $\hat{x}_{k'} < 0$. To frustrate the

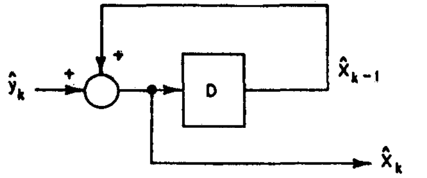

Fig. 8. Inverse linear filter with response $1/(1 - D)$.

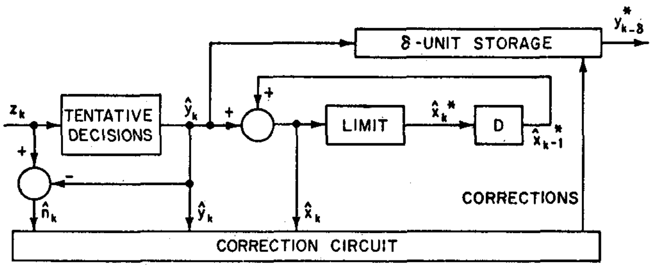

Fig. 9. Error-correction algorithm block diagram.

detection circuit takes a second error in the opposite direction; all single errors are eventually detected.

The error-detection circuit yields additional information beyond simply detecting the error. First, the direction of the excursion obviously tells us the polarity of the error. Second, observe that whenever $\hat{x}_k = 0$ it is certain that no positive error is circulating in the circuit, since that would imply $x_k < 0$; similarly $\hat{x}_k = m - 1$ implies no negative error. When we detect a positive error we can therefore be sure that it occurred in the finite time span since the last time $\hat{x}_k = 0$ and likewise for a negative error.

To localize the error in this time span requires information about the reliability of each tentative decision. Let the apparent error $\hat{n}_k$ be defined as $\hat{n}_k = z_k - \hat{y}_k$; for any reasonable noise distribution the tentative decision most likely to be in error is that for which $\hat{n}_k$ has the largest magnitude with the appropriate polarity. We therefore correct in that place. (The scheme resembles Wagner decoding of a distance-2 block code [40]. It is an improvement over the null-zone scheme of Smith [41], to which it bears the same relation as Wagner decoding does to single-erasure-correction with block codes.)

Fig. 9 shows the circuit that generates $\hat{y}_k$, $\hat{n}_k$, and $\hat{x}_k$, and stores tentative decisions for δ time units awaiting correction. Fig. 10 gives the decision logic in flow-chart form; we use four storage registers NMIN, KMIN, NMAX, and KMAX to store the magnitude and location of the largest positive and negative apparent errors. To implement an equivalent algorithm for $m = 2$ or 4 with $f(D) = 1 - D^2$ with modestly integrated resistor–transistor logic (RTL) circuits (2 flip/flops per IC) took about 50 integrated circuits, including all control logic.

To determine how long the storage time δ should be, we note that the probability that an error at time k will not have been detected before time $k + \tau$ is $[1 - (1/m)]^\tau$, the probability that τ consecutive input symbols are not equal to $m - 1$ or 0, as the case may be. For $m = 4$ and $\delta = 20$, $[1 - (1/m)]^\delta = 0.003$; thus if $\delta = 20$, one out of 300 errors will not be detected in time. This is satisfactory if and only if we are looking for not more than about two orders of magnitude of error rate improvement. That final decisions must be deferred for so long even though ν equals one may be surprising; yet any scheme with shorter delay is condemned to significant suboptimality.

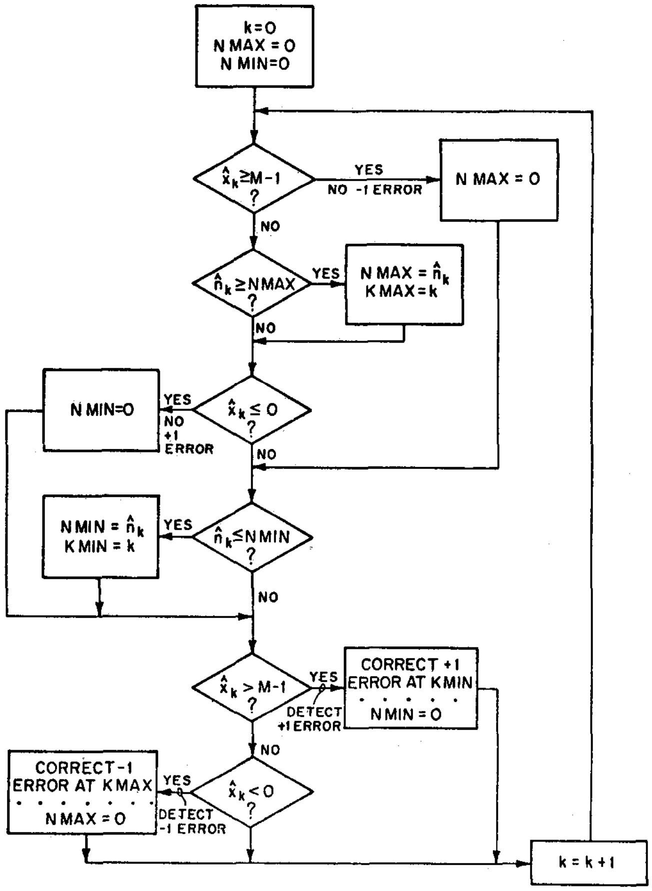

Fig. 10. Flow chart for error correction with partial response $f(D) = 1 - D$.

We now show the equivalence of this simplified algorithm to maximum-likelihood sequence estimation under two assumptions.

Assumption 1: No noise of magnitude $|n_k| \geq 1$ occurs.

Assumption 2: After any single tentative decision error $[|n_k| \geq (\frac{1}{2})]$, no second error occurs before the error is detected.

Under these assumptions a decoding error can occur in the simplified algorithm only if for some k, the following are true.

1) $n_k \geq (\frac{1}{2})$ and $n_{k'} \leq n_k - 1$ for some k' such that a) if $k' < k$, $x_j \neq 0$, $k' \leq j < k$; or b) if $k < k'$, $x_j \neq m - 1$, $k \leq j < k'$; that is, if $x_{k'}$ falls within the time span in which the error could have occurred when the error in x_k is detected.

2) Similarly, if $n_k \leq -(\frac{1}{2})$ and $n_{k'} \geq 1 + n_k$ for some k' such that a) if $k' < k$, $x_j \neq m - 1$, $k' \leq j < k$; or b) if $k < k'$, $x_j \neq 0$, $k \leq j < k'$.

In Fig. 11 we plot the error-causing regions in the $n_k - n_{k'}$ plane, as well as the regions excluded by Assumptions 1 and 2. By comparison with Fig. 4 we see that a maximum-likelihood sequence estimator would make exactly the same decisions.

Conversely, under Assumptions 1 and 2, a maximum-likelihood sequence estimator has at most three survivors at any time, regardless of the number of states: the state sequence corresponding to the tentative decisions $\hat{y}(D)$ (if allowable), that corresponding to a single positive error in the most unreliable past symbol $[\hat{y}^+(D)]$, and that to a single negative error $[\hat{y}^-(D)]$. Fig. 12 shows a typical sequence for $m = 4$; the reduction to a single survivor at time 4 corresponds to an error correction in our simplified algorithm. Since any two sequences containing a tentative decision error differ in only one place from the tentative decision survivor and since likelihoods of two competing paths can be compared on the basis of the magnitude of $\hat{n}_k$ in the single-error place, each single-error survivor can be completely identified by the location and magnitude of its sole apparent error, a fact of which we have taken advantage in our algorithm.

It follows that the probability of error of our simplified algorithm is bounded by the probability of error for maximum-likelihood sequence estimation plus the probability that Assumption 1 or 2 does not hold. The probability of Assumption 1 is $2Q[1/\sigma]$, and of Assumption 2 $K_4(Q[1/2\sigma])^2$, where K_4 is again a constant. In the region of error probabilities of the order of 10^{-3} to 10^{-5}, $Q[1/\sigma]$ is two or more orders of magnitude less than $Q[1/\sigma\sqrt{2}]$ and $(Q[1/2\sigma])^2$ an order of magnitude less, so that the probability that Assumption 1 or 2 does not hold is much less than the probability of error for maximum-likelihood sequence estimation. Hence our simplified algorithm gives effectively the same performance.

The algorithm has been incorporated in a commercially available 9600 bit/s telephone-line modem with 4800 inputs/s, $m = 2$ or 4, and a partial response $f(D) = 1 - D^2$. Despite the quite non-Gaussian and nonwhite character of telephone line disturbances, performance improvements rather similar to the predictions of Fig. 7 have been observed—for error rates in the 10^{-3}–10^{-5} range, we typically see an order of magnitude or so improvement in error rate with a tendency to less improvement at the higher error rates and more at the lower.

Conclusion

We have shown that a maximum-likelihood sequence estimator for a PAM sequence perturbed by finite intersymbol interference and white Gaussian noise can be constructed from a whitened matched filter and the Viterbi algorithm. The combination is simpler to implement than previous "optimum" nonlinear algorithms and is practically feasible if the channel impulse response is not too long. Its performance can be accurately estimated and is shown to be effectively as good as can be attained by any estimator, regardless of the criterion of optimality. Furthermore, in many cases performance is effectively as good as if intersymbol interference were absent.

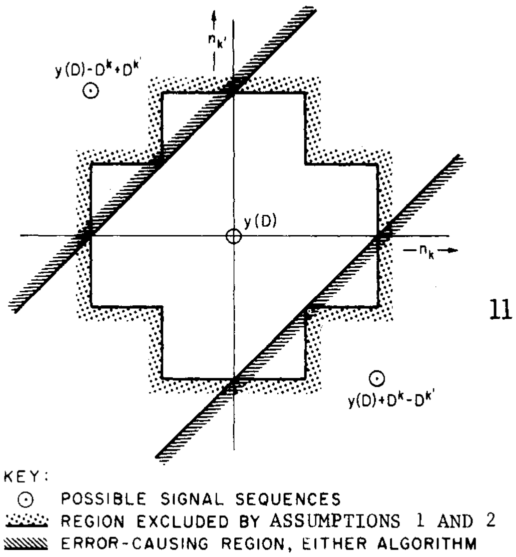
$n_{k'}$
y(D)-D^k+D^k'
y(D)
$-n_k$→
11
y(D)+D^k-D^k'
KEY:
⊙ POSSIBLE SIGNAL SEQUENCES
▨ REGION EXCLUDED BY ASSUMPTIONS 1 AND 2
▧ ERROR-CAUSING REGION, EITHER ALGORITHM

Fig. 11. Decision regions.

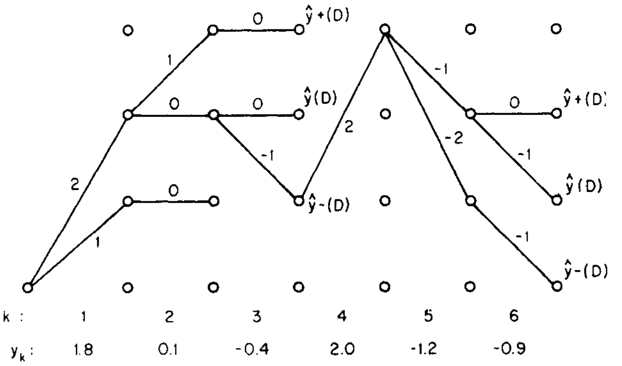

Fig. 12. Maximum-likelihood sequence estimation algorithm under Assumptions 1 and 2.

We have shown that for impulse responses of the type that arise in partial response systems a simple error-correction algorithm approximates the performance of a maximum-likelihood sequence estimator and gives a 3-dB gain in effective signal-to-noise ratio over symbol-by-symbol decisions. The construction of similar practical suboptimum algorithms for other simple responses ought to be a fruitful activity.

In a practical situation, a near-optimum procedure is to use a linear equalizer to shape the channel to some desired channel whose impulse response $f(D)$ is short and whose spectrum is similar to the channel spectrum and then use a Viterbi algorithm that is appropriate for $f(D)$. The desired response $f(D)$ should at least reflect the nulls or near-nulls of the channel's Nyquist spectrum since these cannot be linearly equalized without excessive noise enhancement. From Lemma 1, a channel with ν nulls or near-nulls is approximately equal to the cascade of an equalizable (null-free) channel and a transversal filter characterized by a polynomial $f(D)$ of degree ν. For example, if there is severe attenuation only at the band edges, it is reasonable to choose $f(D) = 1 - D^2$. Analysis of how close one can come to optimal performance with this approach would be useful.

Another practical problem is that the channel response $h(t)$ is usually unknown, so that the receiver must be adaptive. In the approach of the previous paragraph, one can obviously make the linear equalizer the adaptive element. It would be of interest, however, to devise an adaptive version of the maximum-likelihood sequence estimator itself.

On the theoretical side, the greatest deficiency in our results is their reliance on a finite channel response. It can be shown that the brute-force approach of approximating an infinite response by a truncated response of length L and then using the appropriate sequence estimator gives performance that is accurately estimated by (77) and (82) as $L \to \infty$, where $d_{\min}^2$ is still defined as min $\|\varepsilon_x(D)f(D)\|^2$ for the true (infinite) $f(D)$. There must be a better way, however, of dealing with simple infinite responses like $h(t) = e^{-t/\tau}$, $t \geq 0$.

These results can be extended in a number of directions. Extension to quadrature PAM, where phase as well as amplitude is modulated, is achieved straightforwardly by letting the input sequence $x(D)$ be complex, although the analysis is slightly complicated by having to deal with complex $\varepsilon_x(D)$. Colored noise can be handled by prewhitening. The possibility of extensions to handle context-dependent $x(D)$ and other Markov-modelable constraints has already been mentioned. Finally, the use of similar techniques outside digital communications (for example, in magnetic-tape reading [32]) will no doubt be extensive.

Acknowledgment

It is a pleasure to acknowledge the kind interest and helpful comments of J. L. Massey and R. Price.

Appendix I

Determining Weight Distributions

The weight distribution of a particular impulse response is characterized by the generating function

$$g_N(z) = \sum_{d \in D} N_d z^{d^2}, \tag{97}$$

where

$$N_d = \sum_{\mathscr{E} \in E_d} \left[\prod_{i=0}^{n-\nu} \frac{m - |\varepsilon_{xi}|}{m} \right] \tag{98}$$

is the sum over all error events of Euclidean weight d^2 of the probabilities that the corresponding input error sequences $\varepsilon_x(D)$ are allowable. In this Appendix we indicate how a flow graph can be associated with an impulse response such that the transfer function of the flow graph is this generating function.

We have already seen the usefulness of considering the input error sequence $\varepsilon_x(D)$ and the signal error sequence $\varepsilon_y(D) = \varepsilon_x(D)f(D)$ associated with an error event $\mathscr{E}$. Similarly we can define (for finite impulse responses) the state error sequence $\varepsilon_s(D)$ by

$$\begin{aligned}\varepsilon_{st} &= s_{k_1+t} - \hat{s}_{k_1+t} \\ &= (x_{k_1+t-1} - \hat{x}_{k_1+t-1}, x_{k_1+t-2} - \hat{x}_{k_1+t-2}, \cdots, x_{k_1+t-\nu} - \hat{x}_{k_1+t-\nu}) \\ &= (\varepsilon_{x,t-1}, \varepsilon_{x,t-2}, \cdots, \varepsilon_{x,t-\nu}).\end{aligned} \tag{99}$$

Clearly, from the linearity of PAM, the input errors ε_{xt} govern the state-error transitions and the signal errors ε_{yt} are functions of ε_{xt} and ε_{st}, or equivalently of ε_{st} and $\varepsilon_{s,t+1}$.

An error event starts and ends with an all-zero state error. From the input error sequence $\varepsilon_x(D)$ we can determine the path taken through the nonzero state error sequences during the error event. Each possible error event thus corresponds to a unique such path.

Let us then construct a flow diagram in which the initial node is the all-zero state error, the intermediate nodes are the nonzero state errors, and the final node is again the all-zero state error. Let us draw in all possible state-error transitions and label each with the corresponding ε_{xt} and ε_{yt}. To each transition we assign a weight or transfer function equal to

$$\frac{m - |\varepsilon_{xt}|}{m} z^{\varepsilon_{yt}^2}, \tag{100}$$

where z is an indeterminate. Then with any particular path (error event) from the initial node to the final node is associated the transfer function

$$\left[\prod_{i=0}^{n-\nu} \frac{m - |\varepsilon_{xt}|}{m} \right] z^{\Sigma \varepsilon_{yt}^2}. \tag{101}$$

We recognize $\Sigma\, \varepsilon_{yt}^2$ as $d^2(\mathscr{E})$, the Euclidean weight of the error event. Hence the total transfer function of the flow graph, which is the sum of the transfer functions of all paths, is simply

$$\sum_{\mathscr{E} \in E} \left[\prod_{i=0}^{n-\nu} \frac{m - |\varepsilon_{xt}|}{m} \right] z^{d^2(\mathscr{E})} = g_N(z) \tag{102}$$

the desired generating function.

In simple cases, one can solve the flow graph fairly easily to obtain an explicit expression for $g_N(z)$. In more complicated cases, one may merely solve the flow graph modulo z^n for small values of n to determine the coefficients N_d for $d^2 < n$; or one may solve for particular real number values of z to determine the tightness of the asymptotic expressions.

Example 1

Let $f(D) = 1 - D$ and $m = 2$. Then there are only two possible nonzero state errors, $\varepsilon_{st} = \pm 1$. The flow graph is illustrated in Fig. 13. By symmetry the transfer functions from the initial node to each of the nonzero nodes is the same, say $q(z)$. We then have

$$q(z) = \tfrac{1}{2}z + \tfrac{1}{2}(1 + z^4)q(z), \tag{103}$$

which yields

$$q(z) = \frac{z}{1 - z^4}. \tag{104}$$

The transfer function from the initial node to the final node is then

$$g_N(z) = 2zq(z) = \frac{2z^2}{1 - z^4}. \tag{105}$$

Example 2

Let $f(D) = 1 - D$ and $m = 4$. Then there are six nonzero state errors, $\varepsilon_{st} = \pm 1, \pm 2, \pm 3$. Again by symmetry the transfer functions to states differing only in sign are equal, so we need solve for three transfer functions $q_1(z)$, $q_2(z)$, $q_3(z)$. These satisfy the equations

$$\begin{bmatrix} q_1(z) \\ q_2(z) \\ q_3(z) \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix} \left\{ \begin{bmatrix} 1 + z^4 & z + z^9 & z^4 + z^{16} \\ z + z^9 & 1 + z^{16} & z + z^{25} \\ z^4 + z^{16} & z + z^{25} & 1 + z^{36} \end{bmatrix} \times \begin{bmatrix} q_1(z) \\ q_2(z) \\ q_3(z) \end{bmatrix} + \begin{bmatrix} z \\ z^4 \\ z^9 \end{bmatrix} \right\}. \tag{106}$$

Finally, $g_N(z)$ is obtained from

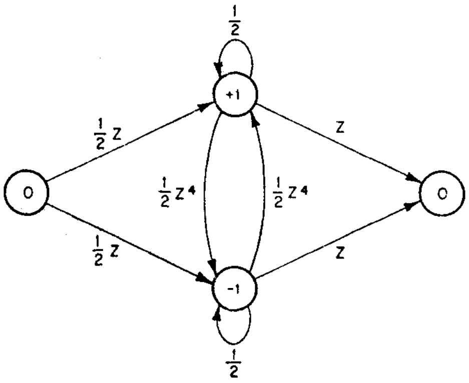

Fig. 13. Flow graph for $f(D) = 1 - D$ and $m = 2$.

$$g_N(z) = 2[zq_1(z) + z^4q_2(z) + z^9q_3(z)]. \tag{107}$$

Solving these equations explicitly involves inversion of a 3 × 3 polynomial matrix. Evaluation of upper bounds for particular values of σ involves substitution of $z = \exp(-1/8\sigma^2)$ and inversion of a 3 × 3 real matrix. The first few terms of $g_N(z)$ can easily be determined recursively by hand to give

$$g_N(z) = 6z^2 + 18z^4 + 90z^6 + 362z^8 + \cdots. \tag{108}$$

Through similar techniques we can obtain flow graphs and generating functions for the bit probability of error, average length of error events, and so forth. The general method is to take the same flow graph and multiply each individual transition transfer function by z_1^{μ}, where μ is the increment to the function $f(\mathcal{E})$ whose average is being evaluated and z_1 is a second indeterminate. For instance, for symbol error probability without precoding we would let $\mu = 1$ if $|\varepsilon_{xi}| \neq 0$ and $\mu = 0$ otherwise, for symbol-error probability with precoding we would let $\mu = 1$ if $|\varepsilon_{yi}| \neq 0$ and $\mu = 0$ otherwise, and for error event length we would let $\mu = 1$ for all transitions. We would then obtain from the flow graph a generating function

$$g(z,z_1) = \sum_{\mathcal{E} \in E} N(d,f) z^{d^2(\mathcal{E})} z_1^{f(\mathcal{E})}, \tag{109}$$

where $N(d,f)$ is the coefficient of the error events of Euclidean weight $d^2(\mathcal{E})$ and f-weight $f(\mathcal{E})$. The partial derivative of this function with respect to z_1 evaluated at $z_1 = 1$,

$$\left.\frac{\partial g(z,z_1)}{\partial z_1}\right|_{z_1=1} = \sum_{\mathcal{E} \in E} fN(d,f) z^{d^2(\mathcal{E})} \tag{110}$$

can be used to obtain upper bounds on

$$\bar{f} = \sum_{\mathcal{E} \in E} fN(d,f) Q[d/2\sigma], \tag{111}$$

be $\bar{f}$ the bit error probability, average length of time in error events, or whatever.

Appendix II

Improving SNR by Preemphasis

Partial-response filters $p(D)$ are finite polynomials with integer-valued coefficients. Suppose that it is permissible to pass the input sequence through a filter of this type before transmission over the channel. Then the received-signal sequence is given by

$$y(D) = x(D)p(D)f(D). \tag{112}$$

The input signal power increases by a factor of $\|p\|^2$, but the output SNR is changed by a factor $\|p(D)f(D)\|^2/\|f\|^2$, which may be less than one.

In particular, consider the case where $d^2_{\min} < \|f\|^2$. Since $d^2_{\min} = \min \|\varepsilon_x(D)f(D)\|^2$, there is some input error sequence $\varepsilon_x(D)$ such that $d^2(\mathcal{E}) = d^2_{\min}$. This will be a finite polynomial with integer-valued coefficients and hence can serve as our partial response filter $p(D)$. Then the output SNR will decrease to $\text{SNR}_{\text{eff}} = \sigma_x{}^2 d^2_{\min}/\sigma^2$. The probability of error can only be improved, since any signal error sequence $\varepsilon_x(D)p(D)f(D)$ possible on this channel is also possible on the original channel, because all multiples of $p(D)f(D)$ are also multiples of $f(D)$. Hence by use of the preemphasis filter we can make the output signal-to-noise ratio equal to SNR_{eff} and thus avoid any significant degradation due to intersymbol interference.

For example, the partial response $f(D) = 1 + 2D + D^2$ has been suggested [2] to obtain a second-order null at the upper Nyquist band edge. Here $\|f\|^2 = 6$, but for $\varepsilon_x(D) = 1 - D$, $\|\varepsilon_x(D)f(D)\|^2 = 4$. The partial response $\varepsilon_x(D)f(D) = 1 + D - D^2 - D^3$ thus gives an improved output SNR as well as a null at DC, which will generally be an additional asset.

Another striking result is obtained when the set $E_{d_{\min}}$ of error events of weight $d^2_{\min}$ is finite. Then there will be at least one signal-error polynomial $\varepsilon_y(D) = \varepsilon_x(D)f(D)$ of weight $d^2_{\min}$ that does not divide any other such polynomial, except trivially $-\varepsilon_y(D)$. Hence if we use the corresponding $\varepsilon_x(D)$ as a preemphasis filter, there will be only two error events of weight $d^2_{\min}$ on the new channel, namely those with $\varepsilon_x(D) = \pm 1$. [This follows as before, from the multiples of $\varepsilon_x(D)f(D)$ being a subset of the multiples of $f(D)$.] Therefore, at moderate-to-high signal-to-noise ratios on the preemphasized channel, both Pr (e) and Pr (E) are approximated by

$$\Pr(e) \simeq \Pr(E) \simeq 2(1 - (1/m))Q[d_{\min}/2\sigma] \tag{113}$$

exactly the same as when intersymbol interference is absent, down to the coefficient $K_3 = 2(m - 1)/m$.

References

[1] A. Lender, "Correlative digital communication techniques," *IEEE Trans. Commun. Technol.*, vol. COM-12, pp. 128–135, Dec. 1964.

[2] E. R. Kretzmer, "Generalization of a technique for binary data communication," *IEEE Trans. Commun. Technol.* (Concise Papers), vol. COM-14, pp. 67–68, Feb. 1966.

[3] A. M. Gerrish and R. D. Howson, "Multilevel partial-response signalling," in *IEEE Int. Conf. Communications Rec.*, Minneapolis, Minn., June 1967, p. 186.

[4] C. W. Helstrom, *Statistical Theory of Signal Detection.* New York: Pergamon, 1960, sect. IV.5.

[5] D. A. George, "Matched filters for interfering signals," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-11, pp. 153–154, Jan. 1965.

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

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

[8] T. Berger and D. W. Tufts, "Optimum pulse amplitude modulation, part I: Transmitter–receiver design and bounds from information theory," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 196–208, Apr. 1967.

[9] R. W. Lucky, J. Salz, and E. J. Weldon, Jr., *Principles of Data Communication.* New York: McGraw-Hill, 1968, ch. 5.

[10] J. G. Proakis and J. H. Miller, "An adaptive receiver for digital signaling through channels with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 484–497, July 1969.

[11] T. Ericson, "Structure of optimum receiving filters in data transmission systems," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-17, pp. 352–353, May 1971.

[12] R. W. Chang and J. C. Hancock, "On receiver structures for channels having memory," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 463–468, Oct. 1966.

[13] K. Abend, T. J. Harley, Jr., B. D. Fritchman, and C. Gumacos, "On optimum receivers for channels having memory," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-14, pp. 819–820, Nov. 1968.

[14] R. R. Bowen, "Bayesian decision procedure for interfering digital signals," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-15, pp. 506–507, July 1969.

[15] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, vol. 58, pp. 779–785, May 1970.

[16] C. G. Hilborn, Jr., "Applications of unsupervised learning to problems of digital communication," in *Proc. 9th IEEE Symp. Adaptive Processes, Decision, and Control*, Dec. 7–9, 1970; also C. G. Hilborn, Jr., and D. G. Lainiotis, "Optimal un-

supervised learning multicategory dependent hypotheses pattern recognition," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 468–470, May 1968.

[17] M. E. Austin, "Decision-feedback equalization for digital communication over dispersive channels," M.I.T. Lincoln Lab., Lexington, Mass., Tech. Rep. 437, Aug. 1967; also Sc.D. thesis, Massachusetts Inst. Technol., Cambridge, May 1967.

[18] D. A. George, R. R. Bowen, and J. R. Storey, "An adaptive decision feedback equalizer," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 281–293, June 1971.

[19] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 260–269, Apr. 1967.

[20] G. D. Forney, Jr., "Review of random tree codes," NASA Ames Res. Cen., Moffett Field, Calif., Contr. NAS2-3637, NASA CR 73176, Final Rep., Appendix A, Dec. 1967.

[21] A. J. Viterbi, "Convolutional codes and their performance in communication systems," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 751–772, Oct. 1971.

[22] G. J. Minty, "A comment on the shortest-route problem," *Oper. Res.*, vol. 5, p. 724, Oct. 1957.

[23] M. Pollack and W. Wiebenson, "Solutions of the shortest-route problem—A review," *Oper. Res.*, vol. 8, pp. 224–230, Mar. 1960.

[24] O. Wing, "Algorithms to find the most reliable path in a network," *IRE Trans. Circuit Theory* (Corresp.), vol. CT-8, pp. 78–79, Mar. 1961.

[25] S. C. Parikh and I. T. Frisch, "Finding the most reliable routes in communication systems," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 402–406, Dec. 1963.

[26] R. Busacker and T. Saaty, *Finite Graphs and Networks: An Introduction with Applications.* New York: McGraw-Hill, 1965.

[27] Y. S. Fu, "Dynamic programming and optimum routes in probabilistic communication networks," in *IEEE Int. Conv. Rec.*, pt. 1, pp. 103–105, Mar. 1965.

[28] J. K. Omura, "On the Viterbi decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 177–179, Jan. 1969.

[29] J. K. Omura, "On optimum receivers for channels with intersymbol interference," (Abstract), presented at the IEEE Int. Symp. Information Theory, Noordwijk, Holland, June 1970.

[30] J. K. Omura, "Optimal receiver design for convolutional codes and channels with memory via control theoretic concepts," unpublished.

[31] H. Kobayashi and D. T. Tang, "On decoding and error control for a correlative level coding system," (Abstract), presented at the IEEE Int. Symp. Information Theory, Noordwijk, Holland, June 1970.

[32] H. Kobayashi, "Application of probabilistic decoding to digital magnetic recording systems," *IBM J. Res. Develop.*, vol. 15, pp. 64–74, Jan. 1971.

[33] H. Kobayashi, "Correlative level coding and maximum-likelihood decoding," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 586–594, Sept. 1971.

[34] T. N. Morrissey, Jr., "A unified analysis of decoders for convolutional codes," Univ. Notre Dame, Notre Dame, Ind., Tech. Rep. EE-687, Oct. 1968, ch. 7; also T. N. Morrissey, Jr., "Analysis of decoders for convolutional codes by stochastic sequential machine methods," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 460–469, July 1970.

[35] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering.* New York: Wiley, 1965, ch. 4.

[36] B. H. Bharucha and T. T. Kadota, "On the representation of continuous parameter processes by a sequence of random variables," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 139–141, Mar. 1970.

[37] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York: Wiley, 1968, problems 2.6.17–18.

[38] G. D. Forney, Jr., "Lower bounds on error probability in the presence of large intersymbol interference," *IEEE Trans. Commun. Technol.* (Corresp.), vol. COM-20, pp. 76–77, Feb. 1972.

[39] J. F. Gunn and J. A. Lombardi, "Error detection for partial-response systems," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 734–737, Dec. 1969.

[40] R. A. Silverman and M. Balser, "Coding for constant-data-rate systems—Part I. A new error-correcting code," *Proc. IRE*, vol. 42, pp. 1428–1435, Sept. 1954.

[41] J. W. Smith, "Error control in duobinary systems by means of null zone detection," *IEEE Trans. Commun. Technol.*, vol. COM-16, pp. 825–830, Dec. 1968.

Rate-Distortion Theory for Context-Dependent Fidelity Criteria

TOBY BERGER, MEMBER, IEEE, AND WENG C. YU

***Abstract*—A lower bound $R_L(D)$ is obtained to the rate-distortion function $R(D)$ of a finite-alphabet stationary source with respect to a context-dependent fidelity criterion. For equiprobable memoryless sources and modular distortion measures, $R(D) = R_L(D)$ for all D. It is conjectured that, for a broad class of finite-alphabet sources and context-dependent fidelity criteria, there exists a critical distortion $D_c > 0$ such that $R(D) = R_L(D)$ for $D \leq D_c$.**

The case of a binary source and span-2 distortion measure is treated in detail. Among other results a coding theorem is proved that establishes that $R(0) = \log(2/r_g)$, where r_g is the golden ratio, $(1 + \sqrt{5})/2$.

I. Introduction

INFORMATION transmission systems usually are designed and analyzed with total disregard for the fact that the seriousness of transmission errors depends critically upon the context in which they occur. For example, reproducing a 3 as a 7 tends to be much more serious in the context 368 → 768 than in the context 863 → 867. Also, it usually is more difficult to detect and correct errors in context when several errors occur close together than when they are more widely dispersed. The minimum capacity necessary to transmit data at a tolerable level of distortion often can be reduced appreciably if the system is designed with the appropriate context-dependent fidelity criterion in mind.

With a view toward taking context into account, Shannon [1] introduced local distortion measures defined as follows. Let the information source produce a sequence of symbols from an alphabet A_M containing M distinct letters. Without loss of generality, we hereafter take $A_M = \{0,1,\cdots M-1\}$. Any function $\rho_g: A_M{}^g \times A_N{}^g \rightarrow [0,\infty)$, where N need not necessarily equal M, is called a local distortion measure of span g. The number $\rho_g(x,y)$ is the penalty, or distortion,

Manuscript received April 21, 1971; revised July 7, 1971. This research was supported in part by NSF Grant GK-14449.

The authors are with the School of Electrical Engineering, Cornell University, Ithaca, N.Y. 14850.