Covert Capacity of Degraded Broadcast Channels

Yossef Steinberg
Technion–Israel Institute of Technology
Haifa, Israel
ysteinbe@technion.ac.il

Michèle Wigger
Université Paris-Saclay, CNRS, CentraleSupélec, L2S
91190 Gif-sur-Yvette, France
michele.wigger@centralesupelec.fr

***Abstract*—We derive the capacity region of the degraded broadcast channel (DBC) subject to the constraint that the communication is not detected by an adversary, the Warden. Our capacity result is in a computable form and numerical results show that time-sharing is suboptimal in general, and improved rates can be obtained through superposition coding.**

***Index Terms*—Broadcast channels, covert communication, degraded broadcast channels.**

I. Introduction

Communication subject to information-theoretic security constraints has a long history in information theory; see, e.g., the seminal work by Wyner [1]. More recently, significant attention has been given to communication systems that are subject to a covertness constraint, i.e., systems in which an external adversary, a so-called Warden, is not allowed to learn even the mere fact that communication is taking place. This requirement is typically enforced by imposing that the Kullback–Leibler divergence (or other measures such as the variational distance) between the warden's actual channel output distribution and the hypothetical output distribution assuming that the transmitter always sends a specific zero symbol x_0 remains below a given threshold $\delta > 0$.

The work [2] first showed that, under the above covertness assumption, reliable communication over memoryless Gaussian channels is possible, provided that the number of communicated information bits scales proportionally to the square root of the number of channel uses. Covert rates are therefore commonly defined as $L := \frac{1}{\sqrt{n\delta}} \log_2 \nu$, for n the blocklength, δ the covertness constraint, and ν the size of the message set. Covert capacities for discrete memoryless channels (DMCs) and Gaussian memoryless channels were determined in [2]–[5], and were shown not to depend on the parameter δ. Moreover, [4] also determined the rate of the secret key shared between encoders and decoders required to achieve this common capacity. This result was extended to all covert rates (not only capacity) in [6].

Covert capacity regions for discrete memoryless multi-access interference channels were determined in [7] and [8], and the required key rates at all covert rates in [6]. Broadcast channels (BC) under a covertness constraint were studied in [9]–[11]. The works in [9], [10] considered a mixed covert/non-covert scenario with a non-covert communication from the transmitter to both receivers and a covert communication from the transmitter to only one of the receivers, which needs to remain undetectable (covert) to the other receiver. In contrast, the work in [11] and this present article both consider a scenario where the transmitter sends individual messages to two broadcast receivers, and the entire communication needs to remain undetectable to an external warden.

The time-sharing region in this setup can be written as [11]:

$$\mathcal{L}^{(\text{TS})} = \left\{ (L_1, L_2) \colon \frac{L_1}{L_1^*} + \frac{L_2}{L_2^*} \leq 1 \right\} \tag{1}$$

for L_1^* and L_2^* the covert capacities of the BC marginal channel transition laws $P_{Y_1|X}$ and $P_{Y_2|X}$. The work in [11] proved that time-sharing is optimal for all BCs with marginal capacities satisfying $L_1^* \geq L_2^*$and

$$\frac{L_1^*}{L_2^*} \geq \sup_{P_X} \frac{I(X;Y_1)}{I(X;Y_2)}. \tag{2}$$

The same work also proved optimality of time-sharing for all Gaussian BCs and binary symmetric BCs.

In this work, we show that optimality of time-sharing does not hold for general (stochastically or physically) degraded channels, and superposition coding can strictly improve over the time-sharing region. In fact, we present a computable characterization of the superposition coding region and show that it achieves capacity for all degraded BCs.

II. Definitions

Let $\mathcal{X}$, $\mathcal{Y}_1$, $\mathcal{Y}_2$, $\mathcal{Z}$ be finite sets. Denote by $\mathcal{P}(\mathcal{X})$ the class of all distributions on $\mathcal{X}$. A discrete memoryless broadcast channel (BC) with two users and a warden is a quintuple $\{\mathcal{X}, P_{Y_1,Y_2,Z|X}, \mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Z}\}$ where $\mathcal{X}$ is the input alphabet, $\mathcal{Y}_k$ is the output alphabet of user k, for $k = 1, 2$, $\mathcal{Z}$ is the output alphabet at the warden and $P_{Y_1,Y_2,Z|X}$ is a transition probability matrix from $\mathcal{X}$ to $\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Z}$. We denote by $P_{Z|X}$ and $P_{Y_k|X}$, for $k = 1, 2$, the conditional marginals of $P_{Y_1,Y_2,Z|X}$, and by $P_X \circ P_{Z|X}$ the distribution on $\mathcal{Z}$ induced by P_X at the input. When P_X is understood from the context, we will use the notation P_Z, and similarly for $P_X \circ P_{Y_k|X}$, P_{Y_k} etc. Let $x_0 \in \mathcal{X}$ stand for the zero symbol, namely, the symbol fed into the channel when no communication is taking place. Define

$$Q_0(z) = P_{Z|X}(z|x_0), \quad z \in \mathcal{Z}, \tag{3}$$

and let $Q_0^{\times n}$ stand for the n-fold product of Q_0, i.e.,

$$Q_0^{\times n}(z^n) = \prod_{i=1}^{n} Q_0(z_i). \tag{4}$$

The goal in covert communication is to transmit information to the receivers while keeping the distribution at the warden output close to $Q_0^{\times n}$. Fix integers ν_k, $k=1,2$ and transmission length n. Let $\mathcal{N}_k = [1:\nu_k]$ stand for the set of messages of user k. The transmitter and legitimate receivers share a random secret key S, taking values in a finite set $\mathcal{S}$. We assume that the key is of sufficiently large randomness, thus do not specify the size of $\mathcal{S}$.

Definition 1: An $(n,\nu_1,\nu_2,\epsilon,\delta)$ covert code for the BC $\{\mathcal{X}, P_{Y_1,Y_2,Z|X}, \mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Z}\}$ with a warden consists of an encoder

$$f: \mathcal{N}_1 \times \mathcal{N}_2 \times \mathcal{S} \to \mathcal{X}^n, \tag{5}$$

and pair of decoders

$$\phi_k: \mathcal{Y}_k^n \times \mathcal{S} \to \mathcal{N}_k, \qquad k=1,2, \tag{6}$$

such that the probabilities of error are bounded by ϵ:

$$P_{e,k} = \frac{1}{\nu_1\nu_2} \sum_{\substack{(m_1,m_2)\\ \in \mathcal{N}_1\times\mathcal{N}_2}} \sum_{s\in\mathcal{S}} P_S(s) P_{Y_k^n|X^n}\big(D_{m_k}^c(s)|f(m_1,m_2,s)\big) \le \epsilon, \quad k=1,2, \tag{7}$$

and the output distribution at the warden approximates $Q_0^{\times n}$ in the divergence sense:

$$D\big(P_{Z^n}||Q_0^{\times n}\big) \le \delta. \tag{8}$$

The set $D_{m_k}(s)$ in (7) is the decoding region of message m_k

$$D_{m_k}(s) = \{y_k^n: \quad \phi_k(y_k^n, s) = m_k\}, \qquad k=1,2, \tag{9}$$

and P_{Z^n} in (8) stands for the distribution of the warden output Z^n induced by the operation of the code:

$$P_{Z^n}(z^n) = \frac{1}{\nu_1\nu_2} \sum_{\substack{(m_1,m_2)\\ \in \mathcal{N}_1\times\mathcal{N}_2}} \sum_{s\in\mathcal{S}} P_S(s) P_{Z^n|X^n}(z^n|f(m_1,m_2,s)). \tag{10}$$

In this work, we focus on BCs that satisfy the following conditions, that are now standard in covert communications [4], [5], [11].

Conditions 1 (Non-redundancy and absolute continuity):

a) The zero symbol is not redundant at the warden output. I.e., $Q_0 \notin \mathrm{CH}[P_{Z|X}(\cdot|x'),\ x' \in \mathcal{X}\setminus\{x_0\}]$, where CH stands for the convex hull.

b) Absolute continuity w.r.t. x_0 symbol at the warden: $P_{Z|X}(\cdot|x) \ll Q_0 \quad \forall x \in \mathcal{X}$.

c) Absolute continuity w.r.t. x_0 symbol at the users output: $P_{Y_k|X}(\cdot|x) \ll P_{Y_k|X}(\cdot|x_0) \quad \forall x\in\mathcal{X}, \quad k=1,2$.

Part a) of Conditions 1 guarantees that the encoder cannot mimic the no communication state with some input distribution P_X, that results with output distribution at the warden being equal to Q_0. If part b) is not satisfied, than there is an input symbol x' that the encoder cannot use, effectively reducing the input alphabet size. For the single user channel, it is shown in [4] that if c) does not hold, the number of covert bits that can be transmitted grows like $\sqrt{n}\log n$. Thus, Conditions 1 are the most pessimistic assumptions that still allow covert communications. For details, see [4, Appendix G] and [11].

The covert rates of the code are defined as

$$L_k = \frac{\log \nu_k}{\sqrt{n\delta}}, \quad k=1,2 \tag{11}$$

A pair of covert rates (L_1,L_2) is called δ-achievable if for any $\epsilon>0$, $\rho>0$ and sufficiently large n there exists an $(n, 2^{\sqrt{n\delta}(L_1-\rho)}, 2^{\sqrt{n\delta}(L_2-\rho)}, \epsilon, \delta)$ covert code for the BC $P_{Y_1,Y_2,Z|X}$. The collection of all δ-achievable pairs is called the covert capacity region, and is denoted by $\mathcal{L}_\delta^*$. As we will see, it does not depend on the value of $\delta>0$.

In this work we derive the covert capacity region for stochastically degraded broadcast channels, where the degradation is between the legitimate users, i.e., we assume that there exists a conditional distribution $P_{Y_2|Y_1}$ such that

$$P_{Y_2|X}(y_2|x) = \sum_{y_1} P_{Y_1|X}(y_1|x) P_{Y_2|Y_1}(y_2|y_1). \tag{12}$$

By the problem definition, $\mathcal{L}_\delta^*$ depends on $P_{Y_1,Y_2,Z|X}$ only via its conditional marginals. Hence in the sequel a BC with a warden is referred to as three channels with common input $\{P_{Y_1|X}, P_{Y_1|X}, P_{Z|X}\}$, where the alphabets are understood from the context. In addition, no distinction has to be made between stochastically and physically degraded models, and they are commonly referred to as degraded channels.

III. Main Results

Let $\mathcal{L}_{n,\delta}^{(I)}$ stand for the collection of nonnegative pairs (L_1,L_2) satisfying

$$L_1 \le \sqrt{n/\delta} I(X;Y_1|U) \tag{13a}$$

$$L_2 \le \sqrt{n/\delta} I(U;Y_2) \tag{13b}$$

where mutual informations are calculated according to $P_{U,X}P_{Y_1Y_2|X}$ for some $P_{U,X}$ so that the induced $P_Z = P_X \circ P_{Z|X}$ satisfies

$$D(P_Z||Q_0) \le \frac{\delta}{n}. \tag{13c}$$

Define

$$\mathcal{L}_\delta^{(I)} = \bigcap_{n\ge 1} \mathcal{L}_{n,\delta}^{(I)}. \tag{13d}$$

Before proceeding to our main result, we state a few properties of the region $\mathcal{L}_\delta^{(I)}$. To exhaust $\mathcal{L}_\delta^{(I)}$ it is enough to restrict the alphabet $\mathcal{U}$ to satisfy

$$|\mathcal{U}| \le |\mathcal{X}| + 1. \tag{14}$$

The bound (14) is proved using the support lemma [12] for every n. Note that the presence of the additional constraint (13c) does not increase the alphabet size of U, because when applying the support lemma to restrict $|\mathcal{U}|$, the distribution of X is preserved, hence also (13c). The details are omitted.

Let $\check{P}_{U,X}^{(n)}$ be a sequence of distributions that achieves a rate pair $(l_1, l_2) \in \mathcal{L}_\delta^{(I)}$. Since the alphabets are finite, $\mathcal{P}(\mathcal{U}\times\mathcal{X})$ is compact, hence $\check{P}_{U,X}^{(n)}$ converges to a limit distribution on a subsequence n_k, $k = 1, 2...$, with $n_k < n_{k+1}$. Define a sequence of distributions $P_{U,X}^{(n)}$ as follows:

$$P_{U,X}^{(n_k)} = \check{P}_{U,X}^{(n_k)} \quad k = 1, 2... \tag{15a}$$

$$P_{U,X}^{(n)} = \check{P}_{U,X}^{(n_k)} \quad n_{k-1} < n \le n_k. \tag{15b}$$

Then $P_{U,X}^{(n)}$ converges, and achieves the same rate pair $(l_1, l_2) \in \mathcal{L}_\delta^{(I)}$. To simplify notation, from this point on we drop the superscript (n) and use just $P_{U,X}$ with the understanding that the distributions depend on n and converge.

Theorem 1: For any discrete memoryless degraded BC with a warden, the following holds:

1) $\mathcal{L}_\delta^* = \mathcal{L}_\delta^{(I)}$.
2) A sequence of distributions $P_{U,X}$ achieves a positive L_2 according to constraint (13b) only if there exists a subset $B \subset \mathcal{U}$ such that

$$P_U(B) > 0 \tag{16a}$$

$$\lim_{n\to\infty} P_U(B) = 0. \tag{16b}$$

See Appendix A.

Note that although the size of $\mathcal{U}$ is finite and fixed, $\mathcal{L}_\delta^{(I)}$ is still not a computable result, since it involves the limit $n \to \infty$. The following computable region $\tilde{\mathcal{L}}^{(I)}$ coincides with $\mathcal{L}_\delta^{(I)}$, which is stated in Theorem 2 and proved in Section IV ahead.

Let $\tilde{\mathcal{L}}^{(I)}$ stand for the collection of nonnegative pairs (L_1, L_2) satisfying:

$$L_1 \le \sqrt{\frac{2}{\chi_2(\nu)}}\Big[(1-\nu)\sum_x \tilde{P}_X^A(x) D\big(P_{Y_1|X}(\cdot|x)||P_{Y_1|X}(\cdot|x_0)\big) + \nu I^B(Y_1; X|U)\Big] \tag{17a}$$

$$L_2 \le \sqrt{\frac{2}{\chi_2(\nu)}}\Big[\nu\sum_x P_X^B(x) D\big(P_{Y_2|X}(\cdot|x)||P_{Y_2|X}(\cdot|x_0)\big) - \nu I^B(Y_2; X|U)\Big], \tag{17b}$$

for some $\nu \in [0, 1]$, auxiliary alphabet $\mathcal{U}$ of size not exceeding $|\mathcal{X}| + 1$, a singleton set $A = \{u_0\} \subset \mathcal{U}$, its complement $B = \mathcal{U}\backslash A$, and pmfs P_{UX}^B over $B \times \mathcal{X}$ and $\tilde{P}_X^A$ over $\mathcal{X}\backslash\{x_0\}$, where in the above mutual informations are calculated with respect to the pmf P_{UX}^B and

$$\chi_2(\nu) := \chi_2\Big((1-\nu)\tilde{P}_Z^A + \nu P_Z^B \| Q_0\Big), \tag{18}$$

where $\chi_2(\cdot\|\cdot)$ denotes the χ_2-distance:

$$\chi_2(P\|Q) := \sum_{z\in\mathcal{Z}} \frac{(P(z) - Q(z))^2}{Q(z)}. \tag{19}$$

Theorem 2: It holds that $\tilde{\mathcal{L}}^I = \mathcal{L}_\delta^{(I)}$.

The proof of above theorem is given in Section IV.

The time-sharing region $\mathcal{L}^{(\mathrm{TS})}$ is obviously included in our region, see Appendix C.

A. Comparison to Time-sharing and Numerical Examples

In [11], it was shown that when (2) holds, time-sharing is optimal and suffices to achieve $\mathcal{L}_\delta$. We will reprove this result using our capacity-expression in (17a) and (17b).

Notice first that (2) in particular holds when X is binary with probability $1-\alpha$ equal to x_0 and with probability α equal to x, for arbitrary $x \in \mathcal{X}\backslash\{x_0\}$ and $\alpha > 0$. Letting $\alpha \to 0$, we can conclude that for any $x \in \mathcal{X}\backslash\{x_0\}$:

$$\frac{L_1^*}{L_2^*} \ge \lim_{\alpha\to 0}\frac{I(X;Y_1)}{I(X;Y_2)} = \frac{D(P_{Y_1|X}(\cdot|x)\|P_{Y_1|X}(\cdot|x_0))}{D(P_{Y_2|X}(\cdot|x)\|P_{Y_2|X}(\cdot|x_0))}, \tag{20}$$

because for above choice of X and when $\alpha \to 0$ we have $I(X;Y_k) = \alpha D(P_{Y_k|X}(\cdot|x)\|P_{Y_k|X}(\cdot|x_0))\cdot(1+o(1))$, for $k = 1, 2$.

If in the following expression we apply above inequality (20) and (2) on the individual summands, we can write

$$\begin{aligned}
&\nu\sum_{u\in B} P_U^B(u)\frac{I^B(Y_1; X|U=u)}{L_1^*} \\
&\quad + (1-\nu)\sum_x \tilde{P}_X^A(x)\frac{D\big(P_{Y_1|X}(\cdot|x)\|P_{Y_1|X}(\cdot|x_0)\big)}{L_1^*} \\
&\le \nu\sum_{u\in B} P_U^B(u)\frac{I^B(Y_2; X|U=u)}{L_2^*} \\
&\quad + (1-\nu)\sum_x \tilde{P}_X^A(x)\frac{D\big(P_{Y_2|X}(\cdot|x)||P_{Y_2|X}(\cdot|x_0)\big)}{L_2^*}.
\end{aligned} \tag{21}$$

Plugging (21) into the upper bound on $\frac{L_1}{L_1^*} + \frac{L_2}{L_2^*}$ obtained from (17a)–(17b), allows to conclude that for channels satisfying (2), any achievable pair (L_1, L_2) lies in the time-sharing region $\mathcal{L}^{(\mathrm{TS})}$ defined in (1). For details, see Appendix D.

Example 1: Consider a ternary input alphabet $\mathcal{X} = \{0, 1, 2\}$ and quaternary output alphabet $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Z} = \{0, 1, 2, 3\}$. Let $x_0 = 0$ and consider the following channel transition laws

$$P_1 = \begin{bmatrix} 0.2 & 0.28 & 0.28 & 0.24 \\ 0.05 & 0.1 & 0.45 & 0.4 \\ 0.07 & 0.37 & 0.4 & 0.16. \end{bmatrix} \tag{22}$$

and

$$P_2 = \begin{bmatrix} 0.1884 & 0.324 & 0.232 & 0.2556 \\ 0.0515 & 0.215 & 0.331 & 0.4025 \\ 0.0744 & 0.399 & 0.326 & 0.2006 \end{bmatrix} \tag{23}$$

for the legitimate receivers and

$$Q = \begin{bmatrix} 0.20 & 0.19 & 0.36 & 0.25 \\ 0.01 & 0.37 & 0.17 & 0.45 \\ 0.42 & 0.35 & 0.05 & 0.18 \end{bmatrix} \tag{24}$$

for the warden. Notice that the channel from Y_1 to Y_2 is stochastically degraded because we can write $P_2 = P_1 \cdot W$ for

$$W = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 \\ 0.02 & 0.8 & 0.12 & 0.06 \\ 0.01 & 0.2 & 0.7 & 0.09 \\ 0 & 0.1 & 0.01 & 0.89 \end{bmatrix}. \tag{25}$$

The covert capacities for the two single-user channels P_1 and P_2 in the presence of the warden Q are $L_1^* = 0.46809$ and $L_2^* = 0.28590$. Figure 1 shows the boundary of the region $\mathcal{L}_\delta$ (solid line) and the boundary of the time-sharing region $\mathcal{L}^{(\mathrm{TS})}$ (dashws line). We observe that for this example, superposition coding improves over time-sharing whenever $L_1 < L_1^*$ or $L_2 < L_2^*$.

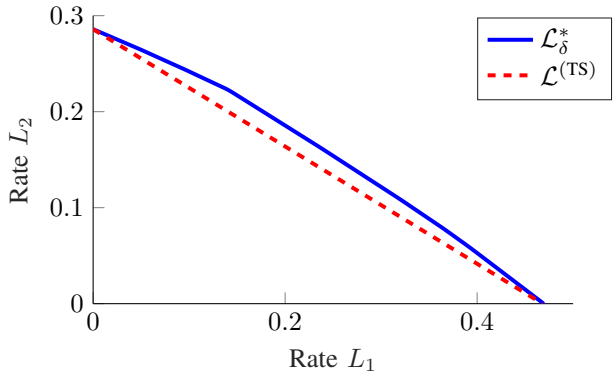

Fig. 1. Illustration of the capacity region $\mathcal{L}_\delta$ (solid line) and the time-sharing region $\mathcal{L}^{(\mathrm{TS})}$ (dashed line).

Example 2: Consider a second example with binary inputs $\mathcal{X} = \{0, 1\}$, for $x_0 = 0$, and ternary outputs $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Z} = \{0, 1, 2\}$. Let the channel to the strong receiver P_1 be a BSC(0.2) and the channel to the warden Q be a BSC(0.4). The channel to the weaker receiver $P_2 = P_1 \cdot W$, for

$$W = \begin{bmatrix} 0.9 & 0.1 \\ c & 1-c \end{bmatrix}, \tag{26}$$

where we study different values of $c \in \{0, 0.1, 0.2, \ldots, 1\}$. Table I shows the maximum coefficient

$$\gamma^* = \max_{(L_1, L_2) \in \mathcal{L}_\delta} \left(\frac{L_1}{L_1^*} + \frac{L_2}{L_2^*} \right) \tag{27}$$

for different values of c. This parameter γ^* captures by how much one can improve over the time-sharing region, for which the parameter cannot exceed 1. The second column of Table I indicates whether the condition $\frac{L_1^*}{L_2^*} \geq \max_{P_X} \frac{I(X;Y_1)}{I(X;Y_2)}$ is satisfied (indicated by 1 in the table) or not (indicated with a 0). It has been shown in [11] that time-sharing is optimal and thus $\gamma^* = 1$ whenever the condition is satisfied. Our results seem to imply that for this example one can improve over time-sharing whenever the condition does not hold.

c	$\frac{L_1^*}{L_2^*} \geq \max_{P_X} \frac{I(X;Y_1)}{I(X;Y_2)}$	γ^*
0.0	1	1.0000
0.1	1	1.0000
0.2	0	1.0047
0.3	0	1.0108
0.4	0	1.0153
0.5	0	1.0178
0.6	0	1.0178
0.7	0	1.0148
0.8	0	1.0078
0.9	1	1.0000
1.0	1	1.0000

TABLE I
RESULTS FOR OUR BINARY-INPUT CHANNEL EXAMPLE.

IV. Proof of Theorem 2: Computable Capacity Characterization

We construct here a sufficiently general joint distribution $P_{U,X}$ that adheres to (13c). A fully-general distribution can be obtained by letting in the following the chosen distributions P_U^B, P_U^A, $P_{X|U}$, and $\tilde{P}_{X|U}$ depend on n. However, a close inspection reveals that the constraints L_1 and L_2 only depend on limiting points of these distributions and not on how they evolve with n. To avoid cumbersome notation, we therefore assume these probability laws to be constant.

Let B be a proper subset of $\mathcal{U}$ and A its complement. Let P_U^B (resp. P_U^A) be a general distribution on B (resp. on A), $P_{X|U}(\cdot|u)$ a conditional distribution on $\mathcal{X}$ for $u \in B$, and $\tilde{P}_{X|U}(\cdot|u)$ a conditional distribution on $\mathcal{X} \setminus x_0$ for $u \in A$. With these definitions, we set

$$\hat{P}_U(u) = (1-\mu_1) P_U^A(u) + \mu_1 P_U^B(u) \tag{28a}$$

$$\hat{P}_{X|U}(x|u) = \begin{cases} (1-\mu_2)\mathbb{1}_{x_0}(x) + \mu_2 \tilde{P}_{X|U}(x|u) & \text{for } u \in A, \\ P_{X|U}(x|u) & \text{for } u \in B, \end{cases} \tag{28b}$$

where

$$\tilde{P}_{X|U}(x_0|u) = 0 \quad \forall u \in A, \tag{28c}$$

and μ_1, μ_2 are small parameters that tend to 0 as $n \to \infty$, at rates to be determined later. The structure we suggest in (28a) and (28b) determines $P_{U,X}$, and thus also $P_{X,Z}$ and P_{U,X,Y_k}, $k = 1, 2$. We define below the notation needed for the characterization of the computable region. The distribution of X is given by

$$\hat{P}_X(x) = \sum_u \hat{P}_{X|U}(x|u) \hat{P}_U(u) \tag{29a}$$

$$= \bar{\mu}_1 \bar{\mu}_2 \mathbb{1}_{x_0}(x) + \bar{\mu}_1 \mu_2 \tilde{P}_X^A(x) + \mu_1 P_X^B(x) \tag{29b}$$

where $\mathbb{1}_{x_0}$ puts mass 1 on x_0, and we use the notation

$$\tilde{P}_X^A(x) \triangleq \sum_{u \in A} \tilde{P}_{X|U}(x|u) P_U^A(u) \tag{29c}$$

$$P_X^B(x) \triangleq \sum_{u \in B} P_{X|U}(x|u) P_U^B(u) \tag{29d}$$

Note that $\tilde{P}_X^A(x_0) = 0$, and $\hat{P}_X \to \mathbb{1}_{x_0}$ as $\mu_1, \mu_2 \to 0$. For simplicity of exposition we also define

$$\tilde{P}_Z^A(z) \triangleq \sum_x P_{Z|X}(z|x) \tilde{P}_X^A(x) \tag{30a}$$

$$P_Z^B(z) \triangleq \sum_x P_{Z|X}(z|x) P_X^B(x) \tag{30b}$$

$$\tilde{P}_{Y_k|U}^A(y_k|u) \triangleq \sum_x P_{Y_k|X}(y_k|x) \tilde{P}_{X|U}(x|u) \quad \text{for } u \in A,$$

$$k=1,2. \quad (30c)$$

$$P^B_{Y_k|U}(y_k|u) \triangleq \sum_x P_{Y_k|X}(y_k|x)P_{X|U}(x|u) \quad \text{for } u\in B,$$

$$k=1,2. \quad (30d)$$

Observe that P^A_U, P^B_U and (29c–30d) do not depend on μ_1,μ_2. Define now the normalized parameters

$$\eta_1 = \sqrt{n/\delta}\mu_1 \quad (31a)$$

$$\eta_2 = \sqrt{n/\delta}\mu_2. \quad (31b)$$

Theorem 2 is obtained by evaluating the region $\mathcal{L}^{(I)}$ for above choice of distributions based on the Taylor expansions of the terms $I(U;Y_2)$, $I(X;Y_1|U)$ and $D(P_Z||Q_0)$ near $\mu_1=0$, $\mu_2=0$.

As proved in Appendix B, this Taylor expansion results in the rate expressions

$$L_1 \le \Big[\sum_x P^B_X(x)D\big(P_{Y_1|X}(\cdot|x)||P_{Y_1|X}(\cdot|x_0)\big) - \sum_{u\in B} P^B_U(u)D\big(P_{Y_1|U}(\cdot|u)||P_{Y_1|X}(\cdot|x_0)\big)\Big]\eta_1 + \sum_x \tilde{P}^A_X(x)D\big(P_{Y_1|X}(\cdot|x)||P_{Y_1|X}(\cdot|x_0)\big)\eta_2 \quad (32a)$$

$$L_2 \le \sum_{u\in B} P^B_U(u)D\big(P_{Y_2|U}(\cdot|u)||P_{Y_2|X}(\cdot|x_0)\big)\eta_1, \quad (32b)$$

while the divergence constraint evaluates to

$$\eta_1^2\chi_2(P^B_Z||Q_0)+\eta_2^2\chi_2(\tilde{P}^A_Z||Q_0) + \eta_1\eta_2\sum_{z\in\mathcal{Z}}\frac{(\tilde{P}^A_Z(z)-Q_0(z))(P^B_Z(z)-Q_0(z))}{Q_0(z)} \le 2. \quad (32c)$$

Without loss in optimality, in the parametrization above we can restrict the set A to be a singleton (=because the result only depends on $\tilde{P}^A_X$)=. Similarly, the rate-constraints are loosest if η_1 and η_2 are chosen so that constraint (32c) is satisfied with equality. We thus reparametrize η_1 and η_2 as $\eta_1 = c\cdot\nu$ and $\eta_2 = c\cdot(1-\nu)$ for $\nu\in[0,1]$ and $c>0$, where the latter should be chosen to ensure equality in (32c) we obtain the characterization in the theorem.

Acknowledgment

This work was supported by the ERC under Grant Agreement 101125691.

Appendix A
Proof of Theorem 1

A. Proof of Part 2)

By the single-user results in [5], it is clear that we can achieve positive rates (L_1,L_2) (apply a simple time-sharing scheme). Therefore, the distributions $P_{U,X}$ that maximize the outer bound in Theorem 1 under the constraint (13c), should yield $(I(X;Y_1|U), I(U;Y_2))$ that decay like $n^{-1/2}$ as $n\to\infty$. We claim that this can be achieved only when the U-marginal of $P_{U,X}$ has a set $B\subset\mathcal{U}$ whose probability decays to 0 as $n\to\infty$.

Proposition 1: $I(U;Y_2)=O(n^{-1/2})$ only if there exists a subset $B\subset\mathcal{U}$ such that

$$P_U(B) > 0 \quad (33a)$$

$$\lim_{n\to\infty} P_U(B) = 0. \quad (33b)$$

Proof: The requirement (13c) implies the following structure on P_X ([5, eq. (32)]):

$$\hat{P}_X(x) = (1-\mu)\mathbb{1}_{x_0}(x) + \mu\tilde{P}_X(x) \quad (34)$$

where $\mathbb{1}_{x_0}(x)$ (resp. $\tilde{P}_X$) puts mass 1 (resp. 0) on x_0, and

$$\mu = O(n^{-1/2}). \quad (35)$$

Due to the Markov chain $U \multimap X \multimap Y_2$, (34) and (35), we have

$$\lim_{n\to\infty} I(U;Y_2) = 0. \quad (36)$$

Therefore we can write

$$I(U;Y_2) = \frac{\partial}{\partial\mu}I(U;Y_2)\Big|_{\mu=0}\mu + O(\mu^2), \quad (37)$$

where

$$\frac{\partial}{\partial\mu}I(U;Y_2) = \frac{\partial}{\partial\mu}\sum_{u,y_2}P_{U,Y_2}(u,y_2)\log\frac{P_{U,Y_2}(u,y_2)}{P_U(u)P_{Y_2}(y_2)} = \sum_{u,y_2}\Big[\frac{\partial}{\partial\mu}P_{U,Y_2}(u,y_2)\Big]\log\frac{P_{Y_2|U}(y_2|u)}{P_{Y_2}(y_2)}. \quad (38)$$

For the proof of (38), see Appendix E-A. Fix $\alpha>0$, and assume that

$$P_U(u') \ge \alpha \quad \forall n \quad (39)$$

for some $u'\in\mathcal{U}$. Then we must have

$$\lim_{n\to\infty} P_{X|U}(x_0|u') = 1, \quad (40)$$

as otherwise (13c) does not hold. Thus, if (39) holds for all $u\in\mathcal{U}$, then (40) holds for all elements of $\mathcal{U}$, resulting in

$$\lim_{n\to\infty} P_{Y_2|U}(y_2|u) = \lim_{n\to\infty}\sum_x P_{Y_2|X}(y_2|x)P_{X|U}(x|u) \quad (41)$$

$$= P_{Y_2|X}(y_2|x_0). \quad (42)$$

Moreover, by (34), we also have

$$\lim_{n\to\infty} P_{Y_2}(y_2) = P_{Y_2|X}(y_2|x_0). \quad (43)$$

By the structure of $\hat{P}_X$ the derivative of P_{U,Y_2} according to μ is bounded, so (38), (42), (43) and (37) yield

$$I(U;Y_2) = O(\mu^2). \quad (44)$$

Therefore, a necessary condition for having $I(U;Y_2)=O(\mu)$ is that some of the elements of $\mathcal{U}$ have vanishing probabilities as $n\to\infty$. □

B. Converse for Part 1)

Assume we have a sequence of $(n, 2^{\sqrt{n\delta}L_1}, 2^{\sqrt{n\delta}L_2}, \epsilon_n, \delta)$ codes with $\lim_{n\to\infty}\epsilon_n = 0$. Denote by M_k the random message for user k, $k = 1, 2$. By Fano's inequality

$$\begin{aligned}
\sqrt{n\delta}L_2(1-\epsilon_n) &\overset{(a)}{=} \log \nu_2(1-\epsilon) && (45)\\
&\le I(M_2; Y_2^n|S) && (46)\\
&= \sum_{i=1}^{n} I(M_2; Y_{2,i}|S, Y_2^{i-1}) && (47)\\
&\le \sum_{i=1}^{n} I(M_2 Y_2^{i-1}; Y_{2,i}|S) && (48)\\
&\le \sum_{i=1}^{n} I(M_2 Y_2^{i-1} Y_1^{i-1}; Y_{2,i}|S) && (49)\\
&\overset{(b)}{=} \sum_{i=1}^{n} I(M_2 Y_1^{i-1}; Y_{2,i}|S) && (50)\\
&\le \sum_{i=1}^{n} I(M_2 Y_1^{i-1} S; Y_{2,i}), && (51)
\end{aligned}$$

where (a) is by (11) and in (b) we use the Markov chain $X \multimap Y_1 \multimap Y_2$. Similarly,

$$\begin{aligned}
\sqrt{n\delta}L_1(1-\epsilon_n) &\le I(M_1; Y_1^n|S, M_2) && (52)\\
&= \sum_{i=1}^{n} I(M_1; Y_{1,i}|S, M_2, Y_1^{i-1}) && (53)\\
&\overset{(a)}{=} \sum_{i=1}^{n} I(M_1 X_i; Y_{1,i}|S, M_2, Y_1^{i-1}) && (54)\\
&\overset{(b)}{=} \sum_{i=1}^{n} I(X_i; Y_{1,i}|S, M_2, Y_1^{i-1}) && (55)
\end{aligned}$$

where (a) holds since X^n is a deterministic function of (M_1, M_2) and (b) due to the Markov chain

$$(M_1, M_2, Y_1^{i-1}, Y_2^{i-1}, S) \multimap X_1 \multimap (Y_{1,i} Y_{2,i}). \quad (56)$$

Define

$$U_i = (M_2, Y_1^{i-1}, S), \quad (57)$$

so that after normalization (51, 55) read

$$L_2(1-\epsilon_n) \le \frac{1}{\sqrt{n\delta}} \sum_{i=1}^{n} I(U_i; Y_{2,i}) \quad (58)$$

$$L_1(1-\epsilon_n) \le \frac{1}{\sqrt{n\delta}} \sum_{i=1}^{n} I(X_i; Y_{1,i}|U_i). \quad (59)$$

The bounds (13a, 13b) follow from (58, 59) by the classical time sharing argument. Inequality (13c) is proved exactly like [5, eq. (9), Theorem 1].

C. Direct Part for Part 1)

The proof of the achievability part of Theorem 1, proceeds along the following steps:

Step 1: Obtain a layered (superposition) version of Feinstein's Lemma [13] for the BC, from the results of [14].

Step 2: Show that distributions $P_{U,X}^{(n)}$ with X marginal satisfying (13c) stabilize the information spectrum expressions of Step 1. I.e., the information spectrum random variables converge, as n tends to ∞, to the mutual information functions of the outer bound.

Fix a joint distribution P_{UVTX} such that the Markov chain $(U, V, T) \multimap X \multimap (Y_1, Y_2)$ holds. We use the following notation for the mutual information random variables [15], [16]:

$$i_{X;Y_1}(X; Y_1) = \log \frac{P_{Y_1|X}(Y_1|X)}{P_{Y_1}(Y_1)} \quad (60a)$$

$$i_{V;Y_1|U}(V; Y_1|U) = \log \frac{P_{Y_1|U}(Y_1|U, V)}{P_{Y_1|U}(Y_1|U)} \quad (60b)$$

and similarly for $i_{U;Y_2}(U; Y_2)$ etc.

Step 1. For convenience, we state here the one-shot coding result of [14] for general BCs. Note that the alphabets are of arbitrary size, hence there is no dependence on n. We use the notation of [14], but do not repeat their definitions, for space considerations.

Theorem 3 (Theorem 10 in [14]): Fix a BC $P_{Y_1,Y_2|X}$, a joint distribution P_{UVT}, a map $f : \mathcal{U} \times \mathcal{V} \times \mathcal{T} \to \mathcal{X}$, and integers M_0, $M_{1,0}$, $M_{2,0}$, N, L, $\hat{N}$ and $\hat{L}$. Set

$$M = M_0 M_{1,0} M_{2,0}, \quad (61a)$$

$$M_1 = M_{1,0} N, \quad (61b)$$

$$M_2 = M_{2,0} L, \quad (61c)$$

$$\tilde{N} = \hat{N} N, \quad (61d)$$

$$\tilde{L} = \hat{L} L. \quad (61e)$$

Then, for any $\gamma > 0$ there exists an $(M_0, M_1, M_2, \epsilon_1, \epsilon_2)$ code for the BC with

$$\begin{aligned}
\max\{\epsilon_1, \epsilon_2\} \le\ & 2\exp(-\gamma) + \exp^{-\exp(\gamma)}\\
&+ P\Big[\Big\{i_{UV;Y_1}(UV; Y_1) \le \log M\tilde{N} + \gamma\Big\}\\
&\quad \cup \Big\{i_{UT;Y_2}(UT; Y_2) \le \log M\tilde{L} + \gamma\Big\}\\
&\quad \cup \Big\{i_{V;Y_1|U}(V; Y_1|U) \le \log \tilde{N} + \gamma\Big\}\\
&\quad \cup \Big\{i_{T;Y_2|U}(T; Y_2|U) \le \log \tilde{L} + \gamma\Big\}\\
&\quad \cup \Big\{i_{V;T|U}(V; T|U) > \log \hat{N}\hat{L} - 2\gamma\Big\}\Big]\\
&+ \frac{\min\Big\{\hat{N}, \hat{L}\Big\} - 1}{\hat{N}\hat{L}(\exp(-\gamma) - \exp(-2\gamma))}. \quad (62)
\end{aligned}$$

For our use, we choose the following random variables and parameters in Theorem 3. For V and T:

$$V = X; \quad (63a)$$

$$T = \emptyset; \quad (63b)$$

and for M_0, $M_{1,0}$, L, $\hat{N}$ and $\hat{L}$:

$$M_0 = M_{1,0} = L = \hat{N} = \hat{L} = 1. \quad (63c)$$

With (63c) we obtain

$$M_1 = N; \tag{63d}$$
$$M_2 = M_{2,0}; \tag{63e}$$
$$M = M_2; \tag{63f}$$
$$\tilde{N} = M_1; \tag{63g}$$
$$\tilde{L} = 1. \tag{63h}$$

Substituting (63) in (62) and using the union bound, we conclude that there exists a $(1, M_1, M_2, \epsilon, \epsilon)$ code for the BC with

$$\begin{aligned}\epsilon \leq\ & 2\exp(-\gamma) + \exp^{-\exp(\gamma)} \\ & + P[i_{X;Y_1}(X;Y_1) \leq \log M_1 M_2 + \gamma] \\ & + P[i_{U;Y_2}(U;Y_2) \leq \log M_2 + \gamma] \\ & + P\big[i_{X;Y_1|U}(X;Y_1|U) \leq \log M_1 + \gamma\big] \end{aligned} \tag{64}$$

where P_X is induced by $P_{U,V,T}$ and the mapping f. Note that here X is not a deterministic function of U, due to (63a). Hence $P_{U,X}$ is a general joint distribution.

We now pass to fixed alphabets and transmission length n. In (64), $\tilde{\gamma}$ is arbitrary. Thus choose an arbitrary $\tilde{\gamma} > 0$ and set

$$\sqrt{n}\tilde{\gamma} = \gamma. \tag{65}$$

Using (64) and the notation of Definition 1, we conclude that for every P_{U^n,X^n} on $\mathcal{U}^n \times \mathcal{X}^n$ such that P_{X^n} satisfies (8), and any $\tilde{\gamma} > 0$, there exists an $(n, \nu_1, \nu_2, \epsilon, \delta)$ covert code for the BC with

$$\begin{aligned}\epsilon \leq\ & 2\exp(-\sqrt{n}\tilde{\gamma}) + \exp^{-\exp(\sqrt{n}\tilde{\gamma})} \\ & + P\left[\frac{1}{\sqrt{n}} i_{X^n;Y_1^n}(X^n;Y_1^n) \leq \frac{1}{\sqrt{n}}\log \nu_1\nu_2 + \tilde{\gamma}\right] \\ & + P\left[\frac{1}{\sqrt{n}} i_{U^n;Y_2^n}(U^n;Y_2^n) \leq \frac{1}{\sqrt{n}}\log \nu_2 + \tilde{\gamma}\right] \\ & + P\left[\frac{1}{\sqrt{n}} i_{X^n;Y_1^n|U}(X^n;Y_1^n|U^n) \leq \frac{1}{\sqrt{n}}\log \nu_1 + \tilde{\gamma}\right]. \end{aligned} \tag{66}$$

This completes Step 1.

Step 2. Let $P_{U,X}$ be any joint distribution satisfying the covertness constraint (13c), which implies:

$$\lim_{n\to\infty} P_X(x_0) = 1. \tag{67}$$

Then, let P_{U^n,X^n} be the n-fold product of $P_{U,X}$:

$$P_{U^n,X^n}(u^n, x^n) = P_{U,X}^{\times n}(u^n, x^n) = \prod_{i=1}^{n} P_{U,X}(u_i, x_i). \tag{68}$$

We show next that the random variables in (66) converge in probability to the corresponding single letter information functions, i.e.,

$$\frac{1}{\sqrt{n}} i_{X^n;Y_1^n}(X^n;Y_1^n) \longrightarrow \sqrt{n} I(X;Y_1) \quad \text{in prob.} \tag{69a}$$
$$\frac{1}{\sqrt{n}} i_{U^n;Y_2^n}(U^n;Y_2^n) \longrightarrow \sqrt{n} I(U;Y_2) \quad \text{in prob.} \tag{69b}$$
$$\frac{1}{\sqrt{n}} i_{X^n;Y_1^n|U^n}(X^n;Y_1^n|U^n) \longrightarrow \sqrt{n} I(X;Y_1|U) \quad \text{in prob.} \tag{69c}$$

The proof of (69a) follows exactly the lines of the proof of [5, eq. (16)] using (67) and is omitted. The proof of (69b) follows these lines as well, using (33). We give it here for completeness. First, note that

$$\mathsf{E}\frac{1}{\sqrt{n}} i_{U^n;Y_2^n}(U^n;Y_2^n) = \mathsf{E}\frac{1}{\sqrt{n}} \log \frac{P_{Y_2|U}^{\times n}(Y_2^n|U^n)}{P_{Y_2}^{\times n}(Y_2^n)} \tag{70}$$
$$= \sqrt{n} I(U;Y_2). \tag{71}$$

Hence by Chebyshev's inequality

$$\begin{aligned} & P\left[\left|\frac{1}{\sqrt{n}} i_{U^n;Y_2^n}(U^n;Y_2^n) - \sqrt{n} I(U;Y_2)\right| \geq \alpha\right] \\ & \leq \frac{1}{\alpha^2}\mathsf{var}\left(\frac{1}{\sqrt{n}}\log\frac{P_{Y_2|U}^{\times n}(Y_2^n|U^n)}{P_{Y_2}^{\times n}(Y_2^n)}\right), \end{aligned} \tag{72}$$

so to prove (69b) it is enough to show that

$$\lim_{n\to\infty} \mathsf{var}\left(\frac{1}{\sqrt{n}}\log\frac{P_{Y_2|U}^{\times n}(Y^n|U^n)}{P_Y^{\times n}(Y^n)}\right) = 0. \tag{73}$$

Indeed

$$\mathsf{var}\left(\frac{1}{\sqrt{n}}\log\frac{P_{Y_2|U}^{\times n}(Y^n|U^n)}{P_Y^{\times n}(Y^n)}\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mathsf{var}\left(\log\frac{P_{Y_2|U}^{\times n}(Y_{2,i}|U_i)}{P_{Y_2}(Y_{2,i})}\right) \tag{74}$$
$$= \mathsf{var}\left(\log\frac{P_{Y_2|U}(Y_2|U)}{P_{Y_2}(Y_2)}\right) \tag{75}$$
$$\leq \mathsf{E}\left[\left(\log\frac{P_{Y_2|U}(Y_2|U)}{P_{Y_2}(Y_2)}\right)^2\right] \tag{76}$$
$$= \sum_{u\in A} P_U(u) \sum_{y_2} P_{Y_2|U}(y_2|u)\left(\log\frac{P_{Y_2|U}(y_2|u)}{P_{Y_2}(y_2)}\right)^2$$
$$+ \sum_{u\in B} P_U(u) \sum_{y_2} P_{Y_2|U}(y_2|u)\left(\log\frac{P_{Y_2|U}(y_2|u)}{P_{Y_2}(y_2)}\right)^2, \tag{77}$$

where we again partition $\mathcal{U}$ into the subsets A and B so that for $u \in B$ we have $\lim_{n\to\infty} P_U(u) = 0$ while for $u \in A$ we have $\lim_{n\to\infty} P_U(u) > 0$.

We now invoke again the arguments in the proof of Proposition 1. The probability of any $u' \in A$ is bounded from below, hence (40), (42) and (43) hold. Thus

$$\lim_{n\to\infty}\left(\log\frac{P_{Y_2|U}(y_2|u)}{P_{Y_2}(y_2)}\right)^2 = (\log 1)^2 = 0 \quad \forall u \in A, \tag{78}$$

and the first sum in the r.h.s of (77) vanishes as $n \to \infty$. Regarding the second sum in the r.h.s of (77), we use (43) to

write

$$\lim_{n\to\infty}\sum_{u\in B}P_U(u)\sum_{y_2}P_{Y_2|U}(y_2|u)\left(\log\frac{P_{Y_2|U}(y_2|u)}{P_{Y_2}(y_2)}\right)^2$$
$$=\lim_{n\to\infty}\sum_{u\in B}P_U(u)\sum_{y_2}P_{Y_2|U}(y_2|u)\left(\log\frac{P_{Y_2|U}(y_2|u)}{P_{Y_2|X}(y_2|x_0)}\right)^2. \tag{79}$$

Since every u is mapped randomly to the elements of $\mathcal{X}$, we have by Part c of Conditions 1

$$\frac{P_{Y_2|U}(y_2|u)}{P_{Y_2|X}(y_2|x_0)}\le\max_{y_2,x}\frac{P_{Y_2|X}(y_2|x)}{P_{Y_2|X}(y_2|x_0)}\le M \tag{80}$$

for some finite, fixed M. Therefore by (79), (80)

$$\lim_{n\to\infty}\sum_{u\in B}P_U(u)\sum_{y_2}P_{Y_2|U}(y_2|u)\left(\log\frac{P_{Y_2|U}(y_2|u)}{P_{Y_2}(y_2)}\right)^2$$
$$\le\lim_{n\to\infty}P_U(B)\log M=0. \tag{81}$$

Using (78) and (81) in (77) yields (73). This establishes (69b).

We proceed to prove (69c). First, observe that following the lines of the proof of (69b), we also have

$$\frac{1}{\sqrt{n}}i_{U^n;Y_1^n}(U^n;Y_1^n)\longrightarrow\sqrt{n}I(U;Y_1)\quad\text{in prob.}, \tag{82}$$

and, by previous results [5]

$$\frac{1}{\sqrt{n}}i_{X^n;Y_1^n}(X^n;Y_1^n)\longrightarrow\sqrt{n}I(X;Y_1)\quad\text{in prob.}. \tag{83}$$

Moreover

$$\frac{1}{\sqrt{n}}i_{X^n;Y_1^n|U^n}(X^n;Y_1^n|U^n)$$
$$=\frac{1}{\sqrt{n}}i_{X^n;Y_1^n}(X^n;Y_1^n)-\frac{1}{\sqrt{n}}i_{U^n;Y_1^n}(U^n;Y_1^n) \tag{84}$$

Then by properties of convergence in probability

$$\frac{1}{\sqrt{n}}i_{X^n;Y_1^n|U^n}(X^n;Y_1^n|U^n)\longrightarrow\sqrt{n}I(X;Y_1)-\sqrt{n}I(U;Y_1)$$
$$=I(X;Y_1|U), \tag{85}$$

proving (69c).

We have shown that for any P_{UX} with marginal P_X satisfying (67), the random variables in (66) converge in probability to the corresponding mutual information functions. This implies that for any $\tilde{\gamma}>0$ and any pairs ν_1,ν_2 such that

$$\frac{\log\nu_1}{\sqrt{n\delta}}\le\sqrt{n/\delta}I(X;Y_1|U)-\tilde{\gamma}/\delta \tag{86a}$$
$$\frac{\log\nu_2}{\sqrt{n\delta}}\le\sqrt{n/\delta}I(U;Y_2)-\tilde{\gamma}/\delta \tag{86b}$$
$$\frac{\log\nu_1\nu_2}{\sqrt{n\delta}}\le\sqrt{n/\delta}I(X;Y_1)-\tilde{\gamma}/\delta. \tag{86c}$$

for some sequence $P_{U,X}$ satisfying the divergence constraint (13c), there exists an $(n,\nu_1,\nu_2,\epsilon,\delta)$ covert code for the BC. The channel is degraded thus (86a), (86b) dominate (86c). Since $\tilde{\gamma}$ is arbitrary, this establishes the direct part. □

Appendix B
Proof of the Computable Region

We employ a Taylor expansions of $I(U;Y_2)$, $I(X;Y_1|U)$ and $D(P_Z||Q_0)$ near $\mu_1=0$, $\mu_2=0$. By (28), (29) and (30) we have for Y_k, $k=1,2$ and Z:

$$P_{X,Y_k}(x,y_k)=P_{Y_k|X}(y_k|x_0)\bar{\mu}_1\bar{\mu}_2\mathbb{1}_{x_0}(x)$$
$$+P_{Y_k|X}(y_k|x)\Big[\bar{\mu}_1\mu_2\tilde{P}_X^A(x)+\mu_1P_X^B(x)\Big], \tag{87a}$$

$$P_{U,Y_k}(u,y_k)=\begin{cases}\bar{\mu}_1\Big[\bar{\mu}_2P_{Y_k|X}(y_k|x_0)\\ \quad+\mu_2\tilde{P}^A_{Y_k|U}(y_k|u)\Big]P_U^A(u) & \text{for } u\in A,\\ \mu_1P^B_{Y_k|U}(y_k|u)P_U^B(u) & \text{for } u\in B.\end{cases} \tag{87b}$$

$$P_{Y_k}(y_k)=\bar{\mu}_1\bar{\mu}_2P_{Y_k|X}(y_k|x_0)$$
$$+\bar{\mu}_1\mu_2\tilde{P}^A_{Y_k}(y_k)+\mu_1P^B_{Y_k}(y_k) \tag{87c}$$
$$P_Z(z)=\bar{\mu}_1\bar{\mu}_2P_{Z|X}(z|x_0)$$
$$+\sum_x P_{Z|X}(z|x)\Big[\bar{\mu}_1\mu_2\tilde{P}_X^A(x)+\mu_1P_X^B(x)\Big]$$
$$=\bar{\mu}_1\bar{\mu}_2P_{Z|X}(z|x_0)+\bar{\mu}_1\mu_2\tilde{P}_Z^A(z)+\mu_1P_Z^B(z) \tag{87d}$$

where in (87c) we used the definitions:

$$\tilde{P}^A_{Y_k}(y_k)=\sum_{u\in A}\tilde{P}^A_{Y_k|U}(y_k|u)P_U^A(u) \tag{87e}$$
$$P^B_{Y_k}(y_k)=\sum_{u\in B}P^B_{Y_k|U}(y_k|u)P_U^B(u) \tag{87f}$$

Note that

$$P_{X,Y_k}(x,y_k)\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=P_{Y_k,X}(y_k,x_0)\mathbb{1}_{x_0}(x) \tag{88a}$$

$$P_{U,Y_k}(u,y_k)\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=\begin{cases}P_{Y_k|X}(y_k|x_0)P_U^A(u) & \text{for } u\in A,\\ 0 & \text{for } u\in B.\end{cases} \tag{88b}$$

$$P_{Y_k}(y_k)\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=P_{Y_k|X}(y_k|x_0) \tag{88c}$$

$$P_U(u)\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=\begin{cases}P_U^A(u) & \text{for } u\in A,\\ 0 & \text{for } u\in B.\end{cases} \tag{88d}$$

$$P_Z(z)\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=Q_0(z) \tag{88e}$$

We obtain the following derivatives of $I(U;Y_k)$:

$$\frac{\partial}{\partial\mu_1}I(U;Y_k)\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=\sum_{u\in B}P_U^B(u)D\Big(P^B_{Y_k|U}(\cdot|u)||P_{Y_k|X}(\cdot|x_0)\Big) \tag{89a}$$

$$\frac{\partial}{\partial\mu_2}I(U;Y_k)\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=0 \tag{89b}$$

$$\frac{\partial}{\partial\mu_1} I(X;Y_k)\Big|_{\substack{\mu_1=0\\ \mu_2=0}} = \sum_x P_X^B(x) D\big(P_{Y_k|X}(\cdot|x)||P_{Y_k|X}(\cdot|x_0)\big) \tag{89c}$$

$$\frac{\partial}{\partial\mu_2} I(X;Y_k)\Big|_{\substack{\mu_1=0\\ \mu_2=0}} = \sum_x \tilde{P}_X^A(x) D\big(P_{Y_k|X}(\cdot|x)||P_{Y_k|X}(\cdot|x_0)\big). \tag{89d}$$

Similarly, evaluating the derivatives of $D(P_Z||Q_0)$ w.r.t. μ_1 and μ_2, we obtain

$$\frac{\partial}{\partial\mu_1} D(P_Z||Q_0)\Big|_{\substack{\mu_1=0\\ \mu_2=0}} = \frac{\partial}{\partial\mu_2} D(P_Z||Q_0)\Big|_{\substack{\mu_1=0\\ \mu_2=0}} = 0 \tag{90}$$

and the Hessian

$$\frac{\partial^2}{\partial\mu_1^2} D(P_Z||Q_0)\Big|_{\substack{\mu_1=0\\ \mu_2=0}} = \chi_2(P_Z^B||Q_0) \tag{91a}$$

$$\frac{\partial^2}{\partial\mu_2^2} D(P_Z||Q_0)\Big|_{\substack{\mu_1=0\\ \mu_2=0}} = \chi_2(\tilde{P}_Z^A||Q_0) \tag{91b}$$

$$\frac{\partial^2}{\partial\mu_1\partial\mu_2} D(P_Z||Q_0)\Big|_{\substack{\mu_1=0\\ \mu_2=0}} = \chi_2(\tilde{P}_Z^A, P_Z^B||Q_0), \tag{91c}$$

where

$$\chi_2(\tilde{P}_Z^A, P_Z^B||Q_0) := \sum_{z\in\mathcal{Z}} \frac{(\tilde{P}_Z^A(z)-Q_0(z))(P_Z^B(z)-Q_0(z))}{Q_0(z)}. \tag{92}$$

The derivations of (89), (90) and (91) are given in Appendix E-B. Thus we can write

$$\begin{aligned} I(X;Y_1|U) &= I(X;Y_1) - I(U;Y_1) \\ &= \mu_1 \sum_x P_X^B(x) D\Big(P_{Y_1|X}^B(\cdot|x)||P_{Y_1|X}(\cdot|x_0)\Big) \\ &\quad +\mu_2 \sum_x \tilde{P}_X^A(x) D\big(P_{Y_1|X}(\cdot|x)||P_{Y_1|X}(\cdot|x_0)\big) \\ &\quad -\mu_1 \sum_{u\in B} P_U^B(u) D\Big(P_{Y_1|U}^B(\cdot|u)||P_{Y_1|X}(\cdot|x_0)\Big) \\ &\quad +o(\mu_1,\mu_2) \end{aligned} \tag{93a}$$

$$\begin{aligned} I(U;Y_2) &= \mu_1 \sum_{u\in B} P_U^B(u) D\Big(P_{Y_2|U}^B(\cdot|u)||P_{Y_2|X}(\cdot|x_0)\Big) \\ &\quad +o(\mu_1,\mu_2) \end{aligned} \tag{93b}$$

where the divergence constraint (13c) entails

$$\begin{aligned} \frac{1}{2}\Big[\mu_1^2\chi_2(P_Z^B||Q_0) + \mu_1\mu_2\chi_2(\tilde{P}_Z^A, P_Z^B||Q_0) \\ +\mu_2^2\chi_2(\tilde{P}_Z^A||Q_0)\Big] + o(\mu_1^2,\mu_2^2) \le \frac{\delta}{n} \end{aligned} \tag{94}$$

Using the normalization (31) in (93) and (94) results in (32). □

Appendix C
Inclusion of Time-Sharing Region $\mathcal{L}^{(\mathrm{TS})} \subseteq \mathcal{L}^{(I)}$

Let P_X^{1*} bet the L_1^*-achieving pmf and P_X^{2*} the L_2^*-achieving pmf. (Both are pmfs over $\mathcal{X}\backslash\{x_0\}$). Let further P_Z^{1*} and P_Z^{2*} be the corresponding output distributions at the warden.

For any $\mu \in [0,1]$, specializing (17a) and (17b) to the choices $\tilde{P}_X^A = P_X^{1*}$ and $P_X^B = P_X^{2*}$ (so each $P_X^{\ell*}$ is only a pmf over $\mathcal{X}\backslash\{x_0\}$), and choosing a deterministic mapping for $P_{U|X}^B$ results in the rate-pair

$$L_1 \le (1-\nu)\frac{\sqrt{2}\sum_x P_X^{1*}(x) D\big(P_{Y_1|X}(\cdot|x)||P_{Y_1|X}(\cdot|x_0)\big)}{\sqrt{\chi_2\big((1-\nu)P_Z^{1*}+\nu P_Z^{2*}\big\|Q_0\big)}} \tag{95}$$

$$= \alpha_1(\nu) L_1^* \tag{96}$$

$$L_2 \le \nu\frac{\sqrt{2}\sum_x P_X^{2*}(x) D\big(P_{Y_2|X}(\cdot|x)||P_{Y_2|X}(\cdot|x_0)\big)}{\sqrt{\chi_2\big((1-\nu)P_Z^{1*}+\nu P_Z^{2*}\big\|Q_0\big)}} \tag{97}$$

$$= \alpha_2(\nu) L_2^* \tag{98}$$

for

$$\alpha_1(\nu) := \frac{(1-\nu)\sqrt{\chi_2\Big(\tilde{P}_Z^{1*}\Big\|Q_0\Big)}}{\sqrt{\chi_2\Big((1-\nu)\tilde{P}_Z^{1*}+\nu P_Z^{2*}\Big\|Q_0\Big)}} \tag{99}$$

$$\alpha_2(\nu) := \frac{\nu\sqrt{\chi_2\Big(\tilde{P}_Z^{2*}\Big\|Q_0\Big)}}{\sqrt{\chi_2\Big((1-\nu)\tilde{P}_Z^{1*}+\nu P_Z^{2*}\Big\|Q_0\Big)}}. \tag{100}$$

Varying ν from 0 to 1 varies α_1 from 1 to 0 and α_2 from 0 to 1. To show that this region includes the time-sharing region it suffices to show that for any $\nu \in [0,1]$:

$$\alpha_1(\nu) + \alpha_2(\nu) \ge 1, \tag{101}$$

which holds because $\alpha_1(\nu), \alpha_2(\nu) > 0$ and because by the convexity of the square-root of the χ_2-distance we have $(\alpha_1(\nu)+\alpha_2(\nu))^2 \ge 1$, as proved by the sequence of Inequalities (102)–(104) on top of the next page, where the last inequality holds because by Cauchy-Schwarz-Inequality:

$$\begin{aligned} &\sqrt{\sum_z \left(\frac{P_Z^{1*}(z)-Q_0(z)}{\sqrt{Q_0(z)}}\right)^2}\sqrt{\sum_z \left(\frac{P_Z^{2*}(z)-Q_0(z)}{\sqrt{Q_0(z)}}\right)^2} \\ &\quad \ge \sum_z \frac{(P_Z^{1*}(z)-Q_0(z))}{\sqrt{Q_0(z)}}\frac{(P_Z^{2*}(z)-Q_0(z))}{\sqrt{Q_0(z)}}. \end{aligned} \tag{105}$$

Appendix D
Optimality of Time-Sharing

For any set of achievable $(L_1, L_2) \in \tilde{\mathcal{L}}^{(I)}$, the set of inequalities (106)–(109) on top of the next page holds for some $B \subset \mathcal{U}$, P_U^A, P_U^B, $\tilde{P}_{X|U}$, $P_{X|U}$ and $\nu \in [0,1]$. where in the first equality we applied (21), and in the second inequality we used the fact that the rate in (108) corresponds to the rate to User 2 achieved by a time-sharing scheme employing pmf $\tilde{P}_X^A$ during $(1-\nu)$-fraction of the time and pmf P_X^B during the remaining time, which cannot exceed L_2^*.

This establishes optimality of time-sharing as proved in [11].

$$(\alpha_1(\nu)+\alpha_2(\nu))^2 = \frac{(1-\nu)^2\chi_2\Big(\tilde{P}_Z^{1*}\|Q_0\Big)+\nu^2\chi_2\Big(\tilde{P}_Z^{2*}\|Q_0\Big)+2\nu(1-\nu)\sqrt{\chi_2\Big(\tilde{P}_Z^{1*}\|Q_0\Big)}\sqrt{\chi_2\Big(\tilde{P}_Z^{2*}\|Q_0\Big)}}{\chi_2\Big((1-\nu)\tilde{P}_Z^{1*}+\nu P_Z^{2*}\|Q_0\Big)} \tag{102}$$

$$= \frac{(1-\nu)^2\chi_2\Big(\tilde{P}_Z^{1*}\|Q_0\Big)+\nu^2\chi_2\Big(\tilde{P}_Z^{2*}\|Q_0\Big)+2\nu(1-\nu)\sqrt{\chi_2\Big(\tilde{P}_Z^{1*}\|Q_0\Big)}\sqrt{\chi_2\Big(\tilde{P}_Z^{2*}\|Q_0\Big)}}{(1-\nu)^2\chi_2\Big(\tilde{P}_Z^{1*}\|Q_0\Big)+\nu^2\chi_2\Big(\tilde{P}_Z^{2*}\|Q_0\Big)+2\nu(1-\nu)\sum_z \frac{(P_Z^{1*}(z)-Q_0(z))(P_Z^{2*}(z)-Q_0(z))}{Q_0(z)}} \tag{103}$$

$$\geq 1, \tag{104}$$

$$\frac{L_1}{L_1^*}+\frac{L_2}{L_2^*} \leq \sqrt{\frac{2}{\chi_2(\nu)}}\left[\nu\sum_{u\in B}P_U^B(u)\frac{I(Y_1;X^B|U=u)}{L_1^*}+(1-\nu)\sum_x \tilde{P}_X^A(x)\frac{D\big(P_{Y_1|X}(\cdot|x)||P_{Y_1|X}(\cdot|x_0)\big)}{L_1^*}\right]$$
$$+\sqrt{\frac{2}{\chi_2(\nu)}}\left[\nu\sum_x P_X^B(x)\frac{D\big(P_{Y_2|X}(\cdot|x)||P_{Y_2|X}(\cdot|x_0)\big)}{L_2^*}-\nu\sum_{u\in B}P_U^B(u)\frac{I(Y_2;X^B|U=u)}{L_2^*}\right] \tag{106}$$

$$= \sqrt{\frac{2}{\chi_2(\nu)}}\left[\nu\sum_{u\in B}P_U^B(u)\frac{I(Y_2;X^B|U=u)}{L_2^*}+(1-\nu)\sum_x \tilde{P}_X^A(x)\frac{D\big(P_{Y_2|X}(\cdot|x)||P_{Y_2|X}(\cdot|x_0)\big)}{L_2^*}\right]$$
$$+\sqrt{\frac{2}{\chi_2(\nu)}}\left[\nu\sum_x P_X^B(x)\frac{D\big(P_{Y_2|X}(\cdot|x)||P_{Y_2|X}(\cdot|x_0)\big)}{L_2^*}-\nu\sum_{u\in B}P_U^B(u)\frac{I(Y_2;X^B|U=u)}{L_2^*}\right] \tag{107}$$

$$= (L_2^*)^{-1}\sqrt{\frac{2}{\chi_2(\nu)}}\Big(\nu\sum_x P_X^B(x)D\big(P_{Y_2|X}(\cdot|x)||P_{Y_2|X}(\cdot|x_0)\big)$$
$$+(1-\nu)\sum_x \tilde{P}_X^A(x)D\big(P_{Y_2|X}(\cdot|x)||P_{Y_2|X}(\cdot|x_0)\big)\Big) \tag{108}$$

$$\leq 1, \tag{109}$$

Appendix E
Derivation of Mutual Informations and Divergence Derivatives

A. Proof of (38)

We have to show that

$$\sum_{u,y_2}P_{U,Y_2}(u,y_2)\frac{\partial}{\partial\mu}\log\frac{P_{U,Y_2}(u,y_2)}{P_U(u)P_{Y_2}(y_2)}=0. \tag{110}$$

Let $P_{U|X}$ be a general conditional distribution. By (34)

$$P_{U,X}(u,x) = (1-\mu)P_{U|X}(u|x_0)\mathbb{1}_{x_0}(x)+\mu P_{U|X}(u|x)\tilde{P}_X(x) \tag{111a}$$

$$P_{U,Y_2}(u,y_2) = (1-\mu)P_{U|X}(u|x_0)P_{Y_2|X}(y_2|x_0)+\mu\sum_x P_{U|X}(u|x)\tilde{P}_X(x)P_{Y_2|X}(y_2|x) \tag{111b}$$

$$P_U(u) = (1-\mu)P_{U|X}(u|x_0)+\mu\sum_x P_{U|X}(u|x)\tilde{P}_X(x) \tag{111c}$$

$$P_{Y_2}(y_2) = (1-\mu)P_{Y_2|X}(y|x_0)+\mu\sum_x \tilde{P}_X(x)P_{Y_2|X}(y|x) \tag{111d}$$

Write

$$\sum_{u,y_2}P_{U,Y_2}(u,y_2)\frac{\partial}{\partial\mu}\log\frac{P_{U,Y_2}(u,y_2)}{P_U(u)P_{Y_2}(y_2)}$$
$$=\sum_{u,y_2}P_{U,Y_2}(u,y_2)\left[\frac{\frac{\partial}{\partial\mu}P_{U,Y_2}(u,y_2)}{P_{U,Y_2}(u,y_2)}-\frac{\frac{\partial}{\partial\mu}P_U(u)}{P_U(u)}-\frac{\frac{\partial}{\partial\mu}P_{Y_2}(y_2)}{P_{Y_2}(y_2)}\right] \tag{112}$$

$$=\frac{\partial}{\partial\mu}\sum_{u,y_2}P_{U,Y_2}(u,y_2)-\sum_{u,y_2}P_{Y_2|U}(y_2|u)\frac{\partial}{\partial\mu}P_U(u)-\sum_{u,y_2}P_{U|Y_2}(u|y_2)\frac{\partial}{\partial\mu}P_{Y_2}(y_2) \tag{113}$$

Using (111c) and (111d) we obtain

$$\sum_{u,y_2}P_{Y_2|U}(y_2|u)\frac{\partial}{\partial\mu}P_U(u)$$
$$=\sum_{u,y_2}P_{Y_2|U}(y_2|u)\left[-P_{U|X}(u|x_0)+\sum_x P_{U|X}(u|x)\tilde{P}_X(x)\right]$$

$$= 1 - 1 = 0 \tag{114a}$$

and

$$\sum_{u,y_2} P_{U|Y_2}(u|y_2)\frac{\partial}{\partial\mu}P_{Y_2}(y_2)$$
$$= \sum_{u,y_2} P_{U|Y_2}(u|y_2)\Big[- P_{Y_2|X}(y_2|x_0)$$
$$+ \sum_x \tilde{P}_X(x)P_{Y_2|X}(y_2|x)\Big] = 0. \tag{114b}$$

Substitution of (114) in (113) yields the desired result. □

B. Proofs of (89)–(91)

1) Proof of (89): We first present general derivative formulas for the mutual information functions.

$$\frac{\partial}{\partial\mu_j} I(U;Y_k)$$
$$= \sum_{u,y_k}\left[\frac{\partial}{\partial\mu_j}P_{U,Y_k}(u,y_k)\right]\log\frac{P_{U,Y_k}(u,y_k)}{P_U(u)P_{Y_k}(y_k)}$$
$$+ \sum_{u,y_k} P_{U,Y_k}(u,y_k)\Bigg[\frac{\frac{\partial}{\partial\mu_j}P_{U,Y_k}(u,y_k)}{P_{U,Y_k}(u,y_k)}$$
$$- \frac{\frac{\partial}{\partial\mu_j}P_U(u)}{P_U(u)} - \frac{\frac{\partial}{\partial\mu_j}P_{Y_k}(y_k)}{P_{Y_k}(y_k)}\Bigg] \tag{115}$$
$$= A_{1,k,j} + A_{2,k,j}, \quad k = 1,2, \quad j = 1,2. \tag{116}$$

with the obvious definitions for $A_{1,j}$ and $A_{2,j}$. Evaluating these terms:

$$A_{2,k,j}$$
$$= -\sum_{u,y_k} P_{U,Y_k}(u,y_k)\Bigg[\frac{\frac{\partial}{\partial\mu_j}P_{U,Y_k}(u,y_k)}{P_{U,Y_k}(u,y_k)}$$
$$- \frac{\frac{\partial}{\partial\mu_j}P_U(u)}{P_U(u)} - \frac{\frac{\partial}{\partial\mu_j}P_{Y_k}(y_k)}{P_{Y_k}(y_k)}\Bigg] \tag{117}$$
$$= -\sum_{u,y_k}\Bigg[P_{Y_k|U}(y_k|u)\frac{\partial}{\partial\mu_j}P_U(u)$$
$$+ P_{U|Y_k}(u|y_k)\frac{\partial}{\partial\mu_j}P_{Y_k}(y_k)\Bigg], \quad k = 1,2, \quad j = 1,2. \tag{118}$$

Evaluating the derivatives in the r.h.s. of (118), we have

$$\frac{\partial}{\partial\mu_1}P_U(u) = -P_U^A(u) + P_U^B(u) \tag{119}$$
$$\frac{\partial}{\partial\mu_1}P_{Y_k}(y_k) = -\bar{\mu}_2 P_{Y_k|X}(y_k|x_0) - \mu_2\tilde{P}_{Y_k}^A(y_k) + P_{Y_k}^B(y_k) \tag{120}$$
$$\frac{\partial}{\partial\mu_2}P_U(u) = 0 \tag{121}$$
$$\frac{\partial}{\partial\mu_2}P_{Y_k}(y_k) = \bar{\mu}_1\Big[\tilde{P}_{Y_k}^A(y_k) - P_{Y_k|X}(y_k|x_0)\Big] \tag{122}$$

resulting in

$$A_{2,k,j} = 0. \tag{123}$$

For $A_{1,k,j}$, we first evaluate the derivative of the joint distribution:

$$\frac{\partial}{\partial\mu_1}P_{U,Y_k}(u,y_k)$$
$$= \begin{cases} -\Big[\bar{\mu}_2 P_{Y_k|X}(y_k|x_0) \\ \qquad +\mu_2\tilde{P}_{Y_k|U}^A(y_k|u)\Big]P_U^A(u) & \text{for } u\in A, \\ P_{Y_k|U}^B(y_k|u)P_U^B(u) & \text{for } u\in B. \end{cases}$$

$$\frac{\partial}{\partial\mu_2}P_{U,Y_k}(u,y_k)$$
$$= \begin{cases} \bar{\mu}_1\Big[-P_{Y_k|X}(y_k|x_0) + \tilde{P}_{Y_k|U}^A(y_k|u)\Big]P_U^A(u) & \text{for } u\in A, \\ 0 & \text{for } u\in B. \end{cases} \tag{124}$$

Therefore

$$A_{1,k,1}\Big|_{\substack{\mu_1=0\\ \mu_2=0}}$$
$$= \sum_{y_k,u\in A}\Big[-\bar{\mu}_2 P_{Y_k|X}(y_k|x_0) - \mu_2\tilde{P}_{Y_k|U}^A(y_k|u)\Big]$$
$$\cdot P_U^A(u)\log\frac{P_{U,Y_k}(u,y_k)}{P_U(u)P_{Y_k}(y_k)}\Big|_{\substack{\mu_1=0\\ \mu_2=0}}$$
$$+ \sum_{y_k,u\in B} P_{Y_k|U}^B(y_k|u)P_U^B(u)\log\frac{P_{Y_k|U}^B(y_k|u)}{P_{Y_k|X}(y_k|x_0)} \tag{125}$$
$$= \sum_{u\in B} P_U^B(u)D\Big(P_{Y_k|U}^B(\cdot|u)||P_{Y_k|X}(\cdot|x_0)\Big) \tag{126}$$

where in the last equality we used (88). Subsituting (126) and (123) in (116) proves (89a).

For (89b), we only have to evaluate $A_{1,k,2}$. By (118) we have

$$A_{1,k,2}\Big|_{\substack{\mu_1=0\\ \mu_2=0}}$$
$$= \sum_{u,y_k}\left[\frac{\partial}{\partial\mu_2}P_{U,Y_k}(u,y_k)\right]\log\frac{P_{U,Y_k}(u,y_k)}{P_U(u)P_{Y_k}(y_k)}\Big|_{\substack{\mu_1=0\\ \mu_2=0}} \tag{127}$$
$$= 0 \tag{128}$$

where we used (88) and the fact that the r.h.s of (124) is bounded. Subsituting (128) and (123) in (116) proves (89b).

The proof of (89c) and (89d) proceed along the same lines as that of (89a). Parallel to (116), we write

$$\frac{\partial}{\partial\mu_j} I(X;Y_k)$$
$$= \sum_{x,y_k}\left[\frac{\partial}{\partial\mu_j}P_{X,Y_k}(x,y_k)\right]\log\frac{P_{X,Y_k}(x,y_k)}{P_X(x)P_{Y_k}(y_k)}$$
$$+ \sum_{x,y_k} P_{X,Y_k}(x,y_k)\Bigg[\frac{\frac{\partial}{\partial\mu_j}P_{X,Y_k}(x,y_k)}{P_{X,Y_k}(x,y_k)}$$

$$-\frac{\frac{\partial}{\partial\mu_j}P_X(x)}{P_X(x)}-\frac{\frac{\partial}{\partial\mu_j}P_{Y_k}(y_k)}{P_{Y_k}(y_k)}\Big]$$
$$=B_{1,k,j}+B_{2,k,j},\qquad k=1,2,\quad j=1,2. \tag{129}$$

and

$$B_{2,k,j}=-\sum_{x,y_k}\Big[P_{Y_k|X}(y_k|x)\frac{\partial}{\partial\mu_j}P_X(x)+P_{X|Y_k}(x|y_k)\frac{\partial}{\partial\mu_j}P_{Y_k}(y_k)\Big],\qquad k=1,2,\quad j=1,2. \tag{130}$$

Evaluation of the derivatives in the r.h.s. of (130) gives

$$\frac{\partial}{\partial\mu_1}P_X(x)=-\Big[\bar\mu_2\mathbb{1}_{x_0}(x)+\mu_2\tilde P_X^A(x)\Big]+P_X^B(x) \tag{131}$$
$$\frac{\partial}{\partial\mu_2}P_X(x)=\bar\mu_1\Big[-\mathbb{1}_{x_0}(x)+\tilde P_X^A(x)\Big]. \tag{132}$$

Using (131) and (120) in (130) yields

$$B_{2,k,1}=\sum_{x,y_k}\Big[P_{Y_k|X}(y_k|x)\Big(-\bar\mu_2\mathbb{1}_{x_0}(x)-\mu_2\tilde P_X^A(x)+P_X^B(x)\Big)+P_{X|Y_k}(x|y_k)\Big(-\bar\mu_2P_{Y_k|X}(y_k|x_0)-\mu_2\tilde P_{Y_k}^A(y_k)+P_{Y_k}^B(y_k)\Big)\Big]=0. \tag{133}$$

Similarly, using (132) and (122) in (130) yields

$$B_{2,k,2}=\sum_{x,y_k}\Big[P_{Y_k|X}(y_k|x)\Big(-\bar\mu_1\mathbb{1}_{x_0}(x)+\bar\mu_1\tilde P_X^A(x)\Big)+P_{X|Y_k}(x|y_k)\Big(\bar\mu_1\tilde P_{Y_k}^A(y_k)-\bar\mu_1P_{Y_k|X}(y_k|x_0)\Big)\Big]=0. \tag{134}$$

Next we evaluate $B_{1,k,j}$. By (87a)

$$\frac{\partial}{\partial\mu_1}P_{X,Y_k}(x,y_k)=\Big[-\bar\mu_2\mathbb{1}_{x_0}(x)-\mu_2\tilde P_X^A(x)+P_X^B(x)\Big]\cdot P_{Y_k|X}(y_k|x) \tag{135}$$
$$\frac{\partial}{\partial\mu_2}P_{X,Y_k}(x,y_k)=\bar\mu_1\Big[-\mathbb{1}_{x_0}(x)+\tilde P_X^A(x)\Big]P_{Y_k|X}(y_k|x), \tag{136}$$

hence

$$B_{1,k,1}=\sum_{x,y_k}\Big[-\bar\mu_2\mathbb{1}_{x_0}(x)-\mu_2\tilde P_X^A(x)+P_X^B(x)\Big]\cdot P_{Y_k|X}(y_k|x)\log\frac{P_{Y_k|X}(y_k|x)}{P_{Y_k}(y_k)} \tag{137}$$
$$\tag{138}$$

Using (88c) we arrive at

$$B_{1,k,1}\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=\sum_x P_X^B(x)D\big(P_{Y_k|X}(\cdot|x)||P_{Y_k|X}(\cdot|x_0)\big). \tag{139}$$

Now (89c) follows from (129), (133) and (139). Similarly, by (136)

$$B_{1,k,2}=\bar\mu_1\sum_{x,y_k}\Big[-\mathbb{1}_{x_0}(x)+\tilde P_X^A(x)\Big]\cdot P_{Y_k|X}(y_k|x)\log\frac{P_{Y_k|X}(y_k|x)}{P_{Y_k}(y_k)} \tag{140}$$

which, using again (88c), implies

$$B_{1,k,2}\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=\sum_x\tilde P_X^A(x)D\big(P_{Y_k|X}(\cdot|x)||P_{Y_k|X}(\cdot|x_0)\big). \tag{141}$$

The proof of (89d) follows from (129), (134) and (141).

2) Proof of (90):

$$\frac{\partial}{\partial\mu_j}D(P_Z||Q_0)=\sum_z\Big(\frac{\partial}{\partial\mu_j}P_Z(z)\Big)\log\frac{P_Z(z)}{Q_0(z)} \tag{142}$$

where, by (87d)

$$\frac{\partial}{\partial\mu_1}P_Z(z)=-\bar\mu_2Q_0(z)-\mu_2\tilde P_Z^A(z)+P_Z^B(z) \tag{143}$$
$$\frac{\partial}{\partial\mu_2}P_Z(z)=-\bar\mu_1Q_0(z)+\bar\mu_1\tilde P_Z^A(z). \tag{144}$$

Note that the derivatives (143), (144) are bounded. Therefore (90) follows from (142) and (88e).

3) Proof of (91):

$$\frac{\partial^2}{\partial\mu_j^2}D(P_Z||Q_0)=\sum_z\Big(\frac{\partial^2}{\partial\mu_j^2}P_Z(z)\Big)\log\frac{P_Z(z)}{Q_0(z)}+\sum_z\Big(\frac{\partial}{\partial\mu_j}P_Z(z)\Big)^2\frac{1}{P_Z(z)} \tag{145}$$

Since the second derivatives of $P_Z(z)$ according to μ_j are bounded, (88e) implies that the first sum in the r.h.s. of (145) is 0. Using (143), (144), in the second sum of (145) we get

$$\frac{\partial^2}{\partial\mu_1^2}D(P_Z||Q_0)\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=\sum_z\frac{(P_Z^B(z)-Q_0(z))^2}{Q_0(z)} \tag{146}$$
$$=\chi_2(P_Z^B||Q_0) \tag{147}$$
$$\frac{\partial^2}{\partial\mu_2^2}D(P_Z||Q_0)\Big|_{\substack{\mu_1=0\\ \mu_2=0}}=\sum_z\frac{(\tilde P_Z^A(z)-Q_0(z))^2}{Q_0(z)} \tag{148}$$
$$=\chi_2(\tilde P_Z^A||Q_0) \tag{149}$$

proving (91a) and (91b). For (91c)

$$\frac{\partial^2}{\partial\mu_1\partial\mu_2}D(P_Z||Q_0)=\sum_z\Big(\frac{\partial^2}{\partial\mu_1\partial\mu_2}P_Z(z)\Big)\log\frac{P_Z(z)}{Q_0(z)}+\sum_z\Big(\frac{\partial}{\partial\mu_1}P_Z(z)\Big)\frac{1}{P_Z(z)}\frac{\partial}{\partial\mu_2}P_Z(z) \tag{150}$$
$$=C_1+C_2 \tag{151}$$

where C_1 (resp. C_2) is the first (resp. second) sum in (151).

Then, using (88e) and the derivatives (143), (144) we obtain

$$C_1\Big|_{\substack{\mu_1=0\\ \mu_2=0}} = 0, \tag{152}$$

$$C_2\Big|_{\substack{\mu_1=0\\ \mu_2=0}} = \sum_z \frac{(P_Z^B(z) - Q_0(z))(\tilde{P}_Z^A(z) - Q_0(z))}{Q_0(z)} \tag{153}$$

$$= \chi_2(\tilde{P}_Z^A, P_Z^B||Q_0) \tag{154}$$

The proof of (91c) follows from (151), (152) and (154). □

REFERENCES

[1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J*, vol. 54, pp. 1355–1387, May 1975.

[2] 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.

[3] 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.

[4] M. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inform. Theory*, vol. 62, pp. 2334–2354, May 2016.

[5] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Inform. Theory*, vol. 62, pp. 3493–3503, Jun 2016.

[6] A. Bounhar, M. Sarkiss, and M. Wigger, "Capacity-key tradeoff in covert communication," in *2025 IEEE Information Theory Workshop (ITW)*, (Sydney, Australia), 2025.

[7] 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.

[8] K.-H. Cho and S.-H. Lee, "Treating interference as noise is optimal for covert communication over interference channels," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 322–332, 2021.

[9] 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.

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

[11] V. Y. F. Tan and S.-H. Lee, "Time-division is optimal for covert communication over some broadcast channels," *IEEE Trans. Inf. Forensics Security*, vol. 14, pp. 1377–1389, May 2019.

[12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U.K.: Cambridge University Press, 2nd ed., 2011.

[13] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 2–22, 1954.

[14] J. Liu, P. Cuff, and S. Verdú, "One-shot mutual covering lemma and marton's inner bound with a common message," in *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2015)*, (Hong Kong), June 2015.

[15] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, Jul 1994.

[16] T. S. Han, *Information Spectrum Methods in Information Theory*. Berlin, Germany: Springer-Verlag, 2002.