On the Capacity of High-Density Recording on a 1-D Granular Magnetic Medium

Arya Mazumdar
Dept. of ECE and Inst. Sys. Res.
University of Maryland,
College Park, MD 20742, USA
Email: arya@umd.edu

Alexander Barg
Dept. of ECE and Inst. Sys. Res.
University of Maryland,
College Park, MD 20742, USA
and IPPI RAS, Moscow, Russia
Email: abarg@umd.edu

Navin Kashyap
Dept. of Mathematics and Statistics
Queen's University,
Kingston, ON K7L3N6, Canada
nkashyap@mast.queensu.ca

***Abstract*—In terabit-density magnetic recording, several bits of data can be replaced by the values of their neighbors in the storage medium. As a result, errors in the medium are dependent on each other and also on the data written. In an earlier paper we studied a simple one-dimensional model of this situation. In our model binary data is sequentially written on the medium and a bit can erroneously change to the immediately preceding value. Here we define a probabilistic finite-state channel model of the storage medium and derive estimates of its capacity. To derive a lower bound we consider a uniform input distribution, and find an exact expression of the *symmetric capacity* of the channel. An upper bound is found by showing that the original channel is a stochastic degradation of another, related channel model whose capacity we compute explicitly.**

I. Introduction

General magnetic recording media are composed of fundamental magnetizable units, called "grains", that do not have a fixed size or shape. Each grain can be magnetized to take on exactly one of the two magnetic polarities. Thus, each grain is in principle capable of storing at most one bit of information. At the same time, in conventional magnetic recording technologies, data is generally stored by dividing the magnetic medium into regularly-spaced bit cells, and writing one bit of data into each of these bit cells. Recently, Wood et al. [5] proposed a new write mechanism, that can magnetize areas commensurate to the size of individual grains. With such a write mechanism and a corresponding readback mechanism in place, the remaining bottleneck to achieving magnetic recording densities as high as 10 Terabits per square inch is that the write and readback mechanisms do not have precise knowledge of the grain boundaries.

In [1] we studied a simple one-dimensional version of a model of errors in binary data that arises as a consequence of reading/writing across the grain boundaries. Restricting ourselves to the one-dimensional case, we defined a combinatorial error model that corresponds to the granular medium described above. Our model of the medium comprises n bit cells, indexed by the integers from 1 to n. The granular structure is described by an increasing sequence of positive integers, $1 = j_1 < j_2 < \cdots < j_s \le n$, where j_i denotes the index of the bit cell at which the ith grain begins. Note that the length of the ith grain is $\ell_i = j_{i+1} - j_i$ (we set $j_{s+1} = n+1$ to be consistent). For notational ease, our model assumes that it is the *first* bit to be written within a grain that sets the polarity of the grain. Thus, for indices j within the ith grain, *i.e.*, for $j_i \le j < j_{i+1}$, we have $y_j = x_{j_i}$. This means that the ith grain introduces an error in the recorded data (*i.e.*, a situation where $y_j \ne x_j$) precisely when $x_j \ne x_{j_i}$ for some j satisfying $j_i < j < j_{i+1}$.

In [1], we considered a medium with grains of length at most 2. (Grains of length 1 do not introduce any errors.) We presented upper and lower bounds on the size of codes capable of correcting a given number of length-2 grains. Most results in [1] can be extended to the case where the maximum length of a grain is an arbitrary constant.

In this paper we consider a probabilistic channel model that corresponds to the one-dimensional combinatorial model of errors discussed above, calling it the "grains channel". We again confine ourselves to length-2 grains. Our objective is to calculate the capacity of the channel. For the lower bound we restrict our attention to uniformly distributed, independent input letters which corresponds to the case of *symmetric information rate* (symmetric capacity) of the channel. We are able to find an exact expression for the SIR as an infinite series which gives a lower bound on the capacity. Computing this expression is one of the main results of this paper (see Thm. 8 below). To estimate capacity from above, we relate the grains channel to an erasure channel in which erasures never occur in adjacent symbols, and are otherwise independent. An expression for its capacity that we find yields an upper bound on the capacity of the grains channel.

We would like to acknowledge a concurrent independent paper by Iyengar, Siegel, and Wolf [4]. The authors of [4] consider a slightly more general channel model that includes the grains channel as a particular case. Paper [4] contains our Prop. 2 and Theorem 4, but not the calculation of the symmetric information rate of the channel of Sect. II-D.

II. Capacity of the Grains Channel

In this section, we define a probabilistic model for the grains channel with grains of length at most 2. This is a binary-input binary-output channel that can make an error only at positions where a length-2 grain ends. In fact, error events are data-dependent: an error occurs at a position where a length-2 grain ends if and only if the channel input at that position differs from the previous channel input. Let us now formally define the channel.

Suppose $\boldsymbol{x} = x_1 x_2 \ldots$ and $\boldsymbol{y} = y_1 y_2 \ldots$ denote the input and output sequence respectively, with $x_i, y_i \in \{0, 1\}$ for all i. We further define the sequence $\boldsymbol{u} = u_1 u_2 \ldots$, where $u_i = 1$ (resp. $u_i = 0$) indicates that a length-2 grain ends (resp. does not end) at position i. We take $\boldsymbol{u}$ to be a first-order Markov chain, independent of the channel input $\boldsymbol{x}$, having transition probabilities $P(u_i|u_{i-1})$ as tabulated below (for some $p \in [0, 1]$):

$$\begin{array}{c|cc} & u_i = 0 & u_i = 1 \\ \hline u_{i-1} = 0 & 1-p & p \\ u_{i-1} = 1 & 1 & 0 \end{array} \quad . \tag{1}$$

The grains channel makes an error at position i (*i.e.*, $x_i \neq y_i$) if and only if $u_i = 1$ and $x_i \neq x_{i-1}$. To be precise,

$$y_i = x_i + u_i(x_i + x_{i-1}), \tag{2}$$

the additions and multiplications above being performed modulo 2. Equivalently,

$$y_i = \begin{cases} x_i & \text{if } u_i = 0 \\ x_{i-1} & \text{if } u_i = 1. \end{cases} \tag{3}$$

We will find it useful to define the error sequence $\boldsymbol{z} = z_1, z_2, z_3, \ldots$, where $z_i = x_i + y_i$, the addition again being modulo-2. Thus,

$$z_i = u_i(x_i + x_{i-1}). \tag{4}$$

The case $i = 1$ is not covered by the above definitions. We will include it once we define a finite-state model of the grains channel.

The grains channel as we have defined above is a special case of a somewhat more general "write channel" model considered in [4].

A. Discrete Finite-State Channels

For easy reference, we record here some important facts about discrete finite-state channels. The material in this section is substantially based upon [2, Section 4.6].

A stationary *discrete finite-state channel (DFSC)* has an input sequence $\boldsymbol{x} = x_1, x_2, x_3, \ldots$, an output sequence $\boldsymbol{y} = y_1, y_2, y_3, \ldots$, and a state sequence $\boldsymbol{s} = s_1, s_2, s_3, \ldots$. 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 s_n takes values in a finite set of states $\mathcal{S}$. The channel is described statistically by specifying a conditional probability assignment $P(y_n, s_n|x_n, s_{n-1})$, which is independent of n. It is assumed that, conditional on x_n and s_{n-1}, the pair y_n, s_n is statistically independent of all inputs x_j, $j < n$, outputs y_j, $j < n$, and states s_j, $j < n-1$. To complete the description of the channel, an initial state s_0, also taking values in $\mathcal{S}$, must be specified.

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

$$\underline{C}_n = n^{-1} \max_{Q^n(\boldsymbol{x}^n)} \min_{s_0 \in \mathcal{S}} I(\boldsymbol{x}^n; \boldsymbol{y}^n \mid s_0)$$
$$\overline{C}_n = n^{-1} \max_{Q^n(\boldsymbol{x}^n)} \max_{s_0 \in \mathcal{S}} I(\boldsymbol{x}^n; \boldsymbol{y}^n \mid s_0).$$

In the above expressions, $I(\boldsymbol{x}^n; \boldsymbol{y}^n \mid s_0)$ is the mutual information between the length-n input $\boldsymbol{x}^n = (x_1, \ldots, x_n)$ and the length-n output $\boldsymbol{y}^n = (y_1, \ldots, y_n)$, given the value of the initial state s_0, and the maximum is taken over probability distributions $Q^n(\boldsymbol{x}^n)$ on the input $\boldsymbol{x}^n$. 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 [2].

The upper and lower capacities coincide for a large class of channels known as *indecomposable* channels. Roughly, an indecomposable DFSC is a DFSC in which the effect of the initial state s_0 dies away with time. Formally, let $q(s_n \mid \boldsymbol{x}^n, s_0)$ denote the conditional probability that the nth state is s_n, given the input sequence $\boldsymbol{x}^n = (x_1, \ldots, x_n)$ and initial state s_0. Evidently, $q(s_n \mid \boldsymbol{x}^n, s_0)$ is computable from the channel statistics. A DFSC is indecomposable if, for any $\epsilon > 0$, there exists an n_0 such that for all $n \geq n_0$, we have

$$|q(s_n \mid \boldsymbol{x}^n, s_0) - q(s_n \mid \boldsymbol{x}^n, s_0')| \leq \epsilon$$

for all s_n, $\boldsymbol{x}^n$, s_0 and s_0'. Theorem 4.6.3 of [2] gives an easy-to-check necessary and sufficient condition for a DFSC to be indecomposable: for some fixed n and each $\boldsymbol{x}^n$, there exists a choice for s_n (which may depend on $\boldsymbol{x}^n$) such that

$$\min_{s_0} q(s_n \mid \boldsymbol{x}^n, s_0) > 0. \tag{5}$$

We note here that the channels we consider in the subsequent sections are indecomposable except in very special cases. For these special cases, it can still be shown that $\underline{C} = \overline{C}$ holds.

We make a few comments about DFSCs for which $\underline{C} = \overline{C}$ holds. We denote by C the common value of $\underline{C}$ and $\overline{C}$. This C, which we refer to simply as the *capacity* of the DFSC, can be expressed alternatively. If we assign a probability distribution to the initial state, so that s_0 becomes a random variable, then $C = \lim_{n\to\infty} C_n$, where

$$C_n = \frac{1}{n} \max_{Q^n(\boldsymbol{x}^n)} I(\boldsymbol{x}^n; \boldsymbol{y}^n \mid s_0). \tag{6}$$

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 s_0.

A further simplification to the expression for capacity is possible. Since $|I(\boldsymbol{x}^n; \boldsymbol{y}^n) - I(\boldsymbol{x}^n; \boldsymbol{y}^n \mid s_0)| \leq \log_2 |\mathcal{S}|$ (see, for example, [2, Appendix 4A, Lemma 1]), we in fact have

$$\mathscr{C} = \lim_{n\to\infty} \frac{1}{n} \max_{Q^n(\boldsymbol{x}^n)} I(\boldsymbol{x}^n; \boldsymbol{y}^n). \tag{7}$$

The capacity of a DFSC is difficult to compute in general. A useful lower bound that is sometimes easier to compute (or at least estimate) is the so-called *symmetric information rate* (SIR) of the DFSC:

$$R = \lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{x}^n; \boldsymbol{y}^n), \tag{8}$$

where the input sequence $\boldsymbol{x}$ is an i.i.d. Bernoulli($1/2$) random sequence.

B. First results

It is easy to see that the grains channel is a DFSC, where the nth state s_n is the pair (u_n, x_n), which takes values in the finite set $\mathcal{S} = \{(0,0),(0,1),(1,0),(1,1)\}$. Again, for completeness, we assume an initial state s_0 that takes values in $\mathcal{S}$.[1]

Proposition 1: The grains channel is indecomposable for $p < 1$.

Proof: We must check that the condition in (5) holds. We take $n = 1$ and $s_1 = (0, x_1)$. Then, $\min_{s_0} q(s_1 \mid x_1, s_0) = \min_{j\in\{0,1\}} P(u_1 = 0 \mid u_0 = j) = 1 - p > 0$. ■

Because of this proposition the capacity of the grains channel for the case $p < 1$ is defined via (6). In fact, as the next proposition shows, the equality $\underline{C} = \overline{C}$ also holds for the grains channel when $p = 1$.

Proposition 2: For the grains channel with $p = 1$, we have $\underline{C} = \overline{C} = \frac{1}{2}$.

Proof: We have, with probability 1,

$$\begin{aligned}\boldsymbol{u} &= u_1, u_2, u_3, u_4, u_5, u_6, \ldots \\ &= \begin{cases} 0,1,0,1,0,1,\ldots & \text{if } u_0 = 1 \\ 1,0,1,0,1,0,\ldots & \text{if } u_0 = 0. \end{cases}\end{aligned}$$

Thus, once the initial state $s_0 = (u_0, x_0)$ is fixed, the output $\boldsymbol{y}$ of the grains channel is a deterministic function of the input $\boldsymbol{x}$:

$$\begin{aligned}\boldsymbol{y} &= y_1, y_2, y_3, y_4, y_5, y_6, \ldots \\ &= \begin{cases} x_1, x_1, x_3, x_3, x_5, x_5, \ldots & \text{if } s_0 = (1, x_0) \\ x_0, x_2, x_2, x_4, x_4, x_6, \ldots & \text{if } s_0 = (0, x_0). \end{cases}\end{aligned}$$

Therefore, for any fixed $s \in \mathcal{S}$, we have $H(\boldsymbol{y}^n \mid \boldsymbol{x}^n, s_0 = s) = 0$, and hence, $I(\boldsymbol{x}^n; \boldsymbol{y}^n \mid s_0 = s) = H(\boldsymbol{y}^n \mid s_0 = s)$. If $\boldsymbol{x}^n$ is a sequence of i.i.d. Bernoulli($1/2$) random variables, then $\min_{s\in\mathcal{S}} H(\boldsymbol{y}^n \mid s_0 = s) = H(\boldsymbol{y}^n \mid s_0 = (0, x_0)) = \lfloor n/2 \rfloor$. It follows that $\underline{C}_n \geq \frac{\lfloor n/2 \rfloor}{n}$, so that $\underline{C} \geq 1/2$. On the other hand, for any input distribution $Q^n(\boldsymbol{x}^n)$, and any $s \in \mathcal{S}$, we have $H(\boldsymbol{y}^n \mid s_0 = s) \leq \lceil n/2 \rceil$. Consequently, $\overline{C}_n \leq \frac{\lceil n/2 \rceil}{n}$, and hence, $\overline{C} \leq 1/2$. We conclude that $\underline{C} = \overline{C} = 1/2$. ■

In view of this discussion, the capacity of the grains channel is defined by (7). From here onward we denote this capacity by $\mathscr{C} = \mathscr{C}(p)$ and use $R^{\mathrm{g}}(p)$ to refer to the SIR of the grains channel.

[1] To be strictly faithful to the granular medium we are modeling, we should restrict s_0 to take values only in $\{(1,0),(1,1)\}$, so that $u_0 = 1$. This would imply $u_1 = 0$, meaning that no length-2 grain ends at the first bit cell of the medium, corresponding to physical reality. But this makes no difference to the asymptotics of the channel, and in particular, to the channel capacity.

C. Upper Bound: BINAEras

Consider a binary-input channel similar to the binary erasure channel, except that erasures in consecutive positions are not allowed. Formally, this is a channel with a binary input sequence $\boldsymbol{x} = x_1, x_2, x_3, \ldots$, with $x_i \in \{0, 1\}$ for all i, and a ternary output sequence $\boldsymbol{y} = y_1, y_2, y_3, \ldots$, with $y_i \in \{0, 1, \varepsilon\}$ for all i, where ε is an erasure symbol. The input-output relationship is determined by a binary sequence $\boldsymbol{u} = u_1, u_2, u_3, \ldots$, which is a first-order Markov chain, independent of the input sequence $\boldsymbol{x}$, with transition probabilities $P(u_i|u_{i-1})$ as in (1). We then have

$$y_i = \begin{cases} x_i & \text{if } u_i = 0 \\ \varepsilon & \text{if } u_i = 1 \end{cases} \tag{9}$$

Since $P(u_i = 1 \mid u_{i-1} = 1) = 0$, adjacent erasures do not occur, so we term this channel the binary-input no-adjacent-erasures (BINAEras) channel. To describe the channel completely, we define an initial state z_0 taking values in $\{0, \varepsilon\}$.

The BINAEras channel is a DFSC for which $\underline{C} = \overline{C}$ holds, and its capacity, which we denote by $C^{\varepsilon}(p)$, can be computed explicitly.

Theorem 3: For the BINAEras channel with parameter $p \in [0, 1]$, we have $\underline{C} = \overline{C} = C^{\varepsilon}(p) \triangleq \frac{1}{1+p}$.

Intuitively, the average erasure probability of a symbol equals $\tilde{p} = \frac{p}{1+p}$, and the capacity $C^{\varepsilon}(p)$ equals $1 - \tilde{p}$. A formal proof is given in Appendix A.

We claim that the grains channel is a stochastically degraded BINAEras channel. Indeed, the grains channel is obtained by cascading the BINAEras channel with a ternary-input channel defined as follows: the input sequence $\boldsymbol{y} = y_1, y_2, y_3, \ldots$, $y_i \in \{0, 1, \varepsilon\}$, is transformed to the output sequence $\boldsymbol{y}' = y'_1, y'_2, y'_3, \ldots$ according to the rule

$$y'_i = \begin{cases} y_i & \text{if } y_i \neq \varepsilon \\ y_{i-1} & \text{if } y_i = \varepsilon \end{cases} \tag{10}$$

To cover the case when $y_1 = \varepsilon$, we set y'_1 equal to some arbitrary $y_0 \in \{0, 1\}$. It is straightforward to verify, via (9), (10) and the fact that $P(u_i = 1 \mid u_{i-1} = 1) = 0$, that the cascade of the BINAEras channel with the above channel has an input-output mapping $x_i \mapsto y'_i$ given by (3). This immediately leads to the following theorem.

Theorem 4: For $p \in [0, 1]$, we have $\mathscr{C}(p) \leq C^{\varepsilon}(p) = \frac{1}{1+p}$.

Remark 1. We remark that any code that corrects t nonadjacent substitution errors (bit flips) also corrects t grain errors. It is therefore tempting to bound the capacity of the grains channel by the capacity of the binary channel with *nonadjacent errors*. Such a channel is defined similarly to the BINAEras channel: the channel noise is controlled by a first-order Markov channel $\boldsymbol{u}$ (1), and $y_i = x_i + u_i$ for all $i \geq 1$. The capacity of this channel is computed as in the BINAEras case and equals $1 - \mathrm{h}(p)/(1+p)$, where $\mathrm{h}(p)$ denotes the binary entropy function. However, a closer examination convinces one that this quantity does not provide a valid lower bound for $\mathscr{C}$.

D. Lower Bound: The Symmetric Information Rate

According to the definition of SIR of the grains channel (8), assume that $\boldsymbol{x}$ is an i.i.d. Bernoulli(1/2) random sequence. With this assumption, the state sequence $\boldsymbol{s}$ is a first-order Markov chain. Also, each output symbol y_n is easily verified to be a Bernoulli(1/2) random variable (but y_n is not independent of y_{n-1}).

We also assume that the initial state s_0 is a random variable distributed according to the stationary distribution of the Markov chain, so that the sequence $\boldsymbol{s}$ is a stationary Markov chain. It follows that the output sequence $\boldsymbol{y}$ is a stationary random sequence, so that the entropy rate $H(Y) := \lim_{n\to\infty} \frac{1}{n} H(\boldsymbol{y}^n)$ exists. It is also worth noting here that the initial distribution assumed on s_0 causes the Markov chain $\boldsymbol{u}$ to be stationary as well. In particular, the random variables u_i, $i \geq 0$, all have the stationary distribution given by $P(u_i = 0) = \frac{1}{1+p}$ and $P(u_i = 1) = \frac{p}{1+p}$.

We have

$$R^{\mathrm{g}} = \lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{x}^n; \boldsymbol{y}^n) \tag{11}$$

$$I(\boldsymbol{x}^n; \boldsymbol{y}^n) = H(\boldsymbol{y}^n) - H(\boldsymbol{y}^n|\boldsymbol{x}^n) = H(\boldsymbol{y}^n) - H(\boldsymbol{z}^n|\boldsymbol{x}^n) \tag{12}$$

As noted above, $H(Y) = \lim_{n\to\infty} \frac{1}{n} H(\boldsymbol{y}^n)$ exists. We also have the following lemma.

Lemma 5: Let $\boldsymbol{x}$ be an i.i.d. uniform Bernoulli sequence, then

$$\lim_{n\to\infty} \frac{1}{n} H(\boldsymbol{z}^n \mid \boldsymbol{x}^n) = \sum_{j=2}^{\infty} H(u_j \mid u_1) \left(\frac{1}{2}\right)^j.$$

Proof: We write

$$H(\boldsymbol{z}^n \mid \boldsymbol{x}^n) = \sum_{i=1}^{n} H(z_i \mid z_1, \ldots, z_{i-1}, \boldsymbol{x}^n).$$

From (4), it is evident that z_i is independent of x_j for $j > i$. Hence,

$$H(\boldsymbol{z}^n \mid \boldsymbol{x}^n) = \sum_{i=1}^{n} H(z_i \mid z_1, \ldots, z_{i-1}, \boldsymbol{x}^i).$$

As a result, by the Cesàro mean theorem,

$$\lim_{n\to\infty} \frac{1}{n} H(\boldsymbol{z}^n \mid \boldsymbol{x}^n) = \lim_{i\to\infty} H(z_i \mid z_1, \ldots, z_{i-1}, \boldsymbol{x}^i),$$

provided the latter limit exists.

To evaluate $H(z_i \mid z_1, \ldots, z_{i-1}, \boldsymbol{x}^i)$, we define the events $A_0 = \{\boldsymbol{x}^i : x_i = x_{i-1}\}$,

$$A_j = \{\boldsymbol{x}^i : x_i \neq x_{i-1} = \cdots = x_{i-j} \neq x_{i-j-1}\},\ 1 \leq j \leq i-2,$$

and $A_{i-1} = \{\boldsymbol{x}^i : x_i \neq x_{i-1} = \cdots = x_1\}$. These events partition the space $\{0,1\}^i$ to which $\boldsymbol{x}^i$ belongs. Note that $\Pr[\boldsymbol{x}^i \in A_j] = (\frac{1}{2})^{j+1}$ for $0 \leq j \leq i-2$, and $\Pr[\boldsymbol{x}^i \in A_{i-1}] = (\frac{1}{2})^{i-1}$.

Now, if $\boldsymbol{x}^i \in A_0$, then by (4), we have $z_i = 0$. Consequently, $H(z_i \mid z_1, \ldots, z_{i-1}, \boldsymbol{x}^i \in A_0) = 0$.

If $\boldsymbol{x}^i \in A_j$ for some $j \in [1, i-2]$, then we have $z_i = u_i$, $z_{i-1} = \cdots = z_{i-j+1} = 0$, and $z_{i-j} = u_{i-j}$. Thus,

$$\begin{aligned} & H(z_i \mid z_1, \ldots, z_{i-1}, \boldsymbol{x}^i \in A_j) \\ = \ & H(u_i \mid z_1, \ldots, z_{i-j-1}, u_{i-j}, \boldsymbol{x}^i \in A_j) \\ \overset{(a)}{=} \ & H(u_i \mid u_{i-j}) \ \overset{(b)}{=} \ H(u_{j+1} \mid u_1). \end{aligned}$$

Equality (a) above is due to the fact that $\boldsymbol{u}$ is a first-order Markov chain independent of $\boldsymbol{x}$, while equality (b) is a consequence of the stationarity of $\boldsymbol{u}$ (which is itself a consequence of the stationarity of the state sequence $\boldsymbol{s}$).

Finally, if $\boldsymbol{x}^i \in A_{i-1}$, then $z_i = u_i$ and $z_{i-1} = \cdots = z_2 = 0$. Thus,

$$H(z_i \mid z_1, \ldots, z_{i-1}, \boldsymbol{x}^i \in A_{i-1}) = H(u_i \mid z_1).$$

Therefore,

$$\begin{aligned} & H(z_i \mid z_1, \ldots, z_{i-1}, \boldsymbol{x}^i) \\ = \ & \sum_{j=0}^{i-1} H(z_i \mid z_1, \ldots, z_{i-1}, \boldsymbol{x}^i \in A_j) \Pr[\boldsymbol{x}^i \in A_j] \\ = \ & \sum_{j=1}^{i-2} H(u_{j+1} \mid u_1) \Big(\frac{1}{2}\Big)^{j+1} + H(u_i \mid z_1) \Big(\frac{1}{2}\Big)^{i-1}. \end{aligned}$$

Letting $i \to \infty$, the lemma follows. ■

The infinite series in Lemma 5 can be expressed in a form more useful for explicit computation.

Lemma 6:

$$\sum_{j=2}^{\infty} \Big(\frac{1}{2}\Big)^j H(u_j|u_1) = \frac{1+p/2}{1+p} \sum_{j=2}^{\infty} \Big(\frac{1}{2}\Big)^j \mathrm{h}\Big(\frac{1-(-p)^j}{1+p}\Big). \tag{13}$$

Proof: Note first that $H(u_j \mid u_1) = H(u_j \mid u_1 = 0) \Pr[u_1 = 0] + H(u_j \mid u_1 = 1) \Pr[u_1 = 1]$. Furthermore, since $u_1 = 1$ implies $u_2 = 0$ with probability 1, we have, for all $j \geq 2$,

$$H(u_j \mid u_1 = 1) = H(u_j \mid u_2 = 0) = H(u_{j-1} \mid u_1 = 0),$$

the last equality following from the stationarity of $\boldsymbol{u}$. Hence, $\sum_{j=2}^{\infty} H(u_j \mid u_1 = 1)(\frac{1}{2})^j = \sum_{j=2}^{\infty} H(u_{j-1} \mid u_1 = 0)(\frac{1}{2})^j = (\frac{1}{2}) \sum_{j=2}^{\infty} H(u_j \mid u_1 = 0)(\frac{1}{2})^j$, since $H(u_1 \mid u_1 = 0) = 0$. Putting it all together, we find that

$$\begin{aligned} & \sum_{j=2}^{\infty} H(u_j \mid u_1) \Big(\frac{1}{2}\Big)^j \\ & = \Big(\Pr[u_1 = 0] + \Big(\frac{1}{2}\Big) \Pr[u_1 = 1]\Big) \\ & \quad \times \sum_{j=2}^{\infty} H(u_j \mid u_1 = 0) \Big(\frac{1}{2}\Big)^j \\ & = \frac{1+p/2}{1+p} \sum_{j=2}^{\infty} H(u_j \mid u_1 = 0) \Big(\frac{1}{2}\Big)^j. \end{aligned}$$

The lemma now follows by observing that $H(u_j \mid u_1 = 0) = \mathrm{h}\left(\frac{1-(-p)^j}{1+p}\right)$, as it can be shown (for example, by induction) that $P(u_j = 0 \mid u_1 = 0) = \frac{1-(-p)^j}{1+p}$ for all $j \geq 1$. ■

The quantity $H(Y)$ can also be expressed as an infinite series.

Proposition 7: The entropy rate of the output process of the grains channel is given by

$$H(Y) = \frac{1}{2(1+p)} \sum_{j=2}^{\infty} \mathrm{h}(\beta_j) \prod_{k=2}^{j-1} (1 - \beta_k), \tag{14}$$

where

$$\beta_j := \Pr[y_{j+1} = 1 \mid y_j = y_{j-1} = \cdots = y_2 = 0, y_1 = 1]$$

is given by the following recursion: $\beta_2 = \frac{1}{2}(1-p)$, and for $j \geq 3$,

$$\beta_j = \frac{1}{2} \left(\frac{1 - (1+p)\beta_{j-1}}{1 - \beta_{j-1}} \right). \tag{15}$$

The proof is given in Appendix B.

Remark: An explicit expression for β_j is as follows:

$$\beta_j = \frac{2(\vartheta_-^j - \vartheta_+^j)}{(3+B+p)\vartheta_-^j - (3-B+p)\vartheta_+^j}, \quad j = 2, 3, \ldots,$$

where $\vartheta_\pm = 1 - \frac{1 \mp B}{p}$ and $B = \sqrt{p^2 + 6p + 1}$.

Together, Eq. (12), Lemmas 5–6, and Proposition 7 provide an *exact expression* for the SIR of the grains channel, and hence a lower bound on the capacity $\mathscr{C}$.

Theorem 8: The capacity $\mathscr{C}(p) \geq \max(1/2, R^{\mathrm{g}}(p))$, where R^{g} is the SIR of the grains channel and is given by the following expression:

$$R^{\mathrm{g}}(p) = \frac{1}{2(1+p)} \sum_{j=2}^{\infty} \left\{ \mathrm{h}(\beta_j) \prod_{k=2}^{j-1} (1 - \beta_k) - \frac{2+p}{2^j} \mathrm{h}\left(\frac{1-(-p)^j}{1+p} \right) \right\}.$$

In the following we derive bounds on $H(Y)$ that can provide easier-to-compute bounds on $\mathscr{C}$.

Lemma 9:

$$\frac{p}{1+p} + \frac{1}{1+p} \mathrm{h}\left(\frac{1+p}{2}\right) \leq H(Y) \leq \frac{1}{2(1+p)} \mathrm{h}\left(\frac{1-p}{2}\right) + \frac{1+2p}{2(1+p)} \mathrm{h}\left(\frac{1+p}{2(1+2p)}\right) \tag{16}$$

Proof: We will prove the lower bound because it is directly relevant to our main problem of bounding $\mathscr{C}$, and omit the proof of the upper bound.

We have $\lim_{n\to\infty} \frac{1}{n} \mathrm{H}(\boldsymbol{y}^n \mid s_0) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n H(y_i \mid \boldsymbol{y}^{i-1}, s_0) = \lim_{i\to\infty} H(y_i \mid \boldsymbol{y}^{i-1}, s_0)$. Since conditioning reduces entropy, we have $H(y_i \mid \boldsymbol{y}^{i-1}) \geq H(y_i \mid \boldsymbol{y}^{i-1}, u_{i-1})$. Now, (2) implies that y_i is conditionally independent of $y_1, \ldots, y_{i-2}$ given u_{i-1}. Hence,

$$H(y_i \mid \boldsymbol{y}^{i-1}, u_{i-1}) = H(y_i \mid y_{i-1}, u_{i-1}).$$

We will show that

$$H(y_i \mid y_{i-1}, u_{i-1}) = \frac{p}{1+p} + \frac{1}{1+p} \mathrm{h}\left(\frac{1+p}{2}\right),$$

which will prove the lemma. This is just a somewhat tedious computation.

We start with the identity

$$H(y_i \mid y_{i-1}, u_{i-1}) = \sum_{(a,b)\in\{0,1\}^2} H(y_i \mid y_{i-1} = a, u_{i-1} = b) \times \Pr[y_{i-1} = a, u_{i-1} = b]. \tag{17}$$

Given $u_{i-1} = 1$, we have (with probability 1) $u_i = 0$, so that $y_i = x_i$. Thus,

$$\begin{aligned} H(y_i \mid y_{i-1} = a, u_{i-1} = 1) &= H(x_i \mid y_{i-1} = a, u_{i-1} = 1) \\ &= H(x_i) = 1. \end{aligned}$$

Next, given $u_{i-1} = 0$, we have $y_{i-1} = x_{i-1}$, so that

$$H(y_i \mid y_{i-1} = a, u_{i-1} = 0) = H(y_i \mid x_{i-1} = a, u_{i-1} = 0),$$

and furthermore,

$$\begin{aligned} &\Pr[y_{i-1} = a, u_{i-1} = 0] \\ &\quad = \Pr[u_{i-1} = 0] \Pr[y_{i-1} = a \mid u_{i-1} = 0] \\ &\quad = \Pr[u_{i-1} = 0] \Pr[x_{i-1} = a \mid u_{i-1} = 0] \\ &\quad = \frac{1}{2} \Pr[u_{i-1} = 0]. \end{aligned}$$

Thus, the right-hand side of (17) simplifies to

$$\Pr[u_{i-1} = 1] + \frac{1}{2} \Pr[u_{i-1} = 0] \times \sum_{a\in\{0,1\}} H(y_i \mid x_{i-1} = a, u_{i-1} = 0). \tag{18}$$

To evaluate $H(y_i | x_{i-1} = 0, u_{i-1} = 0)$, we compute

$$\begin{aligned} &\Pr[y_i = 0 | x_{i-1} = 0, u_{i-1} = 0] \\ &\quad = \sum_{j\in\{0,1\}} \Pr[y_i = 0 | x_{i-1} = 0, u_{i-1} = 0, u_i = j] \\ &\qquad \times \Pr[u_i = j | u_{i-1} = 0]. \end{aligned}$$

Note that, given $u_i = 0$, we have $y_i = x_i$, and hence

$$\begin{aligned} &\Pr[y_i = 0 \mid x_{i-1} = 0, u_{i-1} = 0, u_i = 0] \\ &\quad = \Pr[x_i = 0 \mid x_{i-1} = 0, u_{i-1} = 0, u_i = 0] \\ &\quad = \Pr[x_i = 0] = 1/2. \end{aligned}$$

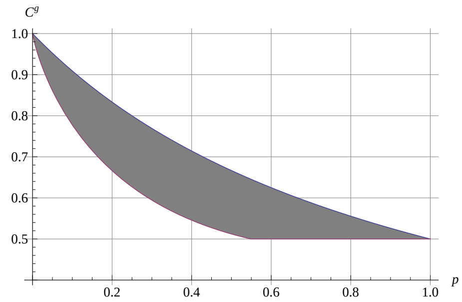
On the other hand, given $u_i = 1$, we have $y_i = x_{i-1}$, from which we obtain

$$\Pr[y_i = 0 \mid x_{i-1} = 0, u_{i-1} = 0, u_i = 1] = 1.$$

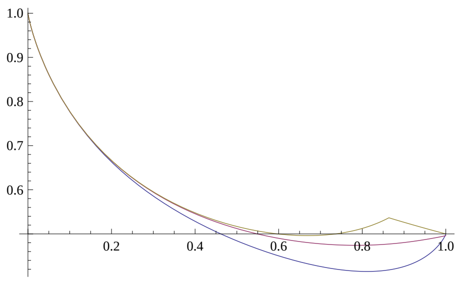
Hence, $\Pr[y_i = 0 \mid x_{i-1} = 0, u_{i-1} = 0] = \frac{1+p}{2}$. An analogous argument shows that $\Pr[y_i = 1 \mid x_{i-1} = 1, u_{i-1} = 0] = \frac{1+p}{2}$. Thus, (18) evaluates to

$$\Pr[u_{i-1} = 1] + \Pr[u_{i-1} = 0] \; \mathrm{h}\left(\frac{1+p}{2}\right). \tag{19}$$

The above expression is precisely the lower bound in the



(a) Bounds on $\mathscr{C}$. The gray area shows the gap between the lower and upper bounds



(b) SIR $R^{\mathrm{g}}(p)$ (13)-(14) and bounds for it (16)

Fig. 1. Bounds on $\mathscr{C}(p)$ and $R^{\mathrm{g}}(p)$ of the grains channel as functions of p.

statement of the lemma. ∎

This lemma together with (11)-(12) and (13) gives the following lower bound on the capacity of the grains channel:

$$\mathscr{C} \geq \frac{p}{1+p} + \frac{1}{1+p}\,\mathrm{h}\left(\frac{1+p}{2}\right) - \frac{1+p/2}{1+p}\sum_{j=2}^{\infty}\mathrm{h}\left(\frac{1-(-p)^j}{1+p}\right)\left(\frac{1}{2}\right)^j.$$

In Figure II-D, we plot the various upper and lower bounds we have for $\mathscr{C}(p)$ and SIR $R^{\mathrm{g}}(p)$ for the grains channel.

E. Zero-Error Capacity

We end with a few remarks on the zero-error capacity of the grains channel. We are interested in the maximum zero-error information rate, $R_0(n)$, achievable over the grains channel with parameter $p \in [0,1]$ and input $\boldsymbol{x}^n$. The case when $p=0$ is trivial (the channel introduces no errors), so we consider $p>0$.

The zero-error analysis depends on the initial state s_0 of the channel. Suppose that s_0 is such that $\Pr[u_1 = 1] > 0$. Then, the state sequence $\boldsymbol{u}^n = 1,0,1,0,\ldots,(n \bmod 2)$ is realized with some positive probability. Corresponding to this state sequence, we have $\boldsymbol{y}^n = x_0, x_2, x_2, x_4, \ldots, x_{2\lfloor n/2\rfloor}$. Thus, at most $\lfloor n/2 \rfloor$ bits can be transmitted without error across this realization of the channel. Hence, $R_0(n) \leq \frac{1}{n}\lfloor n/2\rfloor$. This zero-error information rate can actually be achieved. Consider the binary length-n code

$$\mathcal{R}_n = \{(x_1, x_2, \ldots, x_n) : x_{i-1} = x_i \text{ for all even indices } i\},$$

which has $2^{\lfloor n/2\rfloor}$ codewords. When a codeword from $\mathcal{R}_n$ is sent across *any* realization of the grains channel, the bits at even coordinates remain unchanged. Thus, $\lfloor n/2\rfloor$ bits of information can be transmitted without error, which proves that $R_0(n) = \frac{1}{n}\lfloor n/2\rfloor$.

On the other hand, suppose that the initial state s_0 is such that $\Pr[u_1 = 1] = 0$. Then, the worst-case channel realization is caused by the state sequence $\boldsymbol{u}^n = 0,1,0,1,\ldots,(1+n \bmod 2)$. In this case, the channel is such that the first coordinate of the input sequence is always received without error at the output. A slight modification of the preceding argument now shows that $R_0(n) = \frac{1}{n}\lceil n/2\rceil$.

We have thus proved the following result.

Proposition 10: Consider a grains channel with parameter $p>0$. If the initial state s_0 is such that $\Pr[u_1 = 1] > 0$, then $R_0(n) = \frac{1}{n}\lfloor n/2\rfloor$; otherwise, $R_0(n) = \frac{1}{n}\lceil n/2\rceil$.

In any case, the zero-error capacity of the channel is $C_0 = \lim_{n\to\infty} R_0(n) = \frac{1}{2}$.

Appendix A: Proof of Theorem 3

Observe first that the BINAEras channel is indecomposable for $p<1$. Indeed, for this channel, the condition in (5) reduces to showing that for some fixed n, there exists a choice for u_n such that $\min_{u_0} P(u_n|u_0) > 0$. This condition clearly holds for $n=1$ and $u_1 = 0$: $\min_{j\in\{0,1\}} P(u_1 = 0 \mid u_0 = j) = 1-p > 0$, provided $p<1$. We deal with the indecomposable case in this appendix; when $p=1$, the proof for $\underline{C} = \overline{C} = \frac{1}{2}$ follows, *mutatis mutandis*, the proof of Proposition 2.

When the channel is indecomposable, we have $\underline{C} = \overline{C} = C$. We will show that $C = \frac{1}{1+p}$. Choose the distribution on u_0 to be the stationary distribution of the Markov process $\boldsymbol{u}$, so that $P(u_0 = 0) = \frac{1}{1+p}$ and $P(u_0 = 1) = \frac{p}{1+p}$. Consequently, $\boldsymbol{u}$ is a stationary process, and in particular, for all $i \geq 1$, we have $P(u_i = 0) = \frac{1}{1+p}$ and $P(u_i = 1) = \frac{p}{1+p}$.

Observe that

$$\begin{aligned} I(\boldsymbol{x}^n; \boldsymbol{y}^n \mid u_0) &= H(\boldsymbol{y}^n \mid u_0) - H(\boldsymbol{y}^n \mid \boldsymbol{x}^n, u_0) \\ &\overset{(a)}{=} H(\boldsymbol{y}^n \mid u_0) - H(\boldsymbol{u}^n \mid \boldsymbol{x}^n, u_0) \\ &\overset{(b)}{=} H(\boldsymbol{y}^n \mid u_0) - H(\boldsymbol{u}^n \mid u_0), \end{aligned}$$

with equality (a) above due to the fact that, given $\boldsymbol{x}^n$, the sequences $\boldsymbol{y}^n$ and $\boldsymbol{u}^n$ uniquely determine each other, and equality (b) because $\boldsymbol{u}^n$ is independent of $\boldsymbol{x}^n$. Furthermore, since $\boldsymbol{u}$ is a stationary first-order Markov process, we have $H(\boldsymbol{u}^n \mid u_0) = \sum_{n=1}^{n} H(u_n \mid u_{n-1}) = nH(u_1 \mid u_0) = n\frac{\mathrm{h}(p)}{1+p}$. Hence,

$$C_n = n^{-1}\max_{Q^n(\boldsymbol{x}^n)} H(\boldsymbol{y}^n \mid u_0) - \frac{\mathrm{h}(p)}{1+p}. \tag{20}$$

Now, $H(\boldsymbol{y}^n \mid u_0) = \sum_{i=1}^{n} H(y_i \mid \boldsymbol{y}^{i-1}, u_0)$. Since $\boldsymbol{y}^{i-1}$ completely determines $\boldsymbol{u}^{i-1}$, we have by the data processing inequality [3, Theorem 2.8.1],

$$H(y_i \mid \boldsymbol{y}^{i-1}, u_0) \leq H(y_i \mid \boldsymbol{u}^{i-1}, u_0)$$

We further have

$$H(y_i \mid \boldsymbol{u}^{i-1}, u_0) \le H(y_i \mid u_{i-1})$$
$$= H(y_i \mid u_{i-1}=0)\frac{p}{1+p} + H(y_i \mid u_{i-1}=1)\frac{1}{1+p}$$

Given $u_{i-1}=1$, $\boldsymbol{y}_i$ is a binary random variable (since $u_i = 0$ with probability 1), and thus, $H(y_i \mid u_{i-1}=1) \le 1$. On the other hand, we have $P(y_i = \varepsilon \mid u_{i-1}=0) = P(u_i = 1 \mid u_{i-1}=0) = p$, and so the conditional entropy $H(y_i \mid u_{i-1} = 0)$ is maximized when $P(y_i = 0 \mid u_{i-1}=0) = P(y_i = 1 \mid u_{i-1}=0) = (1-p)/2$. This yields $H(y_i \mid u_{i-1}=1) \le \mathrm{h}(p) + 1 - p$. Putting all the inequalities together, we find that

$$\begin{aligned} H(\boldsymbol{y}^n \mid u_0) &= \sum_{i=1}^{n} H(y_i \mid \boldsymbol{y}^{i-1}, u_0) \\ &\le n\Big(\frac{p}{1+p} + (\mathrm{h}(p)+1-p)\frac{1}{1+p}\Big) \\ &= n\Big(\frac{1+\mathrm{h}(p)}{1+p}\Big) \end{aligned}$$

It is not difficult to check that the above in fact holds with equality when the input sequence $\boldsymbol{x}^n$ is an i.i.d. sequence of Bernoulli(1/2) random variables. Thus,

$$n^{-1} \max_{Q^n(\boldsymbol{x}^n)} H(\boldsymbol{y}^n \mid u_0) = \frac{1+\mathrm{h}(p)}{1+p}.$$

Plugging this into (20), we obtain that $C_n = \frac{1}{1+p}$ for all n, and hence, $C = \frac{1}{1+p}$. ■

APPENDIX B: PROOF OF PROPOSITION 7

To prove Proposition 7, we need to show that $\lim_{i\to\infty} H(y_{i+1} \mid \boldsymbol{y}^i)$ equals the expression in the statement of the proposition. We will work with the identity

$$H(y_{i+1} \mid \boldsymbol{y}^i) = \sum_{\boldsymbol{b}\in\{0,1\}^i} H(y_{i+1} \mid \boldsymbol{y}^i = \boldsymbol{b}) \Pr[\boldsymbol{y}^i = \boldsymbol{b}].$$

From the channel input-output relationship given by (3) and the fact that the input $\boldsymbol{x}$ is an i.i.d. Bernoulli($\frac{1}{2}$) sequence, it is clear that $\Pr[\boldsymbol{y}^i = \boldsymbol{b}] = \Pr[\boldsymbol{y}^i = \bar{\boldsymbol{b}}]$, where $\bar{\boldsymbol{b}} = \boldsymbol{b} + 1^n$ is the sequence obtained by flipping each bit in $\boldsymbol{b}$. It then also follows that $H(y_{i+1} \mid \boldsymbol{y}^i = \boldsymbol{b}) = H(y_{i+1} \mid \boldsymbol{y}^i = \bar{\boldsymbol{b}})$, since $\Pr[y_{i+1} = 1 \mid \boldsymbol{y}^i = \boldsymbol{b}] = \Pr[y_{i+1} = 0 \mid \boldsymbol{y}^i = \bar{\boldsymbol{b}}]$. Hence,

$$H(y_{i+1} \mid \boldsymbol{y}^i) = 2\sum_{\boldsymbol{b}\in B} H(y_{i+1} \mid \boldsymbol{y}^i = \boldsymbol{b}) \Pr[\boldsymbol{y}^i = \boldsymbol{b}], \quad (21)$$

where $B = \{(b_i, \dots, b_1) \in \{0,1\}^i : b_i = 0\}$ is the set of all binary length-i sequences that have a 0 in the leftmost coordinate.

Fix $i \ge 2$. Define, for $2 \le j \le i$, the events

$$B_j = \{\boldsymbol{y}^i : (y_i, y_{i-1}, \dots, y_{i-j+1}) = 0^{j-1}1\},$$

which, together with the event $\{\boldsymbol{y}^i = 0^i\}$, form a partition of B. Here, $0^{j-1}1$ is shorthand for the j-tuple $(0,\dots,0,1)$. We record two facts about B_j. First,

$$\begin{aligned} \Pr[\boldsymbol{y}^i \in B_j] &= \Pr[(y_i, y_{i-1}, \dots, y_{i-j+1}) = 0^{j-1}1] \\ &= \Pr[(y_j, y_{j-1}, \dots, y_1) = 0^{j-1}1], \end{aligned} \quad (22)$$

the last equality stemming from the fact that $\boldsymbol{y}$ is stationary. Second, by the following lemma,

$$H(y_{i+1} \mid \boldsymbol{y}^i = \boldsymbol{b}) = \mathrm{h}(\Pr[y_{i+1} = 1 \mid \boldsymbol{y}^i = \boldsymbol{b}]) \quad (23)$$

is invariant over B_j.

Lemma 11: For $\boldsymbol{b} \in B_j$, $\Pr[y_{i+1} = 1 \mid \boldsymbol{y}^i = \boldsymbol{b}]$ equals

$$\frac{1}{2}\Pr[u_j = 0 \mid (y_{j-1}, y_{j-2}, \dots, y_2) = 0^{j-2},\ (u_1, x_1) = (0,0)].$$

Proof: The proof relies upon the following claim :

Suppose that $y_{k-1} = b$; then, with probability 1, we have $y_k = \bar{b}$ if and only if $s_k := (u_k, x_k) = (0, \bar{b})$.

Indeed, even without the assumption on y_{k-1}, the "if" part holds trivially. For the "only if" part, assume that $y_{k-1} = b$ and $y_k = \bar{b}$. Note that if $u_k = 1$, then with probability 1, we have $u_{k-1} = 0$. Hence, by way of (3), we have $y_k = x_{k-1} = y_{k-1}$. However, $y_{k-1} \ne y_k$ by assumption; so we must have $u_k = 0$. Consequently, $y_k = x_k$, so that $x_k = \bar{b}$.

Consider any $\boldsymbol{b} \in B_j$. From the claim, we have

$$\begin{aligned} \Pr[y_{i+1} = 1 \mid \boldsymbol{y}^i = \boldsymbol{b}] &= \Pr[(u_{i+1}, x_{i+1}) = (0,1) \mid \boldsymbol{y}^i = \boldsymbol{b}] \\ &= \frac{1}{2}\Pr[u_{i+1} = 0 \mid \boldsymbol{y}^i = \boldsymbol{b}], \end{aligned}$$

where we have used the fact that x_{i+1} is independent of $\boldsymbol{y}^i$. Note that, in the event $\boldsymbol{y}^i = \boldsymbol{b}$, we have $y_{i-j+2} = 0$ and $y_{i-j+1} = 1$, so that by the claim again,

$$\Pr[u_{i+1} = 0 \mid \boldsymbol{y}^i = \boldsymbol{b}] = \Pr[u_{i+1} = 0 \mid \boldsymbol{y}^i = \boldsymbol{b}, s_{i-j+2} = (0,0)].$$

Now, given the channel state $s_{i-j+2} = (0,0)$, the random variables u_{i+1}, y_i, $y_{i-1}, \dots, y_{i-j+2}$ are conditionally independent of the past output $\boldsymbol{y}^{i-j+1}$. Furthermore, given $s_{i-j+2} = (0,0)$, the random variable y_{i-j+2} is uniquely determined: $y_{i-j+2} = 0$. Hence,

$$\begin{aligned} &\Pr[u_{i+1} = 0 \mid \boldsymbol{y}^i = \boldsymbol{b}, s_{i-j+2} = 0] = \\ &\qquad \Pr[u_{i+1} = 0 \mid (y_i, \dots, y_{i-j+3}) = 0^{j-2},\ s_{i-j+2} = 0]. \end{aligned}$$

Finally, by the joint stationarity of $\boldsymbol{y}$ and $\boldsymbol{u}$, the right-hand side above is equal to

$$\Pr[u_j = 0 \mid (y_{j-1}, y_{j-2}, \dots, y_2) = 0^{j-2},\ s_1 = 0],$$

which is what we needed to show. ■

In the statement of Proposition 7, we defined $\beta_j = \Pr[y_{j+1} = 1 \mid (y_j, y_{j-1}, \dots, y_1) = 0^{j-1}1]$. Note that if we set $i = j$ in Lemma 11, we get

$$\beta_j = \frac{1}{2}\Pr[u_j = 0 \mid (y_{j-1}, \dots, y_2) = 0^{j-2},\ (u_1, x_1) = (0,0)]. \quad (24)$$

From (21)–(24), and Lemma 11, we have

$$\begin{aligned} H(y_{i+1} \mid \boldsymbol{y}^i) &= 2\sum_{j=2}^{i} \mathrm{h}(\beta_j)\Pr[(y_j,\ldots,y_1) = 0^{j-1}1] \\ &\quad + 2H(y_{i+1} \mid \boldsymbol{y}^i = 0^i)\Pr[\boldsymbol{y}^i = 0^i]. \end{aligned}$$

Letting $i \to \infty$, it is easy to verify that the term at the end of the above expression vanishes, so that

$$H(Y) = 2\sum_{j=2}^{\infty} \mathrm{h}(\beta_j)\Pr[(y_j, y_{j-1},\ldots,y_1) = 0^{j-1}1]. \quad (25)$$

The proof of Proposition 7 will be complete once we prove the next two lemmas.

Lemma 12: For $j \geq 2$, we have

$$\Pr[(y_j, y_{j-1},\ldots,y_1) = 0^{j-1}1] = \frac{1}{4(1+p)}\prod_{k=2}^{j-1}(1-\beta_k)$$

Proof: From the definition of β_j, we readily obtain

$$\Pr[(y_j,\ldots,y_1) = 0^{j-1}1] = \left[\prod_{k=2}^{j-1}(1-\beta_k)\right]\cdot\Pr[(y_2,y_1) = (0,1)].$$

We must show that $\Pr[(y_2,y_1) = (0,1)] = \frac{1}{4(1+p)}$. We skip the details of this elementary calculation. ∎

Lemma 13: $\beta_2 = \frac{1}{2}(1-p)$, and for $j \geq 3$, β_j satisfies the recursion in (15).

Proof: From (24), we have

$$\begin{aligned} \beta_2 &= \frac{1}{2}\Pr[u_2 = 0 \mid (u_1,x_1) = (0,0)] \\ &= \frac{1}{2}\Pr[u_2 = 0 \mid u_1 = 0] = \frac{1}{2}(1-p). \end{aligned}$$

For convenience, define, for $j \geq 2$, $E_j = \{(y_{j-1}, y_{j-2},\ldots,y_2) = 0^{j-2}, (u_1,x_1) = (0,0)\}$, so that $\beta_j = (\frac{1}{2})\Pr[u_j = 0 \mid E_j] = (\frac{1}{2})(1-\gamma_j)$, where $\gamma_j := \Pr[u_j = 0 \mid E_j]$. We shall show that for $j \geq 3$,

$$\gamma_j = \frac{p(1-\gamma_{j-1})}{1+\gamma_{j-1}}. \quad (26)$$

which is equivalent to the recursion in (15).

So, let $j \geq 3$ be fixed. We start with

$$\begin{aligned} \gamma_j &= \sum_{b\in\{0,1\}} \Pr[u_j = 1 \mid u_{j-1} = b]\,\Pr[u_{j-1} = b \mid E_j] \\ &= p\cdot\Pr[u_{j-1} = 0 \mid E_j] \\ &= p\cdot\Pr[u_{j-1} = 0 \mid y_{j-1} = 0, E_{j-1}] \\ &= p\cdot\frac{\Pr[y_{j-1} = 0 \mid u_{j-1} = 0, E_{j-1}](1-\gamma_{j-1})}{\Pr[y_{j-1} = 0 \mid E_{j-1}]} \end{aligned}$$

where we have used $\Pr[u_{j-1} = 0 \mid E_{j-1}] = 1-\gamma_{j-1}$ for the last equality.

Given $u_{j-1} = 0$, we have $y_{j-1} = x_{j-1}$, and since x_{j-1} is independent of u_{j-1} and E_{j-1}, the numerator in the last expression above evaluates to $\frac{1}{2}(1-\gamma_{j-1})$. Thus,

$$\gamma_j = p\cdot\frac{\frac{1}{2}(1-\gamma_{j-1})}{\Pr[y_{j-1} = 0 \mid E_{j-1}]} \quad (27)$$

Turning to the denominator, we write $\Pr[y_{j-1} = 0 \mid E_{j-1}]$ as

$$\begin{aligned} &\sum_{b\in\{0,1\}} \Pr[y_{j-1} = 0 \mid u_{j-1} = b, E_{j-1}]\,\Pr[u_{j-1} = b \mid E_{j-1}] \\ &= \tfrac{1}{2}(1-\gamma_{j-1}) + \Pr[y_{j-1} = 0 \mid u_{j-1} = 1, E_{j-1}]\cdot\gamma_{j-1} \end{aligned} \quad (28)$$

We claim that $\Pr[y_{j-1} = 0 \mid u_{j-1} = 1, E_{j-1}] = 1$. Indeed, given $u_{j-1} = 1$, we have $y_{j-1} = x_{j-2}$. Furthermore, we must have $u_{j-2} = 0$ with probability 1, so that $x_{j-2} = y_{j-2}$. Thus, given $u_{j-1} = 1$, we must have $y_{j-1} = y_{j-2}$ with probability 1. But note that the event E_{j-1} implies $y_{j-2} = 0$: if $j = 3$, this follows from $(u_1,x_1) = (0,0)$, and if $j \geq 4$, this is contained within $(y_{j-2},\ldots,y_2) = 0^{j-3}$. Thus, given $u_{j-1} = 1$ and E_{j-1}, we have $y_{j-1} = y_{j-2} = 0$ with probability 1.

So, carrying on from (28), we get

$$\Pr[y_{j-1} = 0 \mid E_{j-1}] = \frac{1}{2}(1-\gamma_{j-1}) + \gamma_{j-1} = \frac{1}{2}(1+\gamma_{j-1})$$

Feeding this back into (27), we obtain

$$\gamma_j = p\cdot\frac{\frac{1}{2}(1-\gamma_{j-1})}{\frac{1}{2}(1+\gamma_{j-1})}$$

which is the desired recursion (26). ∎

This concludes the proof of Proposition 7.

ACKNOWLEDGEMENT: The work of Arya Mazumdar and Alexander Barg was supported by NSF grants CCF0830699 and CCF0916919. The work of Navin Kashyap was supported in part by a Discovery Grant from NSERC, Canada, and was performed while on sabbatical at the University of Maryland, College Park, and the Indian Institute of Science, Bangalore. The authors would like to thank Haim Permuter for helpful discussions.

REFERENCES

[1] A. Mazumdar, A. Barg, N. Kashyap, “Coding for high density magnetic recording,” *Proc. International Symposium on Information Theory*, Austin, TX, USA, Jun 13–18, 2010, pp. 978–982.

[2] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons, 1968.

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

[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] 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.