Coding for High-Density Recording on a 1-D Granular Magnetic Medium

Arya Mazumdar, Alexander Barg, Navin Kashyap

***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. We consider a simple one-dimensional combinatorial model of this medium. In our model, we assume a setting where binary data is sequentially written on the medium and a bit can erroneously change to the immediately preceding value. We derive several properties of codes that correct this type of errors, focusing on bounds on their cardinality.**

We also define a probabilistic finite-state channel model of the storage medium, and derive lower and upper estimates of its capacity. A lower bound is derived by evaluating the symmetric capacity of the channel, i.e., the maximum transmission rate under the assumption of the uniform input distribution 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 can compute explicitly.

I. Introduction

One of the challenges in achieving ultra-high-density magnetic recording lies in accounting for the effect of the granularity of the recording medium. Conventional magnetic recording media are composed of fundamental magnetizable units, called "grains", that do not have a fixed size or shape. Information is stored on the medium through a write mechanism that sets the magnetic polarities of the grains [9]. There are two types of magnetic polarity, and each grain can be magnetized to take on exactly one of these two polarities. Thus, each grain can store at most one bit of information. Clearly, if the boundaries of the grains were known to the write mechanism and the readback mechanism, then it would be theoretically possible to achieve a storage capacity of one information bit per grain.

There are two bottlenecks to achieving the one-bit-per-grain storage capacity: (i) the existing write (and readback) technologies are not capable of setting (and reading back) the magnetic polarities of a region as small as a single grain; and (ii) the write and readback mechanisms are typically unaware of the shapes and positions of the grains in the medium. In current magnetic recording technologies, writing is generally done by dividing the magnetic medium into regularly-spaced bit cells, and writing one bit of data into each of these bit cells. The bit cells are much larger in size compared to the grains, so that each bit cell comprises many grains. Writing a bit into a bit cell is then a matter of uniformly magnetizing all the grains within the cell; the effect of grains straddling the boundary between two bit cells can be neglected.

Recently, Wood et al. [8] 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.

The authors of [8] went on to consider the information loss caused by the lack of knowledge of grain boundaries. A sample simulation considered a two-dimensional magnetic medium composed of 100 randomly shaped grains, and subdivided into a 14×14 grid of uniformly-sized bit cells. Bits were written in raster-scan fashion onto the grid. At the kth step of the write process, if any grain had more than a 30% (in area) overlap with the bit cell to be written at that step, then that grain was given the polarity value of the kth bit. The polarity of a grain could switch multiple times before settling on a final value. With a readback mechanism that reported the polarity value at the centre of each bit cell, their simulation recorded the proportion of bits that were reported with the wrong polarity. A similar simulation, but with a slightly different assumption on the underlying grain distribution, was reported in [6].

The authors of [8] also considered a simple channel that modeled a one-dimensional granular medium, and computed a lower bound on the capacity of the channel. The one-dimensional medium was divided into regularly-spaced bit cells, and it was assumed that grain boundaries coincided with bit cell boundaries, and that the grains had randomly selected lengths equal to 1, 2 or 3 bit cells. The polarity of a grain is set by the last bit to be written within it. The effect of this is that the last bit to be written in the grain *overwrites* all bits previously written within the same grain.

In this paper, we restrict ourselves to the one-dimensional case, and consider a combinatorial error model that corresponds to the granular medium described above. The medium comprises n bit cells, indexed by the integers from 1 to n. The

Arya Mazumdar is with the Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742 (e-mail: arya@umd.edu).

Alexander Barg is with the Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742 and Institute for Problems of Information Transmission, Moscow, Russia (e-mail: abarg@umd.edu).

Navin Kashyap is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L3N6, Canada (e-mail: nkashyap@mast.queensu.ca).

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 results of this paper were presented in part at the 2010 IEEE International Symposium on Information Theory, Austin, TX, June 13-18, 2010, and at the 48th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 29–October 1, 2010.

granular structure of the medium is described by an increasing sequence of positive integers, $1 = j_1 < j_2 < \cdots < j_s \leq n$, 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).

The effect of a given grain pattern on an n-bit block of binary data $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ to be written onto the medium is represented by an operator ϕ that acts upon $\boldsymbol{x}$ to produce $\phi(\boldsymbol{x}) = (y_1, y_2, \ldots, y_n)$, which is the binary vector that is actually recorded on the medium. 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 $j_i \leq j < j_{i+1}$, 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 $y_j \neq x_j$) precisely when $x_j \neq x_{j_i}$ for some j satisfying $j_i < j < j_{i+1}$. In particular, grains of length 1 do not introduce any errors.

As an example, consider a medium divided into 15 bit cells, with a granular structure consisting of grains of lengths 1 and 2 only, with the length-2 grains beginning at indices 3, 6, 8 and 13. The grains in the medium would transform the vector $\boldsymbol{x} = (100001000010000)$ to (100001100010000) and the vector $\boldsymbol{x} = (000101011100010)$ to $\phi(\boldsymbol{x}) = (000001111100000)$. Note that $\phi(\boldsymbol{x}) \neq \boldsymbol{x}$ iff a 01 or a 10 falls within some grain. In particular, $\phi(\phi(\boldsymbol{x})) = \phi(\boldsymbol{x})$ for any $\boldsymbol{x}$.

In this paper, we consider only the case of granular media composed of grains of length at most 2. Even this simplest possible case brings out the complexity of the problem of coding to correct errors caused by this combinatorial model. Most of the results we present can be extended straightforwardly to the case of magnetic media with a more general grain distribution.

Note that in a medium with grains of length at most 2, it is precisely the length-2 grains that can cause bit errors. We denote by $\Phi_{n,t}$ the set of operators ϕ corresponding to all such media with n bit cells and at most t grains of length equal to 2. Then, for $\boldsymbol{x} \in \{0,1\}^n$, we let $\Phi_{n,t}(\boldsymbol{x}) = \{\phi(\boldsymbol{x}) : \phi \in \Phi_{n,t}\}$, and call two vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \{0,1\}^n$ *t-confusable* if

$$\Phi_{n,t}(\boldsymbol{x}_1) \cap \Phi_{n,t}(\boldsymbol{x}_2) \neq \emptyset.$$

A binary code $\mathcal{C}$ of length n is said to correct t grain errors if no two distinct vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C}$ are t-confusable.

In Sections II and III of this paper, we study properties of t-grain-correcting codes. We derive several bounds on the maximum size of a length-n binary code that corrects t grain errors. Our lower bounds are based on either explicit constructions or existence arguments, while our upper bounds are based on the count of runs of identical symbols in a vector or on a clique partition of the "confusability graph" of the space $\{0,1\}^n$. We also briefly consider list-decodable grain-correcting codes, and derive a lower bound on the maximum cardinality of such codes by means of a probabilistic argument.

In Section IV, we consider a scenario in which the locations of the grains are available to either the encoder or the decoder of the data, and derive estimates of the size of codes in this setting.

In Section V, 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 estimate the capacity of the channel. For a lower bound on the capacity we restrict our attention to uniformly distributed, independent input letters which corresponds to the case of *symmetric information rate* (symmetric capacity or SIR) 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 true capacity. 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. We explicitly compute the capacity of this erasure channel, and observe that the grains channel is a stochastically degraded version of the erasure channel. The capacity of the erasure channel is thus 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] which contains some of our results from Section V. The authors of [4] considered a more general channel model that includes our probabilistic model of the grains channel as a particular case. Their paper contains results that cover our Propositions 12 and 19, as well as our Theorem 14. However, a major contribution of ours that cannot be found in [4] is our Theorem 17, in which we give an exact expression for the SIR of the grains channel.

Throughout the paper, $h(x) = -x\log_2 x - (1-x)\log_2(1-x)$ denotes the binary entropy function.

II. Constructions of grain-correcting codes

As observed above, when the length of the grains does not exceed 2, bit errors are caused only by length-2 grains. Furthermore, it can only be the second bit within such a grain that can be in error. Thus, any code that can correct t bit-flip errors (equivalently, a code with minimum Hamming distance at least 2t+1) is a t-grain-correcting code. In particular, t-grain-correcting codes whose parameters meet the Gilbert-Varshamov bound (see e.g. [7, p. 97]) are guaranteed to exist. But we can sometimes do better than conventional error-correcting codes by taking advantage of the special nature of grain errors.

Observe that the first bit to be written onto the medium can never be in error in the grain model. So, we can construct t-grain-correcting codes $\mathcal{C}$ of length n as follows: take a code $\mathcal{C}'$ of length $n-1$ that can correct t bit-flip errors, and set $\mathcal{C} = (0|\mathcal{C}') \cup (1|\mathcal{C}')$. Here, for $b \in \{0,1\}$, $(b|\mathcal{C}')$ refers to the set of vectors obtained by prefixing b to each codevector of $\mathcal{C}'$. For example, when $n = 2^m$, we can take $\mathcal{C}'$ to be the binary Hamming code of length $2^m - 1$, yielding a 1-grain-correcting code $\mathcal{C}$ of size $|\mathcal{C}| = 2^n/n$. Note that $2^n/n$ exceeds the sphere-packing (Hamming) upper bound, i.e., is greater than the cardinality of the optimal binary single-error-correcting code of length $n = 2^m$.

More generally, again when n is a power of 2, we can take $\mathcal{C}'$ to be a binary BCH code of length $n-1$ that corrects t bit-flip errors. The above construction then yields a t-grain-correcting code $\mathcal{C}$ of length n and size $|\mathcal{C}| \geq 2^n/n^t$.

We next describe a completely different, and remarkably simple, construction of a length-n grain-correcting code that corrects *any* number of grain errors. For even integers $n = 2m$, $m \geq 1$, define the code $\mathcal{R}_n \subset \{0,1\}^n$ as the set

$$\{(x_1 x_2 \dots x_{2m}) \in \{0,1\}^n : x_{i-1} = x_i \text{ for all even indices } i\}. \quad (1)$$

Note that when a codevector from $\mathcal{R}_n$ is written onto a medium composed of grains of length at most 2, the bits at even coordinates remain unchanged. Indeed, a bit at an even index i could be in error only if a grain starts at index $i-1$, causing the bit at index $i-1$ to overwrite the bit at index i. However, the two bits are identical by construction. Thus, $\mathcal{R}_n$ is a code of size $2^{n/2}$ that corrects an arbitrary number of grain errors. This construction can be extended to odd lengths $n = 2m+1$, $m \geq 1$, as follows: $\mathcal{R}_n = (0|\mathcal{R}_{2m}) \cup (1|\mathcal{R}_{2m})$.

III. Bounds on the Size of Grain-Correcting Codes

Let $M(n,t)$ denote the maximum size of a length-n binary code that is t-grain-correcting. The constructions of the previous section show that $M(n,t) \geq 2^{\lceil n/2 \rceil}$ for any n and t, and $M(n,t) \geq 2^n/n^t$ when n is a power of 2. In an attempt to determine the tightness of these lower bounds, we derive below some upper bounds on $M(n,t)$.

A. The One-Bit-Per-Grain Upper Bound

We start with the simplest upper bound, which is based on the fact that each grain in the medium can store at most one bit of information.

Proposition 1 *For $n \geq 2t$, we have $M(n,t) \leq 2^{n-t}$.*

Proof : Consider a medium with exactly t grains of length 2. This medium has $n - 2t$ grains of length 1, so that the total number of grains is $n - t$. Since each grain can hold at most one bit of information, we have $M(n,t) \leq 2^{n-t}$. ∎

The above bound is tight for $n \in \{2t, 2t+1\}$. Indeed, for these values of n, the upper bound evaluates to $2^{\lceil n/2 \rceil}$, which is achieved by the $\mathcal{R}_n$ construction of the previous section. When $n > 2t+1$, bounds better than 2^{n-t} can be obtained using the techniques of Sections III-B and III-C.

B. Upper Bounds Based on Counts of Runs

Denote by $r(\boldsymbol{x})$ the number of runs (maximal subvectors of consecutive identical symbols) in the vector $\boldsymbol{x} \in \{0,1\}^n$. As remarked in Section I, a single grain can change $\boldsymbol{x}$ to a different vector if and only if the grain straddles the boundary between two successive runs in $\boldsymbol{x}$. Thus, $|\Phi_{n,1}(\boldsymbol{x})| = 1 + (r(\boldsymbol{x}) - 1) = r(\boldsymbol{x})$. For $t \geq 2$, the number $|\Phi_{n,t}(\boldsymbol{x})|$ is not readily expressible in a closed form. Nevertheless, we have the following lemma.

Lemma 2

$$|\Phi_{n,t}(\boldsymbol{x})| \geq 1 + \sum_{i=1}^{t} \frac{1}{i!} \prod_{j=0}^{i-1} (r(\boldsymbol{x}) - 1 - 3j).$$

Proof : The right-hand side is a worst-case count of the number of ways in which $i \leq t$ length-2 grains can be placed so that each grain straddles the boundary between successive runs in $\boldsymbol{x}$. The first grain can be placed in $r(\boldsymbol{x}) - 1$ ways; after that, in the worst case (which happens when the first grain falls in the middle of a 1010 or 0101), the next grain can be placed in $(r(\boldsymbol{x}) - 1) - 3$ ways; and so on. ∎

This leads to the following upper bound on $M(n,t)$.

Theorem 3 *For any fixed value of t,*

$$M(n,t) \leq \frac{2^n}{n^t} (t!\, 2^t + 2 + o(1)),$$

where $o(1)$ denotes a term that goes to 0 as $n \to \infty$.

Proof : Let $\mathcal{C}$ be a t-grain-correcting code of length n, and let

$$\mathcal{C}_1 = \left\{\boldsymbol{x} \in \mathcal{C} : |r(\boldsymbol{x}) - n/2| \leq \sqrt{nt \log_2 n}\right\}.$$

For any $\boldsymbol{x} \in \mathcal{C}_1$, we have from Lemma 2,

$$\begin{aligned} |\Phi_{n,t}(\boldsymbol{x})| &\geq \frac{1}{t!}(r(\boldsymbol{x}) - 1 - 3(t-1))^t \\ &\geq \frac{1}{t!}(n/2 - \sqrt{nt \log_2 n} - 1 - 3(t-1))^t \quad (2) \end{aligned}$$

Since $\mathcal{C}_1$ itself is t-grain-correcting, we also have

$$2^n \geq \Big| \bigcup_{\boldsymbol{x} \in \mathcal{C}_1} \Phi_{n,t}(\boldsymbol{x}) \Big| = \sum_{\boldsymbol{x} \in \mathcal{C}_1} |\Phi_{n,t}(\boldsymbol{x})|. \quad (3)$$

It follows from (2) and (3) that

$$|\mathcal{C}_1| \leq \frac{2^{n+t} t!}{n^t} (1 + o(1)).$$

Now, let $\mathcal{C}_2 = \mathcal{C} \backslash \mathcal{C}_1$. We shall bound from above the size of $\mathcal{C}_2$ by the number of vectors $\boldsymbol{x} \in \{0,1\}^n$ such that $|r(\boldsymbol{x}) - n/2| \geq \sqrt{nt \log_2 n}$. Define $\psi : \{0,1\}^n \to \{0,1\}^{n-1}$ by setting

$$\psi((x_1, x_2, \dots, x_n)) = (x_1 \oplus x_2, x_2 \oplus x_3, \dots, x_{n-1} \oplus x_n)$$

where $\oplus$ denotes modulo-2 addition. Then, $r(\boldsymbol{x}) = w_H(\psi(\boldsymbol{x})) + 1$, where $w_H(\cdot)$ denotes Hamming weight. For any given vector $\boldsymbol{y} \in \{0,1\}^{n-1}$, there are exactly two vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 = \mathbf{1} \oplus \boldsymbol{x}_1$ such that $\psi(\boldsymbol{x}_1) = \psi(\boldsymbol{x}_2) = \boldsymbol{y}$. Therefore,

$$\begin{aligned} |\mathcal{C}_2| &\leq 2|\{\boldsymbol{y} \in \mathbb{F}_2^{n-1} : |w_H(\boldsymbol{y}) + 1 - n/2| \geq \sqrt{nt \log_2 n}\}| \\ &\leq 4 \sum_{i=0}^{n/2 - \sqrt{nt \log_2 n}} \binom{n-1}{i} \\ &\leq 4 \exp\left\{(n-1) h\Big(\frac{1}{2} - \frac{2\sqrt{nt \log_2 n} - 1}{2(n-1)}\Big)\right\}, \end{aligned}$$

where $h(z) = -z \log_2 z - (1-z) \log_2 (1-z)$ is the binary entropy function. Since $h(\frac{1}{2} - x) \leq 1 - \frac{2}{\ln 2} x^2$,

$$\begin{aligned} |\mathcal{C}_2| &\leq 4 \exp\left\{(n-1) - \frac{2}{\ln 2} \frac{(2\sqrt{nt \log_2 n} - 1)^2}{4(n-1)}\right\} \\ &\leq 2^{n+1} n^{-t}. \end{aligned}$$

We conclude by noting that $|\mathcal{C}| = |\mathcal{C}_1| + |\mathcal{C}_2|$. ∎

For fixed t, the upper bound of the above theorem is within a constant multiple of the lower bound $M(n,t) \geq 2^n/n^t$, stated earlier as being valid when n is a power of 2.

The bound of Theorem 3 is not useful when t grows linearly with n, say, $t = n\tau$ for $\tau \in (0, 1/2]$. In this case, we define

$$\overline{R}(\tau) = \limsup_{n \to \infty} \frac{\log_2 M(n, \lfloor n\tau \rfloor)}{n}. \quad (4)$$

A simple consequence of Proposition 1 is that $\overline{R}(\tau) \le 1-\tau$ for $\tau \in (0, 1/2]$. A better upper bound on $\overline{R}(\tau)$ for small τ can be established by an argument similar to the proof of Theorem 3.

Proposition 4 *Let $x^* = x^*(\tau)$ be the smallest positive solution of the following equation:*

$$h\Big(\frac{1-x}{2}\Big) + \frac{1-x}{4} h\Big(\frac{4\tau}{1-x}\Big) = 1.$$

For $\tau \le 0.0706$, the following bound holds true:

$$\overline{R}(\tau) \le h\Big(\frac{1-x^*}{2}\Big). \tag{5}$$

Proof : The proof relies on a coarser estimate of $|\Phi_{n,t}(\boldsymbol{x})|$ than the one in Lemma 2. Consider the boundaries between the $(2i-1)$-th and $2i$-th runs in $\boldsymbol{x}$, $i = 1, 2, \ldots, \lfloor r(\boldsymbol{x})/2 \rfloor$. Length-2 grains can be independently placed across these boundaries, leading to the lower bound

$$|\Phi_{n,t}(\boldsymbol{x})| \ge \sum_{i=0}^{t} \binom{\lfloor r(\boldsymbol{x})/2 \rfloor}{i}. \tag{6}$$

For $t = \lfloor \tau n \rfloor$, let $\mathcal{C}$ be a t-grain-correcting code. For some $\delta > 0$, let

$$\mathcal{C}_1 = \Big\{ \boldsymbol{x} \in \mathcal{C} : r(\boldsymbol{x})/2 \ge \Big\lfloor \frac{n}{4}(1-\delta) \Big\rfloor \Big\}$$

The bound (6) implies that for each $\boldsymbol{x} \in \mathcal{C}_1$

$$|\Phi_{n,t}(\boldsymbol{x})| \ge \sum_{i=0}^{t} \binom{\lfloor \frac{n}{4}(1-\delta) \rfloor}{i}.$$

From the above and (3), we obtain

$$|\mathcal{C}_1| \le \frac{2^n}{\sum_{i=0}^{t} \binom{\lfloor \frac{n}{4}(1-\delta) \rfloor}{i}}.$$

The size of the remaining subset of vectors $\mathcal{C}_2 = \mathcal{C} \backslash \mathcal{C}_1$ does not exceed the number of all vectors $\boldsymbol{x}$ with $r(\boldsymbol{x}) \le \frac{n}{2}(1-\delta)$, i.e.,

$$|\mathcal{C}_2| \le \sum_{i=0}^{\lfloor \frac{n}{2}(1-\delta) \rfloor} \binom{n-1}{i} \le 2^{nh(\frac{1-\delta}{2})}.$$

Therefore,

$$|\mathcal{C}| \le \min_{\delta > 0} \Big\{ \frac{2^n}{\sum_{i=0}^{t} \binom{\lfloor \frac{n}{4}(1-\delta) \rfloor}{i}} + 2^{nh(\frac{1-\delta}{2})} \Big\}.$$

When $\tau \le \frac{1-\delta}{8}$, or equivalently, $\delta \le 1 - 8\tau$, the dominant term in the sum in the denominator above is $\binom{\lfloor \frac{n}{4}(1-\delta) \rfloor}{\lfloor \tau n \rfloor}$, which is bounded below by $\frac{1}{\sqrt{8n}} 2^{\frac{n(1-\delta)}{4} h(\frac{4\tau}{1-\delta})}$. From this, we obtain

$$\overline{R}(\tau) \le \min_{0 < \delta \le 1-8\tau} \max \Big\{ 1 - \frac{1-\delta}{4} h\Big(\frac{4\tau}{1-\delta}\Big), h\Big(\frac{1-\delta}{2}\Big) \Big\} \tag{7}$$

Now, for $1 - 8\tau$ to be positive, we need $\tau < 1/8$. For any fixed $\tau \in [0, 1/8)$, and $\delta \in [0, 1-8\tau]$, the function $f(\delta) = 1 - \frac{1-\delta}{4} h\Big(\frac{4\tau}{1-\delta}\Big)$ is an increasing function of δ, while the function $g(\delta) = h(\frac{1-\delta}{2})$ is a decreasing function of δ. At $\delta = 0$, we have $g(\delta) \ge f(\delta)$. If, at $\delta = 1-8\tau$, we have $g(\delta) \le f(\delta)$, then it follows that the minimum over δ in (7) is achieved when $f(\delta) = g(\delta)$. In other words, the minimizing value of δ in this case is precisely the x^* in the statement of the proposition. It is readily verified that at $\delta = 1 - 8\tau$, we have $g(\delta) - f(\delta) = h(4\tau) + 2\tau - 1$, which is negative when $\tau \le 0.0706$. ∎

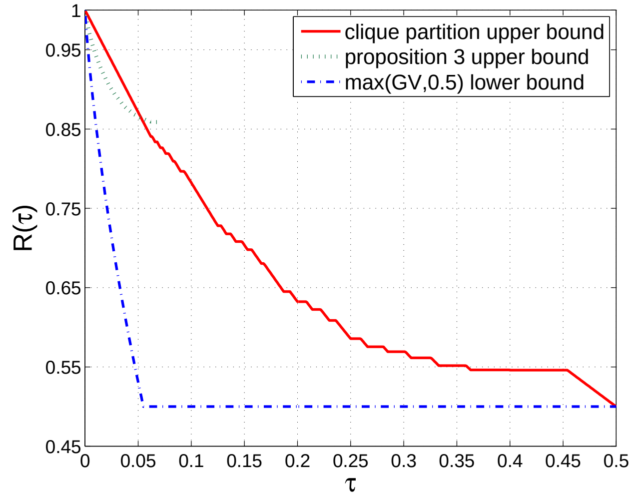


Fig. 1. Upper and lower bounds on the asymptotic coding rate of grain-correcting codes.

Bound (5) is plotted in Fig. 1, along with the asymptotic version of the Gilbert-Varshamov lower bound, which, as observed in Section II, is also valid for grain-correcting codes. The methods of the next subsection yield upper bounds on $\overline{R}(\tau)$ for any $\tau \le 1/2$, but these are harder to evaluate than the bound of Proposition 4.

C. *Upper Bounds Based on Clique Partitions*

A *clique partition* of a graph G is a partition $(V_1, \ldots, V_k)$ of its vertex set V such that the subgraph induced by each V_j, $j = 1, \ldots, k$, is a clique of G. Let $\bar{\chi}(G)$ denote the smallest size (number of parts) of any clique partition of G.

Let $G(n,t)$ be a confusability graph of the code space, defined as follows: the vertex set of $G(n,t)$ is $\{0,1\}^n$, and two distinct vertices $\boldsymbol{x}, \boldsymbol{x}'$ are joined by an edge iff they are t-confusable. For notational simplicity, we denote $\bar{\chi}(G(n,t))$ by $\bar{\chi}_{n,t}$. We do not assume that t is an integer; for non-integer values of t, we set $\bar{\chi}_{n,t} = \bar{\chi}_{n,\lfloor t \rfloor}$.

To state our next result, we need to extend the definition of $M(n,t)$ as follows: $M(0,t) = 1$ for all t.

Proposition 5 *For $m \le n$ and $s \le t$,*

$$M(n,t) \le \bar{\chi}_{m,s}\, M(n-m, t-s).$$

Proof : Let $\mathcal{C} \subseteq \{0,1\}^n$ be a t-grain-correcting code of size $|\mathcal{C}| = M(n,t)$, and let $(V_1, \ldots, V_k)$ be a clique partition of $G(m,s)$ of size $k = \bar{\chi}_{m,s}$. For $j = 1, \ldots, k$, define $\mathcal{C}_j = \{(c_1, \ldots, c_n) \in \mathcal{C} : (c_1, \ldots, c_m) \in V_j\}$. As the V_j's form a partition of $\{0,1\}^m$, the $\mathcal{C}_j$'s form a partition of $\mathcal{C}$. Therefore, it is enough to show that $|\mathcal{C}_j| \le M(n-m, t-s)$ for all j. Let $\mathcal{C}'_j = \{(c_{m+1}, \ldots, c_n) : \exists (c_1, \ldots, c_m, c_{m+1}, \ldots, c_n) \in \mathcal{C}_j\}$. The canonical projection map $\pi : \mathcal{C}_j \to \mathcal{C}'_j$ is a bijection; to see this, it is enough to show that π is injective. If $\pi(\boldsymbol{c}) = \pi(\hat{\boldsymbol{c}})$

for $\boldsymbol{c}, \hat{\boldsymbol{c}} \in \mathcal{C}_j$, then $\boldsymbol{c} = (c_1, \ldots, c_m, c_{m+1}, \ldots, c_n)$ and $\hat{\boldsymbol{c}} = (\hat{c}_1, \ldots, \hat{c}_m, c_{m+1}, \ldots, c_n)$ for some $(c_1, \ldots, c_m)$ and $(\hat{c}_1, \ldots, \hat{c}_m)$ in V_j. But, since the subgraph induced by V_j forms a clique in $G(m, s)$, we have that $(c_1, \ldots, c_m)$ and $(\hat{c}_1, \ldots, \hat{c}_m)$ are s-confusable. Thus, we see that $\boldsymbol{c}, \hat{\boldsymbol{c}}$ are s-confusable (and hence t-confusable since $s \le t$) unless $\boldsymbol{c} = \hat{\boldsymbol{c}}$. Hence, π is a bijection, so that $|\mathcal{C}_j| = |\mathcal{C}'_j|$.

We further claim that $\mathcal{C}'_j \subseteq \{0,1\}^{n-m}$ is a $(t-s)$-grain-correcting code, which would show that $|\mathcal{C}_j| = |\mathcal{C}'_j| \le M(n-m, t-s)$. Indeed, consider any pair of distinct words $\boldsymbol{c}', \boldsymbol{d}' \in \mathcal{C}'_j$. There exist distinct codewords $(\boldsymbol{a}', \boldsymbol{c}')$ and $(\boldsymbol{b}', \boldsymbol{d}')$ in $\mathcal{C}_j$. By definition of $\mathcal{C}_j$, $\boldsymbol{a}'$ and $\boldsymbol{b}'$ are s-confusable. So, if $\boldsymbol{c}'$ and $\boldsymbol{d}'$ were $(t-s)$-confusable, then $(\boldsymbol{a}', \boldsymbol{c}')$ and $(\boldsymbol{b}', \boldsymbol{d}')$ would be t-confusable, which cannot happen for distinct codewords in $\mathcal{C}_j$. Hence, $\mathcal{C}'_j$ is a $(t-s)$-grain-correcting code. ∎

Repeated application of the above proposition yields, for a positive integer $\ell \le \min\{n/m, t/s\}$,

$$M(n,t) \le (\bar{\chi}_{m,s})^{\ell} \, M(n - m\ell, t - s\ell). \tag{8}$$

By choosing ℓ appropriately, we obtain the following corollary to Proposition 5.

Corollary 6 *If $t/n \le s/m$, then*

$$M(n,t) \le (\bar{\chi}_{m,s})^{\lfloor t/s \rfloor} \, 2^{n-t-(m-s)\lfloor t/s \rfloor}. \tag{9}$$

If $s/m < t/n \le 1/2$, then

$$M(n,t) \le (\bar{\chi}_{m,s})^{\ell} \, 2^{n-t-\ell(m-s)} \tag{10}$$

where $\ell = \lfloor \frac{n-2t}{m-2s} \rfloor$.

Proof : Note first that $n - m\ell \ge 2(t - s\ell)$ iff $n - 2t \ge \ell(m-2s)$. When $t/n \le s/m$, set $\ell = \lfloor t/s \rfloor$. Since $t/n \le s/m$ is equivalent to $n - 2t \ge (t/s)(m-2s)$, we have $n - 2t \ge \ell(m-2s)$. Hence, from Proposition 1, we get $M(n - m\ell, t - s\ell) \le 2^{n-m\ell-(t-s\ell)}$. Plugging this into (8), we obtain (9).

When $s/m < t/n \le 1/2$, then set $\ell = \lfloor \frac{n-2t}{m-2s} \rfloor$. It is easily verified that $\ell \le n/m = \min\{n/m, t/s\}$, which is required for (8) to hold. Since we also have $n - 2t \ge \ell(m-2s)$, and hence, $n - m\ell \ge 2(t - s\ell)$, the bound (10) follows directly from Proposition 1. ∎

It is difficult to determine $\bar{\chi}_{m,s}$ exactly for arbitrary m, s. Upper bounds on $\bar{\chi}_{m,s}$ can be found by explicit constructions of clique partitions of $G(m,s)$. Observe that for any $\boldsymbol{y} \in \{0,1\}^m$, the set $\Phi^{-1}_{m,s}(\boldsymbol{y}) := \{\boldsymbol{x} \in \{0,1\}^m : \boldsymbol{y} \in \Phi_{m,s}(\boldsymbol{x})\}$ forms a clique in $G_{m,s}$. Thus, clique partitions of size k can be found by identifying sequences $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k \in \{0,1\}^m$ such that the sets $\Phi^{-1}_{m,s}(\boldsymbol{y}_j)$, $j = 1, \ldots, k$, cover $\{0,1\}^m$. Note that the sets $V_j = \Phi^{-1}_{m,s}(\boldsymbol{y}_j) \setminus \left(\bigcup_{i<j} V_i\right)$, $j = 1, \ldots, k$, then form a clique partition of $G(m,s)$. We implemented the greedy algorithm described below to find such a list of sequences $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$, and hence, a clique partition $V_1, \ldots, V_k$.

Table I lists upper bounds on $\bar{\chi}_{m,s}$ obtained via our implementation of the greedy algorithm. The underlined entries in the table are known to be exact values of $\bar{\chi}_{m,s}$, obtained either from the fact that $\bar{\chi}_{m,s} \ge M(m,s) \ge 2^{\lceil m/2 \rceil}$, or from specialized arguments that we omit here.

Algorithm 1 A greedy algorithm for finding clique partitions in $G(m,s)$.

```
determine the sets Φ^{-1}_{m,s}(y) for all y ∈ {0,1}^m;
set B(y) = Φ^{-1}_{m,s}(y) for all y ∈ {0,1}^m,
  set k = 0;
while there exists a y such that B(y) is non-empty do
    k ← k + 1 ;
    find a y_k such that |B(y_k)| = max_{y∈{0,1}^m} |B(y)|;
    set V_k = B(y_k);
    for each y ∈ {0,1}^m
        B(y) ← B(y) \ V_k;
return V_1, ..., V_k.
```

From Corollary 6 and Table I, we can obtain a suite of upper bounds on $M(n,t)$ valid for various ranges of n and t; for example, the entry for $(m,s) = (10,1)$ in the table yields that $M(n,t) \le 236^t 2^{n-10t}$ for $t/n \le 1/10$. The following upper bound on $\overline{R}(\tau)$, which was defined in (4), is also a direct consequence of Corollary 6.

Corollary 7 *For m, s such that $\tau \le s/m$,*

$$\overline{R}(\tau) \le 1 - \tau \left(\frac{m}{s} - \frac{1}{s} \log_2 \bar{\chi}_{m,s} \right). \tag{11}$$

For m, s such that $s/m < \tau \le 1/2$,

$$\overline{R}(\tau) \le \left(\frac{1 - 2\tau}{m - 2s} \right) \left(\log_2 \bar{\chi}_{m,s} - \frac{m}{2} \right) + \frac{1}{2}. \tag{12}$$

Proof : The bound in (11) follows easily from (9), while the derivation of (12) from (10) could use a note of explanation. To analyze the asymptotics of the bound in (10), we may ignore the floor function in the expression for ℓ. Then, with $\ell = \frac{n-2t}{m-2s}$, we have $n - m\ell = 2(t - s\ell)$, or equivalently, $n - t - \ell(m-s) = (n - m\ell)/2$. From this, the bound in (12) can be readily derived. ∎

When used in conjunction with Table I, the above corollary gives useful upper bounds on $\overline{R}(\tau)$. For instance, using the table entry for $(m,s) = (16,4)$, we find that $\overline{R}(\tau) \le 1 - \tau(4 - \frac{1}{4}\log_2 662) \approx 1 - 1.657\tau$ for $\tau \le 1/4$. Figure 1 plots, for $\tau \in [0, \frac{1}{2}]$, the minimum of all the upper bounds on $\overline{R}(\tau)$ obtainable from Corollary 7 and the entries of Table I. The plot also takes into account the fact that $\overline{R}(\tau)$ is a monotonically decreasing function of τ.

Setting $s = \tau m$ in Corollary 7, we obtain $\overline{R}(\tau) \le \frac{1}{m} \log_2 \bar{\chi}_{m,\tau m}$, and hence,

$$\overline{R}(\tau) \le \inf_m \frac{1}{m} \log_2 \bar{\chi}_{m,\tau m} = \lim_{m \to \infty} \frac{1}{m} \log_2 \bar{\chi}_{m,\tau m}. \tag{13}$$

The last equality above follows from Fekete's lemma (see e.g. [5, p. 85]), noting that $f(m) = \log_2 \bar{\chi}_{m,rm}$ is a subadditive function, i.e., $f(m+n) \le f(m) + f(n)$. The bound in (13) is presently only of theoretical interest, as the infimum (or limit) on the right-hand side is difficult to evaluate in general.

		m														
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
s	1	$\underline{2}$	$\underline{4}$	$\underline{6}$	$\underline{10}$	18	36	66	122	236	428	834	1574	3008	5716	11014
	2			$\underline{4}$	$\underline{8}$	12	18	30	54	92	162	284	530	948	1730	3210
	3					$\underline{8}$	$\underline{16}$	24	34	56	88	138	238	418	716	1266
	4							$\underline{16}$	$\underline{32}$	44	64	98	156	248	392	662

TABLE I

UPPER BOUNDS ON $\bar{\chi}_{m,s}$ OBTAINED BY COMPUTER SEARCH; THE UNDERLINED TABLE ENTRIES ARE KNOWN TO BE EXACT VALUES OF $\bar{\chi}_{m,s}$.

D. A List-Decoding Lower Bound

We briefly venture into the territory of list-decoding in this section, and give a lower bound on the achievable coding rate of a list-L-decodable code. Recall that in the list-decoding setting, the decoder is allowed to produce a list of up to L codewords. Formally, a code $\mathcal{C}$ is *list-L t-grain-correcting* if for any vector $\boldsymbol{x} \in \{0,1\}^n$, $|\{\boldsymbol{c} \in \mathcal{C} : \boldsymbol{x} \in \Phi_{n,t}(\boldsymbol{c})\}| \le L$. In words, for any received vector $\boldsymbol{x} \in \{0,1\}^n$, there are at most L codewords that could get transformed to $\boldsymbol{x}$ by the action of an operator $\phi \in \Phi_{n,t}$.

We will find the following definition useful in what is to follow. For $\phi \in \Phi_{n,t}$, let $\boldsymbol{e}_\phi$ be the vector $(e_1, \ldots, e_n) \in \{0,1\}^n$, with $e_j = 1$ iff ϕ has a length-2 grain beginning at the $(j-1)$th bit cell. Define $\mathcal{E}_{n,t} = \{\boldsymbol{e}_\phi : \phi \in \Phi_{n,t}\}$. Note that $\mathcal{E}_{n,t}$ consists of all binary "error vectors" of length n and Hamming weight at most t such that the first coordinate is always 0 and no two 1's are adjacent. An easy counting argument shows that

$$|\mathcal{E}_{n,t}| = \sum_{i=0}^{t} \binom{n-i}{i}. \tag{14}$$

Denote by $M(n,t;L)$ the maximum size of a list-L t-grain-correcting code of length n, and define for $0 \le \tau \le 1/2$,

$$\underline{R}(\tau; L) = \liminf_{n\to\infty} \frac{\log_2 M(n, \lfloor n\tau \rfloor; L)}{n}.$$

Proposition 8 *We have*

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

and hence,

$$\underline{R}(\tau; L) \ge \frac{L}{L+1} - (1-\tau)h\left(\frac{\tau}{1-\tau}\right)$$

for $\tau \le \frac{1}{2} - \frac{\sqrt{5}}{10} \approx 0.2764$.

Proof : For a vector $\boldsymbol{x} \in \{0,1\}^n$ let us define

$$B(\boldsymbol{x}) = \{\boldsymbol{z} \in \{0,1\}^n : \boldsymbol{x} \in \Phi_{n,t}(\boldsymbol{z})\}.$$

Note that $B(\boldsymbol{x}) \subseteq \{\boldsymbol{x} \oplus \boldsymbol{e} : \boldsymbol{e} \in \mathcal{E}_{n,t}\}$, so that $|B(\boldsymbol{x})| \le |\mathcal{E}_{n,t}| = \sum_{i=1}^{n}\binom{n-i}{i}$.

Let us construct the code by choosing M codewords randomly and uniformly with replacement from $\{0,1\}^n$. For a fixed vector $\boldsymbol{y} \in \{0,1\}^n$, call the choice of any $L+1$ codewords $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{L+1}$ 'bad' if $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{L+1} \in B(\boldsymbol{y})$. Clearly, the expected number of bad choices for a random code $\mathcal{C}$ is less than or equal to

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

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

The bound on $\underline{R}(\tau;L)$ follows from the observation that $\binom{n-i}{i}$ increases with i for $i \le \frac{1}{10}(5n-3-\sqrt{5n^2+10n+9})$. Thus, as long as $t/n \le \frac{1}{2} - \frac{\sqrt{5}}{10}$, the asymptotics of the summation $\sum_{i=0}^{t}\binom{n-i}{i}$ is determined by the term $\binom{n-t}{t}$. ∎

We do not at present have a useful upper bound on $M(n,t;L)$.

IV. GRAIN PATTERN KNOWN TO ENCODER/DECODER

In this section, we assume that the user of the recording system is capable of testing the medium and acquiring information about the structure of its grains. This information is used for the writing of the data on the medium or performing the decoding. Specifically, we assume again a medium with n bit cells and at most t grains of length 2, but now the locations of the grains are available either to the decoder but not the encoder of the data (Scenario I) or, conversely, to the encoder but not the decoder (Scenario II). Accordingly, let $M_i(n,t), i = 1,2$, be the maximum number of messages that can be encoded and decoded without error in each of the two scenarios. Also, for $0 \le \tau \le 1/2$, let

$$\underline{R}_i(\tau) = \liminf_{n\to\infty} \frac{\log_2 M_i(n, \lfloor n\tau \rfloor)}{n}, \quad i = 1,2,$$

be the coding rate achievable in each situation when t grows proportionally with n, with constant of proportionality τ.

For the analysis to follow, we need to recall the definition of $\mathcal{E}_{n,t}$ from Section III-D, and the fact (14) that $|\mathcal{E}_{n,t}| = \sum_{i=0}^{t}\binom{n-i}{i}$.

A. Scenario I

Here, we assume that the locations of the grains are known to the decoder of the data but are not available at the time of writing on the medium. A code $\mathcal{C}$ is said to correct t grains known to the receiver if $\phi(\boldsymbol{x}_1) \ne \phi(\boldsymbol{x}_2)$ for any two distinct vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C}$ and any $\phi \in \Phi_{n,t}$.

An obvious solution for the decoder is to consider as erasures the positions that could be in error, so the encoder can rely on a t-erasure-correcting code. Therefore, by the argument of the Gilbert-Varshamov bound, $M_1(n,t) \geq \frac{2^n}{\sum_{i=0}^{t}\binom{n}{i}}$, and hence, $\underline{R}_1(\tau) \geq 1 - h(\tau)$. However, this lower bound can be improved, as our next proposition shows.

Proposition 9 *We have*

$$M_1(n,t) \geq \frac{2^n}{\sum_{i=0}^{t}\binom{n-i}{i}}.$$

Hence, $\underline{R}_1(\tau) \geq 1-(1-\tau)h(\frac{\tau}{1-\tau})$ *for* $\tau \leq \frac{1}{2} - \frac{\sqrt{5}}{10} \approx 0.2764$.

Proof : We shall construct a code $\mathcal{C}$ of size at least $2^n/|\mathcal{E}_{n,t}|$ by a greedy procedure. We begin with an empty set, choose an arbitrary vector $\boldsymbol{x}_1$ and include it in $\mathcal{C}$. Having picked $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}$, for some $i > 1$, we choose $\boldsymbol{x}_i$ so that

$$\boldsymbol{x}_i \notin \bigcup_{j=1}^{i-1} \{\boldsymbol{x}_j \oplus \boldsymbol{e} : \boldsymbol{e} \in \mathcal{E}_{n,t}\}.$$

We stop when such a choice is not possible. At that point, we will have constructed a code $\mathcal{C}$ that satisfies $|\mathcal{C}| \cdot |\mathcal{E}_{n,t}| \geq 2^n$.

We claim that $\mathcal{C}$ corrects t grains known to the receiver. Suppose not; then there exists a grain pattern $\phi \in \Phi_{n,t}$ such that $\phi(\boldsymbol{x}_i) = \phi(\boldsymbol{x}_j)$ for some $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{C}$, $i > j$. Equivalently, $\boldsymbol{x}_i \oplus \boldsymbol{e} = \boldsymbol{x}_j \oplus \boldsymbol{e}'$ for some error vectors $\boldsymbol{e}, \boldsymbol{e}'$ with $\text{supp}(\boldsymbol{e}), \text{supp}(\boldsymbol{e}') \subseteq \text{supp}(\boldsymbol{e}_\phi)$, where $\text{supp}(\cdot)$ denotes the support of a vector. We then have $\boldsymbol{x}_i = \boldsymbol{x}_j \oplus (\boldsymbol{e} \oplus \boldsymbol{e}')$ with $\boldsymbol{e} \oplus \boldsymbol{e}' \in \mathcal{E}_{n,t}$, which contradicts the construction of $\mathcal{C}$.

As in the proof of Proposition 8, the bound on $\underline{R}_1(\tau)$ follows from the observation that when $t/n \leq \frac{1}{2} - \frac{\sqrt{5}}{10}$, the asymptotics of the summation $\sum_{i=0}^{t}\binom{n-i}{i}$ is determined by the term $\binom{n-t}{t}$. ∎

B. Scenario II

This scenario is similar in spirit to the channel with *localized errors* of Bassalygo et al. [1]. In that setting, both the transmitter and the receiver know that all but t positions of the codevector will remain error-free, and the coordinates of the t positions which can (but need not) be in error are known to the transmitter but not the receiver. Thus, in our Scenario II, the encoder may rely on codes that correct localized errors, which according to [1] gives the bound $\underline{R}_2(\tau) \geq 1 - h(\tau)$. Again, this bound can be improved.

Proposition 10 *We have*

$$M_2(n,t) \geq \frac{1}{2n}\frac{2^n}{\sum_{i=0}^{t}\binom{n-i}{i}}.$$

Hence, $\underline{R}_2(\tau) \geq 1-(1-\tau)h(\frac{\tau}{1-\tau})$ *for* $\tau \leq \frac{1}{2} - \frac{\sqrt{5}}{10} \approx 0.2764$.

Proof : We show that when the encoder knows the error locations, then it can successfully transmit

$$M \geq \frac{1}{2n}\frac{2^n}{|\mathcal{E}_{n,t}|} \tag{15}$$

messages to the decoder, which proves the claimed lower bound on $M_2(n,t)$. We follow the proof of Theorem 3 of [1].

Given a message $i \in \{1,\ldots,M\}$ to be transmitted, the transmitter will use knowledge of the grain pattern ϕ (with $\boldsymbol{e}_\phi \in \mathcal{E}_{n,t}$) to encode i using a suitably chosen vector from a set of binary vectors $\mathcal{X}^i = \{x_j^i : j = 1,\ldots,n\}$. A vector $\boldsymbol{x}_j^i$ is said to be *good* for $\boldsymbol{e} \in \mathcal{E}_{n,t}$ if for any $i \neq i'$ and for any j' we have,

$$d_H(\boldsymbol{x}_j^i \oplus \boldsymbol{e}, \boldsymbol{x}_j^i) < d_H(\boldsymbol{x}_j^i \oplus \boldsymbol{e}, \boldsymbol{x}_{j'}^{i'}),$$

where $d_H(\cdot,\cdot)$ denotes Hamming distance. The family of sets $\mathcal{X}^i$, $i = 1,\ldots,M$, is *good* if for any $i \in \{1,\ldots,M\}$ and for any $\boldsymbol{e} \in \mathcal{E}_{n,t}$, there exists a vector $\boldsymbol{x}_j^i \in \mathcal{X}^i$ that is good for $\boldsymbol{e}$. A good family of sets $\mathcal{X}^i$, $i = 1,\ldots,M$, enables the encoder to transmit any message in $\{1,\ldots,M\}$ with perfect recovery by the decoder. Indeed, given the grain pattern ϕ, the encoder chooses for transmission of message i a vector in $\mathcal{X}^i$ that is good for $\boldsymbol{e}_\phi$.

Thus, we only need to show that for M satisfying (15) there exists a good family of sets $\mathcal{X}^i = \{x_j^i : j = 1,\ldots,n\}$, $i = 1,\ldots,M$. There are 2^{n^2M} families of M sets $\mathcal{X}^i$, each containing at most n binary vectors of length n. Of these, the number of families that are *not* good does not exceed

$$M \cdot |\mathcal{E}_{n,t}| \cdot ((M-1)n|\mathcal{E}_{n,t}|)^n \cdot 2^{n^2(M-1)}.$$

If M satisfies (15) with equality, then this number is less than 2^{n^2M}. Therefore, there exists a good family of sets $\mathcal{X}^i$.

The argument for the lower bound on $\underline{R}_2(\tau)$ is the same as that given for $\underline{R}_1(\tau)$ in the proof of Proposition 9, since the extra multiplicative factor of $\frac{1}{2n}$ does not affect the asymptotic behavior. ∎

To summarize, we obtain a lower bound on $\underline{R}_i(\tau)$, $i = 1$ or 2, of the form

$$\underline{R}_i(\tau) \geq \max\Big\{0.5,\ 1-(1-\tau)h(\frac{\tau}{1-\tau})\Big\}.$$

This is because the rate-1/2 code $\mathcal{R}_n$ defined in (1) is still viable in the context of Scenarios I and II. A straightforward upper bound $\underline{R}_i(\tau) \leq 1-\tau$ follows from the fact that $M_1(n,t)$ and $M_2(n,t)$ cannot exceed 2^{n-t}, which is simply the one-bit-per-grain upper bound.

V. Capacity of the Grains Channel

Thus far in this paper, we have considered a combinatorial model of the one-dimensional granular medium, and given various bounds on the rate of t-grain-correcting codes. We will now switch to a parallel track by defining a natural probabilistic model of a channel corresponding to the one-dimensional granular medium with grains of length at most 2 (the "grains channel"). This is a 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. Our goal is to estimate the Shannon-theoretic capacity for the grains channel model. Let us proceed to formal definitions.

Suppose $\boldsymbol{x} = x_1x_2\ldots.$ and $\boldsymbol{y} = y_1y_2\ldots.$ denote the input and output sequence respectively, with $x_i, y_i \in \{0,1\}$ for all i. We further define the sequence $\boldsymbol{u} = u_1u_2\ldots.$, 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 $p \in [0,1]$):

$$\begin{array}{c|cc} & u_i = 0 & u_i = 1 \\ \hline u_{i-1} = 0 & 1-p & p \\ u_{i-1} = 1 & 1 & 0 \end{array} . \qquad (16)$$

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,

$$y_i = x_i \oplus (x_i \oplus x_{i-1})u_i, \qquad (17)$$

where the operations are being performed modulo 2. Equivalently,

$$y_i = \begin{cases} x_i & \text{if } u_i = 0 \\ x_{i-1} & \text{if } u_i = 1. \end{cases} \qquad (18)$$

We will find it useful to define the error sequence $\boldsymbol{z} = z_1, z_2, z_3, \ldots$, where $z_i = x_i \oplus y_i$. Thus,

$$z_i = u_i(x_i \oplus x_{i-1}). \qquad (19)$$

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 [3, 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 [3].

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 [3] 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

$$\min_{s_0} q(s_n \mid \boldsymbol{x}^n, s_0) > 0. \qquad (20)$$

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

$$C_n = \frac{1}{n} \max_{Q^n(\boldsymbol{x}^n)} I(\boldsymbol{x}^n; \boldsymbol{y}^n \mid s_0). \qquad (21)$$

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, [3, Appendix 4A, Lemma 1]), we in fact have

$$C = \lim_{n\to\infty} \frac{1}{n} \max_{Q^n(\boldsymbol{x}^n)} I(\boldsymbol{x}^n; \boldsymbol{y}^n). \qquad (22)$$

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:

$$R = \lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{x}^n; \boldsymbol{y}^n), \qquad (23)$$

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 11 *The grains channel is indecomposable for* $p < 1$.

Proof : We must check that the condition in (20) holds. We take $n = 1$ and $s_1 = (0, x_1)$. Then, $\min_{s_0} q(s_1 \mid x_1, s_0) = \min_{j\in\{0,1\}} P(u_1 = 0 \mid u_0 = j) = 1 - p > 0$. ∎

As a consequence of the above proposition, the equality $\underline{C} = \overline{C}$ holds for the grains channel when $p < 1$. In fact, this equality also holds for the grains channel when $p = 1$, as the following result shows.

Proposition 12 *For the grains channel with* $p = 1$*, we have* $\underline{C} = \overline{C} = 1/2$.

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 the two propositions above, the capacity of the grains channel is defined by (22). From here onward, we denote this capacity by C^{g}, and use the notation $C^{\mathrm{g}}(p)$ when the dependence on p needs to be emphasized. It is difficult to compute the capacity C^{g} exactly, so we will provide useful upper and lower bounds. We note here for future reference the trivial bound obtained from Proposition 12:

$$C^{\mathrm{g}}(p) \geq C^{\mathrm{g}}(1) = 1/2. \tag{24}$$

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 (16). We then have

$$y_i = \begin{cases} x_i & \text{if } u_i = 0 \\ \varepsilon & \text{if } u_i = 1 \end{cases} \tag{25}$$

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 u_0 taking values in $\{0, 1\}$.

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 13 *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

$$y'_i = \begin{cases} y_i & \text{if } y_i \neq \varepsilon \\ y_{i-1} & \text{if } y_i = \varepsilon \end{cases} \tag{26}$$

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 (25), (26) 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 the equation obtained by replacing y_i with y'_i in (18). This immediately leads to the following theorem.

Theorem 14 *For* $p \in [0, 1]$*, we have* $C^{\mathrm{g}}(p) \leq C^{\varepsilon}(p) = \frac{1}{1+p}$.

Remark: We remark that any code that corrects t nonadjacent substitution errors (bit flips) also corrects t grain errors. It is therefore tempting to bound from below 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 $\boldsymbol{u}$ (16), and $y_i = x_i \oplus u_i$ for all $i \geq 1$. The capacity of this channel is computed as in the BINAEras case and equals $1 - h(p)/(1 + p)$, where $h(p)$ denotes the binary entropy function. However, this does not constitute a valid lower bound for $C^{\mathrm{g}}(p)$: for example, when p is close to 1, we have $1 - h(p)/(1 + p)$ also being close to 1, while $C^{\mathrm{g}}(1) \leq 1/2$ according to Theorem 14.

[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.

D. Lower Bound: The Symmetric Information Rate

In this section, we derive an exact expression for the SIR of the grains channel, which gives a lower bound on the capacity of the channel. In accordance with the definition of SIR (23), 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

$$R^{\mathrm{g}} = \lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{x}^n; \boldsymbol{y}^n) \tag{27}$$

$$I(\boldsymbol{x}^n; \boldsymbol{y}^n) = H(\boldsymbol{y}^n) - H(\boldsymbol{y}^n|\boldsymbol{x}^n) = H(\boldsymbol{y}^n) - H(\boldsymbol{z}^n|\boldsymbol{x}^n) \tag{28}$$

As noted above, $H(Y) = \lim_{n\to\infty} \frac{1}{n} H(\boldsymbol{y}^n)$ exists. In fact, we can give an exact expression for $H(Y)$ in terms of an infinite series.

Proposition 15 *The entropy rate of the output process of the grains channel is given by*

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

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* $j \geq 3$*,*

$$\beta_j = \frac{1}{2}\left(\frac{1-(1+p)\beta_{j-1}}{1-\beta_{j-1}}\right). \tag{29}$$

The lengthy proof of this proposition is given in Appendix B.

Remark: The following explicit expression for $\beta_j, j \geq 2$ can be proved by induction from (29):

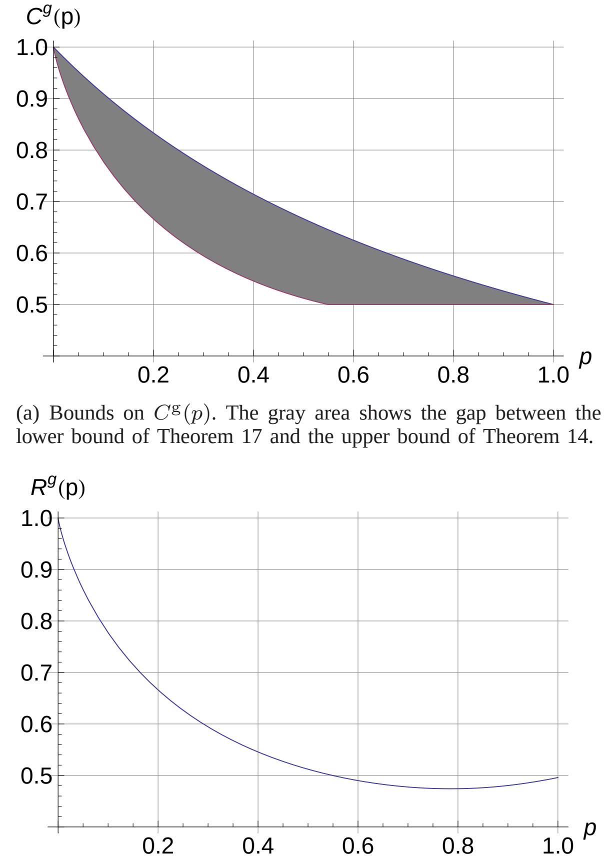
$$\beta_j = \frac{2((\vartheta_-)^j - (\vartheta_+)^j)}{(3+B+p)(\vartheta_-)^j - (3-B+p)(\vartheta_+)^j} \tag{30}$$

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

Our next result shows that $\lim_{n\to\infty} H(\boldsymbol{z}^n \mid \boldsymbol{x}^n)$ also exists, and gives an exact expression for it, again in terms of an infinite series. Appendix B contains a proof of this result.

Proposition 16 *When $\boldsymbol{x}$ is an i.i.d. uniform Bernoulli sequence, we have*

$$\lim_{n\to\infty} H(\boldsymbol{z}^n \mid \boldsymbol{x}^n) = \frac{1+p/2}{1+p} \sum_{j=2}^{\infty} 2^{-j} h\Big(\frac{1-(-p)^j}{1+p}\Big).$$



(a) Bounds on $C^{\mathrm{g}}(p)$. The gray area shows the gap between the lower bound of Theorem 17 and the upper bound of Theorem 14.

(b) The symmetric information rate $R^{\mathrm{g}}(p)$.

Fig. 2. Plots of the upper and lower bounds on the capacity of the grains channel $C^{\mathrm{g}}(p)$, and the SIR of the grains channel $R^{\mathrm{g}}(p)$, as functions of p.

Together, (27), (28), and Propositions 15 and 16 provide an *exact expression* for the SIR of the grains channel. This, along with the trivial bound (24), yields the following lower bound on the capacity C^{g}.

Theorem 17 *The capacity $C^{\mathrm{g}}(p) \geq \max(1/2, R^{\mathrm{g}}(p))$, where $R^{\mathrm{g}}(p)$ is the SIR of the grains channel and is given by the following expression:*

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

with β_j as in (29) or (30).

In Figure 2, we plot the upper and lower bounds on $C^{\mathrm{g}}(p)$ stated in Theorems 14 and 17 as well as the value of $R^g(p)$ from Theorem 17. Observe that the SIR is a strict lower bound on the capacity, at least for $0.56 \leq p < 1$, as $R^{\mathrm{g}}(p) < 1/2$ within this range of p.

The plots are obtained by numerically evaluating $R^{\mathrm{g}}(p)$ by truncating its infinite series at some large value of j. We give here a somewhat crude, but useful, estimate of the error in truncating this series at some index $j = J$, with $J \geq 2$. Define the partial sums

$$S_J = \frac{1+p/2}{1+p} \sum_{j=2}^{J} 2^{-j} h\Big(\frac{1-(-p)^j}{1+p}\Big) \tag{31}$$

$$T_J = \frac{1}{2(1+p)} \sum_{j=2}^{J} h(\beta_j) \prod_{k=2}^{j-1} (1-\beta_k) \tag{32}$$

and note that the Jth partial sum of the $R^{\mathrm{g}}(p)$ series is precisely $T_J - S_J$.

Proposition 18 *The error $|R^g(p) - (T_J - S_J)|$ in truncating the $R^{\mathrm{g}}(p)$ series at an index $j = J$, with $J \geq 2$, is at most*

$$\frac{1}{1+p} \left[(1+p/2)\, 2^{-J} + 2^{-\lfloor (J+1)/2 \rfloor} \right].$$

In particular, for any $p \in [0,1]$, the truncation error is at most $2^{-J} + 2^{-\lfloor (J+1)/2 \rfloor}$.

We defer the proof to Appendix B.

The plot of $R^g(p)$ in Figure 2(a) was generated using $J = 15$ terms of the infinite series, so the plotted curve is within 0.004 of the true R^{g} curve for all p.

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 $\mathcal{R}_n$ defined in (1) 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 19 *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 $C_0 = \lim_{n\to\infty} R_0(n) = 1/2$.

Appendix A: Proof of Theorem 13

Observe first that the BINAEras channel is indecomposable for $p < 1$. Indeed, for this channel, the condition in (20) 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 $\underline{C} = \overline{C} = 1/2$ follows, *mutatis mutandis*, the proof of Proposition 12.

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 $H(\boldsymbol{u}^n \mid u_0) = \sum_{n=1}^{n} H(u_n \mid u_{n-1}) = nH(u_1 \mid u_0) = n\frac{h(p)}{1+p}$. Hence,

$$C_n = n^{-1} \max_{Q^n(\boldsymbol{x}^n)} H(\boldsymbol{y}^n \mid u_0) - \frac{h(p)}{1+p}. \tag{33}$$

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 [2, 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

$$\begin{aligned} &H(y_i \mid \boldsymbol{u}^{i-1}, u_0) \leq H(y_i \mid u_{i-1}) \\ &\quad = H(y_i \mid u_{i-1} = 0)\frac{p}{1+p} + H(y_i \mid u_{i-1} = 1)\frac{1}{1+p} \end{aligned}$$

Given $u_{i-1} = 1$, $\boldsymbol{y}_i$ is a binary random variable (since $u_i = 0$ with probability 1), and thus, $H(y_i \mid u_{i-1} = 1) \leq 1$. 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 $H(y_i \mid u_{i-1} = 1) \leq$

$h(p)+1-p$. Putting all the inequalities together, we find that

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

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,

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

Plugging this into (33), we obtain that $C_n = \frac{1}{1+p}$ for all n, and hence, $C = \frac{1}{1+p}$.

Appendix B: Proofs of Propositions 15, 16 and 18

B.1. Proof of Proposition 15

Since $\lim_{n\to\infty} \frac{1}{n} H(\boldsymbol{y}^n) = \lim_{i\to\infty} H(y_{i+1} \mid \boldsymbol{y}^i)$, we need show that the latter limit 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 (18) and the fact that the input $\boldsymbol{x}$ is an i.i.d. Bernoulli(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,

$$H(y_{i+1} \mid \boldsymbol{y}^i) = 2 \sum_{\boldsymbol{b}\in B} H(y_{i+1} \mid \boldsymbol{y}^i = \boldsymbol{b}) \Pr[\boldsymbol{y}^i = \boldsymbol{b}], \quad (34)$$

where $B = \{(b_i, \ldots, b_1) \in \{0,1\}^i : b_i = 0\}$ 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

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

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 $(0, \ldots, 0, 1)$. We record two facts about B_j. First,

$$\begin{aligned}\Pr[\boldsymbol{y}^i \in B_j] &= \Pr[(y_i, y_{i-1}, \ldots, y_{i-j+1}) = 0^{j-1}1] \\ &= \Pr[(y_j, y_{j-1}, \ldots, y_1) = 0^{j-1}1], \quad (35)\end{aligned}$$

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

$$H(y_{i+1} \mid \boldsymbol{y}^i = \boldsymbol{b}) = h(\Pr[y_{i+1} = 1 \mid \boldsymbol{y}^i = \boldsymbol{b}]) \quad (36)$$

is invariant over B_j.

Lemma 20 *For* $\boldsymbol{b} \in B_j$, $\Pr[y_{i+1} = 1 \mid \boldsymbol{y}^i = \boldsymbol{b}]$ *equals*

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

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 (18), 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

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

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,

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

Now, given the channel state $s_{i-j+2} = (0,0)$, the random variables u_{i+1}, y_i, $y_{i-1}, \ldots, y_{i-j+2}$ 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,

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

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

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

which is what we needed to show. ∎

In the statement of Proposition 15, we defined $\beta_j = \Pr[y_{j+1} = 1 \mid (y_j, y_{j-1}, \ldots, y_1) = 0^{j-1}1]$. Note that if we set $i = j$ in Lemma 20, we get

$$\beta_j = 1/2 \Pr[u_j = 0 \mid (y_{j-1}, \ldots, y_2) = 0^{j-2},\ (u_1, x_1) = (0,0)]. \quad (37)$$

From (34)–(37), and Lemma 20, we have

$$\begin{aligned}H(y_{i+1} \mid \boldsymbol{y}^i) &= 2\sum_{j=2}^{i} h(\beta_j) \Pr[(y_j, \ldots, y_1) = 0^{j-1}1] \\ &\quad + 2H(y_{i+1} \mid \boldsymbol{y}^i = 0^i) \Pr[\boldsymbol{y}^i = 0^i]. \quad (38)\end{aligned}$$

The term at the end of the above expression vanishes as $i \to \infty$, as we show below for completeness.

Lemma 21 $\lim_{i\to\infty} H(y_{i+1} \mid \boldsymbol{y}^i = 0^i) \Pr[\boldsymbol{y}^i = 0^i] = 0.$

Proof : Since $0 \le H(y_{i+1} \mid \boldsymbol{y}^i = 0^i) \le 1$, it is enough to show that $\Pr[\boldsymbol{y}^i = 0^i] = 0$ converges to 0. For this, observe that for any j, if $y_j = 0$, then $(x_{j-1}, x_j) \ne (1,1)$. Hence, if $\boldsymbol{y}^i = 0^i$, then $(x_1, x_2) \ne (1,1)$, $(x_3, x_4) \ne (1,1)$, and so on. Thus, $\Pr[\boldsymbol{y}^i = 0^i] \le (3/4)^{\lfloor i/2 \rfloor}$, which suffices to prove the lemma. ∎

So, letting $i \to \infty$ in (38), we obtain

$$H(Y) = 2\sum_{j=2}^{\infty} h(\beta_j) \Pr[(y_j, y_{j-1}, \ldots, y_1) = 0^{j-1}1]. \quad (39)$$

The proof of Proposition 15 will be complete once we prove the next two lemmas.

Lemma 22 *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 write

$$\Pr[(y_2, y_1) = (0,1)] = \sum_{(a,b)\in\{0,1\}^2} \Pr[(y_2, y_1) = (0,1) \mid (u_2, u_1) = (a,b)] \times \Pr[(u_2, u_1) = (a,b)].$$

Clearly, $\Pr[(u_2, u_1) = (1,1)] = 0$. Also, $\Pr[(y_2, y_1) = (0,1) \mid (u_2, u_1) = (1,0)] = 0$, since, given $(u_2, u_1) = (1,0)$ we must have $y_2 = x_1 = y_1$, by virtue of (18). Next, given $(u_2, u_1) = (0,0)$, we have $(y_2, y_1) = (x_2, x_1)$, and since (x_2, x_1) is independent of (u_2, u_1), we find that

$$\Pr[(y_2, y_1) = (0,1) \mid (u_2, u_1) = (0,0)] = \Pr[(x_2, x_1) = (0,1)] = 1/4.$$

By a similar argument, $\Pr[(y_2, y_1) = (0,1) \mid (u_2, u_1) = (0,1)] = 1/4$. Hence,

$$\Pr[(y_2, y_1) = (0,1)] = (1/4) \Pr[u_2 = 0] = \frac{1}{4(1+p)},$$

as desired. ∎

Lemma 23 $\beta_2 = \frac{1}{2}(1-p)$*, and for $j \geq 3$, β_j satisfies the recursion in (29).*

Proof : From (37), we have

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

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 $\beta_j = (1/2) \Pr[u_j = 0 \mid E_j] = (1/2)(1-\gamma_j)$, where $\gamma_j := \Pr[u_j = 1 \mid E_j]$. We shall show that for $j \geq 3$,

$$\gamma_j = \frac{p(1-\gamma_{j-1})}{1+\gamma_{j-1}}. \tag{40}$$

which is equivalent to the recursion in (29).

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 $1/2(1-\gamma_{j-1})$. Thus,

$$\gamma_j = p \cdot \frac{1/2(1-\gamma_{j-1})}{\Pr[y_{j-1} = 0 \mid E_{j-1}]} \tag{41}$$

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

$$\begin{aligned} &\sum_{b\in\{0,1\}} \Pr[y_{j-1} = 0 \mid u_{j-1} = b, E_{j-1}] \Pr[u_{j-1} = b \mid E_{j-1}] \\ &= 1/2(1-\gamma_{j-1}) + \Pr[y_{j-1} = 0 \mid u_{j-1} = 1, E_{j-1}] \cdot \gamma_{j-1} \end{aligned} \tag{42}$$

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 (42), we get

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

Feeding this back into (41), we obtain

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

which is the desired recursion (40). ∎

This concludes the proof of Proposition 15.

B.2. Proof of Proposition 16

We break the proof into two parts. We first show that

$$\lim_{n\to\infty} \frac{1}{n} H(\boldsymbol{z}^n \mid \boldsymbol{x}^n) = \sum_{j=2}^{\infty} 2^{-j} H(u_j \mid u_1) \tag{43}$$

and subsequently, we prove that

$$\sum_{j=2}^{\infty} 2^{-j} H(u_j|u_1) = \frac{1+p/2}{1+p} \sum_{j=2}^{\infty} 2^{-j} h\Big(\frac{1-(-p)^j}{1+p}\Big). \tag{44}$$

To show (43), we start with

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

From (19), 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 $A_0 = \{\boldsymbol{x}^i : x_i = x_{i-1}\}$,

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

for $1 \leq j \leq i-2$, 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. Since $\boldsymbol{x}$ is an i.i.d. uniform Bernoulli sequence, we have $\Pr[\boldsymbol{x}^i \in A_j] = (1/2)^{j+1}$ for $0 \leq j \leq i-2$, and $\Pr[\boldsymbol{x}^i \in A_{i-1}] = (1/2)^{i-1}$.

Now, if $\boldsymbol{x}^i \in A_0$, then by (19), 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 $z_{i-j} = u_{i-j}$. Therefore,

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

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, for $\boldsymbol{x}^i \in A_{i-1}$, we only need the fact that $\eta_i := H(z_i \mid z_1, \ldots, z_{i-1}, \boldsymbol{x}^i \in A_{i-1})$ lies between 0 and 1.

We thus have

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

Letting $i \to \infty$, we obtain (43).

It remains to prove (44). For this, 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,

$$\begin{aligned} \sum_{j=2}^{\infty} 2^{-j}\, H(u_j \mid u_1 = 1) & = \sum_{j=2}^{\infty} 2^{-j}\, H(u_{j-1} \mid u_1 = 0) \\ & = \frac{1}{2} \sum_{j=2}^{\infty} 2^{-j}\, H(u_j \mid u_1 = 0) \end{aligned}$$

since $H(u_1 \mid u_1 = 0) = 0$. Putting it all together, we find that

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

Finally, observe that $H(u_j \mid u_1 = 0) = h\big(\frac{1-(-p)^j}{1+p}\big)$, as it can be shown (for example, by induction) that $\Pr(u_j = 0 \mid u_1 = 0) = \frac{1-(-p)^j}{1+p}$ for all $j \geq 1$. This proves (44), and with this, the proof of Proposition 16 is complete.

B.3. Proof of Proposition 18

The error in truncating the $R^{\mathrm{g}}(p)$ series at the index $j = J$ is

$$\begin{aligned} & |R^{\mathrm{g}}(p) - (T_J - S_J)| \\ & \leq |H(Y) - T_J| + |\lim_{n\to\infty} H(\boldsymbol{z}^n \mid \boldsymbol{x}^n) - S_J|. \end{aligned} \tag{45}$$

It is easy to bound the second term in (45):

$$\begin{aligned} & |\lim_{n\to\infty} H(\boldsymbol{z}^n \mid \boldsymbol{x}^n) - S_J| \\ & = \frac{1 + p/2}{1 + p} \sum_{j=J+1}^{\infty} 2^{-j} h\Big(\frac{1 - (-p)^j}{1 + p}\Big) \\ & \leq \frac{1 + p/2}{1 + p} \sum_{j=J+1}^{\infty} 2^{-j} \\ & = \left(\frac{1 + p/2}{1 + p}\right) 2^{-J}. \end{aligned} \tag{46}$$

Turning our attention to the first term in (45), we see that

$$\begin{aligned} |H(Y) - T_J| & = \frac{1}{2(1+p)} \sum_{j=J+1}^{\infty} h(\beta_j) \prod_{k=2}^{j-1} (1 - \beta_k) \\ & \leq \frac{1}{2(1+p)} \sum_{j=J+1}^{\infty} \prod_{k=2}^{j-1} (1 - \beta_k). \end{aligned} \tag{47}$$

Now, from the recursion (29), we readily get for $k \geq 3$,

$$1 - \beta_k = \frac{1}{2} \left(\frac{1 - (1-p)\beta_{k-1}}{1 - \beta_{k-1}} \right),$$

and hence,

$$(1 - \beta_k)(1 - \beta_{k-1}) = \frac{1}{2}[1 - (1-p)\beta_{k-1}] \leq \frac{1}{2}.$$

Consequently, if $j = 2m$ for some $m \geq 1$, then

$$\prod_{k=2}^{j-1} (1 - \beta_k) = \prod_{k=1}^{m-1} (1 - \beta_{2k+1})(1 - \beta_{2k}) \leq (1/2)^{m-1},$$

and if $j = 2m + 1$ for some $m \geq 1$, then

$$\prod_{k=2}^{j-1} (1 - \beta_k) \leq \prod_{k=1}^{m-1} (1 - \beta_{2k+1})(1 - \beta_{2k}) \leq (1/2)^{m-1}.$$

Upon replacing the bound in (47) by the looser bound

$$\frac{1}{2(1+p)} \sum_{j=2\lfloor (J+1)/2 \rfloor}^{\infty} \prod_{k=2}^{j-1} (1-\beta_k)$$

so that the summation starts at an even index j, routine algebraic manipulations now yield

$$\begin{aligned} |H(Y) - T_J| &\leq \frac{1}{1+p} \sum_{m=\lfloor (J+1)/2 \rfloor}^{\infty} (1/2)^{m-1} \\ &= \left(\frac{1}{1+p}\right) 2^{-\lfloor (J+1)/2 \rfloor}. \end{aligned}$$

Plugging this and (46) into (45), we obtain Proposition 18.

ACKNOWLEDGEMENT: Navin Kashyap would like to thank Bane Vasic for bringing to his attention the problem considered in this work. The authors would like to thank Haim Permuter for helpful discussions. The authors are also grateful to the reviewer whose suggestion led to a strengthening of the upper bound in Corollaries 6 and 7 in Sect. IIIC.

REFERENCES

[1] L. A. Bassalygo, S. I. Gelfand and M. S. Pinsker, "Coding for channels with localized errors," *Proc. 4th Soviet-Swedish Workshop in Information Theory*, Gotland, Sweden, 1989, pp. 95–99.

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

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

[4] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, "Write channel model for bit-patterned media recording," arXiv:1010.4603, Oct. 2010.

[5] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics,* Cambridge University Press, 1992.

[6] A. Raman Krishnan, R. Radhakrishnan and B. Vasic, "Read Channel Modeling for Detection in Two-Dimensional Magnetic Recording Systems," *IEEE Trans. Magnet.*, vol. 45, no. 10, pp. 3679–3682, Oct. 2009.

[7] R.M. Roth, *Introduction to Coding Theory*, Cambridge University Press, Cambridge, UK, 2006.

[8] 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.

[9] R. I. White, R. M. H. New and R. F. W. Pease, "Patterned media: Viable route to 50Gb/in^2 and up for magnetic recording," *IEEE Trans. Magn.*, vol. 33, no. 1, pt. 2, pp. 990-995, Jan. 1997.