Covert Communication Over a Compound Channel

Sadaf Salehkalaibar, Mohammad Hossein Yassaee, Vincent Y. F. Tan

Abstract

In this paper, we consider fundamental communication limits over a compound channel. Covert communication in the information-theoretic context has been primarily concerned with fundamental limits when the transmitter wishes to communicate to legitimate receiver(s) while ensuring that the communication is not detected by an adversary. This paper, however, considers an alternative, and no less important, setting in which the object to be masked is the state of the compound channel. Such a communication model has applications in the prevention of malicious parties seeking to jam the communication signal when, for example, the signal-to-noise ratio of a wireless channel is found to be low. Our main contribution is the establishment of bounds on the throughput-key region when the covertness constraint is defined in terms of the total variation distance. In addition, for the scenario in which the key length is infinite, we provide a sufficient condition for when the bounds coincide for the scaling of the throughput, which follows the square-root law. Numerical examples, including that of a Gaussian channel, are provided to illustrate our results.

I. Introduction

In modern applications that involve security considerations, it is not only of paramount importance that the message to be transmitted is hidden from malicious parties, but the very act of transmitting a message should be concealed from such parties. This is the study of *covert communication* and has been extensively studied in the information theory community in recent years [1]–[3]. Here, the transmitter feeds a specific symbol, called the *off-symbol*, to the channel when it decides not to send a message, signifying that the transmitter is in the *off-transmission state*. When the transmitter decides to send a *bona fide* message, the channel, together with the codeword associated to the message, induce a certain output distribution that is different from that when a message is not sent. One desires to design codes such that distinguishing between these two output distributions is difficult in the sense that the failure probability is high. The problem of covert communication over a channel with an additive white Gaussian noise (i.e., an AWGN channel) [1], a discrete memoryless channel (DMC) [2], [3], a classical-quantum channel [4], a channel with state [5], in the presence of an adversarial jammer [6], and from the second-order perspective [7], [8] have been considered. It is shown that the maximum amount of information that can be sent scales with the *square root* of the transmission blocklength; this is known as the *square-root law*. The extension to multi-user setups has been considered for multiple-access channels [9], broadcast channels [10] and relay channels [11].

In this work, we consider covert communication over a compound channel with two states $s \in \{1, 2\}$ (see Fig. 1). In a compound channel [12]–[14], the channel law W_s depends on the state which remains fixed during the transmission (see Section 7.2 of [15]). The channel state is known at the transmitter and the receiver. The goal of the communication is to reliably send a message to the receiver subject to a covertness constraint. We define the covertness in this unique setting as follows. An adversary who is observing the channel output cannot reliably distinguish the channel state of communication. The motivation comes from the fact that if the adversary determines the channel state, then there may be a possibility that it interferes with the communication by consuming a small amount of power. For example, consider a wireless channel with fading and for simplicity, assume that the fading takes on one of two states in $\{$"high SNR", "low SNR"$\}$. If a malicious party can deduce that the channel is in the "low SNR" state, then it knows that it is easier to interfere with the communication, possibly by injecting more noise to further disrupt the system by sending jamming signals with a small power consumption. Motivated by this scenario, we consider covert communication over a compound channel. Instead of focusing solely on covertness being measured in terms of the ability of an observer at the receiver being able to distinguish between the absence or presence of message transmission, here we also quantify covertness as the ability of an adversary in distinguishing between the two channel states. That is, we quantify covertness in terms of the probability of success of an adversary in a binary hypothesis test that seeks to distinguish between $s \in \{1, 2\}$. We wish to design coding schemes such that this probability of success is suitably upper bounded by a small constant, meaning that the adversary's test is unreliable. Our main contribution is in the establishment of bounds that characterize the optimal tradeoff between the throughput for deniable (i.e., the adversary cannot learn whether communication is occurring or what the channel state is) and reliable communication as well as the length of the shared key between the transmitter and receiver. In [16], covert communication over compound binary symmetric channels was studied but the setting is different from ours. Specifically, the best throughput to deniably (i.e., the adversary cannot learn

Sadaf Salehkalaibar is with the Department of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran (e-mail: s.saleh@ut.ac.ir).

Mohammad Hossein Yassaee is with Institute for Research in Fundamental Sciences (IPM), Tehran, Iran (e-mail: yassaee@ipm.ir).

Vincent Y. F. Tan is with Department of Electrical and Computer Engineering and the Department of Mathematics, National University of Singapore, (e-mail: vtan@nus.edu.sg).
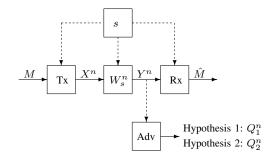
Fig. 1. Covert communication over a compound DMC.

whether communication is occurring) and reliably communicate was established, whereas our main contribution is to establish bounds on the throughput under the constraint that the adversary cannot learn the channel state.

We consider two covertness criteria for this setup.

- For the first criterion, a specific distribution Q_0 over the channel output alphabet $\mathcal{Y}$ is fixed. This is similar to traditional works on covert communication such as [2], [3] in which Q_0 is the distribution that is induced from sending the off-symbol at the input of the channel. For a given blocklength n, the covertness is measured by the Kullback-Leibler (KL)-divergence of the channel output distribution, denoted as Q_s^n, with $Q_0^{\times n}$, which denotes the n-fold product of Q_0. In order to fix such a distribution Q_0, it is assumed that there exists two distributions P_1 and P_2 over the channel input alphabet $\mathcal{X}$ such that they induce the *same* channel output marginal when they are fed into channels W_1 and W_2, respectively and the distribution Q_0 denotes this common marginal. Notice the difference of definition of Q_0 in the compound setup with that of a DMC in [2], [3] where Q_0 denotes the channel output marginal when the off-symbol is fed into the channel. Here, Q_0 serves to *mask* the channel state and it does not necessarily correspond to the output marginal of a specific symbol in the input alphabet. Thus, the off-symbol which indicates that the transmitter is in the off-transmission state, does not necessarily exist. Even though an adversary who is observing the channel output may not be able to infer the state of the communication, he can infer whether the transmitter is sending a message. For this covertness criterion, assuming a shared key of infinite length between the transmitter and receiver, we provide the optimal throughput.
- The second covertness criterion, as we shall see based on the motivation above, is more pertinent to the compound channel. It is defined to be the total variation distance between Q_1^n and Q_2^n, the two output marginal distributions when the state of the channel is $s = 1$ or $s = 2$. Here, we assume that a shared key with bounded length is available between the transmitter and receiver. We provide inner and outer bounds on the throughput-key region and show that the square root law [1] also holds in this setting. The bounds match for a special case and thus for this case, we have characterized the optimal transmission rate given a key with a certain length. The achievable scheme uses two codebooks (each of which is used for communication given a channel state) and a maximum likelihood (ML) decoder at the receiver. The analysis of this scheme consists of three steps. The first step is to show that the average error probability of the scheme when averaged over all codebooks, can be appropriately upper bounded. The second step is to provide a single-letter characterization of the covertness metric. The third step is to show that there exists a deterministic codebook that satisfies the desired constraints on the maximum error probability and also the covertness criterion. The proof of outer (converse) bound is challenging since we need to analyze the total variation distance between two *non-product* distributions Q_1^n and Q_2^n. The single-letter characterization of the covertness metric requires relating the total variation distance to a hypothesis testing problem and using various bounds (Berry-Esseen) to control the false alarm and missed detection probabilities of this test.

We conclude this section with a summary of the main contributions, introduction of notation, and the organization of the paper.

A. Contributions

The main contributions of this paper are as follows:

- The problem of covert communication for a compound DMC is formulated.
- For the case in which covertness is measured in terms of the total variation distance between Q_1^n and Q_2^n, we provide an inner bound on the the optimal transmission throughput-shared key region (Theorem 1). The proof of this theorem involves three steps which are discussed in Lemmas 1–4.
- If we assume that the key is sufficiently long and the error probability of decoding the message at the legitimate receiver is vanishing, we are able to upper bound the optimal transmission throughput for the same covertness criterion (Theorem 2). It is shown that the upper and lower bounds match for a special case (Theorem 3).
- Examples are provided to compare the lower and upper bounds. In particular, a Gaussian setup is also considered where we identify parameter regimes in which the conditions of the optimality result in Theorem 3 are satisfied. The proof of

converse in this theorem needs to be adapted to the Gaussian model since the output alphabet of the Gaussian channel is continuous.

B. Notation

We mostly use the notation of [15]. Random variables are shown by capital letters, e.g., X, Y and their realizations by lower-case letters, e.g., x, y. The alphabets of random variables are denoted by script symbols such as $\mathcal{X}$, $\mathcal{Y}$. The sequence of random variables $(X_i, \ldots, X_j)$ and its realizations $(x_i, \ldots, x_j)$ are abbreviated as X_i^j and x_i^j respectively. We use X^j and x^j instead of X_1^j and x_1^j. The empirical distribution of sequence x^n is also known as its *type*.

The probability mass function (pmf) of a discrete random variable X defined over the channel input alphabet $\mathcal{X}$ denoted using the letter P. The pmf of a discrete random variable Y defined over the channel output alphabet $\mathcal{Y}$ is denoted using the letter Q. The distributions of the sequences of random variables X^n and Y^n are denoted by P^n and Q^n respectively. The notation $P^{\times n}$ denotes the n-fold product distribution, i.e., for every $x^n \in \mathcal{X}^n$, we have

$$P^{\times n}(x^n) := \prod_{i=1}^{n} P(x_i). \tag{1}$$

The expectation and variance operators are written as $\mathbb{E}[.]$ and $\mathbb{V}[.]$, respectively. The notation $\mathbb{1}\{.\}$ denotes the indicator function. The probability of an event $\mathcal{E} \subseteq \mathcal{X}$ is denoted by $\mathbb{P}(\mathcal{E})$.

The KL-divergence, chi-squared distance and total variation distance between two distributions P_1 and P_2 are denoted by $D(P_1\|P_2)$, $\chi_2(P_1\|P_2)$ and $d_{\mathrm{TV}}(P_1, P_2)$, respectively. The Bhattacharyya distance between two distributions P_1 and P_2 is denoted by $F(P_1, P_2)$ and is defined as $F(P_1, P_2) \triangleq \sum_x \sqrt{P_1(x)P_2(x)}$.

The Hamming weight of sequence x^n is denoted by $\mathrm{w}(x^n)$. The binary entropy function is shown by $h_{\mathrm{b}}(.)$. The binary symmetric channel with parameter ϵ is denoted by $\mathrm{BSC}(\epsilon)$. The Bernoulli distribution, i.e., one that is defined over $\{0, 1\}$, is denoted by $\mathrm{Bern}(.)$. The cumulative distribution function of a standard normal distribution is denoted by $\Phi(.)$. All logarithms are with respect to the base 2. The exponential function is written as $\exp(.)$.

C. Organization

The remainder of the paper is organized as follows. Section II describes the covert communication setup over a compound channel. Section III provides the lower and upper bounds on the optimal transmission throughput and some examples to discuss the bounds. The paper is concluded in Section IV and by technical appendices.

II. COVERT COMMUNICATION OVER A COMPOUND CHANNEL

A. System Model

Consider covert communication over a compound DMC with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, see Fig. 1. The compound channel consists of two channel pmfs $W_s(\cdot|\cdot), s \in \{1, 2\}$ where the channel state s is available at both the transmitter and the receiver. The message set is shown by $\mathcal{M}$. The transmitter and receiver share a secret key $K \in \mathcal{K}$. The transmitter sends an n-length input X^n over the channel using a state-dependent encoding function $f_s^{(n)} \colon \mathcal{M} \times \mathcal{K} \to \mathcal{X}^n$ where

$$X^n = f_s^{(n)}(M, K). \tag{2}$$

The receiver uses a decoding function $g_s^{(n)} \colon \mathcal{Y}^n \times \mathcal{K} \to \mathcal{M}$ which maps the channel output to an estimated message as follows:

$$\widehat{M} = g_s^{(n)}(Y^n, K). \tag{3}$$

In the conventional covert communication over a single-state point-to-point channel [2], [3], a fixed distribution which is denoted by Q_0, is chosen to represent the single-letter channel output marginal when the transmitter sends the off-symbol. This symbol is fed to the channel when the transmitter decides not to send a message. The covertness is measured by the "distance" between the channel output marginal when a message is sent and the distribution $Q_0^{\times n}$.

The goal of covert communication over a compound channel is primarily to conceal the channel state. It might be possible to cover message transmission such that an adversary observing the channel output does not know whether an off-symbol or a message is sent. In the following, we consider the different cases where a distribution Q_0 can be fixed for the definition of covert communication over a compound channel.

Case 1 (Only concealing the channel state): In this case, the channel state is masked. However, the message transmission might be revealed to an adversary since an off-symbol does not exist. Specifically, there exist two non-deterministic distributions[1] P_1 and P_2 such that

$$\sum_x P_1(x)W_1(y|x) = \sum_x P_2(x)W_2(y|x), \ \forall y \in \mathcal{Y}. \tag{4}$$

---

[1] A *deterministic distribution* P is one in which there exists a symbol x^* such that $P(x^*) = 1$. A *non-deterministic distribution* is one that is not deterministic.
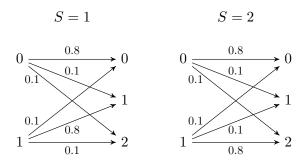
Fig. 2. Example for Case 2.

In this case, we define Q_0 to represent the above common channel output marginal between the two states. Here, an adversary who is observing the channel output, cannot infer the channel state. Thus, the distribution Q_0 serves to *mask* the state of the communication. Here, there does not necessarily exist an input symbol whose channel output marginal is as (4). Thus, the adversary who is observing the channel output may infer whether the transmitter is sending a message.

An example of this case of covert communication is discussed in the following. Suppose that $\mathcal{X} = \mathcal{Y} = \{0,1\}$, where $W_1(.|.)$ and $W_2(.|.)$ are BSC $\left(\frac{1}{4}\right)$ and BSC $\left(\frac{1}{3}\right)$, respectively. The non-deterministic input distributions $P_1, P_2 \sim \text{Bern} \left(\frac{1}{2}\right)$ satisfy Condition (4) and $Q_0(0) = Q_0(1) = 1/2$. Clearly, the distribution Q_0 is not induced by any of the input symbols at both channel states.

Case 2 (Concealing channel state and message transmission): In this case, condition (4) is satisfied only if P_1 and P_2 are deterministic distributions. Thus, there exist symbols "0" and "0'" such that

$$W_1(y|0) = W_2(y|0'), \qquad \forall y \in \mathcal{Y}. \tag{5}$$

Without loss of generality, we can suppose that the symbols 0 and $0'$ are equal, because the transmitter has access to the state[2]. Here, we again call the symbol 0 as *off-symbol*. Let Q_0 denotes the above common channel output marginal between the two channel states. The distribution Q_0 masks the channel state and also represents the scenario in which the transmitter does not send a message at both channel states.

An example of this case is given in Fig. 2 where we assume that $\mathcal{X} = \{0,1\}$ and $\mathcal{Y} = \{0,1,2\}$. Notice that the unique distributions satisfying (4) are given by $P_1(0) = P_2(0) = 1$ and $P_1(1) = P_2(1) = 0$. Thus, the symbol "0" can be regarded as the off-symbol.

Case 3 (Concealing channel state and message transmission partially): Condition (4) is satisfied only if P_1 is deterministic while P_2 can be non-deterministic. In this case, we have:

$$W_1(y|0) = \sum_x P_2(x)W_2(y|x), \qquad \forall y \in \mathcal{Y}. \tag{6}$$

Here, Q_0 denotes the above common channel output marginal. Besides the fact that Q_0 masks the channel state, it also represents the fact that the transmitter has no input to the channel when $S = 1$. A symmetric case can be considered when P_2 is deterministic and P_1 is non-deterministic.

Case 4 (Revealing channel state and message transmission): There does not exist any pair of distributions P_1 and P_2 such that Condition (4) is satisfied. In this case, covert communication over the compound channel is not feasible.

B. Covertness Criteria

As discussed in the above Cases 1–3, a distribution Q_0 over $\mathcal{Y}$ can be fixed. As such we can define the covert rate in a number of different ways depending on the covertness criterion. Two different covertness criteria are considered.

Definition 1: An $(n, |\mathcal{M}|, |\mathcal{K}|)$-code consists of a message set $\mathcal{M}$, a key set $\mathcal{K}$, two encoding functions $f_s^{(n)} : \mathcal{M} \times \mathcal{K} \to \mathcal{X}^n$, and two decoding functions $g_s^{(n)} : \mathcal{Y}^n \times \mathcal{K} \to \mathcal{M}$ for $s \in \{1,2\}$. For each state s, the code induces the following joint pmf on the random variables $(M, K, X^n, Y^n, \widehat{M})$[3]:

$$P_{MKX^nY^n\widehat{M}}(m,k,x^n,y^n,\hat{m}) \triangleq \frac{1}{|\mathcal{M}| \cdot |\mathcal{K}|} W_s^n(y^n|x^n) \cdot \mathbb{1}\left\{f_s^{(n)}(m,k) = x^n\right\} \cdot \mathbb{1}\left\{g_s^{(n)}(y^n,k) = \hat{m}\right\}. \tag{7}$$

---

[2]In fact, if $0 \neq 0'$, we can permute the input symbols for the state $s = 2$, such that the $0'$ is renamed to 0.

[3]The adversary knows the distribution which is used for generating the codebook but does not know the chosen codebook by the transmitter and receiver.

We denote the channel output marginal on $\mathcal{Y}^n$ by Q_s^n. Define:

$$P_{\mathrm{e}}(s,k,m) \triangleq \mathbb{P}\big(\widehat{M} \neq M | S = s, K = k, M = m\big), \tag{8}$$

and

$$\bar{P}_{\mathrm{e}}(s,k) \triangleq \frac{1}{|\mathcal{M}|} \sum_{m=1}^{|\mathcal{M}|} \mathbb{P}\left(\widehat{M} \neq M | M = m, K = k, S = s\right), \qquad \forall s \in \{1,2\},\ k \in \mathcal{K}. \tag{9}$$

For this code, the maximum error probability is defined as

$$P_{\mathrm{e}} \triangleq \max_{s=1,2}\ \max_{k \in \mathcal{K}}\ \max_{m \in \mathcal{M}}\ P_{\mathrm{e}}(s,k,m). \tag{10}$$

Two covertness criteria are considered. The first criterion is that the KL-divergence between the marginal distributions $Q_s^n, s = 1,2$ and the *masking* distribution $Q_0^{\times n}$ is sufficiently small. This criterion imposes more restriction on the concealing the channel state in the sense that not only adversary can not infer the channel state, it also can not distinguish the marginal outputs from the masking distribution. The second criterion relaxes this additional requirement. The second criterion requires that the total variation distance between the channel output marginals of the two states, i.e., $d_{\mathrm{TV}}(Q_1^n, Q_2^n)$ to be small. Thus, adversary can not only infer the channel state.

Therefore, we have two definitions:

- An $(n, |\mathcal{M}|, |\mathcal{K}|, \epsilon, \delta_1, \delta_2, \zeta)_{\mathrm{KL}}$-code is an $(n, |\mathcal{M}|, |\mathcal{K}|)$-code such that the KL-divergence covertness terms are no larger than δ_1 and δ_2, when the state is $s = 1$ and $s = 2$, respectively and the maximum error probability is no larger than ϵ.
- An $(n, |\mathcal{M}|, |\mathcal{K}|, \epsilon, \delta)_{\mathrm{TV}}$-code is an $(n, |\mathcal{M}|, |\mathcal{K}|)$-code such that the total variation distance as the covertness metric is no larger than δ and the maximum error probability is no larger than ϵ.

Definition 2 (KL-divergence criterion): For $(\epsilon, \delta_1, \delta_2, \zeta) \in \mathbb{R}_+^4$, a pair $(L, L_K) \in \mathbb{R}_+^2$ is said to be $(\epsilon, \delta_1, \delta_2, \zeta)$-achievable if there exists a sequence of $(n, |\mathcal{M}|, |\mathcal{K}|, \epsilon, \delta_1, \delta_2, \zeta)_{\mathrm{KL}}$-codes such that

$$\liminf_{n \to \infty} \frac{\log |\mathcal{M}|}{n^\zeta} \geq L, \tag{11}$$

$$\limsup_{n \to \infty} \frac{\log |\mathcal{K}|}{n^\zeta} \leq L_K. \tag{12}$$

The set of all $(\epsilon, \delta_1, \delta_2, \zeta)$-achievable pairs (L, L_K) is denoted as $\mathcal{L}_\zeta^*(\epsilon, \delta_1, \delta_2)$. We also define $L_\zeta^*(\epsilon, \delta_1, \delta_2)$ as follows:

$$L_\zeta^*(\epsilon, \delta_1, \delta_2) = \sup \left\{ L \colon (L, L_K) \in \mathcal{L}^*(\epsilon, \delta_1, \delta_2, \zeta) \text{ for some } L_K \right\}, \tag{13}$$

and

$$L_\zeta^*(0^+, \delta_1, \delta_2) = \lim_{\epsilon \to 0} L_\zeta^*(\epsilon, \delta_1, \delta_2). \tag{14}$$

Remark 1: In this work, we are interested in the cases of $\zeta = \frac{1}{2}$ and $\zeta = 1$. When $L_1^*(\epsilon, \delta_1, \delta_2) > 0$, a positive rate of communication is achievable.

Definition 3 (Total variation distance as the covertness criterion): For $(\epsilon, \delta) \in \mathbb{R}_+^2$, a pair $(L, L_K) \in \mathbb{R}_+^2$ is said to be (ϵ, δ)-achievable if there exists a sequence of $(n, |\mathcal{M}|, |\mathcal{K}|, \epsilon, \delta)_{\mathrm{TV}}$-codes such that (11)–(12) hold for $\zeta = \frac{1}{2}$. The set of all (ϵ, δ)-achievable pairs (L, L_K) is denoted as $\mathcal{L}_{\mathrm{TV}}^*(\epsilon, \delta)$. We also define

$$L_{\mathrm{TV}}^*(\epsilon, \delta) = \sup \left\{ L \colon (L, L_K) \in \mathcal{L}_{\mathrm{TV}}^*(\epsilon, \delta) \text{ for some } L_K \right\}, \tag{15}$$

and

$$L_{\mathrm{TV}}^*(0^+, \delta) = \lim_{\epsilon \to 0} L_{\mathrm{TV}}^*(\epsilon, \delta). \tag{16}$$

C. Basic Results

In this section, we present some basic results on the covert rates to provide more intuition on the model. For simplicity, we assume that an infinite-length key is shared between the transmitter and receiver. The covertness criterion is assumed to be the KL-divergence as in Definition 1. The following proposition states that we can communicate at a positive rate in Case 1. Define:

$$\mathcal{P} \triangleq \{(P_1, P_2) \colon\ (P_1, P_2) \text{ satisfies Condition (4)}\}. \tag{17}$$

Proposition 1: Assuming that Case 1 holds and an infinite-length key is shared between the transmitter and receiver, we have:

$$L_1^*(0^+, \delta_1, \delta_2) = \min_{s \in \{1,2\}} \max \ I(P_s, W_s), \tag{18}$$

where the maximum is over all distributions $(P_1, P_2) \in \mathcal{P}$.

Proof: Similar to [2, Proposition 1]. ∎

An interesting example of Case 1 is the Gaussian setup which we now discuss in the following. Suppose that at each channel state $s \in \{1, 2\}$, the channel is given by:

$$Y = X + Z_s, \qquad Z_s \sim \mathcal{N}(0, \sigma_s^2), \tag{19}$$

where $\sigma_1^2 > \sigma_2^2$, $X, Y, Z_s \in \mathbb{R}$ and X is independent of Z_s. We impose an average power constraint on the input such that $\mathbb{E}\left[|X|^2\right] \leq \mathsf{P}$ for some $\mathsf{P} > 0$. In the following, we discuss the scenario in which there exist two non-deterministic distributions P_1 and P_2 such that Condition (4) holds. Thus, a positive rate can be achieved for this example. Consider Proposition 1 and let the second moment of the distribution P_s be denoted by ρ_s. We know that a Gaussian random variable maximizes $I(P_s, W_s)$ among all distributions of the same second moment. Thus, we get:

$$I(P_s, W_s) \leq \frac{1}{2} \log\left(1 + \frac{\rho_s}{\sigma_s^2}\right). \tag{20}$$

The distributions P_1 and P_2 satisfy (4) if the following holds:

$$\rho_1 + \sigma_1^2 = \rho_2 + \sigma_2^2, \tag{21}$$

which yields the following optimal positive rate (with $\rho_2 = \mathsf{P}, \rho_1 = \mathsf{P} - (\sigma_1^2 - \sigma_2^2) < \mathsf{P}$):

$$L_1^*(0^+, \delta_1, \delta_2) = \frac{1}{2} \log\left(\frac{\mathsf{P} + \sigma_2^2}{\sigma_1^2}\right). \tag{22}$$

Notice the difference of this example with the Gaussian setup of [2, Section V]. In the proposed example, a positive rate can be achieved at each channel state. However, in the Gaussian example of [2, Section V], the maximum communication rate in the presence of covertness constraint is zero, i.e., we cannot communicate at a positive rate. As discussed in the previous section, the difference arises from the fact that the goal of covert communication over a compound channel is to mask the state, while in the conventional model, the aim is to keep the adversary oblivious about the state of transmitter, i.e., whether or not a message is sent.

The following proposition presents the optimal covert communication rate of Case 2 which also states that the maximum rate is zero.

Proposition 2: Assuming that Case 2 holds and an infinite-length key is shared between the transmitter and receiver, we have

$$L_{\frac{1}{2}}^*(0^+, \delta_1, \delta_2) = \min_{s \in \{1,2\}} \max_{\tilde{P}_s : \tilde{P}_s(0)=0} \sqrt{\frac{2\delta_s}{\chi_2(\tilde{Q}_s \| Q_0)}} \cdot D(W_s \| Q_0 | \tilde{P}_s), \tag{23}$$

where $\tilde{Q}_s$ denotes the output marginal of the channel W_s when the input distribution is $\tilde{P}_s$.

Proof: Similar to [2, Theorem 2]. ∎

Recall the example in Fig. 2 which satisfies the condition of Case 2. Evaluation of (23) for the proposed example with $\delta_s = 1$ yields $L_{\frac{1}{2}}^*(0^+, 1, 1) = 1.26$.

To conclude this section, we briefly discuss the advantage of concealing only the channel state as compared to the set-up which requires the strengthen constraint that adversary cannot distinguish between the channel output and the masking distribution. Whenever Case 2 occurs, the latter setup is equivalent to protecting the message transmission from the detection by an adversary, since we have an off-symbol in Case 2. Thus, we compare the message transmission concealing setup versus concealing only the channel state.

Consider the setup of Fig. 3 which satisfies the condition of Case 2. For this example, the KL-divergence $D(Q_s^n \| Q_0^{\times n})$ goes to infinity which implies that it is not feasible to conceal the message transmission. However, as will be discussed in Subsection III-D, concealing the channel state under the total variation covertness criterion is possible in the sense that $L_{\text{TV}}^*(0^+, \delta) > 0$.[4]

Remark 2: In the following results, we assume that all the transition probabilities $W_s(.|x), s \in \{1, 2\}, x \in \mathcal{X}$ are absolutely continuous with respect to $Q_0(.)$ (if it exists). This condition ensures us that examples such as that of Fig. 3 are excluded and all involved information quantities are finite. If this assumption is not required, it will be explicitly mentioned in the text.
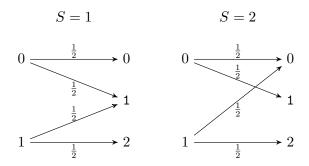
---

[4]This observation is clearly true for the Case 1.

Fig. 3. Example where concealing the message transmission is impossible, while concealing the channel state is feasible.

III. Main Results

In this section, we present our results when the covertness criterion is considered to be the total variation distance between channel output marginals of the two states (see Definition 3). As discussed in the previous section (see Case 1), if there exist two non-deterministic input distributions P_1 and P_2 such that Condition (4) holds, then the optimal rate is given by Proposition 1 and the maximum covert communication rate is positive. Thus, we only focus on Case 2 in the following and assume that Condition 4 is satisfied only if P_1 and P_2 are deterministic input distributions. Moreover, recall "0" is the off-symbol such that $W_1(\cdot|0) = W_2(\cdot|0) \triangleq Q_0(\cdot)$.

A. An Inner Bound to $\mathcal{L}_{TV}^*(\epsilon, \delta)$

In the following, we present an inner bound to $\mathcal{L}_{TV}^*(\epsilon, \delta)$. The proof is given afterwards.
For positive γ_1, γ_2 and any three distributions Q_0, Q_1 and Q_2 over $\mathcal{Y}$, define:

$$\Omega(\gamma_1, \gamma_2; Q_0, Q_1, Q_2) \triangleq \gamma_1^2 \chi_2(Q_1\|Q_0) + \gamma_2^2 \chi_2(Q_2\|Q_0) - 2\gamma_1\gamma_2\rho(Q_1, Q_2, Q_0), \tag{24}$$

where

$$\rho(Q_1, Q_2, Q_0) \triangleq \mathbb{E}_{Q_0}\left[\left(\frac{Q_1(Y)}{Q_0(Y)} - 1\right) \cdot \left(\frac{Q_2(Y)}{Q_0(Y)} - 1\right)\right]. \tag{25}$$

Theorem 1: An inner bound to $\mathcal{L}_{TV}^*(\epsilon, \delta)$, is given by the union of all pairs (L, L_K) such that:

$$L \leq \min_{s\in\{1,2\}} \gamma_s D(W_s\|Q_0|\bar{P}_s), \tag{26}$$

and

$$L_K \geq \max_{s\in\{1,2\}} \gamma_s D(W_s\|Q_0|\bar{P}_s) - \min_{s\in\{1,2\}} \gamma_s D(W_s\|Q_0|\bar{P}_s), \tag{27}$$

where the union is over all input distributions $\bar{P}_s$ such that $\bar{P}_s(0) = 0$ with their output marginals through W_s are denoted by $\bar{Q}_s$ and over all positive γ_1, γ_2 such that

$$\Omega(\gamma_1, \gamma_2; Q_0, \bar{Q}_1, \bar{Q}_2) \leq \left(2\Phi^{-1}\left(\frac{1+\delta}{2}\right)\right)^2. \tag{28}$$

Proof: The achievable scheme is sketched as follows. Two codebooks are generated where each of them is used for communicating over each channel state. The optimal decoder –namely, the ML decoder– is used to determine the message which is sent. The analysis of error probability involves applying Shannon's achievability bound [17, Thm 2] to upper bound the error probability averaged over all codebooks. A main step of the analysis is the single-letter characterization of the covertness metric. Finally, it is shown that there exists a "good" codebook which satisfies desired properties on the maximum error probability and also the covertness criterion.

Preparations: Fix a large blocklength n, choose positive γ_s for $s \in \{1, 2\}$. Let $\bar{P}_s$ be any distribution such that $\bar{P}_s(0) = 0$ and denote the output marginal through channel W_s by $\bar{Q}_s$. Choose positive γ_s such that (28) holds and let

$$\mu_{s,n} \triangleq \frac{\gamma_s}{\sqrt{n}}. \tag{29}$$

Define the input distribution $P_{s,n}$ as

$$P_{s,n}(x) \triangleq \begin{cases} \mu_{s,n}\bar{P}_s(x) & x \neq 0 \\ 1 - \mu_{s,n} & x = 0 \end{cases}. \tag{30}$$

For $s \in \{1, 2\}$, let $Q_{s,n}$ denote the output marginal of the channel W_s when the input distribution is $P_{s,n}$. The distribution $Q_{s,n}$ can be written as follows:

$$Q_{s,n} = (1 - \mu_{s,n})Q_0 + \mu_{s,n}\bar{Q}_s. \tag{31}$$

Denote the n-fold products of $P_{s,n}$ and $Q_{s,n}$ by $P_s^{\times n}$ and $Q_s^{\times n}$, respectively.

Codebook Generation: We generate an i.i.d. codebook $\mathcal{C}_s = \{x_s^n(m, k) \colon m \in \mathcal{M}, k \in \mathcal{K}\}$ according to the input distribution $P_{s,n}$ as defined in (30). Denote the channel output marginal by Q_s^n.

Single-letter Characterization of Covertness Condition: The following two lemmas together provide a single-letter characterization of the covertness condition.

Lemma 1: We have

$$d_{\mathrm{TV}}\left(Q_1^{\times n}, Q_2^{\times n}\right) \leq 2\Phi\left(\frac{1}{2}\sqrt{\Omega(\gamma_1, \gamma_2; Q_0, \bar{Q}_1, \bar{Q}_2)}\right) - 1 + O\left(\frac{1}{\sqrt{n}}\right). \tag{32}$$

Proof: See Appendix A. ∎

Lemma 2: Let κ be an arbitrary positive number. Assume that,

$$\log|\mathcal{M}| + \log|\mathcal{K}| \geq (1 + \kappa)\sqrt{n} \cdot \max_{s \in \{1,2\}} \gamma_s D(W_s \| Q_0 | \bar{P}_s). \tag{33}$$

Then, we have

$$\mathbb{E}\left[d_{\mathrm{TV}}\left(Q_s^n, Q_s^{\times n}\right)\right] = O\left(\frac{1}{\sqrt{n}}\right), \tag{34}$$

whenever $\Omega(\gamma_1, \gamma_2; Q_0, \bar{Q}_1, \bar{Q}_2)$ satisfies (28).

Proof: See Appendix B. ∎

Encoding and Decoding: The transmitter upon observing the message $m \in \mathcal{M}$ and the key $k \in \mathcal{K}$, sends the codeword $x_s^n(m, k)$ over the channel. Given y^n, the decoder based on the secret key k and the state s uses the optimal ML decoder which chooses the message $\hat{m}$ as follows:

$$\hat{m} = \underset{m \in \mathcal{M}}{\operatorname{argmax}} W_s^n(y^n | x_s^n(m, k)) \tag{35}$$

The following lemma provides an upper bound on the average error probability as defined in (9).

Lemma 3: If

$$\frac{\log|\mathcal{M}|}{\sqrt{n}} \leq \min_{s \in \{1,2\}} \sqrt{n} I(P_{s,n}, W_s) - n^{-\frac{1}{6}}, \tag{36}$$

Then,

$$\mathbb{E}\left[\bar{P}_{\mathrm{e}}(s, k)\right] = O\left(n^{-\frac{1}{6}}\right), \qquad \forall s \in \{1, 2\},\ k \in \mathcal{K}. \tag{37}$$

Proof: See Appendix C. ∎

Existence of a Codebook with Desired Properties: The following lemma proves existence of a codebook with desired properties.

Lemma 4: If for some positive κ and ϵ, constraints (33) and (36) hold, then there are exists at least one "good" codebook such that:

$$d_{\mathrm{TV}}\left(Q_s^n, Q_s^{\times n}\right) = o(1), \tag{38}$$

and the maximum error probability as defined in (10), satisfies:

$$P_{\mathrm{e}} = o(1) \tag{39}$$

Proof: See Appendix D. ∎

The proof is concluded by combining lemmas 1 to 4. See Appendix E for details.

∎

Now, we specialize Theorem 1 to the case where the length of key is sufficiently long. Also, we restrict to the case of $\mathcal{X} = \{0, 1\}$ and the output distribution induced by each input symbol is denoted by

$$\tilde{Q}_1(\cdot) \triangleq W_1(\cdot|1), \qquad \tilde{Q}_2(\cdot) \triangleq W_2(\cdot|1). \tag{40}$$

This special case will be used later to provide an optimality result. In this case, we are able to find a closed form for the optimization problem (26) under the constraint (28). The solution of this optimization is given in the following corollary.

Corollary 1: Assume that the length of the key is infinite, i.e., $L_K = \infty$ and $\mathcal{X} = \{0, 1\}$. Then, $L_{\text{TV}}^*(\epsilon, \delta)$ is lower bounded as follows:

$$L_{\text{TV}}^*(\epsilon, \delta) \geq \mathsf{L}, \tag{41}$$

where

$$\mathsf{L} \triangleq 2\Phi^{-1}\left(\frac{1+\delta}{2}\right) \cdot \frac{D(\tilde{Q}_1\|Q_0)D(\tilde{Q}_2\|Q_0)}{\sqrt{\chi_2(\tilde{Q}_1\|Q_0)D(\tilde{Q}_2\|Q_0)^2 - 2\rho(\tilde{Q}_1, \tilde{Q}_2, Q_0)D(\tilde{Q}_1\|Q_0)D(\tilde{Q}_2\|Q_0) + \chi_2(\tilde{Q}_2\|Q_0)D(\tilde{Q}_1\|Q_0)^2}}, \tag{42}$$

if

$$\rho(\tilde{Q}_1, \tilde{Q}_2, Q_0) < \min\left(\frac{\chi_2(\tilde{Q}_1\|Q_0)D(\tilde{Q}_2\|Q_0)}{D(\tilde{Q}_1\|Q_0)}, \frac{\chi_2(\tilde{Q}_2\|Q_0)D(\tilde{Q}_1\|Q_0)}{D(\tilde{Q}_2\|Q_0)}\right), \tag{43}$$

and

$$\mathsf{L} \triangleq 2\Phi^{-1}\left(\frac{1+\delta}{2}\right) \cdot \sqrt{\frac{\min\left(\chi_2(\tilde{Q}_1\|Q_0)D(\tilde{Q}_2\|Q_0)^2, \chi_2(\tilde{Q}_2\|Q_0)D(\tilde{Q}_1\|Q_0)^2\right)}{\chi_2(\tilde{Q}_1\|Q_0)\chi_2(\tilde{Q}_2\|Q_0) - \rho(\tilde{Q}_1, \tilde{Q}_2, Q_0)^2}}, \tag{44}$$

if

$$\rho(\tilde{Q}_1, \tilde{Q}_2, Q_0) \geq \min\left(\frac{\chi_2(\tilde{Q}_1\|Q_0)D(\tilde{Q}_2\|Q_0)}{D(\tilde{Q}_1\|Q_0)}, \frac{\chi_2(\tilde{Q}_2\|Q_0)D(\tilde{Q}_1\|Q_0)}{D(\tilde{Q}_2\|Q_0)}\right). \tag{45}$$

Proof: See Appendix F. ∎

B. An Upper Bound on $L_{TV}^*(0^+, \delta)$ and an Optimality Result

In this section, we provide an optimality result for $L_{\text{TV}}^*(0^+, \delta)$. We consider two simplifications on the system model. First, it is assumed that the length of key is infinite. Second, we restrict to the case of $\mathcal{X} = \{0, 1\}$ where the output distributions are defined in (40). The following theorem presents an upper bound on $L_{\text{TV}}^*(0^+, \delta)$. Its proof is given after the statement of Theorem 3. Define:

$$\Delta \triangleq \mathbb{E}_{Q_0}\left[\left(\frac{\tilde{Q}_1(Y) - \tilde{Q}_2(Y)}{Q_0(Y)}\right)^2\right]. \tag{46}$$

Theorem 2: Assuming

$$\rho(\tilde{Q}_1, \tilde{Q}_2, Q_0) \leq \min\left\{\chi_2(\tilde{Q}_1\|Q_0), \chi_2(\tilde{Q}_2\|Q_0)\right\}, \tag{47}$$

$L_{\text{TV}}^*(0^+, \delta)$ is upper bounded as follows:

$$L_{\text{TV}}^*(0^+, \delta) \leq \frac{2\Phi^{-1}\left(\frac{1+\delta}{2}\right)}{\sqrt{\Delta}} \cdot \max_{s\in\{1,2\}} D(\tilde{Q}_s\|Q_0). \tag{48}$$

The above upper bound is tight in the following special case.

Theorem 3 (Optimality Result): Assuming (47) and

$$D(\tilde{Q}_1\|Q_0) = D(\tilde{Q}_2\|Q_0) \triangleq \mathbb{D}. \tag{49}$$

Then,

$$L_{\text{TV}}^*(0^+, \delta) = \frac{2\Phi^{-1}\left(\frac{1+\delta}{2}\right)}{\sqrt{\Delta}} \cdot \mathbb{D}. \tag{50}$$

Proof: The achievability follows from Corollary 1 and considering the fact that assumptions in (47) and (49) imply the first clause of L in (42) holds. The converse follows from Theorem 2 and using the assumption (49). ∎

Proof of Theorem 2: The proof is sketched as follows. First, we relate the total variation distance to false alarm and missed detection probabilities of a hypothesis testing problem. Then, we upper bound these probabilities which gives a lower bound on the total variation distance. Finally, the proof is concluded by showing that there exists at least one codebook which satisfies the desired property on the total variation distance and also by applying Fano's inequality.

Preparations: For any two codebooks

$$\mathcal{C}_s = \{x_s^n(m,k) : m \in \mathcal{M}, k \in \mathcal{K}\}, \qquad s \in \{1,2\}, \tag{51}$$

we define the following parameters:

$$\mu_{\text{L},n} \triangleq \min_{s \in \{1,2\}} \min_{m \in \mathcal{M}} \min_{k \in \mathcal{K}} \frac{1}{n} \text{w}\left(x_s^n(m,k)\right), \tag{52}$$

$$\mu_{\text{H},n} \triangleq \max_{s \in \{1,2\}} \max_{m \in \mathcal{M}} \max_{k \in \mathcal{K}} \frac{1}{n} \text{w}\left(x_s^n(m,k)\right), \tag{53}$$

and

$$\omega_{\text{L},n} \triangleq n\mu_{\text{L},n}, \qquad \omega_{\text{H},n} \triangleq n\mu_{\text{H},n}. \tag{54}$$

Let

$$\text{D}_s \triangleq \mathbb{E}_{Q_0}\left[\frac{(\tilde{Q}_s(Y) - Q_0(Y))(\tilde{Q}_1(Y) - \tilde{Q}_2(Y))}{Q_0(Y)^2}\right], \tag{55a}$$

$$\Gamma_s \triangleq \mathbb{E}_{Q_0}\left[\frac{(\tilde{Q}_s(Y) - Q_0(Y))(\tilde{Q}_1(Y) - \tilde{Q}_2(Y))^2}{Q_0(Y)^3}\right], \tag{55b}$$

$$\text{V}_s^* \triangleq \Delta + \mu_{\text{H},n} \cdot |\Gamma_s|, \tag{55c}$$

and define

$$\tau \triangleq \frac{n\mu_{\text{L},n}}{2}\left(\text{D}_2 + \text{D}_1\right). \tag{56}$$

Total Variation Distance and a Hypothesis Testing Problem: We relate the total variation distance to a hypothesis testing problem as follows. Consider the following inequality:

$$d_{\text{TV}}(Q_1^n, Q_2^n) \geq 1 - \alpha_n - \beta_n, \tag{57}$$

where α_n and β_n are false alarm and missed detection probabilities of a (possibly suboptimal) hypothesis test for the following setup. Under the null hypothesis $\mathcal{H} = 1$, Y^n is distributed according to Q_1^n and under the alternative hypothesis $\mathcal{H} = 2$, it is distributed according to Q_2^n. Define the following suboptimal test:

$$\mathcal{A}(y^n) \triangleq \mathbb{1}\left\{\sum_{i=1}^n \mathcal{T}_{\text{test}}(y_i) > \tau\right\}, \tag{58}$$

where

$$\mathcal{T}_{\text{test}}(y_i) \triangleq \frac{\tilde{Q}_1(y_i) - \tilde{Q}_2(y_i)}{Q_0(y_i)}. \tag{59}$$

The following lemma provides upper bounds on the false alarm and missed detection probabilities of this hypothesis test.

Lemma 5: We have:

$$\alpha_n \leq 1 - \Phi\left(\frac{1}{2}\mu_{\text{L},n}\sqrt{n\Delta}\right) + \frac{\sqrt{n}\mu_{\text{L},n}\mu_{\text{H},n}|\Gamma_1|}{4\sqrt{2\pi\Delta}} + O\left(\frac{1}{\sqrt{n}}\right). \tag{60}$$

and

$$\beta_n \leq 1 - \Phi\left(\frac{1}{2}\mu_{\text{L},n}\sqrt{n\Delta}\right) + \frac{\sqrt{n}\mu_{\text{L},n}\mu_{\text{H},n}|\Gamma_2|}{4\sqrt{2\pi\Delta}} + O\left(\frac{1}{\sqrt{n}}\right). \tag{61}$$

Proof: See Appendix H. ∎

Combining (60) and (61) with (57), we get:

$$d_{\text{TV}}(Q_1^n, Q_2^n) \geq 2\Phi\left(\frac{1}{2}\mu_{\text{L},n}\sqrt{n\Delta}\right) - 1 - \frac{\sqrt{n}\mu_{\text{L},n}\mu_{\text{H},n}(|\Gamma_1| + |\Gamma_2|)}{4\sqrt{2\pi\Delta}} + O\left(\frac{1}{\sqrt{n}}\right), \tag{62}$$

which can be equivalently written as follows:

$$d_{\text{TV}}(Q_1^n, Q_2^n) \geq 2\Phi\left(\frac{1}{2}\omega_{\text{L},n}\sqrt{\frac{\Delta}{n}}\right) - 1 - \frac{\omega_{\text{L},n}\omega_{\text{H},n}(|\Gamma_1| + |\Gamma_2|)}{4n\sqrt{2n\pi\Delta}} + O\left(\frac{1}{\sqrt{n}}\right). \tag{63}$$

The proof is continued by showing that there exists a sub-codebook which satisfies (63) and its size is at least $\frac{|\mathcal{M}| \cdot |\mathcal{K}|}{\sqrt{n}}$ and it is concluded by applying Fano's inequality. See Appendix I for details. ∎

Remark 3: Assume that (47) and (49) hold. Then, the choice of $\gamma_1 = \gamma_2$ minimizes the RHS of (26) and maximizes the RHS of (27) in Theorem 1. Thus, it can be seen a small key-length suffices for the optimal setup of Theorem 3.
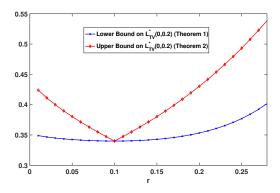
Fig. 4. Lower and upper bounds on $L^*_{\text{TV}}(0, 0.2)$.

C. Examples

In this section, we provide examples that satisfy both conditions of Theorem 3 in (47) and (49) and discuss scenarios in which we can establish optimal covert communication rates.

First, consider the setup of the example in Fig. 2. This example satisfies the constraints (47) and (49). Choosing $\delta = 0.2$ for this example, Theorem 3 yields $L^*_{\text{TV}}(0, 0.2) = 0.3399$. For other values of $r \triangleq \tilde{Q}_2(0) = \tilde{Q}_2(1)$, the upper and lower bounds on $L^*_{\text{TV}}(0, 0.2)$ are plotted in Fig. 4. It can also be verified from the figure that the two bounds match for $r = 0.1$.

Next, consider the following Gaussian setup. Let $\tilde{Q}_s$ and Q_0 be $\mathcal{N}(1, \sigma_s^2)$ and $\mathcal{N}(0, \sigma_0^2)$, respectively, where $\sigma_s^2, \sigma_0^2 \geq 0$ and $\frac{4}{3}\sigma_0^2 \geq \sigma_1^2 \geq \sigma_2^2$. Define the following parameters:

$$D_{\text{G}} \triangleq D(\tilde{Q}_s \| Q_0) = \frac{1}{2} \log \frac{\sigma_0^2}{\sigma_s^2} + \frac{\log e}{2\sigma_0^2} \cdot (\sigma_s^2 + 1), \qquad s \in \{1, 2\}, \tag{64}$$

$$\chi_{\text{G},s} \triangleq \chi_2\left(\tilde{Q}_s \| Q_0\right) = \frac{\sigma_0}{\sqrt{2}\sigma_s} \exp\left(\frac{\frac{1}{2\sigma_0^2}}{1 - \frac{\sigma_s^2}{2\sigma_0^2}}\right) \frac{1}{\sqrt{1 - \frac{\sigma_s^2}{2\sigma_0^2}}} - 1, \tag{65}$$

$$\rho_{\text{G}} \triangleq \rho(\tilde{Q}_1, \tilde{Q}_2, Q_0) = \frac{\sigma_0}{\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(\frac{\frac{1}{2\sigma_0^2}}{1 - \frac{\sigma_1^2\sigma_2^2}{\sigma_0^2(\sigma_1^2 + \sigma_2^2)}}\right) \cdot \frac{1}{\sqrt{1 - \frac{\sigma_1^2\sigma_2^2}{\sigma_0^2(\sigma_1^2 + \sigma_2^2)}}} - 1, \tag{66}$$

$$\Delta_{\text{G}} \triangleq \chi_{\text{G},1} + \chi_{\text{G},2} - 2\rho_{\text{G}}. \tag{67}$$

For this Gaussian setup, the conditions (47) and (49) of Theorem 3 translate to the following:

$$\rho_{\text{G}} \leq \min\{\chi_{\text{G},1}, \chi_{\text{G},2}\}. \tag{68}$$

and

$$\sigma_0^2 = \frac{\log e \cdot (\sigma_2^2 - \sigma_1^2)}{\log \frac{\sigma_2^2}{\sigma_1^2}}, \tag{69}$$

Under the above assumptions, we specialize Theorem 3 to the Gaussian setup.

Corollary 2: For any nonnegative σ_s^2 and σ_0^2 such that

$$\frac{4}{3}\sigma_0^2 \geq \sigma_1^2 \geq \sigma_2^2, \tag{70}$$

and conditions (68)–(69) are satisfied, we have

$$L^*_{\text{TV}}(0^+, \delta) = \frac{2\Phi^{-1}\left(\frac{1+\delta}{2}\right)}{\sqrt{\Delta_{\text{G}}}} \cdot D_{\text{G}}. \tag{71}$$

Proof: See Appendix J. ∎

For the proposed Gaussian setup, Fig. 5 shows the optimal $L^*_{\text{TV}}(0^+, 0.2)$ versus σ_1^2 for different values of σ_2^2. The variance σ_1^2 (resp. σ_2^2) corresponds to the channel state with a larger (resp. smaller) variance of noise. Notice that for each value of σ_2^2, there exists a non-empty interval of σ_1^2 that simultaneously satisfies the three conditions (68)–(70). For $\sigma_2^2 = 0.5$ and $\sigma_1^2 > 0.57$, $L^*_{\text{TV}}(0, 0.2)$ is a decreasing function of σ_1^2. For other values of σ_2^2, a similar observation holds. The optimal covert communication rate increases as σ_2^2 increases.
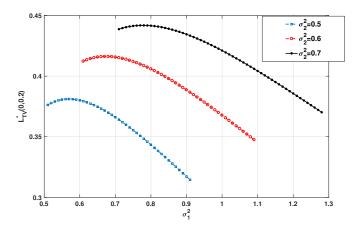
Fig. 5. $L_{\text{TV}}^*(0^+, 0.2)$ for the Gaussian example.

D. Concealing the message transmission vs concealing the channel state

To conclude this section, we revisit the discussion concerning the difference of concealing the message transmission versus concealing the channel state. In this discussion, we remove the assumption stated in Remark 2 and for simplicity, we assume that the length of the key is infinite.

Recall the example of Fig. 3. The following observations are in order:

1) As mentioned before, the assumption of Remark 2 does not hold, because $W_1(2|1) \neq 0$, while $W_1(2|0) = 0$ and also $W_2(2|1) \neq 0$, while $W_2(2|0) = 0$. Thus our results from the previous subsections, do not hold as the assumptions are violated.

2) Consider the achievable scheme of the previous section with the choice $\mu_{1,n} = \mu_{2,n} \triangleq \mu_n$. As shown in Appendix G, under the total variation covertness criterion, if the number of messages $|\mathcal{M}|$ satisfies

$$\log |\mathcal{M}| \leq \frac{\sqrt{\ln \frac{1}{1-\delta}}}{2} \cdot \sqrt{n} \log n + o\left(\sqrt{n} \log n\right), \tag{72}$$

then the desired constraints on the maximum error probability and covertness metric are satisfied. The above result shows that it is possible to communicate $\Omega(\sqrt{n} \log n)$ bits over n channel uses covertly in the sense of concealing of the channel state. However, covert communication in the sense of concealing message transmission is not feasible for this example in the sense that the KL-divergence $D\left(Q_s^n \| Q_0^{\times n}\right)$ tends to infinity as $n \to \infty$.

3) As mentioned in the previous item, the communication rate is at least of the order of $\Omega\left(\frac{\log n}{\sqrt{n}}\right)$, which is somewhat surprising, because the communication rate is usually of order $\Theta\left(\frac{1}{\sqrt{n}}\right)$ for covert communication problems [2], [3], as we have observed in our results in the previous subsections. This again shows the importance of the absolutely continuous assumption in Remark 2.

4) Although we will not discuss the details, it is worth mentioning that unlike the achievable scheme of the previous subsection, the choice $\mu_{1,n} = \mu_{2,n}$ is the unique choice which enables sending a positive number of bits under the covert communication constraint.

IV. CONCLUSION

In this paper, we considered covert communication over a compound channel. In this setting, the covertness is defined as the ability of a malicious party in distinguishing the channel state. We provided inner and outer bounds to the optimal throughput-key region and showed that the square-root law holds for this setup. For a special case, the bounds match and establish the optimal throughput. Some examples, including a Gaussian setup, are given to discuss the bounds.

A promising avenue for future work is to study covert communication over a channel with an i.i.d. state S^n available at the encoder, the decoder or both. The distribution that generates S^n can either be P_{S_1} or P_{S_2}. The main goal in this stochastic–as opposed to deterministic in this paper—setting is to design a code to remain covert such that an adversary who is observing the channel output does not learn the channel state distribution, i.e., P_{S_1} or P_{S_2}. The problem of covert communication over a channel with state has been previously studied in [5]. One can consider an alternative formulation to conceal the distribution of state instead of the presence or absence of message transmission as was done in [5].

APPENDIX A

PROOF OF LEMMA 1

First, we present the following theorem (Berry-Esseen Theorem) which will be used repeatedly in the proofs, later.

Theorem 4 (Thm 1.6 in [18]): Let $X^n \triangleq (X_1, \ldots, X_n)$ be a collection of n independent random variables where each random variable has a zero mean, variance $\sigma_i^2 \triangleq \mathbb{E}\left[X_i^2\right] > 0$ and third absolute moment $T_i \triangleq \mathbb{E}\left[|X_i|^3\right] < \infty$. Define the average variance and average third absolute moment as $\sigma^2 \triangleq \frac{1}{n}\sum_{i=1}^n \sigma_i^2$ and $T \triangleq \frac{1}{n}\sum_{i=1}^n T_i$, respectively. Then, we have:

$$\sup_{a \in \mathbb{R}} \left| \mathbb{P}\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n X_i < a \right) - \Phi(a) \right| \leq \frac{6T}{\sigma^3 \sqrt{n}}. \tag{73}$$

Now, consider the following identity for the total variation distance:

$$d_{\mathrm{TV}}\left(Q_1^{\times n}, Q_2^{\times n}\right) = \mathbb{P}_{Q_1^{\times n}}\left(\log\left(\frac{Q_1^{\times n}}{Q_2^{\times n}}\right) \geq 0 \right) - \mathbb{P}_{Q_2^{\times n}}\left(\log\left(\frac{Q_1^{\times n}}{Q_2^{\times n}}\right) \geq 0 \right) \tag{74}$$

$$= \mathbb{P}_{Q_1^{\times n}}\left(\sum_{i=1}^n \log \frac{Q_{1,n}(Y_i)}{Q_{2,n}(Y_i)} \geq 0 \right) - \mathbb{P}_{Q_2^{\times n}}\left(\sum_{i=1}^n \log \frac{Q_{1,n}(Y_i)}{Q_{2,n}(Y_i)} \geq 0 \right). \tag{75}$$

Now, we bound each of the probabilities in (75). First, we analyze the first probability. The mean and variance of $\sum_{i=1}^n \log \frac{Q_{1,n}(Y_i)}{Q_{2,n}(Y_i)}$ are given by the following:

$$\mathbb{E}\left[\sum_{i=1}^n \log \frac{Q_{1,n}(Y_i)}{Q_{2,n}(Y_i)} \right] = nD(Q_{1,n}\|Q_{2,n}) \triangleq n\mathsf{d}_1, \tag{76}$$

and

$$\mathbb{V}\left[\sum_{i=1}^n \log \frac{Q_{1,n}(Y_i)}{Q_{2,n}(Y_i)} \right] = n\left(\sum_y Q_{1,n}(y)\left(\log \frac{Q_{1,n}(y)}{Q_{2,n}(y)} \right)^2 - D^2(Q_{1,n}\|Q_{2,n}) \right) \tag{77}$$

$$\triangleq n\mathsf{v}_1. \tag{78}$$

Employing Theorem 4, we obtain

$$\mathbb{P}_{Q_1^{\times n}}\left(\sum_{i=1}^n \log \frac{Q_{1,n}(Y_i)}{Q_{2,n}(Y_i)} \geq 0 \right) \leq \Phi\left(\frac{\sqrt{n}\mathsf{d}_1}{\sqrt{\mathsf{v}_1}} \right) + \frac{6\mathsf{t}_1}{\sqrt{n}\mathsf{v}_1^{\frac{3}{2}}}, \tag{79}$$

where

$$\mathsf{t}_1 \triangleq \mathbb{E}_{Q_{1,n}}\left[\left| \log \frac{Q_{1,n}(Y)}{Q_{2,n}(Y)} - D(Q_{1,n}\|Q_{2,n}) \right|^3 \right]. \tag{80}$$

Similarly, we get:

$$\mathbb{P}_{Q_2^{\times n}}\left(\sum_{i=1}^n \log \frac{Q_{1,n}(Y_i)}{Q_{2,n}(Y_i)} \geq 0 \right) \geq 1 - \Phi\left(\frac{\sqrt{n}\mathsf{d}_2}{\sqrt{\mathsf{v}_2}} \right) - \frac{6\mathsf{t}_2}{\sqrt{n}\mathsf{v}_2^{\frac{3}{2}}}, \tag{81}$$

where we define:

$$\mathsf{d}_2 \triangleq D(Q_{2,n}\|Q_{1,n}), \tag{82}$$

$$\mathsf{v}_2 \triangleq \sum_y Q_{2,n}(y)\left(\log \frac{Q_{1,n}(y)}{Q_{2,n}(y)} \right)^2 - D^2(Q_{2,n}\|Q_{1,n}), \tag{83}$$

$$\mathsf{t}_2 \triangleq \mathbb{E}_{Q_{2,n}}\left[\left| \log \frac{Q_{1,n}(Y)}{Q_{2,n}(Y)} + D(Q_{2,n}\|Q_{1,n}) \right|^3 \right]. \tag{84}$$

Now, consider the following set of approximations:

$$\mathsf{d}_1 = D(Q_{1,n}\|Q_{2,n}) \tag{85}$$

$$= \sum_y Q_{1,n}(y) \log \frac{Q_{1,n}(y)}{Q_{2,n}(y)} \tag{86}$$

$$= \mathbb{E}_{Q_{2,n}} \left[\frac{Q_{1,n}}{Q_{2,n}} \log \frac{Q_{1,n}}{Q_{2,n}} \right] \tag{87}$$

$$= \mathbb{E}_{Q_{2,n}} \left[\frac{(1-\mu_{1,n})Q_0 + \mu_{1,n}\bar{Q}_1}{(1-\mu_{2,n})Q_0 + \mu_{2,n}\bar{Q}_2} \log \frac{(1-\mu_{1,n})Q_0 + \mu_{1,n}\bar{Q}_1}{(1-\mu_{2,n})Q_0 + \mu_{2,n}\bar{Q}_2} \right] \tag{88}$$

$$= \mathbb{E}_{Q_{2,n}} \left[\left(1 + \frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{(1-\mu_{2,n})Q_0 + \mu_{2,n}\bar{Q}_2} \right) \log \left(1 + \frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{(1-\mu_{2,n})Q_0 + \mu_{2,n}\bar{Q}_2} \right) \right] \tag{89}$$

$$= \mathbb{E}_{Q_{2,n}} \Bigg[\left(1 + \frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{(1-\mu_{2,n})Q_0 + \mu_{2,n}\bar{Q}_2} \right) \times$$
$$\left(\frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{(1-\mu_{2,n})Q_0 + \mu_{2,n}\bar{Q}_2} - \frac{1}{2} \left(\frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{(1-\mu_{2,n})Q_0 + \mu_{2,n}\bar{Q}_2} \right)^2 \right) \Bigg] + o\left(\frac{1}{n} \right) \tag{90}$$

$$= \frac{1}{2} \mathbb{E}_{Q_{2,n}} \left[\left(\frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{(1-\mu_{2,n})Q_0 + \mu_{2,n}\bar{Q}_2} \right)^2 \right] + o\left(\frac{1}{n} \right) \tag{91}$$

$$= \frac{1}{2} \mathbb{E}_{Q_0} \left[\left(\frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{Q_0} \right)^2 \right] + o\left(\frac{1}{n} \right) \tag{92}$$

$$= \frac{1}{2} \mu_{1,n}^2 \chi_2\left(\bar{Q}_1 \| Q_0 \right) + \frac{1}{2} \mu_{2,n}^2 \chi_2\left(\bar{Q}_2 \| Q_0 \right) - \mu_{1,n}\mu_{2,n}\rho\left(\bar{Q}_1, \bar{Q}_2, Q_0 \right) + o\left(\frac{1}{n} \right), \tag{93}$$

where (89) follows from the Taylor expansion $\log(1 + x) = x - \frac{1}{2}x^2 + o(x^2)$ at $x = 0$; (92) follows because

$$Q_{2,n} = (1-\mu_{2,n})Q_0 + \mu_{2,n}\bar{Q}_2 \to Q_0. \tag{94}$$

Now, define:

$$\bar{\Omega}\left(\mu_{1,n}, \mu_{2,n} \right) \triangleq \mu_{1,n}^2 \chi_2\left(\bar{Q}_1 \| Q_0 \right) + \mu_{2,n}^2 \chi_2\left(\bar{Q}_2 \| Q_0 \right) - 2\mu_{1,n}\mu_{2,n}\rho\left(\bar{Q}_1, \bar{Q}_2, Q_0 \right). \tag{95}$$

The equality (93) can be expressed in terms of $\bar{\Omega}\left(\mu_{1,n}, \mu_{2,n} \right)$ as the following:

$$\mathsf{d}_1 = \frac{1}{2}\bar{\Omega}\left(\mu_{1,n}, \mu_{2,n} \right) + o\left(\frac{1}{n} \right). \tag{96}$$

Next, we consider the following set of approximations:

$$\sum_y Q_1(y) \left(\log \frac{Q_1(y)}{Q_2(y)} \right)^2 = \sum_y \left((1-\mu_{1,n})Q_0(y) + \mu_{1,n}\bar{Q}_1(y) \right) \left(\log \frac{(1-\mu_{1,n})Q_0(y) + \mu_{1,n}\bar{Q}_1(y)}{(1-\mu_{2,n})Q_0(y) + \mu_{2,n}\bar{Q}_2(y)} \right)^2 \tag{97}$$

$$= \sum_y \left((1-\mu_{1,n})Q_0(y) + \mu_{1,n}\bar{Q}_1(y) \right) \left(\log \left(1 + \frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{(1-\mu_{2,n})Q_0(y) + \mu_{2,n}\bar{Q}_2(y)} \right) \right)^2 \tag{98}$$

$$= \sum_y \left((1-\mu_{1,n})Q_0(y) + \mu_{1,n}\bar{Q}_1(y) \right) \left(\frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{(1-\mu_{2,n})Q_0(y) + \mu_{2,n}\bar{Q}_2(y)} \right)^2 + o\left(\frac{1}{n} \right) \tag{99}$$

$$= \sum_y Q_0(y) \left(\frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{Q_0(y)} \right)^2 + o\left(\frac{1}{n} \right) \tag{100}$$

$$= \mathbb{E}_{Q_0} \left[\left(\frac{\mu_{1,n}(\bar{Q}_1 - Q_0) - \mu_{2,n}(\bar{Q}_2 - Q_0)}{Q_0} \right)^2 \right] + o\left(\frac{1}{n} \right) \tag{101}$$

$$= \mu_{1,n}^2 \chi_2\left(\bar{Q}_1 \| Q_0 \right) + \mu_{2,n}^2 \chi_2\left(\bar{Q}_2 \| Q_0 \right) - 2\mu_{1,n}\mu_{2,n}\rho\left(\bar{Q}_1, \bar{Q}_2, Q_0 \right) + o\left(\frac{1}{n} \right) \tag{102}$$

$$= \bar{\Omega}\left(\mu_{1,n}, \mu_{2,n}\right) + o\left(\frac{1}{n}\right), \tag{103}$$

where (99) follows from Taylor expansion $\log(1+x) = x + o(x)$ at $x = 0$, (100) follows because

$$(1 - \mu_{s,n})Q_0(y) + \mu_{s,n}\bar{Q}_s(y) \to Q_0(y),$$

as $\mu_{s,n} \to 0$. Combining (79), (93), (103) and definition of v_1 in (78), we have:

$$\Phi\left(\frac{\sqrt{n}d_1}{\sqrt{v_1}}\right) = \Phi\left(\frac{1}{2}\sqrt{n\bar{\Omega}\left(\mu_{1,n}, \mu_{2,n}\right)}\right) + o\left(\frac{1}{n}\right). \tag{104}$$

Similarly,

$$\Phi\left(\frac{\sqrt{n}d_2}{\sqrt{v_2}}\right) = \Phi\left(\frac{1}{2}\sqrt{n\bar{\Omega}\left(\mu_{1,n}, \mu_{2,n}\right)}\right) + o\left(\frac{1}{n}\right). \tag{105}$$

Following similar steps leading to (103), one can show that $t_s = O\left(\frac{1}{n^{\frac{3}{2}}}\right)$. Combining (75), (79), (81), (104) and (105), and considering the fact that $v_s = \Theta\left(\frac{1}{n}\right)$, we obtain the following:

$$d_{\text{TV}}\left(Q_1^{\times n}, Q_2^{\times n}\right) \le \Phi\left(\frac{\sqrt{n}d_1}{\sqrt{v_1}}\right) + \Phi\left(\frac{\sqrt{n}d_2}{\sqrt{v_2}}\right) - 1 + \frac{6t_1}{\sqrt{n}v_1^{\frac{3}{2}}} + \frac{6t_2}{\sqrt{n}v_2^{\frac{3}{2}}} \tag{106}$$

$$\le 2\Phi\left(\frac{1}{2}\sqrt{n\bar{\Omega}\left(\mu_{1,n}, \mu_{2,n}\right)}\right) - 1 + O\left(\frac{1}{\sqrt{n}}\right). \tag{107}$$

Considering the fact that $\gamma_s = \sqrt{n}\mu_{s,n}$, completes the proof of the lemma.

APPENDIX B
PROOF OF LEMMA 2

First, we state the following lemma.

Lemma 6 (Corollary VII.2 in [19]): For all positive η_s, we have:

$$\mathbb{E}_{\mathcal{C}_s}\left[d_{\text{TV}}\left(Q_s^n, Q_s^{\times n}\right)\right] \le \frac{1}{2}\sqrt{\frac{2^{\eta_s}}{|\mathcal{M}| \cdot |\mathcal{K}|}} + \frac{1}{2}\mathbb{P}_{W_s^n P_s^{\times n}}\left(\sum_{i=1}^n \log \frac{W_s(Y_i|X_i)}{Q_{s,n}(Y_i)} \ge \eta_s\right). \tag{108}$$

Choose

$$\eta_s = n\left(1 + \frac{\kappa}{2}\right)I(P_{s,n}, W_s), \tag{109}$$

for some positive κ. Notice that from [2], the following approximations hold:

$$I(P_s, W_s) = \frac{1}{\sqrt{n}}\gamma_s D(W_s\|Q_0|\bar{P}_s) + o\left(\frac{1}{\sqrt{n}}\right) \tag{110}$$

$$\triangleq \frac{\gamma_s}{\sqrt{n}}\bar{\mathbb{D}}_s + o\left(\frac{1}{\sqrt{n}}\right), \tag{111}$$

Thus,

$$\eta_s = \gamma_s\sqrt{n}\left(1 + \frac{\kappa}{2}\right)\bar{\mathbb{D}}_s + o(\sqrt{n}). \tag{112}$$

Also, we have

$$\sum_x \sum_y P_{s,n}(x)W_s(y|x)\left(\log \frac{W_s(y|x)}{Q_{s,n}(y)}\right)^2 = \frac{\gamma_s}{\sqrt{n}}\sum_{x\neq 0}\sum_y \bar{P}_s(x)W_s(y|x)\left(\log \frac{W_s(y|x)}{Q_0(y)}\right)^2 + o\left(\frac{1}{\sqrt{n}}\right) \tag{113}$$

$$\triangleq \frac{\gamma_s}{\sqrt{n}}\bar{\mathbb{V}}_s + o\left(\frac{1}{\sqrt{n}}\right). \tag{114}$$

Now, consider the probability term in (108) as follows:

$$\mathbb{P}_{W_s^n P_s^{\times n}}\left(\sum_{i=1}^n \log \frac{W_s(Y_i|X_i)}{Q_{s,n}(Y_i)} \ge \eta_s\right) \tag{115}$$

$$= \mathbb{P}_{W_s^n P_s^{\times n}}\left(\sum_{i=1}^n \log \frac{W_s(Y_i|X_i)}{Q_{s,n}(Y_i)} \ge n(1 + \frac{\kappa}{2})I(P_{s,n}, W_s)\right) \tag{116}$$

$$= \mathbb{P}_{W_s^n P_s^{\times n}} \left(\sum_{i=1}^{n} \log \frac{W_s(Y_i|X_i)}{Q_{s,n}(Y_i)} - nI(P_{s,n}, W_s) \geq \frac{n\kappa}{2} I(P_{s,n}, W_s) \right) \tag{117}$$

$$= \mathbb{P}_{W_s^n P_s^{\times n}} \left(\sum_{i=1}^{n} \log \frac{W_s(Y_i|X_i)}{Q_{s,n}(Y_i)} - nI(P_{s,n}, W_s) \geq \frac{\sqrt{n}\kappa\gamma_s \bar{\mathbb{D}}_s}{2} + o(\sqrt{n}) \right) \tag{118}$$

$$\leq \mathbb{P}_{W_s^n P_s^{\times n}} \left(\left| \sum_{i=1}^{n} \log \frac{W_s(Y_i|X_i)}{Q_{s,n}(Y_i)} - nI(P_{s,n}, W_s) \right| \geq \frac{\sqrt{n}\kappa\gamma_s \bar{\mathbb{D}}_s}{2} + o(\sqrt{n}) \right) \tag{119}$$

$$\leq \frac{\sqrt{n}\gamma_s \bar{\mathbb{V}}_s}{4n\kappa^2 \gamma_s^2 \bar{\mathbb{D}}_s^2} + o\left(\frac{1}{\sqrt{n}}\right) \tag{120}$$

$$= O\left(\frac{1}{\sqrt{n}}\right), \tag{121}$$

where the last inequality follows from Chebyshev's inequality and the fact that

$$\mathbb{V}\left[\sum_{i=1}^{n} \log \frac{W_s(Y_i|X_i)}{Q_{s,n}(Y_i)} - nI(P_{s,n}, W_s) \right] \leq n \left(\sum_x \sum_y P_{s,n}(x) W_s(y|x) \left(\log \frac{W_s(y|x)}{Q_{s,n}(y)} \right)^2 \right) \tag{122}$$

$$\leq \sqrt{n}\gamma_s \bar{\mathbb{V}}_s + o(\sqrt{n}). \tag{123}$$

Combining (108) and (121), we get:

$$\mathbb{E}_{\mathcal{C}_s}\left[d_{\mathrm{TV}}\left(Q_s^n, Q_s^{\times n} \right) \right] \leq \frac{1}{2} \sqrt{\frac{2^{\eta_s}}{|\mathcal{M}| \cdot |\mathcal{K}|}} + O\left(\frac{1}{\sqrt{n}}\right). \tag{124}$$

Now, we choose:

$$\log |\mathcal{M}| + \log |\mathcal{K}| \geq \gamma_s \sqrt{n} \left(1 + \kappa \right) \bar{\mathbb{D}}_s. \tag{125}$$

With the above choice, we can further upper bound (124) as follows:

$$\mathbb{E}_{\mathcal{C}_s}\left[d_{\mathrm{TV}}\left(Q_s^n, Q_s^{\times n} \right) \right] = O\left(\frac{1}{\sqrt{n}}\right). \tag{126}$$

This completes the proof of the lemma.

APPENDIX C
PROOF OF LEMMA 3

We provide the analysis of the rate constraint in the following. Since the ML decoder results in the smallest probability of error for a given codebook, we can bound the average probability of error using Shannon's achievability bound (threshold decoding) [17, Thm 2] as follows. For every $s \in \{1, 2\}$ and $k \in \mathcal{K}$, the average error probability satisfies:

$$\mathbb{E}\left[\bar{P}_e(s, k) \right] \leq \mathbb{P}\left(\log \frac{W_s(Y^n|X^n)}{Q_s^{\times n}(Y^n)} \leq \log |\mathcal{M}| + \frac{1}{6} \log n \right) + n^{-\frac{1}{6}}. \tag{127}$$

Recall the definition of $\bar{\mathbb{V}}_s$ from (114) and consider the above probability:

$$\mathbb{P}\left(\log \frac{W_s(Y^n|X^n)}{Q_s^{\times n}(Y^n)} \leq \log |\mathcal{M}| + \frac{1}{6} \log n \right) \leq \mathbb{P}\left(\log \frac{W_s(Y^n|X^n)}{Q_s^{\times n}(Y^n)} \leq nI(P_{s,n}, W_s) - n^{\frac{1}{3}} + \frac{1}{6} \log n \right) \tag{128}$$

$$= \mathbb{P}\left(\log \frac{W_s(Y^n|X^n)}{Q_s^{\times n}(Y^n)} - nI(P_{s,n}, W_s) \leq -n^{\frac{1}{3}} + \frac{1}{6} \log n \right) \tag{129}$$

$$\leq \mathbb{P}\left(\left| \log \frac{W_s(Y^n|X^n)}{Q_s^{\times n}(Y^n)} - nI(P_{s,n}, W_s) \right| \geq n^{\frac{1}{3}} - \frac{1}{6} \log n \right) \tag{130}$$

$$\leq \frac{\sqrt{n}\gamma_s \bar{\mathbb{V}}_s}{(n^{\frac{1}{3}} - \frac{1}{6} \log n)^2} \tag{131}$$

$$= O\left(n^{-\frac{1}{6}} \right), \tag{132}$$

where the last inequality follows from Chebyshev's inequality and the steps leading to (123). This completes the proof of lemma.

APPENDIX D

PROOF OF EXISTENCE OF A CODEBOOK WITH DESIRED PROPERTIES

First, we state the following theorem which will be used later in the proof.

Theorem 5 (McDiarmid's theorem [20]): Let $\{X_k\}_{k=1}^n$ be independent random variables defined on the set $\mathcal{X}$. Consider a random variable $U = f(X^n)$ where f is a function satisfying the following bounded difference property:

$$\sup_{x_1,\ldots,x_n,x_i' \in \mathcal{X}} f(x_1,\ldots,x_i,\ldots,x_n) - f(x_1,\ldots,x_i',\ldots,x_n) \leq d_i, \qquad \forall i \in \{1,\ldots,n\}, \tag{133}$$

for some positive numbers $d_i, 1 \leq i \leq n$. Then, for every $r > 0$,

$$\mathbb{P}(U - \mathbb{E}[U] > r) \leq \exp\left(\frac{-2r^2}{\sum_{i=1}^n d_i^2}\right). \tag{134}$$

Step 1: Probability of Satisfying the Covertness Condition:

Choose positive real numbers $D_n = \frac{1}{\sqrt{\log n}}$ and $a = n^{-\frac{1}{12}}$; and define the following events:

$$\mathcal{B}_s \triangleq \left\{ \forall k \in \mathcal{K}, \ \exists \mathcal{M}_{s,k} \subset \mathcal{M}, \ \mathcal{M}_s = \mathcal{M}_{s,1} \times \ldots \times \mathcal{M}_{s,k} : |\mathcal{M}_{s,k}| = (1-a)|\mathcal{M}|, \right.$$
$$\left. Q_{\mathcal{M}_s}^n(y^n) \triangleq \frac{1}{|\mathcal{M}_s||\mathcal{K}|} \sum_{k \in \mathcal{K}} \sum_{m \in \mathcal{M}_{s,k}} W_s^n(y^n|X_s^n(m,k)), \ d_{\mathrm{TV}}(Q_{\mathcal{M}_s}^n, Q_s^{\times n}) > D_n \right\}, \tag{135}$$

for $s = 1, 2$.

Consider the probability of this event:

$$\mathbb{P}(\mathcal{B}_s) = \mathbb{P}\left(\forall k \in \mathcal{K}, \ \exists \mathcal{M}_{s,k} \subset \mathcal{M} : |\mathcal{M}_{s,k}| = (1-a)|\mathcal{M}|, \ d_{\mathrm{TV}}(Q_{\mathcal{M}_s}^n, Q_s^{\times n}) > D_n\right) \tag{136}$$

$$\overset{(a)}{\leq} \mathbb{P}\left(\forall k \in \mathcal{K}, \ \exists \mathcal{M}_{s,k} \subset \mathcal{M} : |\mathcal{M}_{s,k}| = (1-a)|\mathcal{M}|, \ d_{\mathrm{TV}}(Q_{\mathcal{M}_s}^n, Q_s^{\times n}) - \mathbb{E}[d_{\mathrm{TV}}(Q_{\mathcal{M}_s}^n, Q_s^{\times n})] > \frac{D_n}{2}\right) \tag{137}$$

$$\overset{(b)}{\leq} \sum_{\substack{\mathcal{M}_{s,k} \subset \mathcal{M}, k \in \mathcal{K} : \\ |\mathcal{M}_{s,k}| = (1-a)|\mathcal{M}|}} \mathbb{P}\left(d_{\mathrm{TV}}(Q_{\mathcal{M}_s}^n, Q_s^{\times n}) - \mathbb{E}[d_{\mathrm{TV}}(Q_{\mathcal{M}_s}^n, Q_s^{\times n})] > \frac{D_n}{2}\right), \tag{138}$$

where

- (a) follows because according to (34), the expectation of the total variation distance is upper bounded by $O(n^{-\frac{1}{2}}) \leq \frac{D_n}{2}$ for sufficiently large n, because

$$\log|\mathcal{M}_s| + \log|\mathcal{K}| = \log|\mathcal{M}| + \log|\mathcal{K}| + \log(1-a) \geq (1+\kappa')\sqrt{n} \cdot \max_{s \in \{1,2\}} \gamma_s D(W_s\|Q_0|\bar{P}_s),$$

  for some $\kappa' > 0$ that approaches κ as $n \to \infty$,
- (b) follows from the union bound.

In order to upper bound (138), we use the following lemma due to Liu, *et. al.* [21] (see also [8]), showing that the total variation distance of a random codebook is concentrated around its expectation.

Lemma 7 (Theorem 31 in [21]): Consider a channel with the probability transition W and an input probability distribution P_X. Suppose that P_Y is the output distribution of the channel with the input distribution P_X. Let $\mathcal{C} = \{X_1, \cdots, X_\mathsf{M}\}$ be a random codebook with i.i.d. codewords drawn from P_X and $\hat{P}_Y(.) = \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} W(.|X_m)$ be the induced output distribution by $\mathcal{C}$. Then for any $\Delta > 0$, we have

$$\mathbb{P}(d_{\mathrm{TV}}(\hat{P}_Y, P_Y) - \mathbb{E}[(d_{\mathrm{TV}}(\hat{P}_Y, P_Y)] > \Delta) \leq \exp_e(-2\mathsf{M}\Delta^2). \tag{139}$$

Using the above lemma with the following specifications,

$$\mathcal{C} \leftarrow \mathcal{C}_s = \bigcup_{k \in |\mathcal{K}|} \{X_s^n(m,k) : m \in \mathcal{M}_{s,k}\}, \tag{140}$$

$$\mathsf{M} \leftarrow (1-a)|\mathcal{M}||\mathcal{K}|, \tag{141}$$

$$Y \leftarrow Y^n \tag{142}$$

$$W \leftarrow W_s^n, \tag{143}$$

$$\hat{P}_Y \leftarrow Q_{\mathcal{M}_s}^n, \tag{144}$$

$$P_Y \leftarrow Q_s^{\times n}, \tag{145}$$

$$\Delta \leftarrow \frac{D_n}{2}, \tag{146}$$

we get the following upper bound on the term (138) as follows:

$$\sum_{\substack{\mathcal{M}_{s,k} \subset \mathcal{M}, k \in \mathcal{K}: \\ |\mathcal{M}_{s,k}| = (1-a)|\mathcal{M}|}} \mathbb{P}\left(d_{\mathrm{TV}}(Q^n_{\mathcal{M}_s}, Q^{\times n}_s) - \mathbb{E}\left[d_{\mathrm{TV}}(Q^n_{\mathcal{M}_s}, Q^{\times n}_s)\right] > \frac{D_n}{2}\right)$$

$$\leq \sum_{\substack{\mathcal{M}_{s,k} \subset \mathcal{M}, k \in \mathcal{K}: \\ |\mathcal{M}_{s,k}| = (1-a)|\mathcal{M}|}} \exp\left(-\frac{1}{2}D_n^2(1-a)|\mathcal{M}||\mathcal{K}|\right) \tag{147}$$

$$= \binom{|\mathcal{M}|}{(1-a)|\mathcal{M}|}^{|\mathcal{K}|} \exp\left(-\frac{1}{2}(1-a)D_n^2|\mathcal{M}||\mathcal{K}|\right) \tag{148}$$

$$\overset{(d)}{\leq} \exp\left(-|\mathcal{M}||\mathcal{K}|\left(\frac{1}{2}(1-a)D_n^2 - a\log\frac{e}{a}\right)\right), \tag{149}$$

where (d) follows because

$$\binom{|\mathcal{M}|}{(1-a)|\mathcal{M}|} \leq \left(\frac{e}{a}\right)^{a|\mathcal{M}|}, \tag{150}$$

where the last inequality follows from [22]. Thus, in summary, we get:

$$\mathbb{P}(\mathcal{B}_1 \cup \mathcal{B}_2) \leq 2\exp\left(-|\mathcal{M}||\mathcal{K}|\left(\frac{1}{2}(1-a)D_n^2 - a\log\frac{e}{a}\right)\right) \tag{151}$$

$$\leq \exp\left(-|\mathcal{M}||\mathcal{K}|\frac{1}{4\log n}\right) \to 0. \tag{152}$$

for sufficiently large n.

Step 2: Probability of Satisfying the Reliability Condition:

Now, define the following event:

$$\mathcal{G} = \left\{\forall s \in \{1,2\}, k \in \mathcal{K}: \bar{P}_{\mathrm{e}}(s,k) \leq \mathbb{E}_{\mathcal{C}_s}\left[\bar{P}_{\mathrm{e}}(s,k)\right] + n^{-\frac{1}{6}}\right\}. \tag{153}$$

Using the McDiarmid's theorem, we shall prove the following lemma, which can be thought as dual to Lemma 7,

Lemma 8 (Concentration of average error probability around its expectation): For all $k \in \mathcal{K}$ and any $\Delta > 0$, we have

$$\mathbb{P}\left(\bar{P}_e(s,k) > \mathbb{E}[\bar{P}_e(s,k)] + \Delta\right) \leq \exp_e(-2|\mathcal{M}|\Delta^2). \tag{154}$$

This lemma with $\Delta = n^{-\frac{1}{6}}$ and the union bound imply

$$\mathbb{P}[\mathcal{G}^c] \leq 2|\mathcal{K}|\exp(-2|\mathcal{M}|n^{-\frac{1}{3}}) \to 0, \tag{155}$$

for sufficiently large n, since $\log|\mathcal{K}| = O(\sqrt{n})$.

Proof of Lemma 8: Note that for a random codebook $\mathcal{C}_s$, the random variable $\bar{P}_{\mathrm{e}}(s,k)$ is a function of $|\mathcal{M}|$ independent random variables $X^n_{s,1,k}, \cdots, X^n_{s,|\mathcal{M}|,k}$. In this case, $\bar{P}_{\mathrm{e}}(s,k)$ can be written as follows:

$$\bar{P}_{\mathrm{e}}(s,k) = 1 - \bar{P}_{\mathrm{c}}(s,k) \tag{156}$$

$$= 1 - \frac{1}{|\mathcal{M}|}\sum_{y^n}\max_m W_s^n\left(y^n|X^n_{s,m,k}\right) \tag{157}$$

where $\bar{P}_{\mathrm{c}}(s,k)$ is the probability of correct decoding. Define:

$$g(x^n_{s,1,k}, \cdots, x^n_{1,\mathcal{M},k}) \triangleq \frac{1}{|\mathcal{M}|}\sum_{y^n}\max_m W_s^n\left(y^n|x^n_{s,m,k}\right).$$

We next show that g satisfies the bounded difference property in (133), so $\bar{P}_e(s,k)$ satisfies the bounded difference property as well. Let $(x^n_{s,1,k}, \cdots, x^n_{s,|\mathcal{M}|,k})$ and $(\bar{x}^n_{s,1,k}, \cdots, \bar{x}^n_{s,|\mathcal{M}|,k})$ be two sequences differing only in the i-th coordinate, that is

$$\bar{x}^n_{s,m,k} = \begin{cases} x^n_{s,m,k} & \text{if } m \neq i \\ \text{arbitrary} & \text{if } m = i \end{cases} \tag{158}$$

Let the ML solution be defined as

$$\hat{m}_{y^n} \triangleq \underset{m}{\operatorname{argmax}}\ W_s^n(y^n|x^n_{s,m,k}). \tag{159}$$

We should show that $g(x_{s,1,k}^n, \ldots, x_{s,|\mathcal{M}|,k}^n) - g(\bar{x}_{s,1,k}^n, \ldots, \bar{x}_{s,|\mathcal{M}|,k}^n)$ is bounded. Consider,

$$g(x_{s,1,k}^n, \ldots, x_{s,|\mathcal{M}|,k}^n) - g(\bar{x}_{s,1,k}^n, \ldots, \bar{x}_{s,|\mathcal{M}|,k}^n)$$

$$= \frac{1}{|\mathcal{M}|} \sum_{y^n} \max_m W_s^n\left(y^n | x_{s,m,k}^n\right) - \frac{1}{|\mathcal{M}|} \sum_{y^n} \max_m W_s^n\left(y^n | \bar{x}_{s,m,k}^n\right) \tag{160}$$

$$= \frac{1}{|\mathcal{M}|} \sum_{y^n} W_s^n\left(y^n | x_{s,\hat{m}_{y^n},k}^n\right) - \frac{1}{|\mathcal{M}|} \sum_{y^n} \max_m W_s^n\left(y^n | \bar{x}_{s,m,k}^n\right) \tag{161}$$

$$\leq \frac{1}{|\mathcal{M}|} \sum_{y^n} \left(W_s^n\left(y^n | x_{s,\hat{m}_{y^n},k}^n\right) - W_s^n\left(y^n | \bar{x}_{s,\hat{m}_{y^n},k}^n\right) \right) \tag{162}$$

$$\leq \frac{1}{|\mathcal{M}|} \sum_{y^n : \hat{m}_{y^n}=i} W_s^n\left(y^n | x_{s,\hat{m}_{y^n},k}^n\right) \tag{163}$$

$$\leq \frac{1}{|\mathcal{M}|}, \tag{164}$$

where

- in (161), we used the definition of $\hat{m}_{y^n}$ in (159);
- in (162) follows from the trivial inequality $\max_m W_s^n(y^n | \bar{x}_{s,m,k}^n) \geq W_s^n(y^n | \bar{x}_{s,\hat{m}_{y^n},k}^n)$;
- in (163) follows because for $\hat{m}_{y^n} \neq i$, $x_{s,\hat{m}_{y^n},k}^n = \bar{x}_{s,\hat{m}_{y^n},k}^n$.

Inequality (164) implies that $\bar{P}_e(s,k)$ has the bounded difference property with $d_i = \frac{1}{|\mathcal{M}|}$. Finally, the implication of the McDiarmid's theorem completes the proof. ∎

Step 3: Expurgation: Combining (152) and (155), we get:

$$\mathbb{P}\left[\mathcal{G}^c \cup \mathcal{B}_1 \cup \mathcal{B}_2\right] \leq \mathbb{P}(\mathcal{G}^c) + \mathbb{P}\left(\mathcal{B}_1 \cup \mathcal{B}_2\right) \to 0. \tag{165}$$

Thus, there exists a codebook $\mathcal{C}_s^* \in \mathcal{G} \cap \mathcal{B}_1^c \cap \mathcal{B}_2^c$.

For the codebook $\mathcal{C}_s^*$, for all $s \in \{1,2\}$ and $k \in \mathcal{K}$, we have:

$$\bar{P}_e(s,k) \leq \mathbb{E}\left[\bar{P}_e(s,k)\right] + n^{-\frac{1}{6}} = O\left(n^{-\frac{1}{6}}\right), \tag{166}$$

where we used the bound in Lemma 3.

We then remove $a|\mathcal{M}|$ codewords from the codebook $\mathcal{C}_s^*$ with largest $P_e(s,m,k)$ for all $k \in \mathcal{K}$ and $s \in \{1,2\}$ from the subcodebook $\{X_s^n(m,k) : m \in \mathcal{M}\}$ to get new codebook $\{X_s^n(m,k) : m \in \mathcal{M}_{s,k}\}$ such that $|\mathcal{M}_{s,k}| = (1-a)|\mathcal{M}|$. For these codebooks, the maximum error probability is given by (recall that $a = n^{-\frac{1}{12}}$):

$$\max_m P_e(m,s,k) \leq \frac{1}{a} O\left(n^{-\frac{1}{6}}\right) = O\left(n^{-\frac{1}{12}}\right) = o(1). \tag{167}$$

Further, (152) implies that

$$d_{\text{TV}}(Q_{\mathcal{M}_s}^n, Q_s^{\times n}) \leq D_n = o(1).$$

This completes the proof of existence of a codebook with desired properties.

APPENDIX E
CONCLUSION OF PROOF OF THEOREM 1

Using Lemma 4, if the constraints (33) and (36) hold, then there exists a codebook such that the maximum error probability and the following are satisfied:

$$d_{\text{TV}}(Q_1^n, Q_2^n) \leq d_{\text{TV}}(Q_1^n, Q_1^{\times n}) + d_{\text{TV}}(Q_1^{\times n}, Q_2^{\times n}) + d_{\text{TV}}(Q_2^n, Q_2^{\times n}) \tag{168}$$

$$\overset{(a)}{\leq} 2\Phi\left(\frac{1}{2}\sqrt{\Omega(\gamma_1, \gamma_2)}\right) - 1 + O\left(\frac{1}{\sqrt{n}}\right) + o(1) \tag{169}$$

$$\overset{(b)}{\leq} \delta + o(1), \tag{170}$$

where (a) follows from (38) and Lemma 1 and (b) follows from (28).

Dividing both sides of (33) and (36) by $\sqrt{n}$ and letting $n \to \infty$, we get:

$$L + L_K \geq (1+\kappa) \cdot \max_{s \in \{1,2\}} \gamma_s D(W_s \| Q_0 | \bar{P}_s), \tag{171}$$

and

$$L \leq \min_{s \in \{1,2\}} \gamma_s D(W_s \| Q_0 | \bar{P}_s), \tag{172}$$

where we have used the approximation $I(P_{s,n}, W_s) = \frac{\gamma_s}{\sqrt{n}} D(W_s \| Q_0 | \bar{P}_s) + o\left(\frac{1}{\sqrt{n}}\right)$. Performing Fourier-Motzkin elimination in (171)–(172) and letting $\kappa \to 0$, concludes the proof of the theorem.

APPENDIX F
SOLUTION OF THE OPTIMIZATION PROBLEM (26)

Define:

$$\bar{\chi}_{2,s} \triangleq \chi_2(\tilde{Q}_s \| Q_0), \qquad s \in \{1, 2\}, \tag{173}$$

$\bar{\rho} \triangleq -2\rho(\tilde{Q}_1, \tilde{Q}_2, Q_0)$ and $\bar{\delta} \triangleq \left(2\Phi^{-1}\left(\frac{1+\delta}{2}\right)\right)^2$. Also, recall the definition of $\bar{\mathbb{D}}_s$ from (111). The inequality (28) can be written as follows:

$$\bar{\chi}_{2,1}\gamma_1^2 + \bar{\chi}_{2,2}\gamma_2^2 + \bar{\rho}\gamma_1\gamma_2 \leq \bar{\delta}. \tag{174}$$

The optimization problem in (26) is given by the following:

$$\max_{\gamma_1,\gamma_2} \quad \min_{s} \gamma_s \bar{\mathbb{D}}_s, \tag{175}$$
$$\text{s.t.:} \quad \bar{\chi}_{2,1}\gamma_1^2 + \bar{\chi}_{2,2}\gamma_2^2 + \bar{\rho}\gamma_1\gamma_2 \leq \bar{\delta}. \tag{176}$$

The above optimization problem is equivalently written as:

$$\max_{\gamma_1,\gamma_2} \quad \min_{0 \leq \xi \leq 1} \quad \xi\gamma_1\bar{\mathbb{D}}_1 + (1-\xi)\gamma_2\bar{\mathbb{D}}_2 \tag{177}$$
$$\text{s.t.:} \quad \bar{\chi}_{2,1}\gamma_1^2 + \bar{\chi}_{2,2}\gamma_2^2 + \bar{\rho}\gamma_1\gamma_2 \leq \bar{\delta}. \tag{178}$$

Since the above optimization problem is convex, we have:

$$\min_{0 \leq \xi \leq 1} \quad \max_{\gamma_1,\gamma_2} \quad \xi\gamma_1\bar{\mathbb{D}}_1 + (1-\xi)\gamma_2\bar{\mathbb{D}}_2 \tag{179}$$
$$\text{s.t.:} \quad \bar{\chi}_{2,1}\gamma_1^2 + \bar{\chi}_{2,2}\gamma_2^2 + \bar{\rho}\gamma_1\gamma_2 \leq \bar{\delta}. \tag{180}$$

Define

$$G \triangleq \xi\gamma_1\bar{\mathbb{D}}_1 + (1-\xi)\gamma_2\bar{\mathbb{D}}_2 - \lambda\left(\bar{\chi}_{2,1}\gamma_1^2 + \bar{\chi}_{2,2}\gamma_2^2 + \bar{\rho}\gamma_1\gamma_2 - \bar{\delta}\right). \tag{181}$$

We calculate the derivative of G with respect to γ_1 and γ_2 and let it be zero to get the optimal values γ_1^* and γ_2^*. Thus, we obtain

$$2\bar{\chi}_{2,1}\gamma_1^* + \bar{\rho}\gamma_2^* = \frac{\xi\bar{\mathbb{D}}_1}{\lambda}, \tag{182}$$

$$2\bar{\chi}_{2,2}\gamma_2^* + \bar{\rho}\gamma_1^* = \frac{(1-\xi)\bar{\mathbb{D}}_2}{\lambda}. \tag{183}$$

This yields:

$$\gamma_1^* = \frac{2\bar{\chi}_{2,2}\xi\bar{\mathbb{D}}_1 - \bar{\rho}(1-\xi)\bar{\mathbb{D}}_2}{\lambda(4\bar{\chi}_{2,1}\bar{\chi}_{2,2} - \bar{\rho}^2)} \triangleq \frac{\bar{\gamma}_1}{\lambda}, \tag{184}$$

$$\gamma_2^* = \frac{2\bar{\chi}_{2,1}(1-\xi)\bar{\mathbb{D}}_2 - \xi\bar{\rho}\bar{\mathbb{D}}_1}{\lambda(4\bar{\chi}_{2,1}\bar{\chi}_{2,2} - \bar{\rho}^2)} \triangleq \frac{\bar{\gamma}_2}{\lambda}. \tag{185}$$

From (178), we know that

$$\lambda \geq \sqrt{\frac{\bar{\chi}_{2,1}\bar{\gamma}_1^2 + \bar{\chi}_{2,2}\bar{\gamma}_2^2 + \bar{\rho}\bar{\gamma}_1\bar{\gamma}_2}{\bar{\delta}}}. \tag{186}$$

Thus, we have:

$$\min_{0 \leq \xi \leq 1} \max_{\gamma_1,\gamma_2} \quad \xi\gamma_1\bar{\mathbb{D}}_1 + (1-\xi)\gamma_2\bar{\mathbb{D}}_2 = \min_{0 \leq \xi \leq 1} \sqrt{\frac{\bar{\delta}}{\bar{\chi}_{2,1}\bar{\gamma}_1^2 + \bar{\chi}_{2,2}\bar{\gamma}_2^2 + \bar{\rho}\bar{\gamma}_1\bar{\gamma}_2}} \left(\xi\bar{\gamma}_1\bar{\mathbb{D}}_1 + (1-\xi)\bar{\gamma}_2\bar{\mathbb{D}}_2\right).$$

$$= \min_{0 \leq \xi \leq 1} \sqrt{\frac{4\bar{\delta}\left((1-\xi)^2\bar{\chi}_{2,1}\bar{\mathbb{D}}_2^2 - \xi(1-\xi)\bar{\rho}\bar{\mathbb{D}}_1\bar{\mathbb{D}}_2 + \xi^2\bar{\chi}_{2,2}\bar{\mathbb{D}}_1^2\right)}{(4\bar{\chi}_{2,1}\bar{\chi}_{2,2} - \bar{\rho}^2)}}$$

$$= \begin{cases} \bar{\mathbb{D}}_1 \bar{\mathbb{D}}_2 \sqrt{\dfrac{\bar{\delta}}{\bar{\chi}_{2,1}\bar{\mathbb{D}}_2^2 + \bar{\rho}\bar{\mathbb{D}}_1\bar{\mathbb{D}}_2 + \bar{\chi}_{2,2}\bar{\mathbb{D}}_1^2}}; & \bar{\rho} > -2\min\left(\frac{\bar{\chi}_{2,1}\bar{\mathbb{D}}_2}{\bar{\mathbb{D}}_1}, \frac{\bar{\chi}_{2,2}\bar{\mathbb{D}}_1}{\bar{\mathbb{D}}_2}\right) \\[2em] \sqrt{\dfrac{4\bar{\delta}\min\left(\bar{\chi}_{2,1}\bar{\mathbb{D}}_2^2, \bar{\chi}_{2,2}\bar{\mathbb{D}}_1^2\right)}{4\bar{\chi}_{2,1}\bar{\chi}_{2,2} - \bar{\rho}^2}}; & \bar{\rho} < -2\min\left(\frac{\bar{\chi}_{2,1}\bar{\mathbb{D}}_2}{\bar{\mathbb{D}}_1}, \frac{\bar{\chi}_{2,2}\bar{\mathbb{D}}_1}{\bar{\mathbb{D}}_2}\right) \end{cases} \tag{187}$$

This completes the proof.

APPENDIX G
PROOF OF (72) FOR THE EXAMPLE OF FIG. 3

First, notice that for the proposed example, we have

$$Q_{1,n}(0) = \frac{1-\mu_n}{2}, \quad Q_{1,n}(1) = \frac{1}{2}, \quad Q_{1,n}(2) = \frac{\mu_n}{2}, \tag{188}$$

and

$$Q_{2,n}(0) = \frac{1}{2}, \quad Q_{1,n}(1) = \frac{1-\mu_n}{2}, \quad Q_{1,n}(2) = \frac{\mu_n}{2}, \tag{189}$$

where $\mu_n = \frac{\gamma}{\sqrt{n}}$ for some positive γ such that

$$\gamma \le 2\sqrt{\ln\frac{1}{1-\delta}}. \tag{190}$$

Now, consider the following upper bound on total variation distance in terms of Bhattacharyya distance [23, Lemma 3.3.9]

$$d_{\text{TV}}(Q_1^{\times n}, Q_2^{\times n}) \le \sqrt{1 - F(Q_1^{\times n}, Q_2^{\times n})^2}. \tag{191}$$

Since $Q_1^{\times n}$ and $Q_2^{\times n}$ are product distributions, one can show that:

$$F(Q_1^{\times n}, Q_2^{\times n}) = F(Q_{1,n}, Q_{2,n})^n. \tag{192}$$

Calculation of the Bhattacharyya distance for two pmfs $Q_{1,n}$ and $Q_{2,n}$ yields the following:

$$F(Q_{1,n}, Q_{2,n}) = \frac{1}{2}\left(2\sqrt{1-\mu_n} + \mu_n\right) \tag{193}$$

$$= 1 - \frac{1}{8}\mu_n^2 + o(\mu_n^2) \tag{194}$$

$$= 1 - \frac{\gamma^2}{8n} + o\left(\frac{1}{n}\right), \tag{195}$$

where (194) follows from Taylor series $\sqrt{1-x} = 1 - \frac{x}{2} - \frac{x^2}{8} + o(x^2)$ as $x \to 0$. Combining (191), (192) and (194), we have the following upper bound on total variation distance:

$$d_{\text{TV}}(Q_1^{\times n}, Q_2^{\times n}) \le \sqrt{1 - \left(1 - \frac{\gamma^2}{8n} + o\left(\frac{1}{n}\right)\right)^{2n}} \tag{196}$$

$$= \sqrt{1 - \exp\left(-\frac{\gamma^2}{4}\right)} + o(1) \tag{197}$$

$$\le \delta + o(1), \tag{198}$$

where (197) follows $\left(1 - \frac{x}{2n}\right)^{2n} \to \exp(-x)$ as $n \to \infty$.

To find a lower bound on the optimal throughput, one can repeat the same steps as in Appendices B, C, D and get similar results with the following approximations:

$$I(P_s, W_s) = \frac{1}{2}h_{\text{b}}(\mu_n) \tag{199}$$

$$= \frac{\sqrt{\ln\frac{1}{1-\delta}}}{2} \cdot \frac{\log n}{\sqrt{n}} + o\left(\frac{\log n}{\sqrt{n}}\right), \tag{200}$$

and

$$\sum_x \sum_y P_{s,n}(x)W_s(y|x)\left(\log\frac{W_s(y|x)}{Q_{s,n}(y)}\right)^2 = \frac{1}{2}\left(\mu_n(\log\mu_n)^2 + (1-\mu_n)(\log(1-\mu_n))^2\right) \tag{201}$$

$$= \frac{\sqrt{\ln \frac{1}{1-\delta}}}{4} \cdot \frac{\log^2 n}{\sqrt{n}} + o\left(\frac{\log^2 n}{\sqrt{n}}\right). \tag{202}$$

Thus, Lemmas 3 and 4 state that if we have

$$\log |\mathcal{M}| \leq \frac{\sqrt{\ln \frac{1}{1-\delta}}}{2} \cdot \sqrt{n} \log n + o\left(\sqrt{n} \log n\right), \tag{203}$$

then the desired constraints on the maximum error probability and the (total variation) covertness criterion are satisfied. This completes the proof.

APPENDIX H
PROOF OF LEMMA 5

Let $\mathcal{P}^n$ be the set of all types π_X over $\mathcal{X}^n$ and $\mathcal{C}_{s,\pi_X}$ be the sub-codebook of $\mathcal{C}_s$ consisting of all codewords with the same type π_X. Notice that

$$\mathcal{C}_s = \bigcup_{\pi_X \in \mathcal{P}^n} \mathcal{C}_{s,\pi_X}. \tag{204}$$

Let $P_{X_s^n}$ be the uniform distribution over the codebook $\mathcal{C}_s = \{x_s^n(m,k) : m \in \mathcal{M}, k \in \mathcal{K}\}$, i.e., for every $x^n \in \mathcal{X}^n$:

$$P_{X_s^n}(x^n) = \begin{cases} \frac{1}{|\mathcal{M}| \cdot |\mathcal{K}|} & x^n \in \mathcal{C}_s \\ 0 & \text{else} \end{cases} \tag{205}$$

In the following, we first consider false alarm probability and next the missed detection probability. Consider the false alarm probability as follows:

$$\alpha_n = \mathbb{P}_{Q_1^n}\left(\sum_{i=1}^{n} \mathcal{T}_{\text{test}}(Y_i) \leq \tau\right) \tag{206}$$

$$= \sum_{y^n} Q_1^n(y^n) \mathbb{1}\left\{\sum_{i=1}^{n} \mathcal{T}_{\text{test}}(y_i) \leq \tau\right\} \tag{207}$$

$$= \sum_{y^n} \left(\sum_{x^n \in \mathcal{C}_1} P_{X_1^n}(x^n) W_1^n(y^n|x^n)\right) \mathbb{1}\left\{\sum_{i=1}^{n} \mathcal{T}_{\text{test}}(y_i) \leq \tau\right\} \tag{208}$$

$$= \sum_{x^n \in \mathcal{C}_1} P_{X_1^n}(x^n) \sum_{y^n} W_1^n(y^n|x^n) \mathbb{1}\left\{\sum_{i=1}^{n} \mathcal{T}_{\text{test}}(y_i) \leq \tau\right\} \tag{209}$$

$$= \sum_{\pi_X} \sum_{x^n \in \mathcal{C}_{1,\pi_X}} P_{X_1^n}(x^n) \mathbb{P}_{W_1^n(\cdot|x^n)}\left(\sum_{i=1}^{n} \mathcal{T}_{\text{test}}(Y_i) \leq \tau\right) \tag{210}$$

Since $\mathbb{P}_{W_1^n(\cdot|x^n)}\left(\sum_{i=1}^{n} \mathcal{T}_{\text{test}}(Y_i) \leq \tau\right)$ remains the same for each x^n with the same type, and $\{P_{X_1^n}\{\mathcal{C}_{1,\pi_X}\}\}_{\pi_X \in \mathcal{P}^n}$ is a probability distribution, we have

$$\alpha_n \leq \max_{\pi_X} \mathbb{P}_{W_1^n(\cdot|x_1^n)}\left(\sum_{i=1}^{n} \mathcal{T}_{\text{test}}(Y_i) \leq \tau\right), \tag{211}$$

where x_1^n above refers to any vector with type π_X. Denoting the maximizing type in (211) by π^* and letting x^{*n} be any vector with type π^*, (211) can be written as follows:

$$\alpha_n \leq \mathbb{P}_{W_1^n(\cdot|x^{*n})}\left(\sum_{i=1}^{n} \mathcal{T}_{\text{test}}(Y_i) \leq \tau\right). \tag{212}$$

Notice that π^* relates to the Hamming weight of the codeword x^{*n} as follows:

$$\pi^*(x) = \begin{cases} \frac{\mathsf{w}(x^{*n})}{n} & x = 1 \\ 1 - \frac{\mathsf{w}(x^{*n})}{n} & x = 0 \end{cases} \tag{213}$$

We also define

$$\mu_{1,n} \triangleq \pi^*(1) = \frac{\mathsf{w}(x^{*n})}{n}. \tag{214}$$

Since the channel is memoryless, $\{\mathcal{T}_{\text{test}}(Y_i)\}_{i=1}^n$ are mutually independent so Theorem 4 can be applied to upper bound the probability in (212). For every x^{*n} with type π^*, we calculate $\sum_{i=1}^n \mathbb{E}\left[\mathcal{T}_{\text{test}}(Y_i)\right]$ and $\sum_{i=1}^n \mathbb{V}\left[\mathcal{T}_{\text{test}}(Y_i)\right]$ in the following. First, we calculate the expectation as follows,

$$\sum_{i=1}^n \mathbb{E}_{W_1(\cdot|x_i^*)}\left[\mathcal{T}_{\text{test}}(Y_i)\right] = \sum_{i=1}^n \mathbb{E}_{W_1(\cdot|x_i^*)}\left[\frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)}\right] \tag{215}$$

$$= \sum_{i:x_i^*=1} \mathbb{E}_{\tilde{Q}_1}\left[\frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)}\right] \tag{216}$$

$$= n\mu_{1,n} \sum_y \frac{\tilde{Q}_1(y)(\tilde{Q}_1(y) - \tilde{Q}_2(y))}{Q_0(y)} \tag{217}$$

$$= n\mu_{1,n} \sum_y \frac{(\tilde{Q}_1(y) - Q_0(y))(\tilde{Q}_1(y) - \tilde{Q}_2(y))}{Q_0(y)} \tag{218}$$

$$= n\mu_{1,n}\mathrm{D}_1, \tag{219}$$

where (216) follows because when $x_i^* = 0$, we have $W_1(Y_i|x_i^*) = Q_0(Y_i)$ and the expectation in the summand is equal to zero. Next, we calculate the variance as follows

$$\sum_{i=1}^n \mathbb{V}\left[\mathcal{T}_{\text{test}}(Y_i)\right] = \sum_{i=1}^n \mathbb{E}\left[(\mathcal{T}_{\text{test}}(Y_i))^2\right] - \sum_{i=1}^n (\mathbb{E}\left[\mathcal{T}_{\text{test}}(Y_i)\right])^2, \tag{220}$$

Now, we calculate each expectation term in (220) as follows,

$$\sum_{i=1}^n \mathbb{E}\left[(\mathcal{T}_{\text{test}}(Y_i))^2\right] = \sum_{i=1}^n \mathbb{E}_{W_1(\cdot|x_i^*)}\left[(\mathcal{T}_{\text{test}}(Y_i))^2\right] \tag{221}$$

$$= \sum_{i:x_i^*=1} \mathbb{E}_{\tilde{Q}_1}\left[\left(\frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)}\right)^2\right] + \sum_{i:x_i^*=0} \mathbb{E}_{Q_0}\left[\left(\frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)}\right)^2\right] \tag{222}$$

$$= n\mu_{1,n} \sum_y \frac{\tilde{Q}_1(y)\left(\tilde{Q}_1(y) - \tilde{Q}_2(y)\right)^2}{Q_0^2(y)} + n(1 - \mu_{1,n})\Delta, \tag{223}$$

and

$$\sum_{i=1}^n (\mathbb{E}\left[\mathcal{T}_{\text{test}}(Y_i)\right])^2 = \sum_{i=1}^n \left(\mathbb{E}_{W_1(\cdot|x_i^*)}\left[\mathcal{T}_{\text{test}}(Y_i)\right]\right)^2 \tag{224}$$

$$= \sum_{i=1}^n \left(\mathbb{E}_{W_1(\cdot|x_i^*)}\left[\frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)}\right]\right)^2 \tag{225}$$

$$= \sum_{i:x_i^*=1} \left(\mathbb{E}_{\tilde{Q}_1}\left[\frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)}\right]\right)^2 + \sum_{i:x_i^*=0} \left(\mathbb{E}_{Q_0}\left[\frac{\tilde{Q}_1(Y_i) - \tilde{Q}_2(Y_i)}{Q_0(Y_i)}\right]\right)^2 \tag{226}$$

$$= n\mu_{1,n} \left(\sum_y \frac{\tilde{Q}_1(y)(\tilde{Q}_1(y) - \tilde{Q}_2(y))}{Q_0(y)}\right)^2 + n(1 - \mu_{1,n})\left(\sum_y (\tilde{Q}_1(y) - \tilde{Q}_2(y))\right)^2 \tag{227}$$

$$= n\mu_{1,n} \left(\sum_y \frac{\tilde{Q}_1(y)(\tilde{Q}_1(y) - \tilde{Q}_2(y))}{Q_0(y)}\right)^2 \tag{228}$$

$$= n\mu_{1,n} \left(\sum_y \frac{(\tilde{Q}_1(y) - Q_0(y))(\tilde{Q}_1(y) - \tilde{Q}_2(y))}{Q_0(y)}\right)^2 \tag{229}$$

$$= n\mu_{1,n}\mathrm{D}_1^2. \tag{230}$$

Combining (220), (223) and (230), we get

$$\mathbb{V}\left[Z_n(x^{*n})\right] = n\left(\mu_{1,n} \sum_y \frac{\tilde{Q}_1(y)\left(\tilde{Q}_1(y) - \tilde{Q}_2(y)\right)^2}{Q_0^2(y)} + (1 - \mu_{1,n})\Delta - \mu_{1,n}\mathrm{D}_1^2\right) \tag{231}$$

$$\triangleq n\mathrm{V}_{1,n}. \tag{232}$$

We now bound the sum of the third absolute moments

$$\sum_{i=1}^{n} \mathbb{E}_{W_1(\cdot|x_i^*)} \left[\left| \mathcal{T}_{\text{test}}(Y_i) - \mathbb{E}_{W_1(\cdot|x_i^*)}[\mathcal{T}_{\text{test}}(Y_i)] \right|^3 \right]. \tag{233}$$

We know that $\mathcal{T}_{\text{test}}(y_i) = \frac{\tilde{Q}_1(y_i) - \tilde{Q}_2(y_i)}{Q_0(y_i)}$. Hence,

$$|\mathcal{T}_{\text{test}}(y_i)| \le \frac{2}{\min_{y_i} Q_0(y_i)} \triangleq 2/\eta < \infty. \tag{234}$$

By the triangle inequality,

$$\left| \mathcal{T}_{\text{test}}(Y_i) - \mathbb{E}[\mathcal{T}_{\text{test}}(Y_i)] \right| \le 4/\eta, \quad \text{a.s.} \tag{235}$$

Combining (233) and (235), we obtain

$$\sum_{i=1}^{n} \mathbb{E}_{W_1(\cdot|x_i^*)} \left[\left| \mathcal{T}_{\text{test}}(Y_i) - \mathbb{E}_{W_1(\cdot|x_i^*)}[\mathcal{T}_{\text{test}}(Y_i)] \right|^3 \right] \le \frac{64n}{\eta^3} \triangleq n\mathrm{T}. \tag{236}$$

Notice that $\mathrm{V}_{1,n}$ in (232) is further upper bounded as follows:

$$\mathrm{V}_{1,n} \le \left(\mu_{1,n} \sum_{y} \frac{\tilde{Q}_1(y)\big(\tilde{Q}_1(y) - \tilde{Q}_2(y)\big)^2}{Q_0^2(y)} + (1 - \mu_{1,n})\Delta \right) \tag{237}$$

$$= \Delta + \mu_{1,n}\Gamma_1 \tag{238}$$

$$\le \Delta + \mu_{\mathrm{H},n}|\Gamma_1| \tag{239}$$

$$\triangleq \mathrm{V}_{1,n}^*. \tag{240}$$

Thus, from Theorem 4, we get the following:

$$\mathbb{P}\left(\sum_{i=1}^{n} \mathcal{T}_{\text{test}}(Y_i) \le \tau \right) \le 1 - \Phi\left(\frac{-\tau + n\mu_{1,n}\mathrm{D}_1}{\sqrt{n\mathrm{V}_{1,n}}} \right) + \frac{6\mathrm{T}}{\mathrm{V}_{1,n}^{\frac{3}{2}}\sqrt{n}}. \tag{241}$$

Combining (211) and (241), we get the following:

$$\alpha_n \le \max_{\mu_{1,n}} 1 - \Phi\left(\frac{-\tau + n\mu_{1,n}\mathrm{D}_1}{\sqrt{n\mathrm{V}_{1,n}}} \right) + O\left(\frac{1}{\sqrt{n}} \right), \tag{242}$$

where the maximization is over all $\mu_{1,n}$ such that $\mu_{\mathrm{L},n} \le \mu_{1,n} \le \mu_{\mathrm{H},n}$. Now, consider the missed detection probability as follows:

$$\beta_n = \mathbb{P}_{Q_2^n}\left(\sum_{i=1}^{n} \mathcal{T}_{\text{test}}(Y_i) > \tau \right). \tag{243}$$

Following similar steps leading to (211), we can write the missed detection probability as:

$$\beta_n \le \max_{\mu_{2,n}} 1 - \Phi\left(\frac{\tau - n\mu_{2,n}\mathrm{D}_2}{\sqrt{n\mathrm{V}_{2,n}}} \right) + \frac{6\mathrm{T}}{\mathrm{V}_{2,n}^{\frac{3}{2}}\sqrt{n}}, \tag{244}$$

where the maximization is over all $\mu_{2,n}$ such that $\mu_{\mathrm{L},n} \le \mu_{2,n} \le \mu_{\mathrm{H},n}$ and we define:

$$\mathrm{V}_{2,n} \triangleq \left(\mu_{2,n} \sum_{y} \frac{\tilde{Q}_2(y)\big(\tilde{Q}_1(y) - \tilde{Q}_2(y)\big)^2}{Q_0^2(y)} + (1 - \mu_{2,n})\Delta - \mu_{2,n}\left(\sum_{y} \frac{\tilde{Q}_2(y)\big(\tilde{Q}_1(y) - \tilde{Q}_2(y)\big)}{Q_0(y)} \right)^2 \right). \tag{245}$$

We can further upper bound $\mathrm{V}_{2,n}$ as follows:

$$\mathrm{V}_{2,n} \le \mu_{2,n} \sum_{y} \frac{\tilde{Q}_2(y)\big(\tilde{Q}_1(y) - \tilde{Q}_2(y)\big)^2}{Q_0^2(y)} + (1 - \mu_{2,n})\Delta \tag{246}$$

$$= \Delta + \mu_{2,n}\Gamma_2 \tag{247}$$

$$\le \Delta + \mu_{\mathrm{H},n}|\Gamma_2| \tag{248}$$

$$\triangleq \mathrm{V}_{2,n}^*, \tag{249}$$

The proof is followed by upper bounding the false alarm and missed detection probabilities using the choice of τ in (56):

$$\alpha_n \leq \max_{\mu_{1,n}} 1 - \Phi\left(\frac{-\tau + n\mu_{1,n}\mathrm{D}_1}{\sqrt{n\mathrm{V}_{1,n}}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{250}$$

$$= \max_{\mu_{1,n}} 1 - \Phi\left(\frac{-\frac{n\mu_{\mathrm{L},n}}{2}\left(\mathrm{D}_2 + \mathrm{D}_1\right) + n\mu_{1,n}\mathrm{D}_1}{\sqrt{n\mathrm{V}_{1,n}}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{251}$$

$$= 1 - \Phi\left(\frac{-\frac{n\mu_{\mathrm{L},n}}{2}\left(\mathrm{D}_2 + \mathrm{D}_1\right) + n\mu_{\mathrm{L},n}\mathrm{D}_1}{\sqrt{n\mathrm{V}_{1,n}}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{252}$$

$$\leq 1 - \Phi\left(\frac{\sqrt{n}\mu_{\mathrm{L},n}\left(\mathrm{D}_1 - \mathrm{D}_2\right)}{2\sqrt{\mathrm{V}_{1,n}}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{253}$$

$$= 1 - \Phi\left(\frac{\sqrt{n}\mu_{\mathrm{L},n}\Delta}{2\sqrt{\mathrm{V}_{1,n}}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{254}$$

$$\leq 1 - \Phi\left(\frac{\sqrt{n}\mu_{\mathrm{L},n}\Delta}{2\sqrt{\mathrm{V}_{1,n}^*}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{255}$$

$$= 1 - \Phi\left(\frac{\sqrt{n}\mu_{\mathrm{L},n}\Delta}{2\sqrt{(\Delta + \mu_{\mathrm{H},n}|\Gamma_1|)}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{256}$$

$$\leq 1 - \Phi\left(\frac{1}{2}\mu_{\mathrm{L},n}\sqrt{n}\Delta \cdot \left(1 - \frac{\mu_{\mathrm{H},n}|\Gamma_1|}{2\Delta}\right)\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{257}$$

$$\leq 1 - \Phi\left(\frac{1}{2}\mu_{\mathrm{L},n}\sqrt{n}\Delta\right) + \frac{\sqrt{n}\mu_{\mathrm{L},n}\mu_{\mathrm{H},n}|\Gamma_1|}{4\sqrt{2\pi}\Delta} + O\left(\frac{1}{\sqrt{n}}\right), \tag{258}$$

where

- (253) follows because assumption (47) implies that $\mathrm{D}_1 \geq 0$;
- (254) follows because $\mathrm{D}_1 - \mathrm{D}_2 = \Delta$;
- (255) follows because $\mathrm{V}_{1,n} \leq \mathrm{V}_{1,n}^*$;
- (257) follows from $\frac{1}{\sqrt{1+x}} \geq 1 - \frac{x}{2}$;
- (258) follows because $1 - \Phi(x - y) \leq 1 - \Phi(x) + \frac{y}{\sqrt{2\pi}}$ for all $0 < y < x$.

With the choice of τ in (56), β_n in (242) can be upper bounded as follows:

$$\beta_n \leq \max_{\mu_{2,n}} 1 - \Phi\left(\frac{\tau - n\mu_{2,n}\mathrm{D}_2}{\sqrt{n\mathrm{V}_{2,n}}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{259}$$

$$= \max_{\mu_{2,n}} 1 - \Phi\left(\frac{\frac{n\mu_{\mathrm{L},n}}{2}\left(-\mathrm{D}_2 + \mathrm{D}_1\right) + n(\mu_{\mathrm{L},n} - \mu_{2,n})\mathrm{D}_2}{\sqrt{n\mathrm{V}_{2,n}}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{260}$$

$$\leq 1 - \Phi\left(\frac{\frac{n\mu_{\mathrm{L},n}}{2}\left(-\mathrm{D}_2 + \mathrm{D}_1\right)}{\sqrt{n\mathrm{V}_{2,n}}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{261}$$

$$= 1 - \Phi\left(\frac{n\mu_{\mathrm{L},n}\Delta}{2\sqrt{n\mathrm{V}_{2,n}}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{262}$$

$$\leq 1 - \Phi\left(\frac{n\mu_{\mathrm{L},n}\Delta}{2\sqrt{n\mathrm{V}_{2,n}^*}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{263}$$

$$\leq 1 - \Phi\left(\frac{1}{2}\mu_{\mathrm{L},n}\sqrt{n}\Delta\right) + \frac{\sqrt{n}\mu_{\mathrm{L},n}\mu_{\mathrm{H},n}|\Gamma_2|}{4\sqrt{2\pi}\Delta} + O\left(\frac{1}{\sqrt{n}}\right), \tag{264}$$

where

- (261) follows because assumption (47) implies that $D_2 \leq 0$ and hence, $(\mu_{\mathrm{L},n} - \mu_{2,n})D_2 \geq 0$;
- (262) follows because $D_1 - D_2 = \Delta$;
- (263) follows because $V_{2,n} \leq V_{2,n}^*$;
- (264) follows because $V_{2,n}^* = \Delta + \mu_{\mathrm{H},n}|\Gamma_2|$ and also from the fact that $1 - \Phi(x-y) \leq 1 - \Phi(x) + \frac{y}{\sqrt{2\pi}}$ for all $0 < y < x$.

This completes the proof of lemma.

APPENDIX I
CONCLUSION OF PROOF OF THEOREM 2

In the following, we first show that there exists a sub-codebook which satisfies (63) and its size is at least $\frac{|\mathcal{M}|\cdot|\mathcal{K}|}{\sqrt{n}}$. Define the following set of codewords for some $\gamma > 0$:

$$\mathcal{D}_s \triangleq \left\{ x_s^n \in \mathcal{C}_s \colon \mathrm{w}(x_s^n) \leq 2\sqrt{\frac{n}{\Delta}}\Phi^{-1}\left(\frac{1+\delta}{2} + \gamma + \frac{E}{\sqrt{n}}\right)\right\}, \tag{265}$$

where E is chosen such that

$$E > \frac{|\Gamma_1| + |\Gamma_2|}{8\sqrt{2\pi\Delta}} \cdot \left(\Phi^{-1}\left(\frac{1+\delta}{2} + \gamma\right)\right)^2. \tag{266}$$

Let $\widehat{Q}_s^n$ and $\overline{Q}_s^n$ be the induced output distributions for codes $\mathcal{D}_s$ and $\mathcal{C}_s \backslash \mathcal{D}_s$, respectively. The distribution Q_s^n can be written as follows:

$$Q_s^n = \xi_s \widehat{Q}_s^n + (1 - \xi_s)\overline{Q}_s^n, \tag{267}$$

where $\xi_s \triangleq \frac{|\mathcal{D}_s|}{|\mathcal{M}|\cdot|\mathcal{K}|}$. Without loss of generality, assume that $\xi_1 \geq \xi_2$. Then, we get the following:

$$\delta \geq d_{\mathrm{TV}}\left(Q_1^n, Q_2^n\right) \tag{268}$$

$$= \frac{1}{2}\sum_{y^n}|Q_1^n - Q_2^n| \tag{269}$$

$$= \frac{1}{2}\sum_{y^n}\left|\xi_1\widehat{Q}_1^n + (1-\xi_1)\overline{Q}_1^n - \xi_2\widehat{Q}_2^n - (1-\xi_2)\overline{Q}_2^n\right| \tag{270}$$

$$\geq d_{\mathrm{TV}}\left(\overline{Q}_1^n, \overline{Q}_2^n\right) - \xi_1 d_{\mathrm{TV}}\left(\widehat{Q}_1^n, \overline{Q}_1^n\right) - \xi_2 d_{\mathrm{TV}}\left(\widehat{Q}_2^n, \overline{Q}_2^n\right), \tag{271}$$

where the last inequality follows from the triangle inequality. We know that for any $x_s^n \in \mathcal{C}_s\backslash\mathcal{D}_s$,

$$\mathrm{w}(x_s^n) \geq 2\sqrt{\frac{n}{\Delta}}\Phi^{-1}\left(\frac{1+\delta}{2} + \gamma + \frac{E}{\sqrt{n}}\right), \tag{272}$$

Combining (63) with (272) and considering (266), we can further lower bound the total variation distance as follows:

$$d_{\mathrm{TV}}\left(\overline{Q}_1^n, \overline{Q}_2^n\right) \geq \delta + \frac{2E}{\sqrt{n}} + 2\gamma - \frac{\omega_{\mathrm{L},n}\omega_{\mathrm{H},n}(|\Gamma_1| + |\Gamma_2|)}{4n\sqrt{2n\pi\Delta}} + O\left(\frac{1}{\sqrt{n}}\right) \tag{273}$$

$$\geq \delta + 2\gamma, \tag{274}$$

where the last inequality follows from (266). Uniting (271) and (274) yields

$$\delta \geq \delta + 2\gamma - \xi_1 - \xi_2. \tag{275}$$

If we choose $\gamma = \frac{1}{\sqrt{n}}$, we obtain

$$\xi_1 + \xi_2 \geq \frac{2}{\sqrt{n}}. \tag{276}$$

In summary, from the assumption $\xi_1 \geq \xi_2$, we obtain a set of codewords with size

$$|\mathcal{D}_1| \geq \frac{|\mathcal{M}| \cdot |\mathcal{K}|}{\sqrt{n}}, \tag{277}$$

and the Hamming weight of these codewords is given by

$$\mathrm{w}(x_1^n) \leq 2\sqrt{\frac{n}{\Delta}}\Phi^{-1}\left(\frac{1+\delta}{2} + \frac{E}{\sqrt{n}} + \gamma\right) \tag{278}$$

$$= 2\sqrt{\frac{n}{\Delta}}\Phi^{-1}\left(\frac{1+\delta}{2}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{279}$$

$$\triangleq \psi(n,\delta) + O\left(\frac{1}{\sqrt{n}}\right). \tag{280}$$

Since we have assumed $\xi_1 \geq \xi_2$, inequality (277) can be equivalently written as:

$$\max_s |\mathcal{D}_s| \geq \frac{|\mathcal{M}| \cdot |\mathcal{K}|}{\sqrt{n}} \tag{281}$$

For each $k \in \mathcal{K}$, we denote the sub-codebook of all codewords with the same key k by $\mathcal{C}_s^k$. From the pigeonhole principle, there exists a sub-codebook $\mathcal{C}_s^k$ such that $\max_s |\mathcal{D}_s \cap \mathcal{C}_s^k| \geq \frac{|\mathcal{M}|}{\sqrt{n}}$. Define:

$$\mathcal{D}_{s,i}^k \triangleq \left\{ x_s^n \in \mathcal{D}_s \cap \mathcal{C}_s^k : \mathrm{w}(x_s^n) = i \right\}. \tag{282}$$

The codewords in the sub-codebook $|\mathcal{D}_{s,i}^k|$ have the same type which we denote by π_s^i. This sub-codebook is with maximum probability of error not larger than ϵ. Using Fano's inequality, we can write the following set of inequalities for $s \in \{1,2\}$:

$$(1-\epsilon)\log|\mathcal{D}_{s,i}^k \cap \mathcal{C}_s^k| - 1 \leq I(X_s^n; Y^n) \tag{283}$$

$$\leq \sum_{t=1}^n I(X_{s,t}; Y_t) \tag{284}$$

$$\leq nI(\pi_s^i, W_s) \tag{285}$$

$$= iD(\tilde{Q}_s \| Q_0). \tag{286}$$

Next, we continue to upper bound the size of the message set as follows:

$$\log\frac{|\mathcal{M}|}{\sqrt{n}} \overset{(a)}{\leq} \max_s \log|\mathcal{D}_s \cap \mathcal{C}_s^k| \tag{287}$$

$$\overset{(b)}{=} \max_s \log\left(\sum_{i=0}^{\psi(n,\delta)} |\mathcal{D}_{s,i}^k \cap \mathcal{C}_s^k|\right) \tag{288}$$

$$\overset{(c)}{\leq} \max_s \log\left(\sum_{i=0}^{\psi(n,\delta)} 2^{iD(\tilde{Q}_s \| Q_0)}\right) \tag{289}$$

$$\leq \max_s \log\left(\psi(n,\delta)2^{\psi(n,\delta)\cdot D(\tilde{Q}_s \| Q_0)}\right) \tag{290}$$

$$\leq \psi(n,\delta) \cdot \max_s D(\tilde{Q}_s \| Q_0) + O(1) \tag{291}$$

$$\leq 2\sqrt{\frac{n}{\Delta}}\Phi^{-1}\left(\frac{1+\delta}{2}\right) \cdot \max_s D(\tilde{Q}_s \| Q_0) + O(1), \tag{292}$$

where (a) follows from (281), (b) follows from the definition in (282), (c) follows from inequality (286). Dividing both sides of (292) by $\sqrt{n}$ and taking limits in n completes the proof.

APPENDIX J
PROOF OF COROLLARY 2

For the proposed Gaussian setup, the lower bound to $L_{\mathrm{TV}}^*(\epsilon,\delta)$ in Theorem 3 can be generalized for a continuous alphabet. Several steps in the proof of upper bound to $L_{\mathrm{TV}}^*(0,\delta)$ in Appendix H remain valid for the Gaussian distribution. The only step that should be refined is bounding the sum of third absolute moments in (234). The reason is as follows. In this step of the proof for the discrete memoryless channel, we know that $\eta = \min_y Q_0(y)$ is positive. However, for the Gaussian case, this statement does not hold and η is zero. So, in the following, we first bound the sum of third absolute moments for the Gaussian setup. The evaluation of the involved information quantities with a Gaussian distribution will be presented later.

Bounding the sum of third absolute moments:

Consider the following sum of third absolute moments:

$$\sum_{i=1}^n \mathbb{E}_{W_1(\cdot|x_i^*)}\left[\left|\mathcal{T}_{\mathrm{test}}(Y_i) - \mathbb{E}_{W_1(\cdot|x_i^*)}[\mathcal{T}_{\mathrm{test}}(Y_i)]\right|^3\right] = \sum_{i:x_i^*=0} \mathbb{E}_{W_1(\cdot|x_i^*)}\left[\left|\mathcal{T}_{\mathrm{test}}(Y_i) - \mathbb{E}_{W_1(\cdot|x_i^*)}[\mathcal{T}_{\mathrm{test}}(Y_i)]\right|^3\right]$$

$$+ \sum_{i:x_i^*=1} \mathbb{E}_{W_1(\cdot|x_i^*)}\left[\left|\mathcal{T}_{\mathrm{test}}(Y_i) - \mathbb{E}_{W_1(\cdot|x_i^*)}[\mathcal{T}_{\mathrm{test}}(Y_i)]\right|^3\right] \tag{293}$$

$$= n(1 - \mu_{1,n})\mathbb{E}_{Q_0}\left[\left|\mathcal{T}_{\text{test}}(Y) - \mathbb{E}_{Q_0}[\mathcal{T}_{\text{test}}(Y)]\right|^3\right]$$

$$+ n\mu_{1,n}\mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y) - \mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]\right|^3\right]. \tag{294}$$

We study each expectation term of (294), separately, and show that each expectation term is a constant and does not depend on n. First, consider the following term:

$$\mathbb{E}_{Q_0}\left[\left|\mathcal{T}_{\text{test}}(Y) - \mathbb{E}_{Q_0}[\mathcal{T}_{\text{test}}(Y)]\right|^3\right] \overset{(a)}{=} \mathbb{E}_{Q_0}\left[\left|\mathcal{T}_{\text{test}}(Y)\right|^3\right] \tag{295}$$

$$= \mathbb{E}_{Q_0}\left[\left|\frac{\tilde{Q}_1(Y) - \tilde{Q}_2(Y)}{Q_0(Y)}\right|^3\right] \tag{296}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left|\frac{1}{\sigma_1}\exp\left(-\frac{(y-1)^2}{2\sigma_1^2}\right) - \frac{1}{\sigma_2}\exp\left(-\frac{(y-1)^2}{2\sigma_2^2}\right)\right|^3 \cdot \sigma_0^2\exp\left(\frac{y^2}{\sigma_0^2}\right)dy \tag{297}$$

$$\overset{(b)}{\leq} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{\sigma_1^3}\exp\left(-\frac{3(y-1)^2}{2\sigma_1^2}\right)\cdot\sigma_0^2\exp\left(\frac{y^2}{\sigma_0^2}\right)dy + \kappa_1 \tag{298}$$

$$= \frac{\sigma_0^2}{\sqrt{2\pi}\sigma_1^3}\int_{-\infty}^{\infty}\exp\left(-\frac{3\sigma_0^2 - 2\sigma_1^2}{2\sigma_0^2\sigma_1^2}y^2 + \frac{3}{\sigma_1^2}y - \frac{3}{2\sigma_1^2}\right)dy + \kappa_1 \tag{299}$$

$$\overset{(c)}{=} \kappa_2, \tag{300}$$

where κ_1 and κ_2 are positive constants. Here, (a) follows because $\mathbb{E}_{Q_0}[\mathcal{T}_{\text{test}}(Y)] = 0$, (b) follows because the function $\sigma \mapsto \frac{1}{\sigma}\exp\left(-\frac{(y-1)^2}{2\sigma^2}\right)$ is an increasing function of σ as $y \to \infty$ and also because $\sigma_1^2 \geq \sigma_2^2$, (c) follows because $1.5\sigma_0^2 \geq \frac{4}{3}\sigma_0^2 \geq \sigma_1^2$ and thus the integral converges to a finite positive real number.

Next, we analyze the second term of (294):

$$\mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y) - \mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]\right|^3\right] \leq \mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y)\right|^3\right] + 3\mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y)\right|^2\right]\cdot\left|\mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]\right| \tag{301}$$

$$+ 3\mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y)\right|\right]\cdot\left|\mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]\right|^2 + \left|\mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]\right|^3 \tag{302}$$

$$\leq \mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y)\right|^3\right] + 3\mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y)\right|^3\right]^{\frac{2}{3}}\cdot\left|\mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]\right| \tag{303}$$

$$+ 3\mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y)\right|^3\right]^{\frac{1}{3}}\cdot\left|\mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]\right|^2 + \left|\mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]\right|^3, \tag{304}$$

where inequality (304) follows because $\mathbb{E}\left[\left|\mathcal{T}_{\text{test}}\right|^2\right] \leq \mathbb{E}\left[\left|\mathcal{T}_{\text{test}}\right|^3\right]^{\frac{2}{3}}$ and $\mathbb{E}\left[\left|\mathcal{T}_{\text{test}}\right|\right] \leq \mathbb{E}\left[\left|\mathcal{T}_{\text{test}}\right|^3\right]^{\frac{1}{3}}$. Thus, it remains to prove that $\mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y)\right|^3\right]$ and $\left|\mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]\right|$ are bounded. We first write $\mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]$ as follows:

$$\mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)] = \mathbb{E}_{\tilde{Q}_1}\left[\frac{\tilde{Q}_1(Y) - \tilde{Q}_2(Y)}{Q_0(Y)}\right] \tag{305}$$

$$= \mathbb{E}_{\tilde{Q}_1 - Q_0}\left[\frac{\tilde{Q}_1(Y) - \tilde{Q}_2(Y)}{Q_0(Y)}\right] \tag{306}$$

$$= \mathbb{E}_{\tilde{Q}_1 - Q_0}\left[\frac{\tilde{Q}_1(Y) - Q_0(Y) - (\tilde{Q}_2(Y) - Q_0(Y))}{Q_0(Y)}\right] \tag{307}$$

$$= \chi_{\text{G},1} - \rho_{\text{G}} \tag{308}$$

$$= \kappa_3, \tag{309}$$

where κ_3 is a positive constant. Now, it remains to show that $\mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y)\right|^3\right]$ is also bounded.

$$\mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y)\right|^3\right] = \mathbb{E}_{\tilde{Q}_1}\left[\left|\frac{\tilde{Q}_1(Y) - \tilde{Q}_2(Y)}{Q_0(Y)}\right|^3\right] \tag{310}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left|\frac{1}{\sigma_1}\exp\left(-\frac{(y-1)^2}{2\sigma_1^2}\right) - \frac{1}{\sigma_2}\exp\left(-\frac{(y-1)^2}{2\sigma_2^2}\right)\right|^3 \cdot \sigma_0^3\exp\left(\frac{3y^2}{2\sigma_0^2}\right)\cdot\frac{1}{\sigma_1}\exp\left(-\frac{(y-1)^2}{2\sigma_1^2}\right)dy$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{\sigma_1^3}\exp\left(-\frac{3(y-1)^2}{2\sigma_1^2}\right)\cdot\sigma_0^3\exp\left(\frac{3y^2}{2\sigma_0^2}\right)\cdot\frac{1}{\sigma_1}\exp\left(-\frac{(y-1)^2}{2\sigma_1^2}\right)dy$$

$$\leq \frac{\sigma_0^3}{\sqrt{2\pi}\sigma_1^4} \int_{-\infty}^{\infty} \exp\left(\frac{3y^2}{2\sigma_0^2}\right) \cdot \exp\left(-\frac{2(y-1)^2}{\sigma_1^2}\right) dy + \kappa_4 \tag{311}$$

$$= \frac{\sigma_0^3}{\sqrt{2\pi}\sigma_1^4} \int_{-\infty}^{\infty} \exp\left(-y^2\left(\frac{2}{\sigma_1^2} - \frac{3}{2\sigma_0^2}\right) + y\frac{4}{\sigma_1^2} - \frac{2}{\sigma_1^2}\right) dy + \kappa_4 \tag{312}$$

$$= \kappa_5, \tag{313}$$

where κ_4 and κ_5 are positive constants. The last equality follows because $\sigma_1^2 \leq \frac{4}{3}\sigma_0^2$ and so the integral has a finite value. Combining (294), (300), (304), (309) and (313), we get:

$$\sum_{i=1}^{n} \mathbb{E}_{W_1(\cdot|x_i^*)} \left[\left|\mathcal{T}_{\text{test}}(Y_i) - \mathbb{E}_{W_1(\cdot|x_i^*)}[\mathcal{T}_{\text{test}}(Y_i)]\right|^3\right] \tag{314}$$

$$= n(1 - \mu_{1,n})\mathbb{E}_{Q_0}\left[\left|\mathcal{T}_{\text{test}}(Y) - \mathbb{E}_{Q_0}[\mathcal{T}_{\text{test}}(Y)]\right|^3\right] + n\mu_{1,n}\mathbb{E}_{\tilde{Q}_1}\left[\left|\mathcal{T}_{\text{test}}(Y) - \mathbb{E}_{\tilde{Q}_1}[\mathcal{T}_{\text{test}}(Y)]\right|^3\right]$$

$$\leq n(1 - \mu_{1,n})\kappa_2 + n\mu_{1,n}(\kappa_5 + 3\kappa_5^{2/3}\kappa_3 + 3\kappa_5^{1/3}\kappa_3^2 + \kappa_3^2) \tag{315}$$

$$\leq n(\kappa_2 + \kappa_5 + 3\kappa_5^{2/3}\kappa_3 + 3\kappa_5^{1/3}\kappa_3^2 + \kappa_3^2) \tag{316}$$

$$\triangleq n\mathbf{T}. \tag{317}$$

Thus, the above replaces (236) and the rest of the proof in Appendix H holds.

Evaluation of the involved information quantities:

We now evaluate the information quantities involved in Theorem 3. First, consider the KL-divergence term $D(\tilde{Q}_s\|Q_0)$ as follows:

$$D(\tilde{Q}_s\|Q_0) = -H_{\tilde{Q}_s}(Y) + \mathbb{E}_{\tilde{Q}_s}\left[\log\frac{1}{Q_0(Y)}\right] \tag{318}$$

$$= -\frac{1}{2}\log\left(2\pi e\sigma_s^2\right) + \mathbb{E}_{\tilde{Q}_s}\left[\log\frac{1}{Q_0(Y)}\right] \tag{319}$$

$$= -\frac{1}{2}\log\left(2\pi e\sigma_s^2\right) + \mathbb{E}_{\tilde{Q}_s}\left[\frac{1}{2}\log(2\pi e\sigma_0^2) + \frac{Y^2}{2\sigma_0^2}\log(e)\right] \tag{320}$$

$$= \frac{1}{2}\log\frac{\sigma_0^2}{\sigma_s^2} + \frac{\log e}{2\sigma_0^2}\,\mathbb{E}_{\tilde{Q}_s}\left[Y^2\right] \tag{321}$$

$$= \frac{1}{2}\log\frac{\sigma_0^2}{\sigma_s^2} + \frac{\log e}{2\sigma_0^2}\cdot(\sigma_s^2 + 1) \tag{322}$$

$$= D_{\text{G}}. \tag{323}$$

From (322) and the assumption $D(\tilde{Q}_1\|Q_0) = D(\tilde{Q}_2\|Q_0)$, we get the condition (69). Next, consider the chi-squared distance $\chi_2(\tilde{Q}_s\|Q_0)$ as follows:

$$\chi_2\left(\tilde{Q}_s\|Q_0\right) = \mathbb{E}_{Q_0}\left[\left(\frac{\tilde{Q}_s(Y)}{Q_0(Y)}\right)^2\right] - 1 \tag{324}$$

$$= \frac{\sigma_0^2}{\sigma_s^2}\,\mathbb{E}_{Q_0}\left[\exp\left(-\frac{2(Y-1)^2}{2\sigma_s^2} + \frac{2Y^2}{2\sigma_0^2}\right)\right] - 1 \tag{325}$$

$$= \frac{\sigma_0}{\sqrt{2\pi}\sigma_s^2}\int_{-\infty}^{\infty}\exp\left(-\frac{2(y-1)^2}{2\sigma_s^2} + \frac{y^2}{2\sigma_0^2}\right)dy - 1 \tag{326}$$

$$= \frac{\sigma_0}{\sqrt{2\pi}\sigma_s^2}\int_{-\infty}^{\infty}\exp\left(-y^2\left(\frac{1}{\sigma_s^2} - \frac{1}{2\sigma_0^2}\right) + \frac{2}{\sigma_s^2}y - \frac{1}{\sigma_s^2}\right)dy - 1 \tag{327}$$

$$= \frac{\sigma_0}{\sqrt{2\pi}\sigma_s^2}\int_{-\infty}^{\infty}\exp\left(-y^2\left(\frac{1}{\sigma_s^2} - \frac{1}{2\sigma_0^2}\right) + \frac{2}{\sigma_s^2}y - \frac{1}{\sigma_s^2 - \frac{\sigma_s^4}{2\sigma_0^2}} + \frac{1}{\sigma_s^2 - \frac{\sigma_s^4}{2\sigma_0^2}} - \frac{1}{\sigma_s^2}\right)dy - 1 \tag{328}$$

$$= \frac{\sigma_0}{\sqrt{2\pi}\sigma_s^2}\exp\left(\frac{1}{\sigma_s^2}\left(\frac{1}{1 - \frac{\sigma_s^2}{2\sigma_0^2}} - 1\right)\right)\int_{-\infty}^{\infty}\exp\left(-\left(\sqrt{\frac{1}{\sigma_s^2} - \frac{1}{2\sigma_0^2}}\,y - \sqrt{\frac{1}{\sigma_s^2 - \frac{\sigma_s^4}{2\sigma_0^2}}}\right)^2\right)dy - 1 \tag{329}$$

$$= \frac{\sigma_0}{\sqrt{2\pi}\sigma_s^2}\exp\left(\frac{1}{\sigma_s^2}\left(\frac{1}{1 - \frac{\sigma_s^2}{2\sigma_0^2}} - 1\right)\right)\frac{1}{\sqrt{\frac{1}{\sigma_s^2} - \frac{1}{2\sigma_0^2}}}\int_{-\infty}^{\infty}\exp\left(-u^2\right)du - 1 \tag{330}$$

$$= \frac{\sigma_0}{\sigma_s} \exp\left(\frac{1}{\sigma_s^2}\left(\frac{1}{1 - \frac{\sigma_s^2}{2\sigma_0^2}} - 1\right)\right) \frac{1}{\sqrt{2 - \frac{\sigma_s^2}{\sigma_0^2}}} - 1 \tag{331}$$

$$= \chi_{\mathrm{G},s}. \tag{332}$$

Finally, we evaluate the term $\rho(\tilde{Q}_1, \tilde{Q}_2, Q_0)$ as the following:

$$\rho(\tilde{Q}_1, \tilde{Q}_2, Q_0) = \mathbb{E}_{Q_0}\left[\frac{\tilde{Q}_1(Y)}{Q_0(Y)} \cdot \frac{\tilde{Q}_2(Y)}{Q_0(Y)}\right] - 1 \tag{333}$$

$$= \frac{\sigma_0}{\sqrt{2\pi}\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{(y-1)^2}{2\sigma_1^2} - \frac{(y-1)^2}{2\sigma_2^2} + \frac{y^2}{2\sigma_0^2}\right) dy - 1 \tag{334}$$

$$= \frac{\sigma_0}{\sqrt{2\pi}\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-y^2\left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_0^2}\right) + y\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) - \left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}\right)\right) dy - 1 \tag{335}$$

$$= \frac{\sigma_0}{\sqrt{2\pi}\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-y^2\left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_0^2}\right) + y\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) - \frac{\left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}\right)^2}{\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_0^2}}\right.$$
$$\left. + \frac{\left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}\right)^2}{\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_0^2}} - \left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}\right)\right) dy - 1 \tag{336}$$

$$= \frac{\sigma_0}{\sqrt{2\pi}\sigma_1\sigma_2} \exp\left(\frac{\left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}\right) \cdot \frac{1}{2\sigma_0^2}}{\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_0^2}}\right) \times$$
$$\int_{-\infty}^{\infty} \exp\left(-\left(\sqrt{\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_0^2}}\, y - \frac{\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}}{\sqrt{\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_0^2}}}\right)^2\right) dy - 1 \tag{337}$$

$$= \frac{\sigma_0}{\sqrt{2\pi}\sigma_1\sigma_2} \exp\left(\frac{\left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}\right) \cdot \frac{1}{2\sigma_0^2}}{\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_0^2}}\right) \cdot \frac{1}{\sqrt{\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_0^2}}} \int_{-\infty}^{\infty} \exp\left(-u^2\right) du - 1 \tag{338}$$

$$= \frac{\sigma_0}{\sigma_1\sigma_2} \exp\left(\frac{\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) \cdot \frac{1}{2\sigma_0^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} - \frac{1}{\sigma_0^2}}\right) \cdot \frac{1}{\sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} - \frac{1}{\sigma_0^2}}} - 1 \tag{339}$$

$$= \rho_{\mathrm{G}}. \tag{340}$$

Moreover, consider the fact that

$$\Delta = \chi_2(\tilde{Q}_1 \| Q_0) + \chi_2(\tilde{Q}_2 \| Q_0) - 2\rho(\tilde{Q}_1, \tilde{Q}_2, Q_0) = \Delta_{\mathrm{G}}. \tag{341}$$

The proof is concluded by combining (323) with (341) and (332) with (340).

REFERENCES

[1] B. A. Bash, D. Goekel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Select. Areas Commun.*, vol. 31, pp. 1921–1930., Sept. 2013.

[2] L. Wang, W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. on Info. Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[3] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. on Info. Theory*, vol. 62, no. 5, pp. 2334–2354., 2016.

[4] L. Wang, "Optimal throughput for covert communication over a classical-quantum channel," in *Proc. IEEE Information Theory Workshop*, 2016, pp. 364–368.

[5] S. H. Lee, L. Wang, A. Khisti, and G. W. Wornell, "Covert communication with channel-state information at the transmitter," *IEEE Trans. on Info. Forensics and Security*, vol. 13, pp. 2310–2319, 2018.

[6] Q. Zhang, M. Bakshi, and S. Jaggi, "Covert communication over adversarially jammed channels," *arXiv*, 2018. [Online]. Available: https://arxiv.org/abs/1805.02426

[7] M. Tahmasbi and M. R. Bloch, "Second-order asymptotics of covert communications over noisy channels," in *Proc. IEEE Int. Symp. on Info. Theory*, Jul. 2016, pp. 2224–2228.

[8] ——, "First and second order asymptotics in covert communication," *IEEE Trans. on Info. Theory*, vol. 65, no. 4, pp. 2190–2212, Apr. 2019.

[9] K. S. K. Arumugam and M. R. Bloch, "Keyless covert communication over multiple-access channels," in *Proc. IEEE Int. Symp. on Info. Theory*, Jul. 2016, pp. 2229–2233.

[10] V. Y. F. Tan and S. H. Lee, "Time-division is optimal for covert communication over some broadcast channels," *IEEE Trans. on Info. Forensics and Security*, vol. 14, 2019.

[11] K. S. K. Arumugam, M. R. Bloch, and L. Wang, "Covert communication over a physically degraded relay channel with non-colluding wardens," in *Proc. IEEE Int. Symp. on Info. Theory*, 2018.

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

[13] R. L. Dobrushin, "Optimum information transmission through a channel with unknown parameters," *Radio Eng. Electron.*, vol. 4, no. 12, pp. 1–8, 1959.

[14] J. Wolfowitz, "Simultaneous channels," *Arch. Rational Mech. Anal.,*, vol. 4, no. 1, pp. 371–386, 1960.

[15] A. El Gamal and Y. H. Kim, *Network Information Theory*.   Cambridge University Press, 2011.

[16] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *Proc. IEEE Int. Symp. on Info. Theory*, 2014, pp. 30–34.

[17] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. on Info. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[18] V. Y. F. Tan, *Asymptotic estimates in information theory with non-vanishing error probabilities*.   Foundations and Trends in Communications and Information Theory, 2014, vol. 11, no. 1–2.

[19] P. Cuff, "Distributed channel synthesis," *IEEE Trans. on Info. Theory*, vol. 59, no. 11, 2013.

[20] C. McDiarmid, *On the method of bounded differences, In Surveys in Combinatorics*.   Cambridge University Press, 1989.

[21] J. Liu, P. Cuff, and S. Verdú, "e_γ resolvability," *IEEE Trans. on Info. Theory*, vol. 63, no. 5, pp. 2629–2658, May 2017.

[22] [Online]. Available: https://www.johndcook.com/blog/2008/11/10/bounds-on-binomial-coefficients

[23] R. D. Reiss, *Approximate distributions of order statistics*.   Springer series in statistics, 1989.