Error Event Characterization on 2-D ISI Channels

Ismail Demirkan, *Member, IEEE*, Paul H. Siegel, *Fellow, IEEE*, and Jack K. Wolf, *Life Fellow, IEEE*

***Abstract*—In this paper, we analyze the distance properties of two-dimensional (2-D) intersymbol interference (ISI) channels, in particular the 2-D partial response class–1 (PR1) channel which is an extension of the one-dimensional (1-D) PR1 channel. The minimum squared-Euclidean distance of this channel is proved to be 4 and a complete characterization of the squared-Euclidean distance 4 error events is provided. As for 1-D channels, we can construct error-state diagrams for 2-D channels to help characterize error events. We propose an efficient error event search algorithm operating on the error-state diagram that is applicable to any 2-D channel.**

***Index Terms*—Error events, holographic recording, intersymbol interference (ISI), two-dimensional (2-D) channels.**

I. Introduction

DETECTION and coding for two–dimensional (2-D) channels have been the subject of much research recently because of advances in holographic storage technology. Signal processing and coding aspects of holographic storage systems have been studied by several authors [1], [2]. A generalization of one–dimensional (1-D) detection and coding methods to 2-D channels is not trivial due to the lack of convenient graph-based descriptions of such channels. In particular, there is no simple trellis-based maximum-likelihood (ML) detection algorithm analogous to the 1-D Viterbi algorithm. Ordentlich and Roth proved that the ML sequence detection problem on 2-D intersymbol interference (ISI) channels is NP-complete [3].

However, there are suboptimal detection techniques such as the iterative multistrip (IMS) algorithm for 2-D channels that demonstrate very good error-rate performance and appear to approximate the performance of an ML detector [4]. The IMS algorithm is a message-passing algorithm operating on soft-input soft-output detectors, such as *a posteriori* probability (APP) detectors. It is therefore of interest to identify the dominant 2-D error events, where we define an error event as the difference between the recorded and the decoded data arrays. Empirical evidence has shown that data arrays forming dominant error events for the IMS algorithm generate channel outputs with small squared-Euclidean distance. Therefore, it is important to characterize the 2-D error events with small squared-Euclidean distance, so that 2-D distance-enhancing constrained codes can be designed to improve system performance [2].

Chugg investigated the performance of an ML page detector in the presence of ISI and additive white Gaussian noise (AWGN) [5]. If the channel impulse response has finite support size, then the channel output can be characterized as a Markov random field. The bit error rate performance of an ML page detector can be bounded from above by a union bound, which is computed by using the fundamental error patterns in the channel input arrays.

Karabed *et al.* introduced an analytic method to characterize the distance properties of some 1-D partial-response channels [6]. In this paper, we extend this method to characterize the closed error events of some 2-D channels. In particular, we study the 2-D partial response class–1 (PR1) channel whose impulse response is given by

$$h = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{1}$$

This impulse response can be a good ISI model for holographic storage systems when there is a half-period sampling shift between the read-back signal and the detector sampling points.

The analytic method used for characterizing error events for the 2-D PR1 channel is tedious to apply for most 2-D channels, particularly for the channels whose impulse responses span $p \times q$ arrays where $p, q > 2$. For 1-D channels, efficient search algorithms working on error-state diagrams have been developed to characterize error events for high-order partial response channels [7]. Error-state diagrams for 2-D channels can be generated by fixing the size of the error event in the horizontal or vertical direction. In this paper, we propose a bounded-depth search algorithm for finding closed and connected error events for any 2-D channel. The complexity of the algorithm depends solely on the underlying ISI pattern of the channel.

This paper is organized as follows. In Section II, the description of the 2-D channel model and related definitions and notations are introduced. In Section III, the characterization of minimum distance error events of the 2-D PR1 channel is investigated by studying the channel impulse response in the spectral domain. By generalizing this concept, some distance properties of channels other than the 2-D PR1 channel can be found (see Section IV). The method of precoding is commonly used in 1-D recording systems to invert the ISI effect of the channel. In Section III-B, we discuss the effect of a precoding scheme on error events for the 2-D PR1 channel. For 1-D channels, the probability of error event and bit error can be bounded from above by the union bound. In Section V, we generalize this concept to 2-D channels. Error-state diagrams and a bounded-depth search algorithm are introduced in Section VI. Section VII gives our conclusions.

Manuscript received March 26, 2007; revised October 23, 2008. Current version published February 25, 2009. This work was supported in part by the National Science Foundation under Grant CCR-0514859, and in part by the National Institute of Standards and Technology, Advanced Technology Program, under ATP Award 70NANB3H3031, in cooperation with InPhase Technologies. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Chicago, IL, June/July 2004.

I. Demirkan is with Broadcom Corporation, Longmont, CO 80503 USA (e-mail: demirkan@broadcom.com).

P. H. Siegel and J. K. Wolf are with the Center for Magnetic Recording Research, University of California, San Diego, La Jolla, CA 92093-0401 USA (e-mail: psiegel@ucsd.edu; jwolf@ucsd.edu).

Communicated by I. Dumer, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2008.2011514

II. The 2-D Channel Model

Consider a 2-D channel with bipolar input array x, channel impulse response h, and output $y = x * h$. AWGN η with zero mean and variance σ^2 is added to the channel output array to obtain the received array $r = x * h + \eta$.

For a channel output array y and its estimated array y', the normalized squared-Euclidean distance is defined as

$$d^2(y, y') \triangleq \sum_{i,j} \left[\left(y_{i,j} - y'_{i,j}\right)/2\right]^2$$

which is taken to be ∞ if the sum is unbounded. Normalized squared-Euclidean distances will be referred to as *squared distances*. The quantity $d^2(y, y')$ can be expressed in terms of the corresponding input arrays x and x', respectively

$$d^2(y, y') = d^2(\epsilon * h, 0) \triangleq \|\epsilon * h\|^2$$

where $\epsilon = (x - x')/2$ is an *error event*, whose elements are commonly represented by the symbols $\{0, +, -\}$. The input arrays x and x' are called the *supporting arrays* of ϵ. The *distance* of ϵ is defined as $d(\epsilon) \triangleq \sqrt{d^2(\epsilon * h, 0)}$.

As in the case of 1-D channels, the error events for 2-D channels are classified as either open or closed. An error event is *closed* if the area of the smallest square region containing nonzero differences is bounded. Error events which are not closed are called *open*. Let $\mathcal{E}_{\texttt{closed}}$ be the set of closed error events, $\mathcal{E}_{\texttt{open}}$ be the set of open error events, and $\mathcal{E} = \mathcal{E}_{\texttt{closed}} \cup \mathcal{E}_{\texttt{open}}$ be their union. We define the *minimum closed event distance*

$$d_{<>} \triangleq \min_{\epsilon \neq 0,\, \epsilon \in \mathcal{E}_{\texttt{closed}}} d(\epsilon)$$

and the *minimum event distance*

$$d_{<} \triangleq \min_{\epsilon \neq 0,\, \epsilon \in \mathcal{E}} d(\epsilon).$$

Here 0 represents the all-zero array.

1-D sequences are often represented in the D-transform domain. Likewise, a 2-D array x can be represented in the (D, E)-transform domain as

$$x(D, E) \triangleq \sum_{i,j} x_{i,j} D^i E^j.$$

In this representation, the impulse response of the 2-D PR1 channel is given by $h_1(D, E) = 1 + D + E + DE$ and the channel input–output relationship becomes $y(D, E) = x(D, E)h(D, E)$. The minimum closed event distance of a 2-D channel can be expressed as

$$d_{<>} = \min_{\epsilon(D,E) \neq 0,\, \epsilon \in \mathcal{E}_{\texttt{closed}}} \|h(D, E)\epsilon(D, E)\|.$$

III. Error Event Characterization on the 2-D PR1 Channel

The minimum and near-minimum distance error events play an important role in determining the performance of an ML detector. Table I shows some error events along with their percentage of bit errors for the 2-D PR1 channel and AWGN. The detector used in this simulation is the IMS algorithm working on 10×10 codewords at 12-dB signal-to-noise ratio (SNR) with bit error rate 1.46×10^{-2}. It is clear that the error events with squared distance 4 dominate the performance. This simulation suggests that the minimum squared distance of this channel is 4. In fact, we will prove in the following subsection that this is indeed the case.

TABLE I
Error Events for the 2-D PR1 Channel

ϵ	$d^2(\epsilon)$	%	ϵ	$d^2(\epsilon)$	%
$[+]$	4	14.95	$[+\ -\ +]^T$	4	5.77
$[+\ -]$	4	11.14	$\begin{bmatrix} + & - \\ - & + \end{bmatrix}^T$	4	3.40
$[+\ -]^T$	4	11.34	$\begin{bmatrix} + & 0 \\ 0 & - \end{bmatrix}^T$	6	0.60
$[+\ -\ +]$	4	6.02	$\begin{bmatrix} 0 & + \\ - & 0 \end{bmatrix}^T$	6	0.44

The minimum and near-minimum distance error events can be characterized by studying spectral properties of the channel transfer function and the corresponding limitations on error coefficients [6, Sec. III-A]. Using this method, the minimum distance error events for the 2-D PR1 channel can be completely characterized.

A. Minimum Distance Error Events

Proposition 1: The minimum closed event distance of the 2-D PR1 channel is 2. All distance-2 closed error events are of the form

$$\epsilon = \begin{bmatrix} + & - & \cdots & \epsilon_{0,n-1} \\ - & + & & \\ \vdots & & & \vdots \\ \epsilon_{m-1,0} & & \cdots & \epsilon_{m-1,n-1} \end{bmatrix}$$

and their negatives. Here $\epsilon_{m-1,0} = +$ (resp., $\epsilon_{0,n-1} = +$) if m (resp., n) is odd; otherwise , $\epsilon_{m-1,0} = -$ (resp., $\epsilon_{0,n-1} = -$). The bottom right entry is determined as $\epsilon_{m-1,n-1} = \epsilon_{m-1,0}\epsilon_{0,n-1}$.

Before giving the proof of Proposition 1, we present a few relevant results. Since the proposition is valid for all m and n, it is sufficient to consider error events with span $m \times n$, i.e., all four edges of error events contain at least one nonzero element. For $m = 1$ and/or $n = 1$, the proof is trivial and will not be discussed here. For $m, n \geq 2$, $y(D, E)$ can be expanded as follows:

$$\begin{aligned} y(D, E) &= (1 + D + E + DE) \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} e_{i,j} D^i E^j \\ &\triangleq y_1(D, E) + y_2(D, E) + y_3(D, E) \end{aligned}$$

where $y_1(D, E)$ contains the terms with a single error coefficient, which correspond to the corners of the error event

$$y_1(D, E) = \epsilon_{0,0} + \epsilon_{m-1,0} D^m + \epsilon_{0,n-1} E^n + \epsilon_{m-1,n-1} D^m E^n. \tag{2}$$

Each of the terms in the second group, $y_2(D, E)$, has two error coefficients corresponding to the edges of the error event

$$\begin{aligned} y_2(D, E) = &\sum_{i=1}^{m-1} (\epsilon_{i,0} + \epsilon_{i-1,0}) D^i + \sum_{j=1}^{n-1} (\epsilon_{0,j} + \epsilon_{0,j-1}) E^j \\ &+ \sum_{i=1}^{m-1} (\epsilon_{i,n-1} + \epsilon_{i-1,n-1}) D^i E^n \end{aligned}$$

$$+\sum_{j=1}^{n-1}(\epsilon_{m-1,j}+\epsilon_{m-1,j-1})D^m E^j.$$

Each of the terms in the third group, $y_3(D,E)$, has four error coefficients corresponding to the middle of the error event

$$y_3(D,E)=\sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(\epsilon_{i,j}+\epsilon_{i-1,j}+\epsilon_{i,j-1}+\epsilon_{i-1,j-1})D^iE^j.$$

Considering the terms in $y_1(D,E)$ and $y_2(D,E)$, one can prove the following result.

Lemma 1:

$$\|y_1(d,e)+y_2(d,e)\|^2 \geq 8-(|\epsilon_{0,0}|+|\epsilon_{m-1,0}|+|\epsilon_{0,n-1}|+|\epsilon_{m-1,n-1}|).$$

Proof: If all corners of an error event are nonzero, then $\|y_1(D,E)\|^2 = 4$. If any of the corners is zero, then $\|y_1(D,E)\|^2$ decreases by 1 while $\|y_2(D,E)\|^2$ increases by 2. To understand this, let $\epsilon_{0,0}=0$. Let $j'\leq n-1$ be the index of the first nonzero entry of the first row, i.e., $\epsilon_{0,j'}\neq 0$ and $\epsilon_{0,j}=0$ for $0\leq j\leq j'$. Similarly, we can define $i'\leq m-1$ as the index of the first nonzero entry on the first column. The first term in $y_1(D,E)$ disappears, whereas $y_2(D,E)$ introduces two new terms with single coefficients which are not zero: $\epsilon_{i',0}D^{i'}$ and $\epsilon_{0,j'}E^{j'}$. A similar proof holds for the other corners. Therefore, if all corners are zero, then $\|y_1(D,E)\|^2=0$ but $\|y_2(D,E)\|^2\geq 8$. □

Proof of Proposition 1: Lemma 1 states that the distance of a closed error event is at least 2. This lower bound can be attained when all the corners of the error event are nonzero, and all other terms in $y_2(D,E)$ and $y_3(D,E)$ are zero. Assuming that $\epsilon_{0,0}=+$, the condition $y_2(D,E)=0$ implies that all edges of the error event have to be in alternating form. The other corner coefficients of the error event are not free and are determined as stated in the proposition. The condition $y_3(D,E)=0$ implies that all internal coefficients have to be in alternating form. This concludes the proof of the proposition. □

B. Effect of Precoding

Consider the channel model described in Section II. Let u be a binary user array at the input to the precoder. A precoder complementing the effect of the 2-D PR1 channel is given by

$$a_{i,j}=u_{i,j}\oplus a_{i-1,j}\oplus a_{i,j-1}\oplus a_{i-1,j-1}.$$

The channel input array is obtained by using bipolar modulation, i.e., $x_{i,j}=2a_{i,j}-1$.

A threshold detector provides an estimate of the channel output array, y', by using the received array r. Using y', u can be estimated by

$$u'_{i,j}=\begin{cases}0, & y'_{i,j}=-4,0,4\\ 1, & y'_{i,j}=-2,2.\end{cases}$$

If the threshold detector makes an error, i.e., $y'\neq y$, then it is directly reflected as an error in user arrays.

On the other hand, an APP detector provides soft information about the channel input x. After hard decisions are made on the soft information, an estimate of the channel input array x' is obtained. By the inverse precoder relation, an estimate of the user array u' can be obtained as

$$u'_{i,j}=a'_{i,j}\oplus a'_{i-1,j}\oplus a'_{i,j-1}\oplus a'_{i-1,j-1} \tag{3}$$

where $a'_{i,j}\triangleq(1+x'_{i,j})/2$. In this case, the *user error* event ϵ_u is defined as $\epsilon_u\triangleq u-u'$. In general, ϵ_u and $\epsilon=(x-x')/2$ may have different sizes and they are related to each other in the following manner.

Proposition 2: The user error event can be expressed as $\epsilon_u=\operatorname{sgn}(\epsilon*h)[\epsilon*h \pmod 2]$, where the multiplication between $\operatorname{sgn}(\epsilon*h)$ and $[\epsilon*h \pmod 2]$ is element-wise and $\operatorname{sgn}(0)=0$.

Proof: The estimated user array given in (3) can be expressed as a convolution $u'=(1+x')/2*h \pmod 2$. The same applies to the user array $u=(1+x)/2*h \pmod 2$. Therefore, the user error event is given by

$$\begin{aligned}\epsilon_u&=\frac{1+x}{2}*h \pmod 2-\frac{1+x'}{2}*h \pmod 2\\ &=\operatorname{sgn}(\epsilon*h)[\epsilon*h \pmod 2].\end{aligned}$$ □

A squared distance 4 error event ϵ of size $m\times n$ corresponds to the following user error event of size $(m+1)\times(n+1)$:

$$\epsilon_u=\begin{bmatrix}\epsilon_{0,0} & 0 & \cdots & \epsilon_{0,n-1}\\ 0 & 0 & & \\ \vdots & & & \vdots\\ \epsilon_{m-1,0} & & \cdots & \epsilon_{m-1,n-1}\end{bmatrix}.$$

The Hamming weight of ϵ_u is always 4 due to the one-to-one relationship between y and u.

IV. Extension to Other 2-D Channels

In the previous section, the distance properties of the 2-D PR1 channel are investigated by using the spectral representations of the signals. This method can be extended to other 2-D channels.

Proposition 3: A 2-D channel with impulse response

$$h=\begin{bmatrix}\alpha & \gamma\\ \beta & \delta\end{bmatrix}$$

achieves the matched-filter-bound, i.e.,

$$d_{<>}=\sqrt{\alpha^2+\beta^2+\gamma^2+\delta^2}.$$

The proof of this proposition is similar to that of Proposition 1.

Proposition 4: For a general 2-D channel with impulse response $h=\{h_{i,j}\}_{i=0,j=0}^{p-1,q-1}$, the minimum closed event distance can be bounded from below by

$$d_{<>}\geq\sqrt{h_{0,0}^2+h_{p-1,0}^2+h_{0,q-1}^2+h_{p-1,q-1}^2}.$$

In this case, the matched-filter-bound may not be achieved for some channels, as illustrated in the following example.

Example 1: Consider a 2-D channel with impulse response

$$h = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

The squared distance of $\epsilon(D, E) = 1 - D$ is 6, but the matched-filter-bound of this channel is 9.

V. Bounds on the Probability of Bit Error

Chugg [5] proved that if the entries of input array x (or u for the precoded case) are equally probable and independent, the bit error probability under ML detection and AWGN can be bounded from above by the union bound

$$P_b \leq \sum_d K_d Q\left(\frac{d}{2\sigma}\right).$$

Here $Q(\cdot)$ is the complementary distribution function of a zero-mean and uni-variance Gaussian random variable, and K_d is the average multiplicity of bit errors of distance d, given by

$$K_d = \sum_{\epsilon : d(\epsilon)=d} w(\epsilon) 2^{-w(\epsilon)}$$

where $w(\epsilon)$ is the Hamming weight of ϵ. The error events counted in this upper bound have to be *connected*, as defined below, and only one of the different shifts of ϵ should be taken into account.

Definition 1: An error event ϵ is *connected* if the error event cannot be divided into two separate error events ϵ_1 and ϵ_2 such that $d^2(\epsilon) = d^2(\epsilon_1) + d^2(\epsilon_2)$. The error events which are not connected are called *disconnected.*

Connected error events are referred to as *fundamental error* patterns in [5]. A sufficient condition for an error events to be connected is also given in that paper. An easier way to determine whether the error event is connected or disconnected is based upon the notion of *connected entries* which we now define.

Definition 2: Let $h = \{h_{i,j}\}_{i=0,j=0}^{p-1,q-1}$ be the impulse response of a 2-D channel. Let h' be a $p \times q$ indicator matrix whose entries are defined as

$$h'_{i,j} = \begin{cases} 1, & \text{if } h_{i,j} \neq 0 \\ 0, & \text{otherwise.} \end{cases}$$

For an error event ϵ of size $m \times n$, two nonzero entries (i_1, j_1) and (i_2, j_2) are said to be *connected* to each other if the following condition is true:

$$\|a * h'\|^2 < 2\|h'\|^2$$

where a is an $m \times n$ matrix such that $a_{i_1,j_1} = 1$, $a_{i_2,j_2} = -1$, and other entries of a are zero. If $\|a * h'\|^2 = 2\|h'\|^2$, then the entries (i_1, j_1) and (i_2, j_2) are called *disconnected*. The entries (i_1, j_1) and (i_2, j_2) cannot be connected if $|i_2 - i_1| + 1 > p$ or $|j_2 - j_1| + 1 > q$.

Definition 3: A set of nonzero entries ϵ' in an error event ϵ is called a *connected component* of ϵ if every two nonzero entries of ϵ' are connected directly or via other nonzero entries in ϵ'.

Definitions 1–3 directly imply the following result.

Lemma 2: An error event is connected if and only if it has one connected component.

For a precoded system, K_d becomes

$$K_d = \sum_{\epsilon : d(\epsilon)=d} w(\epsilon_u) 2^{-w(\epsilon)}$$

where ϵ_u is the user error event corresponding to ϵ.

The bit error multiplicity generating function is defined as

$$g_b(z) = \sum_{d \in \mathcal{D}} K_d z^{d^2}$$

where $\mathcal{D}$ is the set of all distances, and K_d and z take nonnegative values.

VI. A Bounded-Depth Search Algorithm

Error-state diagrams for 1-D channels cannot be directly generalized to 2-D channels since there are no convenient graph-based descriptions of such channels. However, when the size of the error event is fixed in either of the dimensions, error-state diagrams can be described as 1-D systems using a higher order alphabet. In this section, we propose a bounded-depth search algorithm for determining closed error events of size $m \times n$ for 2-D channels with impulse response of size $p \times q$. In order to avoid redundant repetitions of error events, the following conditions are imposed on error events: 1) the edges of an error event contain at least one nonzero element and 2) error events are required to be connected.

An error state diagram for a 2-D channel with impulse response $h = \{h_{i,j}\}_{i=0,j=0}^{p-1,q-1}$ can be represented as a labeled graph. A *state* σ in this graph is a sequence of $p-1$ symbols $(s_1, s_2, \ldots, s_{p-1})$ from the alphabet Σ, which is the set of all row vectors of length n with entries $\{0, +, -\}$. Alternatively, the state σ can be represented as a $(p-1) \times n$ array

$$\sigma = \begin{bmatrix} s_1 \\ \vdots \\ s_{p-1} \end{bmatrix}.$$

Therefore, there are $3^{(p-1)n}$ states in the error-state diagram. An edge e has the initial state $\sigma(e) = (s_1, \ldots, s_{p-1})$, the terminal state $\tau(e) = (s_2, \ldots, s_p)$, and the label $\mathcal{L}(e) = s_p$. A path ρ in the error-state diagram is a finite sequence of edges $(e_1, \ldots, e_l)$ such that $\sigma(e_{i+1}) = \tau(e_i)$ for $1 \leq i < l$. A closed error event ϵ of size $m \times n$ corresponds to a path $\rho = (e_1, \ldots, e_{m+p-1})$ in the error-state diagram that starts and ends at the all-zero state $\sigma = (z, \ldots, z)$ without an intermediate visit to that state, where z is the all-zero row vector of length n. The sequence of edge labels for the edges of ρ is given by

$$\mathcal{L}(\rho) = (s_1, \ldots, s_m, \underbrace{z, \ldots, z}_{p-1}).$$

If ρ represents ϵ, the sequence of edge labels $(s_1, \ldots, s_m)$ should give the error event, i.e., $s_i = \epsilon_{i,*}$ for all $1 \leq i \leq m$. There are $p-1$ appended edges with label z for the termination of the error event. Therefore, the relationship between paths and error events is one-to-one. The error event corresponding to the path ρ is denoted as $\epsilon(\rho)$. The distance of a path $d(\rho)$ is defined as the distance of the corresponding error event $d(\epsilon(\rho))$. Note that the error-state diagrams defined in this way are equivalent to 1-D error-state diagrams with memory $p-1$ and alphabet size 3^n.

The bounded-depth search algorithm searches for the closed and connected error events of size $m \times n$ whose distances are not larger than a specified limit $d_{\max}$. Let $\rho^{(l-1)}$ be a path of length $l-1$ for $1 < l \le m+p-1$. If an edge e with label $s \in \Sigma$ is appended to this path, the algorithm checks the following conditions on the extended path $\rho^{(l)} = (\rho^{(l-1)}, e)$.

- $d(\rho^{(l)}) \le d_{\max}$.
- When $l = m$, the error event $\epsilon(\rho^{(l)})$ contains at least one nonzero entry along its edges.
- The error event $\epsilon(\rho^{(l)})$ is connected.

If any of these checks fails, then the path $\rho^{(l)}$ is called *invalid* and the algorithm will not extend the path $\rho^{(l)}$. If all of these checks are satisfied, then the path $\rho^{(l)}$ is called a *valid* path.

Let $C^{(l)}$ and $V^{(l)}$ be the sets of candidate and valid paths at level l for $1 < l \le m+p-1$, respectively. All candidate paths at level l are obtained by extending the valid paths at level $l-1$ as follows.

- For level $l = 1$, the set of candidate paths is

$$C^{(1)} = \{e : \sigma(e) = (z, \ldots, z),\ \mathcal{L}(e) \in \Sigma \backslash \{z\}\}.$$

  Excluding the all-zero state guarantees that the first row of the error event contains at least one nonzero entry.
- For levels $1 < l < m$, the set of candidate paths is

$$C^{(l)} = \left\{(\rho^{(l-1)}, e) : \rho^{(l-1)} \in V^{(l-1)}, \mathcal{L}(e) \in \Sigma\right\}.$$

- For level $l = m$, the set of candidate paths is

$$C^{(m)} = \left\{(\rho^{(m-1)}, e) : \rho^{(m-1)} \in V^{(m-1)}, \mathcal{L}(e) \in \Sigma \backslash \{z\}\right\}.$$

  Excluding the all-zero state guarantees that the last row of the error event contains at least one nonzero entry.
- For levels $m < l \le m+p-1$, the valid paths are extended by the edge with label z to terminate the error event, i.e.,

$$C^{(l)} = \left\{(\rho^{(l-1)}, e) : \rho^{(l-1)} \in V^{(l-1)},\ \mathcal{L}(e) = z\right\}.$$

The algorithm performs three updates and checks on $C^{(l)}$ to obtain $V^{(l)}$ as discussed in the following subsections.

A. Updating and Checking Squared Distance

Let $d(\rho^{(l)})$ and $d(\rho^{(l-1)})$ be the distances of paths $\rho^{(l-1)}$ and $\rho^{(l)}$, respectively. We assume that all entries outside the spans of the error events are simply zero. We can compute the channel output by considering the augmented state $(\sigma(e), s)$, which is padded with zeros

$$\sigma' = \begin{bmatrix} 0^{(p-1)\times(q-1)} & \sigma(e) & 0^{(p-1),(q-1)} \\ 0^{1\times(q-1)} & s & 0^{1,(q-1)} \end{bmatrix}$$

where $0^{m\times n}$ is the all-zero matrix of size $m \times n$. The channel output corresponding to the channel input e is given by

$$\mu_l = \sum_{t=0}^{n-1} \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} h_{i,j} \sigma'_{p-1-i, t+q-1-j}.$$

Therefore, the squared distance of the path $\rho^{(l)}$ is updated as

$$d^2(\rho^{(l)}) = d^2(\rho^{(l-1)}) + \mu_l^2.$$

If $d(\rho^{(l)}) > d_{\max}$, then the distance check on $\rho^{(l)}$ fails.

B. Updating and Checking Edge Indicator Functions

Let $I_1(\rho^{(l)})$ be the left edge indicator function for the path $\rho^{(l)}$ indicating whether the left edge of $\epsilon(\rho^{(l)})$ contains a nonzero entry. The left edge indicator function can be updated as follows:

$$I_1(\rho^{(l)}) = \begin{cases} 1, & \text{if } I_1(\rho^{(l-1)}) = 1 \text{ or } s_0 \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

where s_0 is the first entry of the row vector s. The right edge indicator variable $I_2(\rho^{(l)})$ can be defined and updated similarly. For a path $\rho^{(m)}$ at level $l = m$, if $I_1(\rho^{(m)}) = 0$ or $I_2(\rho^{(m)}) = 0$, then the check on $\rho^{(m)}$ fails.

C. Updating and Checking Connection Map

A connected component of an error event is a set of connected entries in an error event. Connected components of an error event can be distinctly numbered with positive integers. We define the *group number* of a nonzero entry in a path label sequence $\mathcal{L}(\rho^{(l)})$ as the connected component number to which that entry belongs. A *connection map* of a path $\rho^{(l)}$ is an $(l \times n)$ array whose entries are the group numbers of the nonzero entries. Zero entries in $\mathcal{L}(\rho^{(l)})$ are assumed to have group number zero. The maximum group number of $\mathcal{L}(\rho^{(l)})$ is denoted by $\varphi(\rho^{(l)})$. Let $M(\rho^{(l-1)})$ and $M(\rho^{(l)})$ be the connection maps for the paths $\rho^{(l-1)}$ and $\rho^{(l)}$, respectively. For the overlapping edges of $\rho^{(l-1)}$ and $\rho^{(l)}$, the connection groups are the same, i.e, $M(\rho^{(l)})_{i,*} = M(\rho^{(l-1)})_{i,*}$ for $0 \le i < l-1$. For each nonzero entry j in the last row of $M(\rho^{(l)})$, we need to determine which groups are connected to that entry $[M(\rho^{(l)})]_{l-1,j}$. Let P_j be the set of entries in $M(\rho^{(l)})$ connected to the $(l-1, j)$ entry and G_j be the set of group numbers of the entries in P_j. For the $(l-1, j)$th entry, the connection map $M(\rho^{(l)})$ is updated as follows.

- If $G_j = \emptyset$, i.e, the $(l-1, j)$ entry of $\mathcal{L}(\rho^{(l)})$ is not connected to any connected components, then a new group number is assigned to this entry: $[M(\rho^{(l)})]_{l-1,j} = \varphi(\rho^{(l)}) + 1$ and $\varphi(\rho^{(l)}) \leftarrow \varphi(\rho^{(l)}) + 1$.
- If $G_j \neq \emptyset$, then $[M(\rho^{(l)})]_{l-1,j} = \max\{G_j\}$ and for all nonzero entries (a, b), set $[M(\rho^{(l)})]_{a,b} = \max\{G_j\}$. In this way, all connected components whose numbers are in G_j are merged into the single connected component with number $\max\{G_j\}$.

There are two different types of checks associated with connection maps.

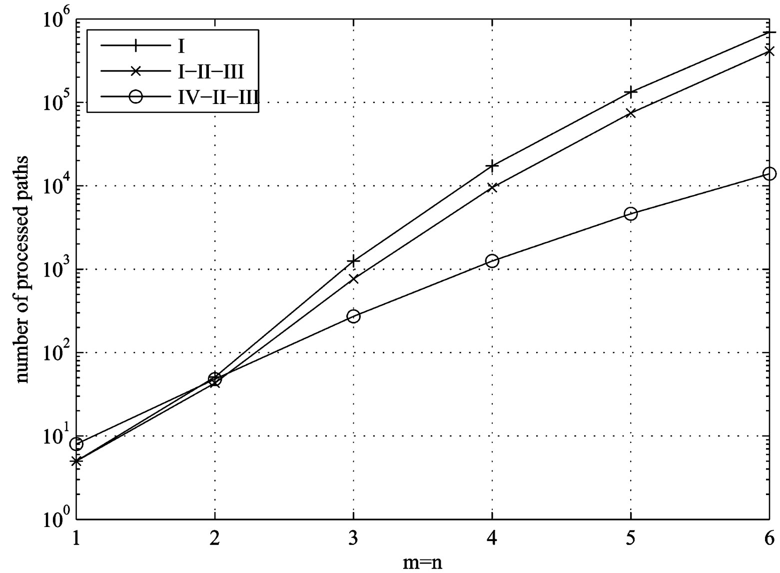

Fig. 1. The complexity of the algorithm for the 2-D PR1 channel for $d^2_{\max} = 6$. Checks I, II, and III refer to the squared distance, edge indicator functions, and the connection map, respectively. Check IV refers to the modified squared distance with $d^2_{\min} = 2$ for all m and n.

- No group can terminate without merging with another group. That is, the rows $[M(\rho^{(l)})]_{i,*}$, $0 \le i \le l - p$, cannot contain a group number which does not occur in the rows $[M(\rho^{(l)})]_{i,*}$, $l - p + 1 \le i < l$. This check is applied for the paths at level $l \le m$.
- All groups have to merge into a single group at level $l = m$; otherwise, the error event will be disconnected. That is, all nonzero entries of $M(\rho^{(m)})$ should be $\varphi(\rho^{(m)})$.

If any of these checks fails, then the algorithm will not expand the path $\rho^{(l)}$.

D. Simulation Results

Simulation results indicate that applying all three checks discussed in previous subsections considerably reduces undesired repetition of error events for a fixed maximum distance $d_{\max}$.

By using simulation results for the 2-D PR1 channel, the bit error multiplicity generating function can be bounded from below by

$$g_b^{\text{unprec}}(z) \ge 3.800z^4 + 34.73z^6 + 311.1z^8 + 2049z^{10}$$
$$g_b^{\text{prec}}(z) \ge 6.301z^4 + 43.30z^6 + 365.9z^8 + 2477z^{10}$$

for the unprecoded and precoded cases, respectively. Here "$\ge$" signifies that the coefficient of each term on the right is less than the corresponding one on the left. The computed values of K_2's for both cases are close to the analytical values.

E. Reduction of Complexity

A measure for the complexity of the bounded-depth search algorithm is the number of paths that are extended and checked. Implementing the checks for the edge indicator functions and the connection map does not significantly reduce the complexity of the algorithm as shown in Fig. 1 for the 2-D PR1 channel. However, the complexity of the algorithm can be reduced by modifying the check corresponding to the squared distance. For some 2-D channels, such as the 2-D PR1 channel, a significant part of the squared distance is caused by the nonzero corner and edge coefficients of the error events. Therefore, for a path $\rho^{(m+p-1)}$ representing an error event, the initial and final levels contribute to its distance considerably.

For an error event of size $m \times n$ represented by the path $\rho^{(m+p-1)}$, the *minimum squared* distance is given by

$$d^2_{\min} \triangleq \min_{\rho^{(m+p-1)}} d^2(\rho^{(m+p-1)}) - d^2(\rho^{(m)})$$

where $\rho^{(m+p-1)}$ is the extension path of $\rho^{(m)}$ with the edges with symbol z. The minimum squared distance only depends on the size of the error event and the channel impulse response. The minimum squared distance can be found by enumerating all paths of length $p-1$, which can start with any state but have to be extended with the edge with symbol z. In this way, the terminal state of the path becomes the all-zero state. Therefore, there are $3^{(p-1)n}$ such paths. The check related to the squared distance can be modified by using different maximum squared distance for different path levels

$$\tilde{d}^2_{\max} = \begin{cases} d^2_{\max} - d^2_{\min}, & \text{for } 1 \le l \le m \\ d^2_{\max}, & \text{for } m < l \le m + p - 1. \end{cases}$$

As shown in Fig. 1, this method reduces the complexity of the algorithm significantly for the 2-D PR1 channel. This method is also observed to be efficient for other channels and $d^2_{\max}$ combinations.

VII. Conclusion

In this paper, minimum distance and near-minimum distance closed error events of the 2-D PR1 channel are characterized. The effect of precoding on error events is also investigated for this channel. Some distance properties valid for the 2-D PR1 channel also apply to 2-D channels with impulse response of size 2×2. Error events for 2-D channels can be generated by using the bounded-depth search algorithm developed here. The results of this algorithm are observed to be consistent with the analytical results.

References

[1] A. H. J. Immink *et al.*, "Signal processing and coding for two-dimensional optical storage," in *Proc. Global Telecommunications Conf. (GLOBECOM'03)*, San Francisco, CA, Dec. 2003, pp. 3904–3908.

[2] I. Demirkan and J. K. Wolf, "Block codes for the hard-square model," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2836–2848, Aug. 2005.

[3] E. Ordentlich and R. Roth, "On the Computational Complexity of 2D Maximum Likelihood Sequence Detection," HP Labs, 2006, Tech. Rep. HPL-2006-69.

[4] M. Marrow, "Detection and Modelling of 2-Dimensional Signals," Ph.D. dissertation, University of California, San Diego, La Jolla, 2004.

[5] K. M. Chugg, "Performance of optimal digital page detection in a two-dimensional ISI/AWGN channel," in *Proc. Asilomar Conf. Signals, Systems and Comp.*, Pacific Grove, CA, Nov. 1996, vol. 2, pp. 958–962.

[6] R. Karabed, P. H. Siegel, and E. Soljanin, "Constrained coding for binary channels with high intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1777–1797, Sep. 1999.

[7] A. D. Weathers, S. A. Altekar, and J. K. Wolf, "Distance spectra for PRML channels," *IEEE Trans. Magn.*, vol. 33, no. 5, pp. 2809–2811, Sep. 1997.

Ismail Demirkan (S'98–M'06) was born in Turkey in 1979. He received the B.S. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey in 2001. He received his M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, San Diego, in 2003 and 2006, respectively.

He is currently employed by Broadcom Corporation, Longmont, CA. His research interests include coding and signal processing for magnetic and optical recording systems.

Paul H. Siegel (M'82–SM'90–F'97) was born in Berkeley, CA, in 1953. He received the B.S. degree in mathematics in 1975 and the Ph.D. degree in mathematics in 1979, both from the Massachusetts Institute of Technology, Cambridge. He held a Chaim Weizmann fellowship during a year of postdoctoral study at the Courant Institute, New York University.

He joined the research staff at IBM in 1980. From 1984 through 1993, he was manager of the Signal Processing and Coding project at the IBM Almaden Research Center in San Jose, CA, and in 1994, he was named Manager of the Mathematics and Related Computer Science Department at Almaden. He was a Visiting Associate Professor at the University of California, San Diego (UCSD) while at the Center for Magnetic Recording Research (CMRR) during the 1989–1990 academic year. He joined the faculty at UCSD in July 1995, and he is currently Professor of Electrical and Computer Engineering in the Jacobs School of Engineering. He is affiliated with the California Institute of Telecommunications and Information Technology (Calit2), the Center for Wireless Communications (CWC), and the CMRR, where he holds an endowed chair and currently serves as Director. His primary research interest is the mathematical foundations of signal processing and coding, especially as applicable to digital data storage and communications. He holds several patents in the area of coding and detection for digital recording systems, and was named a Master Inventor at IBM Research in 1994.

Prof. Siegel was corecipient, with R. Karabed, of the 1992 IEEE Information Theory Society Paper Award for the paper "Matched Spectral Null Codes for Partial Response Channels," and he shared the 1993 IEEE Communications Society Leonard G. Abraham Prize Paper Award with B. Marcus and J. K. Wolf for the paper "Finite-State Modulation Codes for Data Storage." Along with his doctoral students J. B. Soriaga and H. D. Pfister, he received the 2007 IEEE Communications Society, Data Storage Technical Committee Best Paper Award in Signal Processing and Coding for Data Storage for the paper " Determining and Approaching Achievable Rates of Binary Intersymbol Interference Channels using Multistage Decoding." He was a Member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996. He served as Co-Guest Editor of the May 1991 Special Issue on Coding for Storage Devices of the IEEE TRANSACTIONS ON INFORMATION THEORY, Associate Editor for Coding Techniques from 1992 to 1995, and Editor-in-Chief from 2001 to 2004. He was also Guest Editor-in-Chief of the May/September 2001 double- issue of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on The Turbo Principle: From Theory to Practice. He is a member of Phi Beta Kappa. He was elected to the National Academy of Engineering in 2008, "for the invention and development of advanced coding techniques for digital recording systems."

Jack Keil Wolf (S'54–M'60–F'73–LF'97) received his B.S.E.E. degree from the University of Pennsylvania, Philadelphia, in 1956, and the M.S.E., M.A. and Ph.D. degrees from Princeton University, Princeton, NJ, in 1957, 1958, and 1960, respectively.

He was a member of the Electrical Engineering Department at New York University, New York City, from 1963 to 1965, and the Polytechnic Institute of Brooklyn, Brooklyn, NY, from 1965 to 1973. He was Chairman of the Department of Electrical and Computer Engineering at the University of Massachusetts, Amherst, from 1973 to 1975, and was Professor there from 1973 to 1984. Since 1984 he has been a Professor of Electrical and Computer Engineering and a member of the Center for Magnetic Recording Research at the University of California, San Diego (UCSD), La Jolla, CA where his is now the Stephen O. Rice Professor of Magnetics. He also holds a part-time appointment at Qualcomm, Inc., San Diego, CA. His current interest is in signal processing for storage systems.

Dr. Wolf has been a Fellow of the IEEE since 1973. From 1971 to 1972, he was an NSF Senior Postdoctoral Fellow, and from 1979 to 1980 he held a Guggenheim Fellowship. In 1993, he was elected to the National Academy of Engineering. He was the recipient of the 1990 E. H. Armstrong Achievement Award of the IEEE Communications Society and was corecipient of the 1975 IEEE Information Theory Group Paper Award for the paper "Noiseless Coding for Correlated Information Sources" (coauthored with D. Slepian). He served on the Board of Governors of the IEEE Information Theory Group from 1970 to 1976 and from 1980 to 1986. He was president of the IEEE Information Theory Group in 1974. He was International Chairman of Committee C of URSI from 1980 to 1983. He was the recipient of the 1998 IEEE Koji Kobayashi Computers and Communications Award. In May 2000, he received a UCSD Distinguished Teaching Award. In June 2001, he received the 2001 Claude E. Shannon Award from the IEEE Information Theory Society. He was the recipient of the 2004 Richard W. Hamming Medal. He is a member of the National Academy of Engineering.