On 2-D Non-Adjacent-Error Channel Models

Shivkumar K. M. and Navin Kashyap
Dept. of Electrical Communication Engineering
Indian Institute of Science, Bangalore, India
Email: {shiv,nkashyap}@ece.iisc.ernet.in

***Abstract*—In this work, we consider two-dimensional (2-D) binary channels in which the 2-D error patterns are constrained so that errors cannot occur in adjacent horizontal or vertical positions. We consider probabilistic and combinatorial models for such channels. A probabilistic model is obtained from a 2-D random field defined by Roth, Siegel and Wolf (2001). Based on the conjectured ergodicity of this random field, we obtain an expression for the capacity of the 2-D non-adjacent-errors channel. We also derive an upper bound for the asymptotic coding rate in the combinatorial model.**

I. Introduction

Recent advances in storage technology (e.g., [10], [11]) provide means of achieving very high storage densities by writing data ever more densely on the two-dimensional (2-D) surface of the storage medium. This has led to several recent studies of storage channel models in which the data to be stored constitute the input, the media irregularities and the physics of the read/write process account for the noise, and the final data that is retrieved constitute the channel output [3], [4], [7]. Realistic storage channel models are often difficult to handle analytically, as the noise in such channels is, in general, input-dependent and has 2-D correlations.

In this paper, we consider a 2-D additive-noise channel model in which the noise is independent of the input, but which exhibits strong 2-D correlations. Our model is a 2-D extension of the one-dimensional (1-D) channel model considered in a recent work of Mazumdar and Barg [6]. While we do not claim our model to be a realistic model for a storage channel, it serves to illustrate the difficulties involved in analytically handling 2-D correlations.

The channel model considered in [6] was one in which a binary noise vector $\mathbf{n}$ gets added (modulo-2) to a binary input vector $\mathbf{x}$, with $\mathbf{n}$ being independent of $\mathbf{x}$, and additionally having the property that 1s cannot be in adjacent positions of $\mathbf{n}$. Thus, errors cannot occur in adjacent positions; hence, the term *non-adjacent error vector*. They considered both probabilistic and combinatorial (adversarial) models for such a channel. In particular, they showed the surprising fact [6, Prop. 2.3] that, in the combinatorial model, a code can correct all non-adjacent error vectors of Hamming weight at most t iff the code is t-error-correcting in the usual (unconstrained) sense.

To describe the 2-D extension of the 1-D non-adjacent-error model, we need some definitions. For $F \subseteq \mathbb{Z}^2$, an F-*configuration* is a mapping $z : F \to \{0,1\}$, and $\{0,1\}^F$ denotes the set of all F-configurations. The value of $z \in \{0,1\}^F$ at position $(i,j) \in F$ will be denoted by $z_{i,j}$. An F-configuration z satisfies the (2-D) *hard square constraint* [9] if for all $(i,j),(i',j') \in F$ such that $|i-i'|+|j-j'|=1$, either $z_{i,j}=0$ or $z_{i',j'}=0$ (or both). The set of all F-configurations z satisfying the hard square constraint is denoted by $\mathsf{HS}(F)$. Note that if we define a 2-D additive-noise channel with inputs from $\{0,1\}^F$ and error patterns from $\mathsf{HS}(F)$, we get a channel model on F in which errors cannot occur in adjacent positions along the horizontal and vertical directions. A probability distribution on $\mathsf{HS}(F)$ yields a probabilistic channel model, while a combinatorial model is obtained by allowing only those error patterns from $\mathsf{HS}(F)$ that have at most t 1s.

The 2-D situation is significantly different from, and usually a lot harder to handle than, the 1-D set-up. For instance, defining a useful probability measure on $\mathsf{HS(F)}$ is not easy, even when F is a rectangular array. Roth, Siegel and Wolf [9] defined a probability measure $\mu_{m,n}$ on $\mathsf{HS}(\Delta_{m,n})$, where $\Delta_{m,n}$ is an $m \times n$ parallelogram of the form (Figure 1)

$$\Delta_{m,n} = \left\{(i,j) \in \mathbb{Z}^2 : 0 \le i < m,\, 0 \le i+j < n\right\}.$$

This measure can be extended to a measure μ on $\mathsf{HS}(\mathbb{Z}^2)$, which can then be restricted to $\mathsf{HS}(F)$, for any (measurable) $F \subset \mathbb{Z}^2$, to obtain a suitable measure on $\mathsf{HS}(F)$. Consider, for example, the case when F is an $n \times n$ square:

$$\square_n = \left\{(i,j) \in \mathbb{Z}^2 : 0 \le i < n,\, 0 \le j < n\right\}.$$

We will show that the resulting probabilistic channel model has a meaningful notion of channel capacity, if we assume that the measure μ is ergodic. Squares are used for illustrative purposes only — channels defined by any sequence of "well-behaved" subsets F_n (see conditions (F1) and (F2) in Section II-B) can be similarly analyzed.

The 1-D and 2-D cases are quite different within the combinatorial channel model as well. In contrast with the 1-D case [6, Prop. 2.3], a code correcting t hard-square errors need not be a t-error-correcting code. We demonstrate this by giving a simple example. A code $\mathcal{C} \subseteq \{0,1\}^F$ is *t-hard-square error-correcting* (resp. *t-error-correcting*) if for any two different $x_1, x_2 \in \mathcal{C}$, and for any two $e_1, e_2 \in \mathsf{HS}(F)$ (resp. $e_1, e_2 \in \{0,1\}^F$) with Hamming weights at most t, we have

$$x_1 \oplus e_1 \ne x_2 \oplus e_2,$$

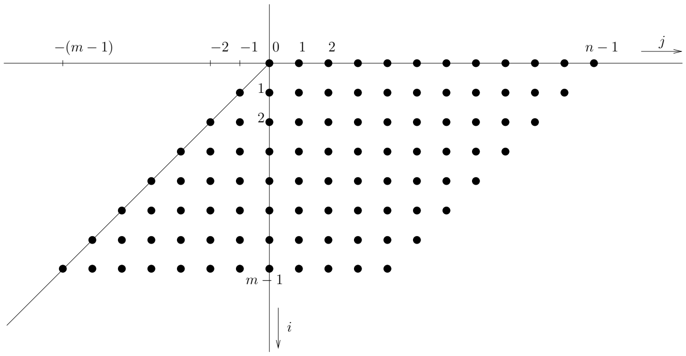

Fig. 1. Parallelogram $\Delta_{m,n}$

where $\oplus$ denotes coordinate-wise modulo-2 addition. Now for the example. Take $F = \{(i,j) : 0 \le i < 2,\ 0 \le j < 3\}$, a 2×3 rectangular array. Let

$$\mathcal{C} = \left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \right\}.$$

It can be easily verified that this code is 2-hard-square error-correcting. To see that it is not 2-error-correcting, take

$$e_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},\ e_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

The Roth-Siegel-Wolf (RSW) measure μ can be used to derive a useful result within the combinatorial channel model as well. Consider a sequence of 2-D channels defined on squares $\square_n$, and let $t_n = \tau n^2$. For any sequence of t_n-hard-square-correcting codes $\mathcal{C}_n \subseteq \{0,1\}^{\square_n}$ of rate R, i.e., with $|\mathcal{C}_n| = \lceil 2^{n^2 R} \rceil$, an upper bound on R can be derived using the RSW measure μ.

The rest of this paper is organized as follows. In Section II, we describe in detail the RSW measure and its properties, including its conjectured ergodicity. Section III presents applications of the RSW measure to the analysis of 2-D channels with hard-square errors.

II. The Roth-Siegel-Wolf (RSW) Measure

We begin by describing the RSW probability measure and some of its properties. The notation and terminology used in this section are largely borrowed from [9].

A. Measures on Finite Parallelograms

Recall, from Section I, the definitions of $\Delta_{m,n}$ and $\mathsf{HS}(\Delta_{m,n})$. *Row* i in $\Delta_{m,n}$ consists of all the locations (i,j) such that $-i \le j < n-i$. *Diagonal* d consists of all locations $(i, d-i)$ such that $0 \le i < m$. Row (diagonal) 0 is referred to as the *horizontal (diagonal) boundary.*

We proceed by defining a probability measure $\mu_{m,n}$ on $\mathsf{HS}(\Delta_{m,n})$. Let Z denote a random $\Delta_{m,n}$-configuration taking values from $\mathsf{HS}(\Delta_{m,n})$. Hereafter, we denote this as $Z \in_{\mu_{m,n}} \mathsf{HS}(\Delta_{m,n})$. Let $Z_{i,j}$ denote its value at location (i,j). For every $z \in \mathsf{HS}(\Delta_{m,n})$,

$$\begin{aligned} \mu_{m,n}(z) = {} & \Pr(Z = z) \\ = {} & \mu_0(z_{0,0}) \cdot \mu_h(z_{0,1}, z_{0,2}, \ldots, z_{0,n-1} | z_{0,0}) \\ & \cdot \mu_d(z_{1,-1}, z_{2,-2}, \ldots, z_{m-1,1-m} | z_{0,0}) \\ & \cdot \prod_{i=1}^{m-1} \prod_{j=-i+1}^{n-1-i} \vartheta\left(z_{i,j} | z_{i,j-1}, z_{i-1,j}, z_{i-1,j+1}\right). \end{aligned} \tag{1}$$

Each component will be defined below:

1) The measure μ_h on the horizontal boundary takes the form of a first-order Markov process:

$$\mu_h(w_1, \ldots, w_{n-1} | w_0) = \prod_{j=1}^{n-1} \Pr(w_j | w_{j-1}), \tag{2}$$

with transition probabilities given by

$$\Pr(w_j = 0 | w_{j-1} = c) = \begin{cases} \alpha & \text{if } c = 0 \\ 1 & \text{if } c = 1. \end{cases} \tag{3}$$

Let this Markov process be termed as the horizontal Markov process M_h.

2) The values of μ_0 are set to the stationary probabilities of M_h, i.e.,

$$\mu_0(0) = 1 - \mu_0(1) = \frac{1}{2 - \alpha}. \tag{4}$$

3) The measure μ_d on the diagonal boundary also takes the form of a first-order Markov process:

$$\mu_d(w_1, \ldots, w_{n-1} | w_0) = \prod_{j=1}^{m-1} \Pr(w_j | w_{j-1}), \tag{5}$$

with transition probabilities given by

$$\Pr(w_j = 0 | w_{j-1} = c) = \begin{cases} \beta_0 & \text{if } c = 0 \\ \beta_1 & \text{if } c = 1. \end{cases} \tag{6}$$

Here, β_0 and β_1 are assigned values in such a way that they are consistent with the stationary distribution along the horizontal boundary:

$$\beta_0 = \frac{\alpha}{\alpha + q_1 - \alpha q_1} \text{ and } \beta_1 = \frac{q_1}{\alpha + q_1 - \alpha q_1}, \tag{7}$$

where q_1 is as given in (8) below.

4) The fourth component of the expression in (1) is defined using two parameters $q_0 \in [0,1)$ and $q_1 \in (0,1]$:

$$\vartheta\left(0 | u, y, v\right) = \begin{cases} q_v, & \text{if } u = y = 0 \\ 1, & \text{otherwise.} \end{cases} \tag{8}$$

We next state two main results from [9] that will be of use to us. For this, we need some definitions: row i in $Z \in_{\mu_{m,n}} \mathsf{HS}(\Delta_{m,n})$ is said to form a first-order Markov chain identical to the horizontal boundary if for $1 - i \le$

$j < n-i$ and every non-empty word $\mathbf{c} = (c_1, c_2, \ldots, c_l)$ of length $l \le i+j$,

$$\Pr\big(Z_{i,j} = 0 \mid (Z_{i,j-1}, \ldots, Z_{i,j-l}) = \mathbf{c}\big) = \begin{cases} \alpha, & c_1 = 0 \\ 1, & c_1 = 1 \end{cases}$$

provided the event on which we condition has positive probability. Similarly, diagonal d in $Z \in_{\mu_{m,n}} \mathsf{HS}(\Delta_{m,n})$ is said to form a first-order Markov chain identical to the diagonal boundary if for $1 \le i \le m$ and every word $\mathbf{c} = (c_1, c_2, \ldots, c_l)$,

$$\Pr\big(Z_{i,d-i} = 0 \mid (Z_{i-1,d-(i-1)}, \ldots, Z_{i-l,d-(i-l)}) = \mathbf{c}\big) = \beta_{c_1}$$

provided the event on which we condition has positive probability.

Theorem 1 ([9], Proposition 2.1)**.** *For $m, n \ge 2$, $q_0 \in (0,1)$ and $q_1 \in (0,1]$, entries in each row (diagonal) form a first-order Markov chain identical to the horizontal (diagonal) boundary if and only if*

$$\alpha = \frac{-q_1 + \sqrt{q_1^2 + 4q_1(1-q_0)}}{2(1-q_0)}. \tag{9}$$

For a measure $\mu_{m,n}$, the entropy $H(\mu_{m,n})$ is defined as

$$H(\mu_{m,n}) = -\sum_{z \in \mathsf{HS}(\Delta_{m,n})} \mu_{m,n}(z) \log_2(\mu_{m,n}(z)).$$

For $\xi \in [0,1]$, $h(\xi) \triangleq -\xi \log_2 \xi - (1-\xi)\log_2(1-\xi)$. In [9] it is shown that, when (9) is satisfied,

$$H(\mu) \triangleq \lim_{m,n\to\infty} \frac{H(\mu_{m,n})}{mn} = \frac{\beta_0}{2-\alpha}\left(\alpha h(q_0) + (1-\alpha)h(q_1)\right). \tag{10}$$

For the rest of this paper we only consider probability measures $\mu_{m,n}$ satisfying the condition given in (9).

B. Extension to a Random Field on $\mathbb{Z}^2$

The measures $\mu_{m,n}$, for fixed q_0 and q_1, are defined on parallelograms $\Delta_{m,n}$. For the channel model that we have in mind, it is necessary to extend this to a measure on $\mathsf{HS}(\mathbb{Z}^2)$, the set of 0/1-configurations on $\mathbb{Z}^2$ that satisfy the hard square constraint. The extension is by a standard application of the Kolmogorov extension theorem [2, Chapter IV.6, Theorem 1], and we only sketch the details.

Define parallelograms Λ_n, $n \in \mathbb{Z}_+$, as follows: $\Lambda_n \triangleq \Delta_{2n+1,2n+1} + (-n, 0)$ is the $(2n+1)\times(2n+1)$ parallelogram obtained by translating $\Delta_{2n+1,2n+1}$ vertically by n coordinates (so that the top-leftmost point moves from $(0,0)$ to $(-n,0)$). The parallelograms Λ_n contain $(0,0)$ as their centre, and form an increasing sequence of nested sets, with $\bigcup_{n=1}^{\infty} \Lambda_n = \mathbb{Z}^2$. The measures $\mu_{2n+1,2n+1}$, originally defined on $\Delta_{2n+1,2n+1}$, may equivalently be defined on the translates Λ_n instead. Formally, $\tilde{\mu}_n \triangleq \mu_{2n+1,2n+1} \circ T_n^{-1}$, where $T_n(i,j) = (i,j) + (-n,0)$ for all $(i,j) \in \mathbb{Z}^2$, is the equivalent measure on Λ_n. Assuming, as we do, that the condition in (9) holds, it is an easy exercise to verify using (1) and Theorem 1 that the measures $\tilde{\mu}_n$, $n = 1, 2, \ldots$, are *consistent*: $\tilde{\mu}_n$, when restricted to Λ_{n-1}, coincides with $\tilde{\mu}_{n-1}$. Therefore, by the Kolmogorov extension theorem, as $n \to \infty$, the measures $\tilde{\mu}_n$ extend to a unique measure μ on the σ-algebra generated by cylinder sets in $\{0,1\}^{\mathbb{Z}^2}$. The measure μ defines a random field $\mathbf{Z} = (Z_{i,j})_{(i,j)\in\mathbb{Z}^2}$, where each $Z_{i,j}$ is a binary-valued random variable. The measure μ and the random field $\mathbf{Z}$ will be called the *RSW measure* and the *RSW random field*, respectively.

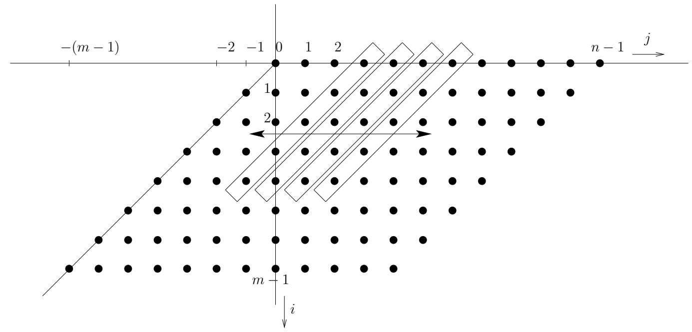

Fig. 2. A horizontal random process, $\mathbf{Z}^{\mathsf{h}}_{[0,4]}$, of diagonal width 5.

The RSW random field $\mathbf{Z}$ has many nice properties. By construction, $\mu(\mathbf{Z} \in \mathsf{HS}(\mathbb{Z}^2)) = 1$. The random variables $(Z_{i,j})_{(i,j)\in\Lambda_n}$ are jointly distributed according to $\tilde{\mu}_n$. It is again an easy exercise to verify using (1) and Theorem 1 that the random variables indexed by any translate of Λ_n are distributed according to $\tilde{\mu}_n$ as well. It follows that the random field $\mathbf{Z}$ is stationary.

From Theorem 1, it follows that for any fixed $i \in \mathbf{Z}$, the "horizontal" random process $\mathbf{Z}^{\mathsf{h}}_i = (Z_{i,j})_{j\in\mathbb{Z}}$ along row i is a stationary first-order Markov process. Similarly, the "diagonal" random process $\mathbf{Z}^{\mathsf{d}}_d = (Z_{i,d-i})_{i\in\mathbb{Z}}$ along diagonal d is a stationary first-order Markov process. In fact, more can be said. For $i, d \in \mathbb{Z}$, and an integer $\ell > 0$, define two vector-valued random processes: the horizontal random process of diagonal width ℓ (see Figure 2),

$$\mathbf{Z}^{\mathsf{h}}_{[i-(\ell-1),i]} = \big([Z_{i,j}, Z_{i-1,j+1}, \ldots, Z_{i-(\ell-1),j+(\ell-1)}]\big)_{j\in\mathbb{Z}},$$

and the diagonal random process of horizontal width ℓ,

$$\mathbf{Z}^{\mathsf{d}}_{[d,d+\ell-1]} = \big([Z_{i,d-i}, Z_{i,d+1-i}, \ldots, Z_{i,d+(\ell-1)-i}]\big)_{i\in\mathbb{Z}}.$$

It can be verified (using (1), (2) and (5)) that both these vector-valued random processes are stationary first-order Markov processes. In fact, they are ergodic.

Lemma 2. *For any $i, d \in \mathbb{Z}$ and $\ell > 0$, the Markov chains $\mathbf{Z}^{h}_{[i-(\ell-1),i]}$ and $\mathbf{Z}^{d}_{[d,d+\ell-1]}$ are ergodic.*

Proof: Let $\mathcal{S} = \{\mathbf{s} \in \{0,1\}^{\ell} : \mathbf{s} \text{ contains no adjacent 1s}\}$ denote the state space of these Markov chains. Clearly the all-zero vector, $\mathbf{0}$, is in $\mathcal{S}$. Since the hard square model places a constraint only on the placement of adjacent 1s, we have $\Pr(\mathbf{s}|\mathbf{0}) > 0$ and $\Pr(\mathbf{0}|\mathbf{s}) > 0$ for all $\mathbf{s} \in \mathcal{S}$. Thus, the Markov chains are irreducible and aperiodic, and hence, ergodic. ∎

The above lemma leads us to believe that the following conjecure is true.

Conjecture 1. *The stationary random field* $\mathbf{Z}$ *is ergodic.*

To be precise, the conjecture states that the (invariant) measure μ is ergodic with respect to the $\mathbb{Z}^2$ action generated by the commuting shifts along the horizontal and diagonal (or for that matter, vertical) directions. A rigorous proof of this eludes us at present.

But, if we assume the conjecture to be true, then we have a law of large numbers and an asymptotic equipartition property (AEP). That is to say, for any "well-behaved" sequence of subsets $F_n \subset \mathbb{Z}^2$, $n = 1, 2, \ldots$, the following hold (see e.g. [5, Theorems 1.3 and 1.4]):

$$\lim_{n\to\infty} \frac{1}{|F_n|} \sum_{(i,j)\in F_n} Z_{i,j} = \mu_0(1) \text{ a.s.} \tag{11}$$

where $\mu_0(1)$ is as given in (4), and for μ-a.e. $\mathbf{z} \in \{0,1\}^{\mathbb{Z}^2}$,

$$\lim_{n\to\infty} -\frac{1}{|F_n|} \log_2 \mu\big(Z_{i,j} = z_{i,j} \,\forall\, (i,j) \in F_n\big) = H(\mu), \tag{12}$$

where $H(\mu)$ is as defined in (10). We should clarify what it means for a sequence (F_n) to be "well-behaved". Theorems 1.3 and 1.4 of [5] show that it is sufficient for (F_n) to satisfy the following two conditions:

(F1) for all $(i,j) \in \mathbb{Z}^2$, the symmetric difference between F_n and its translate $(i,j) + F_n$ should be vanishingly small relative to the cardinality of F_n:

$$\lim_{n\to\infty} \frac{|F_n \,\Delta\, ((i,j) + F_n)|}{|F_n|} = 0;$$

(F2) for some constant $K > 0$ and all n, we must have

$$\left| \bigcup_{k\le n-1} (F_n - F_k) \right| \le K|F_n|,$$

where $F_n - F_k \triangleq \{a - b : a \in F_n, b \in F_k\}$.

The convergence in probability versions of (11) and (12) are: for all $\epsilon > 0$,

$$\Pr\left[\left|\frac{1}{|F_n|} \sum_{(i,j)\in F_n} Z_{i,j} - \mu_0(1)\right| \le \epsilon\right] \overset{n\to\infty}{\longrightarrow} 1 \tag{13}$$

and

$$\Pr\left[\left|-\frac{1}{|F_n|} \log_2 \mu(Z_{F_n}) - H(\mu)\right| \le \epsilon\right] \overset{n\to\infty}{\longrightarrow} 1, \tag{14}$$

where Z_{F_n} is the collection of rvs $(Z_{i,j})_{(i,j)\in F_n}$, and Pr is computed with respect to the measure μ.

It is easy to verify that the $n \times n$ parallelograms $\Delta_{n,n}$ and the $n \times n$ squares $\square_n$ each form a "well-behaved" sequence of subsets of $\mathbb{Z}^2$; observe that condition (F2) is satisfied with $K = 4$ for both sequences. Therefore, assuming Conjecture 1 to be true, (11)–(14) hold with $F_n = \square_n$. Not having a proof for Conjecture 1, we provide some experimental evidence in support of (13) and (14) when $F_n = \square_n$. Recall from Section II-A that the measures $\mu_{m,n}$ are parametrized by q_0 and q_1; therefore, so is the measure μ. For various choices of q_0 and q_1, and $n = 10, 30, 50, 100, 200, 300, 400, 500$, we considered 1000 $\square_n$-configurations z generated according to the probability measure $\pi_n \triangleq \mu|_{\square_n}$. Let $\gamma(n)$ denote the fraction of $\square_n$-configurations z such that $|\frac{1}{n^2}\sum_{i,j} z_{i,j} - \mu_0(1)| \le 0.005$; and let $\theta(n)$ denote the fraction of $\square_n$-configurations z that are "0.005-typical", i.e., $|-\frac{1}{n^2}\log_2 \pi_n(z) - H(\mu)| \le 0.005$. For most values of (q_0, q_1) considered, we found that $\gamma(n) \ge 0.95$ for $n \ge 200$, and $\theta(n) \ge 0.95$ for $n \ge 400$. In almost all cases, $\gamma(500)$ and $\theta(500)$ were very close to 1. This gives us reason to believe that (13) and (14) hold.

We make note of one last fact about the RSW random field $\mathbf{Z}$. Assuming Conjecture 1, the following holds for any "well-behaved" sequence of sets $F_n \subset \mathbb{Z}^2$:

$$\lim_{n\to\infty} \frac{1}{|F_n|} H(Z_{F_n}) = H(\mu), \tag{15}$$

where $H(Z_{F_n})$ refers to the joint entropy of the rvs $(Z_{i,j})_{(i,j)\in F_n}$, and $H(\mu)$ is as defined in (10); see Definition 4.1 and the subsequent discussion in [5].

III. 2-D Hard-Square-Error Channel Models

We will now use the properties of the RSW measure and field noted in the previous section to derive results pertinent to 2-D channel models with additive hard-square errors. Throughout our discussion, we will fix a "well-behaved" sequence of sets $F_n \subset \mathbb{Z}^2$. For concreteness, it may be useful to think of F_n as either $\Delta_{n,n}$ or $\square_n$.

A. Probabilistic Channel Model

Consider a channel $\mathcal{Q}^{(n)}$ defined on F_n, with input $X^{(n)} \in \{0,1\}^{F_n}$ and output $Y^{(n)} \in \{0,1\}^{F_n}$ related by

$$Y^{(n)} = X^{(n)} \oplus Z^{(n)} \tag{16}$$

where $Z^{(n)} \in \mathsf{HS}(F_n)$ is a random variable independent of $X^{(n)}$, distributed according to the RSW measure μ (restricted to F_n); and $\oplus$ denotes coordinate-wise modulo-2 addition. The information-theoretic capacity of the sequence of channels $\mathcal{Q}^{(n)}$, $n = 1, 2, \ldots$, is defined as

$$C = \lim_{n\to\infty} \max_P \frac{1}{|F_n|} I(X^{(n)}; Y^{(n)}), \tag{17}$$

the maximum being taken over probability distributions P on the input $X^{(n)}$. From the fact that $X^{(n)}$ and $Z^{(n)}$ are independent, we obtain that $I(X^{(n)}; Y^{(n)}) = H(Y^{(n)}) - H(Z^{(n)}) \le |F_n| - H(Z^{(n)})$, with equality achieved iff $X^{(n)}$ (and hence, $Y^{(n)}$) is uniformly distributed. Therefore, via (15), we obtain that $C = 1 - H(\mu)$.

A code $\mathcal{C}^{(n)}$ for $\mathcal{Q}^{(n)}$ is a subset of $\{0,1\}^{F_n}$. Its rate is given by $\frac{1}{|F_n|}\log_2 |\mathcal{C}^{(n)}|$. A rate R is said to be *achievable* over the channels $(\mathcal{Q}^{(n)})_{n\in\mathbb{Z}_+}$ if for $n = 1, 2, \ldots$, there exists a code $\mathcal{C}^{(n)}$ with rate $R^{(n)}$ and probability of error[1]

[1] The probability of error is defined with respect to a suitable choice of decoder $g^{(n)} : \{0,1\}^{F_n} \to \mathcal{C}^{(n)}$; see e.g., [1, Chapter 8].

$P_e^{(n)}$ over the channel $\mathcal{Q}^{(n)}$ such that $R^{(n)} \to R$ and $P_e^{(n)} \to 0$, as $n \to \infty$. We then have the following theorem.

Theorem 3. *Assuming Conjecture 1 holds, all rates $R < 1 - H(\mu)$ are achievable. Conversely, if R is an achievable rate, then $R \leq 1 - H(\mu)$.*

The proof of the theorem follows standard information-theoretic arguments. The achievability part is proved by a random coding argument using typical set decoding, for which the AEP in the form of (14) is needed. The converse is an easy application of Fano's inequality, requiring (15).

B. Combinatorial Channel Model

The notion of a t-hard-square error-correcting (t-*hsec*) code $\mathcal{C} \subseteq \{0,1\}^F$, for $F \subset \mathbb{Z}^2$, was introduced in Section I. We consider t-hsec codes for the channels defined by the "well-behaved" sequence of sets F_n. In particular, we are interested in the maximum rate of codes whose error-correction capability, t, is a constant fraction of the blocklength of the code, i.e., $t = \tau|F_n|$ for some fixed $\tau > 0$. Note that when error patterns satisfy the hard square constraint, τ cannot exceed 1/2.

For a fixed $\tau \in [0, 1/2]$, and $n = 1, 2, \ldots$, let $M_n(\tau)$ denote the maximum size of a $(\tau|F_n|)$-hsec code $\mathcal{C} \subseteq \{0,1\}^{F_n}$. Define $R(\tau) = \limsup_{n\to\infty} \frac{1}{|F_n|} \log_2 M_n(\tau)$. We derive here an upper bound on $R(\tau)$.

Fix $\mu_0(1)$ to be $\tau - \delta$ for some small $\delta \in (0, \tau)$. Then, from (4), we have $\alpha = \frac{1-2(\tau-\delta)}{1-(\tau-\delta)}$. It can be easily verified that choosing the parameters (q_0, q_1) of $\mu_{m,n}$ in the following way satisfies the condition in (9):

$$q_0 \in \left[\max(0, \frac{\alpha^2+\alpha-1}{\alpha^2}), 1\right] \text{ and } q_1 = \frac{\alpha^2}{1-\alpha}(1-q_0). \tag{18}$$

Hence, for any choice of (q_0, q_1) satisfying (18), an RSW measure μ exists. Assume that Conjecture 1 holds.

Consider the channels $\mathcal{Q}^{(n)}, n = 1, 2, \ldots$, defined by (16). Since $\mu_0(1) = \tau - \delta$, by (13), the probability that more than $\tau|F_n|$ errors occur in the channel $\mathcal{Q}^{(n)}$ goes to 0 as $n \to \infty$. Hence, any family of $(\tau|F_n|)$-hsec codes can operate over the channels $\mathcal{Q}^{(n)}$ with probability of error going to 0 as $n \to \infty$. Therefore, using Theorem 3, and letting $\delta \to 0$, we conclude that $R(\tau) \leq 1 - H(\mu)$, where μ is parametrized by (q_0, q_1) as in (18), with $\alpha = \frac{1-2\tau}{1-\tau}$. Using the expression for $H(\mu)$ from (10), and tightening the bound by choosing the best possible (q_0, q_1), we obtain

$$R(\tau) \leq 1 - \max_{(q_0,q_1)} \frac{1-2\tau}{1-2\tau+q_1\tau} \left[(1-2\tau)h(q_0) + \tau\, h(q_1)\right], \tag{19}$$

the maximum being taken over all (q_0, q_1) satisfying (18) with $\alpha = \frac{1-2\tau}{1-\tau}$.

We numerically evaluated the upper bound in (19) for $\tau \in [0, 1/2]$, and found it to be decreasing in τ for $\tau \in [0, 0.228]$, and increasing thereafter. Since $R(\tau)$ must be non-increasing in τ, we use the bound $R(\tau) \leq R(0.228)$ for all $\tau \geq 0.228$. The upper bound is plotted in Figure 3.

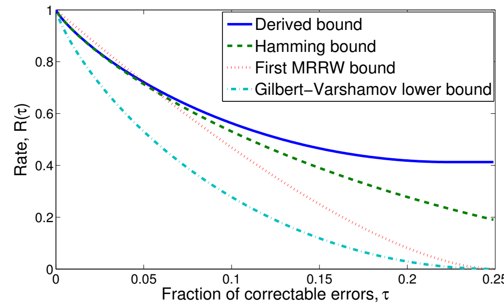

Fig. 3. Comparison of the derived upper bound on $R(\tau)$, calculated from (19), with the asymptotic Hamming bound and the first MRRW upper bound. Also plotted is the asymptotic GV lower bound.

For comparison's sake, also plotted are the asymptotic Hamming and first MRRW upper bounds [8, Chapter 4] for $(\tau|F_n|)$-correcting codes (for unconstrained error patterns), and the asymptotic Gilbert-Varshamov (GV) lower bound.

We believe that the upper bound on $R(\tau)$ can be further tightened. Indeed, we conjecture that asymptotic code rates of hsec codes can be no better than those for codes correcting unconstrained errors. This conjecture stems from an as-yet-unproved assertion that for large values of n, among F_n-configurations of Hamming weight at most $2t$, all but a vanishingly small fraction can be written as a sum of two configurations in $\mathsf{HS}(F_n)$ of weight at most t each. Consequently, the arguments in the proof of [6, Prop. 2.3] will go through in the asymptotic regime.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd ed.*, Wiley, 2006.

[2] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume II*, Delhi: Wiley, 2009.

[3] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, "Write channel model for bit-patterned media recording," *IEEE Trans. Magn.*, vol. 47, no. 1, pp. 35–45, Jan. 2011.

[4] A. Kavcic, X. Huang, B. Vasic, W. Ryan, and M. F. Erden, "Channel modeling and capacity bounds for two-dimensional magnetic recording," *IEEE Trans. Magn.,* vol. 46, no. 3, pp. 812–818, March 2010.

[5] E. Lindenstrauss, "Pointwise theorems for amenable groups," *Electronic Research Announcements of the Amer. Math. Soc.*, vol. 5, pp. 82–90, June 1999.

[6] A. Mazumdar and A. Barg, "Channels with Intermittent errors," in *Proc. IEEE Int. Symp. Inf. Theory,* 2011.

[7] A. Mazumdar, A. Barg, and N. Kashyap, "Coding for high-density recording on a 1-D granular magnetic medium," *IEEE Trans. Inf. Theory,* vol. 57, no. 11, pp. 7403–7417, Nov. 2011.

[8] R. M. Roth, *Introduction to Coding Theory*, Cambridge Univ. Press, 2006.

[9] R. M. Roth, P. H. Siegel, and J. K. Wolf, "Efficient coding schemes for the hard-square model," *IEEE Trans. Inf. Theory,* vol. 47, no. 3, pp. 1166–1176, March 2001.

[10] R. White, R. Newt, and R. Pease, "Patterned media: a viable route to 50 Gbit/in^2 and up for magnetic recording?," *IEEE Trans. Magn.*, vol. 33, no. 1, pp. 990–995, Jan. 1997.

[11] R. Wood, M. Williams, A. Kavcic, and J. Miles, "The feasibility of magnetic recording at 10 terabits per square inch on conventional media," *IEEE Trans. Magn.,* vol. 45, no. 2, pp. 917–923, Feb. 2009.