

Channel Capacity for Adversaries with Computationally Bounded Observations

Eric Ruzomberka, Chih-Chun Wang and David J. Love

Abstract—We study reliable communication over point-to-point adversarial channels in which the adversary can observe the transmitted codeword via some function that takes the n -bit codeword as input and computes an rn -bit output for some given $r \in [0, 1]$. We consider the scenario where the rn -bit observation is *computationally bounded* – the adversary is free to choose an arbitrary observation function as long as the function can be computed using a polynomial amount of computational resources. This observation-based restriction differs from conventional channel-based computational limitations, where in the later case, the resource limitation applies to the computation of the (adversarial) channel error/corruption. For all $r \in [0, 1 - H(p)]$ where $H(\cdot)$ is the binary entropy function and p is the adversary’s error budget, we characterize the capacity of the above channel and find that the capacity is identical to the completely oblivious setting ($r = 0$). This result can be viewed as a generalization of known results on myopic adversaries and on channels with active eavesdroppers for which the observation process depends on a fixed distribution and fixed-linear structure, respectively, that cannot be chosen arbitrarily by the adversary.

Index Terms—Adversarial channels, capacity, arbitrarily varying channels

I. INTRODUCTION

Beginning with Shannon’s seminal paper [2], early channel coding research observed that fundamental coding limits are highly sensitive to channel modeling assumptions. This sensitivity is demonstrated by a gap in capacity between the two classical models: the *Shannon model* in which channel errors are random and follow a known distribution and the *Hamming model* in which error patterns are worst-case for some fixed number of bit errors. In the design of robust codes, the more conservative Hamming model is particularly attractive as it makes no assumptions about the channel distribution and thus any resulting conclusion is *robust* against a wide variety of channel imperfections. The downside of the Hamming model, however, is that it admits a smaller capacity than the Shannon model. In many cases, the gap in capacity is large [3].

This work was supported in part by the Office of Naval Research under ONR Grant N00014-21-1-2472, by NSF Grants CCF-1618475, CCF-1816013, CCF-2008527, CNS-2107363, CIF-2309887 and also by National Spectrum Consortium (NSC) under grant W15QKN-15-9-1004.

E. Ruzomberka is with the Department of Electrical and Computer Engineering, Princeton University, USA (email: er6214@princeton.edu). This work was done while E. Ruzomberka was with the Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, USA. C.-C. Wang and D. J. Love are with the Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, USA (email: chihw,djlove@purdue.edu). A preliminary version of the work was presented at the 2022 IEEE International Symposium on Information Theory [1].

A. Closing the gap

Recent research efforts have made progress in closing this gap by considering settings in between the two classical models. Ideally, the following two properties hold for a good channel model:

Property 1: The channel is *mild* in the sense that its capacity coincides with the Shannon model capacity.

Property 2: The channel inherits conservative aspects of the Hamming model. In particular, the channel may be altered in an arbitrary manner unknown to the communicating parties.

In the following Section I-B, we focus on two different approaches which have had some success towards producing good channel models. These approaches are 1) to bound the channel’s computing power (i.e., computational complexity) [4], [5] and 2) to bound the information known to the channel about the communication scheme [6]–[14].

B. Complexity Bounded Models vs Partially Oblivious Models

Consider a transmitter Alice who wishes to communicate a message m_0 drawn randomly from a set of M possible messages over a noisy channel to a receiver Bob. To protect the message from noise corruption, Alice encodes m_0 into an n -bit codeword \mathbf{x} of rate $R = (1/n) \log M$ and transmits \mathbf{x} over the channel. The channel adds an n -bit error vector \mathbf{e} to \mathbf{x} , and Bob receives the n -bit word $\mathbf{y} = \mathbf{x} \oplus \mathbf{e}$. The channel is controlled by an *adversary* who chooses \mathbf{e} to prevent reliable (unique) decoding by Bob. For an error budget $p \in (0, 1/2)$, the adversary can induce at most pn bit flips, i.e., the Hamming weight of \mathbf{e} must be bounded above by pn . We focus on *deterministic codes* in which the codeword \mathbf{x} is a deterministic function of the message m_0 , and in turn, consider the *average error criterion* in which decoding is permitted to fail over an arbitrarily small fraction of Alice’s messages.¹

We define the Shannon model capacity as $C_{\text{Shannon}(p)} = 1 - H(p)$ for $p \in [0, 1/2]$ where $H(\cdot)$ is the binary entropy function, which coincides with the capacity of a binary symmetric channel with crossover probability $p \in [0, 1/2]$. In general, $C_{\text{Shannon}(p)}$ is an upper bound of any rate achievable by any communication scheme used by Alice and Bob, but

¹Alternatively, one may consider *stochastic codes* in which \mathbf{x} is a function of both m_0 and a private random key known only to Alice. Note that a deterministic code is a degenerate stochastic code where the set of private random keys is empty. Compared to the average error criterion, a stronger decoding criterion which is of interest but not considered here is the *maximum error criterion* in which decoding is permitted to fail for an arbitrarily small fraction of Alice’s keys.

may be tight depending on additional assumptions made about the adversary’s capabilities and limitations. A surprising result of Csiszár and Narayan [13] is that $C_{\text{Shannon}}(p)$ is the channel capacity when the adversary must choose error vector e without knowledge of the codeword x or message m_0 .

In the *computationally bounded model* (first proposed by Lipton [4]), the adversary takes x as input and computes e using limited computational resources, e.g., via an algorithm that takes a bounded number of computational steps. This model has the appeal of sufficiently describing practical channels, including channels with memory and channels governed by natural, but unknown processes. However, the computationally bounded model can be *severe* – an impossibility result of Guruswami and Smith [5] is that the model’s capacity can be less than the Shannon capacity, and can even be 0 when the latter is positive. Thus, Property 1 does not hold for the computationally bounded model.²

Another existing approach is the *partially oblivious model*, where the adversary chooses e based on incomplete side-information about the transmitted codeword x . This model includes myopic channels, e.g., [6]–[8], causal channels, e.g., [9]–[11], channels with active eavesdroppers, e.g., [19], and some arbitrarily varying channels (AVCs), e.g., [12], [13]. We focus on the following setting which captures a special case of the partially oblivious model: for $r \in [0, 1]$ and some observation function $f_n : \{0, 1\}^n \rightarrow \{0, 1\}^{rn}$, the adversary makes an rn -bit observation $f_n(x)$ of codeword x , and in turn, chooses e . We emphasize that the error vector can depend non-causally on the rn -bit observation and, thus, causal channels are not captured by our setting. The special cases $r = 0$ and $r = 1$ correspond to no information (i.e., completely oblivious) and perfect information (i.e., omniscient), respectively.

Property 1 can hold for the partially oblivious model when r is sufficiently small.³ However, Property 2 does not hold for many partially oblivious channels in the literature. While all partially oblivious channels allow error vector e to be chosen in an arbitrarily manner unknown to Alice and Bob, the observation function f_n is not always chosen arbitrarily. For example, a myopic channel in our setting corresponds to f_n being drawn randomly from some distribution known to Alice and Bob. *For Property 2 to hold, however, we must allow f_n to be chosen arbitrarily and require Alice and Bob to devise their communication scheme without knowledge of f_n .* This is equivalent to the adversary choosing a *worst-case* f_n for a fixed r , a model defined and studied by Langberg [14] under

the name of the $(1 - r)$ -*oblivious channel*.⁴ The capacity of the $(1 - r)$ -oblivious channel remains an open problem, where the best known lower bound is given by [14].

C. This Work

In this paper, we define and study a channel model that has qualities of both the computationally bounded model and the partially oblivious model. Roughly speaking, we define this model by requiring the adversary to observe x via an rn -bit observation function f_n that is computationally bounded.

Specifically, for fixed positive integers c and s , the adversary chooses a sequence of observation functions $f_n(\cdot), \forall n \geq 1$ that belongs to $\text{CKT}(r, cn^s)$ – the set of observation functions with n input bits and rn output bits that can be computed by a Boolean circuit with at most cn^s gates. We allow the choice of f_n to be unknown to Alice or Bob. On the other hand, the f_n chosen by the adversary can depend on the codebook of Alice but cannot depend on the actual message being sent. Using the observation function f_n of its choice, the adversary observes $f_n(x)$ and chooses e with no computational bound. Our model differs from the prior works [4], [5], [15]–[18], where in the latter, the channel has a complete view of x but must choose e subject to a computational bound. We refer to our adversary as a $\text{CKT}(r, cn^s)$ -*oblivious adversary*. By construction, Property 2 holds for a channel controlled by a $\text{CKT}(r, cn^s)$ -oblivious adversary due to f_n being unknown to Alice or Bob.

Our computational restriction is modeled after realistic adversarial channels. A channel controlled by a $\text{CKT}(r, cn^s)$ -oblivious adversary closely approximates a $(1 - r)$ -*oblivious channel* [14] (i.e., the definition therein is equivalent to the $\text{CKT}(r, \infty)$ -oblivious adversary) without weakening the power of the adversary too much. Indeed, the adversary is quite strong. To illustrate its strength, if for a sequence of functions $\{f_n\}_{n=1}^{\infty}$ satisfying $\forall c, s \geq 1$ there exists a finite n_0 such that for all $n \geq n_0$ $f_n \notin \text{CKT}(r, cn^s)$, then the sequence is widely regarded to be an *infeasible computation* [20]. The technical value of the computational constraint is to bound the number of observation functions that the adversary can choose from while still including a wide range of important observation functions in the problem formulation.

In this paper, for any fixed finite integers c, s , and all $p \in (0, 1/2)$ and $r \in [0, 1 - H(p))$, we study the channel controlled by a $\text{CKT}(r, cn^s)$ -oblivious adversary with error budget p by characterizing the channel capacity $C(p, r, c, s)$. As our main result, we show that $C(p, r, c, s)$ is exactly $1 - H(p)$, and thus the capacity is independent of parameters c, s, r for the stated parameter regime. It follows that $C(r, p, c, s)$ coincides with the Shannon model capacity $C_{\text{Shannon}}(p)$ and thus Property 1 holds. Furthermore, in this regime, deterministic codes are op-

²Specifically, a channel which uses logarithmic space to process the codeword x has a capacity of 0 when $p \geq 1/4$. In light of this impossibility result, recent studies on the computationally bounded model study either unique decoding when $p \in (0, 1/4)$ [15], [16] or relax the objective of unique decoding and instead consider list-decoding when $p \in (0, 1/2)$ [17], [18]. The works [5], [15]–[18] employ stochastic codes together with pseudorandom sequences to complicate the channels task of computing an effective error pattern e . In contrast to the above works, we consider deterministic codes and unique decoding for all $p \in (0, 1/2)$.

³This fact is an analog to a channel being *sufficiently myopic* (see [7]).

⁴An alternative interpretation of Property 2 is that the adversary may choose a worst-case f_n from some *subset of functions* from $\{0, 1\}^n$ to $\{0, 1\}^{rn}$, and where the subset is known to Alice and Bob. In the $(1 - r)$ -oblivious channel, this subset is the improper subset of all functions. Another channel that satisfies Property 2 under this alternative interpretation is the adversarial wiretap channel of type II [19], in which f_n is chosen from the set of all linear mappings from $\{0, 1\}^n$ to $\{0, 1\}^{rn}$. Depending on the specific application for which the channel model serves, it may be unrealistic to assume that this subset contains only linear mappings.

timal.⁵ This main result was first presented at the International Symposium on Information Theory (ISIT) [1].

The remainder of this paper is organized as follows. In Section II, we present the precise channel model and main result. The main result is discussed in the context of related work on myopic channels, channels controlled by active eavesdroppers, and $(1-r)$ -oblivious channels. In Section III, we present the overview of the proof of the main result and discuss our proof techniques in the context of related work. In Section IV, we present the detailed proof of the main result.

II. CHANNEL MODEL & RESULTS

A. Notation

All vectors are in bold notation. Let $d(\mathbf{z}, \mathbf{z}')$ denote the Hamming distance between two binary vectors \mathbf{z} and \mathbf{z}' . For $t > 0$ and $\mathbf{z} \in \{0, 1\}^n$ we define $\mathcal{B}_t(\mathbf{z}) = \{\mathbf{z}' \in \{0, 1\}^n : d(\mathbf{z}, \mathbf{z}') \leq t\}$ to be the Hamming ball of radius t centered around \mathbf{z} . The functions $\log(\cdot)$ and $\ln(\cdot)$ denote the base 2 and base e logarithms, respectively. For a number $K \geq 1$, let $[K]$ denote the set $\{1, \dots, [K]\}$. For an integer blocklength $n \geq 1$ and rate $R \in (0, 1]$, an (n, Rn) codebook \mathcal{C}_n is a function $\mathcal{C}_n : [2^{Rn}] \rightarrow \{0, 1\}^n$. When useful, we will think of $\mathcal{C}_n = \{\mathcal{C}_n(1), \dots, \mathcal{C}_n(2^{Rn})\}$ as a subset of $\{0, 1\}^n$ and the i th codeword $\mathcal{C}_n(i)$ as a vector in $\{0, 1\}^n$. For a number $\rho > 0$ and a binary vector $\mathbf{a} = (a_1, \dots, a_{\rho n}) \in \{0, 1\}^{\rho n}$, we define the integer representation of \mathbf{a} to be the integer $\text{int}(\mathbf{a}) = 1 + \sum_{j=1}^{\rho n} a_j 2^{j-1} \in [2^{\rho n}]$. For functions $g(n)$ and $h(n)$, we adopt standard “little O”, “big O” and “big Omega” notation: $g = o(h(n))$ if $\lim_{n \rightarrow \infty} \frac{g(n)}{h(n)} = 0$, $g = O(h(n))$ if $\exists k$ s.t. for large enough n , $g(n) \leq kh(n)$, and $g = \Omega(h(n))$ if $\exists k$ s.t. for large enough n , $g(n) \geq kh(n)$.

B. Channel Model

Alice’s Encoding: A transmitter Alice communicates over a noisy channel with a receiver Bob in the following manner. For a rate $R \in (0, 1]$ and integer blocklength $n \geq 1$, Alice randomly draws a message m_0 uniformly from a message set $[2^{Rn}]$. For a (n, Rn) codebook \mathcal{C}_n , Alice encodes m_0 into a codeword $\mathbf{x} \in \{0, 1\}^n$ by computing $\mathbf{x} = \mathcal{C}_n(m_0)$. Since $\mathbf{x} = \mathcal{C}_n(m_0)$ is a deterministic function of m_0 , we say that Alice is using a deterministic code. After encoding, Alice transmits \mathbf{x} into the channel.

Bob’s Decoding: At the channel output, Bob receives word $\mathbf{y} = \mathbf{x} \oplus \mathbf{e}$ where $\mathbf{e} \in \{0, 1\}^n$ is an error vector added by the channel and where the symbol ‘ \oplus ’ denotes the bit-wise XOR.

⁵For the parameter regime $r \geq 1 - H(p)$ and $c \geq 1$, $s \geq 1$, deterministic codes may not be optimal. We remark that our proof techniques, which involve a random coding argument over a set of deterministic codes, only work for the regime $r < 1 - H(p)$. For $r \geq 1 - H(p)$, the channel to the adversary is “less noisy” than the channel to Bob, such that when a deterministic code is used at rate less than $1 - H(p)$, the adversary is likely to decode Alice’s codeword (with high probability over the code selection) and thus the adversary is effectively omniscient (i.e., $r = 1$). For omniscient adversaries, the GV bound of $1 - H(2p)$ [21], [22] is the best-known achievable rate. However, when a stochastic code is used, one may find achievable rates exceeding the GV bound for some values of $r \geq 1 - H(p)$ and $p \in (0, 1/2)$. In fact, in some channel models, stochastic codes are known to achieve rates significantly larger than the GV bound for certain parameters when the channel to the adversary is “less noisy” than the channel to Bob (see, e.g., [23]).

Bob outputs a message estimate \hat{m} based on the received word \mathbf{y} . We say that a decoding error occurs if $\hat{m} \neq m_0$.

Adversary: The channel is controlled by an adversary who has side-information about Alice’s and Bob’s communication scheme but not exact knowledge of the transmitted message m_0 . In particular, the adversary knows Alice’s codebook \mathcal{C}_n and is *partially oblivious* to the transmitted codeword \mathbf{x} . By partially oblivious, we mean that for observation rate $r \in [0, 1]$, the adversary randomly draws a function $f_n : \{0, 1\}^n \rightarrow \{0, 1\}^{rn}$ with probability $U_f(f_n)$ and observes a realization ψ of the random variable $\Psi = \Psi(m_0) = f_n(\mathcal{C}_n(m_0)) = f_n(\mathbf{x})$.⁶ Using its knowledge of \mathcal{C}_n but without knowledge of the realization of m_0 , the adversary randomly draws f_n with probability $U_f(f_n)$. Due to the adversary’s computational bound, for positive integers c, s , $U_f(f_n) = 0$ for all $f_n \notin \text{CKT}(r, cn^s)$ (we provide a rigorous definition of $\text{CKT}(r, cn^s)$ in Section II-C). Neither the actual choice of f_n nor the distribution $U_f(\cdot)$ is revealed to Alice or Bob. As a result, the model falls into the adversarial setting in which the adversary has full freedom of using any specific function (by choosing $U_f(\cdot)$ to be a delta distribution) or any random selection of functions (by choosing $U_f(\cdot)$ to be of general distribution).

Finally, using knowledge of the codebook \mathcal{C}_n and observation function f_n , the adversary chooses the conditional probability $U_{e|\psi}(e|\psi)$ that the error vector e is added to the channel given that it observes $\Psi(m_0) = \psi$. For $p \in (0, 1/2)$, we impose an error budget constraint such that e has a Hamming weight bounded above by pn , i.e., $U_{e|\psi}(e|\psi) = 0$ for all $e \notin \mathcal{B}_{pn}(0)$ and $\psi \in \{0, 1\}^{rn}$. We refer to the above adversary as the $\text{CKT}(r, cn^s)$ -oblivious adversary with error budget p . We note that the distribution $U_f(f_n)$ and $U_{e|\psi}(e|\psi)$ are used so that some randomness can be embedded in the adversary’s action. For simplicity, the reader may assume that the adversary chooses deterministically an observation function $f_n \in \text{CKT}(r, cn^s)$, uses the observation function to observe $\Psi = f_n(\mathcal{C}_n(m_0)) = \psi$, and then chooses deterministically an error vector e under the given error budget p .

C. Adversary’s Computational Bound

For observation rate $r \in [0, 1]$, positive integers c, s, n , we define the set $\text{CKT}(r, cn^s)$. Let $\mathcal{F}_{n,r}$ denote the set of all Boolean functions of the form $f_n : \{0, 1\}^n \rightarrow \{0, 1\}^{rn}$. To define $\text{CKT}(r, cn^s)$, we first define the circuit complexity of a function $f_n \in \mathcal{F}_{n,r}$.

A Boolean circuit B_n is an acyclic directed graph where each node is either an input node (with in-degree 0) or a logic gate (with in-degree 2). All nodes in B_n have out-degree 1 with unbounded fan-out and each logic gate computes an arbitrary Boolean function from $\{0, 1\}^2$ to $\{0, 1\}$. The *size* of B_n is the total number of nodes in B_n (input nodes and logic gates). Note that an observation function $f_n \in \mathcal{F}_{n,r}$ can be computed by some Boolean circuit that takes n bits as input and produces rn bits as output. The *circuit (size) complexity* of an observation function $f_n \in \mathcal{F}_{n,r}$ is the size

⁶The fact that Ψ is a random variable follows from its dependency on the random variable m_0 .

of the smallest size Boolean circuit B_n that can compute f_n . We define $\text{CKT}(r, cn^s)$ to be the set of all functions $f_n \in \mathcal{F}_{n,r}$ with a circuit complexity of at most cn^s . In modern complexity theory, the study of circuit complexity is a common approach for proving lower bounds on the complexity of certain problems [20].

D. Capacity

For an (n, Rn) codebook \mathcal{C}_n , the (average) probability of decoding error is defined as

$$\bar{P}_e(\mathcal{C}_n) = \max_{f_n \in \text{CKT}(r, cn^s)} \mathbb{E}_{\Psi} \left[\max_{e \in \mathcal{B}_{pn}(0)} \mathbb{P}_{m_0}(\hat{m}(e, m_0) \neq m_0 | \Psi = \psi) \right] \quad (1)$$

where the probability measure $\mathbb{P}_{m_0}(\cdot)$ is w.r.t. the distribution of m_0 , and the expectation $\mathbb{E}_{\Psi}[\cdot] = \sum_{\psi \in \{0,1\}^{rn}} (\cdot) \mathbb{P}_{m_0}(\Psi(m_0) = \psi)$. Given the above channel model, we can define achievable rate in the usual way.

Definition 1 (Achievable Rate). *For $p \in (0, 1/2)$, $r \in [0, 1]$, and positive integers c, s , a rate $R \in (0, 1]$ is said to be (c, s) -achievable if for any $\epsilon_e > 0$, there exists an n_0 such that for all $n \geq n_0$, there exists an (n, Rn) codebook \mathcal{C}_n such that $\bar{P}_e(\mathcal{C}_n) \leq \epsilon_e$.*

For $p \in (0, 1/2)$, $r \in [0, 1]$, and positive integers c, s , we define the capacity $C(p, r, c, s)$ of a channel controlled by a $\text{CKT}(r, cn^s)$ -oblivious adversary as the supremum of (c, s) -achievable rates. Let $C(p, r, \infty, \infty)$ denote the capacity of $(1-r)$ -oblivious channel for which there is no constraint on the computational complexity when computing the rn -bit observation, see [14].

E. Main Result

Under the above model, the Shannon capacity is $C_{\text{Shannon}}(p) = 1 - H(p)$ where $H(p) = -p \log p - (1-p) \log(1-p)$ is the binary entropy function [13], [14]. The following result shows that Property 1 holds for our model for a wide range of r .

Theorem 1. *For $p \in (0, 1/2)$, $r \in [0, C_{\text{Shannon}}(p))$, and $c, s \geq 1$, $C(p, r, c, s) = C(p, 0, c, s) = C(p, 0, \infty, \infty) = C_{\text{Shannon}}(p)$.*

We share a few remarks on the above theorem. When $r < C_{\text{Shannon}}(p) = 1 - H(p)$, Theorem 1 implies that the adversary can do no better than to *ignore* its side-information $f_n(\mathbf{x})$ and choose e randomly from the set of all n -bit vectors with Hamming weight pn . Additionally, we note that the largest known lower bound on $C(p, r, \infty, \infty)$ is $1 - H(p) - r$ for $r \in [0, \frac{1-H(p)}{3}]$ [14]. Since $C(p, r, \infty, \infty)$ is a lower bound to $C(p, r, c, s)$, Theorem 1 significantly sharpens the best known lower bound of $C(p, r, c, s)$ to an exactly tight

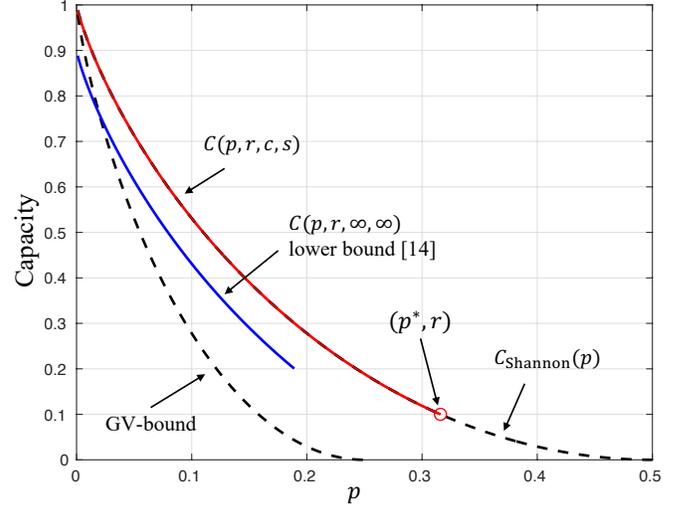


Fig. 1. Capacity when $r = 0.1$ and c, s are finite positive integers. Herein the value p^* satisfies $C_{\text{Shannon}}(p^*) = r = 0.1$.

characterization.⁷ For $r > C_{\text{Shannon}}(p)$, an immediate lower bound of $C(p, r, c, s)$ is given by the Gilbert-Varshamov (GV) bound (i.e. $C(p, r, \infty, \infty) \geq 1 - H(2p)$) [21], [22].⁸ All results discussed thus far are summarized in Fig. 1.

Theorem 1 generalizes a few results on myopic channels and on channels with active eavesdroppers. For $p \in [0, 1/2]$ and $r < C_{\text{Shannon}}(p)$, $C_{\text{Shannon}}(p)$ is known to be the capacity of the *binary-erasure bit-flip myopic channel* where the adversary a) non-causally views \mathbf{x} through a binary erasure channel with erasure probability $1 - r$ (denoted as $\text{BEC}(1 - r)$ in the literature) then b) injects pn bit errors [7, Theorem III.12]. It is clear that this result is generalized by Theorem 1 after observing that a $\text{CKT}(r, cn^s)$ -oblivious adversary can choose f_n randomly in a way that simulates a $\text{BEC}(1 - r)$. Similarly, for $r < C_{\text{Shannon}}(p)$, $C_{\text{Shannon}}(p)$ is known to be the capacity of the *adversarial wiretap channel of type II* where the adversary a) chooses rn indices in $\{1, \dots, n\}$ and observes rn -bits of \mathbf{x} at the chosen indices then b) injects pn bit errors [19, Theorem 4.2]. It is clear that [19, Theorem 4.2] is a special case of Theorem 1 after observing that a $\text{CKT}(r, cn^s)$ -oblivious adversary can choose $f_n(\mathbf{x})$ to output a subset of rn bits of \mathbf{x} .

⁷One can show that $C(p, r, \infty, \infty)$ is strictly less than $C(p, r, c, s)$ for some values of $p \in (0, 1/2)$ and $r < 1 - H(p)$. See Section V for a proof sketch. The intuition behind this result follows from the fact that we have imposed a complexity bound of f while allowing the codebook \mathcal{C}_n to have unbounded complexity. Allowing encoding/decoding to use unlimited computation power while the adversary is $\text{CKT}(r, cn^s)$ -oblivious may give Alice and Bob an advantage compared the setting where both the codebook and observation function have similar complexity constraints.

⁸As discussed above, when $r > C_{\text{Shannon}}(p)$ one may find achievable rates strictly greater than the GV bound when stochastic codes are considered. One such stochastic coding scheme is the following. Suppose that the encoder passes its clean codeword \mathbf{u} through a $\text{BSC}(q)$ (q to be determined) to obtain the transmitted codeword \mathbf{x} . If the effective mutual information between the clean codeword \mathbf{u} and the adversary's observation is less than the rate R (the "right" notion of sufficient myopicity in this scenario), then the above stochastic coding scheme can be shown to achieve rate $R = 1 - H(p')$ where $p' = q(1 - p) + p(1 - q)$. As can be verified numerically, there exists some values of r, p and q such that $r > 1 - H(p)$ and $R > 1 - H(2p)$ (the GV bound).

III. PROOF OUTLINE, OVERVIEW OF PROOF TECHNIQUE

In this section, we outline the proof of Theorem 1 and discuss an overview of our proof technique. A detailed proof of Theorem 1 can be found in Section IV.

A. Achievability Scheme

For our proof of Theorem 1, we construct a specific \mathcal{C}_n .

Encoder Construction: Alice's (n, Rn) codebook \mathcal{C}_n is constructed as follows. Let $\rho \in (R, C_{\text{Shannon}}(p))$. Codebook \mathcal{C}_n is a concatenation of two codebooks: an *outer* $(\rho n, Rn)$ codebook $\mathcal{C}_{\text{out}} : [2^{Rn}] \rightarrow \{0, 1\}^{\rho n}$ and an *inner* $(n, \rho n)$ codebook $\mathcal{C}_{\text{in}} : \{0, 1\}^{\rho n} \rightarrow \{0, 1\}^n$. Encoding proceeds as follows. First, Alice encodes m_0 with \mathcal{C}_{out} where we denote the resulting codeword as $\mathcal{C}_{\text{out}}(m_0)$. Subsequently, Alice encodes $\mathcal{C}_{\text{out}}(m_0)$ with \mathcal{C}_{in} where we denote the resulting codeword as $\mathcal{C}_n(m_0) = \mathcal{C}_{\text{in}}(\mathcal{C}_{\text{out}}(m_0))$. After encoding, Alice transmits the codeword $\mathbf{x} = \mathcal{C}_n(m_0)$ over the channel. We denote the concatenated (n, Rn) codebook as $\mathcal{C}_n = \mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}$.

Decoder Construction: Bob's list decoder is constructed as follows. Given the received word $\mathbf{y} = \mathcal{C}_n(m_0) \oplus \mathbf{e}$, Bob first performs list decoding by forming a list $\mathcal{L}_{\text{in}}(\mathbf{y}, \mathcal{C}_{\text{in}})$ of all words $\mathbf{w} \in \{0, 1\}^{\rho n}$ such that $\mathcal{C}_{\text{in}}(\mathbf{w})$ is contained in the ball $\mathcal{B}_{\rho n}(\mathbf{y})$. After list decoding, Bob refines the list (i.e., Bob performs disambiguation) by removing all words $\mathbf{w} \in \mathcal{L}_{\text{in}}$ that are *not consistent* with \mathcal{C}_{out} : we say that a word \mathbf{w} is consistent with \mathcal{C}_{out} if there exists an $m \in [2^{Rn}]$ such that $\mathbf{w} = \mathcal{C}_{\text{out}}(m)$.

Denote the refined list as \mathcal{L}_{out} and note that $\mathcal{L}_{\text{out}} \subseteq \mathcal{L}_{\text{in}} \subseteq \{0, 1\}^{\rho n}$. After \mathcal{L}_{in} is refined to \mathcal{L}_{out} , a decoding decision is made according to the following rules. If $|\mathcal{L}_{\text{out}}| = 1$, then we have exactly one $m \in [2^{Rn}]$ s.t. $\mathcal{C}_{\text{out}}(m) \in \mathcal{L}_{\text{out}}$, and the decoder outputs $\hat{m} = m$. If \mathcal{L}_{out} is empty or if $|\mathcal{L}_{\text{out}}| > 1$, then the decoder declares an error by setting \hat{m} to an error symbol. We say that a decoding error occurs if $\hat{m} \neq m_0$. However, by the list decoding logic and the adversary error budget constraint ρn , $\mathcal{C}_{\text{out}}(m_0)$ is guaranteed to be in \mathcal{L}_{out} , and so a decoding error occurs if and only if $|\mathcal{L}_{\text{out}}| > 1$.

Probability of Error: For $i = 1, \dots, 2^{\rho n}$, define

$$\mathbf{w}_i(m_0, \mathbf{e}, \mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) = \arg \min_{\mathbf{w} \in \mathcal{W}_i(m_0, \mathbf{e}, \mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}})} \mathbf{int}(\mathbf{w}) \quad (2)$$

such that

$$\mathcal{W}_i(m_0, \mathbf{e}, \mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) = \mathbf{w} \in \{0, 1\}^{\rho n} \setminus \{\mathbf{w}_1, \dots, \mathbf{w}_{i-1}\} \text{ s.t. } d(\mathbf{y}, \mathcal{C}_{\text{in}}(\mathbf{w})) = \arg \min_{\mathbf{w} \in \{0, 1\}^{\rho n} \setminus \{\mathbf{w}_1, \dots, \mathbf{w}_{i-1}\}} d(\mathbf{y}, \mathcal{C}_{\text{in}}(\mathbf{w})).$$

That is, we sort the *message/word vectors* \mathbf{w} according to the distance between the observation \mathbf{y} and the inner codeword $\mathcal{C}_{\text{in}}(\mathbf{w})$, where the term $\mathbf{int}(\mathbf{w})$ in (2) is used to break any tie and ensure that the i th closest codeword to \mathbf{y} is uniquely defined. Note that $\mathbf{w}_i \in \mathcal{L}_{\text{in}}$ iff $i \leq |\mathcal{L}_{\text{in}}|$. Also define

$$\mathcal{I}_{m_0} = \{\mathcal{C}_{\text{out}}(m') : m' \neq m_0\} \quad (3)$$

to be the set of words in $\{0, 1\}^{\rho n}$ that are consistent with \mathcal{C}_{out} but do not correspond to the true message m_0 . Under the code

construction of Section III-A, the probability of decoding error can be written as

$$\begin{aligned} & \bar{P}_e(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) \\ &= \max_{f_n \in \text{CKT}(r, cn^s)} \mathbb{E}_{\Psi} \left[\max_{\mathbf{e} \in \mathcal{B}_{\rho n}(0)} \mathbb{P}_{m_0}(|\mathcal{L}_{\text{out}}| > 1 | \Psi(m_0) = \psi) \right] \\ &= \max_{f_n \in \text{CKT}(r, cn^s)} \mathbb{E}_{\Psi} \left[\max_{\mathbf{e} \in \mathcal{B}_{\rho n}(0)} \mathbb{P}_{m_0} \left(\bigcup_{i=1}^{|\mathcal{L}_{\text{in}}|} \{\mathbf{w}_i \in \mathcal{I}_{m_0}\} \mid \Psi = \psi \right) \right]. \end{aligned} \quad (4)$$

B. Preliminaries

The following preliminary results characterize the list-decodability properties of a random codebook. Let $Q(n, \rho n)$ be the distribution of $(n, \rho n)$ codebooks such that all codewords of \mathcal{C}_{in} are independently and uniformly distributed in $\{0, 1\}^n$.

Definition 2. For $L > 0$, an $(n, \rho n)$ codebook \mathcal{C}_{in} is said to be $[L, p]$ list decodable if $|\mathcal{C}_{\text{in}} \cap \mathcal{B}_{\rho n}(\mathbf{y})| \leq L$ for every $\mathbf{y} \in \{0, 1\}^n$.

Lemma 1. Let $\ell = \ell(n) > 0$ be $\omega(n)$ (i.e., $\lim_{n \rightarrow \infty} \ell(n)/n = \infty$). For large enough n , a codebook \mathcal{C}_{in} drawn from distribution $Q(n, \rho n)$ is $[\ell, p]$ list decodable w.p. greater than $1 - 2^{-\ell(n)/4}$. Proof is in Appendix C.

Similar results hold even if the list size is constant in n .

Lemma 2 ([9, Claim A.15]). Let $\epsilon_\rho \in (0, C_{\text{Shannon}}(p))$ and set $\rho = C_{\text{Shannon}}(p) - \epsilon_\rho$. For $L > \frac{1}{\epsilon_\rho}$ and for large enough n (depending only on ϵ_ρ), an $(n, \rho n)$ codebook \mathcal{C}_{in} drawn from distribution $Q(n, \rho n)$ is $[L, p]$ list decodable w.p. greater than $1 - \frac{1}{n}$.

Lemma 3. Consider an arbitrary 1-to-1 $(\rho n, Rn)$ codebook \mathcal{C}_{out} and randomly draw an $(n, \rho n)$ codebook \mathcal{C}_{in} from distribution $Q(n, \rho n)$. Recall that $\mathcal{C}_n = \mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}$. For any subset $\mathcal{A} \subseteq \{0, 1\}^n$, we have that $\mu = \mathbb{E}_{\mathcal{C}_{\text{in}}} |\mathcal{A} \cap \mathcal{C}_n| = 2^{-(1-R)n} |\mathcal{A}|$, and for $t_L < \mu$ and $t_U > \mu$,

$$\mathbb{P}_{\mathcal{C}_{\text{in}}} (|\mathcal{A} \cap \mathcal{C}_n| < t_L) \leq 2 \exp \left\{ \frac{-(\mu - t_L)^2}{4\mu} \right\}$$

and

$$\mathbb{P}_{\mathcal{C}_{\text{in}}} (|\mathcal{A} \cap \mathcal{C}_n| > t_U) \leq 2 \exp \left\{ \frac{-(t_U - \mu)^2}{4(t_U + \mu)} \right\}.$$

Proof is in Appendix D.

C. Overview of the proof of Theorem 1

For any error budget $p \in (0, 1/2)$ and observation rate parameter $r \in (0, C_{\text{Shannon}}(p))$, the goal of our proof of Theorem 1 is to prove that Alice and Bob can communicate at rate R that is arbitrarily close to $C_{\text{Shannon}}(p)$. Our proof idea is to prove the following slightly different statement instead: for any $p \in (0, 1/2)$ and for any $r \in (0, C_{\text{Shannon}}(p))$, there exists an $R \in (r, C_{\text{Shannon}}(p))$ such that Alice and Bob can communicate at rate R . Such a (seemingly weaker) statement implies Theorem 1 immediately, since for any $r' > r$, the achievable rate under r' is a lower bound of the achievable

rate under r . We can then let $r' \rightarrow C_{\text{Shannon}}(p)$ and use the (seemingly weaker) statement to derive Theorem 1. We now present the setup of our proof.

Setup: The following setup will be used throughout the proof of Theorem 1:

- 1) Fix any error budget $p \in (0, 1/2)$ and observation rate $r \in (0, C_{\text{Shannon}}(p))$, and fix observation complexity bound parameters c, s to be positive integers.
- 2) We can always find parameters $\delta_0, \delta_1, \epsilon_\rho, \epsilon_R > 0$ such that the following two conditions hold:

Condition 1. $r < C_{\text{Shannon}}(p) - \delta_0 - \delta_1 - \epsilon_\rho - \epsilon_R$

Condition 2. $\epsilon_R \in (0, (\frac{5}{13} - \frac{1}{30})\delta_0)$

Set the inner-code rate $\rho = C_{\text{Shannon}}(p) - \epsilon_\rho$ and the inner-outer concatenated code rate $R = \rho - \epsilon_R$. One can easily verify that the above choice of parameters guarantees that $r < R < \rho < C_{\text{Shannon}}(p)$.

- 3) For blocklength $n = 1, 2, \dots$, let the codebook \mathcal{C}_n be the code construction described in Section III-A.
- 4) Fix the $(\rho n, Rn)$ outer codebook \mathcal{C}_{out} to be any 1-to-1 function from $\{0, 1\}^{Rn}$ to $\{0, 1\}^{\rho n}$. Let the $(n, \rho n)$ inner codebook \mathcal{C}_{in} be drawn from distribution $Q(n, \rho n)$. Note that $\mathcal{C}_n = \mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}$ is Alice's (n, Rn) codebook.

We now show that the rate R is (c, s) -achievable by using a *random-coding argument*, i.e., we show that for any $\epsilon_e > 0$ and for large enough n , $\mathbb{P}_{\mathcal{C}_{\text{in}}}(\bar{P}_e(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) > \epsilon_e) < 1$ and thus there exists a sequence of $(\rho n, Rn)$ codebooks \mathcal{C}_{out} and $(n, \rho n)$ codebooks \mathcal{C}_{in} such that $\bar{P}_e(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) \leq \epsilon_e$ for all n large enough.

Random-Coding: In the sequel, we drop the dependency on \mathcal{C}_{out} from all notation due to the outer codebook being fixed. We write $\bar{P}_e(\mathcal{C}_{\text{in}})$ to denote the probability of decoding error evaluated at the (n, Rn) codebook $\mathcal{C}_n = \mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}$.

For $f_n \in \text{CKT}(r, cn^s)$, $\psi \in \{0, 1\}^{rn}$, $e \in \mathcal{B}_{pn}(0)$ and $i \in [2^{\rho n}]$, define

$$q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}) = \mathbb{P}_{m_0}(\mathbf{w}_i \in \mathcal{L}_{\text{in}}, \mathbf{w}_i \in \mathcal{I}_{m_0} | \Psi(m_0) = \psi) \quad (5)$$

to be the probability that word $\mathbf{w}_i(m_0, e, \mathcal{C}_{\text{in}})$ results in a decoding error given that the adversary observes $\Psi(m_0) = \psi$. To apply the random-coding argument, we first apply a simple union bound to $\bar{P}_e(\mathcal{C}_{\text{in}})$ in (4) to bound the quantity above by

$$\bar{P}_e^{\text{ub}}(\mathcal{C}_{\text{in}}) = \max_{f_n \in \text{CKT}(r, cn^s)} \sum_{i=1}^{2^{\rho n}} \mathbb{E}_{\Psi} \left[\max_{e \in \mathcal{B}_{pn}(0)} q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}) \right]. \quad (6)$$

We now prepare to state a sufficient condition for the rate R to be (c, s) -achievable. Recall that ϵ_ρ , δ_0 , and δ_1 are the parameters used to construct \mathcal{C}_{in} and \mathcal{C}_{out} . For integer $L \in (0, 2^{\rho n})$, define the product set $\mathcal{P}(L) = [L] \times \text{CKT}(r, cn^s) \times \{0, 1\}^{rn} \times \mathcal{B}_{pn}(0)$. For $\epsilon_e > 0$, we define the set $\mathcal{H}(L, \epsilon_e)$ to be the set of all $(n, \rho n)$ codebooks \mathcal{C}_{in} such that for all $(i, f_n, \psi, e) \in \mathcal{P}(L)$, either $\mathbb{P}_{m_0}(\Psi(m_0) = \psi) < 2^{-(\delta_0 + \delta_1 - R)n}$ or $q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}) \leq \frac{\epsilon_e}{2L}$.

The intuition behind the definition of $\mathcal{H}(L, \epsilon_e)$ is as follows. For any observation-function/observation-pair (f_n, ψ) , we say that this pair is *informative* if $\mathbb{P}_{m_0}(\Psi(m_0) = \psi) < 2^{-(\delta_0 + \delta_1 - R)n}$. Namely, if the adversary picks f_n and observes $\Psi(m_0) = \psi$, then there are not many other messages $m \neq m_0$

such that $\Psi(m) = \psi$. As a result, the adversary knows that the true message m_0 must be in a very small set of possibilities, thus the name "informative". The set $\mathcal{H}(L, \epsilon_e)$ then considers the set of inner codebooks such that for any (f_n, ψ) that is *not* informative, no matter how the adversary designs the error vector e , with high probability $1 - \epsilon_e/(2L)$, each of the L inner codewords that are closest to $\mathbf{y} = \mathbf{x} \oplus e$ is either outside the Hamming ball $B_{pn}(\mathbf{y})$ or can be ruled out by the outer codebook \mathcal{C}_{out} .

Given the above intuition, we may consider any codebook in $\mathcal{H}(L, \epsilon_e)$ to be a good choice of \mathcal{C}_{in} . The reason is that when the pair (f_n, ψ) is informative, the adversary knows very accurately which message is likely to be m_0 and thus it is hard to keep the error probability small. However, $\mathcal{H}(L, \epsilon_e)$ ensures that under a more favorable situation in which the (f_n, ψ) is not informative, the inner codebook \mathcal{C}_{in} can take advantage of this ambiguity at the adversary and guarantee small error probability for the L closest inner codewords (thus the enumerating index i) and regardless of how the adversary chooses the error vector e .

Lemma 4 (Sufficient Condition for Achievability). *Let $L > 1/\epsilon_\rho$ be a constant. If for any $\epsilon_e > 0$, there exists an n_0 such that for all $n \geq n_0$, the probability $\mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{H}(L, \epsilon_e)) < 1 - 1/n$, then the rate R is (c, s) -achievable.*

Proof. Let $L > 1/\epsilon_\rho$ and let $\epsilon_e > 0$. Consider the probability

$$\mathbb{P}_{\mathcal{C}_{\text{in}}} \left(\mathcal{C}_{\text{in}} \text{ is not } [L, p] \text{ list decodable or } \mathcal{C}_{\text{in}} \notin \mathcal{H}(L, \epsilon_e) \right). \quad (7)$$

By a simple union bound, probability (7) is bounded above by

$$\mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \text{ is not } [L, p] \text{ list dec.}) + \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{H}(L, \epsilon_e)). \quad (8)$$

By Lemma 2, there exists an n_1 such that for all $n \geq n_1$, the first term in equation (8) is bounded above by $1/n$. In turn, since for all $n \geq n_0$ the second term in equation (8) is strictly smaller than $1 - 1/n$, it follows that for all $n \geq \max\{n_0, n_1\}$ probability (7) is strictly less than 1. Thus, for each $n \geq \max\{n_0, n_1\}$, there exists an $(n, \rho n)$ codebook $\mathcal{C}_{\text{in}}^*$ such that $\mathcal{C}_{\text{in}}^*$ is $[L, p]$ list decodable and $\mathcal{C}_{\text{in}}^* \in \mathcal{H}(L, \epsilon_e)$.

The above shows the existence of a special codebook $\mathcal{C}_{\text{in}}^*$. In the following, we show that the error probability evaluated at $\mathcal{C}_{\text{in}}^*$ can be upper bounded analytically. Specifically, since $\mathcal{C}_{\text{in}}^*$ is $[L, p]$ list decodable, we have the identity $q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}^*) = 0$ for all $i > L \geq \max_{\mathbf{y} \in \{0, 1\}^n} |\mathcal{L}_{\text{in}}(\mathbf{y}, \mathcal{C}_{\text{in}}^*)|$, and therefore, $\bar{P}_e^{\text{ub}}(\mathcal{C}_{\text{in}}^*)$ in (6) is equal to

$$\max_{f_n \in \text{CKT}(r, cn^s)} \sum_{i=1}^L \mathbb{E}_{\Psi} \left[\max_{e \in \mathcal{B}_{pn}(0)} q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}^*) \right].$$

For any fixed $f_n \in \text{CKT}(r, cn^s)$ and fixed $i \in [1, L]$, we have

$$\begin{aligned} & \mathbb{E}_{\Psi} \left[\max_{e \in \mathcal{B}_{pn}(0)} q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}^*) \right] \\ &= \sum_{\substack{\psi: (f_n, \psi) \text{ is} \\ \text{not informative}}} \mathbb{P}_{m_0}(\Psi(m_0) = \psi) \cdot \max_{e \in \mathcal{B}_{pn}(0)} q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}^*) \end{aligned} \quad (9)$$

$$+ \sum_{\substack{\psi: (f_n, \psi) \text{ is} \\ \text{informative}}} \mathbb{P}_{m_0}(\Psi(m_0) = \psi) \cdot \max_{e \in \mathcal{B}_{pn}(0)} q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}^*) \quad (10)$$

for which we partition based on the events that ψ and the given f_n are informative or not. Since $\mathcal{C}_{\text{in}}^* \in \mathcal{H}(L, \epsilon_e)$, by the definition of $\mathcal{H}(L, \epsilon_e)$, the first summation (9) can be upper bounded by

$$\sum_{\substack{\psi: (f_n, \psi) \text{ is} \\ \text{not informative}}} \mathbb{P}_{m_0}(\Psi(m_0) = \psi) \cdot \frac{\epsilon_e}{2L} \leq \frac{\epsilon_e}{2L}. \quad (11)$$

Since $q(\cdot)$ is a probability, we have $\max_{e \in \mathcal{B}_{pn}(0)} q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}^*) \leq 1$. By the definition of $\mathcal{H}(L, \epsilon_e)$ the second summation (10) can be upper bounded by

$$\sum_{\substack{\psi: (f_n, \psi) \text{ is} \\ \text{informative}}} \mathbb{P}_{m_0}(\Psi(m_0) = \psi) \leq 2^r 2^{(\delta_0 + \delta_1 - R)n}. \quad (12)$$

By (11) and (12) and by summing over $i = 1, \dots, L$, we have that $\bar{P}_e^{\text{ub}}(\mathcal{C}_{\text{in}}^*)$ is bounded above by

$$\max_{f_n \in \text{CKT}(r, cn^s)} L \left(\frac{\epsilon_e}{2L} + 2^{(r + \delta_0 + \delta_1 - R)n} \right). \quad (13)$$

Following Condition 1, the exponent $r + \delta_0 + \delta_1 - R$ is strictly negative, and thus for large enough n , the quantity (13) is bounded above by ϵ_e . In conclusion, for large enough n , $\bar{P}_e(\mathcal{C}_{\text{in}}^*) \leq \epsilon_e$. \blacksquare

As a result, for $L > 1/\epsilon_\rho$ and $\epsilon_e > 0$, our strategy will be to lower bound the probability $\mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \in \mathcal{H}(L, \epsilon_e))$ and apply Lemma 4. In this strategy, a significant amount of work is needed to show that the following statement holds with probability greater than $1/n$ over the choice of \mathcal{C}_{in} :

$$\max_{\substack{(i, f_n, \psi, e) \in \mathcal{P}(L): \\ (f_n, \psi) \text{ is not informative}}} q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}) \leq \frac{\epsilon_e}{2L}. \quad (14)$$

The first step in this work is to show that each of the $q_i(f_n, \psi, e, \mathcal{C}_{\text{in}})$ terms in (14) has a small expectation (w.r.t \mathcal{C}_{in}), i.e.,

$$\lim_{n \rightarrow \infty} \max_{(i, f_n, \psi, e) \in \mathcal{P}(L)} \mathbb{E}_{\mathcal{C}_{\text{in}}} [q_i(f_n, \psi, e, \mathcal{C}_{\text{in}})] = 0. \quad (15)$$

We prove this result in Lemma 5. The next step is to show that each of the $q_i(f_n, \psi, e, \mathcal{C}_{\text{in}})$ terms is *concentrated* around its expectation $\mathbb{E}_{\mathcal{C}_{\text{in}}} [q_i(f_n, \psi, e, \mathcal{C}_{\text{in}})]$, i.e., for any $\epsilon'_e \in (0, \frac{\epsilon_e}{2L})$,

for large enough n and with probability greater than $1/n$ over the choice of \mathcal{C}_{in} , the following inequality holds:

$$\max_{\substack{(i, f_n, \psi, e) \in \mathcal{P}(L): \\ (f_n, \psi) \text{ not informative}}} (q_i(f_n, \psi, e, \mathcal{C}_{\text{in}}) - \mathbb{E}_{\mathcal{C}_{\text{in}}} [q_i(f_n, \psi, e, \mathcal{C}_{\text{in}})]) \leq \epsilon'_e. \quad (16)$$

The bulk of our proof is dedicated towards this step. Since $\epsilon'_e < \epsilon_e/2L$ with strict inequality, (15) and (16) together imply that (14) holds with probability strictly greater than $1/n$. In the remainder of this overview, we outline our approach for studying the concentration of measure of $q_i(f_n, \psi, e, \mathcal{C}_{\text{in}})$.

Concentration: When confusion can be avoided, we drop the notated dependencies and subscripts of $q_i(f_n, \psi, e, \mathcal{C}_{\text{in}})$ and simply write $q(\mathcal{C}_{\text{in}})$ to emphasize the dependency on \mathcal{C}_{in} . For integer $L > 1/\epsilon_\rho$, fixed $(i, f_n, \psi, e) \in \mathcal{P}(L)$ such that (f_n, ψ) is not informative, and for $n = 1, 2, 3, \dots$, we study the concentration of measure of $q(\mathcal{C}_{\text{in}})$ around its expectation $\mathbb{E}_{\mathcal{C}_{\text{in}}} [q]$ by deriving concentration inequalities from the logarithmic Sobolev inequalities, e.g., [24]. This method of deriving concentration inequalities is also known as the entropy method.

At a high level, the concentration inequalities tell us that if a function g from the set of $(n, \rho n)$ codebooks to the real numbers is *smooth* for *most* $(n, \rho n)$ codebooks, then g is concentrated (around its expectation). To define “most”, we define a subset \mathcal{T} of $(n, \rho n)$ codebooks with the property that $\mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{T}) = \exp\{-2^{\Omega(n)}\}$ (Definition 3). We refer to the set \mathcal{T} as a typical set of $(n, \rho n)$ codebooks and say a codebook \mathcal{C}_{in} is typical if $\mathcal{C}_{\text{in}} \in \mathcal{T}$. To define “smooth”, define the variation of g as

$$V(\mathcal{C}_{\text{in}}) = \sum_{j=1}^{2^{\rho n}} \mathbb{E}_{\mathbf{z}} |g(\mathcal{C}_{\text{in}}) - g(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|^2$$

where codebook $\mathcal{C}_{\text{in}}(j, \mathbf{z})$ is equal to \mathcal{C}_{in} with the j th codeword replaced with the word \mathbf{z} uniformly distributed in $\{0, 1\}^n$. We say that a number $a_G > 0$ is a *global variation coefficient* of g if for any $(n, \rho n)$ codebook \mathcal{C}_{in} , $V(\mathcal{C}_{\text{in}}) \leq a_G$. Similarly, we say that $a_T > 0$ is a *typical variation coefficient* of g if for any $\mathcal{C}_{\text{in}} \in \mathcal{T}$, $V(\mathcal{C}_{\text{in}}) \leq a_T$. Finally, we say that g is smooth for most codebooks if g has typical and global variation coefficients that are both sufficiently small.

Given the above definitions, the following statement summarizes our concentration inequalities: If g has a typical variation coefficient $a_T = \exp\{-2^{\Omega(n)}\}$ and a global variation coefficient $a_G = O(1)$, then⁹

$$\mathbb{P}_{\mathcal{C}_{\text{in}}}(g - \mathbb{E}_{\mathcal{C}_{\text{in}}}[g] > \epsilon'_e) = \exp\{-2^{\Omega(n)}\}.$$

Three remarks are at hand. First, the double-exponential bound ensures that a union bound can be successfully applied to the probability that event (16) occurs (more on this below). Second, the $O(1)$ global variation coefficient prevents the inequalities from blowing up over the set \mathcal{T}^c . Lastly, these inequalities cannot be directly applied in our setting to show that q is concentrated. This last remark follows from the fact that while one can find a typical variation coefficient of q

⁹See Lemma 11 for additional conditions on a_T and a_G .

that is $\exp\{-2^{\Omega(n)}\}$, it is difficult to find a global variation coefficient of q that is $O(1)$. To circumvent this issue, we proceed with the following additional steps.

- For an $(n, \rho n)$ codebook \mathcal{C}_{in} , we define an approximation function $q'(\mathcal{C}_{\text{in}})$ to approximate $q(\mathcal{C}_{\text{in}})$ (Definition 4). The approximation function has the following properties:
 - For any $(n, \rho n)$ codebook \mathcal{C}_{in} , $q'(\mathcal{C}_{\text{in}}) \leq q(\mathcal{C}_{\text{in}})$ which holds with equality if $\mathcal{C}_{\text{in}} \in \mathcal{T}$. Hence, the function q' is a good approximation of q in the sense that $\mathbb{P}_{\mathcal{C}_{\text{in}}}(q \neq q') \leq \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \in \mathcal{T})$.
 - The function q' has a global variation coefficient that is $O(1)$ (Lemma 6). We remark that the concatenated structure of our code construction simplifies the proof of this bound.
- Given a global variation coefficient that is $O(1)$, we show that q' is concentrated, i.e., $\mathbb{P}_{\mathcal{C}_{\text{in}}}(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] > \epsilon'_e) = \exp\{-2^{\Omega(n)}\}$ (Lemma 11).
- We show the concentration of q by proving that our special construction of q' satisfies the following approximation bound:

$$\begin{aligned} & \mathbb{P}_{\mathcal{C}_{\text{in}}}(q - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] > \epsilon'_e) \\ & \leq \mathbb{P}_{\mathcal{C}_{\text{in}}}(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] > \epsilon'_e) + \mathbb{P}_{\mathcal{C}_{\text{in}}}(q \neq q') \\ & \leq \mathbb{P}_{\mathcal{C}_{\text{in}}}(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] > \epsilon'_e) + \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{T}) \end{aligned}$$

(Lemma 10). It follows that $\mathbb{P}_{\mathcal{C}_{\text{in}}}(q - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] > \epsilon'_e) = \exp\{-2^{\Omega(n)}\}$.

To complete the proof that inequality (14) holds with probability greater than $1/n$, we apply a simple union bound:

$$\begin{aligned} & \mathbb{P}_{\mathcal{C}_{\text{in}}}\left(\max_{\substack{(i, f_n, \psi, e) \in \mathcal{P}(L): \\ (f_n, \psi) \text{ not inform.}}} (q - \mathbb{E}[q]) > \epsilon'_e\right) \\ & \leq |\mathcal{P}(L)| \cdot \max_{\substack{(i, f_n, \psi, e) \in \mathcal{P}(L): \\ (f_n, \psi) \text{ not inform.}}} \mathbb{P}_{\mathcal{C}_{\text{in}}}(q - \mathbb{E}[q] > \epsilon'_e) \\ & \leq |\mathcal{P}(L)| \exp\{-2^{\Omega(n)}\} \end{aligned} \quad (17)$$

where $|\mathcal{P}(L)|$ denotes the number of elements in the product space $[L] \times \text{CKT}(r, cn^s) \times \{0, 1\}^{\rho n} \times \mathcal{B}_{pn}(0)$. The final step is to show that (17) is bounded above by $1 - 1/n$, which we show by verifying that $|\mathcal{P}(L)| = 2^{\text{poly}(n)}$ and thus $|\mathcal{P}(L)|$ is growing much slower than the double exponential $\exp\{2^{\Omega(n)}\}$. The key idea in this final step is to use the adversary's computational bound and show that the number of functions in $\text{CKT}(r, cn^s)$ is $2^{\text{poly}(n)}$. We remark that the bounded observation complexity is critical to the proof since if we allow for unbounded circuit complexity, the set $\text{CKT}(r, \infty)$ contains $\exp\{2^{\Omega(n)}\}$ functions.

Prior work: The above approach is inspired by Langberg's framework [14] to study concentration of measure when the function under analysis is smooth over a typical set \mathcal{T} of codebooks. The main technical contribution of [14] is to carefully define \mathcal{T} based on the codebooks' list decodable properties in a way where one can then apply Vu's martingale-type concentration inequalities for non-smooth functions [25]. We follow Langberg's framework by also defining typicality in terms of list decodability. However, we use concentration inequalities derived via the entropy method.

The major technical difference between our work and Langberg's [14] lies at the definition of smoothness. Langberg adopts a Lipschitz criterion of smoothness which states that a function g is smooth if g has a sufficiently small *typical Lipschitz coefficient* $K_T > 0$; K_T is said to be a typical Lipschitz coefficient if for any $\mathcal{C}_{\text{in}} \in \mathcal{T}$, the quantity

$$W(\mathcal{C}_{\text{in}}) = 2^{\rho n} \max_{j \in [2^{\rho n}], \mathbf{z} \in \{0, 1\}^n} |g(\mathcal{C}_{\text{in}}) - g(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|^2$$

is bounded above by K_T . Similarly, a number K_G is said to be a global Lipschitz coefficient if for any $(n, \rho n)$ codebook \mathcal{C}_{in} , $W(\mathcal{C}_{\text{in}}) \leq K_G$. Our work identifies and exploits two advantages of using the variation criterion over the Lipschitz criterion for characterizing smoothness in our setting. First, for a typical codebook $\mathcal{C}_{\text{in}} \in \mathcal{T}$, variation $V(\mathcal{C}_{\text{in}})$ captures more information about the behavior of g locally around codebook \mathcal{C}_{in} than $W(\mathcal{C}_{\text{in}})$. We leverage this additional information to find typical variation coefficients for q' that are smaller than any typical Lipschitz coefficient. Second, for a non-typical codebook $\mathcal{C}_{\text{in}} \notin \mathcal{T}$, the best bound on $W(\mathcal{C}_{\text{in}})$ is $O(2^{\rho n})$. Thus, a good global Lipschitz coefficient of q' is much larger than the $O(1)$ global variation coefficient established in our proof.

Similar to our work and the work of [14], other works adopt proof techniques that are combinatorial in nature. These include the studies by Csiszár and Narayan [12], [13] on ACVs with input and state constraints which adopt a method-of-types approach. We note that the channel controlled by a $\text{CKT}(r, cn^s)$ -oblivious adversary with error budget p can be formulated as an AVC with state constraints. These works also include the study by Dey, Jaggi and Langberg [7] on myopic adversarial channels.¹⁰

Lastly, we remark that our proof techniques and analysis allow for generalization to other channel models. For example, straightforward modifications of our techniques/analysis can allow the bit flip channel from Alice to Bob to be generalized to a q -ary error/erasure channel for $q \geq 2$ in which Alice sends symbols from a q -ary alphabet and the adversary can induce both symbol errors and symbol erasures.

IV. PROOF OF THEOREM 1

In the sequel, we use the setup described in Section III-C.

A. Notation

The following notation will assist in our proof of Theorem 1. For an observation function $f_n \in \text{CKT}(r, cn^s)$ and observation $\psi \in \{0, 1\}^{rn}$, we define the observation set $\mathcal{O}_\psi = \{z \in \{0, 1\}^n : f_n(z) = \psi\}$. Note that observing $\Psi(m_0) \triangleq f_n(\mathcal{C}_n(m_0)) = \psi$ is equivalent to knowing that $\mathcal{C}_n(m_0) \in \mathcal{O}_\psi$. Hence, the following two perspectives are equivalent: a) the adversary draws an observation function f_n and b) the adversary draws a partition $\vec{\mathcal{O}} = (\mathcal{O}_1, \dots, \mathcal{O}_{2^{rn}})$ of the space $\{0, 1\}^n$ consisting of 2^{rn} non-empty observation sets. With an abuse of notation, for $f_n \in \text{CKT}(r, cn^s)$, we write

¹⁰We note that the proof techniques of [7] can provide a simple alternative proof of Theorem 1. We provide an outline of this alternative proof in Appendix A.

$\vec{\mathcal{O}} \in \text{CKT}(r, cn^s)$ to denote the partition of observation sets corresponding to f_n with circuit complexity upper bounded by cn^s . Along the same lines, for $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(2^{\rho n})$ and for an $(n, \rho n)$ codebook \mathcal{C}_{in} , we write $q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ to denote $q_i(f_n, \psi, e, \mathcal{C}_{\text{in}})$.

B. Expectation of q

For any $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(2^{\rho n})$, the following result characterizes the expectation of $q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ when the $(n, \rho n)$ codebook \mathcal{C}_{in} is drawn from distribution $Q(n, \rho n)$.

Lemma 5.

$$\lim_{n \rightarrow \infty} \max_{(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(2^{\rho n})} \mathbb{E}_{\mathcal{C}_{\text{in}}} [q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})] = 0.$$

Proof. Our approach is to find a small upper bound of the expectation $\mathbb{E}_{\mathcal{C}_{\text{in}}} [q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})]$ that is independent of the parameters $i, \vec{\mathcal{O}}, \psi$ and e . Let \emptyset denote the empty set and let $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(2^{\rho n})$. Observe that for an $(n, \rho n)$ codebook \mathcal{C}_{in} , $q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ is bounded above by $\mathbb{P}_{m_0}(\mathcal{I}_{m_0} \cap \mathcal{L}_{\text{in}}(\mathbf{y}, \mathcal{C}_{\text{in}}) \neq \emptyset | \Psi(m_0) = \psi)$. Hence, for any $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(2^{\rho n})$ we have

$$\begin{aligned} & \mathbb{E}_{\mathcal{C}_{\text{in}}} [q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})] \\ & \leq \mathbb{E}_{\mathcal{C}_{\text{in}}} [\mathbb{P}_{m_0}(\mathcal{I}_{m_0} \cap \mathcal{L}_{\text{in}}(m_0, e, \mathcal{C}_{\text{in}}) \neq \emptyset | \Psi(m_0) = \psi)] \\ & = \mathbb{E}_{\mathcal{C}_{\text{in}}} [\mathbb{E}_{m_0 | \Psi = \psi} [\mathbb{1}\{\mathcal{I}_{m_0} \cap \mathcal{L}_{\text{in}}(\mathbf{y}, \mathcal{C}_{\text{in}}) \neq \emptyset\}]] \\ & = \mathbb{E}_{m_0 | \Psi = \psi} [\mathbb{E}_{\mathcal{C}_{\text{in}}} [\mathbb{1}\{\mathcal{I}_{m_0} \cap \mathcal{L}_{\text{in}}(\mathbf{y}, \mathcal{C}_{\text{in}}) \neq \emptyset\}]] \end{aligned}$$

Thus, to prove Lemma 5, it is sufficient to show that the quantity

$$\max_{\substack{(m, i, \vec{\mathcal{O}}, \psi, e) \\ \in [2^{Rn}] \times \mathcal{P}(2^{\rho n})}} \mathbb{E}_{\mathcal{C}_{\text{in}}} [\mathbb{1}\{\mathcal{I}_m \cap \mathcal{L}_{\text{in}}(\mathbf{y}, \mathcal{C}_{\text{in}}) \neq \emptyset\} | m_0 = m] \quad (18)$$

is going to zero in the limit as $n \rightarrow \infty$. This sufficient condition simplifies our problem as the expectation inside the maximum of (18) only depends on parameters m, i and e , but not on $\vec{\mathcal{O}}$ or ψ . In the next steps, we will further bound the expectation in (18) to relax its dependency on parameters i and e by exploiting the list-decodability properties of the random codebook \mathcal{C}_{in} .

Let $m \in [2^{Rn}]$, $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(2^{\rho n})$ and let $\mathbf{y}_m = \mathcal{C}_n(m) \oplus e$. By the definition of the set \mathcal{I}_m and a simple union bound, the quantity $\mathbb{E}_{\mathcal{C}_{\text{in}}} [\mathbb{1}\{\mathcal{I}_m \cap \mathcal{L}_{\text{in}}(\mathbf{y}, \mathcal{C}_{\text{in}}) \neq \emptyset\} | m_0 = m]$ is bounded above by

$$\sum_{m' \in [2^{Rn}] \setminus \{m\}} \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{out}}(m') \in \mathcal{L}_{\text{in}}(\mathbf{y}_m, \mathcal{C}_{\text{in}}))$$

which in turn, by letting \mathcal{E} be the event that \mathcal{C}_{in} is $[n^2 + 1, p]$ list decodable and by the law of total probability, is bounded above by

$$\sum_{m' \in [2^{Rn}] \setminus \{m\}} (\mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{out}}(m') \in \mathcal{L}_{\text{in}}(\mathbf{y}_m, \mathcal{C}_{\text{in}}) | \mathcal{E}) + \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{E}^c)). \quad (19)$$

Note that for $m' \in [2^{Rn}] \setminus \{m\}$, $\mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{out}}(m') \in \mathcal{L}_{\text{in}}(\mathbf{y}_m, \mathcal{C}_{\text{in}}) | \mathcal{E})$ is bounded above by $\frac{n^2}{2^{\rho n} - 1}$ following that the codeword $\mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}(m')$ can be one of at most n^2 codewords of \mathcal{C}_{in} randomly chosen from $2^{\rho n} - 1$ codewords

(nb. we can exclude codeword $\mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}(m)$) contained in $\mathcal{L}_{\text{in}}(\mathbf{y}_m, \mathcal{C}_{\text{in}})$ by the list decodability properties of \mathcal{C}_{in} . Also, by Lemma 1, for large enough n , $\mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{E}^c)$ is bounded above by $2^{-(n^2+1)/4}$. It follows that quantity (19) is bounded above by $\frac{2^{Rn}-1}{2^{\rho n}-1} n^2 + (2^{Rn}-1)2^{-(n^2+1)/4}$ which in turn is going to zero in the limit as $n \rightarrow \infty$ independent of $(m, i, \vec{\mathcal{O}}, \psi, e) \in [2^{Rn}] \times \mathcal{P}(2^{\rho n})$. ■

C. Approximation of q

For $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(2^{\rho n})$, we will not directly show that the quantity $q(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ (as a function of \mathcal{C}_{in}) is concentrated. Instead, we will approximate $q(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ with an approximation function $q'(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ and study the concentration of $q'(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$. We carefully define the approximation function such that $q'(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ has good smoothness properties (and thus $q'(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ concentrates around its mean), and such that we can imply the concentration of $q(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ from the concentration of $q'(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$.

We first define typical codebooks. For the parameter $\delta_0 > 0$, we define typical codebooks for all $\vec{\mathcal{O}} \in \text{CKT}(r, cn^s)$ and $\psi \in \{0, 1\}^{rn}$ such that the observation set \mathcal{O}_ψ is larger than $2^{(1-R)n} 2^{\delta_0 n}$. We remark that the condition $|\mathcal{O}_\psi| \geq 2^{(1-R)n} 2^{\delta_0 n}$ is related to the pair (f_n, ψ) being not informative (see definition of informative in the overview of Section III-C). In our analysis of \bar{P}_e^{ub} , we will let decoding fail for all \mathcal{O}_ψ that are smaller than $2^{(1-R)n} 2^{\delta_0 n}$. Hence, there is no need to define typical codebooks for small observation sets.

Definition 3 (Typical Codebooks). *Suppose that $\vec{\mathcal{O}} \in \text{CKT}(r, cn^s)$ and $\psi \in \{0, 1\}^{rn}$ such that $|\mathcal{O}_\psi| \geq 2^{(1-R)n} 2^{\delta_0 n}$. Set $\delta'_0 = \delta'_0(\vec{\mathcal{O}}, \psi, \epsilon_R, \epsilon_\rho) \geq \delta_0$ to be the unique number such that $|\mathcal{O}_\psi| = 2^{(1-R)n} 2^{\delta'_0 n}$. Set*

$$\ell(\vec{\mathcal{O}}, \psi, \epsilon_R, \epsilon_\rho) = 2^{\frac{4\delta'_0}{13} n}$$

$$t_L(\vec{\mathcal{O}}, \psi, \epsilon_R, \epsilon_\rho) = 2^{-(1-R)n} |\mathcal{O}_\psi| - 2^{\frac{3\delta'_0}{4} n} = 2^{\delta'_0 n} - 2^{\frac{3}{4}\delta'_0 n}$$

and

$$t_U(\vec{\mathcal{O}}, \psi, \epsilon_R, \epsilon_\rho) = 2^{-(1-R)n} |\mathcal{O}_\psi| + 2^{\frac{3\delta'_0}{4} n} = 2^{\delta'_0 n} + 2^{\frac{3}{4}\delta'_0 n}.$$

An $(n, \rho n)$ codebook \mathcal{C}_{in} is said to be typical w.r.t. the parameters $\vec{\mathcal{O}}, \psi, \epsilon_R, \epsilon_\rho$ if \mathcal{C}_{in} is $[\ell, p]$ list decodable and $t_L \leq |\mathcal{O}_\psi \cap \mathcal{C}_n| \leq t_U$ where $\mathcal{C}_n = \mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}$. Define the typical set $\mathcal{T}_{\mathcal{O}_\psi}$ as the set of all $(n, \rho n)$ codebooks that are typical w.r.t. $\vec{\mathcal{O}}, \psi, \epsilon_R, \epsilon_\rho$.

We now provide an equivalent expression of $q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ that is convenient for defining our approximation function. For $m \in [2^{Rn}]$ and for $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(2^{\rho n})$, define

$$\begin{aligned} \phi_{i,m}(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}) & = \mathbb{1}\{\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}) \in \mathcal{I}_m \cap \mathcal{L}_{\text{in}}(\mathbf{y}_m, \mathcal{C}_{\text{in}})\} \mathbb{1}\{\mathcal{C}_n(m) \in \mathcal{O}_\psi\} \end{aligned}$$

and define

$$\Phi_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}) = \sum_{m \in [2^{Rn}]} \phi_{i,m}(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}).$$

For $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(2^{\rho n})$, note that

$$q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}) = \frac{\Phi_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})}{|\mathcal{O}_\psi \cap \mathcal{C}_n|}. \quad (20)$$

Definition 4 (Approximation function). Suppose that $\vec{\mathcal{O}} \in \text{CKT}(r, cn^s)$ and $\psi \in \{0, 1\}^{rn}$ such that $|\mathcal{O}_\psi| \geq 2^{(1-R)n} 2^{\delta_0 n}$. For $i \in [2^{\rho n}]$, $e \in \mathcal{B}_{pn}(0)$ and $(n, \rho n)$ codebook \mathcal{C}_{in} , define the approximation function

$$q'_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}) = \frac{\Phi_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})}{t(\vec{\mathcal{O}}, \psi, \mathcal{C}_{\text{in}})} \quad (21)$$

where $t(\vec{\mathcal{O}}, \psi, \mathcal{C}_{\text{in}}) = \max\{|\mathcal{O}_\psi \cap \mathcal{C}_n|, t_L\}$. Notice that $q'_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}) \leq q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ with equality if $\mathcal{C}_{\text{in}} \in \mathcal{T}_{\mathcal{O}_\psi}$. Furthermore, define the variation of $q'_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ as

$$V'_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}) = \sum_{j=1}^{2^{\rho n}} \mathbb{E}_{\mathbf{z}} \left[\Delta'(j, \mathbf{z}, \mathcal{C}_{\text{in}})^2 \right] \quad (22)$$

where the bounded difference

$$\Delta'(j, \mathbf{z}, \mathcal{C}_{\text{in}}) = |q'_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}) - q'_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}(j, \mathbf{z}))|$$

and the expectation is taken over the random variable \mathbf{z} uniformly distributed in $\{0, 1\}^n$.

The above definition of $q'_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ is carefully set such that $V'_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ is well behaved for all non-typical \mathcal{C}_{in} . This behavior is established in Section IV-E.

D. Combinatorial Preliminaries

In this section, we prove a few claims about the combinatorial properties of the quantities defined thus far. These claims will be used in the following section to characterize the smoothness properties of the approximation function $q'(\cdot)$.

In the sequel, unless otherwise stated, we fix integer $L > 1/\epsilon_\rho$, fix $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(L)$, and allow only the $(n, \rho n)$ codebook \mathcal{C}_{in} to vary. We drop the fixed variables from our notation. We write $q(\mathcal{C}_{\text{in}})$ to denote $q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$. Similarly, we write \mathcal{T} to denote $\mathcal{T}_{\mathcal{O}_\psi}$, $\Phi(\mathcal{C}_{\text{in}})$ to denote $\Phi_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$, $\phi_m(\mathcal{C}_{\text{in}})$ to denote $\phi_{i,m}(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$, $t(\mathcal{C}_{\text{in}})$ to denote $t(\vec{\mathcal{O}}, \psi, \mathcal{C}_{\text{in}})$, $V(\mathcal{C}_{\text{in}})$ to denote $V_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ and $V'(\mathcal{C}_{\text{in}})$ to denote $V'_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$.

The following notation will be used throughout this section. For $m \in [2^{Rn}]$, let $\mathbf{y}_m = \mathcal{C}_n(m) \oplus e$. For $k = 1, \dots, 2^{\rho n}$, let the notation j_k denote the the index $\text{int}(\mathbf{w}_k(m, e, \mathcal{C}_{\text{in}})) \in [2^{\rho n}]$.

For $m \in [2^{Rn}]$ and $(n, \rho n)$ codebook \mathcal{C}_{in} , the first two claims characterize how the i th closest codeword in codebook \mathcal{C}_{in} to received word \mathbf{y}_m (i.e., $\mathcal{C}_{\text{in}}(\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}))$) changes when the j_k^{th} codeword in \mathcal{C}_{in} is replaced with a word $\mathbf{z} \notin \mathcal{B}_{pn}(\mathbf{y}_m)$. The proof of these claims only require the definition (2) of word $\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}})$.

Claim 1. Let \mathcal{C}_{in} be an $(n, \rho n)$ codebook, $m \in [2^{Rn}]$, $\mathbf{z} \notin \mathcal{B}_{pn}(\mathbf{y}_m)$ and $k \in [2^{\rho n}]$. Let \mathcal{C}'_{in} denote the codebook $\mathcal{C}_{\text{in}}(j_k, \mathbf{z})$. If $\mathcal{C}_{\text{in}}(j_k) \notin \mathcal{B}_{pn}(\mathbf{y}_m)$ and

$$\mathcal{C}_{\text{in}}(\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}})) \notin \mathcal{B}_{pn}(\mathbf{y}_m),$$

then

$$\mathcal{C}'_{\text{in}}(\mathbf{w}_i(m, e, \mathcal{C}'_{\text{in}})) \notin \mathcal{B}_{pn}(\mathbf{y}_m). \quad (23)$$

Proof. The location of codewords $\mathcal{C}_{\text{in}}(\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}))$ and $\mathcal{C}'_{\text{in}}(\mathbf{w}_i(m, e, \mathcal{C}'_{\text{in}}))$ around \mathbf{y}_m are illustrated in Fig. 2. We

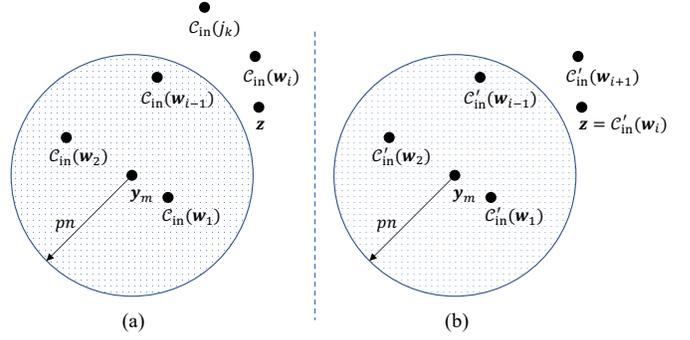


Fig. 2. Location of the i^{th} closest codeword to \mathbf{y}_m (a) before replacing codeword $\mathcal{C}_{\text{in}}(j_k)$ with word \mathbf{z} and (b) after replacing codeword $\mathcal{C}_{\text{in}}(j_k)$ with word \mathbf{z} . In this figure, the codebook \mathcal{C}'_{in} is equal to $\mathcal{C}_{\text{in}}(j_k, \mathbf{z})$.

begin by observing that $\mathcal{C}_{\text{in}}(\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}})) \notin \mathcal{B}_{pn}(\mathbf{y}_m)$ implies that $|\mathcal{C}_{\text{in}} \cap \mathcal{B}_{pn}(\mathbf{y}_m)| \leq i - 1$. Together with the fact that $\mathbf{z} \notin \mathcal{B}_{pn}(\mathbf{y}_m)$ and $\mathcal{C}_{\text{in}}(j_k) \notin \mathcal{B}_{pn}(\mathbf{y}_m)$, it follows that $|\mathcal{C}'_{\text{in}} \cap \mathcal{B}_{pn}(\mathbf{y}_m)| = |\mathcal{C}_{\text{in}} \cap \mathcal{B}_{pn}(\mathbf{y}_m)| \leq i - 1$. This implies equation (23). ■

Claim 2. Let \mathcal{C}_{in} be an $(n, \rho n)$ codebook, $m \in [2^{Rn}]$, $\mathbf{z} \in \{0, 1\}^n$ and $k \in \{i + 1, \dots, 2^{\rho n}\}$ such that $\mathbf{w}_k(m, e, \mathcal{C}_{\text{in}}) \neq \mathcal{C}_{\text{out}}(m)$. If

$$d(\mathbf{y}_m, \mathcal{C}_{\text{in}}(\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}))) < d(\mathbf{y}_m, \mathbf{z}),$$

then $\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}) = \mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}(j_k, \mathbf{z}))$.

Proof. The condition $\mathbf{w}_k(m, e, \mathcal{C}_{\text{in}}) \neq \mathcal{C}_{\text{out}}(m)$ ensures that $\mathbf{y}_m = \mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}(m) \oplus e$ is equal to $\mathcal{C}_{\text{in}}(j_k, \mathbf{z}) \circ \mathcal{C}_{\text{out}}(m) \oplus e$, and thus the center of the ball of radius pn around the received word does not change when the j_k^{th} codeword of \mathcal{C}_{in} is replaced with word \mathbf{z} . For $t = 1, \dots, i$, note that the t^{th} closest codeword in \mathcal{C}_{in} to \mathbf{y}_m (i.e., $\mathcal{C}_{\text{in}}(\mathbf{w}_t(m, e, \mathcal{C}_{\text{in}}))$) is at least as close to \mathbf{y}_m as codeword $\mathcal{C}_{\text{in}}(j_k)$, and is closer to \mathbf{y}_m than word \mathbf{z} . Therefore, by replacing the codeword $\mathcal{C}_{\text{in}}(j_k)$ with the word \mathbf{z} , we do not change the position of the t^{th} closest codeword in \mathcal{C}_{in} to word \mathbf{y}_m . Hence, $\mathbf{w}_t(m, e, \mathcal{C}_{\text{in}}) = \mathbf{w}_t(m, e, \mathcal{C}_{\text{in}}(j_k, \mathbf{z}))$. ■

For $m \in [2^{Rn}]$ and $(n, \rho n)$ codebook \mathcal{C}_{in} , the next two claims build upon the first two claims and characterize how the term $\phi_m(\mathcal{C}_{\text{in}})$ changes (and in turn, how the approximation function $q'(\mathcal{C}_{\text{in}})$ changes) when the j_k^{th} codeword in \mathcal{C}_{in} is replaced with a word $\mathbf{z} \notin \mathcal{B}_{pn}(\mathbf{y}_m)$. In the following section, these claims will help us in bounding the bounded difference $\Delta'(j, \mathbf{z}, \mathcal{C}_{\text{in}})$ and variation $V'(\mathcal{C}_{\text{in}})$.

Claim 3. Let \mathcal{C}_{in} be an $(n, \rho n)$ codebook, $m \in [2^{Rn}]$, $\mathbf{z} \notin \mathcal{B}_{pn}(\mathbf{y}_m)$ and $k \in [2^{\rho n}]$ such that $\mathbf{w}_k(m, e, \mathcal{C}_{\text{in}}) \neq \mathcal{C}_{\text{out}}(m)$. If either $k > i$ or $\mathcal{C}_{\text{in}}(j_k) \notin \mathcal{B}_{pn}(\mathbf{y}_m)$, then

$$\begin{aligned} & \mathbb{1}\{\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}) \in \mathcal{I}_m \cap \mathcal{L}_{\text{in}}(\mathbf{y}_m, \mathcal{C}_{\text{in}})\} \\ &= \mathbb{1}\{\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}(j_k, \mathbf{z})) \in \mathcal{I}_m \cap \mathcal{L}_{\text{in}}(\mathbf{y}_m, \mathcal{C}_{\text{in}}(j_k, \mathbf{z}))\}. \end{aligned} \quad (24)$$

Proof. We consider 2 cases depending on the distance of codeword $\mathcal{C}_{\text{in}}(\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}))$ from received word \mathbf{y}_m . (Case 1): Suppose that $\mathcal{C}_{\text{in}}(\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}})) \notin \mathcal{B}_{pn}(\mathbf{y}_m)$

Second term: Next, we bound term (27). Let the notation $j \in \mathcal{C}_{\text{out}}$ denote an index $j \in [2^{\rho n}]$ that belongs to the set $\{\mathbf{int}(\mathcal{C}_{\text{out}}(1)), \mathbf{int}(\mathcal{C}_{\text{out}}(2)), \dots, \mathbf{int}(\mathcal{C}_{\text{out}}(2^{Rn}))\}$. In term (27), we cut the summation over j by partitioning the set $\{j \in [2^{\rho n}] : \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi}\}$ into the sets $\{j \in [2^{\rho n}] : \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi}, j \notin \mathcal{C}_{\text{out}}\}$ and $\{j \in [2^{\rho n}] : \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi}, j \in \mathcal{C}_{\text{out}}\}$. Since $t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))$ is equal to $t(\mathcal{C}_{\text{in}})$ when $j \notin \mathcal{C}_{\text{out}}$, term (27) is equal to

$$\begin{aligned} & \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi} \text{ and } j \notin \mathcal{C}_{\text{out}}}} \sum_{\mathbf{z} \notin \mathcal{O}_{\psi}} \frac{|\Phi(\mathcal{C}_{\text{in}}) - \Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|}{t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))} 2^{-n} \\ & + \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi} \text{ and } j \in \mathcal{C}_{\text{out}}}} \sum_{\mathbf{z} \notin \mathcal{O}_{\psi}} \left| \frac{\Phi(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}})} - \frac{\Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))}{t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))} \right| 2^{-n}. \end{aligned} \quad (30)$$

To bound quantity (30), the following inequality will prove useful: defining $\Phi \triangleq \Phi(\mathcal{C}_{\text{in}})$, $\Phi' \triangleq \Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))$, $t \triangleq t(\mathcal{C}_{\text{in}})$ and $t' \triangleq t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))$, and using $|t - t'| \leq 1$, we have that

$$\begin{aligned} \left| \frac{\Phi(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}})} - \frac{\Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))}{t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))} \right| &= \begin{cases} \frac{\Phi}{t} - \frac{\Phi'}{t'}, & \frac{\Phi}{t} \geq \frac{\Phi'}{t'} \\ \frac{\Phi'}{t'} - \frac{\Phi}{t}, & \frac{\Phi}{t} < \frac{\Phi'}{t'} \end{cases} \\ &= \begin{cases} \frac{\Phi t' - \Phi' t}{t t'}, & \frac{\Phi}{t} \geq \frac{\Phi'}{t'} \\ \frac{\Phi' t - \Phi t'}{t t'}, & \frac{\Phi}{t} < \frac{\Phi'}{t'} \end{cases} \\ &\leq \begin{cases} \frac{\Phi(t+1) - \Phi' t}{t t'}, & \frac{\Phi}{t} \geq \frac{\Phi'}{t'} \\ \frac{\Phi' t - \Phi(t-1)}{t t'}, & \frac{\Phi}{t} < \frac{\Phi'}{t'} \end{cases} \\ &= \begin{cases} \frac{\Phi - \Phi'}{t'} + \frac{\Phi}{t t'}, & \frac{\Phi}{t} \geq \frac{\Phi'}{t'} \\ \frac{\Phi' - \Phi}{t'} + \frac{\Phi}{t t'}, & \frac{\Phi}{t} < \frac{\Phi'}{t'} \end{cases} \\ &= \frac{|\Phi(\mathcal{C}_{\text{in}}) - \Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|}{t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))} + \frac{\Phi(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}})t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))}. \end{aligned}$$

Following the above inequality, we have that (30) is bounded above by

$$\begin{aligned} & \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi} \text{ and } j \notin \mathcal{C}_{\text{out}}}} \sum_{\mathbf{z} \notin \mathcal{O}_{\psi}} \frac{|\Phi(\mathcal{C}_{\text{in}}) - \Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|}{t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))} 2^{-n} \\ & + \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi} \text{ and } j \in \mathcal{C}_{\text{out}}}} \sum_{\mathbf{z} \notin \mathcal{O}_{\psi}} \frac{|\Phi(\mathcal{C}_{\text{in}}) - \Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|}{t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))} 2^{-n} \\ & + \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi} \text{ and } j \in \mathcal{C}_{\text{out}}}} \sum_{\mathbf{z} \notin \mathcal{O}_{\psi}} \frac{\Phi(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}})t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))} 2^{-n} \end{aligned}$$

which in turn, following that $t(\mathcal{C}_{\text{in}}(j, \mathbf{z}))$ is bounded below by $t(\mathcal{C}_{\text{in}}) - 1$, is bounded above by

$$\begin{aligned} & \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi}}} \sum_{\mathbf{z} \notin \mathcal{O}_{\psi}} \frac{|\Phi(\mathcal{C}_{\text{in}}) - \Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|}{t(\mathcal{C}_{\text{in}}) - 1} 2^{-n} \\ & + \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi} \text{ and } j \in \mathcal{C}_{\text{out}}}} \sum_{\mathbf{z} \notin \mathcal{O}_{\psi}} \frac{\Phi(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}})(t(\mathcal{C}_{\text{in}}) - 1)} 2^{-n} \end{aligned} \quad (31)$$

Following that $\Phi(\mathcal{C}_{\text{in}}) \leq t(\mathcal{C}_{\text{in}})$ and $|\{j \in [2^{\rho n}] : \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi}, j \in \mathcal{C}_{\text{out}}\}| \leq t(\mathcal{C}_{\text{in}})$, we have that (31) is bounded above by

$$\sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi}}} \sum_{\mathbf{z} \notin \mathcal{O}_{\psi}} \frac{|\Phi(\mathcal{C}_{\text{in}}) - \Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|}{t(\mathcal{C}_{\text{in}}) - 1} 2^{-n} + \frac{t(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}}) - 1} \quad (32)$$

Similar to the bounding of term (26), we write Φ as a sum of ϕ_m terms, partition the set $\{\mathbf{z} \notin \mathcal{O}_{\psi}\}$ into $\{\mathbf{z} \in \mathcal{O}_{\psi}^c \cap \mathcal{B}_{pn}(\mathbf{y}_m)\}$ and $\{\mathbf{z} \in \mathcal{O}_{\psi}^c \cap \mathcal{B}_{pn}^c(\mathbf{y}_m)\}$, and apply inequalities $|\mathcal{B}_{pn}(\mathbf{y}_m)| \leq 2^{H(p)n}$, $|\mathcal{O}_{\psi} \cap \mathcal{C}_n| \leq t(\mathcal{C}_{\text{in}})$, and $\sum_{m: \mathcal{C}'_n(m) \in \mathcal{O}_{\psi}} \phi_m(\mathcal{C}_{\text{in}}(j, \mathbf{z})) = \sum_{m: \mathcal{C}'_n(m) \in \mathcal{O}_{\psi}} \phi_m(\mathcal{C}_{\text{in}}(j, \mathbf{z}))$ where $\mathcal{C}'_n = \mathcal{C}_{\text{in}}(j, \mathbf{z}) \circ \mathcal{C}_{\text{out}}$ when $j \in [2^{\rho n}] : \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi}$ and $\mathbf{z} \in \mathcal{O}_{\psi}^c$, to bound equation (32) above by

$$\begin{aligned} & \sum_{\substack{m \in [2^{Rn}]: \\ \mathcal{C}_{\text{in}}(m) \in \mathcal{O}_{\psi}}} \sum_{\substack{\mathbf{z} \in \mathcal{O}_{\psi}^c: \\ \mathbf{z} \in \mathcal{B}_{pn}^c(\mathbf{y}_m)}} \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{O}_{\psi}}} \frac{|\phi_m(\mathcal{C}_{\text{in}}) - \phi_m(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|}{(t(\mathcal{C}_{\text{in}}) - 1) 2^n} \\ & + \frac{2^{-\epsilon_{\rho n}}}{t(\mathcal{C}_{\text{in}}) - 1} + \frac{t(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}}) - 1} \end{aligned}$$

which in turn is bounded above by

$$\left(i + 1 + \frac{2^{-\epsilon_{\rho n}}}{t(\mathcal{C}_{\text{in}})} + 1 \right) \frac{t(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}}) - 1} \leq 2(i + 2) + 2^{-\epsilon_{\rho n} + 1}$$

following Lemma 7 and the inequalities $|\mathcal{O}_{\psi} \cap \mathcal{C}_n| \leq t(\mathcal{C}_{\text{in}})$, $t(\mathcal{C}_{\text{in}}) \geq 1$ and $\frac{t(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}}) - 1} \leq 2$.

Third term: Next, we bound term (28). Using a similar approach to the bounding of term (27), term (28) is bounded above by

$$\begin{aligned} & \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \notin \mathcal{O}_{\psi}}} \sum_{\mathbf{z} \in \mathcal{O}_{\psi}} \frac{|\Phi(\mathcal{C}_{\text{in}}) - \Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|}{t(\mathcal{C}_{\text{in}})} 2^{-n} \\ & + \sum_{\substack{j \in [2^{\rho n}]: \\ (\mathcal{C}_{\text{in}}(j) \notin \mathcal{O}_{\psi}) \cap (j \in \mathcal{C}_{\text{out}})}} \sum_{\mathbf{z} \in \mathcal{O}_{\psi}} \frac{\max\{\Phi(\mathcal{C}_{\text{in}}), \Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))\}}{t^2(\mathcal{C}_{\text{in}})} 2^{-n} \end{aligned} \quad (33)$$

which in turn is bounded above by

$$\begin{aligned} & \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \notin \mathcal{O}_{\psi}}} \sum_{\mathbf{z} \in \mathcal{O}_{\psi}} \sum_{\substack{m \in [2^{Rn}]: \\ \mathcal{C}_n(m) \in \mathcal{O}_{\psi}}} \frac{|\phi_m(\mathcal{C}_{\text{in}}) - \phi_m(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|}{t(\mathcal{C}_{\text{in}})} 2^{-n} \\ & + \sum_{\substack{j \in \mathcal{C}_{\text{out}}: \\ \mathcal{C}_{\text{in}}(j) \notin \mathcal{O}_{\psi}}} \sum_{\mathbf{z} \in \mathcal{O}_{\psi}} \left(\frac{\phi_{m_j}(\mathcal{C}_{\text{in}}(j, \mathbf{z}))}{t(\mathcal{C}_{\text{in}})} + \frac{\Phi_{\max}}{t^2(\mathcal{C}_{\text{in}})} \right) 2^{-n} \end{aligned} \quad (34)$$

where $\Phi_{\max} = \max\{\Phi(\mathcal{C}_{\text{in}}), \Phi(\mathcal{C}_{\text{in}}(j, \mathbf{z}))\}$ and where $m_j = (\mathcal{O}_{\psi} \cap \mathcal{C}_{\text{in}}(j, \mathbf{z}) \circ \mathcal{C}_{\text{out}}) \setminus (\mathcal{O}_{\psi} \cap \mathcal{C}_n)$ (if m_j is the empty set then we define $\phi_{m_j}(\mathcal{C}_{\text{in}}(j, \mathbf{z})) = 0$). In the above expression, we are already familiar with how to bound term (34); using the same approach used to bound (26), term (34) is bounded above by $i + 1 + 2^{-\epsilon_{\rho n}}$. Thus we only need to bound term (35). Since $\Phi_{\max} \leq t(\mathcal{C}_{\text{in}}) + 1$, term (35) is bounded above by $\sum_{j \in [2^{\rho n}]: \mathcal{C}_{\text{in}}(j) \notin \mathcal{O}_{\psi}, j \in \mathcal{C}_{\text{in}}} \sum_{\mathbf{z} \in \mathcal{O}_{\psi}} 3 \frac{2^{-n}}{t(\mathcal{C}_{\text{in}})}$. By $t(\mathcal{C}_{\text{in}}) \geq t_L$ and

the value of t_L given in Definition 3, this in turn is bounded above by

$$\sum_{\substack{j \in [2^{\rho n}] \\ \mathcal{C}_{\text{in}}(j) \notin \mathcal{O}_\psi \text{ and } j \in \mathcal{C}_{\text{out}}} } \sum_{z \in \mathcal{O}_\psi} \frac{3(2^{-n})}{2^{-(1-R)n} |\mathcal{O}_\psi| - 2^{\frac{3\delta'_0}{4}n}}. \quad (36)$$

Following the hypothesis of Lemma 6, for large enough n (which only depends on δ_0), $2^{-(1-R)n} |\mathcal{O}_\psi| - 2^{\frac{3\delta'_0}{4}n}$ is bounded above by $2^{-(1-R)n} |\mathcal{O}_\psi| (1/2)$. Hence, for large enough n , and by the inequality $|\{j \in [2^{\rho n}] : \mathcal{C}_{\text{in}}(j) \notin \mathcal{O}_\psi, j \in \mathcal{C}_{\text{out}}\}| \leq 2^{Rn}$, equation (36) is bounded above by 6. Hence, equation (33) is bounded above by $i + 7 + 2^{-\epsilon_\rho n}$.

Fourth term: Lastly, we bound term (29). Using the same approach to bound term (26), term (29) is bounded above by $i + 1 + 2^{-\epsilon_\rho n}$. The desired result follows by summing together the upper bounds of terms (26) through (29). \blacksquare

The following Lemma will help us find a small typical variation coefficient of q' . The Lemma states that $q'_i(\mathcal{C}_{\text{in}})$ is smooth for all $\mathcal{C}_{\text{in}} \in \mathcal{T}$ in a Lipschitz sense.

Lemma 8. *Define*

$$K_T = K_{T,i}(\vec{\mathcal{O}}, \psi, e) = \frac{2\ell + 3}{t_L - 1}. \quad (37)$$

If $\mathcal{C}_{\text{in}} \in \mathcal{T}$, then $q'(\mathcal{C}_{\text{in}})$ is K_T -Lipshitz, i.e., $\Delta'(j, z, \mathcal{C}_{\text{in}}) \leq K_T$ for all $j \in [2^{\rho n}]$, $z \in \{0, 1\}^n$.

Proof of Lemma 8. Let $\mathcal{C}_{\text{in}} \in \mathcal{T}$, $j \in [2^{\rho n}]$ and $z \in \{0, 1\}^n$. For $m \in [2^{Rn}]$, let $\mathbf{y}_m = \mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}(m) \oplus e$. We first count the number of messages $m \in [2^{Rn}]$ such that $\phi_m(\mathcal{C}_{\text{in}}) \neq \phi_m(\mathcal{C}_{\text{in}}(j, z))$. Since $\mathcal{C}_{\text{in}} \in \mathcal{T}$, \mathcal{C}_{in} is $[\ell, p]$ list decodable and the $(n, \rho n)$ codebook \mathcal{C}'_{in} resulting from a translation of \mathcal{C}_{in} by the vector e (i.e., $\mathcal{C}'_{\text{in}} = \{\mathcal{C}_{\text{in}}(1) \oplus e, \mathcal{C}_{\text{in}}(2) \oplus e, \dots, \mathcal{C}_{\text{in}}(2^{\rho n}) \oplus e\}$) is also $[\ell, p]$ list decodable. Hence, there exists at most 2ℓ messages $m_1, \dots, m_{2\ell}$ such that for $k = 1, \dots, 2\ell$, either $d(\mathbf{y}_{m_k}, \mathcal{C}_{\text{in}}(j)) \leq pn$ or $d(\mathbf{y}_{m_k}, z) \leq pn$. With this observation, we can state the following claim.

Claim 5. *For any message $m \in [2^{Rn}]$ that is not in the set $\{m_1, \dots, m_{2\ell}\}$,*

$$\begin{aligned} & \mathbb{1}\{\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}) \in \mathcal{I}_m \cap \mathcal{L}_{\text{in}}(\mathbf{y}_m, \mathcal{C}_{\text{in}})\} \\ &= \mathbb{1}\{\mathbf{w}_i(m, e, \mathcal{C}_{\text{in}}(j, z)) \in \mathcal{I}_m \cap \mathcal{L}_{\text{in}}(\mathbf{y}_m, \mathcal{C}_{\text{in}}(j, z))\}. \end{aligned} \quad (38)$$

We observe that Claim 5 is a special case of Claim 3. To see this, let $m \in [2^{Rn}] \setminus \{m_1, \dots, m_{2\ell}\}$ and $\mathbf{y}_m = \mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}(m) \oplus e$, and first observe that the ball $\mathcal{B}_{pn}(\mathbf{y}_m)$ contains neither $\mathcal{C}_{\text{in}}(j)$ nor z . Let $k \in [2^{\rho n}]$ such that $j = \mathbf{int}(\mathbf{w}_k(m, e, \mathcal{C}_{\text{in}}))$. Since $\mathcal{C}_{\text{in}}(j) \notin \mathcal{B}_{pn}(\mathbf{y}_m)$, it follows that $\mathbf{w}_k(m, e, \mathcal{C}_{\text{in}}) \neq \mathcal{C}_{\text{out}}(m)$. These conditions are sufficient to satisfy the hypothesis of Claim 3.

Following Claim 5, the number of messages $m \in [2^{Rn}]$ such that $\phi_m(\mathcal{C}_{\text{in}}) \neq \phi_m(\mathcal{C}_{\text{in}}(j, z))$ is bounded above by $2\ell + 2$ (where the 2 is added to account for the 2 possible messages $\{m'_1, m'_2\}$ such that for $k = 1, 2$, $\mathbb{1}\{\mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}(m'_k) \in \mathcal{O}_\psi\} \neq \mathbb{1}\{\mathcal{C}_{\text{in}}(j, z) \circ \mathcal{C}_{\text{out}}(m'_k) \in \mathcal{O}_\psi\}$, which in turn may result in $\phi_{m'_k}(\mathcal{C}_{\text{in}}) \neq \phi_{m'_k}(\mathcal{C}_{\text{in}}(j, z))$). From the triangle inequality, it follows that $|\Phi(\mathcal{C}_{\text{in}}) - \Phi(\mathcal{C}_{\text{in}}(j, z))| \leq 2\ell + 2$.

We are now ready to prove Lemma 8. The proof involves an upper bound of $\Delta'(j, z, \mathcal{C}_{\text{in}}) = \left| \frac{\Phi(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}})} - \frac{\Phi(\mathcal{C}_{\text{in}}(j, z))}{t(\mathcal{C}_{\text{in}}(j, z))} \right|$, which we illustrate by walking through the upper bound of the quantity $\frac{\Phi(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}})} - \frac{\Phi(\mathcal{C}_{\text{in}}(j, z))}{t(\mathcal{C}_{\text{in}}(j, z))}$; the upper bound of the negative of the above quantity follows the same approach. We have that

$$\begin{aligned} & \frac{\Phi(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}})} - \frac{\Phi(\mathcal{C}_{\text{in}}(j, z))}{t(\mathcal{C}_{\text{in}}(j, z))} \\ & \stackrel{(a)}{\leq} \frac{\Phi(\mathcal{C}_{\text{in}})(t(\mathcal{C}_{\text{in}}) + 1)}{t(\mathcal{C}_{\text{in}})t(\mathcal{C}_{\text{in}}(j, z))} - \frac{\Phi(\mathcal{C}_{\text{in}}(j, z))t(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}})t(\mathcal{C}_{\text{in}}(j, z))} \\ & \stackrel{(b)}{\leq} \frac{2\ell + 2}{t(\mathcal{C}_{\text{in}}(j, z))} + \frac{\Phi(\mathcal{C}_{\text{in}})}{t(\mathcal{C}_{\text{in}})t(\mathcal{C}_{\text{in}}(j, z))} \\ & \stackrel{(c)}{\leq} \frac{2\ell + 2}{t(\mathcal{C}_{\text{in}}(j, z))} + \frac{1}{t(\mathcal{C}_{\text{in}}(j, z))} \\ & \stackrel{(d)}{\leq} \frac{2\ell + 3}{t_L - 1} \end{aligned}$$

where inequality (a) follows from $t(\mathcal{C}_{\text{in}}(j, z)) \leq t(\mathcal{C}_{\text{in}}) + 1$, inequality (b) follows from $|\Phi(\mathcal{C}_{\text{in}}) - \Phi(\mathcal{C}_{\text{in}}(j, z))| \leq 2\ell + 2$ and inequality (c) follows from the inequalities $\Phi(\mathcal{C}_{\text{in}}) \leq |\mathcal{C}_n \cap \mathcal{O}_\psi| \leq \max\{|\mathcal{C}_n \cap \mathcal{O}_\psi|, t_L\} \triangleq t(\mathcal{C}_{\text{in}})$, and inequality (d) follows from $t(\mathcal{C}_{\text{in}}(j, z)) \geq t(\mathcal{C}_{\text{in}}) - 1$ and $t(\mathcal{C}_{\text{in}}) \geq t_L$. \blacksquare

An immediate corollary of the previous Lemma is that for all $\mathcal{C}_{\text{in}} \in \mathcal{T}$, $V'(\mathcal{C}_{\text{in}}) \leq 2^{\rho n} K_T^2$. In the following Lemma, this bound is tightened by exploiting the fact that the bounded difference $\Delta'(j, z, \mathcal{C}_{\text{in}})$ inside the definition of $V'(\mathcal{C}_{\text{in}})$ is often much smaller than the Lipschitz coefficient K_T for many $j \in [2^{\rho n}]$ and $z \in \{0, 1\}^n$.

Lemma 9 (Typical Variation Coefficient). *For $\mathcal{C}_{\text{in}} \in \mathcal{T}$,*

$$V'(\mathcal{C}_{\text{in}}) \leq \left(t_U(\ell + 1) + 2^{\delta'_0 n + \epsilon_R n} \right) K_T^2. \quad (39)$$

Proof of Lemma 9. To prove the upper bound on $V'(\mathcal{C}_{\text{in}})$ for $\mathcal{C}_{\text{in}} \in \mathcal{T}$, the proof uses two facts: (a) q'_i is K_T Lipschitz over \mathcal{T} and (b) difference $\Delta'(j, z, \mathcal{C}_{\text{in}})$ is zero for several pairs $(j, z) \in [2^{\rho n}] \times \{0, 1\}^n$. Fact (a) was established above in Lemma 8. We specify and establish Fact (b) below as the following claim.

For $m \in [2^{Rn}]$, define a set $\mathcal{S}_{1,m} = \{\mathbf{v} \in \mathcal{C}_{\text{in}} : \mathbf{v} \in \mathcal{B}_{pn}^c(\mathbf{y}_m)\}$ of all codewords outside the ball $\mathcal{B}_{pn}(\mathbf{y}_m)$ and define a set $\mathcal{S}_{2,m} = \{\mathbf{v} \in \{0, 1\}^n : \mathbf{v} \in \mathcal{B}_{pn}^c(\mathbf{y}_m) \text{ and } \mathbf{v} \in \mathcal{O}_\psi^c\}$ of all words outside both the ball $\mathcal{B}_{pn}(\mathbf{y}_m)$ and observation set \mathcal{O}_ψ . For $k = 1, 2$, define $\mathcal{S}_k = \bigcap_{m: \mathcal{C}_{\text{in}}(m) \in \mathcal{O}_\psi} \mathcal{S}_{k,m}$.

Claim 6 (Fact (b)). *For an $(n, \rho n)$ codebook \mathcal{C}_{in} , for any $j \in [2^{\rho n}]$ such that $\mathcal{C}_{\text{in}}(j) \in \mathcal{S}_1$ and for any $z \in \mathcal{S}_2$, $\Delta'(j, z, \mathcal{C}_{\text{in}}) = 0$.*

We first prove Claim 6, which is a special case of Claim 4. For an $(n, \rho n)$ codebook \mathcal{C}_{in} , let $j \in [2^{\rho n}]$ such that $\mathcal{C}_{\text{in}}(j) \in \mathcal{S}_1$ and let $z \in \mathcal{S}_2$. First, it is easy to verify that this choice of parameters satisfies the hypothesis of Claim 4. Second, note that $\mathcal{C}_{\text{in}}(j)$ is not in $\mathcal{O}_\psi \cap \mathcal{C}_n$ and z is not in \mathcal{O}_ψ . Hence, $\mathcal{O}_\psi \cap \mathcal{C}_n = \mathcal{O}_\psi \cap \mathcal{C}_{\text{in}}(j, z) \circ \mathcal{C}_{\text{out}}$ and thus $t(\mathcal{C}_{\text{in}}) = t(\mathcal{C}_{\text{in}}(j, z))$.

In turn, the difference $\Delta'(j, \mathbf{z}, \mathcal{C}_{\text{in}}) = |q'(\mathcal{C}_{\text{in}}) - q'(\mathcal{C}_{\text{in}}(j, \mathbf{z}))|$ is equal to

$$\left| \sum_{\substack{m \in [2^{Rn}]: \\ \mathcal{C}_n(m) \in \mathcal{O}_\psi}} \frac{\phi_m(\mathcal{C}_{\text{in}}) - \phi_m(\mathcal{C}_{\text{in}}(j, \mathbf{z}))}{t(\mathcal{C}_{\text{in}})} \right|. \quad (40)$$

Following Claim 4, $\phi_m(\mathcal{C}_{\text{in}}) = \phi_m(\mathcal{C}_{\text{in}}(j, \mathbf{z}))$ for all $m \in [2^{Rn}]$, and thus, (40) is zero. This completes the proof of Claim 6.

Let $\mathcal{C}_{\text{in}} \in \mathcal{T}$. Following Claim 6, $V'(\mathcal{C}_{\text{in}})$ is equal to

$$\sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{S}_1^c}} \mathbb{E}_{\mathbf{z}}[\Delta'(j, \mathbf{z}, \mathcal{C}_{\text{in}})^2] + \sum_{\substack{j \in [2^{\rho n}]: \\ \mathcal{C}_{\text{in}}(j) \in \mathcal{S}_1}} \sum_{\mathbf{z} \in \mathcal{S}_2^c} \frac{\Delta'(j, \mathbf{z}, \mathcal{C}_{\text{in}})^2}{2^n}. \quad (41)$$

To finish our proof of Lemma 9, we upper bound (41). Since $\mathcal{C}_{\text{in}} \in \mathcal{T}$, codebook \mathcal{C}_{in} is $[\ell, p]$ list decodable and we have that $|\mathcal{S}_1^c|$ is bounded above by $|\mathcal{O}_\psi \cap \mathcal{C}_n| \ell$. By a simple union bound, $|\mathcal{S}_2^c|$ is bounded above by $|\mathcal{O}_\psi \cap \mathcal{C}_n| 2^{H(p)n} + |\mathcal{O}_\psi|$ and $|\mathcal{S}_1| \leq 2^{\rho n}$. By Lemma 8, $\Delta'(j, \mathbf{z}, \mathcal{C}_{\text{in}}) \leq K_T$ for all $j \in [2^{\rho n}]$ and $\mathbf{z} \in \{0, 1\}^n$, and thus, equation (41) is upper bounded by $(|\mathcal{S}_1^c| + |\mathcal{S}_1| |\mathcal{S}_2^c| 2^{-n}) K_T^2$, which in turn, is upper bounded by

$$|\mathcal{O}_\psi \cap \mathcal{C}_n| \ell K_T^2 + \left(|\mathcal{O}_\psi \cap \mathcal{C}_n| 2^{(H(p)-1+\rho)n} + |\mathcal{O}_\psi| 2^{-(1-\rho)n} \right) K_T^2.$$

Finally, (39) follows by applying the bound $|\mathcal{O}_\psi \cap \mathcal{C}_n| \leq t_U$. ■

F. Concentration of q

The following Lemma shows that if $q'(\mathcal{C}_{\text{in}}) = q(\mathcal{C}_{\text{in}})$ w.h.p. over $Q(n, \rho n)$, then q' concentrated implies that q is concentrated.

Lemma 10. For any $\lambda > 0$,

$$\mathbb{P}_{\mathcal{C}_{\text{in}}}(q - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] > \lambda) \leq \mathbb{P}_{\mathcal{C}_{\text{in}}}(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] > \lambda) + \mathbb{P}_{\mathcal{C}_{\text{in}}}(q \neq q').$$

Proof of Lemma 10. Let \mathcal{C} denote the set of all $(n, \rho n)$ codebooks, and for $\mathcal{C}_{\text{in}} \in \mathcal{C}$ let $Q(\mathcal{C}_{\text{in}})$ denote the probability of drawing \mathcal{C}_{in} . We have that $\mathbb{P}_{\mathcal{C}_{\text{in}}}(q - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] > \lambda)$ is equal to

$$\begin{aligned} & \sum_{\mathcal{C}_{\text{in}} \in \mathcal{C}} \mathbb{1}\{q(\mathcal{C}_{\text{in}}) - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] > \lambda\} Q(\mathcal{C}_{\text{in}}) \\ &= \sum_{\substack{\mathcal{C}_{\text{in}} \in \mathcal{C}: \\ q(\mathcal{C}_{\text{in}}) = q'(\mathcal{C}_{\text{in}})}} \mathbb{1}\{q(\mathcal{C}_{\text{in}}) - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] > \lambda\} Q(\mathcal{C}_{\text{in}}) \\ & \quad + \sum_{\substack{\mathcal{C}_{\text{in}} \in \mathcal{C}: \\ q(\mathcal{C}_{\text{in}}) \neq q'(\mathcal{C}_{\text{in}})}} \mathbb{1}\{q(\mathcal{C}_{\text{in}}) - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] > \lambda\} Q(\mathcal{C}_{\text{in}}) \\ & \stackrel{(a)}{\leq} \sum_{\substack{\mathcal{C}_{\text{in}} \in \mathcal{C}: \\ q(\mathcal{C}_{\text{in}}) = q'(\mathcal{C}_{\text{in}})}} \mathbb{1}\{q'(\mathcal{C}_{\text{in}}) - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] > \lambda\} Q(\mathcal{C}_{\text{in}}) \\ & \quad + \sum_{\substack{\mathcal{C}_{\text{in}} \in \mathcal{C}: \\ q(\mathcal{C}_{\text{in}}) \neq q'(\mathcal{C}_{\text{in}})}} Q(\mathcal{C}_{\text{in}}) \\ & \leq \mathbb{P}_{\mathcal{C}_{\text{in}}}(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] > \lambda) + \mathbb{P}_{\mathcal{C}_{\text{in}}}(q \neq q') \end{aligned}$$

where inequality (a) follows from $\mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] \leq \mathbb{E}_{\mathcal{C}_{\text{in}}}[q]$. ■

We now state and prove our concentration inequality for q' , which follows from a modified logarithmic Sobolev inequality [24, Theorem 2].

Lemma 11. Suppose that $|\mathcal{O}_\psi| \geq 2^{(1-R)n} 2^{\delta_0 n}$, let $i \in [L]$ and suppose that the following Assumptions hold:

- 1) For large enough n (depending only on δ_0, ϵ_ρ and L), there exists a global variation coefficient $a_G \in (0, \infty)$ such that $V'(\mathcal{C}_{\text{in}}) \leq a_G$ for all $(n, \rho n)$ codebooks \mathcal{C}_{in} .
- 2) There exists a typical variation coefficient $a_T \in (0, \min\{1, a_G\})$ such that $V'(\mathcal{C}_{\text{in}}) \leq a_T$ for all $\mathcal{C}_{\text{in}} \in \mathcal{T}$.
- 3) As a sequence in n , the ratio $\frac{a_G}{a_T}$ is $o(-\ln \mathbb{P}(\mathcal{C}_{\text{in}} \notin \mathcal{T}))$.

Then for $\lambda \in (\sqrt{a_T}, 1)$ and for large enough n (depending only on $\delta'_0, \epsilon_\rho, L$),

$$\mathbb{P}_{\mathcal{C}_{\text{in}}}(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] > \lambda) \leq \exp\left\{-\frac{\lambda^2}{8a_T}\right\}. \quad (42)$$

Proof of Lemma 11. The proof follows a conventional “entropy-method” proof for deriving concentration inequalities [24]. A slight modification of the conventional proof is needed to incorporate the typical and global variation coefficients and prevent the inequality from blowing up over a small set of $(n, \rho n)$ codebooks. We begin by restating a modified logarithmic Sobolev inequality in a general form.

Lemma 12 ([24, Theorem 2]). Suppose that X_1, X_2, \dots, X_n are independent random variables taking values in a measurable set \mathcal{X} , and define $Z = g(X_1, \dots, X_n)$ for some given measurable function $g : \mathcal{X}^n \rightarrow \mathbb{R}$. Furthermore, suppose that X'_1, \dots, X'_n are independent copies of X_1, \dots, X_n and define both $Z(j) = g(X_1, \dots, X_{j-1}, X'_j, X_{j+1}, \dots, X_n)$ and

$$V_+ = \sum_{j=1}^n \mathbb{E}_{X'_j} \left[(Z - Z(j))^2 \mathbb{1}\{Z > Z(j)\} \right].$$

Then for $\theta > 0$ and $\mu \in (0, 1/\theta)$,

$$\ln \mathbb{E} \left[e^{\mu(Z - \mathbb{E}[Z])} \right] \leq \frac{\mu\theta}{1 - \mu\theta} \ln \mathbb{E} \left[e^{\frac{\mu V_+}{\theta}} \right].$$

Next, we apply the framework of Lemma 12 to our coding problem. Recall that the $2^{\rho n}$ codewords of \mathcal{C}_{in} are independent random variables that take values in $\mathcal{X} = \{0, 1\}^n$. Setting $g = q'$, it follows that $(Z - Z(j))^2 = \Delta(j, \mathbf{z}, \mathcal{C}_{\text{in}})^2$, and in turn,

$$V_+ = \sum_{j=1}^{2^{\rho n}} \mathbb{E}_{\mathbf{z}} \left[\Delta(j, \mathbf{z}, \mathcal{C}_{\text{in}})^2 \mathbb{1}\{q'(\mathcal{C}_{\text{in}}) > q'(\mathcal{C}_{\text{in}}(j, \mathbf{z}))\} \right].$$

Note that $V_+ \leq V'$ following the indicator bound $\mathbb{1}\{\cdot\} \leq 1$ and the inequalities $\Delta(j, \mathbf{z}, \mathcal{C}_{\text{in}})^2 \leq \Delta(j, \mathbf{z}, \mathcal{C}_{\text{in}}) \leq 1$. Thus, Lemma 12 implies that for $\theta > 0$ and $\mu \in (0, 1/\theta)$

$$\ln \mathbb{E}_{\mathcal{C}_{\text{in}}} \left[e^{\mu(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'])} \right] \leq \frac{\mu\theta}{1 - \mu\theta} \ln \mathbb{E}_{\mathcal{C}_{\text{in}}} \left[e^{\frac{\mu V'}{\theta}} \right]. \quad (43)$$

By Bayes formula and Assumptions 1 and 2, the expectation in the RHS of (43) is bounded such that

$$\begin{aligned} \mathbb{E}_{\mathcal{C}_{\text{in}}} \left[e^{\frac{\mu V'}{\theta}} \right] & \leq e^{\frac{\mu a_T}{\theta}} \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \in \mathcal{T}) + e^{\frac{\mu a_G}{\theta}} \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{T}) \\ & \leq e^{\frac{\mu a_T}{\theta}} + e^{\frac{\mu a_G}{\theta}} \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{T}) \end{aligned} \quad (44)$$

We set θ such that the two terms in the sum of (44) are equal, i.e., set $\theta = \frac{(a_G - a_T)\mu}{-\ln \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{T})}$, and note that $\theta > 0$ following $a_G > a_T > 0$. Given this choice of θ , it follows from (44) that $\mathbb{E}_{\mathcal{C}_{\text{in}}} \left[\exp\left\{\frac{\mu V'}{\theta}\right\} \right] \leq \exp\{\mu a_T / \theta + \ln(2)\}$, and in turn, following (43),

$$\mathbb{E}_{\mathcal{C}_{\text{in}}} \left[e^{\mu(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'])} \right] \leq \exp \left\{ \frac{\mu^2 a_T}{1 - \mu\theta} + \frac{\mu\theta \ln 2}{1 - \mu\theta} \right\}.$$

Applying Markov's inequality to the above inequality yields

$$\mathbb{P}(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] > \lambda) \leq \exp \left\{ \frac{\mu^2 a_T}{1 - \mu\theta} + \frac{\mu\theta \ln 2}{1 - \mu\theta} - \mu\lambda \right\}. \quad (45)$$

To finish the proof, we choose some round numbers to simplify the RHS of (45). We set $\mu = \lambda / (2a_T)$, which is the value that minimizes the RHS of (45) over the optimization variable $\mu \in (0, \infty)$ when θ is treated as a constant fixed at 0. Note that this choice of μ satisfies for large enough n the requirement of Lemma 12 that $\mu \in (0, 1/\theta)$ since the quantity

$$\mu\theta = \frac{\lambda^2}{4a_T^2} \left(\frac{a_G - a_T}{-\ln \mathbb{P}(\mathcal{C}_{\text{in}} \notin \mathcal{T})} \right) \rightarrow 0 \text{ as } n \rightarrow \infty$$

where the limit follows from the inequalities $\lambda \in [0, 1]$ and $a_T < a_G$, and Assumption 3 which states that $\frac{a_G}{-a_T^2 \ln \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{T})}$ tends to 0 as n tends to ∞ . Following a substitution of our choices of θ and μ , the RHS of (45) is equal to

$$\exp \left\{ \frac{\lambda^2}{4a_T} \left(\frac{1}{1 - o(1)} \right) + \frac{o(1)}{1 - o(1)} - \frac{\lambda^2}{2a_T} \right\} \quad (46)$$

which in turn, is bounded above by $\exp \left\{ -\frac{\lambda^2}{8a_T} \right\}$ for large enough n following the condition $\lambda > \sqrt{a_T}$. Together with (45), this yields the desired inequality. ■

The following concentration inequality for q is bootstrapped from the above concentration inequality for q' using Lemma 10.

Lemma 13. *If ϵ_R satisfies Condition 2, $|\mathcal{O}_\psi| \geq 2^{(1-R)n} 2^{\delta_0 n}$ and $i \in [L]$, then for large enough n (depending only on δ_0 , L , and ϵ_ρ),*

$$\mathbb{P}_{\mathcal{C}_{\text{in}}} (q - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] > 1/n) \leq 2 \exp \left\{ -\frac{2^{\frac{\delta_0 n}{30}}}{8n^2} \right\}.$$

Proof of Lemma 13. The proof is a straightforward application of Lemma 6 through Lemma 11, but requires a little accounting to ensure that the chosen parameters check out. Recall that $\delta'_0 \geq \delta_0$ is the unique constant such that $|\mathcal{O}_\psi| = 2^{(1-R)n} 2^{\delta'_0 n}$.

We first bound the probability that \mathcal{C}_{in} is not typical. Note that $\mathbb{E}_{\mathcal{C}_{\text{in}}} |\mathcal{O}_\psi \cap \mathcal{C}_n| = 2^{-(1-R)n} |\mathcal{O}_\psi| = 2^{\delta_0 n}$. Recall from Definition 3 that $\ell = 2^{\frac{4}{13}\delta_0 n}$, $t_U = 2^{\delta_0 n} + 2^{\frac{3}{4}\delta_0 n}$ and $t_L = 2^{\delta_0 n} - 2^{\frac{3}{4}\delta_0 n}$. It follows from Lemma 1 and Lemma 3 that $\mathbb{P}(q' \neq q) \leq \mathbb{P}(\mathcal{C}_{\text{in}} \notin \mathcal{T}) \leq 2^{-2^{\frac{\delta_0 n}{13}}}$ and $\mathbb{P}(|\mathcal{O}_\psi \cap \mathcal{C}_n| < t_L) \leq 2 \exp\{-\frac{2^{\delta_0 n/2}}{4}\}$ for large enough n (depending only on δ_0).

Define a_T^{LB} as the RHS of (39). Next, we bound K_T (defined by equation (37) and a_T^{LB}). Recall that K_T is equal to $\frac{2\ell+1}{t_L-1}$. From substitution of the typical parameters, K_T is

equal to $2^{-\frac{9}{13}\delta'_0 n} + 2$ for large enough n (depending only on δ_0). Similarly, a_T^{LB} is equal to $2^{-\delta'_0 n/13+6} + 2^{-\frac{5}{13}\delta'_0 n + \epsilon_R n + 4}$ for large enough n (depending only on δ_0 and ϵ_R). Hence, for large enough n (depending only on δ_0 and ϵ_R) we can choose any a_T such that

$$a_T \geq a_T^{LB} = 2^{-\delta'_0 n/13+6} + 2^{-\frac{5}{13}\delta'_0 n + \epsilon_R n + 4} \quad (47)$$

and have that $V(\mathcal{C}_{\text{in}}) \leq a_T$ for all $\mathcal{C}_{\text{in}} \in \mathcal{T}$.

Finally, we are ready to apply Lemma 11. We first check that Assumption 3 of Lemma 11 holds. We have that

$$\frac{a_G}{a_T^2 (-\ln \mathbb{P}(\mathcal{C}_{\text{in}} \in \mathcal{T}))} < \frac{a_G}{a_T^2 2^{\frac{\delta'_0 n}{13}}} \quad (48)$$

for large enough n (depending only on δ_0 and ϵ_ρ). Set $a_T = 2^{-\frac{\delta'_0 n}{30}}$; this choice of a_T is possible under Condition 2 and satisfies equation (47). Following Lemma 6, a_G is bounded above by $5L + 14$ for large enough n (depending only on δ_0 , ϵ_ρ , and L). In turn, using $\delta'_0 \geq \delta_0$, the RHS of quantity (48) is bounded above by

$$\frac{5L + 14}{2^{\frac{2\delta_0 n}{195}}}$$

for large enough n (depending on δ_0 , ϵ_ρ and L), and therefore, is $o(1)$ and Assumption 3 holds.

To complete the proof, we apply Lemma 10 to bound the quantity $\mathbb{P}(q - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] > 1/n)$ above by $\mathbb{P}(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] > 1/n) + \mathbb{P}(q' \neq q)$. Lemma 13 follows after applying Lemma 11 to bound $\mathbb{P}(q' - \mathbb{E}_{\mathcal{C}_{\text{in}}}[q'] > 1/n)$ above by $\exp\{-2^{\delta_0 n/30}/(8n^2)\}$ for large enough n (depending only on δ_0 , ϵ_ρ and L), and after observing that $\mathbb{P}(q' \neq q)$ is bounded above by $\mathbb{P}(|\mathcal{O}_\psi \cap \mathcal{C}_n| < t_L) \leq \exp\{-2^{\delta_0 n/30}/(8n^2)\}$. ■

G. Proof of Theorem 1

We are now ready to prove Theorem 1. Let $L > 1/\epsilon_\rho$ be an integer. Our strategy is to apply the sufficient condition (Lemma 4) for the rate R to be (c, s) -achievable. Recall that for one to apply Lemma 4, one must show that for any $\epsilon_e > 0$ and for large enough n , $\mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{H}(L, \epsilon_e)) < 1 - 1/n$. To show this, we construct a set \mathcal{G} of good $(n, \rho n)$ codebooks such that \mathcal{G} is contained in the set $\mathcal{H}(L, \epsilon_e)$ and $\lim_{n \rightarrow \infty} \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{G}) = 0$, and conclude that for large enough n , $\mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{H}(L, \epsilon_e)) \leq \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{G}) < 1 - 1/n$.

For integer $n \geq 1$, we define the set of good $(n, \rho n)$ codebooks using the following sets: Let \mathcal{E}_1 be the set of $(n, \rho n)$ codebooks \mathcal{C}_{in} where there exists some $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(L) : |\mathcal{O}_\psi| \geq 2^{(1-R)n} 2^{\delta_0 n}$ such that $q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})$ is not concentrated, i.e., $q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}) > \mathbb{E}_{\mathcal{C}_{\text{in}}}[q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})] + 1/n$. Let \mathcal{E}_2 be the set of $(n, \rho n)$ codebooks \mathcal{C}_{in} where some small observation set is not typical, i.e., there exists some $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(L) : |\mathcal{O}_\psi| < 2^{(1-R)n} 2^{\delta_0 n}$ such that $|\mathcal{O}_\psi \cap \mathcal{C}_n| > 2^{(\delta_0 + \delta_1)n}$. Finally, let $\mathcal{G} = (\mathcal{E}_1 \cup \mathcal{E}_2)^c$ denote the set of good $(n, \rho n)$ codebooks. We say that an $(n, \rho n)$ codebook \mathcal{C}_{in} is not good if \mathcal{C}_{in} is not in \mathcal{G} . To see that $\mathcal{G} \subseteq \mathcal{H}(L, \epsilon_e)$ for large enough n , we first observe that by Lemma 5 and for large enough n ,

$$\max_{(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(L)} \mathbb{E}_{\mathcal{C}_{\text{in}}}[q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}})] + 1/n \leq \frac{\epsilon_e}{2L}. \quad (49)$$

Then for large enough n such that (49) holds, $\mathcal{C}_{\text{in}} \in \mathcal{G}$ implies that for all $(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(L)$,

$$q_i(\vec{\mathcal{O}}, \psi, e, \mathcal{C}_{\text{in}}) \leq \frac{\epsilon_e}{2L}, \text{ if } |\mathcal{O}_\psi| \leq 2^{(1-R)n} 2^{\delta_0 n} \quad (50)$$

and

$$|\mathcal{O}_\psi \cap \mathcal{C}_n| \leq 2^{(\delta_0 + \delta_1)n}, \text{ if } |\mathcal{O}_\psi| > 2^{(1-R)n} 2^{\delta_0 n}. \quad (51)$$

Following (51), $\mathcal{C}_{\text{in}} \in \mathcal{G}$ implies that for all $(i, \mathcal{O}, \vec{\psi}, e) \in \mathcal{P}(L)$,

$$\begin{aligned} \mathbb{P}_{m_0}(\Psi(m_0) = \psi) &= \mathbb{P}_{m_0}(\mathcal{C}_n(m_0) \in \mathcal{O}_\psi) \\ &= |\mathcal{O}_\psi \cap \mathcal{C}_n| 2^{-Rn} \\ &\leq 2^{(\delta_0 + \delta_1 - R)n}, \text{ if } |\mathcal{O}_\psi| > 2^{(1-R)n} 2^{\delta_0 n}. \end{aligned} \quad (52)$$

Using the definition of set $\mathcal{H}(L, \epsilon_e)$, it is easy to verify that for any $(n, \rho n)$ codebook \mathcal{C}_{in} such that both (50) and (52) hold, we have that $\mathcal{C}_{\text{in}} \in \mathcal{H}(L, \epsilon_e)$.

We now bound the probability that \mathcal{C}_{in} is not good by bounding $\mathbb{P}(\mathcal{E}_1)$ and $\mathbb{P}(\mathcal{E}_2)$. The adversary's computational bound will help us to bound both $\mathbb{P}(\mathcal{E}_1)$ and $\mathbb{P}(\mathcal{E}_2)$. Let S denote the number of unique observation sets in $\text{CKT}(r, cn^s)$, i.e., $S = |\{\mathcal{O} \subseteq \{0, 1\}^n : \vec{\mathcal{O}} \in \text{CKT}(r, cn^s), \mathcal{O} = \mathcal{O}_\psi \text{ for some } \psi \in \{0, 1\}^{rn}\}|$. We can bound S by counting the number of Boolean circuits with cn^s logic gates.

Lemma 14. *For large enough n (depending only on c and s), the number of functions in $\text{CKT}(r, cn^s)$ is bounded above by $2^{n^{s+2}}$, and thus $S = S(r, cn^s) \leq 2^{n^{s+3}}$. Proof is in Appendix E.*

For large enough n (depending only on $\delta_0, L, \epsilon_e, c$ and s),

$$\begin{aligned} \mathbb{P}(\mathcal{E}_1) &= \mathbb{P}_{\mathcal{C}_{\text{in}}} \left(\bigcup_{\substack{(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(L): \\ |\mathcal{O}_\psi| \geq 2^{(1-R+\delta_0)n}}} \{q > \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] + 1/n\} \right) \\ &\leq \sum_{\substack{(i, \vec{\mathcal{O}}, \psi, e) \in \mathcal{P}(L): \\ |\mathcal{O}_\psi| \geq 2^{(1-R+\delta_0)n}}} \mathbb{P}_{\mathcal{C}_{\text{in}}}(q > \mathbb{E}_{\mathcal{C}_{\text{in}}}[q] + 1/n) \end{aligned} \quad (53)$$

$$\leq S 2^n L 2 \exp \left\{ -\frac{2^{\delta_0 n/30}}{8n^2} \right\} \quad (54)$$

where (53) follows from a simple union bound and (54) follows from Lemma 13. Furthermore, for large enough n (depending only on δ_0, δ_1, c and s),

$$\mathbb{P}(\mathcal{E}_2) \leq S 2 \exp \{2^{-\delta_0 n}\}$$

which follows from a simple union bound and Lemma 3. In turn, given the bound on S established in Lemma 14, it is clear that both $\mathbb{P}(\mathcal{E}_1)$ and $\mathbb{P}(\mathcal{E}_2)$ are going to 0 in the limit as $n \rightarrow \infty$. Hence, for $\epsilon_e > 0$ and for large enough n , $\mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{H}(L, \epsilon_e)) \leq \mathbb{P}_{\mathcal{C}_{\text{in}}}(\mathcal{C}_{\text{in}} \notin \mathcal{G}) \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2) < 1 - 1/n$. This completes the proof of Theorem 1.

V. CONCLUSION

In this work, we define and study a binary channel controlled by a $\text{CKT}(r, cn^s)$ -oblivious adversary (an adversary that can observe a fraction r of all bits in the transmitted codeword via some function $f \in \text{CKT}(r, cn^s)$ of bounded complexity and flip a fraction p of all bits). The capacity $C(p, r, c, s)$ of this channel is characterized for the parameter range $r < 1 - H(p)$ (i.e., a sufficiently myopic adversary) under deterministic codes and average error criterion. We give a proof of this result which is based on a new application logarithmic Sobolev inequalities.

An alternative proof of the above result can be stated using the proof techniques for sufficiently myopic channels developed by Dey, Jaggi and Langberg [7]. The advantage of the alternative proof is that it uses a simpler random coding scheme, involves a simpler analysis, and can provide more general results than the proof of Section IV. An outline of this alternative proof is provided in Appendix A.

Lastly, we remark that a $\text{CKT}(r, cn^s)$ -oblivious adversary can be strictly less powerful than a $\text{CKT}(r, \infty)$ -oblivious adversary (i.e., an adversary with no complexity constraint), in the sense that $C(p, r, \infty, \infty)$ is strictly less than $C(p, r, c, s)$ for some values of $p \in (0, 1/2)$ and $r < 1 - H(p)$. A proof sketch is as follows. If no complexity constraint is imposed, then the adversary can choose a function f (dependent on the codebook \mathcal{C}_n) that does the following:

- 1) Take the transmitted codeword x as input. Compute the nearest codeword x' to x and, in turn, compute an "auxiliary" error vector $s \in \{0, 1\}^n$ such that $x \oplus s$ is equal Hamming distance to both x and x' .
- 2) Let $w(s)$ denote the Hamming weight of s . If $w(s)$ is small enough such that the total number of length- n binary vectors of weight $w(s)$ or less (call this number $A_{w(s)}$) is at most 2^{rn} , then "compress" s into an rn bit vector and output this compressed vector. Otherwise, output an error.

Let $\text{LP}(\delta)$ denote the linear programming bound for binary codes with minimum distance δn , and let LP^{-1} denote its inverse. By the linear programming bound (i.e., MRRW bound) [3], x and x' are (with positive probability) within $n \text{LP}^{-1}(R)$ bits where R is the rate, and thus, $w(s)$ is no more than about $\frac{n}{2} \text{LP}^{-1}(R)$ and $A_{w(s)}$ is no more than about $2^{nH(\frac{1}{2} \text{LP}^{-1}(R))}$. Hence, if $r > H(\frac{1}{2} \text{LP}^{-1}(R))$ and $p > \frac{1}{2} \text{LP}^{-1}(R)$, with positive probability, the adversary can reconstruct s from the output of f and, in turn, choose the true error vector $e = s$ to confuse Alice as to whether x or x' was transmitted. As can be verified numerically, the above adversarial strategy can be used to upper bound $C(p, r, \infty, \infty)$ and show that $C(p, r, \infty, \infty) < C(p, r, c, s)$ for a certain range of $r < 1 - H(p)$.

APPENDIX A

ALTERNATIVE PROOF OF THEOREM 1

In this appendix, we present an outline of an alternative proof of Theorem 1. The proof closely follows the achievability proof of Dey, Jaggi and Langberg [7, Theorem III.1] for sufficiently myopic channels. To follow this alternative proof,

we point the reader to this reference, and structure our outline to emphasize the difference between the alternative proof and the proof of [7, Theorem III.1]. For encoding, we replace our concatenated code construction with a simple random code where the codewords of code C_n are i.i.d. uniform in $\{0, 1\}^n$. For decoding, use the Hamming ball decoder as in [7].

- Modify [7, Lemma IV.2] such that for any function $f : \{0, 1\}^n \rightarrow \{0, 1\}^{rn}$ (not necessarily with polynomial circuit complexity) and any rn -bit observation vector ψ , the probability (over random code design) that there are fewer than about $|\mathcal{O}_\psi|2^{(R-1)n}$ codewords compatible with the output of the function f is $\exp\{-2^{\Omega(n)}\}$.
- In [7, Lemma IV.3], instead of analyzing the event that the number of codewords in a set $\mathcal{V} \subseteq \{0, 1\}^n$ exceeds $\Theta(n^2)$, show that this number exceeds $\Theta(n^{s+4})$ with probability $2^{-\Omega(n^{s+4})}$ over random code design.
- In [7, Corollary IV.4], the n^2 is replaced by n^{s+4} , and in [7, Lemma IV.5] the n^4 is replaced by $n^{2(s+4)}$.
- The subsequent arguments in [7, Lemma IV.6] are similarly modified, with n^2 being replaced by n^{s+4} .
- In each of these Lemmas, instead of union bounding over all error vectors (numbering $2^{O(n)}$), one union bounds over all error vectors and circuits in $\text{CKT}(r, cn^s)$ (numbering $O(n^{s+3})$ following Lemma 14).
- In the analysis, allow decoding to fail over small observation sets as described in Section IV-C of this paper. In the event that Alice's transmitted codeword belongs to a small set, this means that adversary has high certainty of Alice's codeword/message upon observing ψ , and thus, may be able design e well-tailored for this codeword/message and induce a decoding error. However, we can ignore this event in the analysis, since such an event is unlikely and thus makes a negligible contribution to the probability of error.

APPENDIX B

A TALAGRAND-TYPE CONCENTRATION INEQUALITY

Let $g(\cdot)$ be a function mapping the set of $(n, \rho n)$ codebooks to $(-\infty, \infty)$. For $b > 0$, g is said to be b -Lipshitz if for any $(n, \rho n)$ codebooks \mathcal{C}_{in} and \mathcal{C}'_{in} differing by at most 1 codeword, then $|g(\mathcal{C}_{\text{in}}) - g(\mathcal{C}'_{\text{in}})| \leq b$. An index set $J(\cdot) \subseteq [2^{\rho n}]$ is said to be a *certificate* of g if for any $(n, \rho n)$ codebook \mathcal{C}_{in} , $g(\mathcal{C}_{\text{in}}) \geq |J(\mathcal{C}_{\text{in}})|$ and $g(\mathcal{C}'_{\text{in}}) \geq g(\mathcal{C}_{\text{in}})$ for any \mathcal{C}'_{in} that agrees with \mathcal{C}_{in} on the codewords indexed in $J(\mathcal{C}_{\text{in}})$. Lastly, for $c > 0$, g is said to be c -certifiable if there exists a certificate J of g such that $|J(\mathcal{C}_{\text{in}})| \leq cg(\mathcal{C}_{\text{in}})$ for all $(n, \rho n)$ codebooks \mathcal{C}_{in} .

Lemma 15 ([26, Theorem 11.3]). *Let $\mathbb{M}[g]$ denote a median of g . For any $t > 0$,*

$$\mathbb{P}_{\mathcal{C}_{\text{in}}}(g - \mathbb{M}[g] > t) \leq 2 \exp \left\{ \frac{-t^2}{4b^2c(\mathbb{M}[g] + t)} \right\}$$

and

$$\mathbb{P}_{\mathcal{C}_{\text{in}}}(g - \mathbb{M}[g] < -t) \leq 2 \exp \left\{ \frac{-t^2}{4b^2c\mathbb{M}[g]} \right\}.$$

APPENDIX C

PROOF OF LEMMA 1

For $\mathbf{y} \in \{0, 1\}^n$, define $g_{\mathbf{y}}(\mathcal{C}_{\text{in}}) = |\mathcal{C}_{\text{in}} \cap \mathcal{B}_{pn}(\mathbf{y})|$. Our goal is to show that $g_{\mathbf{y}}$ is strongly concentrated around its expectation $\mathbb{E}_{\mathcal{C}_{\text{in}}}[g_{\mathbf{y}}]$. Note the following: $g_{\mathbf{y}}$ is 1-Lipschitz and $J(\mathcal{C}_{\text{in}}) = \{k \in [2^{\rho n}] : \mathcal{C}_{\text{in}}(k) \in \mathcal{B}_{pn}(\mathbf{y})\}$ is a certificate of $g_{\mathbf{y}}(\mathcal{C}_{\text{in}})$ where it follows that $g_{\mathbf{y}}$ is 1-certifiable.

Since $g_{\mathbf{y}}(\mathcal{C}_{\text{in}})$ is a binomial random variable, the value $\text{floor}(\mathbb{E}_{\mathcal{C}_{\text{in}}}[g_{\mathbf{y}}])$ is a median. Set $\mathbb{M}[g_{\mathbf{y}}] = \text{floor}(\mathbb{E}_{\mathcal{C}_{\text{in}}}[g_{\mathbf{y}}])$. Note that $R < 1 - H(p)$ implies that $\mathbb{E}_{\mathcal{C}_{\text{in}}}[g_{\mathbf{y}}]$ (which is equal to $\sum_{i=1}^{2^{\rho n}} \mathbb{P}_{\mathcal{C}_{\text{in}}}(x_i \in \mathcal{B}_{pn}(\mathbf{y})) \leq 2^{(R-1+H(p))n}$) is going to zero in n . It follows that for large enough n , $\mathbb{M}[g_{\mathbf{y}}] = 0$.

By Lemma 15, for $\ell > 0$ the probability that $g_{\mathbf{y}} > \ell$ is at most $2^{-\log(e)\frac{\ell}{4}+1}$. In conclusion, $\mathbb{P}_{\mathcal{C}_{\text{in}}}(\exists \mathbf{y} \in \{0, 1\}^n \text{ s.t. } g_{\mathbf{y}}(\mathcal{C}_{\text{in}}) > \ell) = \mathbb{P}_{\mathcal{C}_{\text{in}}}(\cup_{\mathbf{y} \in \{0, 1\}^n} \{g_{\mathbf{y}}(\mathcal{C}_{\text{in}}) > \ell\}) < 2^n 2^{-\log(e)\frac{\ell}{4}+1}$.

APPENDIX D

PROOF OF LEMMA 3

Define $g(\mathcal{C}_{\text{in}}) = |\mathcal{A} \cap \mathcal{C}_n|$. Note the following: $g(\cdot)$ is 1-Lipshitz and $J(\mathcal{C}_{\text{in}}) = \{m \in [2^{Rn}] : \mathcal{C}_{\text{in}} \circ \mathcal{C}_{\text{out}}(m) \in \mathcal{A}\}$ is a certificate of $g(\mathcal{C}_{\text{in}})$ where it follows that $g(\mathcal{C}_{\text{in}})$ is 1-certifiable.

Since $g(\mathcal{C}_{\text{in}})$ is a binomial random variable, the expected value $\mathbb{E}_{\mathcal{C}_{\text{in}}}[g]$ is a median. Set $\mathbb{M}[g] = \mathbb{E}_{\mathcal{C}_{\text{in}}}[g] = 2^{-(1-R)n}|\mathcal{A}|$. The desired result follows from Lemma 15.

APPENDIX E

PROOF OF LEMMA 14

Let W be the number of functions of the form $g_n : \{0, 1\}^n \rightarrow \{0, 1\}$ that can be computed by a Boolean circuit (of n inputs and 1 output) of size cn^s . We first show that $W < 2^{2(s+1)n^{s+1}}$.

Note that each gate can compute one of 16 unique functions from $\{0, 1\}^2$ to $\{0, 1\}$. Furthermore, for a given gate, the number of ways to choose 2 gate inputs from n circuit inputs, $cn^s - 1$ gate outputs, and 2 constant inputs (i.e., 0 and 1) is bounded above by $(n + cn^s + 1)^2$. It follows that W is bounded above by $(16(n + cn^s + 1)^2)^{cn^s}$ which in turn, for large enough n , is bounded above by $(n^{s+1})^{c2n^s} = 2^{2c(s+1)n^s \log n}$. Done.

We now prove Lemma 14. Any function in $\mathcal{F}_{n,r}$ that is computable by a Boolean circuit (of n inputs and rn outputs) of size cn^s can be computed by some rn Boolean circuits (of n inputs and 1 output) each of size cn^s . Hence, the number of functions in $\mathcal{F}_{n,r}$ that can be computed by a Boolean circuit (of n inputs and rn outputs) of size cn^s is bounded above by W^{rn} . We finish the proof by observing that W^{rn} is smaller than $2^{n^{s+2}}$ for large enough n .

ACKNOWLEDGEMENT

We would like to thank the anonymous reviewers and the associate editor for the various constructive feedback during the review process, including the alternative proof of Theorem 1 in Appendix A, the ideas of the proof in Section V confirming $C(p, r, \infty, \infty) < C(p, r, c, s)$ for some range of $r < 1 - H(p)$, and ideas of the stochastic code construction in Section II-E for achieving rates above the GV bound for some range of $r > 1 - H(p)$.

REFERENCES

- [1] E. Ruzomberka, C.-C. Wang, and D. Love, “Channel capacity for adversaries with computationally bounded observations,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, June 2022.
- [2] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell System Technical Journal*, vol. 27, no. 3, pp. 379 – 423, 1948.
- [3] R. J. McEliece, E. R. Rodemich, H. Rumsey, and L. R. Welch, “New Upper Bounds on the Rate of a Code via the Delsarte—MacWilliams Inequalities,” *IEEE Trans. Inf. Theory*, vol. 23, no. 2, pp. 157 – 166, 1977.
- [4] R. J. Lipton, “A new approach to information theory,” in *In Symposium on Theoretical Aspects of Computer Science*, 1994, p. 699–708.
- [5] V. Guruswami and A. Smith, “Optimal rate code constructions for computationally simple channels,” *Journal of the ACM*, vol. 63, no. 4, p. 1–37, 2016.
- [6] A. D. Sarwate, “Coding against myopic adversaries,” in *Proc. IEEE Inf. Theory Workshop*, 2010.
- [7] B. K. Dey, S. Jaggi, and M. Langberg, “Sufficiently myopic adversaries are blind,” *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5718 – 5736, 2019.
- [8] A. J. Budkuley, B. K. Dey, S. Jaggi, M. Langberg, A. D. Sarwate, and C. Wang, “Symmetrizability for Myopic AVCs,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2020, pp. 2103–2107.
- [9] Z. Chen, S. Jaggi, and M. Langberg, “A characterization of the capacity of online (causal) binary channels,” in *ACM symposium on Theory of Computing*, 2015, pp. 287–296.
- [10] B. K. Dey, S. Jaggi, M. Langberg, and A. D. Sarwate, “A bit of delay is sufficient and stochastic encoding is necessary to overcome online adversarial erasures,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2016, pp. 880–884.
- [11] V. Suresh, E. Ruzomberka, C.-C. Wang, and D. J. Love, “Stochastic-Adversarial Channels: Online Adversaries with Feedback Snooping,” *IEEE J. Sel. Areas Inf. Theory*, vol. 3, no. 1, pp. 69–84, 2022.
- [12] I. Csiszár and P. Narayan, “Capacity and Decoding Rules for Classes of Arbitrarily Varying Channels,” *IEEE Trans. Inf. Theory*, vol. 35, no. 4, pp. 752–769, 1989.
- [13] I. Csiszar and P. Narayan, “The Capacity of the Arbitrarily Varying Channel Revisited: Positivity, Constraints,” *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 181 – 193, 1988.
- [14] M. Langberg, “Oblivious communication channels and their capacity,” *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 424 – 429, 2008.
- [15] R. Shaltiel and J. Silbak, “Explicit uniquely decodable codes for space bounded channels that achieve list-decoding capacity,” in *Proc. ACM Symp. Theory Comp.*, June 2021, pp. 1516–1526.
- [16] ———, “Error correcting codes that achieve BSC capacity against channels that are poly-size circuits,” in *Proc. IEEE Symp. Foundations Comp. Science*, Nov 2022, pp. 13–23.
- [17] S. Kopparty, R. Shaltiel, and J. Silbak, “Quasilinear time list-decodable codes for space bounded channels,” in *Proc. IEEE Symp. Foundations Comp. Science*, Nov 2019, pp. 302–333.
- [18] R. Shaltiel and J. Silbak, “Explicit list-decodable codes with optimal rate for computationally bounded channels,” *Comput. Complex.*, vol. 30, no. 3, 2021.
- [19] C. Wang, “On the capacity of the binary adversarial wiretap channel,” in *54th Annual Allerton Conference on Communication, Control, and Computing, Allerton*, 2016, pp. 363–369.
- [20] M. Sipser, *Introduction to the theory of computation*, 2nd ed. Boston: Thompson Course Technology, 2006.
- [21] E. N. Gilbert, “A comparison of signaling alphabets,” *Bell System Technical Journal*, vol. 31, no. 3, p. 504–522, 1952.
- [22] R. R. Varshamov, “Estimate of the number of signals in error correcting codes,” *Dokl. Acad. Nauk*, vol. 117, no. 739–741, 1957.
- [23] B. K. Dey, S. Jaggi, M. Langberg, A. D. Sarwate, and C. Wang, “The interplay of causality and myopia in adversarial channel models,” in *Proc. Int. Symp. Inf. Theory (ISIT)*, June 2019, pp. 1002–1006.
- [24] S. Boucheron, G. Lugosi, and P. Massart, “Concentration inequalities using the entropy method,” *Annals of Probability*, vol. 31, no. 3, pp. 1583–1614, 2003.
- [25] V. H. Vu, “Concentration of Non-Lipschitz Functions and Applications,” in *Random Structures and Algorithms*, vol. 20, no. 3, 2002, pp. 262–316.
- [26] D. P. Dubhashi and A. Panconesi, *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.