Strong Secrecy for Cooperative Broadcast Channels

Ziv Goldfeld, Gerhard Kramer, Haim H. Permuter and Paul Cuff

Abstract

A broadcast channel (BC) where the decoders cooperate via a one-sided link is considered. One common and two private messages are transmitted and the private message to the cooperative user should be kept secret from the cooperation-aided user. The secrecy level is measured in terms of strong-secrecy, i.e., a vanishing information leakage. An inner bound on the capacity region is derived by using a channel-resolvability-based code that *double-bins* the codebook of the secret message, and by using a *likelihood encoder* to choose the transmitted codeword. The inner bound is shown to be tight for semi-deterministic and physically degraded BCs and the results are compared to those of the corresponding BCs without a secrecy constraint. Blackwell and Gaussian BC examples illustrate the impact of secrecy on the rate regions. Unlike the case without secrecy, where sharing information about both private messages via the cooperative link is optimal, our protocol conveys parts of the common and non-confidential messages only. This restriction reduces the transmission rates more than the usual rate loss due to secrecy requirements.

Index Terms

Broadcast channel, conferencing, cooperation, likelihood encoder, physical-layer security, resolvability, strong secrecy.

I. Introduction

User cooperation and security are two essential aspects of modern communication systems. Cooperation can increase transmission rates, whereas security requirements can limit these rates. To shed light on the interaction between these two phenomena, we study broadcast channels (BCs) with one-sided decoder cooperation and one confidential message (Fig. 1). Cooperation is modeled as *conferencing*, i.e., information exchange via a rate-limited link that extends from one receiver (referred to as the *cooperative receiver*) to the other (the *cooperation-aided receiver*). The cooperative receiver possesses confidential information that should be kept secret from the other user.

Secret communication over noisy channels was modeled by Wyner who introduced the degraded wiretap channel (WTC) and derived its secrecy-capacity [1]. Wyner's wiretap code relied on a *capacity-based* approach, i.e., the code

The work of Z. Goldfeld and H. H. Permuter was supported by the Israel Science Foundation (grant no. 684/11), an ERC starting grant and the Cyber Security Research Grant at Ben-Gurion University of the Negev. The work of G. Kramer was supported by an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research. The work of P. Cuff was supported by the National Science Foundation (grants CCF-1350595 and CCF-1116013) and the Air Force Office of Scientific Research (grants FA9550-15-1-0180 and FA9550-12-1-0196). This paper was presented in part at the 2015 IEEE International Symposium on Information Theory, Hong-Kong, and in part at the 2016 International Zurich Seminar on Communications, Zurich, Switzerland.

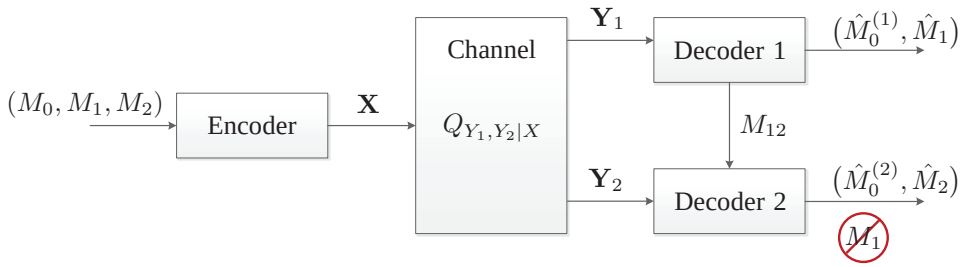


Fig. 1: Cooperative BCs with one confidential message.

is a union of subcodes that operate just below the capacity of the eavesdropper's channel. Csiszár and Körner [2] generalized Wyner's result to a general BC. Multiuser settings with secrecy have since been extensively treated in the literature. Broadcast and interference channels with two confidential messages were studied in [3]–[7]. Gaussian multiple-input multiple-output (MIMO) BCs and WTCs were studied in [8]–[13], while [14]–[16] focus on BCs with an eavesdropper as an external entity from which all messages are kept secret.

The above papers consider the *weak-secrecy* metric, i.e., a vanishing information leakage *rate* to the eavesdropper. Although the leakage rate vanishes asymptotically with the blocklength, the eavesdropper can decipher an increasing number of bits of the confidential message. This drawback was highlighted in [17]–[19] (see also [20]), which advocated using the *information leakage* as a secrecy measure referred to as *strong-secrecy*. We consider strong-secrecy by relying on work by Csiszár [20] and Hayashi [21] to relate the coding mechanism for secrecy to *channel-resolvability*.

The problem of channel resolvability, closely related to the early work of Wyner [22], was formulated by Han and Verdú [23] in terms of total variational (TV) distance. Recently, [24] advocated replacing the TV metric with *unnormalized relative entropy*. In [25], the coding mechanism for the resolvability problem was extended to various scenarios under the name *soft-covering lemma*. These extensions were used to design secure communication protocols for several source coding problems under different secrecy measures [26]–[29]. A *resolvability-based* wiretap code associates with each message a subcode that operates just above the resolvability of the eavesdropper's channel. Using such constructions, [30] extended the results of [2] to strong-secrecy for continuous random variables and channels with memory. In [31] (see also [32, Remark 2.2]), resolvability-based codes were used to establish the strong-secrecy-capacities of the discrete and memoryless (DM) WTC and the DM-BC with confidential messages by using a metric called *effective secrecy*.

Our inner bound on the strong-secrecy-capacity region of the cooperative BC is based on a resolvability-based *Marton* code. Specifically, we consider a state-dependent channel over which an encoder with non-causal access to the state sequence aims to make the conditional probability mass function (PMF) of the channel output given the state a product PMF. The resolvability code coordinates the transmitted codeword with the state sequence by means of multicoding, i.e., by associating with every message a bin that contains enough codewords to ensure joint encoding (similar to a Gelfand-Pinsker codebook). Most encoders use joint typicality tests to determine the

transmitted codeword. We adopt the *likelihood encoder*, recently proposed as a coding strategy for source coding problems [33], as our multicoding mechanism. Doing so significantly simplifies the distribution approximation analysis. We prove that the TV distance between the induced output PMF and the target product PMF approaches zero exponentially fast in the blocklength, which implies convergence in unnormalized relative entropy [34, Theorem 17.3.3].

Next, we construct a BC code in which the relation between the codewords corresponds to the relation between the channel states and the channel inputs in the resolvability problem. To this end we associate with every confidential message a subcode that adheres to the structure of the aforementioned resolvability code. Accordingly, the confidential message codebook is double-binned to allow joint encoding via the likelihood encoder (outer bin layer) and preserves confidentiality (inner bin layer). The bin sizes are determined by the rate constraints for the resolvability problem, which ensures strong-secrecy. The inner bound induced by this coding scheme is shown to be tight for semi-deterministic (SD) and physically-degraded (PD) BCs.

Our protocol uses the cooperation link to convey information about the non-confidential message and the common message. Without secrecy constraints, the optimal scheme shares information on *both* private messages as well as the common message [35]. We show that the restricted protocol results in an additional rate loss on top of standard losses due to secrecy. To this end we compare the achievable regions induced by each cooperation strategy for a cooperative BC *without secrecy*. We show that the restricted protocol does not lose rate when the BC is deterministic or PD, but it is sub-optimal in general.

To the best of our knowledge, we present here the first resolvability-based Marton code. This is also a first demonstration of the likelihood encoder's usefulness in the context of secrecy for channel coding problems. From a broader perspective, our resolvability result is a tool for proving strong-secrecy in settings with Marton coding. As a special case, we derive the secrecy-capacity region of the SD-BC (without cooperation) where the message of the deterministic user is confidential - a new result that has merit on its own. The structure of the obtained region provides insight into the effect of secrecy on the coding strategy for BCs. A comparison between the cooperative PD-BC with and without secrecy is also given.

The results are visualized by considering a Blackwell BC (BBC) [36], [37] and a Gaussian BC. An explicit strong-secrecy-achieving coding strategy for an extreme point of the BBC region is given. Although the BBC's input is ternary, to maximize the transmission rate of the confidential message only a binary subset of the input's alphabet is used. As a result, a zero-capacity channel is induced to the other user, who, therefore, cannot decode any of the secret bits. Further, we show that in the BBC scenario, an improved subchannel (given by the identity mapping) to the legitimate receiver does not increase the strong-secrecy-capacity region.

This paper is organized as follows. Section II provides preliminaries and restates some useful basic properties. In Section III we state a resolvability lemma. Section IV introduces the cooperative BC with one confidential message and gives an inner bound on its strong-secrecy-capacity region. The secrecy-capacity regions for the SD and PD scenarios are then characterized. In Section V the effect of secrecy constraints on the optimal cooperation protocol is discussed. Section VI compares the capacity regions of SD- and PD-BCs with and without secrecy. Blackwell

and Gaussian BCs visualise the results. Finally, proofs are provided in Section VII, while Section VIII summarizes the main achievements and insights of this work.

II. Notation and Preliminaries

We use the following notation. Given two real numbers a, b, we denote by $[a\!:\!b]$ the set of integers $\big\{n \in \mathbb{N}\big|\lceil a\rceil \leq n \leq \lfloor b\rfloor\big\}$. We define $\mathbb{R}_+ = \{x \in \mathbb{R}|x \geq 0\}$. Calligraphic letters denote discrete sets, e.g., $\mathcal{X}$, while the cardinality of a set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. $\mathcal{X}^n$ stands for the n-fold Cartesian product of $\mathcal{X}$. An element of $\mathcal{X}^n$ is denoted by $x^n = (x_1, x_2, \ldots, x_n)$, and its substrings as $x_i^j = (x_i, x_{i+1}, \ldots, x_j)$; when $i = 1$, the subscript is omitted. We define $x^{n\setminus i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. Whenever the dimension n is clear from the context, vectors (or sequences) are denoted by boldface letters, e.g., $\mathbf{x}$. Random variables are denoted by uppercase letters, e.g., X, with similar conventions for random vectors. X_i^j represents the sequence of random variables $(X_i, X_{i+1}, \ldots, X_j)$, while $\mathbf{X}$ stands for X^n. The probability of an event $\mathcal{A}$ is denoted by $\mathbb{P}(\mathcal{A})$, while $\mathbb{P}(\mathcal{A}\big|\mathcal{B})$ denotes conditional probability of $\mathcal{A}$ given $\mathcal{B}$. We use $\mathbb{1}_{\mathcal{A}}$ to denote the indicator function of $\mathcal{A}$. Probability mass functions (PMFs) are denoted by the capital letter P, with a subscript that identifies the random variable and its possible conditioning. For example, for two jointly distributed random variables X and Y, let P_X, $P_{X,Y}$ and $P_{X|Y}$ denote, respectively, the PMF of X, the joint PMF of (X, Y) and the conditional PMF of X given Y. In particular, when X and Y are discrete, $P_{X|Y}$ represents the stochastic matrix whose elements are given by $P_{X|Y}(x|y) = \mathbb{P}\big(X = x|Y = y\big)$. We omit subscripts if the arguments of the PMF are lowercase versions of the random variables. The support of a PMF P and the expectation of a random variable X are denoted by $\mathrm{supp}(P)$ and $\mathbb{E}\big[X\big]$, respectively. We use $\mathbb{E}_P$ and $\mathbb{P}_P$ to indicate that an expectation or a probability are taken taken with respect to a PMF P (when the PMF is clear from the context, the subscript is omitted). If the entries of X^n are drawn in an independent and identically distributed (i.i.d.) manner according to P_X, then for every $\mathbf{x} \in \mathcal{X}^n$ we have $P_{X^n}(\mathbf{x}) = \prod_{i=1}^n P_X(x_i)$ and we write $P_{X^n}(\mathbf{x}) = P_X^n(\mathbf{x})$. Similarly, if for every $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ we have $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$, then we write $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = P_{Y|X}^n(\mathbf{y}|\mathbf{x})$. We often use Q_X^n or $Q_{Y|X}^n$ when referring to product PMFs. The conditional product PMF $Q_{Y|X}^n$ given a specific sequence $\mathbf{x} \in \mathcal{X}^n$ is denoted by $Q_{Y|X=\mathbf{x}}^n$. The empirical PMF $\nu_{\mathbf{x}}$ of a sequence $\mathbf{x} \in \mathcal{X}^n$ is

$$\nu_{\mathbf{x}}(x) \triangleq \frac{N(x|\mathbf{x})}{n} \tag{1}$$

where $N(x|\mathbf{x}) = \sum_{i=1}^n \mathbb{1}_{\{x_i = x\}}$. We use $\mathcal{T}_\epsilon^{(n)}(P_X)$ to denote the set of letter-typical sequences of length n with respect to the PMF P_X and the non-negative number ϵ [38, Ch. 3], [39], i.e., we have

$$\mathcal{T}_\epsilon^{(n)}(P_X) = \Big\{\mathbf{x} \in \mathcal{X}^n \Big| \ \big|\nu_{\mathbf{x}}(x) - P_X(x)\big| \leq \epsilon P_X(x), \ \forall x \in \mathcal{X}\Big\}. \tag{2}$$

Definition 1 (TV Distance and Relative Entropy) *Let P and Q be two PMFs on a countable sample space $\mathcal{X}$*

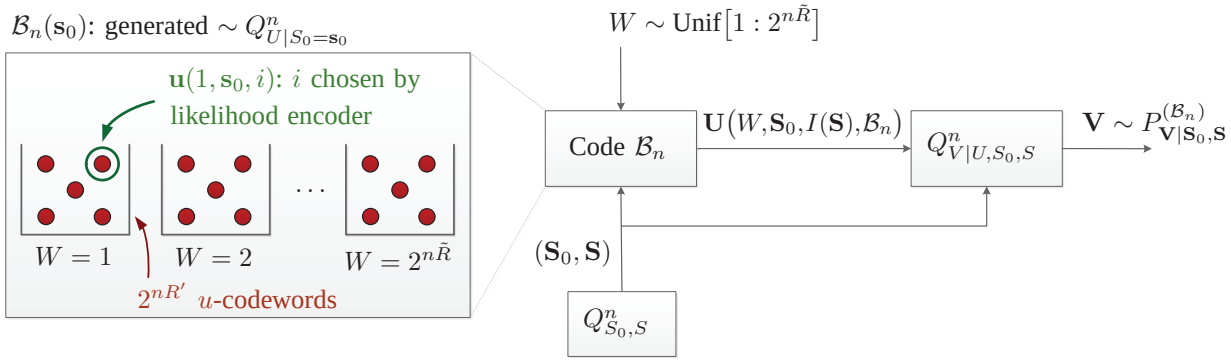


Fig. 2: Coding problem for approximating $P^{(\mathcal{B}_n)}_{\mathbf{V}|\mathbf{S}_0,\mathbf{S}_0}\approx Q^n_{V|S_0,S}$ and the resolvability subcode that is superimposed on $\mathbf{s}_0\in\mathcal{S}_0^n$: Every subcode contains $2^{n(\tilde{R}+R')}$ u-codewords drawn independently according to $Q^n_{U|S_0=\mathbf{s}_0}$. The codewords are partitioned into $2^{n\tilde{R}}$, each associated with a certain $w\in\big[1:2^{n\tilde{R}}\big]$. To transmit $W=w$ the likelihood encoder from (6) is used to choose a u-codeword for the wth bin.

[1]. *The TV distance and the relative entropy between P and Q are*

$$||P-Q||_{TV}=\frac{1}{2}\sum_{x\in\mathcal{X}}\big|P(x)-Q(x)\big| \tag{3}$$

and

$$D(P||Q)=\sum_{x\in\mathrm{supp}(P)}P(x)\log\frac{P(x)}{Q(x)} \tag{4}$$

respectively.

Remark 1 *Pinsker's inequality shows that relative entropy is larger than TV distance. A reverse inequality is sometimes valid. For example, if $P\ll Q$ (i.e., P is absolutely continuous with respect to Q), and Q is an i.i.d. discrete distribution of variables, then (see [25, Equation (29)])* [2]

$$D(P||Q)\in\mathcal{O}\left(\left[n+\log\frac{1}{||P-Q||}\right]||P-Q||\right). \tag{5}$$

In particular, (5) implies that an exponential decay of the TV distance produces an exponential decay of the informational divergence with the same exponent.

III. Conditional Relative Entropy Approximation

A. Problem Definition

Consider a state-dependent discrete memoryless channel (DMC) over which an encoder with non-causal access to the i.i.d. state sequence transmits a codeword (Fig. 2). Each channel state is a pair (S_0,S) of random variables drawn according to $Q_{S_0,S}$. The encoder superimposes its codebook on S_0 and then uses the *likelihood encoder* with

[1]Countable sample spaces are assumed throughout this work.

[2]$f(n)\in\mathcal{O}\big(g(n)\big)$ means that $f(n)\le k\cdot g(n)$, for some k independent of n and sufficiently large n.

respect to S to choose the channel input sequence. The structure of a subcode that is superimposed on some $\mathbf{s}_0 \in \mathcal{S}_0^n$ is also illustrated in Fig. 2. The conditional PMF of the channel output, given the states, should approximate a conditional product distribution in terms of unnormalized relative entropy.

The random variable W is uniformly distributed over $\mathcal{W} = \left[1 : 2^{n\tilde{R}}\right]$ and is independent of $(\mathbf{S}_0, \mathbf{S}) \sim Q_{S_0,S}^n$.

Codebook Construction: For every $\mathbf{s}_0 \in \mathcal{S}_0^n$ generate a codebook $\mathcal{B}_n(\mathbf{s_0})$ that comprises $2^{n\tilde{R}}$ bins. Each bin is associated with a different message $w \in \mathcal{W}$ and contains $2^{nR'}$ u-codewords that are drawn according to $Q_{U|S_0=\mathbf{s}_0}^n \triangleq \prod_{i=1}^n Q_{U|S_0}(\cdot|s_{0,i})$. Let $\mathcal{B}_n = \left\{\mathcal{B}_n(\mathbf{s}_0)\right\}_{\mathbf{s}_0 \in \mathcal{S}_0^n}$ denote this collection of codebooks and denote the codewords in the bin associated with $w \in \mathcal{W}$ by $\left\{\mathbf{u}(\mathbf{s}_0, w, i, \mathcal{B}_n)\right\}_{i \in \mathcal{I}}$, where $\mathcal{I} = \left[1 : 2^{nR'}\right]$.

Encoding and Induced PMF: Consider the *likelihood encoder* described by conditional PMF

$$f^{(\mathrm{LE})}(i|w, \mathbf{s}_0, \mathbf{s}, \mathcal{B}_n) = \frac{Q_{S|U,S_0}^n\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0, w, i, \mathcal{B}_n), \mathbf{s}_0\big)}{\sum\limits_{i' \in \mathcal{I}} Q_{S|U,S_0}^n\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0, w, i', \mathcal{B}_n), \mathbf{s}_0\big)}. \tag{6}$$

Upon observing $(w, \mathbf{s}_0, \mathbf{s})$, an index $i \in \mathcal{I}$ is drawn randomly according to the probability distribution (6). The codeword $\mathbf{u}(\mathbf{s}_0, w, i, \mathcal{B}_n)$ is passed through the DMC $Q_{V|U,S_0,S}^n$. The distribution induced by the codebook $\mathcal{B}_n$ is

$$P^{(\mathcal{B}_n)}(\mathbf{s}_0, \mathbf{s}, w, i, \mathbf{u}, \mathbf{v}) = Q_{S_0,S}^n(\mathbf{s}_0, \mathbf{s}) 2^{-n\tilde{R}} f^{(\mathrm{LE})}(i|w, \mathbf{s}_0, \mathbf{s}, \mathcal{B}_n) \mathbb{1}_{\left\{\mathbf{u}(\mathbf{s}_0, w, i, \mathcal{B}_n) = \mathbf{u}\right\}} Q_{V|U,S_0,S}^n(\mathbf{v}|\mathbf{u}, \mathbf{s}_0, \mathbf{s}). \tag{7}$$

Furthermore, we use $\mathbb{B}_n$ to denote a random codebook that adheres to the above construction.

Lemma 1 (Sufficient Conditions for Approximation) *For any $Q_{S_0,S}$, $Q_{U|S_0,S}$ and $Q_{V|U,S_0,S}$, if $(\tilde{R}, R') \in \mathbb{R}_+^2$ satisfies*

$$R' > I(U; S|S_0) \tag{8a}$$

$$R' + \tilde{R} > I(U; S, V|S_0) \tag{8b}$$

then

$$\mathbb{E}_{\mathbb{B}_n} D\Big(P_{\mathbf{V}|\mathbf{S}_0,\mathbf{S}}^{(\mathbb{B}_n)} \Big|\Big| Q_{V|S_0,S}^n \Big| Q_{S_0,S}^n\Big) \xrightarrow[n\to\infty]{} 0. \tag{9}$$

The proof of Lemma 1 is given in Section VII-A and it shows that the TV distance decays exponentially fast with the blocklength n. By Remark 1 this implies an exponential decay of the desired relative entropy. Another useful property is that the chosen u-codeword is jointly letter-typical with $(\mathbf{S}_0, \mathbf{S})$ with high probability.

Lemma 2 (Typical with High Probability) *If $(\tilde{R}, R') \in \mathbb{R}_+^2$ satisfies* (8)*, then for any $w \in \mathcal{W}$ and $\epsilon > 0$, we have*

$$\mathbb{E}_{\mathbb{B}_n} \mathbb{P}\Big(\big(\mathbf{S}_0, \mathbf{S}, \mathbf{U}(\mathbf{S}_0, w, I, \mathbb{B}_n)\big) \notin \mathcal{T}_\epsilon^{(n)}(Q_{S_0,S,U}) \Big| \mathbb{B}_n\Big) \xrightarrow[n\to\infty]{} 0. \tag{10}$$

The proof of Lemma 2 is given in Section VII-B.

IV. Cooperative Broadcast Channels with One Confidential Message

A. Problem Definition

The cooperative DM-BC with one confidential message is illustrated in Fig. 1. The channel has one sender and two receivers. The sender chooses a triple (m_0, m_1, m_2) of indices uniformly and independently from the set $[1 : 2^{nR_0}] \times [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$ and maps it to a sequence $\mathbf{x} \in \mathcal{X}^n$. The sequence $\mathbf{x}$ is transmitted over a BC with transition probability $Q_{Y_1,Y_2|X}$. If $Q_{Y_1,Y_2|X}$ factors as $\mathbb{1}_{\{Y_1=f(X)\}}Q_{Y_2|X}$ or $Q_{Y_1|X}Q_{Y_2|Y_1}$ then we call the BC SD or PD, respectively. The output sequence $\mathbf{y}_j \in \mathcal{Y}_j^n$, where $j = 1, 2$, is received by decoder j. Decoder j produces a pair of estimates $\left(\hat{m}_0^{(j)}, \hat{m}_j\right)$ of (m_0, m_j). Furthermore, the message m_1 is to be kept secret from Decoder 2. There is a one-sided noiseless cooperation link of rate R_{12} from Decoder 1 to Decoder 2. By conveying a message $m_{12} \in [1 : 2^{nR_{12}}]$ over this link, Decoder 1 can share with Decoder 2 information about $\mathbf{y}_1$, $\left(\hat{m}_0^{(1)}, \hat{m}_1\right)$, or both.

Definition 2 (Code) *An $(n, R_{12}, R_0, R_1, R_2)$ code $\mathcal{C}$ for the BC with cooperation and one confidential message has:*

1) *Four message sets $\mathcal{M}_{12} = \left[1 : 2^{nR_{12}}\right]$ and $\mathcal{M}_j = \left[1 : 2^{nR_j}\right]$, for $j = 0, 1, 2$.*
2) *A stochastic encoder described by a stochastic matrix $f^{(\mathcal{E})}_{\mathbf{X}|M_0,M_1,M_2}$ on $\mathcal{X}^n$.*
3) *A decoder cooperation function $f_{12} : \mathcal{Y}_1^n \to \mathcal{M}_{12}$.*
4) *Two decoding functions $\phi_1 : \mathcal{Y}_1^n \to \mathcal{M}_0 \times \mathcal{M}_1$ and $\phi_2 : \mathcal{Y}_2^n \times \mathcal{M}_{12} \to \mathcal{M}_0 \times \mathcal{M}_2$.*

Definition 3 (Error Probability) *The average error probability for an $(n, R_{12}, R_0, R_1, R_2)$ code $\mathcal{C}_n$ is*

$$P_e(\mathcal{C}_n) = \mathbb{P}_{\mathcal{C}_n}\left((\hat{M}_0^{(1)}, \hat{M}_0^{(2)}, \hat{M}_1, \hat{M}_2) \neq (M_0, M_0, M_1, M_2)\right)$$

where $\mathbb{P}_{\mathcal{C}_n}(\cdot)$ means that the probability is calculated with respect to the joint PMF induced by $\mathcal{C}_n$. Furthermore, $\left(\hat{M}_0^{(1)}, \hat{M}_1\right) = \phi_1(\mathbf{Y}_1)$ and $\left(\hat{M}_0^{(2)}, \hat{M}_2\right) = \phi_2\big(\mathbf{Y}_2, f_{12}(\mathbf{Y}_1)\big)$.

The *information leakage* at receiver 2 is measured by $\mathcal{L}(\mathcal{C}_n) = I_{\mathcal{C}_n}(M_1; M_{12}, \mathbf{Y}_2)$, which is also calculated with respect to PMF induced by $\mathcal{C}_n$.

Definition 4 (Achievability) *A rate tuple $(R_{12}, R_0, R_1, R_2) \in \mathbb{R}_+^4$ is achievable if for any $\epsilon > 0$ there is an $(n, R_{12}, R_0, R_1, R_2)$ code $\mathcal{C}_n$ with*

$$P_e(\mathcal{C}_n) \leq \epsilon \tag{11a}$$

$$\mathcal{L}(\mathcal{C}_n) \leq \epsilon \tag{11b}$$

for n sufficiently large.

The *strong-secrecy-capacity region* $\mathcal{C}_{\mathrm{S}}$ is the closure of the set of the achievable rates.

B. Strong-Secrecy-Capacity Bounds and Results

We state an inner bound on the strong-secrecy-capacity region $\mathcal{C}_{\mathrm{S}}$ of a cooperative BC with one confidential message.

Theorem 3 (Inner Bound) *Let $\mathcal{R}_{\mathrm{I}}$ be the closure of the union of rate tuples $(R_{12}, R_0, R_1, R_2) \in \mathbb{R}_+^4$ satisfying:*

$$R_1 \leq I(U_1; Y_1|U_0) - I(U_1; U_2, Y_2|U_0) \tag{12a}$$

$$R_0 + R_1 \leq I(U_0, U_1; Y_1) - I(U_1; U_2, Y_2|U_0) \tag{12b}$$

$$R_0 + R_2 \leq I(U_0, U_2; Y_2) + R_{12} \tag{12c}$$

$$R_0 + R_1 + R_2 \leq I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2, Y_2|U_0) \tag{12d}$$

where the union is over all PMFs $Q_{U_0,U_1,U_2,X} Q_{Y_1,Y_2|X}$. Then the following inclusion holds:

$$\mathcal{R}_{\mathrm{I}} \subseteq \mathcal{C}_{\mathrm{S}}. \tag{13}$$

Furthermore, $\mathcal{R}_{\mathrm{I}}$ is convex and one may choose $|\mathcal{U}_0| \leq |\mathcal{X}| + 5$, $|\mathcal{U}_1| \leq |\mathcal{X}|$ and $|\mathcal{U}_2| \leq |\mathcal{X}|$.

The proof of Theorem 3 relies on a channel-resolvability-based Marton code and is given in Section VII-C. Two key ingredients allow us keeping M_1 secret while still utilizing the cooperation link to help Receiver 2. First, the cooperation strategy is modified compared to the case without secrecy that was studied in [35], where M_{12} conveyed information about *both* private messages as well as the common message. Here, the confidentiality of M_1 restricts the cooperation message from containing any information about M_1, and therefore, we use an M_{12} that is a function of (M_0, M_2) only. Since the protocol requires Receiver 1 to decode the information it shares with Receiver 2, this modified cooperation strategy results in a rate loss in R_1 when compared to [35]; the loss is expressed in the first mutual information term in (12a) being conditioned on U_0 rather than having U_0 next to U_1.

The second ingredient is associating with each $m_1 \in \mathcal{M}_1$ a resolvability-subcode that adheres to the code construction for Lemmas 1 and 2 described in Section III-A. By doing so, the relations between the codewords in the Marton code correspond to those between the channel states and its input in the resolvability problem. Marton coding combines superposition coding and binning, hence the different roles the state sequences $\mathbf{S}_0$ and $\mathbf{S}$ play in our resolvability setup. Reliability is established with the help of Lemma 2, while invoking Lemma 1 ensures strong-secrecy.

Theorem 4 (SD-BC Secrecy-Capacity) *The strong-secrecy-capacity region $\mathcal{C}_{\mathrm{S}}^{(\mathrm{SD})}$ of a cooperative SD-BC with one confidential message is the closure of the union of rate tuples $(R_{12}, R_0, R_1, R_2) \in \mathbb{R}_+^4$ satisfying:*

$$R_1 \leq H(Y_1|W, V, Y_2) \tag{14a}$$

$$R_0 + R_1 \leq H(Y_1|W, V, Y_2) + I(W; Y_1) \tag{14b}$$

$$R_0 + R_2 \leq I(W, V; Y_2) + R_{12} \tag{14c}$$

$$R_0 + R_1 + R_2 \leq H(Y_1|W,V,Y_2) + I(V;Y_2|W) + I(W;Y_1) \tag{14d}$$

where the union is over all PMFs $Q_{W,V,Y_1,X}Q_{Y_2|X}$ with $Y_1 = f(X)$. Furthermore, $\mathcal{C}_{\mathrm{S}}^{(\mathrm{SD})}$ is convex and one may choose $|\mathcal{W}| \leq |\mathcal{X}| + 3$ and $|\mathcal{V}| \leq |\mathcal{X}|$.

The direct part of Theorem 4 follows from Theorem 3 by setting $U_0 = W$, $U_1 = Y_1$ and $U_2 = V$. The converse is given in Section VII-D.

Theorem 5 (PD-BC Secrecy-Capacity) *The strong-secrecy-capacity region $\mathcal{C}_{\mathrm{S}}^{(\mathrm{PD})}$ of a cooperative PD-BC with on confidential message is the closure of the union of rate tuples $(R_{12}, R_0, R_1, R_2) \in \mathbb{R}_+^4$ satisfying:*

$$R_1 \leq I(X;Y_1|W) - I(X;Y_2|W) \tag{15a}$$

$$R_0 + R_2 \leq I(W;Y_2) + R_{12} \tag{15b}$$

$$R_0 + R_1 + R_2 \leq I(X;Y_1) - I(X;Y_2|W) \tag{15c}$$

where the union is over all PMFs $Q_{W,X}Q_{Y_1|X}Q_{Y_2|Y_1}$. Furthermore, $\mathcal{C}_{\mathrm{S}}^{(\mathrm{PD})}$ is convex and one may choose $|\mathcal{W}| \leq |\mathcal{X}| + 2$.

The achievability of $\mathcal{C}_{\mathrm{S}}^{(\mathrm{PD})}$ follows by setting $U_0 = W$, $U_1 = X$ and $U_2 = 0$ in Theorem 3. For the converse see Section VII-E.

Remark 2 (Converse) *We use two distinct converse proofs for Theorems 4 and 5. In the converse of Theorem 4, the bound in* (14d) *does not involve R_{12} since the auxiliary random variable W_i contains M_{12}. With respect to this choice of W_i, showing that $W - X - (Y_1, Y_2)$ forms a Markov chain relies on the SD property of the channel. For the PD-BC, however, such an auxiliary is not feasible as it violates the Markov relation $W - X - Y_1 - Y_2$ induced by the channel. To circumvent this, in the converse of Theorem 5 we define W_i without M_{12} and use the structure of the channel to keep R_{12} from appearing in* (15c)*. Specifically, this argument relies on the relation $M_{12} = f_{12}(\mathbf{Y}_1)$ and that Y_2 is a degraded version of Y_1, implying that all three messages (M_0, M_1, M_2) are reliably decodable from $\mathbf{Y}_1$ only.*

Remark 3 *The cardinality bounds on the auxiliary random variables in Theorems 3, 4 and 5 are established using the perturbation method [40] and the Carathéodory-Fenchel theorem.*

V. Sub-Optimal Cooperation without Secrecy

The cooperation protocol for the BC with a secret M_1 uses the cooperative link to convey information that is a function of the non-confidential message and the common message. Without secrecy constraints, it was shown in [35] that the best cooperation strategy uses a public message that comprises parts of *both* private messages as well as the common message. To understand whether the cooperation restriction reduces the transmission rates beyond standard losses due to secrecy (which are discussed in Section VI), we compare the achievable regions induced

by each scheme for the cooperative BC *without secrecy*. For simplicity we consider the setting without a common message, i.e., when $R_0 = 0$.

At first glance it is not clear why the cooperative receiver (Decoder 1) should share information about M_1 since the cooperation-aided receiver (Decoder 2) is not required to decode M_1. However, we show that the restricted protocol is sub-optimal in general. For BCs in which Decoder 1 can decode more than nR_{12} bits of M_2 (e.g., PD-BCs), both protocols achieve the same rates and no sharing information about M_1 is needed. However, when Decoder 1 can decode strictly less than nR_{12} bits of M_2, then sharing M_1 achieves higher R_2 values, since now M_1 serves as side information for Decoder 2 in decoding M_2 (note that this side information is also available at the encoder).

The achievable region $\mathcal{R}_{\rm NS}$ for the cooperative BC without secrecy that was characterized in [35] (see also [41], [42]) is the union over the same domain as (12) of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \leq I(U_0, U_1; Y_1) \tag{16a}$$

$$R_2 \leq I(U_0, U_2; Y_2) + R_{12} \tag{16b}$$

$$R_1 + R_2 \leq I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0) \tag{16c}$$

$$R_1 + R_2 \leq I(U_1; Y_1|U_0) + I(U_0, U_2; Y_2) - I(U_1; U_2|U_0) + R_{12}. \tag{16d}$$

This region is tight for SD- and PD-BCs. The cooperation scheme that achieves (16) uses the pair (M_{10}, M_{20}) (where M_{j0} refers to the public part of the message M_j and has rate $R_{j0} \leq R_j$, for $j = 1, 2$) as a public message that is decoded by both users. The public message codebook (generated by i.i.d. realizations of the random variable U_0 in (16)) is partitioned into $2^{nR_{12}}$ bins and is first decoded by User 1. Next, the bin number M_{12} of the decoded public message is shared with User 2 over the cooperative link, which reduces the search space by a factor of $2^{nR_{12}}$. The presence of M_{10} in the public message essentially allows User 1 to achieve rates up $I(U_0, U_1; Y_1)$.

Introducing a secrecy constraint on M_1, we must remove M_{10} from the public message, but we keep the rest of the protocol unchanged. The region $\tilde{\mathcal{R}}_{\rm NS}$ achieved by the restricted cooperation protocol is derived by repeating the steps in the proof of [35, Theorem 6] while setting $R_{10} = 0$. One obtains that $\tilde{\mathcal{R}}_{\rm NS}$ is characterized by the same rate bounds as (16), up to replacing (16a) with

$$R_1 \leq I(U_1; Y_1|U_0) + \Big[I(U_2; Y_2|U_0) - I(U_1; U_2|U_0)\Big]^+ \tag{17}$$

where $[x]^+ = \max\big\{0, x\big\}$. Clearly $\tilde{\mathcal{R}}_{\rm NS} \subseteq \mathcal{R}_{\rm NS}$.

Note that $\tilde{\mathcal{R}}_{\rm NS} = \mathcal{R}_{\rm NS}$ for any BC where setting $U_0 = 0$ in (16) is optimal. In particular, we have the following proposition.

Proposition 6 (Restricted Cooperation is Optimal for Deterministic and PD BCs) *If the BC $Q_{Y_1,Y_2|X}$ is PD or deterministic, then $\tilde{\mathcal{R}}_{\rm NS} = \mathcal{R}_{\rm NS} = \mathcal{C}_{\rm NS}$.*

Proof: For the PD-BC, setting $U_0 = W$, $U_1 = X$ and $U_2 = 0$ into $\tilde{\mathcal{R}}_{\rm NS}$ recovers the region from [43, Equation

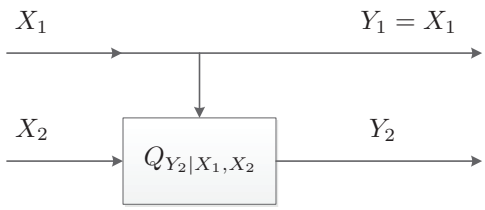


Fig. 3: A SD-BC example.

(17)], which is the capacity region of the cooperative PD-BC. The capacity region of the cooperative deterministic BC (DBC) given in [35, Corollary 12] is recovered from $\tilde{\mathcal{R}}_{\mathrm{NS}}$ by taking $U_0 = 0$, $U_1 = Y_1$ and $U_2 = Y_2$. ■

Proposition 7 (Restricted Cooperation Sub-Optimal) *There exist BCs $Q_{Y_1,Y_2|X}$ for which $\tilde{\mathcal{R}}_{\mathrm{NS}} \subsetneq \mathcal{R}_{\mathrm{NS}}$.*

The proof of Proposition 7 is given in Appendix A, where we construct an example for which the maximal achievable R_1 in both regions is the same, but the highest achievable R_2 while keeping R_1 at its maximum is strictly smaller in $\tilde{\mathcal{R}}_{\mathrm{NS}}$.

We start with a family of SD-BCs as illustrated in Fig. 3, where the channel input is $X = (X_1, X_2)$, the deterministic output is $Y_1 = X_1$ and the probabilistic output Y_2 is generated by the DMC $Q_{Y_2|X_1,X_2}$. All alphabets are binary, i.e., $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}$. The maximal achievable R_1 in both schemes is 1 and we set the capacity of the cooperation link to $R_{12} = 1$. We show that the highest R_2 such that $(R_{12}, R_1, R_2) = (1, 1, R_2) \in \mathcal{R}_{\mathrm{NS}}$ is lower bounded by the capacity of the state-dependent channel $Q_{Y_2|X_1,X_2}$ (with X_1 and X_2 playing the roles of the state and the input, respectively) with non-causal channel state information (CSI) available at the transmitting and receiving ends. This is due to the fact that $R_{12} = 1$ in the permissive protocol allows Decoder 1 to share X_1 with Decoder 2 despite the fact that $X_1^n = M_1$.

The corresponding value of R_2 in $\tilde{\mathcal{R}}_{\mathrm{NS}}$ is then upper bounded by the capacity of the same channel but with non-causal CSI at the transmitter only (also known as a Gelfand-Pinsker (GP) channel). The cooperation link is, in fact, useless in this scenario since $Y_1 = X_1$ and the restricted protocol prohibits exchanging any information about M_1. Thus, the proof boils down to choosing $Q_{Y_2|X_1,X_2}$ as a channel for which the capacity with full CSI is strictly larger than the GP capacity. The binary binary dirty-paper (BDP) channel [44]–[46] qualifies and completes the proof.

VI. Effect of Secrecy on Coding Schemes and Rate Regions for Cooperative BCs

The impact of the secrecy constraint on M_1 on the cooperation strategy and the resulting reduction of transmission rates was discussed in Section V. However, secrecy requirements have additional effects on BC codes regardless of user cooperation. We highlight these effects by comparing the SD and PD versions of the cooperative BC to their corresponding models without secrecy. For simplicity, throughout this section we again assume BCs with private messages only, i.e., $R_0 = 0$.

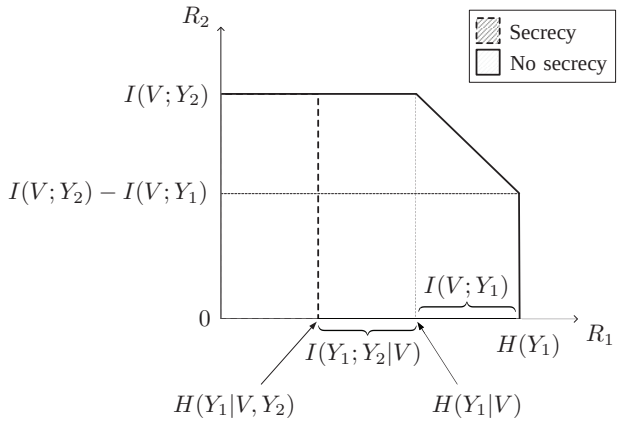


Fig. 4: Capacity region without secrecy vs. strong-secrecy-capacity region where M_1 is confidential for the SD-BC (without cooperation).

A. Semi-Deterministic BCs

1) Capacity Region Comparison: Consider the SD-BC without cooperation (i.e., where $R_{12} = 0$) in which M_1 is secret. By Theorem 4, the strong-secrecy-capacity region of the SD-BC with one confidential message, which was an unsolved problem until this work, is as follows.

Corollary 8 (Secrecy-Capacity for SD-BC without Cooperation) *The strong-secrecy-capacity region $\tilde{\mathcal{C}}_{\mathrm{S}}^{(\mathrm{SD})}$ of the SD-BC with one confidential message is the union of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying:*

$$R_1 \leq H(Y_1|V, Y_2) \tag{18a}$$

$$R_2 \leq I(V; Y_2) \tag{18b}$$

where the union is over all PMFs $Q_{V,Y_1,X} Q_{Y_2|X}$ with $Y_1 = f(X)$.

The region (18) coincides with $\mathcal{C}_{\mathrm{S}}^{(\mathrm{SD})}$ in (14d) (where $R_{12} = R_0 = 0$) by noting that the bound (14d) is redundant because if $Q_{W,V,Y_1,X} Q_{Y_2|X}$ is a PMF for which (14d) is active, then replacing W and V with $\tilde{W} = 0$ and $\tilde{V} = (W, V)$ achieves a larger region. Removing (14d) from $\mathcal{C}_{\mathrm{S}}^{(\mathrm{SD})}$ and setting $\tilde{V} = (W, V)$ recovers (18).

Marton coding achieves the capacity region of the classic SD-BC [47]. The capacity is the union of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying:

$$R_1 \leq H(Y_1) \tag{19a}$$

$$R_2 \leq I(V; Y_2) \tag{19b}$$

$$R_1 + R_2 \leq H(Y_1|V) + I(V; Y_2) \tag{19c}$$

where the union is over the same domain as in Corollary 8.

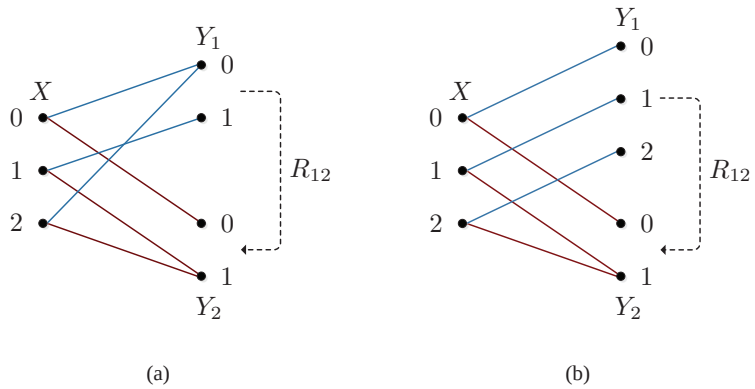


Fig. 5: (a) Cooperative Blackwell BC; (b) Cooperative Blackwell-like PD-BC.

The regions in (18) and (19) (for a fixed PMF) are depicted in Fig. 4. When M_1 is secret, one can no longer operate on both corner points of Marton's region. Rather, the optimal coding scheme is the one with the lower transmission rate to the 1st user. This essentially means that the redundancy in the codebook needed for multicoding befalls solely on User 1 (whose message is to be kept secret). Consequently, a loss of $I(V;Y_1)$, which corresponds to the sizes of the bins used for joint encoding, is inflicted on R_1. An additional rate-loss of $I(Y_1;Y_2|V)$ in R_1 is caused by a second layer of binning used to conceal M_1 from the 2nd user. A coding scheme for the higher corner point of the region without secrecy, i.e., the point $\big(\, H(Y_1)\,,\, I(V;Y_2)-I(V;Y_1)\,\big)$, is not feasible with secrecy since the larger value of R_1 violates the secrecy constraint. A similar effect occurs for the corresponding regions with cooperation.

2) Blackwell BC Example: Suppose the channel from the transmitter to receivers 1 and 2 is the BBC without a common message as illustrated in Fig 5(a) [36], [37]. Using Theorem 4 while setting $R_0 = 0$, the strong-secrecy-capacity region of a deterministic BC (DBC) is the following.

Corollary 9 (Secrecy-Capacity Region for DBC) *The strong-secrecy-capacity region $\mathcal{C}_{\mathrm{S}}^{(\mathrm{D})}$ of a cooperative DBC with one confidential message is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:*

$$R_1 \le H(Y_1|Y_2) \tag{20a}$$

$$R_2 \le H(Y_2) + R_{12} \tag{20b}$$

$$R_1 + R_2 \le H(Y_1, Y_2) \tag{20c}$$

where the union is over all PMFs Q_X.

Corollary 9 follows by arguments similar to those in the proof of [35, Corollary 12]. By parameterizing the input PMF Q_X as

$$Q_X(0) = \alpha \ ,\ Q_X(1) = \beta \ ,\ Q_X(2) = 1 - \alpha - \beta \tag{21}$$

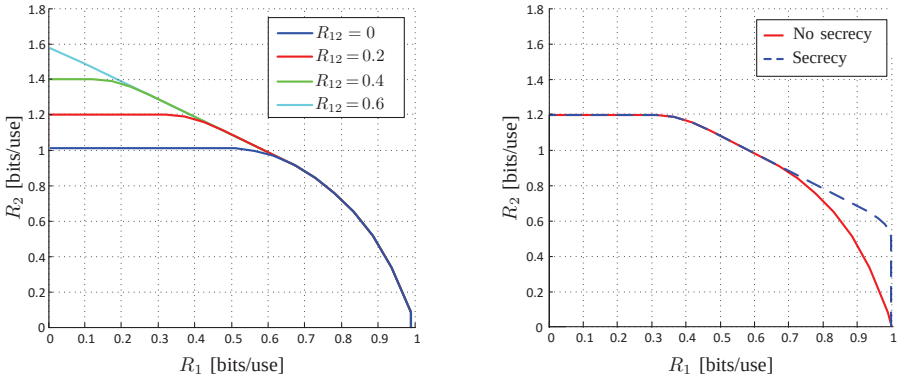


Fig. 6: (a) Projection of the strong-secrecy-capacity region of the cooperative BBC with one confidential message onto the plane (R_1, R_2) for different values of R_{12}; (b) Cooperative BBC with $R_{12} = 0.2$: Strong-secrecy-capacity region where M_1 is confidential vs. Capacity region without secrecy.

where $\alpha, \beta \in \mathbb{R}_+$ and $\alpha + \beta \leq 1$, the strong-secrecy-capacity region of the BBC is:

$$\mathcal{C}_{\mathrm{S}}^{(\mathrm{BBC})} = \bigcup_{\substack{\alpha,\beta\in\mathbb{R}_+,\\ \alpha+\beta\leq 1}} \left\{ (R_{12}, R_1, R_2) \in \mathbb{R}_+^3 \;\middle|\; \begin{array}{l} R_1 \leq (1-\alpha) H_b\left(\frac{\beta}{1-\alpha}\right) \\ R_2 \leq H_b(\alpha) + R_{12} \\ R_1 + R_2 \leq H_b(\alpha) + (1-\alpha) H_b(\frac{\beta}{1-\alpha}) \end{array} \right\}. \tag{22}$$

The projection of $\mathcal{C}_{\mathrm{S}}^{(\mathrm{BBC})}$ onto the plane (R_1, R_2) for different values of R_{12} is shown in Fig. 6(a). For every $R_{12} \in \mathbb{R}_+$, the maximal achievable R_1 in $\mathcal{C}_{\mathrm{S}}^{(\mathrm{BBC})}$ equals 1 [bits/use] (while the corresponding R_2 is zero). The rate triple $(R_{12}, 1, 0)$ is achieved by setting $\alpha = 0$ and $\beta = \frac{1}{2}$ in the bounds in (22). These probability values provide insight into the coding strategy that maximizes the transmission rate to User 1. Namely, the encoder chooses each channel input symbol uniformly from the set $\{1, 2\} \subsetneq \mathcal{X}$. By doing so, Decoder 1 effectively sees a clean binary channel (by mapping every received $Y_1 = 0$ to the input symbol $X = 2$) with capacity 1. Decoder 2, on the other hand, sees a flat channel with zero capacity since both $X = 1$ and $X = 2$ are mapped to $Y_2 = 1$. Thus, Decoder 2 has no information about the transmitted sequence, and therefore, strong-secrecy is achieved while conveying one secured bit to Decoder 1 in each channel use.

Remark 4 *An improved subchannel to the legitimate user does not enlarge the strong-secrecy-capacity region. We illustrate this by considering the Blackwell-like PD-BC (PD-BBC) shown in Fig. 5(b), where $\mathcal{Y}_1 = \mathcal{X}$ and $Y_1 = X$ ($\mathcal{Y}_2$ and the mapping from $\mathcal{X}$ to $\mathcal{Y}_2$ remain as in the BBC). Evaluating the strong-secrecy-capacity region of the PD-BBC reveals that it coincides with $\mathcal{C}_{\mathrm{S}}^{(\mathrm{BBC})}$. This implies that the Q_X that maximizes R_1 while keeping Decoder 2 ignorant of M_1 is $\alpha = 0$ and $\beta = \frac{1}{2}$, which coincides with the input PMF that maximizes R_1 while transmitting over the classical BBC. Thus, to ensure secrecy over the PD-BBC, the encoder overlooks the improved channel to Decoder 1 and ends up not using the symbol $X = 0$.*

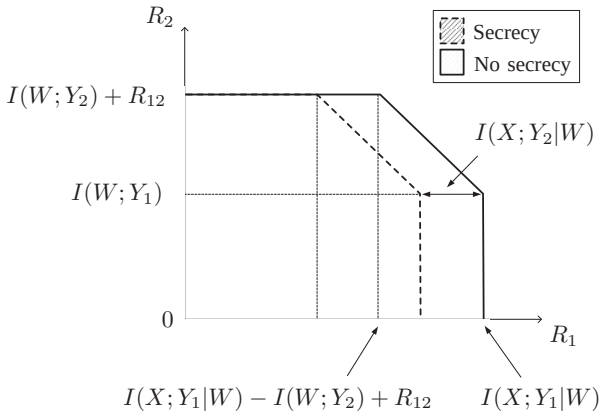


Fig. 7: Capacity region without secrecy vs. strong-secrecy-capacity region where M_1 is confidential for the cooperative PD-BC.

The effect of secrecy on the capacity region of a cooperative BC is illustrated by comparing to the BBC (Fig. 5(a)) without a secrecy constraint. Using the characterization of the capacity region of a cooperative DBC given in [35, Corollary 12] and the parametrization in (21), the capacity region of the cooperative BBC is:

$$\mathcal{C}_{\text{NS}}^{(\text{BBC})} = \bigcup_{\substack{\alpha,\beta\in\mathbb{R}_+,\\ \alpha+\beta\le 1}} \left\{ (R_{12},R_1,R_2)\in\mathbb{R}_+^3 \;\middle|\; \begin{array}{l} R_1 \le H_b(\alpha+\beta) \\ R_2 \le H_b(\alpha)+R_{12} \\ R_1+R_2 \le H_b(\alpha)+(1-\alpha)H_b(\frac{\beta}{1-\alpha}) \end{array} \right\}. \tag{23}$$

Fig. 6(b) compares the regions with and without secrecy. The dashed red line represents the capacity region for the case without secrecy while the blue line depicts the region where M_1 is confidential. Evidently, $\mathcal{C}_{\text{NS}}^{(\text{BBC})}$ is strictly larger than $\mathcal{C}_{\text{S}}^{(\text{BBC})}$. Note that up to approximately $R_1 \approx 0.6597 \triangleq R_1^{(\text{Th})}$, the two regions coincide. Thus, while $R_1 \le R_1^{(\text{Th})}$, concealing M_1 is achieved without any rate loss in R_2. When $R_1 > R_1^{(\text{Th})}$, however, an increased confidential message rate leads to a reduced R_2 value compared to the case without secrecy. Further, if *no secrecy constraint* is imposed on M_1, one can transmit it at its maximal rate of $R_1 = 1$ and still have a positive value of R_2 (up to approximately 0.5148). When M_1 is confidential then $R_1 = 1$ is achievable only if $R_2 = 0$.

B. Physically Degraded BCs

1) Capacity Region Comparison: When the BC is PD, the reduction in R_1 is due to the extra layer of bins in the codebook of M_1 only, while the modified cooperation scheme results in no loss. To see this, consider the capacity region $\mathcal{C}_{\text{NS}}^{(\text{PD})}$ of cooperative PD-BC without a secrecy constraint on M_1 (see [43] and [48]), which is the union over the same domain as (15) of rate triples $(R_{12},R_1,R_2)\in\mathbb{R}_+^3$ satisfying:

$$R_1 \le I(X;Y_1|W) \tag{24a}$$

$$R_2 \le I(W;Y_2)+R_{12} \tag{24b}$$

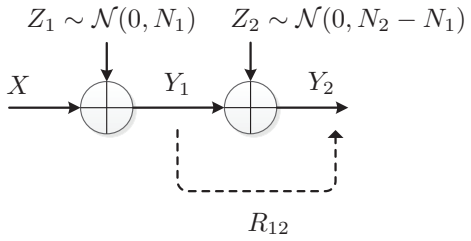


Fig. 8: Cooperative Gaussian PD-BC.

$$R_1 + R_2 \leq I(X; Y_1). \tag{24c}$$

In contrast to the SD case, the only impact of the secrecy requirement on the capacity region is expressed in a rate-loss of $I(X;Y_2|W)$ in R_1 (see (15a) in comparison to (24a)) that is due to the extra layer of bins needed for secrecy. Otherwise, the optimal code construction (and the optimal cooperation protocol) for both problems is the same. The similarity is because, whether M_1 is secret or not, its codebook is superimposed on the codebook of M_2, and decoding M_2 as part of the cooperation protocol comes without cost by the degraded property of the channel (see Proposition 6). Thus, for a fixed $Q_{W,X}Q_{Y_1|X}Q_{Y_2|Y_1}$, if $(R_{12}, R_1, R_2) \in \mathcal{C}_{\mathrm{NS}}^{(\mathrm{PD})}$ then $\Big(R_{12}, \big[R_1 - I(X;Y_2|V)\big]^+, R_2\Big) \in \mathcal{C}_{\mathrm{S}}^{(\mathrm{PD})}$, and vice versa. This relation is illustrated in Fig. 7 for some fixed value of R_{12} and under the assumption that $I(W;Y_2) + R_{12} > I(W;Y_1)$.

2) Gaussian BC Example: Consider next the cooperative Gaussian PD-BC (without a common message) shown in Fig. 8, where for every time instance $i \in [1:n]$, we have

$$Y_{1,i} = X_i + Z_{1,i}, \tag{25a}$$

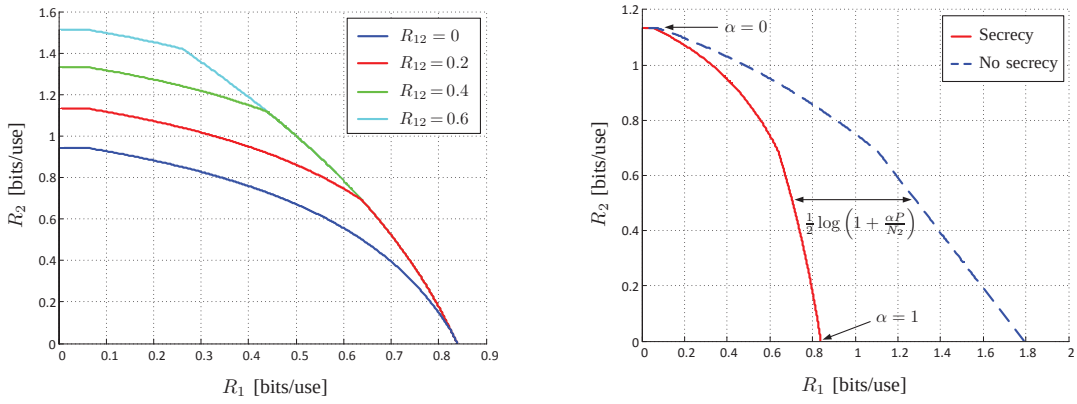
$$Y_{2,i} = X_i + Z_{1,i} + Z_{2,i} \tag{25b}$$

and $\big\{Z_{1,i}\big\}_{i=1}^n$ and $\big\{Z_{2,i}\big\}_{i=1}^n$ are mutually independent sequences of i.i.d. Gaussian random variables with $Z_{1,i} \sim \mathcal{N}(0, N_1)$, $Z_{2,i} \sim \mathcal{N}(0, N_2 - N_1)$ and $N_2 > N_1$, for $i \in [1:n]$. The channel input is subject to an average power constraint

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\big[X_i\big] \leq P. \tag{26}$$

By using continuous alphabets with an input power constraint adaptation of Theorem 5 we characterize the strong-secrecy-capacity region of the cooperative Gaussian PD-BC with one confidential message as

$$\mathcal{C}_{\mathrm{S}}^{(\mathrm{G})} = \bigcup_{\alpha \in [0,1]} \left\{ (R_{12}, R_1, R_2) \in \mathbb{R}_+^3 \;\middle|\; \begin{array}{l} R_1 \leq \frac{1}{2}\log\left(1 + \frac{\alpha P}{N_1}\right) - \frac{1}{2}\log\left(1 + \frac{\alpha P}{N_2}\right) \\ R_2 \leq \frac{1}{2}\log\left(1 + \frac{\bar{\alpha} P}{\alpha P + N_2}\right) + R_{12} \\ R_1 + R_2 \leq \frac{1}{2}\log\left(1 + \frac{P}{N_1}\right) - \frac{1}{2}\log\left(1 + \frac{\alpha P}{N_2}\right) \end{array} \right\}. \tag{27}$$



(a) (b)

Fig. 9: (a) Projection of the strong-secrecy-capacity region of the cooperative Gaussian BC with one confidential message onto the plane (R_1, R_2) for different values of R_{12}; (b) Cooperative Gaussian BC with $R_{12} = 0.2$: Strong-secrecy-capacity region where M_1 is confidential vs. capacity region without secrecy.

The achievability of (27) follows from Theorem 5 with the following choice of random variables:

$$W \sim \mathcal{N}(0, \alpha P) \ , \ \tilde{W} \sim \mathcal{N}(0, \bar{\alpha} P) \ , \ X = W + \tilde{W} \tag{28}$$

where W and $\tilde{W}$ are independent. The optimality of Gaussian inputs is proven in Appendix B.

Setting $P = 11$, $N_1 = 1$ and $N_2 = 4$, Fig. 9(a) shows the strong-secrecy-capacity region of the cooperative Gaussian BC for different R_{12} values, while Fig. 9(b) compares the optimal rate regions when a secrecy constraint on M_1 is and is not present. The red line in both figures coincide and represent the secrecy-capacity region when $R_{12} = 0.2$. The dashed blue line in Fig 9(b) shows the capacity region of the cooperative Gaussian BC without secrecy constraints, which is given by

$$\mathcal{C}_{\mathrm{NS}}^{(\mathrm{G})} = \bigcup_{\alpha \in [0,1]} \left\{ (R_{12}, R_1, R_2) \in \mathbb{R}_+^3 \left| \begin{array}{l} R_1 \leq \frac{1}{2} \log\left(1 + \frac{\alpha P}{N_1}\right) \\ R_2 \leq \frac{1}{2} \log\left(1 + \frac{\bar{\alpha} P}{\alpha P + N_2}\right) + R_{12} \\ R_1 + R_2 \leq \frac{1}{2} \log\left(1 + \frac{P}{N_1}\right) \end{array} \right. \right\}. \tag{29}$$

The derivation of (29) relies on [43, Eq. (17)] and uses standard arguments for proving the optimality of Gaussian inputs.

By the structure of the rate bounds in (27) and (29), for every fixed $\alpha \in [0, 1]$, if $(R_{12}, R_1, R_2) \in \mathcal{C}_{\mathrm{NS}}^{(\mathrm{G})}$, we have

$$\left(R_{12}, R_1 - \frac{1}{2} \log\left(1 + \frac{\alpha P}{N_2}\right), R_2 \right) \in \mathcal{C}_{\mathrm{S}}^{(\mathrm{G})}. \tag{30}$$

This agrees with the discussion in Section VI-B1 as $I(X; Y_2|W) = \frac{1}{2} \log\left(1 + \frac{\alpha P}{N_2}\right)$.

VII. PROOFS

A. Proof of Lemma 1

Note that the factorization in (7) implies that $P^{(\mathcal{B}_n)}_{\mathbf{S}_0,\mathbf{S}} = Q^n_{S_0,S}$, for every codebook $\mathcal{B}_n$. Therefore, to establish Lemma 1 we show that

$$\mathbb{E}_{\mathbb{B}_n} D\Big(P^{(\mathbb{B}_n)}_{\mathbf{S}_0,\mathbf{S},\mathbf{V}}\Big\|Q^n_{S_0,S,V}\Big) \xrightarrow[n\to\infty]{} 0. \tag{31}$$

Lemma 10 (Absolute Continuity) *For any fixed codebook $\mathcal{B}_n$, we have $P^{(\mathcal{B}_n)}_{\mathbf{S}_0,\mathbf{S},\mathbf{V}} \ll Q^n_{S_0,S,V}$, i.e., $P^{(\mathcal{B}_n)}_{\mathbf{S}_0,\mathbf{S},\mathbf{V}}$ is absolutely continues with respect to $Q^n_{S_0,S,V}$.*

The proof of Lemma 10 is relegated to Appendix C. Combining this with Remark 1, a sufficient condition for (31) is that

$$\mathbb{E}_{\mathbb{B}_n}\Big\|P^{(\mathbb{B}_n)}_{\mathbf{S}_0,\mathbf{S},\mathbf{V}} - Q^n_{S_0,S,V}\Big\| \xrightarrow[n\to\infty]{} 0. \tag{32}$$

To evaluate the TV distance in (32), define the *ideal* PMF on $\mathcal{S}_0^n \times \mathcal{S}^n \times \mathcal{W} \times \mathcal{I} \times \mathcal{U}^n \times \mathcal{V}^n$ as

$$\Gamma^{(\mathcal{B}_n)}(\mathbf{s}_0, w, i, \mathbf{u}, \mathbf{s}, \mathbf{v}) = Q^n_{S_0}(\mathbf{s}_0) 2^{-n(\tilde{R}+R')} \mathbb{1}_{\left\{\mathbf{u}(\mathbf{s}_0,w,i,\mathcal{B}_n)=\mathbf{u}\right\}} Q^n_{S|U,S_0}(\mathbf{s}|\mathbf{u},\mathbf{s}_0) Q^n_{V|U,S_0,S}(\mathbf{v}|\mathbf{u},\mathbf{s}_0,\mathbf{s}) \tag{33}$$

with respect to the same codebook $\mathcal{B}_n$ as $P^{(\mathcal{B}_n)}$. Note, however, that Γ describes an encoding process where the choice of the u-codeword from a certain bin is uniform, as opposed to P that uses the likelihood encoder. Furthermore, the structure of Γ implies that the sequence $\mathbf{s}$ is generated by feeding $\mathbf{s}_0$ and the chosen u-codeword into the DMC $Q^n_{S|U,S_0}$.

Using the TV distance triangle inequality, we upper bound the left-hand side (LHS) of (32) by

$$\mathbb{E}_{\mathbb{B}_n}\Big\|P^{(\mathbb{B}_n)}_{\mathbf{S}_0,\mathbf{S},\mathbf{V}} - Q^n_{S_0,S,V}\Big\|_{TV} \le \mathbb{E}_{\mathbb{B}_n}\Big\|P^{(\mathbb{B}_n)}_{\mathbf{S}_0,\mathbf{S},\mathbf{V}} - \Gamma^{(\mathbb{B}_n)}_{\mathbf{S}_0,\mathbf{S},\mathbf{V}}\Big\|_{TV} + \mathbb{E}_{\mathbb{B}_n}\Big\|\Gamma^{(\mathbb{B}_n)}_{\mathbf{S}_0,\mathbf{S},\mathbf{V}} - Q^n_{S_0,S,V}\Big\|_{TV}. \tag{34}$$

By [25, Corollary VII.5], the second expected TV distance on the RHS of (34) decays exponentially fast as $n \to \infty$ if

$$\tilde{R} + R' > I(U; S, V|S_0). \tag{35}$$

For the first term in (34), we use the following relations between Γ and P. For every fixed codebook $\mathcal{B}_n$, we have

$$\Gamma^{(\mathcal{B}_n)}_{I|W,\mathbf{S}_0,\mathbf{S}} = f^{(\mathrm{LE})}_{I|W,\mathbf{S}_0,\mathbf{S},\mathbb{B}_n=\mathcal{B}_n} = P^{(\mathcal{B}_n)}_{I|W,\mathbf{S}_0,\mathbf{S}} \tag{36a}$$

$$\Gamma^{(\mathcal{B}_n)}_{\mathbf{U}|I,W,\mathbf{S}_0,\mathbf{S}} = \mathbb{1}_{\left\{\mathbf{U}=\mathbf{U}(\mathbf{S}_0,W,I,\mathcal{B}_n)\right\}} = P^{(\mathcal{B}_n)}_{\mathbf{U}|I,W,\mathbf{S}_0,\mathbf{S}} \tag{36b}$$

$$\Gamma^{(\mathcal{B}_n)}_{\mathbf{V}|\mathbf{U},I,W,\mathbf{S}_0,\mathbf{S}} = Q^n_{V|U,S_0,S} = P^{(\mathcal{B}_n)}_{\mathbf{V}|\mathbf{U},I,W,\mathbf{S}_0,\mathbf{S}}. \tag{36c}$$

While (36b)-(36c) follow directly from (7) and (33), the justification for (36a) is that for every $(\mathbf{s}_0, \mathbf{s}, w, i) \in \mathcal{S}_0^n \times \mathcal{S}^n \times \mathcal{W} \times \mathcal{I}$, we have

$$\begin{aligned}
\Gamma^{(\mathcal{B}_n)}(i|w, \mathbf{s}_0, \mathbf{s}) &= \frac{\Gamma^{(\mathcal{B}_n)}(\mathbf{s}_0, w, i, \mathbf{s})}{\Gamma^{(\mathcal{B}_n)}(\mathbf{s}_0, w, \mathbf{s})} \\
&= \frac{\sum_{\mathbf{u}} Q_{S_0}^n(\mathbf{s}_0) 2^{-n(\tilde{R}+R')} \mathbb{1}_{\left\{\mathbf{u}(\mathbf{s}_0, w, i, \mathcal{B}_n) = \mathbf{u}\right\}} Q_{S|U,S_0}^n(\mathbf{s}|\mathbf{u}, \mathbf{s}_0)}{\sum_{\mathbf{u}, i'} Q_{S_0}^n(\mathbf{s}_0) 2^{-n(\tilde{R}+R')} \mathbb{1}_{\left\{\mathbf{u}(\mathbf{s}_0, w, i', \mathcal{B}_n) = \mathbf{u}\right\}} Q_{S|U,S_0}^n(\mathbf{s}|\mathbf{u}, \mathbf{s}_0)} \\
&= \frac{Q_{S|U,S_0}^n\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0, w, i, \mathcal{B}_n), \mathbf{s}_0\big)}{\sum_{i'} Q_{S|U,S_0}^n\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0, w, i', \mathcal{B}_n), \mathbf{s}_0\big)} \\
&\overset{(a)}{=} f^{(\mathrm{LE})}(i|w, \mathbf{s}_0, \mathbf{s}, \mathcal{B}_n)
\end{aligned} \tag{37}$$

where (a) follows from (6). The relations in (36) yield

$$\begin{aligned}
\mathbb{E}_{\mathbb{B}_n} \Big|\Big| P_{\mathbf{S}_0,\mathbf{S},\mathbf{V}}^{(\mathbb{B}_n)} - \Gamma_{\mathbf{S}_0,\mathbf{S},\mathbf{V}}^{(\mathbb{B}_n)} \Big|\Big|_{TV} &\leq \mathbb{E}_{\mathbb{B}_n} \Big|\Big| P_{\mathbf{S}_0,\mathbf{S},W,I,\mathbf{U},\mathbf{V}}^{(\mathbb{B}_n)} - \Gamma_{\mathbf{S}_0,\mathbf{S},W,I,\mathbf{U},\mathbf{V}}^{(\mathbb{B}_n)} \Big|\Big|_{TV} \\
&\overset{(a)}{=} \sum_w 2^{-n\tilde{R}} \mathbb{E}_{\mathbb{B}_n} \Big|\Big| P_{\mathbf{S}_0,\mathbf{S},I,\mathbf{U},\mathbf{V}|W=w}^{(\mathbb{B}_n)} - \Gamma_{\mathbf{S}_0,\mathbf{S},I,\mathbf{U},\mathbf{V}|W=w}^{(\mathbb{B}_n)} \Big|\Big|_{TV} \\
&\overset{(b)}{=} \mathbb{E}_{\mathbb{B}_n} \Big|\Big| P_{\mathbf{S}_0,\mathbf{S},I,\mathbf{U},\mathbf{V}|W=1}^{(\mathbb{B}_n)} - \Gamma_{\mathbf{S}_0,\mathbf{S},I,\mathbf{U},\mathbf{V}|W=1}^{(\mathbb{B}_n)} \Big|\Big|_{TV} \\
&\overset{(c)}{=} \mathbb{E}_{\mathbb{B}_n} \Big|\Big| Q_{S_0,S}^n - \Gamma_{\mathbf{S}_0,\mathbf{S}|W=1}^{(\mathbb{B}_n)} \Big|\Big|_{TV}
\end{aligned} \tag{38}$$

where:

(a) is because $\Gamma^{(\mathcal{B}_n)}(w) = P^{(\mathcal{B}_n)}(w) = 2^{-n\tilde{R}}$ for every $w \in \mathcal{W}$ and fixed $\mathcal{B}_n$, and since $\mathbb{B}_n$ is independent of W;

(b) uses the symmetry of the code construction;

(c) is by (36) and because $P_{\mathbf{S}_0,\mathbf{S}}^{(\mathcal{B}_n)} = Q_{S_0,S}^n$ for every fixed $\mathcal{B}_n$.

Invoking [25, Corollary VII.5] once more yields

$$\mathbb{E}_{\mathbb{B}_n} \Big|\Big| Q_{S_0,S}^n - \Gamma_{\mathbf{S}_0,\mathbf{S}|W=1}^{(\mathbb{B}_n)} \Big|\Big|_{TV} \xrightarrow[n\to\infty]{} 0 \tag{39}$$

exponentially fast, as long as

$$R' > I(U; S|S_0). \tag{40}$$

This implies that there exists $c > 0$ such that

$$\mathbb{E}_{\mathbb{B}_n} \Big|\Big| P_{\mathbf{S}_0,\mathbf{S},\mathbf{V}}^{(\mathbb{B}_n)} - Q_{S_0,S,V}^n \Big|\Big|_{TV} \leq e^{-cn}. \tag{41}$$

B. Proof of Lemma 2

This proof uses the following property of the TV distance (see, e.g., [28, Property 1]): Let $\epsilon > 0$ and let $f : \mathcal{X} \to \mathbb{R}$ be a function bounded by $b \in \mathbb{R}$. We have

$$\big|\big|\Pi - \Lambda\big|\big|_{TV} < \epsilon \implies \Big|\mathbb{E}_\Pi f(X) - \mathbb{E}_\Lambda f(X)\Big| < \epsilon b. \tag{42}$$

Fix $\epsilon > 0$ and consider the Γ PMF defined in (33). With respect to the random experiment described by Γ, we have

$$\mathbb{E}_{\mathbb{B}_n}\mathbb{P}_\Gamma\Big(\big(\mathbf{S}_0,\mathbf{S},\mathbf{U}(\mathbf{S}_0,w,I,\mathbb{B}_n)\big)\notin\mathcal{T}_\epsilon^{(n)}(Q_{S_0,S,U})\Big|\mathbb{B}_n\Big)\xrightarrow[n\to\infty]{}0 \tag{43}$$

because $\mathbf{U}(\mathbf{S}_0,w,i,\mathbb{B}_n)\sim Q_{U|S_0}^n$, for every $i\in\mathcal{I}$, and $\mathbf{S}$ is obtained by feeding $\big(\mathbf{S}_0,\mathbf{U}(\mathbf{S}_0,w,i,\mathbb{B}_n)\big)$ into a DMC $Q_{S|U,S_0}^n$. Thus, (43) holds by the law of large numbers (LLN). Further, based on the analysis in Section VII-A, we have

$$\mathbb{E}_{\mathbb{B}_n}\Big|\Big|P_{\mathbf{S}_0,\mathbf{S},\mathbf{V}}^{(\mathbb{B}_n)}-\Gamma_{\mathbf{S}_0,\mathbf{S},\mathbf{V}}^{(\mathbb{B}_n)}\Big|\Big|_{TV}\xrightarrow[n\to\infty]{}0. \tag{44}$$

Finally, let $f:\mathcal{S}_0^n\times\mathcal{S}^n\times\mathcal{U}^n\to\mathbb{R}$ be defined by $f(\mathbf{s}_0,\mathbf{s},\mathbf{u})\triangleq\mathbb{1}_{\left\{(\mathbf{s}_0,\mathbf{s},\mathbf{u})\notin\mathcal{T}_\epsilon^{(n)}(Q_{S_0,S,U})\right\}}$ and consider

$$\begin{aligned}
&\mathbb{E}_{\mathbb{B}_n}\mathbb{P}_P\Big(\big(\mathbf{S}_0,\mathbf{S},\mathbf{U}(\mathbf{S}_0,w,I,\mathbb{B}_n)\big)\notin\mathcal{T}_\epsilon^{(n)}(Q_{S_0,S,U})\Big|\mathbb{B}_n\Big)\\
&\quad=\mathbb{E}_{\mathbb{B}_n}\mathbb{E}_P\Big[f\big(\mathbf{S}_0,\mathbf{S},\mathbf{U}(\mathbf{S}_0,w,I,\mathbb{B}_n)\big)\Big|\mathbb{B}_n\Big]\\
&\quad\le\mathbb{E}_{\mathbb{B}_n}\mathbb{E}_\Gamma\Big[f\big(\mathbf{S}_0,\mathbf{S},\mathbf{U}(\mathbf{S}_0,w,I,\mathbb{B}_n)\big)\Big|\mathbb{B}_n\Big]\\
&\qquad\qquad+\mathbb{E}_{\mathbb{B}_n}\Big|\mathbb{E}_P\Big[f\big(\mathbf{S}_0,\mathbf{S},\mathbf{U}(\mathbf{S}_0,w,I,\mathbb{B}_n)\big)\Big|\mathbb{B}_n\Big]-\mathbb{E}_\Gamma\Big[f\big(\mathbf{S}_0,\mathbf{S},\mathbf{U}(\mathbf{S}_0,w,I,\mathbb{B}_n)\big)\Big|\mathbb{B}_n\Big]\Big|\\
&\quad\overset{(a)}{\le}\mathbb{E}_{\mathbb{B}_n}\mathbb{P}_\Gamma\Big(\big(\mathbf{S}_0,\mathbf{S},\mathbf{U}(\mathbf{S}_0,w,I,\mathbb{B}_n)\big)\notin\mathcal{T}_\epsilon^{(n)}(Q_{S_0,S,U})|\mathbb{B}_n\Big)+\mathbb{E}_{\mathbb{B}_n}\Big|\Big|P_{\mathbf{S}_0,\mathbf{S},\mathbf{V}}^{(\mathbb{B}_n)}-\Gamma_{\mathbf{S}_0,\mathbf{S},\mathbf{V}}^{(\mathbb{B}_n)}\Big|\Big|_{TV}
\end{aligned} \tag{45}$$

where (a) uses (42) and the fact that f is bounded by $b=1$. By (43)-(44), the RHS of (45) approaches 0 as $n\to\infty$.

C. Proof of Theorem 3

Fix a PMF $Q_{U_0,U_1,U_2,X}Q_{Y_1,Y_2|X}$ and $\epsilon>0$.

Codebook Generation: Split each $m_2\in\mathcal{M}_2$ into two sub-messages denoted by (m_{20},m_{22}). The pair $m_p\triangleq(m_0,m_{20})$ is referred to as a *public message* and is to be decoded by both receivers, while m_1 and m_{22}, that serve as *private messages*, are to be decoded by receiver 1 and receiver 2, respectively. The cooperation protocol will use the link to convey information about the decoded m_p from receiver 1 to receiver 2. The rates associated with m_{20} and m_{22} are denoted by R_{20} and R_{22}, while the corresponding alphabets are $\mathcal{M}_{20}$ and $\mathcal{M}_{22}$, respectively. Furthermore, we use $R_p\triangleq R_0+R_{20}$ and $\mathcal{M}_p\triangleq\mathcal{M}_0\times\mathcal{M}_{20}=[1:2^{nR_p}]$. The partial rates R_{20} and R_{22} satisfy

$$R_2=R_{20}+R_{22}. \tag{46}$$

The random variables M_{20} and M_{22} are independent and uniform over $\mathcal{M}_{20}$ and $\mathcal{M}_{22}$, and we set $M_p\triangleq(M_0,M_{20})$. Note that M_p is uniformly distributed over $\mathcal{M}_p$. Moreover, let W be a random variable uniformly distributed over $\mathcal{W}=[1:2^{n\tilde{R}}]$ and independent of (M_0,M_1,M_2) (which is therefore also independent of (M_p,M_1,M_{22})).

Generate a public message codebook[3] $\mathcal{C}_0$ that comprises 2^{nR_p} u_0-codewords $\mathbf{u}_0(m_p,\mathcal{C}_0)$, $m_p\in\mathcal{M}_p$, each drawn according to $Q_{U_0}^n$ independent of all the other u_0-codewords. Partition $\mathcal{M}_p$ into $2^{nR_{12}}$ equal-sized subsets (referred

[3]The subsequent notations for codebooks omit the blocklength n.

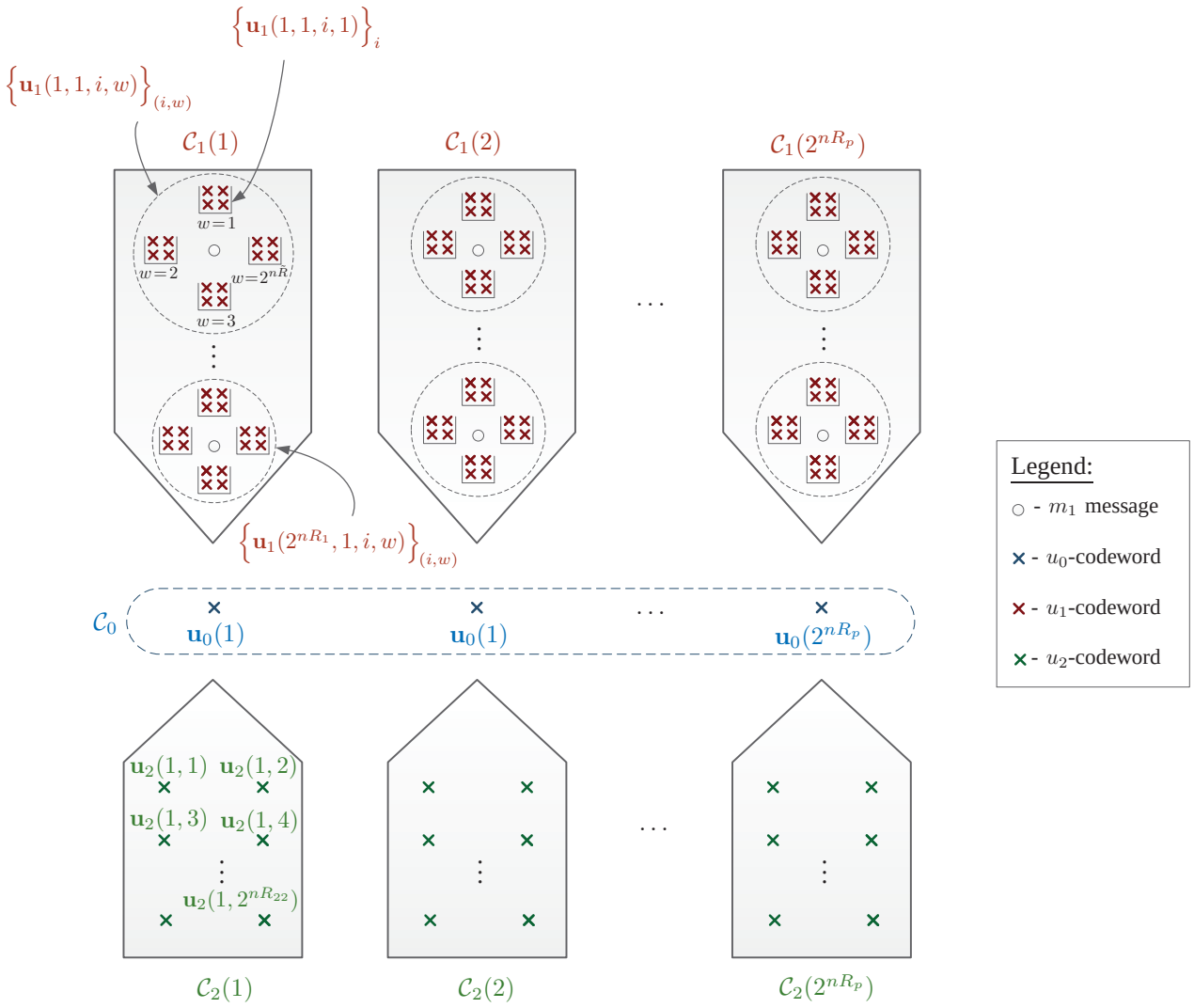


Fig. 10: Codebook structure.

to as "bins") $\mathcal{B}(m_{12})$, where $m_{12} \in \mathcal{M}_{12}$.

For each $m_p \in \mathcal{M}_p$, generate a codebook $\mathcal{C}_1(m_p)$ that comprises $2^{n(R_1+R_1'+\tilde{R})}$ codewords $\mathbf{u}_1$, each drawn according to $Q^n_{U_1|U_0}\big(\,\cdot\,\big|\mathbf{u}_0(m_p,\mathcal{C}_0)\big)$ independent of all the other u_1-codewords. Label these codewords as $\mathbf{u}_1(m_p,m_1,i,w,\mathcal{C}_1)$, where $(m_1,i,w) \in \mathcal{M}_1 \times \mathcal{I} \times \mathcal{W}$ and $\mathcal{I} \triangleq [1:2^{nR_1'}]$. Based on this labeling, the codebook $\mathcal{C}_1(m_p)$ has a u_1-bin associated with every pair $(m_1,w) \in \mathcal{M}_1 \times \mathcal{W}_1$, each containing $2^{nR'}$ u_1-codewords.

For each $m_p \in \mathcal{M}_p$ also generate a codebook $\mathcal{C}_2(m_p)$ that comprises $2^{nR_{22}}$ u_2-codewords, each associated with a private message $m_{22} \in \mathcal{M}_{22}$. Each u_2-codeword is drawn according to $Q^n_{U_2|U_0}\big(\,\cdot\,\big|\mathbf{u}_0(m_p,\mathcal{C}_0)\big)$ independently of all the other u_2-codewords. Denote $\mathcal{C}_2(m_p) \triangleq \big\{\mathbf{u}_2(m_p,m_{22},\mathcal{C}_2)\big\}_{m_{22}\in\mathcal{M}_{22}}$.

The channel input $\mathbf{x}$ associated with a triple $(\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2)$ is generated according to $Q^n_{X|U_0,U_1,U_2}\big(\,\cdot\,\big|\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2\big)$. The structure of the codebook is illustrated in Fig. 10 (for simplicity, the figure does not show the binning of $\mathcal{M}_p$).

Encoding: To transmit a triple (m_0,m_1,m_2), the encoder transforms it into the triple (m_p,m_1,m_{22}), and draws

W uniformly over $\mathcal{W}$. Then, an index $i \in \mathcal{I}$ is chosen by the likelihood encoder, i.e., according to the PMF

$$f_{\mathrm{BC}}^{(\mathrm{LE})}\big(i\big|w, \mathbf{u}_0(m_p, \mathcal{C}_0), \mathbf{u}_2(m_p, m_{22}, \mathcal{C}_2), \mathcal{C}_1\big) = \frac{Q^n_{U_2|U_1,U_0}\big(\mathbf{u}_2(m_p, m_{22}, \mathcal{C}_2)\big|\mathbf{u}_1(m_p, m_1, i, w, \mathcal{C}_1), \mathbf{u}_0(m_p, \mathcal{C}_0)\big)}{\sum_{i' \in \mathcal{I}} Q^n_{U_2|U_1,U_0}\big(\mathbf{u}_2(m_p, m_{22}, \mathcal{C}_2)\big|\mathbf{u}_1(m_p, m_1, i', w, \mathcal{C}_1), \mathbf{u}_0(m_p, \mathcal{C}_0)\big)}. \tag{47}$$

The corresponding sequence $\mathbf{x}$ is transmitted over the BC.

Decoding and Cooperation: Decoder 1: Searches for a unique triple $(\hat{m}_p, \hat{m}_1, \hat{w}) \in \mathcal{M}_p \times \mathcal{M}_1 \times \mathcal{W}$, for which there is an index $\hat{i} \in \mathcal{I}$ such that

$$\Big(\mathbf{u}_0(\hat{m}_p, \mathcal{C}_0), \mathbf{u}_1(\hat{m}_p, \hat{m}_1, \hat{i}, \hat{w}, \mathcal{C}_1), \mathbf{y}_1\Big) \in \mathcal{T}_\epsilon^{(n)}(Q_{U_0,U_1,Y_1}). \tag{48}$$

Upon finding such a unique triple, $(\hat{m}_0^{(1)}, \hat{m}_1) = (\hat{m}_0, \hat{m}_1)$ is declared as the decoded message pair; otherwise, an error is declared.

Cooperation: Having $(\hat{m}_p, \hat{m}_1, \hat{i}, \hat{w})$, Decoder 1 conveys the bin number of $\hat{m}_p$ to Decoder 2 via the cooperation link. That is, Decoder 1 shares with Decoder 2 the index $\hat{m}_{12} \in \mathcal{M}_{12}$ such that $\hat{m}_p \in \mathcal{B}(\hat{m}_{12})$.

Decoder 2: Upon receiving $(\hat{m}_{12}, \mathbf{y}_2)$, Decoder 2 searches for a unique pair $(\hat{\hat{m}}_p, \hat{\hat{m}}_{22}) \in \mathcal{M}_p \times \mathcal{M}_{22}$, such that

$$\Big(\mathbf{u}_0(\hat{\hat{m}}_p, \mathcal{C}_0), \mathbf{u}_2(\hat{\hat{m}}_p, \hat{\hat{m}}_{22}, \mathcal{C}_2), \mathbf{y}_2\Big) \in \mathcal{T}_\epsilon^{(n)}(Q_{U_0,U_2,Y_2}) \tag{49}$$

where $\hat{\hat{m}}_p \in \mathcal{B}(\hat{m}_{12})$. If such a unique pair is found, then $(\hat{m}_0^{(2)}, \hat{m}_2) = \big(\hat{\hat{m}}_0, (\hat{\hat{m}}_{20}, \hat{\hat{m}}_{22})\big)$ is declared as the decoded message; otherwise an error is declared.

The error probability analysis is given in Appendix D and uses Lemma 2 to first show that the above encoding process results in u_0-, u_1- and u_2-sequences that are jointly typical. Then, by standard joint-typicality decoding arguments, reliability is established provided that

$$R' > I(U_1; U_2|U_0) \tag{50a}$$

$$R' + \tilde{R} > I(U_1; U_2, Y_2|U_0) \tag{50b}$$

$$R_1 + R' + \tilde{R} < I(U_1; Y_1|U_0) \tag{50c}$$

$$R_0 + R_{20} + R_1 + R' + \tilde{R} < I(U_0, U_1; Y_1) \tag{50d}$$

$$R_{22} < I(U_2; Y_2|U_0) \tag{50e}$$

$$R_0 + R_2 - R_{12} < I(U_0, U_2; Y_2). \tag{50f}$$

Security Analysis: Let $\mathbb{C}_0$ represent a random public message codebook that adheres to the above construction. Furthermore, let $\mathcal{C}_j \triangleq \{\mathcal{C}_j(m_p)\}_{m_p \in \mathcal{M}_p}$, for $j = 1, 2$, be the private message codebooks 1 and 2, and $\mathbb{C}_1$ and $\mathbb{C}_2$ be the corresponding random codebooks. With some abuse of notation, we use $\mathcal{C} \triangleq (\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2)$ and $\mathbb{C} \triangleq (\mathbb{C}_0, \mathbb{C}_1, \mathbb{C}_2)$. When clear from the context, we omit the functional dependencies of the u_j-codewords, $j = 0, 1, 2$, on the corresponding indices and codebooks, e.g., we write $\mathbf{U}_2$ instead of $\mathbf{U}_2(M_p, M_{22}, \mathcal{C}_2)$.

For any fixed $\mathbb{C}=\mathcal{C}$, let the joint distribution induced by the code be

$$\begin{aligned}
&P^{(\mathcal{C})}(w,m_p,m_1,m_{22},m_{12},\mathbf{u}_0,\mathbf{u}_2,i,\mathbf{u}_1,\mathbf{x},\mathbf{y}_1,\mathbf{y}_2)\\
&\qquad=2^{-n(\tilde{R}+R_p+R_1+R_{22})}\mathbb{1}_{\left\{m_p\in\mathcal{B}(m_{12})\right\}\cap\left\{\mathbf{u}_0=\mathbf{u}_0(m_p,\mathcal{C}_0)\right\}\cap\left\{\mathbf{u}_2=\mathbf{u}_2(m_p,m_{22},\mathcal{C}_2)\right\}}\\
&\qquad\times f_{\mathrm{BC}}^{(\mathrm{LE})}\big(i\big|w,\mathbf{u}_0(m_p,\mathcal{C}_0),\mathbf{u}_2(m_p,m_{22},\mathcal{C}_2),\mathcal{C}_1\big)\mathbb{1}_{\left\{\mathbf{u}_1=\mathbf{u}_1(m_p,m_1,i,w,\mathcal{C}_1)\right\}}\\
&\qquad\times Q^n_{X|U_0,U_1,U_2}(\mathbf{x}|\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2)Q^n_{Y_1,Y_2|X}(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x}).
\end{aligned}\tag{51}$$

Accordingly, we have

$$\begin{aligned}
I_{P^{(\mathcal{C})}}(M_1;M_{12},\mathbf{Y}_2)&\le I_{P^{(\mathcal{C})}}(M_1;M_{12},M_p,M_{22},\mathbf{Y}_2)\\
&\overset{(a)}{=} I_{P^{(\mathcal{C})}}\big(M_1;\mathbf{Y}_2|M_p,M_{22},\mathbf{U}_0(M_p,\mathcal{C}_0),\mathbf{U}_2(M_p,M_{22},\mathcal{C}_2)\big)\\
&\overset{(b)}{=} D\Big(P^{(\mathcal{C})}_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|\Big|P^{(\mathcal{C})}_{\mathbf{Y}_2|M_p,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|P^{(\mathcal{C})}_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big)\\
&\overset{(c)}{\le} D\Big(P^{(\mathcal{C})}_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|\Big|Q^n_{Y_2|U_0,U_2}\Big|P^{(\mathcal{C})}_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big)
\end{aligned}\tag{52}$$

where:

(a) is because M_1 is independent of (M_p,M_{22}), and since M_{12}, $\mathbf{U}_0(M_p,\mathcal{C}_0)$ and $\mathbf{U}_2(M_p,M_{22},\mathcal{C}_2)$ are defined by (M_p,M_{22});

(b) uses the relative entropy chain rule and the independence of M_1 and $\big(M_p,M_{22},\mathbf{U}_0(M_p,\mathcal{C}_0),\mathbf{U}_2(M_p,M_{22},\mathcal{C}_2)\big)$;

(c) is since for every fixed $\mathcal{C}$, we have

$$\begin{aligned}
&D\Big(P^{(\mathcal{C})}_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|\Big|P^{(\mathcal{C})}_{\mathbf{Y}_2|M_p,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|P^{(\mathcal{C})}_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big)\\
&\quad=\sum_{m_p,m_1,m_{22},\mathbf{u}_0,\mathbf{u}_2,\mathbf{y}_2}P^{(\mathcal{C})}(m_p,m_1,m_{22},\mathbf{u}_0,\mathbf{u}_2,\mathbf{y}_2)\\
&\qquad\qquad\times\log\left(\frac{P^{(\mathcal{C})}(\mathbf{y}_2|m_p,m_1,m_{22},\mathbf{u}_0,\mathbf{u}_2)}{P^{(\mathcal{C})}(\mathbf{y}_2|m_{22},\mathbf{u}_0,\mathbf{u}_2)}\cdot\frac{Q^n_{Y_2|U_0,U_2}(\mathbf{y}_2|\mathbf{u}_0,\mathbf{u}_2)}{Q^n_{Y_2|U_0,U_2}(\mathbf{y}_2|\mathbf{u}_0,\mathbf{u}_2)}\right)\\
&\quad=D\Big(P^{(\mathcal{C})}_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|\Big|Q^n_{Y_2|U_0,U_2}\Big|P^{(\mathcal{C})}_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big)\\
&\qquad\qquad+\sum_{m_p,m_1,m_{22},\mathbf{u}_0,\mathbf{u}_2,\mathbf{y}_2}P^{(\mathcal{C})}(m_p,m_1,m_{22},\mathbf{u}_0,\mathbf{u}_2,\mathbf{y}_2)\log\left(\frac{Q^n_{Y_2|U_0,U_2}(\mathbf{y}_2|\mathbf{u}_0,\mathbf{u}_2)}{P^{(\mathcal{C})}(\mathbf{y}_2|m_p,m_{22},\mathbf{u}_0,\mathbf{u}_2)}\right)\\
&\quad=D\Big(P^{(\mathcal{C})}_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|\Big|Q^n_{Y_2|U_0,U_2}\Big|P^{(\mathcal{C})}_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big)\\
&\qquad\qquad-D\Big(P^{(\mathcal{C})}_{\mathbf{Y}_2|M_p,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|\Big|Q^n_{Y_2|U_0,U_2}\Big|P^{(\mathcal{C})}_{M_p,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big).
\end{aligned}$$

By (52), to satisfy (11b) it suffices to show that there is a sufficiently large n for which

$$D\Big(P^{(\mathcal{C})}_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|\Big|Q^n_{Y_2|U_0,U_2}\Big|P^{(\mathcal{C})}_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big)\le\epsilon.\tag{53}$$

Taking the expectation of the RHS of (52) over the ensemble of codebooks, we have

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{C}} D\Big(P^{(\mathbb{C})}_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|\Big|Q^n_{Y_2|U_0,U_2}\Big|P^{(\mathbb{C})}_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big)\\
&\quad= \mathbb{E}_{\mathbb{C}}\Bigg[\sum_{m_0,m_1,m_2,\mathbf{u}_0,\mathbf{u}_2} 2^{-n(R_p+R_1+R_{22})}\mathbb{1}_{\left\{\left(\mathbf{U}_0(m_p,\mathbb{C}_0),\mathbf{U}_0(m_p,m_{22},\mathbb{C}_2)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}}\\
&\qquad\qquad\times D\Big(P^{(\mathbb{C})}_{\mathbf{Y}_2|M_p=m_p,M_1=m_1,M_{22}=m_{22},\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2}\Big|\Big|Q^n_{Y_2|U_0=\mathbf{u}_0,U_2=\mathbf{u}_2}\Big)\Bigg]\\
&\quad\overset{(a)}{=} \sum_{\mathbf{u}_0,\mathbf{u}_2}\mathbb{E}_{\mathbb{C}}\Bigg[\mathbb{1}_{\left\{\left(\mathbf{U}_0(1,\mathbb{C}_0),\mathbf{U}_2(1,1,\mathbb{C}_2)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}}\\
&\qquad\qquad\times D\Big(P^{(\mathbb{C})}_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2}\Big|\Big|Q^n_{Y_2|U_0=\mathbf{u}_0,U_2=\mathbf{u}_2}\Big)\Bigg]\\
&\quad\overset{(b)}{=} \sum_{\mathbf{u}_0,\mathbf{u}_2}\mathbb{E}_{\mathbb{C}_1}\Bigg[\mathbb{E}_{\mathbb{C}_0,\mathbb{C}_2|\mathbb{C}_1}\Bigg[\mathbb{1}_{\left\{\left(\mathbf{U}_0(1,\mathbb{C}_0),\mathbf{U}_2(1,1,\mathbb{C}_2)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}}\\
&\qquad\qquad\times D\Big(P^{(\mathbb{C})}_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2}\Big|\Big|Q^n_{Y_2|U_0=\mathbf{u}_0,U_2=\mathbf{u}_2}\Big)\Big|\mathbb{C}_1\Bigg]\Bigg]\\
&\quad\overset{(c)}{=} \sum_{\mathbf{u}_0,\mathbf{u}_2}\mathbb{E}_{\mathbb{C}_1}\Bigg[D\Big(P^{(\mathbb{C}_1)}_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2}\Big|\Big|Q^n_{Y_2|U_0=\mathbf{u}_0,U_2=\mathbf{u}_2}\Big)\\
&\qquad\qquad\times\mathbb{E}_{\mathbb{C}_0,\mathbb{C}_2|\mathbb{C}_1}\Big[\mathbb{1}_{\left\{\left(\mathbf{U}_0(1,\mathbb{C}_0),\mathbf{U}_2(1,1,\mathbb{C}_2)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}}\Big|\mathbb{C}_1\Big]\Bigg]\\
&\quad\overset{(d)}{=} \sum_{\mathbf{u}_0,\mathbf{u}_2}\mathbb{E}_{\mathbb{C}_1}\Bigg[D\Big(P^{(\mathbb{C}_1)}_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathbb{C}_1}\Big|\Big|Q^n_{Y_2|U_0=\mathbf{u}_0,U_2=\mathbf{u}_2}\Big)\mathbb{E}_{\mathbb{C}_0,\mathbb{C}_2}\Big[\mathbb{1}_{\left\{\left(\mathbf{U}_0(1,\mathbb{C}_0),\mathbf{U}_2(1,1,\mathbb{C}_2)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}}\Big]\Bigg]\\
&\quad\overset{(e)}{=} \mathbb{E}_{\mathbb{C}_1}\Bigg[\sum_{\mathbf{u}_0,\mathbf{u}_2}Q^n_{U_0,U_2}(\mathbf{u}_0,\mathbf{u}_2)D\Big(P^{(\mathbb{C}_1)}_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathbb{C}_1}\Big|\Big|Q^n_{Y_2|U_0=\mathbf{u}_0,U_2=\mathbf{u}_2}\Big)\Bigg]\\
&\quad= \mathbb{E}_{\mathbb{C}_1} D\Big(P^{(\mathbb{C}_1)}_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0,\mathbf{U}_2,\mathbb{C}_1}\Big|\Big|Q^n_{Y_2|U_0,U_2}\Big|Q^n_{U_0,U_2}\Big) \qquad (54)
\end{aligned}
$$

where:

(a) uses the symmetry of the codebook with respect to the messages;

(b) is the law of total expectation (conditioning the inner expectation on $\mathbb{C}_1$);

(c) follows because conditioning on $\mathbb{C}_1$ fixes the relative entropy
$D\Big(P^{(\mathbb{C})}_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathbb{C}_1}\Big|\Big|Q^n_{Y_2|U_0=\mathbf{u}_0,U_2=\mathbf{u}_2}\Big)$;

(d) is since the codebook construction makes $\left(\mathbf{U}_0(1,\mathbb{C}_0),\mathbf{U}_2(1,1,\mathbb{C}_2)\right)$ independent of $\mathbb{C}_1$;

(e) relies on the coding PMF being $Q^n_{U_0,U_2}$.

We now invoke Lemma 1 on the RHS of (54) by viewing $Q^n_{Y_2|U_0,U_1,U_2}$ as a state-dependent DMC from $\mathcal{U}_1$ to $\mathcal{Y}_2$ with state space $\mathcal{U}_0\times\mathcal{U}_2$, and noting that the structure of $\mathcal{C}_1$ and the RHS of (54) fall within the framework of the lemma. Thus, the first two rate bounds in (50) ensure that

$$\mathbb{E}_{\mathbb{C}_1} D\Big(P^{(\mathbb{C}_1)}_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0,\mathbf{U}_2}\Big|\Big|Q^n_{Y_2|U_0,U_2}\Big|Q^n_{U_0,U_2}\Big)\xrightarrow[n\to\infty]{} 0. \qquad (55)$$

By (52) and (54), (55) implies that $\mathbb{E}_{\mathbb{C}}\mathcal{L}(\mathbb{C})\to 0$, as $n\to\infty$.

The Selection Lemma [19, Lemma 2.2] applied to the random variable $\mathbb{C}$ and the functions P_e and $\mathcal{L}$ implies the existence of a realization $\mathcal{C}$ of $\mathbb{C}$ that satisfies (11). Finally, we apply Fourier-Motzkin elimination (FME) on (50) while using (46) and the non-negativity of the involved terms, to eliminate R_{20}, R' and $\tilde{R}$. Since the above linear inequalities have constant coefficients, the FME can be performed by a computer program, e.g., by the FME-IT algorithm [49]. This establishes (12) as an inner bound.

Remark 5 *The main differences between the coding schemes for the cooperative BC with one confidential message and the same channel without secrecy [35] are threefold. First, a randomizer W is used in the secrecy-achieving scheme. Second, the cooperation message M_{12} depends on M_{20} rather than on the pair (M_{10}, M_{20}) (M_{10} refers to the public part of the message M_1). Note that conveying an M_{12} that holds any part of M_1 (in the form of its public part M_{10}) violates the secrecy requirement. Finally, a prefix channel $Q_{X|U_0,U_1,U_2}$ is used to optimize randomness and, in turn, to conceal M_1 from the 2nd receiver.*

D. Converse Proof for Theorem 4

We show that if a rate tuple (R_{12}, R_0, R_1, R_2) is achievable, then there exists a PMF $Q_{W,V,Y_1,X}Q_{Y_2|X}$ with $Y_1 = f(X)$, such that the inequalities in (14) are satisfied. Fix an achievable tuple (R_{12}, R_0, R_1, R_2) and an $\epsilon > 0$, and let $\mathcal{C}_n$ be the corresponding $(n, R_{12}, R_0, R_1, R_2)$ code for some sufficiently large $n \in \mathbb{N}$. The joint distribution induced by $\mathcal{C}_n$ is

$$P\big(m_0, m_1, m_2, \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, m_{12}, (\hat{m}_0^{(1)}, \hat{m}_1), (\hat{m}_0^{(2)}, \hat{m}_2)\big) = 2^{-n(R_0+R_1+R_2)} f^{(\mathcal{E})}(\mathbf{x}|m_0, m_1, m_2)$$
$$\times \mathbb{1}_{\left\{\bigcap_{i=1}^n \left(y_{1,i}=f(x_i)\right)\right\}} Q_{Y_2|X}^n(\mathbf{y}_2|\mathbf{x}) \mathbb{1}_{\left\{m_{12}=f_{12}(\mathbf{y}_1)\right\} \cap \left\{(\hat{m}_0^{(1)}, \hat{m}_1)=\phi_1(\mathbf{y}_1)\right\} \cap \left\{(\hat{m}_0^{(2)}, \hat{m}_2)=\phi_2(\mathbf{y}_2, m_{12})\right\}} \quad (56)$$

Thoughout the converse proof, all multi-letter information measures are calculated with respect to the PMF from (56) or its marginals. By Fano's inequality we have

$$H(M_0, M_1|Y_1^n) \leq 1 + n\epsilon(R_0 + R_1) \triangleq n\epsilon_n^{(1)} \quad (57a)$$

$$H(M_0, M_2|M_{12}, Y_2^n) \leq 1 + n\epsilon(R_0 + R_2) \triangleq n\epsilon_n^{(2)}. \quad (57b)$$

Define

$$\epsilon_n = \max\left\{\epsilon_n^{(1)}, \epsilon_n^{(2)}\right\}. \quad (57c)$$

Moreover, by (11b), we write

$$\begin{aligned}
\epsilon &\geq I(M_1; M_{12}, Y_2^n) \\
&= I(M_1; M_0, M_2, M_{12}, Y_2^n) - I(M_1; M_0, M_2|M_{12}, Y_2^n) \\
&\overset{(a)}{\geq} I(M_1; M_{12}, Y_2^n|M_0, M_2) - H(M_0, M_2|M_{12}, Y_2^n) \\
&\overset{(b)}{\geq} I(M_1; M_{12}, Y_2^n|M_0, M_2) - n\epsilon_n
\end{aligned} \quad (58)$$

where (a) uses the independence of M_1 and (M_0, M_2) and the non-negativity of entropy, while (b) follows from (57). Thus,

$$I(M_1; M_{12}, Y_2^n | M_0, M_2) \leq \epsilon + n\epsilon_n. \tag{59}$$

It follows that

$$\begin{aligned}
nR_1 &= H(M_1) \\
&\overset{(a)}{=} H(M_1|M_{12}, M_0, M_2) + I(M_1; M_{12}|M_0, M_2) \\
&\overset{(b)}{\leq} I(M_1; Y_1^n|M_{12}, M_0, M_2) + I(M_1; M_{12}|M_0, M_2) - I(M_1; M_{12}, Y_2^n|M_0, M_2) + n\delta_n^{(1)} \\
&\overset{(c)}{=} \sum_{i=1}^{n} \Big[I(M_1; Y_1^i, Y_{2,i+1}^n|M_{12}, M_0, M_2) - I(M_1; Y_1^{i-1}, Y_{2,i}^n|M_{12}, M_0, M_2) \Big] + n\delta_n^{(1)} \\
&= \sum_{i=1}^{n} \Big[I(M_1; Y_{1,i}|M_{12}, M_0, M_2, Y_1^{i-1}, Y_{2,i+1}^n) - I(M_1; Y_{2,i}|M_{12}, M_0, M_2, Y_1^{i-1}, Y_{2,i+1}^n) \Big] + n\delta_n^{(1)} \\
&\overset{(d)}{=} \sum_{i=1}^{n} \Big[H(Y_{1,i}|M_2, W_i) - H(Y_{1,i}|M_1, M_2, W_i) - I(M_1; Y_{2,i}|M_2, W_i) \Big] + n\delta_n^{(1)} \\
&\leq \sum_{i=1}^{n} \Big[H(Y_{1,i}|M_2, W_i) - I(Y_{1,i}; Y_{2,i}|M_1, M_2, W_i) - I(M_1; Y_{2,i}|M_2, W_i) \Big] + n\delta_n^{(1)} \\
&= \sum_{i=1}^{n} \Big[H(Y_{1,i}|M_2, W_i) - I(M_1, Y_{1,i}; Y_{2,i}|M_1, M_2, W_i) \Big] + n\delta_n^{(1)} \\
&\leq \sum_{i=1}^{n} H(Y_{1,i}|M_2, W_i, Y_{2,i}) + n\delta_n^{(1)}
\end{aligned} \tag{60}$$

where:

(a) is because M_1 is independent (M_0, M_2);

(b) follows from (57)-(58) and by denoting $\delta_n^{(1)} = 2\epsilon_n + \frac{\epsilon}{n}$;

(c) is a telescoping identity [50, Eqs. (9) and (11)];

(d) is by defining $W_i = (M_{12}, M_0, Y_1^{i-1}, Y_{2,i+1}^n)$.

The common message rate R_0 satisfies

$$\begin{aligned}
nR_0 &= H(M_0) \\
&\overset{(a)}{\leq} I(M_0; Y_1^n) + n\epsilon_n \qquad (61a) \\
&= \sum_{i=1}^{n} I(M_0; Y_{1,i}|Y_1^{i-1}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(M_0, Y_1^{i-1}; Y_{1,i}) + n\epsilon_n \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} I(W_i; Y_{1,i}) + n\epsilon_n \qquad (61b)
\end{aligned}$$

where (a) uses (57) and (b) follows by the definition of W_i. Combining (60) with (61b) yields

$$n(R_0+R_1) \leq \sum_{i=1}^{n} \Big[H(Y_{1,i}|M_2, W_i, Y_{2,i}) + I(W_i; Y_{1,i}) \Big] + n\delta_n^{(2)} \tag{62}$$

where $\delta_n^{(2)} = \delta_n^{(1)} + \epsilon_n$.

For the sum $R_0 + R_2$, we have

$$\begin{aligned} n(R_0+R_2) &= H(M_0, M_2) \\ &\overset{(a)}{\leq} I(M_0, M_2; M_{12}, Y_2^n) + n\epsilon_n \\ &= I(M_0, M_2; Y_2^n|M_{12}) + I(M_0, M_2; M_{12}) + n\epsilon_n \\ &\overset{(b)}{\leq} I(M_0, M_2; Y_2^n|M_{12}) + nR_{12} + n\epsilon_n \\ &= \sum_{i=1}^{n} I(M_0, M_2; Y_{2,i}|M_{12}, Y_{2,i+1}^n) + nR_{12} + n\epsilon_n \\ &\overset{(c)}{\leq} \sum_{i=1}^{n} I(M_2, W_i; Y_{2,i}) + nR_{12} + n\epsilon_n \end{aligned} \tag{63}$$

where:

(a) uses (57);

(b) is by the non-negativity of entropy and since a uniform distribution maximizes entropy;

(c) follows from the definition of W_i and because conditioning cannot increase entropy.

To bound $R_0 + R_1 + R_2$, we begin by writing

$$n(R_0+R_1+R_2) = H(M_0, M_1, M_2) = H(M_1|M_0, M_2) + H(M_2|M_0) + H(M_0). \tag{64}$$

Consider

$$\begin{aligned} H(M_2|M_0) &\overset{(a)}{\leq} I(M_2; Y_2^n|M_{12}, M_0) + I(M_2; M_{12}|M_0) + n\epsilon_n \\ &\overset{(b)}{=} \sum_{i=1}^{n} \Big[I(M_2; Y_{2,i}^n|M_{12}, M_0, Y_1^{i-1}) - I(M_2; Y_{2,i+1}^n|M_{12}, M_0, Y_1^i) \Big] + I(M_2; M_{12}|M_0) + n\epsilon_n \\ &\overset{(c)}{=} \sum_{i=1}^{n} \Big[I(M_2; Y_{2,i+1}^n|M_{12}, M_0, Y_1^{i-1}) + I(M_2; Y_{2,i}|W_i) - I(M_2; Y_{1,i}, Y_{2,i+1}^n|M_{12}, M_0, Y_1^{i-1}) \\ &\qquad + I(M_2; Y_{1,i}|M_{12}, M_0, Y_1^{i-1}) \Big] + I(M_2; M_{12}|M_0) + n\epsilon_n \\ &\overset{(d)}{=} \sum_{i=1}^{n} \Big[I(M_2; Y_{2,i}|W_i) - I(M_2; Y_{1,i}|W_i) \Big] + I(M_2; Y_1^n|M_0) + n\epsilon_n \end{aligned} \tag{65}$$

where:

(a) is by (57) and by the mutual information chain rule;

(b) is a telescoping identity;

(c) follows from the definition of W_i;

(d) is due to the mutual information chain rule and the definition of W_i (second term), and because M_{12} is defined by Y_1^n (third term).

Combining (61a) with (65), yields

$$\begin{aligned} n(R_0+R_2) &\le \sum_{i=1}^{n}\Big[I(M_2;Y_{2,i}|W_i)-I(M_2;Y_{1,i}|W_i)\Big]+I(M_0,M_2;Y_1^n)+2n\epsilon_n \\ &\overset{(a)}{\le} \sum_{i=1}^{n}\Big[I(M_2;Y_{2,i}|W_i)-I(M_2;Y_{1,i}|W_i)+H(Y_{1,i})-H(Y_{1,i}|M_0,M_2,Y_1^{i-1})\Big]+2n\epsilon_n \\ &\overset{(b)}{\le} \sum_{i=1}^{n}\Big[I(M_2;Y_{2,i}|W_i)+I(W_i;Y_{1,i})-I(M_{12},Y_{2,i+1}^n;Y_{1,i}|M_0,M_2,Y_1^{i-1})\Big]+2n\epsilon_n \\ &\overset{(c)}{\le} \sum_{i=1}^{n}\Big[I(M_2;Y_{2,i}|W_i)+I(W_i;Y_{1,i})\Big]+2n\epsilon_n \end{aligned} \tag{66}$$

where:

(a) is because conditioning cannot increase entropy;

(b) uses the definition of W_i;

(c) is by the non-negativity of mutual information.

By inserting (60) and (66) into (64), we bound the sum of rates as

$$n(R_0+R_1+R_2) \le \sum_{i=1}^{n}\Big[H(Y_{1,i}|M_2,W_i,Y_{2,i})+I(M_2;Y_{2,i}|W_i)+I(W_i;Y_{1,i})\Big]+n\delta_n^{(3)} \tag{67}$$

where $\delta_n^{(3)}=\delta_n^{(1)}+2\epsilon_n$.

The bounds in (60), (62), (63) and (66) are rewritten by introducing a time-sharing random variable T that is uniformly distributed over the set $[1:n]$ and is independent of $(M_0,M_1,M_2,X^n,Y_1^n,Y_2^n)$. For instance, (60) is rewritten as

$$\begin{aligned} R_1 &\le \frac{1}{n}\sum_{t=1}^{n} H(Y_{1,t}|M_2,W_t,Y_{2,t})+\delta_n^{(1)} \\ &= \sum_{t=1}^{n}\mathbb{P}\big(T=t\big)H(Y_{1,T}|M_2,W_T,Y_{2,T},T=t)+\delta_n^{(1)} \\ &= H(Y_{1,T}|M_2,W_T,Y_{2,T},T)+\delta_n^{(1)} \end{aligned} \tag{68}$$

Denote $W \triangleq (W_T,T)$, $V \triangleq (M_2,W)$, $X \triangleq X_T$, $Y_1 \triangleq Y_{1,T}$ and $Y_2 \triangleq Y_{2,T}$. This results in the bounds (14) with small added terms such as ϵ_n and $\delta_n^{(1)}$. For large n, we can make these terms approach 0. The converse is completed by showing the PMF of (W,V,X,Y_1,Y_2) factors as $P_{W,V,Y_1,X}P_{Y_2|X}$, which boils down to the Markov relation

$$(W,V,Y_1)-X-Y_2. \tag{69}$$

This is proven in Appendix E-A.

E. Converse Proof for Theorem 5

We show that given an achievable rate tuple (R_{12}, R_0, R_1, R_2), there is a PMF $Q_{W,X}Q_{Y_1|X}Q_{Y_2|Y_1}$ for which (15) holds. Let be (R_{12}, R_0, R_1, R_2) an achievable tuple and fix $\epsilon > 0$. Let $\mathcal{C}_n$ be the corresponding $(n, R_{12}, R_0, R_1, R_2)$ code for some sufficiently large $n \in \mathbb{N}$. The induced joint distribution is coincides (56) up to replacing the DM SD-BC transition matrix $P(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}) = \mathbb{1}_{\left\{\bigcap_{i=1}^n \left(y_{1,i}=f(x_i)\right)\right\}} Q^n_{Y_2|X}(\mathbf{y}_2|\mathbf{x})$ from (56) with this of a PD-BC, i.e., with $P(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}) = Q^n_{Y_1|X}(\mathbf{y}_1|\mathbf{x})Q^n_{Y_2|Y_1}(\mathbf{y}_2|\mathbf{y}_1)$. All multi-letter information measures throughout this proof are calculated with respect to the induced PMF or its marginals. Fano's inequality gives

$$H(M_0, M_1|Y_1^n) \leq 1 + n\epsilon(R_0 + R_1) \triangleq n\kappa_n^{(1)} \tag{70a}$$

$$H(M_0, M_2|M_{12}, Y_2^n) \leq 1 + n\epsilon(R_0 + R_2) \triangleq n\kappa_n^{(2)} \tag{70b}$$

$$H(M_0, M_1, M_2|Y_1^n, Y_2^n) \leq 1 + n\epsilon(R_0 + R_1 + R_2) \triangleq n\kappa_n^{(3)} \tag{70c}$$

and we set

$$\kappa_n = \max\left\{\kappa_n^{(1)}, \kappa_n^{(2)}, \kappa_n^{(3)}\right\} = \kappa_n^{(3)}. \tag{70d}$$

Further, by the strong-secrecy constraint (11b), we have

$$\begin{aligned}
\epsilon &\geq I(M_1; M_{12}, Y_2^n) \\
&= I(M_1; M_0, M_2, M_{12}, Y_2^n) - I(M_1; M_0, M_2|M_{12}, Y_2^n) \\
&\overset{(a)}{\geq} I(M_1; M_{12}, Y_2^n|M_0, M_2) - H(M_0, M_2|M_{12}, Y_2^n) \\
&\overset{(b)}{\geq} I(M_1; Y_2^n|M_0, M_2) - n\kappa_n
\end{aligned} \tag{71}$$

where (a) uses the independence of M_1 and (M_0, M_2) and the non-negativity of entropy, while (b) is by (70) and since conditioning cannot increase entropy. This yields

$$I(M_1; Y_2^n|M_0, M_2) \leq \epsilon + n\kappa_n. \tag{72}$$

We bound

$$\begin{aligned}
nR_1 &= H(M_1) \\
&\overset{(a)}{=} H(M_1|M_0, M_2) \\
&\overset{(b)}{\leq} I(M_1; Y_1^n|M_0, M_2) - I(M_1; Y_2^n|M_0, M_2) + n\eta_n \\
&\overset{(c)}{=} \sum_{i=1}^n \Big[I(M_1; Y_1^i, Y_{2,i+1}^n|M_0, M_2) - I(M_1; Y_1^{i-1}, Y_{2,i}^n|M_0, M_2)\Big] + n\eta_n \\
&\overset{(d)}{=} \sum_{i=1}^n \Big[I(M_1; Y_{1,i}|W_i) - I(M_1; Y_{2,i}|W_i)\Big] + n\eta_n
\end{aligned} \tag{73a}$$

$$\begin{aligned}
&\overset{(e)}{=} \sum_{i=1}^{n} I(M_1; Y_{1,i}|W_i, Y_{2,i}) + n\eta_n \\
&\overset{(f)}{\leq} \sum_{i=1}^{n} I(X_i; Y_{1,i}|W_i, Y_{2,i}) + n\eta_n \\
&\overset{(g)}{\leq} \sum_{i=1}^{n} \Big[I(X_i; Y_{1,i}|W_i) - I(X_i; Y_{2,i}|W_i) \Big] + n\eta_n
\end{aligned} \tag{73b}$$

where:

(a) uses the independence of M_1 and (M_0, M_2);

(b) is by (70) and (71), and by denoting $\eta_n = 2\kappa_n + \frac{\epsilon}{n}$;

(c) is a telescoping identity;

(d) follows by defining $W_i \triangleq (M_0, M_2, Y_1^{i-1}, Y_{2,i+1}^n)$;

(e) and (g) rely on the mutual information chain rule and the PD property of the channel, which implies that $(M_1, X_i) - (W_i, Y_{1,i}) - Y_{2,i}$ forms a Markov chain for all $i \in [1:n]$;

(f) follows since $M_1 - (W_i, X_i, Y_{1,i}) - Y_{2,i}$ forms a Markov chain.

Next, we have

$$\begin{aligned}
n(R_0 + R_2) &= H(M_0, M_2) \\
&\overset{(a)}{\leq} I(M_0, M_2; M_{12}, Y_2^n) + n\kappa_n \\
&\overset{(b)}{\leq} I(M_0, M_2; Y_2^n) + nR_{12} + n\kappa_n \\
&= \sum_{i=1}^{n} I(M_0, M_2; Y_{2,i}|Y_{2,i+1}^n) + nR_{12} + n\kappa_n \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} I(W_i; Y_{2,i}) + nR_{12} + n\kappa_n
\end{aligned} \tag{74}$$

where:

(a) is by (70);

(b) is because entropy is non-negative and is maximized by the uniform distribution;

(c) follows from the definition of W_i and because conditioning cannot increase entropy.

Finally, consider

$$\begin{aligned}
n(R_0 + R_1 + R_2) &= H(M_0, M_1, M_2) \\
&\overset{(a)}{\leq} I(M_0, M_1, M_2; Y_1^n, Y_2^n) - I(M_1; Y_2^n|M_0, M_2) + n\eta_n \\
&\overset{(b)}{=} I(M_0, M_1, M_2; Y_1^n) - I(M_1; Y_2^n|M_0, M_2) + n\eta_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} \Big[I(M_0, M_1, M_2, Y_{2,i+1}^n; Y_{1,i}|Y_1^{i-1}) - I(Y_{2,i+1}^n; Y_{1,i}|M_0, M_1, M_2, Y_1^{i-1}) \\
&\qquad\qquad - I(M_1; Y_{2,i}|M_0, M_2, Y_{2,i+1}^n) \Big] + n\eta_n
\end{aligned}$$

$$\begin{aligned}
&\overset{(d)}{=} \sum_{i=1}^{n} \Big[I(M_0, M_1, M_2, Y_{2,i+1}^n; Y_{1,i}|Y_1^{i-1}) - I(Y_1^{i-1}; Y_{2,i}|M_0, M_1, M_2, Y_{2,i+1}^n) \\
&\qquad\qquad - I(M_1; Y_{2,i}|M_0, M_2, Y_{2,i+1}^n) \Big] + n\eta_n \\
&\leq \sum_{i=1}^{n} \Big[I(M_0, M_1, M_2, Y_1^{i-1}, Y_{2,i+1}^n; Y_{1,i}) - I(M_1, Y_1^{i-1}; Y_{2,i}|M_0, M_2, Y_{2,i+1}^n) \Big] + n\eta_n \\
&\overset{(e)}{\leq} \sum_{i=1}^{n} \Big[I(W_i; Y_{1,i}) + I(M_1; Y_{1,i}|W_i) - I(M_1; Y_{2,i}|W_i) \Big] + n\eta_n \\
&\overset{(f)}{\leq} \sum_{i=1}^{n} \Big[I(W_i; Y_{1,i}) + I(X_i; Y_{1,i}|W_i) - I(X_i; Y_{2,i}|W_i) \Big] + n\eta_n \\
&\overset{(g)}{=} \sum_{i=1}^{n} \Big[I(X_i; Y_{1,i}) - I(X_i; Y_{2,i}|W_i) \Big] + n\eta_n
\end{aligned} \tag{75}$$

where:

(a) uses (70) and the definition of η_n;

(b) is because $(M_0, M_1, M_2) - Y_1^n - Y_2^n$ forms a Markov chain, which is induced by the PD degraded and memoryless property of the channel;

(c) is the mutual information chain rule;

(d) uses the Csiszár sum identity;

(e) follows from the definitions of W_i and because conditioning cannot increase entropy;

(f) is by repeating steps (73a)-(73b);

(g) is by the mutual information chain rule and because $W_i - X_i - Y_{1,i}$ forms a Markov chain (see Appendix E-B for the proof).

By time-sharing arguments similar to those presented in Section VII-D, and by denoting $W \triangleq (W_T, T)$, $X \triangleq X_T$, $Y_1 \triangleq Y_{1,T}$ and $Y_2 \triangleq Y_{2,T}$, we obtain the bounds of (15) with the small added terms κ_n and η_n, which approach 0 as $n \to \infty$. In Appendix E-B we shown that the chain

$$W - X - Y_1 - Y_2 \tag{76}$$

is Markov, which establishes the converse.

VIII. Summary and Concluding Remarks

We considered cooperative BCs with one common and two private messages, where the private message to the cooperative user is confidential. An inner bound on the strong-secrecy-capacity region was established by deriving a channel resolvability lemma and using it as a building block in the BC code. The analysis was made tractable by incorporating the likelihood encoder [33] as the multicoding mechanism.

For the BC, a resolvability-based Marton code with a double-binning of the confidential message codebook was constructed, and the resolvability lemma was invoked to achieve strong-secrecy. The cooperation protocol used the link from Decoder 1 to Decoder 2 to share information on a portion of the non-confidential message and the

common message only. Removing the secrecy constraint on M_1 allows a more flexible cooperation scheme that in general achieves strictly higher transmission rates [35]. The inner bound was shown to be tight for the SD and PD cases. Two separate converse proofs were used because the structure of the joint PMFs describing the regions seem to require distinct choices of auxiliary random variable.

The secrecy results were compared to those of the corresponding BCs without secrecy constraints, and the impact of secrecy on the capacity regions was highlighted. Cooperative Blackwell and Gaussian BCs visualized the results. An explicit coding scheme that achieves strong-secrecy while maximizing the transmission rate of the confidential message over the BBC was given. Further, it was shown that the strong-secrecy-capacity region of the BBC remains unchanged even if the subchannel to the legitimate user is noiseless.

APPENDIX A
PROOF OF PROPOSITION 7

Consider the SD-BC depicted in Fig. 3, where $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}$. Setting $U_0 = W$, $U_1 = Y_1$ and $U_2 = V$ into $\mathcal{R}_{\mathrm{NS}}$ recovers the capacity-region $\mathcal{C}_{\mathrm{NS}}^{(\mathrm{SD})}$ of the cooperative SD-BC [35, Theorem 5], which is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \leq H(Y_1) \tag{77a}$$

$$R_2 \leq I(W, V; Y_2) + R_{12} \tag{77b}$$

$$R_1 + R_2 \leq H(Y_1|W, V) + I(V; Y_2|W) + \min\Big\{I(W; Y_1), I(W; Y_2) + R_{12}\Big\} \tag{77c}$$

where the union is over all PMFs $Q_{V,U,Y_1,X}Q_{Y_2|X}$ for which $Y_1 = f(X)$. Setting $U_0 = W$, $U_1 = Y_1$ and $U_2 = V$ into $\tilde{\mathcal{R}}_{\mathrm{NS}}$, gives a region that coincides with (77), up to replacing (77a) with

$$R_1 \leq H(Y_1|W) + \Big[I(V; Y_2|W) - I(V; Y_1|W)\Big]^+. \tag{78}$$

Denoting the obtained region by $\tilde{\mathcal{R}}_{\mathrm{NS}}^{(\mathrm{SD})}$, we outer bound it by loosening (78) to

$$R_1 \leq H(Y_1|W). \tag{79}$$

Let $\tilde{\mathcal{O}}_{\mathrm{NS}}^{(\mathrm{SD})}$ denote the obtained outer bound on $\tilde{\mathcal{R}}_{\mathrm{NS}}^{(\mathrm{SD})}$. We show that under the considered example $\tilde{\mathcal{O}}_{\mathrm{NS}}^{(\mathrm{SD})} \subsetneq \mathcal{C}_{\mathrm{NS}}^{(\mathrm{SD})}$.

For any $r \in \mathbb{R}_+$, let

$$\mathcal{C}_{\mathrm{NS}}^{(\mathrm{SD})}(r) \triangleq \Big\{(r, R_1, R_2) \in \mathcal{C}_{\mathrm{NS}}^{(\mathrm{SD})}\Big\} \tag{80a}$$

$$\tilde{\mathcal{O}}_{\mathrm{NS}}^{(\mathrm{SD})}(r) \triangleq \Big\{(r, R_1, R_2) \in \tilde{\mathcal{O}}_{\mathrm{NS}}^{(\mathrm{SD})}\Big\} \tag{80b}$$

be the projections of $\mathcal{C}_{\mathrm{NS}}^{(\mathrm{SD})}$ and $\tilde{\mathcal{O}}_{\mathrm{NS}}^{(\mathrm{SD})}$ on the (R_1, R_2) plane for $R_{12} = r$. Set $r = 1$ and note that $R_1 = 1$ is the maximal achievable rate of M_1 in both $\mathcal{C}_{\mathrm{NS}}^{(\mathrm{SD})}$ and $\tilde{\mathcal{O}}_{\mathrm{NS}}^{(\mathrm{SD})}$. Define the maximal achievable R_2 that preserves for

$R_1 = 1$ in each region by

$$R_2^\star \triangleq \max\Big\{R_2 \in \mathbb{R}_+ \Big| (1,1,R_2) \in \mathcal{C}_{\rm NS}^{\rm (SD)}(1)\Big\} \tag{81a}$$

$$\tilde{R}_2^\star \triangleq \max\Big\{R_2 \in \mathbb{R}_+ \Big| (1,1,R_2) \in \tilde{\mathcal{O}}_{\rm NS}^{\rm (SD)}(1)\Big\}. \tag{81b}$$

We next evaluate $R_2^\star$ and $\tilde{R}_2^\star$, and then choose $Q_{Y_2|X_1,X_2}$ for which $R_2^\star > \tilde{R}_2^\star$.

For $\mathcal{C}_{\rm NS}^{\rm (SD)}(1)$, setting $X_1 \sim \text{Ber}\big(\frac{1}{2}\big)$ achieves $R_1 = 1$ since

$$R_1 = H(Y_1) = H(X_1) = 1. \tag{82}$$

Consequently, for $R_2^\star$ we have

$$\begin{aligned}
R_2^\star &\overset{(a)}{=} \max_{\substack{Q_{W,V,X_2|X_1}:\\ (W,V)-(X_1,X_2)-Y_2}} \min\Big\{I(W,V;Y_2)+1\,,\, I(V;Y_2|W)-I(V;Y_1|W)\,,\, I(W,V;Y_2)+H(Y_1|W,V)\Big\}\\
&\overset{(b)}{\geq} \max_{\substack{Q_{V,X_2|X_1}:\\ V-(X_1,X_2)-Y_2}} \min\Big\{I(X_1,V;Y_2)+1\,,\, I(V;Y_2|X_1)-I(V;Y_1|X_1)\,,\, I(X_1,V;Y_2)+H(Y_1|X_1,V)\Big\}\\
&\overset{(c)}{=} \max_{\substack{Q_{V,X_2|X_1}:\\ V-(X_1,X_2)-Y_2}} I(V;Y_2|X_1)
\end{aligned} \tag{83}$$

where (a) uses the structure of $\mathcal{C}_{\rm NS}^{\rm (SD)}$ from (77) and the relation $R_{12} = H(Y_1) = 1$, (b) follows by setting $W = X_1$ and because the Markov chain $(X_1, V) - (X_1, X_2) - Y_2$ implies that $V - (X_1, X_2) - Y_2$ also forms a Markov chain, while (c) is because $I(V;Y_1|X_1) = H(Y_1|X_1,V) = 0$ and $I(V;Y_2|X_1) \leq I(X_1,V;Y_2)$.

Achieving $R_1 = 1$ in $\tilde{\mathcal{O}}_{\rm NS}^{\rm (SD)}(1)$ requires $X_1 \sim \text{Ber}\big(\frac{1}{2}\big)$ that is independent of W. This is since in $\tilde{\mathcal{O}}_{\rm NS}^{\rm (SD)}(1)$ we have

$$R_1 \leq H(Y_1|W) = H(X_1|W) \leq H(X_1) \leq 1 \tag{84}$$

and (X_1, W) as above satisfy (84) with equality. For $\tilde{R}_2^\star$, we have

$$\begin{aligned}
\tilde{R}_2^\star &\overset{(a)}{=} \max_{\substack{Q_W Q_{V,X_2|W,X_1}:\\ (W,V)-(X_1,X_2)-Y_2}} \min\Big\{I(W,V;Y_2)+1\,,\, I(V;Y_2|W)-I(V;Y_1|W)\,,\, I(W,V;Y_2)+H(Y_1|W,V)\Big\}\\
&\overset{(b)}{=} \max_{\substack{Q_W Q_{V,X_2|X_1,W}:\\ (W,V)-(X_1,X_2)-Y_2}} I(V;Y_2|W) - I(V;Y_1|W)\\
&\overset{(c)}{\leq} \max_{w\in\mathcal{W}} \max_{\substack{Q_{V,X_2|X_1,W=w}:\\ V_w-(X_1,X_{2,w})-Y_2}} I(V;Y_2|W=w) - I(V;Y_1|W=w)\\
&\leq \max_{\substack{Q_{V,X_2|X_1}:\\ V-(X_1,X_2)-Y_2}} I(V;Y_2) - I(V;Y_1)
\end{aligned} \tag{85}$$

where:

(a) uses the structure of $\tilde{\mathcal{Q}}_{\rm NS}^{\rm (SD)}$, the independence of W and $X_1 = Y_1$ and the relation $R_{12} = H(Y_1|W) = 1$;

(b) follows by $I(V;Y_2|W) \leq I(W,V;Y_2)$ and the non-negativity of mutual information and entropy.

(c) is by defining $(V_w, X_{2,w})$ to be a pair of random variables jointly distributed with $X_1 \sim \text{Ber}\big(\frac{1}{2}\big)$ according to $Q_{X_1} Q_{V,X_2|X_1,W=w}$, where $w \in \mathcal{W}$.

The lower bound on $R_2^\star$ from (83) is the capacity of the state-dependent channel $Q_{Y_2|X_1,X_2}$ with non-causal CSI available at the transmitting and receiving ends. The upper bound on $\tilde{R}_2^\star$ given in (85) is the capacity of the corresponding GP channel, i.e., with non-causal transmitter CSI only. Thus, to show that $\tilde{R}_2^\star < R_2^\star$ it suffices to choose $Q_{Y_2|X_1,X_2}$ for which the GP capacity is strictly less than the capacity with full CSI. A simple example for which these capacities are different is the binary dirty-paper (BDP) channel. Specifically, let $Q_{Y_2|X_1,X_2}$ be defined by

$$Y_2 = X_2 \oplus X_1 \oplus Z \tag{86}$$

where $\oplus$ denotes modulo 2 addition, $X_1 \sim \text{Ber}\big(\frac{1}{2}\big)$ plays the role of the channel's state, and the noise $Z \sim \text{Ber}(\epsilon)$, with $\epsilon \in \big[0, \frac{1}{2}\big]$ is independent of (X_1, X_2). The input X_2 is subject to a constraint $\frac{1}{n} w_H(\mathbf{x}_2) \le q$, for $q \in \big[0, \frac{1}{2}\big]$, where $w_H : \big\{0,1\big\}^n \to \mathbb{N} \cup \big\{0\big\}$ is the Hamming weight. For the BDP channel, the GP capacity is [44]–[46]

$$C_{\text{GP}}^{(\text{BDP})} = \max_{\substack{Q_{V,X_2|X_1}: \\ V-(X_1,X_2)-Y_2}} I(V;Y_2) - I(V;Y_1) = \text{uce}\Big\{ \big[H_b(q) - H_b(\epsilon)\big]^+ \Big\} \tag{87}$$

where 'uce' is the upper convex envelope operation with respect to q (ϵ is constant). On the other hand, the capacity of the BDP channel with full CSI is [44]–[46]

$$C_{\text{F}-\text{CSI}}^{(\text{BDP})} = \max_{\substack{Q_{V,X_2|X_1}: \\ V-(X_1,X_2)-Y_2}} I(V;Y_2|X_1) = H_b(q * \epsilon) - H_b(\epsilon) \tag{88}$$

where $q * \epsilon = q(1-\epsilon) + (1-q)\epsilon$. Clearly, q and ϵ can be chosen to yield $C_{\text{GP}}^{(\text{BDP})} < C_{\text{F}-\text{CSI}}^{(\text{BDP})}$, which shows that $\mathcal{R}_{\text{NS}}^{(\text{SD})}$ and $\tilde{\mathcal{R}}_{\text{NS}}^{(\text{SD})}$ are not equal in general.

APPENDIX B
CONVERSE PROOF FOR (27)

To prove the optimality of (27), we show that $\mathcal{C}_{\text{S}}^{(\text{PD})} \subseteq \mathcal{C}_{\text{S}}^{(\text{G})}$ ($\mathcal{C}_{\text{S}}^{(\text{PD})}$ and $\mathcal{C}_{\text{S}}^{(\text{G})}$ are given by (15) and (27), respectively). First consider

$$\frac{1}{2}\log(2\pi e N_1) = h(Z_1) = h(Y_1|X) \overset{(a)}{\le} h(Y_1|W) \le h(Y_1) \le \frac{1}{2}\log\big(2\pi e(P+N_1)\big) \tag{89}$$

where (a) is because $W - X - Y_1$ forms a Markov chain. The intermediate-value theorem and (89) imply that there is an $\alpha \in [0,1]$, such that

$$\frac{1}{2}\log\big(2\pi e(\alpha P + N_1)\big). \tag{90}$$

Further, for every $w \in \mathcal{W}$, we have

$$h(Y_2|W=w) = h(Y_1 + Z_2|W=w)$$

$$\begin{aligned} &\overset{(a)}{\geq} \frac{1}{2}\log\left(2^{2h(Y_1|W=w)} + 2^{2h(Z_2|W=w)}\right) \\ &\overset{(b)}{=} \frac{1}{2}\log\left(2^{2h(Y_1|W=w)} + 2\pi e(N_2 - N_1)\right) \end{aligned} \tag{91}$$

where (a) uses the conditional entropypower inequality (EPI), while (b) follows by the independence of Z_2 and W. Using (91), we lower bound $h(Y_2|W)$ in terms of $h(Y_1|W)$ as

$$\begin{aligned} h(Y_2|W) &= \mathbb{E}_W\big[h(Y_2|W)\big] \\ &\overset{(a)}{\geq} \mathbb{E}_W\left[\frac{1}{2}\log\left(2^{2h(Y_1|W)} + 2\pi e(N_2 - N_1)\right)\right] \\ &\overset{(b)}{\geq} \frac{1}{2}\log\left(2^{2\mathbb{E}_W\left[h(Y_1|W)\right]} + 2\pi e(N_2 - N_1)\right) \\ &= \frac{1}{2}\log\left(2^{2h(Y_1|W)} + 2\pi e(N_2 - N_1)\right) \\ &= \frac{1}{2}\log\big(2\pi e(\alpha P + N_2)\big) \end{aligned} \tag{92}$$

where (a) follows from (91), while (b) uses the convexity of the function $x \to \log(2^x + c)$ for $c \in \mathbb{R}_+$ and Jensen's inequality.

Having this, we present the following upper bounds of the information terms in the RHS of (15). For (15a), we have

$$\begin{aligned} I(X;Y_1|W) - I(X;Y_2|W) &\overset{(a)}{=} h(Y_1|W) - h(Y_1|X) - h(Y_2|W) + h(Y_2|X) \\ &\overset{(b)}{\leq} \frac{1}{2}\log\left(1 + \frac{\alpha P}{N_1}\right) - \frac{1}{2}\log\left(1 + \frac{\alpha P}{N_2}\right) \end{aligned} \tag{93}$$

where (a) follows since the chain $W - X - (Y_1, Y_2)$ is Markov, while (b) relies on (90), (92) and on the Gaussian distribution being the maximizer of the differential entropy under a variance constraint. Next, using (92) we bound the RHS of (15b) as

$$I(W;Y_2) + R_{12} = h(Y_2) - h(Y_2|W) + R_{12} \leq \frac{1}{2}\log\left(1 + \frac{\bar{\alpha} P}{\alpha P + N_2}\right) + R_{12}. \tag{94}$$

By repeating arguments similar to those in the derivation of (93), we bound the sum of rates $R_1 + R_2$ as

$$R_1 + R_2 \leq \frac{1}{2}\log\left(1 + \frac{P}{N_1}\right) - \frac{1}{2}\log\left(1 + \frac{\alpha P}{N_2}\right). \tag{95}$$

APPENDIX C
PROOF OF LEMMA 10

For a fixed codebook $\mathcal{B}_n$ and every $(\mathbf{s}_0, \mathbf{s}, \mathbf{v}) \in \mathcal{S}_0^n \times \mathcal{S}^n \times \mathcal{V}^n$, we have

$$P^{(\mathcal{B}_n)}(\mathbf{s}_0, \mathbf{s}, \mathbf{v}) = Q^n_{S_0,S}(\mathbf{s}_0, \mathbf{s}) \sum_{w,i} 2^{-n\tilde{R}} f^{(\mathrm{LE})}(i|w, \mathbf{s}_0, \mathbf{s}, \mathcal{B}_n) Q^n_{V|U,S_0,S}\big(\mathbf{v}\big|\mathbf{u}(\mathbf{s}_0, w, i, \mathcal{B}_n), \mathbf{s}_0, \mathbf{s}\big). \tag{96}$$

Let $(\mathbf{s}_0, \mathbf{s}, \mathbf{v}) \in \mathcal{S}_0^n \times \mathcal{S}^n \times \mathcal{V}^n$ be a triple such that $Q^n_{S_0,S,V}(\mathbf{s}_0, \mathbf{s}, \mathbf{v}) = 0$. Clearly, if $Q^n_{S_0,S}(\mathbf{s}_0, \mathbf{s}) = 0$ then (96)

implies that $P^{(\mathcal{B}_n)}(\mathbf{s}_0,\mathbf{s},\mathbf{v})=0$. Thus, we henceforth assume that $Q^n_{S_0,S}(\mathbf{s}_0,\mathbf{s})>0$ and $Q^n_{V|S_0,S}(\mathbf{v}|\mathbf{s}_0,\mathbf{s})=0$. By expanding

$$Q^n_{V|S_0,S}(\mathbf{v}|\mathbf{s}_0,\mathbf{s})=\sum_{\mathbf{u}\in\operatorname{supp}\left(Q^n_{U|S_0=\mathbf{s}_0,S=\mathbf{s}}\right)} Q^n_{U|S_0,S}(\mathbf{u}|\mathbf{s}_0,\mathbf{s})Q^n_{V|U,S_0,S}(\mathbf{v}|\mathbf{u},\mathbf{s}_0,\mathbf{s}) \tag{97}$$

we have $Q^n_{V|U,S_0,S}(\mathbf{v}|\mathbf{u},\mathbf{s}_0,\mathbf{s})=0$ for every $\mathbf{u}\in\operatorname{supp}\left(Q^n_{U|S_0=\mathbf{s}_0,S=\mathbf{s}}\right)$. Thus, to complete the proof it suffices to show that every u-codeword that is transmitted with positive probability is in $\operatorname{supp}\left(Q^n_{U|S_0=\mathbf{s}_0,S=\mathbf{s}}\right)$.

By the construction of the codebook, every $\mathbf{u}\in\mathcal{B}_n$ also satisfies $\mathbf{u}\in\operatorname{supp}\left(Q^n_{U|S_0=\mathbf{s}_0}\right)$. Moreover, a necessary condition for a codeword $\mathbf{u}(\mathbf{s}_0,w,i,\mathcal{B}_n)$ to be chosen by the encoder with positive probability is $f^{(\mathrm{LE})}(i|w,\mathbf{s}_0,\mathbf{s},\mathcal{B}_n)>0$, which by the definition of the likelihood encoder implies that $Q^n_{S|U,S_0}\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0,w,i,\mathcal{B}_n),\mathbf{s}_0\big)>0$. Combining the above, we have that if a codeword $\mathbf{u}(\mathbf{s}_0,w,i,\mathcal{B}_n)$ is transmitted with positive probability then

$$\begin{aligned}Q^n_{U|S_0,S}\big(\mathbf{u}(\mathbf{s}_0,w,i,\mathcal{B}_n)\big|\mathbf{s}_0,\mathbf{s}\big)&=\frac{Q^n_{S_0,S,U}\big(\mathbf{s}_0,\mathbf{s},\mathbf{u}(\mathbf{s}_0,w,i,\mathcal{B}_n)\big)}{Q^n_{S_0,S}(\mathbf{s}_0,\mathbf{s})}\\&=\frac{Q^n_{S_0}(\mathbf{s}_0)Q^n_{U|S_0}\big(\mathbf{u}(\mathbf{s}_0,w,i,\mathcal{B}_n)\big|\mathbf{s}_0\big)Q^n_{S|U,S_0}\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0,w,i,\mathcal{B}_n),\mathbf{s}_0\big)}{Q^n_{S_0,S}(\mathbf{s}_0,\mathbf{s})}>0.\end{aligned}$$

APPENDIX D
ERROR PROBABILITY ANALYSIS FOR THEOREM 3

By symmetry, we assume that $(M_p,M_1,M_{22},W)=\mathbf{1}=(1,1,1,1)$ is chosen. We also define I to be a random variable that represents the index chosen by the likelihood encoder.

Encoding errors: An encoding error occurs if the u_1-codeword chosen by the likelihood encoder is not jointly typical $\big(\mathbf{U}_0(M_p,\mathbb{C}_0),\mathbf{U}_2(M_p,M_{22},\mathbb{C}_2)\big)$. This is described by the event

$$\mathcal{E}=\Big\{\big(\mathbf{U}_0(1,\mathbb{C}_0),\mathbf{U}_1(1,1,I,1,\mathbb{C}_1),\mathbf{U}_2(1,1,\mathbb{C}_2)\big)\notin\mathcal{T}^{(n)}_\epsilon(Q_{U_0,U_1,U_2})\Big\}. \tag{98}$$

To expand the expected probability of (98) over the ensemble of codebooks, we use the definitions of the random variables $\mathbb{C}_j$, for $j=0,1,2$, and $\mathbb{C}$ given in Section VII-C and denote $\mathcal{T}\triangleq\mathcal{T}^{(n)}_\epsilon(Q_{U_0,U_1,U_2})$. We have

$$\begin{aligned}&\mathbb{E}_{\mathbb{C}}\mathbb{P}\Big(\big(\mathbf{U}_0(M_p,\mathbb{C}_0),\mathbf{U}_1(M_p,M_1,I,W,\mathbb{C}_1),\mathbf{U}_2(M_p,M_{22},\mathbb{C}_2)\big)\notin\mathcal{T}\Big|\mathbb{C}\Big)\\&\quad\overset{(a)}{=}\mathbb{E}_{\mathbb{C}}\mathbb{P}\Big(\big(\mathbf{U}_0(1,\mathbb{C}_0),\mathbf{U}_1(1,1,I,1,\mathbb{C}_1),\mathbf{U}_2(1,1,\mathbb{C}_2)\big)\notin\mathcal{T}\Big|\mathbb{C},(M_p,M_1,M_{22},W)=\mathbf{1}\Big)\\&\quad=\mathbb{E}_{\mathbb{C}}\Bigg[\sum_{i,\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2}\mathbb{1}_{\left\{\left(\mathbf{U}_0(1,\mathbb{C}_0),\mathbf{U}_2(1,1,\mathbb{C}_2)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}}f^{(\mathrm{LE})}_{\mathrm{BC}}(i|1,\mathbf{u}_0,\mathbf{u}_2,\mathbb{C}_1)\\&\qquad\qquad\times\mathbb{1}_{\left\{\mathbf{U}_1(1,1,i,1,\mathbb{C}_1)=\mathbf{u}_1\right\}}\mathbb{1}_{\left\{(\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2)\notin\mathcal{T}\right\}}\Bigg]\\&\quad\overset{(b)}{=}\mathbb{E}_{\mathbb{C}_1}\Bigg[\sum_{i,\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2}Q^n_{U_0,U_2}(\mathbf{u}_0,\mathbf{u}_2)f^{(\mathrm{LE})}_{\mathrm{BC}}(i|1,\mathbf{u}_0,\mathbf{u}_2,\mathbb{C}_1)\mathbb{1}_{\left\{\mathbf{U}_1(1,1,i,1,\mathbb{C}_1)=\mathbf{u}_1\right\}}\mathbb{1}_{\left\{(\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2)\notin\mathcal{T}\right\}}\Bigg]\\&\quad\overset{(c)}{=}\mathbb{E}_{\mathbb{C}_1}\mathbb{P}_{Q^n_{U_0,U_2}}\Big(\big(\mathbf{U}_0,\mathbf{U}_1(1,1,I,1,\mathbb{C}_1),\mathbf{U}_2\big)\notin\mathcal{T}\Big|\mathbb{C}_1\Big)\end{aligned} \tag{99}$$

where:

(a) follows by the symmetry of the code construction;

(b) uses the law of total expectation and the independence of $\big(\mathbf{U}_0(1,\mathbb{C}_0),\mathbf{U}_2(1,1,\mathbb{C}_2)\big)$ and $\mathbb{C}_1$ in a similar manner to steps (b)-(d) in the derivation of (54) in Section VII-C;

(c) is because $\mathbb{P}_{Q^n_{U_0,U_2}}(\cdot)$ denotes that $(\mathbf{U}_0,\mathbf{U}_2)\sim Q^n_{U_0,U_2}$.

By Lemma 2, if (50a)-(50b) hold then the RHS of (99) approaches 0 as $n\to\infty$, which shows that $\mathbb{P}(\mathcal{E})\to 0$ as $n\to\infty$.

Decoding errors: Let $i\in\mathcal{I}$ be the realization of I chosen by the encoder and $\hat{m}_{12}\in\mathcal{M}_{12}$ be the cooperation index conveyed via the link from Decoder 1 to Decoder 2. To account for decoding errors, define the events

$$\mathcal{D}_1(m_p,m_1,\tilde{i},w)=\Big\{\big(\mathbf{U}_0(m_p,\mathbb{C}_0),\mathbf{U}_1(m_p,m_1,\tilde{i},w,\mathbb{C}_1),\mathbf{Y}_1\big)\in\mathcal{T}_\epsilon^{(n)}(Q_{U_0,U_1,Y_1})\Big\} \tag{100a}$$

$$\mathcal{D}_2(m_p,m_{22})=\Big\{\big(\mathbf{U}_0(m_p,\mathbb{C}_0),\mathbf{U}_2(m_p,m_{22},\mathbb{C}_2),\mathbf{Y}_2\big)\in\mathcal{T}_\epsilon^{(n)}(Q_{U_0,U_2,Y_2})\Big\}. \tag{100b}$$

By the union bound, the expected error probability is bounded as

$$\begin{aligned}\mathbb{E}P_e(\mathbb{C})\le\mathbb{P}\big(\mathcal{E}\big)+\Big(1-\mathbb{P}\big(\mathcal{E}\big)\Big)\Bigg[&\underbrace{\mathbb{P}\Big(\mathcal{D}_1^c(1,1,i,1)\Big|\mathcal{E}^c\Big)}_{P_1^{[1]}}+\underbrace{\mathbb{P}\left(\bigcup_{\tilde{m}_p\neq 1}\mathcal{D}_1(\tilde{m}_p,1,i,1)\middle|\mathcal{E}^c\right)}_{P_1^{[2]}}\\&+\underbrace{\mathbb{P}\left(\bigcup_{\substack{\tilde{i},\tilde{m}_1\neq 1,\\ \tilde{w}\neq 1}}\mathcal{D}_1(1,\tilde{m}_1,\tilde{i},\tilde{w})\middle|\mathcal{E}^c\right)}_{P_1^{[3]}}+\underbrace{\mathbb{P}\left(\bigcup_{\substack{\tilde{i},\tilde{m}_p\neq 1\\ \tilde{m}_1\neq 1,\tilde{w}\neq 1}}\mathcal{D}_1(\tilde{m}_p,\tilde{m}_1,\tilde{i},\tilde{w})\middle|\mathcal{E}^c\right)}_{P_1^{[4]}}\\&+\underbrace{\mathbb{P}\Big(\mathcal{D}_2^c(1,1)\Big|\mathcal{E}^c\Big)}_{P_2^{[1]}}+\underbrace{\mathbb{P}\left(\bigcup_{\substack{\tilde{m}_p\neq 1:\\ \tilde{m}_p\in\mathcal{B}(\hat{m}_{12})}}\mathcal{D}_2(\tilde{m}_p,1)\middle|\mathcal{E}^c\right)}_{P_2^{[2]}}+\underbrace{\mathbb{P}\left(\bigcup_{\tilde{m}_{22}\neq 1}\mathcal{D}_2(1,\tilde{m}_{22})\middle|\mathcal{E}^c\right)}_{P_2^{[3]}}\\&+\underbrace{\mathbb{P}\left(\bigcup_{\substack{\tilde{m}_p\neq 1,\tilde{m}_{22}\neq 1:\\ \tilde{m}_p\in\mathcal{B}(\hat{m}_{12})}}\mathcal{D}_2(\tilde{m}_p,\tilde{m}_{22})\middle|\mathcal{E}^c\right)}_{P_2^{[4]}}\Bigg].\end{aligned} \tag{101}$$

Note that $\big\{P_j^{[k]}\big\}_{k=1}^4$ correspond to errors that occur at Decoder j, where $j=1,2$. We proceed with the following steps:

1) $P_j^{[1]}$, for $j=1,2$, approaches 0 as $n\to\infty$ by the LLN.

2) For $P_1^{[3]}$, consider the bound

$$\begin{aligned} P_1^{[3]} &\overset{(a)}{\leq} \sum_{\tilde{i},\tilde{m}_1\neq 1,\tilde{w}\neq 1} 2^{-n\left(I(U_1;Y_1|U_0)-\delta_1^{[3]}(\epsilon)\right)} \\ &\leq 2^{n(R_1+R'+\tilde{R})} 2^{-n\left(I(U_1;Y_1|U_0)-\delta_1^{[3]}(\epsilon)\right)} \\ &= 2^{n\left(R_1+R'+\tilde{R}-I(U_1;Y_1|U_0)+\delta_1^{[3]}(\epsilon)\right)} \end{aligned}$$

where (a) is because for every $\tilde{i} \in \mathcal{I}$, $\tilde{m}_1 \neq 1$ and $\tilde{w} \neq 1$, $\mathbf{U}_1(1, \tilde{m}_1, \tilde{i}, \tilde{w}_j, \mathbb{C}_1)$ is independent of $\mathbf{Y}_1$ while both $\mathbf{U}_1(1, \tilde{m}_1, \tilde{i}, \tilde{w}_j, \mathbb{C}_1)$ and $\mathbf{Y}_1$ are drawn conditioned on $\mathbf{U}_0(1, \mathbb{C}_0)$. Moreover, we have $\delta_1^{[3]}(\epsilon) \to 0$ as $\epsilon \to 0$. Hence, for $P_1^{[3]}$ to vanish as $n \to \infty$, we take:

$$R_1 + R' + \tilde{R} < I(U_1; Y_1|U_0). \tag{102}$$

3) For $P_1^{[4]}$, we have

$$\begin{aligned} P_1^{[4]} &\overset{(a)}{\leq} \sum_{\tilde{i},\tilde{m}_p\neq 1,\tilde{m}_1\neq 1,\tilde{w}\neq 1} 2^{-n\left(I(U_0,U_1;Y_1)-\delta_1^{[4]}(\epsilon)\right)} \\ &\leq 2^{n(R_p+R_1+R'+\tilde{R})} 2^{-n\left(I(U_0,U_1;Y_1)-\delta_1^{[4]}(\epsilon)\right)} \\ &= 2^{n\left(R_p+R_1+R'+\tilde{R}-I(U_0,U_1;Y_1)+\delta_1^{[4]}(\epsilon)\right)} \end{aligned}$$

where (a) follows since for every $\tilde{i} \in \mathcal{I}$, $\tilde{m}_p \neq 1$, $\tilde{m}_1 \neq 1$ and $\tilde{w} \neq 1$, $\mathbf{U}_0(\tilde{m}_p, \mathbb{C}_0)$ and $\mathbf{U}_1(\tilde{m}_p, \tilde{m}_1, \tilde{i}, \tilde{w}, \mathbb{C}_1)$ are jointly letter-typical with each other but independent of $\mathbf{Y}_1$. Again, $\delta_1^{[4]}(\epsilon) \to 0$ as $\epsilon \to 0$, and therefore, we have $P_1^{[4]} \to 0$ as $n \to \infty$ if

$$R_p + R_1 + R' + \tilde{R} < I(U_0, U_1; Y_1). \tag{103}$$

4) By repeating similar steps to upper bound $P_1^{[2]}$, the obtained rate bound is redundant. This is since for every $\tilde{m}_p \neq 1$ the codewords $\mathbf{U}_0(\tilde{m}_p, \mathbb{C}_0)$ and $\mathbf{U}_1(\tilde{m}_p, 1, i, 1, \mathbb{C}_1)$ are independent of $\mathbf{Y}_1$. Hence, to ensure that $P_1^{[2]}$ vanishes as $n \to \infty$, we take

$$R_p < I(U_0, U_1; Y_1) \tag{104}$$

and the RHS coincides with the RHS of (103), while the LHS is with respect to R_p only. Clearly, (103) is the dominating constraint.

5) By a similar argument, we have that $P_2^{[j]}$, for $j = 2, 3, 4$, vanishes with n if

$$R_{22} < I(U_2; Y_2|U_0) \tag{105}$$

$$R_p + R_{22} - R_{12} < I(U_0, U_2; Y_2). \tag{106}$$

Summarizing the above results, and substituting $R_p = R_0 + R_{20}$, we find that the RHS of (101) decays as the

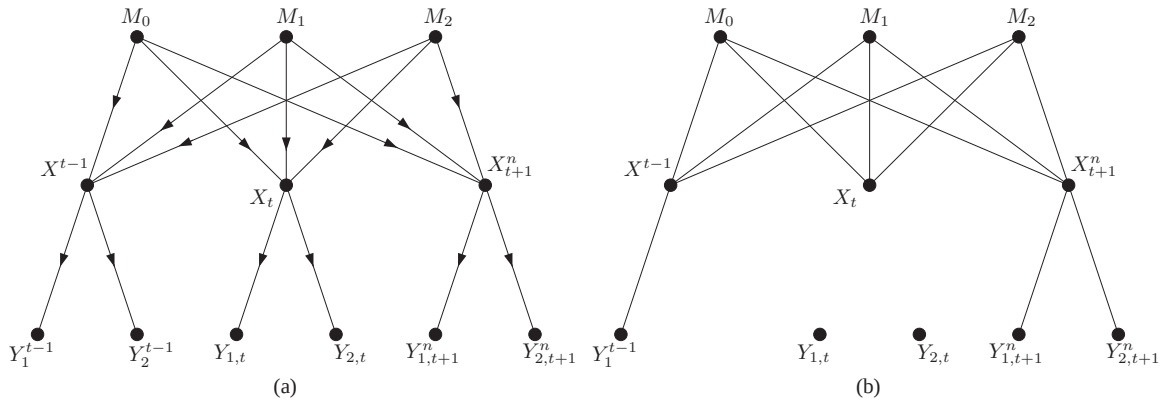


Fig. 11: (a) The FDG that stems from (109): (108) follows since $\mathcal{C} = \{X_t\}$ d-separates $\mathcal{A} = \{Y_{2,t}\}$ from $\mathcal{B} = \{M_1, M_2, Y_1^n, Y_{2,t+1}^n\}$. (b) The undirected graph obtained from the FDG after the manipulations described in Definition [51, Definition 1]. Both FDGs omit the dependence of the channel outputs on the noise.

blocklength $n \to \infty$ if the conditions in (50) are met.

APPENDIX E
PROOF OF THE MARKOV RELATION IN (69) AND (76)

We prove that (69) and (76) form Markov chains by using the notions of d-separation and fd-separation in functional dependence graphs (FDGs), for which we use the formulation from [51].

A. Proof of (69)

By the definitions of the auxiliaries W and V, it suffices to show that

$$(M_0, M_1, M_2, M_{12}, Y_1^{t-1}, Y_{2,t+1}^n, Y_{1,t}) - X_t - Y_{2,t} \tag{107}$$

forms a Markov chain for every $t \in [1:n]$. In fact, we prove the stronger relation

$$(M_0, M_1, M_2, Y_1^n, Y_{2,t+1}^n) - X_t - Y_{2,t} \tag{108}$$

from which (107) follows because M_{12} is a function of Y_1^n. Since the channel is SD, memoryless and without feedback, for every $(m_0, m_1, m_2) \in \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2$, $(x^n, y_1^n, y_2^n) \in \mathcal{X}^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ and $t \in [1:n]$, we have

$$\begin{aligned} P(m_0, m_1, m_2, x^n, y_1^n, y_2^n) = {} & P(m_0)P(m_1)P(m_2)P(x^n|m_0, m_1, m_2)P(y_1^{t-1}|x^{t-1})P(y_2^{t-1}|x^{t-1}) \\ & \times P(y_{1,t}|x_t)P(y_{2,t}|x_t)P(y_{1,t+1}^n|x_{t+1}^n)P(y_{2,t+1}^n|x_{t+1}^n) \end{aligned} \tag{109}$$

Fig. 11(a) shows the FDG induced by (109). The structure of FDGs allows one to establish the conditional statistical independence of sets of random variables by using d-separation. The Markov relation in (108) follows by setting $\mathcal{A} = \{Y_{2,t}\}$, $\mathcal{B} = \{M_0, M_1, M_2, Y_1^n, Y_{2,t+1}^n\}$ and $\mathcal{C} = \{X_t\}$, and noting that $\mathcal{C}$ d-separates $\mathcal{A}$ from $\mathcal{B}$ by applying the manipulations described in [51, Definition 1].

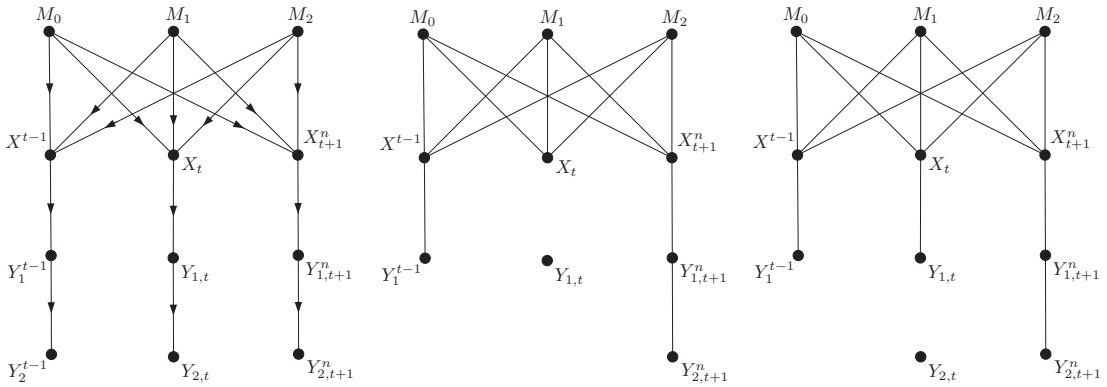


Fig. 12: (a) The FDG that stems from (111): (110) follows since $\mathcal{C}_j$ d-separates $\mathcal{A}_j$ from $\mathcal{B}_j$, for $j = 1, 2$. (b) The undirected graph that corresponds to $\mathcal{A}_1$, $\mathcal{B}_1$ and $\mathcal{C}_1$. (c) The undirected graph that corresponds to $\mathcal{A}_2$, $\mathcal{B}_2$ and $\mathcal{C}_2$. The FDGs omit the dependence of the channel outputs on the noise.

B. Proof of (76)

To prove (76), is suffices to show that Markov relations

$$(M_0, M_2, Y_1^{t-1}, Y_{2,t+1}^n) - X_t - Y_{1,t} \tag{110a}$$

$$(M_0, M_2, Y_1^{t-1}, Y_{2,t+1}^n, X_t) - Y_{1,t} - Y_{2,t} \tag{110b}$$

hold for every $t \in [1 : n]$. By the PD property of the channel, and because it is memoryless and without feedback, for every $(m_0, m_1, m_2) \in \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2$, $(x^n, y_1^n, y_2^n) \in \mathcal{X}^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ and $t \in [1 : n]$, we have

$$\begin{aligned} P(m_0, m_1, m_2, x^n, y_1^n, y_2^n) = {} & P(m_0)P(m_1)P(m_2)P(x^n|m_0, m_1, m_2)P(y_1^{t-1}|x^{t-1})P(y_2^{t-1}|y_1^{t-1}) \\ & \times P(y_{1,t}|x_t)P(y_{2,t}|y_{1,t})P(y_{1,t+1}^n|x_{t+1}^n)P(y_{2,t+1}^n|y_{1,t+1}^n). \end{aligned} \tag{111}$$

The FDG induced by (111) is shown in Fig. 12(a). Set $\mathcal{A}_1 = \{Y_{1,t}\}$, $\mathcal{B}_1 = \{M_0, M_2, Y_1^{i-1}, Y_{2,t+1}^n\}$ and $\mathcal{C}_1 = \{X_t\}$, and $\mathcal{A}_2 = \{Y_{2,t}\}$, $\mathcal{B}_2 = \{M_0, M_2, Y_1^{i-1}, Y_{2,t+1}^n, X_t\}$ and $\mathcal{C}_2 = \{Y_{1,t}\}$. The relations in (110) follow by noting that $\mathcal{C}_j$ d-separates $\mathcal{A}_j$ from $\mathcal{B}_j$, for $j = 1, 2$ by applying the manipulations described in [51, Definition 1].

REFERENCES

[1] A. D. Wyner. The wire-tap channel. *Bell Sys. Techn.*, 54(8):1355–1387, Oct. 1975.

[2] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 24(3):339–348, May 1978.

[3] R. Liu, I. Maric, P. Spasojević, and R. D. Yates. Discrete memoryless interference and broadcast channels with confidential messages: Secrecy rate regions. *IEEE Trans. Inf. Theory*, 54(6):2493–2507, Jun. 2008.

[4] Y. Zhao, P. Xu, Y. Zhao, W. Wei, and Y. Tang. Secret communications over semi-deterministic broadcast channels. In *Fourth Int. Conf. Commun. and Netw. in China (CHINACOM)*, Xian, China, Aug. 2009.

[5] W. Kang and N. Liu. The secrecy capacity of the semi-deterministic broadcast channel. In *Proc. Int. Symp. Inf. Theory*, Seoul, Korea, Jun.-Jul. 2009.

[6] Z. Goldfeld, G. Kramer, and H. H. Permuter. Broadcast channels with privacy leakage constraints. *Submitted for publication to IEEE Trans. Inf. Theory*, 2015. Available on ArXiv at http://arxiv.org/abs/1504.06136.

[7] E. Ekrem and S. Ulukus. Secrecy in cooperative relay broadcast channels. *IEEE Trans. Inf. Theory*, 57(1):137–155, Jan. 2011.

[8] R. Liu and H. Poor. Secrecy capacity region of a multiple-antenna Gaussian broadcast channel with confidential messages. *IEEE Trans. Inf. Theory*, 55(3):1235–1249, Mar. 2009.

[9] T. Liu and S. Shamai. A note on the secrecy capacity of the multiple-antenna wiretap channel. *IEEE Trans. Inf. Theory*, 6(6):2547–2553, Jun. 2009.

[10] R. Liu, T. Liu, H. V. Poor, and S. Shamai. Multiple-input multiple-output Gaussian broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 56(9):4215–4227, Sep. 2010.

[11] A. Khisti and G. W. Wornell. Secure transmission with multiple antennas - part II: The MIMOME channel. *IEEE Trans. Inf. Theory*, 56(11):5515–5532, Nov. 2010.

[12] E. Ekrem and S. Ulukus. The secrecy capacity region of the Gaussian MIMO multi-receiver wiretap channel. *IEEE Trans. Inf. Theory*, 57(4):2083–2114, Apr. 2011.

[13] F. Oggier and B. Hassibi. The secrecy capacity of the MIMO wiretap channel. *IEEE Trans. Inf. Theory*, 57(8):4961–4972, Aug. 2011.

[14] E. Ekrem and S. Ulukus. Secrecy capacity of a class of broadcast channels with an eavesdropper. *EURASIP Journal on Wireless Commun. and Netw.*, 2009(1):1–29, Mar. 2009.

[15] G. Bagherikaram, A. Motahari, and A. Khandani. Secrecy capacity region of Gaussian broadcast channel. In *43rd Annual Conf. on Inf. Sci. and Sys. (CISS) 2009*, pages 152–157, Baltimore, MD, US, Mar. 2009.

[16] M. Benammar and P. Piantanida. Secrecy capacity region of some classes of wiretap broadcast channels. *IEEE Trans. Inf. Theory*, (10):5564–5582, Oct. 2015.

[17] U. Maurer. *Communications and Cryptography: Two Sides of One Tapestry*, chapter The Strong Secret Key Rate of Discrete Random Triples, pages 271–285. Springer US, Norwell, MA, USA, 1994.

[18] U. Maurer and S. Wolf. Information-theoretic key agreement: From weak to strong secrecy for free. In *Lecture Notes in Computer Science*, pages 351–368, 2000.

[19] M. Bloch and J. Barros. *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge Univ. Press, Cambridge, UK, Oct. 2011.

[20] I. Csiszár. Almost independence and secrecy capacity. *Prob. Inf. Trans.*, 32(1):40–47, Jan.-Mar. 1996.

[21] M. Hayashi. General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channels. *IEEE Trans. Inf. Theory*, 52(4):1562–1575, Apr. 2006.

[22] A. D. Wyner. The common information of two dependent random variables. *IEEE Trans. Inf. Theory*, 21(2):163–179, Mar. 1975.

[23] T. Han and S. Verdú. Approximation theory of output statistics. *IEEE Trans. Inf. Theory*, 39(3):752–772, May 1993.

[24] J. Hou and G. Kramer. Informational divergence approximations to product distributions. In *13th Canadian Workshop Inf. Theory*, Toronto, Ontario, Canada, Jun. 2013.

[25] P. W. Cuff. Distributed channel synthesis. *IEEE. Trans. Inf. Theory*, 59(11):7071–7096, Nov. 2013.

[26] C. Schieler and P. Cuff. Rate-distortion theory for secrecy systems. *IEEE Trans. on Inf. Theory*, 66(12):7584–7605, Dec. 2014.

[27] C. Schieler and P. Cuff. The henchman problem: Measuring secrecy by the minimum distortion in a list. *Submitted to IEEE Trans. on Inf. Theory*, 2014. Available on ArXiv at http://arxiv.org/abs/1410.2881.

[28] E. Song, P. Cuff, and V. Poor. A rate-distortion based secrecy system with side information at the decoders. In *Proc. 52nd Annu. Allerton Conf. Commun., Control and Comput.*, Monticell, Illinois, United States, Sep. 2014.

[29] S. Satpathy and P. Cuff. Secure coordination with a two-sided helper. In *Proc. Int. Symp. Inf. Theory (ISIT-2014)*, Honolulu, Hawaii, US, Jun.-Jul. 2014.

[30] M. Bloch and N. Laneman. Strong secrecy from channel resolvability. *IEEE Trans. Inf. Theory*, 59(12):8077–8098, Dec. 2013.

[31] J. Hou and G. Kramer. Effective secrecy: Reliability, confusion and stelth. In *Proc. Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun.-Jul. 2014.

[32] T. S. Han, H. Endo, and M. Sasaki. Reliability and secrecy functions of the wiretap channel under cost constraint. *IEEE Trans. Inf. Theory*, 60(11):6819–6843, Nov. 2014.

[33] E. Song, P. Cuff, and V. Poor. The likelihood encoder for lossy compression. *Accepted for publication in IEEE Trans. Inf. Theory*, Dec. 2015.

[34] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New-York, 2nd edition, 2006.

[35] Z. Goldfeld, H. H. Permuter, and G. Kramer. Duality of a source coding problem and the semi-deterministic broadcast channel with rate-limited cooperation. *Accepted for publication in IEEE Trans. Inf. Theory*, 2015. Available on ArXiv at http://arxiv.org/abs/1405.7812.

[36] E. C. van der Meulen. Random coding theorems for the general discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory*, IT-21(2):180–190, May 1975.

[37] S. I. Gelfand. Capacity of one broadcast channel. *Probl. Pered. Inf. (Problems of Inf. Transm.)*, 13(3):106108, Jul./Sep. 1977.

[38] J. L. Massey. *Applied Digital Information Theory*. ETH Zurich, Zurich, Switzerland, 1980-1998.

[39] A. Orlitsky and J. Roche. Coding for computing. *IEEE Trans. Inf. Theory*, 47(3):903–917, Mar. 2001.

[40] A. Gohari and V. Anantharam. Evaluation of Marton's inner bound for the general broadcast channel. *IEEE Trans. Inf. Theory*, 58(2):608–619, Feb. 2012.

[41] Y. Liang and V. V. Veeravalli. Cooperative relay broadcast channels. *IEEE Trans. Inf. Theory*, 53(3):900–928, Mar. 2007.

[42] Y. Liang and G. Kramer. Rate regions for relay broadcast channels. *IEEE Trans. Inf. Theory*, 53(10):3517–3535, Oct. 2007.

[43] L. Dikstein, H. H. Permuter, and Y. Steinberg. On state dependent broadcast channels with cooperation. *Accepted for publication in IEEE Trans. Inf. Theory*, 2015. Available on ArXiv at http://arxiv.org/abs/1405.5083.

[44] R. Zamir, S. Shamai, and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Inf. Theory*, 48(6):1205–1276, Jun. 2002.

[45] R. J. Barron, B. Chen, and G. W. Wornell. The duality between information embedding and source coding with side information and some applications. *IEEE Trans. Inf. Theory*, 49(5):1159–1180, May 2003.

[46] A. Khina, T. Philosof, U. Erez, and R. Zamir. Binary dirty MAC with common state information. In *Proc. 26-th Convention of Electrical and Electronics Engineers (IEEEI-2010)*, Eilat, Israel, Nov. 2010.

[47] S. I. Gelfand and M. S. Pinsker. Capacity of a broadcast channel with one deterministic component. *Prob. Pered. Inf. (Problems of Inf. Transm.)*, 16(1):17–25, Jan.-Mar. 1980.

[48] R. Dabora and S. D. Servetto. Broadcast channels with cooperating decoders. *IEEE Trans. Inf. Theory*, 52:5438–5454, 2006.

[49] I. B. Gattegno, Z. Goldfeld, and H. H. Permuter. Fourier-Motzkin elimination software for information theoretic inequalities. *IEEE Inf. Theory Society Newsletter*, 65(3):25–28, Sep. 2015.

[50] G. Kramer. Teaching IT: An identity for the Gelfand-Pinsker converse. *IEEE Inf. Theory Society Newsletter*, 61(4):4–6, Dec. 2011.

[51] G. Kramer. Capacity results for the discrete memoryless networks. *IEEE. Trans. Inf. Theory*, 49(1):4–21, Jan. 2003.