Cooperative Binning for Semi-Deterministic Channels With Non-Causal State Information

Ido B. Gattegno, *Student Member, IEEE*, Haim H. Permuter, *Senior Member, IEEE*, Shlomo Shamai (Shitz), *Fellow, IEEE*, and Ayfer Özgür, *Member, IEEE*

***Abstract*— The capacity of the semi-deterministic relay channel (SD-RC) with *non-causal* channel state information (CSI) only at the encoder and decoder is characterized. The capacity is achieved by a scheme based on *cooperative-bin-forward*. This scheme allows cooperation between the transmitter and the relay without the need of the later to decode a part of the message. The transmission is divided into blocks, and each deterministic output of the channel (observed by the relay) is mapped to a bin. The bin index is used by the encoder and the relay to choose the cooperation codeword in the next transmission block. In *causal* settings, the cooperation is independent of the state. In *non-causal* settings, dependence between the relay's transmission and the state can increase the transmission rates. The encoder implicitly conveys partial state information to the relay. In particular, it uses the states of the next block and selects a cooperation codeword accordingly, and the relay transmission depends on the cooperation codeword and, therefore, also on the states. We also consider the multiple access channel with partial cribbing as a semi-deterministic channel. The capacity region of this channel with non-causal CSI is achieved by the new scheme. Examining the result in several cases, we introduce a new problem of a point-to-point (PTP) channel where the state is provided to the transmitter by a state encoder. Interestingly, even though the CSI is also available at the receiver, we provide an example showing that the capacity with non-causal CSI at the state encoder is strictly larger than the capacity with causal CSI.**

***Index Terms*— Cooperative binning, random binning, relay channel, multiple-access channel, semi-deterministic channel.**

I. Introduction

SEMI-DETERMINISTIC models describe a variety of communication problems in which there exists a deterministic link between a transmitter and a receiver [1]. The semi-deterministic relay channel plays an important role in the study of relay channels, as it is the canonical model where various schemes are known to achieve capacity. This work focuses on the semi-deterministic relay channel (SD-RC) and the multiple access channel (MAC) with partial cribbing encoders and non-causal channel state information (CSI) only at the encoder and decoder. The state of a channel may be governed by physical phenomena or by an interfering transmission over the channel, and the deterministic link may also be a function of this state.

The capacity of the relay channel was first studied by van der Meulen [2]. In the relay channel, an encoder receives a message, denoted by M, and sends it to a decoder over a channel with two outputs. A relay observes one of the channel outputs, denoted by Z, and uses past observations to help the encoder deliver the message. The decoder observes the other output, denoted by Y, and uses it to decode the message that was sent by the encoder. Cover and El Gamal [3] established achievable rates for the general relay channel by using a *partial-decode-forward* scheme. If the channel is semi-deterministic (i.e., the output to the relay is a function of the channel inputs), El Gamal and Aref [4] showed that this scheme achieves the capacity. Partial-decode-forward operates as follows: first, the transmission is divided into B blocks, each of length n; in each block b we send a message $M^{(b)}$, at rate R, which is independent of the messages in the other blocks. The message is split; after each transmission block, the relay decodes a part of the message and forwards it to the decoder in the next block using its transmission sequence. Since the encoder also knows the message, it can cooperate with the relay in the next block. The capacity of the SD-RC is given by maximizing $\min\{I(X, X_r; Y), H(Z|X_r) + I(X; Y|X_r, Z)\}$ over the joint probability mass function (PMF) p_{X,X_r}, where X is the input from the encoder and X_r is the input from the relay. The cooperation is expressed in the joint PMF, in which X and X_r are dependent. However, when the channel depends on a state that is unknown to the relay, the partial-decode-forward scheme is suboptimal [5], i.e., it does not achieve the capacity. The partial-decoding procedure at the relay is too restrictive since the relay is not aware of the channel state.

Focusing on the state-dependent SD-RC (Fig. 1), we consider two situations: when the CSI is available in a *causal* or a *non-causal* manner. This family of channels can be thought of as a setup in which both the encoder and the decoder are aware of the disturbance in the channel, but the relay is not. State-dependent relay channels were studied in [5]–[13]; Kolte et al. [5] derived the capacity of state-dependent SD-RC with *causal* CSI and introduced a *cooperative-bin-forward* coding scheme. In the cooperative-bin-forward

Manuscript received March 23, 2017; revised October 16, 2018; accepted May 3, 2019. Date of publication June 10, 2019; date of current version September 13, 2019. I. B. Gattegno and H. H. Permuter were supported in part by the Cyber Security Research Center within the Ben-Gurion University of the Negev, in part by the Heron Consortium via the Ministry of Economy and Science, and in part by the European Research Council (ERC) under Grant 337752. S. Shamai (Shitz) was supported in part by the Heron Consortium via the Ministry of Economy and Science and in part by the European Research Council (ERC) under Grant 694630. A. Özgür was supported in part by NSF under Grant 1514538.

I. B. Gattegno and H. H. Permuter are with the Electrical and Computer Engineering Department, Ben-Gurion University, Beer-Sheva 84105, Israel (e-mail: haimp@bgu.ac.il).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel.

A. Özgür is with the Electrical Engineering Department, Stanford University, Stanford, CA 94305 USA.

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

Digital Object Identifier 10.1109/TIT.2019.2921764

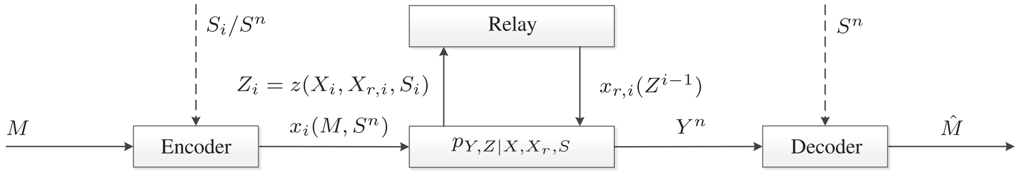

Fig. 1. SD-RC with causal/non-causal CSI at encoder and decoder.

scheme, the relay does not have to explicitly recover the message bits; instead, the encoder and relay agree on a map from the deterministic outputs space $\mathcal{Z}^n$ to a bin index. It thus differs from partial-decode-forward because the relay does not have to spend resources to decode a part of the message; rather, it simply uses the bin index to cooperate. This index is used by the relay to choose the next transmission sequence. Note that this *cooperative-binning* is not a function of the state, and, therefore, the relay does not need to have access to the state to cooperate. The encoder is also aware of this index (since the output is deterministic with respect to (w.r.t.) the state and the inputs) and coordinates with the relay in the next block, despite the lack of state information at the relay. The capacity of this channel is given by maximizing $\min\{I(X, X_r; Y|S), H(Z|X_r, S)+I(X; Y|X_r, Z, S)\}$ over $p_{X_r} p_{X|X_r S}$. Note that X and X_r are dependent, but X_r and S are not. When the state is known causally, coordination between X_r and S is not feasible. At each time i, the encoder can send information to the relay about the states up to time i. The relay can only use *strictly causal* observations Z^{i-1}, which may contain information on S^{i-1} but not on S_i. Furthermore, since the states are distributed independently, information about the past state at the relay does not help to increase the achievable rate.

The main contribution of this paper is that it provides a variant of the cooperative-bin-forward scheme that accounts for non-causal CSI. While the former scheme allows cooperation between the encoder and the relay in the causal setup, the new scheme also allows on top of it coordination between the relay's transmission and the state. Since the CSI is known non-causally by the encoder, partial knowledge of the state at the relay is feasible. The encoder can perform a look-ahead operation and transmit to the relay information about the upcoming states using an auxiliary sequence. The relay can still cooperate with the encoder based on bin indexes, which are chosen by the encoder to maximize the transmission rate. The encoder chooses an index such that (s.t.) it reveals compressed state information to the relay, using an *auxiliary* cooperation codeword. This scheme is also extended to the multiple access channel (MAC) with strictly causal partial cribbing and non-causal CSI.

The MAC with cooperation can also be viewed as a semi-deterministic model due the deterministic nature of the cooperation link. A MAC with partially cooperating encoders, which was introduced by Willems in [14], consists of rate-limited private links between two encoders. Permuter *et al.* [15] showed that for the *state-dependent* setup, the capacity can be achieved by superposition coding and rate-splitting. The cribbing is a different type of cooperation, also introduced by Willems and van der Meulen [16], in which one transmitter has access to (is cribbing) the transmission of the other. In [17], Simeone *et al.* considered cooperative wireless cellular systems and analyzed their performance with cribbing (referred to as in-band cooperation). The results show how cribbing potentially increases the capacity. Asnani and Permuter introduced in [18] a generalization of the cribbing that, termed partial and controlled cribbing, describes a setup in which one encoder has *limited* access to the transmission sequence of the other. The cribbed information is a deterministic function of the transmission sequence. Kopetz *et al.* [19] characterized the capacity region of combined *partial* cribbing and cooperation without states. When states are known *causally* at the first encoder (while the other is cribbing), Kolte *et al.* [5] derived the capacity, which is achieved by cooperative-bin-forward. We show that the variation of this scheme achieves the capacity when the states are known *non-causally*.

The results are examined for several special cases; the first is a point-to-point (PTP) channel where the CSI is available to the transmitter (through a state encoder), and to the receiver. Earlier work on limited CSI was done by Rosenzweig *et al.* [20], where the link from the state encoder to the transmitter is rate-limited. Steinberg [21] derived the capacity of rate-limited state information at the receiver. In our setting, the link between the state encoder and the transmitter is not a rate-limited bit pipe, but rather, a communication channel where the transmitter can observe the output of the state encoder in a causal fashion. We provide an example which illustrates that in this setting the capacity with non-causal CSI available at the state encoder is strictly larger than the capacity with causal CSI at the state encoder even though the receiver also has channel state information.

The remainder of the paper is organized follows. Problem definitions and capacity theorems are given in Section II. Special cases are given in Section III, and the new state-encoder problem and the example are given in Section IV. Proofs for theorems are given in Sections V, VI and VII. In Section VIII, we discuss our conclusions and final remarks.

II. Problem Definition and Main Results

A. Notation

We use the following notation. Calligraphic letters denote discrete sets, e.g., $\mathcal{X}$. Lowercase letters, e.g., x, represent variables. A vector of n variables $(x_1, \ldots, x_n)$ is denoted by x^n. A substring of x^n is denoted by x_i^j, and includes variables $(x_i, \ldots, x_j)$. Whenever the dimensions are clear from the context, the subscript is omitted. Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space where Ω is the sample space, $\mathcal{F}$ is the σ-algebra and $\mathbb{P}$ is the probability measure. Roman face letters denote events in the σ-algebra, e.g., $\mathrm{A} \in \mathcal{F}$. $\mathbb{P}[\mathrm{A}]$ is the probability assigned to A, and $\mathbb{1}[\mathrm{A}]$ is the indicator function, i.e., it indicates whether event A has occurred. Random variables are denoted by uppercase letters, e.g., X, and similar conventions apply for vectors. The probability mass function (PMF) of a random variable, X, is denoted by p_X. If $x \in \mathcal{X}$, then $p_X(x) = \mathbb{P}[X = x]$. Whenever the random variable is clear from the context, we drop the subscript. Similarly, a joint distribution of X and Y is denoted by $p_{X,Y}$ and a conditional PMF by $p_{Y|X}$. Whenever Y is a deterministic function of X, we denote $Y = f(X)$. If X and Y are independent, we denote this as $X \perp\!\!\!\perp Y$, which implies that $p_{X,Y} = p_X p_Y$, and a Markov chain is denoted as $X \leftrightarrow Y \leftrightarrow Z$ and implies that $p_{X,Y,Z} = p_{X,Y} p_{Z|Y}$. The discrete uniform distribution is denoted by $\mathrm{U}[1:m]$, where $[1:n]$ stands for a collection of integers from 1 to n.

An empirical mass function (EMF) is denoted by $\nu(a|x^n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}[x_i = a]$. Sets of typical sequences are denoted by $\mathcal{A}_\epsilon^{(n)}(p_X)$, which is a ϵ-strongly typical set with respect to PMF p_X, and defined by

$$\mathcal{A}_\epsilon^{(n)}(p_X) \triangleq \{x^n : |\nu(a|x^n) - p_X(a)| < \epsilon p_X(a), \quad \forall a \in \mathcal{X}\}.$$

Jointly typical sets satisfy the same definition w.r.t. the joint distribution and are denoted by $\mathcal{A}_\epsilon^{(n)}(p_{X,Y})$. Conditional typical sets are defined as

$$\mathcal{A}_\epsilon^{(n)}(p_{X,Y}|y^n) \triangleq \{x^n : (x^n, y^n) \in \mathcal{A}_\epsilon^{(n)}(p_{X,Y})\}$$

B. Semi-Deterministic Relay Channel

We begin with a *non-causal* state-dependent SD-RC (Fig. 1). This channel depends on a state $S_i \in \mathcal{S}$, which is known non-causally to the encoder and decoder but not to the relay. An encoder sends a message M to the decoder through a channel with two outputs. The relay observes an output Z^n of the channel, which at time i is a deterministic function of the channel inputs, X_i and $X_{r,i}$, and the state (i.e., $Z_i = z(X_i, X_{r,i}, S_i)$). Based on past observations Z^{i-1}, the relay transmits $X_{r,i}$ to assist the encoder. The decoder uses the state information and the channel output Y^n to estimate $\hat{M}$. The channel is memoryless and characterized by the PMF $p_{Y,Z|X,X_r,S}$ where $Z = z(X, X_r, S)$.

Definition 1 (Code for SD-RC) *A (n, R) code $\mathcal{C}_n$ for the SD-RC is defined by*

$$\begin{aligned} &x^n : \left[1 : 2^{nR}\right] \times \mathcal{S}^n \to \mathcal{X}^n \\ &x_{r,i} : \mathcal{Z}^{i-1} \to \mathcal{X}_r \qquad\qquad 1 \le i \le n \\ &\hat{m} : \mathcal{Y}^n \times \mathcal{S}^n \to \left[1 : 2^{nR}\right] \end{aligned}$$

Definition 2 (Achievable rate) *A rate R is achievable if there exists a sequence of (n, R) codes s.t.*

$$P_e(\mathcal{C}_n) \triangleq \mathbb{P}_{\mathcal{C}_n}\left[\hat{m}(Y^n, S^n) \neq M\right] \le \epsilon$$

for any $\epsilon > 0$ and some sufficiently large n.

The capacity is defined to be the supremum of all achievable rates.

Theorem 1 *The capacity of the SD-RC with non-causal CSI (Fig. 1), is given by*

$$\begin{aligned} C = \max\min\Big\{&I(X, X_r; Y|S), \\ &I(X; Y|X_r, Z, S, U) + H(Z|X_r, U, S) - I(U; S)\Big\} \end{aligned}$$

where the maximum is over $p_{U|S} p_{X_r|U} p_{X|X_r,U,S}$ s.t. $I(U; S) \le H(Z|X_r, U, S)$, where $Z = z(X, X_r, S)$ and $|\mathcal{U}| \le \min\{|\mathcal{S}|(|\mathcal{X}||\mathcal{X}_r| - 1) + 2, |\mathcal{S}|(|\mathcal{Y}| - 1) + 2\}$.

The proof for the above theorem is given in Section V. Let us first investigate the capacity and the role of the auxiliary random variable U, which creates coordination between the relay and the states. In the case of SD-RC without states, the capacity can be achievable by a *partial-decode-forward* scheme [3], and it coincides with the theorem by removing U. In the partial-decode-forward, the relay decodes a part of M and then uses these bits for superposition block Markov coding. These bits are used for cooperation between the relay and the encoder. When states are present, decoding message bits reduces the transmission rates.

The original cooperative-bin-forward scheme, which is used in the case of SD-RC with causal CSI, showed that there is no need to decode a part of M because the cooperation can be established by random binning. The bin index is used for superposition coding, but the relay does not and cannot depend on the states. In the non-causal setup, the superposition code can be leveraged to achieve coordination between the relay and the states, where U is a description of the states and is selected by the bin index.

When the link between the encoder and the relay is not deterministic, the cooperative binning is not applicable. The scheme relies on the fact that the encoder and the relay choose the same bin index based on Z^n. When the link is deterministic, the encoder can predict exactly which z^n the relay will observe by using the CSI and the codebook of the relay. When the link is noisy, this cannot be done. In fact, even without states, the setup of the general relay channel is still an open problem.

C. Multiple Access Channel With Partial Cribbing

Consider a MAC with partial cribbing and non-causal state information (Fig. 2). This channel depends on the state (S_1, S_2) that is known to the decoder, and each encoder $w \in \{1, 2\}$ has non-causal access to one state component $S_w \in \mathcal{S}_w$. Each encoder w sends a message M_w over the channel. Encoder 2 is cribbing Encoder 1; the cribbing is strictly causal, partial and controlled by S_1. Namely, the cribbed signal at time i, denoted by Z_i, is a deterministic function of $X_{1,i}$ and $S_{1,i}$. The cribbed information is used by Encoder 2 to assist Encoder 1 during the transmission. Although this setup is relatively close to that of the SD-RC, it is nonetheless different: 1) Encoder 2 plays the role of the relay, but it has its own message to send, and 2) the semi-deterministic link now depends only on $X_{1,i}$ and $S_{1,i}$, while in the relay setup, it is also a function of X_{2_i}, according to this analogy.

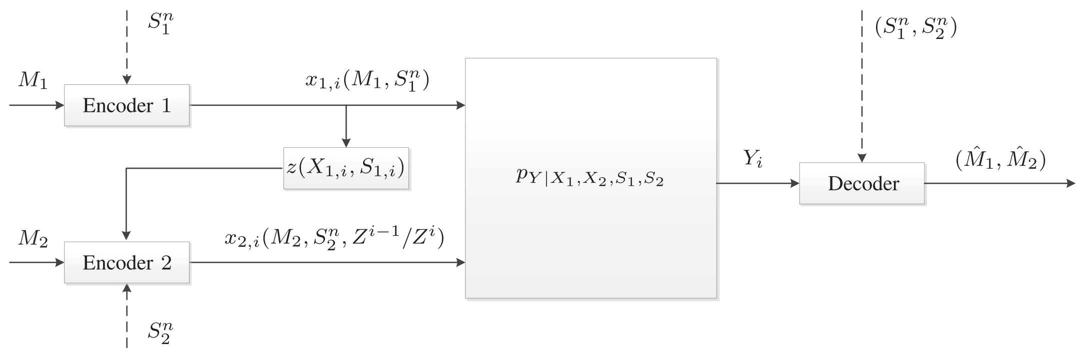

Fig. 2. State-dependent MAC with two state components and one side cribbing. When the cribbing is *strictly* causal, $X_2 = x_{2,i}(M_2, S_2^n, Z^{i-1})$. When the cribbing is causal, $X_2 = x_{2,i}(M_2, S_2^n, Z^i)$.

Definition 3 (Code for MAC) *A* (n, R_1, R_2) *code* $\mathcal{C}_n$ *for the state-dependent MAC with* strictly *causal partial cribbing and two state components is defined by*

$$\begin{aligned} x_1^n &: [1 : 2^{nR_1}] \times S_1^n \to \mathcal{X}_1^n \\ x_{2,i} &: [1 : 2^{nR_2}] \times \mathcal{S}_2^n \times Z^{i-1} \to \mathcal{X}_2 \qquad 1 \le i \le n \\ \hat{m}_1 &: \mathcal{Y}^n \times \mathcal{S}_1^n \times \mathcal{S}_2^n \to [1 : 2^{nR_1}] \\ \hat{m}_2 &: \mathcal{Y}^n \times \mathcal{S}_1^n \times \mathcal{S}_2^n \to [1 : 2^{nR_2}] \end{aligned}$$

for any $\epsilon > 0$ *and some sufficiently large* n.

Definition 4 (Achievable rate-pair) *A rate-pair* (R_1, R_2) *is achievable if there exists a sequence of* (n, R_1, R_2) *codes s.t.*

$$P_e(\mathcal{C}_n) \triangleq \mathbb{P}_{\mathcal{C}_n}\left[(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)\right] \le \epsilon$$

for any $\epsilon > 0$ *and some sufficiently large* n.

The capacity region of this channel is defined to be the closure of the achievable rates region. We note here that a setup with *causal* cribbing (Fig. 2), satisfies the same definitions except $x_{2,i} : [1 : 2^{nR_2}] \times \mathcal{S}_2^n \times Z^i \to \mathcal{X}_2$.

Theorem 2 *The capacity region for discrete memoryless MAC with non-causal CSI and* strictly causal *cribbing in Fig. 2 is given by the set of rate pairs* (R_1, R_2) *that satisfy*

$$\begin{aligned} R_1 \le& I(X_1; Y|X_2, Z, S_1, S_2, U) \\ &+ H(Z|S_1, U) - I(U; S_1|S_2) \\ R_2 \le& I(X_2; Y|X_1, S_1, S_2, U) \\ R_1 + R_2 \le& \min\big\{I(X_1, X_2; Y|S_1, S_2), \\ & I(X_1, X_2; Y|Z, S_1, S_2, U) \\ &+ H(Z|S_1, U) - I(U; S_1|S_2)\big\} \end{aligned} \tag{1a}$$

for PMFs of the form $p_{X_1,U|S_1}p_{X_2|U,S_2}$, *with* $Z = z(X_1, S_1)$, *that satisfies*

$$I(U; S_1|S_2) \le H(Z|S_1, U), \tag{1b}$$

and $|\mathcal{U}| \le \min\big\{|\mathcal{S}_1||\mathcal{S}_2|(|\mathcal{X}_1||\mathcal{X}_2|-1)+3,\ |\mathcal{S}_1||\mathcal{S}_2|(|\mathcal{Y}|-1)+4\big\}$.

Theorem 3 *The capacity region for a discrete memoryless MAC with non-causal CSI and* causal *cribbing in Fig. 2 is given by the set of rate pairs* (R_1, R_2) *that satisfies the equations in* (1) *for PMFs of the form* $p_{X_1,U|S_1}p_{X_2|Z,U,S_2}$.

The proofs for Theorem 2 and Theorem 3 are given in Sections VI and VII, respectively.

The main difference between Theorems 2 and 3 is the conditioning on Z in the PMF $p_{X_2|Z,U,S_2}$. We note here that when S_2 is degenerated, i.e., there is only one state component, the capacity region in each theorem is given by removing S_2 from the inequalities.

The auxiliary random variable U plays a double role in the MAC configuration. The first role is similar to that in the SD-RC, i.e., it creates coordination between X_2 and S_1. This coordination is the result of the non-causal setup, and, it is expressed in the PMF factorization $p_{X1,U|S_1}p_{X_2|U,S_2}$ and can be intuitively explained as in the case of Theorem 1. The second role is to create cooperation between the encoders. For example, revisit the case without states, a MAC with cribbing. In [16], Willems showed that the capacity can be achieved by superposition block Markov coding by selecting the u^n sequence using M_1. Essentially, in each transmission block, cooperation codewords u^n are selected by M_1 of the previous block, and the codewords x_1^n and x_2^n are superimposed on u^n. Then, M_1 of the current block is decoded by the second encoder and used to select u^n for the next block. In [5], Kolte *el al.* shown that there is no need to decode M_1. Instead, random binning is used with superposition coding.

In the following section, we examine the results in cases that emphasize the role of U. These are proved in Appendix A.

III. Special Cases

A. Cases of State-Dependent SD-RC

Case 1: SD-RC without states: When there is no state to the channel, the capacity of SD-RC is given by Cover and El Gamal [4] as

$$C = \max_{p_{X_r,X}} \min\{I(X, X_r; Y), I(X; Y|X_r, Z) + H(Z|X_r)\}.$$

Since there are no states, S can be omitted from the information terms in Theorem 1 and the joint PMF is $p_U p_{X_r|U} p_{X|U} p_{Z,Y|X,X_r}$. Removing U recovers the capacity.

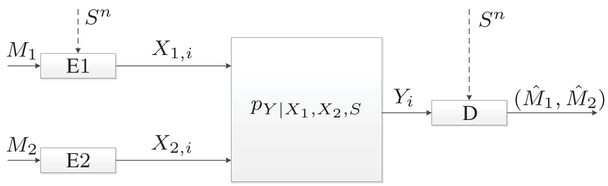

Fig. 3. Case A - MAC with CSI at one encoder.

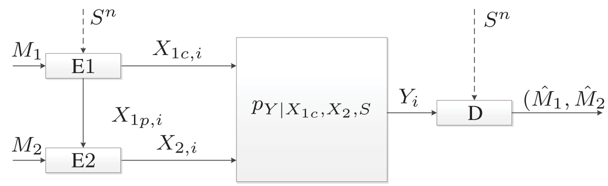

Fig. 4. Case B - MAC with CSI at one encoder and partial cooperation.

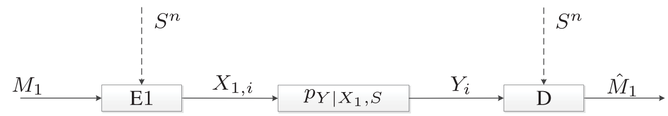

Fig. 5. Case C - PTP with non-causal CSI.

Case 2: SD-RC with causal states: Consider a similar configuration to that in Fig. 1, and assume that the states are known to the encoder in a *causal* manner. Although this is *not* a special case of the non-causal configuration, it further emphasizes the role of U. The capacity for this channel was characterized by Kolte *et al.* [5, Theorem 2] by

$$C = \max_{p_{X_r} p_{X|X_r,S}} \min\{I(X, X_r; Y|S), \\ I(X; Y|X_r, Z, S) + H(Z|X_r, S)\}$$

where $Z = z(X, X_r, S)$. Let us compare this capacity to that with non-causal states. In the latter case, we see that while X and X_r are dependent, X_r and S are not. In the non-causal case (Theorem 1), X_r and S are dependent. The random variable U generates empirical coordination w.r.t. $P_{U|S}$, and it then uses it as common side information at the encoder, the relay and the decoder. When the state is known causally, such dependency cannot be achieved since the states are drawn i.i.d. and the relay observes only past outputs of the channel. The capacity of the causal case is directly achievable by Theorem 1 by removing U.

B. Cases of State-Dependent MAC With Partial Cribbing

We consider here the naive case of one state component, i.e., S_2 is degenerated. This setup emphasizes how U is used in each case and what role it plays. We denote $S \triangleq S_1$ to emphasize this fact.

Case 1: Multiple access channel with states (without cribbing):

Consider the case of a MAC with CSI at Encoder 1 and at the decoder (Fig. 3), which is a special case without cribbing (i.e., z is constant). The capacity region, characterized by Jafar [22, Theorem 5], is defined by the convex hull of all rate pairs (R_1, R_2) satisfying the following inequalities:

$$\begin{aligned} R_1 &\le I(X_1; Y|X_2, S) \\ R_2 &\le I(X_2; Y|X_1, S) \\ R_1 + R_2 &\le I(X_1, X_2; Y|S) \end{aligned} \tag{2}$$

for all $p_{X_1|S} p_{X_2}$.

Case 2: Multiple Access Channel with Partially Cooperating Encoders:

Consider a case of MAC with a private link between the encoders (Fig. 4). In this case, the channel depends only on part of x_1, which we denote by x_{1c}. The other part of x_1, denoted by x_{1p}, is known in a strictly causal manner to Encoder 2.

This setting is different from those described in previous works, which considered rate-limited cooperation. Here we use a sequence with noiseless communication and a fixed alphabet $\mathcal{X}_{1p}$. It turns out that the capacity region of the channel is the same for both strictly causal and non-causal cooperation links. The capacity of both cases when $X_{2,i} = x_{2,i}(M_2, X_{1p}^{i-1})$ and $X_{2,i} = x_{2,i}(M_2, X_{1p}^n)$ is

$$\begin{aligned} R_1 &\le I(X_{1c}; Y|U, S) + R_{12} - I(U; S) \\ R_2 &\le I(X_2; Y|X_{1c}, U, S) \\ R_1 + R_2 &\le \min\{I(X_{1c}, X_2; Y), \\ &\qquad I(X_{1c}, X_2; Y|U) + R_{12} - I(U; S)\} \\ R_{12} &= \log_2 |\mathcal{X}_{1p}| \end{aligned} \tag{3}$$

for $p_{U,X_{1c}|S} p_{X_2|U} p_{Y|X_{1c},X_2,S}$.

Case 3: Point-to-point with non-causal CSI: Consider a configuration of a PTP channel with non-causal CSI (Fig. 5). This is a special case of the MAC, when $R_2 = 0$ and $p_{Y_2|X_1,X_2,S} = p_{Y_2|X_1,S}$. The capacity of this channel was given by Wolfowitz [23, Theorem 4.6.1] as

$$C = \max_{p_{X_1|S}} I(X_1; Y|S). \tag{4}$$

IV. Point-to-Point With State Encoder and Causality Constraint

A. The State Encoder With a Causality Constraint

We introduce a new setting of a PTP channel with a state encoder (SE) and a causality constraint (Fig 6). The SE has non-causal access to CSI and assists the encoder to increase the transmission rate. The causality constraint enforces the encoder to depend on past observations of the SE. This setting is attractive because it models situations in which the state represent interference from another party that can send information about the interference, and, the interference only affects the communication between the encoder and decoder.

The setting is defined for two cases: one with non-causal CSI and the other with causal CSI. Explicitly, the setting with non-causal CSI is defined by a state encoder (E1) $x_{1,i} : \mathcal{S}^n \to \mathcal{X}_1$, an encoder (E2) $x_{2,i} : [1 : 2^{nR_2}] \times \mathcal{X}_1^{i-1} \to \mathcal{X}_2$ and a decoder (D). Note that the encoder depends on strictly causal information from the state encoder. The definition of the second setting, however, is slightly different. First, the state encoder depends on causal CSI, i.e.,

(a) (b)

Fig. 6. Comparison between causal and non-causal CSI. (a) Non-causal CSI, strictly-causal cribbing $x_{2,i}(M_2, X_1^{i-1})$. (b) Causal CSI, causal cribbing $x_{2,i}(M_2, X_1^{i})$.

$x_{1,i} : \mathcal{S}^i \to \mathcal{X}_1$. Second, the encoder can use causal information rather than strictly causal from the state encoder. Namely, $x_{2,i} : [1 : 2^{nR_2}] \times \mathcal{X}_1^i \to \mathcal{X}_2$. We will first discuss the inclusion of the non-causal case in the MAC setting.

To apply the MAC with partial cribbing to this case, consider the following situation with only one state component. Encoder 1 has no access to the channel, i.e., $p_{Y|X_1,X_2,S} = p_{Y|X_2,S}$, and no message to send ($R_1 = 0$). Its only job is to assist Encoder 2 by compressing the CSI and sending it via a private link. The private link is the partial cribbing with $z(x_1, s) = x_1$. When the link between the encoders is non-causal, i.e., when $x_{2,i} = f(M_2, X_1^n)$, using the characterization of Rosenzweig *et al.* [20] with a rate limit of $R_s = \log|\mathcal{X}_1|$ yields

$$C = \max_{\substack{p_{U|S}p_{X_2|U}: \\ I(U;S)\le \log_2|\mathcal{X}_1|}} I(X_2; Y|U, S). \tag{5}$$

When there is a causality constraint, the transmission at time i can only depend on the strictly causal output of the state encoder, i.e., $x_{2,i} = f(M_2, X_1^{i-1})$; nonetheless, the capacity remains.

Briefly explained, the capacity is achieved as follows. The transmission is divided into blocks (block-Markov encoding). In each block, Encoder 1, which serves as the state encoder, sends a compressed version of the states of the next block. After each transmission block, Encoder 2 has a compressed version of the state of the current transmission block and uses it to ensure coherent transmission.

B. An Example - Non-causal CSI Increases Capacity

Here we provide an example to prove the claim that the non-causal CSI in the MAC configuration increases the capacity region in the general case. Consider a model wherein the channel states are coded (Fig. 6). Case (a) is a non-causal case, and (b) is causal. As we previously discussed, the channel in Fig. 6a is a special case of the *non-causal* state-dependent MAC with partial cribbing. Similarly, Fig. 6b is a special case of *causal* state-dependent MAC with partial cribbing [5].

Since this is a point-to-point configuration, it is a bit surprising that the non-causal CSI increases capacity; when the states are perfectly provided to the encoder, the capacity with causal CSI and with non-causal CSI coincide. As we will next show, in the causal case, the size of $\mathcal{X}_1$ can enforce lossy quantization on the state, while in the non-causal case, the states can be losslessly compressed.

For every channel $p_{Y|X_2,S}$ and states distribution p_S,

$$C_{\text{nc}} = \max_{\substack{p_{U|S}p_{X_2|U}: \\ I(U;S)\le \log_2|\mathcal{X}_1|}} I(X_2; Y|S, U),$$

$$C_{\text{c}} = \max_{p_{X_2|X_1}, x1(s)} I(X_2; Y|S, X_1)$$

where C_{nc} and C_{c} are the capacity of non-causal and causal CSI configurations, respectively. Assume that the states distribution is

$$p_S(s) = \begin{cases} \frac{p}{2} & \text{if } s = 0, 1 \\ 1 - p & \text{if } s = 2. \end{cases}$$

For each state there is a different channel: a Z-channel for $s = 0$ and an S-channel for $s = 1$, where both channels share the same parameter α, and a noiseless channel for $s = 2$ (Fig. 7).

The idea is that when the CSI is known non-causally, in contrast to a causal case, we can compress S^n. Assume that X_1 is binary and that p is small enough, for instance, $p = 0.2$, s.t.

$$H(S) < \log_2|\mathcal{X}_1| = 1.$$

Taking $U = S$ satisfies $I(U; S) = H(S) \le 1$ and results in the non-causal capacity

$$C_{\text{nc}} = \frac{p}{2}\left(C_{\text{Z-channel}}(\alpha) + C_{\text{S-channel}}(\alpha)\right) + (1 - p)$$

where

$$\begin{aligned} C_{\text{Z-channel}}(\alpha) &= C_{\text{S-channel}}(\alpha) \\ &= H_b\left(\frac{2^{H_b(\alpha)/\bar{\alpha}}}{1 + 2^{H_b(\alpha)/\bar{\alpha}}}\right) - \frac{H_b(\alpha)/\bar{\alpha}}{1 + 2^{H_b(\alpha)/\bar{\alpha}}}. \end{aligned}$$

On the other hand, the capacity for causal CSI under the same assumptions is

$$\begin{aligned} C_{\text{c}} = \max_{\beta}\Big[&(1 - p)H_b(\beta) + \frac{p}{2}C_{\text{Z-channel}}(\alpha) + \\ &\frac{p}{2}\left(H_b\left(\beta + \bar{\beta}\alpha\right) - \bar{\beta}H_b(\alpha)\right)\Big], \end{aligned}$$

which can be achieved by one of several deterministic functions $x_{1,i}(S^i)$. One such function can be a mapping from $s_i = 2$ to $x_{1,i} = 0$ and $s_i = 0/1$ to $x_{1,i} = 1$. Note that this operation causes a lossy quantization of the CSI. For comparison, we also provide the capacity when there is no CSI at the encoder, which is

$$C_{\text{no-CSI}} = p\left(H_b\left(\frac{1 + \alpha}{2}\right) - 0.5H_b(\alpha)\right) + (1 - p).$$

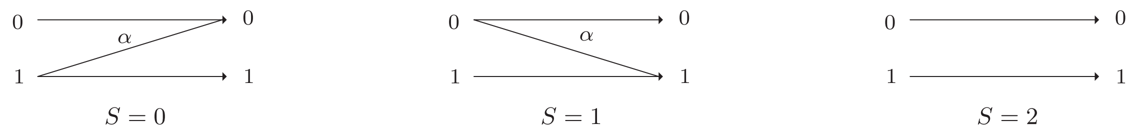

Fig. 7. Example of a state-dependent channel.

TABLE I

CAPACITY OF PTP WITH CODED CSI - NUMERICAL EVALUATIONS FOR $p = 0.2$.

α	**No-CSI**	**Causal CSI**	**Non-causal CSI**
0	1	1	1
0.5	0.8623	0.8633	**0.8644**
1	0.8	0.8	0.8

The capacity of the channels (non-causal, causal, no CSI) for $p = 0.2$ are summarized in Table I. There are two points where all the three channels result in the same capacity. In the first, when $\alpha = 0$, the channel is noiseless for $s = 0, 1, 2$ and the capacity is 1. There is no need for CSI at the encoder, and therefore, the capacity is the same (among the three cases). At the second point, when $\alpha = 1$, the channel is stuck at 0 and stuck at 1 for $s = 0$ and $s = 1$, respectively, and noiseless for $s = 2$. In this case, we can set $p_{X_1}(1) = 0.5$ for every s and achieve the capacity. Therefore, the encoder does not use the CSI in those cases. However, for every $\alpha \in (0, 1)$, the capacity of the non-causal case is strictly larger than those of the others, which confirms that non-causal CSI indeed increases the capacity region.

V. PROOF FOR THEOREM 1

A. Direct

The scheme is based on superposition block Markov coding. The cooperative binning scheme uses the deterministic property of the channel, i.e., the link from the encoder to the relay, to determine exactly which output the relay will observe during the transmission. In turn, this output is mapped to a bin index to perform superposition encoding.

Codebook: Divide the transmission to B block, choose a distribution $p_{X,U|S}p_{X_r|S}$ and split the message of each block, i.e., $m^{(b)} = (m'^{(b)}, m''^{(b)})$, $m'^{(b)} \in 2^{nR'}$, $m''^{(b)} \in 2^{nR''}$. For each block $b \in [1 : B]$, a codebook $\mathcal{C}_n^{(b)}$ is generated as follows:

- *Binning*: Partition the set $\mathcal{Z}^n$ into 2^{nR_B} bins by uniformly and independently choosing an index $\text{bin}^{(b)}(z^n) \sim U\left[1 : 2^{nR_B}\right]$.
- *Cooperation codewords*: Generate 2^{nR_B} u-codewords

$$u^n\left(l^{(b-1)}\right) \sim \prod_{i=1}^{n} p_U(u_i)$$

for $l^{(b-1)} \in [1 : 2^{nR_B}]$.
- *Relay codewords*: For each $u^n \in \mathcal{U}^n$ generate x_r-codeword $x_r^n(u^n) \sim \prod_{i=1}^{n} p_{X_r|U}\left(x_{r,i}|u_i\right)$.

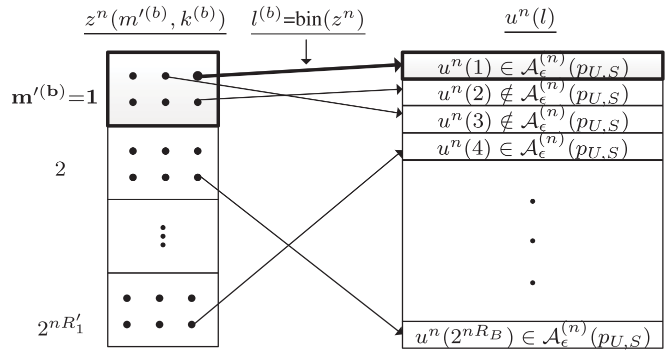

Fig. 8. Encoding procedure: looking for a sequence z^n such that $\text{bin}(z^n)$ points toward a coordinated sequence u^n.

- *z-codewords*: For each $u^n \in \mathcal{U}^n$, $x_r^n \in \mathcal{X}_r^n$ and $s^n \in \mathcal{S}^n$, generate $2^{n(R'+\tilde{R})}$ z-codewords

$$z^n(m'^{(b)}, k^{(b)}|x_r^n, u^n, s^n) \sim \prod_{i=1}^{n} p_{Z|X_r,U,S}(z_i|x_{r,i}, u_i, s_i)$$

for $m'^{(b)} \in [1 : 2^{nR'}]$, $k^{(b)} \in [1 : 2^{n\tilde{R}}]$
- *Transmission codewords*: For each $z^n \in \mathcal{Z}^n$, $u^n \in \mathcal{U}^n$, $x_r^n \in \mathcal{X}_r^n$ and $s^n \in \mathcal{S}^n$ draw $2^{nR''}$ x-codewords

$$x^n(m''^{(b)}|z^n, x_r^n, u^n, s^n) \sim \prod_{i=1}^{n} p_{X|Z,X_r,U,S}(x_i|z_i, x_{r,i}, u_i, s_i)$$

for $m''^{(b)} \in [1 : 2^{nR''}]$.

Encoder: Denote the bin index chosen in block b by $l^{(b)}$ and let $l^{(0)} = m'^{(1)} = m''^{(1)} = m'^{(B)} = m''^{(B)} = k^{(B)} = 1$. Assume that $l^{(b-1)}$ is known due to former operations at the encoder. For $x_r^n(u^n(l^{(b-1)})), u^n(l^{(b-1)})$, we denote

$$z^n(m'^{(b)}, k^{(b)}|l^{(b-1)}, s^{n(b)}) = z^n(m'^{(b)}, k^{(b)}|x_r^n, u^n, s^{n(b)})$$

throughout the proof.

In each block b, the encoder observes $m^{(b)}$. Since the link is between the encoder and the relay is deterministic, it can dictate which sequence the relay will observe during the block. It thus dictates a sequence s.t. the auxiliary codeword u^n of the next block will coordinate with $s^{n(b+1)}$. This procedure is illustrated in Fig. 8 and performed by the following steps.

First, the encoder finds $k^{(b)}$ s.t. for $z^n = z^n(m'^{(b)}, k^{(b)}|l^{(b-1)}, s^{n(b)})$,

$$\left(u^n(\text{bin}(z^n)), s^{n(b+1)}\right) \in \mathcal{A}_\epsilon^{(n)}(p_{S,U})$$

If there exist multiple $k^{(b)}$ that satisfy the above, choose the first one. Then, the encoder sends

$$x^n\left(m''^{(b)}|z^n, x_r^n(u^n(l^{(b-1)})), u^n(l^{(b-1)}), s^{n(b)}\right).$$

and sets $l^{(b)} = \text{bin}\left(z^n(m'^{(b)}, k^{(b)}|l^{(b-1)}, s^{n(b)})\right)$. We abbreviate the notation of the chosen transmission sequence by $x^{n(b)}\left(m''^{(b)}|m'^{(b)}, k^{(b)}, l^{(b-1)}, s^{n(b)}\right)$.

Relay: Assume $l^{(b-1)}$ is known. At block b, send $x_r^n\left(u^n(l^{(b-1)})\right)$. Denote this sequence by $x_r^n\left(l^{(b-1)}\right)$. After the relay observes $z^{n(b)}$, it determines $l^{(b)} = \text{bin}(z^{n(b)})$. Note that the relay is choosing the same index $l^{(b)}$ as the encoder does. Thus, both $x^{n(b)}$ and $x_r^{n(b)}$ are superimposed on the same $u^{n(b)}$.

Decoder: We perform decoding using a sliding window; this is a decoding procedure that decodes from block 1 to $B-1$, and therefore, it reduces the delay for recovering message bits at the decoder.[1] We start at block 2, since the first cooperation sequence is not necessarily coordinated with the states at that block. Moreover, since the first message is fixed, the decoder can imitate the encoder's operation and find $l^{(1)}$.

Assume $l^{(b-1)}$ is known due to previous decoding operations. At block b, the decoder operates in two steps:

1) For each $m'^{(b)}$, it looks for $\hat{k}^{(b)}(m'^{(b)}, l^{(b-1)}, s^{n(b)}, s^{n(b+1)})$ and $\hat{l}^{(b)}(m'^{(b)}, l^{(b-1)}, s^{n(b)}, s^{n(b+1)})$ the same way that the encoder does. We denote these indexes by $\hat{k}^{(b)}\left(m'^{(b)}\right)$ and $\hat{l}^{(b)}\left(m'^{(b)}\right)$.
2) The decoder finds a unique $(\hat{m}'^{(b)}, \hat{m}''^{(b)})$ s.t. (6), as shown at the bottom of the next page are satisfied.

Error analysis: The code $\mathcal{C}_n$ is defined by the block-codebooks and the encoders and decoder functions. We bound the average probability of an error at block b, conditioned on successful decoding in blocks $[1:b-1]$. Without loss of generality, we assume that $M'^{(b)} = 1$ for each $b \in [1:B]$. For each block step of the encoding/decoding procedure, we define the error events as

$$\mathrm{E}_1(b)=\left\{\begin{array}{l}\forall k^{(b)}: \left(U^n(\text{Bin}^{(b)}(Z^n)), S^{n(b+1)}\right) \notin \mathcal{A}_\epsilon^{(n)}\left(p_{S,U}\right)\\ Z^n = Z^n(1, k^{(b)}|L^{(b-1)}, S^{n(b)})\end{array}\right\}$$

$$\mathrm{E}_2(b)=\left\{\begin{array}{l}\exists k, m'^{(b)} \neq 1:\\ \text{Bin}^{(b)}(Z^n(m'^{(b)}, k|L^{(b-1)}, S^{n(b)}))=L^{(b)}\end{array}\right\}$$

$$\mathrm{E}_3(b)=\left\{\begin{array}{l}\text{Condition (6) is not satisfied by}\\ \left(\hat{m}'^{(b)}, \hat{m}''^{(b)}\right)=(1,1)\end{array}\right\}$$

$$\mathrm{E}_4(b)=\left\{\begin{array}{l}\text{Condition (6) is satisfied by some}\\ \left(\hat{m}'^{(b)}, \hat{m}''^{(b)}\right)\neq(1,1)\end{array}\right\}$$

$$\mathrm{E}_5(b)=\left\{\hat{L}^{(b)} \neq L^{(b)}\right\}$$

and the event of error in all blocks up to b as

$$E(b)=\bigcup_{j=1}^{b}\{\mathrm{E}_1(j)\cup\mathrm{E}_2(j)\cup\mathrm{E}_3(j)\cup\mathrm{E}_4(j)\cup\mathrm{E}_5(j)\}$$

Each event E_j, $j \in [1:5]$ stands for an error during the encoding/decoding procedures: $E_1(b)$ is a failure at the lookup procedure at the encoder, $E_2(b)$ is when the decoder associates the wrong message with the right bin index, $E_3(b)$ and $E_4(b)$ are the errors when performing typicality tests at the decoder (eq. (6)) and $E_5(b)$ may cause decoding errors in the next block. The average probability of an error is upper bounded by the probability of failure up to block B, i.e.,

$$\begin{aligned}P_e =&\mathbb{E}_{\mathcal{C}_n}\left[P_e\left(\mathcal{C}_n\right)\right]\\ \leq&\mathbb{P}\left[E(B)\right]\\ \leq&\sum_{b=1}^{B}\underbrace{\mathbb{P}\left[\mathrm{E}_1(b)|\mathrm{E}_5^c(b-1)\right]}_{(1)}\\ &+\underbrace{\mathbb{P}\left[\mathrm{E}_2(b)|\mathrm{E}_5^c(b-1)\right]}_{(2)}+\underbrace{\mathbb{P}\left[\mathrm{E}_3(b)|\mathrm{E}_5^c(b-1),\mathrm{E}_1^c(b)\right]}_{(3)}\\ &+\underbrace{\mathbb{P}\left[\mathrm{E}_4(b)|\mathrm{E}_5^c(b-1),\mathrm{E}_2^c(b),\mathrm{E}_1^c(b)\right]}_{(4)}\\ &+\underbrace{\mathbb{P}\left[\mathrm{E}_5(b)|\mathrm{E}_5^c(b-1),\mathrm{E}_1^c(b),\mathrm{E}_2^c(b),\mathrm{E}_4^c(b)\right]}_{(5)},\end{aligned}$$

where the second inequality follows from union bound and total probability.[2]

We analyze each event separately. Beforehand, let us introduce a lemma that ensures us that $E_1(b)$ does not occur. That is, at the beginning of block b, during the lookup by the encoder, we will find a bin index $\text{Bin}^{(b)}$ s.t. the corresponding cooperation codeword $U^n(\text{Bin}^{(b)})$ is coordinated with $S^{n(b+1)}$.

Lemma 1 (Indirect covering lemma) *Let* $\{Z^n(k)\}_{k\in[1:2^{nR}]}$ *be a collection of sequences, each sequence is drawn i.i.d. according to* $\prod_{i=1}^n p_{Z|V}(z_i|v_i)$. *For every* $z^n \in \mathcal{Z}^n$, *let* $l(z^n) = Bin(z^n) \sim U[1:2^{nR_B}]$. *Then, for every* $v^n \in \mathcal{A}_\epsilon^{(n)}$, *and, for any* $\delta_1, \delta_2 > 0$, *if*

$$\begin{aligned}R <&H(Z|V)-\delta_1\\ R <&R_B-\delta_2,\end{aligned}$$

then,

$$\lim_{n\to\infty}\mathbb{P}\Big[|\{l:\exists k \text{ s.t. } Bin(Z^n(k))=l\}| < 2^{n(R-\delta_n)}|V^n=v^n\Big]=0$$

where $\delta_n \to 0$ *as* $n \to \infty$,

The corresponding proof for this is given in Appendix B. It states that by choosing $R < H(Z|V) - \delta_1$ and $R_B > R + \delta_2$, we guarantee (with high probability) that we can choose approximately $2^{n(R-\Delta_n)}$ different bin indexes by selecting $z^n(k), k \in [1:2^{nR}]$ and hashing the binning function $l(\cdot)$. Note that the lemma does *not* guarantee that the bin index of each z^n is unique; rather, it guarantees a lower bound on the number of different bin indexes that are assigned to z^n sequences. Since each bin is associated with a sequence u^n, we can utilize this property together with a covering lemma [25, Lemma 3.3].

We now proceed to bound the probability of the error events.

1) $\mathbb{P}[\mathrm{E}_1(b)|\mathrm{E}_5^c(b-1)]$: By lemma 1, the probability of seeing fewer than $2^{n(\tilde{R}-\Delta_n)}$ different bin indexes (indexed by l) goes to 0 if $\tilde{R} < H(Z|X_r, U, S) - \delta_1$ and $R_B > \tilde{R} + \delta_2$. This is done by identifying $V = (X_r, U, S)$. Denote

$$\mathrm{A} = \left\{\text{there are less than } 2^{n(\tilde{R}-\Delta_n)}\text{different bin indexes}\right\}$$
$$\mathcal{D} = \left\{l:\ \exists k \text{ such that } \text{Bin}(Z^n(k)) = l\right\}$$

[1] The sliding window technique achieves capacity for the relay channel but not for the MAC [24].

[2] Assume that A and B are two events. Then $\mathbb{P}[\mathrm{A}\cup\mathrm{B}] \leq \mathbb{P}[A]+\mathbb{P}\left[B|A^c\right]$.

Therefore,

$$\begin{aligned}
&\mathbb{P}[\mathrm{E}_1(b)|E_5^c(b-1)] \le \\
&\le \mathbb{P}\big[\mathrm{E}_1(b)|\mathrm{E}_5^c(b-1), \mathrm{A}^c\big] + \mathbb{P}\big[\mathrm{A}|\mathrm{E}_5^c(b-1)\big] \\
&= \mathbb{P}\big[\forall k, \big(U^n(\mathrm{Bin}(Z^n(k))), S^n\big) \notin \mathcal{A}_\epsilon^{(n)}(p_{U,S})|\mathrm{E}_5^c(b-1), \mathrm{A}^c\big] \\
&\quad + \epsilon_n' \\
&\le \mathbb{P}\big[\cap_{l\in\mathcal{D}} (U^n(l), S^n) \notin \mathcal{A}_\epsilon^{(n)}(p_{U,S})|\mathrm{E}_5^c(b-1), \mathrm{A}^c\big] + \epsilon_n' \\
&\le (1 - 2^{-n(I(U;S)+\delta'(\epsilon))})^{2^{n(\tilde{R}-\Delta_n)}} + \epsilon_n' \\
&\le \exp\{-2^{n(\tilde{R}-I(U;S)-\Delta_n-\delta'(\epsilon))}\} + \epsilon_n'
\end{aligned}$$

which tend to 0 when $n \to \infty$ if

$$\tilde{R} > I(U;S) + \Delta_n + \delta'(\epsilon).$$

2) $\mathbb{P}\left[\mathrm{E}_2(b)|\mathrm{E}_5^c(b-1)\right]$: Denote

$$Z^n(m'^{(b)}, k) = Z^n(m'^{(b)}, k|L^{(b-1)}, S^{n(b)}),$$

and let $K^{(b)}$ denote the chosen index by the lookup in block b, i.e., $L^{(b)} = \mathrm{Bin}^{(b)}(Z^n(1, K^{(b)}))$.
Consider

$$\begin{aligned}
&\mathbb{P}\big[\mathrm{E}_2(b)|\mathrm{E}_5^c(b-1)\big] = \\
&= \mathbb{P}\left[\exists k, m'^{(b)} \ne 1 : \mathrm{Bin}^{(b)}(Z^n(m'^{(b)}, k)) = L^{(b)}\right] \\
&= \mathbb{P}\left[\begin{array}{c}\exists k, m'^{(b)} \ne 1 : \\ \mathrm{Bin}^{(b)}(Z^n(m'^{(b)}, k)) = \mathrm{Bin}^{(b)}(Z^n(1, K^{(b)}))\end{array}\right] \\
&\overset{(a)}{\le} \sum_{m'^{(b)}>1,k} \mathbb{P}\left[\mathrm{Bin}^{(b)}(Z^n(m'^{(b)}, k)) = \mathrm{Bin}^{(b)}(Z^n(1, K^{(b)}))\right] \\
&\overset{(b)}{\le} \sum_{m'^{(b)}>1,k} \mathbb{P}\left[\begin{array}{c}\mathrm{Bin}^{(b)}(Z^n(m'^{(b)}, k)) = \mathrm{Bin}^{(b)}(Z^n(1, K^{(b)})) \\ Z^n(m'^{(b)}, k) = Z^n(1, K^{(b)})\end{array}\right] \\
&+ \sum_{m'^{(b)}>1,k} \mathbb{P}\left[\begin{array}{c}\mathrm{Bin}^{(b)}(Z^n(m'^{(b)}, k)) = \mathrm{Bin}^{(b)}(Z^n(1, K^{(b)})) \\ Z^n(m'^{(b)}, k) \ne Z^n(1, K^{(b)})\end{array}\right] \\
&\le \sum_{m'^{(b)}>1,k} \mathbb{P}\left[Z^n(m'^{(b)}, k) = Z^n(1, K^{(b)})\right] \\
&+ \sum_{m'^{(b)}>1,k} \mathbb{P}\left[\begin{array}{c}\mathrm{Bin}^{(b)}(Z^n(m'^{(b)}, k)) = \mathrm{Bin}^{(b)}(Z^n(1, K^{(b)})) \\ \text{conditioned on} \\ Z^n(m'^{(b)}, k) \ne Z^n(1, K^{(b)})\end{array}\right] \\
&\le 2^{n(R'+\tilde{R})} 2^{-n(H(Z|X_r,U,S)-\delta_1(\epsilon))} + 2^{n(R'+\tilde{R})} 2^{-nR_B}
\end{aligned}$$

where (a) follows by union bound and (b) by the law of total probability. Therefore, this probability goes to zero if

$$\begin{aligned}
R' + \tilde{R} &< R_B \\
R' + \tilde{R} &< H(Z|X_r, U, S) - \delta_1(\epsilon)
\end{aligned}$$

TABLE II

STATISTICAL RELATIONS IN THE DECODING PROCEDURE FOR THE SD-RC.

$\hat{m}'^{(b)}$	$\hat{m}''^{(b)}$	PMF (block b)	PMF (block $b+1$)
>1	$*$	$p_{Z,X\|X_r,U,S}\,p_{Y\|X_r,U,S}$	$p_{U,X_r}\,p_{Y\|S}$
1	>1	$p_{X\|Z,X_r,U,S}\,p_{Y\|Z,X_r,U,S}$	$p_{U,X_r,Y\|S}$

3) $\mathbb{P}\left[\mathrm{E}_3(b)|\mathrm{E}_5^c(b-1), \mathrm{E}_1^c(b)\right]$: Since $(U^n(\hat{L}^{(b)}(1)), S^{n(b)}) \in \mathcal{A}_\epsilon^{(n)}(p_{S,U})$, following the conditional typicality lemma [25, Chapter 2.5], the probability of this event goes to zero as n goes to infinity.

4) $\mathbb{P}\left[\mathrm{E}_4(b)|\mathrm{E}_5^c(b-1), \mathrm{E}_2^c(b), \mathrm{E}_1^c(b)\right]$: We distinguish the events in block b and block $b+1$. Note that conditioning on E_2^c ensures us that for $m'^{(b)} \ne 1$ we have $\hat{l}^{(b)}(m'^{(b)}) \ne L^{(b)}$. Therefore, at block $b+1$, for each $m'^{(b)} \ne 1$ the tuple $\left(S^{n(b+1)}, U^n(\hat{l}^{(b)}(m'^{(b)})), X_r^n(\hat{l}^{(b)}(m'^{(b)}))\right)$ is independent of $Y^{n(b+1)}$ given $S^{n(b+1)}$. At block b, $(U^n(\hat{L}^{(b-1)}), X_r^n(\hat{L}^{(b-1)}), S^{n(b)}, Y^{n(b)}) \in \mathcal{A}_\epsilon^{(n)}(p_{S,U,X_r,Y})$ with high probability.
The probability of the error is divided into two different cases (Table II). Applying the standard packing lemma [25, Lemma 3.1] to each gives the following bounds:

$$\begin{aligned}
R &< I(Z,X;Y|X_r,U,S) + I(X_r,U;Y|S) - \delta_3(\epsilon) - \delta_4(\epsilon) \\
R'' &< I(X;Y|Z,X_r,U,S) - \delta_5(\epsilon)
\end{aligned}$$

Note that $I(Z,X;Y|X_r,U,S) + I(X_r,U;Y|S) = I(X_r,X;Y|S)$ since $U \leftrightarrow (Z,X_r,X,S) \leftrightarrow Y$ forms a Markov chain and Z is a function of (X,X_r,S).

5) $\mathbb{P}\left[\mathrm{E}_5(b)|\mathrm{E}_5^c(b-1), \mathrm{E}_1^c(b), \mathrm{E}_2^c(b), \mathrm{E}_4^c(b)\right]$: If previous events did not occur, then the probability of this event is 0.

Following this derivation, the probability of an error goes to zero if

$$\begin{aligned}
R' + \tilde{R} &< R_B \\
R' + \tilde{R} &< H(Z|X_r, U, S) - \delta_1(\epsilon) \\
\tilde{R} &> I(U;S) + \Delta_n + \delta'(\epsilon) \qquad (7)\\
R &< I(X, X_r; Y|S) - \delta_3(\epsilon) - \delta_4(\epsilon) \\
R'' &< I(X;Y|Z, X_r, U, S) - \delta_5(\epsilon) \qquad (8)
\end{aligned}$$

Performing Fourier-Motzkin elimination (see [26]) on the rates in (8) yields

$$\begin{aligned}
R &\le I(X, X_r; Y|S) \\
R &\le I(X;Y|X_r, Z, S, U) + H(Z|X_r, U, S) - I(U;S)
\end{aligned}$$

Cardinality bounds on the auxiliary random variable U are obtained using the convex cover method [25, Appendix C]. □

$$\left(s^{n(b)}, u^n(l^{(b-1)}), x_r^n(l^{(b-1)}), z^n(\hat{m}', \hat{k}(\hat{m}')|l^{(b-1)}, s^{n(b)}), x^n(\hat{m}''^{(b)}|\hat{m}', \hat{k}(\hat{m}'), l^{(b-1)}, s^{n(b)}), y^{n(b)}\right) \in \mathcal{A}_\epsilon^{(n)}(p_{S,U,X_r,X,Z,Y}) \qquad \text{(6a)}$$

$$\left(s^{n(b+1)}, u^n(\hat{l}^{(b)}(\hat{m}'^{(b)})), x_r^n(\hat{l}^{(b)}(\hat{m}'^{(b)})), y^{n(b+1)}\right) \in \mathcal{A}_\epsilon^{(n)}(p_{S,U,X_r,Y}) \qquad \text{(6b)}$$

B. Converse

Assume that the rate R is achievable. The first bound on the rate is obtained by

$$\begin{aligned} nR &= H(M) \\ &\overset{(a)}{=} H(M|S^n) \\ &\overset{(b)}{\leq} I(M;Y^n|S^n) + n\epsilon_n \\ &= \sum_{i=1}^n I(M;Y_i|Y^{i-1},S^n) + n\epsilon_n \\ &\overset{(c)}{=} \sum_{i=1}^n I(M,X^i;Y_i|Y^{i-1},S^n) + n\epsilon_n \\ &\overset{(d)}{=} \sum_{i=1}^n I(M,X^i,X_{r,i};Y_i|Y^{i-1},S^n) + n\epsilon_n \\ &\leq \sum_{i=1}^n I(M,X^i,X_{r,i},Y^{i-1},S^{n\backslash i};Y_i|S_i) + n\epsilon_n \\ &\overset{(e)}{=} \sum_{i=1}^n I(X_i,X_{r,i};Y_i|S_i) + n\epsilon_n \\ &\overset{(f)}{=} n\left(I(X_Q,X_{r,Q};Y_Q|S_Q,Q) + \epsilon_n\right) \\ &\leq n\left(I(Q,X_Q,X_{r,Q};Y_Q|S_Q) + \epsilon_n\right) \\ &\overset{(g)}{=} n\left(I(X_Q,X_{r,Q};Y_Q|S_Q) + \epsilon_n\right) \end{aligned}$$

where:
(a) - since $M \perp\!\!\!\perp S^n$,
(b) - follows by Fano's inequality,
(c) - X^n is a function of (M, S^n),
(d) - $X_{r,i}$ is a composition of functions of X^{i-1}, S^{i-1},[3]
(e) - since $(M, X^{i-1}, Y^{i-1}, S^{n\backslash i}) \leftrightarrow (X_i, X_{r,i}, S_i) \leftrightarrow Y_i$ is a Markov chain,
(f) - by setting $Q \sim U[1:n]$ to be an independent time-sharing random variable,
(g) - since $Q \leftrightarrow (X_Q, X_{r,Q}, S_Q) \leftrightarrow Y_Q$ is a Markov chain.

The second bound is obtained by

$$\begin{aligned} nR &= H(M) \\ &= H(M|S^n) \\ &= H(M,Z^n|S^n) \\ &= H(Z^n|S^n) + H(M|Z^n,S^n) \end{aligned}$$

The first term is bounded by

$$\begin{aligned} H(Z^n|S^n) &= H(Z^n,S^n) - H(S^n) \\ &\overset{(a)}{=} \sum_{i=1}^n \Big(H(Z_i,S_i|Z^{i-1},S^{i-1}) - H(S_i)\Big) \\ &= \sum_{i=1}^n \Big(H(Z_i|Z^{i-1},S^i) + H(S_i|Z^{i-1},S^{i-1}) - H(S_i)\Big) \\ &\overset{(b)}{=} \sum_{i=1}^n \big(H(Z_i|X_{r,i},S_i,U_i) - I(U_i;S_i)\big) \\ &= n\left(H(Z_Q|X_{r,Q},S_Q,U_Q,Q) - I(U_Q;S_Q|Q)\right) \\ &\overset{(c)}{=} n\left(H(Z_Q|X_{r,Q},S_Q,U_Q,Q) - I(U_Q,Q;S_Q)\right) \end{aligned}$$

where:
(a) - since S^n is i.i.d.,
(b) - by setting $U_i \triangleq (S^{i-1}, Z^{i-1})$, and $X_r, i = f(Z^{i-1})$,
(c) - since $Q \perp\!\!\!\perp S_Q$.
Define $U = (U_Q, Q)$, $X = X_Q, X_r = X_{r,Q}, S = S_Q$ and $Z = Z_Q$. The non-negativity of the entropy also imposes distributions $p_{U,S,X_r,Z}$ that comply with

$$I(U;S) \leq H(Z|X_r,S,U)$$

The second term is upper bounded by

$$\begin{aligned} &H(M|Z^n,S^n) \leq I(M;Y^n|Z^n,S^n) + n\epsilon_n \\ &= \sum_{i=1}^n I(M;Y_i|Y^{i-1},Z^n,S^n) + n\epsilon_n \\ &\leq \sum_{i=1}^n I(M,Y^{i-1},Z_{i+1}^n,S_{i+1}^n,X_i;Y_i|Z_i,X_{r,i},S_i,U_i) + n\epsilon_n \\ &\overset{(d)}{=} \sum_{i=1}^n I(X_i;Y_i|Z_i,X_{r,i},S_i,U_i) + n\epsilon_n \\ &\qquad = n\left(I(X_Q;Y_Q|Z_Q,X_{r,Q},S_Q,U_Q,Q) + \epsilon_n\right) \\ &\qquad = n\left(I(X;Y|Z,X_r,S,U) + \epsilon_n\right) \end{aligned}$$

where (d) follows since $(M, Y^{i-1}, Z_{i+1}^n, S_{i+1}^n) \leftrightarrow (X_i, S_i, U_i, Z_i, X_{r,i}) \leftrightarrow Y_i$ is a Markov chain and $X_{r,i} = x_{r,i}(Z^{i-1})$. Therefore,

$$nR \leq n(I(X;Y|Z,X_r,S,U) + H(Z|X_r,S,U) - I(U;S) + \epsilon_n)$$

We need to show that the following conditions hold:

- $Q \perp\!\!\!\perp S_Q$
- The following Markov chains hold
$$(M,X^{i-1},Y^{i-1},S^{n\backslash i}) \leftrightarrow (X_i,X_{r,i},S_i) \leftrightarrow Y_i$$
$$(M,Y^{i-1},Z_{i+1}^n,S_{i+1}^n) \leftrightarrow (X_i,S_i,U_i,Z_i,X_{r,i}) \leftrightarrow Y_i$$
- $p_{Y_Q|X_Q,X_{r,Q},Z_Q,S_Q,U_Q,Q} = p_{Y|X,X_r,Z,S}$
- $Z_Q = z(X_Q, X_{r,Q}, S_Q)$

The first condition follows since S^n is i.i.d. All other conditions can be derived from the factorization of the PMF:

$$\begin{aligned} &p(m,s^n,x^n,x_r^n,z^n,y^n) \\ &\quad = p(m)\prod_{i=1}^n p(s_i)\prod_{i=1}^n 1(x_i|m,s^n)1(x_{r,i}|z^{i-1})\times \\ &\qquad \times 1(z_i|x_i,x_{r,i},s_i)p_{Y|X,X_r,Z,S}(y_i|x_i,x_{r,i},z_i,s_i). \end{aligned}$$

The Markov chains can be readily proven by the factorization above. For each i, $Z_i = z(Z_i, X_{r,i}, S_i)$ and S^n is i.i.d., and therefore, the fourth condition also holds. To conclude, we have

$$\begin{aligned} R &\leq I(X,X_r;Y|S) + \epsilon_n \\ R &\leq I(X;Y|X_r,Z,U,S) + H(Z|X_r,U,S) - I(U;S) + \epsilon_n \end{aligned}$$

with a PMF that factorizes as

$$p_{U|S}\,p_{X_r|U}\,p_{X|X_r,U,S}$$

[3] $X_{r,i}$ is a function of Z^{i-1}, which is a function of X_{i-1}, S_{i-1} and $X_{r,i-1}$. Repeatedly, $X_{r,i-1}$ is a function of X_{i-2}, X_{i-2} and $X_{r,i-1}$ and so on.

that satisfies $I(U;S) \leq H(Z|X_r,U,S)$. This completes the proof for the converse part. ∎

VI. Proof for Theorem 2

Although the MAC setup differs from that of the SD-RC, there is some resemblance between the two. First, the deterministic link between the two encoders implies that Encoder 2 plays the role of the relay. In contrast to the SD-RC, however, in the MAC setup Z is a function only of X_1 and S_1 but not of X_2. Moreover, Encoder 2 now has its own message to send over the channel, and the nature of this message causes X_2 to inherit randomness. Lastly, the MAC setup entails an additional state component, S_2, which is also known at the decoder. Considering the differences and similarities, we will follow basically the same strategy we used with the SD-RC, with the following adaptation: we will account for the different factors by adding superbins that contain multiple cooperation codewords and by generating transmission sequences for M_2. The rest of this section contains the direct part for the proof. The converse part is given in appendix C.

Codebook: Using superposition block Markov coding, we set $m_1^{(1)} = m_2^{(1)} = m_1^{(B)} = m_2^{(B)} = 1$ and each block-codebook $\mathcal{C}_n^{(b)}$ is generated as follows:

– *Binning*: Partition the set $\mathcal{Z}^n$ into $2^{nR'_B}$ bins by uniformly and independently drawing an index
$$\text{bin}^{(b)}(z^n) \sim U[1:2^{nR_u}],\ \forall z^n \in \mathcal{Z}^n$$

– *Cooperation codewords*: Generate $2^{n(R'_B+R''_B)}$ u-codewords
$$u^n(l'^{(b-1)}, l''^{(b-1)}) \sim \prod_{i=1}^{n} p_U(u_i(l^{(b-1)})),$$
for $l'^{(b-1)} \in [1:2^{nR'_B}]$, $l''^{(b-1)} \in [1:2^{nR''_B}]$.

– *Cribbed codewords*: For each $l'^{(b-1)}, l''^{(b-1)}$ and $s_1^n \in \mathcal{S}_1^n$, generate $2^{n(R'_1+\tilde{R})}$ z-codewords,
$$z^n(m_1'^{(b)}, k^{(b)}|l'^{(b-1)}, l''^{(b-1)}, s_1^n) \sim \prod_{i=1}^{n} p_{Z|U,S}(z_i|u_i(l'^{(b-1)}, l''^{(b-1)}), s_{1,i})$$
for $m_1'^{(b)} \in [1:2^{nR'_1}]$, $k^{(b)} \in [1:2^{n\tilde{R}_1}]$.

– *Transmission codewords at Encoder 1*: For each $m_1'^{(b)}, k^{(b)}, l'^{(b-1)}, l''^{(b-1)}$ and $s_1^n \in \mathcal{S}_1^n$ generate $2^{nR''_1}$ codewords
$$x_1{}^n(m_1''^{(b)}|m_1'^{(b)}, k^{(b)}, l'^{(b-1)}, l''^{(b-1)}, s_1^n) \sim \prod_{i=1}^{n} p_{X_1|Z,U,S}(x_{1,i}|z_i(m_1'^{(b)}), u_i(l'^{(b)}, l''^{(b)}, s_{1,i})),$$
for $m_1''^{(b)} \in [1:2^{nR''_1}]$.

– *Transmission codewords at Encoder 2*: For each $l'^{(b-1)}, l''^{(b-1)}$ and $s_2^n \in \mathcal{S}_2^n$, draw 2^{nR_2} codewords
$$x_2^n(m_2^{(b)}|l'^{(b-1)}, l''^{(b-1)}, s_2^n) \sim \prod_{i=1}^{n} p_{X_2|U}(x_{2,i}|u_i^{(b)}(l'^{(b-1)}, l''^{(b-1)}), s_{2,i}),$$
for $m_2 \in [1:2^{nR_2}]$.

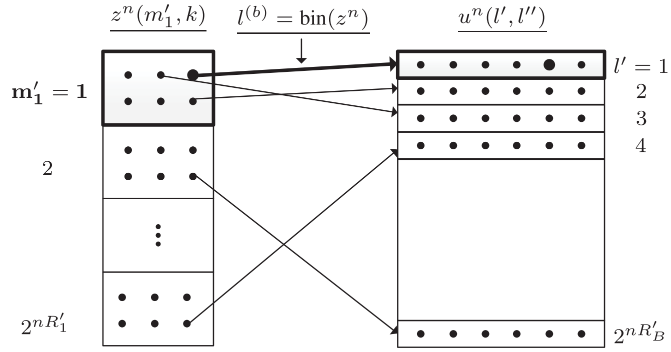

Fig. 9. Choosing a sequence z^n that points toward a bin containing a coordinated sequence u^n. The thick dots are the chosen sequences.

Encoder 1: Since this encoder can predict the cribbed sequence that the second encoder will observe, we perform bin selection by hashing it using the binning function. Instead of selecting the exact cooperation sequence from the collection of generated u-sequences, it selects $l'^{(b)}$, which is the index of the superbin. Given $m_1'^{(b)}$, it looks for $k^{(b)}$ s.t. there exists $\tilde{l}''^{(b)}$ that satisfies
$$\left(u^{n(b+1)}(\tilde{l}'^{(b)}, \tilde{l}''^{(b)}), s_1^{n(b+1)}\right) \in \mathcal{A}_\epsilon^{(n)}(p_{S_1,U}),$$
where $\tilde{l}'^{(b)} = \text{bin}(z^{n(b)}(m_1'^{(b)}, k^{(b)}|l'^{(b-1)}, l''^{(b-1)}, s_1^{n(b)}))$ is the superbin index. This procedure is illustrated in Figure. 9.

Encoder 2: At the end of each block $(b-1)$, the superbin index $l'^{(b-1)}$ is known from the cribbed sequence $z^{n(b-1)}$. To cooperate with first encoder, it first has to extract the exact cooperation u-sequence from the superbin. Hence, look for the first $\tilde{l}''^{(b-1)}$ s.t. $\left(u^{n(b)}(l'^{(b-1)}, l''^{(b-1)}), s_2^{n(b)}\right) \in \mathcal{A}_\epsilon^{(n)}(p_{S_2,U})$. Subsequently, in block b, the second encoder sends $x_2^n(m_2^{(b)}|l'^{(b-1)}, l''^{(b-1)}, s_2^n)$, which conveys the message it is required to send over the channel.

Decoder: The sliding window technique is suboptimal for the MAC setup.[4] Therefore, the decoding procedure is done backwards: we start decoding from block B to block 2. Note that in each block b, all transmissions are superimposed on a u-sequence that is defined by $l^{(b-1)} = (l'^{(b-1)}, l''^{(b-1)})$. When decoding at block b, assume that $l^{(b)} = (l'^{(b)}, l''^{(b)})$ is known from previous decoding operations. The decoder operates in three steps:

1) It seeks each superbin for sequences that are coordinated with $s_2^{n(b)}$. That is, for each $\tilde{l}'^{(b-1)} \in [1:2^{nR'_B}]$, it finds $\tilde{l}''^{(b-1)}(l'^{(b-1)}, s_2^{n(b)})$ the same way that Encoder 2 does.
2) Then, using the decoded $l^{(b)}$ from previous decoded block, it finds functions $\hat{m}_1'^{(b)}(l'^{(b-1)}, s_2^{n(b)})$ and $\hat{k}^{(b)}(\tilde{l}'^{(b-1)}, s_2^{n(b)})$ s.t. $\text{bin}(z^{n(b)}(\hat{m}_1'^{(b)}, \hat{k}^{(b)}|l'^{(b-1)}, \tilde{l}''^{(b-1)}, s_1^{n(b)})) = l^{(b)}$. If there are multiple functions that satisfy the above, choose one uniformly. Note that there are a total of $2^{nR'_B}$ tuples of functions, one for each $\tilde{l}'^{(b-1)} \in [1:2^{nR'_b}]$.

[4] In [24], Laneman and Kraner showed that in some cases of MAC (in contrary to SD-RC), the sliding window technique is inferior to backward decoding in terms of achievable rates.

$$\begin{aligned}\big(s_1^{n(b)}, s_2^{n(b)}, u^n(\hat{l}^{(b-1)}), x_1^n(\hat{m}_1^{\prime\prime(b)}|\hat{m}_1^{\prime(b)}, \hat{k}^{(b)}, \hat{l}^{(b-1)}, s_1^{n(b)}), z^n(\hat{m}_1', \hat{k}|\hat{l}^{(b-1)}, s_1^{n(b)}), x_2^n(\hat{m}_2^{(b)}|\hat{l}^{(b-1)}, s_2^{n(b)}), y^{n(b)}\big)\\ \in \mathcal{A}_\epsilon^{(n)}(p_{S_1,S_2} P_{U,X_1|S_1} p_{X_2|U,S_2} p_{Y,Z|X_1,X_2,S_1,S_2})\end{aligned} \tag{9}$$

3) Finally, it performs a typicality test by searching $(\hat{l}^{\prime(b-1)}, \hat{m}_1^{\prime\prime(b)}, \hat{m}_2^{(b)})$ that satisfy eq. (9), as shown at the top of this page.

Error analysis: Define the events

$$\mathrm{E}_1(b) = \left\{\begin{array}{l}\forall k^{(b)}:\\ (U^{n(b+1)}(\mathrm{Bin}^{(b)}(Z^n)), S_1^{n(b+1)}) \notin \mathcal{A}_\epsilon^{(n)}(p_{S,U})\\ Z^n = Z^n(1, k^{(b)}|L^{(b-1)}, S_1^{n(b)})\end{array}\right\}$$

$$\mathrm{E}_2(b) = \left\{(U^{n(b+1)}(L^{\prime(b)}, L^{\prime\prime(b)}), S_2^{n(b+1)}) \notin \mathcal{A}_\epsilon^{(n)}(p_{S_2,U})\right\}$$

$$\mathrm{E}_3(b) = \left\{\begin{array}{l}\exists l^{\prime\prime(b)} \neq L^{\prime\prime(b)}:\\ (U^{n(b)}(L^{\prime(b-1)}, l^{\prime\prime(b-1)}), S_2^{n(b)}) \in \mathcal{A}_\epsilon^{(n)}(p_{S_2,U})\end{array}\right\}$$

$$\mathrm{E}_4(b) = \left\{\begin{array}{l}\exists k, m_1^{\prime(b)} \neq 1:\\ \mathrm{Bin}^{(b)}(Z^n(m_1^{\prime(b)}, k|L^{(b-1)}, S_1^{n(b)})) = L^{(b)}\end{array}\right\}$$

$$\mathrm{E}_5(b) = \left\{\begin{array}{l}\text{Condition (9)is not satisfied by}\\ (\hat{l}^{\prime(b-1)}, \hat{m}_1^{\prime\prime(b)}, \hat{m}_2^{(b)}) = (L^{(b-1)}, 1, 1)\end{array}\right\}$$

$$\mathrm{E}_6(b) = \left\{\begin{array}{l}\text{Condition (9) is satisfied by some}\\ (\hat{l}^{\prime(b-1)}, \hat{m}_1^{\prime\prime(b)}, \hat{m}_2^{(b)}) \neq (L^{\prime(b-1)}, 1, 1)\end{array}\right\}$$

$$\mathrm{E}_7(b) = \{\hat{L}^{(b)} \neq L^{(b)}\}$$

Define two unions of error events by

$$\tilde{E}_1(b) = \bigcup_{j=1}^{b} \{\mathrm{E}_1(j) \cup \mathrm{E}_2(j) \cup \mathrm{E}_3(j)\}$$

$$\tilde{E}_2(b) = \bigcup_{j=b}^{B} \{\mathrm{E}_4(j) \cup \mathrm{E}_5(j) \cup \mathrm{E}_6(j) \cup \mathrm{E}_7(j)\}$$

where $\tilde{E}_1(b)$ is a union of encoding errors up to block b, and $\tilde{E}_2(b)$ are decoding errors down to block b (backwards). The average probability of an error is upper bounded by

$$\begin{aligned}P_e &= \mathbb{E}_{\mathcal{C}_n}[P_e(\mathcal{C}_n)]\\ &\leq \mathbb{P}\left[\tilde{E}_1(B) \cup \tilde{E}_2(1)\right]\\ &\leq \mathbb{P}\left[\tilde{E}_1(B)\right] + \mathbb{P}\left[\tilde{E}_2(1)|\tilde{E}_1^c(B)\right]\\ &\leq \sum_{b=1}^{B}\Big(\mathbb{P}[\mathrm{E}_1(b)] + \mathbb{P}[\mathrm{E}_2(b)|\mathrm{E}_1^c(b)] + \mathbb{P}[\mathrm{E}_3(b)] + \mathbb{P}[\mathrm{E}_4(b)]\\ &\quad + \mathbb{P}[\mathrm{E}_5(b)|\mathrm{E}_7^c(b+1), \tilde{E}_1^c(B)] + \mathbb{P}[\mathrm{E}_6(b)|\mathrm{E}_7^c(b+1), \tilde{E}_1^c(B)]\\ &\quad + \mathbb{P}[\mathrm{E}_7(b)|\mathrm{E}_5^c(b), \mathrm{E}_6^c(b), \mathrm{E}_7^c(b+1), \tilde{E}_1^c(B)]\Big)\end{aligned}$$

Without loss of generality, assume all messages $m_1^{\prime(b)}, m_1^{\prime\prime(b)}, m_2^{(b)}$ are equal to 1 for all $b \in [1 : B]$.

- $\mathbb{P}[\mathrm{E}_1(b)]$: According to lemma 1 and the covering lemma, this probability goes to zero for $n \to \infty$ if

$$\begin{aligned}\tilde{R} &\leq H(Z|U, S_1)\\ \tilde{R} &< R_B'\\ I(U; S_1) &< R_B'' + \tilde{R}\end{aligned} \tag{10a}$$

- $\mathbb{P}\left[\mathrm{E}_2(b)|\mathrm{E}_1^c(b)\right]$: by Markov lemma [25, Lemma 12.1], this probability also vanishes to zero.
- $\mathbb{P}[\mathrm{E}_3(b)]$: By setting

$$R_B'' < I(U; S_2), \tag{10b}$$

the packing lemma ensures that this probability will reduce to 0.
- $\mathbb{P}[\mathrm{E}_4(b)]$: Following similar derivation as in section V, this probability goes to zero by taking

$$\begin{aligned}R_1' + \tilde{R} &< H(Z|U, S_1)\\ R_1' + \tilde{R} &< R_B'\end{aligned} \tag{10c}$$

- $\mathbb{P}[\mathrm{E}_5(b)|\mathrm{E}_7^c(b+1), \tilde{E}_1^c(B)]$: By the conditional typicality lemma, this probability goes to 0 for $n \to \infty$.
- $\mathbb{P}[\mathrm{E}_6(b)|\mathrm{E}_7^c(b+1), \tilde{E}_1^c(B)]$: This event is bounded by the union of the following events:
  1) $\left(\hat{l}^{\prime(b-1)}, \hat{m}_1^{\prime\prime(b)}, \hat{m}_2^{(b)}\right) = (\neq L^{(b-1)}, *, *)$
  2) $\left(\hat{l}^{\prime(b-1)}, \hat{m}_1^{\prime\prime(b)}, \hat{m}_2^{(b)}\right) = (L^{(b-1)}, > 1, > 1)$
  3) $\left(\hat{l}^{\prime(b-1)}, \hat{m}_1^{\prime\prime(b)}, \hat{m}_2^{(b)}\right) = (L^{(b-1)}, 1, > 1)$
  4) $\left(\hat{l}^{\prime(b-1)}, \hat{m}_1^{\prime\prime(b)}, \hat{m}_2^{(b)}\right) = (L^{(b-1)}, > 1, 1)$

  A standard application of the packing lemma results in

$$\begin{aligned}R_B' + R_1'' + R_2 &< I(X_1, X_2; Y|S_1, S_2) + I(U; S_1)\\ R_1'' + R_2 &< I(X_1, X_2; Y|Z_1, U, S_1, S_2)\\ R_2 &< I(X_2; Y|X_1, U, S_1, S_2)\\ R_1'' &< I(X_1; Y|X_2, Z_1, U, S_1, S_2)\end{aligned} \tag{10d}$$

- $\mathbb{P}[\mathrm{E}_7(b)|\mathrm{E}_5^c(b), \mathrm{E}_6^c(b), \mathrm{E}_7^c(b+1), \tilde{E}_1^c(B)]$: If the preceding events did not occur, then the probability of this event is 0.

Performing FME on (10) yields

$$\begin{aligned}R_1 &< I(X_1; Y|Z, U, X_2, S_1, S_2)\\ &\quad + H(Z|U, S_1) - I(U; S_1|S_2)\\ R_2 &< I(X_2; Y|X_1, U, S_1, S_2)\\ R_1 + R_2 &< I(X_1, X_2; Y|U, Z, S_1, S_2)\\ &\quad + H(Z|U, S_1) - I(U; S_1|S_2)\\ R_1 + R_2 &< I(X_1, X_2; Y|S_1, S_2)\\ I(U; S_1|S_2) &< H(Z|U, S_1),\end{aligned}$$

for all PMFs that factorize as $p_{X_1,U|S_1} p_{X_2|U,S_2}$ and $Z = z(X_1, S_1)$. Note that $I(U; S_1) - I(U; S_2) = I(U; S_1|S_2)$ since $S_2 \leftrightarrow S_1 \leftrightarrow U$ forms a Markov chain. ■

VII. Proof for Theorem 3

The proof for this theorem relies heavily on the proofs from previous sections. The achievability part builds on the cooperative-bin-forward scheme from section VI by combining it with instantaneous relaying (a.k.a. Shannon strategy). To avoid unnecessary repetition, we only show the differences in the achievability part of the proof and the proofs for Markov chains in the converse.

Achievability: Codebook generation is done as in VI, with additional conditioning on Z when drawing $x_2^n(m_2^{(b)}|l^{(b-1)}, s_2^n)$. Namely, the codebook constructed for Encoder 2 is as follows. For each block b, $s_2^n \in \mathcal{S}_2^n$, $z \in \mathcal{Z}$ and $(l'^{(b-1)}, l''^{(b-1)})$, draw 2^{nR_2} codewords

$$x_2^n(m_2^{(b)}|z, u^n(l'^{(b-1)}, l''^{(b-1)}), s_2^n) \sim \prod_{i=1}^{n} p_{X_2|Z,U,S_2}(x_{2,i}|z, u_i(l'^{(b-1)}, l''^{(b-1)}), s_{2,i})$$

In each transmission block, Encoder 1 performs the same operations as before. Encoder 2 also performs the same operation, but at each time i, it transmits $x_{2,i}(m_2^{(b)}|z_i, u^n(l'^{(b-1)}, l''^{(b-1)}), s_2^n)$. The decoder performs backward decoding as before w.r.t. the new codebook. All other operations are preserved and the same error analysis holds. Likewise, the derivation results in the same achievable rate region but under the new PMF factorization $p_{U,X_1|S_1} p_{X_2|Z,U,S_2}$.

Converse: The only difference between the converse and the proof of the previous section is that here we need to show the PMF factorization and prove the new Markov chains. The rate bounds on R_1 and R_2 are the same and are obtained using identical arguments. Continuing the derivation from this point, we need to show that the following Markov chains hold

$$\begin{aligned} &S_{2,i} \leftrightarrow S_{1,i} \leftrightarrow U_i \\ &S_{2,i} \leftrightarrow (S_{1,i}, U_i) \leftrightarrow Z_i \\ &(S_{1,i}, X_{1,i}) \leftrightarrow (S_{2,i}, U_i, Z_i) \leftrightarrow X_{2,i} \end{aligned}$$

Note that now the PMF of the random variables is

$$\begin{aligned} p(m_1, &m_2, s_1^n, s_2^n, x_{1,i}, z^n, x_{2,i}) = \\ &= p(m_1)p(m_2)\left[\prod_{i=1}^{n} p(s_{1,i}, s_{2,i})\right] \times \\ &\quad \times 1(x_{1,i}, z^n|s_1^n, m_1) 1(x_{2,i}|z^i, s_2^n, m_2) \end{aligned}$$

Now $x_{2,i}$ is also a function of z_i and not only z^{i-1}. Therefore, the first two Markov-chains hold due to the same arguments in the previous section. As for the last Markov, consider

$$\begin{aligned} &p(m_1, m_2, s_1^n, s_2^n, x_{1,i}, z^n, x_{2,i}) = \\ &= p(m_1)p(m_2)\left[\prod_{i=1}^{n} p(s_{1,i}, s_{2,i})\right] \times \\ &\quad \times 1(x_{1,i}, z^n|s_1^n, m_1) 1(x_{2,i}|z^i, s_2^n, m_2) \\ &= p(s_1^{i-1}) p(s_{1,i}, s_{2,i}) p(s_{2,i+1}^n) p(s_2^{i-1}|s_1^{i-1}) p(s_{1,i+1}^n|s_{2,i+1}^n) \\ &\quad \times p(x_{1,i}, z^n, m_1|s_1^n) 1(x_{2,i}, m_2|z^i, s_2^n) \\ &= p(s_1^{i-1}) p(s_{1,i}, s_{2,i}) p(s_{2,i+1}^n) p(x_{1,i}, z^n, m_1, s_{1,i+1}^n|s_1^i, s_{2,i+1}^n) \\ &\quad \times p(x_{2,i}, m_2, s_2^{i-1}|z^i, s_{2,i}^n, s_1^{i-1}) \end{aligned}$$

Summing for $(m_1, m_2, , z_{i+1}^n, s_2^{i-1}, s_{1,i+1}^n)$ results in

$$\begin{aligned} &p(s_1^{i-1}) p(s_{1,i}, s_{2,i}) p(s_{2,i+1}^n) \times \\ &\qquad \times p(x_{1,i}, z^i, |s_1^i, s_{2,i+1}^n) p(x_{2,i}|z^i, s_{2,i}^n, s_1^{i-1}) \end{aligned}$$

in which $(S_{1,i}, X_{1,i}) \leftrightarrow (S_{2,i}, S_1^{i-1}, Z^{i-1}, S_{2,i+1}^n, Z_i) \leftrightarrow X_{2,i}$ is Markov. All other arguments regarding the memoryless property of the channel and the time-sharing random variable Q hold. This concludes the proof of Theorem 3. ■

VIII. Conclusions and Final Remarks

Using a variation of the cooperative-bin-forward scheme, we found the capacity of the SD-RC and MAC with partial cribbing when non-causal CSI is given to the decoder and one of the transmitters. Remarkably, in both setups only one auxiliary random variable is used to obtain the capacity region. One cooperation codeword is designated to play the role of creating cooperation and that of compression of the state sequence. It is also evident that in the special case of the MAC, the non-causal access to the state conferred states compression that, consequently, increased the capacity region.

The cooperative-bin-forward scheme relies heavily on the fact that the link for the cooperation, i.e., the link from the encoder to the relay (or the cribbed signal in the MAC), is deterministic. Since the transmitter can predict and dictate the observed output (by the relay), it can coordinate with the relay based on the same bin index. However, it is not known how the cooperative-bin-forward scheme can be generalized to cases in which the link between the encoder and the relay is a general noisy link.

Appendix A
Proofs for Special Cases of MAC

The special cases in section III are captured by Theorem 2. We restate here the region for the case of *one state component* as a reference for the following derivations. The capacity region for discrete memoryless MAC with non-causal CSI in Fig. 2 with one state is given by the set of rate pairs (R_1, R_2) that satisfies

$$\begin{aligned} R_1 &\le I(X_1; Y|X_2, Z, S, U) + H(Z|S, U) - I(U; S) \\ R_2 &\le I(X_2; Y|X_1, S, U) \\ R_1 + R_2 &\le I(X_1, X_2; Y|Z, S, U) + H(Z|S, U) - I(U; S) \\ R_1 + R_2 &\le I(X_1, X_2; Y|S) \end{aligned} \tag{11a}$$

for PMFs of the form $p_{U|S} p_{X_1|S,U} p_{X_2|U}$, with $Z = z(X_1, S)$, that satisfies

$$I(U; S) \le H(Z|U, S), \tag{11b}$$

Case A: Multiple Access Channel with states (without cribbing): This case is captured by Theorem 2 by setting $z(x_1, s) = 0$, $\forall x_1 \in \mathcal{X}_1, s \in \mathcal{S}$ because this configuration lacks cribbing between the encoders. The inequality in (11b) results in $I(U;S) \leq 0$, which enforces U to be independent of S. Thus, the region in (11) becomes

$$\begin{aligned} R_1 &\leq I(X_1;Y|S,U,X_2) \\ R_2 &\leq I(X_2;Y|S,U,X_1) \\ R_1+R_2 &\leq I(X_1,X_2;Y|S,U) \\ R_1+R_2 &\leq I(X_1,X_2;Y|S), \end{aligned} \tag{12}$$

with a PMF of the form $p_U p_{X_1|U,S} p_{X_2|U}$. Note that $U \leftrightarrow (X_1,X_2,S) \leftrightarrow Y$ forms a Markov chain. Therefore, the last inequality is redundant. It also implies that the capacity region in (12) is outer bounded by (2); removing U achieves that outer bound.

Case B: State-dependent MAC with partially cooperating encoders: We investigate the capacity region for the case of the orthogonal cooperation link and channel transmission (Fig. 4). The cooperation link here is *strictly causal* due to the cribbing, i.e., $X_{2,i} = f(M_2, X_{1,p}^n)$. First, note that the region in (3) is an *outer* bound, since it is the capacity region of non-causal cooperation, i.e., when $X_{2,i} = f(M_2, X_{1,p}^n)$. The strictly causal configuration is captured by the cribbing setup when setting $X_1 = (X_{1c}, X_{1p})$, $Z = X_{1,p}$ and the channel transition PMF to $p_{Y|X_{1c},X_2,S}$. Then, the region in (11) becomes

$$\begin{aligned} R_1 &\leq I(X_{1c};Y|S,X_{1p},U,X_2)+H(X_{1p}|U,S)-I(U;S) \\ R_2 &\leq I(X_2;Y|S,U,X_{1c},X_{1p}) \\ R_1+R_2 &\leq I(X_{1c},X_2;Y|S,U,X_{1p})+H(X_{1p}|U,S)-I(U;S) \\ R_1+R_2 &\leq I(X_{1c},X_{1p},X_2;Y|S) \end{aligned} \tag{13a}$$

for PMFs of the form $p_{U|S} p_{X_1|U,S} p_{X_2|U}$ that satisfies

$$I(U;S) \leq H(X_{1c}|U,S). \tag{13b}$$

Note that $I(X_{1c},X_{1p},X_2;Y|S) = I(X_{1c},X_2;Y|S)$ because $X_{1p} \leftrightarrow (X_{1c},X_2,S) \leftrightarrow Y$ is a Markov chain. We identify the rate $H(X_{1p}|U,S)$ as the cooperation rate R_{12}. Let $p_{X_1|U,S} = p_{X_{1p}|U,S} p_{X_{1c}|U,S}$, and $p_{X_{1p}|U=u,S=s}$ be a uniform distribution for every $(u,s) \in \mathcal{U} \times \mathcal{S}$. By doing so, $H(X_{1p}|U,S) = \log_2|X_{1p}|$ and $I(X_1;Y|X_2,U,S,X_{1p}) = I(X_{1c};Y|X_2,U,S)$. The latter holds since $X_{1p} \leftrightarrow (X_{1c},X_2,S) \leftrightarrow Y$ is a Markov chain and X_{1c} is independent of X_{1p}. By denoting $R_{12} = \log_2|\mathcal{X}_{1p}|$, the regions in (3) and (13) coincide.

Case C: Point-to-point with non-causal CSI: First, note that the channel depends only on X_1 and S. Encoder 1 sends a message over the channel, and the states are revealed to it non-causally at the beginning of the transmission. Encoder 2, however, has no message to send; in fact, it cannot send anything over the channel since the channel's output is not affected by X_2 at all. Therefore, the rate R_2 is 0. This configuration is captured by the MAC when

$$\begin{aligned} R_2 &= 0 \\ p_{Y|X_1,X_2,S} &= p_{Y|X_1,S}. \end{aligned} \tag{14}$$

Inserting (14) into Theorem 2 derives with

$$\begin{aligned} R_1 &\leq I(X_1;Y|S,U,Z,X_2)+H(Z|U,S)-I(U;S) \\ R_1 &\leq I(X_1,X_2;Y|S,U,Z)+H(Z|U,S)-I(U;S) \\ R_1 &\leq I(X_1,X_2;Y|S) \end{aligned} \tag{15a}$$

with $p_{S,U,X_1} p_{X_2|U} p_{Z,Y|X_1,S}$ that satisfies

$$I(U;S) \leq H(Z|U,S). \tag{15b}$$

Due to the Markov chains $X_2 \leftrightarrow (X_1,S) \leftrightarrow Y$, $X_2 \leftrightarrow (X_1,U,S,Z) \leftrightarrow Y$ and $X_2 \leftrightarrow (U,S,Z) \leftrightarrow Y$, the following identities hold

$$\begin{aligned} I(X_1,X_2;Y|S) &= I(X_1;Y|S) \\ I(X_1,X_2;Y|S,U,Z) &= I(X_1;Y|S,U,Z,X_2) \\ I(X_1,X_2;Y|S,U,Z) &= I(X_1;Y|U,Z,S) \end{aligned}$$

Therefore, the region in (15) reduces to

$$\begin{aligned} R_1 &\leq I(X_1;Y|S,U,Z)+H(Z|U,S)-I(U;S) \\ R_1 &\leq I(X_1;Y|S) \\ I(U;S) &\leq H(Z|U,S) \end{aligned}$$

This region is smaller than or equal to (4); if we drop the first and last inequalities, we get the expression for capacity.[5] Taking U to be constant results in $I(U;S) = 0$, which makes the last inequality redundant. The first inequality is also redundant because

$$\begin{aligned} &I(X_1;Y|S,Z)+H(Z|S) = \\ &= I(X_1,Z;Y|S)-I(Z;Y|S)+H(Z|S) \\ &= I(X_1,Z;Y|S)+H(Z|Y,S) \\ &= I(X_1;Y|S)+H(Z|Y,S), \end{aligned}$$

which is larger than the right-hand side of the second inequality.

Point-to-point with state encoder and output causality constraint: This configuration is captured by the MAC with cribbing by setting

$$\begin{aligned} R_1 &= 0 \\ p_{Y|X_1,X_2,S} &= p_{Y|X_2,S} \\ z(x_1,s) &= x_1. \end{aligned}$$

The region in (11) reduces to

$$\begin{aligned} R_2 &\leq I(X_2;Y|S,U,X_1) \\ R_2 &\leq I(X_2;Y|S,U,X_1)+H(X_1|U,S)-I(U;S) \\ R_2 &\leq I(X_1,X_2;Y|S) \\ I(U;S) &\leq H(X_1|U,S). \end{aligned}$$

with $p_{U,X_1|S} p_{X_2|U} p_{Y|X_2,S}$. Notice that $I(X_2;Y|S,U,X_1) \leq I(X_1,X_2,U;Y|S)$, and both $(U,X_1) \leftrightarrow (X_2,S) \leftrightarrow Y$ and $X_1 \leftrightarrow (X_2,S) \leftrightarrow Y$ are Markov chains. Therefore, the third inequality is redundant. Moreover, from the constraint $I(U;S) \leq H(X_1|U,S)$, it follows that $I(X_2;Y|S,U,X_1) \leq$

[5]The expressions for the capacity after dropping the constraints are not exactly the same, since the PMF domains are different. However, the capacities coincide due to the objective and maximization.

$I(X_2;Y|U,X_1,S)+H(X_1|U,S)-I(U;S)$; thus, the second inequality is also redundant. The Markov chains $Y \leftrightarrow (U,S) \leftrightarrow X_1$ and $Y \leftrightarrow (X_2,S,U) \leftrightarrow X_1$ imply that $I(X_2;Y|S,U,X_1)=I(X_2;Y|S,U)$. Therefore, the region is further reduced to

$$\begin{aligned} R_2 &\le I(X_2;Y|S,U)\\ I(U;S) &\le H(X_1|U,S). \end{aligned}$$

Note that $H(X_1|U,S) \le \log_2|\mathcal{X}_1|$, and therefore, this region is upper bounded by the capacity. By taking $p_{X_1|U,S}$ to be uniform distribution for every $(u,s)\in\mathcal{U}\times\mathcal{S}$, the conditional entropy $H(X_1|U,S)$ is equal to $\log_2|\mathcal{X}_1|$ and we achieve the capacity.

Appendix B
Proof for Indirect Covering Lemma

We bound the following probability

$$P_e^{(n)} \triangleq \mathbb{P}\Big[|\{l:\exists k \text{ s.t. } \mathrm{Bin}(Z^n(k))=l\}|<2^{n(R-\delta_n)}|V^n=v^n\Big] \le \Delta_n$$

and ensure that both δ_n and Δ_n go to zero as n goes to infinity.

Assume $v^n \in \mathcal{A}_\epsilon^{(n)}(p_V)$ and recall that $\mathrm{Bin}(z^n) \overset{\text{i.i.d}}{\sim} \mathrm{U}[1:2^{nR_B}]$. Thus,

$$\begin{aligned} &\mathbb{P}\Big[\{Z^n(k)\}=\{z^n(k)\}_k, \{\mathrm{Bin}(Z^n(k))\}=\{\mathrm{bin}(z^n(k))\}_k|V^n=v^n\Big]\\ &\qquad = \prod_{k=1}^{2^{nR}} p^n_{Z|V}(z^n(k)|v^n)2^{-nR_B} \end{aligned}$$

where $p^n_{Z|V}(z^n(k)|v^n)=\prod_{i=1}^n p_{Z|V}(z_i(k)|v_i)$.

Define the sets

$$\begin{aligned} \mathcal{D}_1 &\triangleq \Big\{k : \big(Z^n(k),v^n\big)\in\mathcal{A}_{\epsilon'}^{(n)}(p_{Z,V}|v^n)\Big\}\\ \mathcal{D}_2 &\triangleq \Big\{k : Z^n(k)\neq Z^n(j), \forall j\neq k \text{ and } k,j\in\mathcal{D}_1\Big\}\\ \mathcal{D}_3 &\triangleq \Big\{k : \mathrm{Bin}\big(Z^n(k)\big)\neq \mathrm{Bin}\big(Z^n(j)\big), \forall j\neq k \text{ and } k,j\in\mathcal{D}_2\Big\} \end{aligned}$$

and the events

$$\begin{aligned} \mathrm{E}_1 &\triangleq |\mathcal{D}_1| < 2^{n(R-\delta_n^{(1)})}\\ \mathrm{E}_2 &\triangleq |\mathcal{D}_2| < 2^{n(R-\delta_n^{(2)})}\\ \mathrm{E}_3 &\triangleq |\mathcal{D}_3| < 2^{n(R-\delta_n^{(3)})} \end{aligned}$$

By the definition of E_3 and the law of total probability, it follows that

$$\begin{aligned} P_e^{(n)} &\le \mathbb{P}\left[\mathrm{E}_3|V^n=v^n\right]\\ &\le \underbrace{\mathbb{P}\left[\mathrm{E}_1|V^n=v^n\right]}_{(1)} + \underbrace{\mathbb{P}\left[\mathrm{E}_2|\mathrm{E}_1^c, V^n=v^n\right]}_{(2)}\\ &\quad + \underbrace{\mathbb{P}\left[\mathrm{E}_3|\mathrm{E}_2^c,\mathrm{E}_1^c, V^n=v^n\right]}_{(3)} \end{aligned}$$

We bound each probability separately.

1) Define $\theta_k = \mathbb{1}\Big[(Z^n(k),v^n)\in\mathcal{A}_{\epsilon'}^{(n)}(p_{Z,V})\Big]$, and note that $\theta_k \overset{\text{i.i.d}}{\sim} \mathrm{Bernoully}(\rho_n)$, where $1-\tilde{\delta}_n \le \rho_n \le 1$ and $\tilde{\delta}_n \to 0$ as $n\to\infty$. Therefore, for any $\delta'>0$ we have

$$\begin{aligned} \mathbb{P}\left[\mathrm{E}_1|V^n=v^n\right] &= \mathbb{P}\left[|\mathcal{D}_1| < 2^{n(R-\delta_n^{(1)})}|V^n=v^n\right]\\ &\overset{(a)}{=} \mathbb{P}\left[\Sigma_{k=1}^{2^{nR}}\theta_k < 2^{nR}\rho_n(1-\delta')|V^n=v^n\right]\\ &\overset{(b)}{\le} 2^{-2^{nR}\rho_n\delta'^2/2}\\ &= \Delta_n^{(1)} \end{aligned}$$

where:
(a) - by setting $\delta_n^{(1)} \triangleq -\frac{1}{n}\log_2\left(\rho_n(1-\delta')\right) \xrightarrow[n\to\infty]{} 0$,
(b) - by Chernoff's inequality [25, Appendix B], and $\Delta_n^{(1)} \to 0$ as $n\to\infty$.

2) First, note that given E_1^c, we have that with a probability of one, $|\mathcal{D}_1| > 2^{n(R-\delta_n^{(1)})}$. Define the *normalized* amount of *non-unique* sequences in $\mathcal{D}_1$,

$$C_2 = \frac{1}{|\mathcal{D}_1|}\sum_{k\in\mathcal{D}_1}\mathbb{1}\left[\exists j\neq k : Z^n(j)=Z^n(k), \in\mathcal{D}_1\right]$$

By this definition, it follows that $|\mathcal{D}_2| = |\mathcal{D}_1|(1-C_2)$. First, we bound the expected value of C_2 by

$$\begin{aligned} &\mathbb{E}[C_2|V^n=v^n,\mathrm{E}_1^c] =\\ &= \sum_{d_1}\mathbb{P}[\mathcal{D}_1=d_1|V^n=v^n,\mathrm{E}_1^c]\mathbb{E}[C_2|V^n=v^n,\mathrm{E}_1^c,\mathcal{D}_1=d_1]\\ &= \sum_{d_1:|d_1|>2^{n(R-\delta_n^{(1)})}}\mathbb{P}[\mathcal{D}_1=d_1|V^n=v^n,\mathrm{E}_1^c]\frac{1}{|d_1|}\times\\ &\qquad \sum_{k\in d_1}\mathbb{P}[\exists j\neq k : Z^n(j)=Z^n(k), j\in\mathcal{D}_1|V^n=v^n,\mathrm{E}_1^c]\\ &\le \sum_{d_1:|d_1|>2^{n(R-\delta_n^{(1)})}}\mathbb{P}[\mathcal{D}_1=d_1|V^n=v^n,\mathrm{E}_1^c]\frac{1}{\cancel{|d_1|}}\sum_{k\in d_1}\cancel{|d_1|}2^{-n(H(Z|V)-\epsilon')}\\ &= \sum_{d_1:|d_1|>2^{n(R-\delta_n^{(1)})}}\mathbb{P}\left[\mathcal{D}_1=d_1|V^n=v^n,\mathrm{E}_1^c\right]|d_1|2^{-n(H(Z|V)-\epsilon'+\delta_n^{(1)})}\\ &\le 2^{n(R-H(Z|V)+\epsilon'+\delta_n^{(1)})} \end{aligned}$$

Therefore, for any $\gamma_1'>0$ it follows by Markov's inequality that

$$\begin{aligned} \mathbb{P}\left[C_2 > 2^{-n\gamma_1'}|E_1^c, V^n=v^n\right] &\le 2^{n(R-H(Z|V)+\delta_n^{(1)}+\epsilon'+\gamma_1')}\\ &= \Delta_n^{(2)} \end{aligned}$$

where $\Delta_n^{(2)}\to 0$ as $n\to\infty$ if $R<H(Z|V)-\gamma_1$, $\gamma_1=\delta_n^{(1)}+\epsilon'+\gamma_1'$. By setting $\delta_n^{(2)}=\delta_n^{(1)}-\frac{1}{n}\log_2\left(1-2^{-n\gamma_1'}\right)$, we have

$$\mathbb{P}\left[E_2|E_1^c, V^n=v^n\right] \le \Delta_2^{(n)}.$$

and $\delta_n^{(2)}\to 0$ as $n\to\infty$.

3) We follow similar arguments as those for the previous bound. Define

$$C_3 = \frac{1}{|\mathcal{D}_2|}\sum_{k\in\mathcal{D}_2}\mathbb{1}\left[\exists j\neq k : \mathrm{Bin}(Z^n(j))=\mathrm{Bin}(Z^n(k)), j\in\mathcal{D}_2\right]$$

and recall that the probability of each bin index is independent of the realization of $\{Z^n(k)\}_k$. It follows that

$$\mathbb{E}\left[C_3|\mathrm{E}_2^c, \mathrm{E}_1^c, V^n = v^n\right] \leq 2^{n(R-R_B+\delta_n^{(2)})}.$$

By Markov's inequality, for any $\gamma_2' > 0$

$$\begin{aligned}
&\mathbb{P}[\mathrm{E}_3|\mathrm{E}_2^c, \mathrm{E}_1^c, V^n = v^n] = \\
=&\mathbb{P}[|\mathcal{D}_3| < 2^{n(R-\delta_n^{(3)})}|\mathrm{E}_2^c, \mathrm{E}_1^c, V^n = v^n] \\
=&\mathbb{P}[C_3 > 2^{-n\gamma_2'}|\mathrm{E}_2^c, \mathrm{E}_1^c, V^n = v^n] \\
\leq&2^{n(R-R_B+\delta_n^{(2)}+\gamma_2')} \\
=&\Delta_n^{(3)}
\end{aligned}$$

where $\Delta_n^{(3)} \to 0$ and $\delta_n^{(3)} = \delta_n^{(2)} - \frac{1}{n}(1-2^{-n\gamma_2'}) \to 0$ as $n \to \infty$, if $R < H(Z|V) - \gamma_2$ where $\gamma_2 = \delta_n^{(2)} + \gamma_2'$.

Finally, for any $\gamma_1, \gamma_2 >$ and a sufficiently large n, if

$$\begin{aligned}
R &< H(Z|V) - \gamma_1 \\
R &< H(Z|V) - \gamma_2
\end{aligned}$$

then

$$P_e^{(n)} \leq \Delta_n^{(1)} + \Delta_n^{(2)} + \Delta_n^{(3)}$$

where $\delta_n^{(i)}, \Delta_n^{(i)}$ tends to 0 when $n \to \infty$ for $i = 1, 2, 3$. ■

Appendix C
Converse for MAC

Let $U_i \triangleq (Z^{i-1}, S_1^{i-1}, S_{2,i+1}^n)$. We have

$$\begin{aligned}
&H(Z^n|S_1^n, S_2^n) = \\
=&H(Z^n, S_1^n, S_2^n) - H(S_1^n, S_2^n) \\
\overset{(a)}{=}&\sum_{i=1}^n \left[H(Z_i, S_{1,i}, S_{2,i}|Z^{i-1}, S_1^{i-1}, S_{2,i+1}^n) - H(S_{1,i}, S_{2,i})\right] \\
\overset{(b)}{=}&\sum_{i=1}^n \left[H(Z_i|S_{1,i}, S_{2,i}, U_i) + H(S_{1,i}, S_{2,i}|U_i) - H(S_{1,i}, S_{2,i})\right] \\
=&\sum_{i=1}^n \left[H(Z_i|S_{1,i}, S_{2,i}, U_i) - I(U_i; S_{1,i}, S_{2,i})\right] \\
\leq&\sum_{i=1}^n \left[H(Z_i|S_{1,i}, S_{2,i}, U_i) - I(U_i; S_{1,i}|S_{2,i})\right] \\
\overset{(c)}{=}&n\left[H(Z_Q|S_{1,Q}, S_{2,Q}, U_Q, Q) - I(U_Q; S_{1,Q}|S_{2,Q}, Q)\right]
\end{aligned}$$

where (a) follows since S_1^n and S_2^n are drawn i.i.d. in pairs, (b) follows by our definition of U_i and (c) is derived by setting $Q \sim \mathrm{U}[1:n]$ to be a time sharing random variable. Note that the following Markov chains hold:

$$\begin{aligned}
&S_{2,i} \leftrightarrow S_{1,i} \leftrightarrow U_i \\
&S_{2,i} \leftrightarrow (S_{1,i}, U_i) \leftrightarrow Z_i \\
&(S_{1,i}, X_{1,i}) \leftrightarrow (S_{2,i}, U_i) \leftrightarrow X_{2,i}
\end{aligned} \tag{16}$$

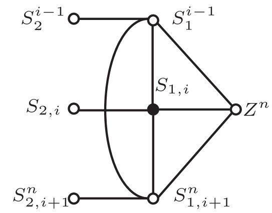

Fig. 10. Proof for Markov chains $S_{2,i} \leftrightarrow S_{1,i} \leftrightarrow U_i$ and $S_{2,i} \leftrightarrow (S_{1,i}, U_i) \leftrightarrow Z_i$ using an undirected graphical technique [27]. The undirected graph corresponds to the PMF $p(s_1^n, s_2^n, z^n) = p(s_1^{i-1}, s_2^{i-1})p(s_{1,i}, s_{2,i})p(s_{1,i+1}^n, s_{2,i+1}^n)p(z^n|s_1^n)$. The Markov chains follow since all paths from $S_{2,i}$ to all other nodes go through $S_{1,i}$.

Recall that the PMF on $(m_1, m_2, s_1^n, s_2^n, x_{1,i}, z^n, x_{2,i})$ is

$$\begin{aligned}
&p(m_1, m_2, s_1^n, s_2^n, x_{1,i}, z^n, x_{2,i}) = \\
&\quad= p(m_1)p(m_2)\left[\prod_{i=1}^n p(s_{1,i}, s_{2,i})\right] \times \\
&\quad\quad \times 1(x_{1,i}, z^n|s_1^n, m_1)1(x_{2,i}|z^{i-1}, s_2^n, m_2)
\end{aligned}$$

Note that since X_1^n is a deterministic function of (M_1, S_1^n), so is Z^n. Therefore, the Markov chain $(S_{1,i}, X_{1,i}) \leftrightarrow (S_{2,i}, U_i) \leftrightarrow X_{2,i}$ is readily proven from the PMF. As for the other Markovs in (16), we use an undirected graphical technique in Figure 10. It is also straightforward to show that $S_{2,Q} \leftrightarrow (S_{1,Q}, U_Q, Q) \leftrightarrow Z_Q$ holds. Therefore,

$$\begin{aligned}
&H(Z^n|S_1^n, S_2^n) = \\
=&n\left[H(Z_Q|S_{1,Q}, S_{2,Q}, U_Q, Q) - I(U_Q; S_{1,Q}|S_{2,Q}, Q)\right] \\
=&n\left[H(Z_Q|S_{1,Q}, U_Q, Q) - I(U_Q; S_{1,Q}|S_{2,Q}, Q)\right]
\end{aligned} \tag{17}$$

Note that due to this identity, $I(U_Q; S_{1,Q}|S_{2,Q}, Q) \leq H(Z_Q|S_{1,Q}, U_Q, Q)$. We proceed to bound R_1 and R_2. Note that by Fano's inequality,

$$H(M_1, M_2|Y^n, S_1^n, S_2^n) \leq n\epsilon_n$$

where $\epsilon_n \to 0$ when $n \to \infty$. Bounding R_1 yields

$$\begin{aligned}
nR_1 =&H(M_1) \\
\overset{(a)}{=}&H(M_1|S_1^n, S_2^n) \\
\overset{(b)}{=}&H(M_1, Z^n|S_1^n, S_2^n) \\
=&H(M_1|Z^n, S_1^n, S_2^n) + H(Z^n|S_1^n, S_2^n) \\
\overset{(c)}{=}&H(M_1|Z^n, S_1^n, S_2^n, M_2) + H(Z^n|S_1^n, S_2^n) \\
\leq&I(M_1; Y^n|Z^n, S_1^n, S_2^n, M_2) + H(Z^n|S_1^n, S_2^n) + n\epsilon_n
\end{aligned}$$

where:
(a) - follows since $M_1 \perp\!\!\!\perp (S_1^n, S_2^n)$
(b) - follows since $Z^n = f(M_1, S_1^n)$,
(c) - follows since $M_2 \perp\!\!\!\perp (M_1, Z^n, S_1^n, S_2^n)$. It follows that

$$\begin{aligned}
&I(M_1; Y^n|Z^n, S_1^n, S_2^n, M_2) = \\
&\quad= \sum_{i=1}^n I(M_1; Y_i|Y^{i-1}, Z^n, S_1^n, S_2^n, M_2) \\
&\quad\overset{(d)}{=} \sum_{i=1}^n I(M_1, X_{1,i}; Y_i|X_{2,i}, Y^{i-1}, Z^n, S_1^n, S_2^n, M_2)
\end{aligned}$$

$$\overset{(e)}{\leq}\sum_{i=1}^{n} I(X_{1,i};Y_i|X_{2,i},Z^i,S_1^i,S_{2,i}^n)$$
$$=\sum_{i=1}^{n} I(X_{1,i};Y_i|X_{2,i},Z_i,S_{1,i},S_{2,i},U_i)$$
$$=nI(X_{1,Q};Y_Q|X_{2,Q},Z_Q,S_{1,Q},S_{2,Q},U_Q,Q)$$

where (d) follows since $X_{1,i}$ is a function of (M_1, S_1^n) and (e) follows by moving $(M_2, Y^{i-1}, Z_{i+1}^n, S_{1,i+1}^n, S_2^{i-1})$ from the conditioning to the left-hand side of the mutual information. Since the channel is memoryless and without feedback, $(M_2, Y^{i-1}, Z_{i+1}^n, S_{1,i+1}^n, S_2^{i-1}) \leftrightarrow (X_{1,i}, X_{2,i}, S_{1,i}, S_{2,i}) \leftrightarrow Y_i$ holds.

We derive with the bound

$$R_1 \leq n[I(X_{1,Q};Y_Q|X_{2,Q},Z_Q,S_{1,Q},S_{2,Q},Q)$$
$$+H(Z_Q|S_{1,Q},U_Q,Q)-I(U_Q;S_{1,Q}|S_{2,Q})+\epsilon_n]$$

Following similar steps, we have

$$nR_2 = H(M_2)$$
$$=H(M_2|S_1^n,S_2^n,M_1)$$
$$\leq I(M_2;Y^n|S_1^n,S_2^n,M_1)+n\epsilon_n$$
$$=\sum_{i=1}^{n} I(M_2;Y_i|Y^{i-1},S_1^n,S_2^n,M_1)+n\epsilon_n$$
$$=\sum_{i=1}^{n} I(M_2,X_{2,i};Y_i|Y^{i-1},S_1^n,S_2^n,M_1,X_{1,i})+n\epsilon_n$$
$$\leq\sum_{i=1}^{n} I\left(\begin{matrix}Y^{i-1},S_{1,i+1}^n,S_2^{i-1}\\ M_1,M_2,X_{2,i};Y_i\end{matrix}\middle|S_1^i,S_{2,i}^n,X_{1,i}\right)+n\epsilon_n$$
$$=\sum_{i=1}^{n} I(X_{2,i};Y_i|S_1^i,S_{2,i}^n,X_{1,i})+n\epsilon_n$$
$$=nI(X_{2,Q};Y_Q|X_{1,Q},S_{1,Q},S_{2,Q},Q)+n\epsilon_n$$

The sum-rate R_1+R_2 is upper bounded by

$$n(R_1+R_2)=H(M_1)+H(M_2)$$
$$=H(M_1,M_2)$$
$$=H(M_1,M_2|S_1^n,S_2^n)$$
$$=H(M_1,M_2,Z^n|S_1^n,S_2^n)$$
$$=H(M_1,M_2|Z^n,S_1^n,S_2^n)+H(Z^n|S_1^n,S_2^n)$$
$$\leq I(M_1,M_2;Y^n|Z^n,S_1^n,S_2^n)+H(Z^n|S_1^n,S_2^n)+n\epsilon_n$$

where

$$I(M_1,M_2;Y^n|Z^n,S_1^n,S_2^n)=$$
$$=\sum_{i=1}^{n} I(M_1,M_2;Y_i|Y^{i-1},Z^n,S_1^n,S_2^n)$$
$$\leq\sum_{i=1}^{n} I\left(\begin{matrix}M_1,M_2,S_{1,i+1}^n,S_2^{i-1},Y^{i-1}\\ Z^{i-1},X_{1,i},X_{2,i};Y_i\end{matrix}\middle|Z^i,S_1^i,S_{2,i}^n\right)$$
$$=\sum_{i=1}^{n} I(X_{1,i},X_{2,i};Y_i|Z_i,S_{1,i},S_{2,i},U_i)$$
$$=nI(X_{1,Q},X_{2,Q};Y_Q|Z_Q,S_{1,Q},S_{2,Q},U_Q,Q)$$

and therefore, it follows from the identity in (17) and the above that

$$n(R_1+R_2)\leq n\big[I(X_{1,Q},X_{2,Q};Y_Q|Z_Q,S_{1,Q},S_{2,Q},U_Q,Q)+$$
$$H(Z_Q|S_{1,Q},U_Q,Q-I(U_Q;S_{1,Q}|S_{2,Q}))+\epsilon_n\big]$$

and the second upper bound by:

$$n(R_1+R_2)=H(M_1,M_2)$$
$$=H(M_1,M_2|S_1^n,S_2^n)$$
$$=H(M_1,M_2,Z^n|S_1^n,S_2^n)$$
$$\leq I(M_1,M_2;Y^n|S_1^n,S_2^n)+n\epsilon_n$$
$$=\sum_{i=1}^{n} I(M_1,M_2;Y_i|Y^{i-1},S_1^n,S_2^n)+n\epsilon_n$$
$$\leq\sum_{i=1}^{n} I(M_1,M_2,Y^{i-1},S_1^{n\backslash i},S_2^{n\backslash i},X_{1,i},X_{2,i};Y_i|S_{1,i},S_{2,i})+n\epsilon_n$$
$$=\sum_{i=1}^{n} I(X_{1,i},X_{2,i};Y_i|S_{1,i},S_{2,i})+n\epsilon_n$$
$$=n\left(I(X_{1,Q},X_{2,Q};Y_Q|S_{1,Q},S_{2,Q},Q)+\epsilon_n\right)$$
$$\leq n\left(I(X_{1,Q},X_{2,Q};Y_Q|S_{1,Q},S_{1,Q})+\epsilon_n\right)$$

where the last inequality is due to the Markov chain $Q \leftrightarrow (X_{1,Q}, X_{2,Q}, S_Q) \leftrightarrow Y_Q$. Thus, we obtain the following region

$$R_1 \leq I(X_{1,Q};Y_Q|X_{2,Q},Z_Q,S_{1,Q},S_{2,Q},Q)$$
$$+H(Z_Q|S_{1,Q},U_Q,Q)-I(U_Q;S_{1,Q}|S_{2,Q})$$
$$R_2 \leq I(X_{2,Q};Y_Q|X_{1,Q},S_{1,Q},S_{2,Q},Q)$$
$$R_1+R_2 \leq I(X_{1,Q},X_{2,Q};Y_Q|Z_Q,S_{1,Q},S_{2,Q},U_Q,Q)$$
$$+H(Z_Q|S_{1,Q},U_Q,Q)-I(U_Q;S_{1,Q}|S_{2,Q})$$
$$R_1+R_2 \leq I(X_{1,Q},X_{2,Q};Y_Q|S_{1,Q},S_{1,Q})$$
$$0 \leq H(Z_Q|S_{1,Q},U_Q,Q)-I(U_Q;S_{1,Q}|S_{2,Q},Q)$$

for PMFs of the form

$$p(q)p_{S_1,S_2}(s_{1,q},s_{2,q})p(u_q,x_{1,q}|s_{1,q},q)p(x_{2,q}|u_q,s_{2,q})\times$$
$$\times\, p_{Y|X_1,X_2,S_1,S_2}(y_q|x_{1,q},x_{2,q},s_{1,q},s_{2,q}).$$

The rest of the proof (regarding the removal of the time sharing random variable Q) is straight-forward using similar arguments to those employed the SD-RC in section V-B. Therefore, by letting $U=(U_Q,Q), X_1=X_{1,Q}, X_2=X_{2,Q}, Y=Y_Q, Z=Z_Q, S_1=S_{1,Q}$ and $S_{2,Q}$, we obtain the capacity region in Theorem 2. The upper bound of the cardinality of the auxiliary variable U is obtained by the convex cover method. ■

REFERENCES

[1] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "A deterministic approach to wireless relay networks," 2007, *arXiv:0710.3777*. [Online]. Available: https://arxiv.org/abs/0710.3777

[2] E. C. van Der Meulen, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, no. 1, pp. 120–154, 1971. [Online]. Available: http://www.jstor.org/stable/1426331

[3] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[4] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 3, p. 536, May 1982.

[5] R. Kolte, A. Özgür, and H. Permuter, "Cooperative binning for semi-deterministic channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1231–1249, Mar. 2016.

[6] B. Akhbari, M. Mirmohseni, and M. R. Aref, "Compress-and-forward strategy for relay channel with causal and non-causal channel state information," *IET Commun.*, vol. 4, no. 10, pp. 1174–1186, Jul. 2010.

[7] M. N. Khormuji, A. El Gamal, and M. Skoglund, "State-dependent relay channel: Achievable rate and capacity of a semideterministic class," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2629–2638, May 2013.

[8] M. N. Khormuji and M. Skoglund, "The relay channel with partial causal state information," in *Proc. Int. Symp. Inf. Theory Appl.*, Dec. 2008, pp. 1–6.

[9] Z. Deng, F. Lang, B.-Y. Wang, and S.-M. Zhao, "Capacity of a class of relay channel with orthogonal components and non-causal channel state," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2696–2700.

[10] A. Zaidi, S. Shamai (Shitz), P. Piantanida, and L. Vandendorpe, "Bounds on the capacity of the relay channel with noncausal state at the source," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2639–2672, May 2013.

[11] A. Zaidi, S. P. Kotagiri, J. N. Laneman, and L. Vandendorpe, "Cooperative relaying with state available noncausally at the relay," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2272–2298, May 2010.

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

[13] I. E. Aguerri and D. Gündüz, "Capacity of a class of state-dependent orthogonal relay channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1280–1295, Mar. 2016.

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

[15] H. H. Permuter, S. Shamai (Shitz), and A. Somekh-Baruch, "Message and state cooperation in multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6379–6396, Oct. 2011.

[16] F. M. J. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.

[17] O. Simeone *et al.*, "Cooperative wireless cellular systems: An information-theoretic view," *Found. Trends Commun. Inf. Theory*, vol. 8, nos. 1–2, pp. 1–177, 2012.

[18] H. Asnani and H. H. Permuter, "Multiple-access channel with partial and controlled cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2252–2266, Apr. 2013.

[19] T. Kopetz, H. H. Permuter, and S. Shamai (Shitz), "Multiple access channels with combined cooperation and partial cribbing," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 825–848, Feb. 2016.

[20] A. Rosenzweig, Y. Steinberg, and S. Shamai (Shitz), "On channels with partial channel state information at the transmitter," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1817–1830, May 2005.

[21] Y. Steinberg, "Coding for channels with rate-limited side information at the decoder, with applications," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4283–4295, Sep. 2008.

[22] S. Jafar, "Capacity with causal and noncausal side information: A unified view," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5474, Dec. 2006.

[23] J. Wolfowitz, *Coding Theorems of Information Theory* (Ergebnisse der Mathematik und Ihrer Grenzgebiete). Berlin, Germany: Springer, 1964.

[24] J. N. Laneman and G. Kraner, "Window decoding for the multiaccess channel with generalized feedback," in *Proc. Int. Symp. Inf. Theory*, Jun./Jul. 2004, p. 281.

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

[26] I. B. Gattegno, Z. Goldfeld, and H. H. Permuter, "Fourier-motzkin elimination software for information theoretic inequalities," *IEEE Inf. Theory Soc. Newslett.*, vol. 65, no. 3, pp. 25–28, Sep. 2015. [Online]. Available: *arXiv:1610.03990*. [Online]. Available: https://wwwee.ee.bgu.ac.il/~fmeit/

[27] 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.

Ido B. Gattegno (S'15) received his B.Sc. (cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 2014 and 2015, respectively. Between 2005 and 2008, he served in the infantry of the Israeli Defense Forces.

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 a professor, Luck-Hille Chair in Electrical Engineering. Haim also serves as head of the communication, cyber/ and information track in his department. 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 served on the editorial boards of the IEEE TRANSACTIONS ON INFORMATION THEORY in 2013–2016.

Shlomo Shamai (Shitz) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975-1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Life Fellow, an URSI Fellow, a member of the Israeli Academy of Sciences and Humanities and a foreign member of the US National Academy of Engineering. He is the recipient of the 2011 Claude E. Shannon Award, the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering and the 2017 IEEE Richard W. Hamming Medal. He is a co-recipient of the 2018 Third Bell Labs Prize for Shaping the Future of Information and Communications Technology.

He has been awarded the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and is a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 and 2015 European Commission FP7, Network of Excellence in Wireless COMmunications (NEWCOM++, NEWCOM#) Best Paper Awards, the 2010 Thomson Reuters Award for International Excellence in Scientific Research, the 2014 EURASIP Best Paper Award (for the EURASIP Journal on Wireless Communications and Networking), the 2015 IEEE Communications Society Best Tutorial Paper Award and the 2018 IEEE Signal Processing Best Paper Award. Dr. Shamai (Shitz) is listed as a Highly Cited Researcher (Computer Science) for the years 2013–2018. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He has also served on the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY and on the IEEE Information Theory Society Nominations and Appointments Committee, and serves on the IEEE Information Theory Society, Shannon Award Committee.

Ayfer Özgür (M'06) received her B.Sc. degrees in electrical engineering and physics from Middle East Technical University, Turkey, in 2001 and the M.Sc. degree in electrical engineering from the same university in 2004. From 2001 to 2004, she worked as a hardware engineer for the SAGE Research Lab of the National Science Foundation in Turkey. She received her Ph.D. degree in 2009 from the Information Processing Group at EPFL, Switzerland. In 2010 and 2011, she was a post-doctoral scholar at the same institution. She is currently an Assistant Professor in the Electrical Engineering Department at Stanford University where she is a Hoover and Gabilan Fellow. Her current research interests include distributed communication and learning, wireless systems, and information theory. Dr. Ozgur received the EPFL Best Ph.D. Thesis Award in 2010, the NSF CAREER award in 2013, the Okawa Foundation Research Award and the IEEE CTTC Early Achievement Award in 2018.