Upper Bounds on the Capacity of Binary Channels With Causal Adversaries

Bikash Kumar Dey, *Member, IEEE*, Sidharth Jaggi, *Member, IEEE*, Michael Langberg, *Member, IEEE*, and Anand D. Sarwate, *Member, IEEE*

Abstract—In this paper, we consider the communication of information in the presence of a *causal* adversarial jammer. In the setting under study, a sender wishes to communicate a message to a receiver by transmitting a codeword $\mathbf{x} = (x_1, \ldots, x_n)$ bit-by-bit over a communication channel. The sender and the receiver do not share common randomness. The adversarial jammer can view the transmitted bits x_i one at a time and can change up to a p-fraction of them. However, the decisions of the jammer must be made in a *causal* manner. Namely, for each bit x_i, the jammer's decision on whether to corrupt it or not must depend only on x_j for $j \leq i$. This is in contrast to the "classical" adversarial jamming situations in which the jammer has no knowledge of $\mathbf{x}$, or knows $\mathbf{x}$ completely. In this study, we present upper bounds (that hold under both the average and maximal probability of error criteria) on the capacity which hold for both deterministic and stochastic encoding schemes.

Index Terms—Arbitrarily varying channels (AVCs), channel coding, jamming.

I. INTRODUCTION

ALICE wishes to transmit a message u to Bob over a binary-input binary-output channel. To do so, she encodes u into a length-n binary vector $\mathbf{x}$ and transmits it over the channel. However, the channel is controlled by a malicious adversary Calvin who may observe the transmissions, and attempts to jam communication by flipping up to a p fraction of the bits transmitted by Alice. Since he must act in a causal manner, Calvin's decisions on whether or not to flip the bit x_i must be a function solely of the bits $x_1, \ldots, x_i$ he has observed thus far. This

communication scenario models jamming by an adversary who is limited in his jamming capability (perhaps due to limited transmit energy) and is causal. This *causality* assumption is reasonable for many communication channels, both wired and wireless. Calvin can only corrupt a bit when it is transmitted (and thus its error is based on its view so far). To decode the transmitted message, Bob waits until all the bits have arrived.

In this paper, we investigate the information-theoretic limits of communication in this setting. We stress that in our model, Calvin knows everything that both Alice and Bob do—there is no *shared secret or common randomness* (a model where such a shared secret *may* be allowed has been considered in the literature pertaining to *arbitrarily varying channels (AVCs)*, discussed further in Section I-A). However, we make no assumptions about the computational tractability of Alice, Bob, or Calvin's encoding, decoding and jamming processes. Our main contribution in this study is a *converse* that helps to make progress toward a better understanding of the communication *rates* (average number of bits per channel use) achievable against a causal adversary. Specifically, we describe and analyze a novel jamming strategy for Calvin and show that it (upper) bounds the rate of communication regardless of the coding strategy used by Alice and Bob. This jamming strategy results in Calvin being able to force Bob's *average* probability of decoding error over all of Alice's messages to be bounded away from zero (and hence correspondingly also his *maximum* probability of error).

A. Previous and Related Work

Many of the following works deal with related channels; we restrict our discussion mostly to the binary-input binary-output case, except where specifically indicated otherwise.

Coding theory model: A very strong class of adversarial channels is one where Calvin is *omniscient*—he knows Alice's entire codeword $\mathbf{x}$ prior to transmission and can tailor the pattern of up to pn bit flips to each specific transmission. This is the "worst case noise" model studied in coding theory. In this model, there is no randomness in code design, and it is desired that Bob *always* decodes correctly. For binary channels, characterizing the capacity has been an open problem for several decades. The best known upper bound is due to McEliece *et al.* [2] as the solution of an LP, and the best known achievable scheme corresponds to codes suggested by Gilbert [3] and Varshamov [4], which

achieve a rate of $1 - H(2p)$. Improving either of these bounds would be a significant breakthrough.[1]

Information theory model: A much weaker class of adversarial channels is one where Calvin generates bit flips in an i.i.d. manner with probability p and Bob must decode correctly "with high probability" over the randomness in Calvin's bit flips. The original work of Shannon [8] effectively characterized the capacity of this *binary symmetric channel* $\mathsf{BSC}(p)$. The capacity $1 - H(p)$ in Shannon's setting (for crossover probability p) is strictly greater than that of the coding theory model.

Causal adversarial model: The class of channels considered in this study, i.e., that of *causal adversaries*, falls in between the above two extremes. In one direction, this is because a causal adversary is certainly no stronger than an omniscient adversary, since he cannot tailor his jamming strategy to take into account Alice's future transmissions. Indeed, the work of Haviv and Langberg [9] indicates that (for $2p < H^{-1}(1/2) \simeq 0.11$) rates strictly better than those achievable by Gilbert–Varshamov codes [3], [4] against an omniscient adversary are achievable against a causal adversary. However, since it is still unknown whether Gilbert–Varshamov codes are optimal against omniscient adversaries, it is unknown whether causal adversaries are indeed strictly weaker than omniscient adversaries. Nonetheless, the Gilbert–Varshamov bound and the bound of [9] indicate that for $p < 1/4$, the capacity under causal adversaries is bounded away from zero.

In the other direction, the causal adversarial model under study is at least as strong as the information-theoretic model in which Calvin generates bit flips in an i.i.d. manner. Specifically, if $p \le 1/2$, for any $\delta_p > 0$ and sufficiently long block-length n a causal adversary can ignore the transmitted codeword seen so far and just mimic the behavior of a binary symmetric channel $\mathsf{BSC}(p - \delta_p)$—with high probability, he does not exceed his budget of pn bit flips. Similarly, if $p > 1/2$, Calvin simply mimics the behavior of a $\mathsf{BSC}(1/2)$. This implies that when communicating in the presence of causal adversaries with jamming capabilities that are parametrized by p, $1 - H(p)$ is an upper bound on the achievable rate for $p \le 1/2$, and no positive rate is achievable for $p > 1/2$. Improving over this naïve upper bound (and hence narrowing the gap to the lower bound of [9]) is the focus of the paper.

The improved upper bounds we present hold for general coding schemes that allow Alice to encode a message u to one of several possible codewords $\mathbf{x} \in \{\mathbf{x}(u, r)\}$, where r is a random source available to Alice but unknown to either Bob or Calvin. Such general coding schemes are referred to as *stochastic* coding schemes. We stress that in such schemes, there is no shared randomness between Alice and Bob, and the source of randomness in Alice's encoder is solely known to Alice.

---

[1]As is often the case, results for channels over "large" alphabets are significantly easier. In the "intermediate" alphabet-size regime, wherein the alphabet is of size at least 49, advances in algebraic-geometry codes over the last three decades (see [5] for a survey) have resulted in codes exceeding the Gilbert–Varshamov bound. For alphabets larger than n, the bound of $1 - 2p$ due to Singleton [6] is known to be achievable in a computationally efficient manner via Reed–Solomon codes [7].

Arbitrarily Varying Channels: Our model is a variant of the AVC model [10]. The AVC model where the adversary has access to the entire codeword was considered by Ahlswede and Wolfowitz [11], [12] but received little attention since [13, Problem 2.6.21]. General AVC models have been extended to include channels with constraints on the adversary (such as pn bit flips) for cases where the adversary has no access to the codeword [14], or has access to the full codeword [15]. For binary channels in which the jammer has knowledge of the entire codeword $\mathbf{x}$, Langberg [16] showed that $O(\log n)$ bits of common randomness is sufficient to achieve the optimal rate of $1 - H(p)$ (and the work in [17] investigated computationally efficient constructions of such codes). However, issues of causality have only been studied in the context of randomized coding (when the encoder and decoder share common randomness), but not for deterministic codes or stochastic encoding.

Delayed adversaries: The *delayed adversary* model was studied in [18] and [19]. In this model, the jammer's decision on whether to corrupt x_i must depend only on x_j for $j \le i - Dn$ for a delay parameter $D \in [0, 1]$. The case of $D = 0$ is exactly the causal setting studied in this paper, and that of $D = 1$ corresponds to the "oblivious adversary" studied by Langberg [18]. In this oblivious adversary setting, the work in [20] demonstrates computationally efficient code constructions that achieve information-theoretically rate-optimal throughput of $1 - H(p)$ for all $p < 1/2$.

In a different line of work, Dey *et al.* [19] showed that for a large class of channels, the capacity for delay $D > 0$ equals that of the constrained AVC model [21]. In particular, a positive delay implies that the optimal rate $1 - H(p)$ is achievable against a delayed adversary over a binary-input binary-output channel. In this paper, we show that a causal adversary is strictly stronger than a delayed adversary with $D > 0$ for all $p > 0.0804$. For p smaller than this value, our techniques do not help separate the capacity regions of these two models.

Causal and delayed adversaries for "large alphabets": In the *large alphabet* setting (where the alphabet size is allowed to grow without bound with increasing block-length), Dey *et al.* [22] give a full characterization of the capacity region of several variants of both the causal adversary and the delayed adversary models. They further give computationally efficient codes achieving every point in the capacity regions for the models considered. In general, in the large alphabet regime, code design is easier than in the binary alphabet regime (that is the primary focus of this study) since with large alphabets, a "few random hashes" can be hidden inside each symbol with asymptotically negligible rate-loss. These hashes aid the decoder in detecting the adversarial attack pattern and correcting for it. In the binary alphabet setting, this technique is not applicable—this is one of the bottlenecks in further narrowing the gaps between outer and inner bounds for the model considered in this study.

Previous attacks: This study continues our preliminary work on binary causal channels [23] (and a related result of Guruswami and Smith [20]), which proposed an upper bound using the so called "wait-and-push" attack. This study improves
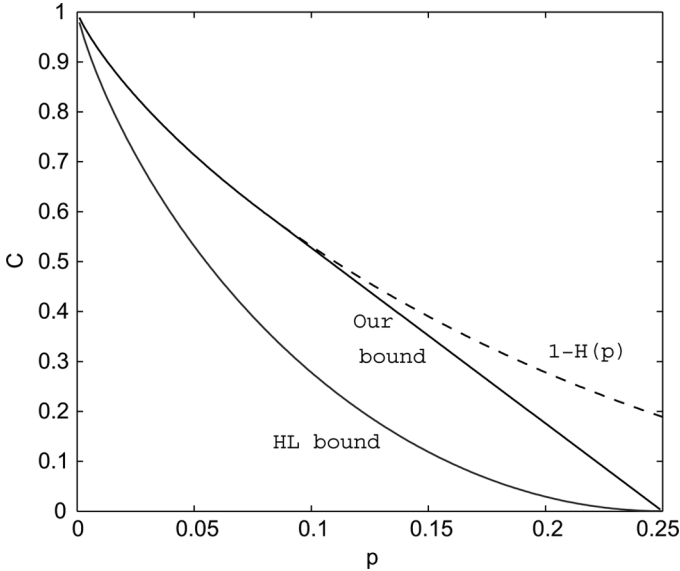
Fig. 1. We plot previous bounds related to the channel at hand compared to our bound. The *upper* bound of $1 - H(p)$ corresponds to the binary symmetric channel. The *lower* bound [9] (denoted HL) is based on an evaluation of the parameters specified by Haviv and Langberg [9] and it slightly improves on the Gilbert–Varshamov bound $1 - H(2p)$. Our improved bound appears in between.

on this earlier work in two aspects—specifically the bound presented is tighter and holds also for stochastic encoding.[2]

B. Main Result

Our improved bounds are given in the following theorem and are depicted (in comparison with the previous bounds) in Fig. 1. For any $\bar{p} \in [0, p]$, let $\alpha(p, \bar{p}) = 1 - 4(p - \bar{p})$. In what follows, $C(p)$ is the capacity of the causal channel under study. For precise definitions and model, see Section II.

Theorem 1: For $p \in [0, 1/4]$, the capacity $C(p)$ of a binary causal adversary channel with constraint p satisfies

$$C(p) \leq \min_{\bar{p} \in [0, p]} \left[\alpha(p, \bar{p}) \left(1 - H \left(\frac{\bar{p}}{\alpha(p, \bar{p})} \right) \right) \right].$$

For $p > 1/4$, the capacity $C(p) = 0$.

A few remarks are in order. Notice that in the regime $\bar{p} \leq p \leq 1/4$, it holds that $\bar{p} \leq \alpha(p, \bar{p})$, and thus, $\frac{\bar{p}}{\alpha(p, \bar{p})}$ in the expression

[2]For completeness, we specify the two major differences between this paper and [23]. First, we propose a different two-phase attack ("babble-and-push") which gives a tighter outer bound than the previous attack ("wait-and-push"). In "wait-and-push," Calvin passively eavesdrops in the first phase uses this information to design an error vector to confuse Bob in the second phase. In our new attack, Calvin instead injects noise in the first phase to increase Bob's uncertainty about Alice's transmissions. However, we must carefully choose the number of bit flips Calvin injects in this "babble" phase to obtain a tighter outer bound, because Calvin must trade-off between using bit flips to increase Bob's uncertainty and to push to an alternative codeword in the second phase. The second improvement in this paper is that we prove that the "babble-and-push" attack works even when Alice and Bob use stochastic encoding (i.e., for each message u she has, Alice may choose to transmit one of multiple possible codewords $\mathbf{x}(u)$, with an arbitrary random distribution over the set of codewords). Our bounds therefore hold for general codes, as opposed to previous work [23], where the outer bound was proved for codes in which each message u corresponded to a unique $\mathbf{x}(u)$ *deterministically* chosen by Alice.

of Theorem 1 is at most of value 1. We show in Appendix A that the optimum $\bar{p}$ in the computation of $C(p)$ is

$$\min \left\{ p, \frac{3(1 - 4p)}{\left(2 + \sqrt[3]{1592 + 24\sqrt{33}} + \sqrt[3]{1592 - 24\sqrt{33}} \right)} \right\}$$

$$\simeq \min \left\{ p, \frac{1 - 4p}{8.4445} \right\}.$$

Namely, for p greater than approximately 0.0804, the capacity $C(p)$ is bounded away from $1 - H(p)$, and for p less than this value our bound equals $1 - H(p)$ (in the latter case we get $\bar{p} = p$). For $p = 1/4$, the new strategy we propose for Calvin shows that no positive rate is achievable; when $p > 1/4$, Calvin can simply mimic the case $p = 1/4$.

C. Techniques and Proof Overview

To prove Theorem 1, we show that no matter which encoding/decoding scheme is used by Alice and Bob, there exists a strategy for Calvin that does not allow communication at rate higher than $C(p)$. Specifically, we demonstrate that whenever Alice and Bob attempt to communicate at a rate higher than $C(p)$, there exists a causal jamming strategy (that in general depends on Alice and Bob's encoding/decoding strategy) that allows Calvin to enforce a constant probability of error bounded away from zero. More precisely, for any block-length n, any $\varepsilon > 0$ and any encoding/decoding scheme of Alice and Bob of rate $\left(C(p) + \sqrt{\frac{c}{n}} \right) + \varepsilon$, Calvin can cause a decoding error probability of at least $\varepsilon^{O(1/\varepsilon)}$.

At a high level, Calvin uses a two-phase *"babble-and-push"* strategy. In the first phase of $\ell(p)$ channel uses, Calvin "babbles" by behaving like a $\mathsf{BSC}(\bar{p})$ for some $\bar{p}$ chosen as a function of p. In the second phase of $n - \ell(p)$ channel uses, Calvin randomly selects a codeword from Alice and Bob's codebook that is consistent with what Bob has received so far. Calvin then "randomly pushes" the remaining part of Alice's codeword toward his selected codeword (i.e., in every location in the "push" phase where Alice's codeword bit differs from his selected codeword, he adds a with probability half). A decoding error occurs if Calvin is able to push the transmitted codeword half the distance toward the codeword selected by Calvin (via a standard symmetrization argument [14]).

Roughly speaking, the first phase allows Calvin to gain information regarding which codeword was transmitted by Alice, while the second phase allows Calvin to use this information in order to design a corresponding symmetrization-based jamming strategy.

In Section III, we present the proof of our main result, that of the outer bound on the capacity of online adversaries. Section IV then improves on this result (by giving a tighter bound on the probability of error) for the special case of deterministic encoders (rather than the general stochastic encoders considered in Section III).

II. MODEL AND PRELIMINARIES

We first reprise some standard notation. Let $d_H(\cdot, \cdot)$ denote the *Hamming distance* function between two vectors (number of locations in which two vectors of the same length differ). The

Hamming weight $wt_H(\mathbf{x})$ of a vector $\mathbf{x}$ is the Hamming distance between that vector and the all-zero vector. Let $\log(.)$ denote the binary logarithm, here and throughout. As is common, the notation $H(A)$ is used to denote the (binary) *entropy* of a random variable A, $H(A|B)$ to denote the *conditional entropy of A given B*, and $I(A; B)$ to denote the *mutual information between A and B*. Also, for any real number $x \in (0, 1)$, $H(x)$ denotes the *binary entropy function*. Properties of and inequalities between these functions are referenced at the point in the text where needed. The *indicator function* $\mathbf{1}(condition)$ takes value 1 if $condition$ is true, and 0 otherwise.

Let the input and the output alphabets of the channel be $\mathcal{X}$ and $\mathcal{Y}$, respectively. For any positive integer k, let $[k] = \{1, 2, \ldots, k\}$. We let $\mathcal{U} = [2^{nR}]$ denote Alice's message set, and $\mathbf{U}$ denote the message random variable uniformly distributed in $\mathcal{U}$. A *deterministic code* of rate R and block-length n is a pair of maps $\mathcal{C}_d = (\Phi, \Psi)$ where $\Phi : \mathcal{U} \to \mathcal{X}^n$ and $\Psi : \mathcal{Y}^n \to \mathcal{U}$ are deterministic maps. The map Φ is called the encoder and the map Ψ is called the decoder.

A *code with stochastic encoding and decoding* of rate R and block-length n is a pair of maps $\mathcal{C}_s = (\Phi, \Psi)$ where $\Phi : \mathcal{U} \to \mathcal{X}^n$ and $\Psi : \mathcal{Y}^n \to \mathcal{U}$ are probabilistic maps. The random map Φ gives a probability distribution $\rho(\cdot|u)$ on $\mathcal{X}^n$ for every $u \in \mathcal{U}$. The mapping $\Psi(\mathbf{y})$ is a random variable taking values from $\mathcal{U}$. The encoding Φ is equivalently represented by first picking a random variable $\mathfrak{R}$ from a set $\mathscr{R}$ according to a conditional distribution $\rho_{\mathfrak{R}|\mathbf{U}}(\cdot|u)$, and then applying a deterministic encoder map $\Phi : \mathcal{U} \times \mathscr{R} \to \mathcal{X}^n$. Note that our definition does not preclude there existing pairs (u, r) and (u', r') such that $\Phi(u, r) = \Phi(u', r')$. As we are addressing upper bounds on the capacity $C(p)$ in this study, it is crucial to prove our results in the stochastic setting above—any bounds proved in the stochastic setting also hold in the deterministic setting.

A *causal adversarial strategy* of block-length n is a sequence of (possibly random) mappings $\mathsf{Adv} = \{f_\mathcal{C}^{(i)} : i \in [n]\}$. Here, each $f_\mathcal{C}^{(i)} : \mathcal{X}^i \times \mathcal{E}^{i-1} \to \mathcal{E}$ depends on $\mathcal{C}$, and for each time, $i \in [n]$ chooses an *action at time i*, $e_i = f_\mathcal{C}^{(i)}(x_1, \ldots, x_i) \in \mathcal{E}$—the inputs to $f_\mathcal{C}^{(i)}$ are the past and current channel inputs $(x_1, x_2, \ldots, x_i)$ and its own previous actions $(e_1, e_2, \ldots, e_{i-1})$. The resulting *channel output at time i* is $y_i = x_i + e_i$. In our setting, $\mathcal{E} = \{0, 1\}$. The strategy obeys constraint p if the Hamming weight $\|\mathbf{e}\| = \sum_{i=1}^{n} e_i$ of $\mathbf{e} = (e_1, \ldots, e_n)$ is at most pn over the randomness in the message, encoder, and strategy. For a given adversarial strategy and an input codeword $\mathbf{x}$, the strategy produces a (possibly random) $\mathbf{e}$ and the output is $\mathbf{y} = \mathbf{x} \oplus \mathbf{e}$. Let $\Pr_{\mathsf{Adv}}(\mathbf{y}|\mathbf{x})$ denote the probability of an output $\mathbf{y}$ given an input $\mathbf{x}$ under the strategy Adv where this strategy might depend on (u, r) via the adversary's causal observations of $\mathbf{x}$—to simplify notation we henceforth do not make this explicit. When the block-length is understood from the context, let $\mathsf{Adv}(p)$ denote all adversarial strategies obeying constraint p.

The (average) probability of error for a code with stochastic encoding and decoding is given by

$$\bar{\varepsilon} = \max_{\mathsf{Adv} \in \mathsf{Adv}(p)} \frac{1}{2^{Rn}} \sum_{u=1}^{2^{Rn}} \sum_{r \in \mathscr{R}} \rho_{\mathfrak{R}|\mathbf{U}}(r|u)$$
$$\sum_{\mathbf{y}} \Pr_{\mathsf{Adv}}(\mathbf{y}|\Phi(u, r)) \Pr(\Psi(\mathbf{y}) \neq u), \quad (1)$$

where the probability $\Pr(\Psi(\mathbf{y}) \neq u)$ is over any randomness in the decoder (but there is *no* conditioning on r since shared randomness between the encoder and the decoder is not allowed). We can interpret the errors as the error in expectation over Alice choosing a message $\mathbf{U} = u$ and a codeword $\mathbf{x} = \Phi(u, r)$ according to the conditional distribution $\rho(\mathbf{x}|u)$.

A rate R is *achievable* against a causal adversary under average error if for every $\delta > 0$ there exist infinitely many block-lengths $\{n_i\}$, such that for each n_i there is an n_i block-length (stochastic) code of rate at least R and average probability of error at most δ. The supremum of all achievable rates is the capacity. We denote by $C(p)$ the capacity of the channel corresponding to adversaries parametrized by p.

Consider a code of block-length n, rate R, and error probability δ. We can, without loss of generality (w.l.o.g.), assume that the encoding probabilities $\{\rho(\mathbf{x}|u) : \mathbf{x} \in \{0, 1\}^n, u \in [2^{nR}]\}$ are rational. To see why this is the case, note that for any small $\eta > 0$ we can find rational numbers $\{\tilde{\rho}(\mathbf{x}|u)\}$ such that $\rho(\mathbf{x}|u) - \eta \leq \tilde{\rho}(\mathbf{x}|u) \leq \rho(\mathbf{x}|u)$. Now consider a code with encoding probabilities $q(\mathbf{x}|u) = \tilde{\rho}(\mathbf{x}|u)$ for $\mathbf{x} \neq \mathbf{0}$ and assign the remaining probability to $\mathbf{0}$. Under the same decoder, this code has error probability at most $\delta + 2^{n+nR}\eta$, but since η was arbitrary, the error is at most 2δ.

Now, for a given stochastic code, let N be the least common multiple of the denominators of $\rho(\mathbf{x}|u)$ for all $\mathbf{x}$, u. Each codeword $\mathbf{x}$ of u can be treated as $N\rho(\mathbf{x}|u)$ copies of the same codeword with conditional probability $1/N$ each. So we can equivalently associate a random variable $\mathfrak{R}$ with $|\mathscr{R}| = N$ s.t. the conditional distribution $\rho_{\mathfrak{R}|\mathbf{U}}(\cdot|u)$ is uniform, and the encoding map $\Phi(u, \cdot)$ is not necessarily injective. Since we consider the uniform message distribution, henceforth, w.l.o.g., we assume that the joint distribution $\rho_{\mathbf{U},\mathfrak{R}}(\cdot, \cdot)$ is uniform.

We use a version of Plotkin's bound [24] in our proof. This result gives an upper bound on the number of codes in any binary code with a given minimum distance.

Theorem 2 (Plotkin Bound [24]): There are at most $\frac{2d_{\min}}{2d_{\min} - n}$ codewords in any binary code of block-length n with minimum distance $d_{\min} > n/2$.

III. PROOF OF THEOREM 1

In this section, we analyze an adversarial attack for the general case of stochastic encoders and decoders. For fully deterministic codes, the analysis is more combinatorial and the error bounds are somewhat better, as shown in Section IV.

Let $p \in [0, 1/4]$ and let $\bar{p} \leq p$. Without loss of generality, we assume that pn is an integer—if not, Calvin can simply choose the largest p' smaller than p such that $p'n$ is an integer. Asymptotically in n, the effect of this quantization on our outer bound is negligible. Let $\varepsilon > 0$. In what follows, we prove that the rate of communication over the causal adversarial channel (with parameter p) is bounded by

$$R \leq C + \varepsilon, \quad (2)$$

where

$$C = \alpha(p, \bar{p}) \left(1 - H\left(\frac{\bar{p}}{\alpha(p, \bar{p})}\right)\right) \quad (3)$$

and

$$\alpha = \alpha(p, \bar{p}) = 1 - 4(p - \bar{p}) \qquad (4)$$

as defined in Theorem 1. Namely, if (2)–(4) is violated, for any sufficiently large block-length n, and any (n-block stochastic) code $\mathcal{C}_s = (\Phi, \Psi)$ shared by Alice and Bob, there exists an adversarial jammer Adv that can impose a constant decoding error. The decoding error we obtain will depend on $\varepsilon > 0$.

For $\bar{p} = p$, the adversary can generate a noise sequence to simulate a BSC with crossover probability arbitrarily close to p, which yields an upper bound of $1 - H(p)$ on the capacity. We, therefore, assume that $p - \bar{p} > 0$ and that $\varepsilon < 2(p - \bar{p})$. We show that for such $\varepsilon > 0$, there cannot exist a sequence of codes (each with rate at least $C + \varepsilon$) of increasing block-length n, such that the probability of error of these codes converges to 0 asymptotically in n. To do so, we will consider block-lengths $n \geq \Omega(\varepsilon^{-2})$. Note that this argument does not provide lower bounds on the error of codes of a given block-length, but instead shows a bound on the capacity. We elaborate on this point at the end of the proof.

Our converse bound is based on a particular two-phase adversarial strategy for Calvin that we call "babble-and-push." Let $\ell = (\alpha + \varepsilon/2)n$ and w.l.o.g. assume $\ell \in \mathbb{N}$. For a vector $\mathbf{z}$ of length n, let $\mathbf{z}_1 = (z_1, z_2, \ldots, z_\ell)$ and $\mathbf{z}_2 = (z_{\ell+1}, z_{\ell+2}, \ldots, z_n)$. In what follows, $\mathbf{z}_1$ will correspond to the first phase of Calvin's attack, while $\mathbf{z}_2$ corresponds to the second phase. For $p > 0$, the strategy is given as follows.

1) **("Babble")** Calvin chooses a random subset Γ of $\bar{p}n$ indices uniformly from the set of all ($\bar{p}n$)-sized subsets of $\{1, 2, \ldots, \ell\}$. For $i \in \Gamma$, Calvin flips bit x_i; that is, for $i \in \{1, 2, \ldots, \ell\}$, $e_i = 1$ for $i \in \Gamma$ and $e_i = 0$ for $i \notin \Gamma$.

2) **("Push")** Calvin constructs the set of (u, r) that have encodings $\mathbf{x}(u, r) = \Phi(u, r)$ that are *close* to $\mathbf{y}_1 = y_1, \ldots, y_\ell$. Namely, Calvin constructs the set

$$B_{\mathbf{y}_1} = \{(u, r) : d_H(\mathbf{y}_1, \mathbf{x}_1(u, r)) = \bar{p}n\} \qquad (5)$$

and selects an element $(u', r') \in B_{\mathbf{y}_1}$ uniformly at random. Calvin then considers the corresponding codeword $\mathbf{x}' = \Phi(u', r')$. Given the selected $\mathbf{x}'$, for $i > \ell$, if $x_i \neq x_i'$, Calvin sets e_i equiprobably to 0 or 1 until $\sum_{j=1}^{i} e_j = pn$ or $i = n$. Note that, under our assumption (w.l.o.g.) of uniform $\rho_{\mathbf{U}, \mathfrak{R}}$, the *a posteriori* distribution of Alice's choice (u, r) given $\mathbf{y}_1$ is also uniform in $B_{\mathbf{y}_1}$.

We start by proving the following technical lemma that we use in our proof.

Lemma 3: Let V be a random variable on a discrete finite set $\mathcal{V}$ with entropy $H(V) \geq \lambda$, and let $V_1, V_2, \ldots, V_m$ be i.i.d. copies of V. Then

$$\Pr(\{V_i : i = 1, \ldots, m\} \text{ are all distinct})$$
$$\geq \left(\frac{\lambda - 1 - \log m}{\log |\mathcal{V}|} \right)^{m-1}. (6)$$

Proof: Fix $i \leq m$ and a set $v_1, v_2, \ldots, v_i \in \mathcal{V}$. Let $A_i = \{v_1, \ldots, v_i\}$ and let $W_i = \mathbf{1}(V_{i+1} \in A_i)$, where $\mathbf{1}(.)$ denotes

the *indicator function*. We can write the distribution of V as a mixture:

$$\Pr[V_{i+i} = v] = \sum_{j \in \{0,1\}} \Pr[W_i = j] \cdot \Pr[V_{i+i} = v | W_i = j].$$

We can *bound from above* the entropy of V as

$$H(V_{i+1}) \leq H(V_{i+1} | W_i) + H(W_i)$$
$$= \sum_{j \in \{0,1\}} \Pr[W_i = j] H(V_{i+1} | W_i = j) + H(W_i).$$

Since conditioning reduces entropy and the support of V_{i+1} conditioned on $W_i = 1$ is at most i, we have

$$\lambda \leq 1 + \log i + \Pr[W_i = 0] \log |\mathcal{V}|.$$

Namely

$$\Pr[W_i = 0] \geq \frac{\lambda - 1 - \log i}{\log |\mathcal{V}|} \geq \frac{\lambda - 1 - \log m}{\log |\mathcal{V}|}.$$

But the event that each V_i is distinct is equivalent to the event that for each $i \in \{2, \ldots, m\}$, W_i is 0. ∎

To prove the upper bound, we now present a series of claims. Let $\mathbf{X}$ denote the random variable corresponding to Alice's input codeword and let $\mathbf{Y}$ be the output of the channel. Thus, $\mathbf{X}_1 \in \{0, 1\}^\ell$ is Alice's input during the "babble" phase of length ℓ and $\mathbf{X}_2$ is her input during the "push" phase; the randomness comes from the message $\mathbf{U}$ and the stochastic encoding. Similarly, $\mathbf{Y}_1$ is the random variable corresponding to the ℓ bits received by Bob during the "babble" phase, and $\mathbf{Y}_2$ the $n - \ell$ bits of the "push" phase. Let Adv denote the "babble-and-push" adversarial strategy.

Let

$$A_0 = \{\mathbf{y}_1 : H(\mathbf{U} | \mathbf{Y}_1 = \mathbf{y}_1) \geq n\varepsilon/4\},$$

where the entropy $H(\mathbf{U} | \mathbf{Y}_1 = \mathbf{y}_1)$ is measured over the randomness of the encoder, the message, and any randomness in Calvin's action during the "babble" phase. Further, let the event E_0 be defined as

$$E_0 = \{\mathbf{Y}_1 \in A_0\}. \qquad (7)$$

Claim 4: For the "babble-and-push" attack Adv

$$\Pr_{\text{Adv}}(E_0) \geq \varepsilon/4. \qquad (8)$$

Proof: By the data-processing inequality ($\mathbf{U} \to \mathbf{X}_1 \to \mathbf{Y}_1$ form a Markov chain and hence $I(\mathbf{U}; \mathbf{Y}_1) \leq I(\mathbf{X}_1; \mathbf{Y}_1)$), and the choice of Calvin's strategy, we have

$$I(\mathbf{U}; \mathbf{Y}_1) \leq I(\mathbf{X}_1; \mathbf{Y}_1)$$
$$\leq \ell(1 - H(\bar{p}n/\ell))$$
$$= (\alpha n + \varepsilon n/2) \left(1 - H \left(\frac{\bar{p}}{\alpha + \varepsilon/2} \right) \right).$$

Therefore

$$
\begin{aligned}
H(\mathbf{U}|\mathbf{Y}_1) &\geq H(\mathbf{U}) - n(\alpha + \varepsilon/2)\left(1 - H\left(\frac{\bar{p}}{\alpha + \varepsilon/2}\right)\right) \\
&\geq n\left(\varepsilon + \alpha\left(1 - H\left(\frac{\bar{p}}{\alpha}\right)\right)\right) \\
&\quad - n(\alpha + \varepsilon/2)\left(1 - H\left(\frac{\bar{p}}{\alpha + \varepsilon/2}\right)\right) \\
&= n\varepsilon/2 + n\left((\alpha + \varepsilon/2)H\left(\frac{\bar{p}}{\alpha + \varepsilon/2}\right)\right. \\
&\quad \left. - \alpha H\left(\frac{\bar{p}}{\alpha}\right)\right) \\
&\geq n\varepsilon/2.
\end{aligned}
$$

Here, the first inequality follows from the definition of conditional entropy, the second from the assumption underlying this proof by contradiction that nR (and hence $H(\mathbf{U})$) violates (2)–(4), and the third from the fact that the function $\alpha H(\bar{p}/\alpha)$ is monotonically increasing in α since the function's derivative with respect to α equals $\log(\alpha/(\alpha - \bar{p}))$ which is always positive. Thus, the expected value of $H(\mathbf{U}|\mathbf{Y}_1 = \mathbf{y}_1)$ over $\mathbf{y}_1$ is at least $n\varepsilon/2$, and the maximum value of $H(\mathbf{U}|\mathbf{Y}_1 = \mathbf{y}_1)$ is nR. Applying the Markov inequality to the random variable $nR - H(\mathbf{U}|\mathbf{Y}_1 = \mathbf{y}_1)$, we see that

$$
\begin{aligned}
\Pr\left[nR - H(\mathbf{U}|\mathbf{Y}_1 = \mathbf{y}_1) > nR - n\varepsilon/4\right] \\
< \frac{nR - n\varepsilon/2}{nR - n\varepsilon/4} \\
= 1 - \frac{\varepsilon/4}{R - \varepsilon/4},
\end{aligned}
$$

and hence

$$
\Pr\left[H(\mathbf{U}|\mathbf{Y}_1 = \mathbf{y}_1) \geq n\varepsilon/4\right] \geq \frac{\varepsilon/4}{R - \varepsilon/4}.
$$

Using the fact that $R \leq 1$ yields the result. ∎

Now consider drawing m pairs (U_i, R_i) from $B_{\mathbf{y}_1}$ i.i.d. $\sim \rho_{\mathbf{U}, \Re|\mathbf{y}_1}$ (which happens to be uniform). Note that the marginal distribution of U_i is also i.i.d. $\sim \rho_{\mathbf{U}|\mathbf{y}_1}$, which is not necessarily uniform. Let

$$
E_1 = \{\{U_1, U_2, \ldots, U_m\} \text{ are all distinct}\}. \tag{9}
$$

Claim 5: Let $\rho_{\mathbf{U}|\mathbf{y}_1}$ be the conditional distribution of $\mathbf{U}$ given $\mathbf{y}_1$ under Adv. Let $U_1, U_2, \ldots, U_m$ be m random variables drawn i.i.d. according to $\rho_{\mathbf{U}|\mathbf{y}_1}$. Then, for large enough n

$$
\Pr(E_1 \mid E_0) \geq (\varepsilon/5)^{m-1}. \tag{10}
$$

Proof: The proof follows from Claim 4 and Lemma 3 by using $\lambda = n\varepsilon/4$, $V = \mathbf{U}$, and the fact that there are at most 2^n messages, so $|\mathcal{V}| \leq 2^n$. The lower bound in (6) then becomes $\left(\frac{n\varepsilon/4 - 1 - \log m}{n}\right)^{m-1}$. For fixed m, there exists a sufficiently large n such that $\varepsilon/4 - (1 + \log m)/n > \varepsilon/5$. ∎

The preceding two claims establish a lower bound on the probability that $\mathbf{Y}_1$ takes a value such that the distribution of the message $\mathbf{U}$ conditioned on $\mathbf{Y}_1$ has sufficient entropy. For such values $\mathbf{y}_1$ of $\mathbf{Y}_1$, we now use the fact that Alice's pair (u, r) is uniform in $B_{\mathbf{y}_1}$ to analyze the probability that Calvin's "push" attack succeeds. Let $\mathbf{U}'$ and $\mathbf{X}'$ denote the random choice of Calvin's message and codeword in the "push" phase. We show that the following two events occur with probability bounded away from zero:

$$
\begin{aligned}
E_2 &= \{\mathbf{U}' \neq \mathbf{U}\} \tag{11} \\
E_3 &= \{d_H(\mathbf{X}_2, \mathbf{X}_2') \leq 2(p - \bar{p})n - \varepsilon n/8\}. \tag{12}
\end{aligned}
$$

The first event is that Calvin chooses a different message than Alice and the second is that he chooses a codeword that is close enough to Alice's. The occurrence of the first event ensures that the codeword Calvin chooses to try to confuse Bob into thinking might have been transmitted corresponds to a message u' different than Alice's actual message u. The occurrence of the second event ensures that the two codewords chosen ($\mathbf{x}_2$ chosen by Alice, and $\mathbf{x}_2'$ by Calvin) are "close enough" for Calvin to be able to push Bob's received codeword halfway between $\mathbf{x}_2$ and $\mathbf{x}_2'$.

Claim 6: For the "babble-and-push" attack Adv

$$
\Pr_{\text{Adv}}(E_2 \text{ and } E_3 \mid E_0) \geq \varepsilon^{O(1/\varepsilon)}. \tag{13}
$$

Proof: Conditioned on E_0, the realization $\mathbf{y}_1$ satisfies $H(\mathbf{U}|\mathbf{Y}_1 = \mathbf{y}_1) \geq \varepsilon n/4$. We first use Claim 5 to lower bound the probability that E_2 holds. First consider randomly sampling a set of mutually independent pairs $S = \{(u_i, r_i) : i \in [m]\}$ uniformly from $B_{\mathbf{y}_1}$, and let $\mathbf{X}^i$ be the codeword for (u_i, r_i).

Claim 5 shows that with probability at least $(\varepsilon/5)^{m-1}$, all the messages in S are distinct. In particular, this shows that

$$
\Pr_{\text{Adv}}(E_2 \mid E_0) \geq (\varepsilon/5).
$$

Turning to E_3, applying Claim 5 for general m shows that the probability that m draws from the conditional distribution $\rho_{\mathbf{U}|\mathbf{y}_1}$ yield unique messages is lower bounded by $(\varepsilon/5)^{m-1}$. Plotkin's bound [24] (reprised in Theorem 2) shows that there do not exist binary error-correcting codes of block-length $n - \ell$ and minimum distance d with more than $\frac{2d}{2d - (n-\ell)}$ codewords. Setting $m = 17/\varepsilon$, this bound implies that with probability at least $(\varepsilon/5)^{m-1}$ there must exist codewords $\mathbf{x}, \mathbf{x}'$ corresponding to (u, r) and (u', r'), respectively, (with $u \neq u'$) within a distance d that satisfies

$$
\frac{17}{\varepsilon} \leq \frac{2d}{2d - (n - \ell)}.
$$

Solving for d and using $\ell = (1 - 4(p - \bar{p}) + \varepsilon/2)n$ shows that d satisfies

$$
d \leq 2(p - \bar{p})n\frac{17}{17 + \varepsilon} - \frac{\varepsilon n}{4}\frac{17}{17 + \varepsilon} < 2(p - \bar{p})n - \varepsilon n/8.
$$

Let $\Delta = 2(p - \bar{p})n - \varepsilon n/8$.

Let γ be the fraction of pairs (u, r) and (u', r') in $B_{\mathbf{y}_1}$ that satisfy E_2 and E_3. We would like to lower bound γ. A union

bound shows that the probability over the selection of S gives the upper bound

$$\Pr\left(\bigcup_S \{d_H(\mathbf{X}^i, \mathbf{X}^j) < \Delta\} \text{ and } \{\mathbf{U}^i \neq \mathbf{U}^j\}\right) \leq m^2\gamma. \tag{14}$$

However, the earlier argument shows that by selecting $m = 17/\varepsilon$ pairs in S, we get a lower bound of $(\varepsilon/5)^{m-1}$ on the probability that (a) all $\{\mathbf{U}^i\}$ are distinct, and (b) at least one pair $\mathbf{X}^i$, $\mathbf{X}^j$ has distance less that Δ:

$$\Pr\left(\{\text{all } \mathbf{U}^i \in S \text{ are distinct}\} \text{ and } \bigcup_S \{d_H(\mathbf{X}^i, \mathbf{X}^j) < \Delta\}\right)$$
$$\geq (\varepsilon/5)^{m-1}. \tag{15}$$

As the event analyzed in (14) includes that analyzed in (15), we have that

$$\gamma \geq \frac{1}{m^2}\left(\frac{\varepsilon}{5}\right)^{m-1} = \frac{17^2}{\varepsilon^2}\left(\frac{\varepsilon}{5}\right)^{17/\varepsilon - 1}.$$

Therefore, by the definition of γ, we conclude our assertion. ∎

The next step is to show that Calvin does not "run out" of bit flips during the second "push" phase of his attack. This follows directly from Chernoff's bound [25].

We now analyze Calvin's action during the "push" phase. This action can be viewed as being equivalent to the following two stages. In the first stage, $d_H(\mathbf{X}_2, \mathbf{X}_2')$ bits are drawn i.i.d. Bernoulli-$(1/2)$—these bits comprise the *intended error vector* $\hat{\mathbf{e}}$. However, Calvin may not have the power to impose this intended vector in the push phase if the weight of $\hat{\mathbf{e}}$ is too large. In general, the bit flips in Calvin's actual error vector $\mathbf{e}_2$ correspond to the components of $\hat{\mathbf{e}}$ up to the point that he runs out of his bit-budget.

Let d be the distance between the $\mathbf{X}_2$ chosen by Alice and $\mathbf{X}_2'$ chosen by Calvin, and let the event E_4 be defined as

$$E_4 = \left\{wt_H(\hat{\mathbf{e}}) \in \left(\frac{d}{2} - \frac{\varepsilon n}{16}, \frac{d}{2} + \frac{\varepsilon n}{16}\right)\right\}. \tag{16}$$

Claim 7: For the "babble-and-push" attack Adv

$$\Pr_{\text{Adv}}\left(E_4 \mid E_2, E_3\right) \geq 1 - 2^{-\Omega(\varepsilon^2 n)}. \tag{17}$$

Proof: As d is the distance between the $\mathbf{X}_2$ chosen by Alice and $\mathbf{X}_2'$ chosen by Calvin, without any constraint, Calvin would flip $d/2$ locations in expectation. Conditioned on E_2 and E_3, we have the following upper bound:

$$\frac{d}{2} \leq (p - \bar{p})n - \varepsilon n/16.$$

Assume that $d/2 = (p-\bar{p})n - \varepsilon n/16$ (for smaller values of d the bound is only tighter). By Chernoff's bound [25], the probability that the number of bit flips in $\hat{e}$ (i.e., the Hamming weight of $\hat{e}$) deviates from the expectation by more than $\varepsilon n/16$ is at most $2^{-\Omega(\varepsilon^2 n)}$. ∎

Note that the number of bit flips in the first phase of the algorithm is exactly $\bar{p}n$, and thus, Claim 7 implies that with high probability, the total number of bit flips in $\hat{\mathbf{e}}$ in the second phase will not exceed $\frac{d}{2} + \frac{\varepsilon n}{16} \leq (p - \bar{p})n$ and will not be significantly

less than that expected (i.e., less than $\frac{d}{2} - \frac{\varepsilon n}{16}$—in this case Bob might be able to conclude that $\mathbf{X}_2'$ was not transmitted). If this is not the case, our analysis assumes Calvin (in the worst case for him) fails to jam Alice's transmission to Bob.

Theorem 8: For any code with stochastic encoding of rate $R = C + \varepsilon$, under Calvin's "babble-and-push" strategy, the average error probability $\bar{\varepsilon}$ is lower bounded by $\varepsilon^{O(1/\varepsilon)}$.

Proof: The main idea behind the proof of our outer bound is that conditioned on events E_0, E_2, E_3, and E_4, (whose probabilities of occurrence are analyzed in Claims 4, 6, and 7), Calvin can "symmetrize" the channel [14]. That is, Calvin can choose to inject bit flips in a manner so that Bob is unable to distinguish between two possible codewords $\mathbf{x}$ and $\mathbf{x}'$ (corresponding to different messages u and u') transmitted by Alice. Calvin does this by ensuring (with probability bounded away from zero) that the codeword received by Bob, $\mathbf{y}$, is equally likely to be either $\mathbf{x} + \mathbf{e}$ or $\mathbf{x}' + \mathbf{e}'$ for two valid pairs $(\mathbf{x}, \mathbf{e})$ and $(\mathbf{x}', \mathbf{e}')$ of transmitted codewords and bit-flip vectors.

Let (u, r) denote the message and randomness of Alice, $\mathbf{y}_1$ be the received codeword in the "babble" phase, and (u', r') be the message and randomness chosen by Calvin for the "push" phase. Let $\rho(\mathbf{y}_1, u, r, u', r')$ be the joint distribution of these variables under Alice's uniform choice of (u, r) and Calvin's attack. For each $\mathbf{y}$, let $\rho(\mathbf{y}|\mathbf{y}_1, u, r, u', r')$ be the conditional distribution of $\mathbf{y}$ under Calvin's attack.

The error probability can be written as

$$\bar{\varepsilon} = \sum_{\mathbf{y}_1, u, r, u', r'} \rho(\mathbf{y}_1, u, r, u', r')$$
$$\sum_{\mathbf{y}_2} \rho(\mathbf{y}|\mathbf{y}_1, u, r, u', r') \Pr(\Psi(\mathbf{y}) \neq u).$$

Let $\mathcal{F}$ be the set of tuples $(\mathbf{y}_1, u, r, u', r')$ satisfying events E_0, E_2, and E_3. Claims 4 and 6 show that $\rho(\mathcal{F}) \geq (\varepsilon/4) \cdot \varepsilon^{O(1/\varepsilon)}$. For $(\mathbf{y}_1, u, r, u', r') \in \mathcal{F}$, we have that $u \neq u'$, and that $\mathbf{x}_2(u, r)$ and $\mathbf{x}_2(u', r')$ are sufficiently close.

Assuming E_4 holds, if $\mathbf{y}_2$ results from $\mathbf{x}_2$ via $\hat{\mathbf{e}}$, then $\mathbf{y}_2$ may also have resulted from $\mathbf{x}_2'$ via $\hat{\mathbf{e}}^C$ (the binary complement of $\hat{\mathbf{e}}$). Since $\hat{\mathbf{e}}$ is generated via i.i.d. Bernoulli-$(1/2)$ components, $\hat{\mathbf{e}}$ and $\hat{\mathbf{e}}^C$ have the same probability.

Thus, the conditional distribution is symmetric:

$$\rho(\mathbf{y}|\mathbf{y}_1, u, r, u', r') = \rho(\mathbf{y}|\mathbf{y}_1, u', r', u, r). \tag{18}$$

Then, for $(\mathbf{y}_1, u, r, u', r') \in \mathcal{F}$, by Claim 7,

$$\sum_{\mathbf{y}_2 \in \mathcal{G}} \rho(\mathbf{y}_2|\mathbf{y}_1, u, r, u', r') \geq 1 - 2^{-\Omega(\varepsilon^2 n)}.$$

Now, returning to the overall error probability, let $\rho(\mathbf{y}_1)$ be the unconditional probability of Bob receiving $\mathbf{y}_1$ in the "babble" phase, where the probability is taken over Alice's uniform choice of (u, r) and Calvin's random babble $\mathbf{e}_1$. Since the *a posteriori* distribution of (u, r) and (u', r') given $\mathbf{y}_1$ are independent and both uniform in $B_{\mathbf{y}_1}$, the joint distribution can be written as

$$\rho(\mathbf{y}_1, u, r, u', r') = \rho(\mathbf{y}_1) \cdot \frac{1}{|B_{\mathbf{y}_1}|^2}$$
$$= \rho(\mathbf{y}_1, u', r', u, r).$$

Recall that for any $\mathbf{y}_2 \in \mathcal{G}$,

$$\rho(\mathbf{y}_2|\mathbf{y}_1, u, r, u', r') = \rho(\mathbf{y}_2|\mathbf{y}_1, u', r', u, r).$$

Thus

$$2\bar{\varepsilon} \geq \sum_{\mathcal{F}} \rho(\mathbf{y}_1, u, r, u', r')$$

$$\left(\sum_{\mathbf{y}_2 \in \mathcal{G}} \rho(\mathbf{y}_2|\mathbf{y}_1, u, r, u', r') \Pr(\Psi(\mathbf{y}_1, \mathbf{y}_2) \neq u) \right.$$

$$\left. + \sum_{\mathbf{y}_2 \in \mathcal{G}} \rho(\mathbf{y}_2|\mathbf{y}_1, u', r', u, r) \Pr(\Psi(\mathbf{y}_1, \mathbf{y}_2) \neq u') \right)$$

$$\geq \sum_{\mathcal{F}} \rho(\mathbf{y}_1, u, r, u', r') \sum_{\mathbf{y}_2 \in \mathcal{G}} \rho(\mathbf{y}_2|\mathbf{y}_1, u, r, u', r')$$

$$(\Pr(\Psi(\mathbf{y}_1, \mathbf{y}_2) \neq u) + \Pr(\Psi(\mathbf{y}_1, \mathbf{y}_2) \neq u'))$$

$$\geq \sum_{\mathcal{F}} \rho(\mathbf{y}_1, u, r, u', r') \sum_{\mathbf{y}_2 \in \mathcal{G}} \rho(\mathbf{y}_2|\mathbf{y}_1, u, r, u', r') \qquad (19)$$

$$\geq \varepsilon/4 \cdot \varepsilon^{O(1/\varepsilon)} \cdot \left(1 - 2^{-\Omega(\varepsilon^2 n)}\right).$$

∎

Our analysis implies a refined statement of Theorem 1. Namely, let c be a sufficiently large constant. For any block-length n, any $\varepsilon > 0$, and any encoding/decoding scheme of Alice and Bob of rate $\left(C(p) + \sqrt{\frac{c}{n}}\right) + \varepsilon$, Calvin can cause a decoding error probability of at least $\varepsilon^{O(1/\varepsilon)}$.

IV. IMPROVED BOUNDS FOR DETERMINISTIC CODES

We now present an alternative analysis for the case of deterministic encoding. Without loss of generality, we assume each codeword corresponds to a unique message in $\mathcal{U}$ so there are 2^{Rn} distinct equiprobable codewords in $\{\mathbf{x}(u) : u \in \mathcal{U}\}$, with a unique codeword for each message. The attack is the same as in Section III. Apart from the simpler proof, the analysis below gives a decoding error $\bar{\varepsilon}$ proportional to ε, which improves over the decoding error presented for stochastic encoding appearing in the body of this work.

Using the notation of Section III, for any vector $\mathbf{y}_1$, consider the set

$$B_{\mathbf{y}_1} = \{(\mathbf{x}, \mathbf{e}_1) : \mathbf{x}_1 + \mathbf{e}_1 = \mathbf{y}_1, \ \mathbf{e}_1 = \bar{p}n\}. \qquad (20)$$

Here, $\mathbf{e}_1$ represents the potential error vector that Calvin imposes in the first stage of its attack on the transmitted codeword $\mathbf{x}$. Notice that the set $B_{\mathbf{y}_1}$ defined above is analogous to the set $B_{\mathbf{y}_1}$ defined in Section III. Namely, for any message u, a pair $(u, r) \in B_{\mathbf{y}_1}$ in Section III corresponds to a pair $(\mathbf{x}(u), \mathbf{e}_1) \in B_{\mathbf{y}_1}$ defined above. We note that in the definition above, $B_{\mathbf{y}_1} \cap B_{\mathbf{y}_1'} = \phi$ for $\mathbf{y}_1 \neq \mathbf{y}_1'$ as we assume all codewords to be distinct.

Claim 9: With probability at least $1/2$ over the codeword $\mathbf{x}$ sent by Alice and the actions of Calvin in the first stage of his attack, the set $B_{\mathbf{y}_1}$ is of size at least $2^{\varepsilon n/4}/2$.

Proof: The proof is obtained by the following counting argument. The number of possible sets $B_{\mathbf{y}_1}$ is exactly $2^\ell =$

$2^{\alpha n + \varepsilon n/2}$. The number of pairs $(\mathbf{x}, \mathbf{e}_1)$ for a codeword $\mathbf{x}$ and an error vector $\mathbf{e}_1$ (to be applied in the first stage by Calvin) is

$$2^{Rn} \binom{\ell}{\bar{p}n} \geq 2^{Rn} \binom{\alpha n}{\bar{p}n}$$

$$\geq 2^{Rn} \cdot 2^{\alpha n H(\bar{p}/\alpha) - \varepsilon n/4}$$

$$\geq 2^{\alpha n + 3\varepsilon n/4}.$$

Here, the first inequality follows from the fact that $\ell \leq \alpha n$, the second inequality from the standard bound $2^{nH(k/n)}/(n+1) \leq \binom{n}{k}$ (for instance [26, Th. 11.1.3]) and the fact that n is sufficiently large with respect to $1/\varepsilon$ and hence $2^{-\varepsilon n/4}$ is smaller than any polynomial in $1/(n+1)$, and the third inequality from the starting assumption that R is at least $\varepsilon + \alpha(1 - H(\bar{p}/\alpha))$. Thus, the *average* size of a set $B_{\mathbf{y}_1}$ is at least $2^{\varepsilon n/4}$. Consider all the sets $B_{\mathbf{y}_1}$ of size less than half the average $2^{\varepsilon n/4}/2$. The total number of codewords in the union of these sets is at most

$$2^\ell \cdot 2^{\varepsilon n/4}/2 \leq 2^{Rn} \binom{\ell}{\bar{p}n} \cdot \frac{1}{2},$$

which is half the number of $(\mathbf{x}, \mathbf{e}_1)$ pairs. As each pair is chosen with the same probability, we conclude that with probability at least $1/2$, the pair $(\mathbf{x}, \mathbf{e}_1)$ appears in a set $B_{\mathbf{y}_1}$ which is of size at least $2^{\varepsilon n/4}/2$. This completes the proof of our assertion. ∎

We now show that Claim 9 above implies that the transmitted codeword $\mathbf{x}$ and the codeword $\mathbf{x}'$ chosen by Calvin are distinct and of *small* Hamming distance apart with a positive probability (independent of n).

Claim 10: Conditioned on Claim 9, with probability at least $\frac{\varepsilon}{64}$, $\mathbf{x} \neq \mathbf{x}'$ and $d_H(\mathbf{x}_2, \mathbf{x}_2') < 2(p - \bar{p})n - \varepsilon n/8$.

Proof: Consider the undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in which the vertex set $\mathcal{V}$ consists of the set $B_{\mathbf{y}_1}$ and two nodes $\mathbf{x}$ and $\mathbf{x}'$ are connected by an edge if $d_H(\mathbf{x}_2, \mathbf{x}_2') \leq d = 2(p - \bar{p})n - \varepsilon n/8$. The set of codewords defined by the suffixes of an independent set $\mathcal{I}$ in $\mathcal{G}$ corresponds to a binary error-correcting code with block-length $n - \ell = 4(p - \bar{p})n - \varepsilon n/2$ of size $|\mathcal{I}|$ and minimum distance d.

By Plotkin's bound [24] (reprised in Theorem 2), there do not exist binary error correcting codes with more than $\frac{2d}{2d - (4(p-\bar{p})n - \varepsilon n/2)} + 1$ codewords. Thus, $\mathcal{I}$, any maximal independent set in $\mathcal{G}$, must satisfy

$$|\mathcal{I}| \leq \frac{2(2(p - \bar{p})n - \varepsilon n/8)}{2(2(p - \bar{p})n - \varepsilon n/8) - 4(p - \bar{p})n + \varepsilon n/2} + 1$$

$$= \frac{16(p - \bar{p})}{\varepsilon} \leq \frac{16}{\varepsilon}. \qquad (21)$$

By Turán's theorem [27], any undirected graph $\mathcal{G}$ on $|\mathcal{V}|$ vertices and average degree Δ has an independent set of size at least $|\mathcal{V}|/(\Delta + 1)$. This, along with (21), implies that the average degree of our graph $\mathcal{G}$ satisfies

$$\frac{|\mathcal{V}|}{\Delta + 1} \leq |\mathcal{I}| \leq \frac{16}{\varepsilon}.$$

This, in turn, implies that

$$\Delta \geq \frac{\varepsilon|\mathcal{V}|}{16} - 1 \geq \frac{\varepsilon|\mathcal{V}|}{32}.$$

The second inequality holds for our setting of n, since $|\mathcal{V}|$ is of size at least $2^{\varepsilon n/4}$. To summarize the above discussion, we have shown that our graph G has *large* average degree of size $\Delta \geq \frac{\varepsilon|\mathcal{V}|}{32p}$. We now use this fact to analyze Calvin's attack.

By the definition of deterministic codes, any valid codeword in $\mathcal{X}^n$ is transmitted with equal probability. Also, by definition, both $\mathbf{x}$ (the transmitted codeword) and $\mathbf{x}'$ (the codeword chosen by Calvin) are in $\mathcal{V} = B_{\mathbf{y}_1}$. Hence, both $\mathbf{x}$ and $\mathbf{x}'$ are uniform in $B_{\mathbf{y}_1}$. This implies that with probability $|\mathcal{E}|/|\mathcal{V}|^2$, the nodes corresponding to codewords $\mathbf{x}$ and $\mathbf{x}'$ are distinct and connected by an edge in $\mathcal{G}$. This, in turn, implies that with probability $|\mathcal{E}|/|\mathcal{V}|^2$, $\mathbf{x} \neq \mathbf{x}'$ and $d_H(\mathbf{x}, \mathbf{x}') < 2(p - \bar{p})n - \varepsilon n/8$, as required. Now

$$\frac{|\mathcal{E}|}{|\mathcal{V}|^2} = \frac{\Delta|\mathcal{V}|}{2|\mathcal{V}|^2} \geq \frac{\varepsilon}{64}.$$

∎

The preceding claims provide the same guarantees as Claim 6 appearing in the body of the paper, and so Claim 7 follows. Namely, w.h.p., Calvin does not "run out" of his budget of pn bit flips. We conclude by proving that given the analysis above, Bob cannot distinguish between the case in which $\mathbf{x}$ or $\mathbf{x}'$ were transmitted, using a similar symmetrization argument.

Theorem 11: For any code with deterministic encoding and decoding of rate $R = C + \varepsilon$, under Calvin's "babble-and-push" strategy, the average error probability $\bar{\varepsilon}$ is lower bounded by $\frac{\varepsilon}{256}\left(1 - 2^{-\Omega(\varepsilon^2 n)}\right)$.

Proof: Let u be the message chosen by Alice, u' be the message chosen by Calvin, $\rho(\mathbf{y}_1, u, u')$ be the joint distribution of the output during the "babble" phase and these two messages, and $\rho(\mathbf{y}|\mathbf{y}_1, u, u')$ be the conditional distribution of the output on the result of the "babble" phase.

Let $\mathcal{G}'$ be the set of $\mathbf{y}_2$ such that Claim 7 is satisfied for $\mathbf{x}_2(u)$. As in the arguments of Theorem 8, Calvin's attack is symmetric, so that

$$\rho(\mathbf{y}_1, u, u') = \rho(\mathbf{y}_1, u', u)$$

and therefore we have

$$\sum_{\mathbf{y}_2 \in \mathcal{G}'} \rho(\mathbf{y}|\mathbf{y}_1, u, u') \geq 1 - 2^{-\Omega(\varepsilon^2 n)}.$$

Let $\mathcal{F}$ be the set of tuples $(\mathbf{y}_1, u, u')$ satisfying Claims 9 and 10. Following the analysis in Theorem 8, from (20) and applying Claims 9 and 10, we have

$$2\bar{\varepsilon} \geq \sum_{\mathcal{F}} \rho(\mathbf{y}_1, u, u')\left(1 - 2^{-\Omega(\varepsilon^2 n)}\right)$$
$$\geq \frac{\varepsilon}{128}\left(1 - 2^{-\Omega(\varepsilon^2 n)}\right).$$

Dividing both sides by 2 yields the result. ∎

V. CONCLUDING REMARKS

In this paper, we presented a novel upper bound on the rates achievable on binary additive channels with a causal adversary.

This model is weaker than the traditional worst case error model studied in coding theory, but is stronger than an i.i.d. model for the noise. Indeed, our results show the binary symmetric channel capacity $1 - H(p)$ is not achievable against causal adversaries. By contrast, previous work shows that a delay of Dn (with D a positive constant in $(0, 1]$) for the adversary allows Alice and Bob to communicate at rate $1 - H(p)$. Thus, the causal adversary is strictly more powerful than the delayed adversary (which, in turn, is no stronger than i.i.d. noise).

To show our bound, we demonstrated a new "babble-and-push" attack. The adversary increases the uncertainty at the decoder during the "babble" phase, enabling it to choose an alternative codeword during the "push" phase. The "push" phase succeeds because the adversary can effectively symmetrize the channel. We demonstrate that the upper bound presented herein holds against arbitrary codes, rather than simply against deterministic codes, as is common in the coding theory literature. Since our analysis pertains to adversarial jamming rather than random noise, the proof techniques presented may be of independent interest in the more general setting of AVCs.

APPENDIX
MINIMIZATION IN THEOREM 1

Let us denote $\bar{p}$ by x, and write the bound as a function of x as

$$f(x) = (1 - 4p + 4x)\left(1 - H\left(\frac{x}{1 - 4p + 4x}\right)\right).$$

So

$$f'(x) = 4 - \frac{d}{dx}\Bigg\{(1 - 4p + 4x)$$
$$\Bigg[-\frac{x}{1 - 4p + 4x}\log\frac{x}{1 - 4p + 4x}$$
$$-\frac{1 - 4p + 3x}{1 - 4p + 4x}\log\frac{1 - 4p + 3x}{1 - 4p + 4x}\Bigg]\Bigg\}$$
$$= 4 - \frac{d}{dx}\Bigg\{-x\log\frac{x}{1 - 4p + 4x}$$
$$-(1 - 4p + 3x)\log\frac{1 - 4p + 3x}{1 - 4p + 4x}\Bigg\}$$
$$= 4 - \Bigg\{-\log\frac{x}{1 - 4p + 4x} - \frac{(1 - 4p + 4x) - 4x}{1 - 4p + 4x}$$
$$-3\log\frac{1 - 4p + 3x}{1 - 4p + 4x}$$
$$-\frac{3(1 - 4p + 4x) - 4(1 - 4p + 3x)}{1 - 4p + 4x}\Bigg\}$$
$$= 4 + \Bigg\{\log\frac{x}{1 - 4p + 4x} + 3\log\frac{1 - 4p + 3x}{1 - 4p + 4x}\Bigg\}$$
$$= 4 + \log\frac{x(1 - 4p + 3x)^3}{(1 - 4p + 4x)^4}.$$

First, we check for any roots of $f'(x)$ in $0 \leq x \leq p$.

$$f'(x) = 0$$
$$\iff \frac{x(1 - 4p + 3x)^3}{(1 - 4p + 4x)^4} = \frac{1}{16}$$
$$\iff ((1 - 4p + 3x) + x)^4 = 16x(1 - 4p + 3x)^3$$
$$\iff (1 - 4p + 3x)^4 - 12x(1 - 4p + 3x)^3$$
$$+ 6x^2(1 - 4p + 3x)^2 + 4x^3(1 - 4p + 3x) + x^4 = 0.$$

We now substitute, for brevity, $a = (1 - 4p + 3x)/x$.

$$a^4 - 12a^3 + 6a^2 + 4a + 1 = 0$$
$$\iff (a - 1)(a^3 - 11a^2 - 5a - 1) = 0.$$

We now consider two cases. If $p = 0.25$, we have that $f(x) = 0$ for $x = 0$. Thus, setting $x = 0$ will yield the minimum value for $p = 1/4$.

For $p < 0.25$, we study the minimum value given $x > 0$. When $x > 0$ and $p < 0.25$, it holds that $1 - 4p > 0$ and $a > 3$. Thus, for $f'(x)$ to be zero, we require that $a^3 - 11a^2 - 5a - 1 = 0$, which can be found via the general formula for cubic equations (for instance, [28, Chap. 6]) to have one real solution and two complex conjugate solutions. The only real solution is

$$a_0 = \frac{1}{3}\left(11 + \sqrt[3]{1592 + 24\sqrt{33}} + \sqrt[3]{1592 - 24\sqrt{33}}\right)$$
$$\simeq 11.4445,$$

giving $x = (1 - 4p)/(a_0 - 3) \simeq (1 - 4p)/8.4445$. However, this value is greater than p for $p < 1/(a_0 + 1) \simeq 1/12.4445$. For $p \in [1/(a_0 + 1), 0.25]$, this solution is in the range $[0, p]$.

Now we will see that $f'(x)$ is negative for $0 < x < (1 - 4p)/(a_0 - 3)$.

For $p < 0.25$, and $0 < x < (1 - 4p)/(a_0 - 3)$, we have

$$a = \frac{1 - 4p + 3x}{x} > a_0 \simeq 11.4445,$$

so

$$a^3 - 11a^2 - 5a - 1 \simeq (a - a_0)(a^2 + 0.4445a + 0.087) > 0,$$

but since $(a - 1)(a^3 - 11a^2 - 5a - 1) > 0$, we have $f'(x) < 0$. By the continuity of the objective function, $f(0) = \lim_{x \to 0} f(x)$. So, $f(x)$ is decreasing in $0 \leq x < (1 - 4p)/a_0 - 3$, and thus, the optimum $\bar{p}$ is given by

$$\bar{p} = \min\left\{p, \frac{1 - 4p}{a_0 - 3}\right\}$$
$$= \min\left\{p, \frac{3(1 - 4p)}{\left(2 + \sqrt[3]{1592 + 24\sqrt{33}} + \sqrt[3]{1592 - 24\sqrt{33}}\right)}\right\}$$
$$\simeq \min\left\{p, \frac{1 - 4p}{8.4445}\right\}.$$

ACKNOWLEDGMENT

REFERENCES

[1] B. K. Dey, S. Jaggi, M. Langberg, and A. D. Sarwate, "Improved upper bounds on the capacity of binary channels with causal adversaries," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 681–685.

[2] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr, and L. R. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 2, pp. 157–166, Mar. 1977.

[3] E. N. Gilbert, "A comparison of signalling alphabets," *Bell Syst. Tech. J.*, vol. 31, pp. 504–522, 1952.

[4] R. R. Varshamov, "Estimate of the number of signals in error correcting codes," *Dokl. Acad. Nauk*, vol. 117, pp. 739–741, 1957.

[5] M. A. Tsfasman, S. G. Vludut, and D. Nogin, *Algebraic Geometric Codes: Basic Notions*, ser. Mathematical surveys and monographs. Providence, RI, USA: Amer. Math. Soc., 2007, vol. 139.

[6] R. Singleton, "Maximum distance q-nary codes," *IEEE Trans. Inf. Theory*, vol. IT-10, no. 2, pp. 116–118, Apr. 1964.

[7] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. Soc. Ind. Appl. Math.*, vol. 8, no. 2, pp. 300–304, 1960.

[8] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 (623–656), Oct. 1948.

[9] I. Haviv and M. Langberg, "Beating the Gilbert-Varshamov bound for online channels," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, 2011, pp. 1297–1301.

[10] D. Blackwell, L. Breiman, and A. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, no. 3, pp. 558–567, 1960.

[11] R. Ahlswede and J. Wolfowitz, "The capacity of a channel with arbitrarily varying channel probability functions and binary output alphabet," *Zeitschrift für Wahrscheinlichkeit und verwandte Gebiete*, vol. 15, no. 3, pp. 186–194, 1970.

[12] R. Ahlswede, "Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback," *Zeitschrift für Wahrscheinlichkeit und verwandte Gebiete*, vol. 25, pp. 239–252, 1973.

[13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest, Hungary: Akadémi Kiadó, 1982.

[14] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 181–193, Mar. 1988.

[15] A. D. Sarwate and M. Gastpar, "Rateless codes for AVC models," *IEEE Trans. Inf. Theory* vol. 56, no. 7, pp. 3105–3114, Jul. 2010.

[16] M. Langberg, "Private codes or succinct random codes that are (almost) perfect," presented at the 45th Ann. IEEE Symp. Found. Comput. Sci. Rome, Italy, 2004.

[17] A. Smith, "Scrambling adversarial errors using few random bits, optimal information reconciliation, and better private codes," in *Proc. 2007 ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 395–404.

[18] M. Langberg, "Oblivious communications channels and their capacity," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 424–429, Jan. 2008.

[19] M. Langberg, S. Jaggi, and B. Dey, "Coding against delayed adversaries," in *Proc. IEEE Int. Symp. Inf. Theory*, 2010, pp. 285–289.

[20] V. Guruswami and A. Smith, "Codes for computationally simple channels: Explicit constructions with optimal rate," in *Proc. Symp. Found. Comput. Sci.*, 2010, pp. 723–732.

[21] I. Csiszár and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Trans. Inf. Theory*, vol. 34, no. 1, pp. 27–34, Jan. 1988.

[22] B. K. Dey, S. Jaggi, and M. Langberg, "Codes against online adversaries," presented at the 47th Annu. Allerton Conf. Commun., Control, Comput., Monticello, IL, USA, Sep./Oct. 2009.

[23] M. Langberg, S. Jaggi, and B. Dey, "Binary causal-adversary channels," in *Proc. IEEE Int. Symp. Inf. Theory*, 2009, pp. 2723–2727.

[24] M. Plotkin, "Binary codes with specified minimum distance," *IRE Trans. Inf. Theory*, vol. 6, no. 4, pp. 445–450, 1960.

[25] N. Alon and J. H. Spencer, *The Probabilistic Method*. New York, NY, USA: Wiley-Interscience, 2011, vol. 73.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley-Interscience, 2006.

[27] P. Turán, "On the theory of graphs," *Colloq. Math.*, vol. 3, pp. 19–30, 1954.

[28] W. Dunham, *Journey through Genius: The Great Theorems of Mathematics*. New York, NY, USA: Wiley, 1990.

Bikash Kumar Dey (S'00–M'04) received his B.E. degree in Electronics and Telecommunication Engineering from Bengal Engineering College, Howrah, India, in 1996. He received his M.E. degree in Signal Processing and Ph.D. in Electrical Communication Engineering from the Indian Institute of Science in 1999 and 2003 respectively.

From August 1996 to June 1997, he worked at Wipro Infotech Global R&D. In February 2003, he joined Hellosoft India Pvt. Ltd. as a Technical Member. In June 2003, he joined the International Institute of Information Technology, Hyderabad, India, as Assistant Professor. In May 2005, he joined the Department of Electrical Engineering of Indian Institute of Technology Bombay where he works as Associate Professor. His research interests include information theory, coding theory, and wireless communication.

He was awarded the Prof. I.S.N. Murthy Medal from IISc as the best M.E. student in the Department of Electrical Communication Engineering and Electrical Engineering for 1998–1999 and Alumni Medal for the best Ph.D. thesis in the Division of Electrical Sciences for 2003–2004.

Sidharth Jaggi (M'00) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 2000, and the M.S. and Ph.D. degrees in 2001 and 2006, respectively, from the California Institute of Technology, Pasadena.

He is an Assistant Professor in Information Engineering at the Chinese University of Hong Kong. His primary research interests are in network coding, coding theory and information theory.

Michael Langberg (M'07) received his B.Sc. in mathematics and computer science from Tel-Aviv University in 1996, and his M.Sc. and Ph.D. in computer science from the Weizmann Institute of Science in 1998 and 2003 respectively.

Between 2003 and 2006, he was a postdoctoral scholar in the Computer Science and Electrical Engineering departments at the California Institute of Technology. He is an Associate Professor in the Mathematics and Computer Science department at the Open University of Israel. His research interests include information theory, combinatorics, and algorithm design.

Anand D. Sarwate (S'99–M'09) received B.S. degrees in electrical engineering and computer science and mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002 and the M.S. and Ph.D. degrees in electrical engineering in 2005 and 2008, respectively, from the University of California, Berkeley.

From 2008 to 2011 he was a postdoctoral researcher at the Information Theory and Applications Center at the University of California, San Diego. He is a Research Assistant Professor at the Toyota Technological Institute at Chicago. His research interests include information theory, distributed signal processing, machine learning, and privacy.

Dr. Sarwate received the Laya and Jerome B. Wiesner Student Art Award from MIT, and the Samuel Silver Memorial Scholarship Award and Demetri Angelakos Memorial Achievement Award from the EECS Department at University of California at Berkeley. He was awarded an NDSEG Fellowship from 2002 to 2005. He is a member of Phi Beta Kappa and Eta Kappa Nu.