Strong Secrecy for Cooperative Broadcast Channels

Ziv Goldfeld, *Student Member, IEEE*, Gerhard Kramer, *Fellow, IEEE*, Haim H. Permuter, *Senior Member, IEEE*, and Paul Cuff, *Member, IEEE*

***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 with 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. An example that illustrates this loss is provided.**

***Index Terms*— Broadcast channel, channel resolvability, conferencing, cooperation, likelihood encoder, physical-layer security, 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.

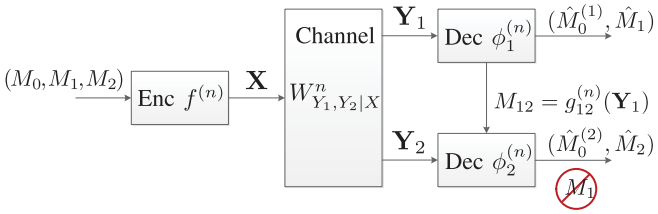


Fig. 1. Cooperative BCs with one confidential message.

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 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 variation (TV). 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].

Manuscript received January 6, 2016; revised July 3, 2016; accepted September 11, 2016. Date of publication October 26, 2016; date of current version December 20, 2016. Z. Goldfeld and H. H. Permuter were supported in part by the Cyber Security Research Center within the Ben-Gurion University of the Negev, in part by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant under Grant 337752, and in part by the Israel Science Foundation. G. Kramer was supported by an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research. P. Cuff was supported in part by the National Science Foundation under Grant CCF-1350595 and Grant CCF-1116013 and in part by the Air Force Office of Scientific Research under Grant FA9550-15-1-0180 and Grant FA9550-12-1-0196. This paper was presented at the 2015 IEEE International Symposium on Information Theory and the 2016 International Zurich Seminar on Communications.

Z. Goldfeld and H. H. Permuter are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beersheba 8499000, Israel (e-mail: gziv@post.bgu.ac.il; haimp@bgu.ac.il).

G. Kramer is with the Institute for Communications Engineering, Technical University of Munich, D-80333 Munich, Germany (gerhard.kramer@tum.de).

P. Cuff is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: cuff@princeton.edu).

Communicated by A. Khisti, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2016.2622058

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 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 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, Th. 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 show this 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 suboptimal 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 (BW-BC) [36], [37] and a Gaussian BC. An explicit strong secrecy-achieving coding strategy for an extreme point of the BW-BC region is given. Although the BW-BC'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 BW-BC 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. Notations and Preliminary Definitions

A. Notations

We use the following notations. As customary $\mathbb{N}$ is the set of natural numbers (which does not include 0), while $\mathbb{R}$ denotes the reals. We further define $\mathbb{R}_+ = \{x \in \mathbb{R} | x \geq 0\}$ and $\mathbb{R}_{++} = \mathbb{R}_+ \setminus \{0\}$. Given two real numbers a, b, we denote by $[a : b]$ the set of integers $\left\{n \in \mathbb{N} \middle| \lceil a \rceil \leq n \leq \lfloor b \rfloor\right\}$. Calligraphic letters denote sets, e.g., $\mathcal{X}$, the complement of $\mathcal{X}$ is denoted by $\mathcal{X}^c$, while $|\mathcal{X}|$ stands for its cardinality. $\mathcal{X}^n$ denoted the n-fold Cartesian product of $\mathcal{X}$. An element of $\mathcal{X}^n$ is denoted by $x^n = (x_1, x_2, \ldots, x_n)$; whenever the dimension n is clear from the context, vectors (or sequences) are denoted by boldface letters, e.g., $\mathbf{x}$. A substring of $\mathbf{x} \in \mathcal{X}^n$ is denoted by $x_i^j = (x_i, x_{i+1}, \ldots, x_j)$, for $1 \leq i \leq j \leq n$; when $i = 1$, the subscript is omitted. We also define $x^{n\setminus i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

Let $\left(\mathcal{X}, \mathcal{F}, \mathbb{P}\right)$ be a probability space, where $\mathcal{X}$ is the sample space, $\mathcal{F}$ is the σ-algebra and $\mathbb{P}$ is the probability measure. Random variables over $\left(\mathcal{X}, \mathcal{F}, \mathbb{P}\right)$ are denoted by uppercase letters, e.g., X, with conventions for random vectors similar to those for deterministic sequences. The probability of an event $\mathcal{A} \in \mathcal{F}$ is denoted by $\mathbb{P}(\mathcal{A})$, while $\mathbb{P}(\mathcal{A}|\mathcal{B})$ denotes conditional probability of $\mathcal{A}$ given $\mathcal{B}$. We use $\mathbb{1}_{\mathcal{A}}$ to denote the indicator function of $\mathcal{A}$. The set of all probability mass

functions (PMFs) on a finite set $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$, i.e.,

$$\mathcal{P}(\mathcal{X}) = \left\{ P : \mathcal{X} \to [0,1] \middle| \sum_{x\in\mathcal{X}} P(x) = 1 \right\}. \tag{1}$$

PMFs are denoted by the uppercase letters such as P or Q, with a subscript that identifies the random variable and its possible conditioning. For example, for a discrete probability space $\left(\mathcal{X}, \mathcal{F}, \mathbb{P}\right)$ and two correlated random variables X and Y over that space, we use P_X, $P_{X,Y}$ and $P_{X|Y}$ to denote, respectively, the marginal PMF of X, the joint PMF of (X, Y) and the conditional PMF of X given Y. In particular, $P_{X|Y}$ represents the stochastic matrix whose elements are given by $P_{X|Y}(x|y) = \mathbb{P}\big(X = x|Y = y\big)$. Expressions such as $P_{X,Y} = P_X P_{Y|X}$ are to be understood as $P_{X,Y}(x, y) = P_X(x)P_{Y|X}(y|x)$, for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Accordingly, when three random variables X, Y and Z satisfy $P_{X|Y,Z} = P_{X|Y}$, they form a Markov chain, which we denote by $X - Y - Z$. We omit subscripts if the arguments of a PMF are lowercase versions of the random variables. The support of a PMF P and the expectation of a random variable $X \sim P$ are denoted by $\text{supp}(P)$ and $\mathbb{E}_P\big[X\big]$, respectively; when the distribution of X is clear from the context we write its expectation simply as $\mathbb{E}\big[X\big]$. Similarly, H_P and I_P denote entropy and mutual information that are calculated with respect to an underlying PMF P.

For a discrete measurable space $(\mathcal{X}, \mathcal{F})$, a PMF $Q \in \mathcal{P}(\mathcal{X})$ gives rise to a probability measure on $(\mathcal{X}, \mathcal{F})$, which we denote by $\mathbb{P}_Q$; accordingly, $\mathbb{P}_Q\big(\mathcal{A}\big) = \sum_{x\in\mathcal{A}} Q(x)$, for every $\mathcal{A} \in \mathcal{F}$. For a sequence of random variables X^n, 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 $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = P_{Y|X}^n(\mathbf{y}|\mathbf{x})$. The conditional product PMF $P_{Y|X}^n$ given a specific sequence $\mathbf{x} \in \mathcal{X}^n$ is denoted by $P_{Y|X}^n(\cdot|\mathbf{x})$.

Let $\mathcal{X}$ be a finite set. The empirical PMF $\nu_{\mathbf{x}}$ of a sequence $\mathbf{x} \in \mathcal{X}^n$ is

$$\nu_{\mathbf{x}}(x) \triangleq \frac{N(x|\mathbf{x})}{n}, \tag{2}$$

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

$$\mathcal{T}_\delta^n(P) = \left\{ \mathbf{x} \in \mathcal{X}^n \middle| \ \big|\nu_{\mathbf{x}}(x) - P(x)\big| \le \delta P(x), \quad \forall x \in \mathcal{X} \right\}. \tag{3}$$

B. Measures of Distribution Proximity

Definition 1 (Relative Entropy): Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let P and Q be two probability measures on $\mathcal{F}$, with $P \ll Q$ (i.e., P is absolutely continuous with respect to Q). The relative entropy between P and Q is

$$D(P||Q) = \int_{\mathcal{X}} dP \log\left(\frac{dP}{dQ}\right), \tag{4}$$

where $\frac{dP}{dQ}$ denotes the Radon-Nikodym derivative of P with respect to Q. If the sample space $\mathcal{X}$ is countable, (4) *reduces to*

$$D(P||Q) = \sum_{x\in supp(P)} P(x) \log\left(\frac{P(x)}{Q(x)}\right). \tag{5}$$

Definition 2 (Total Variation): Let $(\mathcal{X}, \mathcal{F})$ be a measurable and P and Q be two probability measures on $\mathcal{F}$. The total variation between P and Q is

$$||P - Q||_{\mathsf{TV}} = \sup_{\mathcal{A}\in\mathcal{F}} \big|P(\mathcal{A}) - Q(\mathcal{A})\big|. \tag{6}$$

If the sample space $\mathcal{X}$ is countable, (6) *reduces to*

$$||P - Q||_{\mathsf{TV}} = \frac{1}{2} \sum_{x\in\mathcal{X}} \big|P(x) - Q(x)\big|. \tag{7}$$

Remark 1 (TV Dominates Relative Entropy): Pinsker's inequality shows that relative entropy is larger than TV. A reverse inequality is sometimes valid. For example, if $\mathcal{X}$ is a finite set, $\{P_n\}_{n\in\mathbb{N}}$ is a sequence of distributions with $P_n \in \mathcal{P}(\mathcal{X}^n)$, $Q \in \mathcal{P}(\mathcal{X})$ and $P_n \ll Q^n$ for every $n \in \mathbb{N}$, then[1] *(see [25, eq. (29)])*

$$D(P_n||Q^n) \in \mathcal{O}\left(\left[n + \log\frac{1}{||P_n - Q^n||_{\mathsf{TV}}}\right] ||P_n - Q^n||_{\mathsf{TV}}\right). \tag{8}$$

In particular, (8) *implies that an exponential decay of the TV in n produces an (almost, up to a $\frac{\log n}{n}$ term) exponential decay of the relative entropy with the same exponent.*

III. A Channel Resolvability Lemma for Strong Secrecy

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 $Q_{S_0,S} \in \mathcal{P}(\mathcal{S}_0 \times \mathcal{S})$. The encoder superimposes its codebook on S_0 and then uses a *likelihood encoder* with 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 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. A formal description of the setup is as follows.

Let $\mathcal{S}_0$, $\mathcal{S}$, $\mathcal{U}$ and $\mathcal{V}$ be finite sets. Fix any $Q_{S_0,S,U,V} \in \mathcal{P}(\mathcal{S}_0 \times \mathcal{S} \times \mathcal{U} \times \mathcal{V})$ and let W be a random variable uniformly distributed over[2] $\mathcal{W}_n = \big[1 : 2^{n\tilde{R}}\big]$ that is independent of $(\mathbf{S}_0, \mathbf{S}) \sim Q_{S_0,S}^n$.

A. Codebook

For every $\mathbf{s}_0 \in \mathcal{S}_0^n$, let $\mathsf{B}_n(\mathbf{s}_0) \triangleq \big\{\mathbf{U}(\mathbf{s}_0, w, i)\big\}_{(w,i)\in\mathcal{W}_n\times\mathcal{I}_n}$, where $\mathcal{I}_n = \big[1 : 2^{nR'}\big]$, be a collection of $2^{n(\tilde{R}+R')}$ conditionally independent random vectors of length n, each distributed according to $Q_{U|S_0}^n(\cdot|\mathbf{s}_0)$. A realization of $\mathsf{B}_n(\mathbf{s}_0)$, for $\mathbf{s}_0 \in \mathcal{S}_0^n$,

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

[2] To simplify notation, from here on we assume that quantities of the form 2^{nR}, where $n \in \mathbb{N}$ and $R \in \mathbb{R}_+$, are integers. Otherwise, simple modifications of some of the subsequent expressions using floor operations are needed.

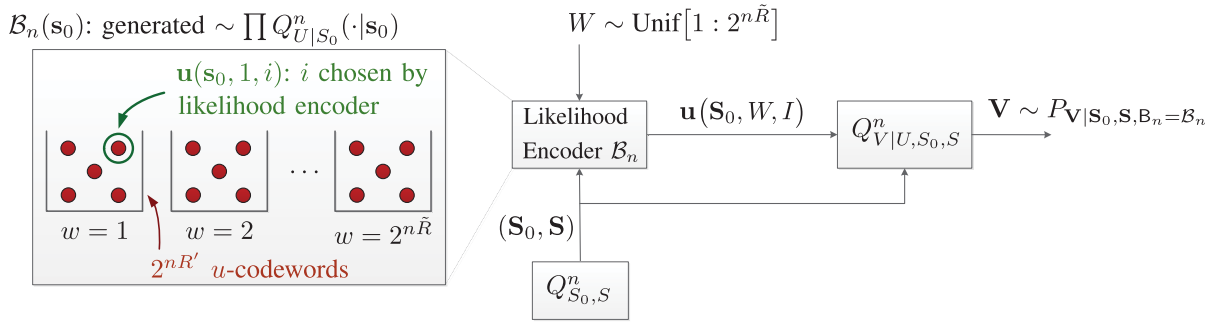


Fig. 2. Coding problem for approximating $P_{\mathbf{V}|\mathbf{S}_0,\mathbf{S},\mathsf{B}_n=\mathcal{B}_n} \approx Q^n_{V|S_0,S}$ under a resolvability codebook that is superimposed on $\mathbf{s}_0 \in \mathcal{S}_0^n$: For each $\mathbf{s}_0 \in \mathcal{S}_0^n$, the codebook $\mathcal{B}_n(\mathbf{s}_0)$ contains $2^{n(\tilde{R}+R')}$ u-codewords drawn independently according to $Q^n_{U|S_0}(\cdot|\mathbf{s}_0)$. The codewords are partitioned into $2^{n\tilde{R}}$ bins, each associated with a certain $w \in [1 : 2^{n\tilde{R}}]$. The u-codeword that is fed into the channel is selected by first randomly and uniformly drawing a bin index W from $[1 : 2^{n\tilde{R}}]$, and then drawing I from $[1 : 2^{nR'}]$ by means of the likelihood encoder from (10).

is denoted by $\mathcal{B}_n(\mathbf{s}_0) \triangleq \{\mathbf{u}(\mathbf{s}_0, w, i)\}_{(w,i)\in\mathcal{W}_n\times\mathcal{I}_n}$. Each codebook $\mathcal{B}_n(\mathbf{s}_0)$ can be thought of as comprising $2^{n\tilde{R}}$ bins, each associated with a different message $w \in \mathcal{W}_n$ and contains $2^{nR'}$ u-codewords. We also denote $\mathsf{B}_n \triangleq \{\mathsf{B}_n(\mathbf{s}_0)\}_{\mathbf{s}_0\in\mathcal{S}_0^n}$, which is referred to as the random resolvability codebook. A possible value of B_n is denoted by $\mathcal{B}_n$ and we set $\mathfrak{B}_n$ as the collection of all such possible values.

The above codebook construction induces a PMF $\lambda \in \mathcal{P}(\mathfrak{B}_n)$ over the codebook ensemble. For every $\mathcal{B}_n \in \mathfrak{B}_n$, we have

$$\lambda(\mathcal{B}_n) = \prod_{\mathbf{s}_0\in\mathcal{S}_0^n} \prod_{\substack{(w,i)\\ \in\mathcal{W}_n\times\mathcal{I}_n}} Q^n_{U|S_0}\big(\mathbf{u}(\mathbf{s}_0, w, i)\big|\mathbf{s}_0\big). \tag{9}$$

B. Encoding and Induced PMF

For each codebook $\mathcal{B}_n \in \mathfrak{B}_n$, consider the *likelihood encoder* described by conditional PMF

$$\hat{P}^{(\mathcal{B}_n)}(i|w, \mathbf{s}_0, \mathbf{s}) = \frac{Q^n_{S|U,S_0}\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0, w, i), \mathbf{s}_0\big)}{\sum\limits_{i'\in\mathcal{I}_n} Q^n_{S|U,S_0}\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0, w, i'), \mathbf{s}_0\big)}. \tag{10}$$

Upon observing $(w, \mathbf{s}_0, \mathbf{s})$, an index $i \in \mathcal{I}_n$ is drawn randomly according to (10). The codeword $\mathbf{u}(\mathbf{s}_0, w, i) \in \mathcal{B}_n(\mathbf{s}_0)$ is passed through the DMC $Q^n_{V|U,S_0,S}$. For a fixed codebook $\mathcal{B}_n \in \mathfrak{B}_n$, the induced joint distribution is

$$\begin{aligned} P^{(\mathcal{B}_n)}(\mathbf{s}_0, \mathbf{s}, w, i, \mathbf{u}, \mathbf{v}) = {} & Q^n_{S_0,S}(\mathbf{s}_0, \mathbf{s}) 2^{-n\tilde{R}} \hat{P}^{(\mathcal{B}_n)}(i|w, \mathbf{s}_0, \mathbf{s}) \\ & \times \mathbb{1}_{\{\mathbf{u}=\mathbf{u}(\mathbf{s}_0,w,i)\}} Q^n_{V|U,S_0,S}(\mathbf{v}|\mathbf{u}, \mathbf{s}_0, \mathbf{s}). \end{aligned} \tag{11}$$

Accounting for the random codebook generation, we also set

$$P(\mathcal{B}_n, \mathbf{s}_0, \mathbf{s}, w, i, \mathbf{u}, \mathbf{v}) = \lambda(\mathcal{B}_n) P^{(\mathcal{B}_n)}(\mathbf{s}_0, \mathbf{s}, w, i, \mathbf{u}, \mathbf{v}). \tag{12}$$

Lemma 1 (Sufficient Conditions for Approximation): For any $Q_{S_0,S,U,V} \in \mathcal{P}(\mathcal{S}_0 \times \mathcal{S} \times \mathcal{U} \times \mathcal{V})$, if $(\tilde{R}, R') \in \mathbb{R}_+^2$ satisfies

$$R' > I(U; S|S_0) \tag{13a}$$

$$R' + \tilde{R} > I(U; S, V|S_0), \tag{13b}$$

then

$$\mathbb{E}_{\mathsf{B}_n} D\Big(P_{\mathbf{V}|\mathbf{S}_0,\mathbf{S},\mathsf{B}_n}\Big\|Q^n_{V|S_0,S}\Big|Q^n_{S_0,S}\Big) \xrightarrow[n\to\infty]{} 0. \tag{14}$$

The proof of Lemma 1 (see Section VII-A) shows that the TV decays exponentially fast with the blocklength n. By Remark 1 this implies an almost 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 (13), then for any $w \in \mathcal{W}_n$ and $\epsilon > 0$, we have

$$\mathbb{E}_{\mathsf{B}_n}\mathbb{P}_P\Big(\big(\mathbf{S}_0, \mathbf{S}, \mathbf{U}(\mathbf{S}_0, w, I)\big) \notin \mathcal{T}^n_\epsilon(Q_{S_0,S,U})\Big|\mathsf{B}_n\Big) \xrightarrow[n\to\infty]{} 0. \tag{15}$$

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

IV. Cooperative Broadcast Channels With One Confidential Message

A. Problem Definition

The $\big(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, W_{Y_1,Y_2|X} : \mathcal{X} \to \mathcal{P}(\mathcal{Y}_1 \times \mathcal{Y}_2)\big)$ cooperative DM-BC with one confidential message is illustrated in Fig. 1. The channel has one sender and two receivers. The sender uniformly chooses a triple (m_0, m_1, m_2) of indices from the product set $[1 : 2^{nR_0}] \times [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$ and maps it to a sequence $\mathbf{x} \in \mathcal{X}^n$, which is the channel input (the mapping may be random). The sequence $\mathbf{x}$ is transmitted over a BC with transition probability $W_{Y_1,Y_2|X} : \mathcal{X} \to \mathcal{P}(\mathcal{Y}_1 \times \mathcal{Y}_2)$. 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 $(\hat{m}_0^{(j)}, \hat{m}_j)$ of (m_0, m_j). Furthermore, the message m_1 is to be kept secret from Decoder 2 and there is a one-sided noiseless cooperation link of rate R_{12} that extends 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 $\mathbf{y}_1$, $(\hat{m}_0^{(1)}, \hat{m}_1)$, or both.

Remark 2 (Specific Classes of BCs): We sometimes specialize to the following classes of BCs:

- *Semi-Deterministic BCs: A BC is SD if its channel transition matrix factors as $W_{Y_1,Y_2|X} = \mathbb{1}_{\{Y_1=y_1(X)\}} W_{Y_2|X}$, where $y_1 : \mathcal{X} \to \mathcal{Y}_1$ and $W_{Y_2|X} : \mathcal{X} \to \mathcal{P}(\mathcal{Y}_2)$.*

- *Physically-Degraded BCs:* *A BC is PD if its channel transition matrix factors as* $W_{Y_1,Y_2|X} = W_{Y_1|X}W_{Y_2|Y_1}$, *where* $W_{Y_1|X}: \mathcal{X} \to \mathcal{P}(\mathcal{Y}_1)$ *and* $W_{Y_2|Y_1}: \mathcal{Y}_1 \to \mathcal{P}(\mathcal{Y}_2)$.
- *Deterministic BCs:* *A BC is deterministic if its channel transition matrix factors as* $W_{Y_1,Y_2|X} = \mathbb{1}_{\{Y_1=y_1(X)\}\cap\{Y_2=y_2(X)\}}$, *where* $y_j: \mathcal{X} \to \mathcal{Y}_j$, *for* $j = 1, 2$.

Definition 3 (Code): *An* $(n, R_{12}, R_0, R_1, R_2)$ *code* c_n *for the BC with cooperation and one confidential message has:*

1) *Four message sets* $\mathcal{M}_{12}^{(n)} = \left[1 : 2^{nR_{12}}\right]$ *and* $\mathcal{M}_j^{(n)} = \left[1 : 2^{nR_j}\right]$, *for* $j = 0, 1, 2$.
2) *A stochastic encoder* $f^{(n)}: \mathcal{M}_0^{(n)} \times \mathcal{M}_1^{(n)} \times \mathcal{M}_2^{(n)} \to \mathcal{P}(\mathcal{X}^n)$.
3) *A decoder cooperation function* $g_{12}^{(n)}: \mathcal{Y}_1^n \to \mathcal{M}_{12}^{(n)}$.
4) *Two decoding functions* $\phi_1^{(n)}: \mathcal{Y}_1^n \to \mathcal{M}_0 \times \mathcal{M}_1^{(n)}$ *and* $\phi_2^{(n)}: \mathcal{M}_{12}^{(n)} \times \mathcal{Y}_2^n \to \mathcal{M}_0^{(n)} \times \mathcal{M}_2^{(n)}$.

The joint distribution induced by an $(n, R_{12}, R_0, R_1, R_2)$ code c_n is:

$$P^{(c_n)}\left(m_0, m_1, m_2, \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, m_{12}, \left(\hat{m}_0^{(1)}, \hat{m}_1\right), \left(\hat{m}_0^{(2)}, \hat{m}_2\right)\right) = \left(\prod_{j=0,1,2} \frac{1}{|\mathcal{M}_j^{(n)}|}\right) f^{(n)}(\mathbf{x}|m_0, m_1, m_2) W_{Y_1,Y_2|X}^n(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}) \times \mathbb{1}_{\left\{\hat{m}_{12}=g_{12}^{(n)}(\mathbf{y}_1), \left(\hat{m}_0^{(1)}, \hat{m}_1\right)=\phi_1^{(n)}(\mathbf{y}_1), \left(\hat{m}_0^{(2)}, \hat{m}_2\right)=\phi_2^{(n)}(m_{12}, \mathbf{y}_2)\right\}}. \tag{16}$$

The performance of c_n is evaluated in terms of its rate tuple (R_{12}, R_0, R_1, R_2), the average decoding error probability and the strong secrecy metric.

Definition 4 (Average Error Probability): *The average error probability for an* $(n, R_{12}, R_0, R_1, R_2)$ *code* c_n *is*

$$P_e(c_n) = \mathbb{P}_{P^{(c_n)}}\left(\bigcup_{j=1,2}\left\{\left(\hat{M}_0^{(j)}, \hat{M}_j\right) \neq (M_0, M_j)\right\}\right), \tag{17}$$

where $\left(\hat{M}_0^{(1)}, \hat{M}_1\right) = \phi_1^{(n)}(\mathbf{Y}_1)$ *and* $\left(\hat{M}_0^{(2)}, \hat{M}_2\right) = \phi_2^{(n)}\left(g_{12}^{(n)}(\mathbf{Y}_1), \mathbf{Y}_2\right)$.

Definition 5 (Information Leakage): *The information leakage at receiver 2 under an* $(n, R_{12}, R_0, R_1, R_2)$ *code* c_n *is*

$$\ell(c_n) = I_{P^{(c_n)}}(M_1; M_{12}, Y_2^n), \tag{18}$$

where the subscript $P^{(c_n)}$ *indicates that the mutual information term is calculated with respect to the marginal PMF* $P_{M_1,M_{12},\mathbf{Y}_2}^{(c_n)}$ *of the induced joint distribution from* (16).

Definition 6 (Achievability): $(R_{12}, R_0, R_1, R_2) \in \mathbb{R}_+^4$ *is achievable if for any* $\epsilon > 0$ *there exists an* $(n, R_{12}, R_0, R_1, R_2)$ *code* c_n, *such that*

$$P_e(c_n) \leq \epsilon \tag{19a}$$

$$\ell(c_n) \leq \epsilon. \tag{19b}$$

Definition 7 (Secrecy-Capacity Region): *The strong secrecy-capacity region* $\mathcal{C}_\mathsf{S}$ *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 $\mathcal{C}_\mathsf{S}$ of a cooperative BC with one confidential message.

Theorem 1 (Inner Bound): *Let* $W_{Y_1,Y_2|X}$ *be a transition probability of a BC and let* $\mathcal{R}_\mathsf{I}$ *be the closure of the union of rate tuples* $(R_{12}, R_0, R_1, R_2) \in \mathbb{R}_+^4$ *satisfying:*

$$R_1 \leq I(U_1; Y_1|U_0) - I(U_1; U_2, Y_2|U_0) \tag{20a}$$

$$R_0 + R_1 \leq I(U_0, U_1; Y_1) - I(U_1; U_2, Y_2|U_0) \tag{20b}$$

$$R_0 + R_2 \leq I(U_0, U_2; Y_2) + R_{12} \tag{20c}$$

$$R_0 + R_1 + R_2 \leq I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2, Y_2|U_0) \tag{20d}$$

where the union is over all PMFs $Q_{U_0,U_1,U_2,X} \in \mathcal{P}(\mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X})$, *each inducing a joint distribution* $Q_{U_0,U_1,U_2,X}W_{Y_1,Y_2|X}$. *Then the following inclusion holds:*

$$\mathcal{R}_\mathsf{I} \subseteq \mathcal{C}_\mathsf{S}. \tag{21}$$

Furthermore, $\mathcal{R}_\mathsf{I}$ *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 1 relies on a channel-resolvability-based Marton code and is given in Section VII-C. Two key ingredients allow us to keep 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 the decoded $\left(\hat{M}_0^{(2)}, \hat{M}_2\right)$ 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 (20a) 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 construction for Lemmas 1 and 2 described in Section III. 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 state sequences $\mathbf{S}_0$ and $\mathbf{S}$ play different roles in our resolvability setup. Reliability is established with the help of Lemma 2, while Lemma 1 essentially produces strong secrecy.

The inner bound from Theorem 1 is tight for SD- and PD-BCs, giving rise to the new strong secrecy-capacity results stated in Theorems 2 and 3.

Theorem 2 (SD-BC Secrecy-Capacity): *The strong secrecy-capacity region* $\mathcal{C}_\mathsf{S}^{(\mathsf{SD})}$ *of a cooperative SD-BC* $\mathbb{1}_{\{Y_1=y_1(X)\}}W_{Y_2|X}$ *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:

$$R_1 \leq H(Y_1|W,V,Y_2) \tag{22a}$$
$$R_0 + R_1 \leq H(Y_1|W,V,Y_2) + I(W;Y_1) \tag{22b}$$
$$R_0 + R_2 \leq I(W,V;Y_2) + R_{12} \tag{22c}$$
$$R_0 + R_1 + R_2 \leq H(Y_1|W,V,Y_2) + I(V;Y_2|W) + I(W;Y_1) \tag{22d}$$

where the union is over all PMFs $Q_{W,V,Y_1,X} \in \mathcal{P}(\mathcal{W} \times \mathcal{V} \times \mathcal{Y}_1 \times \mathcal{X})$ *with* $Y_1 = y_1(X)$*, each inducing a joint distribution* $Q_{W,V,Y_1,X} W_{Y_2|X}$*. Furthermore,* $\mathcal{C}_{\mathsf{S}}^{(\mathsf{SD})}$ *is convex and one may choose* $|\mathcal{W}| \leq |\mathcal{X}| + 3$ *and* $|\mathcal{V}| \leq |\mathcal{X}|$*.*

The direct part of Theorem 2 follows from Theorem 1 by setting $U_0 = W$, $U_1 = Y_1$ and $U_2 = V$. The converse is proven in Section VII-D.

Theorem 3 (PD-BC Secrecy-Capacity): The strong secrecy-capacity region $\mathcal{C}_{\mathsf{S}}^{(\mathsf{PD})}$ *of a cooperative PD-BC* $W_{Y_1|X} W_{Y_2|Y_1}$ *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:*

$$R_1 \leq I(X;Y_1|W) - I(X;Y_2|W) \tag{23a}$$
$$R_0 + R_2 \leq I(W;Y_2) + R_{12} \tag{23b}$$
$$R_0 + R_1 + R_2 \leq I(X;Y_1) - I(X;Y_2|W) \tag{23c}$$

where the union is over all PMFs $Q_{W,X} \in \mathcal{P}(\mathcal{W} \times \mathcal{X})$*, each inducing a joint distribution* $Q_{W,X} W_{Y_1|X} W_{Y_2|Y_1}$*. Furthermore,* $\mathcal{C}_{\mathsf{S}}^{(\mathsf{PD})}$ *is convex and one may choose* $|\mathcal{W}| \leq |\mathcal{X}| + 2$*.*

The achievability of $\mathcal{C}_{\mathsf{S}}^{(\mathsf{PD})}$ is a consequence of Theorem 1 by taking $U_0 = W$, $U_1 = X$ and $U_2 = 0$. For the converse see Section VII-E.

Remark 3 (Converse): We use two distinct converse proofs for Theorems 2 and 3. In the converse of Theorem 2, the bound in (22d) *does not involve* R_{12} *since the auxiliary random variable* W_i *contains* M_{12}*. With respect to this choice of* W_i *(see* (77)*), 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 3 we define* W_i *without* M_{12} *and use the structure of the channel to keep* R_{12} *from appearing in* (23c)*. Specifically, this argument relies on the relation* $M_{12} = g_{12}^{(n)}(\mathbf{Y}_1)$ *and on* Y_2 *being a degraded version of* Y_1 *(which implies that all three messages* (M_0, M_1, M_2) *can be reliably decoded from* $\mathbf{Y}_1$ *only).*

Remark 4 (Weak Versus Strong Secrecy): The results of Theorems 1, 2 and 3 remain unchanged if the strong secrecy requirement (see (18) *and* (19b)*) is replaced with the weak secrecy constraint. As weak secrecy refers to a vanishing normalized information leakage, to formally define the corresponding achievability, one should replace the left-hand side (LHS) of* (19b) *with* $\frac{1}{n}\ell(c_n)$*. To see that the results of the preceding theorems coincide under both metrics, first notice that strong secrecy implies weak secrecy (which validates the claim from Theorem 1). Furthermore, the converse proofs of Theorems 2 and 3 (given in Sections VII-D and VII-E, respectively) are readily reformulated under the weak secrecy metric by replacing* ϵ *with* $n\epsilon$ *in* (75)-(76) *and* (88)-(89)*.*

Remark 5 (Cardinality Bounds): The cardinality bounds on the auxiliary random variables in Theorems 1, 2 and 3 are established using the perturbation method [39] and the Eggleston-Fenchel-Carathéodory theorem [40, Th. 18].

V. Restricted Cooperation Scheme Is Sub-Optimal Without Secrecy Constraints

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 restricted protocol 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*. The formal description of this BC instance (see [35]) closely follows the definitions from Section IV-A up to removing the security requirement (19b) from Definition 6 of achievability. For simplicity we consider the setting without a common message, i.e., when $R_0 = 0$.

To isolate the (possible) rate-loss due to the restricted cooperation scheme used in this paper from other losses due to secrecy, we subsequently describe an adaptation of our coding scheme to the case where M_1 is not confidential. Namely, we remove the secrecy requirement on M_1 but still limit the cooperation protocol to share information on M_2 only. This results in an achievable scheme for the cooperative BC with no security requirements, and the induced achievable region is compared with the result from [35].

At first glance it might seem that even without secrecy requirements, the restricted cooperation protocol is optimal. After all, why should the cooperative receiver (Decoder 1) share information about M_1 with the cooperation-aided receiver (Decoder 2), which is not required to decode it? Yet, we show that this intuitive argument fails and 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 M_1 need not be shared. 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 $\mathcal{R}_{\mathsf{NS}}$ for the cooperative BC $W_{Y_1,Y_2|X}$ without secrecy that was characterized in [35] (see also [41], [42]) is the union over the same domain as (20) of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \leq I(U_0, U_1; Y_1) \tag{24a}$$
$$R_2 \leq I(U_0, U_2; Y_2) + R_{12} \tag{24b}$$
$$R_1 + R_2 \leq I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0) \tag{24c}$$
$$R_1 + R_2 \leq I(U_1; Y_1|U_0) + I(U_0, U_2; Y_2) - I(U_1; U_2|U_0) + R_{12}. \tag{24d}$$

The cooperation scheme that achieves (24) 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. samples of the random variable U_0 in (24)) is partitioned into $2^{nR_{12}}$ bins and is first decoded by User 1. The partitioning is defined by a mapping $m_{12} : \left[1 : 2^{nR_{10}}\right] \times \left[1 : 2^{nR_{20}}\right] \to \mathcal{M}_{12}^{(n)}$ and the bin number $m_{12}\big((\hat{M}_{10}, \hat{M}_{20})\big)$ of the decoded public message is shared with User 2 over the cooperative link. This reduces the search space by a factor of $2^{nR_{12}}$. The dependence of the public message on $\hat{M}_{10}$ essentially allows User 1 to achieve rates up $I(U_0, U_1; Y_1)$.

The cooperation protocol used in this work (constructed to account for the secrecy constraint on M_1) removes M_{10} from the public message, while keeping the rest of the protocol unchanged. The region $\tilde{\mathcal{R}}_{\mathsf{NS}}$ achieved by the restricted cooperation protocol is derived by repeating the steps in the proof of [35, Th. 6] while setting $R_{10} = 0$. One obtains that $\tilde{\mathcal{R}}_{\mathsf{NS}}$ is characterized by the same rate bounds as (24), up to replacing (24a) with

$$R_1 \leq I(U_1; Y_1|U_0) + \Big[I(U_2; Y_2|U_0) - I(U_1; U_2|U_0)\Big]^+ \tag{25}$$

where $[x]^+ = \max\{0, x\}$. Since $\tilde{\mathcal{R}}_{\mathsf{NS}}$ is achieved by specializing the scheme that achieves $\mathcal{R}_{\mathsf{NS}}$ (i.e., setting $R_{10} = 0$ therein), we have that $\tilde{\mathcal{R}}_{\mathsf{NS}} \subseteq \mathcal{R}_{\mathsf{NS}}$.

Note that $\tilde{\mathcal{R}}_{\mathsf{NS}} = \mathcal{R}_{\mathsf{NS}}$ for any BC where setting $U_0 = 0$ in (24) is optimal. In particular, we have the following proposition.

Proposition 4 (Optimality of Restricted Protocol): If a BC $W_{Y_1,Y_2|X}$ is PD or deterministic, i.e., it satisfies $W_{Y_1,Y_2|X} = W_{Y_1|X}W_{Y_2|Y_1}$ or $W_{Y_1,Y_2|X} = \mathbb{1}_{\{Y_1=y_1(X)\}\cap\{Y_2=y_2(X)\}}$, respectively, then $\tilde{\mathcal{R}}_{\mathsf{NS}} = \mathcal{R}_{\mathsf{NS}} = \mathcal{C}_{\mathsf{NS}}$.

Proof: For the PD-BC, setting $U_0 = W$, $U_1 = X$ and $U_2 = 0$ into $\tilde{\mathcal{R}}_{\mathsf{NS}}$ recovers the region from [43, eq. (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 $\tilde{\mathcal{R}}_{\mathsf{NS}}$ by taking $U_0 = 0$, $U_1 = Y_1$ and $U_2 = Y_2$. ∎

Proposition 5 (Restricted Protocol can be Sub-Optimal): There exist BCs $W_{Y_1,Y_2|X}$ for which $\tilde{\mathcal{R}}_{\mathsf{NS}} \subsetneq \mathcal{R}_{\mathsf{NS}}$.

The proof of Proposition 5 is given in Appendix VIII, 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 $\tilde{\mathcal{R}}_{\mathsf{NS}}$.

We start with a family of BCs as illustrated in Fig. 3, where the channel input is $X = (X_1, X_2)$, the output Y_1 is produced by feeding X_1 into a binary symmetric channel (BSC) with crossover probability[3] 0.1, while Y_2 is generated by the DMC $W_{Y_2|X_1,X_2}$. 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 the capacity of the aforementioned BSC, i.e., $c \triangleq 1 - H_b(0.1)$,

[3]The actual value of the crossover probability is of no real importance as long as it is not 0.5.

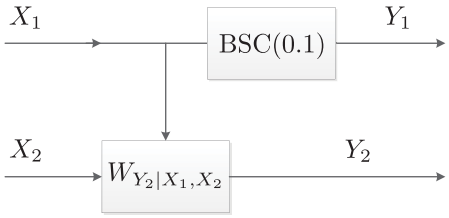


Fig. 3. A semi-orthogonal BC.

where $H_b : [0, 1] \to [0, 1]$ is the binary entropy function. Setting the capacity of the cooperation link to $R_{12} = c$, we show that the highest R_2 such that $(R_{12}, R_1, R_2) = (c, c, R_2) \in \mathcal{R}_{\mathsf{NS}}$ is lower bounded by the capacity of the state-dependent channel $W_{Y_2|X_1,X_2}$ (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 because $R_{12} = c$ in the permissive protocol allows Decoder 1 to share the decoded $\mathbf{X}_1$ with Decoder 2 despite its dependence on M_1.

The corresponding value of R_2 in $\tilde{\mathcal{R}}_{\mathsf{NS}}$ 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 the entire capacity of the BSC was used to reliably convey bits of M_1, on which the restricted protocol prohibits exchanging information. Thus, the proof boils down to choosing $W_{Y_2|X_1,X_2}$ as a channel for which the capacity with full CSI is strictly larger than the GP capacity. The binary dirty-paper (BDP) channel [44]–[46] qualifies and completes the proof.

VI. Effect of Secrecy on the Capacity-Region of Cooperative Broadcast Channels

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 affect BC codes even when no user cooperation is allowed. Thus, when considering a scenario that combines secrecy and cooperation, both these effects occur simultaneously. We highlight this 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., $R_0 = 0$.

A. Semi-Deterministic Broadcast Channels

1) Capacity Region Comparison: Consider the SD-BC without cooperation (i.e., where $R_{12} = 0$) in which M_1 is secret. By Theorem 2, 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 6 (Non-Cooperative SD-BC Secrecy-Capacity): The strong secrecy-capacity region $\tilde{\mathcal{C}}_{\mathsf{S}}^{(\mathsf{SD})}$ of the SD-BC $\mathbb{1}_{\{Y_1=y_1(X)\}}W_{Y_2|X}$ with one confidential message is the union

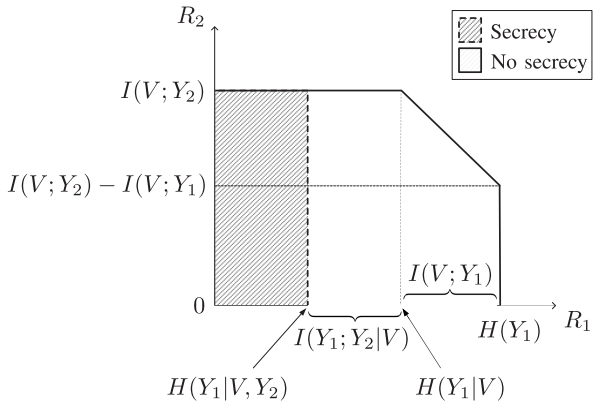


Fig. 4. Capacity region without secrecy vs. strong secrecy-capacity region where M_1 is confidential for the SD-BC (without cooperation).

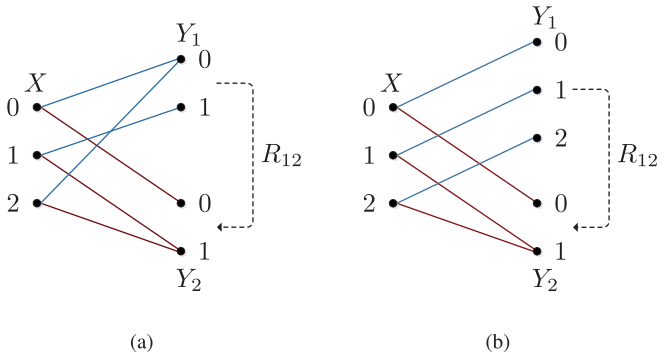


Fig. 5. (a) Cooperative Blackwell BC; (b) Cooperative Blackwell-like PD-BC.

of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ *satisfying:*

$$R_1 \le H(Y_1|V, Y_2) \tag{26a}$$
$$R_2 \le I(V; Y_2) \tag{26b}$$

where the union is over all PMFs $Q_{V,Y_1,X} \in \mathcal{P}(\mathcal{V} \times \mathcal{Y}_1 \times \mathcal{X})$ *with* $Y_1 = y_1(X)$*, each inducing a joint distribution* $Q_{V,Y_1,X} W_{Y_2|X}$.

The region (26) coincides with $\mathcal{C}_{\mathsf{S}}^{(\mathsf{SD})}$ in (22d) (where $R_{12} = R_0 = 0$) by noting that the bound (22d) is redundant because if $Q_{W,V,Y_1,X}$ is a PMF for which (22d) is active, then replacing W and V with $\tilde{W} = 0$ and $\tilde{V} = (W, V)$ achieves a larger region. Removing (22d) from $\mathcal{C}_{\mathsf{S}}^{(\mathsf{SD})}$ and setting $\tilde{V} = (W, V)$ recovers (26).

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:

$$R_1 \le H(Y_1) \tag{27a}$$
$$R_2 \le I(V; Y_2) \tag{27b}$$
$$R_1 + R_2 \le H(Y_1|V) + I(V; Y_2) \tag{27c}$$

where the union is over the same domain as in Corollary 6.

The regions in (26) and (27) (for a fixed $Q_{W,Y_1,X}$) 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 BW-BC without a common message as illustrated in Fig 5(a) [36], [37]. Noting that the BW-BC is deterministic, we set $R_0 = 0$ into the region from Theorem 2 to characterize the strong secrecy-capacity region of a DBC as follows.

Corollary 7 (DBC Secrecy-Capacity): The strong secrecy-capacity region $\mathcal{C}_{\mathsf{S}}^{(\mathsf{D})}$ *of a cooperative DBC* $\mathbb{1}_{\{Y_1 = y_1(X)\} \cap \{Y_2 = y_2(X)\}}$ *with one confidential message is the union of rate triples* $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ *satisfying:*

$$R_1 \le H(Y_1|Y_2) \tag{28a}$$
$$R_2 \le H(Y_2) + R_{12} \tag{28b}$$
$$R_1 + R_2 \le H(Y_1, Y_2) \tag{28c}$$

where the union is over all input distributions $Q_X \in \mathcal{P}(\mathcal{X})$.

Corollary 7 follows by arguments similar to those in the proof of [35, Corollary 12]. By parameterizing the input PMF Q_X as

$$Q_X(0) = \alpha \ , \quad Q_X(1) = \beta \ , \quad Q_X(2) = 1 - \alpha - \beta \tag{29}$$

where $\alpha, \beta \in \mathbb{R}_+$ and $\alpha + \beta \le 1$, the strong secrecy-capacity region $\mathcal{C}_{\mathsf{S}}^{(\mathsf{BW})}$ of the BW-BC is the union of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying:

$$R_1 \le (1 - \alpha) H_b\left(\frac{\beta}{1 - \alpha}\right) \tag{30a}$$
$$R_2 \le H_b(\alpha) + R_{12} \tag{30b}$$
$$R_1 + R_2 \le H_b(\alpha) + (1 - \alpha) H_b\left(\frac{\beta}{1 - \alpha}\right) \tag{30c}$$

where the union is over all $\alpha, \beta \in \mathbb{R}_+$ with $\alpha + \beta \le 1$.

The projection of $\mathcal{C}_{\mathsf{S}}^{(\mathsf{BW})}$ 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 $\mathcal{C}_{\mathsf{S}}^{(\mathsf{BW})}$ 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 (30). 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 6 (Clean Channel to User 1 Does Not Help): *An improved subchannel to the legitimate user does not enlarge the strong secrecy-capacity region. We illustrate this by considering the BW-like PD-BC 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 BW-BC). Evaluating the strong secrecy-capacity region of the BW-like PD-BC reveals that it coincides with* $\mathcal{C}_{\mathsf{S}}^{(\mathsf{BW})}$. *This implies that the* Q_X *that maximizes* R_1 *while keeping Decoder 2 ignorant of* M_1 *has* $\alpha = 0$ *and* $\beta = \frac{1}{2}$, *which coincides with the input PMF that maximizes* R_1 *while transmitting over the classic BW-BC. Thus, to ensure secrecy over the BW-like PD-BC, the encoder overlooks the improved channel to Decoder 1 and ends up not using the symbol* $X = 0$.

The effect of secrecy on the capacity region of a cooperative BC is illustrated by comparing to the BW-BC (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 (29), the capacity region $\mathcal{C}_{\mathsf{NS}}^{(\mathsf{BW})}$ of the cooperative BW-BC is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

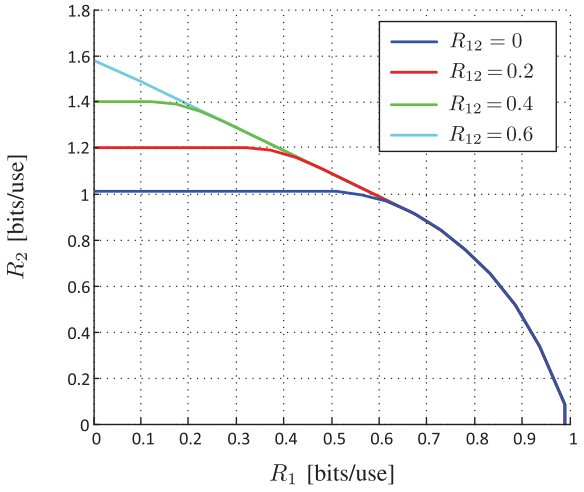
$$R_1 \le H_b(\alpha + \beta) \tag{31a}$$

$$R_2 \le H_b(\alpha) + R_{12} \tag{31b}$$

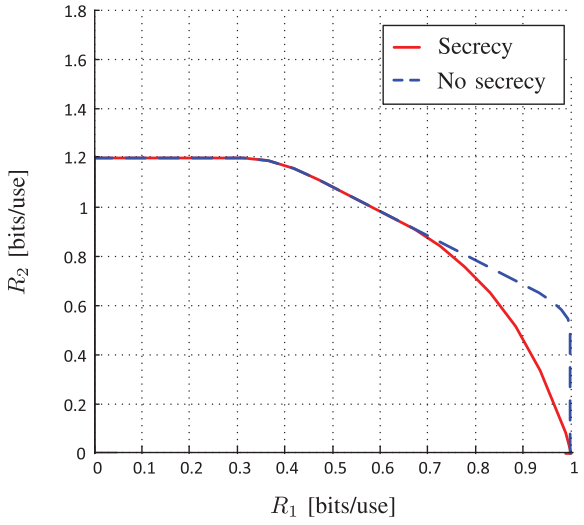
$$R_1 + R_2 \le H_b(\alpha) + (1-\alpha) H_b\left(\frac{\beta}{1-\alpha}\right) \tag{31c}$$

where the union is over all $\alpha, \beta \in \mathbb{R}_+$ with $\alpha + \beta \le 1$.

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, $\mathcal{C}_{\mathsf{NS}}^{(\mathsf{BW})}$ is strictly larger than $\mathcal{C}_{\mathsf{S}}^{(\mathsf{BW})}$. Note that up to approximately $R_1 \approx 0.6597 \triangleq R_1^{(\mathsf{Th})}$, the two regions coincide. Thus, as long as $R_1 \le R_1^{(\mathsf{Th})}$, concealing M_1 is achieved without any rate loss in R_2. When $R_1 > R_1^{(\mathsf{Th})}$, on the other hand, 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 $R_2 = 0$.



(a)



(b)

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

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 (in accordance with Proposition 4). To see this, consider the capacity region $\mathcal{C}_{\mathsf{NS}}^{(\mathsf{PD})}$ of the cooperative PD-BC without a secrecy constraint on M_1 (see [43], [48]), which is the union over the same domain as (23) of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \le I(X; Y_1|W) \tag{32a}$$

$$R_2 \le I(W; Y_2) + R_{12} \tag{32b}$$

$$R_1 + R_2 \le I(X; Y_1). \tag{32c}$$

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 (23a) in comparison to (32a)) 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. Thus, for a fixed $Q_{W,X}$, if

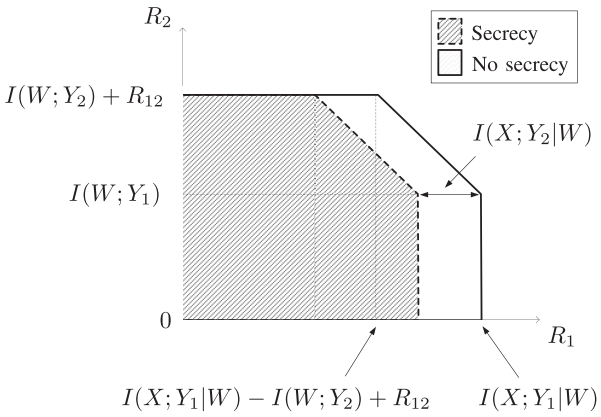


Fig. 7. Capacity region without secrecy vs. strong secrecy-capacity region where M_1 is confidential for the cooperative PD-BC.

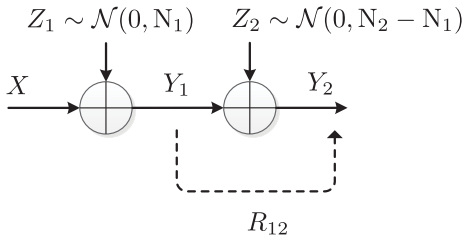


Fig. 8. Cooperative Gaussian PD-BC.

$(R_{12}, R_1, R_2) \in \mathcal{C}_{\mathsf{NS}}^{(\mathsf{PD})}$ then $\left(R_{12}, \left[R_1 - I(X;Y_2|W)\right]^+, R_2\right) \in \mathcal{C}_{\mathsf{S}}^{(\mathsf{PD})}$, 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

$$Y_{1,i} = X_i + Z_{1,i}, \tag{33a}$$

$$Y_{2,i} = X_i + Z_{1,i} + Z_{2,i} \tag{33b}$$

and $\{Z_{1,i}\}_{i=1}^n$ and $\{Z_{2,i}\}_{i=1}^n$ are mutually independent sequences of i.i.d. Gaussian random variables with $Z_{1,i} \sim \mathcal{N}(0, \mathrm{N}_1)$, $Z_{2,i} \sim \mathcal{N}(0, \mathrm{N}_2 - \mathrm{N}_1)$ and $\mathrm{N}_2 > \mathrm{N}_1$, for $i \in [1:n]$. The channel input is subject to an average power constraint

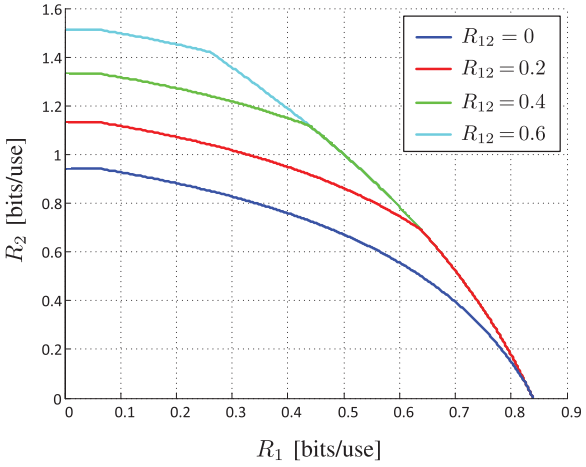
$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[X_i\right] \le \mathrm{P}. \tag{34}$$

By using continuous alphabets with an input power constraint adaptation of Theorem 3 we characterize the strong secrecy-capacity region $\mathcal{C}_{\mathsf{S}}^{(\mathsf{G})}$ of the cooperative Gaussian PD-BC with one confidential message as the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

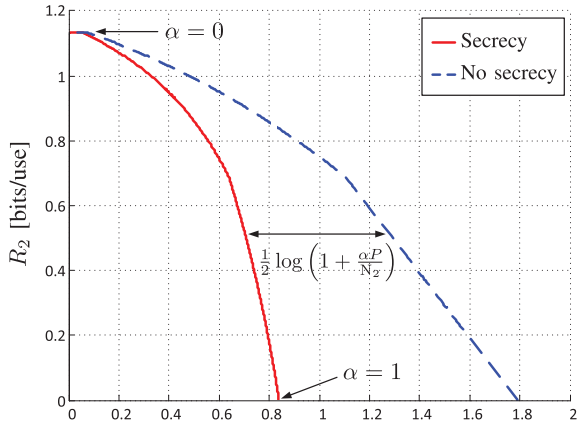
$$R_1 \le \frac{1}{2}\log\left(1 + \frac{\alpha \mathrm{P}}{\mathrm{N}_1}\right) - \frac{1}{2}\log\left(1 + \frac{\alpha \mathrm{P}}{\mathrm{N}_2}\right) \tag{35a}$$

$$R_2 \le \frac{1}{2}\log\left(1 + \frac{\bar{\alpha}\mathrm{P}}{\alpha \mathrm{P} + \mathrm{N}_2}\right) + R_{12} \tag{35b}$$

$$R_1 + R_2 \le \frac{1}{2}\log\left(1 + \frac{\mathrm{P}}{\mathrm{N}_1}\right) - \frac{1}{2}\log\left(1 + \frac{\alpha \mathrm{P}}{\mathrm{N}_2}\right) \tag{35c}$$



(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.

where the union is over all $\alpha \in [0, 1]$.

The achievability of (35) follows from Theorem 3 with the following choice of random variables:

$$W \sim \mathcal{N}(0, \alpha \mathrm{P}) \ , \quad \tilde{W} \sim \mathcal{N}(0, \bar{\alpha}\mathrm{P}) \ , \quad X = W + \tilde{W} \tag{36}$$

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

Setting $\mathrm{P} = 11$, $\mathrm{N}_1 = 1$ and $\mathrm{N}_2 = 4$, 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 $\mathcal{C}_{\mathsf{NS}}^{(\mathsf{G})}$ of the cooperative Gaussian BC without secrecy constraints, which is given by the union over all $\alpha \in [0, 1]$ of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \le \frac{1}{2}\log\left(1 + \frac{\alpha \mathrm{P}}{\mathrm{N}_1}\right) \tag{37a}$$

$$R_2 \le \frac{1}{2}\log\left(1+\frac{\bar{\alpha}\mathrm{P}}{\alpha\mathrm{P}+\mathrm{N}_2}\right)+R_{12} \tag{37b}$$

$$R_1+R_2 \le \frac{1}{2}\log\left(1+\frac{\mathrm{P}}{\mathrm{N}_1}\right) \tag{37c}$$

The derivation of (37) relies on [43, eq. (17)] and uses standard arguments for proving the optimality of Gaussian inputs.

By the structure of the rate bounds in (35) and (37), for every fixed $\alpha \in [0,1]$, if $(R_{12}, R_1, R_2) \in \mathcal{C}_{\mathsf{NS}}^{(\mathsf{G})}$, we have

$$\left(R_{12}, R_1-\frac{1}{2}\log\left(1+\frac{\alpha\mathrm{P}}{\mathrm{N}_2}\right), R_2\right) \in \mathcal{C}_{\mathsf{S}}^{(\mathsf{G})}. \tag{38}$$

This agrees with the discussion in Section VI-B.1 as $I(X;Y_2|W)=\frac{1}{2}\log\left(1+\frac{\alpha\mathrm{P}}{\mathrm{N}_2}\right)$.

VII. Proofs

A. Proof of Lemma 1

Recall that the factorization in (12) implies that $P_{\mathbf{S}_0,\mathbf{S},W,I,\mathbf{U},\mathbf{V}|\mathsf{B}_n=\mathcal{B}_n} = P^{(\mathcal{B}_n)}_{\mathbf{S}_0,\mathbf{S},W,I,\mathbf{U},\mathbf{V}}$, where $\mathcal{B}_n \in \mathfrak{B}_n$ and the RHS is given in (11). Throughout this proof we use $P^{(\mathcal{B}_n)}_{\mathbf{S}_0,\mathbf{S},W,I,\mathbf{U},\mathbf{V}}$ when the codebook $\mathcal{B}_n \in \mathfrak{B}_n$ is fixed, and prefer $P_{\mathbf{S}_0,\mathbf{S},W,I,\mathbf{U},\mathbf{V}|\mathsf{B}_n}$ when the codebook is random. Furthermore, on account of the factorization in (11) we have $P^{(\mathcal{B}_n)}_{\mathbf{S}_0,\mathbf{S}} = Q^n_{S_0,S}$, for each $\mathcal{B}_n \in \mathfrak{B}_n$. Therefore, to establish Lemma 1 we show that

$$\mathbb{E}_{\mathsf{B}_n} D\Big(P_{\mathbf{S}_0,\mathbf{S},\mathbf{V}|\mathsf{B}_n}\Big\|Q^n_{S_0,S,V}\Big) \xrightarrow[n\to\infty]{} 0. \tag{39}$$

Lemma 3 (Absolute Continuity): For any $\mathcal{B}_n \in \mathfrak{B}_n$, 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 3 is relegated to Appendix VIII. Combining this with Remark 1, a sufficient condition for (39) is that

$$\mathbb{E}_{\mathsf{B}_n}\Big\|P_{\mathbf{S}_0,\mathbf{S},\mathbf{V}|\mathsf{B}_n} - Q^n_{S_0,S,V}\Big\| \xrightarrow[n\to\infty]{} 0 \tag{40}$$

at an exponential rate.

To evaluate the TV in (40), for any $\mathcal{B}_n \in \mathfrak{B}_n$, define the *ideal* PMF on $\mathcal{S}_0^n \times \mathcal{S}^n \times \mathcal{W}_n \times \mathcal{I}_n \times \mathcal{U}^n \times \mathcal{V}^n$ as

$$\Gamma^{(\mathcal{B}_n)}(\mathbf{s}_0, w, i, \mathbf{u}, \mathbf{s}, \mathbf{v}) = Q^n_{S_0}(\mathbf{s}_0)2^{-n(\tilde{R}+R')}\mathbb{1}_{\left\{\mathbf{u}=\mathbf{u}(\mathbf{s}_0,w,i)\right\}} Q^n_{S,V|U,S_0}(\mathbf{s},\mathbf{v}|\mathbf{u},\mathbf{s}_0) \tag{41a}$$

and further set

$$\Gamma(\mathcal{B}_n, \mathbf{s}_0, w, i, \mathbf{u}, \mathbf{s}, \mathbf{v}) = \lambda(\mathcal{B}_n)\Gamma^{(\mathcal{B}_n)}(\mathbf{s}_0, w, i, \mathbf{u}, \mathbf{s}, \mathbf{v}). \tag{41b}$$

Note that Γ describes an encoding process where the choice of the u-codeword from a certain bin is uniform, as opposed to P in (11) that uses a 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 triangle inequality, we upper bound the LHS of (40) by

$$\begin{aligned}\mathbb{E}_{\mathsf{B}_n}\Big\|P_{\mathbf{S}_0,\mathbf{S},\mathbf{V}|\mathsf{B}_n} - Q^n_{S_0,S,V}\Big\|_{\mathsf{TV}} &\le \mathbb{E}_{\mathsf{B}_n}\Big\|P_{\mathbf{S}_0,\mathbf{S},\mathbf{V}|\mathsf{B}_n} - \Gamma_{\mathbf{S}_0,\mathbf{S},\mathbf{V}|\mathsf{B}_n}\Big\|_{\mathsf{TV}} \\ &\quad + \mathbb{E}_{\mathsf{B}_n}\Big\|\Gamma_{\mathbf{S}_0,\mathbf{S},\mathbf{V}|\mathsf{B}_n} - Q^n_{S_0,S,V}\Big\|_{\mathsf{TV}}.\end{aligned} \tag{42}$$

By [25, Corollary VII.5], the second expected TV on the RHS of (42) decays exponentially fast as $n\to\infty$ if

$$\tilde{R}+R' > I(U;S,V|S_0). \tag{43}$$

For the first term in (42), we use the following relations between Γ and P. For every $\mathcal{B}_n \in \mathfrak{B}_n$, we have

$$\Gamma^{(\mathcal{B}_n)}_{I|W,\mathbf{S}_0,\mathbf{S}} = \hat{P}^{(\mathcal{B}_n)}_{I|W,\mathbf{S}_0,\mathbf{S}} = P^{(\mathcal{B}_n)}_{I|W,\mathbf{S}_0,\mathbf{S}} \tag{44a}$$

$$\Gamma^{(\mathcal{B}_n)}_{\mathbf{U}|I,W,\mathbf{S}_0,\mathbf{S}} = \mathbb{1}_{\left\{\mathbf{U}=\mathbf{u}(\mathbf{S}_0,W,I)\right\}} = P^{(\mathcal{B}_n)}_{\mathbf{U}|I,W,\mathbf{S}_0,\mathbf{S}} \tag{44b}$$

$$\Gamma^{(\mathcal{B}_n)}_{\mathbf{V}|\mathbf{U},I,W,\mathbf{S}_0,\mathbf{S}} = Q^n_{V|U,S_0,S} = P^{(\mathcal{B}_n)}_{\mathbf{V}|\mathbf{U},I,W,\mathbf{S}_0,\mathbf{S}}. \tag{44c}$$

While (44b)-(44c) follow directly from (11) and (41b), the justification for (44a) is that for every $(\mathcal{B}_n, \mathbf{s}_0, \mathbf{s}, w, i) \in \mathfrak{B}_n \times \mathcal{S}_0^n \times \mathcal{S}^n \times \mathcal{W}_n \times \mathcal{I}_n$, we have

$$\begin{aligned}\Gamma^{(\mathcal{B}_n)}(i|w,\mathbf{s}_0,\mathbf{s}) &= \frac{\Gamma^{(\mathcal{B}_n)}(\mathbf{s}_0,w,i,\mathbf{s})}{\Gamma^{(\mathcal{B}_n)}(\mathbf{s}_0,w,\mathbf{s})} \\ &= \frac{\sum_{\mathbf{u}} Q^n_{S_0}(\mathbf{s}_0)2^{-n(\tilde{R}+R')}\mathbb{1}_{\left\{\mathbf{u}=\mathbf{u}(\mathbf{s}_0,w,i)\right\}} Q^n_{S|U,S_0}(\mathbf{s}|\mathbf{u},\mathbf{s}_0)}{\sum_{\mathbf{u},i'} Q^n_{S_0}(\mathbf{s}_0)2^{-n(\tilde{R}+R')}\mathbb{1}_{\left\{\mathbf{u}=\mathbf{u}(\mathbf{s}_0,w,i')\right\}} Q^n_{S|U,S_0}(\mathbf{s}|\mathbf{u},\mathbf{s}_0)} \\ &= \frac{Q^n_{S|U,S_0}\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0,w,i),\mathbf{s}_0\big)}{\sum_{i'} Q^n_{S|U,S_0}\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0,w,i'),\mathbf{s}_0\big)} \\ &\overset{(a)}{=} \hat{P}^{(\mathcal{B}_n)}(i|w,\mathbf{s}_0,\mathbf{s})\end{aligned} \tag{45}$$

where (a) follows from (10). The relations in (44) yield

$$\begin{aligned}\mathbb{E}_{\mathsf{B}_n}\Big\|P_{\mathbf{S}_0,\mathbf{S},\mathbf{V}|\mathsf{B}_n} - \Gamma_{\mathbf{S}_0,\mathbf{S},\mathbf{V}|\mathsf{B}_n}\Big\|_{\mathsf{TV}} &\le \mathbb{E}_{\mathsf{B}_n}\Big\|P_{\mathbf{S}_0,\mathbf{S},W,I,\mathbf{U},\mathbf{V}|\mathsf{B}_n} - \Gamma_{\mathbf{S}_0,\mathbf{S},W,I,\mathbf{U},\mathbf{V}|\mathsf{B}_n}\Big\|_{\mathsf{TV}} \\ &\overset{(a)}{=} \mathbb{E}_{\mathsf{B}_n}\Big\|P_{\mathbf{S}_0,\mathbf{S},I,\mathbf{U},\mathbf{V}|W=1,\mathsf{B}_n} - \Gamma_{\mathbf{S}_0,\mathbf{S},I,\mathbf{U},\mathbf{V}|W=1,\mathsf{B}_n}\Big\|_{\mathsf{TV}} \\ &\overset{(b)}{=} \mathbb{E}_{\mathsf{B}_n}\Big\|Q^n_{S_0,S} - \Gamma_{\mathbf{S}_0,\mathbf{S}|W=1,\mathsf{B}_n}\Big\|_{\mathsf{TV}}\end{aligned} \tag{46}$$

where:
(a) is because $\Gamma^{(\mathcal{B}_n)}(w) = P^{(\mathcal{B}_n)}(w) = 2^{-n\tilde{R}}$, for every $w \in \mathcal{W}_n$ and $\mathcal{B}_n \in \mathfrak{B}_n$, the independence of B_n and W, and the symmetry of the codebook construction with respect to W;
(b) is by (44) and because $P^{(\mathcal{B}_n)}_{\mathbf{S}_0,\mathbf{S}} = Q^n_{S_0,S}$ for every $\mathcal{B}_n \in \mathfrak{B}_n$.

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

$$\mathbb{E}_{\mathsf{B}_n}\Big\|Q^n_{S_0,S} - \Gamma_{\mathbf{S}_0,\mathbf{S}|W=1,\mathsf{B}_n}\Big\|_{\mathsf{TV}} \xrightarrow[n\to\infty]{} 0 \tag{47}$$

exponentially fast, as long as

$$R' > I(U;S|S_0). \tag{48}$$

This implies that there exists $\gamma > 0$ such that

$$\mathbb{E}_{\mathsf{B}_n}\Big\|P_{\mathbf{S}_0,\mathbf{S},\mathbf{V}|\mathsf{B}_n} - Q^n_{S_0,S,V}\Big\|_{\mathsf{TV}} \le e^{-n\gamma}. \tag{49}$$

B. Proof of Lemma 2

The proof uses the following property of the TV (see, e.g., [28, Property 1]): Let μ, ν be two probability measures on a measurable space $(\mathcal{X}, \mathcal{F})$ and $g : \mathcal{X} \to \mathbb{R}$ be a measurable function bounded by $b \in \mathbb{R}$. We then have

$$\left|\mathbb{E}_\mu g - \mathbb{E}_\nu g\right| \leq b \cdot \left|\left|\mu - \nu\right|\right|_{\mathsf{TV}}. \tag{50}$$

Fix $\epsilon > 0$ and consider the Γ PMF defined in (41b). With respect to the random experiment described by Γ, we have

$$\mathbb{E}_{\mathsf{B}_n}\mathbb{P}_\Gamma\Big(\big(\mathbf{S}_0, \mathbf{S}, \mathbf{U}(\mathbf{S}_0, w, I)\big) \notin \mathcal{T}_\epsilon^n(Q_{S_0,S,U})\Big|\mathsf{B}_n\Big) \xrightarrow[n\to\infty]{} 0 \tag{51}$$

because $\mathbf{U}(\mathbf{S}_0, w, i) \sim Q_{U|S_0}^n$, for every $i \in \mathcal{I}_n$, and $\mathbf{S}$ is obtained by feeding $\big(\mathbf{S}_0, \mathbf{U}(\mathbf{S}_0, w, i)\big)$ into the DMC $Q_{S|U,S_0}^n$. Thus, (51) holds by the weak law of large numbers (WLLN). Further, basic properties of the TV and the analysis in Section VII-A (see (46)) imply

$$\begin{aligned}
&\mathbb{E}_{\mathsf{B}_n}\left|\left|P_{\mathbf{S}_0,\mathbf{S},\mathbf{U}|\mathsf{B}_n} - \Gamma_{\mathbf{S}_0,\mathbf{S},\mathbf{U}|\mathsf{B}_n}\right|\right|_{\mathsf{TV}} \\
&\quad\leq \mathbb{E}_{\mathsf{B}_n}\left|\left|P_{\mathbf{S}_0,\mathbf{S},W,I,\mathbf{U},\mathbf{V}|\mathsf{B}_n} - \Gamma_{\mathbf{S}_0,\mathbf{S},W,I,\mathbf{U},\mathbf{V}|\mathsf{B}_n}\right|\right|_{\mathsf{TV}} \xrightarrow[n\to\infty]{} 0.
\end{aligned} \tag{52}$$

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

$$\begin{aligned}
&\mathbb{E}_{\mathsf{B}_n}\mathbb{P}_P\Big(\big(\mathbf{S}_0, \mathbf{S}, \mathbf{U}(\mathbf{S}_0, w, I)\big) \notin \mathcal{T}_\epsilon^n(Q_{S_0,S,U})\Big|\mathsf{B}_n\Big) \\
&\quad= \mathbb{E}_{\mathsf{B}_n}\mathbb{E}_P\Big[g_n\big(\mathbf{S}_0, \mathbf{S}, \mathbf{U}(\mathbf{S}_0, w, I)\big)\Big|\mathsf{B}_n\Big] \\
&\quad\leq \mathbb{E}_{\mathsf{B}_n}\mathbb{E}_\Gamma\Big[g_n\big(\mathbf{S}_0, \mathbf{S}, \mathbf{U}(\mathbf{S}_0, w, I)\big)\Big|\mathsf{B}_n\Big] \\
&\qquad+ \mathbb{E}_{\mathsf{B}_n}\Big|\mathbb{E}_P\Big[g_n\big(\mathbf{S}_0, \mathbf{S}, \mathbf{U}(\mathbf{S}_0, w, I)\big)\Big|\mathsf{B}_n\Big] \\
&\qquad- \mathbb{E}_\Gamma\Big[g_n\big(\mathbf{S}_0, \mathbf{S}, \mathbf{U}(\mathbf{S}_0, w, I)\big)\Big|\mathsf{B}_n\Big]\Big| \\
&\quad\overset{(a)}{\leq} \mathbb{E}_{\mathsf{B}_n}\mathbb{P}_\Gamma\Big(\big(\mathbf{S}_0, \mathbf{S}, \mathbf{U}(\mathbf{S}_0, w, I)\big) \notin \mathcal{T}_\epsilon^n(Q_{S_0,S,U})\Big|\mathsf{B}_n\Big) \\
&\qquad+ \mathbb{E}_{\mathsf{B}_n}\left|\left|P_{\mathbf{S}_0,\mathbf{S},\mathbf{U}|\mathsf{B}_n} - \Gamma_{\mathbf{S}_0,\mathbf{S},\mathbf{U}|\mathsf{B}_n}\right|\right|_{\mathsf{TV}}
\end{aligned} \tag{53}$$

where (a) uses (50) and g_n being bounded by $b = 1$, for any $n \in \mathbb{N}$. By (51)-(52), the RHS of (53) approaches 0 as $n \to \infty$.

C. Proof of Theorem 1

Fix $n \in \mathbb{N}$, $\epsilon, \delta > 0$, a PMF $Q_{U_0,U_1,U_2,X} \in \mathcal{P}(\mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X})$ and denote $Q_{U_0,U_1,U_2,X,Y_1,Y_2} \triangleq Q_{U_0,U_1,U_2,X} W_{Y_1,Y_2|X}$. In the following we omit the blocklength n from our notations of the involved sets of indices, e.g., we write $\mathcal{M}_0$ instead of $\mathcal{M}_0^{(n)}$, etc. Furthermore, we assume that quantities of the form 2^{nR}, where $n \in \mathbb{N}$ and $R \in \mathbb{R}_+$, are integers.

1) Message Splitting: 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 $\mathcal{M}_p \triangleq \mathcal{M}_0 \times \mathcal{M}_{20}$. Since $|\mathcal{M}_p| = 2^{nR_p}$, with some abuse of notation, we also use $\mathcal{M}_p = \left[1 : 2^{nR_p}\right]$. The partial rates R_{20} and R_{22} satisfy

$$R_2 = R_{20} + R_{22}. \tag{54}$$

With respect to the above, the random variable M_2 is split into two independent random variables M_{20} and M_{22} that are uniform over $\mathcal{M}_{20}$ and $\mathcal{M}_{22}$, respectively. The random variable $M_p \triangleq (M_0, M_{20})$ is uniformly distributed over $\mathcal{M}_p$. Moreover, let W be a random variable uniformly distributed over $\mathcal{W} = \left[1 : 2^{n\tilde{R}}\right]$ and independent of (M_0, M_1, M_2) (which implies its independence of (M_p, M_1, M_{22})).

2) Cooperation Protocol Preliminaries: Fix a partitioning[4] of $\mathcal{M}_p$ into $2^{nR_{12}}$ equal-sized subsets (referred to as "bins") $\mathcal{B}_n(m_{12})$, where $m_{12} \in \mathcal{M}_{12}$. Let $\hat{m}_{12} : \mathcal{M}_p \to \mathcal{M}_{12}$ be the function that associates with each public message $m_p \in \mathcal{M}_p$ its bin index $\hat{m}_{12}(m_p)$, i.e., $m_p \in \mathcal{B}_n\big(\hat{m}_{12}(m_p)\big)$.

3) Codebook $\mathcal{C}_n$: Let $\mathsf{C}_0^{(n)} \triangleq \big\{\mathbf{U}_0(m_p)\big\}_{m_p \in \mathcal{M}_p}$ be a random public message codebook that comprises 2^{nR_p} i.i.d. random vectors $\mathbf{U}_0(m_p)$, each distributed according to $Q_{U_0}^n$. A realization of $\mathsf{C}_0^{(n)}$ is denoted by $\mathcal{C}_0^{(n)} \triangleq \big\{\mathbf{u}_0(m_p)\big\}_{m_p \in \mathcal{M}_p}$.

Fix a public message codebook $\mathcal{C}_0^{(n)}$. For every $m_p \in \mathcal{M}_p$, let $\mathsf{C}_1^{(n)}(m_p) \triangleq \big\{\mathbf{U}_1(m_p, m_1, w, i)\big\}_{(m_1,w,i) \in \mathcal{M}_1 \times \mathcal{W} \times \mathcal{I}}$, where $\mathcal{I} \triangleq \left[1 : 2^{nR'}\right]$, be a random codebook of confidential messages to User 1, consisting of conditionally independent random vectors each distributed according to $Q_{U_1|U_0}^n\big(\cdot \,\big|\mathbf{u}_0(m_p)\big)$. A realization of $\mathsf{C}_1^{(n)}(m_p)$ is denoted by $\mathcal{C}_1^{(n)}(m_p) \triangleq \big\{\mathbf{u}_1(m_p, m_1, w, i)\big\}_{(m_1,w,i) \in \mathcal{M}_1 \times \mathcal{W} \times \mathcal{I}}$. Based on this labeling, each $\mathsf{C}_1^{(n)}(m_p)$, $m_p \in \mathcal{M}_p$, can be thought of as having a u_1-bin associated with every pair $(m_1, w) \in \mathcal{M}_1 \times \mathcal{W}$, each containing $2^{nR_1'}$ u_1-codewords.

Next, for each $m_p \in \mathcal{M}_p$, the corresponding random codebook of private message 2 is $\mathsf{C}_2^{(n)}(m_p) \triangleq \big\{\mathbf{U}_2(m_p, m_{22})\big\}_{m_{22} \in \mathcal{M}_{22}}$, and comprises $2^{nR_{22}}$ conditionally independent random vectors distributed according to

[4]The partitioning may be preformed in any prescribed manner and it is not part of the random coding experiment.

$$\mu(\mathcal{C}_n) = \prod_{m_p \in \mathcal{M}_p} Q_{U_0}^n\big(\mathbf{u}_0(m_p)\big) \prod_{\substack{(m_p^{(1)}, m_1, w, i) \\ \in \mathcal{M}_p \times \mathcal{M}_1 \times \mathcal{W} \times \mathcal{I}}} Q_{U_1|U_0}^n\Big(\mathbf{u}_1\big(m_p^{(1)}, m_1, w, i\big)\Big|\mathbf{u}_0\big(m_p^{(1)}\big)\Big) \prod_{\substack{(m_p^{(2)}, m_{22}) \\ \in \mathcal{M}_p \times \mathcal{M}_{22}}} Q_{U_2|U_0}^n\Big(\mathbf{u}_2\big(m_p^{(2)}, m_{22}\big)\Big|\mathbf{u}_0\big(m_p^{(2)}\big)\Big) \tag{55}$$

$Q^n_{U_2|U_0}(\,\cdot\,|\mathbf{u}_0(m_p))$. We use $\mathcal{C}_2^{(n)}(m_p) \triangleq \{\mathbf{u}_2(m_p, m_{22})\}_{m_{22}\in\mathcal{M}_{22}}$ to denote a possible outcome of $\mathsf{C}_2^{(n)}(m_p)$.

For $j = 1, 2$, we denote $\mathsf{C}_j^{(n)} \triangleq \left\{\mathsf{C}_j^{(n)}(m_p)\right\}_{m_p\in\mathcal{M}_p}$, and its realization by $\mathsf{C}_j^{(n)}$. A random codebook is denoted by $\mathsf{C}_n = \left\{\mathsf{C}_0^{(n)}, \mathsf{C}_1^{(n)}, \mathsf{C}_2^{(n)}\right\}$, while $\mathcal{C}_n = \left\{\mathcal{C}_0^{(n)}, \mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}\right\}$ denotes a fixed codebook (a possible realization of C_n). Denoting the set of all possible values of C_n by $\mathfrak{C}_n$, the above codebook construction induces a PMF $\mu \in \mathcal{P}(\mathfrak{C}_n)$ over the codebook ensemble. For every $\mathcal{C}_n \in \mathfrak{C}_n$, we have (55) from the bottom of the previous page.

For a fixed codebook $\mathcal{C}_n \in \mathfrak{C}_n$ we next describe its associated encoding function $f^{(\mathcal{C}_n)}$, cooperation function $g_{12}^{(\mathcal{C}_n)}$ and decoding functions $\phi_j^{(\mathcal{C}_n)}$, for $j = 1, 2$.

4) Encoder $f^{(\mathcal{C}_n)}$: To transmit a triple $(m_0, m_1, m_2) \in \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2$, the encoder transforms it into the triple $(m_p, m_1, m_{22}) \in \mathcal{M}_p \times \mathcal{M}_1 \times \mathcal{M}_{22}$, and draws W uniformly over $\mathcal{W}$; denote the realization of W by $w \in \mathcal{W}$. Given (m_p, m_1, m_{22}, w), an index $i \in \mathcal{I}$ is then randomly selected by the likelihood encoder according to

$$P_{\mathsf{LE}}^{(\mathcal{C}_n)}\big(i\,\big|w, \mathbf{u}_0(m_p), \mathbf{u}_2(m_p, m_{22})\big) = \frac{Q^n_{U_2|U_1,U_0}\big(\mathbf{u}_2(m_p, m_{22})\big|\mathbf{u}_1(m_p, m_1, w, i), \mathbf{u}_0(m_p)\big)}{\sum\limits_{i'\in\mathcal{I}} Q^n_{U_2|U_1,U_0}\big(\mathbf{u}_2(m_p, m_{22})\big|\mathbf{u}_1(m_p, m_1, w, i'), \mathbf{u}_0(m_p)\big)}. \tag{56}$$

The structure of $P_{\mathsf{LE}}^{(\mathcal{C}_n)}$ adheres to the setup of Lemmas 1-2 from Section III and, in particular, to the stochastic choice of indices therein as described in (10).

Denoting by $i \in \mathcal{I}$ the index selected by $P_{\mathsf{LE}}^{(\mathcal{C}_n)}$, the channel input sequence is then randomly generated according to the conditional product distribution $Q^n_{X|U_0,U_1,U_2}(\,\cdot\,|\mathbf{u}_0(m_p), \mathbf{u}_1(m_p, m_1, w, i), \mathbf{u}_2(m_p, m_{22}))$.

5) Decoding and Cooperation: For a fixed codebook $\mathcal{C}_n \in \mathfrak{C}_n$, we define the following:

- **Decoder $\boldsymbol{\phi}_1^{(\mathcal{C}_n)}$:** 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 exists an index $\hat{i} \in \mathcal{I}$ such that
$$\Big(\mathbf{u}_0(\hat{m}_p), \mathbf{u}_1(\hat{m}_p, \hat{m}_1, \hat{w}, \hat{i}), \mathbf{y}_1\Big) \in \mathcal{T}_\epsilon^n(Q_{U_0,U_1,Y_1}). \tag{57}$$
If such a unique triple is found set $\phi_1^{(\mathcal{B}_n)}(\mathbf{y}_1) = (\hat{m}_0, \hat{m}_1)$, where $\hat{m}_0$ is taken from $\hat{m}_p = (\hat{m}_0, \hat{m}_{22})$; otherwise, set $\phi_1^{(\mathcal{C}_n)}(\mathbf{y}_1) = (1, 1)$.
- **Cooperation $\boldsymbol{g}_{\mathbf{12}}^{(\boldsymbol{\mathcal{C}_n})}$:** Having $(\hat{m}_p, \hat{m}_1, \hat{w}, \hat{i})$, Decoder 1 conveys the bin number of $\hat{m}_p$, i.e., $\hat{m}_{12}(\hat{m}_p) \in \mathcal{M}_{12}$, to Decoder 2 via the cooperation link. That is, $g_{12}^{(\mathcal{C}_n)}(\mathbf{y}_1) = \hat{m}_{12}(\hat{m}_p)$.
- **Decoder $\boldsymbol{\phi}_2^{(\boldsymbol{\mathcal{C}_n})}$:** Upon observing $\big(\hat{m}_{12}(\hat{m}_p), \mathbf{y}_2\big)$, 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
$$\Big(\mathbf{u}_0(\hat{\hat{m}}_p), \mathbf{u}_2(\hat{\hat{m}}_p, \hat{\hat{m}}_{22}), \mathbf{y}_2\Big) \in \mathcal{T}_\epsilon^n(Q_{U_0,U_2,Y_2}) \tag{58}$$
where $\hat{\hat{m}}_p \in \mathcal{B}_n\big(\hat{m}_{12}(\hat{m}_p)\big)$. If such a unique pair is found, set $\phi_2^{(\mathcal{C}_n)}\big(\hat{m}_{12}(\hat{m}_p), \mathbf{y}_2\big) = \big(\hat{\hat{m}}_0, \hat{\hat{m}}_2\big)$, where $\hat{\hat{m}}_2 = (\hat{\hat{m}}_{20}, \hat{\hat{m}}_{22})$ in which $\hat{\hat{m}}_0$ and $\hat{\hat{m}}_{20}$ are specified by $\hat{\hat{m}}_p = \big(\hat{\hat{m}}_0, \hat{\hat{m}}_{20}\big)$; otherwise, set $\phi_2^{(\mathcal{B}_n)}\big(\hat{m}_{12}(\hat{m}_p), \mathbf{y}_2\big) = (1, 1)$.

6) Induced Code and Joint Distribution: The tuple $\Big(f^{(\mathcal{C}_n)}, g_{12}^{(\mathcal{C}_n)}, \phi_1^{(\mathcal{C}_n)}, \phi_2^{(\mathcal{C}_n)}\Big)$ defined with respect to the codebook $\mathcal{C}_n \in \mathfrak{C}_n$ constitutes an $(n, R_{12}, R_0, R_1, R_2)$ code c_n for the cooperative BC. Thus, for every codebook $\mathcal{C}_n \in \mathfrak{C}_n$, the induced joint distribution is given in (59) at the bottom of this page, where the random variables $\mathbf{U}_0$, $\mathbf{U}_1$ and $\mathbf{U}_2$ are the chosen codewords at the conclusion of the encoding process (from which the input $\mathbf{X}$ to the BC is generated).

Taking the random codebook generation into account, we also set (60) from the bottom of this page, where $\mu \in \mathcal{P}(\mathfrak{C}_n)$ is described in (55). The PMF P induces a probability measure $\mathbb{P} \triangleq \mathbb{P}_P$, with respect to which the subsequent analysis is preformed. Specifically, all the mutli-letter information measures in the sequel are taken with respect to P from (60), while single-letter information terms are always calculated with respect to $Q_{U_0,U_1,U_2,X,Y_1,Y_2}$.

7) Expected Average Error Probability Analysis: By virtue of Lemma 2 we first show that under the proper rate constraints, the above encoding process results in u_0-, u_1- and u_2-sequences that are jointly typical. The rest of the analysis goes through via classic joint typicality arguments. The details of the analysis are relegated to Appendix VIII, where it is shown that

$$\mathbb{E}P_{\mathsf{e}}(\mathsf{C}_n) \le \eta(n, \delta, \delta'), \tag{61}$$

$$\begin{aligned} &P^{(\mathcal{C}_n)}\Big(m_p, m_1, m_{22}, w, m_{12}, \mathbf{u}_0, \mathbf{u}_2, i, \mathbf{u}_1, \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \big(\hat{m}_0^{(1)}, \hat{m}_1\big), \big(\hat{m}_0^{(2)}, \hat{m}_2\big)\Big) \\ &\quad= 2^{-n(R_p+R_1+R_{22}+\tilde{R})}\mathbb{1}_{\left\{m_{12}=\hat{m}_{12}(m_p), \mathbf{u}_0=\mathbf{u}_0(m_p), \mathbf{u}_2=\mathbf{u}_2(m_p,m_{22})\right\}} P_{\mathsf{LE}}^{(\mathcal{C}_n)}\big(i\,\big|w, \mathbf{u}_0(m_p), \mathbf{u}_2(m_p, m_{22})\big)\mathbb{1}_{\left\{\mathbf{u}_1=\mathbf{u}_1(m_p,m_1,w,i)\right\}} \\ &\qquad\times Q^n_{X|U_0,U_1,U_2}(\mathbf{x}|\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2) Q^n_{Y_1,Y_2|X}(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x})\mathbb{1}_{\left\{\left(\hat{m}_0^{(1)}, \hat{m}_1\right)=\phi_1^{(\mathcal{C}_n)}(\mathbf{y}_1), \left(\hat{m}_0^{(2)}, \hat{m}_2\right)=\phi_2^{(\mathcal{C}_n)}(m_{12}, \mathbf{y}_2)\right\}} \end{aligned} \tag{59}$$

$$\begin{aligned} &P\Big(m_p, m_1, m_{22}, w, m_{12}, \mathbf{u}_0, \mathbf{u}_2, i, \mathbf{u}_1, \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \big(\hat{m}_0^{(1)}, \hat{m}_1\big), \big(\hat{m}_0^{(2)}, \hat{m}_2\big)\Big) \\ &\quad= \mu(\mathcal{C}_n) P^{(\mathcal{C}_n)}\Big(m_p, m_1, m_{22}, w, m_{12}, \mathbf{u}_0, \mathbf{u}_2, i, \mathbf{u}_1, \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \big(\hat{m}_0^{(1)}, \hat{m}_1\big), \big(\hat{m}_0^{(2)}, \hat{m}_2\big)\Big) \end{aligned} \tag{60}$$

where $\delta' \in (0, \delta)$ and $\lim_{n\to\infty} \eta(n, \delta, \delta') = 0$ for all $0 < \delta' < \delta$, if

$$R' > I(U_1; U_2|U_0) \tag{62a}$$
$$R' + \tilde{R} > I(U_1; U_2, Y_2|U_0) \tag{62b}$$
$$R_1 + \tilde{R} + R' < I(U_1; Y_1|U_0) - \tau_\delta \tag{62c}$$
$$R_p + R_1 + \tilde{R} + R' < I(U_0, U_1; Y_1) - \tau_\delta \tag{62d}$$
$$R_{22} < I(U_2; Y_2|U_0) - \tau_\delta \tag{62e}$$
$$R_p + R_{22} - R_{12} < I(U_0, U_2; Y_2) - \tau_\delta. \tag{62f}$$

with $\tau_\delta \to 0$ as $\delta \to 0$ and $\tau_{\delta'} \to 0$ as $\delta' \to 0$. To clarify, the δ' that appears in the upper bound on the expected error probability from (61) is a consequence of the Conditional Typicality Lemma [49, Sec. 2.5]. Namely, the lemma considers conditioning on sequences that are jointly letter-typical with respect to a slightly smaller gap δ' than the original δ.

8) Security Analysis: As in the proof of Lemma 1 from Section VII-A, throughout this proof we use $P^{(\mathcal{C}_n)}$ when the codebook $\mathcal{C}_n \in \mathfrak{C}_n$ is fixed, and $P_{\cdot|\mathsf{C}_n}$ when the codebook is random (see (59)-(60)). Fix a codebook $\mathcal{C}_n \in \mathfrak{C}_n$ and let $I_{\mathcal{C}_n}$ denote the a mutual information taken with respect to $P^{(\mathcal{C}_n)}$. Consider the following upper bound on the information leakage.

$$\begin{aligned} &I_{\mathcal{C}_n}(M_1; M_{12}, \mathbf{Y}_2) \\ &\quad\le I_{\mathcal{C}_n}(M_1; M_{12}, M_p, M_{22}, \mathbf{Y}_2) \\ &\quad\overset{(a)}{=} I_{\mathcal{C}_n}\big(M_1; \mathbf{Y}_2|M_p, M_{22}, \mathbf{U}_0, \mathbf{U}_2\big) \\ &\quad\overset{(b)}{\le} D\Big(P^{(\mathcal{C}_n)}_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big\|Q^n_{Y_2|U_0,U_2}\Big|P^{(\mathcal{C}_n)}_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big) \end{aligned} \tag{63}$$

where:
(a) is because M_1 is independent of (M_p, M_{22}), and since $M_{12} = \hat{m}_{12}(M_p)$, $\mathbf{U}_0 = \mathbf{u}_0(M_p)$ and $\mathbf{U}_2 = \mathbf{u}_2(M_p, M_{22})$ are defined by (M_p, M_{22});
(b) follows by the relative entropy chain rule and because for every $\mathcal{C}_n \in \mathfrak{C}_n$, the definition of relative entropy gives (64) from the bottom of this page.

Taking the expectation of the RHS of (63) over the ensemble of codebooks, we get (65) from the bottom of this page, where (a) uses the symmetry of the codebook with respect to the messages, while (b) is by the law of total expectation conditioning the inner expectation on $\mathsf{C}^{(n)}_{0,2} \triangleq \Big\{\mathsf{C}^{(n)}_0, \mathsf{C}^{(n)}_2\Big\}$.

Next, we adjust the RHS of (65) so that it corresponds to the setup of Lemma 1. To this end, note that when $\mathcal{C}_n \in \mathfrak{C}_n$ is fixed, $P^{(\mathcal{C}_n)}_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2}$ is well-defined only if $\mathbf{u}_0 = \mathbf{u}_0(1)$ and $\mathbf{u}_2 = \mathbf{u}_2(1)$. For any other $\mathbf{u}_0$ and $\mathbf{u}_2$, we may set this conditional distribution as any arbitrary PMF on $\mathcal{Y}_2^n$, since this does not affect the joint distribution from (59). Accordingly, if $\mathbf{u}_0 \neq \mathbf{u}_0(1)$ or $\mathbf{u}_2 \neq \mathbf{u}_2(1,1)$, we define

$$P^{(\mathcal{C}_n)}_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2} = Q^n_{Y_2|U_0,U_1}\big(\cdot\,\big|\mathbf{u}_0, \mathbf{u}_2\big). \tag{66}$$

Having this, note that for any $(\mathbf{u}_0, \mathbf{u}_2) \in \mathcal{U}_0^n \times \mathcal{U}_2^n$ and a fixed $\mathsf{C}^{(n)}_{0,2} = \mathcal{C}^{(n)}_{0,2} \triangleq \Big\{\mathcal{C}^{(n)}_0, \mathcal{C}^{(n)}_2\Big\}$, we have (67) from the top of the next page. In the derivation of (67) (a) follows from (66) and because conditioned on $\mathbf{U}_0(1)$ and $\mathbf{U}_2(1,1)$, $P_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathsf{C}_n}$ is independent of all the other codewords in $\mathsf{C}_{0,2}$. Furthermore, $P_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathsf{C}_n}$ is actually a function of the codebook $\mathsf{C}^{(n)}_1(1)$, rather than the entire collection C_n.

Some further definitions are required in order to rigorously justify the application of Lemma 1. For each $\mathbf{u}_0 \in \mathcal{U}_0^n$, let $\tilde{\mathsf{C}}_n(\mathbf{u}_0) \triangleq \big\{\tilde{\mathbf{U}}_1(\mathbf{u}_0, w, i)\big\}_{(w,i)\in\mathcal{W}\times\mathcal{I}}$, be a collection of i.i.d. random vectors of length n, each distributed according to $Q^n_{U_1|U_0}(\cdot|\mathbf{u}_0)$. The collection $\tilde{\mathsf{C}}_n \triangleq \Big\{\tilde{\mathsf{C}}_n(\mathbf{u}_0)\Big\}_{\mathbf{u}_0\in\mathcal{U}_0^n}$ is independent of C_n and is distributed according to

$$\tilde{\lambda}(\tilde{\mathcal{C}}_n) = \prod_{\mathbf{u}_0\in\mathcal{U}_0^n}\prod_{\substack{(w,i)\\ \in\mathcal{W}\times\mathcal{I}}} Q^n_{U_1|U_0}\big(\tilde{\mathbf{u}}_1(\mathbf{u}_0, w, i)\big|\mathbf{u}_0\big), \tag{68}$$

where, as before, $\tilde{\mathcal{C}}_n(\mathbf{u}_0) \triangleq \big\{\tilde{\mathbf{u}}_1(\mathbf{u}_0, w, i)\big\}_{(w,i)\in\mathcal{W}\times\mathcal{I}}$ stands for a realization of $\tilde{\mathsf{C}}_n(\mathbf{u}_0)$. For each $(\mathbf{u}_0, \mathbf{u}_2) \in \mathcal{U}_0^n \times \mathcal{U}_2^n$ and

$$\begin{aligned} &D\Big(P^{(\mathcal{B}_n)}_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big\|P^{(\mathcal{B}_n)}_{\mathbf{Y}_2|M_p,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big|P^{(\mathcal{B}_n)}_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big) \\ &\quad\le D\Big(P^{(\mathcal{B}_n)}_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big\|Q^n_{Y_2|U_0,U_2}\Big|P^{(\mathcal{B}_n)}_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big) - D\Big(P^{(\mathcal{B}_n)}_{\mathbf{Y}_2|M_p,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big\|Q^n_{Y_2|U_0,U_2}\Big|P^{(\mathcal{B}_n)}_{M_p,M_{22},\mathbf{U}_0,\mathbf{U}_2}\Big) \end{aligned} \tag{64}$$

$$\begin{aligned} &\mathbb{E}_{\mathsf{C}_n} D\Big(P_{\mathbf{Y}_2|M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2,\mathsf{C}_n}\Big\|Q^n_{Y_2|U_0,U_2}\Big|P_{M_p,M_1,M_{22},\mathbf{U}_0,\mathbf{U}_2,\mathsf{C}_n}\Big) \\ &\quad= \mathbb{E}_{\mathsf{C}_n}\Bigg[\sum_{m_p,m_1,m_{22},\mathbf{u}_0,\mathbf{u}_2} 2^{-n(R_p+R_1+R_{22})}\mathbb{1}_{\{(\mathbf{U}_0(m_p),\mathbf{U}_2(m_p,m_{22}))=(\mathbf{u}_0,\mathbf{u}_2)\}} \\ &\qquad\qquad\times D\Big(P_{\mathbf{Y}_2|M_p=m_p,M_1=m_1,M_{22}=m_{22},\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathsf{C}_n}\Big\|Q^n_{Y_2|U_0,U_1}(\cdot|\mathbf{u}_0,\mathbf{u}_2)\Big)\Bigg] \\ &\quad\overset{(a)}{=} \sum_{\mathbf{u}_0,\mathbf{u}_2}\mathbb{E}_{\mathsf{C}_n}\Big[\mathbb{1}_{\{(\mathbf{U}_0(1),\mathbf{U}_2(1,1))=(\mathbf{u}_0,\mathbf{u}_2)\}} D\Big(P_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathsf{C}_n}\Big\|Q^n_{Y_2|U_0,U_1}(\cdot|\mathbf{u}_0,\mathbf{u}_2)\Big)\Big] \\ &\quad\overset{(b)}{=} \sum_{\mathbf{u}_0,\mathbf{u}_2}\mathbb{E}_{\mathsf{C}^{(n)}_{0,2}}\Big[\mathbb{1}_{\{(\mathbf{U}_0(1),\mathbf{U}_2(1,1))=(\mathbf{u}_0,\mathbf{u}_2)\}}\mathbb{E}_{\mathsf{C}^{(n)}_1|\mathsf{C}^{(n)}_{0,2}}\Big[D\Big(P_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathsf{C}_n}\Big\|Q^n_{Y_2|U_0,U_1}(\cdot|\mathbf{u}_0,\mathbf{u}_2)\Big)\Big]\Big] \end{aligned} \tag{65}$$

$$\begin{aligned}
&\mathbb{E}_{\mathsf{C}_1^{(n)}\left|\mathsf{C}_{0,2}^{(n)}=\mathcal{C}_{0,2}^{(n)}\right.}\left[D\Big(P_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathsf{C}_n}\Big|\Big|Q^n_{Y_2|U_0,U_1}(\cdot|\mathbf{u}_0,\mathbf{u}_2)\Big)\right]\\
&\quad=\mathbb{E}_{\mathsf{C}_1^{(n)}\left|\mathsf{C}_{0,2}^{(n)}=\mathcal{C}_{0,2}^{(n)}\right.}\Bigg[\mathbb{1}_{\left\{\left(\mathbf{u}_0(1),\mathbf{u}_2(1,1)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}}D\Big(P_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathsf{C}_n}\Big|\Big|Q^n_{Y_2|U_0,U_1}(\cdot|\mathbf{u}_0,\mathbf{u}_2)\Big)\\
&\qquad+\mathbb{1}_{\left\{\left(\mathbf{u}_0(1),\mathbf{u}_2(1,1)\right)\neq(\mathbf{u}_0,\mathbf{u}_2)\right\}}D\Big(P_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathsf{C}_n}\Big|\Big|Q^n_{Y_2|U_0,U_1}(\cdot|\mathbf{u}_0,\mathbf{u}_2)\Big)\Bigg]\\
&\quad\overset{(a)}{=}\mathbb{E}_{\mathsf{C}_1^{(n)}\left|\mathbf{U}_0(1)=\mathbf{u}_0(1),\mathbf{U}_2(1,1)=\mathbf{u}_2(1,1)\right.}\Bigg[\mathbb{1}_{\left\{\left(\mathbf{u}_0(1),\mathbf{u}_2(1,1)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}}\\
&\qquad\times D\Big(P_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathsf{C}_1^{(n)}(1)}\Big|\Big|Q^n_{Y_2|U_0,U_1}(\cdot|\mathbf{u}_0,\mathbf{u}_2)\Big)\Bigg]
\end{aligned}\tag{67}$$

a corresponding $\tilde{\mathcal{C}}_n(\mathbf{u}_0)$, define a conditional PMF

$$\begin{aligned}
&\tilde{P}^{(\tilde{\mathcal{C}}_n)}(w,i,\tilde{\mathbf{u}}_1,\mathbf{y}_2|\mathbf{u}_0,\mathbf{u}_2)\\
&\quad=2^{-n\tilde{R}}\tilde{P}^{(\tilde{\mathcal{C}}_n)}(i|w,\mathbf{u}_0,\mathbf{u}_2)\mathbb{1}_{\left\{\tilde{\mathbf{u}}_1=\tilde{\mathbf{u}}_1(\mathbf{u}_0,w,i)\right\}}\\
&\qquad\times Q^n_{Y_2|U_0,U_1,U_2}(\mathbf{y}_2|\mathbf{u}_0,\tilde{\mathbf{u}}_1,\mathbf{u}_2),
\end{aligned}\tag{69}$$

where $\tilde{P}^{(\tilde{\mathcal{C}}_n)}(i|w,\mathbf{u}_0,\mathbf{u}_2)$ is defined exactly like $\hat{P}^{(\mathcal{B}_n)}(i|w,\mathbf{s}_0,\mathbf{s})$ from (10), up to renaming $\mathbf{s}_0$, $\mathbf{s}$, $\mathbf{u}$ and $\mathcal{B}_n$ therein to $\mathbf{u}_0$, $\mathbf{u}_2$, $\tilde{\mathbf{u}}_1$ and $\tilde{\mathcal{C}}_n$, respectively. Also define

$$\tilde{P}(\tilde{\mathcal{C}}_n,w,i,\tilde{\mathbf{u}}_1,\mathbf{y}_2|\mathbf{u}_0,\mathbf{u}_2)=\tilde{\lambda}(\tilde{\mathcal{C}}_n)\tilde{P}^{(\tilde{\mathcal{C}}_n)}(w,i,\tilde{\mathbf{u}}_1,\mathbf{y}_2|\mathbf{u}_0,\mathbf{u}_2).\tag{70}$$

For any $(\mathbf{u}_0,\mathbf{u}_2)\in\mathcal{U}_0^n\times\mathcal{U}_2^n$, the RHS of (67) is further upper bounded by

$$\mathbb{E}_{\tilde{\mathsf{C}}_n}D\Big(\tilde{P}_{\mathbf{Y}_2|\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\tilde{\mathsf{C}}_n}\Big|\Big|Q^n_{Y_2|U_0,U_1}(\cdot|\mathbf{u}_0,\mathbf{u}_2)\Big).\tag{71}$$

This follows by removing the indicator function and because when $\mathbf{u}_0(1)=\tilde{\mathbf{u}}_0$ and $\mathcal{C}_1^{(n)}(1)=\tilde{\mathcal{C}}_n(\tilde{\mathbf{u}}_0)$, the distributions $P_{\mathbf{Y}_2|M_p=1,M_1=1,M_{22}=1,\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\mathsf{C}_1^{(n)}(1)=\mathcal{C}_1^{(n)}(1)}$ and $\tilde{P}_{\mathbf{Y}_2|\mathbf{U}_0=\tilde{\mathbf{u}}_0,\mathbf{U}_2=\mathbf{u}_2,\tilde{\mathsf{C}}_n(\tilde{\mathbf{u}}_0)=\tilde{\mathcal{C}}_n(\tilde{\mathbf{u}}_0)}$ are equal as PMFs on $\mathcal{Y}_2^n$. Since (71) falls within the framework of Lemma 1 we can make this expectation arbitrarily small provided that (62a)-(62b) hold.

Inserting (65), (67) and (71) back into (63), yields

$$\begin{aligned}
&\mathbb{E}_{\mathsf{C}_n}\ell(\mathsf{C}_n)\\
&\quad I(M_1;M_{12},\mathbf{Y}_2|\mathsf{C}_n)\\
&\quad\leq\sum_{\mathbf{u}_0,\mathbf{u}_2}\mathbb{E}_{\mathsf{C}_{0,2}}\mathbb{1}_{\left\{\left(\mathbf{U}_0(1),\mathbf{U}_2(1,1)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}}\\
&\qquad\times\mathbb{E}_{\tilde{\mathsf{C}}_n}D\Big(\tilde{P}_{\mathbf{Y}_2|\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\tilde{\mathsf{C}}_n}\Big|\Big|Q^n_{Y_2|U_0,U_1}(\cdot|\mathbf{u}_0,\mathbf{u}_2)\Big)\\
&\quad\overset{(a)}{=}\mathbb{E}_{\tilde{\mathsf{C}}_n}\Bigg[\sum_{\mathbf{u}_0,\mathbf{u}_2}Q^n_{U_0,U_2}(\mathbf{u}_0,\mathbf{u}_2)\\
&\qquad\times D\Big(\tilde{P}_{\mathbf{Y}_2|\mathbf{U}_0=\mathbf{u}_0,\mathbf{U}_2=\mathbf{u}_2,\tilde{\mathsf{C}}_n}\Big|\Big|Q^n_{Y_2|U_0,U_1}(\cdot|\mathbf{u}_0,\mathbf{u}_2)\Big)\Bigg]\\
&\quad=\mathbb{E}_{\tilde{\mathsf{C}}_n}D\Big(\tilde{P}_{\mathbf{Y}_2|\mathbf{U}_0,\mathbf{U}_2,\tilde{\mathsf{C}}_n}\Big|\Big|Q^n_{Y_2|U_0,U_2}\Big|Q^n_{U_0,U_2}\Big)
\end{aligned}\tag{72}$$

where (a) is since Q_{U_0,U_2} is the coding PMF, which gives $\mathbb{P}_\mu\Big(\mathbf{U}_0(1)=\mathbf{u}_0,\mathbf{U}_2(1,1)=\mathbf{u}_2\Big)=Q^n_{U_0,U_2}(\mathbf{u}_0,\mathbf{u}_2)$. Invoking Lemma 1 on the RHS of (72), while viewing $Q_{Y_2|U_0,U_1,U_2}$ as a state-dependent DMC from $\mathcal{U}_1$ to $\mathcal{Y}_2$ with state space $\mathcal{U}_0\times\mathcal{U}_2$, we see that (62a)-(62b) give

$$\mathbb{E}_{\tilde{\mathsf{C}}_n}D\Big(\tilde{P}_{\mathbf{Y}_2|\mathbf{U}_0,\mathbf{U}_2,\tilde{\mathsf{C}}_n}\Big|\Big|Q^n_{Y_2|U_0,U_2}\Big|Q^n_{U_0,U_2}\Big)\xrightarrow[n\to\infty]{}0.\tag{73}$$

The Selection Lemma [50, Lemma 5] (see also [19, Lemma 2.2]) applied to the sequence of random variables $\{\mathsf{C}_n\}_{n\in\mathbb{N}}$ and the functions P_e and ℓ implies the existence of a sequence of codebooks $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, each giving rise to a code c_n such that $P_e(c_n)\leq\epsilon$ and $\ell(c_n)\leq\epsilon$, for n sufficiently large. Finally, we apply Fourier-Motzkin elimination (FME) on (62) while using (54) 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 [51]. This produces the rate bounds from (20) with small subtracted terms such as τ_δ. Since $\delta>0$ and $\delta'\in(0,\delta)$ can be chosen arbitrarily small (which shrinks τ_δ), this concludes the proof of Theorem 1.

Remark 7 (BC Code and Resolvability Lemma Analogy): Lemma 1 is key in the security analysis of the proposed coding scheme. In the following, we relate the cooperative BC code construction and the setup of our resolvability lemma. Having (63)*, the main idea is to adjust the relative entropy on the RHS so that it corresponds to the lemma. This is done by viewing the u_0- and the u_2-codewords from the BC codebook as a pair of states of the subchannel $Q_{Y_2|U_0,U_1,U_2}$ to Decoder 2, where the u_1-codewords plays the role of the channel's input. The validity of this analogy stems from the structure of the BC codebook, where for each $(m_p,m_1)\in\mathcal{M}_p\times\mathcal{M}_1$, the set $\{\mathbf{U}_1(m_p,m_1,w,i)\}_{(w,i)\in\mathcal{W}\times\mathcal{I}}$ forms a resolvability codebook just like in Lemma 1. This resolvability codebook is superimposed on $\mathbf{U}_0(m_p)$, while the transmitted u_1-codeword is correlated with $\mathbf{U}_2(m_p,m_{22})$ by means of the likelihood encoder* (56)*. The correspondence between the coding scheme presented in this section and the setup of Lemma 1 is summarized in Table I.*

The main challenge in applying the resolvability for the BC code is accounting for the relative entropy from the RHS of (63) *being conditioned on the induced joint distribution of $\mathbf{U}_0$ and $\mathbf{U}_2$, while the lemma conditions it on a product distribution. However, as the derivation between Equation* (63)-(73) *shows, under the expectation over the ensemble of codebooks, the induced distribution in the conditioning can be*

TABLE I
CORRESPONDENCE BETWEEN THE CODING SCHEME FOR THE COOPERATIVE BC AND THE SETUP OF THE RESOLVABILITY LEMMA 1

	Cooperative BC Code	**Resolvability Lemma**
State-dependent DMC	$Q_{Y_2\|U_0,U_1,U_2}$	$Q_{V\|U,S_0,S}$
Channel states	$(\mathbf{U}_0, \mathbf{U}_2)$	$(\mathbf{S}_0, \mathbf{S})$
Channel input	$\mathbf{U}_1$	$\mathbf{U}$
Resolvability codebook	$\big\{\mathbf{U}_1(m_p, m_1, w, i)\big\}_{(w,i)=(1,1)}^{(2^{n\tilde{R}}, 2^{nR'})}$, for each $(m_p, m_1) \in \mathcal{M}_p \times \mathcal{M}_1$	$\big\{\mathbf{U}(\mathbf{s}_0, w, i)\big\}_{(w,i)=(1,1)}^{(2^{n\tilde{R}}, 2^{nR'})}$
Codebook generation	$\sim Q_{U_1\|U_0}^n\big(\,\cdot\,\big|\mathbf{u}(m_p)\big)$	$\sim Q_{U\|S_0}^n(\cdot\|\mathbf{s}_0)$
Likelihood encoder	$P_{\mathsf{LE}}^{(\mathcal{C}_n)}(i\|w, \mathbf{u}_0, \mathbf{u}_2)$ from (56) - Correlates $(\mathbf{U}_0, \mathbf{U}_1)$ with $\mathbf{U}_2$	$\hat{P}^{(\mathcal{B}_n)}(i\|w, \mathbf{s}_0, \mathbf{s})$ from (10) - Correlates $(\mathbf{S}_0, \mathbf{U})$ with $\mathbf{S}_2$
Rate bounds	$R' > I(U_1; U_2\|U_0)$ $R' + \tilde{R} > I(U_1; U_2, Y_2\|U_0)$	$R' > I(U; S\|S_0)$ $R' + \tilde{R} > I(U; S, V\|S_0)$
Implied asymptotic behaviour	$I(M_1; M_{12}, \mathbf{Y}_2\|\mathsf{C}_n) \to 0$ as $n \to \infty$	$\mathbb{E}_{\mathsf{B}_n} D\Big(P_{\mathbf{V}\|\mathbf{S}_0,\mathbf{S},\mathsf{B}_n}\Big\|Q_{V\|S_0,S}^n\Big|Q_{S_0,S}^n\Big) \to 0$ as $n \to \infty$

converted to the product PMf Q_{U_0,U_2}^n (according to which the codebooks $\mathbf{U}_0$ and $\mathbf{U}_2$ are drawn).

Remark 8 (Comparison to the Scheme Without Secrecy): 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. In the non-secret scenario $Q_{X|U_0,U_1,U_2}$ can be replaced with a deterministic function.

D. Converse Proof for Theorem 2

We show that if a rate tuple (R_{12}, R_0, R_1, R_2) is achievable, then there exists a PMF $Q_{W,V,Y_1,X} \in \mathcal{P}(\mathcal{W} \times \mathcal{V} \times \mathcal{Y}_1 \times \mathcal{X})$ with $Y_1 = y_1(X)$, such that the inequalities in (22) are satisfied with respect to the joint distribution $Q_{W,V,Y_1,X} W_{Y_2|X}$. Fix an achievable tuple (R_{12}, R_0, R_1, R_2), an $\epsilon > 0$, and let c_n be the corresponding $(n, R_{12}, R_0, R_1, R_2)$ code for some sufficiently large $n \in \mathbb{N}$ such that (19) holds. All subsequent multi-letter information measures are calculated with respect to the PMF induced by c_n from (16), with the SD-BC $W_{Y_1,Y_2|X}^n(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}) = \mathbb{1}_{\bigcap_{i=1}^n \{y_{1,i} = y_1(x_i)\}} W_{Y_2|X}^n(\mathbf{y}_2|\mathbf{x})$. By Fano's inequality we have

$$H(M_0, M_1|Y_1^n) \le 1 + n\epsilon(R_0 + R_1) \triangleq n\epsilon_n^{(1)} \tag{74a}$$

$$H(M_0, M_2|M_{12}, Y_2^n) \le 1 + n\epsilon(R_0 + R_2) \triangleq n\epsilon_n^{(2)}. \tag{74b}$$

Define

$$\epsilon_n = \max\big\{\epsilon_n^{(1)}, \epsilon_n^{(2)}\big\}. \tag{74c}$$

Moreover, (19b) implies

$$\begin{aligned} \epsilon &\ge I(M_1; M_{12}, Y_2^n) \\ &= I(M_1; M_0, M_2, M_{12}, Y_2^n) - I(M_1; M_0, M_2|M_{12}, Y_2^n) \\ &\overset{(a)}{\ge} I(M_1; M_{12}, Y_2^n|M_0, M_2) - H(M_0, M_2|M_{12}, Y_2^n) \\ &\overset{(b)}{\ge} I(M_1; M_{12}, Y_2^n|M_0, M_2) - n\epsilon_n \end{aligned} \tag{75}$$

where (a) uses the independence of M_1 and (M_0, M_2) and the non-negativity of entropy, while (b) follows from (74). Thus,

$$I(M_1; M_{12}, Y_2^n|M_0, M_2) \le \epsilon + n\epsilon_n. \tag{76}$$

It follows that

$$\begin{aligned} nR_1 &= H(M_1) \\ &\overset{(a)}{=} H(M_1|M_{12}, M_0, M_2) + I(M_1; M_{12}|M_0, M_2) \\ &\overset{(b)}{\le} I(M_1; Y_1^n|M_{12}, M_0, M_2) + I(M_1; M_{12}|M_0, M_2) \end{aligned}$$

$$
\begin{aligned}
&-I(M_1; M_{12}, Y_2^n|M_0, M_2) + n\delta_n^{(1)} \\
&\overset{(c)}{=} \sum_{i=1}^{n} \Big[I(M_1; Y_1^i, Y_{2,i+1}^n|M_{12}, M_0, M_2) \\
&\qquad - I(M_1; Y_1^{i-1}, Y_{2,i}^n|M_{12}, M_0, M_2) \Big] + n\delta_n^{(1)} \\
&= \sum_{i=1}^{n} \Big[I(M_1; Y_{1,i}|M_{12}, M_0, M_2, Y_1^{i-1}, Y_{2,i+1}^n) \\
&\qquad - I(M_1; Y_{2,i}|M_{12}, M_0, M_2, Y_1^{i-1}, Y_{2,i+1}^n) \Big] + n\delta_n^{(1)} \\
&\overset{(d)}{=} \sum_{i=1}^{n} \Big[H(Y_{1,i}|M_2, W_i) - H(Y_{1,i}|M_1, M_2, W_i) \\
&\qquad - I(M_1; Y_{2,i}|M_2, W_i) \Big] + n\delta_n^{(1)} \\
&\leq \sum_{i=1}^{n} \Big[H(Y_{1,i}|M_2, W_i) - I(Y_{1,i}; Y_{2,i}|M_1, M_2, W_i) \\
&\qquad - I(M_1; Y_{2,i}|M_2, W_i) \Big] + n\delta_n^{(1)} \\
&= \sum_{i=1}^{n} \Big[H(Y_{1,i}|M_2, W_i) \\
&\qquad - I(M_1, Y_{1,i}; Y_{2,i}|M_1, M_2, W_i) \Big] + n\delta_n^{(1)} \\
&\leq \sum_{i=1}^{n} H(Y_{1,i}|M_2, W_i, Y_{2,i}) + n\delta_n^{(1)} \qquad (77)
\end{aligned}
$$

where:
(a) is because M_1 is independent (M_0, M_2);
(b) follows from (74)-(75) and by denoting $\delta_n^{(1)} = 2\epsilon_n + \frac{\epsilon}{n}$;
(c) is a telescoping identity [52, eqs. (9) and (11)];
(d) defines $W_i = (M_{12}, M_0, Y_1^{i-1}, Y_{2,i+1}^n)$.

The common message rate R_0 satisfies

$$
\begin{aligned}
nR_0 &= H(M_0) \\
&\overset{(a)}{\leq} I(M_0; Y_1^n) + n\epsilon_n \qquad (78a) \\
&= \sum_{i=1}^{n} I(M_0; Y_{1,i}|Y_1^{i-1}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(M_0, Y_1^{i-1}; Y_{1,i}) + n\epsilon_n \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} I(W_i; Y_{1,i}) + n\epsilon_n \qquad (78b)
\end{aligned}
$$

where (a) uses (74) and (b) follows by the definition of W_i. Combining (77) with (78b) yields

$$
n(R_0 + R_1) \leq \sum_{i=1}^{n} \Big[H(Y_{1,i}|M_2, W_i, Y_{2,i}) + I(W_i; Y_{1,i}) \Big] + n\delta_n^{(2)} \qquad (79)
$$

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

For the sum $R_0 + R_2$, we have

$$
\begin{aligned}
n(R_0+R_2) &= H(M_0, M_2) \\
&\overset{(a)}{\leq} I(M_0, M_2; M_{12}, Y_2^n) + n\epsilon_n \\
&= I(M_0, M_2; Y_2^n|M_{12}) + I(M_0, M_2; M_{12}) + n\epsilon_n \\
&\overset{(b)}{\leq} I(M_0, M_2; Y_2^n|M_{12}) + nR_{12} + n\epsilon_n \\
&= \sum_{i=1}^{n} I(M_0, M_2; Y_{2,i}|M_{12}, Y_{2,i+1}^n) + nR_{12} + n\epsilon_n \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} I(M_2, W_i; Y_{2,i}) + nR_{12} + n\epsilon_n \qquad (80)
\end{aligned}
$$

where:
(a) uses (74);
(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

$$
\begin{aligned}
n(R_0+R_1+R_2) &= H(M_0, M_1, M_2) \\
&= H(M_1|M_0, M_2) + H(M_2|M_0) + H(M_0). \qquad (81)
\end{aligned}
$$

Consider now

$$
\begin{aligned}
&H(M_2|M_0) \\
&\overset{(a)}{\leq} I(M_2; Y_2^n|M_{12}, M_0) + I(M_2; M_{12}|M_0) + n\epsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} \Big[I(M_2; Y_{2,i}^n|M_{12}, M_0, Y_1^{i-1}) \\
&\qquad - I(M_2; Y_{2,i+1}^n|M_{12}, M_0, Y_1^i) \Big] + I(M_2; M_{12}|M_0) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} \Big[I(M_2; Y_{2,i+1}^n|M_{12}, M_0, Y_1^{i-1}) \\
&\qquad + I(M_2; Y_{2,i}|W_i) - I(M_2; Y_{1,i}, Y_{2,i+1}^n|M_{12}, M_0, Y_1^{i-1}) \\
&\qquad + I(M_2; Y_{1,i}|M_{12}, M_0, Y_1^{i-1}) \Big] + I(M_2; M_{12}|M_0) + n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n} \Big[I(M_2; Y_{2,i}|W_i) - I(M_2; Y_{1,i}|W_i) \Big] \\
&\qquad + I(M_2; Y_1^n|M_0) + n\epsilon_n \qquad (82)
\end{aligned}
$$

where:
(a) uses (74) and 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 (78a) with (82), yields

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

$$-I(M_{12}, Y_{2,i+1}^n; Y_{1,i}|M_0, M_2, Y_1^{i-1})\Big] + 2n\epsilon_n$$
$$\overset{(c)}{\leq} \sum_{i=1}^{n}\Big[I(M_2; Y_{2,i}|W_i) + I(W_i; Y_{1,i})\Big] + 2n\epsilon_n \tag{83}$$

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 (77) and (83) into (81), we bound the sum of rates as

$$n(R_0+R_1+R_2) \leq \sum_{i=1}^{n}\Big[H(Y_{1,i}|M_2, W_i, Y_{2,i}) + I(M_2; Y_{2,i}|W_i) + I(W_i; Y_{1,i})\Big] + n\delta_n^{(3)} \tag{84}$$

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

The bounds in (77), (79), (80) and (83) 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, (77) is rewritten as

$$\begin{aligned} R_1 &\leq \frac{1}{n}\sum_{t=1}^{n} H(Y_{1,t}|M_2, W_t, Y_{2,t}) + \delta_n^{(1)} \\ &= \sum_{t=1}^{n}\mathbb{P}(T=t)H(Y_{1,T}|M_2, W_T, Y_{2,T}, T=t) + \delta_n^{(1)} \\ &= H(Y_{1,T}|M_2, W_T, Y_{2,T}, T) + \delta_n^{(1)}. \end{aligned} \tag{85}$$

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 (22) 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 $Q_{W,V,Y_1,X}W_{Y_2|X}$ and satisfies $Y_1 = y_1(X)$. As the functional relation between Y_1 and X is straightforward, it remains to be shown that

$$(W, V, Y_1) - X - Y_2 \tag{86}$$

forms a Markov chain. This is proven in Appendix VIII-D.

E. Converse Proof for Theorem 3

We show that given an achievable rate tuple (R_{12}, R_0, R_1, R_2), there exists a PMF $Q_{W,X} \in \mathcal{P}(\mathcal{W}\times\mathcal{X})$ for which (23) holds with respect to the joint distribution $Q_{W,X}W_{Y_1|X}W_{Y_2|Y_1}$. Let (R_{12}, R_0, R_1, R_2) be an achievable tuple and fix $\epsilon > 0$. Let c_n be the corresponding $(n, R_{12}, R_0, R_1, R_2)$ code for some sufficiently large $n \in \mathbb{N}$ such that (19) holds. The induced joint distribution is again given by (16), but now the transition matrix is of a PD-BC, i.e., $W_{Y_1,Y_2|X}^n(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}) = W_{Y_1|X}^n(\mathbf{y}_1|\mathbf{x})W_{Y_2|Y_1}^n(\mathbf{y}_2|\mathbf{y}_1)$. Fano's inequality gives

$$H(M_0, M_1|Y_1^n) \leq 1 + n\epsilon(R_0+R_1) \triangleq n\kappa_n^{(1)} \tag{87a}$$
$$H(M_0, M_2|M_{12}, Y_2^n) \leq 1 + n\epsilon(R_0+R_2) \triangleq n\kappa_n^{(2)} \tag{87b}$$
$$H(M_0, M_1, M_2|Y_1^n, Y_2^n) \leq 1 + n\epsilon(R_0+R_1+R_2) \triangleq n\kappa_n^{(3)} \tag{87c}$$

and we set

$$\kappa_n = \max\left\{\kappa_n^{(1)}, \kappa_n^{(2)}, \kappa_n^{(3)}\right\} = \kappa_n^{(3)}. \tag{87d}$$

Further, by the strong secrecy constraint (19b), we have

$$\begin{aligned} \epsilon &\geq I(M_1; M_{12}, Y_2^n) \\ &= I(M_1; M_0, M_2, M_{12}, Y_2^n) - I(M_1; M_0, M_2|M_{12}, Y_2^n) \\ &\overset{(a)}{\geq} I(M_1; M_{12}, Y_2^n|M_0, M_2) - H(M_0, M_2|M_{12}, Y_2^n) \\ &\overset{(b)}{\geq} I(M_1; Y_2^n|M_0, M_2) - n\kappa_n \end{aligned} \tag{88}$$

where (a) uses the independence of M_1 and (M_0, M_2) and the non-negativity of entropy, while (b) is by (87) and since conditioning cannot increase entropy. This yields

$$I(M_1; Y_2^n|M_0, M_2) \leq \epsilon + n\kappa_n. \tag{89}$$

We bound

$$\begin{aligned} nR_1 &= H(M_1) \\ &\overset{(a)}{=} H(M_1|M_0, M_2) \\ &\overset{(b)}{\leq} I(M_1; Y_1^n|M_0, M_2) - I(M_1; Y_2^n|M_0, M_2) + n\eta_n \\ &\overset{(c)}{=} \sum_{i=1}^{n}\Big[I(M_1; Y_1^i, Y_{2,i+1}^n|M_0, M_2) \\ &\quad - I(M_1; Y_1^{i-1}, Y_{2,i}^n|M_0, M_2)\Big] + n\eta_n \\ &\overset{(d)}{=} \sum_{i=1}^{n}\Big[I(M_1; Y_{1,i}|W_i) - I(M_1; Y_{2,i}|W_i)\Big] + n\eta_n \end{aligned} \tag{90a}$$

$$\begin{aligned} &\overset{(e)}{=} \sum_{i=1}^{n} I(M_1; Y_{1,i}|W_i, Y_{2,i}) + n\eta_n \\ &\overset{(f)}{\leq} \sum_{i=1}^{n} I(X_i; Y_{1,i}|W_i, Y_{2,i}) + n\eta_n \\ &\overset{(g)}{\leq} \sum_{i=1}^{n}\Big[I(X_i; Y_{1,i}|W_i) - I(X_i; Y_{2,i}|W_i)\Big] + n\eta_n \end{aligned} \tag{90b}$$

where:
(a) uses the independence of M_1 and (M_0, M_2);
(b) is by virtue of (87)-(88) 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

$$\begin{aligned} n(R_0+R_2) &= H(M_0, M_2) \\ &\overset{(a)}{\leq} I(M_0, M_2; M_{12}, Y_2^n) + n\kappa_n \\ &\overset{(b)}{\leq} I(M_0, M_2; Y_2^n) + nR_{12} + n\kappa_n \\ &= \sum_{i=1}^{n} I(M_0, M_2; Y_{2,i}|Y_{2,i+1}^n) + nR_{12} + n\kappa_n \\ &\overset{(c)}{\leq} \sum_{i=1}^{n} I(W_i; Y_{2,i}) + nR_{12} + n\kappa_n \end{aligned} \tag{91}$$

where:
(a) is by (87);
(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

$$\begin{aligned}
&n(R_0+R_1+R_2)\\
&= H(M_0,M_1,M_2)\\
&\overset{(a)}{\leq} I(M_0,M_1,M_2;Y_1^n,Y_2^n)-I(M_1;Y_2^n|M_0,M_2)+n\eta_n\\
&\overset{(b)}{=} I(M_0,M_1,M_2;Y_1^n)-I(M_1;Y_2^n|M_0,M_2)+n\eta_n\\
&\overset{(c)}{=} \sum_{i=1}^{n}\Big[I(M_0,M_1,M_2,Y_{2,i+1}^n;Y_{1,i}|Y_1^{i-1})\\
&\qquad -I(Y_{2,i+1}^n;Y_{1,i}|M_0,M_1,M_2,Y_1^{i-1})\\
&\qquad\qquad -I(M_1;Y_{2,i}|M_0,M_2,Y_{2,i+1}^n)\Big]+n\eta_n\\
&\overset{(d)}{=} \sum_{i=1}^{n}\Big[I(M_0,M_1,M_2,Y_{2,i+1}^n;Y_{1,i}|Y_1^{i-1})\\
&\qquad -I(Y_1^{i-1};Y_{2,i}|M_0,M_1,M_2,Y_{2,i+1}^n)\\
&\qquad\qquad -I(M_1;Y_{2,i}|M_0,M_2,Y_{2,i+1}^n)\Big]+n\eta_n\\
&\leq \sum_{i=1}^{n}\Big[I(M_0,M_1,M_2,Y_1^{i-1},Y_{2,i+1}^n;Y_{1,i})\\
&\qquad -I(M_1,Y_1^{i-1};Y_{2,i}|M_0,M_2,Y_{2,i+1}^n)\Big]+n\eta_n\\
&\overset{(e)}{\leq} \sum_{i=1}^{n}\Big[I(W_i;Y_{1,i})+I(M_1;Y_{1,i}|W_i)-I(M_1;Y_{2,i}|W_i)\Big]+n\eta_n\\
&\overset{(f)}{\leq} \sum_{i=1}^{n}\Big[I(W_i;Y_{1,i})+I(X_i;Y_{1,i}|W_i)-I(X_i;Y_{2,i}|W_i)\Big]+n\eta_n\\
&\overset{(g)}{=} \sum_{i=1}^{n}\Big[I(X_i;Y_{1,i})-I(X_i;Y_{2,i}|W_i)\Big]+n\eta_n \qquad (92)
\end{aligned}$$

where:
(a) uses (87) 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 (see, e.g., [52, eq. (3)]);
(e) follows from the definitions of W_i and because conditioning cannot increase entropy;
(f) is by repeating steps (90a)-(90b);
(g) is by the mutual information chain rule and because $W_i - X_i - Y_{1,i}$ forms a Markov chain (see Appendix VIII-E 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 (23) with the small added terms κ_n and η_n, which approach 0 as $n \to \infty$. In Appendix VIII-E we show that the chain

$$W - X - Y_1 - Y_2 \qquad (93)$$

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 for the BC code. A resolvability-based Marton code for the BC 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 seems 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 illustrated the results. An explicit coding scheme that achieves strong secrecy while maximizing the transmission rate of the confidential message over the BW-BC was given. Further, it was shown that the strong secrecy-capacity region of the BW-BC remains unchanged even if the subchannel to the legitimate user is noiseless.

Appendix A
Proof of Proposition 5

Let $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0,1\}$. Consider the BC $W_{Y_1|X_1}W_{Y_2|X_1,X_2}$ from Fig. 3, where $W_{Y_1|X_1}$ is a BSC with transition probability 0.1 and $W_{Y_2|X_1,X_2}$ is an arbitrary channel from $\{0,1\}^2$ to $\{0,1\}$ to be specified later.

For simplicity of notation we relabel $U_0 = W$, $U_1 = U$ and $U_2 = V$ in $\mathcal{R}_{\mathsf{NS}}$, which becomes the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \leq I(W,U;Y_1) \qquad (94a)$$

$$R_2 \leq I(W,V;Y_2) + R_{12} \qquad (94b)$$

$$\begin{aligned}
R_1+R_2 \leq\ & I(U;Y_1|W)+I(V;Y_2|W)-I(U;V|W)\\
&+\min\Big\{I(W;Y_1),\, I(W;Y_2)+R_{12}\Big\} \qquad (94c)
\end{aligned}$$

where the union is over all PMFs $Q_{W,U,V,X_1,X_2} \in \mathcal{P}(\mathcal{W}\times\mathcal{V}\times\mathcal{V}\times\mathcal{X}_1\times\mathcal{X}_2)$, each inducing a joint distribution $Q_{W,U,V,X_1,X_2,Y_1,Y_2} \triangleq Q_{W,U,V,X_1,X_2}W_{Y_1|X_1}W_{Y_2|X_1,X_2}$. Setting $U_0 = W$, $U_1 = U$ and $U_2 = V$ into $\tilde{\mathcal{R}}_{\mathsf{NS}}$, gives a region described by the same rate bounds as (94), up to replacing (94a) with

$$R_1 \leq I(U;Y_1|W) + \Big[I(V;Y_2|W) - I(U;V|W)\Big]^+. \qquad (95)$$

We outer bound $\tilde{\mathcal{R}}_{\mathsf{NS}}$ by loosening (95) to

$$R_1 \leq I(U;Y_1|W). \qquad (96)$$

Let $\tilde{\mathcal{O}}_{\mathsf{NS}}$ denote the obtained outer bound on $\tilde{\mathcal{R}}_{\mathsf{NS}}$. We show that under the considered example $\tilde{\mathcal{O}}_{\mathsf{NS}} \subsetneq \mathcal{R}_{\mathsf{NS}}$.

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

$$\mathcal{R}_{\mathsf{NS}}(r) \triangleq \left\{(R_1, R_2) \in \mathbb{R}_+^2 \middle| (r, R_1, R_2) \in \mathcal{R}_{\mathsf{NS}}\right\} \tag{97a}$$

$$\tilde{\mathcal{O}}_{\mathsf{NS}}(r) \triangleq \left\{(R_1, R_2) \in \mathbb{R}_+^2 \middle| (r, R_1, R_2) \in \tilde{\mathcal{O}}_{\mathsf{NS}}\right\} \tag{97b}$$

be the projections of $\mathcal{R}_{\mathsf{NS}}$ and $\tilde{\mathcal{O}}_{\mathsf{NS}}$ on the (R_1, R_2) plane for $R_{12} = r$. Let $c = 1 - H_b(0.1)$, where $H_b : [0, 1] \to [0, 1]$ is the binary entropy function, and note that $R_1 = c$ is the maximal achievable rate of M_1 in both $\mathcal{C}_{\mathsf{NS}}(c)$ and $\tilde{\mathcal{O}}_{\mathsf{NS}}(c)$. Define the supremum of all achievable R_2 that preserve $R_1 = c$ in each region by

$$R_2^\star \triangleq \sup\left\{R_2 \in \mathbb{R}_+ \middle| (c, R_2) \in \mathcal{R}_{\mathsf{NS}}(c)\right\} \tag{98a}$$

$$\tilde{R}_2^\star \triangleq \sup\left\{R_2 \in \mathbb{R}_+ \middle| (c, R_2) \in \tilde{\mathcal{O}}_{\mathsf{NS}}(c)\right\}. \tag{98b}$$

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

For $\mathcal{R}_{\mathsf{NS}}(c)$, setting $W = X_1 \sim \mathrm{Ber}\left(\frac{1}{2}\right)$ achieves

$$R_1 = I(W, U; Y_1) \overset{(a)}{=} I(X_1; Y_1) = c \tag{99}$$

where (a) follows because $U - X_1 - Y_1$ forms a Markov chain. Consequently, for $R_2^\star$ we have

$$\begin{aligned} R_2^\star &\overset{(a)}{=} \sup_{\substack{Q_{U,V,X_2|X_1}: \\ (U,V)-(X_1,X_2)-Y_2}} \min\begin{Bmatrix} I(X_1, V; Y_2) + c, \\ I(X_1, V; Y_2) - I(U; V|X_1) \end{Bmatrix} \\ &\overset{(b)}{\geq} \sup_{\substack{Q_{V,X_2|X_1}: \\ V-(X_1,X_2)-Y_2}} I(V; Y_2|X_1) \end{aligned} \tag{100}$$

where (a) uses the structure of $\mathcal{R}_{\mathsf{NS}}$ from (94) and the relations $R_{12} = I(X_1; Y_1) = c$ and $W = X_1$, while (b) is by setting $U = X_1$ and due to the non-negativity of mutual information.

For $\tilde{\mathcal{O}}_{\mathsf{NS}}(c)$, first note that R_1 is upper bounded by c since

$$I(U; Y_1|W) \overset{(a)}{\leq} I(W, U; Y_1) \overset{(b)}{\leq} I(X_1; Y_1) \overset{(c)}{\leq} c. \tag{101}$$

However, $R_1 = c$ is also achievable: (a) becomes an inequality if and only if Y_1 is independent of W; (b) is an equality if and only if $X_1 - (W, U) - Y_1$ forms a Markov chain (this step also uses the Markov relation $(W, U) - X_1 - Y_1$; (c) holds with equality if and only if $X_1 \sim \mathrm{Ber}\left(\frac{1}{2}\right)$.

Now, since Y_1 and X_1 are connected by a BSC, the independence of Y_1 and W implies that X_1 and W are also independent. To see this observe that the independence of Y_1 and W means that

$$Q_{Y_1|W}(0|w) = Q_{Y_1|W}(0|w'), \quad \forall (w, w') \in \mathcal{W}^2, \tag{102}$$

and assume by contradiction that a similar relation does not hold for X_1 and W. Namely, assume that there exists a pair $(w, w') \in \mathcal{W}^2$ such that

$$Q_{X_1|W}(0|w) \neq Q_{X_1|W}(0|w'). \tag{103}$$

Denote $Q_{X_1|W}(0|w) = \alpha$ and $Q_{X_1|W}(0|w') = \alpha'$, where $\alpha, \alpha' \in [0, 1]$ and $\alpha \neq \alpha'$. Consider the following:

$$\begin{aligned} &Q_{Y_1|W}(0|w) \\ &\overset{(a)}{=} Q_{X_1|W}(0|w)Q_{Y_1|X_1}(0|0) + Q_{X_1|W}(1|w)Q_{Y_1|X_1}(0|1) \\ &= 0.9\alpha + 0.1(1 - \alpha) \\ &= 0.1 + 0.8\alpha. \end{aligned} \tag{104}$$

By repeating similar steps for $Q_{Y_1|W}(0|w')$, we get

$$Q_{Y_1|W}(0|w') = 0.1 + 0.8\alpha'. \tag{105}$$

Combining (104)-(105) with (102) gives that $\alpha = \alpha'$, which is a contradiction. Therefore X_1 and W must be independent.

Furthermore, recall that from the equality in step (b) of (101) the chain $X_1 - (W, U) - Y_1$ is Markov, i.e.,

$$\begin{aligned} &Q_{X_1,Y_1|W,U}(x_1, y_1|w, u) \\ &\quad = Q_{X_1|W,U}(x_1|w, u)Q_{Y_1|W,U}(y_1|w, u) \end{aligned} \tag{106}$$

for all $(w, u, x_1, y_1) \in \mathcal{W} \times \mathcal{U} \times \mathcal{X}_1 \times \mathcal{Y}_1$. Since $(W, U) - X_1 - Y_1$ is also a Markov chain, we have that $Q_{X_1,Y_1|W,U}$ also factors as

$$Q_{X_1,Y_1|W,U}(x_1, y_1|w, u) = Q_{X_1|W,U}(x_1|w, u)Q_{Y_1|X_1}(y_1|x_1) \tag{107}$$

for all $(w, u, x_1, y_1) \in \mathcal{W} \times \mathcal{U} \times \mathcal{X}_1 \times \mathcal{Y}_1$. Therefore, for every $(w, u, x_1, y_1) \in \mathcal{W} \times \mathcal{U} \times \mathcal{X}_1 \times \mathcal{Y}_1$, either $Q_{X_1|W,U}(x_1|w, u) = 0$ or $Q_{Y_1|W,U}(y_1|w, u) = Q_{Y_1|X_1}(y_1|x_1)$. In particular, for $(x_1, y_1) = (1, 1)$ and any $(w, u) \in \mathcal{W} \times \mathcal{U}$, either

$$Q_{X_1|W,U}(1|w, u) = 0 \tag{108a}$$

or

$$Q_{Y_1|W,U}(1|w, u) = Q_{Y_1|X_1}(1|1) = 0.9. \tag{108b}$$

If (108b) is true, then

$$\begin{aligned} &Q_{Y_1|W,U}(1|w, u) \\ &\overset{(a)}{=} Q_{X_1|W,U}(0|w, u)Q_{Y_1|X_1}(1|0) \\ &\quad + Q_{X_1|W,U}(1|w, u)Q_{Y_1|X_1}(1|1) \\ &= 0.1 \cdot Q_{X_1|W,U}(0|w, u) + 0.9 \cdot Q_{X_1|W,U}(1|w, u) \\ &= 0.1 + 0.8 \cdot Q_{X_1|W,U}(1|w, u) \end{aligned} \tag{109}$$

where (a) uses the Markov chain $(W, U) - X_1 - Y_1$. When combined with (108b), this gives

$$Q_{X_1|W,U}(1|w, u) = 1. \tag{110}$$

Thus, for any $(w, u) \in \mathcal{W} \times \mathcal{U}$ either (108a) or (110) is true, which implies that X_1 is a deterministic function of (W, U).

Having this, we upper bound $\tilde{R}_2^\star$ as follows.

$$\begin{aligned}
\tilde{R}_2^\star &\overset{(a)}{=} \sup_{\substack{Q_W Q_{U,V,X_2|W,X_1}:\\ (W,U,V)-(X_1,X_2)-Y_2}} \min \left\{ \begin{array}{l} I(W,V;Y_2)+c, \\ I(V;Y_2|W)-I(U;V|W), \\ I(U;Y_1|W)+I(W,V;Y_2) \\ \qquad -I(U;V|W) \end{array} \right\} \\
&\overset{(b)}{=} \sup_{\substack{Q_W Q_{U,V,X_2|X_1,W}:\\ (W,U,V)-(X_1,X_2)-Y_2}} I(V;Y_2|W)-I(U;V|W) \\
&\overset{(c)}{=} \sup_{\substack{Q_W Q_{U,V,X_2|X_1,W}:\\ (W,U,V)-(X_1,X_2)-Y_2}} I(V;Y_2|W)-I(U,X_1;V|W) \\
&\leq \sup_{\substack{Q_W Q_{V,X_2|X_1,W}:\\ (W,V)-(X_1,X_2)-Y_2}} I(V;Y_2|W)-I(V;X_1|W) \\
&\overset{(d)}{\leq} \max_{w\in\mathcal{W}} \sup_{\substack{Q_{V,X_2|X_1,W=w}:\\ V_w-(X_1,X_{2,w})-Y_2}} I(V;Y_2|W=w)-I(V;X_1|W=w) \\
&\leq \sup_{\substack{Q_{V,X_2|X_1}:\\ V-(X_1,X_2)-Y_2}} I(V;Y_2)-I(V;X_1) \qquad (111)
\end{aligned}$$

where:
(a) uses the structure of $\tilde{\mathcal{O}}_{\mathsf{NS}}$, the independence of W and X_1 and the relation $R_{12}=I(W,U;Y_1)=c$;
(b) follows by the non-negativity of mutual information;
(c) is because X_1 is determined by (W,U);
(d) follows by defining $(V_w, X_{2,w})$ to be a pair of random variables jointly distributed with $X_1 \sim \mathrm{Ber}\left(\frac{1}{2}\right)$ 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 (100) is the capacity of the state-dependent channel $W_{Y_2|X_1,X_2}$ with non-causal CSI X_1^n available at both the transmitting and receiving ends. The upper bound on $\tilde{R}_2^\star$ given in (111) 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 $W_{Y_2|X_1,X_2}$ 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 $W_{Y_2|X_1,X_2}$ be defined by

$$Y_2 = X_2 \oplus X_1 \oplus Z \qquad (112)$$

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

$$\begin{aligned}
C_{\mathsf{GP}}^{(\mathsf{BDP})} &= \max_{\substack{Q_{V,X_2|X_1}:\\ V-(X_1,X_2)-Y_2}} I(V;Y_2)-I(V;Y_1) \\
&= \mathsf{uce}\left\{\left[H_b(q)-H_b(\epsilon)\right]^+\right\} \qquad (113)
\end{aligned}$$

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]

$$C_{\mathsf{F-CSI}}^{(\mathsf{BDP})} = \max_{\substack{Q_{V,X_2|X_1}:\\ V-(X_1,X_2)-Y_2}} I(V;Y_2|X_1) = H_b(q*\epsilon)-H_b(\epsilon) \qquad (114)$$

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

Appendix B
Converse Proof for (35)

To prove the optimality of (35), we show that $\mathcal{C}_{\mathsf{S}}^{(\mathsf{PD})} \subseteq \mathcal{C}_{\mathsf{S}}^{(\mathsf{G})}$ ($\mathcal{C}_{\mathsf{S}}^{(\mathsf{PD})}$ and $\mathcal{C}_{\mathsf{S}}^{(\mathsf{G})}$ are given by (23) and (35), respectively). First note that on one hand

$$h(Y_1|W) \overset{(a)}{\geq} h(Y_1|X) = h(Z_1) = \frac{1}{2}\log(2\pi e \mathrm{N}_1) \qquad (115a)$$

where (a) is because $W-X-Y_1$ forms a Markov chain, while on the other hand

$$h(Y_1|W) \leq h(Y_1) \leq \frac{1}{2}\log\left(2\pi e(\mathrm{P}+\mathrm{N}_1)\right). \qquad (115b)$$

The intermediate value theorem and (115) imply that there is an $\alpha \in [0,1]$ such that

$$h(Y_1|W) = \frac{1}{2}\log\left(2\pi e(\alpha \mathrm{P}+\mathrm{N}_1)\right). \qquad (116)$$

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

$$\begin{aligned}
h(Y_2|W=w) &= h(Y_1+Z_2|W=w) \\
&\overset{(a)}{\geq} \frac{1}{2}\log\left(2^{2h(Y_1|W=w)}+2^{2h(Z_2|W=w)}\right) \\
&\overset{(b)}{=} \frac{1}{2}\log\left(2^{2h(Y_1|W=w)}+2\pi e(\mathrm{N}_2-\mathrm{N}_1)\right) \\
&\triangleq \lambda(w) \qquad (117)
\end{aligned}$$

where (a) uses the conditional entropy-power inequality (EPI), while (b) follows by the independence of Z_2 and W. Using (117), we lower bound $h(Y_2|W)$ in terms of $h(Y_1|W)$ as

$$\begin{aligned}
h(Y_2|W) &\overset{(a)}{\geq} \mathbb{E}_W \lambda(W) \\
&\overset{(b)}{\geq} \frac{1}{2}\log\left(2^{2h(Y_1|W)}+2\pi e(\mathrm{N}_2-\mathrm{N}_1)\right) \\
&= \frac{1}{2}\log\left(2\pi e(\alpha \mathrm{P}+\mathrm{N}_2)\right) \qquad (118)
\end{aligned}$$

where (a) follows from (117), while (b) uses the convexity of the function $x \mapsto \log(2^x+c)$ for $c \in \mathbb{R}_+$ and Jensen's inequality.

We next present upper bounds on the information terms on the RHS of (23). For (23a), we have

$$\begin{aligned}
&I(X;Y_1|W)-I(X;Y_2|W) \\
&\overset{(a)}{=} h(Y_1|W)-h(Y_1|X)-h(Y_2|W)+h(Y_2|X) \\
&\overset{(b)}{\leq} \frac{1}{2}\log\left(1+\frac{\alpha \mathrm{P}}{\mathrm{N}_1}\right)-\frac{1}{2}\log\left(1+\frac{\alpha \mathrm{P}}{\mathrm{N}_2}\right) \qquad (119)
\end{aligned}$$

where (a) follows since the chain $W-X-(Y_1,Y_2)$ is Markov, while (b) relies on (116), (118) and on the Gaussian distribution maximizing the differential entropy under a variance constraint. Next, using (118) we bound the RHS of (23b) as

$$I(W;Y_2)+R_{12} = h(Y_2)-h(Y_2|W)+R_{12} \le \frac{1}{2}\log\left(1+\frac{\bar{\alpha}\mathrm{P}}{\alpha\mathrm{P}+\mathrm{N}_2}\right)+R_{12}. \quad (120)$$

By repeating arguments similar to those in the derivation of (119), we bound the sum of rates R_1+R_2 as

$$R_1+R_2 \le \frac{1}{2}\log\left(1+\frac{\mathrm{P}}{\mathrm{N}_1}\right)-\frac{1}{2}\log\left(1+\frac{\alpha\mathrm{P}}{\mathrm{N}_2}\right). \quad (121)$$

APPENDIX C
PROOF OF LEMMA 3

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

$$P^{(\mathcal{B}_n)}(\mathbf{s}_0,\mathbf{s},\mathbf{v}) = Q^n_{S_0,S}(\mathbf{s}_0,\mathbf{s})\,2^{-n\tilde{R}} \sum_{(w,i)\in\mathcal{W}_n\times\mathcal{I}_n} \hat{P}^{(\mathcal{B}_n)}(i|w,\mathbf{s}_0,\mathbf{s}) \times Q^n_{V|U,S_0,S}\big(\mathbf{v}\big|\mathbf{u}(\mathbf{s}_0,w,i),\mathbf{s}_0,\mathbf{s}\big). \quad (122)$$

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 (122) 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

$$Q^n_{V|S_0,S}(\mathbf{v}|\mathbf{s}_0,\mathbf{s}) = \sum_{\mathbf{u}\in\mathrm{supp}\left(Q^n_{U|S_0=\mathbf{s}_0,S=\mathbf{s}}\right)} Q^n_{U|S_0,S}(\mathbf{u}|\mathbf{s}_0,\mathbf{s})\,Q^n_{V|U,S_0,S}(\mathbf{v}|\mathbf{u},\mathbf{s}_0,\mathbf{s}) \quad (123)$$

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

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

$$Q^n_{U|S_0,S}\big(\mathbf{u}(\mathbf{s}_0,w,i)\big|\mathbf{s}_0,\mathbf{s}\big) = \frac{Q^n_{S_0,S,U}\big(\mathbf{s}_0,\mathbf{s},\mathbf{u}(\mathbf{s}_0,w,i)\big)}{Q^n_{S_0,S}(\mathbf{s}_0,\mathbf{s})} = \frac{Q^n_{S_0}(\mathbf{s}_0)\,Q^n_{U|S_0}\big(\mathbf{u}(\mathbf{s}_0,w,i)\big|\mathbf{s}_0\big)\,Q^n_{S|U,S_0}\big(\mathbf{s}\big|\mathbf{u}(\mathbf{s}_0,w,i),\mathbf{s}_0\big)}{Q^n_{S_0,S}(\mathbf{s}_0,\mathbf{s})} > 0.$$

APPENDIX D
ERROR PROBABILITY ANALYSIS FOR THEOREM 1

Since we evaluate the expected value (over the codebook ensemble) of the error probability and because the code is symmetric with respect to the uniformly distributed tuple (M_p,M_1,M_{22},M), we may assume that $(M_p,M_1,M_{22},W)=(1,1,1,1)$. For any event $\mathcal{A}$ from the σ-algebra over which $\mathbb{P}$ is defined, denote

$$\mathbb{P}_1 \triangleq \mathbb{P}\big(\mathcal{A}\big|M_p=1,M_{11}=1,W_1=1,M_{22}=1,W_2=1\big).$$

A. Encoding Error

An encoding error occurs if the u_1-codeword chosen by the likelihood encoder is not jointly typical with $\big(\mathbf{U}_0(M_p),\mathbf{U}_2(M_p,M_{22})\big)$. Based on the aforementioned symmetry, for any $\delta' \in (0,\delta)$, we set the event of an encoding error as

$$\mathcal{E} = \Big\{\big(\mathbf{U}_0(1),\mathbf{U}_1(1,1,1,I),\mathbf{U}_2(1,1)\big) \notin \mathcal{T}^n_{\delta'}(Q_{U_0,U_1,U_2})\Big\}. \quad (124)$$

Abbreviating $\mathcal{T} \triangleq \mathcal{T}^n_{\delta'}(Q_{U_0,U_1,U_2})$ and recalling that $\mathsf{C}^{(n)}_{0,2} \triangleq \Big\{\mathsf{C}^{(n)}_0,\mathsf{C}^{(n)}_2\Big\}$, we have

$$\begin{aligned}
&\mathbb{P}_1(\mathcal{E}) \\
&= \mathbb{E}_{\mathsf{C}_n}\mathbb{P}_1\Big(\big(\mathbf{U}_0(1),\mathbf{U}_1(1,1,1,I),\mathbf{U}_2(1,1)\big) \notin \mathcal{T}\Big|\mathsf{C}_n\Big) \\
&= \mathbb{E}_{\mathsf{C}_n}\Bigg[\sum_{i,\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2} \mathbb{1}_{\left\{\left(\mathbf{U}_0(1),\mathbf{U}_2(1,1)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}} \times P^{(\mathsf{C}_n)}_{\mathsf{LE}}(i|1,\mathbf{u}_0,\mathbf{u}_2)\mathbb{1}_{\left\{\mathbf{U}_1(1,1,1,i)=\mathbf{u}_1\right\}}\mathbb{1}_{\left\{(\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2)\notin\mathcal{T}\right\}}\Bigg] \\
&\overset{(a)}{=} \mathbb{E}_{\mathsf{C}^{(n)}_{0,2}}\Bigg[\sum_{\substack{i,\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2:\\(\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2)\notin\mathcal{T}}} \mathbb{1}_{\left\{\left(\mathbf{U}_0(1),\mathbf{U}_2(1,1)\right)=(\mathbf{u}_0,\mathbf{u}_2)\right\}} \times\mathbb{E}_{\mathsf{C}^{(n)}_1\left|\mathsf{C}^{(n)}_{0,2}\right.}\Big[P^{(\mathsf{C}_n)}_{\mathsf{LE}}(i|1,\mathbf{u}_0,\mathbf{u}_2)\mathbb{1}_{\left\{\mathbf{U}_1(1,1,1,i)=\mathbf{u}_1\right\}}\Big]\Bigg] \\
&\overset{(b)}{=} \mathbb{E}_{\tilde{\mathsf{C}}_n}\Bigg[\sum_{\substack{i,\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2:\\(\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2)\notin\mathcal{T}}} Q^n_{U_0,U_2}(\mathbf{u}_0,\mathbf{u}_2)\tilde{P}^{(\tilde{\mathsf{C}}_n)}(i|1,\mathbf{u}_0,\mathbf{u}_2) \times\mathbb{1}_{\left\{\tilde{\mathbf{U}}_1(\mathbf{u}_0,1,i)=\mathbf{u}_1\right\}}\Bigg] \\
&\overset{(c)}{=} \mathbb{E}_{\tilde{\mathsf{C}}_n}\,\mathbb{P}_{Q^n_{U_0,U_2}\times\tilde{P}}\Big(\big(\mathbf{U}_0,\tilde{\mathbf{U}}_1\big(\mathbf{U}_0,1,I\big),\mathbf{U}_2\big) \notin \mathcal{T}\Big|\tilde{\mathsf{C}}_n\Big).
\end{aligned} \quad (125)$$

In the above derivation (a) applies the law of total expectation in a similar fashion as in (65) (an inner expectation over $\mathsf{C}^{(n)}_1$ conditioned on $\mathsf{C}^{(n)}_{0,2}$, and an outer expectation over the possible

values of $\mathsf{C}_{0,2}^{(n)}$), while (c) uses (70). To justify step (b), for every $\mathcal{C}_n \in \mathfrak{C}_n$, we define (analogously to (66))

$$P_{\mathsf{LE}}^{(\mathcal{C}_n)}(i|1, \mathbf{u}_0, \mathbf{u}_2) = 0 \tag{126}$$

whenever $\mathbf{u}_0 \neq \mathbf{u}_0(1)$ or $\mathbf{u}_2 \neq \mathbf{u}_2(1,1)$, and note that for every fixed $\mathcal{C}_{0,2}^{(n)}$, we have (127) on the bottom of this page, where the last step follows by intersecting the event of interest with $\Big\{\big(\mathbf{u}_0(1), \mathbf{u}_2(1,1)\big) = (\mathbf{u}_0, \mathbf{u}_2)\Big\}$ (otherwise the probability is zero due to (126)) and, once again, using (70). Inequality (b) then follows by removing the intersection with the aforementioned event and because $\tilde{\mathsf{C}}_n$ and C_n are independent. Since the PMF $Q_{U_0,U_2}^n \tilde{P}_{\tilde{\mathsf{C}}_n, W, I, \tilde{\mathbf{U}}_1|\mathbf{U}_0,\mathbf{U}_2}$ is merely a relabeling of the induced distribution (12) in our resolvability setup, Lemma 2 implies that the RHS of (125) approaches 0 as $n \to \infty$, as long as (62a)-(62b) are satisfied.

B. Decoding Errors

To account for decoding errors, define the events in (128) at the bottom of this page.

C. Expected Average Error Probability

By the union bound, the expectation of the average error probability over the codebook ensemble[5] is bounded as (129) at the top of the next page. Note that $P_0^{[1]}$ is the probability of an encoding error, while $P_0^{[2]}$ and $P_j^{[k]}$, for $k \in [1:4]$, correspond to decoding errors of Decoder $j = 1, 2$. We proceed with the following steps:

1) The encoding error analysis shows that $P_0^{[1]} \to 0$ as $n \to \infty$ if (62a)-(62b).
2) The Conditional Typicality Lemma [49, Sec. 2.5] implies that $P_0^{[2]} \to 0$ as n grows. More precisely, there exists a function $\beta(n, \delta, \delta')$ with $\lim_{n\to\infty} \beta(n, \delta, \delta') = 0$ for any $0 < \delta' < \delta$, such that $P_0^{[2]} \leq \beta(n, \delta, \delta')$. The interested reader may refer to, e.g., [53, Th. 3.16] for precise expressions.
3) The definitions in (128) clearly give $P_j^{[1]} = 0$ for $j = 1, 2$ and every $n \in \mathbb{N}$.
4) For $P_1^{[3]}$, we have

$$\begin{aligned} P_1^{[3]} &\overset{(a)}{\leq} \sum_{\substack{(\tilde{m}_1, \tilde{w}) \neq (1,1),\\ \tilde{i} \in \mathcal{I}}} 2^{-n\left(I(U_1;Y_1|U_0) - \tau_1^{[3]}(\delta)\right)} \\ &\leq 2^{n(R_1 + \tilde{R} + R')} 2^{-n\left(I(U_1;Y_1|U_0) - \tau_1^{[3]}(\delta)\right)} \\ &= 2^{n\left(R_1 + \tilde{R} + R' - I(U_1;Y_1|U_0) + \tau_1^{[3]}(\delta)\right)} \end{aligned}$$

where (a) follows since for any $(\tilde{m}_1, \tilde{w}) \neq (1,1)$ and $\tilde{i} \in \mathcal{I}$, $\mathbf{U}_1(1, \tilde{m}_1, \tilde{w}, \tilde{i})$ is independent of $\mathbf{Y}_1$ while both of them are drawn conditioned on $\mathbf{U}_0(1)$. Moreover, $\tau_1^{[3]}(\delta) \to 0$ as $\delta \to 0$. Hence, for the probability $P_1^{[3]}$ to vanish as $n \to \infty$, we take:

$$R_1 + \tilde{R} + R' < I(U_1; Y_1|U_0) - \tau_1^{[3]}(\delta). \tag{130}$$

5) For $P_1^{[4]}$, consider

$$\begin{aligned} P_1^{[4]} &\overset{(a)}{\leq} \sum_{\substack{(\tilde{m}_p, \tilde{m}_1, \tilde{w}) \neq (1,1,1),\\ \tilde{i} \in \mathcal{I}}} 2^{-n\left(I(U_0,U_1;Y_1) - \tau_1^{[4]}(\delta)\right)} \\ &\leq 2^{n(R_p + R_1 + \tilde{R} + R')} 2^{-n\left(I(U_0,U_1;Y_1) - \tau_1^{[4]}(\delta)\right)} \\ &= 2^{n\left(R_p + R_1 + \tilde{R} + R' - I(U_0,U_1;Y_1) + \tau_1^{[4]}(\delta)\right)} \end{aligned}$$

where (a) follows since for any $(\tilde{m}_p, \tilde{m}_1, \tilde{w}) \neq (1,1,1)$ and $\tilde{i} \in \mathcal{I}$, $\mathbf{U}_0(\tilde{m}_p)$ and $\mathbf{U}_1(\tilde{m}_p, \tilde{m}_1, \tilde{w}, \tilde{i})$ are correlated with one another but independent of $\mathbf{Y}_1$. As before, $\tau_1^{[4]}(\delta) \to 0$ as $\delta \to 0$, and we have $P_1^{[4]} \to 0$ as $n \to \infty$ if

$$R_p + R_1 + \tilde{R} + R' < I(U_0, U_1; Y_1) - \tau_1^{[4]}(\delta). \tag{131}$$

6) Similar steps as in the upper bound of $P_1^{[3]}$ show that the rate bound that ensures that $P_1^{[2]} \to 0$ as $n \to \infty$ is redundant. This is since for every $\tilde{m}_p \neq 1$ and $\tilde{i} \in \mathcal{I}$, the codewords $\mathbf{U}_0(\tilde{m}_p)$ and $\mathbf{U}_1(\tilde{m}_p, 1, 1, \tilde{i})$ are independent of $\mathbf{Y}_1$. Hence, the condition

$$R_p < I(U_0, U_1; Y_1) - \tau_1^{[2]}(\delta) \tag{132}$$

where $\lim_{\delta\to 0} \tau_1^{[2]}(\delta) = 0$ suffices for $P_1^{[2]}$ to vanish. However, up to the vanishing terms, the RHS of (132) coincides with the RHS of (131), while the left-hand side (LHS) of (132) is with respect to R_p only. Clearly,

[5]We slightly abuse notation in writing $\mathbb{E}P_e(\mathsf{C}_n)$ because P_e is actually a function of the code c_n rather than the codebook $\mathcal{C}_n$. We favor this notation for its simplicity and remind the reader that $\mathcal{C}_n$ uniquely defines c_n.

$$\begin{aligned} \mathbb{E}_{\mathsf{C}_1^{(n)}\left|\mathsf{C}_{0,2}^{(n)}=\mathcal{C}_{0,2}^{(n)}\right.}\left[P_{\mathsf{LE}}^{(\mathsf{C}_n)}(i|1, \mathbf{u}_0, \mathbf{u}_2)\mathbb{1}_{\left\{\mathbf{U}_1(1,1,1,i)=\mathbf{u}_1\right\}}\right] &= \mathbb{E}_{\mathsf{C}_1^{(n)}\left|\mathsf{C}_{0,2}^{(n)}=\mathcal{C}_{0,2}^{(n)}\right.}\mathbb{P}_1\Big(I = i, \mathbf{U}_1(1,1,1,i) = \mathbf{u}_1\Big|\mathsf{C}_1^{(n)}, \mathsf{C}_{0,2}^{(n)} = \mathcal{C}_{0,2}^{(n)}\Big) \\ &\leq \mathbb{E}_{\tilde{\mathsf{C}}_n}\mathbb{P}_{\tilde{P}}\Big(I = i, \tilde{\mathbf{U}}_1(\mathbf{u}_0, 1, i) = \mathbf{u}_1\Big|W = 1, \mathbf{U}_0 = \mathbf{u}_0, \mathbf{U}_2 = \mathbf{u}_2, \tilde{\mathsf{C}}_n\Big) \end{aligned} \tag{127}$$

$$\mathcal{D}_0 = \Big\{\big(\mathbf{U}_0(1), \mathbf{U}_1(1,1,1,I), \mathbf{U}_2(1,1), \mathbf{Y}_1, \mathbf{Y}_2\big) \in \mathcal{T}_\delta^n(Q_{U_0,U_1,U_2,Y_1,Y_2})\Big\} \tag{128a}$$

$$\mathcal{D}_1(m_p, m_1, w) = \Big\{\big(\mathbf{U}_0(m_p), \mathbf{U}_1(m_p, m_1, w, I), \mathbf{Y}_1\big) \in \mathcal{T}_\delta^n(Q_{U_0,U_1,Y_1})\Big\} \tag{128b}$$

$$\mathcal{D}_2(m_p, m_{22}) = \Big\{\big(\mathbf{U}_0(m_p), \mathbf{U}_2(m_p, m_{22}), \mathbf{Y}_2\big) \in \mathcal{T}_\delta^n(Q_{U_0,U_2,Y_2})\Big\} \tag{128c}$$

$$
\begin{aligned}
\mathbb{E}P_e(\mathsf{C}_n) &\le \mathbb{P}_1\left(\mathcal{E}\cup\mathcal{D}_0^c\cup\mathcal{D}_1(1,1,1,I)^c\cup\mathcal{D}_2(1,1)^c\cup\left\{\bigcup_{\substack{(\tilde m_p,\tilde m_1,\tilde w)\\ \neq(1,1,1)}}\mathcal{D}_1(\tilde m_p,\tilde m_1,\tilde w,I)\right\}\cup\left\{\bigcup_{\substack{(\tilde m_p,\tilde m_{22})\neq(1,1):\\ \tilde m_p\in\mathcal{B}_n\left(\hat m_{12}(1)\right)}}\mathcal{D}_2(\tilde m_p,\tilde m_{22})\right\}\right)\\
&\le \mathbb{P}_1\left(\mathcal{E}\right)+\mathbb{P}_1\left(\mathcal{D}_0^c\cap\mathcal{E}^c\right)+\mathbb{P}_1\left(\mathcal{D}_1(1,1,1,I)^c\cap\mathcal{D}_0\right)+\mathbb{P}_1\left(\bigcup_{(\tilde m_p,\tilde m_1,\tilde w)\neq(1,1,1)}\mathcal{D}_1(\tilde m_p,\tilde m_1,\tilde w,I)\right)\\
&\quad+\mathbb{P}_1\left(\mathcal{D}_2(1,1)^c\cap\mathcal{D}_0\right)+\mathbb{P}_1\left(\bigcup_{\substack{(\tilde m_p,\tilde m_{22})\neq(1,1):\\ \tilde m_p\in\mathcal{B}_n\left(\hat m_{12}(1)\right)}}\mathcal{D}_2(\tilde m_p,\tilde m_{22})\right)\\
&\le \underbrace{\mathbb{P}_1\left(\mathcal{E}\right)}_{P_0^{[1]}}+\underbrace{\mathbb{P}_1\left(\mathcal{D}_0^c\cap\mathcal{E}^c\right)}_{P_0^{[2]}}+\underbrace{\mathbb{P}_1\left(\mathcal{D}_1(1,1,1,I)^c\cap\mathcal{D}_0\right)}_{P_1^{[1]}}+\underbrace{\sum_{\tilde i\in\mathcal{I}}P(\tilde i)\mathbb{P}_1\left(\bigcup_{\tilde m_p\neq1}\mathcal{D}_1(\tilde m_p,1,1,\tilde i)\right)}_{P_1^{[2]}}\\
&\quad+\underbrace{\mathbb{P}_1\left(\bigcup_{\substack{(\tilde m_1,\tilde w)\neq(1,1),\\ \tilde i\in\mathcal{I}}}\mathcal{D}_1(1,\tilde m_1,\tilde w,\tilde i)\right)}_{P_1^{[3]}}+\underbrace{\mathbb{P}_1\left(\bigcup_{\substack{(\tilde m_p,\tilde m_1,\tilde w)\neq(1,1,1),\\ \tilde i\in\mathcal{I}}}\mathcal{D}_1(\tilde m_p,\tilde m_1,\tilde w,\tilde i)\right)}_{P_1^{[4]}}+\underbrace{\mathbb{P}_1\left(\mathcal{D}_2(1,1)^c\cap\mathcal{D}_0\right)}_{P_2^{[1]}}\\
&\quad+\underbrace{\mathbb{P}_1\left(\bigcup_{\substack{\tilde m_p\neq1:\\ \tilde m_p\in\mathcal{B}_n\left(\hat m_{12}(1)\right)}}\mathcal{D}_2(\tilde m_p,1)\right)}_{P_2^{[2]}}+\underbrace{\mathbb{P}_1\left(\bigcup_{\tilde m_{22}\neq1}\mathcal{D}_2(1,\tilde m_{22})\right)}_{P_2^{[3]}}+\underbrace{\mathbb{P}_1\left(\bigcup_{\substack{(\tilde m_p,\tilde m_{22})\neq(1,1):\\ \tilde m_p\in\mathcal{B}_n\left(\hat m_{12}(1)\right)}}\mathcal{D}_2(\tilde m_p,\tilde m_{22})\right)}_{P_2^{[4]}}
\end{aligned}
\tag{129}
$$

(131) is the dominating constraint.

7) By similar arguments, we find that $P_2^{[j]}$, for $j=2,3,4$, vanish with n if

$$R_{22} < I(U_2;Y_2|U_0)-\tau_2^{[3]}(\delta) \tag{133}$$

$$R_p+R_{22}-R_{12} < I(U_0,U_2;Y_2)-\tau_2^{[4]}(\delta) \tag{134}$$

where $\tau_2^{[3]}(\delta),\tau_2^{[4]}(\delta)\to0$ as $\delta\to0$.

Summarizing the above results, by setting

$$\tau_\delta \triangleq \max\left\{\tau_j^{[k]}(\delta)\right\}_{\substack{j=1,2,\\ k=3,4}} \tag{135}$$

we find that the RHS of (129) decays as $n\to\infty$ for any $0<\delta'<\delta$ if the conditions in (62) are met.

Appendix E
Proof of the Markov Relation in (86) and (93)

We prove that (86) and (93) 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 [54]. Throughout this appendix all probabilities are taken with respect to the PMF $P^{(c_n)}$ that is induced by c_n and given in (16). For brevity, we omit the superscript and write P instead of $P^{(c_n)}$.

D. Proof of (86)

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

$$(M_0,M_2,M_{12},Y_1^{t-1},Y_{2,t+1}^n,Y_{1,t})-X_t-Y_{2,t} \tag{136}$$

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

$$(M_0,M_2,Y_1^n,Y_{2,t+1}^n)-X_t-Y_{2,t} \tag{137}$$

from which (136) follows because M_{12} is a function of Y_1^n. Since the channel is SD, memoryless and without feedback, for every $(m_0,m_1,m_2)\in\mathcal{M}_0^{(n)}\times\mathcal{M}_1^{(n)}\times\mathcal{M}_2^{(n)}$, $(x^n,y_1^n,y_2^n)\in\mathcal{X}^n\times\mathcal{Y}_1^n\times\mathcal{Y}_2^n$ and $t\in[1:n]$, we have

$$
\begin{aligned}
&P(m_0,m_1,m_2,x^n,y_1^n,y_2^n)\\
&\quad=P(m_0)P(m_1)P(m_2)P(x^n|m_0,m_1,m_2)\\
&\qquad\times P\left(y_1^{t-1}\middle|x^{t-1}\right)P\left(y_2^{t-1}\middle|x^{t-1}\right)P(y_{1,t}|x_t)\\
&\qquad\times P(y_{2,t}|x_t)P\left(y_{1,t+1}^n\middle|x_{t+1}^n\right)P\left(y_{2,t+1}^n\middle|x_{t+1}^n\right).
\end{aligned}
\tag{138}
$$

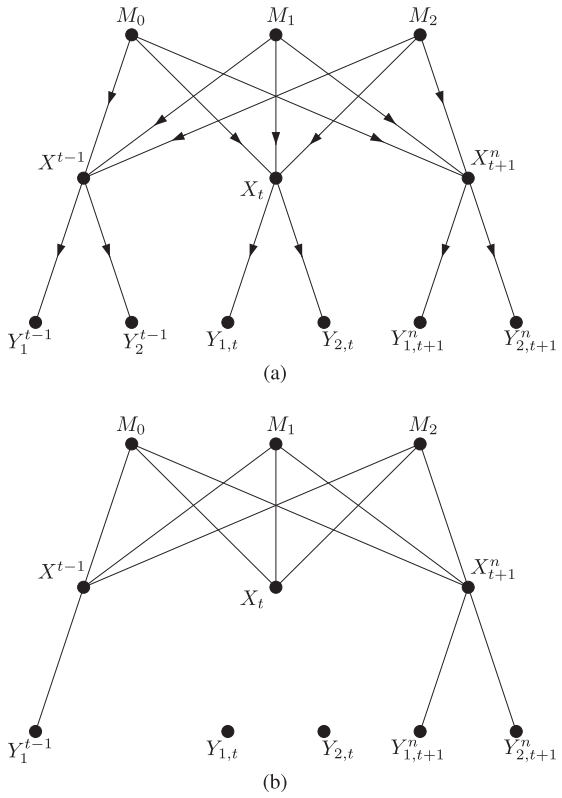


Fig. 10. (a) The FDG that stems from (138): (137) follows since $\mathcal{C} = \{X_t\}$ d-separates $\mathcal{A} = \{Y_{2,t}\}$ from $\mathcal{B} = \{M_0, M_2, Y_1^n, Y_{2,t+1}^n\}$. (b) The undirected graph obtained from the FDG after the manipulations described in Definition [54, Definition 1]. Both FDGs omit the dependence of the channel outputs on the noise.

Fig. 10(a) shows the FDG induced by (138). 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 (137) follows by setting $\mathcal{A} = \{Y_{2,t}\}$, $\mathcal{B} = \{M_0, M_2, Y_1^n, Y_{2,t+1}^n\}$ and $\mathcal{C} = \{X_t\}$, and noting that $\mathcal{C}$ d-separates $\mathcal{A}$ from $\mathcal{B}$ by applying the manipulations described in [54, Definition 1].

E. Proof of (93)

To prove (93), is suffices to show that Markov relations

$$(M_0, M_2, Y_1^{t-1}, Y_{2,t+1}^n) - X_t - Y_{1,t} \tag{139a}$$

$$(M_0, M_2, Y_1^{t-1}, Y_{2,t+1}^n, X_t) - Y_{1,t} - Y_{2,t} \tag{139b}$$

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

$$\begin{aligned} &P(m_0, m_1, m_2, x^n, y_1^n, y_2^n) \\ &\quad = P(m_0)P(m_1)P(m_2)P(x^n|m_0, m_1, m_2) \\ &\qquad \times P\big(y_1^{t-1}\big|x^{t-1}\big)P\big(y_2^{t-1}\big|y_1^{t-1}\big)P(y_{1,t}|x_t) \\ &\qquad \times P(y_{2,t}|y_{1,t})P\big(y_{1,t+1}^n\big|x_{t+1}^n\big)P\big(y_{2,t+1}^n\big|y_{1,t+1}^n\big). \end{aligned} \tag{140}$$

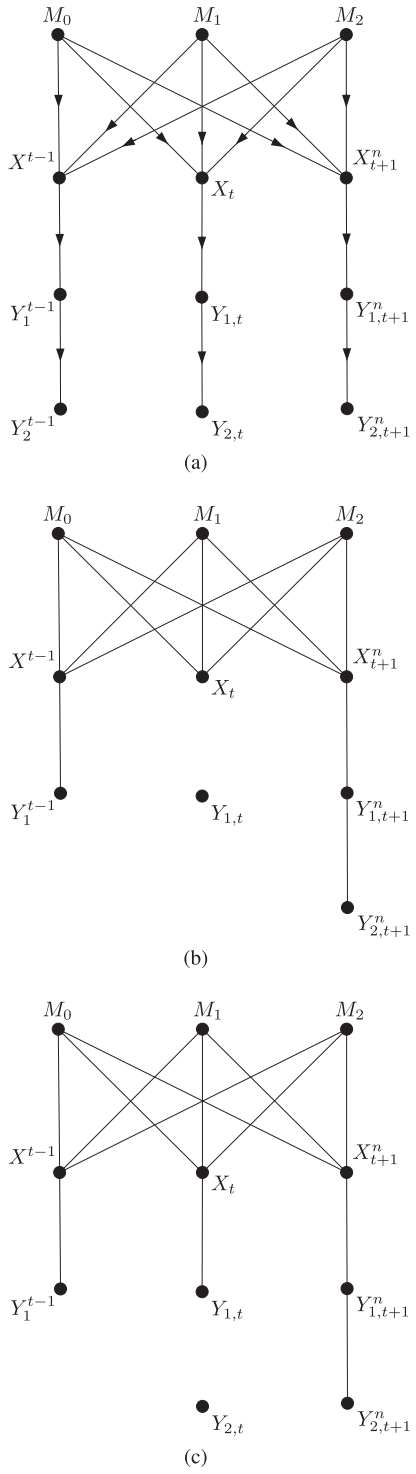


Fig. 11. (a) The FDG that stems from (140): (139) follows since $\mathcal{C}_j$ d-separates $\mathcal{A}_j$ from $\mathcal{B}_j$, for $j = 1, 2$. (b) The undirected graph that corresponds to $\mathcal{A}_1$, $\mathcal{B}_1$ and $\mathcal{C}_1$. (c) The undirected graph that corresponds to $\mathcal{A}_2$, $\mathcal{B}_2$ and $\mathcal{C}_2$. The FDGs omit the dependence of the channel outputs on the noise.

The FDG induced by (140) is shown in Fig. 11(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 (139) 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 [54, Definition 1].

REFERENCES

[1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[2] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 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*, vol. 54, no. 6, pp. 2493–2507, Jun. 2008.

[4] Y. Zhao, P. Xu, Y. Zhao, W. Wei, and Y. Tang, "Secret communications over semi-deterministic broadcast channels," in *Proc. 4th Int. Conf. Commun. Netw. China (CHINACOM)*, Xi'an, China, Aug. 2009, pp. 1–4.

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

[6] Z. Goldfeld, G. Kramer, and H. H. Permuter, "Broadcast channels with privacy leakage constraints," *IEEE Trans. Inf. Theory*, 2015. [Online]. Available: http://arxiv.org/abs/1504.06136

[7] E. Ekrem and S. Ulukus, "Secrecy in cooperative relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 137–155, Jan. 2011.

[8] R. Liu and H. V. Poor, "Secrecy capacity region of a multiple-antenna gaussian broadcast channel with confidential messages," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1235–1249, Mar. 2009.

[9] T. Liu and S. Shamai (Shitz), "A note on the secrecy capacity of the multiple-antenna wiretap channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2547–2553, Jun. 2009.

[10] R. Liu, T. Liu, H. V. Poor, and S. Shamai (Shitz), "Multiple-input multiple-output Gaussian broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4215–4227, Sep. 2010.

[11] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—Part II: The MIMOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 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*, vol. 57, no. 4, pp. 2083–2114, Apr. 2011.

[13] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4961–4972, Aug. 2011.

[14] E. Ekrem and S. Ulukus, "Secrecy capacity of a class of broadcast channels with an eavesdropper," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, no. 1, pp. 1–29, Mar. 2009.

[15] G. Bagherikaram, A. S. Motahari, and A. K. Khandani, "Secrecy capacity region of Gaussian broadcast channel," in *Proc. 43rd Annu. Conf. Inf. Sci. Syst. (CISS)*, Baltimore, MD, USA, Mar. 2009, pp. 152–157.

[16] M. Benammar and P. Piantanida, "Secrecy capacity region of some classes of wiretap broadcast channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5564–5582, Oct. 2015.

[17] U. M. Maurer, "The strong secret key rate of discrete random triples," in *Communications and Cryptography: Two Sides of One Tapestry*. Norwell, MA, USA: Springer, 1994, pp. 271–285.

[18] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Proc. 19th Int. Conf. Theory Appl. cryptograph. Tech. (EUROCRYPT)*, 2000, pp. 351–368.

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

[20] I. Csiszár, "Almost independence and secrecy capacity," *Problems Inf. Transmiss.*, vol. 32, no. 1, pp. 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 channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, Apr. 2006.

[22] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[23] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[24] J. Hou and G. Kramer, "Informational divergence approximations to product distributions," in *Proc. 13th Can. Workshop Inf. Theory*, Toronto, ON, Canada, Jun. 2013, pp. 76–81.

[25] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[26] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7584–7605, Dec. 2014.

[27] C. Schieler and P. Cuff, "The Henchman problem: Measuring secrecy by the minimum distortion in a list," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3436–3450, Jun. 2016.

[28] E. C. Song, P. Cuff, and H. V. Poor, "A rate-distortion based secrecy system with side information at the decoders," in *Proc. 52nd Annu. Allerton Conf. Commun., Control Comput.*, Monticell, IL, USA, Sep./Oct. 2014, pp. 755–762.

[29] S. Satpathy and P. Cuff, "Secure coordination with a two-sided helper," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun./Jul. 2014, pp. 406–410.

[30] M. R. Bloch and J. N. Laneman, "Strong secrecy from channel resolvability," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8077–8098, Dec. 2013.

[31] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *Proc. Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun./Jul. 2014, pp. 601–605.

[32] T. S. Han, H. Endo, and M. Sasaki, "Reliability and secrecy functions of the wiretap channel under cost constraint," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6819–6843, Nov. 2014.

[33] E. C. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1836–1849, Apr. 2016.

[34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New-York, NY, USA: Wiley, 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," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 2285–2307, May 2016.

[36] E. V. D. Meulen, "Random coding theorems for the general discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 180–190, Mar. 1975.

[37] S. I. Gel'fand, "Capacity of one broadcast channel," *Problems Inf. Transmiss.*, vol. 13, no. 3, p. 106108, Jul./Sep. 1977.

[38] J. L. Massey, *Lecture Notes for Applied Digital Information Theory*, ETH Zurich, Zürich, Switzerland, 1980–1998.

[39] A. A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 608–619, Feb. 2012.

[40] H. G. Eggleston, *Convexity*, 6th ed. Cambridge, U.K.: Cambridge Univ. Press, 1958.

[41] Y. Liang and V. V. Veeravalli, "Cooperative relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 900–928, Mar. 2007.

[42] Y. Liang and G. Kramer, "Rate regions for relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3517–3535, Oct. 2007.

[43] L. Dikstein, H. H. Permuter, and Y. Steinberg, "On state-dependent degraded broadcast channels with cooperation," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2308–2323, May 2016.

[44] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, 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*, vol. 49, no. 5, pp. 1159–1180, May 2003.

[46] A. Khina, T. Philosof, U. Erez, and R. Zamir, "Binary dirty MAC with common state information," in *Proc. 26th Conv. Electron. Electron. Eng. (IEEEI)*, Eilat, Israel, Nov. 2010, pp. 496–500.

[47] S. I. Gel'fand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Problems Inf. Transmiss.*, vol. 16, no. 1, pp. 17–25, Jan./Mar. 1980.

[48] R. Dabora and S. D. Servetto, "Broadcast channels with cooperating decoders," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5438–5454, Dec. 2006.

[49] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[50] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Semantic-security capacity for wiretap channels of type II," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3863–3879, Jul. 2016.

[51] I. B. Gattegno, Z. Goldfeld, and H. H. Permuter, "Fourier–Motzkin elimination software for information theoretic inequalities," *IEEE Inf. Theory Soc. Newslett.*, vol. 65, no. 3, pp. 25–28, Sep. 2015.

[52] G. Kramer, "Teaching IT: An identity for the Gelfand-Pinsker converse," *IEEE Inf. Theory Soc. Newslett.*, vol. 61, no. 4, pp. 4–6, Dec. 2011.

[53] G. Kramer, *Lecture Notes for Multi-User Information Theory*. Munich, Germany: Technical Univ. Munich, 2012.
[54] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

Ziv Goldfeld (S'13) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 2012 and 2014, respectively. He is currently a student in the direct Ph.D. program for honor students in Electrical and Computer Engineering at that same institution. Between 2003 and 2006, he served in the intelligence corps of the Israeli Defense Forces.

Ziv is a recipient of several awards, among them are the Dean's List Award, the Basor Fellowship, the Lev-Zion fellowship, IEEEI-2014 best student paper award, a Minerva Short-Term Research Grant (MRG), and a Feder Family Award in the national student contest for outstanding research work in the field of communications technology.

Gerhard Kramer (S'91–M'94–SM'08–F'10) received the Dr. sc. techn. (Doktor der technischen Wissenschaften) degree from the Swiss Federal Institute of Technology (ETH), Zurich, in 1998.

From 1998 to 2000, he was with Endora Tech AG, Basel, Switzerland, as a Communications Engineering Consultant. From 2000 to 2008, he was with Bell Labs, Alcatel-Lucent, Murray Hill, NJ, as a Member of Technical Staff. He joined the University of Southern California (USC), Los Angeles, in 2009. Since 2010, he has been a Professor and Head of the Institute for Communications Engineering at the Technical University of Munich (TUM), Munich, Germany.

Dr. Kramer served as the 2013 President of the IEEE Information Theory Society. He has won several awards for his work and teaching, including an Alexander von Humboldt Professorship in 2010 and a Lecturer Award from the Student Association of the TUM Electrical and Computer Engineering Department in 2015. He has been a member of the Bavarian Academy of Sciences and Humanities since 2015.

Haim H. Permuter (M'08–SM'13) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and the Ph.D. degree in Electrical Engineering from Stanford University, California in 2008.

Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. Since 2009 he is with the department of Electrical and Computer Engineering at Ben-Gurion University where he is currently an associate professor.

Prof. Permuter is a recipient of several awards, among them the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and the U.S.-Israel Binational Science Foundation Bergmann Memorial Award. Haim is currently serving on the editorial board of the IEEE Transactions on Information Theory.

Paul Cuff (S'08–M'10) received the B.S. degree in electrical engineering from Brigham Young University, Provo, UT, in 2004 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University in 2006 and 2009. Since 2009 he has been an Assistant Professor of Electrical Engineering at Princeton University.

As a graduate student, Dr. Cuff was awarded the ISIT 2008 Student Paper Award for his work titled Communication Requirements for Generating Correlated Random Variables and was a recipient of the National Defense Science and Engineering Graduate Fellowship and the Numerical Technologies Fellowship. As faculty, he received the NSF Career Award in 2014 and the AFOSR Young Investigator Program Award in 2015.