Reduced-State Sequence Estimation with Set Partitioning and Decision Feedback

M. VEDAT EYUBOĞLU, MEMBER, IEEE, AND SHAHID U. H. QURESHI, FELLOW, IEEE

Abstract—**A reduced-state sequence estimator for linear intersymbol interference (ISI) channels is described. The estimator uses a conventional Viterbi algorithm (VA) with decision feedback to search a reduced-state "subset trellis" which is constructed using set partitioning principles. The complexity of maximum likelihood sequence estimation (MLSE) due to the length of the channel memory and the size of the signal set is systematically reduced. An error probability analysis shows that a good performance/complexity tradeoff can be obtained. In particular, our results indicate that the required complexity to achieve the performance of MLSE is independent of the size of the signal set for large enough signal sets. Simulation results are provided for two partial-response systems. In addition, we describe a simple technique for quadrature partial-response signaling (QPRS) that eliminates the "quasi-catastrophic" nature of the ML trellis.**

I. INTRODUCTION

In this paper, we are concerned with linearly modulated uncoded data transmission systems subject to severe intersymbol interference (ISI). It is well known that in such systems, maximum likelihood sequence estimation (MLSE) [1], implemented with the Viterbi algorithm (VA) [2], can provide a significant improvement in detection performance compared to the decision-feedback equalizer (DFE). However, in general, the implementation complexity of MLSE is roughly M^K times that of a DFE where K is the length of the overall channel impulse response and M is the size of the signal set.

A considerable amount of research has been undertaken to achieve the performance of the MLSE at reduced complexity [3]–[11]. Most of the earlier work concentrated on preprocessing techniques to reduce the channel impulse response to a shorter length [3]–[6]. In [3], a linear equalizer was used to force the overall impulse response to a desired shape of short length. Falconer and Magee [4] and Beare [5] investigated the optimization of the desired impulse response for a fixed given length. Later, Lee and Hill [6] proposed the use of a DFE to truncate the channel impulse response, in an effort to reduce noise enhancement in the linear equalizer. Recently, Duel and Heegard [11] incorporated the decision-feedback mechanism into the sequence estimator to increase the overall reliability of the feedback decisions. As we will see, this estimator (which was developed independently) is a special case of the algorithm that will be presented in this paper. We will refer to it as a decision-feedback sequence estimator (DFSE).

For bandwidth-efficient modulation systems which use large signal sets, reducing the length of the channel impulse response alone is often not sufficient to reduce complexity. In MLSE, the VA searches a trellis with M^K states (this trellis will be called an ML trellis), and therefore has to keep track of M^K paths. For large values of M, the complexity can be large, even for very small K. The works of Vermuelen and Hellman [7] and Foschini [8] have shown, however, that the performance of MLSE can be achieved by following a smaller number of the more likely paths. More recently, Clark *et al.* [9], [10] and Wesolowski [19] have described interesting sequence estimation algorithms which demonstrate the potential of this approach.

In contrast to the algorithms presented in [9], which can be characterized as ad hoc, the algorithms to be described in this paper involve a highly structured reduced-state sequence estimator (RSSE) which can achieve nearly the performance of MLSE at significantly reduced complexity (RSSE was first reported in [20]). The RSSE structure is very general and can be used for modulation formats with large alphabet size and/or for channels with long memory. The primary idea is the construction of trellises with a reduced number of states. These states are formed by combining the states of the ML trellis using Ungerboeck-like set partitioning principles [12], [13]. The RSSE is then implemented using the VA to search this reduced-state trellis. It has the following characteristics.

1) Good Performance/Complexity Tradeoff: This is achieved by a) retaining the structure of MLSE to preserve ease of implementation, b) using Ungerboeck-like set partitioning in constructing the reduced-state trellis to reduce performance degradation, and c) using built-in decision feedback in branch metric computations. In contrast to the Lee and Hill [6] and Wesolowski [19] schemes, this RSSE does not utilize a DFE, and therefore does not suffer from the relatively poor performance of DFE on severely distorted channels. Analysis and simulations indicate that substantial complexity reduction can be obtained with little loss in performance. In particular, we will show that quadrature $1 \pm D$ partial-response systems can be decoded using only a two-state trellis with only a very small loss relative to the MLSE, independent of the size of the signal set; further, we will show that binary $(1 \pm D)^2$ partial-response systems can be decoded using a two-state trellis with 4.77 dB better performance than a zero-forcing DFE.

2) Analyzability: In contrast to most earlier RSSE work, analysis of the performance of our estimator is as straightforward as that of MLSE, when the effects of error propagation are neglected. We will show that the first error event probability can be accurately characterized by a minimum distance parameter $d'_{\min}$ which can be related to the minimum (free) distance $d_{\min}$ of MLSE.

3) Flexibility: The performance and complexity of RSSE are controlled by K parameters. By choosing them appropriately, we can obtain a tradeoff between desired performance and complexity, ranging between that of a zero-forcing DFE and that of MLSE.

It is well known that for $1 \pm D^N$ partial-response signaling, MLSE can achieve the same effective minimum distance as ISI-free transmission. However, as the signal set size is increased, the performance of MLSE deteriorates. This fact is true even with precoding. In fact, as we will show, MLSE for a $1 + D$ partial-response system with $M = 16$ and precoding performs about 1 dB worse than ISI-free transmission. This

Paper approved by the Editor for Digital Communications of the IEEE Communications Society. Manuscript received January 9, 1987; revised July 10, 1987. This paper was presented in part at GLOBECOM'86, Houston, TX, December 1986.

The authors are with the Codex Corporation, Mansfield, MA 02048.

IEEE Log Number 8718270.

phenomenon has frequently been overlooked. Recently, it was also observed by Forney and Calderbank [17] who correctly connect it to the fact that the ML trellis in these systems is what they call quasi-catastrophic. In Section IV, we will discuss this phenomenon, and then show that in $1 \pm D^N$ quadrature partial-response signaling (QPRS), it can be eliminated by a method that can be characterized as a shift of the carrier frequency with respect to the center frequency of the transmitted signal.

The paper is organized as follows. In Section II, we briefly describe the system model. Our RSSE structure is described in Section III. Section IV contains analysis and simulation results. This section starts with the general analysis of the RSSE, followed by the discussion on the quasi-catastrophic behavior of MLSE that was mentioned above and ends with a brief analysis of the DFSE. Finally, in Section V, we summarize our results and indicate some directions for future research.

II. System Model

We consider a generic uncoded quadrature amplitude modulation (QAM) system [14], as shown in Fig. 1. During each signaling interval $nT \leq t \leq (n+1)T$, q input bits are collected and mapped into a complex-valued symbol x_n selected from a two-dimensional signal set with $M = 2^q$ signal points. [We assume that the bit sequence is independent and identically distributed (i.i.d.).] As is often the case in practice, the signal points are assumed to be chosen from a rectangular lattice with odd-integer coordinates ($\pm 1, \pm 3, \cdots$). (RSSE can be easily generalized to other lattices.) The real and imaginary coordinates of x_n are filtered and modulated on in-phase and quadrature carriers which are superimposed for transmission over a linear additive white Gaussian noise (AWGN) channel. The received signal is passed through a matched filter, demodulated with correct carrier phase into its quadrature components, and sampled at the symbol rate with correct timing phase.

Further, a discrete-time noise-whitening filter can be used to generate a sequence $\{r_n\}$ such that [1]

$$r_n = x_n + (p_n, f) + w_n \tag{1}$$

where $\{w_n\}$ represents a complex white Gaussian noise sequence with zero mean and variance $2\mu^2$. The expression (p_n, f) is an inner product between the state vector

$$p_n = [x_{n-1}, x_{n-2}, \cdots, x_{n-K}], \tag{2}$$

which has the K most recent transmitted symbols as its elements, and the vector f representing the complex postcursor ISI coefficients (assumed known) associated with the overall channel impulse response, i.e.,

$$f = [f_1, f_2, \cdots, f_K]. \tag{3}$$

(Without loss of generality, f_0 is assumed to be unity.) The length K of channel memory is assumed to be finite.

In contrast to MLSE, the performance of an RSSE may be affected by the phase response of the noise-whitening filter. If the z transform $H(z)$ of the noise autocorrelation sequence at the matched filter output has no zeros on the unit circle, the noise-whitening filter can be chosen as an anticausal but stable filter with poles corresponding to the zeros of $H(z)$, which are outside the unit circle. This ensures that the overall channel response is minimum phase, and shifts the signal energy towards the earliest sample of the channel response. In particular, the minimum phase condition implies that the energy in the first K' samples is maximized for *every* K' [18]. Zeros on the unit circle can be handled by combining the matched filter and the noise-whitening filter into a whitened matched filter [1]. Recall that a whitened matched filter is also the optimum feedforward filter for a zero-forcing DFE. The

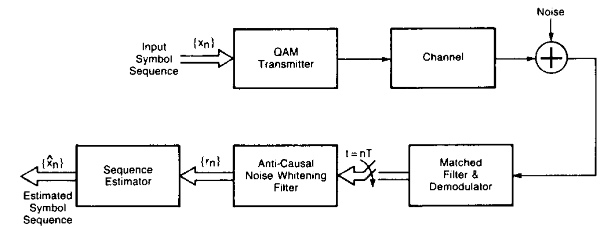

Fig. 1. QAM transmission system.

DFE uses K previous decisions to cancel the postcursor ISI contributions. However, even in the absence of error propagation, it is strictly suboptimum, since in the decision process, it ignores the signal energy embedded in the ISI terms.

III. Reduced-State Sequence Estimation

Recall that in MLSE, the trellis states are defined as $p_n = [x_{n-1}, x_{n-2}, \cdots, x_{n-K}]$, i.e., the state of a sequence at a given time is equal to its K most recent symbols. Since each element in the state vector can take one of M values, the ML trellis has M^K states and there are M transitions to and from each state.

To reduce the number of states, for each element x_{n-k} in the vector p_n, we define a two-dimensional set partitioning denoted as $\Omega(k)$. Specifically, for the kth element x_{n-k}, the signal set is partitioned into J_k subsets where J_k can range anywhere from 1 to M (further specifics of set partitioning will be described shortly). The index of the subset of a symbol x_i in the partitioning $\Omega(k)$ is, in general denoted as $a_i(k)$, which can be taken as an integer between 0 and $J_k - 1$. We constrain the set partitionings such that a) the numbers J_k are nonincreasing (i.e., $J_1 \geq J_2 \geq \cdots \geq J_K$), and b) the partitioning $\Omega(k)$ is a further partition of the subsets of $\Omega(k+1)$ for each k between 1 and $K - 1$. These restrictions allow us to define the *subset state* of a sequence at time n as

$$t_n = [a_{n-1}(1), a_{n-2}(2), \cdots, a_{n-K}(K)], \tag{4}$$

representing the subsets of the K most recent symbols in the respective partitionings. Conditions a) and b) assure that, given the current state t_n and the subset $a_n(1)$ of the current symbol, the next state t_{n+1} of a sequence is uniquely determined. Therefore, the subset states define a proper trellis, which we will call a subset trellis, representing all possible sequences $\{x_n\}$.

Since in (4), $a_{n-k}(k)$ can take one of J_k values, the total number of states in the subset trellis is given by the product of the J_k. Furthermore, there are again, in principle, M transitions from each state, one for each possible value of the current symbol x_n. However, for any current state, there are only J_1 distinct next states, one for each subset in the partitioning $\Omega(1)$ [clearly, two paths originating from the same state will undergo the same state transition if their current symbols x_n and x_n' belong to the same subset in the partitioning $\Omega(1)$]. When $J_1 < M$, the trellis thus has parallel transitions that start in a common state and end in a common state. We may say that from each state there are J_1 "subset transitions," each consisting of as many parallel transitions as the number of symbols in the corresponding subset.

Now, we discuss the set partitioning in more detail. This should be selected with the objective of optimizing the performance/complexity tradeoff for the RSSE. Note that each subset state consists of the union of a number of ML states. Therefore, in the subset trellis, certain paths will merge earlier than in the ML trellis. Hence, the set partitioning should be such that these "early merging" paths can be reliably distinguished at the point of merging without any further delay. It appears that good results are generally obtained if, for each partitioning $\Omega(k)$, the minimum intrasubset Euclidean distance, denoted by Δ_k, is maximized. In [12], Ungerboeck

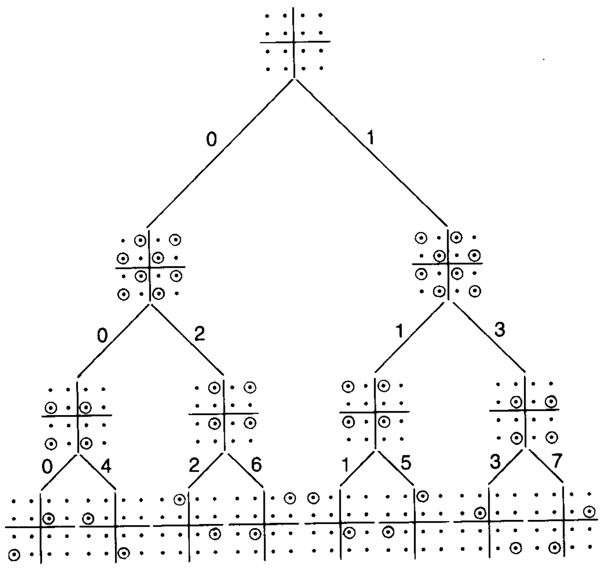

Fig. 2. Ungerboeck partition tree for the rectangular 16-QAM signal set.

showed that when J_k is a power of two, maximum Δ_k can be obtained by successive two-way partitions, as illustrated in Fig. 2 for a 16-point signal set partitioned into two, four, and eight subsets. Forney points out [21] that more generally, there is a J_k-way partition of the rectangular lattice with maximum $\Delta_k^2 = 4J_k$ as long as J_k is the magnitude of a Gaussian integer β, squared (norm) a complex number with integer real and imaginary parts. If G is the original rectangular lattice, then the number β induces the desired partition $G/\beta G$. The sublattice βG is also rectangular, except it is rotated by the phase of β. Of course, when J_k is not a power of two, the subsets will not have the same number of symbols; however, as will become apparent, this will be of little concern.

As an example, in Fig. 3(a), we show the four-state subset trellis for the case $K = 1$, $J_1 = 4$. Similarly, using the subset labeling of Fig. 2, in Fig. 3(b) we show the eight-state subset trellis for $K = 2$, $J_1 = 4$, and $J_2 = 2$. In both figures, parallel transitions are shown by single lines and the corresponding subsets are indicated on the left. Note that the size of the signal set affects only the number of parallel transitions. Otherwise, the trellises are fully specified by the selection of J_k, $k = 1, \cdots, K$.

Now, we describe the use of the VA to search a subset trellis. First recall that in MLSE, for transitions originating from a state p_n, the VA computes branch metrics according to

$$b[r_n; p_n, x_n] = |r_n - (p_n, f) - x_n|^2. \tag{5}$$

The branch metrics depend on the K most recent symbols, which are uniquely specified by the ML state p_n. In a subset trellis, however, the states specify only the subsets to which these symbols belong. Therefore, we modify the VA slightly by introducing decision feedback into the branch metric computations. Specifically, for transitions from a state t_n, we use the branch metrics

$$b[r_n; t_n, x_n] = |r_n - (\hat{p}_n(t_n), f) - x_n|^2 \tag{6}$$

where $\hat{p}_n(t_n)$ represents the K most recent symbols stored in the *path history* associated with the state t_n. Note that in MLSE, path histories usually contain the surviving state sequences leading to the current states. In our RSSE, however, when $J_1 < M$, it may be more appropriate to store the actual

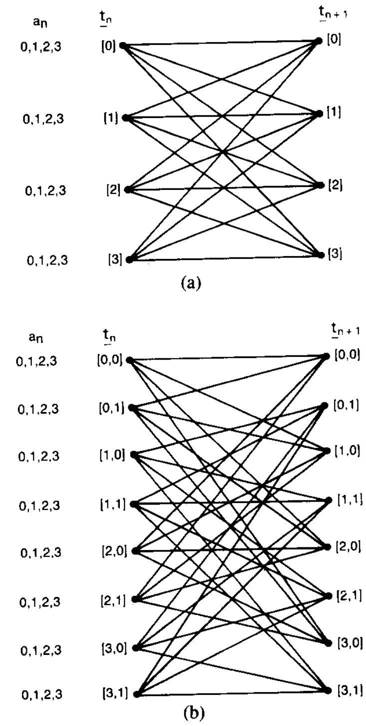

Fig. 3. Subset trellises (a) $K = 1$, $J_1 = 4$, (b) $K = 2$, $J_1 = 4$, $J_2 = 2$.

surviving symbols since there is no one-to-one correspondence between state and symbol sequences.

When the subset trellis has parallel transitions (i.e., $J_1 < M$), then for each subset transition, the VA can first select the symbols with the minimum branch metric; in effect, for each state, the VA makes delay-free decisions between symbols within the subsets of $\Omega(1)$, using past decisions obtained from path histories as feedback. When the partitioning $\Omega(1)$ has enough symmetry, this can be done without explicitly computing the branch metric for every symbol. For example (see Fig. 2), in an Ungerboeck partition of a rectangular signal set, signal points in a subset always lie on a rectangular grid. (This holds more generally for any partition $G/\beta G$ where β is a Gaussian integer.) Therefore, signal points with minimum branch metric can be determined by simple slicing operations. That is, for each state, only J_1 explicit branch metric computations will be necessary.

In each iteration, this first step in the VA reduces the number of possible extensions of the N path histories from NM to NJ_1. This is then reduced to N by selecting the paths with minimum accumulated metric for each possible next state t_{n+1} (i.e., the RSSE retains one "survivor" path for each possible combination of the K most recent subsets). Note that reducing the number of possible extensions in this manner requires only $N(J_1 - 1)$ binary comparisons. This is considerably simpler than selecting the N extensions with minimum accumulated metric, as in the M algorithm or its variations [9]. Decisions are made, as usual, after some delay by tracing back the path history of the state with the smallest accumulated metric. The required delay should not be greater than in MLSE since (as will become apparent) the error events in RSSE are no longer than those in MLSE.

If $J_k = 1$ for all k, the RSSE degenerates into a zero-forcing DFE. If $J_k = M$ for all k, the RSSE becomes an MLSE. Thus, by choice of the J_k, a tradeoff of performance versus complexity between a DFE and MLSE can be obtained.

In practice, there are still many applications (e.g., binary transmission) where the principal source of MLSE complexity is the length of channel memory. In this case, a reduced-state trellis can be formed by simply truncating the ML state vector to some suitable length $K' < K$ and defining the reduced state vector as $p_n' = [x_{n-1}, x_{n-2}, \cdots, x_{n-K'}]$. In our subset formulation, this is equivalent to choosing $J_k = M$ for k between 1 and K' and $J_k = 1$ for k between $K' + 1$ and K. This is a special case of the RSSE, which will be called a decision feedback sequence estimator (DFSE). To explore this case further, we write the received sequence $\{r_n\}$ in the form

$$r_n = x_n + (p_n', f') + (p_n'', f'') + w_n \tag{7}$$

where the vectors f' and f'' are defined according to

$$f = [f' \,|\, f''] = [f_1, \cdots, f_{K'} \,|\, f_{K'+1}, \cdots, f_K] \tag{8}$$

and

$$p_n = [p_n' \,|\, p_n''] = [x_{n-1}, \cdots, x_{n-K'} \,|\, x_{n-K'+1}, \cdots, x_{n-K}]. \tag{9}$$

The term (p_n'', f'') represents residual ISI which is eliminated using decision feedback. In contrast to the approach taken in [6], which uses a DFE separate from the sequence estimator, in a DFSE, the feedback mechanism is incorporated into the sequence estimator. To see this, we rewrite the branch metric in the form

$$b[r_n; p_n', x_n] = |r_n - (\hat{p}_n''(p_n'), f'') - (p_n', f') - x_n|^2. \tag{10}$$

Here, the term $(\hat{p}_n''(p_n'), f'')$ represents an estimate of the residual ISI for the state p_n'. The feedback decisions $\hat{p}_n''(p_n')$ are obtained from the path histories. This guarantees that the branch metric of the correct path is the ML metric, as long as it is not discarded in favor of some incorrect path. Hence, in contrast to the approach taken in [6], the first-error probability of the sequence estimator does not depend on a DFE, which can exhibit a relatively poor performance on bad channels. Of course, the primary difference between MLSE and DFSE is that in the (reduced) trellis used by the DFSE, two paths merge earlier (when they agree in K' rather than K most recent symbols). As we will see, this generally increases the first-error probability for discarding the correct path; however, the performance should be superior to that of the preprocessing techniques described in [3]–[6].

IV. Error Probability Performance

A. General Analysis of RSSE

With RSSE, the VA will make an error for the first time at time j if up to that time, the correct path $\{x_n\}$ accumulates a worse ML metric than some incorrect path $\{x_n'\}$ that merges with $\{x_n\}$ in the subset trellis at time j. Note that in a subset trellis, two paths merge if the subsets of the K most recent symbols are the same, while in an ML trellis, merging occurs only if these symbols are identical. Also, in an RSSE, once the correct path is discarded, there is a possibility of error propagation since the probability of discarding the correct path again may increase for the next few symbols. We should note that error propagation can occur only if the paths $\{x_n\}$ and $\{x_n'\}$ are different in one or more of the K most recent symbols, i.e., if $x_{n-k} \neq x_{n-k}'$ for some k between 1 and K.

The exact error probability of the RSSE is somewhat difficult to analyze because of decision feedback. However, significant insight can be gained by examining the probability P_e of a first-error event occurring at some time j. We prefer to assess the effects of error propagation through simulations.

If $\{x_n\}$ is the correct sequence and $\{x_n'\}$ is the estimated sequence and they are "merged" at times $n = i$ and $n = j$ and unmerged in between, we represent such an error event by the vector $e = [e_i, e_{i+1}, \cdots, e_{j-1}]$ where the $e_n = x_n - x_n'$ are error symbols. We assume that the ending position j is fixed, while the starting position $i < j$ is arbitrary. The set of all such error events, denoted as E', depends on the subset trellis in use, and in general will be different from the set of error events in the ML trellis. Specifically, an error event e is in the set E' if and only if the following hold.

1) The first element e_i is nonzero.

2) The last K elements, e_{j-K} through e_{j-1}, satisfy the "merging condition," i.e., e_{j-k} is equal to $x_{j-k} - x_{j-k}'$ where the correct symbol x_{j-k} and the estimated symbol x_{j-k}' belong to the same subset in the partitioning $\Omega(k)$ for $k = 1, 2, \cdots, K$.

3) No earlier K elements satisfy the merging condition.

We define the squared distance $d^2(e)$ of an error event e as

$$d^2(e) = \sum_{n=i}^{j-1} \left| e_n + \sum_{k=1}^{K} e_{n-k} f_k \right|^2 \tag{11}$$

with $e_n = 0$ for $n < i$. (Here we assume that, up to time j, the correct path has not been discarded by the VA.) Then, using union bound arguments, at sufficiently large SNR, P_e can be approximated by

$$P_e \approx C' Q(d_{\min}'/2\mu) \tag{12}$$

where $d_{\min}'$ is the minimum distance, i.e.,

$$d_{\min}' = \min_{e \in E'} d(e), \tag{13}$$

and C' is an error coefficient equal to the average (averaged over all possible transmitted sequences) number of error events at distance $d_{\min}'$.

Note that when $J_k = M$ for all k, (12) gives the MLSE error event probability. For this case, we define $C' = C$ and $d_{\min}' = d_{\min}$. To establish the relationship between $d_{\min}$ and $d_{\min}'$, we first observe that in the MLSE, the merging condition 2) implies that the error events must have their last K elements equal to zero. In RSSE, however, the subset trellis will generally have a set E_{12}' of what may be called "early merging error events" which are not present in the ML trellis. We denote the minimum distance of such error events as $d_{\min}(E_{12}')$. Then, it is straightforward to show that

$$d_{\min}' = \min\,(d_{\min}, d_{\min}(E_{12}')). \tag{14}$$

Note that certain error events in the ML trellis may not belong to the set E' because of condition 3) above. However, either these error events have a distance greater than $d_{\min}$ or there will be an error event in the set E_{12}' which has a smaller distance. Therefore (14) holds.

When $J_1 < M$, it is useful to represent the set E_{12}' by the disjoint union

$$E_{12}' = E_1' \cup E_2' \tag{15}$$

where E_1' represents the (early merging) error events of length 1, associated with parallel transitions. The minimum distance $d_{\min}(E_1')$ is the minimum intrasubset distance Δ_1 for the partitioning $\Omega(1)$. For a partition induced by a Gaussian integer, it is easily shown that

$$d_{\min}(E_1') = 2\sqrt{J_1}. \tag{16}$$

Then, (14) can be written as

$$d_{\min}' = \min\,(d_{\min}, 2\sqrt{J_1}, d_{\min}(E_2')) \tag{17}$$

where $d_{\min}(E_2')$ is the minimum distance among the remaining early merging error events that are longer than one symbol interval. Therefore, to achieve the same minimum distance as the MLSE, the number of subsets J_1 should be chosen to be no less than $d^2_{\min}/4$.

Example: Consider a channel of length $K = 1$ with f_1 real and with $|f_1| < 1$ (minimum phase condition). The size of the signal set is arbitrary (say $M \geq 4$). The minimum distance $d_{\min}$ of the MLSE, given by

$$d^2_{\min} = 4(1 + f_1^2), \tag{18}$$

is achieved by error events $\boldsymbol{e} = [\pm 2, 0]$ and $\boldsymbol{e} = [\pm 2j, 0]$.

Suppose that the RSSE is based on a J_1-way Ungerboeck partitioning. Clearly, $d^2_{\min} < 8$, and therefore using (16), we find that $d_{\min}(E_1') > d_{\min}$ as long as $J_1 \geq 2$. Thus, a two-way partition is sufficient to guarantee that the distance between parallel transitions is greater than the minimum distance of MLSE. Next, we consider other critical early merging error events. These are of the form $\boldsymbol{e} = [e_{j-2}, e_{j-1}]$ where $e_{j-2} = \pm 2$ or $e_{j-2} = \pm 2j$ and e_{j-1} is an intrasubset error with $|e_{j-1}|^2 \geq 4J_1$. Using (11), it is then straightforward to show that if $J_1 \geq 2$, $d(\boldsymbol{e}) > d_{\min}$; thus, we can conclude that the distance of early merging error events is always greater than $d_{\min}$, provided that $J_1 \geq 2$. Hence, we have shown that with just two states, RSSE can achieve the same asymptotic performance as MLSE, i.e., $d'_{\min} = d_{\min}$ and $C' = C$.

When $f_1 = \pm 1$, it can be shown that two states are again sufficient to achieve $d'_{\min} = d_{\min}$; however, at least four states are needed to achieve the MLSE error coefficient.

These results may not hold in general when f_1 is complex valued. For example, when $f_1 = e^{\pm j\pi/4}$, it takes a five-state RSSE based on a five-way partitioning ($\beta = 2 + j$) to achieve the MLSE performance.

Note that in the above example, the asymptotic performance of the RSSE does not depend on the size of the signal set. In fact, we conjecture that for any channel (not necessarily of memory length 1), the complexity of RSSE required to achieve MLSE performance is independent of the size of the signal set, at least for large enough signal sets, and only depends on the channel coefficients.

The case $K = 1$, $f_1 = \pm 1$ also corresponds to $1 \pm D$ QPRS over an ideal channel (here and in what follows, we assume that in partial-response systems, all the shaping is done at the transmitter). We have simulated our RSSE for $1 + D$ QPRS with precoding, transmitting four bits per symbol, i.e., with $M = 16$. As shown in Fig. 4 with four states, there is no noticeable degradation (at large SNR's) relative to the MLSE, and with two states, the degradation is only 0.2 dB. Note that in this example, the critical error events are primarily ML error events. Therefore, the effects of error propagation should be negligible. (In all simulation results, a long decoding delay is assumed for MLSE and RSSE.)

B. MLSE: Avoiding Quasi-Catastrophic Behavior

In Fig. 4, it can be observed that MLSE performance for $1 + D$ QPRS is about 1.0 dB worse that ISI-free transmission (at $P_s = 10^{-4}$), although there is no difference in the effective minimum distance. This can be explained as follows. In the ML trellis, there are an infinite number of error events of the form $\boldsymbol{e} = [e_i, -e_i, e_i, -e_i, \cdots, 0]$ (with error symbols having alternating signs and a zero element at the end) where $e_i = \pm 2$ or $e_i = \pm 2j$, all achieving the minimum distance of 8. The average number of such error events and hence the error coefficient C is finite. However, C increases with the size of the signal set as the likelihood of an error symbol e_n increases with enlarging signal set boundaries. In fact, $C = 4(\sqrt{M} - 1)$, whereas for ISI-free transmission, the error coefficient is at most 4. The error coefficient for symbol error probability is $2C$ if precoding is used. Without precoding, it can be significantly larger. Note that precoding does not affect the error coefficient for the error event probability. Another undesirable effect of this "quasi-catastrophicity" is that it necessitates an infinite decoding delay to cover potentially infinitely long minimum distance error events. When a finite delay is used, the effective minimum distance will be that of a DFE.

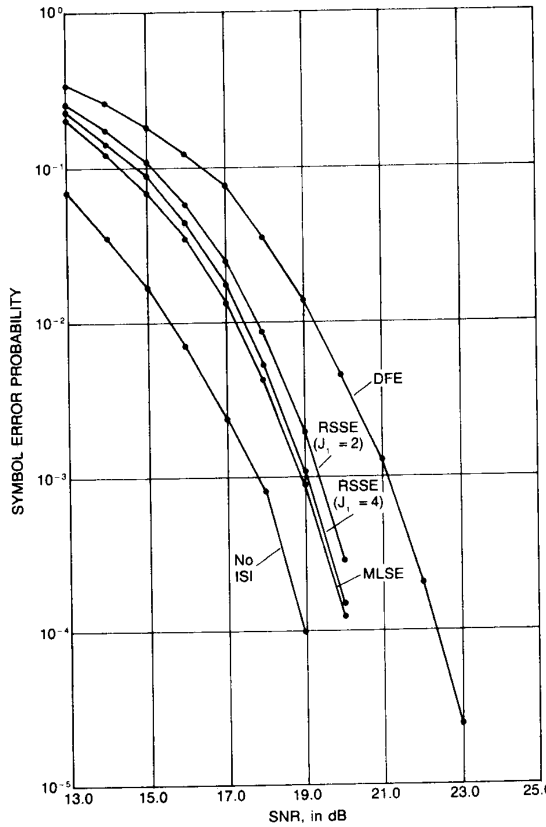

Fig. 4. Simulated performance for $(1 + D)$ quadrature partial-response signaling with precoding ($M = 16$).

We now show that this quasi-catastrophic behavior can be eliminated by simply rotating the signal set in the transmitter by $\pi/4$ rad every symbol interval. (Rotating the signal set to improve MLSE performance has been previously advocated by Acampora [15] in a more general context.) That is, the symbol transmitted at time n is $x_n e^{j(\pi/4)n}$ where x_n is chosen from the original (fixed) signal set. Such a rotation does not affect the transmitted spectrum since the transmitted sequence is still uncorrelated. Rotating the signal set in this manner can also be viewed as shifting the carrier frequency by 1/8T Hz, while the spectrum of the transmitted signal remains unchanged. The received signal is processed by the receiver front end in the same way as described in Section II, and at the output of the noise whitening filter, we have

$$\begin{aligned} r_n' &= x_n e^{j(\pi/4)n} + x_{n-1} e^{j(\pi/4)(n-1)} + w_n' \\ &= e^{j(\pi/4)n}(x_n + x_{n-1} e^{-j(\pi/4)} + w_n). \end{aligned} \tag{19}$$

It can be observed that by multiplying r_n' by $e^{-j(\pi/4)n}$ prior to sequence estimation, we obtain an equivalent channel with transfer function $1 + e^{-j\pi/4}D$ with input symbols chosen from the original signal set. It is easy to see that for this equivalent channel, the MLSE minimum squared distance is again 8. Furthermore, as in ISI-free transmission, the minimum dis-

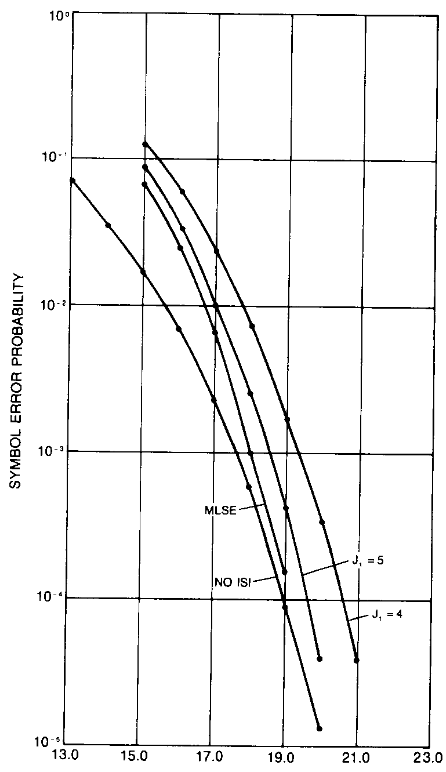

Fig. 5. Simulated performance for $(1 + D)$ quadrature partial-response signaling with carrier frequence offset $(M = 16)$.

tance is achieved only by the error events $e = [\pm 2, 0]$ and $e = [\pm 2j, 0]$. Therefore, the asymptotic performance of MLSE should be as good as ISI-free transmission. Since long critical error events are avoided, the decoding delay is also significantly reduced. It can be shown that signal set rotation also reduces the peak-to-average ratio of the transmitted signal. This may reduce the effects of any nonlinear distortion which may be present in the transmitter or in the channel.

The simulation results shown in Fig. 5 illustrate that with rotation, the performance of MLSE asymptotically approaches that of ISI-free transmission, as predicted above. We also show in Fig. 5 the performance of RSSE in this case. It can be observed that a five-state RSSE is needed to approach MLSE performance. The five-state RSSE uses the five-way partitioning shown in Fig. 6. Note that with five-way partitioning, the minimum intrasubset distance Δ_1 is larger than that of Ungerboeck's four-way partitioning by 0.97 dB.

The more general class of $1 \pm D^N$ QPRS can be treated in a similar manner using the appropriate rotation and demodulation strategy to obtain an equivalent $1 + e^{-j\pi/4} D^N$ channel, and similar conclusions can be drawn.

It should be mentioned that when the signal set is square and the channel is real (as in conventional QPRS), both MLSE and RSSE can be formulated as two interleaved one-dimensional partial-response systems to reduce complexity. However, this is not possible with rotation since in the equivalent channel, the in-phase and quadrature components become cross coupled.

C. Analysis of DFSE

Now, we consider the case of the DFSE with truncated memory $K' < K$. Here, the set E' consists of error events (all of the early merging type) which have their last K' elements equal to zero. From (11), it can be seen that additional zero

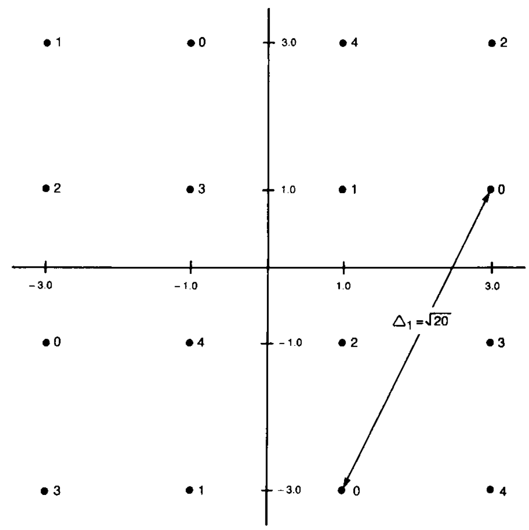

Fig. 6. Five-way set partitioning for the rectangular 16-QAM signal set.

elements at the end of an error event can only increase the distance. Therefore, $d'_{\min}$ can only increase as K' is increased towards K. For $K' = K$, we have $d'_{\min} = d_{\min}$, while for $K' = 0$ (where DFSE becomes a zero-forcing DFE), $d'_{\min}$ is the minimum Euclidean distance between symbols in the signal set.

For a DFSE, the relationship given by (14) is trivial. Instead, when the minimum distance error events of the MLSE are known, a simple upper bound on $d'_{\min}$ can be found by subtracting the maximum possible contribution of the last $K - K'$ zero elements of these error events from d_{min}.

Example: Consider binary signaling $(x_n = \pm 1)$ over a length $K = 2$ channel with real coefficients $f_1 \geq 0, f_2 \geq 0$. We assume $f_1 < 1 + f_2$ and $f_2 < 1$ (minimum phase condition). In this case, it can be shown that the minimum distance error events in the DFSE $(K' = 1)$ are the same as the minimum distance error events of the MLSE, except for a missing zero element at the end. Therefore,

$$d'^2_{\min} = d^2_{\min} - 4f_2^2. \tag{20}$$

Thus, in this particular example, $d'_{\min}$ achieves the bound mentioned above. Also note that the reduction in squared minimum distance is equal to the energy in the "neglected" portion of the channel impulse response (this may not be expected to be true in general).

Now, consider a $(1 \pm D)^2$ partial-response system with binary signaling (i.e., $f_1 = \pm 2, f_2 = 1$). In this case, DFSE is only 1.25 dB worse than the (four-state) MLSE, and it is 4.77 dB better than a zero-forcing DFE. These estimates closely agree with the simulation results shown in Fig. 7 for the $(1 + D)^2$ channel. Also shown in Fig. 7 is the performance that would be obtained if the true path always used the correct feedback decisions in the branch metric computations. This illustrates the effects of error progagation. Note that error propagation affects the DFSE less than it affects the DFE; in particular, at high SNR, the effect is negligible. This encouraging result might have been expected since in DFE, both ISI terms are cancelled with feedback, while in DFSE, only the second ISI term is subtracted with feedback.

V. Summary and Conclusions

We have shown that effective reduced-state sequence estimation for linear ISI channels can be achieved with an MLSE-like structure. We have used Ungerboeck set partitioning principles to construct trellises with a reduced number of

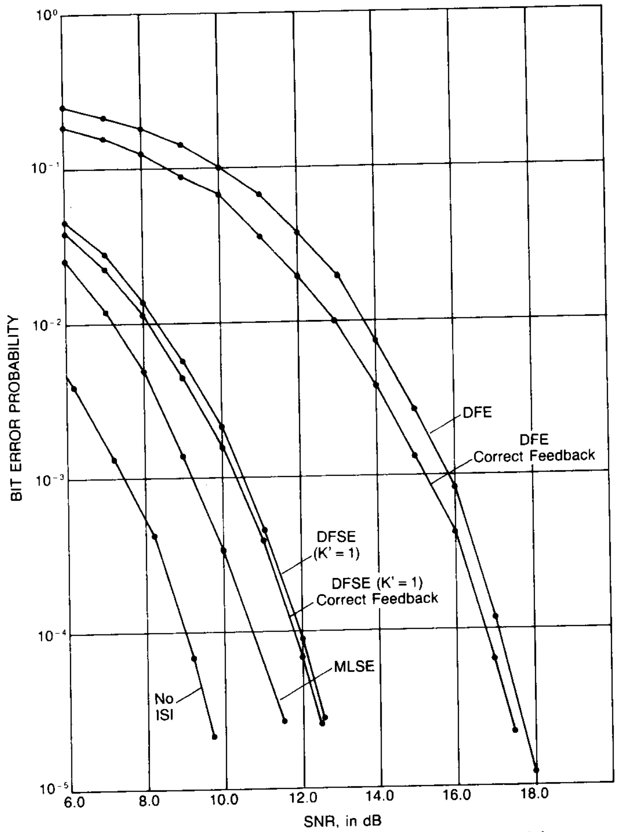

Fig. 7. Simulated performance for $(1 + D)^2$ binary partial-response signaling.

states. Some simple examples are given to demonstrate the performance of our RSSE structure.

In this paper, no attempt has been made to extensively compare RSSE to various previously proposed detection techniques, e.g., those described in [6], [9], and [10]. However, the results presented here (as well as further results to be reported in the future) look very encouraging.

If the channel response is unknown, the RSSE can be made adaptive by use of a channel estimation algorithm. Although the development assumes a whitened matched filter as the received front-end filter, for null-free channels, an appropriately modified version of the RSSE can also operate at the output of a zero-forcing linear equalizer. It should also be noted that Duel and Heegard [11] recognized the application of the DFSE to recursive channels with infinite impulse response.

In Section III, we constrained the set partitioning in order to obtain a well-defined subset trellis. However, at least in principle, this is not necessary. For any partitioning of the set of all possible values of the $2K$-dimensional ML state vector p_n, the state transitions can be defined in real time based on symbols stored in the path histories. Of course, this may increase implementation complexity and, in general, will not be preferred.

Finally, the RSSE structure can also be used for near-MLSE decoding of trellis-coded modulation systems [12], [13] with large signal sets operating in the presence of ISI. This important extension will be described in the future.

Acknowledgment

The authors are greatly indebted to G. D. Forney, Jr. for his close attention and many valuable comments and suggestions which improved this paper significantly.

References

[1] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory,* vol. IT-18, pp. 363-378, May 1972.

[2] ——, "The Viterbi algorithm," *Proc. IEEE,* vol. 61, pp. 268-278, Mar. 1973.

[3] S. U. Qureshi and E. E. Newhall, "An adaptive receiver for data transmission over time-dispersive channels," *IEEE Trans. Inform. Theory,* vol. IT-19, pp. 448-457, July 1973.

[4] D. D. Falconer and F. R. Magee, Jr., "Adaptive channel memory truncation for maximum-likelihood sequence estimation," *Bell Syst. Tech. J.,* vol. 52, pp. 1541-1562, Nov. 1973.

[5] C. T. Beare, "The choice of the desired impulse response in combined linear-Viterbi algorithm equalizers," *IEEE Trans. Commun.,* vol. COM-26, pp. 1301-1327, Aug. 1978.

[6] W. U. Lee and F. S. Hill, "A maximum-likelihood sequence estimator with decision-feedback equalization," *IEEE Trans. Commun.,* vol. COM-25, pp. 971-979, Sept. 1977.

[7] F. L. Vermuelen and M. E. Hellman, "Reduced-state Viterbi decoding for channels with intersymbol interference," in *Proc. ICC '74,* Minneapolis, MN, June 1974.

[8] G. J. Foschini, "A reduced-state variant of maximum-likelihood sequence detection attaining optimum performance for high signal-to-noise ratio performance," *IEEE Trans. Inform. Theory,* vol. IT-24, pp. 505-509, Sept. 1977.

[9] A. P. Clark, S. N. Abdullah, S. J. Jayasinghe, and K. H. Sun, "Pseudobinary and pseudoquaternary detection processes for linearly distorted multilevel QAM signals," *IEEE Trans. Commun.,* vol. COM-33, pp. 639-645, July 1985.

[10] A. P. Clark and M. Clayden, "Pseudobinary Viterbi detector," *Proc. IEE,* vol. 131, part F, pp. 208-218, Apr. 1984.

[11] A. Duel and C. Heegard, "Delayed decision-feedback sequence estimation," presented at the Allerton Conf. Commun., Contr., Comput., Oct. 1985.

[12] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory,* vol. IT-28, pp. 55-67, Jan. 1982.

[13] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for bandlimited channels," *IEEE J. Select. Areas Commun.,* vol. SAC-2, pp. 632-647, Sept. 1984.
[14] S. U. Qureshi, "Adaptive equalization," *Proc. IEEE,* vol. 73, pp. 1349-1387, Sept. 1985.
[15] A. S. Acampora, "Analysis of maximum-likelihood sequence estimation performance for quadrature amplitude modulation," *Bell Syst. Tech. J.,* vol. 60, pp. 865-885, July-Aug. 1981.
[16] J. B. Anderson and S. Mohan, "Sequential decoding algorithms: A survey and cost analysis," *IEEE Trans. Commun.,* vol. COM-32, pp. 169-176, Feb. 1984.
[17] G. D. Forney, Jr. and A. R. Calderbank, "Coset codes for partial response channels; or coset codes with spectral nulls," submitted to *IEEE Trans. Inform. Theory,* May 1987.
[18] A. W. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1975.
[19] K. Wesolowski, "An efficient DFE & ML suboptimum receiver for data transmission over dispersive channels using two-dimensional signal constellations," *IEEE Trans. Commun.,* vol. COM-35, pp. 336-339, Mar. 1987.
[20] M. V. Eyuboğlu and S. U. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," in *Proc. GLOBECOM '86,* Houston, TX, Dec. 1986, pp. 1023-1028.
[21] G. D. Forney, Jr., private communications, June 1987.

★

M. Vedat Eyuboğlu (S'80-M'84) was born in Istanbul, Turkey, on May 22, 1955. He received the B.S. degree from Bogazici University, Istanbul, in 1978 and the M.S. and Ph.D. degrees from Rensselaer Polytechnic Institute, Troy, NY, in 1980 and 1984, respectively, all in electrical engineering. At Rennselaer Polytechnic Institute, he studied data communication receivers with antenna arrays, particularly optimum receiver structures and their adaptive implementation.

In November 1983 he joined Codex Corporation, Mansfield, MA, where he is currently a Principal Engineer leading a high-speed voiceband modem project. His current research interests are in modulation, coding, advanced equalization, and their applications to various communication systems. He has pending and issued patents in this field.

★

Shahid U. H. Qureshi (S'68-M'73-SM'81-F'87) was born in Peshawar, Pakistan, on September 22, 1945. He received the B.Sc. degree from the University of Engineering and Technology, Lahore, Pakistan, in 1967, the M.Sc. degree from the University of Alberta, Edmonton, Canada, in 1970, and the Ph.D. degree from the University of Toronto, Toronto, Ont., Canada, in 1973, all in electrical engineering.

He received the President of Pakistan's Award for outstanding student in the graduate stage in 1968 and held the Canadian Commonwealth Scholarship from 1968 to 1972. In 1967 and 1968 he was a Lecturer with the Department of Electrical Engineering, University of Engineering and Technology, Lahore, Pakistan. In 1973 he joined Codex Corporation, Mansfield, MA, as a member of the Technical Staff in the Research Department. He was elected an Associate Member of the Motorola Science Advisory Board in 1978 and was named a Motorola Dan Noble Fellow in 1984 in recognition of leadership in modem research and development. Since 1980, he has represented Codex in the U.S. Modem Working Party (now the EIA TR30.1) which studies data communication over the telephone network, and in the CCITT's Study Group XVII. He is currently Vice President of Research and Advanced Development at Codex, responsible for technology development in the areas of networking, software, data transmission, ISDN, voice coding, and digital signal processing.

Dr. Qureshi was elected an IEEE Fellow "for contributions to the architecture and commercial development of high-speed voice-band modems," and was awarded the IEEE 1987 Fink Award for his paper on adaptive equalization.