Constrained Coding for Binary Channels with High Intersymbol Interference

Razmik Karabed, Paul H. Siegel, *Fellow, IEEE,* and Emina Soljanin, *Member, IEEE*

***Abstract*—Partial-response (PR) signaling is used to model communications channels with intersymbol interference (ISI) such as the magnetic-recording channel and the copper-wire channel for digital subscriber lines. Coding for improving noise immunity in higher order partial-response channels, such as the "extended" class-4 channels denoted EPR4, E^2PR4, E^3PR4, has become an important subject as the linear densities in magnetic recording approach those at which these partial-response channels are the best models of real channels. In this paper, we consider partial-response channels for which ISI is so severe that the channels fail to achieve the matched-filter bound (MFB) for symbol error rate, assuming maximum-likelihood decoding. We show that their performance can be improved to the MFB by high-rate codes based on constrained systems, some of which may even simplify the Viterbi detectors relative to the uncoded channels. We present several examples of high-rate constrained codes for E^2PR4 and E^3PR4 channels and evaluate their performance by simulation.**

***Index Terms*—Constrained coding, magnetic recording, partial response.**

I. Introduction

PARTIAL response (PR) signaling is used to model communications channels with intersymbol interference (ISI) such as the magnetic-recording channel and the copper-wire channel for digital subscriber lines. This paper is mainly concerned with high-density, binary-input, magnetic-recording systems. Partial-response channels with the transfer function of the form $h_N(D) = (1-D)(1+D)^N, N \geq 1$, have been shown to closely match magnetic-recording channels for a range of linear recording densities [59]. Detectors for these channels, known as PRML detectors, employ PR equalization followed by maximum-likelihood (ML) sequence detection matched to the equalized channel. Most of the original generation of commercially available PRML detectors employed equalization to the PR4 channel $(N=1)$ and the Viterbi ML detection matched to this channel with additive white Gaussian noise [63], [56]. At higher linear densities, the channel transfer function changes and the PR4 equalization alters the spectral density of the noise by enhancing the high-frequency components. It has been shown, analytically [51] and by simulation [47], that this in turn causes a loss in performance of the Viterbi detector, which is an optimal algorithm in the case of white noise. This loss is referred to as the equalization loss. To reduce the equalization loss, a PR polynomial $h_N(D)$ that more closely matches the channel should be chosen as the equalization target. For current linear recording densities the appropriate choice is $(1-D)(1+D)^2$ or EPR4, and at higher densities $(1-D)(1+D)^3$ or E^2PR4 becomes better. Future systems may incorporate E^3PR4 equalization, which corresponds to $(1-D)(1+D)^4$. This paper is particularly concerned with systems using E^2PR4 and E^3PR4.

The general problem of coding for binary-input constrained PR channels has been approached in several ways. One approach, initiated by Wolf and Ungerboeck in [66], uses error-control codes with good Hamming distance on a precoded channel. (A related technique, employing a channel post-coder, was presented by Calderbank, Heegard, and Lee [14].) This approach was justified by showing that the minimum squared-Euclidean distance of a code on a precoded ISI channel of the form $(1 \pm D^N)$ is lower-bounded by its minimum Hamming distance. Codes with good Hamming distance are known and can be found in numerous tables available in the literature, as for example in Lin and Costello [30, p. 331]. Hole [20] and Hole and Ytrehus [21] have refined this technique and found codes with improved properties and reduced detector complexity for the precoded PR4 channel.

A second approach, as described by Karabed and Siegel in [25], is to use codes with spectral nulls at the frequencies of the spectral nulls of the channel. This design approach was justified by showing that matching of the code and the channel spectral nulls provides a substantial increase in the minimum Euclidean distance for an important class of partial response channels. These codes are known as *matched spectral null* (MSN) codes.

Several approaches were proposed for simplified decoding of MSN codes on the PR4 channel. Knudson *et al.* [28] introduced a form of concatenated detection for MSN-coded PR4, sometimes referred to as post-processing. (A related form of post-processing was applied to EPR4 detection by Wood [68] and Knudson [16].) Fredrickson *et al.* [18] and Mittelholzer *et al.* [37] developed MSN codes and related codes with decoders characterized by time-varying trellis structures. Another approach to simplification of MSN encoders and decoders on the PR4 channel was also recently proposed by Soljanin in [53] and [54].

Based upon these techniques, a number of proposals for performance-improving codes for the PR4 channel have been

Manuscript received March 29, 1997; revised February 18, 1999. The work of P. H. Siegel was supported in part by the National Science Foundation under Grant NCR-9612802.

R. Karabed is with the Siemens Microelectronics, Scotts Valley, CA 95066 USA.

P. H. Siegel is with the University of California at San Diego, La Jolla, CA 92093 USA.

E. Soljanin is with the Mathematical Sciences Research Center, Bell Labs, Lucent Technologies, Murray Hill, NJ 07974 USA.

Communicated by N. Seshadri, Associate Editor for Coding Techniques.

Publisher Item Identifier S 0018-9448(99)05882-4.

0018–9448/99$10.00 © 1999 IEEE

made, and VLSI prototypes of MSN codes, in particular, have been experimentally evaluated [46], [13]. However, none has yet been incorporated widely into commercial magnetic-recording systems. The primary reason has been the continuing emphasis in the disk storage industry upon increasing the linear density (i.e., the density along recorded tracks). As mentioned above, this has led to the introduction of EPR4 equalization, as well as to the use of rate $16/17$, $(0, G/I)$ constrained codes. For trellis-coded PR4 systems, which, for reasons of complexity, would be limited to a rate $8/9$ code, the gain in minimum distance is generally offset by the equalization loss at these higher linear densities; see, for example, [47], [9].

Although MSN codes are in principle applicable to higher order partial-response channels, the complexity of the encoding and decoding functions, as well as that of the Viterbi detector, present even more of an obstacle than in the case of the dicode or PR4 channels. Therefore, alternative approaches to designing, implementing, and detecting MSN codes have been proposed.

For example, Soljanin and Agazzi [55] described an MSN coding scheme for the EPR4 channel that was closely related to the MSN coding scheme proposed by Knudson *et al.* for the dicode and PR4 channels [28]. The system used interleaved encoding of balanced codewords to generate spectral nulls at zero frequency and the Nyquist frequency. The concatenated decoder comprised an EPR4 Viterbi detector, followed by a post-processor designed to detect and correct the most probable errors producing violations of the spectral null conditions in the detected codewords. This suboptimal scheme allows much simpler implementation of the detector with negligible performance loss.

Despite these advances, there remained major obstacles to deploying trellis-coded PRML systems based upon these code design approaches. The purpose of this paper is to present the details of an alternative approach to designing distance-enhancing, constrained codes for higher order PR channels, as outlined in [26], [51], [49], and [52]. The design methodology finds its motivation in the properties of the well-known $(d, k) = (1, \infty)$ constraint. Wood [67] observed that a $(d, k) = (1, k)$ code increased the minimum squared-Euclidean distance of the partial response class 2 (PR2) constraint, corresponding to the PR polynomial $h(D) = (1 + D)^2$, by a factor of 1.5. In [6], Behrens and Armstrong observed that the $(d, k) = (1, 7)$ constraint increased the minimum squared-Euclidean distance on the E^{2}PR4 channel by a factor of 2.5. Moreover, incorporating the $(d, k) = (1, \infty)$ constraint into the detector trellis actually reduced the required number of trellis states.

Karabed and Siegel [26] recognized that the distance-enhancing properties of the $(1, \infty)$ constraint on the E^{2}PR4 channel could be explained by means of a characterization of the channel input error sequences that generated output error events with small Euclidean distance. Using such a characterization, they identified a family of constraints with the same distance-enhancing properties as the $(d, k) = (1, \infty)$ constraint, but with substantially higher capacity. The constraints were defined by specifying a "forbidden list" $\mathcal{F}$ of code strings whose omission ensured that the small distance error events would never occur. In particular, the constraint with $\mathcal{F} = \{101\}$, designated by $X_{\{101\}}$, had Shannon capacity $C \approx 0.8113$, permitting the design of a rate $4/5$ sliding-block code that provided substantially the same performance improvement on the E^{2}PR4 channel as the rate $2/3$ $(1, 7)$ code. The constraint also reduced the number of detector states relative to uncoded E^{2}PR4, although not to the same extent as the $(d, k) = (1, \infty)$ constraint.

(McEwen and Wolf [36] extended the code design in [6] in a different direction. They designed $(1, k)$ codes for E^{2}PR4 with squared distance 18 by using the modulo-N construction—-a variation of MSN codes—introduced by Fredrickson [17].)

Soljanin [51] investigated on-track and off-track distance properties of PR4, EPR4, and E^{2}PR4 channels, observing independently that the on-track and off-track performance could be improved by constraining the inputs to prevent the occurrence of dominant input error sequences. In particular, Soljanin showed that a code which limited the number of consecutive transitions in the recorded sequence to at most three, such as the binary complement of the tape-industry standard rate $8/9$ $(0, 3)$ code [44] provided off-track performance gains on the EPR4 channel. She also demonstrated that the $(1, \infty)$ constraint, by limiting the number of consecutive transitions to one, also improved the off-track performance of the E^{2}PR4 channel.

Building upon the results in [26] and [51], this paper provides the details of a general framework for constructing distance-enhancing codes for ISI channels, with a particular emphasis upon higher order partial-response channels, as outlined in [49] and [52]. Conceptually, the proposed code design technique involves the following steps: a) Determine the set of input error sequences $\mathcal{E}$ corresponding to the most probable error events for the specified ISI channel. b) Determine a system of constrained sequences $\mathcal{C}$ such that the difference between any two sequences in $\mathcal{C}$ does not belong to $\mathcal{E}$. c) Construct an efficient, practical encoder and decoder for this constraint. d) Design a sequence detector incorporating the channel and code constraints.

The previously described MSN and runlength-limited trellis coding techniques fall within this framework, although steps a) and b), and the identification of broader classes of distance-enhancing constraints they reveal, were not explicitly carried out in their development. When the proposed approach is applied to the E^{2}PR4 and higher order channels, it is particularly effective, because these channels possess a property that simplifies the design of high-rate, distance-enhancing codes, as we now briefly describe.

The minimum (merged) distance of the uncoded binary channel with transfer function $h(D)$ is defined by

$$d^2_{<>} = \min_{\epsilon(D) \neq 0} \|h(D)\epsilon(D)\|^2$$

where

$$\epsilon(D) = \sum_{i=0}^{l-1} \epsilon_i D^i, \qquad \epsilon_i \in \{-1, 0, 1\},\ \epsilon_0 = 1,\ \epsilon_{l-1} \neq 0$$

is the polynomial corresponding to a normalized input error sequence $\boldsymbol{\epsilon} = \{\epsilon_i\}_{i=0}^{l-1}$ of length $l < \infty$, and the squared

norm of a polynomial is defined as the sum of its squared coefficients. The minimum distance is bounded from above by $||h(D)||^2$, the energy in the channel impulse response. This quantity is the signal energy at the output of a filter matched to the channel impulse response. The inequality

$$d^2_{<>} \leq ||h(D)||^2 \tag{1}$$

is therefore known as the *matched-filter bound* (MFB), and we will denote the upper bound by

$$d^2_{\mathrm{MFB}} \stackrel{\mathrm{def}}{=} ||h(D)||^2.$$

When (1) is satisfied with equality, we say that the channel *achieves the MFB*. This occurs when the error sequence of length $l = 1$, i.e., $\epsilon(D) = 1$, is in the set

$$\arg \min_{\epsilon(D)\neq 0} ||h(D)\epsilon(D)||^2. \tag{2}$$

For the E^2PR4 and higher order partial-response channels, the inequality in (1) is strict, $d^2_{\min} < ||h(D)||^2$. We say that they fail to achieve the MFB. For these channels, one can construct distance-enhancing constrained codes that increase the minimum distance to the MFB, preventing the occurrence of error sequences $\epsilon(D)$ for which

$$d^2_{<>} \leq ||h(D)\epsilon(D)||^2 < ||h(D)||^2. \tag{3}$$

In fact, the constraint can achieve the distance gain by precluding specified error strings that have length $l \geq 2$. If the set of forbidden error strings is finite, then it is a relatively straightforward task to identify distance-enhancing "forbidden-list" constraints and to construct practical, efficient codes using, for example, the state splitting algorithm of Adler, Coppersmith, and Hassner [1], as was done in [26].

Moreover, as illustrated by the $(d, k) = (1, \infty)$ and $X_{\{101\}}$ constraints, when the channel itself has large memory, the memory of the constraint needed to remove the forbidden strings may be comparable to, or smaller than, the channel memory. This implies that the complexity of a maximum-likelihood sequence detector trellis will be on the order of, or even smaller than, that of the uncoded channel. In the remainder of this paper, the details and applications of this code design methodology are presented. Section II presents the models for higher order class-4 partial-response channels and a summary of their error-probability performance, under the standard, simplifying assumption of additive, white Gaussian noise. Section III is concerned with the analysis of distance properties of ISI channels, in particular of E^2PR4 and E^3PR4 channels. We refer to the procedure of determining the input error sequences corresponding to specified channel output distances as "distance properties analysis" or "error-event characterization." We will illustrate two approaches to this procedure. The first, an analytic approach, makes use of the spectral properties of the channel transfer function and moment conditions that are consequently satisfied by the channel output sequences. The second approach [7], [62] uses a modified form of the channel error-state diagram [61] and a bounded depth search algorithm that is readily amenable to computer implementation. (This computational algorithm has recently been refined and applied to more extensive characterization of error events for high order partial response channels [2]–[4].)

In Section IV, we discuss the problem of identifying distance-enhancing constraints from the error event characterization, as well as the design of constrained codes, with particular attention to the E^2PR4 and E^3PR4 channels. The distance-enhancing properties of previously developed performance-improving codes are discussed. We then describe several new classes of constraints based upon lists of forbidden input error sequences and corresponding lists of forbidden channel input sequences. Among these are time-varying forbidden lists that permit the design of efficient high-rate encoders. We also describe the close connections to the recent, independent results of Moon and Brickner [8] on maximum-transition-run (MTR) codes and Bliss [8], [9] on the "Tribit Mod 2" constraint. Finally, we address practical issues pertaining to encoder and decoder design, detector structure, and error probability performance. The Appendix provides a brief summary of relevant concepts and results from the theory of constrained systems and codes.

II. The ISI Channel

A. Basic Channel Model

We consider a discrete-time model for the magnetic recording channel with possibly constrained input $\boldsymbol{a} = \{a_n\} \in \mathcal{C} \subseteq \{-1, 1\}^\infty$, impulse response $\{h_n\}$, and output $\boldsymbol{y} = \{y_n\}$ given by

$$y_n = \sqrt{E} \sum_m a_m h_{n-m} + \eta_n \tag{4}$$

where

$$h(D) = \sum_n h_n D^n = (1 - D)(1 + D)^3 \qquad (E^2\text{PR4})$$

or

$$h(D) = \sum_n h_n D^n = (1 - D)(1 + D)^4 \qquad (E^3\text{PR4}),$$

η_n are independent Gaussian random variables with zero mean and variance σ^2, and E is a constant related to the output-voltage amplitude. The quantity E/σ^2 is referred to as the signal-to-noise ratio (SNR) per track. We refer to the symbol set $\{0, 1\}$ as the *binary* alphabet, and the symbol set $\{1, -1\}$ as the *bipolar* alphabet. For simplicity of notation, we shall sometimes use $-$ to denote -1, and $+$ to denote 1.

The optimal detector for the channel described above performs maximum-likelihood sequence estimation (MLSE) over $\mathcal{C}$, i.e., it determines an $\hat{\boldsymbol{a}} \in \mathcal{C}$ satisfying

$$\min_{\boldsymbol{a}\in\mathcal{C}} \Omega(\boldsymbol{a}) = \Omega(\hat{\boldsymbol{a}}) \tag{5}$$

where $\Omega(\boldsymbol{a})$ is the well-known log-likelihood function for channels with intersymbol interference [45, pp. 548–554]

$$\Omega(\boldsymbol{a}) = \sum_n y_n \sum_m a_m h_{n-m} - \frac{\sqrt{E}}{2} \sum_n \left(\sum_m a_m h_{n-m} \right)^2.$$

The search for an $\hat{\boldsymbol{a}} \in \mathcal{C}$ satisfying (5) is performed by means of a Viterbi detector whose complexity depends on the channel impulse response $\{h_n\}$ and the code constraint $\mathcal{C}$. The complexity is reflected in the detector trellis structure, whose characteristics we discuss below.

B. Graph Representation of ISI Channels

The input–output relationship for the PR channel with system polynomial $h_N(D) = (1-D)(1+D)^N$, $N \geq 1$, is given by a graph we will denote by G_N.

The states $V(\mathsf{G}_N)$ correspond to the length-$(N+1)$ channel memory determined by the previous $N+1$ input bits, $a_{i-1}, \cdots, a_{i-(N+1)}$. An edge e is defined by an initial state

$$\sigma(e) = a_{i-1}, \cdots, a_{i-(N+1)}$$

and a terminal state

$$\tau(e) = a_i, a_{i-1}, \cdots, a_{i-N}$$

that is determined by a next input bit a_i. The input/output edge labels, denoted a_i/y_i, represent the input a_i and the corresponding channel output symbol y_i, given by

$$y_i = \sum_{k=0}^{N+1} \left[\binom{N}{k} - \binom{N}{k-1} \right] a_{i-k}. \tag{6}$$

(We will on occasion use the same designation for a graph with input/output edge labels as for the same graph with output edge labels only. The meaning will be clear from the context.)

Let Y_N denote the constrained system consisting of all output sequences produced by the paths corresponding to walks along the graph G_N. The graph G_N with edges labeled by channel output symbols is the Shannon cover for Y_N, which is the deterministic presentation of Y_N with the minimal number of states. In fact, no graph representation of Y_N has fewer states. For this reason, Viterbi detection for these channels generally will use the trellis derived from G_N, a trellis that can be justifiably referred to as the minimum-complexity trellis for Y_N.

An important property of the system Y_N is that it is almost-finite-type (AFT), but not finite-type (FT), as defined in the Appendix. This is reflected in the Shannon cover by the existence of bi-infinite sequences that are generated by more than one distinct path. These sequences are often referred to as quasi-catastrophic sequences [25]. The effect of these sequences on Viterbi detector performance in the presence of finite path memory in the detector will be discussed in the next subsection. The following results characterize these sequences and the paths in G_N that generate them.

Lemma II.1: Let $\boldsymbol{a} = \{a_n\}$ and $\boldsymbol{a}' = \{a'_n\}$, $\boldsymbol{a}' \neq \boldsymbol{a}$, be input sequences over the bipolar alphabet with normalized error sequence $\boldsymbol{\epsilon} = \{\epsilon_n\}$, $\epsilon_n = (a_n - a'_n)/2$. Then $\boldsymbol{a}$ and $\boldsymbol{a}'$ produce the same output sequence $\boldsymbol{y}$; that is, $\boldsymbol{\epsilon} * \boldsymbol{h}_N = 0$, if and only if, for all $n \in \mathbb{Z}$

$$\epsilon_0 = \epsilon_{2n} \tag{7}$$

$$\epsilon_1 = \epsilon_{2n+1}. \tag{8}$$

Proof: The conditions in (7) and (8) are clearly sufficient. To prove necessity, note that the sequence $\boldsymbol{\epsilon}$ must be periodic [25]. Since the frequency response $H_N(f)$ of the ISI channel $h_N(D)$ satisfies $H_N(f) \neq 0$ for $f \neq \pi n$, $n \in \mathbb{Z}$, it follows that the transform $E(f)$ of $\boldsymbol{\epsilon}$ can be nonzero only for these frequencies. In other words,

$$E(f) = \sum_{\boldsymbol{n} \in \mathbb{Z}} E_0 \delta(f - \pi 2n) + E_1 \delta(f - \pi(2n+1)).$$

Thus the period of $\boldsymbol{\epsilon}$ must be a divisor of 2, so $\boldsymbol{\epsilon}$ has period 1 or 2. Conditions (7) and (8) follow. □

Remark: As pointed out by an anonymous referee, this result can also be proved by an induction on N. Generalizations of this result to other channel polynomials are considered in [3] and [4].

Corollary II.2: For all $N \geq 1$, the nonzero, bi-infinite error sequences annihilated by the channel filter $h_N(D)$ are

$$\boldsymbol{\epsilon} = \begin{cases} \pm \quad \cdots 111 \cdots \\ \pm \quad \cdots 1010 \cdots \\ \pm \quad \cdots 1 - 11 - 1 \cdots. \end{cases}$$

Proof: The result follows directly from Lemma II.1. □

The set of sequences identified in the corollary, which can be regarded as the kernel of the mapping on error input sequences induced by the channel, will be denoted $\mathcal{K}^{\{0,\pm 1\}}$. We can now identify the quasi-catastrophic sequences $\mathcal{Q}_N$ in the channel output Y_N, as well as the channel input sequences, which we denote $\mathcal{IQ}_N$, that generate them. This is accomplished by determining bipolar input pairs $\boldsymbol{a}$ and $\boldsymbol{a}'$ such that $\boldsymbol{a} = \boldsymbol{a}' + 2\boldsymbol{\epsilon}$ for some input error sequence $\boldsymbol{\epsilon} \in \mathcal{K}^{\{0,\pm 1\}}$.

Proposition II.3: The quasi-catastrophic sequences $\mathcal{Q}_N$ in G_N are generated by the set of bipolar input sequences $\mathcal{IQ}_N$ of the form

$$\boldsymbol{a} = \pm \cdots 1 z_0 1 z_1 1 z_2 1 \cdots$$

where the interleaved bipolar sequence $\boldsymbol{z} = \{z_n\}$ may be chosen arbitrarily.

Proof: The proof follows immediately from Corollary II.2 and the restriction that the sequences $\boldsymbol{a}$ and $\boldsymbol{a}'$ be bipolar. □

Remarks:

a) The set $\mathcal{IQ}_N$ is independent of N, for $N \geq 1$.

b) The all-0's output sequence $\boldsymbol{y} = \cdots 000 \cdots$ quasi-catastrophic in G_N, and is generated by the input sequences $\boldsymbol{a} = \pm \cdots 1 - 11 - 1 \cdots$ and the period-2 input sequences $\boldsymbol{a} = \pm \cdots 1111 \cdots$.

c) The Interleaved Non-Return to Zero (INRZI) precoded $(0, G/I)$ constraints [34] eliminate all of the sequences in $\mathcal{IQ}_N$, and therefore prevent the generation of quasi-catastrophic outputs from the channels $h_N(D)$.

C. Constrained-Input ISI Channel Model

We now extend the results on quasi-catastrophic sequences to input-restricted PR channels, where the channel inputs must belong to a constrained system $\mathcal{C}$. Let $\mathcal{C}$ be a constrained system over the bipolar alphabet, with Shannon cover $\mathsf{G}(\mathcal{C})$. (See the Appendix for the definition of the Shannon cover.) Let

$Y_N(\mathcal{C})$ be the constrained system of output sequences produced by the channel filter $h_N(D)$ when inputs belong to $\mathcal{C}$.

The Shannon cover of $Y_N(\mathcal{C})$, which we denote $\mathsf{G}_N(\mathcal{C})$, is obtained as follows. First, define a labeled graph H with states

$$V(\mathsf{H}) = \{(s; a_0, a_1, \cdots, a_N)\}$$

and edges

$$E(\mathsf{H}) = \{(s; a_0, a_1, \cdots, a_N) \rightarrow (t; a_1, \cdots, a_{N+1})\}$$

where s, t are states in $\mathsf{G}(\mathcal{C})$, the sequence $a_0, \cdots, a_N$ is generated by a path in $\mathsf{G}(\mathcal{C})$ that terminates in the state s, and t is a state in $\mathsf{G}(\mathcal{C})$, and there is an edge $s \rightarrow t$ with label a_{N+1} in $\mathsf{G}(\mathcal{C})$. The label $y(e)$ associated with edge e is the appropriate output from the channel $h_N(D)$ given by

$$y(e) = \sum_{k=0}^{N+1} \left[\binom{N}{k} - \binom{N}{k-1} \right] a_{N+1-k}.$$

The resulting graph is deterministic ($h_0 = 1$), irreducible, and has outdegree no more than 2 at any state. It is possible that it is not reduced, however; that is, there may be states with identical follower sets. If this is the case, one simply merges states having identical follower sets until a reduced, deterministic graph is obtained. This will be the Shannon cover $\mathsf{G}_N(\mathcal{C})$ of $Y_N(\mathcal{C})$.

Let $\mathcal{Q}_N(\mathcal{C})$ denote the quasi-catastrophic sequences in $\mathsf{G}_N(\mathcal{C})$, and let $\mathcal{IQ}_N(\mathcal{C})$ denote the constrained input sequences that generate them. The following proposition establishes a relationship between these sequences and their counterparts in the uncoded channel, when $\mathcal{C}$ is FT.

Proposition II.4: Let $\mathcal{C}$ be an FT-constrained system. Then

$$\mathcal{Q}_N(\mathcal{C}) \subset \mathcal{Q}_N \tag{9}$$

and

$$\mathcal{IQ}_N(\mathcal{C}) \subset \mathcal{IQ}. \tag{10}$$

Proof: If $\mathcal{C}$ is FT with memory M, and if $M - 1 \leq N$, then G_N contains a subgraph that is labeled-graph isomorphic to $\mathsf{G}_N(\mathcal{C})$; that is,

$$\mathsf{G}_N(\mathcal{C}) \subset \mathsf{G}_N. \tag{11}$$

This follows from the fact that each state of G_N is represented by a unique binary $N+1$ block, and each edge corresponds to a unique binary block of length $N+2$. Therefore, forbidding a block of length $l \leq M+1 < N+2$ will eliminate those states containing the forbidden block, along with their incoming and outgoing edges, whereas forbidding a block of length $l = N+2$ eliminates the corresponding edge from G_N. From this relationship of Shannon covers, we can immediately conclude that, for this case, the claimed inclusion relationships (9) and (10) hold.

In fact, these inclusion relationships hold even when $M - 1 > N$, although the inclusion (11) does not necessarily hold. To see this, note that each constrained input sequence is represented by a unique path in H and, consequently, in $\mathsf{G}_N(\mathcal{C})$. Therefore, any pair of paths in $\mathsf{G}_N(\mathcal{C})$ that generate the same output label sequence must correspond to distinct input sequences $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{C}$, whose difference $\boldsymbol{\epsilon} = (\boldsymbol{a} - \boldsymbol{b})/2$ satisfies $\boldsymbol{\epsilon} * \boldsymbol{h_N} = 0$. The result now follows from Proposition II.3. □

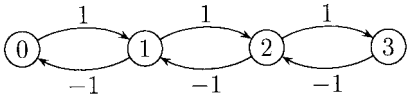


Fig. 1. Shannon cover for RDS-constrained system.

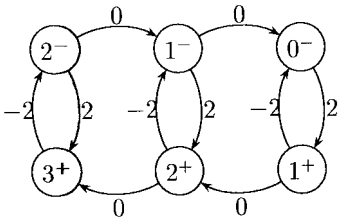


Fig. 2. Shannon cover $\mathsf{G}_0(\mathcal{C})$ for RDS-constrained dicode $(1 - D)$ system.

Remark: If $\mathcal{C}$ is AFT, but not FT, the quasi-catastrophic sequences $\mathcal{Q}_N(\mathcal{C})$ and their pre-images $\mathcal{IQ}(\mathcal{C})$ need not satisfy the conditions stated above. For example, let $\mathcal{C}$ be the charge-constrained system with running-digital sums (RDS) restricted to four values. The Shannon cover is shown in Fig. 1.

Consider the dicode channel filter $h_0(D) = 1 - D$, for which $\boldsymbol{y} = \{\boldsymbol{0}\}$, the all-0's sequence, is the only quasi-catastrophic output sequence. The Shannon cover $\mathsf{G}(\mathcal{C})$ of the output sequences produced by this channel with inputs constrained to $\mathcal{C}$ is shown in Fig. 2. The quasi-catastrophic sequences in $Y_0(\mathcal{C})$ are generated by the set of input sequences corresponding to paths in $\mathsf{G}(\mathcal{C})$ that visit only a proper subset of states. Since some of these quasi-catastrophic sequences are nonzero, neither of the inclusion relationships (9) and (10) hold.

D. Distance Measures and Error Probability

Given two channel output sequences, y_1 and y_2, the normalized squared Euclidean distance is given by

$$d^2(\boldsymbol{y_1}, \boldsymbol{y_2}) \triangleq \sum_i \left[(y_{1,i} - y_{2,i})/2 \right]^2$$

which is taken to be ∞ if the sum is unbounded. The quantity $d^2(\boldsymbol{y_1}, \boldsymbol{y_2})$ can be expressed in terms of the inputs $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$ that generate $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ via the relationship

$$\begin{aligned} d^2(\boldsymbol{y_1}, \boldsymbol{y_2}) &= d^2((\boldsymbol{y_1} - \boldsymbol{y_2}), \boldsymbol{0}) \\ &= d^2((\boldsymbol{a_1} - \boldsymbol{a_2}) * \boldsymbol{h}, \boldsymbol{0}). \end{aligned}$$

We can, therefore, characterize the error events in terms of the corresponding normalized input error sequences $\boldsymbol{\epsilon} = (\boldsymbol{a_1} - \boldsymbol{a_2})/2$.

The classification of error events involves two notions of distance. Both are relevant to the evaluation of the performance of a sequence detector which applies the Viterbi algorithm on a trellis derived from an output-labeled graph G that is deterministic and reduced, such as the Shannon cover graph.

Given a pair of right semi-infinite output sequences $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ that are generated by paths in G that diverge from a

common state s at time i, the normalized output error sequence $\boldsymbol{w} = (\boldsymbol{y_1} - \boldsymbol{y_2})/2$ is called a *closed* event if the paths remerge at some later time j, and remain identical beyond time j. On the other hand, the error sequence $\boldsymbol{w}(D)$ is called an *open* event if there exists a sequence of times $t_k, k \geq 1$ with $t_{k+1} > t_k$ such that the paths differ at times t_k for all k. Let W_{closed} be the set of closed events, W_{open} be the set of open events, and $W = W_{\text{closed}} \cup W_{\text{open}}$ be their union. Then we define the *minimum merged event squared distance*

$$d^2_{<>}(\mathsf{G}) = \min_{\boldsymbol{w}=\boldsymbol{y_1}-\boldsymbol{y_2} \in W_{\text{closed}}} d^2(\boldsymbol{y_1}, \boldsymbol{y_2})$$

and the *minimum event squared distance*

$$d^2_{<}(\mathsf{G}) = \min_{\boldsymbol{w}=\boldsymbol{y_1}-\boldsymbol{y_2} \in W} d^2(\boldsymbol{y_1}, \boldsymbol{y_2}).$$

It follows from these definitions that $d^2_{<}(\mathsf{G}) \leq d^2_{<}(\mathsf{G})$ [25]. Now, let $\boldsymbol{a_1} = \{a_{1,n}\}$ and $\boldsymbol{a_2} = \{a_{2,n}\}$ be two allowable sequences in a constrained system $\mathcal{C}$, and let

$$\boldsymbol{\epsilon} = \{\epsilon_n = (a_{1,n} - a_{2,n})/2\}$$

be their normalized error sequence. We define the set of all admissible nonzero error sequences

$$\mathcal{E}(\mathcal{C}) = \{\boldsymbol{\epsilon} \in \{-1, 0, 1\}^\infty | \boldsymbol{\epsilon} \neq \boldsymbol{0}, \boldsymbol{\epsilon} = (\boldsymbol{a_1} - \boldsymbol{a_2})/2,\ \boldsymbol{a}, \boldsymbol{b} \in \mathcal{C}\}. \tag{12}$$

Similarly, if $\mathcal{C}^*$ is a constrained system properly contained in $\mathcal{C}$, the set of all error sequences where at least one of $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$ belongs to $\mathcal{C}^*$ is given by

$$\mathcal{E}(\mathcal{C}, \mathcal{C}^*) = \epsilon \in \{-1, 0, 1\}^\infty | \boldsymbol{\epsilon} \neq \boldsymbol{0}, \boldsymbol{\epsilon} = (\boldsymbol{a_1} - \boldsymbol{a_2})/2, \\ \boldsymbol{a_1}, \boldsymbol{a_2} \in \mathcal{C},\ \text{with } \boldsymbol{a_1} \in \mathcal{C}^* \text{ or } \boldsymbol{a_2} \in \mathcal{C}^*\}. \tag{13}$$

This set of error sequences is relevant when the image of an encoder into $\mathcal{C}$ is a proper subset $\mathcal{C}^*$, but the detector structure is derived from $\mathsf{G}_N(\mathcal{C})$. If we restrict the sequences $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$ to differ in only a finite number of positions, we denote the corresponding sets of input error sequences by $\mathcal{E}_{\text{closed}}(\mathcal{C})$ and $\mathcal{E}_{\text{closed}}(\mathcal{C}, \mathcal{C}^*)$.

Let $d^2(\boldsymbol{\epsilon})$ be the normalized squared-Euclidean distance between the channel outputs generated by $\boldsymbol{a}$ and $\boldsymbol{b}$, given by

$$d^2(\boldsymbol{\epsilon}) = \sum_n \left(\sum_m \epsilon_m\, h_{n-m} \right)^2.$$

Note that, for the constrained channel represented by $\mathsf{G}_N(\mathcal{C})$

$$d^2_{<>}(\mathsf{G}_N(\mathcal{C})) = \min_{\boldsymbol{\epsilon} \in \mathcal{E}_{\text{closed}}(\mathcal{C})} d^2(\boldsymbol{\epsilon})$$

and

$$d^2_{<}(\mathsf{G}_N(\mathcal{C})) = \min_{\boldsymbol{\epsilon} \in \mathcal{E}(\mathcal{C})} d^2(\boldsymbol{\epsilon}).$$

At moderate-to-high signal-to-noise ratios, the parameter $d_{<>}(\mathsf{G}_N(\mathcal{C}))$ is the measure of minimum distance that determines, up to a constant factor, the average probability of an error event in the ML detector, P_{event}, when substituted for $d_{\min}$ in the well-known approximation

$$P_{\text{event}} \approx N_{\min} Q(d_{\min} \sqrt{\text{SNR}}).$$

Here $N_{\min}$ is the error coefficient and $Q(x)$ is the complementary error function. We often refer to the ratio of merged event distances $d^2_{<>}(\mathsf{G}_N(\mathcal{C}))/d^2_{<>}(\mathsf{G}_N)$ as the coding gain of code $\mathcal{C}$ over the uncoded channel $h_N(D)$.

In Section IV, we will impose finite-type constraints to eliminate all closed events $\boldsymbol{\epsilon}$ with $d^2(\boldsymbol{\epsilon}) < d^2_{\text{MFB}}$, implying that $d^2_{<>}(\mathsf{G}_N(\mathcal{C})) = d^2_{\text{MFB}}$, where d_{MFB} is defined by (1). We then construct finite-state encoders with finite-type image $\mathcal{C}^* \subseteq \mathcal{C}$, providing the desired coding gain. However, a consequence of the presence of the quasicatastrophic sequences in G_3 and G_4 is the existence of open error events $\boldsymbol{\epsilon}$ with $d^2(\boldsymbol{\epsilon}) < d^2_{\text{MFB}}$. One way to eliminate the open events would be to demand that the constraint $\mathcal{C}$ satisfy

$$\mathcal{C} \cap \mathcal{IQ} = \phi.$$

Alternatively, the constraint can satisfy the weaker condition that $d_{<}(\mathsf{G}_N(\mathcal{C})) = d_{\text{MFB}}$. Yet another approach is to construct the code $\mathcal{C}^*$ so that $d(\boldsymbol{\epsilon}) \geq d_{\text{MFB}}$, for all events $\boldsymbol{\epsilon} \in \mathcal{E}(\mathcal{C}, \mathcal{C}^*)$. This can be accomplished, for example, by imposing the even stronger condition

$$\mathcal{C}^* \cap \mathcal{IQ} = \phi.$$

These approaches are illustrated by examples in Section IV.

Another parameter of interest, particularly in the context of a recording system that uses an outer algebraic error-correcting code, is the maximum length of a burst error due to a minimum merged distance error event. More precisely, let $\boldsymbol{a}$ and $\boldsymbol{b}$ be channel input sequences. If there exist integers $i_1 \leq i_2$ such that $a_i \neq b_i$ for $i = i_1$ and $i = i_2$, while $a_i = b_i$ for $i < i_1$ and for $i > i_2$, we define their span, *span* $(\boldsymbol{a}, \boldsymbol{b}) = i_2 - i_1 + 1$. For an input-constrained channel presented by $\mathsf{G}{=}\mathsf{G}_N(\mathcal{C})$, the quantity of interest is obtained by maximizing *span* $(\boldsymbol{a}, \boldsymbol{b})$ over the set of all pairs $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{C}$ whose normalized difference $\boldsymbol{\epsilon}$ satisifies $d^2(\boldsymbol{\epsilon}) = d^2_{<>}(\mathsf{G})$. If a code $\mathcal{C}^* \subseteq \mathcal{C}$ is used, then the maximization can be restricted to the set of all pairs $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{C}$ where at least one of the sequences is constrained to belong to $\mathcal{C}^*$.

III. Distance Properties of ISI Channels

In this section, we analyze distance properties of two channels of practical interest, E^2PR4 and E^3PR4. In Section III-A, we use analytic methods to characterize closed error events $\boldsymbol{\epsilon}$ in G_3 and G_4 with $d^2(\boldsymbol{\epsilon}) < d^2_{\text{MFB}}$. The characterization is complete for G_3. For G_4, it is only partial; nevertheless, it suffices to define a class of distance-enhancing constraints in Section IV. In Section III-B, we develop and apply a finite-depth search algorithm on the error-state diagram to completely characterize open events with $d^2(\boldsymbol{\epsilon}) < d^2_{\text{MFB}}$ for G_3, as well as to characterize closed events with $d(\epsilon) = d_{\text{MFB}}$ for this channel.

A. Analytic Derivations

To examine distance properties of channels, it is often advantageous to work in the transform domain where each sequence $\{s_n\}$ has a corresponding function $s(D) = \Sigma_n\, s_n D^n$. We recall from Section I that the minimum merged squared

distance of the uncoded channel with transfer function $h_N(D)$ can be expressed as

$$d^2_{<>}(\mathsf{G}_N) = \min_{\epsilon(D)\neq 0} \|h_N(D)\epsilon(D)\|^2$$

where

$$\epsilon(D) = \sum_{i=0}^{l-1} \epsilon_i D^i, \qquad \epsilon_i \in \{-1,0,1\},\ \epsilon_0 = 1,\ \epsilon_{l-1} \neq 0$$

is the polynomial corresponding to a normalized error sequence $\boldsymbol{\epsilon} = \{\epsilon_i\}_{i=0}^{l-1}$ of length l.

It is known that $d^2_{<>}(\mathsf{G}_3) = 6$ for the E^{2}PR4 channel, and this is attained for the input error event $\epsilon(D) = 1 - D + D^2$. It is also known that $d^2_{<>}(\mathsf{G}_4) = 12$ for the E^{3}PR4 channel, and this is likewise attained for $\epsilon(D) = 1 - D + D^2$ [60]. Therefore, neither of these channels achieves the MFB. Recall that the MFB is determined by $\epsilon(D) = 1$, for which

$$\|(1-D)(1+D)^3 \cdot 1\|^2 = 10$$

and

$$\|(1-D)(1+D)^4 \cdot 1\|^2 = 28.$$

1) The E^{2}PR4 Channel: Since

$$\|(1-D)(1+D)^3 \cdot 1\|^2 = 10$$

we need to find all error polynomials $\epsilon(D)$ for which

$$\|(1-D)(1+D)^3\epsilon(D)\|^2 = 6,$$

as well as all error polynomials $\epsilon(D)$ for which

$$\|(1-D)(1+D)^3\epsilon(D)\|^2 = 8.$$

We consider the polynomial

$$\begin{aligned} y(D) &= (1-D)(1+D)^3\epsilon(D) \\ &= (1 + 2D - 2D^3 - D^4) \\ &\quad \cdot (1 + \epsilon_1 D + \epsilon_2 D^2 + \cdots + \epsilon_{l-3}D^{l-3} \\ &\quad + \epsilon_{l-2}D^{l-2} + \epsilon_{l-1}D^{l-1}). \end{aligned}$$

It is easy to check that for all input error sequences of length $l \leq 2$, $\|y(D)\|^2 \geq 10$. For error sequences of length $l \geq 3$, the polynomial $y(D)$ is of the form

$$\begin{aligned} y(D) = 1 &+ (\epsilon_1 + 2)D + (\epsilon_2 + 2\epsilon_1)D^2 + D^3 z(D) \\ &+ (-2\epsilon_{l-2} - \epsilon_{l-3})D^{l+1} + (-2\epsilon_{l-1} - \epsilon_{l-2})D^{l+2} \\ &+ (-\epsilon_{l-1})D^{l+3} \end{aligned} \tag{14}$$

where $z(D)$ is a polynomial with degree of at most $l-3$. Since $\epsilon_{l-1} \neq 0$, we have

$$\|y(D)\|^2 \geq 3 + \|z(D)\|^2 + 3. \tag{15}$$

We also know that $y(D)$ has a first-order zero at $D = 1$ and a third-order zero at $D = -1$.

Proposition III.1: The error polynomial $\epsilon(D) = 1 - D + D^2$ is the unique error polynomial for which

$$\|(1-D)(1+D)^3\epsilon(D)\|^2 = 6.$$

Proof: Because of (15), $\|y(D)\|^2 = 6$ only if $z(D) = 0$. Therefore,

$$\begin{aligned} y(D) = 1 &+ D - D^2 + (-2\epsilon_{l-2} - \epsilon_{l-3})D^{l+1} \\ &+ (-2\epsilon_{l-1} - \epsilon_{l-2})D^{l+2} + (-\epsilon_{l-1})D^{l+3}. \end{aligned}$$

For $y(1) = 0$, we need

$$y(D) = 1 + D - D^2 + D^{l+1} - D^{l+2} - D^{l+3}.$$

For $y(-1) = 0$, we need

$$y(D) = 1 + D - D^2 + D^{2k} - D^{2k+1} - D^{2k+2}.$$

For $y'(-1) = 0$, we need

$$y(D) = 1 + D - D^2 + D^4 - D^5 - D^6.$$

Note that

$$\begin{aligned} y(D) &= 1 + D - D^2 + D^4 - D^5 - D^6 \\ &= (1-D)(1+D)^3(1 - D + D^2) \end{aligned}$$

and, therefore, $\epsilon(D) = 1 - D + D^2$ is the only error polynomial for which $\|(1-D)(1+D)^3\epsilon(D)\|^2 = 6$. □

Proposition III.2: The error polynomials

$$\epsilon(D) = 1 - D + D^2 + D^5 - D^6 + D^7$$

and

$$\epsilon(D) = \sum_{i=0}^{l-1} (-1)^i D^i$$

for $l \geq 4$ are the only input error polynomials for which $\|(1-D)(1+D)^3\epsilon(D)\|^2 = 8$.

Proof: Because $y(1) = 0$ and $y(-1) = 0$, it follows from (14) that every polynomial $y(D)$ for which $\|y(D)\|^2 = 8$ may have at most one coefficient with absolute value equal to 2. Therefore, the corresponding error polynomial $\epsilon(D)$ either starts by $1 - D + D^2$ or ends by $\pm(D^{l-3} - D^{l-2} + D^{l-1})$ or both. This, together with the previous result, implies that any code that removes $\epsilon(D) = \pm(1 - D + D^2)$ from the set of all possible input error sequences provides a coding gain of $10\log_{10}(10/6)$ decibels.

A way to characterize all input error sequences for which $\|(1-D)(1+D)^3\epsilon(D)\|^2 = 8$, is the following. From (15), $\|y(D)\|^2 = 8$ only if $\|z(D)\|^2 = 2$, i.e.,

$$z(D) = \alpha D^m + \beta D^n, \qquad \alpha, \beta \in \{1, -1\}.$$

The sequence $\{y_i\}$ of the coefficients of polynomial $y(D)$ has a third-order null at the Nyquist frequency. It follows from the moment equations that

$$\sum_{i=0}^{l+3} (-1)^i i^k y_i = 0, \qquad k = 0, 1, 2. \tag{16}$$

We consider two cases.

Case 1: For $e_{l-1} = -1$, we have

$$y(D) = 1 + D - D^2 - D^m - D^n - D^{l+1} + D^{l+2} + D^{l+3}$$

since for $y(1) = 0$ we need $\alpha = \beta = -1$.

a) For l even, we need m even and n odd in order to satisfy (16) for $k = 0$. Because of that, (16) for $k = 1$ and $k = 2$ gives the system of equations

$$l - 3 - n + m = 0 \tag{17}$$

$$l^2 - 9 + n^2 - m^2 = 0 \tag{18}$$

whose only solution under the assumptions above is $n = 3$ and $m = l$. Therefore. $\epsilon(D) = \Sigma_{i=0}^{l-1} (-1)^i D^i$, for even l, $l \geq 4$.

b) For l odd, we need m odd and n odd order to satisfy (16) for $k = 0$. Because of that, (16) for $k = 1$ gives $l+3+m+n = 0$, which has no solution over positive integers.

Case 2: For $e_{l-1} = 1$, we have

$$y(D) = 1 + D - D^2 + D^m - D^n + D^{l+1} - D^{l+2} - D^{l+3}$$

since for $y(1) = 0$, we need $\alpha = -\beta = 1$.

a) For l even, we need m even and n odd in order to satisfy (16) for $k = 0$. Because of that, (16) for $k = 1$ and $k = 2$ gives the system of equations

$$-l - 3 + n + m = 0 \tag{19}$$

$$-l^2 - 1 + n^2 + m^2 = 0 \tag{20}$$

whose only solution under the assumptions above is $m = 4$, $n = 7$, $l = 8$. Therefore, $\epsilon(D) = 1 - D + D^2 + D^5 - D^6 + D^7$.

b) For l odd, we need m even and n even in order to satisfy (16) for $k = 0$. Because of that, (16) for $k = 1$ and $k = 2$ gives the system of equations

$$l - 3 - n + m = 0 \tag{21}$$

$$l^2 - 9 + n^2 - m^2 = 0 \tag{22}$$

whose only solution under the assumptions above is $n = 3$ and $m = l$. Therefore, $\epsilon(D) = \Sigma_{i=0}^{l-1} (-1)^i D^i$, for odd l, $l \geq 5$.

We conclude that the error polynomials for which

$$\|(1 - D)(1 + D)^3 \epsilon(D)\|^2 = 8$$

are

$$\epsilon(D) = 1 - D + D^2 + D^5 - D^6 + D^7$$

and

$$\epsilon(D) = \sum_{i=0}^{l-1} (-1)^i D^i$$

for $l \geq 4$. □

2) The E³PR4 Channel: Since $\|(1-D)(1+D)^4 \cdot 1\|^2 = 28$, we are interested in finding all error polynomials $\epsilon(D)$ for which

$$12 \leq \|(1 - D)(1 + D)^4 \epsilon(D)\|^2 < 28. \tag{23}$$

The following proposition gives a partial characterization of the closed input error events ϵ with $d^2(\epsilon) < 28$.

Proposition III.3: The closed input error sequences with the polynomial $1 + \epsilon_1 D + \cdots + \epsilon_{l-1} D^{l-1}$ satisfying inequalities (23) have length $l \geq 3$ and contain either the string $1\ -1\ 1$ or $-1\ 1\ -1$.

Proof: The proof relies upon the following observations.

Observation 1: It follows from

a) $\|(1 - D)(1 + D)^4 (1 - D)\|^2 = 28$ and

b) $\|(1-D)(1+D)^4(1+D+e_2D^2+\cdots+e_{l-1}D^{l-1})\|^2 = \|1 + 4D + (5 + \epsilon_2)D^2 + \cdots\|^2 > 33$

that an error sequence satisfying (23) must have length $l \geq 3$ and not start with string $1\ 1$. Because of the symmetry of $(1 - D)(1 + D)^4$, an error sequence satisfying (23) must not end with $1\ 1$ or $-1\ -1$.

Observation 2: It follows from

a) $y(D) = (1 - D)(1 + D)^4(1 + D^2 + e_3 D^3 + \cdots + e_{l-1} D^{l-1}) = 1 + 3D + 3D^2 + \cdots + (-3\epsilon_{l-1} - \epsilon_{l-2}) D^{l+3} + (-\epsilon_{l-1}) D^{l+4}$ and

b) $\Sigma_i\ y_i = 0$

that

$$\|(1-D)(1+D)^4(1+D^2+\epsilon_3 D^3+\cdots+\epsilon_{l-1}D^{l-1})\|^2 \geq 28.$$

Therefore, an error sequence satisfying (23) must not start with $1\ 0\ 1$ or end with either $1\ 0\ 1$ or $-1\ 0\ -1$.

Observation 3: It follows from

$$\|(1 - D)(1 + D)^4(1 - D^2 + \epsilon_3 D^3 + \cdots + \epsilon_{l-1} D^{l-1})\|^2 = \|1 + 3D + D^2 + (\epsilon_3 - 5)D^3 + \cdots\|^2 > 27$$

that an error sequence satisfying (23) must not start with the string $1\ 0\ -1$ or end with $1\ 0\ -1$ or $-1\ 0\ 1$.

Observation 4: It follows from

$$\begin{aligned} y(D) = 1 &+ (\epsilon_1 + 3)D + (\epsilon_2 + 3\epsilon_1 + 2)D^2 \\ &+ (\epsilon_3 + 3\epsilon_2 + 2\epsilon_1 - 2)D^3 + D^3 z(D) \\ &+ (2\epsilon_{l-1} - 2\epsilon_{l-2} - 3\epsilon_{l-3} - \epsilon_{l-4})D^{l+1} \\ &+ (-2\epsilon_{l-1} - 3\epsilon_{l-2} - \epsilon_{l-3})D^{l+2} \\ &+ (-3\epsilon_{l-1} - \epsilon_{l-2})D^{l+3} + (-\epsilon_{l-1})D^{l+4} \end{aligned}$$

that for an error sequence of length $l \geq 4$ starting with $1\ 0\ 0$ or $1\ -1\ 0$ and ending with $0\ 0\ 1$ or $0\ 1\ -1$ or $0\ 1\ -1$, we have $\|y(D)\|^2 \geq 30$.

It follows from these observations that an error sequence satisfying (23) must contain either the string $1\ -1\ 1$ or $-1\ 1\ -1$. □

B. Search Algorithm

A convenient tool in studying distance properties of ISI channels is the error-state diagram [61]. For the family of binary-input PR channels $h_N(D)$, $N \geq 1$, the error-state diagram, denoted E_N, has 3^{N+1} states $(e_0, \cdots, e_N)$, $e_i \in \{0, \pm 1\}$. An edge e has initial state $\sigma(e) = (e_0, \cdots, e_N)$, terminal state $\tau(e) = (e_1, \cdots, e_{N+1})$, and edge label

$$\mu_i = \sum_{k=0}^{N+1} \left[\binom{N}{k} - \binom{N}{k-1} \right] e_{i-k}.$$

The "fundamental" closed error events of length l and distance d correspond to paths $e_0, \cdots, e_l$ in E_N that start and end at state $\sigma = (0, \cdots, 0)$, with no intermediate visit to that state, and edge label sequence satisfying

$$\sum_{i=0}^{l} \mu_i^2 = d^2.$$

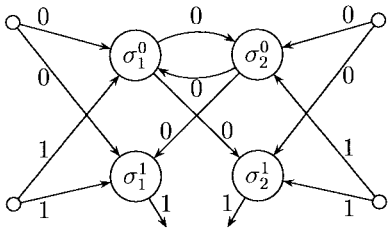


Fig. 3. Local picture of out-splitting of E_N.

Similarly, the open error events of length l and distance d correspond to those paths in E_N that start at state $\sigma = (0,\cdots,0)$ and end at one of the states $\tau = (\tau_0,\cdots,\tau_N)$ satisfying

$$\tau_{2i} = \tau_0, \qquad 0 \le i \le \left\lceil \frac{N}{2} \right\rceil$$

$$\tau_{2i+1} = \tau_1, \qquad 0 \le i \le \left\lceil \frac{N-1}{2} \right\rceil$$

with no intermediate visits to state σ, and that accumulate distance d. The terminal states τ lie along the *zero-cycles*, paths that generate error sequences $\mathcal{K}^{\{0,\pm 1\}}$ producing the all-0's output from the channel filter. Recall from Corollary II.2 that the paths generating the all-0's sequence in E_N are the period-1 cycles at states $\sigma = (0,\cdots,0)$, $\sigma = (1,\cdots,1)$, and $\sigma = (-1,\cdots,-1)$, as well as the period-2 cycles at states $\sigma = \pm(e_0,-e_0,\cdots,(-1)^N e_0)$, where $e_0 \in \{\pm 1\}$ or $\sigma = [\pm(1/2,-1/2,\cdots,(-1)^N/2)] - (1/2,1/2,\cdots,1/2)$. Note that all of the zero cycles are mutually disjoint in E_N. The characterization of error events of specified distance d is complicated by the existence of these zero cycles. In principle, an infinite-depth tree search on E_N might be necessary for certain values of d. However, the properties of the zero cycles suggest a simple modification to E_N, producing a graph E_N^*, that will allow complete characterization of the error events for specified distance d in a search of bounded depth on E_N^*.

The graph E_N^* effectively removes the zero cycles. This modified error-state diagram is obtained in three steps:

Step 1: Eliminate the self-loops (period-1 cycles) at states $\sigma = (0,\cdots,0)$, $\sigma = (1,\cdots,1)$, and $\sigma = (-1,\cdots,-1)$.

Step 2: Out-split states $\sigma_1 = (e_0, e_1, \cdots, e_N)$ and $\sigma_2 = (e_1,\cdots,e_N,e_0)$ comprising a period-2 zero cycle. Partition the outgoing edges E_1 from σ_1 as $E_1 = E_1^0 \cup E_1^1$, where E_1^0 contains the edge with label 0, and $E_1^1 = E_1 - E_1^0$. The descendant states are denoted σ_1^0 and σ_1^1. Then partition the edges E_2 from state σ_2 in an analogous fashion. The resulting local picture is shown in Fig. 3. Denote the new graph by E'_N.

Step 3: Remove from E'_N the edges constituting the period-2 zero cycle involving the offspring states σ_1^0 and σ_2^0.

The resulting graph E_N^* has exactly $3^{N+1}+6$ states. It no longer contains any zero cycles, so there exists a positive integer l_N that provides an upper bound on the length of any run of 0's in the generated sequences. Consequently, any path in E_N^*, starting and ending at state $\sigma = (0,0,\cdots,0)$, that generates a closed error event with distance d in G_N must have length no more than $d^2(l_N+1)$. Similarly, any path in E_N^*, starting at state $\sigma = (0,0,\cdots,0)$ and ending in a state σ_1^0 or σ_2^0, that generates an open error event with distance d in G_N, must satisfy the same bound on length. By keeping track of the state sequences in E_N^* corresponding to channel output sequences, one can use a bounded-depth search based on E_N^* to determine and characterize all closed and open error events of specified distance d in G_N.

Two variants of this algorithm are described in [4]. Tables of closed and open error events through a range of small distances for PR4, EPR4, E^{2}PR4, and E^{3}PR4, along with the corresponding tables for partial-response channels with $h(D) = (1+D)^N$, for $1 \le N \le 3$, may be found in [2] and [4]. Here, we present only a subset of the closed and open error events for the E^{2}PR4 and E^{3}PR4 channels.

Table I lists the input error sequences for closed events on the E^{2}PR4 channel having squared distance $d^2(\boldsymbol{\epsilon}) \le d^2_{\mathrm{MFB}} = 10$.

Table II lists the input error sequences for open events on the E^{2}PR4 channel having squared distance $d^2(\boldsymbol{\epsilon}) < d^2_{\mathrm{MFB}} = 10$.

Table III lists the input error sequences for closed events on the E^{3}PR4 channel having squared-distance $d^2(\boldsymbol{\epsilon}) < d^2_{\mathrm{MFB}} = 28$.

TABLE I
CLOSED ERROR EVENTS FOR E^{2}PR4 CHANNEL, $d^2 \le 10$

d^2	$\boldsymbol{\epsilon}$
6	+−+0000
8	+−+00+−+0000
	+−+−(+−)0000
	+−+−(+−)+0000
10	+−+0−+−0000
	+−+00+−+00+−+0000
	+−+00+00+−+0000
	+00+−+0000
	+−+00+−+−(+−)+0000
	+−+00+−+−(+−)0000
	+−+000+−+0000
	+−+00+0000
	+0000
	+−+−(+−)+00+−+0000
	+−+−(+−)00−+−0000

TABLE II
OPEN ERROR EVENTS FOR THE E^{2}PR4 CHANNEL, $d^2 < 10$

d^2	$\boldsymbol{\epsilon}$
4	+−+−[+−]$^\infty$
6	+−+00+−+−[+−]$^\infty$
	+−+00+0+[0+]$^\infty$
	+0+[0+]$^\infty$
8	+−+−(+−)+00+−+−[+−]$^\infty$
	+−+0−+−+[−+]$^\infty$
	+−+−(+−)00−0−[0−]$^\infty$
	+−+00+−+00+−+−[+−]$^\infty$
	+00+−+−[+−]$^\infty$
	+−+00+00+0+[0+]$^\infty$
	+−+0+0[+0]$^\infty$
	+−+−(+−)+00+0+[0+]$^\infty$
	+−+−(+−)00−+−+[−+]$^\infty$
	+−+000+−+−[+−]$^\infty$
	+−+00+00+−+−[+−]$^\infty$
	+−+00+−+00+0+[0+]$^\infty$
	+00+0+[0+]$^\infty$

TABLE III
CLOSED ERROR EVENTS FOR E^3PR4 CHANNEL, $d^2 < 28$

d^2	ϵ
12	$+-+00000$
16	$+-+00+-+00000$ $+-+000+-+00000$
20	$+-+-+(-+)00000$ $+-+-+-(+-)00000$ $+-+0-+-00000$ $+-+00+00+-+00000$ $+-+00+-+00+-+00000$ $+-+000+-+00+-+00000$ $+-+000+-+000+-+00000$
22	$+-+-00000$ $+-+0000+-+00000$ $+-+00+00+00+-+00000$
24	$+00+-+00000$ $+-+00+-+-+(-+)00000$ $+-+00+-+-+-(+-)00000$ $+-+000+-+-+(-+)00000$ $+-+000+-+-+-(+-)00000$ $+-+00+-+0-+-00000$ $+-+000+-+0-+-00000$ $+-+00+00+-+00+-+00000$ $+-+00+-+00+-+00+-+00000$ $+-+000+-+00+00+-+00000$ $+-+000+-+00+-+00+-+00000$ $+-+00+-+000+-+00+-+00000$ $+-+000+-+000+-+00+-+00000$ $+-+000+-+00+-+000+-+00000$ $+-+00+00+00+00+-+00000$ $+-+000+-+000+-+000+-+00000$
26	$+-0+-+00000$ $+-+0-+-+-(+-)00000$ $+-+0-+-+-+(-+)00000$ $+-+00+-+-00000$ $+-+0-+0-+-00000$ $+-+000+-+-00000$ $+00+00+-+00000$ $+-+0000-+-00000$ $+-+000+00+-+00000$ $+-+0000+-+00+-+00000$ $+-+0000+-+000+-+00000$ $+-+00+00+00+-+00+-+00000$ $+-+000+-+00+00+00+-+00000$ $+-+00+00+00+00+00+-+00000$

Table IV lists the input error sequences for open events on the E^3PR4 channel having squared distance $d^2(\boldsymbol{\epsilon}) < 18$.

In the error sequence tables, the symbol "$+$" is used to designate "1," "$-$" is used to designate "-1." A parenthesized string (s) denotes any nonnegative integer number of repetitions of the string s, while $[s]^\infty$ denotes the infinite repetition of string s.

IV. CONSTRAINED CODE DESIGN

In this section, we demonstrate how distance properties analysis and error-event characterization may be used to determine constrained systems and construct codes that improve distance properties of certain ISI channels (in particular, E^2PR4 and E^3PR4) up to their matched filter bounds. The analysis is useful in clarifying the known distance-enhancing properties of certain runlength-limited constraints and matched-spectral-null constraints. We examine a variety of distance-enhancing constraints and codes for the E^2PR4 and E^3PR4 channels. For the purposes of comparison and completeness, we also include discussions of previously known codes, as well as related, independently proposed performance-improving codes that fit into the framework that is established in this paper. We first review some basic conventions used in describing modulation codes. The description of a binary recording code is affected by the type of modulation used on the channel. There are two important modulation methods commonly used on magnetic recording channels: *non-return-to-zero* (NRZ) and *non-return-to-zero-inverse* (NRZI), as we now briefly review. Saturation recording of binary information on a magnetic medium is accomplished by converting an input stream of binary data into a spatial stream of bit cells along a track where each cell is fully magnetized in one of two possible directions. In NRZ modulation, the directions of cell magnetizations correspond directly to digits in the input data stream. In previous sections we have represented the digits using the bipolar alphabet $\{-1, 1\}$. In this section, for notational convenience, we will use the binary alphabet $\{0, 1\}$ to represent the magnetization directions, with "0" and "1" corresponding to "-1" and "$+1$," respectively. Also, as in the tables of the previous section, we will at times designate the bipolar error symbol "1" by "$+$" and the bipolar error symbol "-1" by "$-$," for notational convenience. In NRZI modulation, a binary digit 1 denotes the presence of a transition from one direction of magnetization to another, while a binary digit 0 denotes the absence of such a transition. This modulation is implemented by translating (or "precoding") a binary data sequence $\boldsymbol{a} = a_0, a_1, \cdots$ into an intermediate binary sequence $\boldsymbol{x} = x_0, x_1, \cdots$ which is then mapped to the magnetization pattern according to the NRZ convention. The sequence $\boldsymbol{x}$ is determined by the rule

$$x_i = x_{i-1} \oplus a_i$$

where $\oplus$ represents addition modulo 2, with a specified initial condition for x_{-1}.

The constrained systems that we consider will often have a convenient definition in terms of a finite list of forbidden strings $\mathcal{F}$ over the binary alphabet. The corresponding constraint will be denoted $X_{\mathcal{F}}^{\{0,1\}}$. We will also consider "time-

TABLE IV
OPEN ERROR EVENTS FOR THE E^3PR4 CHANNEL, $d^2 \leq 18$

d^2	ϵ
10	$+-+-+-+[-+]^\infty$
14	$+-+000+-+-+-+[+-]^\infty$ $+-+00+-+-+[-+]^\infty$ $+-+00+0+0[+0]^\infty$
16	$+-+0-+-+-[+-]^\infty$ $+-+00+00+0+0[+0]^\infty$
18	$+-+000+-+000+-+-+[-+]^\infty$ $+-+00+-+000+-+-+[-+]^\infty$ $+-+00+00+-+-+[-+]^\infty$ $+-+00+-+00+0+0[+0]^\infty$ $+-+000+-+00+-+-+[-+]^\infty$ $+-+00+-+00+-+-+[-+]^\infty$ $+-+000+-+00+0+0[+0]^\infty$ $+-+00+00+00+0+0[+0]^\infty$

varying" constraints defined by a list $\mathcal{F}^e$ of strings forbidden to begin at even time indices and a list $\mathcal{F}^o$ of strings forbidden to begin at odd time indices. These "forbidden-list" constraints will often have an interpretation in terms of restrictions on runlengths, such as: a) the minimum or maximum runlength of 0's, denoted d_0 and k_0, respectively, between two consecutive 1's; b) the minimum or maximum runlength of 1's, denoted d_1 and k_1, respectively, between two consecutive 0's; or c) the related runlength constraints imposed on strings beginning at even or odd time indices, d_b^e, k_b^e and $d_b^o, k_b^o, b \in \{0,1\}$.

For certain constraints, there will be a preferred description in terms of either NRZ or NRZI modulation. When clarification is needed, we will indicate the intended modulation by a subscript on the forbidden list. We illustrate this with an example.

Example IV.1: One family of constrained systems of finite type commonly used for coding in magnetic recording systems are (d,k)-constrained systems with NRZI modulation. These systems obey runlength-limited (RLL) constraints, where the runlength of 0's between consecutive 1's is constrained to be at least d and at most k. The constrained system defined by the forbidden list $\mathcal{F} = \{11\}_{\mathrm{NRZI}}$ is equivalent to the $(d,k) = (1,\infty)$ constraint. In other words, changes in the direction of magnetization must be separated by at least two bit cells. The constrained system that imposes the same constraint on magnetization patterns when NRZ modulation is used directly is determined by the forbidden list $\mathcal{F} = \{101,010\}_{\mathrm{NRZ}}$.

In this section, we will be interested in constraints $\mathcal{C}$ whose error sequences $\mathcal{E}(\mathcal{C})$ do not contain any substrings in a finite list $\mathcal{L}$ of forbidden input error strings. That is, denoting by $X_{\mathcal{L}}^{\{\pm,0\}}$ the system of error sequences over the alphabet $\{\pm 1, 0\}$ determined by the forbidden list $\mathcal{L}$, we consider input constraints $\mathcal{C}$ such that $\mathcal{E}(\mathcal{C}) \subseteq X_{\mathcal{L}}^{\{\pm,0\}}$.

More specifically, $\mathcal{L}$ will be chosen so that $X_{\mathcal{L}}^{\{\pm,0\}}$ contains no input error sequences corresponding to closed error events $\boldsymbol{\epsilon}$ with $d(\boldsymbol{\epsilon}) < d_{\mathrm{MFB}}$.

We will also discuss issues pertaining to the construction of efficient codes $\mathcal{C}^*$ such that $\mathcal{E}(\mathcal{C},\mathcal{C}^*)$ further avoids input error sequences corresponding to open error events $\boldsymbol{\epsilon}$ with $d^2(\boldsymbol{\epsilon}) < d^2_{\mathrm{MFB}}$. This will ensure that the Viterbi detector based upon G_N can provide near-maximum-likelihood performance with finite path memory. Finally, for a selection of codes, we compare their simulated performance on the E^2PR4 and E^3PR4 channels.

A. Finite-Type Distance-Enhancing Constraints

Given the condition $\mathcal{E}(\mathcal{C}) \subseteq X_{\mathcal{L}}^{\{\pm,0\}}$, we will identify finite collections $\mathcal{F}$ of blocks over the alphabet $\{0,1\}$ so that

$$\mathcal{C} \subseteq X_{\mathcal{F}}^{\{0,1\}} \Rightarrow \mathcal{E}(\mathcal{C}) \subseteq X_{\mathcal{L}}^{\{\pm,0\}}. \tag{24}$$

In the case of data sequences $\boldsymbol{a}$ and $\boldsymbol{b}$ represented over the binary alphabet, the corresponding error sequence is defined by $\boldsymbol{\epsilon} = \boldsymbol{a} - \boldsymbol{b}$. It follows that, whenever $\epsilon_i \neq 0, \epsilon_i = 2a_i - 1$. Therefore, the collection $\mathcal{F}$ of blocks forbidden in code sequences can be constructed based on the collection $\mathcal{L}$ of blocks forbidden in error sequences. Specifically, for each block $\boldsymbol{w}_{\mathcal{E}} \in \mathcal{L}$ of length l, we construct a list $\mathcal{F}_{\boldsymbol{w}_{\mathcal{E}}}$ of blocks of the same length l according to the rule

$$\mathcal{F}_{\boldsymbol{w}_{\mathcal{E}}} = \{\boldsymbol{w}_{\mathcal{C}} \in \{0,1\}^l | w_{\mathcal{C}}^i = (w_{\mathcal{E}}^i + 1)/2 \text{ for all } i \text{ for which } w_{\mathcal{E}}^i \neq 0\}.$$

Then the collection $\mathcal{F}$ obtained as

$$\mathcal{F} = \bigcup_{\boldsymbol{w}_{\mathcal{E}} \in \mathcal{L}} \mathcal{F}_{\boldsymbol{w}_{\mathcal{E}}}$$

satisfies condition (24). Furthermore, if $\mathcal{F}$ contains blocks $\boldsymbol{w}_{\mathcal{C}}$ and $\overline{\boldsymbol{w}_{\mathcal{C}}}$, which are binary complements of each other, then one of these blocks can be removed from $\mathcal{F}$. The constraint then becomes less restrictive and allows construction of higher rate codes, although certain issues pertaining to code phase invariance may then need to be addressed.

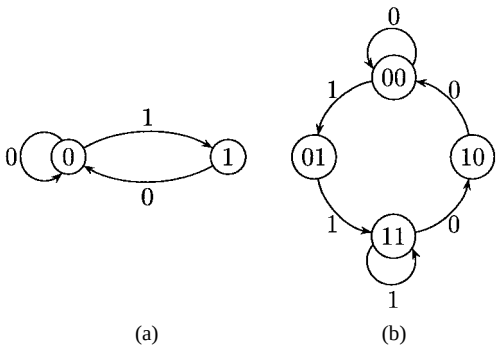


Fig. 4. Two equivalent constraints: (a) $\mathcal{F} = \{11\}$ NRZI and (b) $\mathcal{F} = \{101, 010\}$ NRZ.

Remark: The problem of determining the maximum capacity achievable by a constrained system $\mathcal{C}$ satisfying

$$\mathcal{E}(\mathcal{C}) \subseteq X_{\mathcal{L}}^{\{\pm,0\}}$$

has not been completely solved. A general approach, as well as definitive results for certain classes of forbidden lists $\mathcal{L}$, are presented in [38].

1) $\mathcal{L} = \{\pm[+-+]\}$: A $d_0 = 1$ NRZI Code: The results in Section III on distance properties of the E^2PR4 and E^3PR4 channels imply that constraints $\mathcal{C}$ for which

$$\mathcal{E}(\mathcal{C}) \subseteq X_{\mathcal{L}}^{\{\pm,0\}}$$

where $\mathcal{L} = \{\pm[+-+]\}$ provide the gain in distance of $\|h_N(D)\|/d_{\min}$ determined by the MFB. Moreover, for the E^2PR4 channel, the constraint eliminates all closed error events with $d^2(\boldsymbol{\epsilon}) = 10$ except for the events $\epsilon = [+]^{\infty}$. Examination of Table II in the previous section shows that the only open input error events not containing the strings $+-+$ or $-+-$ are those terminating in $\pm[+0]^{\infty}$. Therefore, if

$$\mathcal{E}(\mathcal{C}) \subseteq X_{\mathcal{L}}^{\{\pm,0\}}$$

where $\mathcal{L} = \{\pm[+-+], \pm[+0+]\}$, the constraint will also eliminate all open events with $d^2(\boldsymbol{\epsilon}) < 10$ on E^2PR4.

The constrained system with $d_0 = 1$ and NRZI modulation is the well-known $(d,k) = (1,\infty)$ constraint and is represented by the graph shown in Fig. 4(a). The corresponding NRZ constraint is an $X_{\{101,010\}}^{\{0,1\}}$ system whose Shannon cover is shown in Fig. 4(b).

We first show that this constraint eliminates all closed input error sequences with distance below the MFB on both E^2PR4 and E^3PR4.

Lemma IV.2: Any constraint $\mathcal{C} \subseteq X^{\{0,1\}}_{\{101,010\}}$ satisfies

$$\mathcal{E}(\mathcal{C}) \subseteq X^{\{\pm,0\}}_{\{+-+,-+-\}}$$

and, therefore, eliminates all closed input error sequences $\boldsymbol{\epsilon}$ with $d^2(\boldsymbol{\epsilon}) < 10$ on E^2PR4 and $d^2(\boldsymbol{\epsilon}) < 28$ on E^3PR4.

Proof: From the definition of error sequences we see that $\epsilon_s\epsilon_{s+1}\epsilon_{s+2} = +-+$ implies $a_sa_{s+1}a_{s+2} = 101$ and $b_sb_{s+1}b_{s+2} = 010$, and $\epsilon_s\epsilon_{s+1}\epsilon_{s+2} = -+-$ implies $a_sa_{s+1}a_{s+2} = 010$ and $b_sb_{s+1}b_{s+2} = 101$. Therefore, any constraint $\mathcal{C}$ such that $\mathcal{C} \subseteq X^{\{0,1\}}_{\{101,010\}}$ guarantees that

$$\mathcal{E}(\mathcal{C}) \subseteq X^{\{-,0,+\}}_{\{+-+,-+-\}}. \qquad \square$$

The following corollary is immediate.

Corollary IV.3: Any code satisfying the constraint of the previous lemma provides a gain in distance of $10\log_{10}(10/6)$ decibels on the E^2PR4 and $10\log_{10}(28/12)$ decibels on the E^3PR4 channel. Moreover, in both cases, the maximum span of an input error sequence corresponding to a minimum distance event is 1 bit.

The constraint also eliminates all open error events $\boldsymbol{\epsilon}$ with $d^2(\boldsymbol{\epsilon}) < 10$ on the E^2PR4 channel, as we now show.

Lemma IV.4: Any code $\mathcal{C}$ which is a subset of the constrained system $X^{\{0,1\}}_{\{101,010\}}$ satisfies the following:

$$\mathcal{E}(\mathcal{C}) \subseteq X^{\{\pm,0\}}_{\{+-+,-+-,+0+,-0-\}}.$$

Proof: From the definition of error sequences we see that $\epsilon_s\epsilon_{s+1}\epsilon_{s+2} = +0+$ implies $a_sa_{s+1}a_{s+2} = 1a1$ and $b_sb_{s+1}b_{s+2} = 0a0$. Similarly, $\epsilon_s\epsilon_{s+1}\epsilon_{s+2} = -0-$ implies $a_sa_{s+1}a_{s+2} = 0a0$ and $b_sb_{s+1}b_{s+2} = 1a1$. In either case, regardless of the choice of a, one of each pair of strings will be of the form 101 or 010. The claim follows. $\square$

Corollary IV.5: Any code $\mathcal{C}$ which is a subset of the constrained system $X^{\{0,1\}}_{\{101,010\}}$ eliminates all open and closed error events with $d^2 < 10$ on the E^2PR4 channel. Hence, $d^2_< = 10$.

Proof: Examination of Table II in the previous section shows that the only open input error events not containing the strings $+-+$ or $-+-$ are those terminating in $\pm[+0]^\infty$. The preceding lemma shows that the $X^{\{0,1\}}_{\{101,010\}}$ constraint eliminates the input error strings $+0+$ and $-0-$, and, therefore, the remaining open events. $\square$

The constrained systems defined by forbidden lists $\mathcal{F} = \{11\}_{\mathrm{NRZI}}$ and $\mathcal{F} = \{101,010\}_{\mathrm{NRZ}}$ allow construction of codes with finite-state encoders and sliding-block decoders with rates of up to about 0.6942. It was first observed by Behrens and Armstrong [6] that the industry standard rate $2/3$, $(1,7)$ NRZI code provided a coding gain for the E^2PR4 channel. The interpretation of the behavior of this code in terms of channel error event characterization, given above, was presented in Karabed and Siegel in [26] and Soljanin in [51].

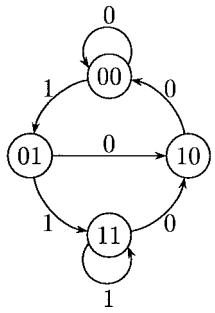


Fig. 5. Graph representation of $\mathcal{F} = \{101\}$ NRZ constrained system.

The graph $\mathsf{G}_3(X^{\{0,1\}}_{\{101,010\}})$ was shown in [6] to require ten states, of which four had only a single incoming edge. It can also be shown that the truncation depth for $d^2 = 10$ of this constraint is no more than 10, indicating that a path memory of length 10 can be used to achieve the expected coding gains.

The $\mathcal{C} = X^{\{0,1\}}_{\{101,010\}}$ input-constrained channel still generates quasi-catastrophic output sequences, specifically, the all-0's sequence, although this sequence does not play a role in any of the error events with $d^2 < 10$. However, through the introduction of the $k = 7$ runlength limitation, the rate $2/3\,(1,7)$ code $\mathcal{C}^*$ eliminates all remaining open error events of any distance by imposing the condition

$$\mathcal{E}(\mathcal{C},\mathcal{C}^*) \cap \mathcal{K}_N^{\{\pm 1,0\}} = \phi.$$

2) $\mathcal{L} = \{\pm[+-+]\}$: An $\mathcal{F} = \{101\}$ NRZ Code: The constrained system with $\mathcal{F} = \{101\}$ and NRZ modulation is represented by the graph shown in Fig. 5.

Lemma IV.6: A constraint $\mathcal{C} \subseteq X^{\{0,1\}}_{\{101\}}$ satisfies the following:

$$\mathcal{E}(\mathcal{C}) \subseteq X^{\{\pm,0\}}_{\{+-+,-+-,+0+0,-0-0\}}.$$

It, therefore, eliminates all closed error events $\boldsymbol{\epsilon}$ with $d^2(\boldsymbol{\epsilon}) < d^2_{\mathrm{MFB}}$ on the E^2PR4 and E^3PR4 channels. On the E^2PR4 channel, it also removes all but one of the closed events with $d^2(\boldsymbol{\epsilon}) = 10$, as well as all open error events with $d^2(\boldsymbol{\epsilon}) < 10$.

Proof: From the discussion above, we see that $\epsilon_s\epsilon_{s+1}\epsilon_{s+2} = +-+$ and $\epsilon_s\epsilon_{s+1}\epsilon_{s+2} = -+-$ implies that one of the sequences $\boldsymbol{a}$ or $\boldsymbol{b}$ must contain the string 101. Therefore, a code $\mathcal{C}$ such that $\mathcal{C} \subseteq X^{\{0,1\}}_{\{101\}}$ or $\mathcal{C} \subseteq X^{\{0,1\}}_{\{010\}}$ guarantees that

$$\mathcal{E}(\mathcal{C}) \subseteq X^{\{\pm,0\}}_{\{+-+,-+-\}}.$$

Although it is no longer true that the error strings $+0+$ and $-0-$ are eliminated, one can easily see that the strings $+0+0+$ and $-0-0-$ are avoided. In the former case, $\epsilon_s\epsilon_{s+1}\epsilon_{s+2}\epsilon_{s+3}\epsilon_{s+4} = 10101$ implies that

$$a_sa_{s+1}a_{s+2}a_{s+3}a_{s+4} = 1c_11c_21$$

and

$$b_sb_{s+1}b_{s+2}b_{s+3}b_{s+4} = 0c_10c_20.$$

If either $c_1 = 0$ or $c_2 = 0$, then $\boldsymbol{a}$ contains 101. If not, then $c_1 = c_2 = 1$, in which case $\boldsymbol{b}$ contains the forbidden string. This means that the open events terminating in $\pm[+0]^\infty$ are forbidden. □

We conclude that this constraint has the same distance-enhancing properties as the $d_0 = 1$ NRZI constraint.

Corollary IV.7: The $X^{\{0,1\}}_{\{101\}}$ constraint provides a coding gain of $10\log_{10}(10/6)$ decibels on the E^{2}PR4 channel and $10\log_{10}(28/12)$ decibels on the E^{3}PR4 channel. Moreover, $d^2_<(\mathsf{G}_3(\mathcal{C})) = 10$ and the maximum span of an input error sequence corresponding to a minimum distance event is 1 bit.

Remark: The graph $\mathsf{G}_3(X^{\{0,1\}}_{\{101\}})$ was shown in [26] to require 12 states, of which three had only a single incoming edge. It can also be shown that the truncation depth for $d^2 = 10$ of this constraint is exactly 10, indicating that a path memory of length 10 can be used to achieve the expected coding gains.

The constrained system $X^{\{0,1\}}_{\{101\}}$ has capacity approximately 0.8113. A rate $4/5$ $X^{\{0,1\}}_{\{101\}}$ code was described by Karabed and Siegel in [26]. As with the $X^{\{0,1\}}_{\{101,010\}}$ constraint, the $X^{\{0,1\}}_{\{101\}}$ input-constrained PR channels generate only the all-zeros quasi-catastrophic output sequence. The rate $4/5$ code $\mathcal{C}^*$ constructed in [26] imposes additional runlength constraints k_1 and k_0, thereby preventing the generation of any open error events with squared distance smaller than 10. The rate $4/5$ code construction was based upon application of the ACH state-splitting algorithm to a nondeterministic presentation of the constraint, which facilitated the incorporation of the run-length constraints. The finite-state encoder has seven encoder states. If the additional runlength constraints are not required, a rate $4/5$ code with only 4-states may be constructed from the Shannon cover.

3) $\mathcal{L} = \{\pm[+-+-], \pm[+-+00], \pm[+0+0+0]\}$: An $\mathcal{F} = \{0101, 11101\}$ NRZ Code: As mentioned in [26], one can find FT constraints with larger capacity than those discussed in the preceding sections that achieve the MFB. For example, the constraint with $\mathcal{F} = \{0101, 11101\}$, with capacity approximately 0.8590, eliminates all open and closed error events with $d^2(\boldsymbol{\epsilon}) < 10$ on the E^{2}PR4 channel, as shown in the following lemma, taken from [26]. This constraint, therefore, permits construction of codes with rates $5/6$ or $6/7$, for example.

Lemma IV.8: Let $\mathcal{F} = \{0101, 11101\}$, A constraint $\mathcal{C} \subseteq X^{\{0,1\}}_{\mathcal{F}}$ satisfies $\mathcal{E}(\mathcal{C}) \subseteq X^{\{\pm,0\}}_{\mathcal{L}}$ where

$$\mathcal{L} = \{\pm[+-+-], \pm[+-+00], \pm[+0+0+0]\}.$$

It, therefore, eliminates all closed error events with $d^2(\boldsymbol{\epsilon}) < d^2_{\mathrm{MFB}}$ on the E^{2}PR4 and E^{3}PR4 channels. On the E^{2}PR4 channel, it also removes all but one of the closed events with $d^2(\boldsymbol{\epsilon}) = 10$, as well as all open error events with $d^2(\boldsymbol{\epsilon}) < 10$.

Proof: Suppose $\boldsymbol{\epsilon} = \boldsymbol{a} - \boldsymbol{b}$. If $\epsilon_i\epsilon_{i+1}\epsilon_{i+2}\epsilon_{i+3} = +-+-$, then $a_ia_{i+1}a_{i+2}a_{i+3} = 1010$ and $b_ib_{i+1}b_{i+2}b_{i+3} = 0101$. Clearly, this situation cannot arise if the input string 0101 is forbidden. A similar argument applies to the case $-+-+$. If $\epsilon_i\epsilon_{i+1}\epsilon_{i+2}\epsilon_{i+3}\epsilon_{i+4} = 00+-+$, then we must have $a_ia_{i+1}a_{i+2}a_{i+3}a_{i+4} = c_1c_0101$ and $b_ib_{i+1}b_{i+2}b_{i+3}b_{i+4} = c_1c_010$. There are several cases to consider. If $a_0 = 0$, then $\boldsymbol{a}$ would contain the forbidden string 0101. If $a_0 = 1$ and $a_1 = 0$, then $\boldsymbol{b}$ would contain 0101. If $a_0 = 1$ and $a_1 = 1$, then $\boldsymbol{a}$ would contain 11101. The string $00-+-$ is handled similarly. Finally, if

$$\epsilon_i\epsilon_{i+1}\epsilon_{i+2}\epsilon_{i+3}\epsilon_{i+4}\epsilon_{i+5}\epsilon_{i+6} = +0+0+0+$$

we must have

$$a_ia_{i+1}a_{i+2}a_{i+3}a_{i+4}a_{i+5}a_{i+6} = 1c_01c_11c_21$$

and

$$b_ib_{i+1}b_{i+2}b_{i+3}b_{i+4}b_{i+5}b_{i+6} = 0c_00c_10c_20.$$

The cases $b_0 = 0, b_1 = b_2 = 1$, and $b_0 = b_1 = 1$ force the string 0101 to occur in $\boldsymbol{b}$. The case $b_0 = b_1 = 0$ forces 0101 to occur in $\boldsymbol{a}$. Finally, the cases $b_0 = 1, b_1 = 0$ and $b_0 = 0, b_1 = 1, b_2 = 0$ force the appearance of 11101 in $\boldsymbol{a}$. The lemma follows. □

Remark: The graph $\mathsf{G}_3(X^{\{0,1\}}_{\mathcal{F}})$ was shown in [26] to require 15 states, of which three had only a single incoming edge. It can also be shown that the truncation depth for $d^2 = 10$ of this constraint is exactly 10, indicating that a path memory of length 10 can be used to achieve the expected coding gains.

4) $\mathcal{L} = \{\pm[0+-+0]\}$: A $k_1 = 2$ NRZI Code: Recall from the analysis of distance properties of the E^{2}PR4 channel that, for closed events $\epsilon(D)$

$$\|\epsilon(D)(1-D)(1+D)^3\|^2 = 6, \qquad \text{iff } \epsilon(D) = 1 - D + D^2 \tag{25}$$

and

$$\|\epsilon(D)(1-D)(1+D)^3\|^2 = 8 \tag{26}$$

iff

$$\epsilon(D) = 1 - D + D^2 + D^5 - D^6 + D^7 \tag{27}$$

or

$$\epsilon(D) = \sum_{i=0}^{l-1} (-1)^i D^i, \qquad \text{for } l \geq 4. \tag{28}$$

Therefore, another way to eliminate these closed error sequences is to limit the runs of transitions to five, to eliminate the events satisfying (28), and insure that

$$\mathcal{E}(\mathcal{C}) \subseteq X^{\{\pm,0\}}_{\{0+-+0,0-+-0\}}$$

to eliminate the events satisfying (25) or (27).

Fig. 6 explains why the NRZI constraint with forbidden list $\mathcal{F} = \{111\}_{\mathrm{NRZI}}$, that is, the $k_1 = 2$ constraint with NRZI modulation, achieves the MFB. The figure shows all pairs of NRZ binary sequences generating the error strings in $\mathcal{L}$. It is easy to verify that the NRZ constraint corresponding to the $k_1 = 2$ NRZI constraint, that is, with forbidden list

$$\begin{array}{ll} \boldsymbol{a}: & \mathbf{0}\,\boxed{\mathbf{1}\;\mathbf{0}\;\mathbf{1}}\,0 \qquad 1\,\boxed{\mathbf{1}\;\mathbf{0}\;\mathbf{1}}\,\mathbf{0} \\ \boldsymbol{b}: & 0\,\boxed{0\;1\;0}\,0 \qquad 1\,\boxed{0\;1\;0}\,0 \\ \\ \boldsymbol{a}: & \mathbf{0}\,\boxed{\mathbf{1}\;\mathbf{0}\;\mathbf{1}}\,1 \qquad 1\,\boxed{1\;0\;1}\,1 \\ \boldsymbol{b}: & 0\,\boxed{0\;1\;0}\,1 \qquad \mathbf{1}\,\boxed{\mathbf{0}\;\mathbf{1}\;\mathbf{0}}\,1 \end{array}$$

Fig. 6. Possible pairs of sequences for which error subsequence $\epsilon(D) = \pm D^s(1 - D + D^2) + 0 \cdot D^{s+3}$ may occur.

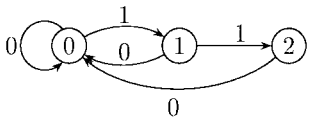


Fig. 7. Graph representation of $k_1 = 2$ NRZI constrained system.

TABLE V
ENCOSER TABLE FOR RATE $4/5$ $k_1 = 2$ BLOCK CODE

input	output
0000	00001
0001	00101
0010	01001
0011	01101
0100	10001
0101	10101
0110	00010
0111	00100
1000	00110
1001	01000
1010	01010
1011	01100
1100	10000
1101	10010
1110	10100
1111	10110

$\mathcal{F} = \{0101, 1010\}_{\mathrm{NRZ}}$, forbids at least one sequence in each pair.

The constrained system with $k_1 = 2$ and NRZI modulation is represented by the graph shown in Fig. 7. The constraint allows construction of codes with finite-state encoders and sliding-block decoders with rate of up to about 0.8791. A rate $4/5$ block code is shown in Table V.

The graph $\mathsf{G}_3(X^{\{0,1\}}_{\{1010,0101\}})$ requires 14 states, of which two have only a single incoming edge. This constraint does not eliminate all open error events with $d^2(\boldsymbol{\epsilon}) < 10$, namely, the $d^2 = 6$ open event corresponding to the input error sequence $\boldsymbol{\epsilon} = [+0]^\infty$ and the $d^2 = 8$ open event corresponding to $\boldsymbol{\epsilon} = 100[+0]^\infty$. Thus unlike the two previous examples, this constrained channel graph does not have finite truncation depth. However, the rate $4/5$ block code imposes additional "incidental" constraints that prevent the occurrence of these events, ensuring that the generalized truncation depth [25] is finite. The required path memory will nevertheless be greater than in either of the previous two constraints. In particular, the pair of sequences (shown at the bottom of this page), one of which $\boldsymbol{x}$, is a valid code sequence, accumulate only $d^2 = 8$ over 16 symbols, showing that the truncation depth is at least 15 symbols.

The codewords in Table V are precisely the binary complement of the Group Code Recording (GCR) code, which was a tape-industry standard [22, p. 120]. (Recall that Soljanin had observed that the binary complement of the rate $8/9$, $(d, k) = (0, 3)$ code improved the off-track performance of the EPR4 channel.)

Remark: This constraint was independently introduced and first published by Moon and Brickner [8], who called it a *maximum-transition-run (MTR)* constraint. The rate $4/5$ code in Table V was also found independently [41]. See also the patent [10]. Subsequent design and performance analysis of codes satisfying the MTR constraint is discussed in recent papers [27], [35], [11], and [29].

5) Further Codes from Finite-Type Constraints: The following lemma confirms that the constraint with $\mathcal{F} = \{0101, 111010, 1110111\}_{\mathrm{NRZ}}$ eliminates all of the $d^2(\boldsymbol{\epsilon}) < 10$ events, both open and closed, on the E^{2}PR4 channel. The constraint has capacity approximately 0.8791, like the MTR constraint, permitting the construction of a code with rate $7/8$.

Lemma IV.9: Let $\mathcal{F} = \{0101, 111010, 1110111\}$. A constraint $\mathcal{C} \subseteq X^{\{0,1\}}_{\mathcal{F}}$ satisfies $\mathcal{E}(\mathcal{C}) \subseteq X^{\{\pm,0\}}_{\mathcal{L}}$ where

$$\mathcal{L} = \{\pm[+ - +-], \pm[00 + - + 00], \pm[+0 + 0 + 0+]\}.$$

It, therefore, eliminates all closed error events with $d^2(\boldsymbol{\epsilon}) < d^2_{\mathrm{MFB}}$ on the E^{2}PR4. On the E^{2}PR4 channel, it also removes all but one of the closed events with $d^2(\boldsymbol{\epsilon}) = 10$, as well as all open error events with $d^2 < 10$.

Proof: First consider error sequences containing the string $\pm[+ - +-]$. It is immediate that these cannot occur if the constraint forbids the string 0101. Next, suppose the sequence contains $00 + - + 00$. The constrained sequences must then contain strings

$$a_i a_{i+1} a_{i+2} a_{i+3} a_{i+4} a_{i+5} a_{i+6} = c_0 c_1 101 c_2 c_3$$

and

$$b_i b_{i+1} b_{i+2} b_{i+3} b_{i+4} b_{i+5} b_{i+6} = c_0 c_1 010 c_2 c_3.$$

We now consider several cases. If $c_1 = 0$, then $\boldsymbol{a}$ contains 0101. If $c_1 = 1$ and $c_0 = 0$, then $\boldsymbol{b}$ contains 0101. If $c_1 = 1$

NRZI $\boldsymbol{x}$	0	0	1	0	1	0	0	1	1	0	0	1	1	0	0	1
NRZI $\boldsymbol{y}$	0	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0
NRZ $\boldsymbol{a}$	1	1	0	0	1	1	1	0	1	1	1	0	1	1	1	0
NRZ $\boldsymbol{b}$	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0
ϵ	0	+	0	0	+	0	+	0	+	0	+	0	+	0	+	0
d^2	0	1	4	0	1	1	1	0	0	0	0	0	0	0	0	0

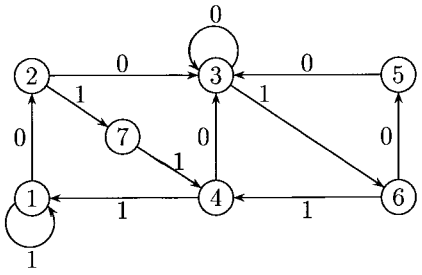


Fig. 8. Shannon cover for $X_{\{0101,111010\}}$ constraint.

and $c_0 = 1$, we must consider possible values of c_2 and c_3. In particular, if $c_2 = 1$, then $\boldsymbol{b}$ contains 0101. If $c_2 = 0$ and $c_3 = 1$, then $\boldsymbol{a}$ contains 0101. If $c_2 = 0$ and $c_3 = 0$, then $\boldsymbol{a}$ contains 1110100, Note that these cases cover all closed input error sequences, as well as those open input error sequences not containing $\pm[+0]^{\infty}$. A similar case-by-case analysis shows that forbidding the string 1110111 as well will prevent the occurrence of $+0+0+0+$ and, therefore, eliminate the remaining open events. □

The lemma also shows that the forbidden list $\mathcal{F} = \{0101, 111010\}_{\mathrm{NRZ}}$ suffices to eliminate all events with $d^2(\boldsymbol{\epsilon}) < 10$ except for the same open events that were allowed by the MTR constraint. The capacity of the constraint is approximately 0.8960, thereby permitting the design of a code with rate $8/9$. This is significant because, in practice, a code rate of $8/9$ or higher is often desirable for PRML systems. The Shannon cover for this constraint is shown in Fig. 8.

The memory of this constraint is 6, so the graph $\mathsf{G}_3(\mathcal{C})$ is no longer a subgraph of G_3. However, it is easy to confirm that the graph requires only 16 states. Application of Franaszek's algorithm for finding an approximate eigenvector of G^9 suitable for construction of a rate $8/9$ code yields

$$\boldsymbol{v} = [6\,5\,6\,6\,3\,5\,3].$$

Through suitable state-splitting and state-merging operations, one can construct a practical sliding-block code. The code design must also eliminate the sequence $11101110\cdots$ in order to prevent the open events with $d^2 < 10$ that were not avoided by the constraint.

Moision *et al.* [40] identified NRZI constraints that eliminate all closed events with $d^2(\boldsymbol{\epsilon}) < 10$ and support a code rate $8/9$. The distance-enhancing properties of one such constraint, corresponding to the forbidden error string list

$$\mathcal{L} = \{\pm[0+-+00]\}$$

are confirmed in the following lemma.

Lemma IV.10: Consider the finite-type NRZI constraint determined by $\mathcal{F} = \{1111, 11100\}_{\mathrm{NRZI}}$. A constraint $\mathcal{C} \subseteq X_{\mathcal{F}}^{\{0,1\}}$ satisfies $\mathcal{E}(\mathcal{C}) \subseteq X_{\mathcal{L}}^{\{\pm,0\}}$ where $\mathcal{L} = \{\pm[0+-+00]\}$. Therefore it eliminates all closed error events with $d^2(\boldsymbol{\epsilon}) < d^2_{\mathrm{MFB}}$ on the E^{2}PR4. It also removes all but one of the closed events with $d^2(\boldsymbol{\epsilon}) = 10$, as well as all open error events with $d^2(\boldsymbol{\epsilon}) < 10$ except those allowed by the MTR constraint.

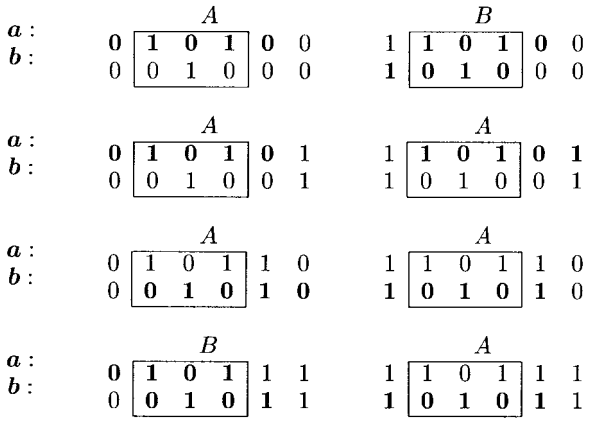


Fig. 9. Possible pairs of sequences for which error event $\epsilon(D) = \pm D^s(1 - D + D^2) + 0 \cdot D^{s+3} + 0 \cdot D^{s+4}$ may occur.

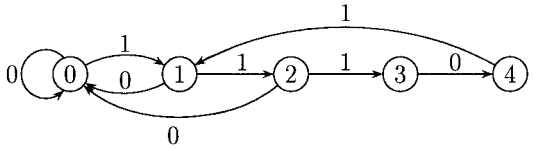


Fig. 10. Shannon cover for NRZI constraint $X_{\{1111,11100\}}$ constraint.

Proof: The proof follows from examination of Fig. 9, which represents all possible pairs of binary sequences corresponding to the error strings in the set $\mathcal{L}$. Within each pair of sequences denoted by type A, one can find a run of four transitions. Within each pair denoted by type B, one can find four symbols corresponding to three successive transitions, followed by two symbols corresponding to no transition. Examination of the tables of error events confirms that the constraint eliminates the specified closed and open error events. □

The Shannon cover for this constraint is shown in Fig. 10.

The graph underlying the detector trellis requires 18 states. Application of Franaszek's algorithm for finding an approximate eigenvector of G^9 suitable for construction of a rate $8/9$ code yields

$$\boldsymbol{v} = [2\,2\,1\,0\,1].$$

Remark: In [40], it is shown that this NRZI constraint, along with others with higher capacity, improve the performance of a Lorentzian channel equalized to an EPR4 target. The analysis of the system makes use of a characterization of input error sequences in terms of an "effective distance" d_{eff} at the channel output, a measure that includes the effect of the equalizer-induced noise coloration at the detector input.

B. Strictly AFT Distance–Enhancing Constraints

In this section we consider distance-enhancing constraints that are strictly almost-finite-type (AFT). We first make a few remarks about spectral-null constraints defined by moment conditions, and then consider a new class of constraints described by time-varying lists of forbidden input strings.

1) MSN Codes Revisited: As was the case for the NRZI $d_0 = 1$ constraint, we can use the error event characteriza-

tion to understand the distance-enhancing properties of MSN constraints [25]. The E^{2}PR4 channel with system polynomial $h(D) = (1-D)(1+D)^3$ has a transfer function with a first-order spectral null at frequency $f = 0$ and a third-order null at $f = 1/2$. Similarly, the E^{3}PR4 channel, with polynomial $h(D) = (1-D)(1+D)^4$, has a transfer function with a first-order spectral null at frequency $f = 0$ and a fourth-order null at $f = 1/2$. The MSN coding theorem indicates that the use of a code with a Kth-order null at $f = 1/2$, for $K \geq 0$, on $h_N(D)$ will generate output sequences with $d^2_{<>} \geq 2(K+N)$. In particular, for E^{2}PR4, a second-order null at $f = 1/2$ will ensure that $d^2_{<>} \geq 10$. This can be confirmed by examination of the closed input error sequences in Table I. Moreover, it can be seen that a first-order null at $f = 1/2$ may not suffice to ensure $d^2_{<>} \geq 10$. However, the combination of a first-order null at $f = 0$ and a first-order null at $f = 1/2$ will guarantee $d^2_{<>} \geq 10$. These observations are proved in the following lemma.

Lemma IV.11: If $\mathcal{C}$ is a constraint with a second-order null at $f = 1/2$, or first-order nulls at both $f = 0$ and $f = 1/2$, the system presented by $\mathsf{G}_3(\mathcal{C})$ will have minimum merged distance $d^2_{<>} \geq 10$.

Proof: The closed input error sequences $\boldsymbol{\epsilon} = \epsilon_0, \cdots, \epsilon_n$, $\boldsymbol{\epsilon} \in \mathcal{E}(\mathcal{C})$ generated by a constraint with a first-order null at $f = 1/2$ must satisfy the moment condition

$$\sum_{k=0}^{n} (-1)^k \epsilon_k = 0. \tag{29}$$

If the constraint has a second-order null at $f = 1/2$, these error sequences satisfy the additional constraint

$$\sum_{k=0}^{n} (-1)^k k \epsilon_k = 0. \tag{30}$$

From Table I, it is easy to see that this pair of conditions is not satisfied by any of the closed error input sequences for $d^2(\boldsymbol{\epsilon}) < 10$. In fact, with the exception of the $d^2 = 8$ sequences having the form $\boldsymbol{\epsilon} = \pm[+-00+-]$, the sequences do not satisfy condition (29) for a first-order null at $f = 1/2$. The error sequences of a constraint with a first-order null at $f = 0$ satisfy

$$\sum_{k=0}^{n} \epsilon_k = 0. \tag{31}$$

Since the exceptional sequences $\boldsymbol{\epsilon} = +-00+-$ do not satisfy (31), it follows that first-order nulls at both frequencies suffice to eliminate all closed input error sequences with $d^2(\boldsymbol{\epsilon}) < 10$. □

2) $\mathcal{L} = \{\pm[+-+]\}$: An $\mathcal{F}^e = \{101\}$ and $\mathcal{F}^o = \{010\}$ NRZ Constraint: In this subsection and the next, we consider time-varying forbidden list constraints. (Time-varying constraints were also considered by Fredrickson [17].) These constraints specify lists $\mathcal{F}^e$ and $\mathcal{F}^o$ of strings forbidden to begin at even or odd time positions, respectively. For notational convenience, we will denote a string $\boldsymbol{a}$ forbidden to begin at an even or odd time position by $\boldsymbol{a}^e$ or $\boldsymbol{a}^o$, respectively. These constraints are represented by graphs with period 2.

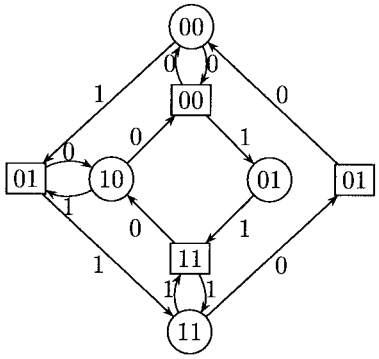


Fig. 11. Graph representation of $\mathcal{F}^e = \{101\}$ and $\mathcal{F}^o = \{010\}$ NRZ constrained system.

Winkler [64] discovered a strictly AFT constraint $\mathcal{C}$ for which $\mathcal{E}(\mathcal{C}) \subseteq X^{\{\pm,0\}}_{\{+-+,-+-\}}$. The constraint reflects the fact that if code sequences contain neither the string 101 starting at an even position nor the string 010 starting at an odd position, then their normalized differences cannot contain strings $+-+$ and $-+-$. A graph representing the system with forbidden list $\mathcal{F} = \{101^e, 010^o\}_{\mathrm{NRZ}}$, that is, with $\mathcal{F}^e = \{101\}_{\mathrm{NRZ}}$ and $\mathcal{F}^o = \{010\}_{\mathrm{NRZ}}$, is shown in Fig. 11.

Squares denote states where the next label generated will occupy an even position, and circles do likewise for odd positions. This constrained system allows construction of codes with finite-state encoders and sliding-block decoders with rate bounded by its capacity, which is precisely equal to that of the MTR constraint, approximately 0.8791. In fact, as described in [38], there is a simple relationship between the two constrained systems.

3) $\mathcal{L} = \{\pm[0+-+00]\}$: A $k^o_1 = 3$ and $k^e_1 = 2$ NRZI Code: In this subsection, we describe an NRZI constraint that imposes different maximum transition runlength limitations according to the parity of the position at which the run begins. This constraint may be viewed as a variation on the MTR constraint described earlier. This time-varying MTR (TMTR) constraint, has capacity approximately 0.916 and permits construction of a rate $8/9$ code. Moreover, as will be shown, a simple block code is possible.

Lemma IV.12: Consider the finite-type NRZI constrained system $\mathcal{C}$ defined by the conditions $k^o_1 = 3$ and $k^e_1 = 2$. Then $\mathcal{E}(\mathcal{C}) \subseteq X^{\{\pm,0\}}_{\{+-+00,-+-00\}}$. Therefore, the constraint eliminates all closed error events with $d^2(\boldsymbol{\epsilon}) < d^2_{\mathrm{MFB}}$ on the E^{2}PR4 channel. On the E^{2}PR4 channel, it also removes all but one of the closed events with $d^2(\boldsymbol{\epsilon}) = 10$, as well as all open error events with $d^2(\boldsymbol{\epsilon}) < 10$ except those allowed by the MTR constraint.

Proof: The lemma follows from the case-by-case analysis depicted in Fig. 9, previously used in the proof of Lemma IV.10. We can distinguish two types—denoted by A and B in the figure—of pairs of code sequences involved in forming an error event. In a pair of type A, at least one of the sequences has a transition run of length 4. In a pair of type B, both sequences have transition runs of length 3, but for one of them the run starts at an even position and for the other at

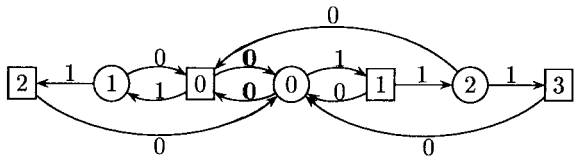


Fig. 12. Graph representation of $k_1^o = 3$ and $k_1^e = 2$ NRZI constrained system.

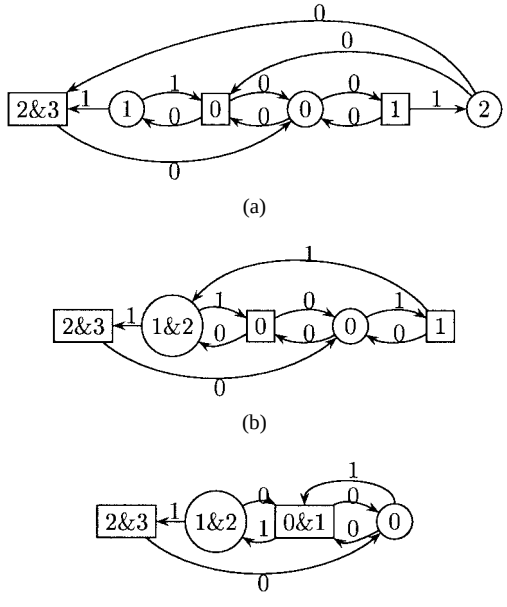


Fig. 13. Merging of states for $k_1^o = 3$ and $k_1^e = 2$ constrained system. (a) After merging states ② and ③. (b) After merging states ① and ②. (c) After merging states ⓪ and ①.

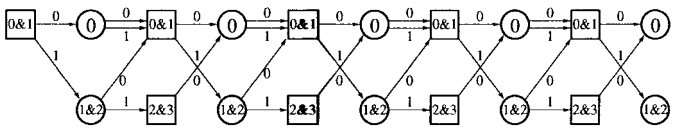


Fig. 14. Trellis diagram for $k_1^3 = 2$ and $k_1^e = 2$ constrained system.

an odd position. Therefore, a $k_1^o = 3$ and $k_1^e = 2$ NRZI code eliminates all possibilities. □

A graph representing the $k_1^o = 3$ and $k_1^e = 2$ NRZI constrained system is shown in Fig. 12. This graph representation can be simplified by means of state merging as described in the Appendix. The simplification is accomplished in three steps by merging two states in each step, as shown in Fig. 13.

The adjacency matrix of the graph is given by

$$A = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

where the rows and columns correspond to the states $0, 0\&1, 1\&2, 2\&3$, respectively, and the capacity satisfies $C(A) > 0.916\cdots$. The 8th and 18th powers of the adjacency matrix, shown below, indicate that the construction of rate $7/8$ and rate $16/18$ block codes is straightforward

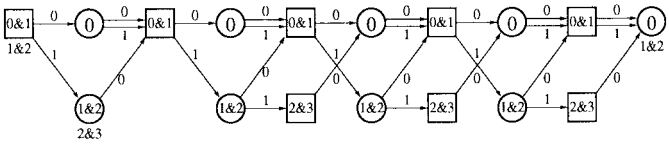


Fig. 15. Trellis diagram for a rate $8/9$ $k_1^o = 3$ and $k_1^e = 2$ constrained block code.

$$A^8 = \begin{bmatrix} 100 & 0 & 78 & 0 \\ 0 & \boxed{139 > 2^7} & 0 & 39 \\ 78 & 0 & 61 & 0 \\ 0 & 78 & 0 & 22 \end{bmatrix}$$

$$A^{18} = \begin{bmatrix} 57284 & 0 & 44726 & 0 \\ 0 & \boxed{79647 > 2^{16}} & 0 & 22363 \\ 44726 & 0 & 34921 & 0 \\ 0 & 44726 & 0 & 12558 \end{bmatrix}$$

To design a rate $8/9$ block code we consider the trellis diagram (representation in time) of the graph shown in Fig. 13(c). The trellis is time-varying and is shown in Fig. 14, starting from an even state with at most one transition and ending in one of the two odd states after nine steps. Therefore, the situation at the end can be described by a single state with zero, one, or two transitions as shown in Fig. 15. If these length 9 strings are to be freely concatenated, then in the starting state there can be zero, one, or two transitions. This leads to zero, two, or three transitions in one of the states in the second step. This implies that the only allowed branch from that state must be labeled by a 0, which completes the diagram in Fig. 15. The number of paths through this trellis is 267, out of which $2^8 = 256$ can be chosen for a rate $8/9$ block codebook. The design of specific encoding and decoding functions for such a block code have been presented by Bliss [8] and Young [70].

A 16-state, time-varying trellis may be used for detection of this rate $8/9$ TMTR code on the E^2PR4 channel. Further details pertaining to VLSI implementation of the detector are discussed by Young in [70].

Remark: The $k_1^o = 3$ and $k_1^e = 2$ TMTR constraint and a simple rate $8/9$ block code were independently found and first published by Bliss [8]. Related results were also found independently by R. Wood [69], H. Thapar [58], and K. Knudson Fitzpatrick and C. Modlin [15]. In [15], a rate $9/10$ block code satisfying a time-varying MTR-type constraint and with similar distance-enhancing properties is also presented.

Remark: It has been shown that the $k_1^o = 3$ and $k_1^e = 2$ TMTR constraint also achieves the MFB on the E^3PR4 channel [39].

A $k_1^o = 2$ and $k_1^e = 1$ TMTR constraint has been shown to achieve the MFB on the (PR2) channel, with $h(D) = (1+D)^2$, as well as on the "extended PR2 (EPR2)" channel, with $h(D) = (1+D)^3$. The capacity of this constraint is aproximately 0.7925, and a simple rate $3/4$ block code has been described in [39] and [11].

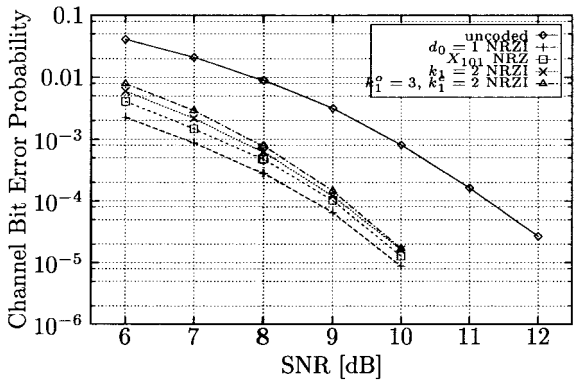


Fig. 16. Performance of uncoded and coded E^2PR4 systems.

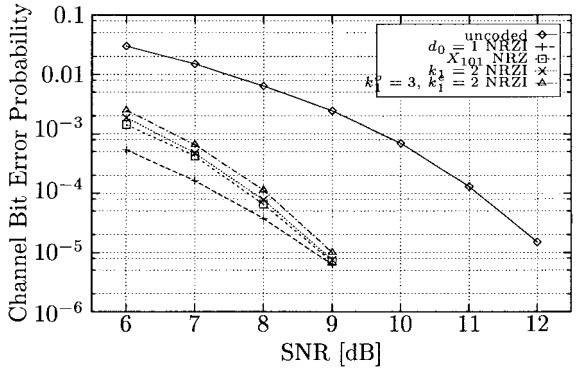


Fig. 17. Performance of uncoded and coded E^3PR4 systems.

C. Code Performance

We tested by computer simulation a rate $2/3\,(1,7)$ NRZI code, the rate $4/5$ $X_{\{101\}}$ NRZ code described in [26], the rate $4/5$ $k_1 = 2$ NRZI code, and the rate $8/9$ $k_1^o = 3$ and $k_1^e = 2$ TMTR code. The error probability performance of the codes on the E^2PR4 channel is shown in Fig. 16. The small differences in performance for high SNR reflect the different error event multiplicities, or "error coefficients," induced by the constraints and the specific codes. In general, a code based on $X_{\{101\}}$ NRZ constrained system should have lower error coefficient since the sequences satisfying the $X_{\{101\}}$ NRZ constraint also satisfy $k_1 = 2$ NRZI (MTR) constraint. (It should be noted, however, that sequences generated by a particular code based on the $X_{\{101\}}$ NRZ constraint may not not belong to a particular code based on $k_1 = 2$ NRZI constraint. For example, the code described in [26] permits symbol runlengths of 13 whereas the $k_1 = 2$ NRZI described above permits runlengths of up to 9. Thus the specific code construction can potentially affect the system error coefficients.) The $(1,7)$ code has the smallest error coefficient for the minimum-distance events. It allows minimum-distance error events only at places where a string of 0's (longer than 2) changes into a string of 1's (longer than 2) and *vice versa*.

In the higher error-rate region, the coding gain increases because the codes remove some of the non-minimum-distance error events, as well. For example, based on the analysis in [51], it is easy to see that error sequences of the form $1 + D^2 + D^4 + \cdots + D^{2k}, k \geq 1$, have squared distance 12 on the E^2PR4 channel. The $d_1 = 1$ NRZI constrained system does not permit any of these error events, and the other systems permit only those for which $k = 1$.

The error probability performance of the codes on the E^3PR4 channel is shown in Fig. 17. As in the case of the E^2PR4 channel, an evaluation of the error event multiplicities may be used to interpret the observed performance of the different constraints and codes.

Appendix
Constrained Systems and Codes

In this section we review some necessary concepts, terminology, and results from the theory of symbolic dynamics and constrained codes. For a more comprehensive treatment, see [31], [33], and [34].

A. Labeled Graphs and Constraints

A *labeled graph* $\mathsf{G} = (V, E, L)$ consists of a finite set of states $V = V(\mathsf{G})$, a finite set of directed edges $E = E(\mathsf{G})$, and an edge labeling $L = L(\mathsf{G})$ that assigns to each edge a symbol in a finite alphabet $\mathcal{A}$. Each edge e has an initial state $\sigma(e)$ and a terminal state $\tau(e)$. When there is no possible ambiguity, a labeled graph may be called simply a graph. A path in G is a finite sequence of edges in G in which the initial state of an edge having a predecessor corresponds to the terminal state of that predecessor. A *constrained system* X, sometimes referred to as a *constraint*, is the set of all symbol strings generated by the labeling of paths in G. This system is sometimes denoted by $X(\mathsf{G})$. The system $X(\mathsf{G})$ is said to be *presented* by G.

A *labeled graph* G is *irreducible* if there is a path in G from any specified starting state i to any specified destination state j; otherwise it is *reducible*. A constrained system X is irreducible if it can be presented by an irreducible graph. In coding applications, the graphs of primary interest are irreducible.

A graph is *deterministic* if at each state the labels on outgoing edges are distinct. Any constrained system X can be presented by a deterministic graph. A graph has *finite local anticipation* a if there exists a nonnegative integer a such that, for every state i, all of the paths of length $a+1$ that start at i and generate the same sequence begin with the same edge. The *local anticipation* $a(\mathsf{G})$ is the smallest such integer. A deterministic graph has local anticipation $a = 0$. If no such a exists, we say that the graph has *infinite local anticipation*. Definitions of *finite local memory* m and *local memory* are similarly stated by considering paths that end at a given state. Thus a graph has local memory m if m is the smallest nonnegative integer such that, for every state i, all of the paths of length $m+1$ that end at i and generate the same sequence end with the same edge. A graph has finite local memory if it has local memory m for some nonnegative integer m, otherwise it has *infinite local memory*. A constrained system X is *almost finite type* (AFT) if it can be presented by a graph with finite local anticipation and finite local memory.

A labeled graph presenting a constrained system X has *memory* m and *anticipation* a if, given any sequence $\overline{x} = x_{-m}, \cdots, x_0, \cdots, x_a \in X$, the set of paths $\overline{e} = e_{-m}, \cdots, e_0, \cdots, e_a$ that generate $\overline{x}$ all agree in the edge

e_0. A graph with this property is also called (m,a)-definite. Definite graphs are said to have *finite memory and anticipation*. A constrained system X is *finite type* (FT) if it can be presented by a graph G with finite memory and anticipation. It can be shown that, if G has finite memory and anticipation, then it has finite local memory and finite local anticipation. Therefore, any constrained system which is FT is also AFT. The reverse is not true. An FT system X over alphabet $\mathcal{A}$ can always be characterized by a finite list of forbidden strings $\mathcal{F} = \{w_1, \cdots, w_N\}$ of symbols in $\mathcal{A}$. An FT system defined in such way will be denoted by $X_{\mathcal{F}}^{\mathcal{A}}$. In Section IV we have considered several constraints that are naturally described in this fashion.

A simple, but crucial, observation is that there may be many different graphs presenting a given constraint X. Two operations that play a role in the transformation of one presentation to another are state-splitting and state-merging. An out-splitting of a state s in G is defined as follows. Partition the set of edges with initial state s, denoted $E(s)$, into N disjoint subsets

$$E(s) = E_s^1 \cup \cdots \cup E_s^N.$$

The graph H created by out-splitting according to this partition has state set $V(\mathsf{H}) = \{u \in V(\mathsf{G}) | u \neq s\} \cup \{s^1, \cdots, s^N\}$, where the states $s^i, i = 1, \cdots, N$, are the descendants of s. For a state $u \in V(\mathsf{H}), u \neq s$, edges are inherited from the ancestor state in $V(\mathsf{G})$, with any edge $e \in E(\mathsf{G})$ satisfying $\tau(e) = s$ replicated N times, yielding edges e^i with terminal states s^i, for $i = 1, \cdots, N$. The outgoing edges $E(s)$ are partitioned among the descendant states, with an edge $e \in E_s^i$ being assigned to state s^i. Replication of any edge descending from an edge e in $E(\mathsf{G})$ with $\tau(e) = s$ obeys the same rule as before. An in-splitting of a state in G is described by reversing the role of outgoing and incoming edges in the definition of an out-splitting of a state.

In is often desirable in practice to present a constrained system by a graph with the smallest possible number of states. It is desirable as well to design an encoder graph with the smallest possible number of states. Sometimes a graph can be reduced by means of an operation called out-merging. A condition for out-merging two states can be described in terms of their follower sets. The follower set of state s, denoted by $F(s)$ is the set of all words of all lengths generated by paths in G starting from s. The inclusion relation on follower sets imposes a partial ordering of states, namely, $s \prec t$ if $F(s) \subseteq F(t)$. When $s \prec t$, the out-merger of s and t creates a new graph H from G by first eliminating all edges in $E(t)$, redirecting into state s all remaining edges coming into state t, and then eliminating state t. If G is deterministic, then $F(s) \subseteq F(t)$ implies $F(s) = F(t)$, and we say that states s and t are follower-set equivalent. In this situation, states s and t can be merged. A deterministic graph G is said to be *reduced* if, for any pair of states s and $t, F(s) \neq F(t)$. Among the deterministic graphs presenting an irreducible constrained system X, there is a unique (up to labeled graph isomorphism) irreducible, deterministic, reduced presentation. This presentation has the smallest number of states of any deterministic presentation. We refer to this graph as the *Shannon cover*. The Shannon cover of a constraint often is the graph underlying the maximum-likelihood detector trellis structure for input-constrained channels.

The qth power of G, denoted G^q, is the graph with the same state set $V(\mathsf{G})$ as G, and with an edge corresponding to each path of length q, labeled by the length-q string generated by the path. If X is the constrained system presented by G, then the system presented by G^q, denoted X^q, is called the qth power of X. The qth higher edge graph $\mathsf{G}^{[q]}$ is the graph with state set corresponding to the paths of length $q-1$ in G, and edges corresponding to paths of length q in G. For example, the path $e_1e_2\cdots e_q$ has initial state $e_1e_2\cdots e_{q-1}$ and terminal state $e_2\cdots e_q$. The label on this edge is the string generated by the path $e_1e_2\cdots e_q$. If X is the constrained system presented by G, then the system presented by G^q, denoted X^q, is called the qth higher order system of X.

When an FT system $X_{\mathcal{F}}^{\mathcal{A}}$ over alphabet $\mathcal{A}$ is defined by a finite list of forbidden strings $\mathcal{F} = \{w_1, \cdots, w_N\}$ of symbols in $\mathcal{A}$, the Shannon cover may be constructed in a number of ways. Let $l(w)$ denote the length of a finite string w and $l_{\max}$ be the length of the longest string in $\mathcal{F}$, that is,

$$l_{\max} = \max_{w \in \mathcal{F}} l(w).$$

Then the Shannon cover can be obtained from the higher edge graph of order $l_{\max}$ of the unconstrained binary system by, first, eliminating all edges whose labels contain one of the forbidden strings and, then, merging states whose follower sets are identical [33]. Another construction, with roots in automata theory, is described in [50]. There, a finite-state automaton generating the complement of the desired constraint is derived from the forbidden list. The automaton is made deterministic, if necessary, using the subset construction [31]. Finally, the roles of accepting and nonaccepting states are reversed. From the resulting automaton, the Shannon cover can then be derived. A third approach uses a sequence of localized edge graph constructions in the form of in-splittings, beginning with the single-state graph of the unconstrained binary system. This allows the selective deletion of edges corresponding to forbidden strings. For some of the constraints considered in later portions of this paper, the latter two procedures lead very simply to the Shannon cover.

B. Finite-State Constrained Codes

A *rate p/q finite-state encoder* is a synchronous finite-state machine that accepts an input block of p user bits and generates a length q binary codeword depending on the input block and the current internal state of the encoder. Given a graph in which each state has 2^p outgoing edges labeled by length q binary codewords, a rate p/q finite-state encoder is easily defined by assigning one of the 2^p possible data symbols to a distinct outgoing edge from each state. That encoder graph can then serve as a *state-dependent decoder*, a synchronous finite-state machine that accepts as input codewords of length q and generates a block of p user bits depending on the internal state, the input codeword, and finitely many upcoming codewords. It is desirable that the decoder be *noncatastrophic*, i.e., that any input symbol error gives rise to a finite number of

output (decoded) symbol errors. In practice, it is preferable to confine the output errors to a certain bounded time interval. A *sliding-block* decoder makes a decision on a given received word on the basis of a local context of the word in the received sequence: the word itself as well as a fixed number m' of immediately preceding words and a fixed number a' of immediately following words. Therefore, a single error at the input to a sliding-block decoder can only affect the decoding of words that fall in a *window* of length $m' + a' + 1$.

The maximum rate of a code into a constrained system $X(\mathsf{G})$ is determined by its *Shannon capacity*. The Shannon capacity or simply *capacity* of a constrained system X, denoted by $C(X)$, is defined as

$$C(X) = \lim_{n\to\infty} \frac{\log_2 N(n;X)}{n}$$

where $N(n;X)$ is the number of sequences of length n in X. It measures the growth rate of the number of strings of length n in the system. The capacity of a constrained system presented by a graph G with finite local anticipation can be easily computed from the *adjacency matrix* (or *state transition matrix*) $A(\mathsf{G})$ of the underlying graph. Specifically

$$C(X(\mathsf{G})) = \log_2 \lambda(A(\mathsf{G}))$$

where $\lambda(A)$ is the largest real eigenvalue of a matrix A.

As mentioned above, Shannon showed that the capacity of constrained system $X(\mathsf{G})$ represents an upper bound of the achievable rate of any finite-state code into systems $X(\mathsf{G})$ [48]. Adler, Coppersmith, and Hassner proved that the rate equal to the capacity is achievable when the capacity is a rational number, and that for FT systems encoders have sliding block decoders [1]. Karabed and Marcus extended this result to the AFT constrained systems [24].

In [1] and [24], the construction of rate $p : q$ finite-state encoders is accomplished by a sequence of out-splittings, guided by an approximate eigenvector $\boldsymbol{v}$ of the qth power of the adjacency matrix of G^q, that is, a vector satisfying the inequality

$$A^q\boldsymbol{v} \geq 2^p\boldsymbol{v}.$$

The existence of such a vector is ensured by the Perron–Frobenius theory of nonnegative matrices. Note that if $\boldsymbol{v}$ is an all-1's vector, then each state of graph G^q has at least 2^p outgoing edges. If an all-1's approximate eigenvector cannot be found for A^q, then G^q can be transformed by an iterative procedure into a graph representing the same constrained system $X(\mathsf{G}^q)$ whose adjacency matrix has an all-1's approximate eigenvector.

These results show that for an FT- or AFT-constrained system, there exist finite-state codes of rates up to the constrained system capacity. A *block code for length* q is a list of q-blocks in $X(\mathsf{G})$ that can be concatenated without violating the constraints defined by G. A block code for length q is *optimal* if there is no other block code for that length with more codewords in it. A method to construct optimal block codes for finite memory constraints was described by Freiman and Wyner in [19]. Several important implementation issues, such as quasi-catastrophic error propagation [25] and possibilities for simplified, suboptimal decoding [55], favor block codes.

ACKNOWLEDGMENT

The authors wish to thank to S. Altekar, B. Moision, H. Thapar, P. Winkler, A. J. Van Wijngaarden, and R. Wood for helpful discussions.

REFERENCES

[1] R. Adler, D. Coppersmith, and M. Hassner, "Algorithms for sliding-block codes," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 5–22, Jan. 1983.

[2] S. Altekar, "Detection and coding techniques for magnetic recording channels," Ph.D. dissertation, Univ. Calif. San Diego, June 1997.

[3] S. A. Altekar, M. Berggren, B. E. Moision, P. H. Siegel, and J. K. Wolf, "Error-event characterization on partial-response channels," in *Proc. 1997 IEEE Int. Symp. Information Theory* (Ulm, Germany, June 29–July 4), p. 461.

[4] ———, "Error-event characterization on partial-response channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 241–247, Jan. 1999.

[5] J. Ashley, "A linear bound for sliding-block decoder window size," *IEEE Trans. Inform. Theory*, vol. 34, pp. 389–399, May 1988.

[6] R. Behrens and A. Armstrong, "An advanced read/write channel for magnetic disk storage," in *Proc. 26th Asilomar Conf. Signals, Systems, and Computers* (Pacific Grove, CA, Oct. 1992), pp. 956–960.

[7] M. Berggren, "An algorithm for characterizing error event sequences in partial response channels," Diploma Project, Swiss Fed. Inst. Technol. Zurich, Switzerland, Aug. 1996.

[8] W. G. Bliss, "An 8/9 rate time-varying trellis code for high density magnetic recording," in *Proc. 1997 IEEE Int. Magnetics Conf.* (New Orleans, LA, Apr. 1–4); also. *IEEE Trans. Magn.*, vol. 33, pp. 2746–2748, Sept. 1997.

[9] W. G. Bliss, S. She, and L. Sundell, "The performance of generalized maximum transition run trellis codes," in *Proc. 1997 IEEE Magnetic Recording Conf. (TMRC'97)*; also *IEEE Trans. Magn.*, vol. 34, pt. 1, pp. 85–90, Jan. 1998.

[10] B. Brickner and J. Moon, "Method and apparatus for implementing maximum transition codes," U.S. Patent Pending 60/014 954, filed Oct. 1996.

[11] ———, "Investigation of error propagation in DFE and MTR$(1/2;k)$ coding for ultra-high density," Tech. Rep., Commun. and Data Storage Lab, Univ. Minnesota, Minneapolis, July 10, 1997.

[12] ———, "Design of a rate 6/7 maximum transition run code," in *Proc. 1997 IEEE Int. Magnetics Conf.* (New Orleans, LA, Apr. 1–4); also, *IEEE Trans. Magn.*, vol. 33, pt. 1, pp. 2749–2751, Sept. 1997.

[13] V. Braun, K. A. S. Immink, M. A. Ribiero, and G. J. van den Enden, "On the application of sequence estimation algorithms in the Digital Compact Cassette (DCC)," *IEEE Trans. Consumer Electron.*, vol. 40, pp. 992–998, Nov. 1994.

[14] A. R. Calderbank, C. Heegard, and T. A. Lee, "Binary convolutional codes with application to magnetic recording," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 797–715, Nov. 1986.

[15] K. Knudson Fitzpatrick and C. S. Modlin, "Time-varying MTR codes for high density magnetic recording," in *Poc. 1997 IEEE Global Telecommun. Conf. (GLOBECOM '97)* (Phoenix, AZ, Nov. 1997), pp. 1250–1253.

[16] K. Knudson Fitzpatrick, "A reduced complexity EPR4 post-processor," in *Proc. 1997 IEEE Magnetic Recording Conf. (TMRC'97)*, also, *IEEE Trans. Magn.*, vol. 34, pt. 1, pp. 135–140, Jan. 1998.

[17] L. Fredrickson, "Time-varying modulo N trellis codes for input restricted partial response channels," U.S. Patent 5 257 272, Oct. 26, 1993.

[18] L. Fredrickson, R. Karabed, J. W. Rae, P. H. Siegel, H. Thapar, and R. Wood, "Improved trellis coding for partial response channels," *IEEE Trans. Magn.*, vol. 31, pp. 1141–48, Mar. 1995.

[19] C. Freiman and A. D. Wyner, "Optimum block codes for noiseless input restricted channels," *Inform. Contr.*, vol. 7, pp. 398–415, 1964.

[20] K. J. Hole, "Punctured convolutional codes for the $1-D$ partial-response channel," *IEEE Trans. Inform. Theory*, vol. 37, pt. 2, pp. 808–817, May 1991.

[21] K. J. Hole and Ø. Ytrehus, "Improved coding techniques for partial-response channels," *IEEE Trans. Inform. Theory*, vol. 40, pp. 482–493, Mar. 1994.

[22] K. A. S. Immink, *Coding Techniques for Digital Recorders*. Englewood Cliffs, NJ: Prentice-Hall International (U.K.), 1991.

[23] H. Kamabe, "Minimum scope for sliding-block decoder mappings," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1335–1340, Nov. 1989.
[24] R. Karabed and B. H. Marcus, "Sliding-block coding for input-restricted channels," *IEEE Trans. Inform. Theory*, vol. 34, pp. 2–26, Jan. 1988.
[25] R. Karabed and P. H. Siegel, "Matched spectral-null codes for partial response channels," *IEEE Trans. Inform. Theory*, vol. 37, pt. II, pp. 818–855, May 1991.
[26] ——, "Coding for higher order partial response channels," in *Proc. 1995 SPIE Int. Symp. Voice, Video, and Data Communications* (Philadelphia, PA, Oct. 1995), vol. 2605, pp. 115–126.
[27] Z. Keirn and S. McCarthy,"Experimental performance comparison of FDTS/DF detectors: $8/9(0,k)$ versus $4/5$ MTR codes,"in *Proc. 1997 IEEE Int. Magnetics Conf.* (New Orleans, LA, Apr. 1–4); also, *IEEE Trans. Magn.*, vol. 33, pt. 1, pp. 2731–2733, Sept. 1997.
[28] K. J. Knudson, J. K. Wolf, and L. B. Milstein, "A concatenated decoding scheme for $(1-D)$ partial response with matched spectral-null coding," in *Proc. 1993 IEEE Global Telecommun. Conf. (GLOBECOM '93)* (Houston, TX, Nov. 1993), pp. 1960–1964.
[29] J. Lee and V. Madisetti,"A rate 8/10 (0,6) MTR code and its encoder/decoder," in *Proc. 1997 IEEE Int. Magnetics Conf.* (New Orleans, LA, Apr. 1–4); also, *IEEE Trans. Magn.*, vol. 33, pt. 1, pp. 2731–2733, Sept. 1997.
[30] S. Lin and D. J. Costello, Jr, *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
[31] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 1995.
[32] B. Marcus and R. Roth, "Bounds on the number of states in encoder graphs for input-constrained channels," *IEEE Trans. Inform. Theory*, vol. 37, pt. II, pp. 742–758, May 1991.
[33] B. Marcus, R. Roth, and P. Siegel, "Constrained systems and coding for recording channels," in *Handbook of Coding Theory*. Amsterdam, The Netherlands: Elsevier, 1998.
[34] B. H. Marcus, P. H. Siegel, and J. K. Wolf, "Finite-state modulation codes for data storage," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 5–37, Jan. 1992.
[35] S. McCarthy and Z. Keirn, "Distance enhancing code for E2PRML: Performance comparison using spinstand data," in *Proc. 1997 IEEE Int. Magnetics Conf.* (New Orleans, LA, Apr. 1–4); also, *IEEE Trans. Magn.*, vol. 33, pt. 1, pp. 2734–2736, Sept. 1997.
[36] P. A. McEwen and J. K. Wolf, "Trellis codes for $(1,k)$ E^{2}PR4ML with squared distance 18," in *Proc. 1996 IEEE Int. Magnetic Conf.* (Seattle, WA, April 1996); also, *IEEE Trans. Magn.*, vol. 32, pt. 1, pp. 3995–3997, Sept. 1996.
[37] T. Mittelholzer, P. A. McEwen, S. A. Altekar, and J. K. Wolf, "Finite truncation depth trellis codes for the dicode channel," in *Proc. 1996 IEEE Int. Magnetics Conf.* (San Antonio, TX, Apr. 18–21); also, *IEEE Trans. Magn.*, vol. 31, pt. 1, pp. 3027–3029, Nov. 1995.
[38] B. E. Moision, A. Orlitsky, and P. H. Siegel, "Bounds on the rate of codes that forbid specified difference sequences," submitted to the 1999 IEEE Global Telecommunications Conference (GLOBECOM'99).
[39] B. E. Moision, P. H. Siegel, and E. Soljanin, "Distance-enhancing codes for digital recording," in *Proc. 1997 IEEE Magnetic Recording Conf. (TMRC'97)*; also, *IEEE Trans. Magn.*, vol. 34, pt. 1, pp. 69–74, Jan. 1998.
[40] ——, "Error event characterization and coding for the equalized Lorentzian channel," in *Proc. 1998 IEEE Int. Symp. Information Theory* (Cambridge, MA, Aug. 16–21), p. 77.
[41] J. Moon, "Maximum transition run codes for data storage systems," in *Proc. 1996 IEEE Int. Magnetics Conf.* (Seattle, WA, Apr. 9–12); also, *IEEE Trans. Magn.*, vol. 32, pt. I, pp. 3992–3994, Sept. 1996.
[42] J. Moon and B. Brickner, "Maximum transition run codes for data storage systems," *IEEE Trans. Magn.*, vol. 32, pt. I, pp. 3992–3994, Sept. 1996.
[43] ——, "Design of a rate 5/6 maximum transition run code," in *Proc. 1997 IEEE Int. Magnetics Conf.* (New Orleans, LA, Apr. 1–4); also, *IEEE Trans. Magn.*, vol. 33, pp. 2749–2751, Sept. 1997.
[44] A. M. Patel, "Improved encoder and decoder for a byte-oriented rate 8/9 (0,3) code," *IBM Tech. Discl. Bull.*, vol. 28, p. 1938, 1985.
[45] J. G. Proakis, *Digital Communications*. New York, NY: McGraw-Hill, 1983.
[46] J. Rae, G. Christiansen, S.-M. Shih, H. Thapar, R. Karabed, and P. Siegel, "Design and performance of a VLSI 120 Mb/s trellis-coded partial-response channel," *IEEE Trans. Magn.*, vol. 31, pp. 1208–1214, Mar. 1995,
[47] N. Sayiner, "Impact of the track density versus linear density trade-off on the read channel: TCPR4 versus EPR4," in *Proc. 1995 SPIE Int. Symp. Voice, Video, and Data Communications* (Philadelphia, PA, Oct. 1995), vol. 2605, pp. 84–91.
[48] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 10, pp. 379–423, 623–656, Oct. 1948.
[49] P. H. Siegel "Coded modulation for binary partial response channels: state-of-the-art," in *Proc. 1996 Information Theory Workshop* (Haifa, Israel, June 9–13).
[50] N. T. Sindhushayana, "Symbolic dynamics, automata theory and the theory of coding: A comparative study and application," M.S. thesis, Cornell Univ. Ithaca, NY, Feb. 1992.
[51] E. Soljanin, "On-track and off-track distance properties of class 4 partial response channels," in *Proc. 1995 SPIE Int. Symp. Voice, Video, and Data Communications* (Philadelphia, PA, Oct. 1995), vol. 2605, pp. 92–102.
[52] ——, "On coding for binary partial-response channels that don't achieve the matched-filter-bound," in *Proc. 1996 Information Theory Workshop* (Haifa, Israel, June 9–13, 1996).
[53] ——, "Method and apparatus for generating dc-free sequences," U.S. Patent 5 608 397, Mar. 4, 1997.
[54] ——, "A coding scheme for generating dc-free sequences and their decoding on the $1-D$ channel," in *Proc. 1997 IEEE Int. Magnetics Conf.* (New Orleans, LA, Apr. 1–4); also, *IEEE Trans. Magn.*, vol. 33, pt. I, pp. 2755–2757, Sept. 1997.
[55] E. Soljanin and O. E. Agazzi "An interleaved coding scheme for $(1-D)(1+D)^2$ partial response with concatenated decoding" in *Proc. 1996 IEEE Global Telecommun. Conf. (GLOBECOM'96* (London, U.K., Nov. 18–22), vol. 1.3, pp. 353–357, 1996.
[56] J. Sonntag *et al.*, "A high speed, low power PRML read channel device," *IEEE Trans. Magn.*, vol. 31, pp. 1186–95, Mar. 1995.
[57] G. Strang, *Linear Algebra and its Applications*. Orlando, FL: Harcourt Brace Jovanovich, 1988.
[58] H. Thapar, private communication.
[59] H. Thapar and A. Patel, "A class of partial-response systems for increasing storage density in magnetic recording," *IEEE Trans. Magn.*, vol. 23, pp. 3666–3668, Sept. 1987.
[60] G. Vannucci and G. Foschini, "The minimum distance for digital magnetic recording partial responses," *IEEE Trans. Inform. Theory*, vol. 37, pt. II, pp. 955–960, May 1991.
[61] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.
[62] A. D. Weathers, S. A. Altekar, and J. K. Wolf, "Distance spectra for PRML channels," in *Proc. 1997 IEEE Int. Magnetics Conf.* (New Orleans, LA, Apr. 1–4); also, *IEEE Trans. Magn.*, vol. 33, pt. I, pp. 2809–2811, Sept. 1997.
[63] D. Welland *et al.*, "Implementation of a digital read/write channel with EEPR4 detection," *IEEE Trans. Magn.*, vol. 31, pp. 1180–85, Mar. 1995.
[64] J. P. Winkler, private communication, July 1996.
[65] J. K. Wolf, private communication, July 1996.
[66] J. K. Wolf and G. Ungerboeck, "Trellis coding for partial-response channels," *IEEE Trans. Commun.*, vol. 34, no. 8, pp. 765–773, Aug. 1986.
[67] R. Wood, "New detector for $1,k$ codes equalized to class II partial response," *IEEE Trans. Magn.*, vol. 25, pp. 4075–4077, Sept. 1989.
[68] ——, "Turbo PRML: A compromise EPRML detector," in *Proc. 1993 IEEE Int. Magnetics Conf.* (Stockholm, Sweden, Apr. 13–16); also, *IEEE Trans. Magn.*, vol. 29, pt. II, pp. 4018–4020, Nov. 1993.
[69] ——, private communication.
[70] A. H. Young, "Implementation issues of 8/9 distance-enhancing constrained codes for EEPR4 channel," M.S. thesis, Univ. Calif,, San Diego, June 1997.