Duality of a Source Coding Problem and the Semi-Deterministic Broadcast Channel With Rate-Limited Cooperation

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

Abstract—The Wyner–Ahlswede–Körner (WAK) empirical-coordination problem where the encoders cooperate via a finite-capacity one-sided link is considered. The coordination-capacity region is derived by combining several source coding techniques, such as Wyner–Ziv coding, binning, and superposition coding. Furthermore, a semi-deterministic (SD) broadcast channel (BC) with one-sided decoder cooperation is considered. Duality principles relating the two problems are presented, and the capacity region for the SD-BC setting is derived. The direct part follows from an achievable region for a general BC that is tight for the SD scenario. A converse is established by using telescoping identities. The SD-BC is shown to be operationally equivalent to a class of relay-BCs, and the correspondence between their capacity regions is established. The capacity region of the SD-BC is transformed into an equivalent region that is shown to be dual to the admissible region of the WAK problem in the sense that the information measures defining the corner points of both regions coincide. Achievability and converse proofs for the equivalent region are provided. For the converse, we use a probabilistic construction of auxiliary random variables that depends on the distribution induced by the codebook. Several examples illustrate the results.

Index Terms—Channel and source duality, cooperation, empirical coordination, multiterminal source coding, relay-broadcast channel, semi-deterministic broadcast channel.

I. INTRODUCTION

COOPERATION can substantially improve the performance of a network. A common form of cooperation permits information exchange between the transmitting and receiving ends via rate-limited links, generally referred to as *conferencing* [1]. In this work, conferencing is incorporated in
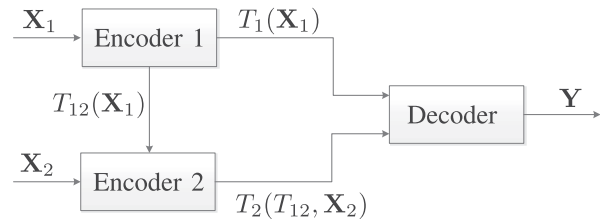
Fig. 1. The WAK source coding problem.

a special case of the fundamental two-encoder multiterminal source coding problem (cf., e.g., [2], [3]). Solutions for several special cases of the two-encoder source coding problem have been provided. Among these are the Slepian-Wolf (SW) [4], Wyner-Ziv (WZ) [5], Gaussian quadratic [6] and Wyner-Ahlswede-Körner (WAK) [7], [8] problems. The last setting refers to two correlated sources that are separately compressed, and their compressed versions are conveyed to the decoder, which reproduces only one of the sources in a lossless manner. We consider the WAK problem with conferencing (Fig. 1) in which a pair of correlated sources (X_1, X_2) are compressed by two encoders that are connected via a one-sided rate-limited link that extends from the 1st encoder to the 2nd. The compressed versions are conveyed to the decoder that outputs an empirical coordination sequence Y from which X_1 can be reproduced in a lossless manner.

Source coordination is an alternative formulation for lossy source coding. *Strong coordination* was considered by Wyner [9], while *empirical coordination* was studied in [10]–[12]. Cuff *et al.* extended these results to the multiuser case [13]. Rather than sending data from one point to another with a fidelity constraint, in a coordination problem all network nodes should develop certain joint statistics. Moreover, it was shown in [13] that rate-distortion theory is a special case of source coordination. In this work, we consider empirical coordination, a problem in which the terminals, upon observing correlated sources, generate sequences with a desired empirical joint distribution. A closely related empirical coordination problem was presented by Bereyhi *et al.* [14], who considered a triangular multiterminal network. In this setting, each of the two terminals receives a different correlated source that it compresses and conveys to the decoder. The decoder outputs a sequence that achieves the desired coordination. Moreover, the
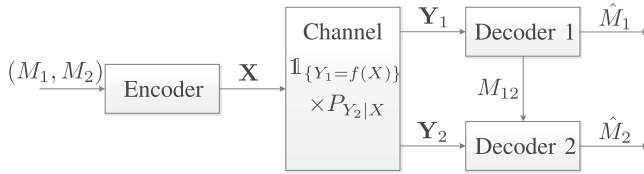
Fig. 2.   SD-BC with one-sided decoder cooperation.

encoders in [14] may share information via a one-sided cooperation link (see [15] and references therein for additional work involving cooperation in source coding problems). The main contributions of [14] comprise inner and outer bounds on the optimal rate region.

The WAK problem with cooperation considered here is a special case of the triangular multiterminal network in [14] where the sequence X_1 is losslessly reproduced from the output coordination sequence. We derive a single-letter characterization of the coordination-capacity region for this problem. The direct proof unifies several concepts in source coding by relying on WZ coding [5], binning [16] and superposition coding [17]. Note that in the classical WAK problem, where the encoders are non-cooperative, coordination of the output with the side information (i.e., the sequence X_2 in Fig. 1) is achieved even though it is not required. Therefore, adding such a coordination constraint to the classic WAK problem does not alter its solution, which can be obtained as a special case of the rate region we give here. The non-cooperative version of the problem in Fig. 1, i.e., where one of the sources is losslessly reproduced while coordination with the other source *is* required, was studied by Berger and Yeung in [18].

To explore duality, we consider a channel coding problem (Fig. 2) that we show is *dual* to the WAK problem of interest. By interchanging the roles of the encoders and decoder of the WAK problem, we obtain a semi-deterministic (SD) broadcast channel (BC) where the decoders cooperate via a rate-limited link. This duality naturally extends the well-known duality between point-to-point (PTP) source and channel coding problems. PTP duality has been widely treated in the literature since it was observed by Shannon in 1959 [19] (see [20]–[22] and references therein). Multiuser duality, however, remains obscure, despite the attention it attracted in the last decade [15], [23]–[25]. We provide principles according to which the two problems can be transformed from one to the other. Moreover, we show that the admissible rate regions of the considered SD-BC and WAK problems are dual. The duality is in the sense that the information measures that define the corner points of both regions coincide, which extends the relation between dual results in the PTP situation.

Cooperative communication over noisy channels was extensively treated in the literature since it was introduced by Willems in the context of a multiple-access channel (MAC), in which the encoders are able to hold a conference [1]. The Gaussian case was solved by Bross *et al.* in [26], followed by several works involving the compound MAC [27], [28]. Cooperation between receivers in a broadcast channel (BC) was introduced by Dabora and Servetto [29]. Liang and Veeravalli generalized the work in [29] by examining the problem of a relay-BC (RBC) [30]. In both [29] and [30],

the capacity region of the physically degraded BC (PD-BC) is characterized. Here we combine cooperation in a SD-BC setting.

The SD-BC without cooperation was solved by Gelfand and Pinsker [31]. The coding scheme was based on Marton's scheme for BCs [32] (see [33] for a generalization of [31] to the state-dependent case). We derive the capacity region of the SD-BC with cooperation by first deriving an inner bound on the capacity region of the cooperative general BC. The achievable scheme combines rate-splitting with Marton and superposition coding. The cooperation protocol uses binning to increase the transmission rate to the cooperation-aided user. The inner bound is then reduced to the SD-BC case and shown to be tight by providing a converse. The presented converse proof takes a simple and compact form by leveraging telescoping identities [34].

There is a close relation between the SD-BC with cooperation and a class of SD-RBCs considered in [35]. We show that a SD-RBC with an orthogonal and deterministic relay is operationally equivalent to the SD-BC with cooperation (see [36] for a related work on the equivalence between PTP channels in a general network and noiseless bit-pipes with the same capacity). Consequently, the capacity regions of the two problems are the same. However, there are several advantages to our approach. First, we present a capacity achieving coding scheme over a *single* transmission block, while [35] relies on transmitting many blocks and applying backward decoding. Thus, our scheme avoids the delay introduced by backward decoding. Second, our converse proof is considerably simpler than in [35]. Finally, considering the SD-BC with a one-sided conferencing link between the decoders gives insight into multiuser channel-source duality [37].

To show the duality between the optimal rate regions of the considered source and channel coding problems, an alternative characterization of the capacity region of the SD-BC is given. The corner points of the alternative region satisfy the correspondence to those of the coordination-capacity region of the WAK problem. The structure of the alternative expression motivates a converse proof technique that generalizes classical techniques. Specifically, our converse uses auxiliary random variables that are not only *chosen as a function of the joint distribution induced by each codebook*, but that are *constructed in a probabilistic manner* (see [33] for a deterministic codebook-dependent construction of auxiliaries). Allowing a probabilistic construction of the auxiliary random variables introduces additional optimization parameters (i.e., a probability distribution). By optimizing over the probability values, an upper bound on the alternative formulation of the capacity region is tightened to coincide with the achievable region. Probabilistic arguments of a similar nature were previously used in the literature [38]–[40]. The novelty of our approach is the incorporation of such arguments in a converse proof to describe the optimal choice of auxiliaries. Moreover, a closed form formula for the optimal probability values is derived as part of the converse and highlights the dependence of the choice of auxiliaries on the codebook.

This paper is organized as follows. In Section II we describe the two models of interest - the WAK problem with

encoder cooperation and the SD-BC with decoder cooperation. In Section III, we state capacity results for the WAK and BC models. In Section IV we analyse the duality between the two problems and their capacity regions. In Section V we discuss the relation of the considered SD-BC to a class of SD-RBCs. Section VI presents special cases of the capacity region of the SD-BC, and each case is shown to preserve a dual relation to the corresponding reduced source coding problem. Finally, Section VII summarizes the main achievements and insights of this work.

II. PRELIMINARIES AND PROBLEM DEFINITIONS

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

For every sequence $\mathbf{x} \in \mathcal{X}^n$, the empirical PMF of $\mathbf{x}$ is

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

where $N(a|\mathbf{x}) = \sum_{i=1}^n \mathbb{1}_{\{x_i = a\}}$. We use $\mathcal{T}_\epsilon^n(P_X)$ to denote the set of letter-typical sequences of length n with respect to the PMF P_X and the non-negative number ϵ [41, Ch. 3], [42],

i.e., we have

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

Furthermore, for a PMF $P_{X,Y}$ over $\mathcal{X} \times \mathcal{Y}$ and a fixed sequence $\mathbf{y} \in \mathcal{Y}^n$, we define

$$\mathcal{T}_\epsilon^n(P_{X,Y}|\mathbf{y}) = \left\{ \mathbf{x} \in \mathcal{X}^n \middle| (\mathbf{x}, \mathbf{y}) \in \mathcal{T}_\epsilon^n(P_{X,Y}) \right\}. \tag{3}$$

A. The WAK Source Coordination Problem With One-Sided Encoder Cooperation

Consider the source coding problem illustrated in Fig. 1. Two source sequences $\mathbf{x}_1 \in \mathcal{X}_1^n$ and $\mathbf{x}_2 \in \mathcal{X}_2^n$ are available at Encoder 1 and Encoder 2, respectively. The sources are drawn in a pairwise i.i.d. manner according to the PMF Q_{X_1, X_2}.[1] Each encoder communicates with the decoder by sending a message via a noiseless communication link of limited rate. The rate of the link between Encoder j and the decoder is R_j and the corresponding message is t_j, where $j = 1, 2$. Moreover, Encoder 1 can communicate with Encoder 2 over a one-sided communication link of rate R_{12}.

Definition 1 (Coordination Code): An (n, R_{12}, R_1, R_2) coordination code $\mathcal{L}_n$ for the WAK source coordination problem with one-sided encoder cooperation has:
1) *Three message sets: $\mathcal{T}_{12} = [1 : 2^{nR_{12}}]$, $\mathcal{T}_1 = [1 : 2^{nR_1}]$ and $\mathcal{T}_2 = [1 : 2^{nR_2}]$.*
2) *An encoder cooperation function:*

$$f_{12} : \mathcal{X}_1^n \to \mathcal{T}_{12}. \tag{4a}$$

3) *Two encoding functions:*

$$f_1 : \mathcal{X}_1^n \to \mathcal{T}_1 \tag{4b}$$

$$f_2 : \mathcal{X}_2^n \times \mathcal{T}_{12} \to \mathcal{T}_2. \tag{4c}$$

4) *A decoding function:*

$$\phi : \mathcal{T}_1 \times \mathcal{T}_2 \to \mathcal{Y}^n. \tag{4d}$$

Definition 2 (Total Variation): Let $\mathcal{X}$ be a countable space[2] and let $P, Q \in \mathcal{P}(\mathcal{X})$. The total variation (TV) distance between P and Q is

$$||P - Q||_{TV} = \frac{1}{2} \sum_{a \in \mathcal{X}} |P(a) - Q(a)|. \tag{5}$$

Let $\mathcal{Q}$ be the set of PMFs defined in (6), as shown at the top of the next page.

Definition 3 (Coordination Error): Let $P_{X_1,X_2,Y} \in \mathcal{Q}$ and $\delta > 0$. The coordination error $e_\delta(P_{X_1,X_2,Y}, \mathcal{L}_n)$ of an (n, R_{12}, R_1, R_2) coordination code $\mathcal{L}_n$ with respect to $P_{X_1,X_2,Y}$ is given in (7), as shown at the top of the next page.

Definition 4 (Coordination Achievability): Let $P_{X_1,X_2,Y} \in \mathcal{Q}$. A rate triple (R_{12}, R_1, R_2) is $P_{X_1,X_2,Y}$-achievable if for every $\epsilon, \delta > 0$ there is a sufficiently large

---

[1]We usually use Q to denote a PMF that is fixed as part of the problem's definition, while P is used for PMFs that we optimize over.

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

$$\mathcal{Q} = \left\{ P_{X_1,X_2,Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) \middle| \begin{array}{l} \exists f : \mathcal{Y} \to \mathcal{X}_1, \ P_{X_1,X_2,Y} = Q_{X_2} P_{Y|X_2} \mathbb{1}_{\{X_1 = f(Y)\}}, \\ \sum_{y \in \mathcal{Y}} P_{X_1,X_2,Y}(x_1, x_2, y) = Q_{X_1,X_2}(x_1, x_2), \forall (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 \end{array} \right\} \tag{6}$$

$$e_\delta(P_{X_1,X_2,Y}, \mathcal{L}_n) \triangleq \mathbb{P}_{\mathcal{L}_n}\left(\left\| \nu_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}} - P_{X_1,X_2,Y} \right\|_{TV} \geq \delta \right) = \sum_{\substack{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}) \in \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n: \\ \|\nu_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}} - P_{X_1,X_2,Y}\|_{TV} \geq \delta}} Q^n_{X_1,X_2}(\mathbf{x}_1, \mathbf{x}_2) \mathbb{1}_{\left\{ \phi\left(f_1(\mathbf{x}_1), f_2\left(\mathbf{x}_2, f_{12}(\mathbf{x}_1)\right)\right) = \mathbf{y} \right\}}$$

$$\tag{7}$$

$$e(\mathcal{C}_n) \triangleq \mathbb{P}_{\mathcal{C}_n}\left((\hat{M}_1, \hat{M}_2) \neq (M_1, M_2) \right) = 2^{-n(R_1 + R_2)} \sum_{(m_1,m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} \sum_{\substack{(\mathbf{y}_1,\mathbf{y}_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2: \\ \psi(\mathbf{y}_1) \neq m_1 \ \text{or} \\ \psi(\mathbf{y}_2, g_{12}(\mathbf{y}_1)) \neq m_2}} Q^n_{Y_1,Y_2|X}\left(\mathbf{y}_1, \mathbf{y}_1 \middle| g(m_1, m_2)\right)$$

$$\tag{9}$$

$n \in \mathbb{N}$ and an (n, R_{12}, R_1, R_2) coordination code $\mathcal{L}_n$ such that $e_\delta(P_{X_1,X_2,Y}, \mathcal{L}_n) \leq \epsilon$.

Definition 5 (Coordination-Capacity Region): The coordination-capacity region $\mathcal{R}_{\mathrm{WAK}}(P_{X_1,X_2,Y})$ with respect to a PMF $P_{X_1,X_2,Y} \in \mathcal{Q}$ is the closure of the set of $P_{X_1,X_2,Y}$-achievable rate triples (R_{12}, R_1, R_2).

B. SD-BCs With One-Sided Decoder Cooperation

The SD-BC with cooperation is illustrated in Fig. 2. The channel has one sender and two receivers. The sender chooses a pair (m_1, m_2) of indices uniformly and independently from the $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$ and maps them to a sequence $\mathbf{x} \in \mathcal{X}^n$, which is the channel input. The sequence $\mathbf{x}$ is transmitted over a BC with transition probability $Q_{Y_1,Y_2|X} = \mathbb{1}_{\{Y_1 = f(X)\}} Q_{Y_2|X}$. The output sequence $\mathbf{y}_j \in \mathcal{Y}_j^n$, where $j = 1, 2$, is received by decoder j. Decoder j produces an estimate of m_j, which is denoted by $\hat{m}_j$. There is a one-sided noiseless cooperation link of rate R_{12} from Decoder 1 to Decoder 2. By conveying a message $m_{12} \in [1 : 2^{nR_{12}}]$ over this link, Decoder 1 can share with Decoder 2 information about $\mathbf{y}_1$, $\hat{m}_1$, or both.

Definition 6 (Code): An (n, R_{12}, R_1, R_2) code $\mathcal{C}_n$ for the SD-BC with one-sided decoder cooperation has:

1) Three message sets $\mathcal{M}_{12} = [1 : 2^{nR_{12}}]$, $\mathcal{M}_1 = [1 : 2^{nR_1}]$ and $\mathcal{M}_2 = [1 : 2^{nR_2}]$.
2) An encoding function:

$$g : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}^n. \tag{8a}$$

3) A decoder cooperation function:

$$g_{12} : \mathcal{Y}_1^n \to \mathcal{M}_{12}. \tag{8b}$$

4) Two decoding functions:

$$\psi_1 : \mathcal{Y}_1^n \to \mathcal{M}_1 \tag{8c}$$

$$\psi_2 : \mathcal{Y}_2^n \times \mathcal{M}_{12} \to \mathcal{M}_2. \tag{8d}$$

Definition 7 (Error Probability): The average error probability $e(\mathcal{C}_n)$ of an (n, R_{12}, R_1, R_2) code $\mathcal{C}_n$ is given in (9), as shown at the top of this page.

Definition 8 (Achievability): A rate triple (R_{12}, R_1, R_2) is achievable if for any $\epsilon > 0$ there is a sufficiently large $n \in \mathbb{N}$ and an (n, R_{12}, R_1, R_2) code $\mathcal{C}_n$ such that $e(\mathcal{C}_n) \leq \epsilon$.

Definition 9 (Capacity Region): The capacity region $\mathcal{C}_{\mathrm{BC}}$ of the SD-BC with one-sided encoder cooperation is the closure of the set of achievable rate triples (R_{12}, R_1, R_2).

III. MAIN RESULTS

We state our main results as the coordination-capacity region of the WAK source coordination problem (Section II-A) and the capacity region of the SD-BC with cooperation (Section II-B).

Theorem 1 (WAK Problem Coordination-Capacity): The coordination-capacity region $\mathcal{R}_{\mathrm{WAK}}(P_{X_1,X_2,Y})$ of the WAK source coordination problem with one-sided encoder cooperation with respect to a PMF $P_{X_1,X_2,Y} \in \mathcal{Q}$ is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_{12} \geq I(V; X_1 | X_2) \tag{10a}$$

$$R_1 \geq H(X_1 | V, U) \tag{10b}$$

$$R_2 \geq I(U; X_2 | X_1, V) \tag{10c}$$

$$R_1 + R_2 \geq H(X_1 | V, U) + I(V, U; X_1, X_2) \tag{10d}$$

where the union is over all PMFs $Q_{X_1,X_2} P_{V|X_1}$ $P_{U|X_2,V} P_{Y|X_1,U,V}$ that have $P_{X_1,X_2,Y}$ as a marginal. Moreover, $\mathcal{R}_{\mathrm{WAK}}(P_{X_1,X_2,Y})$ is convex and one may choose $|\mathcal{V}| \leq |\mathcal{X}_1| + 3$ and $|\mathcal{U}| \leq |\mathcal{V}| \cdot |\mathcal{X}_2| + 3$.

See Appendix A for the proof of Theorem 1.

Remark 2: For a fixed PMF in Theorem 1, the triples (R_{12}, R_1, R_2) at the corner points of $\mathcal{R}_{\mathrm{WAK}}(P_{X_1,X_2,Y})$ are (see Fig. 3)

$$\left(I(V; X_1 | X_2), H(X_1), I(U; X_2 | X_1, V) \right) \tag{11a}$$

$$\left(I(V; X_1 | X_2), H(X_1 | V, U), I(U; X_2 | V) + I(V; X_1) \right). \tag{11b}$$

The corner point in (11b) is achieved using the coding scheme from [14] by setting $V = 0$ in [14, Th. 1]. However, the rate triple (11a) does not seem to be achievable for that scheme.

Remark 3: The cardinality bounds on the auxiliary random variables V and U in Theorem 1 are established by standard application of the Eggleston-Fenchel-Carathéodory theorem [43, Th. 18] twice. The details are omitted.

The source coordination problem defined in Section II-A can be transformed into an equivalent rate-distortion problem. This is done by substituting $\mathbf{Y}$, the output of the coordination
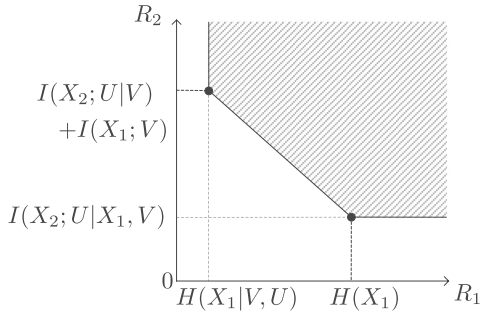
Fig. 3. Corner points of the coordination-capacity region of the WAK coordination problem with cooperation at the hyperplane where $R_{12} = I(X_1; V|X_2)$.

problem, with the pair $(\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2)$, where $\hat{\mathbf{X}}_1$ is a lossless reconstruction of the source sequence $\mathbf{X}_1$, while $\hat{\mathbf{X}}_2$ satisfies the distortion constraint

$$\mathbb{E}\left[\sum_{i=1}^{n} d(X_{2,i}, \hat{X}_{2,i})\right] \leq D \qquad (12)$$

where $d : \mathcal{X}_2 \times \hat{\mathcal{X}}_2 \to \mathbb{R}_+$ is a single-letter distortion measure and $D \in \mathbb{R}_+$ is the distortion constraint. The two models are equivalent in the sense that the rate bounds that describe the optimal rate regions of both problems are the same; the domain over which the union is taken, however, is slightly modified. This gives rise to the following corollary.

Corollary 4 (WAK Problem Rate-Distortion Region): The rate-distortion region $\mathcal{R}_{\mathrm{WAK}}(D)$ for the equivalent rate-distortion problem is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying (10)*, where the union is over all PMFs $Q_{X_1,X_2} P_{V|X_1} P_{U|X_2,V}$ and the reconstructions $\hat{X}_2$ that are a functions of (X_1, U, V) such that $\mathbb{E}[d(X_2, \hat{X}_2)] \leq D$.*

The proof of Corollary 4 is similar to that of Theorem 1 and is omitted. We next state the capacity region of the SD-BC with cooperation.

Theorem 5 (SD-BC Capacity Region): The capacity region $\mathcal{C}_{\mathrm{BC}}$ of the SD-BC with one-sided decoder cooperation is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \leq H(Y_1) \qquad (13a)$$
$$R_2 \leq I(V, U; Y_2) + R_{12} \qquad (13b)$$
$$R_1 + R_2 \leq H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1) \qquad (13c)$$
$$R_1 + R_2 \leq H(Y_1|V, U) + I(V, U; Y_2) + R_{12} \qquad (13d)$$

where the union is over all PMFs $P_{V,U,Y_1,X} Q_{Y_2|X}$ for which $Y_1 = f(X)$. Moreover, $\mathcal{C}_{\mathrm{BC}}$ is convex and one may choose $|\mathcal{V}| \leq |\mathcal{X}| + 3$ and $|\mathcal{U}| \leq |\mathcal{X}|$.

The proof of Theorem 5 is relegated to Appendix B. The achievable scheme combines Marton and superposition coding with rate-splitting and binning. The rather simple converse proof is due to the telescoping identity [34, eqs. (9) and (11)].

Remark 6: The derivation of the capacity region in Theorem 5 strongly relies on the SD nature of the channel. Since $Y_1 = f(X)$, the encoder has full control over the message that is conveyed via the cooperation link. This allows one to design the cooperation protocol at the encoding stage without assuming a particular Markov relation on the coding random variables. Our approach differs from the one taken in [29], where an inner bound on the capacity region of a BC with two-sided conferencing links between the decoders was derived. In [29], the decoders cooperate by conveying to each other a compressed versions of their received channel outputs (via a WZ-like coding mechanism). Doing so forced the authors to restrict their coding PMF to satisfy certain Markov relations that must not hold in general. Consequently, the inner bound in [29] is not tight for the SD-BC considered here.

Remark 7: The SD-BC with cooperation is strongly related to the SD-RBC that was studied in [35]. The SD-BC with cooperation is operationally equivalent to a reduced version of the SD-RBC, in which the relay channel is orthogonal and deterministic. Section V gives a detailed discussion on the relation between the two problems.

Remark 8: The cardinality bounds on the auxiliary random variables in Theorem 5 are established using the perturbation method [44] and a standard application of the Eggleston-Fenchel-Carathéodory theorem. The details are omitted.

Remark 9: The SD-BC with decoder cooperation and the WAK problem with encoder cooperation are duals. A full discussion on the duality between the problems is given in the following section.

IV. CHANNEL AND SOURCE DUALITY

We examine the WAK coordination problem with encoder cooperation (Fig. 1) and the SD-BC with decoder cooperation (Fig. 2) from a duality perspective. We show that the two problems and their solutions are dual to one another in a manner that naturally extends PTP duality [20]–[22]. In the PTP scenario, two lossy source (or equivalently, source coordination) and channel coding problems are said to be dual if interchanging the roles of the encoder and the decoder in one problem produces the other problem. The solutions of such problems are dual in that they require an optimization of an information measure of the same structure, up to renaming the random variables involved. Solving one problem provides insight into the solution of the other. However, how duality extends to the multiuser case is still obscure.

In the context of multiuser lossy source coding, we favor the framework of source coordination over rate-distortion, since the former provides a natural perspective on the similarities of the two problems. Source coordination inherently accounts for the probabilistic relations among *all* the sequences involved in the problem's definition. However, in a coordination problem, both the input and output (coordination) PMFs are fixed, while in a channel coding problem, the input PMF is optimized. Therefore, for convenience, throughout this section we consider channel codes with codewords of fixed composition, as defined in the following (see also [15]).

Definition 10 (Fixed-Type Codes, Achievability and Capacity): An $(n, R_{12}, R_1, R_2, Q_X^\star)$ fixed-type code $\mathcal{C}_n^\star$ for the SD-BC with one-sided decoder cooperation consists of three integer sets, an encoding function, a decoder cooperation function, and two decoding functions as defined in (8)*.*

For any $\delta > 0$, the average error probability $e_\delta(Q_X^\star, \mathcal{C}_n^\star)$ of an $(n, R_{12}, R_1, R_2, Q_X^\star)$ fixed-type code $\mathcal{C}_n^\star$ is defined

TABLE I
DUALITY TRANSFORMATION PRINCIPLES: THE WAK PROBLEM WITH COOPERATION VS. THE SD-BC WITH COOPERATION

WAK Problem with Encoder Cooperation	SD-BC with Decoder Cooperation
Decoder inputs / Encoder outputs: $t_j \in \left[1 : 2^{nR_j}\right], \; j = 1, 2, 12$	Encoder inputs / Decoder outputs: $m_j \in \left[1 : 2^{nR_j}\right], \; j = 1, 2, 12$
Encoder inputs / Sources: $\mathbf{X}_1, \; \mathbf{X}_2$	Decoder inputs / Channel outputs: $\mathbf{Y}_1, \; \mathbf{Y}_2$
Decoder output / Coordination sequence: $\mathbf{Y}$	Encoder output / Channel input: $\mathbf{X}$
Encoding functions: $f_1 : \mathcal{X}_1^n \to \mathcal{T}_1,$ $f_2 : \mathcal{X}_2^n \times \mathcal{T}_{12} \to \mathcal{T}_2$	Decoding functions: $\psi_1 : \mathcal{Y}_1^n \to \mathcal{M}_1,$ $\psi_2 : \mathcal{Y}_2^n \times \mathcal{M}_{12} \to \mathcal{M}_2$
Encoder cooperation function: $f_{12} : \mathcal{X}_1^n \to \mathcal{T}_{12}$	Decoder cooperation function: $g_{12} : \mathcal{Y}_1^n \to \mathcal{M}_{12}$
Decoding function: $\phi : \mathcal{T}_1 \times \mathcal{T}_2 \to \mathcal{Y}^n$	Encoding function: $g : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}^n$

in (14), *as shown at the bottom of this page, where* $\hat{M}_1 = \psi_1(\mathbf{Y}_1)$ *and* $\hat{M}_2 = \psi_2(\mathbf{Y}_2, g_{12}(\mathbf{Y}_1))$.

A rate triple (R_{12}, R_1, R_2) *is achievable if for any* $\epsilon, \delta > 0$, *there is a sufficiently large* $n \in \mathbb{N}$ *and an* $(n, R_{12}, R_1, R_2, Q_X^\star)$ *fixed-type code* $\mathcal{C}_n^\star$ *such that* $e_\delta(Q_X^\star, \mathcal{C}_n^\star) \leq \epsilon$. *The definition of the capacity region is standard (see, e.g., [45]).*

Note that for fixed-composition codes [46]–[49] and for codes that are drawn in an i.i.d. manner according to $Q_X^\star$, the TV distance in (14) is arbitrarily small with high probability. Moreover, the capacity region of the SD-BC with cooperation and a fixed-type code is similar to that stated in Theorem 5. The only difference between the regions is the domain of PMFs over which the union is taken. Specifically, for the BC with a fixed-type code, the union is taken over all PMFs $P_{V,U,Y_1} P_{X|V,U,Y_1} Q_{Y_2|X}$ that have $Q_X^\star \mathbb{1}_{\{Y_1 = f(X)\}} Q_{Y_2|X}$ as a marginal.

The WAK and SD-BC problems with cooperation are obtained from each other by interchanging the roles of their encoder(s) and decoder(s) and renaming the random variables involved. A full description of the duality transformation principles is given in Table I. The duality is also evident in that the input and output sequences in both problems are jointly typical with respect to a PMF of the same form. Namely, in the source coding problem, the triple $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ is coordinated with respect to the PMF

$$Q_{X_2} P_{Y|X_2} \mathbb{1}_{\{X_1 = f(Y)\}} = P_Y \mathbb{1}_{\{X_1 = f(Y)\}} P_{X_2|Y}. \qquad (15)$$

The corresponding triple of sequences $(\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2)$ in the channel coding problem are jointly typical with high probability

with respect to the PMF

$$Q_X^\star \mathbb{1}_{\{Y_1 = f(X)\}} Q_{Y_2|X}. \qquad (16)$$

By renaming the random variables according to Table I, the two PMFs in (15) and (16) coincide.

The duality between the two problems extends beyond the correspondence presented above. The coordination-capacity region of the WAK problem (Theorem 1) and the capacity region of the SD-BC (Theorem 5) are also dual to one another. To see this, the following lemma gives an alternative characterization of the capacity region $\mathcal{C}_{BC}$.

Lemma 10 (SD-BC Capacity Alternative Characterization): Let $\mathcal{C}_{BC}^{(D)}$ be the region defined by the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_{12} \geq I(V; Y_1) - I(V; Y_2) \qquad (17a)$$
$$R_1 \leq H(Y_1) \qquad (17b)$$
$$R_2 \leq I(V, U; Y_2) + R_{12} \qquad (17c)$$
$$R_1 + R_2 \leq H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1) \qquad (17d)$$

where the union is over the domain stated in Theorem 5. Then:

$$\mathcal{C}_{BC}^{(D)} = \mathcal{C}_{BC}. \qquad (18)$$

See Appendix D for a proof of Lemma 10 based on bidirectional inclusion arguments.

Remark 11: $\mathcal{C}_{BC}^{(D)}$ *can be established as the capacity region of the SD-BC with cooperation by providing achievability and converse proofs. We refer the reader to [50] for a full description of the achievability scheme. The proof of the converse is given in Appendix E. The converse is established via a novel*

$$e_\delta(Q_X^\star, \mathcal{C}_n^\star) \triangleq \mathbb{P}_{\mathcal{C}_n^\star}\left(\left\{ (\hat{M}_1, \hat{M}_2) \neq (M_1, M_2) \right\} \cup \left\{ \bigcup_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} \left\{ ||\nu_{g(m_1, m_2), \mathbf{Y}_1, \mathbf{Y}_2} - Q_X^\star Q_{Y_1, Y_2|X}||_{TV} \geq \delta \right\} \right\} \right) \qquad (14)$$
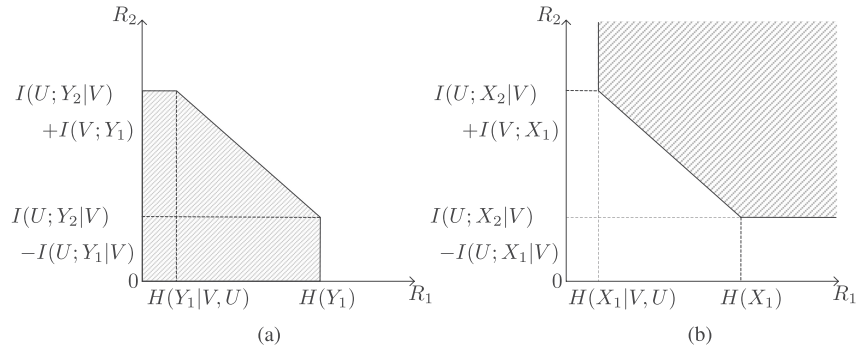
Fig. 4.   Corner point correspondence between: (a) the capacity region of the SD-BC with cooperation; (b) the coordination-capacity region of the WAK coordination problem with cooperation. The regions are depicted at the hyperplanes where $R_{12} = I(V; Y_1) - I(V; Y_2)$ and $R_{12} = I(V; X_1) - I(V; X_2)$, respectively.

approach, in which the auxiliaries are not only chosen as a (possibly different) function of the joint distribution induced by each code, but they are also constructed in a probabilistic manner. The need for this probabilistic construction stems from the unique structure of the region $\mathcal{C}_{\mathrm{BC}}^{(\mathrm{D})}$. Specifically, the lower bound on R_{12} in (17a) (which is typical to source coding problems where the random source sequences are memoryless) and the fact that $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have memory are the underlying reasons for the usefulness of the approach. Depending on the distribution that stems from the code, a deterministic choice of auxiliaries may result in a $I(V; Y_1) - I(V; Y_2)$ that is too large. By a stochastic choice of the auxiliaries, we circumvent this difficulty and dominate the quantity $I(V; Y_1) - I(V; Y_2)$ to satisfy (17a).

The converse proof boils down to two key steps. First, we derive an outer bound on the achievable region $\mathcal{C}_{\mathrm{BC}}^{(\mathrm{D})}$ that is described by three auxiliary random variables (A, B, C). Then, by probabilistically choosing (V, U) from (A, B, C), we show that the outer bound is tight. The second step implies that the outer bound is an alternative formulation of the capacity region. Capacity proofs that rely on alternative descriptions for which the converse is provided have been previously used (see, e.g., [51], [52]). However, the proof of equivalence typically relies on operational arguments rather than on a probabilistic identification of auxiliaries. Probabilistic arguments of a similar nature to those we present here were also used before [38]–[40]. For instance, in [38], such arguments were used to prove the equivalence between two representations of the compress-and-forward inner bound for the relay channel via time-sharing. Such arguments were also leveraged in [39] to characterize the admissible rate-distortion region for the multiterminal source coding problem under logarithmic loss. The novelty of our approach stems from combining these two concepts and essentially using a probabilistic construction to define the auxiliary random variables and establish the tightness of the outer bound. We derive a closed form formula for the optimal probability values, that highlights the dependence of the the auxiliaries on the distribution induced by the code.

The duality between $\mathcal{R}_{\mathrm{WAK}}(P_{X_1, X_2, Y})$ in (10) and $\mathcal{C}_{\mathrm{BC}}^{(\mathrm{D})}$ in (17) is expressed as a correspondence between the

information measures at their corner points. The values of (R_{12}, R_1, R_2) at the corner points of the coordination-capacity region of the WAK problem are

$$\left(I(V; X_1|X_2), H(X_1), I(U; X_2|X_1, V) \right) \tag{19a}$$

$$\left(I(V; X_1|X_2), H(X_1|V, U), I(U; X_2|V) + I(V; X_1) \right) \tag{19b}$$

while the triple (R_{12}, R_1, R_2) at the corner points of capacity region of the SD-BC with cooperation are

$$\left(I(V; Y_1) - I(V; Y_2), H(Y_1), I(U; Y_2|V) - I(U; Y_1|V) \right)$$

$$\left(I(V; Y_1) - I(V; Y_2), H(Y_1|V, U), I(U; Y_2|V) + I(V; Y_1) \right). \tag{20}$$

We show that (19) and (20) correspond by first rewriting the value of R_{12} in (19) as

$$R_{12} = I(V; X_1|X_2) \stackrel{(a)}{=} I(V; X_1) - I(V; X_2) \tag{21}$$

where (a) is due to the Markov relation $V - X_1 - X_2$. Moreover, the value of R_2 in (19a) is rewritten as

$$R_2 = I(U; X_2|X_1, V) \stackrel{(a)}{=} I(U; X_2|V) - I(U; X_1|V) \tag{22}$$

where (a) is since $U - (X_2, V) - X_1$ forms a Markov chain. By substituting (21)-(22) into (19) and renaming the random variables according to Table I, the corner points of both regions coincide (see Fig. 4).

Chronologically, upon observing the duality between the two problem settings, we solved the WAK problem first. Then, based on past experience (cf., e.g., [15], [25]), our focus turned to the dual SD-BC with cooperation. Since the capacity region is defined by the corner points of a union of polytopos, the structure of the capacity region for the SD-BC was evident. Thus, duality was key in obtaining the results of this work. We note that the relation between our result for the SD-BC with cooperation and the SD-RBC (that is discussed in the following section) was observed only at a later stage.

V.   RELATION TO THE SD-RBC

The SD-BC with cooperation is strongly related to the SD-RBC that was studied in [35]. A general RBC is illustrated
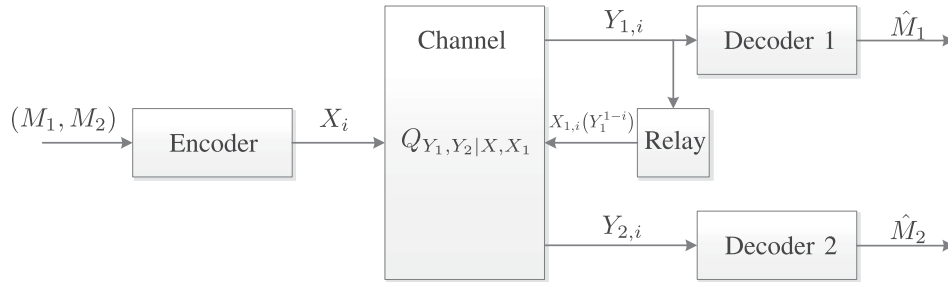
Fig. 5. A general RBC.

in Fig. 5 (for the full definition see [35, Sec. II]). The RBC is SD if the PMF $Q_{Y_1|X,X_1}$ only takes on the values 0 or 1. To see the correspondence between the SD-RBC and the BC of interest, let $Y_2 = (Y_{21}, Y_{22})$ and let the channel transition PMF factorize as

$$Q_{Y_1,Y_{21},Y_{22}|X,X_1} = Q_{Y_{21}|X}\mathbb{1}_{\{Y_1=f(X)\}}Q_{Y_{22}|X_1}. \qquad (23)$$

(23) implies that the channel from the encoder to the decoders is orthogonal to the one between the decoders. Suppose the relay channel is deterministic with capacity R_{12} and let $Y_{22} = f_R(X_1)$. The SD-RBC obtained under these assumptions is referred to as the R_{12}-reduced SD-RBC and its capacity region is denoted by $\mathcal{C}_{RBC}(R_{12})$. As stated in the following lemma, the R_{12}-reduced SD-RBC is operationally equivalent to the SD-BC with cooperation. By operational equivalence, we mean that for every achievable rate tuple in one problem, there exists a code (that achieves these rates) that can be transformed into a code (with the same rates) for the other problem. The transformation mechanism treats the code for each model as a black-box and is described as part of the proof of Lemma 12 given in Appendix G.

Lemma 12 (Operational Equivalence): For every $(R_1, R_2) \in \mathcal{C}_{RBC}(R_{12})$, *there is an* (n, R_1, R_2) *code* $\mathcal{C}_n^{(RBC)}(R_{12})$ *for the R_{12}-reduced SD-RBC that can be transformed into an* (n, R_{12}, R_1, R_2) *code* $\mathcal{C}_n^{(BC)}$ *for the SD-BC with cooperation, and vice versa. Namely, for every* $(R_{12}, R_1, R_2) \in \mathcal{C}_{BC}$, *there is an* (n, R_{12}, R_1, R_2) *code* $\mathcal{C}_n^{(BC)}$ *for the SD-BC with cooperation that can be transformed into an* (n, R_1, R_2) *code* $\mathcal{C}_n^{(RBC)}(R_{12})$ *for the R_{12}-reduced SD-RBC.*

Lemma 12 implies that the capacity regions of the SD-BC with cooperation and the R_{12}-reduced SD-RBC coincide. Using the result of [35, Th. 8], the capacity region $\mathcal{C}_{RBC}(R_{12})$ of the R_{12}-reduced SD-RBC is the union of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying:

$$R_1 \leq H(Y_1|X_1)$$
$$R_2 \leq I(V, U, X_1; Y_{21}) + H(Y_{22}|Y_{21})$$
$$R_1 + R_2 \leq H(Y_1|V, U, X_1) + I(U; Y_{21}|V, X_1) + I(V; Y_1|X_1)$$
$$R_1 + R_2 \leq H(Y_1|V, U, X_1) + I(V, U, X_1; Y_{21}) + H(Y_{22}|Y_{21})$$
$$(24)$$

where the union is over all PMFs P_{V,U,X,X_1} $Q_{Y_{21}|X}\mathbb{1}_{\{Y_1=f(X)\}}\mathbb{1}_{\{Y_{22}=f_R(X_1)\}}$. In Appendix F we simplify

the region in (24) and show that it coincides with the capacity region of the SD-BC with cooperation from Theorem 5.

The advantage of the approach taken in this work compared to that in [35] is threefold. First, we achieve capacity over a *single* transmission block, while the scheme in [35] (which, as a consequence of Lemma 12, can also be used for the SD-BC with cooperation) transmits a large number of blocks and applies backward decoding. The substantial delay introduced by a backward decoding process implies the superiority of our scheme for practical uses. The reduction of the multi-block coding scheme in [35] to our single-block scheme is consistent with the results in [53]. The authors of [53] showed that for the primitive relay channel (i.e., a relay channel with a noiseless link from the relay to the receiver), the decode-and-forward and compress-and-forward multi-block coding schemes can be applied with only a single transmission block. The second advantage of our approach is the simple and concise converse proof that follows using telescoping identities [34, eqs. (9) and (11)]. Finally, focusing on the SD-BC with cooperation (rather than the SD-RBC) highlights the duality with the cooperative WAK source coordination problem (as discussed in Section IV), and gives insight into the relations between multiuser channel and source coding problems.

VI. SPECIAL CASES

We consider special cases of the capacity region of the SD-BC with decoder cooperation and show that the dual relation discussed in Section IV is preserved for each special case.

A. Deterministic BCs With Decoder Cooperation

Corollary 13 (Deterministic BC Capacity Region): The capacity region of a deterministic BC (DBC) is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ *satisfying:*

$$R_1 \leq H(Y_1)$$
$$R_2 \leq H(Y_2) + R_{12}$$
$$R_1 + R_2 \leq H(Y_1, Y_2) \qquad (25)$$

where the union is over all input PMFs P_X.

Proof: Achievability follows from Theorem 5 by taking $V = 0$ and $U = Y_2$. A converse follows by the Cut-Set bound. ∎

The DBC is dual to the SW source coding problem with one-sided encoder cooperation (see [54], [55]). The SW setting

is obtained from the WAK coordination problem by also adding a lossless reproduction requirement to the second source. A proper choice of the auxiliary random variables reduces $\mathcal{R}_{WAK}(P_{X_1,X_2,Y})$ to the optimal rate region for the SW problem, which is the set of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}^3_+$ satisfying:

$$R_1 \geq H(X_1|X_2) - R_{12}$$
$$R_2 \geq H(X_2|X_1)$$
$$R_1 + R_2 \geq H(X_1, X_2) \qquad (26)$$

(see Appendix H for the derivation of (26)). Examining the regions from (25) and (26), reveals the correspondence between their corner points.

B. PD-BCs With Decoder Cooperation

Corollary 14 (PD-BC Capacity Region): The capacity region $\mathcal{C}_{PD}$ for the PD-BC with $Y_1 = X$ coincides with the results in [29] and [40] and is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}^3_+$ satisfying:

$$R_1 \leq H(X|U) \qquad (27a)$$
$$R_2 \leq I(U; Y_2) + R_{12} \qquad (27b)$$
$$R_1 + R_2 \leq H(X) \qquad (27c)$$

where the union is over all PMFs $P_{U,X} Q_{Y_2|X}$.

Proof: The capacity region of the PD-BC was originally derived in [29] where it was described as the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}^3_+$ satisfying:

$$R_1 \leq I(X; Y_1|U)$$
$$R_2 \leq I(U; Y_2) + R_{12}$$
$$R_2 \leq I(U; Y_1) \qquad (28)$$

where the union is over all PMFs $P_{U,X} Q_{Y_1|X} Q_{Y_2|Y_1}$.

An equivalent characterization of region in (28) was later given in [40] as the union over the domain stated above of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}^3_+$ satisfying:

$$R_1 \leq I(X; Y_1|U)$$
$$R_2 \leq I(U; Y_2) + R_{12}$$
$$R_1 + R_2 \leq I(X; Y_1). \qquad (29)$$

Since a SD-BC in which $Y_1 = X$ is also PD, substituting $Y_1 = X$ into (29) yields the region from Corollary 14. By substituting $Y_1 = X$, setting $U = 0$, and relabeling V as U in the capacity of the SD-BC with cooperation stated in Theorem 5, we obtain an achievable region given by the union over the domain stated in Corollary 14 of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}^3_+$ satisfying:

$$R_2 \leq I(U; Y_2) + R_{12} \qquad (30a)$$
$$R_1 + R_2 \leq H(X). \qquad (30b)$$

Denote the region in (30) by $\mathcal{R}_{SD}$. Since $\mathcal{R}_{SD}$ is an achievable region, clearly $\mathcal{R}_{SD} \subseteq \mathcal{C}_{PD}$. On the other hand, the opposite inclusion $\mathcal{C}_{PD} \subseteq \mathcal{R}_{SD}$ also holds, because the rate bound (27a) does not appear in $\mathcal{R}_{SD}$, while (27b)-(27c) and the domain over which the union is taken are preserved. ∎

The dual source coding problem for the PD-BC with cooperation where $Y_1 = X$ is a model in which the output

sequence is a lossless reproduction of $\mathbf{X}_1$. The latter setting is a special case of the WAK problem with cooperation, that is obtained by taking f (the coordination function) to be the identity function. The corresponding coordination-capacity region is given by (10) (with a slight modification of the domain over which the union is taken). However, an equivalent coordination-capacity region that is characterized by a single auxiliary random variable has yet to be derived. Since the capacity region of the PD-BC with cooperation where $Y_1 = X$ is described using a single auxiliary (as in (27)), the lack of such a characterization for the region of the dual problem makes the comparison problematic. Nonetheless, recalling that the capacity region of the considered PD-BC is also given by (13) while substituting $Y_1 = X$ emphasizes that the duality holds.

VII. SUMMARY AND CONCLUDING REMARKS

We considered the WAK empirical coordination problem with one-sided encoder cooperation and derived its coordination-capacity region. The capacity-achieving coding scheme combined WZ coding, binning and superposition coding. Furthermore, a SD-BC in which the decoders can cooperate via a one-sided rate-limited link was considered and its capacity region was found. Achievability was established by deriving an inner bound on the capacity region of a general BC that was shown to be tight for the SD scenario. The coding strategy that achieved the inner bound combined rate-splitting, Marton and superposition coding, and binning (used for the cooperation protocol). The converse for the SD case leveraged telescoping identities that resulted in a concise and a simple proof. The relation between the SD-BC with cooperation and the SD-RBC was examined. The two problems were shown to be operationally equivalent under proper assumptions and the correspondence between their capacity regions was established.

The cooperative WAK and SD-BC problems were inspected from a channel-source duality perspective. Transformation principles between the two settings that naturally extend duality relations between PTP models were presented. It was shown that the duality between the WAK and the SD-BC problems induces a duality between their capacities that is expressed in a correspondence between the corner points of the two regions. To this end, the capacity region of the SD-BC was restated as an alternative expression. The converse was based on a novel approach where the construction of the auxiliary random variables is probabilistic and depends on the distribution induced by the code. The probabilistic construction introduced additional optimization parameters (the probability values) that were used to tighten the outer bound to coincide with the alternative achievable region. To conclude the discussion, several special cases of the BC setting and their corresponding capacity regions were inspected.

APPENDIX A
PROOF OF THEOREM 1

A. Achievability

For any $P_{X_1,X_2,Y} \in \mathcal{Q}$, the direct proof is based on a coding scheme that achieves the corner points of $\mathcal{R}_{WAK}(P_{X_1,X_2,Y})$.

The corner points are stated in (19a)-(19b) and illustrated in Fig 3. Fix a PMF $P_{X_1,X_2,Y} \in \mathcal{Q}$, $\epsilon, \delta > 0$ and a PMF $P_{X_1,X_2,V,U,Y} = Q_{X_1,X_2} P_{V|X_1} P_{U|X_2,V} P_{Y|X_1,U,V}$ that has $P_{X_1,X_2,Y}$ as a marginal. Recall that $P_{X_1,X_2,Y}$ factors as $Q_{X_2} P_{Y|X_2} \mathbb{1}_{\{X_1 = f(Y)\}}$ and that it has the source PMF Q_{X_1,X_2} as a marginal.

The error probability analysis of the subsequently described coding scheme follows by standard random coding arguments. Namely, we evaluate the expected error probability over the ensemble of codebooks and use the union bound to account for each error event separately. Being standard, the details are omitted and only the consequent rate bounds required for reliability are stated.

1) Codebook Generation: A codebook $\mathcal{C}_V$ that comprises 2^{nR_V} codewords $\mathbf{v}(i)$, where $i \in [1 : 2^{nR_V}]$, each generated according to P_V^n. The codebook $\mathcal{C}_V$ is randomly partitioned into $2^{nR_{12}}$ bins indexed by $t_{12} \in [1 : 2^{nR_{12}}]$ and denoted by $\mathcal{B}_V(t_{12})$. For every $i \in [1 : 2^{nR_V}]$ a codebook $\mathcal{C}_U(i)$ is generated. Each codebook $\mathcal{C}_U(i)$ is assembled of 2^{nR_U} codewords $\mathbf{u}(i,j)$, $j \in [1 : 2^{nR_U}]$, generated according to $P_{U|V=\mathbf{v}(i)}^n$. Each $\mathcal{C}_U(i)$ codebook is randomly partitioned into $2^{nR_2'}$ bins $\mathcal{B}_U(i, t_2')$, where $t_2 \in [1 : 2^{nR_2'}]$. Moreover, the set $\mathcal{T}_\epsilon^n(Q_{X_1})$ is partitioned into $2^{nR_1'}$ bins $\mathcal{B}_{X_1}(t_1')$, where $t_1' \in [1 : 2^{nR_1'}]$. To achieve (19a), consider the following scheme:

2) Encoding at Encoder 1: Upon receiving $\mathbf{x}_1$, Encoder 1 searches a pair of indices $(i, t_1') \in [1 : 2^{nR_V}] \times [1 : 2^{nR_1'}]$ such that $(\mathbf{x}_1, \mathbf{v}(i)) \in \mathcal{T}_\epsilon^n(P_{X_1,V})$ and $\mathbf{x}_1 \in \mathcal{B}_{X_1}(t_1')$. A concatenation of i and t_1' is conveyed to the decoder. The bin index of $\mathbf{v}(i)$, i.e., the index $t_{12} \in [1 : 2^{nR_{12}}]$ such that $\mathbf{v}(i) \in \mathcal{B}_V(t_{12})$, is conveyed to Encoder 2 via the cooperation link. Taking

$$R_V > I(V; X_1) \tag{31}$$

ensures that such a codeword $\mathbf{v}(i)$ is found with high probability.

3) Decoding at Encoder 2: Given the source sequence $\mathbf{x}_2$ and the bin index t_{12}, Encoder 2 searches for an index $\hat{i} \in [1 : 2^{nR_V}]$ such that $\mathbf{v}(\hat{i}) \in \mathcal{B}_V(t_{12})$ and $(\mathbf{x}_2, \mathbf{v}(\hat{i})) \in \mathcal{T}_\epsilon^n(P_{X_2,V})$. Reliable decoding follows by taking

$$R_V - R_{12} < I(V; X_2). \tag{32}$$

4) Encoding at Encoder 2: After decoding $\mathbf{v}(\hat{i})$, Encoder 2 searches for an index $j \in [1 : 2^{nR_U}]$, such that $\mathbf{u}(\hat{i}, j) \in \mathcal{C}_U(\hat{i})$ and $(\mathbf{x}_2, \mathbf{v}(\hat{i}), \mathbf{u}(\hat{i}, j)) \in \mathcal{T}_\epsilon^n(P_{X_2,V,U})$. The bin number of the chosen $\mathbf{u}(\hat{i}, j)$, that is, the index $t_2' \in [1 : 2^{nR_2'}]$ such that $\mathbf{u}(\hat{i}, j) \in \mathcal{B}_U(\hat{i}, t_2')$, is conveyed to the decoder. If

$$R_U > I(U; X_2|V) \tag{33}$$

then a codeword $\mathbf{u}(\hat{i}, j)$ as needed is found with high probability.

5) Decoding and Output Generation: Upon receiving (i, t_1') from Encoder 1 and t_2' from Encoder 2, the decoder first identifies the codeword $\mathbf{v}(i) \in \mathcal{C}_V$ that is associated with i. Then it searches the bin $\mathcal{B}_{X_1}(t_1')$ for a sequence $\hat{\mathbf{x}}_1$ such that

$(\mathbf{v}(i), \hat{\mathbf{x}}_1) \in \mathcal{T}_\epsilon^n(P_{X_1,V})$. A reliable lossless reconstruction of $\mathbf{x}_1$ follows provided that

$$R_1' > H(X_1|V). \tag{34}$$

Given $(\mathbf{v}(i), \hat{\mathbf{x}}_1)$, the decoder searches for an index $\hat{j} \in [1 : 2^{nR_U}]$, such that $\mathbf{u}(i, \hat{j}) \in \mathcal{B}_U(i, t_2')$ and $(\hat{\mathbf{x}}_1, \mathbf{v}(i), \mathbf{u}(i, \hat{j})) \in \mathcal{T}_\epsilon^n(P_{X_1,V,U})$. To ensure error-free decoding, we take

$$R_U - R_2' < I(U; X_1|V). \tag{35}$$

Finally, an output sequence $\mathbf{y}$ is generated according to $P_{Y|X_1=\hat{\mathbf{x}}_1, U=\mathbf{u}(i,\hat{j}), V=\mathbf{v}(i)}^n$. The structure of the joint PMF implies that the output sequence admits the desired coordination constraint.

By taking $(R_1, R_2) = (R_1' + R_V, R_2')$ and applying the Fourier-Motzkin elimination (FME) on (31)-(35), we obtain the rate bounds

$$R_{12} > I(V; X_1) - I(V; X_2) = I(V; X_1|X_2)$$
$$R_1 > H(X_1|V) + I(V; X_1) = H(X_1)$$
$$R_2 > I(U; X_2|V) - I(U; X_1|V) = I(U; X_2|X_1, V) \tag{36}$$

which imply that (19a) is achievable.

To establish the achievability of (19b) requires no binning of the codebooks $\mathcal{C}_U(i)$, where $i \in [1 : 2^{nR_V}]$.

6) Encoding at Encoder 1: Given $\mathbf{x}_1$, Encoder 1 finds $\mathbf{v}(i) \in \mathcal{C}_V$ in a similar manner and conveys its bin index t_{12} to Encoder 2. Moreover, it conveys the bin index of the received $\mathbf{x}_1$, say t_1', to the decoder. Again, by having (31), such a codeword $\mathbf{v}(i)$ is found with high probability.

7) Decoding at Encoder 2: Performed in a similar manner as before. We again take (32) to ensure reliable decoding of $\mathbf{v}(i)$. As before, the decoded codeword is denoted by $\mathbf{v}(\hat{i})$.

8) Encoding at Encoder 2: Encoder 2 finds a codeword $\mathbf{u}(\hat{i}, j) \in \mathcal{C}_U(\hat{i})$ in a manner similar to that presented in the previous scheme. Now, however, it sends to the decoder a concatenation of $\hat{i}$ and j. This decoding process has a vanishing probability of error if (33) holds.

9) Decoding and Output Generation: Upon receiving t_1' and $(\hat{i}, j)$ from Encoder 1 and 2, respectively, the decoder first finds the $\mathbf{v}(\hat{i}) \in \mathcal{C}_V$ that is associated with $\hat{i}$ and the $\mathbf{u}(\hat{i}, j) \in \mathcal{C}_U(\hat{i})$ that is associated with $(\hat{i}, j)$. Given $(\mathbf{v}(\hat{i}), \mathbf{u}(\hat{i}, j))$, it searches the bin $\mathcal{B}_{X_1}(t_1')$ for a sequence $\hat{\mathbf{x}}_1$ such that $(\hat{\mathbf{x}}_1, \mathbf{v}(\hat{i}), \mathbf{u}(\hat{i}, j)) \in \mathcal{T}_\epsilon^n(P_{X_1,V,U})$. A reliable lossless reconstruction of $\mathbf{x}_1$ is ensured provided

$$R_1' > H(X_1|V, U). \tag{37}$$

Finally, an output sequence $\mathbf{y}$ is generated in the same manner as in the coding scheme for (19a).

Taking $(R_1, R_2) = (R_1', R_V + R_U)$ and applying FME on (31)-(33) and (37) yields the following bounds:

$$R_{12} > I(V; X_1) - I(V; X_2) = I(V; X_1|X_2)$$
$$R_1 > H(X_1|V, U)$$
$$R_2 > I(V; X_1) + I(U; X_2|V). \tag{38}$$

This concludes the proof of achievability for (19b).

$$P_{\mathbf{X}_1,\mathbf{X}_2,T_{12},T_1,T_2,\mathbf{Y}}(\mathbf{x}_1,\mathbf{x}_2,t_{12},t_1,t_2,\mathbf{y}) = Q^n_{X_1,X_2}(\mathbf{x}_1,\mathbf{x}_2)\mathbb{1}_{\{t_{12}=f_{12}(\mathbf{x}_1)\}\cap\{t_1=f_1(\mathbf{x}_1)\}\cap\{t_2=f_2(\mathbf{x}_2,t_{12})\}\cap\{\mathbf{y}=\phi(t_1,t_2)\}} \tag{39}$$

B. Converse-Fixed

We show that given an achievable rate triple (R_{12}, R_1, R_2), there exists a PMF $P_{X_1,X_2,V,U,Y} = Q_{X_1,X_2}P_{V|X_1}P_{U|X_2,V}P_{Y|X_1,U,V}$ that has $Q_{X_2}P_{Y|X_2}\mathbb{1}_{\{X_1=f(Y)\}}$ as a marginal, such that the inequalities in (13) are satisfied. Fix an achievable tuple (R_{12}, R_1, R_2) and $\delta, \epsilon > 0$, and let $\mathcal{L}_n$ be the corresponding coordination code for some sufficiently large $n \in \mathbb{N}$. The joint distribution on $\mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{T}_{12} \times \mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{Y}^n$ induced by $\mathcal{L}_n$ is given in (39) at the top of the page. All subsequent multi-letter information measures are calculated with respect to $P_{\mathbf{X}_1,\mathbf{X}_2,T_{12},T_1,T_2,\mathbf{Y}}$ or its marginals.

Since (R_{12}, R_1, R_2) is achievable, X_1^n can be reconstructed at the decoder with a small probability of error. By Fano's inequality we have

$$H(X_1^n|T_1,T_2) \le (1 + \epsilon n R) \triangleq n\epsilon_n \tag{40}$$

where $\epsilon_n = \frac{1}{n} + \epsilon R$.

Next, by the structure of the single-letter PMF $P_{X_1,X_2,V,U,Y}$, we rewrite the mutual information measure in (10c) as

$$R_2 \ge I(U; X_2|X_1, V) \overset{(a)}{=} I(V; X_2|X_1) + I(U; X_2|X_1, V)$$
$$= I(V, U; X_2|X_1). \tag{41}$$

where (a) is because $V - X_1 - X_2$ forms a Markov chain.

For the lower bound on R_{12}, consider

$$nR_{12} \ge H(T_{12})$$
$$\overset{(a)}{\ge} I(T_{12}; X_1^n|X_2^n)$$
$$= \sum_{i=1}^n I(T_{12}; X_{1,i}|X_{1,i+1}^n, X_2^{n\backslash i}, X_{2,i})$$
$$\overset{(b)}{=} \sum_{i=1}^n I(T_{12}, X_{1,i+1}^n, X_2^{n\backslash i}; X_{1,i}|X_{2,i})$$
$$\ge \sum_{i=1}^n I(T_{12}, X_{1,i+1}^n, X_2^{i-1}; X_{1,i}|X_{2,i})$$
$$\overset{(c)}{=} \sum_{i=1}^n I(V_i; X_{1,i}|X_{2,i}) \tag{42}$$

where (a) is because T_{12} is determined by X_1^n and since conditioning cannot increase entropy, (b) is since (X_1^n, X_2^n) are pairwise i.i.d., and (c) defines $V_i \triangleq (T_{12}, X_{1,i+1}^n, X_2^{i-1})$, for every $i \in [1:n]$.

Next, for R_1 we have

$$nR_1 \ge H(T_1)$$
$$\ge H(T_1|T_{12}, T_2) \tag{43}$$
$$\overset{(a)}{=} I(T_1; X_1^n|T_{12}, T_2)$$
$$= H(X_1^n|T_{12}, T_2) - H(X_1^n|T_{12}, T_1, T_2)$$

$$\overset{(b)}{\ge} \sum_{i=1}^n H(X_{1,i}|T_{12}, T_2, X_{1,i+1}^n) - n\epsilon_n$$
$$\ge \sum_{i=1}^n H(X_{1,i}|T_{12}, T_2, X_{1,i+1}^n, X_2^{i-1}) - n\epsilon_n$$
$$\overset{(c)}{=} \sum_{i=1}^n H(X_{1,i}|V_i, U_i) - n\epsilon_n \tag{44}$$

where (a) is because T_1 is determined by X_1^n, (b) uses (40) and the mutual information chain rule, while in (c) we define $U_i \triangleq T_2$, for every $i \in [1:n]$, and use the definition of V_i.

To bound R_2 consider

$$nR_2 \ge H(T_2)$$
$$\ge H(T_2|X_1^n)$$
$$\ge I(T_2; X_2^n|X_1^n)$$
$$= \sum_{i=1}^n I(T_2; X_{2,i}|X_1^{n\backslash i}, X_2^{i-1}, X_{1,i})$$
$$\overset{(a)}{=} \sum_{i=1}^n I(T_2, X_1^{n\backslash i}, X_2^{i-1}; X_{2,i}|X_{1,i})$$
$$\overset{(b)}{=} \sum_{i=1}^n I(T_2, T_{12}, X_1^{n\backslash i}, X_2^{i-1}; X_{2,i}|X_{1,i})$$
$$\overset{(c)}{\ge} \sum_{i=1}^n I(V_i, U_i; X_{2,i}|X_{1,i}) \tag{45}$$

where (a) is because (X_1^n, X_2^n) are pairwise i.i.d., (b) is because T_{12} is determined by X_1^n, while (c) follows since conditioning cannot increase entropy and from the definitions of V_i and U_i.

For the sum of rates, we have

$$n(R_1 + R_2)$$
$$\ge H(T_1, T_2)$$
$$\overset{(a)}{=} I(T_1, T_2; X_1^n, X_2^n)$$
$$\overset{(b)}{=} H(X_1^n) - H(X_1^n|T_1, T_2) + I(T_{12}, T_2; X_2^n|X_1^n)$$
$$\overset{(c)}{\ge} \sum_{i=1}^n \left[H(X_{1,i}) + I(T_{12}, T_2, X_1^{n\backslash i}, X_2^{i-1}; X_{2,i}|X_{1,i}) \right] - n\epsilon_n$$
$$\ge \sum_{i=1}^n \left[H(X_{1,i}) + I(T_{12}, T_2, X_{1,i+1}^n, X_2^{i-1}; X_{2,i}|X_{1,i}) \right] - n\epsilon_n$$
$$\overset{(d)}{=} \sum_{i=1}^n \left[H(X_{1,i}|V_i, U_i) + I(V_i, U_i; X_{1,i}, X_{2,i}) \right] - n\epsilon_n \tag{46}$$

where:
(a) is because (T_1, T_2) are determined by (X_1^n, X_2^n);
(b) is since X_1^n defines (T_{12}, T_1);
(c) uses (40), the mutual information chain rule and the pairwise i.i.d. nature of (X_1^n, X_2^n);

(b) uses the mutual information chain rule and the definition of (V_i, U_i).

The upper bounds in (42), (44), (45) and (46) are rewritten by introducing a time-sharing random variable T that is independent of $(X_1^n, X_2^n, T_{12}, T_1, T_2, Y^n)$ and is uniformly distributed over $[1:n]$. The rate bound on R_{12} is rewritten as

$$R_{12} \geq \frac{1}{n} \sum_{t=1}^{n} I(V_t; X_{1,t} | X_{2,t}, T = t) \qquad (47)$$

$$= \sum_{t=1}^{n} \mathbb{P}(T = t) I(V_t; X_{1,t} | X_{t,q}, T = t) \qquad (48)$$

$$= I(V_T; X_{1,T} | X_{2,T}, T) \qquad (49)$$

$$\overset{(a)}{=} I(V_T, T; X_{1,T} | X_{2,T}) \qquad (50)$$

where (a) follows because T is independent of the pair $(X_{1,T}, X_{2,T})$ (see property 1 in [13, Section IIV-B]). By rewriting (44), (45) and (46) in an analogous manner, the region obtained is convex. This follows from the presence of the time-sharing random variable T in the conditioning of all the mutual information and entropy terms.

Next, define $X_1 \triangleq X_{1,T}$, $X_2 \triangleq X_{2,T}$, $V \triangleq (V_T, T)$, $U \triangleq U_T$ and $Y \triangleq Y_T$. Notice that $(X_1, X_2) \sim Q_{X_1, X_2}$ and then use the time-mixing property from [13, Section IV-B, Property 2] to get

$$R_{12} \geq I(V; X_1 | X_2)$$
$$R_1 \geq H(X_1 | V, U) - \epsilon_n$$
$$R_2 \geq I(V, U; X_2 | X_1)$$
$$R_1 + R_2 \geq H(X_1 | V, U) + I(V, U; X_1, X_2) - \epsilon_n. \qquad (51)$$

To complete the converse, the following Markov relations must be shown to hold.

$$V - X_1 - X_2 \qquad (52a)$$
$$U - (X_2, V) - X_1 \qquad (52b)$$
$$Y - (X_1, U, V) - X_2. \qquad (52c)$$

We prove that the Markov relations in (52) hold for every $t \in [1:n]$. Upon doing so, showing that the relations hold in their single-letter (as stated in (52)) is straightforward.

For (52a), recall that $V_t = (T_{12}, X_{1,t+1}^n, X_2^{t-1})$, for every $t \in [1:n]$, and consider

$$0 \leq I(T_{12}, X_{1,t+1}^n, X_2^{t-1}; X_{2,t} | X_{1,t})$$
$$\overset{(a)}{\leq} I(X_1^{n \setminus t}, X_2^{t-1}; X_{2,t} | X_{1,t}) \overset{(b)}{=} 0$$

where (a) is because conditioning cannot increase entropy and sicen T_{12} is determined by X_1^n, while (b) uses the pairwise i.i.d. nature of (X_1^n, X_2^n). Thus (52a) holds.

To establish (52b), we use Lemma 1 in [56]. Since $U_t = T_2$, for every $t \in [1:n]$, we have

$$0 \leq I(T_2; X_{1,t} | X_{2,t}, T_{12}, X_{1,t+1}^n, X_2^{t-1})$$
$$\leq I(T_2; X_{1,t}, X_1^{t-1} | X_{2,t}, T_{12}, X_{1,t+1}^n, X_2^{t-1}). \qquad (53)$$

Set

$$A_1 = X_{1,t+1}^n, \quad A_2 = (X_{1,t}, X_1^{t-1}),$$
$$B_1 = X_{2,t+1}^n, \quad B_2 = (X_{2,t}, X_2^{t-1}).$$

Accordingly, (53) is rewritten as

$$0 \leq I(T_2; A_2 | T_{12}, A_1, B_2). \qquad (54)$$

Noting that (A_1, A_2) and (T_{12}, B_1, B_2) determine T_{12} and T_2, respectively, and that $P_{A_1, A_2, B_1, B_2} = P_{A_1, B_1} P_{A_2, B_2}$. The result of [56, Lemma 1, Conclusion 2] thus implies

$$0 \leq I(T_2; A_2 | T_{12}, A_1, B_2) = 0 \qquad (55)$$

which establishes (52b).

For (52c) note the the structure of the joint PMF from (39) implies that for any $i \in [1:n]$, the marginal distribution of $(X_1^n, X_2^n, T_2, T_{12}, Y^n)$ factors as:

$$P(x_1^n, x_2^n, t_2, t_{12}, y^n)$$
$$= P(x_2^{i-1}) P(x_{1,i}, x_{2,i}) P(x_{1,i+1}^n, x_{2,i+1}^n)$$
$$\times \mathbb{1}_{\{t_2 = f_2(x_2^n, t_{12})\}} P(x_1^{i-1}, t_{12}, y^n | x_2^{i-1}, x_{1,i}^n, t_2). \qquad (56)$$

Consequently, by further marginalizing over X_1^{i-1}, we get

$$P(x_{1,i}^n, x_2^n, t_2, t_{12}, y^n)$$
$$= P(x_2^{i-1}) P(x_{1,i}, x_{2,i}) P(x_{1,i+1}^n, x_{2,i+1}^n) \mathbb{1}_{\{t_2 = f_2(x_2^n, t_{12})\}}$$
$$\times P(t_{12} | x_2^{i-1}, x_{1,i}^n, t_2) P(y^n | x_2^{i-1}, x_{1,i}, x_{1,i+1}^n, t_2, t_{12}). \qquad (57)$$

The structure of the conditional distribution of Y^n given $(X_{1,i}^n, X_2^n, T_2, T_{12})$ implies that

$$Y^n - (T_2, T_{12}, X_{1,i+1}^n, X_2^{i-1}, X_{1,i}) - (X_{2,i}, X_{2,i+1}^n) \qquad (58)$$

forms a Markov chain, and in particular we have

$$Y_i - (T_2, T_{12}, X_{1,i+1}^n, X_2^{i-1}, X_{1,i}) - X_{2,i} \qquad (59)$$

for every $i \in [1:n]$. Taking $\delta, \epsilon \to 0$ and $n \to \infty$ concludes the converse.

APPENDIX B
PROOF OF THEOREM 5

A. Achievability

To establish achievability, we show that for any fixed $\epsilon > 0$, a PMF

$$P_{V, U, Y_1} P_{X | V, U, Y_1} Q_{Y_2 | X} \qquad (60)$$

for which $Y_1 = f(X)$, and a rate triple (R_{12}, R_1, R_2) that satisfies (13), there is a sufficiently large $n \in \mathbb{N}$ and a corresponding (n, R_{12}, R_1, R_2) code $\mathcal{C}_n$, such that $e(\mathcal{C}_n) \leq \epsilon$. We first derive an achievable region for a general BC with a one-sided conferencing link between the decoders with a channel transition matrix $Q_{Y_1, Y_2 | X}$. The region is described using three auxiliaries (rather than two). Then, by a proper choice of the auxiliaries, we achieve $\mathcal{C}_{BC}$. Fix a PMF

$$P_{V, U_1, U_2, X, Y_1, Y_2} = P_{V, U_1, U_2, X} Q_{Y_1, Y_2 | X} \qquad (61)$$

and an $\epsilon > 0$, and consider the following coding scheme.

1) Codebook Generation: Split each message m_j, $j = 1, 2$, into two sub-messages denoted by (m_{j0}, m_{jj}). The pair $m_0 \triangleq (m_{10}, m_{20})$ is referred to as a *public message* while m_{jj} serve as *private message* j. The rates associated with m_{j0} and m_{jj}, $j = 1, 2$, are denoted by R_{j0} and R_{jj}, while the corresponding alphabets are $\mathcal{M}_{j0}$ and $\mathcal{M}_{jj}$, respectively. Accordingly, we have

$$R_j = R_{j0} + R_{jj}, \quad j = 1, 2. \tag{62}$$

We also denote $R_0 \triangleq R_{10} + R_{20}$ and $\mathcal{M}_0 \triangleq \mathcal{M}_{10} \times \mathcal{M}_{20}$. The random variables M_0 and M_{jj}, for $j = 1, 2$, are associated with the public message and private message j, respectively. Furthermore, M_0, M_{11} and M_{22} are independent and uniform over $\mathcal{M}_0$, $\mathcal{M}_{11}$ and $\mathcal{M}_{22}$, respectively.

Partition $\mathcal{M}_0$ into $2^{nR_{12}}$ equal-sized bins $\mathcal{B}(m_{12})$, where $m_{12} \in \mathcal{M}_{12}$. Generate a public message codebook, denoted by $\mathcal{C}_V$, that comprises 2^{nR_0} v-codewords $\mathbf{v}(m_0)$, $(m_0) \in \mathcal{M}_0$, each drawn according to P_V^n independent of all the other v-codewords.

For each $\mathbf{v}(m_0) \in \mathcal{C}_V$, generate two codebooks $C_{U_j}(m_0)$, $j = 1, 2$, each comprises $2^{n(R_{jj}+R_j')}$ codewords $\mathbf{u}_j$ that are independently drawn according to $P_{U_j|V=\mathbf{v}(m_0)}^n$. The u_j-codewords in $C_{U_j}(m_0)$ are labeled as $\mathbf{u}_j(m_0, m_{jj}, i_j)$, where $(m_{jj}, i_j) \in \mathcal{M}_{jj} \times \mathcal{I}_j$ and $\mathcal{I}_j = [1 : 2^{nR_j'}]$. Based on this labeling, the codebook $C_{U_j}(m_0)$ has a u_j-bin associated with every $m_{jj} \in \mathcal{M}_{jj}$, each containing $2^{nR_j'}$ u_j-codewords.

2) Encoding: To transmit the message pair $(m_1, m_2) = \big((m_{10}, m_{11}), (m_{20}, m_{22})\big)$, the encoder searches for a pair $(i_1, i_2) \in \mathcal{I}_1 \times \mathcal{I}_2$ that satisfies (63), as shown at the bottom of this page, where $\mathbf{v}(m_0) \in \mathcal{C}_V$ and $\mathbf{u}_j(m_0, m_j, i_j) \in \mathcal{C}_{U_j}(m_0)$, for $j = 1, 2$. If the set of appropriate index pairs contains more than one element, the encoder chooses the component-wise minimal pair; if the set is empty, the encoder sets $(i_1, i_2) = (1, 1)$. The channel input sequence $\mathbf{x}$ is then randomly generated according to $P_{X|V,U_1,U_2}^n$ and is transmitted over the channel.

3) Decoding and Cooperation (Decoder 1): Searches for a unique pair $(\hat{m}_0, \hat{m}_{11}) \in \mathcal{M}_0 \times \mathcal{M}_{11}$ for which there is an index $\hat{i}_1 \in \mathcal{I}_1$, such that

$$\Big(\mathbf{v}(\hat{m}_0), \mathbf{u}_1(\hat{m}_0, \hat{m}_{11}, \hat{i}_1), \mathbf{y}_1\Big) \in \mathcal{T}_\epsilon^n (P_{V,U_1,Y_1}) \tag{64}$$

where $\mathbf{v}(\hat{m}_0) \in \mathcal{C}_V$ and $\mathbf{u}_1(\hat{m}_0, \hat{m}_{11}, \hat{i}_1) \in \mathcal{C}_{U_1}(\hat{m}_0)$. If such a unique triple is found, then $\hat{m}_1 = (\hat{m}_0, \hat{m}_{11})$ is declared as the decoded message; otherwise, an error is declared.

Cooperation: Given $(\hat{m}_0, \hat{m}_{11}, \hat{i}_1)$, Decoder 1 conveys the bin number of $\hat{m}_0$ to Decoder 2 via the cooperation link. Namely, Decoder 1 shares with Decoder 2 the index $\hat{m}_{12} \in \mathcal{M}_{12}$, such that $\hat{m}_0 \in \mathcal{B}(\hat{m}_{12})$.

Decoder 2: Upon receiving $\hat{m}_{12}$ from Decoder 1 and $\mathbf{y}_2$ from the channel, Decoder 2 searches for a unique pair $(\hat{\hat{m}}_0, \hat{\hat{m}}_{22}) \in \mathcal{M}_0 \times \mathcal{M}_{22}$ for which there is an $\hat{\hat{i}}_2 \in \mathcal{I}_2$, such

that

$$\Big(\mathbf{v}(\hat{\hat{m}}_0), \mathbf{u}_2(\hat{\hat{m}}_0, \hat{\hat{m}}_{22}, \hat{\hat{i}}_2), \mathbf{y}_2\Big) \in \mathcal{T}_\epsilon^n (P_{V,U_2,Y_2}) \tag{65}$$

where $\hat{\hat{m}}_0 \in \mathcal{B}(\hat{m}_{12})$, $(\hat{\hat{m}}_0) \in \mathcal{C}_V$ and $\mathbf{u}_2(\hat{\hat{m}}_0, \hat{\hat{m}}_{22}, \hat{\hat{i}}_2) \in \mathcal{C}_{U_2}(\hat{\hat{m}}_0)$. If such a unique triple is found, then $\hat{\hat{m}}_2 \triangleq (\hat{\hat{m}}_{20}, \hat{\hat{m}}_{22})$ is declared as the decoded message; otherwise, an error is declared.

By standard error probability analysis (see Appendix C) and existence arguments, an (n, R_{12}, R_1, R_2) code $\mathcal{C}_n$ that achieves reliability is extracted provided that

$$R_1' + R_2' > I(U_1; U_2|V)$$
$$R_{11} + R_1' < I(U_1; Y_1|V)$$
$$R_{20} + R_1 + R_1' < I(V, U_1; Y_1)$$
$$R_{22} + R_2' < I(U_2; Y_2|V)$$
$$R_{10} + R_2 + R_2' - R_{12} < I(V, U_2; Y_2). \tag{66}$$

Applying FME on (66) while using (62) yields the rate bounds

$$R_1 < I(V, U_1; Y_1)$$
$$R_2 < I(V, U_2; Y_2) + R_{12}$$
$$R_1 + R_2 < I(V, U_1; Y_1) + I(U_2; Y_2|V) - I(U_1; U_2|V)$$
$$R_1 + R_2 < I(U_1; Y_1|V) + I(V, U_2; Y_2) - I(U_1; U_2|V) + R_{12}. \tag{67}$$

By setting $U_1 = Y_1$ and $U_2 = U$, the bounds in (67) reduce to (13). Note that this choice of auxiliaries is valid as they satisfy the Markov relations stated in Theorem 5. This shows that $\mathcal{C}_{BC}$ is achievable.

Remark 15: The cooperation protocol described in the proof is reminiscent of the WZ coding technique. The cooperation link is used to convey a bin *of the common message codeword* $\mathbf{v}$ *(rather than the codeword itself) from 1st decoder to the 2nd. As part of the joint typicality decoding rule in (65), the channel input* $\mathbf{y}_2$ *is used as correlated side information to isolate the actual* v*-codeword from the bin. This correlation is induced from the channel transition probability and the underlying Markov relations (with respect to the PMF in (61)).*

B. Converse

We show that if a rate triple (R_{12}, R_1, R_2) is achievable, then there exists a PMF $P_{V,U,Y_1,X} Q_{Y_2|X}$ for which $Y_1 = f(X)$, such that the inequalities in (13) are satisfied. Fix an achievable tuple (R_{12}, R_1, R_2) and an $\epsilon > 0$, and let $\mathcal{C}_n$ be the corresponding (n, R_{12}, R_1, R_2) code for some sufficiently large $n \in \mathbb{N}$. The joint distribution on $\mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{X}^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n \times \mathcal{M}_{12} \times \mathcal{M}_1 \times \mathcal{M}_2$ induced by $\mathcal{C}_n$ is given in (68), as shown at the top of the next page. All subsequent multi-letter information measures are calculated with respect to the PMF from (68) or its marginals.

$$\Big(\mathbf{v}(m_0), \mathbf{u}_1(m_0, m_1, i_1), \mathbf{u}_2(m_0, m_{22}, i_2)\Big) \in \mathcal{T}_\epsilon^n (P_{V,U_1,U_2}). \tag{63}$$

$$P_{M_1,M_2,\mathbf{X},\mathbf{Y}_1,\mathbf{Y}_2,M_{12},\hat{M}_1,\hat{M}_2}(m_1, m_2, \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, m_{12}, \hat{m}_1, \hat{m}_2) = 2^{-n(R_1+R_2)} \mathbb{1}_{\left\{\mathbf{x}=g(m_1,m_2)\right\} \cap \left\{\cap_{i=1}^{n}\left(y_{1,i}=f(x_i)\right)\right\}} Q_{Y_2|X}^n(\mathbf{y}_2|\mathbf{x})$$
$$\times \mathbb{1}_{\left\{m_{12}=g_{12}(\mathbf{y}_1)\right\} \cap \left\{\hat{m}_1=\psi_1(\mathbf{y}_1)\right\} \cap \left\{\hat{m}_2=\psi_2(\mathbf{y}_2,m_{12})\right\}} \tag{68}$$

Since $e(\mathcal{C}_n) \le \epsilon$, Fano's inequality gives

$$H(M_1|Y_1^n) \le 1 + \epsilon n R_1 \triangleq n\epsilon_n^{(1)} \tag{69a}$$
$$H(M_2|M_{12}, Y_2^n) \le 1 + \epsilon n R_2 \triangleq n\epsilon_n^{(2)} \tag{69b}$$

where $\epsilon_n^{(j)} \triangleq \frac{1}{n} + \epsilon R_j$, for $j = 1, 2$. Define

$$\epsilon_n = \max\left\{\epsilon_n^{(1)}, \epsilon_n^{(2)}\right\}. \tag{69c}$$

It follows that

$$\begin{aligned}
nR_1 &= H(M_1) \\
&\overset{(a)}{\le} I(M_1; Y_1^n) + n\epsilon_n \\
&\overset{(b)}{=} I(X^n; Y_1^n) + n\epsilon_n \\
&\overset{(c)}{\le} \sum_{i=1}^{n} H(Y_{1,i}) + n\epsilon_n
\end{aligned} \tag{70}$$

where (a) uses (69), (b) is by the Markov chain $M_1 - X^n - Y_1^n$ and the Data Processing Inequality, while (c) follows because Y_1^n is a function of X^n and since conditioning cannot increase entropy.

To bound R_2 consider

$$\begin{aligned}
nR_2 &= H(M_2) \tag{71} \\
&\overset{(a)}{\le} I(M_2; M_{12}, Y_2^n) + n\epsilon_n \\
&= I(M_2; Y_2^n|M_{12}) + I(M_2; M_{12}) + n\epsilon_n \tag{72} \\
&\overset{(b)}{\le} \sum_{i=1}^{n} I(M_2; Y_{2,i}|M_{12}, Y_{2,i+1}^n) + nR_{12} + n\epsilon_n \\
&\overset{(c)}{\le} \sum_{i=1}^{n} I(V_i, U_i; Y_{2,i}) + nR_{12} + n\epsilon_n \tag{73}
\end{aligned}$$

where (a) uses (69), (b) is because a uniform distribution maximizes entropy, while (c) defines $V_i \triangleq (M_{12}, Y_1^{i-1}, Y_{2,i+1}^n)$ and $U_i \triangleq M_2$, for every $i \in [1:n]$.

For the sum of rates, we first write

$$n(R_1 + R_2) = H(M_1, M_2) = H(M_2) + H(M_1|M_2). \tag{74}$$

By the independence of M_1 and M_2 and by (69), we have

$$H(M_1|M_2) \le H(Y_1^n|M_2) + n\epsilon_n. \tag{75}$$

Moreover, we bound $H(M_2)$ as

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

$$\begin{aligned}
&\quad - I(M_2; Y_{1,i}, Y_{2,i+1}^n|M_{12}, Y_1^{i-1}) \\
&\quad + I(M_2; Y_{1,i}|M_{12}, Y_1^{i-1})\Big] + I(M_2; M_{12}) + n\epsilon_n \\
&= \sum_{i=1}^{n}\Big[I(U_i; Y_{2,i}|V_i) \\
&\quad - I(M_2; Y_{1,i}|M_{12}, Y_1^{i-1}, Y_{2,i+1}^n)\Big] \\
&\quad + I(M_2; M_{12}, Y_1^n) + n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n}\Big[I(U_i; Y_{2,i}|V_i) - I(U_i; Y_{1,i}|V_i)\Big] \\
&\quad + I(M_2; Y_1^n) + n\epsilon_n
\end{aligned} \tag{76}$$

where:
(a) is by repeating steps (71)-(72) in the upper bounding of R_2;
(b) uses a telescoping identity [34, eq. (9) and (11)];
(c) uses the definitions of V_i and U_i;
(d) again uses the definition of V_i and U_i (second term) and the Markov relation $M_{12} - Y_1^n - M_2$ (third term).

Inserting (75) and (76) into (74) results in

$$\begin{aligned}
&n(R_1 + R_2) \\
&\le \sum_{i=1}^{n}\Big[I(U_i; Y_{2,i}|V_i) - I(U_i; Y_{1,i}|V_i)\Big] + H(Y_1^n) + 2n\epsilon_n \tag{77} \\
&\le \sum_{i=1}^{n}\Big[H(Y_{1,i}|V_i, U_i) + I(U_i; Y_{2,i}|V_i) + I(V_i; Y_{1,i})\Big] + 2n\epsilon_n. \tag{78}
\end{aligned}$$

Finally, note that

$$\begin{aligned}
&H(Y_1^n) - \sum_{i=1}^{n} H(Y_{1,i}|V_i) \\
&\overset{(a)}{=} \sum_{i=1}^{n} I(Y_{2,i+1}^n; Y_{1,i}|M_{12}, Y_1^{i-1}) + I(M_{12}; Y_1^n) \\
&\overset{(b)}{\le} \sum_{i=1}^{n} I(Y_1^{i-1}; Y_{2,i}|M_{12}, Y_{2,i+1}^n) + H(M_{12}) \\
&\overset{(c)}{\le} \sum_{i=1}^{n} I(V_i; Y_{2,i}) + nR_{12} \tag{79}
\end{aligned}$$

where (a) is the mutual information chain rule and the definition of V_i, (b) is the Csiszár sum identity, and (c) is because conditioning cannot increase entropy and since a uniform distribution maximizes it.

By plugging (79) into (77), we obtain

$$n(R_1 + R_2) \leq \sum_{i=1}^{n} \left[H(Y_{1,i}|V_i, U_i) + I(V_i, U_i; Y_{2,i}) \right] + nR_{12} + 2n\epsilon_n. \tag{80}$$

The upper bounds in (70), (73), (78) and (80) can be rewritten by introducing a time-sharing random variable T that is independent of $(M_1, M_2, X^n, Y_1^n, Y_2^n, M_{12})$ and is uniformly distributed over $[1 : n]$. For instance, the bound in (73) is rewritten as

$$R_2 \leq \frac{1}{n} \sum_{t=1}^{n} I(V_t, U_t; Y_{2,t}) + R_{12} + \epsilon_n$$

$$= \sum_{t=1}^{n} \mathbb{P}(T = t) I(V_t, U_t; Y_{2,t}|T = t) + R_{12} + \epsilon_n$$

$$= I(V_T, U_T; Y_{2,T}|T) + R_{12} + \epsilon_n$$

$$\leq I(T, V_T, U_T; Y_{2,T}) + R_{12} + \epsilon_n. \tag{81}$$

By rewriting the rate bounds (70), (78) and (80) in a similar manner, the region obtained is convex. Next, let $X \triangleq X_T$, $Y_1 \triangleq Y_{1,T}$, $Y_2 \triangleq Y_{2,T}$, $V \triangleq (V_T, T)$ and $U \triangleq U_T$. We have

$$R_1 \leq H(Y_1) + \epsilon_n$$

$$R_2 \leq I(V, U; Y_2) + R_{12} + \epsilon_n$$

$$R_1 + R_2 \leq H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1) + 2\epsilon_n$$

$$R_1 + R_2 \leq H(Y_1|V, U) + I(V, U; Y_2) + R_{12} + 2\epsilon_n. \tag{82}$$

To complete the proof we need to show that the PMF of (V, U, X, Y_1, Y_2) factors as $P_{V,U,Y_1,X} Q_{Y_2|X}$, which boils down to the Markov relation

$$(V, U, Y_1) - X - Y_2. \tag{83}$$

The proof of (83) is given in Appendix I. Taking $\epsilon \to 0$ and $n \to \infty$ establishes the converse.

APPENDIX C
ERROR PROBABILITY ANALYSIS FOR THEOREM 5

Recall that (M_0, M_{11}, M_{22}) is a triple of random variables that represents the transmitted messages. Since the analysis considers the expected error probability over the ensemble of codebooks, by the symmetry of the codebook construction we may assume that $(M_0, M_{11}, M_{22}) = \mathbf{1} \triangleq (1, 1, 1)$. With some abuse of notation, we denote by i_j the index chosen by the encoder from the u_j-bin that is associated with the transmitted messages (recall that for a fixed codebook i_1 and i_2 are deterministically defined by the transmitted messages).

A. Encoding Error

An encoding error occurs if the v-, u_1- and u_2-codewords chosen by the encoder are not jointly typical. This is described by the event stated in (84), as shown at the bottom of the page, where

$$\left(\mathbf{V}(1), \mathbf{U}_1(1, 1, \hat{i}_1), \mathbf{U}_2(1, 1, \hat{i}_2)\right) \sim P_V^n P_{U_1|V}^n P_{U_2|V}^n$$

and (i_1, i_2) are chosen according to the encoding rule from Subsection B in Appendix B. Namely, an encoding error occurs if there is no pair of indices $(\hat{i}_1, \hat{i}_2) \in \mathcal{I}_1 \times \mathcal{I}_2$ that satisfies (63). By the Multivariate Covering Lemma [57, Lemma 8.2], $\mathbb{P}(\mathcal{E}) \to 0$ as $n \to \infty$ if we have

$$R_1' + R_2' > I(U_1; U_2|V). \tag{85}$$

B. Decoding Errors

To account for decoding errors, for any $(m_0, m_{jj}, \hat{i}_j) \in \mathcal{M}_0 \times \mathcal{M}_{jj} \times \mathcal{I}_j$ and $j = 1, 2$, define the following event

$$\mathcal{D}_j(m_0, m_{jj}, \hat{i}_j)$$
$$= \left\{ \left(\mathbf{V}(m_0), \mathbf{U}_j(m_0, m_{jj}, \hat{i}_j), \mathbf{Y}_j\right) \in \mathcal{T}_\epsilon^n(P_{V,U_j,Y_j}) \right\} \tag{86}$$

where $\left(\mathbf{V}(m_0), \mathbf{U}_j(m_0, m_{jj}, \hat{i}_j)\right) \sim P_V^n P_{U_j|V}^n$ and $\mathbf{Y}_j$ is distributed according to the channel transition probability conditioned on the input sequence that corresponds to $(m_0, m_{11}, m_{22}) = \mathbf{1}$ and (i_1, i_2).

Let C_n be a random variable that represents a random codebook that adheres to the scheme from Appendix B. By the union bound, the average error probability over the ensemble of codebooks is bounded as shown in (87) at the top of the next page. Note that $\{P_j^{[k]}\}_{k=1}^4$ correspond to decoding errors by Decoder j, where $j = 1, 2$. We proceed with the following steps:

1) $P_j^{[1]}$, for $j = 1, 2$, vanishes to 0 as $n \to \infty$ by the law of large numbers.
2) To upper bound $P_j^{[2]}$, $j = 1, 2$, consider:

$$P_j^{[2]} \overset{(a)}{\leq} \sum_{\tilde{i}_j, \tilde{m}_{jj} \neq 1} 2^{-n\left(I(U_j; Y_j|V) - \delta_j^{[2]}(\epsilon)\right)}$$

$$\leq 2^{n(R_{jj} + R_j')} 2^{-n\left(I(U_j; Y_j|V) - \delta_j^{[2]}(\epsilon)\right)}$$

$$= 2^{n\left(R_{jj} + R_j' - I(U_j; Y_j|V) + \delta_j^{[2]}(\epsilon)\right)}$$

where (a) follows since for every $\tilde{m}_{jj} \neq 1$ and $\tilde{i}_j \in \mathcal{I}_j$, $\mathbf{U}_j(1, \tilde{m}_{jj}, \tilde{i}_j)$ is independent of $\mathbf{Y}_j$ while both of them are drawn conditioned on $\mathbf{V}(1)$. Moreover, $\delta_j^{[2]}(\epsilon) \to 0$ as $\epsilon \to 0$. Hence, to ensure that $P_j^{[2]}$ vanishes as $n \to \infty$, we take:

$$R_{jj} + R_j' < I(U_j; Y_j|V) - \delta_j^{[2]}(\epsilon), \quad j = 1, 2. \tag{88}$$

$$\mathcal{E} = \bigcap_{(\hat{i}_1, \hat{i}_2) \in \mathcal{I}_1 \times \mathcal{I}_2} \left\{ \left(\mathbf{V}(1), \mathbf{U}_1(1, 1, \hat{i}_1), \mathbf{U}_2(1, 1, \hat{i}_2)\right) \notin \mathcal{T}_\epsilon^n(P_{V,U_1,U_2}) \right\} \tag{84}$$

$$\mathbb{E}e(\mathcal{C}_n) \leq \mathbb{P}(\mathcal{E}) + \left(1 - \mathbb{P}(\mathcal{E})\right)\left(\sum_{j=1}^{2}\left[\underbrace{\mathbb{P}\left(\mathcal{D}_j^c(1,1,i_j)\big|\mathcal{E}^c\right)}_{P_j^{[1]}} + \underbrace{\mathbb{P}\left(\bigcup_{\tilde{i}_j,\tilde{m}_{jj}\neq 1}\mathcal{D}_j(1,\tilde{m}_{jj},\tilde{i}_j)\Big|\mathcal{E}^c\right)}_{P_j^{[2]}}\right]\right.$$

$$+ \underbrace{\mathbb{P}\left(\bigcup_{\tilde{m}_0\neq 1}\mathcal{D}_1(\tilde{m}_0,1,i_1)\Big|\mathcal{E}^c\right)}_{P_1^{[3]}} + \underbrace{\mathbb{P}\left(\bigcup_{\substack{\tilde{i}_1,\tilde{m}_0\neq 1,\\\tilde{m}_{11}\neq 1}}\mathcal{D}_1(\tilde{m}_0,\tilde{m}_{11},\tilde{i})\Big|\mathcal{E}^c\right)}_{P_1^{[4]}} + \underbrace{\mathbb{P}\left(\bigcup_{\substack{\tilde{m}_0\neq 1:\\\tilde{m}_0\in\mathcal{B}(m_{12})}}\mathcal{D}_2(\tilde{m}_0,1,i_2)\Big|\mathcal{E}^c\right)}_{P_2^{[3]}}$$

$$\left.+ \underbrace{\mathbb{P}\left(\bigcup_{\substack{\tilde{i}_2,\tilde{m}_0\neq 1,\tilde{m}_{22}\neq 1:\\\tilde{m}_0\in\mathcal{B}(m_{12})}}\mathcal{D}_2(\tilde{m}_0,\tilde{m}_{22},\tilde{i}_2)\Big|\mathcal{E}^c\right)}_{P_2^{[4]}}\right) \tag{87}$$

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

$$P_1^{[4]} \overset{(a)}{\leq} \sum_{\substack{\tilde{i}_1,\tilde{m}_0\neq 1,\\\tilde{m}_{11}\neq 1}} 2^{-n\left(I(V,U_1;Y_1)-\delta_1^{[4]}(\epsilon)\right)}$$

$$\overset{(b)}{\leq} 2^{n(R_{20}+R_1+R_1')}\cdot 2^{-n\left(I(V,U_1;Y_1)-\delta_1^{[4]}(\epsilon)\right)}$$

$$= 2^{n\left(R_{20}+R_1+R_1'-I(V,U_1;Y_1)+\delta_1^{[4]}(\epsilon)\right)}$$

where (a) follows since for every $(\tilde{m}_0,\tilde{m}_{11}) \neq \mathbf{1}$ and $\tilde{i}_1 \in \mathcal{I}_1$, $\mathbf{V}(\tilde{m}_0)$ and $\mathbf{U}_1(\tilde{m}_0,\tilde{m}_{11},\tilde{i}_1)$ are drawn together by independent of $\mathbf{Y}_1$, while (b) uses $R_0 = R_{10} + R_{20}$. Again, $\delta_1^{[4]}(\epsilon) \to 0$ as $\epsilon \to 0$, and therefore, we have that $P_1^{[4]} \to 0$ as $n \to \infty$ if

$$R_{20} + R_1 + R_1' < I(V,U_1;Y_1) - \delta_1^{[4]}(\epsilon). \tag{89}$$

4) By repeating similar arguments as before while keeping in mind that the search space of m_0 at Decoder 2 is of size $2^{n(R_0-R_{12})}$ (as a consequence of the binning of $\mathcal{M}_0$ and the cooperation protocol), we have that $P_2^{[4]}$ decays with n provided that

$$R_{10} + R_2 + R_2' - R_{12} < I(V,U_2;Y_2) - \delta_2^{[4]}(\epsilon) \tag{90}$$

where $\delta_2^{[4]}(\epsilon) \to 0$ as $\epsilon \to 0$.

5) By repeating similar steps to upper bound $P_1^{[3]}$, the obtained rate bound is redundant. This is since for every $\tilde{m}_0 \neq 1$ and $\tilde{i}_1 \in \mathcal{I}_1$, the sequences $\mathbf{V}(\tilde{m}_0)$ and $\mathbf{U}_1(\tilde{m}_0,1,\tilde{i}_1)$ are independent of $\mathbf{Y}_1$. Hence, to ensure that $P_1^{[3]}$ vanishes to 0 as $n \to \infty$, we take

$$R_{10} + R_{20} < I(V,U_1;Y_1) - \delta_1^{[3]}(\epsilon) \tag{91}$$

where $\delta_1^{[3]}(\epsilon) \to 0$ as $\epsilon \to 0$. But the right-hand side (RHS) of (91) coincides with the RHS of (89), while the left-hand side (LHS) is with respect to $R_{10} + R_{20}$ only. Clearly, (89) is the dominating constraint.

In a similar manner one finds that the rate bound that ensures that $P_2^{[3]}$ can be made arbitrarily small with n is redundant (due to (90)).

Summarizing the above results, we get that the RHS of (87) decays as the blocklength $n \to \infty$ if the conditions in (66) are met. By standard existence arguments, a vanishing expected average error probability (over the ensemble of codes) ensures that there exists a reliable (n, R_{12}, R_1, R_2) code $\mathcal{C}_n$ for all rate triples that satisfy (66).

APPENDIX D
PROOF OF LEMMA 10

To show $\mathcal{C}_{BC}^{(D)} \subseteq \mathcal{C}_{BC}$, let $(R_{12}, R_1, R_2) \in \mathcal{C}_{BC}^{(D)}$ be a rate triple achieved by (V, U, X). Setting $V^\star = V$ and $U^\star = U$, implies that the same rate triple (R_{12}, R_1, R_2) is contained in $\mathcal{C}_{BC}$, as it is achieved by $(V^\star, U^\star, X)$ (since substituting (17a) into (17d) yields (13d)).

To see that $\mathcal{C}_{BC} \subseteq \mathcal{C}_{BC}^{(D)}$, let $(R_{12}, R_1, R_2) \in \mathcal{C}_{BC}$ be a rate triple achieved by (V, U, X). Further assume that

$$R_{12} < I(V;Y_1) - I(V;Y_2) \tag{92}$$

(otherwise, all four inequalities in (17) clearly hold). Accordingly, there is a real number $\gamma > 0$, such that

$$R_{12} = I(V;Y_1) - I(V;Y_2) - \gamma. \tag{93}$$

Define $V^\star \triangleq (\Theta, \widetilde{V})$, where $\Theta \sim \text{Ber}(\lambda)$, $\lambda \in [0,1]$, is a binary random variable independent of (V, U, X) that takes values in $\mathcal{O} = \{\theta_1, \theta_2\}$, and

$$\widetilde{V} = \begin{cases} V, & \Theta = \theta_1 \\ \emptyset, & \Theta = \theta_2. \end{cases} \tag{94}$$

Furthermore, set

$$\lambda = \frac{I(V;Y_1) - I(V;Y_2) - \gamma}{I(V;Y_1) - I(V;Y_2)} \tag{95}$$

and $U^\star = (V, U)$.

With respect to this choice of $(V^\star, U^\star)$, consider

$$
\begin{aligned}
I(V^\star; Y_1) - I(V^\star; Y_2) &= \lambda\Big[I(V; Y_1) - I(V; Y_2) \Big] \\
&\stackrel{(a)}{=} I(V; Y_1) - I(V; Y_2) - \gamma \\
&\stackrel{(b)}{=} R_{12} \qquad\qquad\qquad (96)
\end{aligned}
$$

where (a) uses the choice of λ in (95) and (b) follows from (93). Thus, (17a) holds.

Next, by the definition of $U^\star$ and because (13a)-(13b) are valid, we obtain (17b)-(17c). It remains to be shown that (17d) holds. Consider the following:

$$
\begin{aligned}
&H(Y_1|V^\star, U^\star) + I(U^\star; Y_2|V^\star) + I(V^\star; Y_1) \\
&= H(Y_1|V^\star, U^\star) + I(V^\star, U^\star; Y_2) + I(V^\star; Y_1) - I(V^\star; Y_2) \\
&\stackrel{(a)}{=} H(Y_1|V, U) + I(V, U; Y_2) + R_{12} \\
&\stackrel{(b)}{\geq} R_1 + R_2 \qquad\qquad\qquad (97)
\end{aligned}
$$

where (a) uses the definition of $U^\star$ and (96), while (b) is by (13d). Consequently (17d) is valid and the inclusion $\mathcal{C}_{\text{BC}} \subseteq \mathcal{C}_{\text{BC}}^{(D)}$ follows.

APPENDIX E
EXPLICIT CONVERSE FOR LEMMA 10

The converse for Theorem 5 is established using a novel approach that generalizes the classical technique used for converse proofs. Our approach relies on two key properties. First, the construction of the auxiliary random variables depends on the distribution induced by the code. Second, the auxiliaries are constructed in a probabilistic manner.

We show that if a rate triple (R_{12}, R_1, R_2) is achievable, then there is a PMF $P_{V,U,Y_1,X} Q_{Y_2|X}$ for which $Y_1 = f(X)$, such that the inequalities in (17) are satisfied. To do so, we first state an upper bound on $\mathcal{C}_{\text{BC}}^{(D)}$ and then establish its inclusion in $\mathcal{C}_{\text{BC}}^{(D)}$. The upper bound is stated in the following lemma.

Lemma 16 (Upper Bound on the Capacity Region): Let $\mathcal{R}_O$ be the region defined by the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$
\begin{aligned}
R_{12} &\geq I(A; Y_1|C) - I(C; Y_2|A) & (98a) \\
R_1 &\leq H(Y_1|B, C) & (98b) \\
R_2 &\leq I(B; Y_2|A) + R_{12} & (98c) \\
R_1 + R_2 &\leq H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_1|C) & (98d)
\end{aligned}
$$

where the union is over all PMFs $P_{A,B,C,Y_1,X} Q_{Y_2|X}$ for which $Y_1 = f(X)$. The following inclusion holds:

$$
\mathcal{C}_{\text{BC}}^{(D)} \subseteq \mathcal{R}_O. \qquad\qquad (99)
$$

Proof: By similar arguments to those given in Subsection B of Appendix B, since (R_{12}, R_1, R_2) is achievable and by Fano's inequality, we have

$$
\begin{aligned}
H(M_1|Y_1^n) &\leq n\epsilon_n & (100a) \\
H(M_2|M_{12}, Y_2^n) &\leq n\epsilon_n & (100b)
\end{aligned}
$$

where ϵ_n is defined as in (69). It follows that

$$
\begin{aligned}
nR_{12} &\geq H(M_{12}) \\
&\stackrel{(a)}{=} I(M_{12}; Y_1^n) \\
&\stackrel{(b)}{=} \sum_{i=1}^{n} \Big[I(M_{12}, Y_{2,i+1}^n; Y_1^i) - I(M_{12}, Y_{2,i}^n; Y_1^{i-1}) \Big] \\
&= \sum_{i=1}^{n} \Big[I(M_{12}, Y_{2,i+1}^n; Y_{1,i}|Y_1^{i-1}) \\
&\qquad\quad - I(Y_1^{i-1}; Y_{2,i}|M_{12}, Y_{2,i+1}^n) \Big] \\
&\stackrel{(c)}{=} \sum_{i=1}^{n} \Big[I(A_i; Y_{1,i}|C_i) - I(C_i; Y_{2,i}|A_i) \Big] \qquad (101)
\end{aligned}
$$

where (a) is because M_{12} is defined by Y_1^n, (b) is a telescoping identity, while (d) is by defining $A_i \triangleq (M_{12}, Y_{2,i+1}^n)$ and $C_i \triangleq Y_1^{i-1}$, for every $i \in [1:n]$.

For the upper bound on R_1, consider

$$
\begin{aligned}
nR_1 &= H(M_1) \\
&= H(M_1|M_2) \\
&\stackrel{(a)}{\leq} I(M_1; Y_1^n|M_2) + n\epsilon_n \\
&\stackrel{(b)}{=} H(Y_1^n|M_2) - H(Y_1^n|M_1, M_2, X^n) + n\epsilon_n \\
&\stackrel{(c)}{=} \sum_{i=1}^{n} H(Y_{1,i}|M_2, Y_1^{i-1}) + n\epsilon_n \\
&\stackrel{(d)}{=} \sum_{i=1}^{n} H(Y_{1,i}|B_i, C_i) + n\epsilon_n \qquad (102)
\end{aligned}
$$

where (a) uses (100a), (b) is since X^n is a function of (M_1, M_2), (c) is because Y_1^n is determined by X^n, while (d) defines $B_i \triangleq M_2$, for every $i \in [1:n]$, and uses the definition of C_i.

To bound R_2 we have

$$
\begin{aligned}
nR_2 &\stackrel{(a)}{\leq} I(M_2; Y_2^n|M_{12}) + I(M_2; M_{12}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(M_2; Y_{2,i}|M_{12}, Y_{2,i+1}^n) + H(M_{12}) + n\epsilon_n \\
&\stackrel{(b)}{\leq} \sum_{i=1}^{n} I(B_i; Y_{2,i}|A_i) + nR_{12} + n\epsilon_n \qquad (103)
\end{aligned}
$$

where (a) is by repeating steps (71)-(72) in Appendix B, while (b) is by the definition of (A_i, B_i) and because a uniform distribution maximizes entropy.

Finally, for the sum of rates, we begin from step (77) in Appendix B and note that the auxiliaries in Appendix B can be rewritten in terms of (A_i, B_i, C_i) as $V_i = (A_i, C_i)$ and $U_i = B_i$. We thus have

$$
\begin{aligned}
n(R_1 + R_2) &\leq \sum_{i=1}^{n} \Big[I(B_i; Y_{2,i}|A_i, C_i) - I(B_i; Y_{1,i}|A_i, C_i) \Big] \\
&\qquad + H(Y_1^n) + 2n\epsilon_n \\
&\stackrel{(a)}{=} \sum_{i=1}^{n} \Big[H(Y_{1,i}|A_i, B_i, C_i) + I(B_i; Y_{2,i}|A_i, C_i) \\
&\qquad + I(A_i; Y_{1,i}|C_i) \Big] + 2n\epsilon_n \qquad (104)
\end{aligned}
$$

where (a) is from the mutual information chain rule and the definition of (A_i, B_i, C_i).

By standard time-sharing arguments, we rewrite the bounds in (101)-(104) as

$$R_{12} \geq I(A; Y_1|C) - I(C; Y_2|A)$$
$$R_1 \leq H(Y_1|B, C) + \epsilon_n$$
$$R_2 \leq I(B; Y_2|A) + R_{12} + \epsilon_n$$
$$R_1 + R_2 \leq H(Y_1|A, B, C) + I(B; Y_2|A, C)$$
$$+ I(A; Y_1|C) + 2\epsilon_n \quad (105)$$

which are the bounds from (98) with small added terms such as ϵ_n. Taking $\epsilon \to 0$ and $n \to \infty$, these terms approach 0. The proof is completed by showing that the Markov relations stated in Lemma 16 hold. This follows by arguments similar to those presented in Appendix B. ∎

Based on Lemma 16, the inclusion relation stated in the following lemma completes the proof of the converse.

Lemma 17 (Tightness of Upper Bound): The following inclusion holds:

$$\mathcal{R}_O \subseteq \mathcal{C}_{BC}^{(D)}. \quad (106)$$

Proof: Let $(R_{12}, R_1, R_2) \in \mathcal{R}_O$ be achieved by a given tuple of random variables (A, B, C, X). We show that there exists a pair of random variables (V, U), such that $(R_{12}, R_1, R_2) \in \mathcal{C}_{BC}^{(D)}$ and is achieved by (V, U, X). We define (V, U) as follows. Let $\Theta \sim \mathrm{Ber}(\lambda)$, $\lambda \in [0, 1]$, be a binary random variable independent of (A, B, C, X) that takes values in $\mathcal{O} = \{\theta_1, \theta_2\}$. Define the random variable

$$\widetilde{V} = \begin{cases} (A, C), & \Theta = \theta_1 \\ \emptyset, & \Theta = \theta_2. \end{cases} \quad (107)$$

Set $V \triangleq (\Theta, \widetilde{V})$ and

$$U = (A, B, C) \quad (108)$$

and note that (V, U) preserves the Markov structure

$$(Y_1, Y_2) - X - (U, V) \quad (109)$$

since, as stated in Lemma 16, $(Y_1, Y_2) - X - (A, B, C)$ forms a Markov chain.

First, consider the case when

$$I(A, C; Y_1) - I(A, C; Y_2) \leq 0. \quad (110)$$

By setting $\lambda = 1$ we have

$$I(V; Y_1) - I(V; Y_2) \leq 0 \overset{(a)}{\leq} R_{12} \quad (111)$$

where (a) is since $R_{12} \geq 0$, which establishes (17a). (17b) holds since $H(Y_1|B, C) \leq H(Y_1)$.

For (17c), note that the definition of (V, U) in (107)-(108) implies that

$$(A, B, C, X, Y_1, Y_2) - U - V \quad (112)$$

forms a Markov chain. Consequently, we obtain

$$I(V, U; Y_2) = I(A, B, C; Y_2) \quad (113)$$

which yields

$$I(V, U; Y_2) + R_{12} \geq I(B; Y_2|A) + R_{12} \overset{(a)}{\geq} R_2 \quad (114)$$

where (a) uses (98c). This shows that (17c) also holds.

For the sum rate, we rewrite (17d) as

$$H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1)$$
$$= H(Y_1|V, U) + I(V, U; Y_2) + I(V; Y_1) - I(V; Y_2) \quad (115)$$

and obtain an explicit expression for each of the information measures in the RHS of (115) in terms of (A, B, C, X). Based on similar arguments to those presented before, we have

$$H(Y_1|V, U) = H(Y_1|A, B, C) \quad (116)$$

while the other two information measures in (115) were previously evaluated in (111) and (113). Inserting (111), (113) and (116) into (115) results in

$$H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1)$$
$$\overset{(a)}{\geq} H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_1|C)$$
$$\overset{(b)}{\geq} R_1 + R_2 \quad (117)$$

where (a) is because $\lambda = 1$ and the mutual information chain rule, while (b) uses (98d). This satisfies (17d).

To conclude the proof it is left to consider the case where

$$I(A, C; Y_1) - I(A, C; Y_2) > 0. \quad (118)$$

This time set

$$\lambda = \min \left\{ 1, \left(\frac{I(A; Y_1|C) - I(A, C; Y_2) + I(A; Y_2)}{I(A; Y_1|C) - I(A, C; Y_2) + I(C; Y_1)} \right)^+ \right\} \quad (119)$$

where $(x)^+ = \max\{0, x\}$, and consider the following.

$$I(V; Y_1) - I(V; Y_2)$$
$$= \lambda \left[I(A, C; Y_1) - I(A, C; Y_2) \right] \quad (120)$$
$$= \lambda \left[I(A; Y_1|C) - I(A, C; Y_2) + I(C; Y_1) \right] \quad (121)$$
$$\overset{(a)}{\leq} I(A; Y_1|C) - I(C; Y_2|A) \overset{(b)}{\leq} R_{12} \quad (122)$$

where (b) relies (98a), while step (a) is justified as follows. If $\lambda = 1$ we have

$$I(A; Y_2) \geq I(C; Y_1). \quad (123)$$

Using (123), we rewrite (121) as

$$\lambda \left[I(A; Y_1|C) - I(A, C; Y_2) + I(C; Y_1) \right]$$
$$\overset{(a)}{=} I(A; Y_1|C) - I(C; Y_2|A) + I(C; Y_1) - I(A; Y_2)$$
$$\overset{(b)}{\leq} I(A; Y_1|C) - I(C; Y_2|A)$$

where (a) follows because $\lambda = 1$ and by the mutual information chain rule, while (b) is by (123). On the other hand, if

$$\lambda = \frac{I(A; Y_1|C) - I(A, C; Y_2) + I(A; Y_2)}{I(A; Y_1|C) - I(A, C; Y_2) + I(C; Y_1)} \quad (124)$$

then

$$I(A; Y_2) < I(C; Y_1) \qquad (125)$$

and we rewrite (121) as

$$\lambda\Big[I(A; Y_1|C) - I(A, C; Y_2) + I(C; Y_1)\Big]$$
$$\overset{(a)}{=} I(A; Y_1|C) - I(A, C; Y_2) + I(A; Y_2)$$
$$= I(A; Y_1|C) - I(C; Y_2|A)$$

where (a) uses (124). The case $\lambda = 0$ is trivial, and we omit the derivation of (a) in (122). We conclude that (17a) is satisfied. (17b)-(17c) follow by the same arguments presented above, while for (17d) we have

$$H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1)$$
$$\overset{(a)}{=} H(Y_1|A, B, C) + I(A, B, C; Y_2)$$
$$+ \lambda\Big[I(A, C; Y_1) - I(A, C; Y_2)\Big]$$
$$\overset{(b)}{\geq} H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_1|C)$$
$$\overset{(c)}{\geq} R_1 + R_2 \qquad (126)$$

where (a) is by (112) and (115), (c) uses (98d), while the derivation of (b) relies on evaluating the terms of interest for the three possible values of λ. First, by (118), $\lambda = 0$ if and only if

$$I(A; Y_1|C) \leq I(C; Y_2|A) \qquad (127)$$

which implies

$$H(Y_1|A, B, C) + I(A, B, C; Y_2)$$
$$+ \lambda\Big[I(A, C; Y_1) - I(A, C; Y_2)\Big]$$
$$= H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A, C; Y_2)$$
$$\geq H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(C; Y_2|A)$$
$$\geq H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_2|C). \qquad (128)$$

If $\lambda = 1$, by the mutual information chain rule we have

$$H(Y_1|A, B, C) + I(A, B, C; Y_2)$$
$$+ \lambda\Big[I(A, C; Y_1) - I(A, C; Y_2)\Big]$$
$$\geq H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_1|C). \qquad (129)$$

Finally, if λ is as in (124), we obtain

$$H(Y_1|A, B, C) + I(A, B, C; Y_2)$$
$$+ \lambda\Big[I(A, C; Y_1) - I(A, C; Y_2)\Big]$$
$$= H(Y_1|A, B, C) + I(A; Y_2) + I(B; Y_2|A, C) + I(A; Y_1|C)$$
$$\geq H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_1|C). \qquad (130)$$

We find that (17d) is also satisfied, thus concluding that (17) holds for the choice of (V, U) and λ stated in (107)-(108) and (119), respectively. This implies that $\mathcal{R}_O \subseteq \mathcal{C}_{BC}^{(D)}$. ∎

Lemma 17 completes the converse and characterizes the region in (17) as the capacity region of the SD-BC with cooperation.

Remark 18: The definition of V in (107) is probabilistic and through λ depends on the joint distribution of (A, B, C, X) that is induced by the code.

APPENDIX F
DERIVATION OF THE REGION IN (13) FROM (24)

Denote the region in (24) by $\mathcal{R}$. Note that $\mathcal{C}_{BC}$ is achievable from $\mathcal{R}$ by taking X_1 to be independent of (V, U, X) and applying a coding scheme where the transmission rate via the relay channel is R_{12}. This implies that $\mathcal{C}_{BC} \subseteq \mathcal{R}$.

To see that $\mathcal{R} \subseteq \mathcal{C}_{BC}$ recall that [35, Proof of Theorem 8 relies on Theorem 4 in that same work, which characterized an upper bound on the capacity region of a general RBC. In the proof of Theorem 4 (see [35, Appendix II]), the auxiliary random variables V_i and U_i are defined as

$$V_i \triangleq (M_0, Y_1^{i-1}, Y_{2,i+1}^n); \quad U_i \triangleq (M_2, Y_1^{i-1}, Y_{2,i+1}^n). \quad (131)$$

M_0 is a common message that was also considered in [35]. Since $X_{1,i}$ is a function of Y_1^{i-1}, it is also a function of V_i (and\or U_i) for every $i \in [1 : n]$. In particular, this implies that X_1 is a function of V. Consequently, the information measures defining $\mathcal{R}$ are then upper bounded as follows. For R_1 we have

$$R_1 \leq H(Y_1|X_1) \leq H(Y_1). \qquad (132)$$

For the R_2 consider

$$R_2 \leq I(V, U, X_1; Y_{21}) + H(Y_{22}|Y_{21})$$
$$\overset{(a)}{\leq} I(V, U; Y_{21}) + H(Y_{22})$$
$$\overset{(b)}{\leq} I(V, U; Y_{21}) + R_{12} \qquad (133)$$

where (a) is because X_1 is defined by V and since conditioning cannot increase entropy, while (b) follows because the relay channel is deterministic with capacity R_{12}.

For the first bound on the sum of rates, we have

$$R_1 + R_2 \leq H(Y_1|V, U, X_1) + I(U; Y_{21}|V, X_1) + I(V; Y_1|X_1)$$
$$\overset{(a)}{\leq} H(Y_1|V, U) + I(U; Y_{21}|V) + I(V; Y_1). \qquad (134)$$

Here (a) is justified similarly to step (a) in (133).

Finally, the second bound on $R_1 + R_2$ is upper bounded as

$$R_1 + R_2 \leq H(Y_1|V, U, X_1) + I(V, U, X_1; Y_{21}) + H(Y_{22}|Y_{21})$$
$$\overset{(a)}{\leq} H(Y_1|V, U) + I(V, U; Y_{21}) + H(Y_{22})$$
$$\overset{(b)}{\leq} H(Y_1|V, U) + I(V, U; Y_{21}) + R_{12}. \qquad (135)$$

Again, (a) and (b) follow by the same arguments as (a) and (b) in (133).

To complete the proof, it remains to be shown that taking the union only over PMFs in which X_1 is independent of (V, U, X) exhausts the entire region. This follows since the rate bounds in (132)-(135) do not involve X_1 nor Y_{22}. Relabeling Y_{21} as Y_2 shows that $\mathcal{R} \subseteq \mathcal{C}_{BC}$ and completes the proof.

APPENDIX G
PROOF OF LEMMA 12

Fix $(R_1, R_2) \in \mathcal{C}_{RBC}(R_{12})$ and let $\{\mathcal{C}_n^{(RBC)}(R_{12})\}_{n \in \mathbb{N}}$ be the sequence of (n, R_1, R_2) codes for the R_{12}-reduced SD-RBC that adhere to the coding scheme described in [35, Appendix I]. Accordingly, $e(\mathcal{C}_n^{(RBC)}(R_{12})) \to 0$ as

$n \to \infty$ and the induced codewords, channel inputs, and channel outputs are jointly-typical with high probability.[3] Since the channel from Decoder 1 to Decoder 2 is deterministic, there are approximately $2^{nH(Y_{22})}$ different possible relay channel outputs $\mathbf{y}_{22}$. Recall that the capacity of the orthogonal and deterministic relay of the R_{12}-reduced SD-RBC is exactly R_{12}, i.e., $H(Y_{22}) = R_{12}$. For every sequence $\mathbf{y}_{22} \in \mathcal{T}_\epsilon^n(P_{Y_{22}})$ (here $\epsilon > 0$ corresponds to the margin between the region achieved by the nth code in the sequence and (R_1, R_2)), define the following subset of $\mathbf{x}_1$ codewords:

$$\mathcal{V}(\mathbf{y}_{22}) = \left\{ \mathbf{x}_1 \in \mathcal{C}_n^{(\text{RBC})}(R_{12}) \middle| f_R(x_{1,i}) = y_{22,i}, \forall i \in [1:n] \right\}. \tag{136}$$

Consider a SD-BC with cooperation and associate a cooperation message m_{12}, where $m_{12} \in \mathcal{M}_{12}$, with every set $\mathcal{V}(\mathbf{y}_{22})$. To use $\mathcal{C}_n^{(\text{RBC})}$ for the SD-BC with cooperation, Decoder 1 waits for the n-symbol transmission to end and then shares with Decoder 2 the message m_{12} associated with a set $\mathcal{V}(\mathbf{y}_{22})$ that contains the intended $\mathbf{x}_1$ codeword (i.e., such that $\mathbf{x}_1 \in \mathcal{V}(\mathbf{y}_{22})$). Given m_{12}, Decoder 2 recovers the sequence $\mathbf{y}_{22}$ and proceeds with the decoding process of the R_{12}-reduced SD-RBC coding scheme. This results in a sequence of (n, R_{12}, R_1, R_2) codes $\{\mathcal{C}_n^{(\text{BC})}\}_{n \in \mathbb{N}}$ for the SD-BC with cooperation.

Next, fix $(R_{12}, R_1, R_2) \in \mathcal{C}_{\text{BC}}$ and let $\{\mathcal{C}_n^{(\text{BC})}\}_{n \in \mathbb{N}}$ be the sequence of (n, R_{12}, R_1, R_2) codes for the SD-BC with cooperation described in Appendix B. Consider an R_{12}-reduced SD-RBC and map each cooperation message $m_{12} \in \mathcal{M}_{12}$ to a codeword $\mathbf{x}_1(m_{12})$. Since the capacity of the channel between the decoders is R_{12}, there is a sufficient number of different codewords $\mathbf{x}_1$ (i.e., sufficient to cover the space of cooperation messages $\mathcal{M}_{12} = [1:2^{nR_{12}}]$) that can be conveyed via this channel. To use $\mathcal{C}_n^{(\text{BC})}$ for the R_{12}-reduced SD-RBC, transmit B blocks, each of length n, and denote the messages transmitted by $(m_1^{(b)}, m_2^{(b)}) \in \mathcal{M}_1 \times \mathcal{M}_2$, where $b \in [1:B]$. In the subsequent coding scheme, the transmission of the 1st block is discarded, while during every block $b \geq 2$, the messages $(m_1^{(b-1)}, m_2^{(b-1)})$ are reliably transmitted over the channel. Accordingly, the scheme forfeits the decoding of the messages $(m_1^{(B)}, m_2^{(B)})$, which implies that the average rate pair $\left(\frac{B-1}{B}R_1, \frac{B-1}{B}R_2\right)$, over B blocks, is achievable. By taking $B \to \infty$, the transmission rates approach (R_1, R_2).

The coding scheme for the R_{12}-reduced SD-RBC during block $b \geq 2$ is as follows. First, note that the channel output $\mathbf{y}_1^{(b-1)}$ at Decoder 1 during the previous block is known at the relay at the beginning of block b. Thus, during block b, the encoder transmits the codeword $\mathbf{x}$ that corresponds to the message pair $(m_1^{(b-1)}, m_2^{(b-1)})$, while the relay transmits the codeword $\mathbf{x}_1\left(m_{12}^{(b-1)}(\mathbf{y}_1^{(b-1)})\right)$. At the end of transmission block b, Decoder 2 uses the induced relay output $\mathbf{y}_{22}^{(b)}$ to reliably decode $m_{12}^{(b-1)}(\mathbf{y}_1^{(b-1)})$. Both decoders then proceed with the decoding process for the SD-BC with cooperation to

decode the messages $(m_1^{(b-1)}, m_2^{(b-1)})$. By taking n to infinity, this coding scheme achieves $\left(\frac{B-1}{B}R_1, \frac{B-1}{B}R_2\right)$, over B blocks, for the R_{12}-reduced SD-RBC.

APPENDIX H
DERIVATION OF THE REGION IN (26)

We prove that the admissible region for the SW problem with one-sided encoder cooperation is (26), and that (26) is obtained from $\mathcal{R}_{\text{WAK}}(P_{X_1,X_2,Y})$ stated in Theorem 5. To this end set $U = X_2$ and evaluate the rate bounds in (10) to get

$$R_{12} \geq I(V; X_1|X_2)$$
$$R_1 \geq H(X_1|X_2) - I(V; X_1|X_2)$$
$$R_2 \geq H(X_2|X_1)$$
$$R_1 + R_2 \geq H(X_1, X_2). \tag{137}$$

The structure of the region in (26) implies $R_{12} \leq H(X_1|X_2)$. Thus, it suffices to show that for every $0 \leq R_{12} \leq H(X_1|X_2)$ there is a random variable V that admits the Markov property $V - X_1 - X_2$ such that $I(V; X_1|X_2) = R_{12}$. Since $R_{12} \leq H(X_1|X_2)$, there is a real number $\gamma \geq 0$ such that

$$R_{12} = H(X_1|X_2) - \gamma. \tag{138}$$

Set the auxiliary random variable $V \triangleq (\Theta, \widetilde{V})$, where $\Theta \sim \text{Ber}(\lambda)$, $\lambda \in [0, 1]$, is a binary random variable independent of (X_1, X_2) that takes values in $\mathcal{O} = \{\theta_1, \theta_2\}$, and

$$\widetilde{V} = \begin{cases} X_1, & \Theta = \theta_1 \\ 0, & \Theta = \theta_2. \end{cases} \tag{139}$$

Taking

$$\lambda = \frac{H(X_1|X_2) - \gamma}{H(X_1|X_2)} \tag{140}$$

results in

$$I(V; X_1|X_2) = \lambda I(X_1; X_1|X_2) + \bar{\lambda} I(0; X_1|X_2)$$
$$= H(X_1|X_2) - \gamma$$
$$= R_{12}$$

and implies the achievability of (26).

The converse follows by the generalized Cut-Set bound [58, Th. 1] and characterizes (26) as the admissible rate region for the SW problem with one-sided encoder cooperation.

APPENDIX I
PROOF OF THE MARKOV RELATION IN (83)

We present two proofs for the Markov relation in (83), each based on a different graphical method. The first uses the sufficient condition via undirected graphs that was introduced in [59]. The second approach relies on the notion of d-separation in functional dependence graphs (FDGs), for which we use the formulation from [60].

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

$$(M_1, M_2, M_{12}, Y_1^{t-1}, Y_{2,t+1}^n, Y_{1,t}) - X_t - Y_{2,t} \tag{141}$$

---

[3] $e(\mathcal{C}_n)$ stands for the error probability of the code $\mathcal{C}_n$ defined analogously to (9).
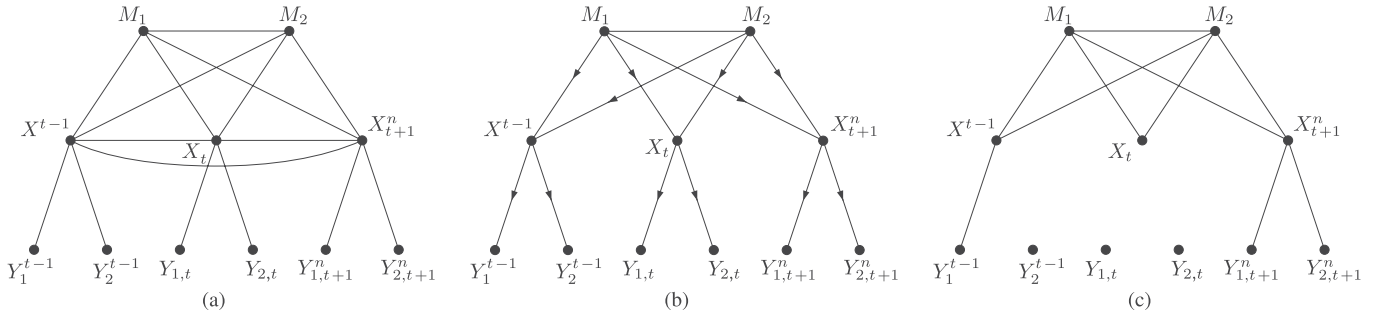
Fig. 6. (a) The undirected graph that corresponds to the PMF from (143): the relation (142) holds because all paths from $Y_{2,t}$ to $(M_1, M_2, Y_1^n, Y_{2,t+1}^n)$ pass through X_t. (b) The FDG that stems from (143): (142) follows since $\mathcal{C} = \{X_t\}$ d-separates $\mathcal{A} = \{Y_{2,t}\}$ from $\mathcal{B} = \{M_1, M_2, Y_1^n, Y_{2,t+1}^n\}$. (c) The undirected graph obtained from the FDG after the manipulations described in Definition 11.

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

$$(M_1, M_2, Y_1^n, Y_{2,t+1}^n) - X_t - Y_{2,t} \qquad (142)$$

from which (141) follows because M_{12} is determined by Y_1^n. Since the channel is SD, memoryless and without feedback, for every $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, $(x^n, y_1^n, y_2^n) \in \mathcal{X}^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ and $t \in [1 : n]$, the structure of the joint PMF from (68) gives

$$P(m_1, m_2, x^n, y_1^n, y_2^n)$$
$$= P(m_1)P(m_2)P(x^n|m_1, m_2)P(y_1^{t-1}|x^{t-1})P(y_2^{t-1}|x^{t-1})$$
$$\times P(y_{1,t}|x_t)P(y_{2,t}|x_t)P(y_{1,t+1}^n|x_{t+1}^n)P(y_{2,t+1}^n|x_{t+1}^n).$$
$$(143)$$

Given (143), the Markov relation in (142) follows by using either of two subsequently explained methods.

C. Via Undirected Graph

Fig. 6(a) shows the undirected graph that stems from the PMF in (143) with respect to the principles described in [59]. Namely, the nodes of the graph correspond to the random variables in (143). All the nodes that are associated with random variables that appear together in any of the terms in the factorization of (143) are connected by edges. For instance, the term $P(x^n|m_1, m_2)$ induces edges that connect the nodes of M_1, M_2, X^{t-1}, X_t and X_{t+1}^n with one another. The Markov chain in (142) follows from Fig. 6(a), since all paths from $Y_{2,t}$ to $(M_1, M_2, Y_1^n, Y_{2,t+1}^n)$ pass through X_t.

D. Via Functional Dependence Graph and d-Separation

Fig. 6(b) shows the FDG induced by (143). The structure of FDGs allows one to establish the conditional statistical independence of sets of random variables using the notion of d-separation.

Definition 11 (d-separation [60]): Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be disjoint subsets of the vertices of an FDG $\mathcal{G}$. $\mathcal{C}$ is said to d-separate $\mathcal{A}$ from $\mathcal{B}$ if there is no path between a vertex in $\mathcal{A}$ and a vertex in $\mathcal{B}$ after the following manipulations of the graph have been performed.

1) *Consider the subgraph $\mathcal{G}_{\mathcal{ABC}}$ of $\mathcal{G}$ consisting of the vertices in $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, as well as the edges and vertices*

encountered when moving backward one or more edges starting from any of the vertices in $\mathcal{A}$, $\mathcal{B}$ or $\mathcal{C}$.

2) *In $\mathcal{G}_{\mathcal{ABC}}$, delete all edges coming out of the vertices in $\mathcal{C}$. Call the resulting graph $\mathcal{G}_{\mathcal{AB}|\mathcal{C}}$.*

3) *Remove the arrows on the remaining edges of $\mathcal{G}_{\mathcal{AB}|\mathcal{C}}$ to obtain an undirected graph.*

The Markov relation from (143) follows by setting $\mathcal{A} = \{Y_{2,t}\}$, $\mathcal{B} = \{M_1, M_2, Y_1^n, Y_{2,t+1}^n\}$ and $\mathcal{C} = \{X_t\}$, and noting that $\mathcal{C}$ d-separates $\mathcal{A}$ from $\mathcal{B}$ [60]. To see this, in Fig. 6(c) we show the undirected graph obtained from the FDG in Fig. 6(b) by applying the manipulations described in Definition 11 with respect to the specified choices of $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$.

Neither of the methods is a special case of the other. While the first method (via undirected graphs) involves graphs with more edges, the derivation of the Markov relations using such graphs is more direct. The second method (via FDGs and d-separation) requires manipulating the original FDG. However, the FDG is typically simpler than its undirected counterpart.

ACKNOWLEDGMENT

REFERENCES

[1] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 441–445, May 1983.

[2] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications* (CISM Cource and Lecture), vol. 229, G. Longo, Ed. Springer-Verlag, 1978, pp. 171–231.

[3] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School Elect. Eng., Cornell Univ., Ithaca, NY, USA, May 1978.

[4] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[5] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[6] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[7] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, May 1975.

[8] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, Nov. 1975.

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

[10] G. Kramer and S. A. Savari, "Quantum data compression with commuting density operators," in *Proc. IEEE Int. Symp. Inf. Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 115.

[11] G. Kramer and S. A. Savari, "Communicating probability distributions," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 518–525, Feb. 2007.

[12] V. Anantharam and V. Borkar, "Common randomness and distributed control: A counterexample," *Syst. Control Lett.*, vol. 56, pp. 568–572, Jul. 2007.

[13] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4181–4206, Sep. 2010.

[14] A. Bereyhi, M. Bahrami, M. Mirmohseni, and M. R. Aref, "Empirical coordination in a triangular multiterminal network," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2149–2153.

[15] H. Asnani, H. H. Permuter, and T. Weissman, "Successive refinement with decoder cooperation and its channel coding duals," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5511–5533, Sep. 2013.

[16] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inf. Theory*, vol. 22, no. 2, pp. 226–228, Mar. 1975.

[17] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 2–14, Jan. 1972.

[18] T. Berger and R. W. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 228–236, Mar. 1989.

[19] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criteria," *IRE Nat. Conv. Rec.*, vol. 4, pp. 142–163, Jan. 1959.

[20] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1629–1638, Jun. 2002.

[21] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1181–1203, May 2003.

[22] A. Gupta and S. Verdu, "Operational duality between lossy compression and channel coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3171–3179, Jun. 2011.

[23] W. Yu, "Duality and the value of cooperation in distributed source and channel coding problems," in *Proc. 41st Annu. Allerton Conf. Commun. Control Comput.*, Monticell, IL, USA, 2003, pp. 1054–1063.

[24] H. Wang and P. Viswanath, "Fixed binning schemes: An operational duality between channel and source coding problems with side information," in *Proc. Int. Symp. Inf. Theory*, Chicago, IL, USA, Jun. 2004, p. 197.

[25] L. Dikstein, H. H. Permuter, and S. Shamai (Shitz), "MAC with action-dependent state information at one encoder," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 173–188, Jan. 2015.

[26] S. I. Bross, A. Lapidoth, and M. A. Wigger, "The Gaussian MAC with conferencing encoders," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 2702–2706.

[27] O. Simeone D. Gündüz, H. V. Poor, A. J. Goldsmith, and S. Shamai (Shitz), "Compound multiple-access channels with partial cooperation," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2425–2441, Jun. 2009.

[28] M. Wiese, H. Boche, I. Bjelakovic, and V. Jungnickel, "The compound multiple access channel with partially cooperating encoders," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3045–3066, May 2011.

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

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

[31] S. I. Gelfand 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.

[32] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.

[33] A. Lapidoth and L. Wang, "The state-dependent semideterministic broadcast channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2242–2251, Apr. 2013.

[34] 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.

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

[36] R. Koetter, M. Effros, and M. Médard, "A theory of network equivalence—Part I: Point-to-point channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 972–995, Feb. 2011.

[37] Z. Goldfeld, H. H. Permuter, and G. Kramer, "The Ahlswede–Körner coordination problem with one-sided encoder cooperation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun./Jul. 2014, pp. 1341–1345.

[38] R. Dabora and S. D. Servetto, "On the role of estimate-and-forward with time sharing in cooperative communication," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4409–4431, Oct. 2008.

[39] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.

[40] L. Dikstein, H. H. Permuter, and Y. Steinberg, "On state dependent broadcast channels with cooperation," *IEEE Trans. Inf. Theory*, to be published. [Online]. Available: http://arxiv.org/abs/1405.5083.

[41] J. L. Massey, "Applied digital information theory," ETH Zurich, Zürich, Switzerland, Tech. Rep. 19, 1998.

[42] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[43] H. G. Eggleston, *Convexity*, Cambridge, U.K.: Cambridge Univ. Press, 1958.

[44] 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.

[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[46] G. Dueck and J. Korner, "Reliability function of a discrete memoryless channel at rates above capacity," *IEEE Trans. Inf. Theory*, vol. 25, no. 1, pp. 82–85, Jan. 1979.

[47] J. Pokorny and H.-M. Wallmeier, "Random coding bound and codes produced by permutations for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 741–750, Nov. 1985.

[48] Y.-S. Liu and B. L. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 376–386, Mar. 1996.

[49] C. Chang, "Interference channel capacity region for randomized fixed-composition codes," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep. 2009, pp. 280–287.

[50] Z. Goldfeld, H. H. Permuter, and G. Kramer, "Semi-deterministic broadcast channels with cooperation," in *Proc. 28th Conv. Electr. Electron. Eng. (IEEEI)*, Eilat, Israel, Dec. 2014, pp. 1–5.

[51] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 60–64, Jan. 1977.

[52] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 166–169, Mar. 1979.

[53] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008.

[54] S. C. Draper, B. J. Frey, and F. R. Kschischang, "On interacting encoders and decoders in multiuser settings," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Chicago, IL, USA, Jun./Jul. 2004, p. 118.

[55] T. D. Nguyen and S. Lasaulce, "Capacity region of the deterministic broadcast channel with cooperative decoders," LSS Tech. Rep., Oct. 2005.

[56] A. H. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 735–740, Nov. 1985.

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

[58] A. Cohen, S. Avestimehr, and M. Effros, "On networks with side information," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Seoul, South Korea, Jun./Jul. 2009, pp. 1343–1347.

[59] H. H. Permuter, Y. Steinberg, and T. Weissman, "Two-way source coding with a helper," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2905–2919, Jun. 2010.

[60] 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 is a Feder Award, the Dean's List Award, the Basor Fellowship for honor students in the direct Ph.D. program, the Lev-Zion fellowship and the Minerva Short-Term Research Grant (MRG).

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.

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 full member of the Bavarian Academy of Sciences and Humanities since 2015.