Binary Causal-Adversary Channels

M. Langberg
Computer Science Division
Open University of Israel
Raanana 43107, Israel
mikel@openu.ac.il

S. Jaggi
Department of Information Engineering
Chinese University of Hong Kong
Shatin, N.T., Hong Kong
jaggi@ie.cuhk.edu.hk

B. K. Dey
Department of Electrical Engineering
Indian Institute of Technology Bombay
Mumbai, India, 400 076
bikash@ee.iitb.ac.in

Abstract— In this work 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 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 an *online* or *causal* manner. Namely, for each bit x_i the jammer's decision on whether to corrupt it or not (and on how to change it) must depend only on x_j for $j \leq i$. This is in contrast to the "classical" adversarial jammer which may base its decisions on its complete knowledge of $\mathbf{x}$. We present a non-trivial upper bound on the amount of information that can be communicated. We show that the achievable rate can be asymptotically no greater than $\min\{1 - H(p), (1 - 4p)^+\}$. Here $H(.)$ is the binary entropy function, and $(1 - 4p)^+$ equals $1 - 4p$ for $p \leq 0.25$, and 0 otherwise.

I. INTRODUCTION

Consider the following adversarial communication scenario. A sender Alice wishes to transmit a message u to a receiver Bob. To do so, Alice encodes u into a codeword $\mathbf{x}$ and transmits it over a *binary channel*. The codeword $\mathbf{x} = x_1, \ldots, x_n$ is a binary vector of length n. However, Calvin, a malicious adversary, can observe $\mathbf{x}$ and corrupt up to a p-fraction of the n transmitted bits, *i.e.*, pn bits.

In the classical adversarial channel model, e.g., [4], it is usually assumed that Calvin has full knowledge of the entire codeword $\mathbf{x}$, and based on this knowledge (together with the knowledge of the code shared by Alice and Bob) Calvin can maliciously plan what error to impose on $\mathbf{x}$. We refer to such an adversary as an *omniscient* adversary. For binary channels, the optimal rate of communication in the presence of an omniscient adversary has been an open problem in classical coding theory for several decades. The best known lower bound is given by the Gilbert-Varshamov bound [10], [18], which implies that Alice can transmit at rate $1 - H(2p)$ to Bob. Conversely, the tightest upper bound was given by McEliece et al. [12], and has a positive gap from the lower bound for all $p \in (0, 1/4)$ (see Fig. 1).

In this work we initiate the analysis of coding schemes that allow communication against certain adversaries that are weaker than the omniscient adversary. We consider adversaries that behave in a *causal* or *online* manner. Namely, for
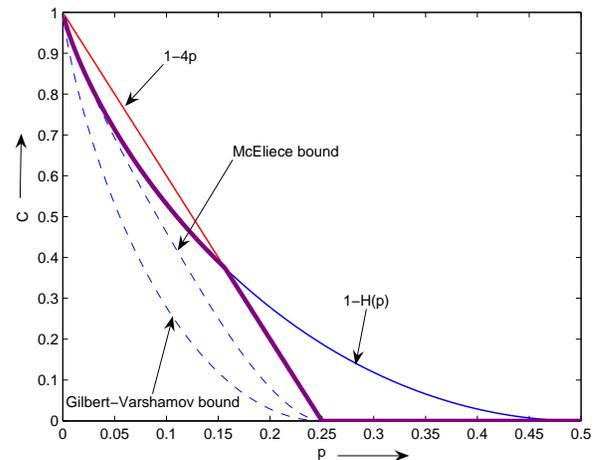
Fig. 1. Bounds on capacity of the adversarial channel. The bold line in purple is our upper bound of $\min\{1 - H(p), (1 - 4p)^+\}$.

each bit x_i, we assume that Calvin decides whether to change it or not (and if so, how to change it) based on the bits x_j, for $j \leq i$ alone, *i.e.*, the bits that he has already observed. In this case we refer to Calvin as a *causal* adversary.

Causal adversaries arise naturally in practical settings, where adversaries typically have no *a priori* knowledge of Alice's message u. In such cases they must simultaneously learn u based on Alice's transmissions, and jam the corresponding codeword $\mathbf{x}$ accordingly. This *causality* assumption is reasonable for many communication channels, both wired and wireless, where Calvin is not co-located with Alice. For example consider the scenario in which the transmission of $\mathbf{x} = x_1, \ldots, x_n$ is done during n channel uses over time, where at time i the bit x_i is transmitted over the channel. 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. As in the omniscient model, Calvin is restricted in the number of bits pn he can corrupt. This might be because of limited processing power or limited transmit energy.

Recently, the problem of codes against causal adversaries was considered and solved by the authors [6] for *large-q channels*, *i.e.*, channels where Alice's codeword $\mathbf{x} = x_1, \ldots, x_n$ is considered to be a vector of length n over a field of "large" size q. Each *symbol* x_i may represent a large packet of bits in practice. Calvin is allowed to arbitrarily

corrupt a p-fraction of the symbols, rather than bits. A tight characterization of the rate-region for various scenarios is given in [6], and computationally efficient codes that achieve these rate-regions are presented. However, the techniques used in characterizing the rate-region of causal adversaries over large-q channels do not work over binary channels. This is because each symbol in a large-q channel can contain within it a "small" hash that can be used to verify the symbol. This is the crux of the technique used to achieve the lower bounds in [6]. We currently do not know how to extend this method to binary channels. Conversely, for upper bounds, the geometry of the space of length-n codewords over large-q alphabets is significantly different than that corresponding to binary alphabets. For instance, for large-q channels the volume of an n-sphere of radius αn ($0 \leq \alpha \leq 1$) over F_q is $\sim q^{n\alpha}$, This leads to simpler bounds for large-q channels.

In this work we initiate the study of binary causal-adversary channels, and present two upper bounds on their capacity: $1 - H(p)$, and $(1 - 4p)^+$. The upper bound of $1 - H(p)$ is very "natural". Namely, it is not hard to verify that if Calvin attacks Alice's transmission by simulating the well-studied Binary Symmetric Channel [4], he can force a communication rate of no more than $1 - H(p)$. The upper bound of $(1 - 4p)^+$ presented in this work is non-trivial for both its implications and its proof techniques. The bound demonstrates that at least for some values of p, the achievable rate is bounded away from $1 - H(p)$. For $p \in (p_0, 0.5)$, $1 - 4p$ is strictly less than $1 - H(p)$ (here p_0 is the value of p satisfying $H(p) = 4p$, and can be computed to be approximately $0.15642\ldots$). In fact for $p \in (0.25, 0.5)$ our bound implies that no communication at positive rate is possible, which is much stronger than the result obtained by the upper bound of $1 - H(p)$ (see Fig. 1). Our proof techniques include a combination of tools from the fields of Extremal Combinatorics (e.g. Turán's theorem [17]), and classical Coding Theory (e.g. the Plotkin bound [14], [2]).

II. MODEL

For any integer i let $[i]$ denote the set $\{1, \ldots, i\}$. Let $R \geq 0$ be Alice's *rate*. An (n, Rn)-*code* $\mathcal{C}$ is defined by Alice's encoder and Bob's corresponding decoder, as below.

Alice: Alice's message u is assumed to be a random variable $\mathbf{U}$ with entropy Rn, over alphabet $\mathcal{U}$. We consider two types of encoding schemes for Alice.

For *deterministic codes*, Alice's message $\mathbf{U}$ is assumed to be uniformly distributed over $\mathcal{U} = [2^{Rn}]$. Her *deterministic encoder* is a deterministic function $f_D(.)$ that maps every u in $[2^{Rn}]$ to a vector $\mathbf{x}(u) = (x_1, \ldots, x_n)$ in $\{0, 1\}^n$. Alice's *codebook* $\mathcal{X}$ is the collection $\{\mathbf{x}(u)\}$ of all possible transmitted codewords.

More generally, Alice and Bob may use *probabilistic codes*. For such codes, the random variable $\mathbf{U}$ corresponding to Alice's message p_U may have an arbitrary distribution p_U (with entropy Rn) over an arbitrary alphabet $\mathcal{U}$. Alice's *codebook* $\mathcal{X}$ is an arbitrary collection $\{\mathcal{X}(u)\}$ of subsets of $\{0, 1\}^n$. For each subset $\mathcal{X}(u) \subset \mathcal{X}$, there is a corresponding

codeword random variable $\mathbf{X}(u)$ with *codeword distribution* $p_{X(u)}$ over $\mathcal{X}(u)$. For any value $\mathbf{U} = u$ of the message, Alice's encoder choses a codeword from $\mathcal{X}(u)$ randomly from the distribution $p_{X(u)}$. Alice's message distribution p_U, codebook $\mathcal{X}$, and all the codebook distributions $p_{X(u)}$ are all known to both Bob and Calvin, but the values of the random variables $\mathbf{U}$ and $\mathbf{X}(.)$ are unknown to them. If $\mathcal{X}(u) = \{\mathbf{x}(u, r) : r \in \Lambda_u\}$, then the transmitted codeword $\mathbf{X}(\mathbf{U})$ has the probability distribution given by $\Pr[\mathbf{X}(\mathbf{U}) = \mathbf{x}(u, r)] = p_U(u)p_{X(u)}(\mathbf{x}(u, r))$. Let p be the overall distribution of codewords $\mathbf{x} = \mathbf{x}(u, r)$ of Alice. It holds that $p(\mathbf{x}(u, r)) = p_U(u)p_{X(u)}(\mathbf{x})$ and $p(\mathbf{x}) = \sum_U p_U(u)p_{X(u)}(\mathbf{x})$.

Calvin/Channel: Calvin possesses n *jamming functions* $g_i(.)$ and n arbitrary jamming random variables $\mathbf{J}_i$ that satisfy the following constraints.

Causality constraint: For each $i \in [n]$, the jamming function $g_i(.)$ maps $\mathbf{x}^i = (x_1, \ldots, x_i)$ and $\mathbf{J}^i = (\mathbf{J}_1, \ldots, \mathbf{J}_i)$ to an element of $\{0, 1\}$.

Power constraint: The number of indices $i \in [n]$ for which the value of $g_i(.)$ equals 1 is at most pn. That is, for all $\mathbf{x}^n, \mathbf{J}^n, \sum_i g_i(\mathbf{x}^i, \mathbf{J}^i) \leq pn$.

The *output* of the channel is the set of bits $y_i = x_i \oplus g_i(\mathbf{x}^i, \mathbf{J}^i)$ for $i = 1, \ldots, n$.

Bob: Bob's *decoder* is a (potentially) probabilistic function $h(.)$ of the received vector $\mathbf{y}$. It maps the vectors $\mathbf{y} = (y_1, \ldots y_n)$ in $\{0, 1\}^n$ to the messages in $\mathcal{U}$.

Code parameters: Bob is said to make a *decoding error* if the message u' he decodes differs from the message u encoded by Alice. The *probability of error* for a given message u is defined as the probability, over Alice, Calvin and Bob's random variables, that Bob makes a decoding error. The *probability of error* of the code $\mathcal{C}$ is defined as the average over all $u \in \mathcal{U}$ of the probability of error for message u.

We define two types of rates and corresponding capacities.

The rate R is said to be *weakly achievable* if for every $\varepsilon > 0$, $\delta > 0$ and every sufficiently large n there exists an $(n, (R - \delta)n)$-code that allows communication with probability of error at most ε. The supremum over n of the weakly achievable rates is called the *weak capacity* and is denoted by C^{w}.

The rate R is said to be *strongly achievable*[1] if for every $\delta > 0$, $\exists \alpha > 0$ so that for sufficiently large n there exists an $(n, (R-\delta)n)$-code that allows communication with probability of error at most $e^{-\alpha n}$. The supremum over n of the strongly achievable rates is called the *strong capacity* and is denoted by C^{s}.

Remark: Since a rate that is strongly achievable is always weakly achievable but the converse is not true in general, $C^{\text{w}} \geq C^{\text{s}}$.

---

[1]This definition is motivated by the extensive literature on error exponents in information theory – for large classes of information-theoretic problems, e.g. [9], [5], the probability of error of the coding scheme is required to decay exponentially in block length.

III. Related work and our results

To the best of our knowledge, communication in the presence of a causal adversary has not been explicitly addressed in the literature (other than our prior work for causal adversaries over large-q channels). Nevertheless, we note that the model of causal channels, being a natural one, has been "on the table" for several decades and the analysis of the online/causal channel model appears as an open question in the book of Csiszár and Korner [5] (in the section addressing Arbitrary Varying Channels [1]). Various variants of causal adversaries have been addressed in the past, for instance [1], [11], [15], [16], [13] – however the models considered therein differ significantly from ours.

At a high level, we show that for causal adversaries, for a large range of p (for all $p > 0.25$), the maximum achievable rate equals that of the classical "omniscient" adversarial model (i.e., 0). This may at first come as a surprise, as the online adversary is weaker than the omniscient one, and hence one may suspect that it allows a higher rate of communication.

We have two main results. Theorem 1 gives an upper bound on the weak capacity C^{w} if Alice's encoder is deterministic. Theorem 2 gives an upper bound on the strong capacity C^{s} in the more general case where Alice's encoder is probabilistic. Due to certain limitations of our proof techniques, we do not present any bounds on the weak capacity in the latter setting. The upper bound in both cases equals $\min\{1 - H(p), (1 - 4p)^+\}$.

Theorem 1 (Deterministic encoder): For deterministic codes, $C^{\mathrm{s}} \leq C^{\mathrm{w}} \leq \min\{1 - H(p), (1 - 4p)^+\}$.

Theorem 2 (Probabilistic encoder): For probabilistic codes, $C^{\mathrm{s}} \leq \min\{1 - H(p), (1 - 4p)^+\}$.

We note that under a very weak notion of capacity in which one only requires the success probability to be bounded away from zero (instead of approaching 1), the capacity of the omniscient channel, and thus the binary causal-adversary channel, approaches $1 - H(p)$. This follows by the fact that for n sufficiently large and $\ell \geq 4$ there exists (n, Rn) codes which are (ℓ, pn) *list decodable* with $R = 1 - H(p)(1 + 1/\ell)$ [7]. Communicating using an (ℓ, pn) list decodable code allows Bob to decode a list of size ℓ of messages which includes the message transmitted by Alice. Choosing a message uniformly at random from his list, Bob decodes correctly with probability at least $1/\ell$.

A. Outline of proof techniques

The upper bound of $1 - H(p)$ follows directly by describing an attack for Calvin wherein he approximately simulates a BSC(p) (Binary Symmetric Channel [4] with crossover probability p). More precisely, for each $i \in [n]$ and any sufficiently small $\varepsilon > 0$, Calvin flips x_i with probability $p - \varepsilon$ until he runs out of his budget of pn bit-flips. By the Chernoff bound [3], with very high probability he does not run out of his budget, and is therefore indistinguishable from a BSC($p - \varepsilon$). But it is well-known [4] that in this case the optimal rate of communication from Alice to Bob

is $1 - H(p - \varepsilon)$. Taking the limit when $\varepsilon \to 0$ implies our bound.

The upper bound of $(1 - 4p)^+$ is more involved. For the case where Alice's encoder is deterministic, the proof of Theorem 1 has the following overall structure. Assume for sake of contradiction that Alice attempts to communicate at rate greater than $R = (1 - 4p)^+$. To prove our upper bound we design the following *wait-and-push* attack for Calvin.

Calvin starts by waiting for Alice to transmit approximately Rn bits. As Alice is assumed to communicate at rate greater than R, the set of Alice's codewords $\mathcal{X}'$ *consistent* with the bits Calvin has seen so far is "large" with "high probability". Calvin constructs $\mathcal{X}'$ and chooses a codeword $\mathbf{x}'$ uniformly at random from $\mathcal{X}'$. He then actively "pushes" $\mathbf{x}$ in the direction of $\mathbf{x}'$ by flipping, with probability $1/2$, each future x_i that differs from x'_i. If Calvin succeeds in pushing $\mathbf{x}$ to a word $\mathbf{y}$ roughly midway between $\mathbf{x}$ and $\mathbf{x}'$, a careful analysis demonstrates that regardless of Bob's decoding strategy, Bob is unable to determine whether Alice transmitted $\mathbf{x}$ or $\mathbf{x}'$ — causing a decoding error of $1/2$ in this case. So, to prove our bound, we must show that with constant probability (independent of the block length n) Calvin will indeed succeed in pushing $\mathbf{x}$ to $\mathbf{y}$. Namely, that Alice's codeword $\mathbf{x}$ and the codeword chosen at random by Calvin $\mathbf{x}'$ are of distance at most $2pn$. Roughly speaking, we prove the above by a detailed analysis of the distance structure of the set of codewords in any code using tools from extremal combinatorics and coding theory.

The case where Alice's encoder may be randomized is more technically challenging, and is considered in Theorem 2. At a high level, the strategy of Calvin for a probabilistic encoder follows that outlined for the deterministic case. However, there are two main difficulties in its extended analysis. Firstly, the symmetry between $\mathbf{x}$ and $\mathbf{x}'$ no longer exists. Namely, the fact that Bob may not be able to distinguish which of the two were transmitted by Alice does not necessarily cause a significant decoding error, since the probability of $\mathbf{x}'$ being transmitted by Alice may well be significantly smaller than the probability that $\mathbf{x}$ was transmitted. Secondly, the fact that both $\mathbf{x}$ and $\mathbf{x}'$ may correspond to the same message u places the entire scheme in jeopardy. As it now no longer matters if Bob decodes to $\mathbf{x}$ or $\mathbf{x}'$, in both cases the decoded message will be that sent by Alice.

To overcome these difficulties, we describe a more intricate analysis of Calvin's attack. Roughly speaking, we prove that a "large" subset $\mathcal{X}''$ of $\mathcal{X}'$ behaves "well". Any $\mathbf{x}'$ chosen uniformly at random from $\mathcal{X}'$, with "significant" probability, is in $\mathcal{X}''$, and has three properties corresponding to those when Alice uses a deterministic encoder. That is, $\mathbf{x}'$ is sufficiently *close* to $\mathbf{x}$ as desired, it has approximately the same probability of transmission that $\mathbf{x}$ does (thus preserving the needed symmetry), and it also corresponds to a message that differs from that corresponding to $\mathbf{x}$. All in all, we show that the above three properties hold with probability $1/\mathrm{poly}(n)$, which suffices to bound the strong capacity of

the channel at hand (but not the weak capacity).

In case of a randomized encoder of Alice, we assume that the messages may have nonuniform distribution, and also any message is encoded into one of a set of possible codewords as per some probability distribution in that set. One may think of various other ways of encoding, for example the following, to confuse Calvin. But as we discuss in the next paragraph, such schemes are also covered in our setup.

Multiple codebooks: In this scheme, Alice maintains a set of codes $C_1, C_2, \ldots, C_L$. For transmitting a message u, she randomly selects the code C_i with probability q_i. If the set of messages is $\mathcal{U} = \{1, 2, \ldots, M\}$ with a probability distribution given by $p_i \triangleq Pr\{u = i\}$, and the code C_r contains the codewords $\{\mathbf{x}(u, r) \mid u = 1, 2, \ldots, M\}$, then in our setup, the corresponding codebook for the message u will be $\mathcal{X}(u) = \{\mathbf{x}(u, r) \mid r = 1, 2, \ldots, L\}$. This codebook may have less than L codewords due to common codewords in the original codes. The induced probability distribution in this codebook of u is given by $Pr\{\mathbf{x}|u\} = \sum_{r:\mathbf{x}(u,r)=\mathbf{x}} q_r$.

If Alice picks a code and uses it to encode several messages, even then she does not gain anything. First, if she uses the same code to encode too many messages (and calvin knows the encoding scheme, as assumed), then both Bob and Calvin will know the code used after receiving or 'reading' some codewords. On the other hand, if a randomly chosen code is used only to encode a block of few messages this is equivalent to using a longer ('superblock') code in our setup. The only difference is that the probability of error analysed in our set up is the probability of error in decoding the 'superblocks' rather than the smaller blocks/codewords.

The proofs of the upper bounds corresponding to $1 - H(p)$ have already been sketched in Section III-A. Hence we only provide proofs of the upper bounds corresponding to $(1 - 4p)^+$ in Theorems 1 and 2.

IV. PROOF OF THEOREM 1

Let $R = (1 - 4p)^+ + \varepsilon$ for some $\varepsilon > 0$. Let $\log(.)$ denote the binary logarithm, here and throughout. By assumption for deterministic codes, Alice's message space $\mathcal{U}$ is of size 2^{Rn}. Here we assume for that 2^{Rn} in an integer. This implies that the set $\mathcal{X}$ of Alice's transmitted codewords is of size 2^{Rn}. [2]

We now present Calvin's attack. We show that for any fixed $\varepsilon > 0$, regardless of Bob's decoding strategy, there is a decoding error with constant probability (namely, the error probability is independent of n). Calvin's attack is in two stages. First Calvin *passively* waits until Alice transmits $\ell = (R - \varepsilon/2)n$ bits over the channel. Let $\mathbf{x}^\ell \in \{0, 1\}^\ell$ be the value of the codeword observed so far. He then considers the set of codewords that are consistent with the observed $\mathbf{x}^\ell$. Namely, Calvin constructs the set $\mathcal{X}|_{\mathbf{x}^\ell} = \{\mathbf{x} = x_1, \ldots, x_n \in$

---

[2] In fact, $\mathcal{X}$ may be smaller, however we note that for codes of optimal rate, $|\mathcal{X}|$ is of size *exactly* 2^{Rn}. If $|\mathcal{X}| < 2^{Rn}$, then for some transmitted codeword $\mathbf{x}$ at least two messages u and u' must both be encoded to $\mathbf{x}$. On receiving $\mathbf{x}$, Bob's probability of error is maximal – it is at least $1/2$. Therefore changing the codebook so as to encode u' as some $\mathbf{x}' \notin \mathcal{X}$ cannot increase the probability of decoding error.

$\mathcal{X} \mid x_1, \ldots, x_\ell = \mathbf{x}^\ell\}$. He then chooses an element $\mathbf{x}' \in \mathcal{X}|_{\mathbf{x}^\ell}$ uniformly at random. In the second stage, Calvin follows a *random bit-flip strategy*. That is, for each remaining bit x_i' of $\mathbf{x}'$ that differs from the corresponding bit x_i of $\mathbf{x}$ transmitted, he flips the transmitted bit with probability $1/2$, until he has either flipped pn bits, or until $i = n$.

We analyze Calvin's attack by a series of claims. We first show that with high probability (w.h.p.) the set $\mathcal{X}|_{\mathbf{x}^\ell}$ is *large*.

Claim 4.1: With probability at least $1 - 2^{-\varepsilon n/4}$, the set $\mathcal{X}|_{\mathbf{x}^\ell}$ is of size at least $2^{\varepsilon n/4}$.

Proof: The number of messages u for which $\mathcal{X}|_{\mathbf{x}^\ell(u)}$ is of size less than $2^{\varepsilon n/4}$ is at most the number of distinct prefixes $\mathbf{x}^\ell$ times $2^{\varepsilon n/4}$, which in turn is at most $2^{\ell + \varepsilon n/4} = 2^{(R - \varepsilon/4)n}$. ∎

Now assume that the message u is such that its corresponding set $\mathcal{X}|_{\mathbf{x}^\ell(u)}$ is of size at least $2^{\varepsilon n/4}$. We now show that this 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 4.2: Conditioned on Claim 4.1, with probability at least $\frac{\varepsilon}{64p}$, $\mathbf{x} \neq \mathbf{x}'$ and $d_H(\mathbf{x}, \mathbf{x}') < 2pn - \varepsilon n/8$.

Proof: Consider the undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in which the vertex set $\mathcal{V}$ consists of the set $\mathcal{X}|_{\mathbf{x}^\ell}$ and two nodes are connected by an edge if their Hamming distance is less than $d = 2pn - \varepsilon n/8$. An *independent set* $\mathcal{I}$ in $\mathcal{G}$ corresponds to a subset of codewords in $\{0, 1\}^n$ that are all (pairwise) at distance greater than d.

Since the codewords in $\mathcal{X}|_{\mathbf{x}^\ell}$ all have the same prefix $\mathbf{x}^\ell$, one may consider only the suffix (of length $n - \ell = 4pn - \varepsilon n/2$) of the codewords in $\mathcal{X}|_{\mathbf{x}^\ell}$. Here we assume $p \leq 0.25$, minor modifications in the proof are needed for larger p. The set of vectors defined by the suffixes in an independent set $\mathcal{I}$ of $\mathcal{G}$ now corresponds to a binary error-correcting code of length $4pn - \varepsilon n/2$, with $|\mathcal{I}|$ codewords and minimum distance d.

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

$$|\mathcal{I}| \leq \frac{2(2pn - \varepsilon n/8)}{2(2pn - \varepsilon n/8) - 4pn + \varepsilon n/2} + 1 = \frac{16p}{\varepsilon} \quad (1)$$

By Turán's theorem [17], any undirected graph $\mathcal{G}$ of size $|\mathcal{V}|$ and average degree Δ has an independent set of size at least $|\mathcal{V}|/(\Delta + 1)$. This, along with (1) implies that the average degree of our graph $\mathcal{G}$ satisfies

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

This in turn implies that

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

The second inequality is for large enough n, since $|\mathcal{V}|$ is of size at least 2^{Rn}. 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 codeword in $\mathcal{X}$ is transmitted with equal probability. Also, by definition both $\mathbf{x}$ (the transmitted codeword) and $\mathbf{x}'$ (the codeword chosen by Calvin) are in $\mathcal{V} = \mathcal{X}|_{\mathbf{x}^\ell}$. Hence both $\mathbf{x}$ and $\mathbf{x}'$ are uniform in $\mathcal{X}|_{\mathbf{x}^\ell}$. 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 $d_H(\mathbf{x}, \mathbf{x}') < 2pn - \varepsilon n/8$, as required. Now

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

■

Conditioned on Claim 4.2, Calvin's codeword $\mathbf{x}'$ is very close to Alice's transmitted codeword $\mathbf{x}$. Specifically, $d_H(\mathbf{x}, \mathbf{x}') \in (0, 2pn - \varepsilon n/8)$. We now show that if Calvin follows the random bit-flip strategy, from Bob's perspective (w.h.p.), both $\mathbf{x}$ or $\mathbf{x}'$ were equally likely to have been transmitted by Alice.

We first show that during Calvin's random bit-flip process, w.h.p., Calvin does not "run out" of his budget of pn bit flips.

Claim 4.3: Conditioned on Claim 4.2, with probability at least $1 - 2^{-\Omega(\varepsilon^2 n)}$

$$d_H(\mathbf{x}, \mathbf{y}) \in \left(\frac{d}{2} - \frac{\varepsilon n}{16}, \frac{d}{2} + \frac{\varepsilon n}{16} \right). \quad (2)$$

Proof: The expected number of locations flipped by Calvin is $d/2 \leq pn - \varepsilon n/16$. Assume that $d/2 = pn - \varepsilon n/16$ (for smaller values of d the bound is only tighter). By Sanov's theorem [4, Theorem 12.4.1], the probability that the number of bits flipped by Calvin deviates from the expectation $d/2$ by more than $\varepsilon n/16$ is at most $e^{-\Omega(\varepsilon^2 n^2/d)} \leq e^{-\Omega(\varepsilon^2 n)}$ for large enough n. ■

It should be noted that $d/2 + \varepsilon n/16 \leq pn$, and so $d_H(\mathbf{x}, \mathbf{y}) \leq d/2 + \varepsilon n/16$ implies that the number of bits flipped by Calvin does not exceed pn. Since Calvin possibly flips only the bits of $\mathbf{x}$ which differ from the corresponding bits in $\mathbf{x}'$, (2) also implies

$$d_H(\mathbf{x}', \mathbf{y}) \in \left(\frac{d}{2} - \frac{\varepsilon n}{16}, \frac{d}{2} + \frac{\varepsilon n}{16} \right). \quad (3)$$

We conclude by proving that if the number of bits flipped by Calvin lies in the range $(d/2 - \varepsilon n/16, d/2 + \varepsilon n/16)$, then indeed Bob cannot distinguish between the case in which $\mathbf{x}$ or $\mathbf{x}'$ were transmitted.

Claim 4.4: Conditioned on Claim 4.3 Bob makes a decoding error with probability at least $1/2$.

Proof: By Bayes' Theorem [8], if Bob receives $\mathbf{y}$, the *a posteri* probability that Alice transmitted $\mathbf{x}$, denoted $p(\mathbf{x}|\mathbf{y})$, equals $p(\mathbf{y}|\mathbf{x})p(\mathbf{x})/p(\mathbf{y})$. Here $p(\mathbf{x})$ is the probability (over her encoding strategy) that Alice transmits $\mathbf{x}$, $p(\mathbf{y}|\mathbf{x})$ is the probability (over Calvin's random bit-flipping strategy) that Bob receives $\mathbf{y}$ given that Alice transmits $\mathbf{x}$, and $p(\mathbf{y})$ is the resulting probability that Bob receives $\mathbf{y}$. Similarly,

$p(\mathbf{x}'|\mathbf{y}) = p(\mathbf{y}|\mathbf{x}')p(\mathbf{x}')/p(\mathbf{y})$. Taking the ratio and noting that for deterministic codes $p(\mathbf{x}) = p(\mathbf{x}')$, we have

$$p(\mathbf{x}|\mathbf{y})/p(\mathbf{x}'|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})/p(\mathbf{y}|\mathbf{x}'). \quad (4)$$

Since Calvin's random bit-flip strategy involves him flipping bits of $\mathbf{x}$ (which are different from the corresponding bits of $\mathbf{x}'$) with probability $1/2$, for all $\mathbf{y}$ satisfying (2), the probabilities $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{y}|\mathbf{x}')$ are equal. This observation and (4) together imply $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x}'|\mathbf{y})$. Thus, Bob cannot distinguish whether $\mathbf{x}$ or $\mathbf{x}'$ were transmitted. Namely, on the pair of events in which Alice transmits $\mathbf{x}$ and Calvin chooses $\mathbf{x}'$ and in which Alice transmits $\mathbf{x}'$ and Calvin chooses $\mathbf{x}$, no matter which decoding process Bob uses, he will have an average decoding error of at least $1/2$. This suffices to prove our assertion. ■

Thus a decoding error happens if the conditions of Claims 4.1, 4.2, 4.3 and 4.4 are all satisfied. This happens with probability at least $\left(1 - 2^{-\varepsilon n/4}\right)\left(\frac{\varepsilon}{64p}\right)\left(1 - 2^{-\Omega(\varepsilon^2 n)}\right)\left(\frac{1}{2}\right) \geq \left(\frac{1}{2}\right)\left(\frac{\varepsilon}{64p}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) \geq \frac{\varepsilon}{512p}$ for large enough n. ■

V. PROOF OF THEOREM 2

We start by proving the following technical Lemma that we use in our proof. Let q be an arbitrary probability distribution over an index set $I = \{1, \ldots, k\}$. Let $\mathbf{A_1}, \ldots, \mathbf{A_k}$ be arbitrary discrete random variables with probability distributions $q_1, \ldots, q_k$ over alphabets $\mathcal{A}_1, \ldots, \mathcal{A}_k$ respectively. Let $k_i = |\mathcal{A}_i|$. Let $\mathbf{A}$ be a random variable that equals the random variable $\mathbf{A_i}$ with probability $q(i)$. Then the following Lemma describing an elementary property of the entropy function $H(.)$ is useful in the proof of Theorem 2.

Lemma 5.1: The entropies of $\mathbf{A}, \mathbf{A_1}, \ldots, \mathbf{A_k}$ and q satisfy $H(\mathbf{A}) \leq \sum_{i=1}^{k} q(i)H(\mathbf{A_i}) + H(q)$, with equality if and only if for each i, i' for which both $q(i)$ and $q(i')$ are positive it holds that $\Pr_{q_i, q_{i'}}[\mathbf{A_i} = \mathbf{A_{i'}}] = 0$.

Proof: For any $a \in \mathcal{A}$, the probability $\Pr\{\mathbf{A} = a\} = p(a)$ of occurrence of a, equals $\sum_{i:a \in \mathcal{A}_i} q(i)q_i(a)$. Hence

$$\begin{aligned} H(\mathbf{A}) &= -\sum_{a \in \bigcup_i \mathcal{A}_i} p(a) \log(p(a)) \\ &\leq -\sum_{i=1}^{k} \sum_{j=1}^{k_i} q(i)q_i(j) \log(q(i)q_i(j)) \quad (5) \\ &= \sum_{i=1}^{k} \sum_{j=1}^{k_i} q(i) \left(q_i(j) \log(q_i(j)) \right) \\ &\quad + \sum_{i=1}^{k} \sum_{j=1}^{k_i} q_i(j) \left(q(i) \log(q(i)) \right) \\ &= \sum_{i=1}^{k} q(i)H(\mathbf{A_i}) + H(q). \end{aligned}$$

Here (5) follows from Jensen's inequality, e.g. [4], with equality if and only if for each positive $\Pr\{\mathbf{A} = a\}$, there is a unique i such that $q(i)q_i(j) > 0$ (here $a_i(j) = a$). ■

We now turn to prove Theorem 2. Recall our notation: let $\mathbf{U}$ be the random variable corresponding to Alice's message and p_U its distribution (with entropy Rn). Throughout we assume the message set $\mathcal{U}$ (the support of $\mathbf{U}$) is at most of size 2^n. Let $\mathcal{X}$ be Alice's codebook. $\mathcal{X}$ is a collection $\{\mathcal{X}(u)\}$ of subsets of $\{0,1\}^n$. For each subset $\mathcal{X}(u) \subset \mathcal{X}$, there is a corresponding codeword random variable $\mathbf{X}(u)$ with codeword distribution $p_{X(u)}$ over $\mathcal{X}(u)$. For any value $\mathbf{U} = u$ of the message, Alice's encoder choses a codeword from $\mathcal{X}(u)$ randomly from the distribution $p_{X(u)}$. Alice's message distribution p_U, codebook $\mathcal{X}$, and all the codebook distributions $p_{X(u)}$ are all known to both Bob and Calvin, but the values of the random variables $\mathbf{U}$ and $\mathbf{X}(.)$ are unknown to them. If $\mathcal{X}(u) = \{\mathbf{x}(u,r) : r \in \Lambda_u\}$, then the transmitted codeword $\mathbf{X}(\mathbf{U})$ has the probability distribution given by $\Pr[\mathbf{X}(\mathbf{U}) = \mathbf{x}(u,r)] = p_U(u)p_{X(u)}(\mathbf{x}(u,r))$. Let p the the overall distribution of codewords $\mathbf{x} = \mathbf{x}(u,r)$ of Alice. It holds that $p(\mathbf{x}(u,r)) = p_U(u)p_{X(u)}(\mathbf{x})$ and $p(\mathbf{x}) = \sum_{\mathcal{U}} p_U(u)p_{X(u)}(\mathbf{x})$.

For any $\varepsilon > 0$, let $R = (1 - 4p)^+ + \varepsilon$. We start by specifying Calvin's attack. Calvin uses a very similar attack to the one described in the proof of Theorem 1. That is, Calvin first *passively* waits until Alice transmits $\ell = (R - \varepsilon/2)n$ bits over the channel. Let $\mathbf{x}^\ell \in \{0,1\}^\ell$ be the value of the codeword observed so far. He then considers the set of codewords $\mathbf{x}(u,r)$ *consistent* with the observed $\mathbf{x}^\ell$. Here and throughout this section, we denote codewords by their corresponding message u and index r in $\mathcal{X}(u)$. As it may be that $\mathbf{x}(u,r)$ is exactly the same codeword as $\mathbf{x}(u',r')$, the sets in the definitions to follow and in this section are in a sense *multisets*. Namely, Calvin constructs the set $\mathcal{X}|_{\mathbf{x}^\ell} = \{\mathbf{x}(u,r) = x_1, \ldots, x_n \in \mathcal{X} \mid x_1, \ldots, x_\ell = \mathbf{x}^\ell\}$. Let $p(\mathbf{x}^\ell) = p(\mathcal{X}|_{\mathbf{x}^\ell})$ be the probability, under the probability distribution p, corresponding to the event that Calvin observes $\mathbf{x}^\ell$ in the first ℓ transmissions. Let $p_{U|_{\mathbf{x}^\ell}}$ and $p_{X(u)|_{\mathbf{x}^\ell}}$ be the probability distributions p_U and $p_{X(u)}$ also respectively conditioned on the same event. Calvin then chooses an element $\mathbf{x}'(u',r') \in \mathcal{X}|_{\mathbf{x}^\ell}$ with probability[3] $p_{U|_{\mathbf{x}^\ell}}(u')p_{X(u')|_{\mathbf{x}^\ell}}(\mathbf{x}'(u',r'))$. In the second stage he then follows exactly the same *random bit-flip strategy* as in the proof of Theorem 1.

Recall that in the proof of Theorem 1, our goal was to prove that with some constant probability, the distance between $\mathbf{x}(u,r)$ and $\mathbf{x}'(u',r')$ is approximately $2pn$. Loosely speaking, this allows the success of Calvin's attack (i.e., imply a decoding error). Following the same outline of proof, we now show that with probability $1/\text{poly}(n)$ the codeword $\mathbf{x}'(u',r')$ chosen by Calvin has the following three properties:

- It's corresponding message differs from that corresponding to $\mathbf{x}(u,r)$ (i.e., $u \neq u'$).
- $\mathbf{x}'(u',r')$ is *close* to $\mathbf{x}(u,r)$ and thus Calvin will be able

[3]This is one significant difference from the attack in the proof of Theorem 1 – there Calvin chooses each $\mathbf{x}'$ uniformly at random from the corresponding consistent set.

to "push" $\mathbf{x}(u,r)$ to a codeword $\mathbf{y}$ at approximately the same distance from $\mathbf{x}(u,r)$ and $\mathbf{x}'(u',r')$.
- Given $\mathbf{y}$, Bob is unable to distinguish whether $\mathbf{x}(u,r)$ or $\mathbf{x}'(u',r')$ was transmitted.

To this end, we partition the set $\mathcal{X}|_{\mathbf{x}^\ell}$ into n^2 disjoint subsets $\mathcal{X}_{ij}$ for $i,j \in \{1,2,\ldots,n\}$. Let $p(\mathcal{X}_{ij})$ be the probability mass of $\mathcal{X}_{ij}$. Let $p_{U|_{ij}}$ and $p_{X(u)|_{ij}}$ be the probability distributions p_U and $p_{X(u)}$ respectively conditioned on the event that Alice transmitted $\mathbf{x}(u,r)$ in $\mathcal{X}_{ij}$. The partition $\mathcal{X}_{ij}$ is obtained in two steps – first we partition $\mathcal{X}|_{\mathbf{x}^\ell}$ into n subsets $\mathcal{X}_i$, then we partition each $\mathcal{X}_i$ into n sets $\mathcal{X}_{ij}$. We also use the probability distribution $p(\mathcal{X}_i)$, $p_{U|_i}$ and $p_{X(u)|_i}$ defined accordingly. All in all, we prove the existence of a subset $\mathcal{X}_{ij}$ with the following properties

- $H(p_{U|_{ij}})$ is "large".
- $p(\mathcal{X}_{ij})$ is large with respect to $p(\mathbf{x}^\ell)$.
- For any $\mathbf{x}(u,r) \in \mathcal{X}_{ij}$ it holds that $p(\mathbf{x}(u,r))$ has approximately the same value.
- $p_{U|_{ij}}$ is approximately uniform on its support.

Roughly speaking, proving these properties on $\mathcal{X}_{ij}$ reduces us to the case of a deterministic encoder (addressed in Theorem 1) and allows us to complete our proof.

We now present our proof for the existence of $\mathcal{X}_{ij}$ as specified above. We first show that with positive probability the set $\mathcal{X}|_{\mathbf{x}^\ell}$ has *high* entropy.

Claim 5.1: With probability at least $\varepsilon/4$, $H(p_{U|_{\mathbf{x}^\ell}}) \geq \varepsilon n/4$.

Proof: Let q be the probability distribution over $\{0,1\}^\ell$ for which $q(\mathbf{x}^\ell) = p(\mathbf{x}^\ell)$ for all possible $\mathbf{x}^\ell \in \{0,1\}^\ell$. Let $q_{\mathbf{x}^\ell}$ be the probability distribution $p_{U|_{\mathbf{x}^\ell}}$. Now using Lemma 5.1 we obtain

$$H(p_U) \leq \sum_{\mathbf{x}^\ell} q(\mathbf{x}^\ell)H(p_{U|_{\mathbf{x}^\ell}}) + H(q). \tag{6}$$

By our definitions $H(p_U) = Rn$. Moreover, $H(q) \leq \ell = (R - \varepsilon/2)n$ (since q is defined over an alphabet of size 2^ℓ). Thus (6) becomes

$$\sum_{\mathbf{x}^\ell} q(\mathbf{x}^\ell)H(p_{U|_{\mathbf{x}^\ell}}) \geq Rn - (R - \varepsilon/2)n = \varepsilon n/2.$$

As the average of $H(p_{U|_{\mathbf{x}^\ell}})$ is at least $\varepsilon n/2$, then $H(p_{U|_{\mathbf{x}^\ell}}) \geq \varepsilon n/4$ with probability at least $\varepsilon/4$ (by a Markov type inequality, here we use the fact that $H(p_{U|_{\mathbf{x}^\ell}}) \leq n$). ∎

We now define the sets $\mathcal{X}_i$. For $i = 1, \ldots, n-1$, let $\mathcal{X}_i$ be the set of codewords in $\mathcal{X}|_{\mathbf{x}^\ell}$ for which $p(\mathbf{x}(u,r))/p(\mathbf{x}^\ell)$ is in the range $(2^{-3i}, 2^{-3i+3}]$. The set $\mathcal{X}_n$ is defined to be the set of codewords in $\mathcal{X}|_{\mathbf{x}^\ell}$ for which $p(\mathbf{x}(u,r))/p(\mathbf{x}^\ell)$ is in the range $[0, 2^{-3n+3}]$. Let $p(\mathcal{X}_i)$ be the probability mass of $\mathcal{X}_i$. Namely $p(\mathcal{X}_i) \simeq 2^{-3i}|\mathcal{X}_i|p(\mathbf{x}^\ell)$. Let q be the distribution over $\{1,2,\ldots,n\}$ taking i w.p. $p(\mathcal{X}_i)/p(\mathbf{x}^\ell)$. Notice that $H(q) \leq \log(n) = o(n)$ (as its support is of size n). Conditioning on Claim 5.1 and using Lemma 5.1 it can be verified that

Claim 5.2:

$$\sum_i q(i)H(p_{U|_i}) \geq H(p_{U|_{\mathbf{x}^\ell}}) - H(q) \geq \varepsilon n/8 \tag{7}$$

Consider sets $\mathcal{X}_i$ with (relative) mass $q(i) \geq 1/n^2$. It holds that

$$\sum_{i \leq n-1; q(i) \geq 1/n^2} q(i) H(p_{U|_i}) \geq \varepsilon n/16$$

The above follows from the fact that $\sum_{i \leq n-1; q(i) \leq 1/n^2} q(i) H(p_{U|_i}) + q(n) H(p_{U|_i}) \leq \sum_{i \leq n-1; q(i) \leq 1/n^2} n/n^2 + 2^{-n+3} n \leq 2$ (for sufficiently large n). Here we use the fact that $q(n) \leq |\mathcal{X}_i| 2^{-3n+3}$.

We conclude the existence of a set $\mathcal{X}_i$ such that $q(i) \geq 1/n^2$ and $H(p_{U|_i}) \geq \varepsilon n/16$. We now further partition $\mathcal{X}_i$. For $j = 1, \ldots, n-1$, let $\mathcal{X}_{ij}$ be the set of codewords $\mathbf{x}(u, r)$ in $\mathcal{X}_i$ for which $p_{U|_i}(u)$ is in the range $(2^{-3j}, 2^{-3j+3}]$. $\mathcal{X}_{in}$ is defined to be the set of codewords $\mathbf{x}(u, r)$ in $\mathcal{X}_i$ for which $p_{U|_i}(u)$ is in the range $[0, 2^{-3n+3}]$. Let $p(\mathcal{X}_{ij})$ be the probability mass of $\mathcal{X}_{ij}$. Namely $p(\mathcal{X}_{ij}) \simeq 2^{-3i} |\mathcal{X}_{ij}| p(\mathbf{x}^\ell)$. Let q' be the distribution over $\{1, 2, \ldots, n\}$ taking j w.p. $p(\mathcal{X}_{ij})/p(\mathcal{X}_i)$. Notice that $H(q') \leq \log(n) = o(n)$ (as its support is of size n). As before, conditioning on Claim 5.2 and using Lemma 5.1 it can be verified that (for the index i specified above),

Claim 5.3:

$$\sum_j q'(j) H(p_{U|_{ij}}) \geq H(p_{U|_i}) - H(q') \geq \varepsilon n/32 \quad (8)$$

Again, consider sets $\mathcal{X}_{ij}$ with mass $q'(i) \geq 1/n^2$. It holds that

$$\sum_{j \leq n-1; q'(j) \geq 1/n^2} q'(j) H(p_{U|_{ij}}) \geq \varepsilon n/64$$

We conclude the existence of a set $\mathcal{X}_{ij}$ such that

- $H(p_{U|_{ij}}) \geq \varepsilon n/64$.
- $p(\mathcal{X}_{ij}) \geq p(\mathbf{x}^\ell)/n^4$.
- For any $\mathbf{x}(u, r) \in \mathcal{X}_{ij}$ it holds that $p(\mathbf{x}(u, r))$ is approximately $2^{-3i} p(\mathbf{x}^\ell)$.
- For any $\mathbf{x}(u, r) \in \mathcal{X}_{ij}$ it holds that $p_{U|_{ij}}(u)$ is approximately equal.

The set $\mathcal{X}_{ij}$ is exactly what we are looking for. Roughly speaking, by Claim 5.1, with probability at least $\varepsilon/4$ Calvin views a prefix $\mathbf{x}^\ell$ for which $H(p_{U|_{\mathbf{x}^\ell}}) \geq \varepsilon n/4$. Conditioning on this event, both Alice and Calvin choose codewords $\mathbf{x}(u, r), \mathbf{x}'(u', r')$ in $\mathcal{X}_{ij}$ with probability at least $1/n^8$.

We now sketch to remainder of the proof which closely follows that of Theorem 1. We partition $\mathcal{X}_{ij}$ into *groups* of messages $\mathcal{X}_{ij}(u)$ consisting of all codewords in $\mathcal{X}_{ij}$ corresponding to u. Recall that each codeword $\mathbf{x}(u, r) \in \mathcal{X}_{ij}$ has approximately the same probability $p(\mathbf{x}(u, r))$, and for each $\mathbf{x}(u, r) \in \mathcal{X}_{ij}$ it holds that $p_{U|_{ij}}(u)$ is approximately the same value. This implies that each group $\mathcal{X}_{ij}(u) \subseteq \mathcal{X}_{ij}$ has approximately the same size. Moreover, as $H(p_{U|_{ij}}) \geq \varepsilon n/64$ it holds that there are at least $2^{\varepsilon n/64}$ non-empty subsets $\mathcal{X}_{ij}(u)$ in $\mathcal{X}_{ij}$.

So, all in all, $\mathcal{X}_{ij}$ has a very symmetric structure: it includes *many* groups, each consisting of elements with the same transmission probability, and each of approximately the same size and mass (w.r.t. p). This reduces us to the case considered in Theorem 1 in which our subset $\mathcal{X}|_{\mathbf{x}^\ell}$

included many messages, each with the same probability, details follow.

Consider the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in which the vertex set $\mathcal{V}$ consists of the set $\mathcal{X}_{ij}$ and two nodes are connected by an edge if their Hamming distance is less than $d = 2pn - \varepsilon n/8$.

Now, it is can be verified (using analysis almost identical to that given in the proof of Theorem 1) that

1) With probability at least $1 - 2^{-\Omega(\varepsilon n)}$ the codewords $\mathbf{x}(u, r)$ and $\mathbf{x}'(u', r')$ satisfy $u \neq u'$. Here one needs to take into consideration the slight difference in the group sizes and the probabilities for each codeword.

2) With probability $\Omega\left(\frac{\varepsilon}{p}\right)$ the vertices in $\mathcal{G}$ corresponding to $\mathbf{x}(u, r)$ and $\mathbf{x}'(u', r')$ are connected by an edge.

3) During Calvin's random bit-flip process, with high probability of $1 - 2^{-\Omega(\varepsilon^2 n)}$, Calvin does not "run out" of his budget of pn bit flips.

4) Conditioning on the above, Bob cannot distinguish between the case in which $\mathbf{x}(u, r)$ or $\mathbf{x}'(u', r')$ were transmitted.

5) Finally, on the pair of events in which Alice transmits $\mathbf{x}(u, r)$ and Calvin chooses $\mathbf{x}'(u', r')$, and Alice transmits $\mathbf{x}'(u', r')$ and Calvin chooses $\mathbf{x}(u, r)$, no matter which decoding process Bob uses, he has an average decoding error that is bounded away from zero. Here again we take into account the slight differences between $p(\mathbf{x}(u, r))$ and $p(\mathbf{x}'(u', r'))$.

To summarize, Calvin causes a decoding error with probability $\Omega(\mathrm{poly}(\varepsilon)/\mathrm{poly}(n)) = \Omega(1/\mathrm{poly}(n))$ as desired. This concludes our proof. ∎

VI. CONCLUSIONS

We analyze the capacity of the causal-adversarial channel and show (for both deterministic and probabilistic encoders) that the capacity is bounded by above by $\min\{1 - H(p), (1 - 4p)^+\}$. For a large range of p (for all $p > 0.25$), the maximum achievable rate equals that of the *stronger* classical "omniscient" adversarial model (*i.e.*, 0).

Several questions remain open. In this work we do not address achievability results (i.e., the construction of codes). It would be very interesting to obtain codes for the causal-adversary channel which obtain rate greater than that know for the "omniscient" adversarial model (*i.e.*, the Gilbert-Varshamov bound) for $p < 0.25$. As we do not believe that the upper bound of $(1 - 4p)^+$ presented in this work is actually tight, such codes, if they exist, may give a hint to the correct capacity.

As done in our work on large alphabets [6], one may also consider the more general channel model in which for a *delay* parameter $d \in (0, 1)$, the jammer's decision on the corruption of x_i must depend solely on x_j for $j \leq i - dn$. This might correspond to the scenario in which the error transmission of the adversarial jammer is delayed due to certain computational tasks that the adversary needs to perform. The capacity of the causal channel with delay is an intriguing problem left open in this work.

REFERENCES

[1] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacities of certain channel classes under random coding. *The Annals of Mathematical Statistics*, 31(3):558–567, 1960.

[2] A. E. Brouwer. Bounds on the size of linear codes. In V. S. Pless and W. C. Huffman, editors, *Handbook of Coding Theory*, volume 1, chapter 4, pages 295–461. Elsevier Science, New York, NY, USA, 1998.

[3] H. Chernoff. Measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–507, 1952.

[4] T. M. Cover and J. A. Thomas. *Elements of information theory, 2nd edition*. Wiley-Interscience, New York, NY, USA, 2006.

[5] I. Csiszar and J. G. Korner. *Information Theory: coding theorems for discrete memoryless systems*. Academic Press, Inc, Orlando, FL, USA, 1982.

[6] B. K. Dey, S. Jaggi, and M. Langberg. Codes against online adversaries. *Manuscript*. Available at http://arxiv.org/abs/0811.2850.

[7] P. Elias. Error-correcting codes for list decoding. *IEEE Transactions on Information Theory*. 37(1):512, 1991.

[8] W. Feller. *An Introduction to Probability Theory and Its Applications, Volume II (2nd ed.)*. John Wiley & Sons, New York, 1972.

[9] R. G. Gallager. *Information Theory and Reliable Communication*. J. Wiley and Sons, New York, 1968.

[10] E. N. Gilbert. A comparison of signalling alphabets. *Bell Systems Technical Journal*, 31:504–522, 1952.

[11] S. Jaggi, M. Langberg, T. Ho, and M. Effros. Correction of Adversarial Errors in Networks. In *proceedings of IEEE International Symposium on Information Theory (ISIT)*, pages 1455–1459, 2005.

[12] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch. New upper bounds on the rate of a code via the delsarte-macwilliams inequalities. *IEEE Trans. Inform. Theory*, 23(2):157–166, March 1977.

[13] L. Nutman and M. Langberg. Adversarial Models and Resilient Schemes for Network Coding. *In proceedings of IEEE International Symposium on Information Theory*, pages 171–175, 2008.

[14] M. Plotkin. Binary codes with specified minimum distance. *IRE Trans. Inform. Theory*, 6:445–450, 1960.

[15] A. Sahai and S. Mitter. The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link, Part I: scalar systems. *IEEE Transactions on Information Theory*, 52(8):3369–3395, 2006.

[16] A. Sarwate. Robust and adaptive communication under uncertain interference. *PhD thesis, Berkeley*, 2008.

[17] P. Turán. On the Theory of Graphs., Colloq . Math . 3 (1954), 19-30. *Colloq. Math.* 3:19-30, 1954.

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