Mixing a Covert and a Non-Covert User

Abdelaziz Bounhar
LTCI, Telecom Paris, IP Paris
91120 Palaiseau, France
abdelaziz.bounhar@telecom-paris.fr

Mireille Sarkiss
SAMOVAR, Telecom SudParis, IP Paris
91011 Evry, France
mireille.sarkiss@telecom-sudparis.eu

Michèle Wigger
LTCI, Telecom Paris, IP Paris
91120 Palaiseau, France
michele.wigger@telecom-paris.fr

***Abstract*—This paper establishes the fundamental limits of a two-user single-receiver system where communication from User 1 (but not from User 2) needs to be undetectable to an external warden. Our fundamental limits show a tradeoff between the highest rates (or square-root rates) that are simultaneously achievable for the two users. Moreover, coded time-sharing for both users is fundamentally required on most channels, which distinguishes this setup from the more classical setups with either only covert users or only non-covert users. Interestingly, the presence of a non-covert user can be beneficial for improving the covert capacity of the other user.**

I. Introduction

Covert communication refers to any communication setup where users wish to convey information while ensuring low probability of detection by other users, adversaries or network monitoring nodes. Such setups are relevant in future IoT and sensor networks. For instance in healthcare applications, sensors in a hospital may transmit sensitive data and this data should be reliably decoded by authorized devices while staying undetectable by any unauthorized one. The work in [1] first characterized the fundamental limits of covert communications over AWGN channels. It showed that it is possible to communicate covertly as long as the message is subject the so-called *square-root law*, i.e., the number of communicated bits scales like $\mathcal{O}(\sqrt{n})$, for n indicating the number of channel uses. Recently, it has been establised in various works [1]–[4] that the fundamental limits of covert communication is indeed characterized by this *square-root law*. While [4] assumed the existence of a sufficiently large secret key allowing covertness, [3] derived the exact growth rate of this secret key and established conditions where it is not needed. These results were also extended to keyless setups over binary symmetric channels (BSCs) [5] and over Multiple Access Channels (MACs) [6], and to asymptotically keyless setups [7]. Higher covert-rates than indicated by the *square-root law* were shown to be achievable in scenarios where the warden has uncertainty about the channel statistics [8]–[11] or in the presence of a jammer [12]–[14]. More closely related to this paper are [15]–[18] which consider extensions to Broadcast Channels (BCs) and MACs. In particular, [15] characterized the limits of covert communication over a BC when the transmitter sends a common non-covert message to two receivers and a covert message to only one of them by embedding the covert codeword into the non-covert codeword. Extensions to scenarios with a fixed codebook for the common message or with several receivers were presented in [16] [17], [18].

In this paper, we consider a Discrete Memoryless Multiple Access Channel (DMMAC) with two users communicating with a legitimate receiver. More specifically, sharing a secret key with this receiver, User 1 wishes to communicate covertly without being detected by an external warden. On the other side, User 2 transmits a non-covert message to the same legitimate receiver. The covertness constraint imposes in this case that the communication of the covert user must resemble communication of the non-covert user rather than pure noise. We establish the fundamental limits on the set of achievable triples of non-covert-rate, covert-square-root-rate, and key-rate. Compared to the previous related results [3], [4], [15], [16], our setup required extra non-trivial steps especially in the asymptotic analysis and converse proof.

We show through numerical examples that coded time-sharing improves the covert user square-root rate under a key-rate constraint. Moreover, we observe a tradeoff between the rates and square-root-rates of the non-covert and covert users, which illustrates the dependence of the covert rate on the channel parameters, emphasizing the influence of the non-covert codewords on the achievable covert-square-root-rate. We also show that the covert user's square-root-rate can be improved in the presence of a non-covert user. This conclusion resembles the previous conclusions in [15], [17], [18], which showed e.g., that the probability of detection vanishes faster when one increases the number of non-covert users. In our setup we consider only one non-covert user for simplicity. However, our proofs can be extended to any number of non-covert users.

Notation: We follow standard information theory notations. We note by $|\mathcal{S}|$ the cardinality of a set $\mathcal{S}$. Random variables are denoted by upper case letters (e.g., X), while their realizations are denoted by lowercase (e.g. x). We write X^n and x^n for the tuples $(X_1, \ldots, X_n)$ and $(x_1, \ldots, x_n)$, respectively, for any positive integer $n > 0$. For a distribution P on $\mathcal{X}$, we note its product distribution on $\mathcal{X}^n$ by $P^{\otimes n}(x^n) = \prod_{i=1}^n P(x_i)$. For two distributions P and Q on same alphabet $\mathcal{X}$, the chi-squared test is denoted $\chi_2(P\|Q) = \sum_{x\in\mathcal{X}} \frac{(P(x)-Q(x))^2}{P(x)}$, the divergence by $\mathbb{D}(P\|Q) = \sum_{x\in\mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$, and we write $P \ll Q$ whenever $Q(x) = 0$ implies $P(x) = 0$ for all $x \in \mathcal{X}$. We use $\mathbb{H}(\cdot)$, $\mathbb{H}(\cdot|\cdot)$ and $\mathbb{I}(\cdot;\cdot)$ for the entropy, conditional entropy and mutual information of random variables. The

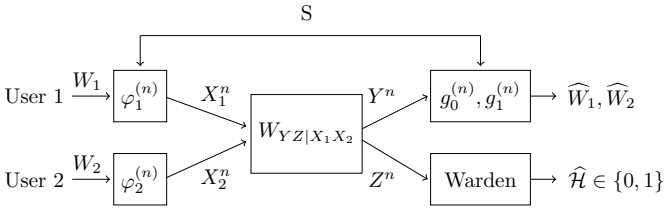

Fig. 1: Multi-access communication where communication of User 1 has to remain undetectable to an external warden.

type of a sequence $x^n \in \mathcal{X}^n$ is defined as $\pi_{x^n}(a) = |\{t\colon x_t = a\}|/n$ and the strongly-typical set $\mathcal{T}_\mu^{(n)}(P_X)$ [19, Definition 2.8] is the subset of sequences $x^n \in \mathcal{X}^n$ that satisfy $|\pi_{x^n}(a) - P_X(x)| \leq \mu$ for all $a \in \mathcal{X}$ and whenever $P_X(x) = 0$ then also $\pi_{x^n}(a) = 0$. We further abbreviate *probability mass function* by *pmf*. Finally, the logarithm and exponential functions are in base e.

II. Problem Setup

Consider the setup depicted in Figure 1 where two users communicate to a legitimate receiver in the presence of a warden. User 1 wishes to communicate covertly, i.e., the warden cannot detect its communication. User 2 does not mind being detected by the warden, and we shall even assume that the warden knows the message transmitted by User 2. We thus have two hypotheses $\mathcal{H} = 0$ and $\mathcal{H} = 1$, where under $\mathcal{H} = 0$ only User 2 sends a message while under $\mathcal{H} = 1$ both users send a message to the legitimate receiver. For simplicity we assume that User 1 produces inputs in the binary alphabet $\mathcal{X}_1 = \{0, 1\}$. User 2's input alphabet $\mathcal{X}_2$ is finite but arbitrary otherwise. The legitimate receiver and the warden observe channel outputs in the finite alphabets $\mathcal{Y}$ and $\mathcal{Z}$. These outputs are produced by a discrete and memoryless interference channel with transition law $W_{YZ|X_1X_2}$, see Figure 1.

Define the message and key sets

$$\mathcal{M}_1 \triangleq \{1, \ldots, M_1\} \tag{1}$$
$$\mathcal{M}_2 \triangleq \{1, \ldots, M_2\} \tag{2}$$
$$\mathcal{K} \triangleq \{1, \ldots, K\} \tag{3}$$

for given numbers M_1, M_2, and K and let the messages W_1 and W_2 and the key S be uniform over $\mathcal{M}_1, \mathcal{M}_2$, and $\mathcal{K}$, respectively. The key S is known to User 1 and to the legitimate receiver, message W_1 is known to User 1 only, and message W_2 to User 2 and the warden. Under $\mathcal{H} = 0$, User 1 sends the all-zero sequence

$$X_1^n = 0^n, \tag{4}$$

whereas User 2 applies some encoding function $\varphi_2^{(n)}\colon \mathcal{M}_2 \to \mathcal{X}_2^n$ to its message W_2 and sends the resulting codeword

$$X_2^n = \varphi_2^{(n)}(W_2) \tag{5}$$

over the channel. Under $\mathcal{H} = 1$, User 1 applies some encoding function $\varphi_1^{(n)}\colon \mathcal{M}_1 \times \mathcal{K} \to \mathcal{X}_1^n$ to its message W_1 and the secret key S and sends the resulting codeword

$$X_1^n = \varphi_1^{(n)}(W_1, S) \tag{6}$$

over the channel. User 2 constructs its channel inputs in the same way as before, see (5), since it is not necessarily aware of whether $\mathcal{H} = 0$ or $\mathcal{H} = 1$. For readability we will also write $x_1^n(w_1, s)$ and $x_2^n(w_2)$ instead of $\varphi_1^{(n)}(w_1, s)$ and $\varphi_2^{(n)}(w_2)$.

The legitimate receiver, which knows the hypothesis $\mathcal{H}$, decodes the desired messages W_2 (under $\mathcal{H} = 0$) or (W_1, W_2) (under $\mathcal{H} = 1$) based on its observed outputs Y^n and the key S. Thus, under $\mathcal{H} = 0$ it uses a decoding function $g_0^{(n)}\colon \mathcal{Y}^n \times \mathcal{K} \to \mathcal{W}_2$ to produce the single guess

$$\widehat{W}_2 = g_0^{(n)}(Y^n) \tag{7}$$

and under $\mathcal{H} = 1$ it uses a decoding function $g_1^{(n)}\colon \mathcal{Y}^n \to \mathcal{W}_2 \times \mathcal{W}_1$ to produce the pair of guesses

$$(\widehat{W}_1, \widehat{W}_2) = g_1^{(n)}(Y^n, S). \tag{8}$$

Decoding performance of a tuple of encoding and decoding functions $(\varphi_1^{(n)}, \varphi_2^{(n)}, g_0^{(n)}, g_1^{(n)})$ is measured by the error probabilities under the two hypotheses:

$$P_{e1} \triangleq \Pr\Big(\widehat{W}_2 \neq W_2 \text{ or } \widehat{W}_1 \neq W_1 \Big| \mathcal{H} = 1\Big) \tag{9}$$
$$P_{e0} \triangleq \Pr\Big(\widehat{W}_2 \neq W_2 \Big| \mathcal{H} = 0\Big). \tag{10}$$

Communication is subject to a covertness constraint at the warden, which observes the channel outputs Z^n as well as the correct message W_2. (Obviously, covertness assuming that the warden knows W_2 implies also covertness in the setup where it does not know W_2.) For each $w_2 \in \mathcal{M}_2$ and $W_2 = w_2$, we define the warden's output distribution under $\mathcal{H} = 1$

$$\widehat{Q}_{\mathcal{C},w_2}^n(z^n) \triangleq \frac{1}{M_1 K} \sum_{(w_1,s)} W_{Z|X_1X_2}^{\otimes n}(z^n | x_1^n(w_1, s), x_2^n(w_2)), \tag{11}$$

and under $\mathcal{H} = 0$

$$W_{Z|X_1X_2}^{\otimes n}(z^n | 0^n, x_2^n(w_2)), \tag{12}$$

and the divergence between these two distributions:

$$\delta_{n,w_2} \triangleq \mathbb{D}\Big(\widehat{Q}_{\mathcal{C},w_2}^n \Big\| W_{Z|X_1X_2}^{\otimes n}(\cdot | 0^n, x_2^n(w_2))\Big), \quad w_2 \in \mathcal{M}_2. \tag{13}$$

(Standard arguments [20] can be used to relate this divergence to the warden's detection error probabilities.)

Definition 1: A triple (r_1, r_2, k) is achievable if there exists a sequence (in the blocklength n) of triples (M_1, M_2, K) and encoding/decoding functions $(\varphi_1^{(n)}, \varphi_2^{(n)}, g_0^{(n)}, g_1^{(n)})$ satisfying

$$\lim_{n\to\infty} \delta_{n,w_2} = 0, \qquad \forall w_2 \in \mathcal{M}_2, \tag{14}$$
$$\lim_{n\to\infty} P_{ei} = 0, \qquad i \in \{0, 1\}. \tag{15}$$

and

$$r_1 \leq \liminf_{n\to\infty} \frac{\log(M_1)}{\sqrt{n \frac{1}{M_2} \sum_{w_2=1}^{M_2} \delta_{n,w_2}}}, \tag{16}$$

$$r_2 \leq \liminf_{n\to\infty} \frac{\log(M_2)}{n}, \tag{17}$$

$$k \geq \limsup_{n\to\infty} \frac{\log(K)}{\sqrt{n \frac{1}{M_2} \sum_{w_2=1}^{M_2} \delta_{n,w_2}}}. \tag{18}$$

III. Main Result and Examples

We shall assume that for any $x_2 \in \mathcal{X}_2$:

$$W_{Y|X_1X_2}(\cdot \mid 1, x_2) \ll W_{Y|X_1X_2}(\cdot \mid 0, x_2), \tag{19a}$$
$$W_{Y|X_1X_2}(\cdot \mid 1, x_2) \neq W_{Y|X_1X_2}(\cdot \mid 0, x_2), \tag{19b}$$
$$W_{Z|X_1X_2}(\cdot \mid 1, x_2) \ll W_{Z|X_1X_2}(\cdot \mid 0, x_2), \tag{19c}$$
$$W_{Z|X_1X_2}(\cdot \mid 1, x_2) \neq W_{Z|X_1X_2}(\cdot \mid 0, x_2). \tag{19d}$$

Notice that if (19d) is violated, then in all channel uses where User 2 sends symbol x_2, User 1 can trivially transmit information without being detected. Applying this to a sub-linear fraction of channel uses, the rate of User 2 is unchanged and User 1 can achieve infinite covert rate $r_1 = \infty$. If (19b) is violated, then User 1 cannot transmit any information to the receiver over all channel uses where User 2 sends symbol x_2. If (19c) is violated, then with high probability the warden can detect communication from User 1 on the channel uses where User 2 sends x_2. Define

$$D_Y(x_2) \triangleq \mathbb{D}\big(W_{Y|X_1X_2}(\cdot|1, x_2) \,||\, W_{Y|X_1X_2}(\cdot|0, x_2)\big) \tag{20}$$
$$D_Z(x_2) \triangleq \mathbb{D}\big(W_{Z|X_1X_2}(\cdot|1, x_2) \,||\, W_{Z|X_1X_2}(\cdot|0, x_2)\big) \tag{21}$$
$$\chi_{2,Y}(x_2) \triangleq \chi_2\big(W_{Y|X_1X_2}(\cdot|1, x_2) \,||\, W_{Y|X_1X_2}(\cdot|0, x_2)\big) \tag{22}$$
$$\chi_{2,Z}(x_2) \triangleq \chi_2\big(W_{Z|X_1X_2}(\cdot|1, x_2) \,||\, W_{Z|X_1X_2}(\cdot|0, x_2)\big). \tag{23}$$

A. Main Results

Theorem 1: Let $\mathcal{T} \triangleq \{1, 2\}$ and let the pair of random variables (T, X_2) be distributed according to any pmf P_{TX_2} over the alphabets $\mathcal{T} \times \mathcal{X}_2$. Let also $\{\omega_n\}_{n=1}^{\infty}$ be a sequence satisfying

$$\lim_{n\to\infty} \omega_n = 0 \tag{24a}$$
$$\lim_{n\to\infty} \big(\omega_n \sqrt{n} - \log n\big) = \infty, \tag{24b}$$

and let $\epsilon_1, \epsilon_2 \in [0, 1]$.

Then, there exists a sequence of encoding and decoding functions $\{(\varphi_1^{(n)}, \varphi_2^{(n)}, g_0^{(n)}, g_1^{(n)})\}_n$ with message and key sizes M_1, M_2, K so that for any $\epsilon > 0$ and $\xi, \xi_1, \xi_2 \in (0, 1)$ and all sufficiently large blocklengths n the following conditions hold:

$$P_{ei} \leq \epsilon, \qquad i \in \{0, 1\}, \tag{25}$$
$$\delta_{n,w_2} \leq \epsilon \qquad \forall w_2 \in \mathcal{M}_2, \tag{26}$$
$$\log(M_2) = (1 - \xi) n I(X_2; Y \mid X_1 = 0, T), \tag{27}$$
$$\log(M_1) = (1 - \xi_1)\omega_n \sqrt{n} \mathbb{E}_{P_{TX_2}}[\epsilon_T D_Y(X_2)], \tag{28}$$
$$\log(M_1) + \log(K) = (1 + \xi_2)\omega_n \sqrt{n} \mathbb{E}_{P_{TX_2}}[\epsilon_T D_Z(X_2)]. \tag{29}$$

Proof: Section IV describes a coding scheme achieving the desired performance. The analysis of the scheme is similar to the analysis in [3], and sketched in Appendix A. A sketch of the converse proof is given in Section V. ∎

Lemma 1: For any choice of the pmf P_{TX_2}, of the positive numbers ϵ_1, ϵ_2, and the sequence ω_n as in Theorem 1 there exists a sequence of encoding and decoding functions $\{(\varphi_1^{(n)}, \varphi_2^{(n)}, g_0^{(n)}, g_1^{(n)})\}_n$ satisfying the conditions in the theorem and moreover

$$\frac{1}{M_2} \sum_{w_2=1}^{M_2} \delta_{n,w_2} = (1 + o(1)) \frac{\omega_n^2}{2} \mathbb{E}_{P_{TX_2}}\big[\epsilon_T^2 \cdot \chi_{2,Z}(X_2)\big] \tag{30}$$

for a function $o(1)$ that tends to 0 as $n \to \infty$.

Proof: By inspecting the proof of Theorem 1, see Appendix B. ∎

Theorem 2: A rate-triple (r_1, r_2, k) is achievable, if, and only if, for some pmf P_{TX_2} over $\mathcal{T} \times \mathcal{X}_2$ and $\epsilon_1, \epsilon_2 \in [0, 1]$ the following three inequalities hold:

$$r_2 \leq \mathbb{I}(X_2; Y \mid X_1 = 0, T), \tag{31}$$
$$r_1 \leq \sqrt{2} \frac{\mathbb{E}_{P_{TX_2}}[\epsilon_T D_Y(X_2)]}{\sqrt{\mathbb{E}_{P_{TX_2}}[\epsilon_T^2 \cdot \chi_{2,Z}(X_2)]}}, \tag{32}$$
$$k \geq \sqrt{2} \frac{\mathbb{E}_{P_{TX_2}}[\epsilon_T (D_Z(X_2) - D_Y(X_2))]}{\sqrt{\mathbb{E}_{P_{TX_2}}[\epsilon_T^2 \cdot \chi_{2,Z}(X_2)]}}, \tag{33}$$

where for the right-hand sides of (32) and (33) we define $0/0 = 0$.

Proof: The "if" direction follows directly from Theorem 1 and Lemma 1. The "only-if" part is proved in Section V. ∎

Lemma 2: The set of three-dimensional vectors (r_1, r_2, k) satisfying inequalities (31)–(33) for some choice of pmfs P_{TX_2} and values $\epsilon_1, \epsilon_2 \in [0, 1]$ is a convex set.

Proof: See Appendix C. ∎

Remark 1: Whenever $\mathbb{E}_{P_{TX_2}}[\epsilon_T (D_Z(X_2) - D_Y(X_2))] < 0$ for any choice of the pmf P_{TX_2}, the condition (33) is always satisfied and no secret key is needed for covert communication.

Remark 2: For $\mathcal{X}_2 = \{x_2\}$ a singleton, we recover the result in [3] for the channel $W_{Y|X_1X_2}(\cdot|\cdot, x_2)$. In this case it suffices to choose T deterministic, i.e., $|\mathcal{T}| = 1$ and the expression in (32)–(33) further simplify in the sense that the ϵ_T-terms in the fraction can be reduced and the final expression does not depend on ϵ_T anymore.

B. Numerical Examples

Example 1: Consider input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ and channels (where rows indicate pairs (x_1, x_2) in lexocographic order and columns the y- or z-values)

$$W_{Y|X_1X_2} = \begin{bmatrix} 0.20 & 0.30 & 0.20 & 0.30 \\ 0.10 & 0.20 & 0.30 & 0.40 \\ 0.25 & 0.45 & 0.10 & 0.20 \\ 0.35 & 0.25 & 0.20 & 0.20 \end{bmatrix}, \tag{34a}$$

$$W_{Z|X_1X_2} = \begin{bmatrix} 0.30 & 0.20 & 0.10 & 0.40 \\ 0.30 & 0.20 & 0.15 & 0.35 \\ 0.35 & 0.15 & 0.20 & 0.30 \\ 0.23 & 0.27 & 0.20 & 0.30 \end{bmatrix}. \tag{34b}$$

Notice that these channels satisfy Conditions (19). Figure 2, illustrates the rate-region in Theorem 2 for key-rates $k \leq 0.5$ (in red) and the corresponding reduced rate-region when one restricts to deterministic Ts (dashed blue). (This latter

corresponds to the performance of a scheme without coded time-sharing.)

Example 2: Consider input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ and channels

$$W_{Y|X_1X_2} = \begin{bmatrix} 0.35 & 0.11 & 0.31 & 0.23 \\ 0.03 & 0.55 & 0.40 & 0.01 \\ 0.51 & 0.02 & 0.17 & 0.30 \\ 0.04 & 0.33 & 0.62 & 0.01 \end{bmatrix}, \tag{35a}$$

$$W_{Z|X_1X_2} = \begin{bmatrix} 0.30 & 0.50 & 0.08 & 0.12 \\ 0.21 & 0.32 & 0.39 & 0.08 \\ 0.16 & 0.28 & 0.37 & 0.19 \\ 0.48 & 0.10 & 0.38 & 0.04 \end{bmatrix}. \tag{35b}$$

Notice that these channels satisfy Conditions (19). Figure 3, illustrates the rate-region in Theorem 2 for key-rates $k \leq 0.3$ (dashed blue line) and for key-rates $k \leq 0.8$ (red line).

Example 3: Consider the same channel law in (35). Figure 4, illustrates the largest possible covert-user square-root rate r_1, i.e., when one optimizes over P_{X_2}, in function of the key-rate k (red line). The same relation is also plotted under the restriction that User 2 sends the constant symbol $X_2 = 0$ (dashed blue line) or $X_2 = 1$ (dotted black line). This shows that non-constant channel inputs X_2 at User 2 achieve better performance than any of the two constant channel inputs. In this sense, the presence of User 2 in the system actually increases the covert capacity of User 1.

Considering the inputs X_2^n a state sequence that influences the channel, above observations imply that a state-dependent channel can have higher covert square-root rate for a given key-rate then any of the marginal channels that result when one fixes the channel state. State-dependent covert-communication was also considered in [9]–[11], [14].

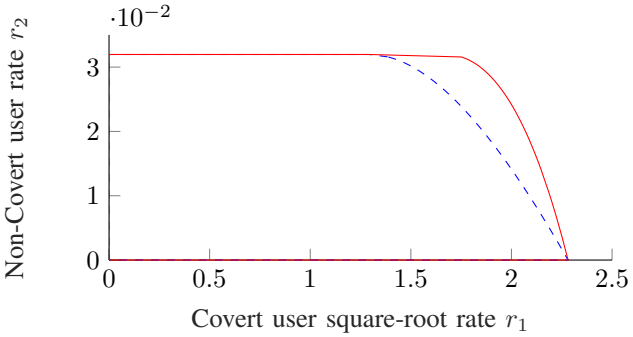

Fig. 2: Rate-region (r_1, r_2) for the channels in (34) and key-rate $k \leq 0.5$ (red line) and a degenerate region when one restricts to $|\mathcal{T}| = 1$ (dashed blue line).

IV. Coding Scheme Achieving Theorem 1

Preparations: Fix a pmf P_{TX_2}, a pair ϵ_1, ϵ_2 and a sequence $\{\omega_n\}$ as in the theorem. For each $t \in \mathcal{T}$, define the conditional pmf

$$P_{X_{1,n}|T}(1|t) = 1 - P_{X_{1,n}|T}(0|t) = \epsilon_t \frac{\omega_n}{\sqrt{n}} \tag{36}$$

and let $\mu_n \triangleq n^{-1/3}$ and the type π_{t^n} over $\mathcal{T}^n$ satisfy

$$|\pi(t) - P_T(t)| \leq \mu_n \quad \forall t \in \mathcal{T}, \tag{37}$$

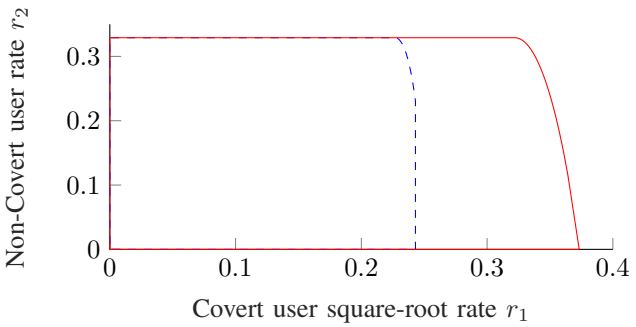

Fig. 3: Rate-region (r_1, r_2) in Theorem 2 for the channels in (35) and key-rates $k \leq 0.3$ (dashed blue line) or $k \leq 0.8$ (red line).

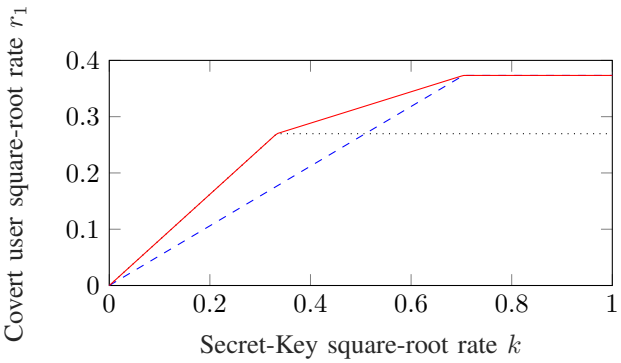

Fig. 4: Maximum covert square-root rate r_1 in function of the key-rate k when one optimizes over P_{X_2T} (red line) and when $X_2 = 0$ or $X_2 = 1$ deterministically (dashed blue and dotted black lines).

as well as $\pi(t) = 0$ whenever $P_T(t) = 0$. Fix a large blocklength n and let $t^n = (t_1, \ldots, t_n)$ be of type π_{t^n}.

Define the joint pmf

$$P^{(n)}_{TX_1X_2Y} \triangleq P_{TX_2} P_{X_{1,n}|T} W_{Y|X_1X_2}. \tag{38}$$

Let P_{TX_2Y} denote the (T, X_2, Y)-marginal of the pmf $P^{(n)}_{TX_1X_2Y}$ and notice that the following asymptotic quantity exists because $\frac{\omega_n}{\sqrt{n}} \to 0$:

$$P^*_{TX_2Y}(t, x_2, y) \triangleq \lim_{n\to\infty} P^{(n)}_{TX_2Y}(t, x_2, y) \tag{39}$$

$$= P_{TX_2}(t, x_2) W_{Y|X_1X_2}(y|0, x_2). \tag{40}$$

Codebook generation: For User 1, generate a codebook $\mathcal{C}_1 = \{x_1^n(1,1), \ldots, x_1^n(2^{M_1}, 2^K)\}$ by drawing the i-th entry of codeword $x_1^n(w_1, s)$ according to the pmf $P_{X_{1,n}|T}(\cdot|t_i)$ independent of all other entries.

For User 2, generate a codebook $\mathcal{C}_2 = \{x_2^n(1), \ldots, x_2^n(2^{M_2})\}$ by drawing the i-th entry of codeword $x_2^n(w_2)$ according to the pmf $P_{X_2|T}(\cdot|t_i)$ independent of all other entries.

The realisation of the codebook is revealed to all parties.

Encoding and Decoding: If $\mathcal{H} = 1$ User 1 sends the codeword $x_1^n(W_1, S)$, and if $\mathcal{H} = 0$ it sends $x_1^n = 0^n$. User 2 sends codeword $x_2^n(W_2)$.

The legitimate receiver, who observes $Y^n = y^n$ and knows the secret key S and the hypothesis $\mathcal{H}$, performs successive decoding starting with message W_2 followed by message W_1. More specifically, it first looks for a unique index w_2 satisfying

$$(t^n, x_2^n(w_2), y^n) \in \mathcal{T}_{\mu_n}^n(P_{TX_2Y}). \tag{41}$$

If such a unique index w_2 exists, the receiver sets $\widehat{W}_2 = w_2$. Otherwise it declares an error and stops.

If $\mathcal{H} = 1$, the receiver also looks for a unique index w_1 satisfying

$$(x_1^n(w_1, S), x_2^n(\widehat{W}_2), y^n) \in \mathcal{A}_\eta^n, \tag{42}$$

where

$$\mathcal{A}_\eta^n \triangleq \left\{(x_1^n, x_2^n, y^n)\colon \log\left(\frac{W_{Y|X_1X_2}^{\otimes n}(y^n|x_1^n, x_2^n)}{W_{Y|X_1X_2}^{\otimes n}(y^n|0^n, x_2^n)}\right) \geq \eta\right\} \tag{43}$$

and $\eta \triangleq (1-\xi_1/2)\sqrt{n}\omega_n \mathbb{E}_{P_{TX_2}}[\epsilon_T D_Y(X_2)]$.

V. Converse Proof to Theorem 2

Similarly to [3, Theorem 3] and [15], it can be shown that:

$$\frac{1}{n}\log(M_2) \leq \frac{1}{1-P_{e1}}\mathbb{I}(X_{2,T}; Y_T \mid X_{1,T}, T) + \frac{1}{n}\mathbb{H}_b(P_{e1,0}) \tag{44}$$

and by Appendices D-B and D-C:

$$\frac{\log(M_1)}{\sqrt{\frac{n}{M_2}\sum_{w_2=1}^{M_2}\delta_{n,w_2}}} \leq \frac{\sqrt{2}}{1-P_{e1}} \frac{\mathbb{E}_{P_{TX_2}}[\alpha_{n,T} D_Y(X_2)] + \frac{1}{n}}{\sqrt{\mathbb{E}_{P_{TX_2}}\left[(1-\sqrt{\alpha_{n,T}})\alpha_{n,T}^2 \chi_{2,Z}(X_2)\right]}} \tag{45}$$

$$= \frac{\sqrt{2}}{1-P_{e1}} \frac{\mathbb{E}_{P_{TX_2}}[\gamma_{n,T} D_Y(X_2)] + \frac{1}{n}}{\sqrt{\mathbb{E}_{P_{TX_2}}\left[(1-\sqrt{\alpha_{n,T}})\gamma_{n,T}^2 \chi_{2,Z}(X_2)\right]}}, \tag{46}$$

where here we define T to be uniform over $\{1, \ldots, n\}$ independent of the inputs and the channel and $\alpha_{n,t}$ denotes the fraction of 1-symbols in the t-th positions of the x_1-codewords:

$$\alpha_{n,t} \triangleq \frac{1}{M_1 K}\sum_{w_1=1}^{M_1}\sum_{s=1}^{K} \mathbb{1}\{x_{1,t}(w_1, s) = 1\}, \tag{47}$$

for $x_{1,t}(w_1, s)$ denoting the t-th symbol of codeword $x_1^n(w_1, s)$. In (46) we used the normalized definition

$$\gamma_{n,t} \triangleq \frac{\alpha_{n,t}}{\mathbb{E}_T[\alpha_{n,T}]}, \qquad t \in \{1, \ldots, n\}. \tag{48}$$

The new parameters $\gamma_{n,t}$ are well defined because $\mathbb{E}_T[\alpha_{n,T}]$ equals the fraction of 1-entries in the codebook $\{x_1^n(W_1, S)\}$ and is thus non-zero because otherwise no communication is going on. Moreover, by Jensen's Inequality, $\mathbb{E}_T[\gamma_{n,T}^2] \geq (\mathbb{E}_T[\gamma_{n,T}])^2 = 1$ and the covertness constraint implies that $\alpha_{n,t} \to 0$ for any t (proof omitted, but similar to [3]).

It then follows by Assumptions (19), that the right-hand side of (46) lies in a bounded interval, and consequently there exists a subsequence of blocklengths so that (44) and (46) converge. By the continuity of the expressions and in view of the achievability result, it can be concluded that there exists a sequence of coding schemes achieving the same asymptotic expressions. In the remainder of this proof we restrict attention to these coding schemes, for which we can conclude that (see also the arguments in Appendix D-B) for any number $\phi_2 \in (0, 1)$ and sufficiently large blocklengths n:

$$\sqrt{\frac{n}{M_2}\sum_{w_2}\delta_{n,w_2}} \leq \frac{n}{(1-\phi_2)}\sqrt{\mathbb{E}_{P_{TX_2}}\left[(1-\sqrt{\alpha_{n,T}})\frac{\alpha_{n,T}^2}{2}\cdot\chi_{2,Z}(X_2)\right]}. \tag{49}$$

Combining (49) and the lower bound on $\log M_1 + \log K$ derived in Appendix D-D, it can then be concluded that for sufficiently large blocklengths n:

$$\frac{\log(M_1) + \log(K)}{\sqrt{n\frac{1}{M_2}\sum_{w_2=1}^{M_2}\delta_{n,w_2}}} \geq (1-\phi_2')\frac{\sqrt{2}\cdot\mathbb{E}_{P_{TX_2}}\left[\gamma_{n,T} D_Z(X_2)\right]}{\sqrt{\mathbb{E}_{P_{TX_2}}\left[\gamma_{n,T}^2\cdot\chi_{2,Z}(X_2)\right]}}, \tag{50}$$

where ϕ_2' can again be chosen as an arbitrary positive number. Here we used again the fact that $\alpha_{n,t} \to 0$ as $n \to \infty$.

By the Fenchel-Eggleston strenghtening of Carathéodory's theorem it can be shown that in Constraints (44), (46), and (50), one can restrict to random variables T over alphabets of size 4. This allows to obtain the desired asymptotic results by letting $n \to \infty$, as we explain in Appendix D-E.

The final step is to show that no loss in optimality is incurred by restricting T to be of cardinality 2, see Appendix D-F.

VI. Summary and Discussion

We characterized the fundamental limits of a system mixing a covert user and a non-covert user both communicating to the same receiver, which also shares a common key with the covert user of a given rate. Our results show a tradeoff between the three quantities: the covert user's square-root-rate, the non-covert user's rate and the key rate. They also show necessity of a coded time-sharing strategy at the two users, similarly as in multi-access scenarios without covertness constraints. Finally, our results also prove that the presence of the non-covert user can increase the covert-capacity of the other user under a stringent key-rate constraint.

While our results are for multiple-access channels with a single covert and non-covert users, extensions to multiple users seems feasible. Further interesting research directions include studies less standard models for the users or the channels such as fading channels, channels with states, or non-synchronized transmissions.

REFERENCES

[1] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on awgn channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, 2013.

[2] B. A. Bash, D. Goeckel, D. Towsley, and S. Guha, "Hiding information in noise: fundamental limits of covert wireless communication," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 26–31, 2015.

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

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

[5] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *2013 IEEE International Symposium on Information Theory*, 2013, pp. 2945–2949.

[6] K. S. K. Arumugam and M. R. Bloch, "Covert communication over a k -user multiple-access channel," *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7020–7044, 2019.

[7] H. Zivarifard, M. R. Bloch, and A. Nosratinia, "Keyless covert communication via channel state information," *IEEE Transactions on Information Theory*, vol. 68, no. 8, pp. 5440–5474, 2022.

[8] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *2014 IEEE Information Theory Workshop (ITW 2014)*, 2014, pp. 30–34.

[9] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell, "Covert communication with channel-state information at the transmitter," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2310–2319, 2018.

[10] H. ZivariFard, M. Bloch, and A. Nosratinia, "Keyless covert communication in the presence of non-causal channel state information," in *2019 IEEE Information Theory Workshop (ITW)*, 2019, pp. 1–5.

[11] H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Keyless covert communication via channel state information," *CoRR*, vol. abs/2003.03308, 2020. [Online]. Available: https://arxiv.org/abs/2003.03308

[12] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 6193–6206, 2017.

[13] O. Shmuel, A. Cohen, and O. Gurewitz, "Multi-antenna jamming in covert communication," *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4644–4658, 2021.

[14] H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Covert communication via non-causal cribbing from a cooperative jammer," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 202–207.

[15] K. S. Kumar Arumugam and M. R. Bloch, "Embedding covert information in broadcast communications," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2787–2801, 2019.

[16] D. Kibloff, S. M. Perlaza, and L. Wang, "Embedding covert information on a given broadcast code," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 2169–2173.

[17] S. W. Kim and H. Q. Ta, "Covert communications over multiple overt channels," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1112–1124, 2022.

[18] H. Q. Ta, K. Ho-Van, D. B. Da Costa, S. W. Kim, and H. Oh, "Covert communications over non-orthogonal multiple overt channels," *IEEE Access*, vol. 10, pp. 122 361–122 375, 2022.

[19] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Ed.* Wiley, 2006.

APPENDIX A
SKETCH OF THE ANALYSIS OF THE CODING SCHEME IN SECTION IV

A. Error Probability Analysis of the Proposed Scheme

By standard steps, e.g., [20, Chapter 15], we can conclude that

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{C}}[P_{e0}] = 0 \tag{51}$$

whenever

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log M_2 \leq I_{P^*}(X_2; Y \mid T) \tag{52}$$

$$= I_P(X_2; Y \mid X_1 = 0, T). \tag{53}$$

Next, notice that upon defining

$$P_{e1,1} \triangleq \Pr\Big(\widehat{W}_1 \neq W_1 \Big| \mathcal{H} = 1, \widehat{W}_2 = W_2\Big) \tag{54}$$

$$P_{e1,2} \triangleq \Pr\Big(\widehat{W}_2 \neq W_2 \Big| \mathcal{H} = 1\Big), \tag{55}$$

we have that

$$P_{e1} = P_{e1,1} + P_{e1,2}. \tag{56}$$

As in the analysis of P_{e0}, we deduce that under Condition (53)

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{C}}[P_{e1,2}] = 0. \tag{57}$$

The analysis of $\mathbb{E}_{\mathcal{C}}[P_{e1,1}]$ is an extension of [3], we only provide the major steps here.

By the symmetry of the code construction, we have

$$\mathbb{E}_{\mathcal{C}}(P_{e1,1}) = \sum_{i=2}^{M_1} \Pr[(X_1^n(i,1), X_2^n(w_2), Y^n) \in \mathcal{A}_\eta^n] + \Pr[(X_1^n(1,1), X_2^n(w_2), Y^n) \notin \mathcal{A}_\eta^n] \tag{58}$$

and one can show that

$$\sum_{i=2}^{M_1} \Pr[(X_1^n(i,1), X_2^n(w_2), Y^n) \in \mathcal{A}_\eta^n]$$

$$\leq \sum_{i=2}^{M_1} e^{-\eta} \prod_{t\in\mathcal{T}} \left(\mathop{\mathbb{E}}_{P_{X_2Y|T=t}} \left[\frac{W_{Y|X_1}^{(t)}(Y|X_2)}{W_{Y|X_1X_2}(Y|0,X_2)} \right] \right)^{n\pi(t)} \tag{59}$$

$$\leq M_1 e^{-\eta} e^{-\omega_n^2\left((1-\Delta)\mathbb{E}_{P_T}[\epsilon_T^2] + \mu_n \sum_{t\in\mathcal{T}} \epsilon_t^2\right)} \tag{60}$$

where we define

$$W_{Y|X_2}^{(t)}(y|x_2) \triangleq \sum_{x_1\in\mathcal{X}_1} W_{Y|X_1X_2}(y|x_1,x_2) P_{X_{1,n}|T}(x_1|t). \tag{61}$$

Since $\big((1-\Delta)\mathbb{E}_{P_T}[\epsilon_T^2] + \mu_n \sum_{t\in\mathcal{T}} \epsilon_t^2\big)$ is bounded and $\omega_n^2 \to 0$, the above sum in (60) tends to 0 only if

$$\lim_{n\to\infty} (\log M_1 - \eta) = -\infty. \tag{62}$$

It can be shown by Hoeffding's inequality that the second summand in the right hand side of (58) vanishes exponentially fast whenever

$$\eta < \sqrt{n}\omega_n \mathbb{E}_{P_{TX_2}}[\epsilon_T[D_Y(X_2)]]. \tag{63}$$

We therefore can conclude that $\mathbb{E}_{\mathcal{C}}[P_{e1,1}]$ is vanishing if (28) is satisfied, which concludes the error probability analysis.

B. Channel Resolvability Analysis

Recall the distribution $\widehat{Q}_{\mathcal{C},w_2}^n$ in (11). We show that

$$\mathbb{E}_{\mathcal{C}}\Big[\mathbb{D}\Big(\widehat{Q}_{\mathcal{C},w_2}^n \Big\| W_{Z|X_1X_2}^{\otimes n}(\cdot|0^n, x_2^n)\Big)\Big] \to 0, \qquad \forall w_2 \in \mathcal{M}_2, \tag{64}$$

if (29) is satisfied. Similarly to (61), we define the conditional output distributions

$$W_{Z|X_2}^{(t)}(z|x_2) \triangleq \sum_{x_1\in\mathcal{X}_1} P_{X_{1,n}|T}(x_1 \mid t) W_{Z|X_1X_2}(z|x_1,x_2) \tag{65}$$

and the product distribution

$$\tilde{W}_{Z^n|X_2^n}(z^n|x_2^n) \triangleq \prod_{i=1}^n W_{Z|X_2}^{(t_i)}(z|x_2). \tag{66}$$

Fix a message $w_2 \in \mathcal{W}_2$ and a codeword $x_2^n(w_2)$. For ease of notation, we write x_2^n instead of $x_2^n(w_2)$.

Start by noticing that by following analogous steps as in [3], it can be shown that the inequalities in (69) and (70) on top of the next page hold for sufficiently large values of n, where for any $t \in \mathcal{T}$ and $x_2 \in \mathcal{X}_2$ we define

$$\lambda_{w_2,t}^{(n)}(x_2) \triangleq \frac{\big|\{j \in [n] \colon x_{2,j}(w_2) = x_2,\ t_j = t\}\big|}{n} \tag{67}$$

as well as

$$\eta_0 \triangleq \min_{z,x_2 \in \mathrm{supp}(W_{Z|X_1X_2}(z|0,x_2))} W_{Z|X_1X_2}(z|0,x_2) \tag{68}$$

for an arbitrary small $\xi_2 > 0$.

Since $\omega_n \to 0$ as $n \to \infty$, we deduce from (69) and (70) that the limit in (64) vanishes for a specific message w_2 whenever

$$\lim_{n\to\infty} \mathbb{D}\Big(\widehat{Q}_{\mathcal{C},w_2}^n \Big\| \tilde{W}_{Z^n|X_2^n}(\cdot|X_2^n)\Big) = 0. \tag{71}$$

We shall prove the stronger statement that the divergence in (71) vanishes exponentially fast in the blocklength. In the remainder of this subsection we consider the average (over the codebooks) expected divergence

$$\mathbb{E}_{\mathcal{C}}\Big[\mathbb{D}\Big(\widehat{Q}_{\mathcal{C},w_2}^n \Big\| \tilde{W}_{Z^n|X_2^n}(\cdot|X_2^n(w_2))\Big)\Big]. \tag{72}$$

Following similar steps as in [3], one can show inequality (73) on top of the next page, where we define

$$\tau \triangleq (1+\xi_2)\omega_n\sqrt{n}\mathbb{E}_{P_{TX_2}}[\epsilon_T D_Z(X_2)]. \tag{74}$$

We notice that the first summand in (73) tends to 0 because ne^{-na} decays for any positive $a > 0$. If moreover,

$$\overline{\lim_{n\to\infty}} \log M_1 + \log K - (1+\xi_2)\omega_n\sqrt{n}\mathbb{E}_{P_{TX_2}}[\epsilon_T D_Z(X_2)] = -\infty, \tag{75}$$

then also the second summand tends to 0 exponentially fast. This concludes the resolvability analysis.

$$\Big|\mathbb{D}\Big(\widehat{Q}^n_{\mathcal{C},w_2}\|W^{\otimes n}_{Z|X_1X_2}(\cdot|0^n,x_2^n)\Big)-\mathbb{D}\Big(\tilde{W}_{Z^n|X_2^n}(\cdot|x_2^n)\|W^{\otimes n}_{Z|X_1X_2}(\cdot|0^n,x_2^n)\Big)\Big|\leq \mathbb{D}\Big(\widehat{Q}^n_{\mathcal{C},w_2}\|\tilde{W}_{Z^n|X_2^n}(\cdot|x_2^n)\Big)\left(1+\frac{n}{2}\log\left(\frac{1}{\eta_0}\right)\right). \tag{69}$$

$$\frac{\omega_n^2}{2}\cdot\sum_{(x_2,t)\in\mathcal{X}_2\times\mathcal{T}}\left(1-\sqrt{\epsilon_t\frac{\omega_n}{\sqrt{n}}}\right)\epsilon_t^2\lambda^{(n)}_{w_2,t}(x_2)\cdot\chi_{2,Z}(x_2)\leq\mathbb{D}\Big(\tilde{W}_{Z^n|X_2^n}\|W^{\otimes n}_{Z|X_1X_2}(\cdot|0^n,x_2^n)\Big)$$
$$\leq\frac{\omega_n^2}{2}\cdot\sum_{(x_2,t)\in\mathcal{X}_2\times\mathcal{T}}\left(1+\sqrt{\epsilon_t\frac{\omega_n}{\sqrt{n}}}\right)\epsilon_t^2\lambda^{(n)}_{w_2,t}(x_2)\cdot\chi_{2,Z}(x_2), \tag{70}$$

$$\mathbb{E}_{\mathcal{C}}\Big[\mathbb{D}\Big(\widehat{Q}^n_{\mathcal{C},w_2}\|\tilde{W}_{Z^n|X_2^n}(\cdot|X_2^n(w_2)\Big)\Big]\leq n\mathbb{E}_{P_T}\left[\log\left(\frac{2}{(1-\epsilon_T\frac{\omega_n}{\sqrt{n}})\eta_0}\right)\right]\exp\Big(-\frac{n^2}{B}\Big(\kappa\mathbb{E}_{P_{TX_2}}\Big[\epsilon_T\frac{\omega_n}{\sqrt{n}}D_Z(X_2)\Big]\Big)^2\Big)+\frac{e^\tau}{M_1K}. \tag{73}$$

C. Summary of the Analysis

Since $\lim_{n\to\infty}\pi(\cdot)=P_T(\cdot)$, our findings (53), (58) and (75) allows us to conclude the existence of a sequence of encoding and decoding functions $\{(\varphi_1^{(n)},\varphi_2^{(n)},g_0^{(n)},g_1^{(n)})\}_n$ with message and key sizes M_1,M_2,K so that for any $\epsilon>0$ and $\xi,\xi_1,\xi_2\in(0,1)$ and all sufficiently large blocklengths n, (25)–(29) hold.

Appendix B
Proof of Lemma 1

Consider the random code-construction from Section IV, which we analyzed in Appendix A. Combining (70) and (69) with (73), and after averaging over the message w_2, we obtain that under condition (75) for any realization of this code construction:

$$\frac{1}{M_2}\sum_{w_2=1}^{M_2}\delta_{n,w_2}$$
$$\leq e^{-\zeta_2\omega_n}\left(1+n\log\left(\frac{1}{\eta_0}\right)\right)$$
$$+\frac{\omega_n^2}{2}\sum_{x_2,t}\left[\lambda_t(x_2)\epsilon_t^2\left(1+\sqrt{\epsilon_t\frac{\omega_n}{\sqrt{n}}}\right)\chi_{2,Z}(x_2)\right]. \tag{76}$$

where ζ_2 is an appropriate positive constant and

$$\lambda_t(x_2)\triangleq\frac{1}{M_2}\sum_{w_2=1}^{M_2}\lambda^{(n)}_{w_2,t}(x_2). \tag{77}$$

In a similar way:

$$\frac{1}{M_2}\sum_{w_2=1}^{M_2}\delta_{n,w_2}$$
$$\geq -e^{-\zeta_2\omega_n}\left(1+n\log\left(\frac{1}{\eta_0}\right)\right)$$
$$+\frac{\omega_n^2}{2}\sum_{x_2,t}\left[\lambda_t(x_2)\epsilon_t^2\left(1-\sqrt{\epsilon_t\frac{\omega_n}{\sqrt{n}}}\right)\chi_{2,Z}(x_2)\right]. \tag{78}$$

Since both P_{e0} and P_{e1} vanish as $n\to\infty$, and by the decoding rule in (41), we can conclude that the sequence of codes in Theorem 1 satisfies for each $(t,x_2)\in\mathcal{T}\times\mathcal{X}_2$:

$$\big|\lambda_t(x_2)-P_{TX_2}(t,x_2)\big|\to 0. \tag{79}$$

We thus conclude from (76) and (78) that

$$\frac{1}{M_2}\sum_{w_2=1}^{M_2}\delta_{n,w_2}=(1+o(1))\frac{\omega_n^2}{2}\mathbb{E}_{P_{TX_2}}\left[\epsilon_T^2\chi_{2,Z}(X_2)\right] \tag{80}$$

for a function $o(1)$ that tends to 0 as $n\to\infty$. This concludes the proof of the lemma.

Appendix C
Proof of Lemma 2

Fix two pmfs P_{TX_2} and Q_{TX_2} as well as two tuples (ϵ_1,ϵ_2) and (δ_1,δ_2) in $[0,1]^2$. Then, choose $\lambda\in[0,1]$ and set $\nu>0$ so that

$$\nu^2\triangleq\frac{\mathbb{E}_{P_T}\left[\epsilon_T^2\mathbb{E}_{P_{X_2|T}}[\chi_{2,Z,T}]\right]}{\mathbb{E}_{Q_T}\left[\delta_T^2\mathbb{E}_{Q_{X_2|T}}[\chi_{2,Z,T}]\right]}. \tag{81}$$

Also form the new pmf R_{TX_2} by choosing

$$R_T(t)=\begin{cases}\lambda\cdot P_T(t) & t\in\{1,2\}\\ (1-\lambda)\cdot Q_T(t-2) & t\in\{3,4\}\end{cases} \tag{82}$$

and

$$R_{X_2|T}(x_2|t)=\begin{cases}P_{X_2|T}(x_2|t) & t\in\{1,2\}\\ Q_{X_2|T}(x_2|t-2) & t\in\{3,4\}.\end{cases} \tag{83}$$

Moreover,

$$\gamma_t\triangleq\begin{cases}\epsilon_t, & t\in\{1,2\}\\ \nu\cdot\delta_{t-2}, & t\in\{3,4\}\end{cases} \tag{84}$$

Let (r_1,r_2,k), (r_1',r_2',k'), and $(\tilde{r}_1,\tilde{r}_2,\tilde{k})$ be the triples given by the right-hand sides of (31)–(33) when evaluated for

P_{TX_2} and (ϵ_1, ϵ_2), for Q_{TX_2} and (δ_1, δ_2), and for R_{TX_2} and $(\gamma_1, \ldots, \gamma_4)$. We shall show that

$$\lambda \begin{pmatrix} r_1 \\ r_2 \\ k \end{pmatrix} + (1-\lambda) \begin{pmatrix} r_1' \\ r_2' \\ k' \end{pmatrix} = \begin{pmatrix} \tilde{r}_1 \\ \tilde{r}_2 \\ \tilde{k} \end{pmatrix}. \tag{85}$$

The desired equality for the r_2-component is directly obtained by the linearity of conditional mutual information and because it does not depend on the ϵ-, δ-, and γ-tuples. To see the equality for the other two components, notice that for any functions f and g from $\mathcal{X}_2$ to $\mathbb{R}$ satisfying

$$\nu^2 \triangleq \frac{\mathbb{E}_{P_{TX_2}}\big[\epsilon_T^2 g(X_2)\big]}{\mathbb{E}_{Q_{TX_2}}[\delta_T^2 g(X_2)]} \tag{86}$$

we have (90)

$$\lambda \frac{\mathbb{E}_{P_{X_2T}}[\epsilon_T f(X_2)]}{\sqrt{\mathbb{E}_{P_{X_2T}}[\epsilon_T^2 g(X_2)]}} + (1-\lambda) \frac{\mathbb{E}_{Q_{X_2T}}[\delta_T f(X_2)]}{\sqrt{\mathbb{E}_{Q_{X_2T}}[\delta_T^2 g(X_2)]}} \tag{87}$$

$$= \lambda \frac{\mathbb{E}_{P_{X_2T}}[\epsilon_T f(X_2)]}{\sqrt{\mathbb{E}_{P_{X_2T}}[\epsilon_T^2 g(X_2)]}} + (1-\lambda) \frac{\mathbb{E}_{Q_{X_2T}}[\nu\delta_T f(X_2)]}{\sqrt{\mathbb{E}_{Q_{X_2T}}[\nu^2\delta_T^2 g(X_2)]}} \tag{88}$$

$$\overset{(a)}{=} \frac{\lambda \mathbb{E}_{P_{X_2T}}[\epsilon_T f(X_2)] + (1-\lambda)\ \mathbb{E}_{Q_{X_2T}}[\nu\delta_T f(X_2)]}{\sqrt{\lambda \mathbb{E}_{P_{X_2T}}[\epsilon_T^2 g(X_2)] + (1-\lambda)\mathbb{E}_{Q_{X_2T}}[\nu^2\delta_T^2 g(X_2)]}} \tag{89}$$

$$= \frac{\mathbb{E}_{R_{X_2T}}[\gamma_T f(X_2)]}{\sqrt{\mathbb{E}_{R_{X_2T}}[\gamma_T^2 g(X_2)]}} \tag{90}$$

where (a) holds because by the definition of ν we have

$$\mathbb{E}_{Q_{X_2T}}\big[\nu^2\delta_T^2 g(X_2)\big]$$

$$= \mathbb{E}_{P_{X_2T}}\big[\epsilon_T^2 g(X_2)\big] \tag{91}$$

$$= \lambda \mathbb{E}_{P_{X_2T}}\big[\epsilon_T^2 g(X_2)\big] + (1-\lambda)\mathbb{E}_{Q_{X_2T}}\big[\nu^2\delta_T^2 g(X_2)\big]. \tag{92}$$

Appendix D
Details of the Converse Proof

A. Auxiliary Lemmas

The following two lemmas will be used in various proofs of this section. They are simple extensions of the lemmas in [3, Lemma 1]. Their proofs are thus omitted.

Lemma 3: Let $(T, X_1, X_2, Y) \sim P_{TX_2} P_{X_{1,n}|T} W_{Y|X_1X_2}$ for some pmfs P_{TX_2} and $P_{X_{1,n}|T}$ satisfying $\alpha_{n,t} \triangleq P_{X_{1,n}|T=t}(1) \to 0$ as $n \to \infty$. For any $n \in \mathbb{N}^*$ we have

$$\mathbb{I}(X_1; Y \mid X_2, T)$$
$$= \mathbb{E}_{P_{TX_2}}\Big[\alpha_{n,T} D_Y(X_2) - \mathbb{D}(W_{Y|X_2} \| W_{Y|X_1X_2}(\cdot|0, X_2))\Big]. \tag{93}$$

Lemma 4: Consider for each blocklength n a pmf $P_{X_{1,n}}$ over the binary alphabet $\mathcal{X}_1$ satisfying $\alpha_n \triangleq P_{X_{1,n}}(1) \to 0$ as $n \to \infty$. Define for each $x_2 \in \mathcal{X}_2$:

$$W_{Z|X_2}(z|x_2) \triangleq \alpha_n W_{Z|X_1X_2}(z|1, x_2) + (1-\alpha_n) W_{Z|X_1X_2}(z|0, x_2). \tag{94}$$

Then, for all sufficiently large values of n:

$$(1-\sqrt{\alpha_n})\frac{\alpha_n^2}{2}\chi_{2,Z}(x_2) \le \mathbb{D}(W_{Z|X_2}(\cdot|x_2) \| W_{Z|X_1X_2}(\cdot|0, x_2)) \le (1+\sqrt{\alpha_n})\frac{\alpha_n^2}{2}\chi_{2,Z}(x_2). \tag{95}$$

B. Upper bound on $\log(M_1)$

Since W_1 is uniformly distributed over $[M_1]$, we have:

$$\log(M_1)$$
$$= \mathbb{H}(W_1) \tag{96}$$
$$= \mathbb{H}(W_1 \mid W_2, S) \tag{97}$$
$$= \mathbb{I}(W_1; Y^n \mid W_2, S) + \mathbb{H}(W_1 \mid Y^n, W_2, S) \tag{98}$$
$$\overset{(a)}{\le} \mathbb{I}(W_1; Y^n \mid W_2, S, X_2^n) + \mathbb{H}_b(P_{e1}) + P_{e1}\log(M_1) \tag{99}$$
$$= \frac{1}{1-P_{e1}}(\mathbb{I}(W_1; Y^n \mid W_2, S, X_2^n) + \mathbb{H}_b(P_{e1,1})) \tag{100}$$
$$\overset{(b)}{\le} \frac{1}{1-P_{e1}}\bigg(\sum_i^n \mathbb{H}(Y_i \mid X_{2,i}) - \mathbb{H}(Y_i \mid X_{1,i}, X_{2,i}) + \mathbb{H}_b(P_{e1})\bigg) \tag{101}$$
$$= \frac{1}{1-P_{e1}}\bigg(n\sum_{i=1}^n \frac{1}{n}\mathbb{I}(X_{1,i}; Y_i \mid X_{2,i}) + \mathbb{H}_b(P_{e1,1})\bigg), \tag{102}$$

where (a) holds by Fano's inequality and because $X_2^n = \varphi_2^{(n)}(W_2)$ and (b) holds respectively by the chain rule and because conditioning cannot increase entropy.

Defining T as a uniform random variable over $[|1, n|]$ and independent of all other random variables, we can rewrite (102) as:

$$\log(M_1) \le \frac{1}{1-P_{e1}}(n\mathbb{I}(X_{1,T}; Y_T \mid X_{2,T}, T) + \mathbb{H}_b(P_{e1})) \tag{103}$$
$$\le \frac{1}{1-P_{e1}} n\bigg(\mathbb{E}[\alpha_{n,T} \cdot D_Y(X_{2,T})] + \frac{1}{n}\bigg), \tag{104}$$

where the last step is obtained by applying Lemma 3 for each realization of T, by the nonnegativity of divergence, and by upper-bounding the binary entropy by 1.

C. Lower bound on $\frac{1}{M_2}\sum_{w_2=1}^{M_2}\delta_{n,w_2}$

Recalling the definition of $\widehat{Q}^n_{\mathcal{C},w_2}(z^n)$ in (11) and letting $x_{2,i}(w_2)$ denote the i-th component of codeword w_2, we obtain for a specific code $\mathcal{C}$:

$$\frac{1}{M_2}\sum_{w_2=1}^{M_2} \mathbb{D}\Big(\widehat{Q}^n_{\mathcal{C},w_2} \| W_{Z|X_1X_2}^{\otimes n}(\cdot|0^n, x_2^n(w_2))\Big) \tag{105}$$
$$\overset{(a)}{=} \frac{1}{M_2}\sum_{w_2=1}^{M_2}\sum_{i=1}^n \mathbb{D}\Big(\widehat{Q}^{(i)}_{\mathcal{C},w_2} \| W_{Z|X_1X_2}(\cdot|0, x_{2,i}(w_2))\Big) \tag{106}$$
$$\overset{(b)}{=} \frac{1}{M_2}\sum_{w_2=1}^{M_2}\sum_{i=1}^n \mathbb{D}\Big(\overline{Q}_{\alpha_{n,i},w_2} \| W_{Z|X_1X_2}(\cdot|0, x_{2,i}(w_2))\Big) \tag{107}$$
$$= \sum_{i=1}^n \sum_{x_2} P_{X_{2,i}}(x_2)\mathbb{D}\Big(\overline{Q}_{\alpha_{n,i},w_2} \| W_{Z|X_1X_2}(\cdot|0, x_2)\Big) \tag{108}$$

$$\stackrel{(c)}{\geq} n\mathbb{E}_{P_{TX_{2,T}}}\left[(1-\sqrt{\alpha_{n,T}})\frac{\alpha_{n,T}^2}{2}\chi_{2,Z}(X_{2,T})\right], \tag{109}$$

where we defined T uniform over $[n]$ independent of all other random variables and the last step holds for sufficiently large values of n. Here, (a) holds by the memoryless nature of the channel and upon defining $x_{1,i}(w_1,s)$ as the i-th symbol of codeword $x_1^n(w_1,s)$ and

$$\widehat{Q}^{(i)}_{\mathcal{C},w_2}(z_i) \triangleq \frac{1}{M_1K}\sum_{w_1=1}^{M_1}\sum_{s=1}^{K} W_{Z|X_1X_2}(z_i|x_{1,i}(w_1,s),x_{2,i}(w_2)); \tag{110}$$

(b) holds by recalling the definition in (47):

$$\alpha_{n,i} \triangleq \frac{1}{M_1K}\sum_{w_1=1}^{M_1}\sum_{s=1}^{K}\mathbb{1}\{x_{1,i}(w_1,s)=1\}, \tag{111}$$

and defining

$$\overline{Q}_{\alpha,w_2} \triangleq \alpha W_{Z|X_1X_2}(\cdot|1,x_{2,i}(w_2)) + (1-\alpha)W_{Z|X_1X_2}(\cdot|0,x_{2,i}(w_2)); \tag{112}$$

and (c) holds by Lemma 4 and because the covertness constraint implies that $\alpha_{n,t}\to 0$ for any t. (Proof omitted due to lack of space.)

D. Lower bound on $\log(M_1)+\log(K)$

We start with the lower bound

$$\log(M_1)+\log(K) \geq \mathbb{I}(W_1,S;Z^n \mid X_2^n) \tag{113}$$
$$\stackrel{(a)}{\geq} \mathbb{I}(X_1^n;Z^n \mid X_2^n), \tag{114}$$

where (a) holds because $X_1^n = x_1^n(W_1,S)$ is a function of W_1 and S.

To single-letterize the mutual information $\mathbb{I}(X_1^n;Z^n \mid X_2^n)$, we abbreviate the covertness constraint using the definition of δ_{n,w_2} in (13). Then notice that by

$$\mathbb{E}_{W_2}[\delta_{n,W_2}] = \mathbb{E}_{W_2}\left[\sum_{z^n}\widehat{Q}^n_{\mathcal{C},W_2}(z^n)\log\left(\frac{1}{W^{\otimes n}_{Z|X_1X_2}(z^n|0^n,X_2^n(W_2))}\right)\right] - H(Z^n|X_2^n) \tag{115}$$

we can obtain (119), see (118) and use the nonnegativity of divergence, and where we defined

$$\tilde{W}_{Z|X_2}(z|x_2) \triangleq \sum_{x_1\in\mathcal{X}_1} W_{Z|X_1X_2}(z \mid x_1,x_2)P_{X_{1,T}}(x_1) \tag{116}$$
$$= W_{Z|X_1X_2}(z \mid 0,x_2)\alpha_{n,T} + W_{Z|X_1X_2}(z \mid 1,x_2)(1-\alpha_{n,T}). \tag{117}$$

Recalling then the definition of $\alpha_{n,i}$ in (111) and we obtain the bound:

$$\mathbb{I}(X_1^n;Z^n \mid X_2^n) \geq n\mathbb{I}(X_{1,T};Z_T \mid X_{2,T},T) - \mathbb{E}_{W_2}[\delta_{n,W_2}]. \tag{120}$$

Combining (120) with (114) and applying Lemma 3 followed by Lemma 4 for each realization of T, continue with

$$\log(M_1)+\log(K) \geq n\mathbb{E}_{P_{TX_2}}\left[\alpha_{n,T}D_Z(X_2) - \frac{\alpha_{n,T}^2}{2}\chi_{2,Y}(X_2)(1-\sqrt{\alpha_{n,T}})\right] + \mathbb{E}_{W_2}[\delta_{n,W_2}]. \tag{121}$$

Notice that by assumption, $\delta_{n,w_2}\to 0$ for any w_2 and thus

$$\lim_{n\to\infty}\mathbb{E}_{W_2}[\delta_{n,W_2}] = 0. \tag{122}$$

Moreover, the second term in the expectation is dominated by the first term because the covertness constraint $\delta_{n,w_2}\to 0$ for any w_2 implies that $\alpha_{n,t}\to 0$ (proof omitted.)

Combining these observations with (121) and the lower bound on $\frac{1}{M_2}\sum_{w_2=1}^{M_2}\delta_{n,w_2}$ establishes the desired result.

E. Asymptotic Analysis

To conclude the proof, we notice that by the Bolzano-Weierstrass Theorem there exists an increasing subsequence $\{n_i\}$ so that $\{P_{X_{2,T}}(\cdot|t)\}$ and $\{P_T(\cdot)\}$ converge on this subsequence. If also $\gamma_{n_i,t}$ converges for each value of $t\in\mathcal{T}\triangleq\{1,\ldots,4\}$, then in view of bounds (44), (46), and (50), the desired bounds (31)–(33) follow immediately by considering the convergence points γ_t of these sequences and defining

$$\epsilon_t := \frac{\gamma_t}{\max_{t'\in\mathcal{T}}\gamma_t'}. \tag{123}$$

Otherwise, if some of the $\gamma_{n_i,t}$ diverge to ∞, we notice that for each of these t-values the probability $P_T(t)\to 0$ as $n\to\infty$ and because by definition the expectation $\mathbb{E}[\gamma_{n,t}]=1$ one of the following three cases applies:

1.) $P_T(t)\gamma_{n_i,t}\to 0$ and $P_T(t)\gamma_{n_i,t}^2\to 0$;
2.) $P_T(t)\gamma_{n_i,t}\to 0$ and $\lim_{n_i\to\infty}P_T(t)\gamma_{n_i,t}^2 = c$ for $c\in(0,\infty)$;
3.) $P_T(t)\gamma_{n_i,t}\in[0,1]$ and $P_T(t)\gamma_{n_i,t}^2\to\infty$.

All t-values satisfying case 1.) can simply be ignored since they do not change the bounds. Whenever there exists a t-value in case 3.), then bounds (46) and (50) are 0 and the result is trivial. In case 2.) we can modify the probabilities $P_T(t)$ and the parameters $\gamma_{n_i,t}$ to values in a bounded interval $[a,b]$ for $b>a>0$, while still approximating the bounds (44), (46), and (50) arbitrarily closely. We then fall back to the case where all sequences $\gamma_{n_i,t}$ converge.

F. A tighter cardinality bound for the auxiliary random variable T

Since by Lemma 2 the set of all valid vectors (r_1,r_2,k) is convex, its dominant boundary points are all maximizers of the objective function defined by

$$\sqrt{2}\frac{(\mu_1+\mu_3)\mathbb{E}_{P_{TX_2}}[\epsilon_T D_Y(X_2)]}{\sqrt{\mathbb{E}_{P_{TX_2}}[\epsilon_T^2\chi_{2,Z}(X_2)]}}$$

$$\mathbb{I}(X_1^n; Z^n \mid X_2^n)$$
$$= \mathbb{H}(Z^n \mid X_2^n) - \mathbb{H}(Z^n \mid X_1^n, X_2^n) + \mathbb{E}_{W_2}\left[\sum_z \widehat{Q}^n_{\mathcal{C},w_2}(z^n) \log\left(\frac{1}{W^{\otimes n}_{Z|X_1X_2}(z^n|0^n, X_2^n(W_2))}\right)\right] - \mathbb{H}(Z^n \mid X_2^n) - \mathbb{E}_{W_2}[\delta_{n,W_2}] \tag{118}$$
$$\geq n\sum_{t=1}^{n}\sum_{(x_1,x_2,z)} P_{X_{1,T},X_{2,T},Z_T,T}(x_1,x_2,z,t)\log\left(\frac{W_{Z|X_1X_2}(z \mid x_1,x_2)}{W_{Z|X_1X_2}(z|0,x_2)}\right) - \mathbb{E}_{W_2}[\delta_{n,W_2}] - n\mathbb{D}(\tilde{W}_{Z|X_2}\|W_{Z|X_1X_2}(\cdot|0,x_2)). \tag{119}$$

$$-\sqrt{2}\frac{\mu_3\mathbb{E}_{P_{TX_2}}[\epsilon_T D_Z(X_2)]}{\sqrt{\mathbb{E}_{P_{TX_2}}[\epsilon_T^2\chi_{2,Z}(X_2)]}}$$
$$+\mu_2\mathbb{I}(X_2;Y \mid X_1=0,T), \tag{124}$$

for some positive values $\mu_1, \mu_2, \mu_3 \geq 0$. Fix any triple $\mu_1, \mu_2, \mu_3 \geq 0$ and any positive constant $c > 0$. Then, for each pair (P_{X_2}, ϵ), where $\epsilon \in [0,1]$, define the two-dimensional vector $\boldsymbol{v} = (v_1, v_2)$ with first component

$$v_1 = \sqrt{2}\frac{(\mu_1+\mu_3)\epsilon\mathbb{E}_{P_{X_2}}[D_Y(X_2)]}{\sqrt{c}} - \sqrt{2}\frac{\mu_3\epsilon\mathbb{E}_{P_{X_2}}[D_Z(X_2)]}{\sqrt{c}}$$
$$+\mu_2\mathbb{I}(X_2;Y \mid X_1=0), \tag{125}$$

and second component

$$v_2 = \epsilon^2\mathbb{E}_{P_{X_2}}[\chi_{2,Z}(X_2)]. \tag{126}$$

By the Fenchel-Eggleston strengthening of Carathéodory's theorem, we can conclude that each point in the convex hull of the two-dimensional vectors can be obtained as an average of 2 vectors. As a consequence, if for some pmf P_{TX_2} and tuple $\epsilon_1, \ldots, \epsilon_4$ we choose

$$c = \mathbb{E}_{P_{TX_2}}\left[\epsilon_T^2 \cdot \chi_{2,Z}(X_2)\right], \tag{127}$$

we can conclude that there exists a new pair $\tilde{P}_{TX_2}$ and $(\tilde{\epsilon}_1, \tilde{\epsilon}_2)$ with T only over the alphabet $\{1,2\}$ and so that the term in (124) evaluates to the same value as for the original pmf P_{TX_2} and tuple $(\epsilon_1, \ldots, \epsilon_4)$.