# Superposition coding in rewritable channels

L.A. Lastras-Montaño,
IBM T. J. Watson Research Center
e-mail: lastrasl@us.ibm.com

T. Mittelholzer
Zurich Research Laboratory
e-mail: tmi@zurich.ibm.com

M. M. Franceschini
IBM T. J. Watson Research Center
e-mail: franceschini@us.ibm.com

***Abstract*—We introduce the notions of superposition coding and sequential decoding in the context of rewritable channels. Using these concepts we will show that for $\kappa_1 > \kappa_0$, $C(\kappa_1) \geq C(\kappa_0)+\log\left(\frac{\kappa_1}{\kappa_0}\right)$, where $C(\cdot)$ is the capacity of the rewritable channel for a given cost. A consequence of this result is that $C(\kappa) \geq C(1) + \log \kappa$, where $C(1)$ is the classical channel capacity with no rewrite iterations. Thus this result provides a connection between rewritable and classical channel theory. We also derive a general upper bound on capacity which can be written as an offset plus the logarithm of the average number of write iterations. Closed form bounds on rewritable channel capacity will be given for Gaussian rewritable channels.**

## I. Introduction

The theory of rewritable channels [1][2][3][4][5] studies the tradeoff between the information storage capacity of a medium and the cost for writing into it. One of the key elements of this theory is the concept of *write noise* in which a write attempt results in general in a written value different from the one intended.

In order to cope with write noise, a *write controller* reads what has been written into the medium after a write attempt, to determine what it is that has been actually stored. If the write controller judges a stored value to be unsuitable for transmitting effectively a message to a subsequent reader of the memory, then it can retry the write attempt, possibly adjusting its input to the memory so as increase the probability of a favorable outcome. Because rewriting is key to this procedure, we call this type of medium a rewritable channel. As one increases the average number of write iterations allowed when imprinting the memory with a message one also increases the storage capacity of the medium, since loosely speaking, the write iterations create a feedback loop that improves the memory channel's signal to noise ratio.

A main motivation for introducing and studying rewritable channels is that there are important storage channels with these general characteristics, particularly in the form of Phase-Change-Memory [6] [7] and FLASH [8]. In the latter, multiple bits/cell storage is attained in practical devices by employing an iterative feedback algorithm that is referred to as a "write-and-verify" strategy. Also in demonstrations of multiple bit/cell PCM [9][10][11], similar techniques are exploited.

A key element of a rewritable channel is a description of the statistical behavior of a *memory cell*. Most of the progress in rewritable channel theory has been focused on the case that a write attempt outcome is given by the input plus a random variable uniformly distributed on a real interval, which we call the uniform noise rewritable channel[1].

In this article we present new results on rewritable channel theory that hold for much more general rewritable channels. These results are necessarily not as sharp as those available for more specific settings - for example the uniform noise rewritable channel capacity with an average cost constraint has been recently characterized completely [5]. Nonetheless, the results we give in here have several satisfying properties. For example, we have found that rewritable channel capacity is always at least the classical channel capacity of the medium (with no rewrites) plus the logarithm of the allowed average number of write attempts. We also present a new upper bound on rewritable channel capacity particularly relevant for peak input constrained rewritable channels. This upper bound can be written as an offset plus the logarithm of the average number of iterations. Thus to a first order, the qualitative behavior of the capacity/cost functions of an important class of rewritable channels has now been found.

One of the key techniques that we describe for obtaining capacity lower bounds is based on the notion of *superposition coding*, in which a decoder employs *sequential decoding* in order to fully decode the message encoded in the memory. While these techniques are well known in information theory in the context of multiuser communications [12], they have not been applied, to our knowledge, to the problem of storage of information. In this context, we use them to create two virtual memories out of a physical memory. The first virtual memory appears to a decoder as a classical memory in which a single write (no rewrite iterations) was used to encode information. Upon decoding the message conveyed in the first virtual memory, the second virtual memory is decoded which contains additional message bits. While the amount of information that one may store in first virtual memory by definition does not change with the average number of iterations, the same is not necessarily true of the second virtual memory, whose capacity, as we will show, grows at least as the logarithm of the average number of iterations. The total rewritable channel capacity is then the sum of these two capacities.

Most of the insight behind the use of superposition coding in rewritable storage channels can be found in an example for the Gaussian rewritable channel that we present at the

[1]We remark that the recent submission [4] contains results both for the uniform noise setting as well as more general rewritable channels

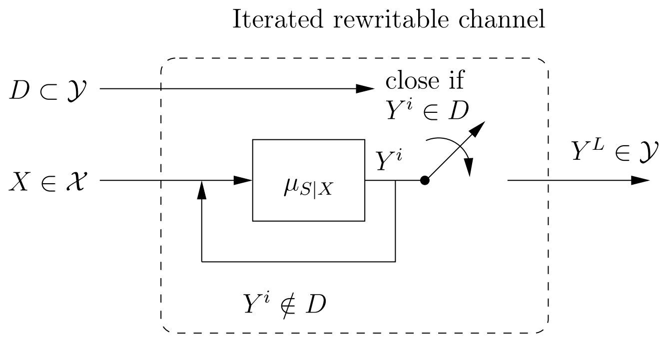


Fig. 1. Information theoretic model for the rewritable channel. The basic rewritable channel characteristics are captured by the conditional probability law $\mu_{S|X}$.

beginning of Section III. The proof of the main result is also given in Section III, with an important partial result proved in the paper's Appendix in order to improve the readability of the result. The upper bound on capacity is presented in Section IV; in Section V we illustrate the results with specific examples involving the Gaussian rewritable channel.

## II. Preliminaries

A memory is modeled as having *n memory cells*, which are the basic entities that are able to store information as a result of an input stimulus, and from which information may be retrieved during a read action. In order to specify a rewritable channel, one provides

- An input alphabet $\mathcal{X}$, which is the space of possible input *stimuli* to a memory cell.
- An output alphabet $\mathcal{Y}$ which contains the possible set of values that one may read from a memory cell.
- A conditional probability law $\mu_{S|X}$ which establishes the probabilistic relation between an input stimulus $x \in \mathcal{X}$, and a resulting internal cell state $s \in \mathcal{Y}$ that a memory cell will take on during a single write attempt.

We assume in this article that $\mathcal{X}, \mathcal{S} \subset \mathcal{R}$ and that the internal state is directly observable without noise by a reader, which is the reason we do not need to specify a different alphabet for the set of possible states. We also assume that the channel $\mu_{S|X}$ is known to the encoder and decoder of the memory, and that each memory cell obeys this same probabilistic relation. We also assume that when conditioned on the value of their input stimuli, the resulting states of the memory cells are statistically independent.

The conditional probability law is interpreted to be a function $\mu_{S|X}$ assigns probabilities to events which in turn are subsets of $\mathcal{Y}$. We employ the notation

$$\mu_{S|X}(\delta|x)$$

to denote the probability that the memory cell state falls in the set $\delta \subset \mathcal{Y}$ when the input stimulus to the cell is $X = x$.

We refer the reader to Figure 1, which illustrates how the *iterated rewritable channel* (also termed simply the *iterated channel*) is derived from the basic law $\mu_{S|X}$. The iterated channel has two inputs: the stimulus $X$ and the encoding set $D$. For every $i \geq 1$, let $Y^i$ be the result of passing the random variable $X$ through an instance of the channel $\mu_{S|X}$; the statistical independence assumption then says that conditioned on $X$, $\{Y^i\}_{i=1}^{\infty}$ is an i.i.d. process. The role of the encoding set $D$ is to define a criterion for accepting or rejecting the outcome of a write attempt; more precisely, the first instance in which the memory cell's state is inside of $D$, we will exit the iterative rewrite loop. Define

$$L = \min\{i \geq 1 : Y^i \in D\}$$

The random variable $L$ is the number of write attempts one incurs when the inputs to the rewritable channel are the stimulus $X$ and the encoding set $D$. The average cost for writing is then $E\{L\}$, where $E\{\cdot\}$ denotes the mathematical expectation operator.

The storage capacity of the rewritable channel when the average number of allowed write attempts is at most $\kappa$ is given by the formula [1]

$$C(\kappa) = \sup_{X \in \mathcal{X}, D \subset \mathcal{Y}, E\{L\} \leq \kappa} I(X, D; Y^L), \tag{1}$$

where the maximization is interpreted to be with respect to all possible marginal input distributions for $X$ and $D$. In some problems we may be interested in further restricting the class of input distributions so as to incorporate a *stimulus cost constraint*, not to be confused with the constraint on the number of iterations. In particular, let $\rho : \mathcal{X} \to \mathcal{R}$ be a stimulus cost function, then the capacity of the rewritable channel with an constraint of $\kappa$ average iterations and $\rho^*$ average stimulus cost is given by

$$C(\kappa) = \sup_{X \in \mathcal{X}, E\{\rho(X)\} \leq \rho^*, D \subset \mathcal{Y}, E\{L\} \leq \kappa} I(X, D; Y^L). \tag{2}$$

Throughout this proof, we shall make use of tools from conditional expectation theory; in particular we will rely on the fact that for any two random variables $A, B$, $E\{A\} = E\{E\{A|B\}\}$. We shall also utilize the concept of an *absolutely continuous* probability measure. If $\mu$ and $\nu$ are two measures defined over the same measurable space, $\mu$ is absolutely continuous with respect to $\nu$ if for every measurable set $a$ the statement $\nu(a) = 0$ implies $\mu(a) = 0$. We will use the notation $\lambda(\cdot)$ to denote the Lebesgue measure on the real line, which in essence measures the length of a set. We emphasize that these concepts are used to add rigor to the proof but nonetheless the main ideas of the proofs may be grasped without recourse to these tools.

## III. Lower bounds using superposition coding

The role that superposition coding and sequential decoding play in rewritable channels derives from a simple observation based on the chain rule for mutual information. The rewritable channel model that we study in this article (see Figure 1) has two inputs: a stimulus signal $X$ that is used as a physical input to the cell, as well as a set $D$ that determines when it is that the iterative write algorithm will finish. For a given marginal distribution on $X, D$, the corresponding storage rate is given by

$$I(X, D; Y^L) = I(X; Y^L) + I(D; Y^L|X).$$

The rewriting of the mutual information as a sum of two mutual informations emphasizes the fact that *sequential decoding* can be used when interpreting the contents of a memory. In particular, if we can build a decoder for recovering the first $I(X; Y^L)$ bits by, in a loose sense, retrieving $X$, then in principle we can recover an additional $I(D; Y^L|X)$ bits by building a decoder for the channel whose input is $D$, output is $Y^L$, under the assumption that both the encoder and decoder know $X$.

One can take this analogy further and construct a mechanism for storing information at a cost $\kappa_1 > \kappa_0$ by using an existing technique for storing information at cost $\kappa_0$ as a scaffold. These observations are the basis of the following result:

*Theorem 1:* Assume a rewritable channel with conditional probability law $\mu_{S|X}$. Further assume that for every $x \in \mathcal{X}$, $\mu_{S|X}(\cdot|x)$ is absolutely continuous with respect to the Lebesgue measure. For any $1 \leq \kappa_0 < \kappa_1$, the capacity/cost function $C(\cdot)$ of the rewritable channel under an average number of iterations constraint satisfies

$$C(\kappa_1) \geq C(\kappa_0) + \log\left(\frac{\kappa_1}{\kappa_0}\right).$$

*Note:* This result holds for both the setting in which there is no stimulus cost constraint and when there is one. The associated capacity characterizations can be found in (1) and (2), respectively.

In order to motivate the arguments used to prove Theorem 1 we discuss informally the superposition coding concept in the following subsection, using an example with Gaussian write noise.

### A. An example using the Gaussian rewritable channel

Suppose that the memory cell write noise statistics $\mu_{S|X}$ are such that the state of the cell after a write attempt is equal to the input stimulus plus an additive offset that is a unit variance Gaussian random variable. We shall further assume that we will require an average input stimulus constraint of the form $E\{X^2\} \leq \rho^*$; this does not necessarily relate to any physically meaningful constraint on any memory and only used to make it easier for the reader to grasp our point. Let $M > 1$ an integer, and partition the Gaussian density function in $M$ bins each having the same probability. For the sake of simplicity, we will assume that these bins are open intervals, possibly stretching out to $-\infty$ or $+\infty$. Figure 2 shows a specific such partition for the case $M = 5$. Define these open intervals as

$$\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_M$$

where $\mathcal{B}_i \subset \mathcal{R}$.

Let $n$ be the number of cells in a codeword of a code for the rewritable channel. Let $\epsilon > 0$ be a parameter, and let $C_\mathcal{G} \subset \mathcal{R}^n$ be a good code for the classical Gaussian channel with unit

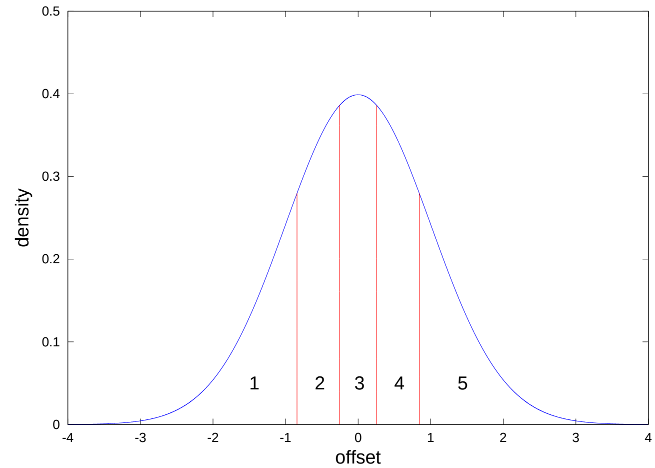


Fig. 2. An example of a partition for the unit variance, zero mean Gaussian density which has $M = 5$ bins, each with probability $1/5$. These bins are identified with a distinct integer.

noise power and average input power constraint $\rho^*$. We assume that $C_\mathcal{G}$ satisfies

$$\frac{1}{n}\log|C_\mathcal{G}| \geq C(1) - \epsilon \tag{3}$$

Thus, by good we mean a capacity achieving code with vanishing probability of decoding error. Next, let $C_\mathcal{B} \subset \{\mathcal{B}_1, \cdots, \mathcal{B}_M\}^n$ be a code with

$$\frac{1}{n}\log|C_\mathcal{B}| \geq M - \epsilon \tag{4}$$

and additionally with the property for every $\mathbf{b} \in C_\mathcal{B}$,

$$\sum_{i=1}^{M}\left|\frac{N(\mathcal{B}_i|\mathbf{b})}{n} - \frac{1}{M}\right| < \epsilon \tag{5}$$

where $N(\mathcal{B}_i|\mathbf{b})$ stands for the number of instances of $\mathcal{B}_i$ within the vector $\mathbf{b}$. We assume $n$ to be large enough so that (3), (4) and (5) can be accomplished.

We now refer the reader to Figure 3. Split a message to be written into the memory into two sections, one with rate $C(1) - \epsilon$ bits/cell, and the other one with rate $\log(M) - \epsilon$ bits/cell. Select a codeword $\tilde{\mathbf{x}} \in C_\mathcal{G}$ according to the first message section, and select a codeword $\tilde{\mathbf{b}} \in C_\mathcal{B}$ according to the second message section.

Next define

$$\tilde{\mathbf{d}} = \tilde{\mathbf{b}} + \tilde{\mathbf{x}}$$

where the addition is meant to shift each set in $\tilde{\mathbf{b}}$ with an additive offset given by the corresponding entry in $\tilde{\mathbf{x}}$.

Next, the vectors $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{d}}$ are used to select, for each of the $n$ memory cells, the corresponding input stimulus and the encoding set. Figure 3 illustrates a sample codeword formed in this manner.

Because of the property (5) and due to the construction of the partition of the Gaussian with $M$ equal area bins, it is not very difficult to see that if $\epsilon$ is small enough, then

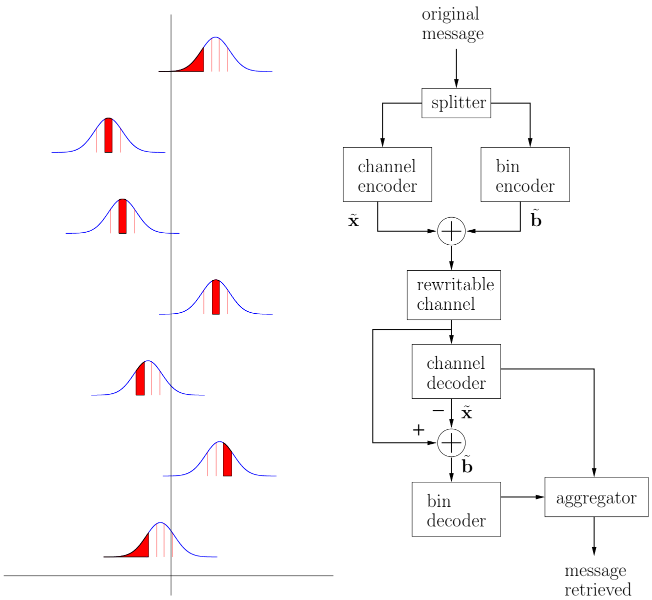


Fig. 3. In the left we give an example of a codeword of a superposition code for the Gaussian rewritble channel with $n = 7$ and $M = 5$. The center of each Gaussian is given by a vector $\tilde{\mathbf{x}} \subset C_{\mathcal{G}}$, while the bin shown as a filled area is selected by a vector $\tilde{\mathbf{b}} \subset C_{\mathcal{B}}$; the codeword is given by $\tilde{\mathbf{x}} + \tilde{\mathbf{b}}$. In the right we show a general diagram of the encoding/decoding procedures for a superposition code.

a decoder can retrieve, with very high probability[2] the first message section using a decoder for the code $C_{\mathcal{G}}$. This implies that the decoder has also learned $\tilde{\mathbf{x}}$ and thus it may subtract it from the codeword read from the memory in order to deduce $\tilde{\mathbf{b}}$, from which the second message section can be retrieved.

By using the formula for the capacity of the Gaussian channel, we have then argued that this superposition code allows us to store approximately

$$\frac{1}{2}\log\left(1+\frac{1}{\sigma_W^2}\right)+\log(M) \tag{6}$$

bits/cell at a cost of $M$ average iterations (since each bin has probability $1/M$ on a given write attempt). Thus $C(M)$ is lower bounded by (6). In what follows, we prove the more general statement in Theorem 1 using a generalization of the above.

### B. Proof preliminaries

The key structure in the proof rests on the idea that a capacity achieving input distribution for cost $\kappa_0$ will be used to construct another (non necessarily capacity achieving) input distribution for cost $\kappa_1$. Because of this, it will be necessary to introduce two related iterated channels. Both iterated channels will have a common stimulus input random variable $X$. Following Section II, let $\{Y^i\}_{i=1}^{+\infty}$ be a random process that is obtained by passing $X$ through statistically independent copies of the channel $\mu_{S|X}$. It is clear that the random process $\{Y^i\}_{i=1}^{+\infty}$ will be also shared between the two iterated channels.

[2]The high probability qualifier can be easily converted into complete certainty in this model; we do not give the details of the corresponding construction in this article.

The iterated channel for cost $\kappa_0$ will use for the encoding set the input random variable $D_0 \subset \mathcal{Y}$, with the associated number of write iterations being

$$L_0 = \min\{i \geq 1 : Y^i \in D_0\}.$$

Similarly, the iterated channel for cost $\kappa_1$ will use for the encoding set the input random variable $D_1 \subset \mathcal{Y}$, and as before, the associated number of write iterations is defined as

$$L_1 = \min\{i \geq 1 : Y^i \in D_1\}.$$

Because in both instances the same input $X$ is chosen, the average stimulus cost constraint (if any) for the input random variables constructed for iteration cost $\kappa_1$ will be identical to the one for iteration cost $\kappa_0$. We are now ready to develop the proof of the result.

### C. Proof of Theorem 1

Let $\epsilon > 0$, and let $X, D_0$ be random variables such that for the rewritable channel with conditional probability law $\mu_{S|X}$, output process $\{Y^i\}_{i=1}^{\infty}$ and random cost $L_0$, the average cost satisfies

$$E\{L_0\} \leq \kappa_0 \tag{7}$$

and furthermore

$$I(X, D_0; Y^{L_0}) \geq C(\kappa_0) - \epsilon$$

We additionally assume that any existing average stimulus cost constraint is satisfied by $X, D_0$. If $\kappa_0 = 1$, then we choose $D_0 = \mathcal{Y}$, and we choose $X$ so as to achieve the classical capacity of the channel $\mu_{S|X}$ within $\epsilon$ bits/cell. For any $d \subset \mathcal{Y}$ and $x \in \mathcal{X}$ with $\mu_{S|X}(d|x) > 0$, let $\mu_{Y^{L_0}|X,D_0}(\cdot,|x,d)$ denote the conditional probability law of $Y^{L_0}$ given $X$ and $D_0$. Using the definition of the iterated channel, it can be easily shown that for $a \subset \mathcal{Y}$,

$$\mu_{Y^{L_0}|X,D_0}(a|x,d) = \frac{\mu_{S|X}(a \cap d|x)}{\mu_{S|X}(d|x)} \tag{8}$$

Let $a \subset \mathcal{Y}$ be a set with Lebesgue measure zero, this is, $\lambda(a) = 0$. Because $\mu_{S|X}(\cdot|x)$ is absolutely continuous, it follows that $\mu_{S|X}(a|x) = 0$ and since $\mu_{S|X}(a \cap d|x) \leq \mu_{S|X}(a|x) = 0$, we thus conclude that whenever it is defined, the measure $\mu_{Y^{L_0}|X,D_0}(\cdot|x,d)$ is also absolutely continuous with respect to the Lebesgue measure.

It is a known fact from probability theory that the cumulative distribution function of a random variable whose probability law is absolutely continuous with respect to the Lebesgue measure is a continuous function in the standard real analysis sense. Applied to the setting at hand, we see that if $\mu_{S|X}(d|x) > 0$, the cumulative distribution function

$$F_{Y^{L_0}|X,D_0}(\xi|x,d) = \frac{\mu_{S|X}((-\infty,\xi] \cap d|x)}{\mu_{S|X}(d|x)} \tag{9}$$

is a continuous function of $\xi$. Define the *gain* as

$$g \stackrel{\Delta}{=} \kappa_1/\kappa_0 \tag{10}$$

which necessarily satisfies $g > 1$. For any $\phi \in (0,1)$, define the set $\gamma(\phi) \subset (0,1)$ as follows:

$$\gamma(\phi) = \begin{cases} (\phi, \phi + 1/g) & \text{if } \phi + 1/g < 1 \\ (\phi, 1) \cup (0, -1 + \phi + 1/g) & \text{if } \phi + 1/g \geq 1 \end{cases} .$$

Next, for any $d \subset \mathcal{Y}$, $x \in \mathcal{X}$ and $\phi \in (0,1)$, let

$$\pi_{d,x}(\phi) = \left\{ \xi \in d : F_{Y^{L_0}|X,D_0}(\xi | x, d) \in \gamma(\phi) \right\}$$

We now make use of the following basic result (see for example Billingsley [13], Section 14).

*Lemma 1:* Let $A$ be a real valued random variable with a continuous cumulative distribution function $F_A(\xi) = \mu_A((-\infty, \xi])$. Then

$$\mu_A(\{a : F_A(a) \leq u\}) = u.$$

Using this lemma, we can then see that

$$\frac{\mu_{S|X}(\pi_{d,x}(\phi)|x)}{\mu_{S|X}(d|x)} = \frac{1}{g}. \tag{11}$$

Now let $\Phi$ be a random variable taking values on the alphabet $[0,1]$ that additionally is statistically independent from $X$ and $D_0$. The encoding set for cost $\kappa_1$ is now defined as

$$D_1 = \pi_{D_0,X}(\Phi) \tag{12}$$

We now consider $X, D_1$ to be the input distribution to the rewritable channel $\mu_{S|X}$. The associated average cost can be evaluated with

$$\begin{aligned} E\{L_1\} &\overset{(a)}{=} E\{E\{L_1|X, D_1\}\} \\ &\overset{(b)}{=} E\left\{\frac{1}{\mu_{S|X}(D_1|X)}\right\} \\ &\overset{(c)}{=} E\left\{\frac{1}{\mu_{S|X}(\pi_{D_0,X}(\Phi)|X)}\right\} \\ &\overset{(d)}{=} E\left\{\frac{g}{\mu_{S|X}(D_0|X)}\right\} \\ &\overset{(e)}{\leq} \kappa_1. \end{aligned} \tag{13}$$

In this development, (a) follows from the basic properties of conditional expectation, (b) follows from the fact that the mean of a geometric distribution with a trial success probability $p$ is $1/p$, (c) follows from the definition of $D_1$ in (12), and (e) follows from the assumption (7). The step (d) follows from (11), which does not have any fundamental restriction on $x, d, \phi$, and hence will also hold in the case the arguments are random variables.

We remark that the result $E\{L_1\} \leq \kappa_1$ holds regardless of our choice for the marginal distribution of $\Phi$. Nonetheless, in what follows we will consider two explicit choices for the random variable $\Phi$. The first choice works only when $g > 1$ is an integer. In this choice, $\Phi$ is a discrete random variable uniformly distributed on the set

$$\left\{0, \frac{1}{g}, \frac{2}{g}, \cdots, \frac{g-1}{g}\right\}$$

In the second choice, which works for all $g > 1$, $\Phi$ will be a random variable uniformly distributed on the interval $(0,1)$. In both cases, as stated previously, $\Phi$ will be statistically independent of $X$ and $D_0$.

Since the second choice works for all $g > 1$, it suffices to prove the theorem. Nonetheless the first choice is associated with a far simpler decoding scheme and thus we believe there is value in including it in this proof. Note that the example in subsection III-A uses a scheme based on the first choice.

Using the chain rule for mutual information, write

$$\begin{aligned} I(X, D_0, D_1; Y^{L_1}) &= I(X, D_0; Y^{L_1}) + I(D_1; Y^{L_1}|X, D_0) \\ &= I(X, D_1; Y^{L_1}) + I(D_0; Y^{L_1}|X, D_1) \end{aligned}$$

Due to the construction of the iterated channels, it is not difficult to see that the following Markov chain holds:

$$Y^{L_1} \to (X, D_1) \to D_0 \tag{14}$$

Therefore we have

$$I(X, D_1; Y^{L_1}) = I(X, D_0; Y^{L_1}) + I(D_1; Y^{L_1}|X, D_0).$$

The following fundamental lemma characterizes the two quantities on the right. It is proof is included in the Appendix in order to improve the flow of this paper.

*Lemma 2:* If $\Phi$ is chosen as a uniform random variable on $(0,1)$, we have

$$\begin{aligned} I(D_1; Y^{L_1}|X, D_0) &\geq \log(g) \\ I(X, D_0; Y^{L_1}) &= I(X, D_0; Y^{L_0}) \end{aligned}$$

The same result holds if $g > 1$ is an integer and $\Phi$ is chosen to be uniformly distributed on $\{0, 1/g, \cdots, (g-1)/g\}$.

In light of this result, we then find that

$$I(X, D_1; Y^{L_1}) \geq I(X, D_0; Y^{L_0}) + \log(g) \tag{15}$$

$$\geq C(\kappa_0) - \epsilon + \log(g) \tag{16}$$

Using (13) and the characterization of rewritable channel capacity in (1) or (2), as appropriate, we have $C(\kappa_1) \geq I(X, D_1; Y^{L_1})$. Finally,

$$C(\kappa_1) \geq C(\kappa_0) + \log\left(\frac{\kappa_1}{\kappa_0}\right) - \epsilon$$

where we have used the definition of $g$ in (10). Since this holds for every $\epsilon > 0$, we have proved the theorem.

## IV. A Logarithmic Capacity Upper Bound

The lower bound on capacity that we have derived using superposition coding concepts indicates that for a general class of rewritable channels, improvement in capacity as one increases the number of iterations is at least logarithmic. As it turns out, a converse statement can also be made: for any rewritable channel satisfying certain assumptions, capacity is upper bounded by an offset added to a logarithm. This result will be most useful in the case that the input stimulus has a peak value constraint; we leave for a later research effort the problem of sharpening this bound so as to take into account a possible stimulus cost constraint.

Recall that a rewritable channel, within the limited scope of this paper, is defined by a write noise conditional probability law $\mu_{S|X}$. We will assume for this section that there exists a density $f_{S|X}$ such that for every $x, s \in \mathcal{X}$,

$$\mu_{S|X}((-\infty, s]|x) = \int_{-\infty}^{s} f_{S|X}(\xi|x)\lambda(\xi)$$

which is equivalent to the absolute continuity assumption in Theorem 1. Define the function $f_{\sup} : \mathcal{Y} \to [0, \infty)$ as

$$f_{\sup}(y) = \sup_{x \in \mathcal{X}} f_{S|X}(y|x) \tag{17}$$

This function plays a central role in our result, similar to the role of the $a_{\min}(\cdot)$ function in the data dependent uniform noise rewritable channel result of [2]. The present result can be considered a generalization of the upper bound in [2].

*Theorem 2:* For a given rewritable channel with write noise conditional density $f_{S|X}$, assume that

$$\int_{\mathcal{Y}} f_{\sup}(y) d\lambda(y) < +\infty$$

where $\lambda(\cdot)$ denotes the Lebesgue measure. Then

$$C(\kappa) \leq \log\left(\kappa \int_{\mathcal{Y}} f_{\sup}(y) d\lambda(y)\right)$$

*Proof:* To simplify our development, define

$$\Gamma = \int_{\mathcal{Y}} f_{\sup}(y) d\lambda(y).$$

In this proof we will use the definitions for $X, D$, the random process $\{Y^i\}_{i=1}^{\infty}$ and the random cost $L$ established in Section II. To clarify our development, we will add a subscript to the expectation operator that identifies the random variables described by the probability measure that we use in the averaging.

Let $\epsilon > 0$ and let $X, D$ be input random variables such that

$$\begin{aligned} I(X, D; Y^L) &\geq C(\kappa) - \epsilon & (18) \\ E_L\{L\} = E_{X,D}\left\{\frac{1}{\mu_{S|X}(D|X)}\right\} &\leq \kappa. & (19) \end{aligned}$$

Then

$$\begin{aligned}
& I(X, D; Y^L) \\
&\overset{(a)}{=} E_{Y^L,X,D}\left\{\log \frac{f_{Y^L|X,D}(Y^L|X, D)}{f_{Y^L}(Y^L)}\right\} \\
&= E_{Y^L,X,D}\left\{\log \frac{f_{Y^L|X,D}(Y^L|X, D)}{f_{Y^L}(Y^L)} \frac{f_{\sup}(Y^L)}{\Gamma} \frac{\Gamma}{f_{\sup}(Y^L)}\right\} \\
&= E_{Y^L,X,D}\left\{\log \frac{f_{Y^L|X,D}(Y^L|X, D)}{f_{\sup}(Y^L)}\Gamma\right\} + E_{Y^L}\left\{\log \frac{f_{\sup}(Y^L)/\Gamma}{f_{Y^L}(Y^L)}\right\} \\
&\overset{(b)}{=} E_{Y^L,X,D}\left\{\log \frac{f_{Y^L|X,D}(Y^L|X, D)}{f_{\sup}(Y^L)}\Gamma\right\} - D\left(f_{Y^L} \middle\| f_{\sup}/\Gamma\right) \\
&\overset{(c)}{\leq} E_{Y^L,X,D}\left\{\log \frac{f_{Y^L|X,D}(Y^L|X, D)}{f_{\sup}(Y^L)}\Gamma\right\} \\
&\overset{(d)}{=} E_{Y^L,X,D}\left\{\log \frac{f_{S|X}(Y^L|X)}{\mu_{S|X}(D|X) f_{\sup}(Y^L)}\Gamma\right\} \\
&\overset{(e)}{\leq} E_{X,D}\left\{\log \frac{\Gamma}{\mu_{S|X}(D|X)}\right\} \\
&\overset{(f)}{\leq} \log\left(\Gamma E_{X,D}\left\{\frac{1}{\mu_{S|X}(D|X)}\right\}\right) \\
&\overset{(g)}{\leq} \log(\Gamma\kappa)
\end{aligned}$$

where (a) follows from the definition of mutual information, in (b) $D(\cdot\|\cdot)$ stands for the Kullback-Liebler (K-L) divergence and (c) follows from the nonnegativity for the K-L divergence. The step (d) follows from the definition of the operation of the rewritable channel and the associated random variables $Y^L, X, D$ from which it can be deduced that

$$f_{Y^L|X,D}(Y^L|X, D) = \begin{cases} \frac{f_{S|X}(Y^L|X)}{\mu_{S|X}(D|X)} & \text{if } Y^L \in D \\ 0 & \text{otherwise} \end{cases}$$

The inequality (e) follows from the definition of the $f_{\sup}(\cdot)$ function in (17) and (f) follows from Jensen's inequality. Finally, (g) follows from the theorem's assumption in (19). We have then determined that

$$C(\kappa) \leq \log(\Gamma\kappa) + \epsilon.$$

The proof is finalized by observing that $\epsilon$ can be made arbitrarily small. ■

## V. Discussion

A particularly useful consequence of Theorem 1 is obtained by setting $\kappa_0 = 1$: for all $\kappa > 1$,

$$C(\kappa) \geq C(1) + \log(\kappa)$$

where $C(1)$ is simply the classical channel capacity of the channel $\mu_{S|X}$. As an example, suppose that $\mathcal{X} = \mathcal{R}$ and the law $\mu_{S|X}(\cdot|x)$ describes an additive Gaussian noise channel with variance $\sigma_W^2$. If we further assume that the input $X$ to the rewritable channel must satisfy an average stimulus cost constraint of the form

$$E\left\{X^2\right\} \leq \sigma_X^2$$

then Theorem 1 implies that

$$C(\kappa) \geq \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_W^2}\right) + \log(\kappa).$$

Let us now assume that $\mathcal{X} = [-\max(\mathcal{X}), \max(\mathcal{X})]$ with $\max(\mathcal{X}) < +\infty$ and that we do not have any average input stimulus constraint. As before, we assume that the medium has additive Gaussian write noise with variance $\sigma_W^2$. Then Theorem 1 combined with the results of Raginsky [14] on the channel capacity of Gaussian channels with small peak power constraints imply that as long as $\max(\mathcal{X}) \leq 1.05\sigma_W$, then

$$C(\kappa) \geq \frac{1}{\mu^*}\frac{1}{2}\log\left(1 + \frac{\max(\mathcal{X})^2}{\sigma_W^2}\right) + \log(\kappa) \tag{20}$$

where $\mu^*$ is a constant satisfying $\mu^* \leq 5/4$. On the other hand, note that for this setting,

$$f_{\sup}(y) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_W^2}} & \text{for } y \in [-\max(\mathcal{X}), \max(\mathcal{X})] \\ \frac{\exp(-(y-\max(\mathcal{X}))^2/(2\sigma_W^2))}{\sqrt{2\pi\sigma_W^2}} & \text{for } y > \max(\mathcal{X}) \\ \frac{\exp(-(y+\max(\mathcal{X}))^2/(2\sigma_W^2))}{\sqrt{2\pi\sigma_W^2}} & \text{for } y < -\max(\mathcal{X}) \end{cases}$$

Therefore,

$$\int_{\mathcal{R}} f_{\sup}(y) d\lambda(y) = 1 + \sqrt{\frac{2}{\pi}} \frac{\max(\mathcal{X})}{\sigma_W} < +\infty$$

Thus we can use Theorem 2 to obtain, for every $\sigma_W > 0$,

$$C(\kappa) \leq \log\left(1 + \sqrt{\frac{2}{\pi}} \frac{\max(\mathcal{X})}{\sigma_W}\right) + \log(\kappa)$$

as a counterpart to (20).

More generally, the specialization of Theorem 2 to additive channels results in the following corollary:

*Corollary 1:* For an additive rewritable channel with noise density $f_W$, and a peak input stimulus constraint $X \in [\min(\mathcal{X}), \max(\mathcal{X})]$, we have

$$C(\kappa) \leq \log\left(1 + (\max(\mathcal{X}) - \min(\mathcal{X})) \sup_{\xi \in \mathcal{Y}} f_W(\xi)\right) + \log(\kappa)$$

The proof of this is omitted as it follows from the reasoning shown for the Gaussian example.

Because both our upper and lower bounds are of the form of an offset plus the logarithm of the average number of iterations, it is reasonable to ask whether it can be hoped that further research in this area can show that capacity is simply an offset added to a logarithmic growth. It turns out that this is not true, as the exact capacity/cost function of the uniform noise rewritable channel has been found recently [5] and such function has a section that grows faster than a logarithm. Nonetheless, in [5] it is shown that there is always (for the specific modeled considered therein) a critical value of the cost such that for all costs equal or larger than that cost, capacity is precisely an offset added to a logarithmic growth. Thus we believe that the ideas present in this article are a good step in the direction of finding general formulas for rewritable channel capacity.

We end this discussion with a side note. Theorem 1 has been presented as a lower bound mostly because of the important link it establishes with classical channel capacity. Nonetheless, because of its structure, Theorem 1 is in reality both an upper and lower bound on $C(\kappa)$:

*Corollary 2:* For any $\kappa \geq b \geq 1$,

$$C(\kappa/b) + \log b \leq C(\kappa) \leq C(b\kappa) - \log b$$

This observation can be exploited to extend an upper bound on capacity that holds for a $\kappa_1$ to all costs $\kappa_0 < \kappa_1$. Because our upper bound on capacity is already of the form of an offset plus the logarithm of the cost, we cannot use this result to further improve our bounds; future research might be able to take advantage of this technique.

## VI. Concluding remarks

It is our hope that the application of superposition coding concepts to storage channels will motivate the research community to investigate the problem of rewritable channel capacity using the broader repertoire of tools available in other subfields of information theory. Many important problems remain open even within the limited scope of the class of rewritable channels considered in this paper, including the task of finding a formula for rewritable Gaussian channel capacity. Also of interest is whether the ideas presented in here have relevance in the practice of memory storage. We expect to address these matters in subsequent research efforts.

## VII. Appendix - Proof of Lemma 2

*1) Proof of $I(D_1; Y^{L_1}|X, D_0) \geq \log(g)$:* By using the definition of $D_1$ in (12), we have

$$I(D_1; Y^{L_1}|X, D_0) = I(\pi_{D_0,X}(\Phi); Y^{L_1}|X, D_0) \quad (21)$$

We next argue that

$$I(\pi_{D_0,X}(\Phi); Y^{L_1}|X, D_0) = I(\Phi; Y^{L_1}|X, D_0) \quad (22)$$

which can be shown if we demonstrate that the function $\pi_{d,x}(\cdot)$ is invertible for any $d, x$ with $\mu_{S|X}(d|x) > 0$. Suppose that there exist $\phi_1, \phi_2 \in (0, 1)$ such that $\phi_1 \neq \phi_2$ and

$$\pi_{d,x}(\phi_1) = \pi_{d,x}(\phi_2)$$

Let $\nu$ be any open interval satisfying $\nu \subset \psi(\phi_2) \setminus \psi(\phi_1)$. Such $\nu$ must exist since $\phi_1 \neq \phi_2$. Then the set

$$\left\{\xi \in d : F_{Y^{L_0}|X,D_0}(\xi|x, d) \in \nu\right\} \subset \pi_{d,x}(\phi_2) = \pi_{d,x}(\phi_1) \quad (23)$$

must be empty as otherwise we arrive to a contradiction. Choose two points $u_1, u_2 \in \nu$ with $u_1 < u_2$. Since the function $F_{Y^{L_0}|X,D_0}(\cdot|x, d)$ is continuous, there exists $\xi_1 < \xi_2$ such that

$$F_{Y^{L_0}|X,D_0}(\xi_i|x, d) = u_i \qquad i \in \{1, 2\}$$

The set $(\xi_1, \xi_2) \cap d$ must be nonempty, since

$$\begin{aligned} &\mu_{Y^{L_0}|X,D_0}((\xi_1, \xi_2) \cap d|x, d) \\ &= F_{Y^{L_0}|X,D_0}(\xi_2|x, d) - F_{Y^{L_0}|X,D_0}(\xi_1|x, d) \\ &= u_2 - u_1 > 0 \end{aligned}$$

but $(\xi_1, \xi_2) \cap d$ is a subset of the set in the left of (23) which in turn is empty. Since this is a contradiction, it establishes the invertibility of $\pi_{d,x}(\phi)$.

In summary, if we know $D_0, X$ and $\pi_{D_0,X}(\Phi)$ we can retrieve $\Phi$; clearly also if we know $D_0, X$ and $\Phi$ we can construct $\pi_{D_0,X}(\Phi)$. This establishes (22). Combining (21) and (22), we get

$$I(D_1; Y^{L_1}|X) = I(\Phi; Y^{L_1}|X, D_0) \quad (24)$$

We now need to specialize the result according to the choice of $\Phi$. If $\Phi$ is chosen as a discrete random variable, then it is readily seen that

$$\begin{aligned} I(\Phi; Y^{L_1}|X, D_0) &= H(\Phi|X, D_0) - H(\Phi|X, D_0, Y^{L_1}) \\ &= H(\Phi) - H(\Phi|X, D_0, Y^{L_1}) \\ &\overset{(d)}{=} H(\Phi) \\ &= \log(g) \end{aligned} \quad (25)$$

where (d) follows from the fact that knowledge of $X, D_0$ reveals a partition of $D_0$ in the form of

$$\left\{\pi_{D_0,X}\left(\frac{j}{g}\right)\right\}_{j=0}^{g-1}$$

An element of this partition was chosen by $\Phi$ as the encoding set for the iterated channel for cost $\kappa_1$. But further knowledge $Y^{L_1}$ selects a unique element of this partition, thus revealing $\Phi$.

On the other hand, if $\Phi$ is chosen as uniformly distributed on $[0, 1]$, then we employ differential entropy instead to obtain

$$\begin{aligned} I(\Phi; Y^{L_1}|X, D_0) &= h(\Phi|X, D_0) - h(\Phi|X, D_0, Y^{L_1}) \\ &\overset{(e)}{=} -h(\Phi|X, D_0, Y^{L_1}) \\ &\overset{(f)}{\geq} -\log\left(\frac{1}{g}\right) = \log(g) \end{aligned} \quad (26)$$

where (e) follows from the assumption that $\Phi$ is statistically independent from $X$ and $D_0$ and from the fact that the differential entropy of a random variable uniformly distributed on an interval of unit length is zero. The step (f) can be deduced as follows: knowledge of $X, D_0$ and $Y^{L_1}$ reveals that

$$\Phi \in \{\phi : Y^{L_1} \in \pi_{D_0,X}(\phi)\}$$

On the other hand,

$$|\{\phi : Y^{L_1} \in \pi_{D_0,X}(\phi)\}| = 1/g.$$

Finally, recall that the differential entropy of a random variable with a bounded support is upper bounded by the logarithm of the length of the support. This establishes (f) and hence the the first part of the Lemma.

*2) Proof of $I(X, D_0; Y^{L_1}) = I(X, D_0; Y^{L_0})$:* We claim that for either of the two choices for $\Phi$ (the first choice being valid only when $g$ is an integer), the joint distribution of $(X, D_0, Y^{L_0})$ is identical to the joint distribution of $(X, D_0, Y^{L_1})$, hence implying the result of the lemma. An examination of the probability law of $Y^{L_0}$ and $Y^{L_1}$ conditioned on specific values of $X, D_0$ will suffice for this purpose.

We do not discuss the case when $g$ is an integer and that $\Phi$ is chosen to be uniformly distributed over the discrete alphabet $\{0, 1/g, \cdots, (g-1)/g\}$, since it is quite easy to see particularly after reading the following proof for the case in which $\Phi$ is uniformly distributed in the real interval $(0, 1)$. Let $\mathcal{I}(\cdot)$ be the indicator function of a boolean event, this is, it is equal to one if the event in the argument is true and equal to zero otherwise. The following result will be convenient in establishing our desired result.

*Lemma 3:* Let $F_A(\xi)$ be the cumulative distribution function of a random variable $A$ whose probability law $\mu_A$ is absolutely continuous with respect to the Lebesgue measure. Further define

$$\pi(\phi) = \{a : F_A(a) \in \gamma(\phi)\}$$

Then for any $\mu_A$-measurable set $\delta$

$$\int_0^1 \mu_A(\delta \cap \pi(\phi))d\lambda(\phi) = \mu_A(\delta)/g.$$

where the integral above is with respect to the Lebesgue measure.

*Proof:* Let $F_A(\delta)$ denote the image of the set $\delta$ through the function $F$, this is,

$$F_A(\delta) = \{F_A(\xi) : \xi \in \delta\}$$

Next write

$$\begin{aligned} \int_0^1 \mu_A(\delta \cap \pi(\phi))d\lambda(\phi) &= \int_0^1 \int_{F(\delta)} \mathcal{I}(\xi \in \psi(\phi))d\lambda(\xi)d\lambda(\phi) \\ &= \int_{F(\delta)} \int_0^1 \mathcal{I}(\xi \in \psi(\phi))d\lambda(\phi)d\lambda(\xi) \\ &= \int_{F(\delta)} \frac{1}{g} d\lambda(\xi) \\ &= \mu_A(\delta)/g. \end{aligned}$$

We now proceed with the main proof. Let $\delta \subset d_0$. Then

$$\begin{aligned} &\mu_{Y^{L_1}|X,D_0}(\delta|x, d_0) \\ &= \int_0^1 \mu_{Y^{L_1}|X,D_0,\Phi}(\delta|x, d_0, \phi)d\lambda(\phi) \\ &= \int_0^1 \frac{\mu_{Y^{L_0}|X,D_0}(\delta \cap \pi_{d_0,x}(\phi)|x, d_0)}{\mu_{Y^{L_0}|X,D_0}(\pi_{d_0,x}(\phi)|x, d_0)} d\lambda(\phi) \\ &= g\int_0^1 \mu_{Y^{L_0}|X,D_0}(\delta \cap \pi_{d_0,x}(\phi)|x, d_0)d\lambda(\phi) \end{aligned}$$

where the last equality follows from (8) and (11). At this moment, we invoke Lemma 3 to deduce that

$$\begin{aligned} &\int_0^1 \mu_{Y^{L_0}|X,D_0}(\delta \cap \pi_{d_0,x}(\phi)|x, d_0)d\lambda(\phi) \\ &\quad = (1/g)\mu_{Y^{L_0}|X,D_0}(\delta|x, d_0) \end{aligned}$$

This proves the lemma.

## References


[1] L.A. Lastras-Montaño; M.M. Franceschini; T. Mittelholzer; M. Sharma. Rewritable storage channels. In *2008 International Symposium on Information Theory and its Applications*, pages 1–6, Dec 2008.

[2] T. Mittelholzer, M. Franceschini, L.A. Lastras-Montaño, I. Elfadel, M. Sharma. Rewritable channels with data-dependent noise. In *2009 International Conference on Communications*, pages 1–6, June 2009.

[3] Franceschini, M.; Lastras-Montaño, L.A.; Mittelholzer, T.; Sharma, M. The role of feedback in rewritable storage channels [lecture notes]. *IEEE Signal Processing Magazine*, pages 190–194,222, November 2009.

[4] T. Mittelholzer; L.A. Lastras-Montaño; M. Sharma; M. Franceschini. Rewritable storage channels with limited number of rewrite iterations. In *Submitted to ISIT 2010*, July 2010.

[5] L.A. Lastras-Montaño; M. M. Franceschini, T. Mittelholzer. The capacity of the uniform noise rewritable channel with average cost. In *Submitted to ISIT 2010*, July 2010.

[6] S. Raoux, G. W. Burr, M. J. Breitwisch, C. T. Rettner, Y. C. Chen, R. M. Shelby, M. Salinga, D. Krebs, S. H. Chen, and H. L. Lung. Phase-change random access memory: A scalable technology. *IBM Journal of Research and Development*, 52:465–480, 2008.

[7] 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*, 28(2), 2010. http://arxiv.org/abs/1001.1164.

[8] B. Ricco M. Grossi, M. Lanzoni. Program schemes for multilevel flash memories. *Proceedings of the IEEE*, 91(4), April 2003.

[9] T. Nirschl, J. B. Philipp, T. D. Happ, G. W. Burr, B. Rajendran, M. H. Lee, A. Schrott, M. Yang, M. Breitwisch, and C. F. Chen. Write strategies for 2 and 4-bit multi-level phase-change memory. *IEEE International Electron Devices Meeting, 2007. IEDM 2007.*, pages 461–464, 2007.

[10] F. Bedeschi, R. Fackenthal, C. Resta, E. M. Donze, M. Jagasivamani, E. Buda, F. Pellizzer, D. Chow, A. Cabrini, and G. M. A. Calvi. A multi-level-cell bipolar-selected phase-change memory. *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*, pages 428–625, 2008.

[11] D.-H. Kang et al. Two-bit cell operation in diode-switch phase change memory cells with 90nm technology. In *Proc. VLSI'08*, Hyderabad, January 4-8 2008.

[12] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.

[13] P. Billingsley. *Probability and Measure*. John Wiley & Sons, Los Alamitos, CA, 1995.

[14] M. Raginsky. On the information capacity of gaussian channels under small peak power constraints. In *Proc. Forty-Sixth Annual Allerton Conference*, pages 286–293, September 2008.