Channel Modeling and Capacity Bounds for Two-Dimensional Magnetic Recording

Aleksandar Kavcic, Xiujie Huang, Bane Vasic, William Ryan and M. Fatih Erden

***Abstract*—Two-Dimensional Magnetic Recording (TDMR) differs from traditional track recording in that the bit-size-to-grain-size ratio is drastically reduced so that the channel bits are roughly of the same sizes as those of the magnetic media grains. This is envisioned to be achieved by shingled writing, a write-process in which the corner-writer partially overwrites the previously written adjacent track, effectively writing very closely spaced narrow tracks with no guard bands. Since the tracks are very narrow, they can be read by an array of read-elements whose spacing is equal to the narrow track pitch, creating a 2-dimensional (2D) array of readback signals (the two dimensions being the cross-track and the down-track dimensions). In the absence of an array of read elements, the same 2D readback can be obtained by progressive scans of a single read element that reads one narrow track at a time and stores the readback signal in a 2D array. In this paper, we present simple magnetic grain media models and shingled-write process models that capture the essence of TDMR. We assume that the granular recording medium consists of randomly shaped tiles (each tile represents a grain), randomly covering the medium plane. We then derive a suitable 2-dimensional read/write process model. Using proper information-theoretic inequalities and bounding techniques, we derive methods to bound the capacities of TDMR channels. The results reveal that information capacities in excess of 0.6 user bits per grain are possible to attain over TDMR channels.**

I. Introduction

Two-dimensional magnetic recording (TDMR) was first suggested by Wood et.al. [1] as an alternative technology for approaching magnetic recording densities of 10Tb/in^2 and higher. Other proposed technologies include heat-assisted magnetic recording (HAMR) [2], [3], microwave-assisted magnetic recording (MAMR) [4] and bit-patterned magnetic recording (BPMR) [5]. The basic difference between TDMR on one hand, and HAMR, MAMR and BPMR on the other, is that TDMR does not require the media to be redesigned to accommodate the technology change. TDMR attempts to extend the use of conventional media, but requires radical changes in writing, reading, signal processing and data access software integration. However, TDMR should not be seen as a strict alternative to HAMR, MAMR or BPMR. Instead, we emphasize that TDMR could be utilized as an enhancement in conjunction with either HAMR, MAMR or BPMR.

It was suggested in [1] that the integral component of TDMR would be shingled writing. Shingled writing refers to a write process in which the writer partially overwrites the adjacent track, resulting in a very narrow track width. It was envisioned in [1] that shingled writing will be utilized in order to achieve track widths that are comparable to the grain diameter. In fact, the bold assumption made in [1] is that the average area of a grain will be roughly twice as large as the area of a channel bit. Obviously, this will create technological problems that will need to be addressed: 1) accurate head positioning and timing recovery will be much more difficult to achieve, and 2) very high levels of 2D intersymbol interference will necessitate the use of sophisticated 2D signal processing. Our goal here is not to address these difficult issues, but rather to focus on building a working mathematical model for the TDMR channel. For excellent expositions of TDMR and the associated technological challenges, we refer the reader to [1], [6], [7], [8], [9], [10].

In the earliest exposition of TDMR [1], a 1-dimensional (1D) model was used to argue that recording densities above 0.5 bits/grain are achievable. This was cause for early optimism, and subsequent research on channel models [7], [8], [9], code design [7] and systems issues [10] followed. In this paper, we develop a true 2D channel model and bound the capacity (achievable information rate) for such a model. The paper covers two basic topics.

1) We develop a fairly general channel model for TDMR. The model is based on representing magnetic grains as random-sized tiles that cover the magnetic medium plane. We further go on to show that the medium (consisting of randomly shaped tiles) can be modeled as a two-dimensional (2D) *causal* Markov field, with only a minimal loss in accuracy. This finding vastly simplifies (and accelerates) the creation of large sample sizes that are typically needed in statistical studies (such as bit error rate simulations). Further, with proper notation for the random *tiling* process, we develop a fairly general input-output relationship (channel law) that is capable of capturing irregular and fractional tile shapes, 2D intersymbol interference and additive readback noise.
2) For a very simple TDMR channel, we develop a methodology to compute bounds on the achievable information rates. We show that if the average grain area is roughly equal to the area of 2 channel bits, the channel capacity is at least 0.6 bits/grain! This is a very promising result given that the receiver (reader) need not know the exact positions nor sizes of the grains, yet the information rate is only 40% lower than the ultimate theoretical limit of 1 bit/grain.

The paper is organized as follows. In Section II, for clarity

This research was supported by the Information Storage Industry Consortium (INSIC)

The authors would like to thank Roger Wood for technical guidance, support and encouragement throughout the duration of this project.

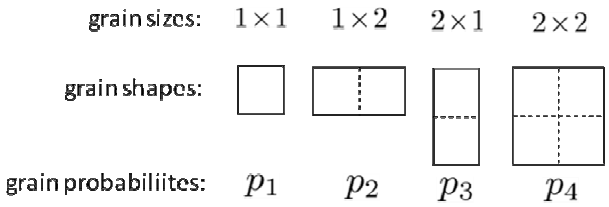

Fig. 1. A simple tiling model: Tile sizes and their probabilities of occurrence.

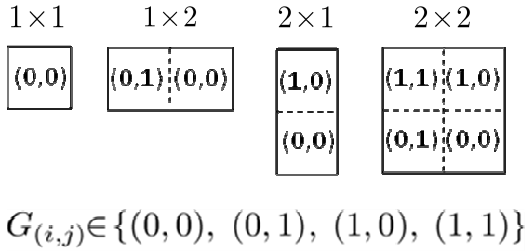

Fig. 2. Quadrant notation: Each quadrant is labeled by a pair (a, b), where $a \in \{0, 1\}$ and $b \in \{0, 1\}$.

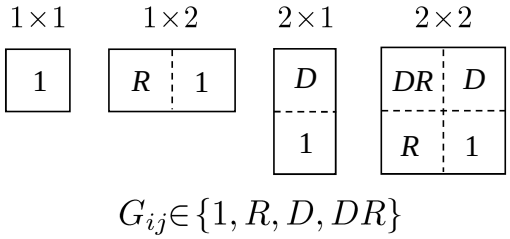

Fig. 3. Operator notation: Each quadrant is labeled by an operator of the form $D^a R^b$, where $a \in \{0, 1\}$ and $b \in \{0, 1\}$.

of presentation, we give only a very simple tiling model and show how to represent the model using causal 2D Markov fields. A much more comprehensive model (that incorporates irregular and fractional tile shapes, 2D intersymbol interference and additive readback noise) is given in Appendix A. In Section III, we establish the proper input-output relationship (channel law) that mathematically captures both the shingled writing process and the effect of grain randomness. In the same section, we demonstrate that causal 2D Markov fields are reasonably accurate models for isotropic random granular media. Finally, in Section IV, we develop and evaluate some upper and lower bounds on the channel capacity of the TDMR model exposed in Sections II and III

II. Granular Tiling Model

A. Simple granular tiling model

We assume that the granular recording medium consists of randomly shaped tiles (each tile represents a grain), randomly covering the medium plane. A tile of size $m \times n$ means that the width of the tile is equal to m-times the width of the channel bit, and the length of the tile equals to n-times the length of the channel bit. Here m and n can be any positive rational numbers. For clarity of presentation, we shall assume that the grain tiles can be of 4 possible sizes: 1×1, 1×2, 2×1 and 2×2, that occur with probabilities p_1, p_2, p_3 and p_4, respectively (see Figure 1). In Appendix A, we relax this simplifying assumption and consider tiling models in which the grain dimensions m and n can be non-integers.

Let us now consider how to create a random tiling of a medium whose area is $N \times N$ channels bits (symbols). Here, we assume that N is a very large number. Within the $N \times N$ area, let the total number of tiles of sizes 1×1, 1×2, 2×1 and 2×2 be N_1, N_2, N_3 and N_4, respectively. If we assume that N is a very large number, we have (with probability 1)

$$N^2 = N_1 + 2N_2 + 2N_3 + 4N_4. \quad (1)$$

Hence, if N is very large, we have for $i \in \{1, 2, 3, 4\}$ (with probability 1)

$$N_i = p_i(N_1 + N_2 + N_3 + N_4). \quad (2)$$

By solving the system of linear equations (1) and (2), we can obtain the numbers N_1, N_2, N_3 and N_4 in terms of p_1, p_2, p_3, p_4 and N. To create a medium whose tiles occur with probabilities p_1 through p_4, we utilize the following algorithm:

1) Randomly place N_4 tiles of size 2×2 in the square of size $N \times N$.
2) Next, randomly place N_2 and N_3 tiles of sizes 1×2 and 2×1, respectively, in the remaining slots of the $N \times N$ square.
3) Place the remaining N_1 tiles of size 1×1 in the remaining slots.

B. Granular tiling field notations

We divide each tile into subtiles, called **quadrants**, where each quadrant is of size 1×1. Thus a tile of size 1×1 has only one quadrant, a tile of size 1×2 (or 2×1) has 2 quadrants, and a tile of size 2×2 has 4 quadrants.

Quadrant notation: We label each quadrant by a pair (a, b), where $a \in \{0, 1\}$ and $b \in \{0, 1\}$. The quadrant labels for each of the tiles are given in Figure 2. As the tiles cover the media grid, each grid location (i, j) (where $1 \le i \le N$ and $1 \le j \le N$) is covered by a quadrant of a tile. For example, from Figure 2, if location (i, j) is occupied by the upper-left-hand quadrant of a 2×2 tile whose label is $(a, b) = (1, 1)$, we write $G_{(i,j)} = (1, 1)$.

Operator notation: Let the pair (a, b) be the same as in the quadrant notation, but now we label each quadrant using the operator notation $D^a R^b$ given in Figure 3, where $a \in \{0, 1\}$ and $b \in \{0, 1\}$. To each grid location (i, j) (where $1 \le i \le N$ and $1 \le j \le N$) we assign a tile label. For example, from Figure 3, if location (i, j) is occupied by the left-hand quadrant of a 1×2 tile, then we write $G_{(i,j)} = R$.

C. Causal Markov granular tiling field

Thus far, we have considered a random tiling as described in Subsection II-A. However, creating a realization of such a randomly tiled medium requires creating a full $N \times N$ plane of tiles. If N is very large (such as required in large statistical simulations), this is very impractical. For this reason,

we consider an alternative causal Markov tiling field. To do that, let us first introduce the notion of causality in 2D random fields. Relative to any position (i,j) in a 2D grid, we define the set of *past* field values as follows.

- The set of past field values relative to position (i,j), denoted by $\mathcal{G}(i,j)$ is defined as $\mathcal{G}(i,j) = \left\{G_{(k,\ell)} | (k<i) \text{ or } (k=i \text{ and } \ell<j)\right\}$.

A *causal* 2D Markov random field is a field $G_{(i,j)}$ for which the conditional probability mass (or density) function satisfies

$$p\left(G_{(i,j)} \middle| \mathcal{G}(i,j)\right) = p\left(G_{(i,j)} \middle| \mathcal{F}(i,j)\right)$$

where $\mathcal{F}(i,j)$ is a *finite-size* subset of $\mathcal{G}(i,j)$.

We next give two examples of causal tiling fields that will be of interest int his paper.

- A 3-rd order 2D causal Markov granular tiling field $G_{(i,j)}$ is a field for which

$$p\left(G_{(i,j)} \middle| \mathcal{G}(i,j)\right) = p\left(G_{(i,j)} \middle| \mathcal{F}^{(3)}(i,j)\right)$$

  where $\mathcal{F}^{(3)}(i,j)$ is a set of exactly 3 elements, i.e., $\left|\mathcal{F}^{(3)}(i,j)\right| = 3$. The exact choice of $\mathcal{F}^{(3)}(i,j)$ is arbitrary and is up to the designer of the model. Here we choose

$$\mathcal{F}^{(3)}(i,j) = \left\{G_{(i-1,j)}, G_{(i-1,j+1)}, G_{(i,j-1)}\right\}.$$

- A 4-th order 2D causal Markov granular tiling field $G_{(i,j)}$ is a field for which

$$p\left(G_{(i,j)} \middle| \mathcal{G}(i,j)\right) = p\left(G_{(i,j)} \middle| \mathcal{F}^{(4)}(i,j)\right)$$

  where $\mathcal{F}^{(4)}(i,j)$ is a set of exactly 4 elements, i.e., $\left|\mathcal{F}^{(4)}(i,j)\right| = 4$. The exact choice of $\mathcal{F}^{(4)}(i,j)$ is arbitrary and is up to the designer of the model. Here we choose

$$\mathcal{F}^{(4)}(i,j) = \left\{G_{(i-1,j)}, G_{(i-1,j+1)}, G_{(i-1,j+2)}, G_{(i,j-1)}\right\}.$$

Note that a causal Markov field can never be equivalent to the non-causal field described in Section II-A because the Markov approximation inevitably introduces approximation errors. However, if the order of the causal Markov field is chosen large enough, the approximation error can be made arbitrarily small. In the next section, after introducing the proper channel input-output model, we will show (by virtue of an example) that an appropriately chosen measure of the approximation error of a 3-rd order causal Markov field is already very small. Further, we will demonstrate that the approximation error of a 4-th order causal Markov field is even smaller.

III. Channel Input-Output Model

A. Shingled writing and input-output relationship

The basic assumption in TDMR is that shingled writing will be employed [1], [6]. Shingled writing creates tracks whose track widths are only a fraction of the write head width. This is achieved by partially overwriting previously written tracks, see [1], [6] for further explanations.

Since the grains can have dimensions that are larger than either the channel bit length or the track width, it is possible that the writer overwrites the same grain several times before the grain assumes its final (remanent) magnetized state. The remanent polarization state of the grain is always equal to the direction of the latest head field applied to the grain. Hence, the exact write sequence is very important when determining the channel input-output relationship. Here, we will assume that the bits are written onto the medium in a raster-scan fashion, which we explain next.

Let $X_{(i,j)} \in \{-1,1\}$ denote the channel input bit (head field) that is applied to the grid position (i,j). The write-head starts at track $i=1$ and sequentially writes over positions $(1,1)$, $(1,2)$, $(1,3)$, and so on. Next, the writer is positioned over track $i=2$ and writes sequentially over positions $(2,1)$, $(2,2)$, $(2,3)$, and so on. This process continues until all tracks are written.

Now, since the grains are sometimes larger than the channel symbol dimensions, we must account for the possibility that a grain senses the write field of adjacent values $X_{(i,j)}$. The channel output $Y_{(i,j)} \in \{-1,1\}$ is defined as the remanent magnetization direction at the end of the write process. Obviously, the field $Y_{(i,j)}$ must depend on the realization of the channel input $X_{(i,j)}$ as well the realization of the granular tiling field $G_{(i,j)}$. The exact form of the input-output relationship (channel law) depends on the notation chosen for the granular tiling field $G_{(i,j)}$. We give the channel law for both the quadrant notation and the operator notation.

Channel law - quadrant notation: The channel law is particularly easy to capture using the quadrant notation:

$$Y_{(i,j)} = X_{(i,j)+G_{(i,j)}} \tag{3}$$

where $G_{(i,j)} \in \{(0,0),(0,1),(1,0),(1,1)\}$. Note that in (3) the granular tiling field $G_{(i,j)}$ acts as a random timing offset that alters the index of $X_{(i,j)}$.

Channel law - operator notation: Define the bivariate channel-input polynomial $X(D,R) \triangleq \sum_{i,j} X_{(i,j)} D^i R^j$. Similarly, let $Y(D,R) \triangleq \sum_{i,j} Y_{(i,j)} D^i R^j$ be the channel-output polynomial. Using the operator notation $G_{(i,j)} \in \{1, D, R, DR\}$, the channel law takes the following form

$$Y(D,R) = \sum_{i,j} Y_{(i,j)} D^i R^j \tag{4}$$

where

$$Y_{(i,j)} = \begin{cases} X_{(i,j)} & \text{if } G_{(i,j)} = 1 \\ X_{(i,j+1)} & \text{if } G_{(i,j)} = R \\ X_{(i+1,j)} & \text{if } G_{(i,j)} = D \\ X_{(i+1,j+1)} & \text{if } G_{(i,j)} = DR \end{cases}.$$

B. Isotropy of 2D causal Markov granular tiling fields

Let $X_{(i,j)}$ be a 2D random field consisting of independent and identically distributed symbols such that $\mathrm{P}\left\{X_{(i,j)}=1\right\} = \mathrm{P}\left\{X_{(i,j)}=-1\right\} = 1/2$. Assume that the tiling field $G_{(i,j)}$ is the one described in Section II-A with parameters $p_1 = 0.2$, $p_2 = p_3 = 0.35$ and $p_4 = 0.1$. Let $Y_{(i,j)}$ be the channel output when $X_{(i,j)}$ is the channel input and

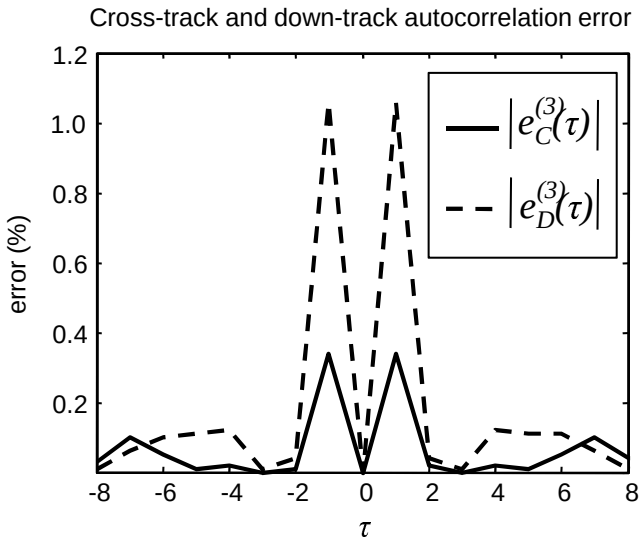

Fig. 4. Cross-track and down-track autocorrelation error for the 3-rd order causal Markov field approximation.

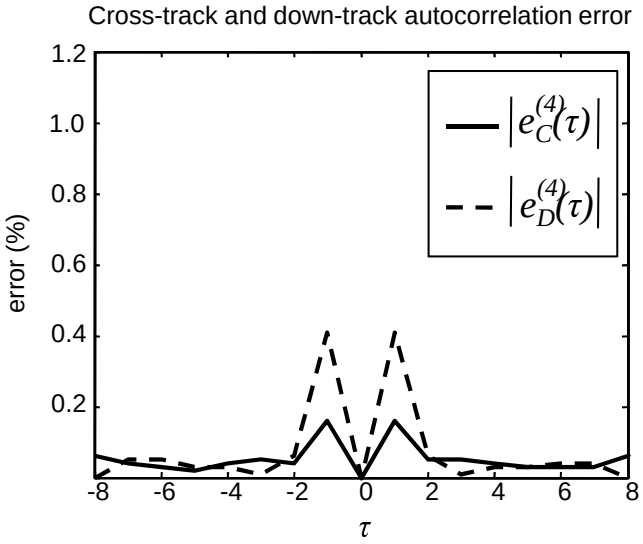

Fig. 5. Cross-track and down-track autocorrelation error for the 4-th order causal Markov field approximation.

$G_{(i,j)}$ is the granular tiling field. We define the cross-track and down-track autocorrelation functions, respectively, as

$$A_C(\tau) = \mathrm{E}\left[Y_{(i+\tau,j)}Y_{(i,j)}\right] \quad (5)$$

$$A_D(\tau) = \mathrm{E}\left[Y_{(i,j+\tau)}Y_{(i,j)}\right]. \quad (6)$$

Obviously, since by construction $Y_{(i,j)}$ has equal statistics in the cross-track and down-track directions, we have $A_C(\tau) = A_D(\tau)$.

Now, let $G^{(3)}_{(i,j)}$ be the 3-rd order causal Markov field that best approximates $G_{(i,j)}$, meaning that the 3-rd order conditional marginals of $G_{(i,j)}$ and $G^{(3)}_{(i,j)}$ are equal. Let $Y^{(3)}_{(i,j)}$ be the channel output when $X_{(i,j)}$ is the channel input and $G^{(3)}_{(i,j)}$ is the granular tiling field. The corresponding cross-track and down-track autocorrelation functions are

$$A^{(3)}_C(\tau) = \mathrm{E}\left[Y^{(3)}_{(i+\tau,j)}Y^{(3)}_{(i,j)}\right] \quad (7)$$

$$A^{(3)}_D(\tau) = \mathrm{E}\left[Y^{(3)}_{(i,j+\tau)}Y^{(3)}_{(i,j)}\right]. \quad (8)$$

Next, define the normalized cross-track and down-track autocorrelation errors as

$$e^{(3)}_C(\tau) = \frac{A^{(3)}_C(\tau) - A_C(\tau)}{A_C(0)} \quad (9)$$

$$e^{(3)}_D(\tau) = \frac{A^{(3)}_D(\tau) - A_D(\tau)}{A_D(0)}. \quad (10)$$

The magnitudes of $e^{(3)}_C(\tau)$ and $e^{(3)}_D(\tau)$ are plotted in Figure 4. From the figure we see that the process $Y^{(3)}_{(i,j)}$ (and therefore also $G^{(3)}_{(i,j)}$) is anisotropic because $e^{(3)}_C(\tau) \neq e^{(3)}_D(\tau)$. However, we also see from Figure 4 that both the cross-track and the down-track autocorrelation errors are within 1.1%, which is already an acceptable margin of error.

Next, we construct the best-fitting 4-th order causal Markov tiling field $G^{(4)}_{(i,j)}$, and define the associated cross-track and down-track autocorrelation functions $A^{(4)}_C(\tau)$ and $A^{(4)}_D(\tau)$. Correspondingly, we define the normalized cross-track and down-track autocorrelation errors

$$e^{(4)}_C(\tau) = \frac{A^{(4)}_C(\tau) - A_C(\tau)}{A_C(0)} \quad (11)$$

$$e^{(4)}_D(\tau) = \frac{A^{(4)}_D(\tau) - A_D(\tau)}{A_D(0)}. \quad (12)$$

Figure 5 shows the magnitudes of $e^{(4)}_C(\tau)$ and $e^{(4)}_D(\tau)$. We notice that the autocorrelation error dropped to within 0.4%. We further notice that $\left|e^{(4)}_C(\tau)\right| \leq \left|e^{(3)}_C(\tau)\right|$ and $\left|e^{(4)}_D(\tau)\right| \leq \left|e^{(3)}_D(\tau)\right|$. This illustrates that as the Markov order m increases, the anisotropy error between $G^{(m)}_{(i,j)}$ and $G_{(i,j)}$ decreases, which supports the claim that causal Markov tiling fields are reasonable low-complexity approximations of isotropic granular tiling fields.

IV. Channel Capacity Bounds

In this section, we shall assume that the probabilities of the 1×2 and 2×1 tiles are equal

$$p_2 = p_3.$$

Further, we shall assume that the average grain size equals to the area of 2 channel symbols (channel bits), i.e.,

$$2 = p_1 + 2p_2 + 2p_3 + 4p_4. \quad (13)$$

Let $X_{N\times N}$ denote the 2D channel input sequence of size $N \times N$, and let $Y_{N\times N}$ denote the 2D channel output sequence of size $N\times N$. As usual, the definition of the channel capacity is the maximal achievable mutual information rate between the channel input and the channel output

$$C_{[\mathrm{b/ch.use}]} = \lim_{N\to\infty} \frac{1}{N^2} \max_{\mathrm{P}(X_{N\times N})} I\left(X_{N\times N}; Y_{N\times N}\right). \quad (14)$$

The capacity in (14) is measured in *bits per channel use*. Alternatively, given that the average grain size is 2 symbols (13), we may express the channel capacity in terms of *bits per grain* as

$$C_{[\mathrm{b/grain}]} = 2 \cdot C_{[\mathrm{b/ch.use}]}. \tag{15}$$

A. Trivial bounds

The obvious upper bound on the capacity is

$$C_{[\mathrm{b/grain}]} \leq 1$$

because a single grain cannot store more than 1 bit of information. The next proposition establishes an easy lower bound.

Proposition 1:

$$C_{[\mathrm{b/grain}]} \geq \frac{1}{2}.$$

Proof: Let us adopt the following transmission scheme

$$X_{(i,j)} = \begin{cases} X_{(i,j)} & \text{if } i \text{ is even} \quad \text{and} \quad j \text{ is even} \\ X_{(i-1,j)} & \text{if } i \text{ is odd} \quad \text{and} \quad j \text{ is even} \\ X_{(i,j-1)} & \text{if } i \text{ is even} \quad \text{and} \quad j \text{ is odd} \\ X_{(i-1,j-1)} & \text{if } i \text{ is odd} \quad \text{and} \quad j \text{ is odd} \end{cases}.$$

Clearly such a transmission scheme results form a repetition code of rate $r = 1/4$ that satisfies $X_{(2m,2n)} = X_{(2m+1,2n)} = X_{(2m,2n+1)} = X_{(2m+1,2n+1)}$. If this transmission scheme is utilized, one can easily verify that $Y_{(2m,2n)} = X_{(2m,2n)}$ for any realization of the tiling field $G_{(i,j)}$. Hence, all channel symbols $X_{(2m,2n)}$ can be recovered free of errors, and therefore the channel capacity must be at least as high as the employed code rate $r = 1/4$, i.e.,

$$C_{[\mathrm{b/ch.use}]} \geq r = \frac{1}{4} \qquad \text{or} \qquad C_{[\mathrm{b/grain}]} \geq \frac{1}{2}.$$

■

B. Lower bounds using K adjacent tracks

Let us employ the following transmission scheme. We write K adjacent tracks (say, tracks numbered 1 through K) and do not even attempt to write the $(K+1)$-st track (because attempting to write the $(K+1)$-st track may overwrite some grains that are shared between the K-th and $(K+1)$-st track). Then we read the K written tracks. Let $C^{(K)}$ be the maximal information rate achievable by this transmission scheme over K adjacent tracks. Then we have the following lower bound[1]

$$C \geq \frac{K}{K+1} C^{(K)}.$$

The problem in the bound above is that there exists no known method to compute the exact value $C^{(K)}$. However, we may utilize Markov chain Monte-Carlo techniques to compute lower bounds on $C^{(K)}$. Since the K-track transmission scheme constitutes a finite-state channel, we can employ Blahut-Arimoto-like Markov process optimization techniques [11], [12] to get numerical lower bounds on $C^{(K)}$. Let $C_M^{(K)}$ be the maximal information rate achievable by an M-th order Markov process. $C_M^{(K)}$ is easily computed by the generalized Blahut-Arimoto optimization algorithm in [11], [12] applied to the K-track finite-state channel. Clearly, $C^{(K)} \geq C_M^{(K)}$, and we have the following lower bound

$$L_M^{(K)} \triangleq \frac{K}{K+1} C_M^{(K)} \leq \frac{K}{K+1} C^{(K)} \leq C.$$

[1]Here, we assume that C and $C^{(K)}$ are either both expressed in bits per channel use, or both expressed in bits per grain. Hence, in the formulas that follow, we no longer specify whether information rates are expressed in bits per channel use or bits per grain.

C. Upper bounds using a genie-aided $(K+1)$-st track

When computing the lower bound, we utilized only K tracks. Assume now that a genie helps us utilize the $(K+1)$-st track in the following manner. Instead of the original $(K+1)$-st track, a genie creates a different (new) $(K+1)$-st track for us to use. The genie takes the 1×1 and 1×2 tiles from the original $(K+1)$-st track and concatenates them to create the new $(K+1)$-st track. We are now free to use the new $(K+1)$-st track. The probability of occurrence of the 1×1 tile in the new track is $p_1' = p_1/(p_1+p_2)$ and the probability of occurrence of the 1×2 tile in the new track is $p_2' = p_2/(p_1+p_2)$. Now, let C' be the capacity of the new genie-designed $(K+1)$-st track. Clearly, we then have an upper bound

$$\frac{K \cdot C^{(K)} + C'}{K+1} \geq C.$$

Since neither $C^{(K)}$ nor C' are known exactly, we utilize delayed-feedback numerical techniques [13] to further upper bound $C^{(K)}$ and C'. Let these upper bounds be denoted by $C^{(K)}(d)$ and $C'(d)$, respectively, where d denotes the delay parameter, see [13] for details. Then, clearly we have the following upper bound

$$U_d^{(K)} \triangleq \frac{K \cdot C^{(K)}(d) + C'(d)}{K+1} \geq \frac{K \cdot C^{(K)} + C'}{K+1} \geq C.$$

We note that, in any practical scenario, the delay parameter d must be chosen very small or else the numerical methods in [13] are prohibitively complex.

D. Numerical capacity bound values

It is clear that the developed upper and lower bounds on capacity are asymptotically tight as the parameters M, d and K become infinitely large, i.e.,

$$\lim_{K,M\to\infty} L_M^{(K)} = C = \lim_{K,d\to\infty} U_d^{(K)}.$$

However, obtaining the bounds for very large values of K, M and d is computationally prohibitive, so we resort to reasonably small values K, M and d.

Figure 6 shows the lower bound $L_6^{(2)}$ and the upper bound $U_2^{(2)}$, for the entire range of possible probabilities p_2 under the assumptions

$$\begin{aligned} 1 &= p_1 + p_2 + p_3 + p_4 \\ p_2 &= p_3 \\ 2 &= p_1 + 2p_2 + 2p_3 + 4p_4. \end{aligned}$$

One can see from Figure 6 that the capacity is higher than 0.6 bits/grain over the entire range of tiling probabilities (assuming that the average tile size is 2 channel bits). We underscore that

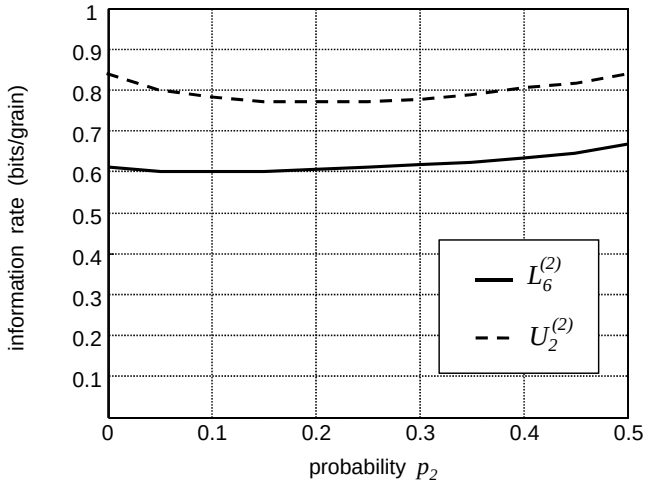

Fig. 6. Upper and lower bound on capacity. For the lower bound, we used $M = 6$ and $K = 2$. For the upper bound, we used $d = 2$ and $K = 2$.

this result is obtained for a very simple media tiling model, but nonetheless this result is very encouraging for the following reasons.

- No knowledge of the positions nor sizes of grain (tile) realizations is available to the reader (receiver). That is, the receiver only knows the probabilities p_1, p_2, p_3 and p_4 of tile occurrences.
- The upper capacity limit when the reader (receiver) knows the position and size of each grain (such as in envisioned patterned-media systems [5]) is 1 bit per grain. Here we showed that a reasonably high information rate of 0.6 bits per grain is achievable under vastly simpler channel assumptions.

The reader of this paper should understand these findings as *preliminary* encouraging results. To get more realistic estimates of the ultimate information theoretic limits, more research is needed in the following directions.

1) More accurate channel models - Appendix A may provide a path towards this goal.
2) Better 2D capacity bounding methods - an encouraging development may be the modified Gibbs sampling methodology using disjoint trees [14].
3) Faster simulations and efficient error correction codes - see [7], [8], [9] for research directions along these lines.

V. Conclusion

In this paper, we developed a simple model that describes random magnetic grains as tiles. We further showed that causal 2D Markov fields may be utilized to very accurately describe the granular tiling process. Next, we developed an input-output model (channel law) that captures the effects of shingled writing over a tiled magnetic medium. For such a channel law, we constructed upper and lower bounds on the channel capacity and evaluated the bounds for a medium for which the average tile (grain) area equals twice the channel bit area. The surprising result was that the achievable information rate (channel capacity) is higher than 0.6 bits/grain. This is a very high information rate considering that the receiver has no knowledge of the exact grain sizes nor positions. While these results are obtained for a very simple tiling model, they are very encouraging because they confirm earlier speculations [1] that the limits of TDMR are around 0.5 bits/grain.

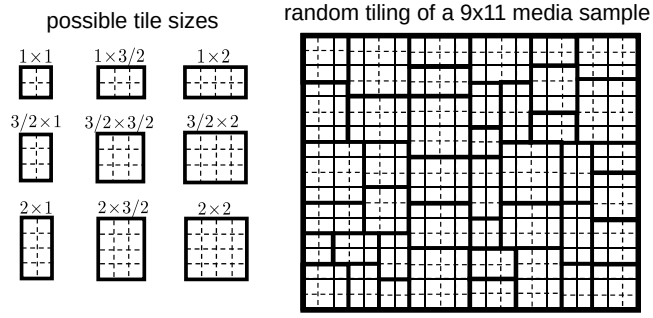

Fig. 7. Left: Possible tile sizes. Right: Random tile covering of a medium plane.

Appendix A: Non-integer tile sizes

In this Appendix, we consider tiling models whose tile sizes are non-integers. For simplicity, let us first assume a simple tiling model in which the tiles can occur in sizes $m\times n$, where

$$m \in \left\{1, \frac{3}{2}, 2\right\} \quad \text{and} \quad n \in \left\{1, \frac{3}{2}, 2\right\}.$$

Figure 7 shows the possible tile sizes and a sample realization of a media plane tiling.

Since all tile dimensions are multiples of $1/2$, we divide each of the possible 9 tiles into sub-quadrants of sizes $1/2 \times 1/2$. Figure 8 shows the sub-quadrants along with the assigned sub-quadrant labels. Note that the tiling field $G_{(i,j)}$ is now defined over the sub-quadrant grid, over which we can now apply a channel law similar to (3). The only remaining issue is the proper conversion from the symbol grid to the sub-quadrant grid, which we address next.

As in Section III, denote by $X_{(i,j)}$, the channel-input field (defined on the symbol grid). To convert $X_{(i,j)}$ to the sub-quadrant grid, we simply upsample $X_{(i,j)}$ by a factor of 2 in each dimension to get

$$A_{(i,j)} = X_{\left(\lfloor \frac{i}{2} \rfloor, \lfloor \frac{j}{2} \rfloor\right)}$$

where $A_{(i,j)}$ is the equivalent field representation on the sub-quadrant grid. We next apply the channel law on the sub-quadrant grid (in quadrant notation) as

$$B_{(i,j)} = A_{(i,j)+G_{(i,j)}}.$$

Finally, we must downsample by a factor of 2 in each dimension to convert the channel-output signal back to the symbol grid as follows

$$Y_{(i,j)} = \frac{1}{4} \sum_{\alpha=0}^{1} \sum_{\beta=0}^{1} B_{(2i+\alpha, 2j+\beta)}.$$

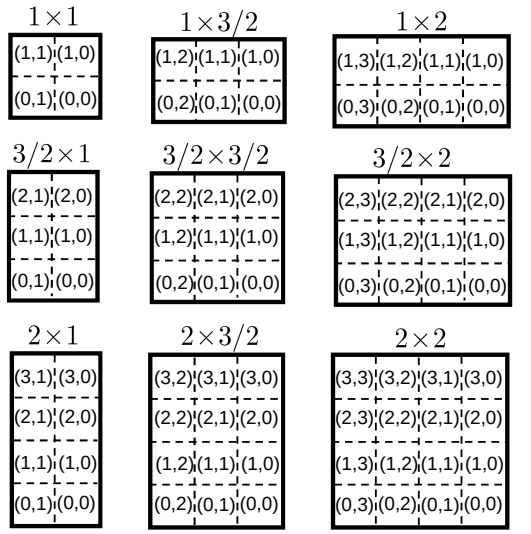

Fig. 8. Sub-quadrants and sub-quadrant labels.

In a general case, if the smallest sub-quadrant is of dimension $1/k \times 1/k$, the proper channel model is captured by the following sequence of operations

1) Upsample by a factor of k in each dimension to get the equivalent channel-input field on the sub-quadrant grid
$$A_{(i,j)} = X_{\left(\lfloor \frac{i}{k} \rfloor, \lfloor \frac{j}{k} \rfloor\right)}.$$
2) Apply the channel law (in quadrant notation) to the sub-quadrant grid
$$B_{(i,j)} = A_{(i,j)+G_{(i,j)}}.$$
3) Downsample by a factor of k in each dimension to map the readback signal onto the symbol grid
$$Y_{(i,j)} = \frac{1}{k^2} \sum_{\alpha=0}^{k-1} \sum_{\beta=0}^{k-1} B_{(ki+\alpha, kj+\beta)}. \qquad (16)$$

In an even more comprehensive model, we can incorporate additive readback noise so that equation (16) takes the form

$$Y_{(i,j)} = \frac{1}{k^2} \sum_{\alpha=0}^{k-1} \sum_{\beta=0}^{k-1} B_{(ki+\alpha, kj+\beta)} + W_{(i,j)}$$

where $W_{(i,j)}$ is the additive noise field. This model can be even further generalized by incorporating arbitrary 2D intersymbol interference into the readback signal as follows

$$Y_{(i,j)} = \sum_{\alpha,\beta} H_{(\alpha,\beta)} B_{(ki+\alpha, kj+\beta)} + W_{(i,j)}$$

where $H_{(\alpha,\beta)}$ is a 2D read-head response. Thus, we conclude that the simple media model (Section II) and the channel law (Section III) can easily be extended to allow non-integer tile dimensions, 2D intersymbol interference and additive readback noise.

REFERENCES

[1] 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, pp. 917–923, Feb. 2008.

[2] R. E. Rottmayer, S. Batra, D. Buechel, W. A. Challener, J. H. Y. Kubota, L. Li, B. Lu, C. Mihalcea, K. Mountfield, K. Pelhos, C. Peng, T. Rausch, M. A. Seigler, D. Weller, and X. Yang, "Heat-assisted magnetic recording," *IEEE Transactions on Magnetics*, vol. 42, pp. 2417–2421, October 2006.

[3] M. H. Kryder, E. C. Gage, T. W. McDaniel, W. A. Challener, R. E. Rottmayer, G. Ju, Y.-T. Hsia, and M. F. Erden, "Heat assisted magnetic recording," *Proc. IEEE*, vol. 96, pp. 1810–1835, November 2008.

[4] J.-G. Zhu, X. Zhu, and Y. Tang, "Microwave assisted magnetic recording," *IEEE Transactions on Magnetics*, vol. 44, pp. 125–131, January 2008.

[5] B. Terris, T. Thomson, and G. Hu, "Patterned media for future magnetic data storage," in *Microsystem Technologies*, vol. 13, pp. 189–196, Springer Verlag, Berlin, November 2006.

[6] I. Tagawa and M. Williams, "High density data-storage using shingle-write." presented at IEEE INTERMAG, May 2009 (submitted for publication in *IEEE Transactions on Magnetics*).

[7] B. Vasic, A. R. Krishnan, R. Radhakrishnan, A. Kavcic, W. Ryan, and M. F. Erden, "Read channel modeling and detection for two-dimensional magnetic recording systems." presented at IEEE INTERMAG, May 2009 (submitted for publication in *IEEE Transactions on Magnetics*).

[8] K. Chan, J. J. Miles, E. Hwang, M. Lin, R. Negi, B. Kumar, and J. Zhu, "TDMR platform simulations and experiments." presented at IEEE INTERMAG, May 2009 (submitted for publication in *IEEE Transactions on Magnetics*).

[9] K. S. Chan, R. Radhakrishnan, K. Eason, R. M. Elidrissi1, J. Miles, B. Vasic, and A. R. Krishnan, "Channel models and detectors for two-dimensional magnetic recording (TDMR)." to be presented at IEEE TMRC, October 2009 (submitted for publication in *IEEE Transactions on Magnetics*).

[10] G. Gibson and M. Polte, "Directions for shingled-write and TDMR system architectures: Synergies with solid-state disks." presented at IEEE INTERMAG, May 2009 (submitted for publication in *IEEE Transactions on Magnetics*).

[11] A. Kavčić, "On the capacity of Markov sources over noisy channels," in *Proceedings IEEE Global Communications Conference 2001*, (San Antonio, Texas), pp. 2997–3001, November 2001.

[12] P. O. Vontobel, A. Kavcic, D. M. Arnold, and H.-A. Loeliger, "A generalization of the BlahutArimoto algorithm to finite-state channels," *IEEE Transactionson Information Theory*, vol. 54, pp. 1887–1918, 2008.

[13] X. Huang, A. Kavcic, X. Ma, and D. Mandic, "Upper bounds on the capacities of non-controllable finite-state channels using dynamic programming methods," in *IEEE Intrnational Symposium on Information Theory*, (Seoul, Korea), pp. 2346–2350, July 2009.

[14] H.-A. Loeliger and M. Molkaraie, "Simulation-based estimation of the partition function and the information rate of two-dimensional models," in *IEEE Intrnational Symposium on Information Theory*, (Toronto, Canada), pp. 1113–1117, July 2008.