

Rewritable Storage Channels with Hidden State

Ramji Venkataramanan, *Member, IEEE*, Sekhar Tatikonda, *Senior Member, IEEE*,
Luis Lastras-Montaño, *Senior Member, IEEE*, Michele Franceschini, *Member, IEEE*

Abstract—Many storage channels admit reading and rewriting of the content at a given cost. We consider rewritable channels with a hidden state which models the unknown characteristics of the memory cell. In addition to mitigating the effect of write noise, rewrites can help the write controller obtain a better estimate of the hidden state. The paper has two contributions. The first is a lower bound on the capacity of a general rewritable channel with hidden state. The lower bound is obtained using a coding scheme that combines Gelfand-Pinsker coding with superposition coding. The rewritable AWGN channel is discussed as an example. The second contribution is a simple coding scheme for a rewritable channel where the write noise and hidden state are both uniformly distributed. It is shown that this scheme is asymptotically optimal as the number of rewrites gets large.

Index Terms—Rewritable channels, storage channels, PCM

I. INTRODUCTION

IN nonvolatile memory technologies, the write mechanism is commonly impaired by write noise due to which the value written on a cell is different from the one intended. An important feature of many of these technologies such as Flash [2], Phase Change Memory [3] and Resistive RAM [4], [5] is that they allow rewriting, i.e. the value written on a memory cell can be read and rewritten if necessary. Rewrites can increase the storage capacity but are costly since they are time consuming and degrade the memory. Hence there is a fundamental trade-off between the number of writes and the amount of information that can be stored in a memory cell.

Before proceeding, we note that the rewritable channels discussed in this paper are distinct from rewriting codes for Write-Once-Memories and Flash, see e.g. [6], [7]: in the latter, the goal is to store *different* messages in each write with a constraint that cell values can only be increased; in our case, we use multiple writes to improve the fidelity of storing a *single* message. A formal definition of a coding scheme for a rewritable channel will be given later in this section.

Given a memory array of n cells, the goal is to maximize the number of distinct messages that can be reliably encoded in the array, subject to a constraint on the average or maximum number of writes per cell. The cells are assumed to be statistically independent. Rewritable channels were introduced in [8] and subsequently studied in [9]–[12] under an average

cost constraint. Maximum cost constrained rewritable channels were considered in [13]–[16].

In practice, a memory cell is an amalgam of physical components which reacts to inputs in some way that designers hope to model as well as possible. However, there are always some unknown characteristics of the cell due, for example, to fabrication variability. These characteristics, which may be too costly to learn, introduce an extra degree of uncertainty into the value written on the cell. In this paper, we model this effect with the channel $P_{Y|X,S}$ where X is the input stimulus, S is a hidden (unknown) state parameter of the cell and Y is the value stored in the cell. S is assumed to have known distribution P_S . The alphabets of X, Y, S are denoted by $\mathcal{X}, \mathcal{Y}, \mathcal{S}$, respectively. We consider two canonical examples of rewritable channels in this paper:

- 1) *Uniform noise channel with state*: The channel model is

$$Y = X + W + S. \quad (1)$$

$X \in [0, 1]$ is the input stimulus, the write noise W is uniformly distributed in $[-a/2, a/2]$, and the state S is uniformly distributed in $[0, B]$. a, B are known positive constants. The basic version of this channel ($S = 0$) was introduced in [8] as a simple model that captures some essential features of non-volatile memories such as analog inputs and bounded write noise.

- 2) *AWGN channel with state*: The channel is again described by the additive model (1). The write noise and the state are Gaussian random variables: W is distributed as $\mathcal{N}(0, N)$ and S is distributed as $\mathcal{N}(0, \sigma_s^2)$.¹ The input is constrained in terms of either the average or peak power per write.

One motivation for this model has been drawn from phenomena observed in real Phase Change Memory (PCM) cells. This type of memory encodes information as a resistance value of the cell, which in turn is influenced by the relative degree to which a memory cell is in amorphous or crystalline state. To reach a given value of resistance, one “programs” the memory cell by applying a pulse with a given set of parameters chosen according to the desired resistance value. However, the actual resistance value one obtains can vary significantly (see, for example, [17]). Nevertheless, as [17, Fig. 1] shows, there is a regime of input stimulus where the logarithm of the achieved resistance of a cell can be approximately predicted as an additive offset to the logarithm of the resistance of another cell, with the same input stimulus. This naturally leads to the notion of additive offset S in the model (1). The W term is included to model additional write noise which is independent across write attempts.

¹ $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with mean μ and variance σ^2 .

This work was supported in part by NSF grant CCF-1017744 and an IBM faculty award. A part of the paper was presented at the 2012 IEEE International Symposium on Information Theory [1].

R. Venkataramanan is with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK (e-mail: ramji.v@eng.cam.ac.uk).

S. Tatikonda is with the Department of Electrical Engineering, Yale University, New Haven CT 06511, USA (e-mail: sekhar.tatikonda@yale.edu).

L. Lastras and M. Franceschini are with IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: {lastrasl, franceschini}@us.ibm.com).

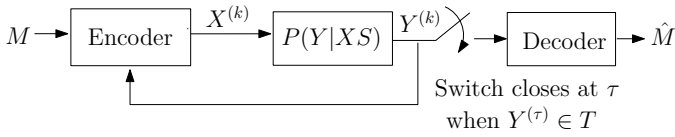


Fig. 1: Rewritable channel with state.

Finally, the independence of each write attempt on a cell (conditioned on the unknown offset) can be justified as follows. A programming pulse in PCM can either melt a cell or not. For example, a RESET pulse which is intended to put the cell in a high resistance state or the family of partial-RESET pulses (see [18]) are both pulses that melt a cell. On the other hand, a pulse that partially anneals a PCM cell does not melt it. Programming pulses that cause melting of the cell will have an outcome that is not highly correlated to the prior state of the cell, because the organization of the atoms prior to the pulse is forgotten during the melting phase. This in turn implies that the independent write noise model is a good one for such programming schemes for Phase Change Memory.

A key feature of the state S is that it stays *fixed* across write attempts in each cell. Conditioned on $S = s$, the value stored in the cell after each attempt depends only on the most recent input stimulus – it is determined according to $P_{Y|XS}(\cdot|X, S = s)$, where X is the current input. In the additive model (1), this means that each write attempt on a cell is affected with an independent realization of the random variable W , while S stays fixed across write attempts.

In this paper, we consider rewritable channels with a constraint on the average number of writes per cell. Given a constraint κ on the average write cost, the goal is to determine the capacity $C(\kappa)$ and design coding schemes to achieve rates close to $C(\kappa)$.

We consider the following class of coding schemes for the rewritable channel. To write on cell i , the write controller applies stimuli $X_i^{(1)}, X_i^{(2)}, \dots$ until the output falls within a target region T_i , where T_i is a subset of the output space. The k th write stimulus $X_i^{(k)}$ can depend on the outputs of the previous stimuli, denoted $Y_i^{(1)}, \dots, Y_i^{(k-1)}$. Formally, a rate R code over an array of n cells is defined by:

- An encoder mapping which maps a message in $\{1, \dots, 2^{nR}\}$ to a sequence $((X_1, T_1), \dots, (X_n, T_n))$, where T_i is the target region for cell i , and $X_i = (X_i^{(1)}, X_i^{(2)}, \dots)$ is the input strategy for cell i .
- A decoder which maps the output sequence (Y_1, \dots, Y_n) to $\{1, \dots, 2^{nR}\}$.

For cell i , the number of writes needed for the output to fall within region T_i is a random variable, denoted $\tau_i(X_i, T_i)$, where $X_i = (X_i^{(1)}, X_i^{(2)}, \dots)$ is the input strategy. Fig. 1 shows the schematic of the rewritable channel model.

The average write-cost of the code is $\frac{1}{n} \sum_{i=1}^n \mathbb{E} \tau_i(X_i, T_i)$. Due to the statistical independence of the cells, the capacity for an average cost constraint κ is [8], [11]

$$C(\kappa) = \sup_{X, T: \mathbb{E} \tau(X, T) \leq \kappa} I(XT; Y). \quad (2)$$

The capacity formula in (2) is not easy to compute in general. This is because the optimization is over adaptive

strategies where each input stimulus can depend on the outcomes of the previous stimuli. Adaptive strategies are particularly useful in channels with hidden state because we get a better estimate of the state with each write, which can be used to generate the next stimulus.² For intuition, consider two extreme cases:

- When $\kappa = 1$, we are allowed only one write attempt and the hidden state is treated as an additional noise variable.
- When the average cost constraint $\kappa \rightarrow \infty$, we can spend a number of write attempts to get a very good estimate of the state, and use the remaining writes to store information by designing the input stimulus to nullify the effect of the state. Thus we expect the storage rate to approach the no-state capacity when κ is very large.

For $1 < \kappa < \infty$, the challenge is to simultaneously learn the state while attempting to store information at a high rate.

The main contributions of this paper are as follows.

- 1) In Section II, we derive a capacity lower bound for continuous-output rewritable channels with state. The scheme used to obtain this bound involves state estimation phase followed by a coding phase. The writing strategy in the coding phase combines two techniques from multi-user information theory: Gelfand-Pinsker coding [19], [20] and superposition coding [21]. The AWGN rewritable channel is discussed as an example.
- 2) In Section III, we focus on the uniform noise channel and present a coding scheme that is computationally simple and amenable to practical implementation. The scheme implicitly combines state estimation and coding, and is shown to be asymptotically optimal as the number of rewrites gets large.

The rewritable channel considered in this paper is a stylized model relevant to technologies like Phase Change Memory and Resistive RAM which have analog outputs. Both these memory technologies are known to be affected by variability across devices [3], [22], which to the first order can be modeled as a hidden state, for the reasons outlined earlier in this section. Though relaxing assumptions such as noiseless reads would make the model more realistic, we believe that the current model gives useful insights regarding how rewrites can be harnessed to improve the storage density of these memories.

Notation: We use upper-case letters to denote random variables and bold-face notation for random vectors. Entropy and mutual information are measured in bits, and logarithms are with base 2 unless otherwise mentioned.

II. LOWER BOUND ON THE REWRITE CAPACITY

For an average write cost κ , we design a scheme consisting of two phases: an estimation phase of l ($< \kappa$) writes to learn the state S , and a coding phase requiring an average of $\kappa - l$ writes. For the coding phase, we combine two techniques: 1) Gelfand-Pinsker coding [19] which achieves the optimal rate given the state estimate if we are allowed only a single write for the coding phase (i.e., $\kappa - l = 1$), and 2) Superposition

²For memoryless rewritable channels without state, we can restrict the input strategies to be non-adaptive, i.e. repeatedly apply the same stimulus to a cell until the target region is hit. See [11], [12].

coding [21] to store an additional $\log(\kappa - l)$ bits/cell when $\kappa > l + 1$.

Before presenting the general result, we describe the coding scheme for the AWGN rewritable channel to highlight the main ideas.

A. The AWGN channel with hidden state

The channel is defined by (1) with the write noise $W \sim \mathcal{N}(0, N)$ and the state $S \sim \mathcal{N}(0, \sigma_s^2)$. We assume that there is an average power-constraint P , i.e., in each write attempt the average power of the input stimulus across the n cells is at most P .

State Estimation: The first step is to construct an estimate of the state of each cell using l writes. Due to the symmetry of the channel model, this can be done by applying any input, say c , for l writes and recording the outputs $Y^{(1)}, \dots, Y^{(l)}$ which are generated according to

$$\begin{aligned} Y^{(1)} &= c + W^{(1)} + S \\ &\vdots \\ Y^{(l)} &= c + W^{(l)} + S \end{aligned} \quad (3)$$

where $W^{(1)}, \dots, W^{(l)}$ are independent $\mathcal{N}(0, N)$ random variables. The minimum-mean squared error (MMSE) estimate of S given the observations $Y^{(1)}, \dots, Y^{(l)}$ is

$$\hat{S}(l) = \mathbb{E}[S | Y^{(1)}, \dots, Y^{(l)}] = \frac{\sigma_s^2}{l\sigma_s^2 + N} \sum_{j=1}^l (Y^{(j)} - c). \quad (4)$$

Encoding: The write channel for the $(l+1)$ th write can be expressed as

$$Y^{(l+1)} = X^{(l+1)} + \hat{S}(l) + (S - \hat{S}(l)) + W^{(l+1)}. \quad (5)$$

The estimate $\hat{S}(l)$ is known to the encoder prior to the $(l+1)$ th write. Further, $(S - \hat{S}(l))$ is independent of $\hat{S}(l)$ due to the orthogonality principle [23] and the joint Gaussianity of $(S, \hat{S}(l))$.

Let us first consider the case where we use only a single write after the estimation period. For write $l+1$, (5) describes a channel with state $\hat{S}(l)$ known to the encoder and effective channel noise $S - \hat{S}(l) + W^{(l+1)}$ which is independent of $\hat{S}(l)$. The effective channel noise is a Gaussian random variable distributed as $\mathcal{N}(0, N_{\text{eff},l})$ where

$$\begin{aligned} N_{\text{eff},l} &= \mathbb{E}[(S - \hat{S}(l) + W^{(l+1)})^2] \\ &= \mathbb{E}[(S - \hat{S}(l))^2] + \mathbb{E}[(W^{(l+1)})^2] = \frac{\sigma_s^2 N}{l\sigma_s^2 + N} + N. \end{aligned} \quad (6)$$

The optimal coding scheme for this channel is the ‘writing on dirty-paper’ scheme of Costa [20]. The key idea is to incorporate part of the known state $\hat{S}(l)$ into the codeword. This is done by building a codebook over an auxiliary random variable $U \sim \mathcal{N}(0, P + \alpha^2 \sigma_{s,l}^2)$ where

$$\alpha = \frac{P}{P + N_{\text{eff},l}}, \quad (7)$$

$$\sigma_{s,l}^2 = \mathbb{E}[\hat{S}(l)^2] = \frac{l\sigma_s^4}{l\sigma_s^2 + N}. \quad (8)$$

Let the storage rate be R bits/cell. We build a U -codebook with 2^{nR_1} codewords whose elements are generated i.i.d according to $\mathcal{N}(0, P + \alpha^2 \sigma_{s,l}^2)$. The value of $R_1 > R$ will be specified below. The codebook is divided into 2^{nR} bins, with each bin containing $2^{n(R_1 - R)}$ codewords. Each bin represents a message in the set $\{1, \dots, 2^{nR}\}$. To transmit message m , the encoder attempts to find a codeword \mathbf{U} within bin m such that $(\mathbf{U} - \alpha \hat{\mathbf{S}}(l))$ is nearly orthogonal to $\hat{\mathbf{S}}(l)$. Formally, the encoder finds a codeword \mathbf{U} that is jointly typical³ [24] with $\hat{\mathbf{S}}(l)$ according to the distribution described by

$$U = X + \alpha \hat{S}(l) \quad (9)$$

where $X \sim \mathcal{N}(0, P)$ and $\hat{S}(l) \sim \mathcal{N}(0, \sigma_{s,l}^2)$ are independent Gaussians. From rate-distortion theory, this step will be successful if the number of sequences in each bin $2^{n(R_1 - R)}$ is larger than $2^{nI(U; \hat{S}(l))}$, where the mutual information computed using the joint distribution described by (9). The codeword written on the n cells is

$$\mathbf{X}^{(l+1)} = \mathbf{U} - \alpha \hat{\mathbf{S}}(l). \quad (10)$$

Note that $\mathbf{X}^{(l+1)}$ has average power nearly P , due to (9). The sequence stored on the cells is

$$\mathbf{Y}^{(l+1)} = \mathbf{X}^{(l+1)} + \mathbf{S} + \mathbf{W}^{(l+1)}. \quad (11)$$

The decoder’s task is to decode the codeword \mathbf{U} from $\mathbf{Y}^{(l+1)}$. The corresponding bin index then gives the message. If the encoding operation is successful, $(\mathbf{U}, \mathbf{Y}^{(l+1)})$ are jointly typical according to

$$\begin{aligned} U &= X + \alpha \hat{S}(l), \\ Y &= X + \hat{S}(l) + (S - \hat{S}(l) + W), \end{aligned} \quad (12)$$

where $X \sim \mathcal{N}(0, P)$, $\hat{S}(l) \sim \mathcal{N}(0, \sigma_{s,l}^2)$, and $(S - \hat{S}(l) + W) \sim \mathcal{N}(0, N_{\text{eff},l})$ are mutually independent random variables. If we use a maximum-likelihood or joint typicality decoder [24], the codeword \mathbf{U} can be successfully decoded if $R_1 < I(U; Y)$.

Combining this with the earlier bound $R_1 - R > I(U; \hat{S}(l))$, we conclude that rates

$$R < I(U; Y) - I(U; \hat{S}(l)) \quad (13)$$

are achievable. Computing this with the joint distribution specified in (12), we obtain that any rate

$$R < \frac{1}{2} \log \left(1 + \frac{P}{N_{\text{eff},l}} \right) \quad (14)$$

is achievable.

If we restrict ourselves to a single write after the estimation period, (14) gives the optimal rate. This is because even when the encoder and decoder *both* know $\hat{\mathbf{S}}(l)$ a priori, the maximum rate is given by (14). (The decoder can simply cancel off any effect of $\hat{\mathbf{S}}(l)$ from the stored value.) When $\hat{\mathbf{S}}(l)$ is not available at the decoder, the Costa coding scheme nullifies its effect by incorporating part of it into the codeword \mathbf{U} .

Superposition: When $\kappa > l + 1$, we have available more than one write after the estimation period. We use superposition

³Roughly speaking, sequences (\mathbf{U}, \mathbf{S}) are jointly typical according to distribution P if their empirical joint distribution is close to i.i.d. P .

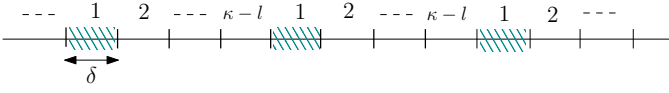


Fig. 2: Target regions for superposition coding. The shaded regions together form the target region for one message.

coding to store additional bits using the remaining writes. For the sake of intuition, temporarily assume that κ is an integer.

The idea is to partition the output space (\mathbb{R} for the AWGN channel) into $(\kappa - l)$ different regions such that the output is equally likely to fall in each of these regions *regardless* of the input. This is done in the following way. We divide the real line into intervals of length δ , and assign to the intervals labels $1, \dots, (\kappa - l)$ in succession, as shown in Figure 2. Target region 1 is the union of the intervals marked 1, target region 2 is the union of the intervals marked 2, and so on. Formally, we define the target regions T_j for $j = 1, \dots, \kappa - l$ as

$$T_i = \cup_{i \in \mathbb{Z}} \left[((\kappa - l)i + j)\delta - \delta, ((\kappa - l)i + j)\delta \right]. \quad (15)$$

We let $\delta \rightarrow 0$ for reasons explained below.

For each cell $i \in \{1, \dots, n\}$, the additional information stored is represented by a message m_i drawn uniformly from the set $\in \{1, \dots, \kappa - l\}$. At the end of estimation period, the controller uses the state sequence estimate $\hat{S}(l)$ to determine the codeword $\mathbf{U} = (U_1, \dots, U_n)$ of the Costa scheme. It then repeatedly applies stimulus X_i to cell i until the output falls in the target region m_i . Recall that

$$X_i = U_i - \alpha \hat{S}_i(l).$$

In each write attempt, the noise realization is an independent realization of a $\mathcal{N}(0, N)$ random variable. However, the state estimation error $S - \hat{S}(l)$ remains *constant* across attempts and is unknown to the decoder. Defining each target region as a collection of disjoint infinitesimal intervals ensures that the output in each write attempt is equally likely to lie in each of the $(\kappa - l)$ target regions, regardless of the value of $S - \hat{S}(l)$. Thus the number of writes required to obtain an output in the desired region m_i is a geometric random variable with mean $(\kappa - l)$.

The total number of writes (including the estimation period) for cell i is denoted τ_i and the final value stored in cell i is $Y_i^{(\tau_i)}$. The discussion above shows that

$$\mathbb{E}[\tau_i] = l + \kappa - l = \kappa. \quad (16)$$

Decoding: The decoder observes the stored sequence

$$\mathbf{Y}^{(\tau)} = (Y_1^{(\tau_1)}, \dots, Y_n^{(\tau_n)})$$

and attempts to decode the codeword \mathbf{U} . The key observation is that $(\mathbf{U}, \mathbf{Y}^{(l+1)})$ and $(\mathbf{U}, \mathbf{Y}^{(\tau)})$ have the same joint distribution. This is because for each cell i , the write stimulus and the channel state remain the same for writes $(l+1)$ through τ_i , and the noise realizations $W^{(l+1)}, \dots, W^{(\tau_i)}$ are i.i.d. $\mathcal{N}(0, N)$. Thus $(\mathbf{U}, \mathbf{Y}^{(\tau)})$ is jointly typical according to (12), and the codeword \mathbf{U} can be reliably decoded if R satisfies (14). The target region containing the output of cell i directly gives the message $m_i \in \{1, \dots, \kappa - l\}$ stored in the superposition phase.

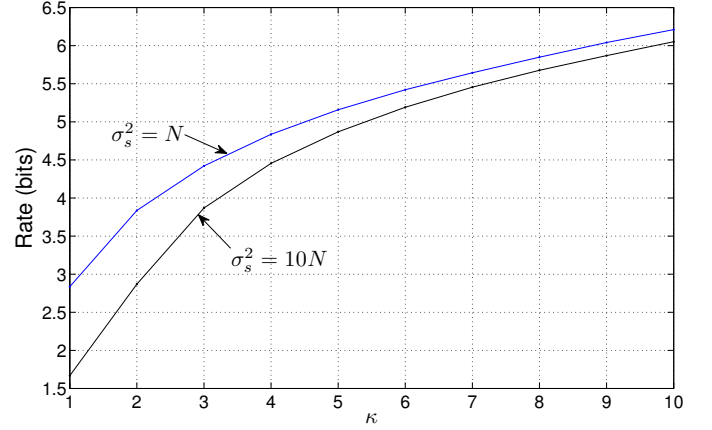


Fig. 3: Lower bound of Proposition 1 for $\frac{P}{N} = 100$. The top curve is for $\sigma_s^2 = N$ and the bottom one for $\sigma_s^2 = 10N$.

When κ is not an integer, we can vary the number of target regions across the cells in order to achieve a write-cost of κ . For example, if $\kappa = 1 + \lambda$, for $\lambda \in (0, 1)$, we can code a fraction λ of the n cells at average cost 2 and the remaining $n(1 - \lambda)$ cells at average cost 1 to obtain an overall cost of $\kappa = 1 + \lambda$. Thus the straight line joining the value of the lower bound at $\kappa = 1$ and $\kappa = 2$ is a lower bound for $\kappa \in (1, 2)$. In general, the convex hull of the rates achieved at the integer points can be achieved through ‘cost-sharing’ between cells.

The performance achieved of two-step coding strategy described above is summarized in the following proposition.

Proposition 1: Consider the channel described by (1) with state $S \sim \mathcal{N}(0, \sigma_s^2)$, noise $W \sim \mathcal{N}(0, N)$ and an average power constraint P on the input. With average cost $\kappa \geq 1$, the rewrite capacity satisfies

$$C(\kappa) \geq \text{conv} \left(\max_{l \in \{0, 1, \dots, \lfloor \kappa \rfloor - 1\}} \left\{ \frac{1}{2} \log \left(1 + \frac{P}{N_{\text{eff}, l}} \right) + \log \lfloor \kappa - l \rfloor \right\} \right) \quad (17)$$

where conv denotes the convex hull and

$$N_{\text{eff}, l} = N \left(1 + \frac{\sigma_s^2}{l\sigma_s^2 + N} \right).$$

Figure 3 shows the capacity lower bound for $\frac{P}{N} = 100$ and $1 \leq \kappa \leq 10$. Curves are plotted for $\sigma_s^2 = N$ and $\sigma_s^2 = 10N$. For the second case, the maximum in (17) is attained with an estimation period of $l = 0$ for $\kappa = 1$, $l = 1$ for $2 \leq \kappa \leq 9$, and $l = 2$ for $\kappa = 10$.

B. Lower bound for General Channels

We consider channels whose output support is continuous valued, i.e., for $\forall(x, s) \in \mathcal{X} \times \mathcal{S}$, $P_{Y|XS}(\cdot|x, s)$ is absolutely continuous with respect to the Lebesgue measure. This assumption is necessary because the definition of target regions for superposition coding in Section II-A implicitly assumes continuous valued outputs. The superposition idea can be extended to many discrete channels as well, but we do not pursue this here in order to keep the exposition simple.

For a channel with average write cost κ , the two-step strategy involves: a) Designing a suitable estimator to estimate the state sequence using l writes, and b) storing information in the remaining $(\kappa - l)$ writes using Gelfand-Pinsker coding and superposition. The Costa coding scheme for the AWGN channel is a special case of the Gelfand-Pinsker scheme for general memoryless channels with state, with the state known a priori at the encoder.

Theorem 1: Consider a channel $P_{Y|XS}(\cdot|x, s)$ that is absolutely continuous with respect to the Lebesgue measure for all $(x, s) \in \mathcal{X} \times \mathcal{S}$. With average cost $\kappa \geq 1$, the capacity satisfies

$$C(\kappa) \geq \text{conv} \left(\max_{l \in \{0, \dots, \lfloor \kappa \rfloor - 1\}} \max_{\mathcal{P}} \left\{ I(U; Y) - I(U; \hat{S}(l)) + \log[\kappa - l] \right\} \right) \quad (18)$$

where \mathcal{P} is the set of joint distributions of $(S, \hat{S}(l), U, X, Y)$ of the form

$$P_S \cdot P_{\hat{S}(l)|S} \cdot P_{U|\hat{S}(l)} \cdot \mathbf{1}_{X=f(U, \hat{S}(l))} \cdot P_{Y|XS}.$$

Proof: See Appendix.

Remarks:

- 1) In the set of joint distributions \mathcal{P} the state distribution P_S and the channel law $P_{Y|XS}$ are fixed. The maximization over \mathcal{P} is therefore over the choice of estimator $\hat{S}(l)$, auxiliary distribution $P_{U|\hat{S}(l)}$, and function f to generate the channel input X from $(U, \hat{S}(l))$.
- 2) The MMSE estimator is optimal for the AWGN average-power constrained channel, but in general the optimal estimator depends on the channel law and the input constraints.

We conclude this section with a brief discussion of the shortcomings of the two-step coding scheme discussed in this section. First, dedicating l writes to estimating the state and then coding is not optimal in general. A scheme that simultaneously performs estimation and coding in each write attempt is likely to yield higher rates, but such a scheme may also be harder to analyze.

The information is stored in the cell array in coded in two ways: through Gelfand-Pinsker coding and superposition coding. Each of these poses a different challenge for practical implementation. In the Costa/Gelfand-Pinsker scheme we used joint typicality or maximum-likelihood decoding, both of which have complexity that grows exponentially in n , and thus computationally infeasible for a large array of n cells. For the AWGN case, there has been progress towards feasible decoders using structured codebooks such as those based on lattices [25], [26].

For superposition coding, we need the reads to be very accurate as the width δ of the intervals is made small (cf. Figure 2). This is important during encoding (so that the controller knows when to stop writing) as well as decoding (for the decoder to know which target region the output lies in). This problem can be handled by using an outer error-correcting code to correct errors that arise due to imperfect reads.

In the following section, we design a coding scheme that addresses all the above issues for the uniform-noise channel.

III. UNIFORM NOISE CHANNEL WITH HIDDEN STATE

The channel is described by (1), with the write noise W is uniformly distributed in $[-a/2, a/2]$ and the state S is uniformly distributed in $[0, B]$. a, B are assumed to be known positive constants. For ease of analysis, we will assume that $B < a$.

We present two code constructions, each of which gives a lower bound on the rewrite capacity. The first is sub-optimal but gives insight into features of good coding strategies. The second construction yields a better lower bound which is asymptotically optimal, i.e., it is arbitrarily close to the no-state capacity for sufficiently large cost constraint. The second scheme implicitly performs simultaneous state estimation and coding; further, it is computationally simple and robust to small inaccuracies in the reading process.

A. Code Construction 1

For the uniform noise rewrite channel without state given by $Y = X + W$, the basic coding idea is that with an average of κ rewrites, we can shrink the effective width of the noise interval to a/κ . The average-cost capacity was obtained in [11].

Fact 1: [11] For $\kappa \geq \kappa_0 \triangleq \lceil \frac{1+a}{a} \rceil / (\frac{1+a}{a})$, the rewrite capacity with average cost constraint κ is

$$C(\kappa) = \log \left(\frac{1+a}{a} \kappa \right).$$

When $\frac{1+a}{a} \kappa$ is an integer, the capacity is achieved by simply dividing the space $[-a/2, 1+a/2]$ into equal-sized intervals of length a/κ and choosing the target region T to be one of these intervals with equal probability. The input X is any point which maximizes the probability of the output falling in the region T . When $\frac{1+a}{a} \kappa$ is not an integer, the capacity is achieved by a careful generalization of the above idea, described in [11].

When there is an unknown state offset $S \in [0, B]$, the idea is to define each target region such that there is exactly one subset of width a/κ that can be accessed with a *fixed* input and an average of κ writes, irrespective of the offset.

Proposition 2: For the uniform noise rewrite channel with hidden state and average cost $\kappa \geq 2$,

$$C(\kappa) \geq \log \left(\kappa \left[\frac{1+a+B}{a+B} \right] \right). \quad (19)$$

Proof: The target regions: Refer Fig. 4. The output space $[-a/2, 1+a/2+B]$ is first divided into intervals of length $(a+B)$ each. There are $N = \lfloor \frac{1+a+B}{a+B} \rfloor$ such intervals, denoted $\mathcal{Z}_i, 0 \leq i \leq N-1$. If $\frac{1+a+B}{a+B}$ is not an integer, the remaining output space $(N(a+B), 1+a/2+B]$ is discarded.

For clarity, consider the case where κ is an integer. Divide each interval \mathcal{Z}_i into κ target regions. The target regions for the first interval are defined as shown in Fig. 4. Target region 1 is the interval $[-\frac{a}{2}, -\frac{a}{2} + \frac{a}{\kappa}] \cup [\frac{a}{2}, \frac{a}{2} + \frac{a}{\kappa}]$; region 2 is the interval

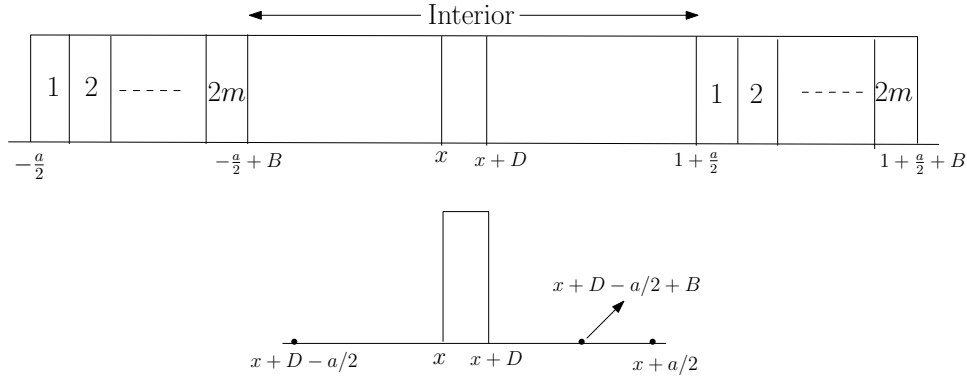


Fig. 5: Construction 2: The output space is divided into interior and exterior target regions. The bottom figure shows the an interior target region $[x, x + D]$. The state shifts the input $x + D - a/2$ to the right by an amount at most B . For all $b \in [0, B]$, the probability of hitting the target region in each write attempt is D/a as long as $D < a - B$.

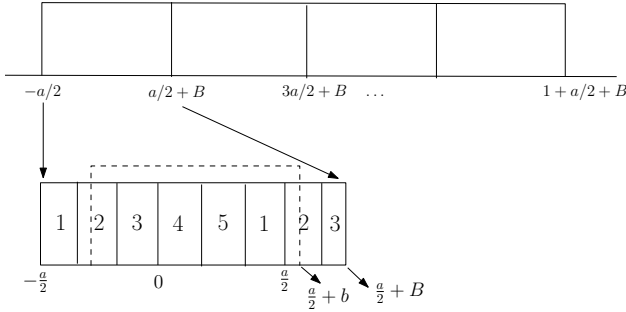


Fig. 4: Each interval of width $a + B$ is divided into $\kappa = 5$ target regions. The target regions in the interval $(-a/2, a/2+B)$ are shown. To write on a region in this interval, the stimulus 0 is applied. The dashed lines indicate the part of the interval accessible with $S = b$.

$[-\frac{a}{2} + \frac{a}{\kappa}, -\frac{a}{2} + \frac{2a}{\kappa}] \cup [\frac{a}{2} + \frac{a}{\kappa}, \frac{a}{2} + \frac{2a}{\kappa}]$, and so on. Similarly, κ target regions are defined for each of the N intervals.

Encoding: To reach target region t in interval Z_i for $t \in \{1, \dots, \kappa\}$, $i \in \{0, \dots, N-1\}$, apply input $X = (a+B)i$ until the output falls within region t in interval i . With this input, the accessible part of the target region has width exactly a/κ for any value of $S \in [0, B]$. This is illustrated in the bottom part of Fig. 4. Regardless of the offset, the probability of the output falling within the target region on any write attempt is $\frac{a/\kappa}{a}$. The average number of rewrites is therefore κ .

The total number of target regions is $N\kappa$ and by assigning them equal probability, the rate is

$$I(XT; Y) = I(T; Y) = H(T) = \log(N\kappa) \quad (20)$$

where the first equality holds because the input X is a function of the target region T , and the second equality is due to T being uniquely determined by Y .

The general case where κ is not an integer can be handled by an extension of the above scheme using the techniques in [11]. ■

Remark: When $\frac{1+a+B}{a+B}$ is an integer, (19) can be written as

$$C(\kappa) \geq \log\left(\kappa \frac{1+a}{a}\right) - \log\left(\frac{1+B/a}{1+B/(1+a)}\right) \quad (21)$$

The first term above is the capacity when $S = 0$, or when S is precisely known at the encoder. The second term is the loss incurred by the coding scheme due the state being unknown.

In the above construction, we designed the target regions so that each one can be accessed with equal probability regardless of the value of S . We did not use the rewrites to do any state estimation. The sub-optimality of this strategy is seen by observing that even when the number of rewrites κ is very large, the lower bound of Proposition 2 is strictly less than the capacity with $S = 0$, given by the first term in (21). The next construction remedies this deficiency.

B. Code Construction 2

As shown in Fig. 5, divide the output space $[-a/2, 1 + a/2 + B]$ into two regions: the interval $[-a/2 + B, 1 + a/2]$ called the ‘interior’, and the remaining space $[-a/2, -a/2 + B] \cup [1 + a/2, 1 + a/2 + B]$ called the ‘exterior’.

Interior target regions: Divide the interior into intervals (target regions) of width D . The key observation is that if $D < a - B$, regardless of the value of S , each interior target region is fully accessible with an average of a/D write attempts with a fixed input. As illustrated in the bottom part of Fig. 5, to access the interior target region $[x, x + D]$, apply the stimulus $(x + D - a/2)^+$. To fully access the region with offset b , we need

$$(x + D - a/2)^+ + b - a/2 < x$$

which holds for all $b \in [0, B]$ as long as $D < a - B$.

Exterior target regions: As shown in Fig. 5, define $2m$ exterior target regions for an integer $m \geq 1$. For $i = 1, \dots, 2m$, the i th exterior target region, labeled E_i , is

$$\left[-\frac{a}{2} + (i-1)\frac{B}{2m}, -\frac{a}{2} + i\frac{B}{2m}\right] \cup \left[1 + \frac{a}{2} + (i-1)\frac{B}{2m}, 1 + \frac{a}{2} + i\frac{B}{2m}\right].$$

With this construction, we present a coding scheme that achieves the following lower bound on the rewrite capacity.

Theorem 2: For $\kappa \geq 2$

$$C(\kappa) \geq \max_{p, D, m} h(p) + p \log \left[\frac{1+a-B}{D} \right] + (1-p) \log 2m$$

where the maximum is over $p \in [0, 1]$, $D \in (0, a - B)$ and integers $m \geq 1$ that satisfy

$$(1-p)\frac{a}{B} \left[2m+1 + \frac{1}{m} \sum_{i=1}^m \left(\ln \frac{i-\delta_i}{(1-\delta_i)^2} + \frac{\delta_i}{1-\delta_i} \ln \delta_i \right) \right] + p\frac{a}{D} \leq \kappa \quad (22)$$

where the optimal $\delta_i \in [0, 1]$ for $i = 1, 2, \dots, m$ is determined by the following equation:

$$2(1-\delta_i)^2 + 3(i-1)(1-\delta_i) + (i-\delta_i) \ln \delta_i = 0. \quad (23)$$

The optimal δ_i for a few values of i are listed in Table I.

Remark: The δ_i in the above theorem can be chosen to be arbitrary values in $[0, 1]$. Picking δ_i that satisfy (23) minimizes the number of rewrites, given by the left side of (22). For example, (22) can be replaced by a simpler condition obtained by setting $\delta_i = 0$ for all i :

$$p\frac{a}{D} + (1-p)\frac{a}{B} \left(2m+1 + \frac{\ln m!}{m} \right) \leq \kappa. \quad (24)$$

The proof of the theorem is given in the next section. Fig. 6 shows the lower bound of Theorem 2 with $a = 1/3$ and $B = a/2$ for various values of κ . We also plot the capacity with no hidden state ($S = 0$), which was obtained in [11] and is an upper bound to the rewrite capacity of the current model with hidden state.

We now show that the above lower bound converges to the no-state capacity as the rewrite constraint $\kappa \rightarrow \infty$.

Corollary 1: The rate $R(\kappa)$ achieved by Theorem 2 satisfies

$$\left| \log \left(\frac{1+a}{a} \kappa \right) - R(\kappa) \right| \rightarrow 0 \text{ as } \kappa \rightarrow \infty$$

Proof: Choose $D = a/\kappa$ and $m = \frac{B}{2D} = \kappa \frac{B}{2a}$. Note that for all $B < a$, $D = a/\kappa < (a - B)$ for sufficiently large κ . With this choice and setting $\delta_i = 0$ for all i , the average number of rewrites given by the left-side of (24) becomes

$$p\kappa + (1-p)\frac{\kappa}{2} \left(2 + \frac{1}{m} + \frac{\ln m!}{m^2} \right) = \kappa(1 + \epsilon_\kappa),$$

where $\epsilon_\kappa = O\left(\frac{\log \kappa}{\kappa}\right)$ goes to 0 as $\kappa \rightarrow \infty$. Then with $p = \frac{1+a-B}{1+a}$, we see that

$$R(\kappa(1 + \epsilon_\kappa)) = \log \left(\frac{1+a}{a} \kappa \right).$$

Therefore $R(\kappa) = \log \left(\frac{1+a}{a} \frac{\kappa}{1+\epsilon_\kappa} \right)$. ■

Remarks :

- 1) The coding scheme for the uniform noise channel (described in Section IV) stores information cell-by-cell, and therefore has low encoding and decoding complexity ($O(1)$ time per cell). In contrast, each codeword in the Gelfand-Pinsker scheme of Section II is defined over a large array of n cells, which makes the encoding and decoding computationally hard.
- 2) All results in this section generalize to the case where the hidden state uniformly distributed over a different support set of width B that is different from $[0, B]$.

TABLE I: Optimal value of δ_i for $1 \leq i \leq m$

i	1	2	3	4	5	6
δ_i	0.2032	0.1038	0.0858	0.0782	0.0740	0.0713

- 3) The coding scheme of construction 1 can directly be used when $B > a$. Construction 2 needs modification – the interior target region will not completely lie within the noise support when $B > a$. This can be addressed by shifting the input stimulus by an appropriate amount if it is determined that the offset $b > a$. This is similar in principle to the switching strategy used for the exterior regions.

IV. PROOF OF THEOREM 2

To highlight the main ideas, we start with a simplified coding scheme for the case of $m = 1$, i.e., two exterior target regions.

A. Two Exterior Target Regions

Coding Scheme: Fix $p \in [0, 1]$. For each cell, an interior target region is picked with probability p and an exterior region is picked with probability $(1-p)$. All interior target regions are equally likely, as are the exterior regions. Formally, each interior region has probability $p\frac{D}{1+a-B}$ and the two exterior regions each have probability $(1-p)/2$. Refer Fig. 7. To write on interior region $[x, x+D]$, repeatedly apply stimulus $(x+D-a/2)^+$ until the output falls within the region.

To write on exterior region E_1 : Apply stimulus 1 until either the output falls in $(1+a/2, 1+a/2+B/2)$, or it is less than $1-a/2+B/2$. If the former occurs, stop. Otherwise, apply stimulus 0 until the output falls in $(-a/2, -a/2+B/2)$. The intuition is that the right bin of E_1 is fully accessible with stimulus 1 if the offset lies in the interval $[B/2, 1]$. We switch to the left bin of E_1 if we detect that the offset lies outside this interval.

To write on exterior region E_2 : Apply stimulus 0 until either the output falls in $(-a/2+B/2, -a/2+B)$ or it is greater than $a/2+B/2$. If the former occurs, stop. Otherwise, switch to applying stimulus 1 until the output falls in $(1+a/2+B/2, 1+a/2+B)$. If the offset lies in the interval $[0, B/2]$, the left bin of E_2 is fully accessible with stimulus 0. We switch to the right bin of E_2 if we detect that the offset lies outside this interval.

The decoding procedure is simple. The index of the target region containing the output is the decoded message for each cell.

Analysis: Since we have two exterior target regions with probability $(1-p)/2$, and $\lfloor \frac{1+a-B}{D} \rfloor$ interior regions each with probability $p/\lfloor \frac{1+a-B}{D} \rfloor$, the rate of information stored in each cell is calculated to be

$$H(T) = h(p) + p \log \left\lfloor \frac{1+a-B}{D} \right\rfloor + (1-p) \log 2. \quad (25)$$

Next we compute the average number of writes and set it equal to κ .

$$\begin{aligned} \kappa &= p \mathbb{E}[\# \text{ writes} \mid \text{interior}] + (1-p) \mathbb{E}[\# \text{ writes} \mid \text{exterior}] \\ &= p\frac{a}{D} + (1-p) \mathbb{E}[\# \text{ writes} \mid \text{exterior}]. \end{aligned} \quad (26)$$

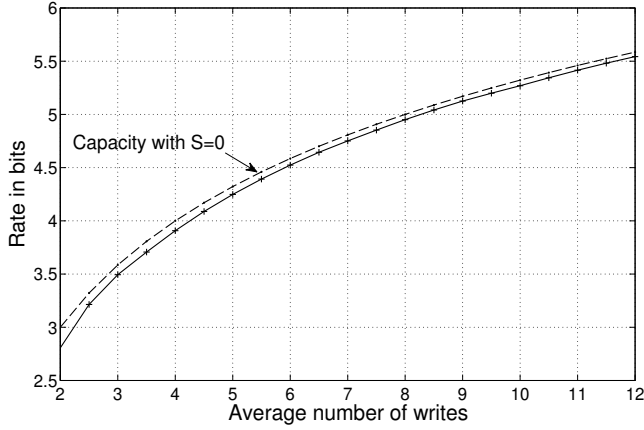


Fig. 6: Achievable rate of Theorem 2 with noise width $a = 1/3$ and S uniformly distributed in $[0, B]$ with $B = a/2$.

By symmetry,

$$\begin{aligned} \mathbb{E}[\# \text{ writes} \mid \text{exterior}] &= \mathbb{E}[\# \text{ writes} \mid \text{ext. region } E_1] \\ &= \int_0^B \mathbb{E}[\# \text{ writes} \mid E_1, S = b] \frac{1}{B} db \\ &= \int_0^{B/2} \mathbb{E}[\# \text{ writes} \mid E_1, S = b] \frac{1}{B} db + \int_{B/2}^B \frac{a}{B/2} \frac{1}{B} db \end{aligned} \quad (27)$$

since the right bin of E_1 is fully accessible with stimulus 1 when $S \in [B/2, B]$. We now show that for all $b \in [0, B/2]$,

$$\mathbb{E}[\# \text{ writes} \mid E_1, S = b] = 4a/B. \quad (28)$$

Recall that for E_1 , we first apply stimulus 1 until either the output falls in either $(1 + a/2, 1 + a/2 + B/2)$ or it is less than $1 - a/2 + B/2$. For $b \in [0, B/2]$, the probability of the first event occurring in any write attempt is b/a , and that of the second event occurring is $(B/2 - b)/a$. Hence the probability of the first step being completed in each write attempt, is

$$\frac{b}{a} + \frac{B/2 - b}{a} = \frac{B}{2a}. \quad (29)$$

Therefore the average number of writes for the first step of E_1 is $2a/B$ for all $b \in [0, B/2]$. The probability of the first step ending by obtaining an output less than $1 - a/2 + B/2$ is

$$\frac{(B/2 - b)/a}{b/a + (B/2 - b)/a} = \frac{B/2 - b}{B/2}.$$

When this event occurs, the average number of additional writes required (by applying stimulus 0) is $a/(B/2 - b)$. Thus for $b \in [0, B/2]$, the average number of writes for writing on E_1 is

$$\mathbb{E}[\# \text{ writes} \mid E_1, S = b] = \frac{2a}{B} + \frac{B/2 - b}{B/2} \cdot \frac{a}{B/2 - b} = \frac{4a}{B}. \quad (30)$$

Substituting in (27), we obtain

$$\mathbb{E}[\# \text{ writes} \mid \text{exterior}] = \frac{4a}{B} \frac{1}{2} + \frac{2a}{B} \frac{1}{2} = \frac{3a}{B}. \quad (31)$$

Using this in (26), we get

$$\kappa = p \frac{a}{D} + (1 - p) \frac{3a}{B} \quad (32)$$

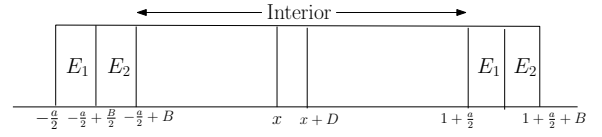


Fig. 7: Construction 2 with two exterior target regions.

which corresponds to (22) with $m = 1$ and $\delta_1 = 0$. We now modify the scheme slightly to reduce the average number of rewrites to the level stated in Theorem 2:

$$\kappa = p \frac{a}{D} + (1 - p) \left(\frac{3a}{B} + \ln \frac{1}{1 - \delta_1} + \frac{\delta_1}{1 - \delta_1} \ln \delta_1 \right) \quad (33)$$

with δ_1 given by Table I.

Optimizing the Switching Strategy: To write on exterior region 1 in Fig. 7, the above coding scheme switches from stimulus 1 to stimulus 0 when an output less than $1 - \frac{a}{2} + \frac{B}{2}$ is obtained. Such an output indicates that the value of the hidden state S is less than $\frac{B}{2}$ which implies that the right bin of E_1 – the region $[1 + \frac{a}{2}, 1 + \frac{a}{2} + \frac{B}{2}]$ – is not fully accessible with stimulus 1; so the scheme switches to targeting the left bin of E_1 with stimulus 0. This switching strategy is not optimal. Consider a more general switching strategy of the following form: switch from stimulus 1 to 0 once you obtain an output less than $1 - \frac{a}{2} + \frac{B}{2}(1 - \delta_1)$ for some $\delta_1 \in [0, 1]$. This corresponds to switching once you detect that S is less than $\frac{B(1 - \delta_1)}{2}$. We now determine the optimum value of δ_1 that minimizes the average number of rewrites. By symmetry, the switching strategy for exterior regions E_2 is to switch from stimulus 0 to 1 when you get an output greater than $\frac{a}{2} + \frac{B}{2}(1 + \delta_1)$.

The average number of rewrites for region E_1 is

$$\mathbb{E}[\# \text{ writes} \mid \text{region } E_1] = \int_0^B \mathbb{E}[\# \text{ writes} \mid E_1, S = b] \frac{1}{B} db \quad (34)$$

where $\mathbb{E}[\# \text{ writes} \mid E_1, S = b]$ with the new switching strategy can be calculated to be

$$\begin{aligned} &2ma/B && \text{for } \frac{B}{2} \leq b \leq B, \\ &a/b && \text{for } \frac{B}{2}(1 - \delta_1) \leq b < \frac{B}{2}, \\ &\frac{2a}{B(1 - \delta_1)} \left(1 + \frac{(1 - \delta_1)B - 2b}{B - 2b} \right) && \text{for } 0 \leq b < \frac{B}{2}(1 - \delta_1). \end{aligned} \quad (35)$$

Substituting this in (34) and calculating the integral, we obtain

$$\mathbb{E}[\# \text{ writes} \mid E_1] = \frac{a}{B} \left(3 + \ln \frac{1}{1 - \delta_1} + \frac{\delta_1}{1 - \delta_1} \ln \delta_1 \right). \quad (36)$$

Using (36) in (26) completes the proof for $m = 1$.

B. $2m$ Exterior Target Regions

Coding Scheme: Refer Fig. 5. Writing on the interior regions is the same as before: For interior region $[x, x + D]$, repeatedly apply stimulus $(x + D - a/2)^+$. To write on exterior region i , $1 \leq i \leq 2m$:

- If $1 \leq i \leq m$, write 1 until the output falls in region E_i or it is less than $1 - \frac{a}{2} + \frac{B(i - \delta_i)}{2m}$. In the first case, stop.

In the second case, switch to writing 0 until the output falls in the left bin of region i .

- If $m+1 \leq i \leq 2m$, write 0 until the output falls in region E_i or it is greater than $\frac{a}{2} + \frac{B(i-1+\delta_{2m+1-i})}{2m}$. In the first case, stop. In the second case, switch to writing 1 until the output falls in the right bin of region i .

Analysis: The rate calculation is straightforward, the only change from the previous subsection being that each of the exterior target regions now represents $\log 2m$ bits of information. The average number of writes for an interior target region is a/D . For an exterior region, we calculate it separately for each $E_i, i = 1, \dots, m$ as follows. Note that by symmetry, E_i and E_{2m+1-i} have the same average cost. We have

$$\mathbb{E}[\# \text{ writes} \mid \text{region } E_i] = \int_0^B \mathbb{E}[\# \text{ writes} \mid E_i, S = b] \frac{1}{B} db \quad (37)$$

where $\mathbb{E}[\# \text{ writes} \mid E_i, S = b]$ can be calculated to be

$$\begin{aligned} 2ma/B & \quad \text{for } \frac{Bi}{2m} \leq b \leq B, \\ \frac{2ma}{2mb - (i-1)B} & \quad \text{for } \frac{B(i-\delta_i)}{2m} \leq b < \frac{Bi}{2m}, \\ \frac{2ma}{B(1-\delta_i)} \left(1 + \frac{(i-\delta_i)B-2mb}{iB-2mb} \right) & \quad \text{for } \frac{B(i-1)}{2m} \leq b < \frac{B(i-\delta_i)}{2m}, \\ \frac{2ma}{(i-\delta_i)B-2mb} + \frac{2ma}{B} & \quad \text{for } 0 \leq b < \frac{B(i-1)}{2m}. \end{aligned} \quad (38)$$

Using this in (34) and calculating the integral, we obtain that for $1 \leq i \leq m$:

$$\mathbb{E}[\# \text{ writes} \mid E_i] = \frac{a}{B} \left(2m + 1 + \ln \frac{i - \delta_i}{(1 - \delta_i)^2} - \frac{\delta_i}{1 - \delta_i} \ln \frac{1}{\delta_i} \right). \quad (39)$$

The average number of write attempts for an exterior region is therefore

$$\frac{a}{B} \left[2m + 1 + \frac{1}{m} \sum_{i=1}^m \left(\ln \frac{i - \delta_i}{(1 - \delta_i)^2} - \frac{\delta_i}{1 - \delta_i} \ln \frac{1}{\delta_i} \right) \right].$$

For each $i \in \{1, \dots, m\}$, it is easily verified that the $\delta_i \in [0, 1]$ that minimizes (39) satisfies (23). This completes the proof.

V. CONCLUSION

In a channel with unknown parameters (modeled by a hidden state), rewrites increase the capacity in two ways: 1) by mitigating the effect of write noise, and 2) by enabling the write controller to get progressively better estimates of the state. For the uniform noise channel, one of the key observations was that the hidden state does not affect coding in the interior region. This idea could be generalized to other channels where the output has bounded support.

There are many open questions to be explored. One is obtaining a capacity upper bound, which is challenging as we need to consider all adaptive input strategies. The general capacity lower bound can be improved via a scheme that does simultaneous coding and estimation; the challenge here lies in analyzing such a coding scheme to get a computable expression for the achieved rate. Another goal is to modify the superposition scheme so that it robust to small amounts of

read noise. As discussed at the end of Section II, the current scheme requires the reads to be highly accurate.

The channel model analyzed here is motivated by non-volatile memories such as Phase Change Memory and Resistive RAM. The coding schemes illustrate how information-theoretic techniques like superposition can be used to increase the storage density. Though the schemes presented are for analog storage channels, the ideas can be extended to finite alphabet channels which arise in technologies such as Magnetic RAM [27]. For memories such as Flash, a more relevant channel model would be one in which a write can only increase the previous value stored in the cell; extending the ideas in this paper to such a model is an important direction for future work. Another interesting possibility is extending the model to account for stochastic variation of the cell contents over time, a phenomenon which is encountered in most memory technologies and which manifests as a read noise at the ‘‘receiver’’. We believe that developing efficient rewritable schemes for such realistic models will have a significant impact on several non-volatile memory technologies.

APPENDIX

Proof of Theorem 1:

Fix an estimation period $l \in \{0, \dots, \lfloor \kappa - 1 \rfloor\}$, an estimator $\hat{S}(l)$, a distribution $P_{U|\hat{S}(l)}$, and a function f to generate the channel input $X = f(U, \hat{S}(l))$. This defines a joint distribution of $(S, \hat{S}(l), U, X, Y)$ in the set \mathcal{P} .

Construct a codebook consisting of 2^{nR_1} codewords, whose elements are picked i.i.d. according to P_U , the marginal distribution of the auxiliary random variable U . This codebook is partitioned in 2^{nR} bins where

$$R < I(U; Y) - I(U; \hat{S}(l)). \quad (40)$$

$R_1 > R$ will be specified later.

The output space of each cell is divided into $\lfloor \kappa - l \rfloor$ target regions, as described in Section II-A (see Fig. 2).

Encoding: The message to be stored in the n -cell array consists of two parts (m_1, m_2) , where $m_1 \in \{1, \dots, 2^{nR}\}$ and $m_2 \in \{1, \dots, (\lfloor \kappa - l \rfloor)^n\}$. Let $\hat{\mathbf{S}}(l)$ be the state estimate obtained using the first l writes. To encode the first part of the message, we choose a codeword \mathbf{U} from the m_1 th bin such that the pair $(\mathbf{U}, \hat{\mathbf{S}}(l))$ is jointly typical [24, Section 8.2] according to the distribution described by the following joint density:

$$P_{U, \hat{S}(l)}(u, \hat{s}) = \int_{\mathcal{S}} P_S(s) P_{\hat{S}(l)|S}(\hat{s}|s) P_{U|\hat{S}(l)}(u|\hat{s}) ds. \quad (41)$$

From rate-distortion theory [24], such a codeword \mathbf{U} can be found with high probability if

$$R_1 - R > I(U; \hat{S}(l)). \quad (42)$$

(42) gives a lower bound on the minimum number of codewords in each bin ($2^{n(R_1 - R)}$) required for successful encoding. The input stimulus $\mathbf{X} = \{X_i\}_{i=1}^n$ is generated symbol by symbol as $X_i = f(U_i, \hat{S}_i(l))$.

The second part of the message is conveyed through superposition coding. For cell i , apply stimulus X_i until the output

falls in the appropriate target region. For any realization of the input stimulus and state, the output is equally likely to fall in each of the target regions with probability $\frac{1}{\lfloor \kappa - l \rfloor}$. Hence the average number of writes required after the estimation period is $\lfloor \kappa - l \rfloor$, and the average total writes per cell is $l + \lfloor \kappa - l \rfloor$.

Decoding: The decoder attempts to find a codeword $\hat{\mathbf{U}}$ that is jointly typical with the stored sequence \mathbf{Y} according to (41). If there is a unique such codeword, its bin is decoded as the message m_1 . The target region containing the output of each cell gives the message m_2 . The codeword \mathbf{U} can be successfully decoded if the rate of the codebook satisfies

$$R_1 < I(U; Y). \quad (43)$$

Combining (42) and (43), we conclude that \mathbf{U} can be successfully encoded and decoded if (40) is satisfied.

We have thus shown that a total of $R + \log \lfloor \kappa - l \rfloor$ bits/cell can be reliably stored and decoded with average write cost $l + \lfloor \kappa - l \rfloor$ as long as R satisfies (40).

REFERENCES

- [1] R. Venkataramanan, S. Tatikonda, L. Lastras-Montano, and M. Franceschini, "Coding strategies for the uniform noise rewritable channel with hidden state," *Proc. IEEE Int. Symp. Inf. Thy.*, 2012.
- [2] M. Grossi, M. Lanzoni, and B. Ricco, "Program schemes for multilevel flash memories," *Proc. of the IEEE*, vol. 91, pp. 594 – 601, April 2003.
- [3] G. W. Burr, M. J. Breitwisch, M. Franceschini, D. Garetto, K. Gopalakrishnan, B. Jackson, B. Kurdi, C. Lam, L. A. Lastras, A. Padilla, B. Rajendran, S. Raoux, and R. Shenoy, "Phase change memory technology," *Journal of Vacuum Science & Technology B*, vol. 28, no. 2, 2010.
- [4] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80–83, 2008.
- [5] R. Waser and M. Aono, "Nanoionics-based resistive switching memories," *Nature Materials*, vol. 6, no. 11, pp. 833–840, 2007.
- [6] E. Yaakobi, S. Kayser, P. H. Siegel, A. Vardy, and J. K. Wolf, "Codes for write-once memories," *IEEE Transactions on Information Theory*, vol. 58, no. 9, pp. 5985–5999, 2012.
- [7] E. Yaakobi, H. Mahdaviifar, P. H. Siegel, A. Vardy, and J. K. Wolf, "Rewriting codes for flash memories," 2012. Online: <http://arxiv.org/abs/1210.7515>.
- [8] L. Lastras-Montano, M. Franceschini, T. Mittelholzer, and M. Sharma, "Rewritable storage channels," in *Int. Symp. on Inf. Theory and its Applications*, 2008.
- [9] T. Mittelholzer, M. Franceschini, L. Lastras-Montano, I. Elfadel, and M. Sharma, "Rewritable channels with data-dependent noise," in *IEEE Int. Conf. on Comm.*, 2009.
- [10] L. Lastras-Montano, T. Mittelholzer, and M. Franceschini, "Superposition coding in rewritable channels," in *Information Theory and Applications Workshop (ITA)*, 2010.
- [11] L. Lastras-Montano, M. Franceschini, and T. Mittelholzer, "The capacity of the uniform noise rewritable channel with average cost," in *Proc. IEEE Int. Symp. on Inf. Theory*, 2010.
- [12] L. Lastras-Montano, M. Franceschini, T. Mittelholzer, and M. Sharma, "On the capacity of memoryless rewritable storage channels," *Submitted to IEEE Trans. Inf. Thy.*, 2012.
- [13] T. Mittelholzer, L. Lastras-Montano, M. Sharma, and M. Franceschini, "Rewritable storage channels with limited number of rewrite iterations," in *Proc. IEEE Int. Symp. on Inf. Theory*, June 2010.
- [14] C. Bunte and A. Lapidoth, "On the storage capacity of rewritable memories," in *IEEE Conv. of EE Engineers in Israel*, Nov. 2010.
- [15] C. Bunte and A. Lapidoth, "Computing the capacity of rewritable memories," in *Proc. IEEE Int. Symp. on Inf. Theory*, 2011.
- [16] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. on Inf. Theory*, vol. 56, pp. 5396 –5411, Nov. 2010.
- [17] A. Jagmohan, L. Lastras-Montano, M. Franceschini, M. Sharma, and R. Cheek, "Coding for multilevel heterogeneous memories," in *Proc. IEEE International Conf. on Communications (ICC)*, 2010.
- [18] N. Papandreou, H. Pozidis, A. Pantazi, A. Sebastian, M. Breitwisch, C. Lam, and E. Eleftheriou, "Programming algorithms for multilevel phase-change memory," in *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS)*, 2011.
- [19] S. Gelfand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [20] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, 1983.
- [21] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2524–2530, Oct. 1998.
- [22] H. Lee, Y. Chen, P. Chen, P. Gu, Y. Hsu, S. Wang, W. Liu, C. Tsai, S. Sheu, P. Chiang, et al., "Evidence and solution of over-RESET problem for HfO_x based resistive memory with sub-ns switching speed and high endurance," in *IEEE Int. Electron Devices Meeting (IEDM)*, pp. 19–7, 2010.
- [23] S. M. Kay, *Fundamentals of Statistical Signal Processing*, vol. 1. Prentice Hall, 1993.
- [24] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley-interscience, 2006.
- [25] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, 2005.
- [26] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.
- [27] M. Hosomi, H. Yamagishi, T. Yamamoto, K. Bessho, Y. Higo, K. Yamane, H. Yamada, M. Shoji, H. Hachino, C. Fukumoto, H. Nagao, and H. Kano, "A novel nonvolatile memory with spin torque transfer magnetization switching: Spin-ram," in *IEEE Int. Electron Devices Meeting, IEDM Technical Digest.*, pp. 459–462, 2005.



Ramji Venkataramanan (S'06-M'09) received the B.Tech degree in Electrical Engineering from the Indian Institute of Technology, Madras in 2002, and the Ph.D degree in Electrical Engineering (Systems) from the University of Michigan, Ann Arbor in 2008. He held post-doctoral positions at Stanford University and Yale University before joining the University of Cambridge in 2013, where he is a University Lecturer in the Department of Engineering and a Fellow of Trinity Hall. His current research interests are in information theory, coding, and statistical inference.



Sekhar Tatikonda (S'92-M'00-SM'13) received the B.S. (1993), M.S. (1995), and Ph.D. (2000) in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology, Cambridge, MA. He was a postdoctoral fellow at the University of California, Berkeley from 2000-2002. In 2002, he joined the Electrical Engineering Department at Yale University, New Haven, CT, where he is currently an associate professor. His research interests are in communication theory, information theory, stochastic control, distributed optimization, statistical machine learning, and inference.



Luis Lastras-Montano (S'99-M'01-SM'06) received the Bachelor in Science degree in Electrical Engineering from the Universidad Autónoma de San Luis Potosí, México in 1995, and the M.Sc. and Ph.D. degrees in Electrical Engineering from Cornell University in 1998 and 2000. Since 2000 he has been with the IBM T. J. Watson Research Center in Yorktown Heights, NY, where he is a Research Staff Member, a Member of the IBM Academy of Technology, a Master Inventor and a manager of a research group focused on computer memory. His current research interests are in information theory and coding theory, in particular applied to computer memories.



Michele Franceschini (S'02-M'08) received the M.Sc and Ph.D degrees from the University of Parma, Italy. From 2006 to 2008 he was a research associate at University of Parma. From 2008 to 2009 he was a postdoc at IBM T. J. Watson Research Center, Yorktown Heights, NY. Since 2009, he has been a Research Staff Member at IBM T.J. Watson Research Center. His research interests are in the general areas of communication theory and information theory, with emphasis on iterative detection techniques, code design and signal processing for non-volatile memories.