

Channels with Intermittent Errors

Arya Mazumdar*

Alexander Barg*[‡]

Abstract—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. We study coding for binary channels in which out of any two consecutive transmitted bits at most one can be affected by errors. We consider a set of basic coding problems for such channels, deriving estimates on the size of optimal codes and also providing some constructions. We further consider a probabilistic model of noise with nonadjacent errors, as well as a generalization to errors separated by at least $s = 2, 3, \dots$ error-free channel uses.

Index Terms—Non-adjacent errors, bounds on codes, list decoding, channel capacity, linear codes

I. INTRODUCTION

Constrained systems, in particular, channels with data-dependent noise and channels with memory play an important role in the analysis of performance of magnetic recording devices and other storage systems [6]. Recent works [4], [7] considered a model of errors for a high-density magnetic recording channel, in which the action of errors on the recorded data depends on the contents of the memory cells. Under this model, errors can occur in a cell (bit) only if its contents is different from the contents of the previous cell. Another feature of this model, imposed by the nature of the read/write process in memory, is that errors never occur in adjacent cells. In this paper we assume the last property as a definition of the error process and analyze the problem of coding against errors in the combinatorial and probabilistic context.

II. CODES CORRECTING NON-ADJACENT ERRORS

The following general definition characterizes codes correcting a given set of errors.

Definition 2.1: A code $\mathcal{C} \subset \{0, 1\}^n$ is said to correct errors from a set $E \subset \{0, 1\}^n$ if for all $\mathbf{x} \neq \mathbf{x}' \in \mathcal{C}$ and for any $\mathbf{e}, \mathbf{e}' \in E$,

$$\mathbf{x} + \mathbf{e} \neq \mathbf{x}' + \mathbf{e}',$$

where the addition is modulo 2. Vectors from the set E are called *correctable errors* for \mathcal{C} .

A vector $\mathbf{e} = (e_1, \dots, e_n) \in \{0, 1\}^n$ will be called a *non-adjacent error vector* if for all $1 \leq i < n$, $e_i = 1$ implies $e_{i+1} = 0$. As usual, the weight (multiplicity) of error $w_H(\mathbf{e})$ is equal to the number of ones in \mathbf{e} .

Let $E_{n,t} = \{\mathbf{e} \in \{0, 1\}^n : w_H(\mathbf{e}) \leq t\}$ and let $\mathcal{E}_{n,t} \subset E_{n,t}$ be the set of all non-adjacent error vectors of weight less than or equal to t . We say that a code \mathcal{C} is *t-non-adjacent error-correcting* if in the above definition $E = \mathcal{E}_{n,t}$. In the standard scenario of coding theory, a t -error-correcting code corresponds to choosing $E = E_{n,t}$, i.e., all the errors of multiplicity not exceeding some $t > 0$.

We begin with computing the cardinality of the set $\mathcal{E}_{n,t}$. The proof of the next proposition is obtained by a simple counting argument, omitted here.

Proposition 2.2:

$$|\mathcal{E}_{n,t}| = \sum_{i=0}^t \binom{n-i+1}{i}. \quad (1)$$

Let $h(z) = -z \log_2 z - (1-z) \log_2(1-z)$ be the binary entropy function. Observe that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{E}_{n,\tau n}| = (1-\tau) h\left(\frac{\tau}{1-\tau}\right), \quad (2)$$

for $\tau \leq \frac{1}{2} - \frac{\sqrt{5}}{10} \approx 0.2764$. This asymptotic formula follows from the observation that $\binom{n-i+1}{i}$ increases on i for $i \leq \frac{1}{10}(5n+3-\sqrt{5n^2+10n+9})$. Thus, as long as $t/n \leq \frac{1}{2} - \frac{\sqrt{5}}{10}$, the asymptotics of the summation $\sum_{i=0}^t \binom{n-i+1}{i}$ is determined by the term $\binom{n-t+1}{t}$, and (1) follows by standard estimates of binomial coefficients (for example, see [5, p. 310]).

A. Bounds on non-adjacent error correcting codes. Clearly a t -error-correcting code is t -non-adjacent error-correcting. Somewhat surprisingly, the converse is also true.

Proposition 2.3: A code \mathcal{C} is t -non-adjacent error-correcting if and only if it is t -error-correcting.

Proof: Observe that $\mathcal{E}_{n,t} + \mathcal{E}_{n,t} = E_{n,2t}$, or in other words, any vector of weight $\leq 2t$ can be written as a sum of two non-adjacent errors of weight $\leq t$. At the same time, also $E_{n,t} + E_{n,t} = E_{n,2t}$.

Suppose that a code \mathcal{C} is t -non-adjacent error-correcting, but there exist uncorrectable errors $\mathbf{e}, \mathbf{e}' \in E_{n,t}$, i.e., for some $\mathbf{x} \neq \mathbf{x}' \in \mathcal{C}$, we have $\mathbf{x} + \mathbf{e} = \mathbf{x}' + \mathbf{e}'$. This implies that $\mathbf{x} = \mathbf{x}' + (\mathbf{e}' + \mathbf{e}) = \mathbf{x}' + \mathbf{z}$, where $\mathbf{z} \in E_{n,2t}$. Then write $\mathbf{z} = \mathbf{e}'_1 + \mathbf{e}_1$, where $\mathbf{e}'_1, \mathbf{e}_1 \in \mathcal{E}_{n,t}$. This implies the equality $\mathbf{x} + \mathbf{e}_1 = \mathbf{x}' + \mathbf{e}'_1$, which is a contradiction. ■

It would seem that the problem of correcting t non-adjacent errors is equivalent to correcting arbitrary t errors. The following theorem proves that this is not the case.

Theorem 2.4: There exists a linear code \mathcal{C} of rate

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{C}| \geq \frac{1}{2}(1 - h(2\tau)).$$

This work was supported by NSF grants CCF0916919, CCF0830699, and DMS0807411.

*Department of ECE and Institute for Systems Research, University of Maryland, College Park, 20742, USA. E-mails: {arya,abarg}@umd.edu.

[‡]Dobrushin Mathematical Laboratory, Institute for Problems of Information Transmission, Russian Academy of Science, Moscow, Russia.

that can correct any $t = \tau n$ non-adjacent errors with a polynomial time decoding algorithm.

Proof: We will only consider the case of n even (the case of odd n is established in a similar way). Let \mathcal{C}' be a linear $[n/2, k, t+1]$ code¹. Construct a code \mathcal{C} from \mathcal{C}' by repeating each coordinate in \mathcal{C}' twice: namely, for each codeword $\mathbf{x}' = (x'_1, \dots, x'_{n/2}) \in \mathcal{C}'$, form the vector $\mathbf{x} \in (x_1, \dots, x_n) \in \mathcal{C}$ by putting $x_{2i} = x_{2i-1} = x'_i$ for $1 \leq i \leq n/2$. Clearly, the code \mathcal{C} is linear and $|\mathcal{C}| = |\mathcal{C}'|$.

The code \mathcal{C} is decoded by the following two-step procedure. Suppose that a vector $\mathbf{z} = (z_1, \dots, z_n) \in \{0, 1\}^n$ is received from the channel. In the first step, we construct a vector $\mathbf{z}' = (z'_1, \dots, z'_{n/2}) \in \{0, 1, \epsilon\}^{n/2}$ from \mathbf{z} ; here ϵ is the erasure symbol. For $1 \leq i \leq n/2$, we set $z'_i = z_{2i}$ if $z_{2i} = z_{2i-1}$, or otherwise we set $z'_i = \epsilon$.

If there are at most t non-adjacent errors then \mathbf{z}' will contain at most t erasure symbols. The code \mathcal{C}' will then be used to correct them which is possible by a polynomial-time procedure.

To estimate the cardinality of the code \mathcal{C} , take a linear code \mathcal{C}' of distance $t+1$ that attains the Gilbert-Varshamov lower bound. Then

$$|\mathcal{C}'| \geq \frac{2^{n/2}}{\sum_{i=0}^t \binom{n/2-1}{i}}.$$

The estimate of the theorem is obtained by using a standard asymptotic estimate for the sum of binomial coefficients. ■

This theorem justifies the claim made before its statement: for $\tau = 1/4(1 - \theta)$, where θ is small, the estimate of rate established in it behaves as $\theta^2(\log_2 e)/4$ while the rate of the best known polynomial-time decodable τn -error-correcting codes approaches 0 in proportion to θ^3 .

B. List decoding. The gap between error correcting codes and non-adjacent error correcting codes becomes even more pronounced if we consider decoding into a list. Formally, a code \mathcal{C} is *list-of- L t -non-adjacent-error-correcting* ((L, t) -NAECC) if for any vector $\mathbf{x} \in \{0, 1\}^n$,

$$|\{\mathbf{c} \in \mathcal{C} : \mathbf{x} + \mathbf{c} \in \mathcal{E}_{n,t}\}| \leq L.$$

In words, for any received vector $\mathbf{x} \in \{0, 1\}^n$, there are at most L codewords that could be transformed to \mathbf{x} by the action of an error $\mathbf{e} \in \mathcal{E}_{n,t}$.

As usual, let $M(n, t; L)$ be maximum size of an (L, t) -NAECC of length n . For $0 \leq \tau \leq 1/2$ define

$$\underline{R}(\tau; L) = \liminf_{n \rightarrow \infty} (1/n) \log_2 M(n, \lfloor n\tau \rfloor; L)$$

and

$$\overline{R}(\tau; L) = \limsup_{n \rightarrow \infty} (1/n) \log_2 M(n, \lfloor n\tau \rfloor; L).$$

The following theorem contains elementary upper and lower bounds on the size of list codes.

¹Throughout the paper we write use the notation $\mathcal{C}[n, k, d]$ to refer to a linear binary code of length n , dimension k and minimum Hamming distance d .

Theorem 2.5:

$$\frac{2^{\frac{nL}{L+1}}}{\sum_{i=0}^t \binom{n-i+1}{i}} \leq M(n, t; L) \leq \frac{L2^n}{\sum_{i=0}^t \binom{n-i+1}{i}}.$$

Proof: We begin with the lower bound. Let us construct the code by choosing M codewords randomly and uniformly with replacement from $\{0, 1\}^n$. For a fixed vector $\mathbf{y} \in \{0, 1\}^n$, call the choice of any $L+1$ codewords $\mathbf{c}_1, \dots, \mathbf{c}_{L+1}$ ‘bad’ if $\mathbf{c}_1, \dots, \mathbf{c}_{L+1} \in \{\mathbf{y} + \mathbf{e} : \mathbf{e} \in \mathcal{E}_{n,t}\}$. Clearly, the expected number of bad choices for a random code \mathcal{C} is less than or equal to

$$2^n \binom{M}{L+1} \left(\frac{|\mathcal{E}_{n,t}|}{2^n} \right)^{L+1} < \left(M \sum_{i=0}^t \binom{n-i+1}{i} \right)^{L+1} 2^{-nL},$$

where we have used the equality of (1). Take $M = 2^{nL/(L+1)} / \sum_{i=0}^t \binom{n-i+1}{i}$, then the ensemble-average number of bad $(L+1)$ -tuples is less than 1. Therefore there exists a code of size M in which all the $(L+1)$ -tuples of codewords are good. This implies the lower bound on $M(n, t; L)$.

For the upper bound on $M(n, t; L)$ we again use the probabilistic method. Let \mathcal{C} be an (L, t) -NAECC and let \mathbf{y} be a vector randomly and uniformly chosen from $\{0, 1\}^n$. For a fixed codeword $\mathbf{c} \in \mathcal{C}$, the probability,

$$\Pr(\mathbf{c} + \mathbf{y} \in \mathcal{E}_{n,t}) = \frac{|\mathcal{E}_{n,t}|}{2^n}.$$

Therefore,

$$E(|\{\mathbf{c} \in \mathcal{C} : \mathbf{c} + \mathbf{y} \in \mathcal{E}_{n,t}\}|) = \frac{|\mathcal{C}| |\mathcal{E}_{n,t}|}{2^n}.$$

This implies that there exists at least one vector $\mathbf{y} \in \{0, 1\}^n$ such that

$$L \geq |\{\mathbf{c} \in \mathcal{C} : \mathbf{c} + \mathbf{y} \in \mathcal{E}_{n,t}\}| > \frac{|\mathcal{C}| |\mathcal{E}_{n,t}|}{2^n}.$$

This proves the upper bound on $M(n, t; L)$. ■

The above theorem along with (2) implies that whenever $\tau \leq \frac{1}{2} - \frac{\sqrt{5}}{10}$,

$$\begin{aligned} 1 - (1 - \tau) h\left(\frac{\tau}{1 - \tau}\right) + \frac{1}{L+1} &\leq \underline{R}(\tau; L) \leq \overline{R}(\tau; L) \quad (3) \\ &\leq 1 - (1 - \tau) h\left(\frac{\tau}{1 - \tau}\right) + \lim_{n \rightarrow \infty} \frac{\log L}{n}. \end{aligned}$$

An equivalent results for list decoding of ordinary binary code is well known (it is present in some form in [2]): the estimates (3) are valid for list-of L t -error-correcting codes if $(1 - \tau) h\left(\frac{\tau}{1 - \tau}\right)$ is replaced by $h(\tau)$. Concavity implies that $(1 - \tau) h\left(\frac{\tau}{1 - \tau}\right) < h(\tau)$, so we have proved that there exist codes of higher rates for list decoding in the case of non-adjacent errors than in the case of usual errors of the same multiplicity. Thus, non-adjacent errors are less adversarial than unrestricted errors (as should be expected), so codes of higher rates are possible. Later in the paper we will see that this claim also holds true for probabilistic error correction.

III. BIGGER GAPS: INTERMITTENT ERRORS

In this section we consider the case when the gap between two errors is at least $s = 2, 3, \dots$ bits. Let $\mathcal{E}_{n,t}^s$ be the set of all binary vectors of length n and weight t , where between any two ones, there are at least s zeros. Clearly, $n > (s+1)(t-1)$. Codes correcting t intermittent errors satisfy Definition 2.1 with $E = \mathcal{E}_{n,t}^s$ and will be called (s, t) *intermittent-error correcting*.

Proposition 3.1:

$$|\mathcal{E}_{n,t}^s| = \sum_{i=0}^t \binom{n - (i-1)s}{i}. \quad (4)$$

Proof: Let us count the number of binary vectors of weight i which any two ones are separated by at least s zeros ((s, ∞) patterns in the language of constrained systems). Trying to place i ones in n cells so that every two are separated by $\geq s$ empty cells leaves $n - (i-1)s$ cells for the ones themselves. Any placement of ones in these cells gives rise to a valid vector, which implies the claimed count. ■

As shown earlier, $(1, t)$ -intermittent error correcting codes is equivalent to t -error-correcting code. As s increases from one, we expect to be able to construct codes of much higher rates for intermittent errors than for usual errors. Let $M_s(n, t)$ be the size of the largest possible (s, t) -intermittent error correcting code of length n .

Theorem 3.2:

$$\frac{2^n}{\sum_{l=0}^2 \binom{n}{l} + \sum_{l=3}^{2t} \binom{s+1}{2}^l \binom{n-(l-3)(s-1)}{l}} \leq M_s(n, t) \leq \frac{2^n}{\sum_{l=0}^t \binom{n-(l-1)s}{l}}.$$

Proof: The upper bound follows by an elementary “sphere-packing” argument, while the lower bound is similar to the Gilbert bound on error-correcting codes. Namely, suppose that $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$ are two codewords that can be confused by the decoder. Then there exist $\mathbf{e}_1, \mathbf{e}_2 \in \mathcal{E}_{n,t}^s$ such that $\mathbf{c}_1 = \mathbf{c}_2 + \mathbf{e}$, where $\mathbf{e} = \mathbf{e}_1 + \mathbf{e}_2$. Let

$$\mathcal{D} = \mathcal{E}_{n,t}^s + \mathcal{E}_{n,t}^s = \{\mathbf{e}_1 + \mathbf{e}_2 : \mathbf{e}_1, \mathbf{e}_2 \in \mathcal{E}_{n,t}^s\},$$

then we obtain that $M_s(n, t) \geq 2^n/|\mathcal{D}|$. It remains to show that

$$|\mathcal{D}| \leq \sum_{l=0}^2 \binom{n}{l} + \sum_{l=3}^{2t} \binom{s+1}{2}^l \binom{n-(l-3)(s-1)}{l}.$$

Any vector $\mathbf{e} \in \mathcal{D}$ must satisfy the following two properties:

- 1) $0 \leq w_H(\mathbf{e}) \leq 2t$,
- 2) If the positions of the ones in \mathbf{e} are j_1, \dots, j_l , $l = w_H(\mathbf{e}) \geq 3$, then for all $1 \leq i \leq l-3$, the vector $(e_{j_i}, e_{j_{i+1}}, \dots, e_{j_{i+3}})$ must contain at least $s-1$ zeros (to verify this, consider that $\mathbf{e} = \mathbf{e}_1 + \mathbf{e}_2$, and that the smallest number of zeros is obtained when the ones in \mathbf{e}_1 and \mathbf{e}_2 are in disjoint locations).

Let us count the number of vectors that satisfy property 2. Consider a binary vector of length $n - (l-3)(s-1)$ and

weight l . As in the proof of the previous Proposition, the count of such vectors is $\binom{n-(l-3)(s-1)}{l}$. Further, the remaining part of the vector \mathbf{e} (i.e., $(e_{j_i}, e_{j_{i+1}}, \dots, e_{j_{i+3}})$) has the form $(10 \dots 010 \dots 010 \dots 1)$, where the two ones in the middle are placed in an arbitrary way. This gives $\binom{s+1}{2}$ possibilities for any choice of $j_i, 1 \leq i \leq l-3$. Taking account of the fact that $0 \leq l \leq 2t$, we obtain the final estimate. ■

Observe that for $s = 1$ we recover the standard GV bound, as is to be expected. Note also that the calculations for list decoding of Thm. 2.5 can be generalized to $s > 2$ without difficulty.

A. Constructions

In this section we make brief remarks on constructing (s, t) -intermittent error-correcting codes.

Construction 1: Let $s = 2^m - 1$ for some integer m , and suppose that s divides n . (The construction can be extended without much effort for general s .)

Consider a Hamming code \mathcal{H} with parameters $[s, s-m, 3]$. Consider a direct concatenation of n/s copies of \mathcal{H} ,

$$\mathcal{C} = \{(\mathbf{c}_1 | \mathbf{c}_2 | \dots | \mathbf{c}_{n/s}) : \mathbf{c}_i \in \mathcal{H}, 1 \leq i \leq n/s\}.$$

The rate of the code \mathcal{C} is equal to

$$\frac{2^m - m - 1}{2^m - 1} = 1 - \frac{\log(s+1)}{s}.$$

The code \mathcal{C} is an (s, t) -intermittent error-correcting for any $t \geq 0$. This is because each Hamming sub-block of a codeword of \mathcal{C} has to correct at most one error.

In the next section we consider a probabilistic model of the channel with intermittent errors. The above construction establishes that the *zero-error capacity* for such a channel is at least $1 - \log(s+1)/s$.

Construction 2: The construction of Theorem 2.4 can be viewed as a concatenation of a binary code \mathcal{A} of length $n/2$ and a $[2, 1, 2]$ repetition code \mathcal{B} . Taking instead a Reed-Solomon code \mathcal{A} of length $q = 2^{s-1}$ and distance $t+1$ over \mathbb{F}_q and a $[s, s-1, 2]$ binary single parity-check code \mathcal{B} , we construct a concatenated code \mathcal{C} that corrects t intermittent errors. This is because every block of s symbols will include at most one error which will be detected by the code \mathcal{B} . The set of $\leq t$ erasures resulting from inner decoding will be corrected by the code \mathcal{A} . Here the rate of the inner code \mathcal{B} is $\frac{s-1}{s}$, and the rate of the outer code \mathcal{A} is $\frac{2^{s-1}-t}{2^{s-1}}$. The code \mathcal{C} will have length $n = s2^{s-1}$ and rate $R = \left(1 - \frac{1}{s}\right)\left(1 - \frac{t}{2^{s-1}}\right)$. It is possible to have more constructions using code concatenation techniques.

In conclusion, note that both constructions presented here have polynomial-time encoding and decoding procedures.

IV. CHANNEL CAPACITY

Thus far in this paper, we have considered a combinatorial model of the non-adjacent errors. We will now switch to a parallel track by defining a natural probabilistic model of a

channel corresponding to the intermittent error channel defined above. This is a binary-output channel that can make an error only at positions that are at least $s + 1$ bits away. Our goal is to estimate the Shannon-theoretic capacity for this channel model. Let us proceed to formal definitions.

We consider a binary-input channel similar to the binary symmetric channel, except that errors must be separated. Formally, this is a channel with a binary input sequence $\mathbf{x} = x_1, x_2, x_3, \dots$, with $x_i \in \{0, 1\}$ for all i , and a binary output sequence $\mathbf{y} = y_1, y_2, y_3, \dots$, with $y_i \in \{0, 1\}$ for all i . The input-output relationship is determined by a binary sequence $\mathbf{u} = u_1, u_2, u_3, \dots$, which is a Markov chain, independent of the input sequence \mathbf{x} , with transition probabilities $P(u_i | u_{i-1}, u_{i-2}, \dots, u_{i-s})$ defined as follows:

$$\begin{array}{c|cc} & u_i = 0 & u_i = 1 \\ \hline u_{i-l} = 0, \quad \forall 1 \leq l \leq s & 1 - p & p \\ \exists u_{i-l} = 1, \text{ for some } 1 \leq l \leq s & 1 & 0. \end{array} \quad (5)$$

For any i the output of the channel is connected to the input by the equation

$$y_i = x_i + u_i.$$

We call this channel the binary-input intermittent (BINInter(s, p)) channel.

The BINInter(s, p) is a *discrete finite state channel (DFSC)* [3]. A stationary *discrete finite-state channel (DFSC)* has an input sequence $\mathbf{x} = x_1, x_2, x_3, \dots$, an output sequence $\mathbf{y} = y_1, y_2, y_3, \dots$, and a state sequence $\sigma = \sigma_1, \sigma_2, \sigma_3, \dots$. Each x_n is a symbol from a finite input alphabet \mathcal{X} , each y_n is a symbol from a finite output alphabet \mathcal{Y} , and each state σ_n takes values in a finite set of states \mathcal{S} . To complete the description of the channel, an initial state σ_0 , also taking values in \mathcal{S} , must be specified. For a detailed discussion of DFSCs we refer the reader to [3].

For a DFSC, we define the *lower (or pessimistic) capacity* $\underline{C} = \lim_{n \rightarrow \infty} \underline{C}_n$, and *upper (or optimistic) capacity* $\overline{C} = \lim_{n \rightarrow \infty} \overline{C}_n$, where

$$\begin{aligned} \underline{C}_n &= n^{-1} \max_{Q^n(\mathbf{x}^n)} \min_{\sigma_0 \in \mathcal{S}} I(\mathbf{x}^n; \mathbf{y}^n | \sigma_0) \\ \overline{C}_n &= n^{-1} \max_{Q^n(\mathbf{x}^n)} \max_{\sigma_0 \in \mathcal{S}} I(\mathbf{x}^n; \mathbf{y}^n | \sigma_0). \end{aligned}$$

In the above expressions, $I(\mathbf{x}^n; \mathbf{y}^n | \sigma_0)$ is the mutual information between the length- n input $\mathbf{x}^n = (x_1, \dots, x_n)$ and the length- n output $\mathbf{y}^n = (y_1, \dots, y_n)$, given the value of the initial state σ_0 , and the maximum is taken over probability distributions $Q^n(\mathbf{x}^n)$ on the input \mathbf{x}^n . The limits in the above definitions of \underline{C} and \overline{C} are known to exist. Clearly, $\underline{C}_n \leq \overline{C}_n$ for all n , and thus, $\underline{C} \leq \overline{C}$. The capacities \underline{C} and \overline{C} have an operational meaning in the usual Shannon-theoretic sense — see Theorems 4.6.2 and 5.9.2 in [3].

The upper and lower capacities coincide for a large class of channels known as *indecomposable* channels. Theorem 4.6.3 of [3] gives an easy-to-check necessary and sufficient condition for a DFSC to be indecomposable: for some fixed n and each \mathbf{x}^n , there exists a choice for σ_n (which may depend on

\mathbf{x}^n) such that

$$\min_{\sigma_0} q(\sigma_n | \mathbf{x}^n, \sigma_0) > 0. \quad (6)$$

We make a few comments about DFSCs for which $\underline{C} = \overline{C}$ holds, denoting by C the common value of \underline{C} and \overline{C} . This C , which we call simply the *capacity* of the DFSC, can be expressed alternatively. If we assign a probability distribution to the initial state, so that σ_0 becomes a random variable, then $C = \lim_{n \rightarrow \infty} C_n$, where

$$C_n = \frac{1}{n} \max_{Q^n(\mathbf{x}^n)} I(\mathbf{x}^n; \mathbf{y}^n | \sigma_0). \quad (7)$$

Clearly, $\underline{C}_n \leq C_n \leq \overline{C}_n$ for all n , so that C , as defined above, is indeed the common value of \underline{C} and \overline{C} . Note that this is independent of the choice of the probability distribution on σ_0 .

It is easy to see that the BINInter channel is a DFSC, where the n th state σ_n is a number $i \in \{0, 1, 2, \dots, s\} \triangleq \mathcal{S}$. Suppose that $\ell \geq 0$ is the smallest number such that $u_{n-\ell} = 1$. If no such ℓ exists then set $\ell = \infty$. Whenever $\ell \leq s$, we set $\sigma_n = \ell$. For $\ell > s$, we set $\sigma_n = s$. For completeness, we introduce an initial state σ_0 that takes values in \mathcal{S} . When $n < s$, $\sigma_n = \min(n + \sigma_0, \ell, s)$. The main result of this section is the following.

Theorem 4.1: For the BINInter(s, p) channel with parameter $p \in [0, 1]$, we have $\underline{C} = \overline{C} = C(p) \triangleq 1 - \frac{h(p)}{1+sp}$. We begin with two lemmas.

Lemma 4.2: For the BINInter(s, p) channel with parameter $p = 1$, we have $\underline{C} = \overline{C} = C(1) = 1$.

Proof: Since $p = 1$, once the initial state σ_0 is fixed, the output \mathbf{y} of the BINInter channel becomes a deterministic function of the input \mathbf{x} (i.e. the \mathbf{u} sequence is fixed with probability 1). Therefore, for any fixed $a \in \mathcal{S}$, we have $H(\mathbf{y}^n | \mathbf{x}^n, \sigma_0 = a) = 0$, and hence, $I(\mathbf{x}^n; \mathbf{y}^n | \sigma_0 = a) = H(\mathbf{y}^n | \sigma_0 = a)$. If \mathbf{x}^n is a sequence of i.i.d. Bernoulli($1/2$) random variables, then $\min_{a \in \mathcal{S}} H(\mathbf{y}^n | \sigma_0 = a) = n$. It follows that $\underline{C}_n \geq 1$, so that $\underline{C} \geq 1$. On the other hand $\underline{C} \leq \overline{C} \leq 1$. ■

Next, we look into the case when $p < 1$.

Lemma 4.3: The BINInter(s, p) channel is indecomposable for $p < 1$.

Proof: We must check that the condition in (6) holds. Take $n = s$ and $\sigma_n = s$, then $\min_{\sigma_0} q(\sigma_n | \mathbf{x}^n, \sigma_0) = (1 - p)^s > 0$. ■

Now we proceed to prove Thm. 4.1.

Proof of Thm. 4.1: We assume that $p < 1$, so the channel is indecomposable and its capacity is defined by (7).

The state sequence $\sigma = \sigma_0, \sigma_1, \dots; \sigma_i \in \mathcal{S} = \{0, 1, \dots, s\}$ forms a first-order Markov chain whose (nonzero) transition probabilities are given by

$$\begin{aligned} \Pr(\sigma_n = i + 1 | \sigma_{n-1} = i) &= 1, \quad 0 \leq i \leq s - 1 \\ \Pr(\sigma_n = 0 | \sigma_{n-1} = s) &= p, \\ \Pr((\sigma_n = s | \sigma_{n-1} = s)) &= 1 - p \end{aligned}$$

All other transition probabilities are zero. The stationary distribution of this Markov chain is the following:

$$\Pr(\sigma_n = s) = \frac{1}{1 + sp}$$

$$\Pr(\sigma_n = i) = \frac{p}{1 + sp}, \quad 0 \leq i \leq s - 1.$$

We assume that the initial state σ_0 follows this distribution as well.

We have

$$\begin{aligned} I(\mathbf{x}^n; \mathbf{y}^n | \sigma_0) &= H(\mathbf{y}^n | \sigma_0) - H(\mathbf{y}^n | \mathbf{x}^n, \sigma_0) \\ &\stackrel{(a)}{=} H(\mathbf{y}^n | \sigma_0) - H(\mathbf{u}^n | \mathbf{x}^n, \sigma_0) \\ &\stackrel{(b)}{=} H(\mathbf{y}^n | \sigma_0) - H(\sigma^n | \sigma_0), \end{aligned}$$

where (a) is due to the fact that, given \mathbf{x}^n , the sequences \mathbf{y}^n and \mathbf{u}^n uniquely determine each other, and (b) follows because \mathbf{u}^n is independent of \mathbf{x}^n . Also given σ_0 , the sequences \mathbf{u} and σ completely determine each other. Further, since σ is a stationary first-order Markov process, we have

$$H(\sigma^n | \sigma_0) = \sum_{n=1}^n H(\sigma_n | \sigma_{n-1}) = nH(\sigma_1 | \sigma_0) = n \frac{h(p)}{1 + ps}.$$

Hence,

$$C_n = n^{-1} \max_{Q^n(\mathbf{x}^n)} H(\mathbf{y}^n | \sigma_0) - \frac{h(p)}{1 + ps}. \quad (8)$$

Clearly, $H(\mathbf{y}^n | u_0) \leq n$. However, if \mathbf{x}^n is a sequence of i.i.d. Bernoulli($1/2$) random variables, then $H(\mathbf{y}^n | \sigma_0) = n$. Therefore,

$$\lim_{n \rightarrow \infty} C_n = 1 - \frac{h(p)}{1 + ps}.$$

As our final result, we show that capacity of the BINInter(s, p) can be achieved by binary linear codes.

Theorem 4.4: Let $C = 1 - \frac{h(p)}{1 + ps}, \epsilon > 0$. There exists a sequence of binary linear codes of growing length n and rate $R = C - \epsilon$ for which the error probability of decoding on the BINInter(s, p) is less than

Proof: (outline) We will prove the theorem by constructing a sequence of linear codes for which the set of “typical errors” that occur in the channel is formed of vectors with distinct syndromes.

Let $R = C - \epsilon$ and let n be an integer (the code length). Consider the ensemble of linear codes defined by random parity-check matrices of dimensions $(1 - R)n \times n$ with Bernoulli($1/2$) independent entries. The rate of any code in the ensemble is at least R . The probability that two vectors $\mathbf{x} \neq \mathbf{y} \in \{0, 1\}^n$ have the same syndrome equals

$$\Pr(H\mathbf{x} \neq H\mathbf{y}) = \Pr(H(\mathbf{x} + \mathbf{y}) = 0) = \frac{1}{2^{n-Rn}}.$$

If such an event occurs, the errors \mathbf{x}, \mathbf{y} contribute to the decoding error event. Now let E be an (unspecified) set of error vectors. By the Lovász Local Lemma (see [1]), if

$$\frac{2e|E|}{2^{n-Rn}} \leq 1,$$

then there exists a matrix H with no two errors in the set E colliding. The code with this parity-check matrix will have low error probability of decoding if errors outside E have a small probability of occurring in the channel.

Now consider the set of vectors of length n (errors) generated by the random process described in (5). We claim that the set of typical vectors generated by this process has cardinality

$$|E| \leq \frac{1}{2e} 2^{n \left(\frac{h(p)}{1+ps} + \epsilon \right)} = \frac{1}{2e} 2^{n-Rn}$$

(estimating the cardinality is a standard but tedious calculation which will be omitted). Concluding, there exists a linear code of rate $R = C - \epsilon$ for which these errors will be decoded correctly, and thus the overall decoding error probability will be arbitrarily small. ■

Acknowledgment. The authors are grateful to Navin Kashyap and Gilles Zémor for useful discussions of this work.

REFERENCES

- [1] N. Alon and J. Spencer, *The Probabilistic Method*, J. Wiley & Sons, 2000.
- [2] P. Elias, “Error-correcting codes for list decoding,” *IEEE Trans. Information Theory*, vol. 37, no. 1, pp. 5–12, Jan. 1991.
- [3] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons, 1968.
- [4] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, “Write channel model for bit-patterned media recording,” in *Proc. 21st Magnetic Recording Conference (TMRC 2010)*, La Jolla, CA, USA, Aug. 16–18, 2010.
- [5] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam, 1977.
- [6] B. H. Marcus, R. M. Roth, and P.H. Siegel, “Constrained systems and coding for recording channels” *Handbook of coding theory*, Vol. II, pp. 1635–1764, North-Holland, Amsterdam, 1998.
- [7] A. Mazumdar, A. Barg, and N. Kashyap “Coding for high-density recording on a 1-D granular magnetic medium,” arXiv1012:1895.