Multiple Access Channels With Combined Cooperation and Partial Cribbing

Tal Kopetz, *Student Member, IEEE*, Haim H. Permuter, *Senior Member, IEEE*, and Shlomo (Shitz) Shamai, *Fellow, IEEE*

***Abstract*— In this paper, the multiple access channel (MAC) with combined cooperation and partial cribbing is studied, and its capacity region is characterized. Cooperation means that each of the two encoders sends a message to the other via a rate-limited link prior to transmission, while partial cribbing means that each of the two encoders obtains a deterministic function of the other encoder's output with or without delay. Prior work in this field dealt separately with cooperation and partial cribbing, but by combining these two methods, we can achieve significantly higher rates. Surprisingly, the capacity region of the MAC with combined cooperation and partial cribbing can be expressed using only one auxiliary random variable (RV) similar to the capacity regions of the MAC with cooperation and with partially cribbing encoders. The reason is that in an optimal coding scheme, the encoders use both cooperation and partial cribbing to generate a common message between the encoders. Furthermore, the Gaussian MAC with combined one-sided cooperation and quantized cribbing is studied. For this model, an achievability scheme is given. This scheme shows how many cooperation or quantization bits are required to practically achieve the capacity region of the Gaussian MAC with full message cooperation or perfect cribbing. To ratify the main results, two additional models are studied. In both models, only one auxiliary RV is needed. The first is a rate distortion dual setting for the MAC with degraded message set and combined cooperation and cribbing. The second is a state-dependent MAC with cooperation, where the state is known at a partially cribbing encoder and at the decoder. However, there are cases where more than one auxiliary RV is needed, e.g., when the cooperation and the cribbing are not used for the same purposes. The MAC with an action-dependent state is presented, where the action is based on the cooperation but not on the cribbing. Therefore, in this case, more than one auxiliary RV is needed. As a result, when the common information shared by the two encoders is used unevenly by the users in the channel, more than one auxiliary RV is needed to express the capacity region.**

***Index Terms*— Action, block Markov coding, cooperation, duality, double rate splitting, Gaussian MAC, Gelfand-Pinsker coding, multiple access channels, partial cribbing, state.**

Manuscript received April 3, 2014; revised May 1, 2015; accepted September 7, 2015. Date of publication November 11, 2015; date of current version January 18, 2016. This work was supported in part by the European Commission in the framework of the Seventh Framework Programme Network of Excellence in Wireless Communications (NEWCOM#), in part by the European Research Council Starting Grant, in part by the Israel Science Foundation, and in part by the S. and N. Grand Research Fund. This paper was presented at the 2014 IEEE International Symposium on Information Theory.

T. Kopetz and H. Permuter are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beersheba 8410501, Israel (e-mail: kopetz@post.bgu.ac.il; haimp@bgu.ac.il).

S. Shamai is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel (e-mail: sshlomo@ee.technion.ac.il).

Communicated by T. Liu, Associate Editor for Shannon Theory.

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

Digital Object Identifier 10.1109/TIT.2015.2499759

I. Introduction

The MAC with cooperating encoders was first studied by Willems [1]–[3], who introduced two separate approaches to cooperating encoders. In the first, the two encoders use a rate-limited cooperation link to cooperate and share as much of their private messages as possible. In the second approach, named cribbing, each encoder "listens" to and obtains the output of the other. Capacity regions for the two approaches were established separately by Willems. Furthermore, the cribbing setting was generalized in [4] to partial cribbing, which means that each of the two encoders obtains a deterministic function of the other encoder's output. The partial cribbing is especially important in the continuous alphabet, such as in the Gaussian MAC, because perfect cribbing in a continuous alphabet means full cooperation between the encoders regardless of the cribbing delay.

In this paper, we combine cooperation and partial cribbing and use them simultaneously to obtain better performance and a larger capacity region. In a MAC with combined cooperation and partial cribbing, depicted in Fig. 1, Encoder 1 and Encoder 2 obtain messages M_{21} and M_{12} prior to transmission. As for the cribbing part, we address two cases. In Case A, the cribbing is done strictly causally by both encoders, i.e., Encoder 1 forms $X_{1,i}$ as a function of (M_1, M_{21}, Z_2^{i-1}) and Encoder 2 forms $X_{2,i}$ as a function of (M_2, M_{12}, Z_1^{i-1}) where $Z_{1,i}$ and $Z_{2,i}$ are deterministic functions of $X_{1,i}$ and $X_{2,i}$, respectively. In Case B, the cribbing is done strictly causally by Encoder 1 and causally by Encoder 2, i.e., Encoder 1 forms $X_{1,i}$ as a function of (M_{21}, Z_2^{i-1}) and Encoder 2 forms $X_{2,i}$ as a function of (M_{12}, Z_1^{i}). The idea behind the approach is that the deterministic function, Z_1, is on a sliding scale where one end is $Z_{1,i} = X_{1,i}$ (the actual output) and the other end is when $Z_{1,i}$ is a constant, which does not give any information about $X_{1,i}$. The same applies for Z_2. In this research, it was our goal to obtain a generic capacity region for a scheme with both cooperation and partial cribbing.

Cooperation and cribbing carry practical implications. In [5, Ch. 8], Simone et al. considered cooperative wireless cellular systems and analyzed their performance with separate cooperation and cribbing (referred to as Out-of-Band cooperation and In-Band cooperation, respectively). The results show how cooperation and cribbing separately increase capacity in wireless cellular systems. In the expected 3GPP Release 12, a standard called Proximity Services (ProSE) will be added to the LTE-Advanced "grab bag" of technologies [6].

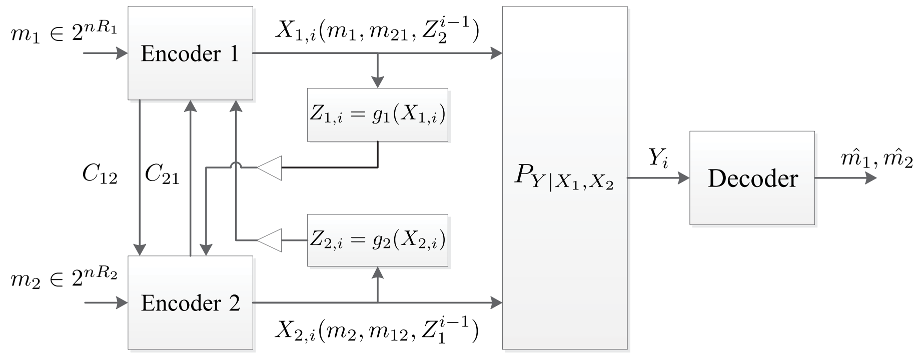

Fig. 1. MAC with combined cooperation and partial cribbing. Encoder 1 and Encoder 2 obtain messages M_{21} and M_{12} prior to transmission. The cribbing is done strictly causally by both encoders. This setting corresponds to Case A.

The ProSE protocol will address issues of spectrum utilization, overall throughput, and energy consumption, while enabling new peer to peer and location based applications and services, all of which will be applied using cooperation between "nearby" users in the network. The communication between the users can be attained by using mobile ad hoc networks (Out-of-Band/Cooperation) or by using the same band as the cell sites (In-Band/Cribbing). The settings of combined cooperation and cribbing considered in this paper give the fundamental limits and insights on how to design optimal coding for communication systems where the users have cognition capabilities, and therefore, "listen" to each other's signals and, in addition, cooperate with each other via dedicated links. We show that combining cribbing and cooperation is straightforward since it does not require any additional auxiliary RV compared with either cribbing or cooperation exclusively. Therefore, the combination of cooperation and cribbing should be considered in future cooperative wireless cellular systems such as ProSE.

In this paper, we solve the general model that incorporates both cooperation and partial cribbing, which enables the encoders to form a common message consisting of information that is known to both of them. In [9], Slepian and Wolf studied the MAC with a common message and two private messages and found that its capacity region is generally larger than that of the MAC with only private messages. Moreover, the impact of the common message on the capacity region can be captured using only one auxiliary RV. Accordingly, the capacity regions for separate cooperation and partial cribbing [1], [4] consist of an auxiliary RV, U. One of the results in our work is that the combination of the models does not require an additional auxiliary RV, as it is possible to use only one auxiliary RV that represents the common information. This implies that if for the MAC with partial cribbing we have a "good code", namely, a code that achieves the capacity region, then by performing minor modifications, namely, increasing the common message rate, we can construct a "good code" for the MAC in which cooperation and partial cribbing are combined. The coding techniques we use in this paper include block Markov coding (introduced by Willems), joint typicality decoding, backward decoding, and double rate splitting, the last of which is necessary because we need to split the original message twice: one part will be obtained through the cooperation link and the other part using partial cribbing.

Combining cooperation and cribbing was first considered by Bracher and Lapidoth [10] in the context of feedback and state information. In [10], however, only strictly-causal perfect cribbing was considered, and in our paper we consider the cases of strictly causal and causal partial cribbing. We show that it is preferable to compare cooperation with partial cribbing because of their similarity, i.e., in both cases, only part of the private message is shared.

After establishing our main results, we present the Gaussian MAC with combined one-sided cooperation and partial cribbing. One can see that an outer bound for the capacity region of this setting is when Encoder 2 knows the message of Encoder 1. Inspired by the work of Asnani and Permuter [4] and Bross et al. [11], we describe an achievability scheme that coincides with this outer bound in some cases.

To obtain further support for our results, we consider the channel-coding rate-distortion duality that was first presented by Shannon in [12]. We provide a duality between a MAC with a degraded message set and combined cooperation and cribbing and the rate distortion model known as "Successive Refinement (SR) With Decoder Cooperation" presented in [13]. The decoder cooperation is through a dedicated link and partial cribbing. Indeed, the rate region of combined cooperation and partial cribbing in the SR problem consists of one auxiliary RV.

Using only one auxiliary RV for both cooperation and partial cribbing is not always possible. To see this, we study the impact of cooperation and cribbing on state-dependent MACs where the state may provide a refined characterization of the channel, which is a widely accepted approach in the literature to state-dependent channels. We address two different state-dependent MACs with cooperation and cribbing (see [10], [14] for further reading). The first is a MAC with cooperation and channel state known non-causally at a partially cribbing encoder and at the decoder. In this case, we use our results to establish a capacity region with a single auxiliary RV, only one of which is needed since the purpose of both cooperation and partial cribbing is to generate a common message between the encoders. The second is a MAC where an action-dependent state is known non-causally at a cribbing encoder, at which a one-sided cooperation link is attained. Action-dependent states, introduced by Weissman in [15], comprise an action based on the private message of the cribbing encoder and the message from the cooperation link. Because the purpose of the cooperation is not only to generate a common message but also to contribute to the action and affect the channel state, in the case of the action-dependent state, one auxiliary RV will not suffice.

The remainder of the paper is organized as follows: In Section II, we define the MAC with combined cooperation and partial cribbing and provide its capacity region for two cases. The first is for strictly causal partial cribbing (Case A) and the second is for mixed causal and strictly causal partial cribbing (Case B). Thereafter, the proof for both cases is provided. In Section III, we give an achievability scheme for the Gaussian MAC with combined one-sided cooperation and partial cribbing. In Section IV, we establish the duality between the MAC with combined cooperation and partial cribbing at the encoders and the SR problem with combined cooperation and partial cribbing at the decoders. We show that one auxiliary RV is needed to characterize the rate region of the SR problem. In Section V, we give an example of a state-dependent MAC with combined cooperation and partial cribbing where only one auxiliary RV is needed. In Section VI, we study the case of the MAC with an action-dependent state where more than one auxiliary RV is needed and explain why. In Section VII we conclude the paper and state open problems such as noncausal partial cribbing and combined cooperation and cribbing in the interference channel.

II. The MAC With Combined Cooperation and Partial Cribbing

A. Definitions and Main Results

Let us consider the MAC with combined cooperation and partial cribbing depicted in Fig. 1. The MAC setting consists of two transmitters (encoders) and one receiver (decoder). Each transmitter $l \in \{1, 2\}$ chooses an index m_l uniformly from the set $\{1, \dots, 2^{nR_l}\}$ and independently of the other transmitter. RVs are denoted by capital letters, and their realizations by the respective lower case letters. X_m^n denotes the random vector $(X_m, \dots, X_n)$ and X^n denotes the random vector $(X_1, \dots, X_n)$. The input to the channel from Encoder $l \in \{1, 2\}$ is denoted by $\{X_{l,1}, X_{l,2}, X_{l,3}, \dots\}$. Encoder 1 and Encoder 2 obtain deterministic functions of the form $Z_{2,i} = g_2(X_{2,i})$ and $Z_{1,i} = g_1(X_{1,i})$, respectively. We address two cases in this setting:

- *Case A:* Both Encoder 1 and Encoder 2 obtain $Z_{2,i}$ and $Z_{1,i}$, respectively, with unit delay.
- *Case B:* Encoder 1 obtains $Z_{2,i}$ with unit delay and Encoder 2 obtains $Z_{1,i}$ without delay.

Additionally, Encoder 1 obtains a message $m_{21} \in \{1, \dots, 2^{nC_{21}}\}$ from Encoder 2 and Encoder 2 obtains a message $m_{12} \in \{1, \dots, 2^{nC_{12}}\}$ from Encoder 1. Both messages are obtained prior to the transmission of (X_1^n, X_2^n) through the channel. The output of the channel is denoted by $\{Y_1, Y_2, Y_3, \dots\}$. The channel is characterized by a conditional probability $P(y_i|x_{1,i}, x_{2,i})$. The channel probability does not depend on the time index i and is memoryless, i.e.,

$$P(y_i|x_1^i, x_2^i, y^{i-1}) = P(y_i|x_{1,i}, x_{2,i}), \tag{1}$$

where the superscripts denote sequences as follows: $x_l^i = (x_{l,1}, x_{l,2}, \dots, x_{l,i}), l \in \{1, 2\}$. Since the settings in this paper do not include feedback from the receiver to the transmitters, i.e., $P(x_{1,i}, x_{2,i}|x_1^{i-1}, x_2^{i-1}, y^{i-1}) = P(x_{1,i}, x_{2,i}|x_1^{i-1}, x_2^{i-1})$, equation (1) implies that

$$P(y_i|x_1^n, x_2^n, y^{i-1}) = P(y_i|x_{1,i}, x_{2,i}). \tag{2}$$

Definition 1: A $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, 2^{nC_{21}}, n)$ *code* for the MAC with combined cooperation and partial cribbing, as shown in Fig. 1, consists of encoding functions at Encoder 1 and Encoder 2

$$f_{12} : \{1, \dots, 2^{nR_1}\} \mapsto \{1, \dots, 2^{nC_{12}}\}, \tag{3}$$
$$f_{21} : \{1, \dots, 2^{nR_2}\} \mapsto \{1, \dots, 2^{nC_{21}}\}, \tag{4}$$
$$f_{1,i} : \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nC_{21}}\} \times \mathcal{Z}_2^{i-1} \mapsto \mathcal{X}_{1,i}, \tag{5}$$
$$f_{2,i}^A : \{1, \dots, 2^{nR_2}\} \times \{1, \dots, 2^{nC_{12}}\} \times \mathcal{Z}_1^{i-1} \mapsto \mathcal{X}_{2,i}, \tag{6}$$
$$f_{2,i}^B : \{1, \dots, 2^{nR_2}\} \times \{1, \dots, 2^{nC_{12}}\} \times \mathcal{Z}_1^{i} \mapsto \mathcal{X}_{2,i}, \tag{7}$$

where $i \in \{1, \dots, n\}$, and a decoding function

$$g : \mathcal{Y}^n \mapsto \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\}. \tag{8}$$

The average probability of error for a $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, 2^{nC_{21}}, n)$ code is defined in (9) at the bottom of the page.

A rate (R_1, R_2) is said to be *achievable* for the MAC with combined cooperation and partial cribbing if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, 2^{nC_{21}}, n)$ codes s.t. $P_e^{(n)} \to 0$. The *capacity region* of the MAC is the closure of all achievable rates. The following theorem describes the capacity region of a MAC with combined cooperation and partial cribbing.

Let us define the regions $\mathcal{R}^A$ and $\mathcal{R}^B$ that are contained in the set of nonnegative two-dimensional real numbers, which we henceforth denote by $\mathbb{R}_+^2$. The region $\mathcal{R}^A$ is defined in (10) at the bottom of the page. The region $\mathcal{R}^B$ is defined with the

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{m_1, m_2} \Pr\{g(Y^n) \neq (m_1, m_2)|(m_1, m_2) \text{ sent}\}. \tag{9}$$

$$\mathcal{R}^A = \left\{ \begin{array}{l} R_1 \leq I(X_1; Y|X_2, Z_1, U) + H(Z_1|U) + C_{12}, \\ R_2 \leq I(X_2; Y|X_1, Z_2, U) + H(Z_2|U) + C_{21}, \\ R_1 + R_2 \leq I(X_1, X_2; Y|U, Z_1, Z_2) + H(Z_1, Z_2|U) + C_{12} + C_{21}, \\ R_1 + R_2 \leq I(X_1, X_2; Y), \text{ for} \\ P(u)P(x_1|u)\mathbb{1}_{z_1=g_1(x_1)}P(x_2|u)\mathbb{1}_{z_2=g_2(x_2)}P(y|x_1, x_2). \end{array} \right\} \tag{10}$$

same set of inequalities as in (10), but the second inequality, i.e., the rate bound on R_2, is replaced by

$$R_2 \leq I(X_2; Y|X_1, Z_2, U) + H(Z_2|U, Z_1) + C_{21}, \quad (11)$$

and the joint distribution is of the form

$$P(u)P(x_1|u)\mathbb{1}_{z_1=g_1(x_1)}P(x_2|u, z_1)\mathbb{1}_{z_2=g_2(x_2)}P(y|x_1, x_2). \quad (12)$$

Theorem 1 (Capacity Region of the MAC With Combined Cooperation and Partial Cribbing): The capacity regions of the MAC with combined cooperation and strictly causal (Case A) and mixed strictly causal and causal (Case B) partial cribbing, as described in Def. 1, are $\mathcal{R}^A$ and $\mathcal{R}^B$, respectively.

We note that $H(Z_1|U) = I(Z_1; X_1|U)$, which practically represents the capacity between Encoder 1 and Encoder 2. Thus the cribbing term, $I(Z_1; X_1|U)$, plays the same role (in a quantitative sense) as the cooperation link term, C_{12}; both are capacities between Encoder 1 and Encoder 2. Similarly, the role of $I(Z_2; X_2|U)$ to C_{21} and of $I(Z_1, Z_2; X_1, X_2|U)$ to $C_{12} + C_{21}$. Hence, the important feature here is the mutual information of the cooperation, whether the cooperation is done by cribbing or by dedicated links, since they both behave similarly.

Remark 1 In Willems' definition of conferencing [1], a conference between two encoders consists of K subsequent pairs of communications, (V_1^K, V_2^K), emitted by both encoders simultaneously prior to transmission over the channel. Each cooperation transmission relies on the private message and past received transmissions, i.e., $V_{1,i} = f(M_1, V_2^{i-1})$ and $V_{2,i} = f(M_2, V_1^{i-1})$. The number of information bits transmitted after K rounds is nC_{12} and nC_{21} at Encoder 2 and Encoder 1, respectively. Since cooperation occurs prior to transmission, it can be modeled as a rate-limited link between the encoders and the capacity region is the same for both definitions. Furthermore, without cribbing, it is implicit that having cooperation prior to transmission yields a capacity region which is greater than or equal to the capacity region when cooperation occurs during transmission. However, when cribbing is available, this assumption is not straightforward. In Appendix A we prove that when partial cribbing is available, cooperation prior to transmission and during transmission yield the same capacity region.

In the case of a common message, the capacity region is as follows. We define the rate regions $\mathcal{R}_0^A$ and $\mathcal{R}_0^B$ exactly as $\mathcal{R}^A$ and $\mathcal{R}^B$, respectively, but the last inequality in (10), i.e., $R_1 + R_2 \leq I(X_1, X_2; Y)$, is replaced by

$$R_0 + R_1 + R_2 \leq I(X_1, X_2; Y). \quad (13)$$

Theorem 2 (Capacity Region in the Case of a Common Message): The capacity regions of the MAC with combined cooperation and strictly causal (Case A) and mixed strictly causal and causal (Case B) partial cribbing and a common message are $\mathcal{R}_0^A$ and $\mathcal{R}_0^B$, respectively.

The proof for this Theorem is given in Appendix B.

B. Proof of Theorem 1

1) Converse for Case A:: Given an achievable rate (R_1, R_2), we need to show that there exists a joint distribution of the form $P(u)P(x_1|u)\mathbb{1}_{z_1=g_1(x_1)}P(x_2|u)\mathbb{1}_{z_2=g_2(x_2)}P(y|x_1, x_2)$ such that the inequalities (10) are satisfied. Since (R_1, R_2) is an achievable rate-pair, there exists a $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, 2^{nC_{21}}, n)$ code with an arbitrarily small error probability $P_e^{(n)}$. By Fano's inequality,

$$H(M_1, M_2|Y^n) \leq n(R_1 + R_2)P_e^{(n)} + H(P_e^{(n)}). \quad (14)$$

We set

$$(R_1 + R_2)P_e^{(n)} + \frac{1}{n}H(P_e^{(n)}) \triangleq \epsilon_n, \quad (15)$$

where $\epsilon_n \to 0$ as $P_e^{(n)} \to 0$. Hence,

$$H(M_1|Y^n, M_2) \leq H(M_1, M_2|Y^n) \leq n\epsilon_n, \quad (16)$$
$$H(M_2|Y^n, M_1) \leq H(M_1, M_2|Y^n) \leq n\epsilon_n. \quad (17)$$

For R_1 we have the following:

$$\begin{aligned}
nR_1 &= H(M_1) && (18)\\
&\overset{(a)}{=} H(M_1, M_{12}, Z_1^n|M_2) && (19)\\
&\overset{(b)}{=} H(M_{12}|M_2) + H(Z_1^n|M_{12}, M_2)\\
&\quad + H(M_1|Z_1^n, M_{12}, M_2) && (20)\\
&= H(M_{12}) + H(Z_1^n|M_{12}, M_{21}, M_2)\\
&\quad + H(M_1|Z_1^n, M_{12}, M_2)\\
&\quad + H(M_1|Y^n, Z_1^n, M_{12}, M_2)\\
&\quad - H(M_1|Y^n, Z_1^n, M_{12}, M_2) && (21)\\
&\overset{(c)}{\leq} H(M_{12}) + H(Z_1^n|M_{12}, M_{21}, M_2)\\
&\quad + I(M_1; Y^n|Z_1^n, M_{12}, M_2, M_{21}) + n\epsilon_n && (22)\\
&\overset{(d)}{=} H(M_{12}) + \sum_{i=1}^{n}[H(Z_{1,i}|Z_1^{i-1}, M_{12}, M_{21}, M_2)\\
&\quad + I(M_1; Y_i|Y^{i-1}, Z_1^n, M_{12}, M_2, M_{21})] + n\epsilon_n && (23)\\
&\overset{(e)}{=} H(M_{12}) + \sum_{i=1}^{n}[H(Z_{1,i}|Z_1^{i-1}, Z_2^{i-1}, M_{12}, M_{21}, M_2)\\
&\quad + I(M_1, X_{1,i}; Y_i|Y^{i-1}, Z_1^n, Z_2^{i-1}, M_{12}, M_2, M_{21})]\\
&\quad + n\epsilon_n && (24)\\
&\overset{(f)}{\leq} H(M_{12}) + \sum_{i=1}^{n}[H(Z_{1,i}|Z_1^{i-1}, Z_2^{i-1}, M_{12}, M_{21})\\
&\quad + I(X_{1,i}; Y_i|X_{2,i}, Z_1^i, Z_2^{i-1}, M_{12}, M_{21})] + n\epsilon_n, && (25)
\end{aligned}$$

where (a) follows since messages M_1 and M_2 are independent and since $(M_{12}, Z_1^n) = f(M_1, M_2)$, (b) and (d) follow from the chain rule, (c) follows from Fano's inequality and because M_{21} is a function of M_2, (e) follows since Z_2^{i-1} is a function of (M_{12}, M_2, Z_1^{i-2}) and $X_{1,i}$ is a function of (M_1, M_{21}, Z_2^{i-1}), and step (f) follows since conditioning reduces entropy and from the Markov chain $Y_i - (X_{1,i}, X_{2,i}, M_{12}, M_{21}, Z_1^i, Z_2^{i-1}) - (M_1, M_2, Y^{i-1}, Z_{1,i+1}^n)$. We set the following RV

$$U_i \triangleq (Z_1^{i-1}, Z_2^{i-1}, M_{12}, M_{21}), \quad (26)$$

and obtain

$$R_1 \leq C_{12} + \frac{1}{n}\sum_{i=1}^{n}[H(Z_{1,i}|U_i) + I(X_{1,i}; Y_i|X_{2,i}, Z_{1,i}, U_i)] + \epsilon_n. \quad (27)$$

Similarly to (27), we obtain

$$R_2 \leq C_{21} + \frac{1}{n}\sum_{i=1}^{n}[H(Z_{2,i}|U_i) + I(X_{2,i}; Y_i|X_{1,i}, Z_{2,i}, U_i)] + \epsilon_n. \quad (28)$$

Now, consider

$$\begin{aligned}
n(R_1 + R_2) &= H(M_1, M_2) && (29)\\
&\overset{(a)}{=} H(M_1, M_2, Z_1^n, Z_2^n, M_{12}, M_{21}) && (30)\\
&\overset{(b)}{=} H(M_{12}) + H(M_{21}|M_{12}) \\
&\quad + H(Z_1^n, Z_2^n|M_{12}, M_{21}) \\
&\quad + H(M_1, M_2|Z_1^n, Z_2^n, M_{12}, M_{21}) && (31)\\
&\overset{(c)}{\leq} H(M_{12}) + H(M_{21}) \\
&\quad + H(Z_1^n, Z_2^n|M_{12}, M_{21}) \\
&\quad + I(M_1, M_2; Y^n|Z_1^n, Z_2^n, M_{12}, M_{21}) + n\epsilon_n && (32)\\
&\overset{(d)}{\leq} nC_{12} + nC_{21} \\
&\quad + \sum_{i=1}^{n}[H(Z_{1,i}, Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}, M_{12}, M_{21}) \\
&\quad + I(M_1, M_2; Y_i|Y^{i-1}, Z_1^n, Z_2^n, M_{12}, M_{21})] + n\epsilon_n && (33)\\
&\overset{(e)}{=} nC_{12} + nC_{21} \\
&\quad + \sum_{i=1}^{n}[H(Z_{1,i}, Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}, M_{12}, M_{21}) \\
&\quad + I(M_1, X_{1,i}, M_2, X_{2,i}; Y_i|Y^{i-1}, Z_1^n, Z_2^n, M_{12}, M_{21})] \\
&\quad + n\epsilon_n && (34)\\
&\overset{(f)}{=} nC_{12} + nC_{21} \\
&\quad + \sum_{i=1}^{n}[H(Z_{1,i}, Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}, M_{12}, M_{21}) \\
&\quad + I(X_{1,i}, X_{2,i}; Y_i|Z_1^i, Z_2^i, M_{12}, M_{21})] + n\epsilon_n, && (35)
\end{aligned}$$

where (a) follows from the fact that $(M_{12}, M_{21}, Z_1^n, Z_2^n) = f(M_1, M_2)$, (b) and (d) follow from the chain rule, (c) follows from Fano's inequality and because M_{21} is independent of M_{12}, (e) follows from the fact that $(X_{1,i}, X_{2,i}) = f(M_1, M_2)$, and step (f) follows from the Markov chain $Y_i - (X_{1,i}, X_{2,i}, Z_1^i, Z_2^i, M_{12}, M_{21}) - (M_1, M_2, Y^{i-1}, Z_{1,i+1}^n, Z_{2,i+1}^n)$. From the definition of the RV U, we obtain

$$R_1 + R_2 \leq C_{12} + C_{21} + \frac{1}{n}\sum_{i=1}^{n}[H(Z_{1,i}, Z_{2,i}|U_i) + I(X_{1,i}, X_{2,i}; Y_i|Z_{1,i}, Z_{2,i}, U_i)] + \epsilon_n. \quad (36)$$

Furthermore, consider

$$\begin{aligned}
n(R_1 + R_2) &= H(M_1, M_2) && (37)\\
&= H(M_1, M_2) + H(M_1, M_2|Y^n) \\
&\quad - H(M_1, M_2|Y^n) && (38)\\
&\overset{(a)}{\leq} I(M_1, M_2; Y^n) + n\epsilon_n && (39)\\
&\overset{(b)}{=} I(X_1^n, X_2^n; Y^n) + n\epsilon_n && (40)\\
&\overset{(c)}{=} \sum_{i=1}^{n} I(X_1^n, X_2^n; Y_i|Y^{i-1}) + n\epsilon_n && (41)\\
&\overset{(d)}{=} \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i) + n\epsilon_n, && (42)
\end{aligned}$$

where (a) follows from Fano's inequality, (b) follows from the fact that (X_1^n, X_2^n) is a deterministic function of (M_1, M_2) and from the Markov chain $Y^n - (X_1^n, X_2^n) - (M_1, M_2)$, (c) follows from the chain rule, and step (d) follows from the memoryless property of the channel. Thus we obtain

$$R_1 + R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i) + \epsilon_n. \quad (43)$$

Finally, to prove the Markov chains, we will use the graphic method as in [16, Sec. II]. In this method we choose a joint distribution, draw each RV of the joint distribution as an edge and connect the edges according to the Markov chains corresponding to the joint distribution. For example, if we have the joint distribution $P(a, b, c) = P(a)P(b|a)p(c|b)$, i.e., the Markov chain $A - B - C$ holds, then edge a will be connected to edge b, edge b will be connected to edges (a, c), and edge c will be connected to edge b. We now prove the following Markov chains:

- $Z_{2,i} - U_i - Z_{1,i}$ - Using the undirected graph in Fig. 2, we can see that the Markov Chain $Z_{2,i} - (M_{12}, M_{21}, Z_1^{i-1}, Z_2^{i-1}) - Z_{1,i}$ holds since we cannot get from node $Z_{2,i}$ to node $Z_{1,i}$ without going through nodes $(M_{12}, M_{21}, Z_1^{i-1}, Z_2^{i-1})$.
- $X_{1,i} - (U_i, Z_{1,i}) - Z_{2,i}$ - Using the undirected graph in Fig. 2, we can see that the Markov Chain $X_{1,i} - (M_{12}, M_{21}, Z_1^i, Z_2^{i-1}) - Z_{2,i}$ holds since we cannot get from node $X_{1,i}$ to node $Z_{2,i}$ without going through nodes $(M_{12}, M_{21}, Z_1^i, Z_2^{i-1})$.
- $X_{2,i} - (U_i, Z_{2,i}) - X_{1,i}$ - Using the undirected graph in Fig. 2, we can see that the Markov Chain $X_{2,i} - (M_{12}, M_{21}, Z_1^{i-1}, Z_2^i) - X_{1,i}$ holds since we cannot get from node $X_{2,i}$ to node $X_{1,i}$ without going through nodes $(M_{12}, M_{21}, Z_1^{i-1}, Z_2^i)$.
- $Y_i - (X_{1,i}, X_{2,i}) - (Z_{1,i}, Z_{2,i}, U_i)$ - Follows since the channel output at time i depends on the history (X_1^i, X_2^i) only through $(X_{1,i}, X_{2,i})$.

Finally, let Q be an RV independent of (X_1^n, X_2^n, Y^n) and uniformly distributed over the set $\{1, 2, 3, \dots, n\}$. We define the RVs $U \triangleq (Q, U_Q)$, $X_1 \triangleq X_{1,Q}$, $X_2 \triangleq X_{2,Q}$, and $Y \triangleq Y_Q$ to obtain the region given in (10). This completes the converse for Case A. ■

Converse for Case B: For the first, third and fourth inequalities we repeat the same approach as for Case A. For the bound on R_2 we have the following:

$$\begin{aligned}
nR_2 &= H(M_2) && (44)\\
&\overset{(a)}{=} H(M_2, M_{21}, Z_2^n|M_1) && (45)\\
&\overset{(b)}{=} H(M_{21}|M_1) + H(Z_2^n|M_{21}, M_1) \\
&\quad + H(M_2|Z_2^n, M_{21}, M_1) && (46)
\end{aligned}$$

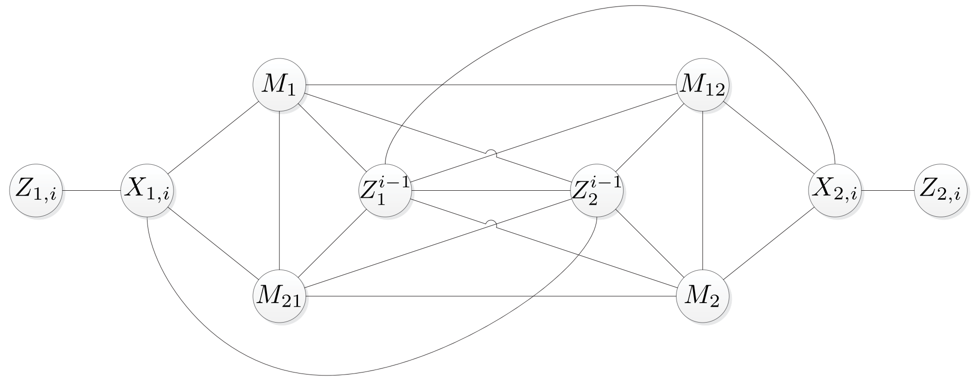

Fig. 2. Proof of the Markov Chain $X_{2,i} - (M_{12}, M_{21}, Z_1^{i-1}, Z_2^{i-1}) - X_{1,i}$ using the undirected graphical technique [16, Sec. II]. This graph corresponds to the joint distribution $P(m_1)P(m_2)P(m_{12}|m_1)P(m_{21}|m_2)\prod_{k=1}^{i-1} P(z_{1,k}|m_1, m_{21}, z_2^{k-1})P(z_{2,k}|m_2, m_{12}, z_1^{k-1})P(x_{1,i}|m_{21}, m_1, z_2^{i-1})$ $P(x_{2,i}|m_{12}, m_2, z_1^{i-1})P(z_{1,i}|x_{1,i})P(z_{2,i}|x_{2,i})$.

$$
\begin{aligned}
&= H(M_{21}) + H(Z_2^n|M_{12}, M_{21}, M_1) \\
&\quad + H(M_2|Z_2^n, M_{21}, M_1) + H(M_2|Y^n, Z_2^n, M_{21}, M_1) \\
&\quad - H(M_2|Y^n, Z_2^n, M_{21}, M_1) \qquad (47)\\
&\overset{(c)}{\leq} H(M_{21}) + H(Z_2^n|M_{12}, M_{21}, M_1) \\
&\quad + I(M_2; Y^n|Z_2^n, M_{12}, M_1, M_{21}) + n\epsilon_n \qquad (48)\\
&\overset{(d)}{=} H(M_{21}) + \sum_{i=1}^{n}[H(Z_{2,i}|Z_2^{i-1}, M_{12}, M_{21}, M_1) \\
&\quad + I(M_2; Y_i|Y^{i-1}, Z_2^n, M_{12}, M_1, M_{21})] + n\epsilon_n \qquad (49)\\
&\overset{(e)}{=} H(M_{21}) + \sum_{i=1}^{n}[H(Z_{2,i}|Z_1^i, Z_2^{i-1}, M_{12}, M_{21}, M_1) \\
&\quad + I(M_2, X_{2,i}; Y_i|Y^{i-1}, Z_2^n, Z_1^i, M_{12}, M_1, M_{21})] + n\epsilon_n \qquad (50)\\
&\overset{(f)}{\leq} H(M_{21}) + \sum_{i=1}^{n}[H(Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}, M_{12}, M_{21}, Z_{1,i}) \\
&\quad + I(X_{2,i}; Y_i|X_{1,i}, Z_2^i, Z_1^{i-1}, M_{12}, M_{21})] + n\epsilon_n, \qquad (51)
\end{aligned}
$$

where (a) follows since messages M_1 and M_2 are independent and since $(M_{21}, Z_2^n) = f(M_1, M_2)$, (b) and (d) follow from the chain rule, (c) follows from Fano's inequality and because M_{12} is a function of M_1, (e) follows since Z_1^i is a function of (M_{21}, M_1, Z_2^{i-1}) and $X_{2,i}$ is a function of (M_2, M_{12}, Z_1^{i-1}), and step (f) follows since conditioning reduces entropy and from the Markov chain $Y_i - (X_{1,i}, X_{2,i}, M_{12}, M_{21}, Z_1^{i-1}, Z_2^i) - (M_1, M_2, Y^{i-1}, Z_{1,i}, Z_{2,i+1}^n)$. From the definition of the RV U, we obtain

$$
\begin{aligned}
R_2 \leq C_{21} + \frac{1}{n}\sum_{i=1}^{n}[&H(Z_{2,i}|U_i, Z_{1,i}) \\
&+ I(X_{2,i}; Y_i|X_{1,i}, Z_{2,i}, U_i)] + \epsilon_n. \qquad (52)
\end{aligned}
$$

Additionally, we need to show the Markov chain $X_{2,i} - (U_i, Z_{1,i}, Z_{2,i}) - X_{1,i}$ rather than $X_{2,i} - (U_i, Z_{2,i}) - X_{1,i}$ as in Case A. Since for Case A the Markov chain $X_{2,i} - (M_{12}, M_{21}, Z_1^{i-1}, Z_2^i) - X_{1,i}$ holds, then $X_{2,i} - (M_{12}, M_{21}, Z_1^i, Z_2^i) - X_{1,i}$ also holds. ■

2) Achievability for Case A: To prove the achievability of the capacity region, we need to show that for a fixed distribution of the form $P(u)P(x_1|u)\mathbb{1}_{z_1=g_1(x_1)}$ $P(x_2|u)\mathbb{1}_{z_2=g_2(x_2)}P(y|x_1, x_2)$ and for (R_1, R_2) that satisfy the inequalities in (10), there exists a sequence of $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, 2^{nC_{21}}, n)$ codes for which $P_e^{(n)} \to 0$ as $n \to \infty$.

The idea behind this proof is to convert the cooperation problem into a setting that corresponds to the MAC with a common message and partially cribbing encoders considered in [4] and rely on its capacity region to show that the cooperation capacity region is indeed achievable. This is done by sharing as much as possible of the original private messages, (m_1, m_2), through the communication links to create a common message; the unshared parts of the original messages serve as the new private messages. By doing so, the coding scheme of the setting with a common message can be employed. The capacity region found in [4] for the MAC with a common message and partially cribbing encoders is

$$
\begin{aligned}
\tilde{R}_1 &\leq H(Z_1|U) + I(X_1; Y|X_2, Z_1, U), \\
\tilde{R}_2 &\leq H(Z_2|U) + I(X_2; Y|X_1, Z_2, U), \\
\tilde{R}_1 + \tilde{R}_2 &\leq I(X_1, X_2; Y|U, Z_1, Z_2) + H(Z_1, Z_2|U), \\
\tilde{R}_0 + \tilde{R}_1 + \tilde{R}_2 &\leq I(X_1, X_2; Y). \qquad (53)
\end{aligned}
$$

Let us define the following rates

$$
\begin{aligned}
\tilde{R}_0 &= C_{12} + C_{21}, \qquad (54)\\
\tilde{R}_1 &= R_1 - C_{12}, \qquad (55)\\
\tilde{R}_2 &= R_2 - C_{21}, \qquad (56)
\end{aligned}
$$

i.e., we defined the common message as the messages that are transmitted through the cooperation links. With respect to these definitions, the inequalities in (53) can be rewritten as (57), given at the bottom of the next page. This region is equivalent to the region in (10). ■

Achievability for Case B: The achievability of Case B is very similar to that of Case A with minor modifications since $Z_{1,i}$ is known causally at Encoder 2. The encoding part is the same, but now the codewords Z_2^n and X_2^n are generated according to a code-tree (or a Shannon's strategy).

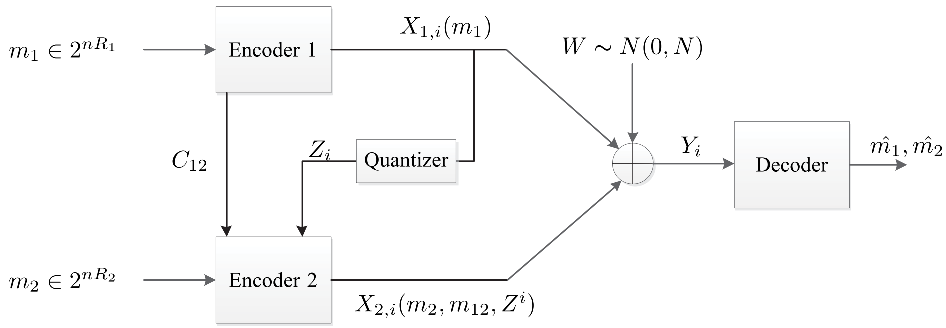

Fig. 3. Gaussian MAC with one-sided combined cooperation and quantized cribbing. Message M_{12} is sent prior to transmission and Z_i is known causally at Encoder 2.

At every time index $i \in \{1, \ldots, n\}$, for every $z_1 \in \mathcal{Z}_1$ we generate $Z_{2,i}$ and $X_{2,i}$ according to the distribution $p(z_2|u, z_1)$ and $p(x_2|u, z_1, z_2)$, respectively. This will result in $2^{n(R_2'+R_2'')}$ code-trees.

Decoding is done backwards as in Case A, but here the decoder looks for X_2^n differently. Since at block b the decoder already decoded z_1^n, it knows the correct path on the code-tree, i.e., it can identify the correct leaf in the code-tree. First, the decoder follows the chosen path on the the tree to find the correct z_2^n. Then, it looks for a typical X_2^n from the set of codewords that correspond to the chosen path on the tree of X_2. The rest of the proof is the same as in Case A. ■

III. Gaussian MAC With Combined Cooperation and Quantized Cribbing

We now consider a Gaussian MAC where $Y = X_1 + X_2 + W$ and $W \sim N(0, N)$, depicted in Fig. 3.

We assume that the power constraint over the outputs of Encoder 1 and Encoder 2 is

$$\frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X_{l,i}^2\right] \leq P_l, \quad \text{for } l = 1, 2. \tag{58}$$

Prior to transmission, Encoder 1 sends a message M_{12} to Encoder 2. In addition, Encoder 2 cribs causally from Encoder 1 and obtains Z_i, which is a scalar quantization of the signal $X_{1,i}$. First, we examine an inner bound to the capacity region. Since Encoder 2 can ignore the cribbed symbols and the messages that it obtains from Encoder 1, the capacity region of the Gaussian MAC without cooperation and quantized cribbing is contained in the capacity region with combined cooperation and quantized cribbing. Hence, we have the following inner bound:

$$\begin{aligned} R_1 &\leq \frac{1}{2}\log(1 + \frac{P_1}{N}), \\ R_2 &\leq \frac{1}{2}\log(1 + \frac{P_2}{N}), \\ R_1 + R_2 &\leq \frac{1}{2}\log(1 + \frac{P_1 + P_2}{N}). \end{aligned} \tag{59}$$

On the other hand, an outer bound is obtained when there is full cooperation or perfect cribbing, i.e., Encoder 2 obtains the message m_1 before sending X_2. The capacity region in this case is

$$\begin{aligned} R_2 &\leq \frac{1}{2}\log(1 + \frac{P_2}{N}(1-\rho^2)), \\ R_1 + R_2 &\leq \frac{1}{2}\log(1 + \frac{P_1 + 2\rho\sqrt{P_1P_2} + P_2}{N}). \end{aligned} \tag{60}$$

We now present an achievable region by choosing a joint distribution for the region $\mathcal{R}^B$. The choice of the distribution of each RV is inspired by the work of Asnani and Permuter [4] and Bross et al. [11]. In [4], an achievable region for the Gaussian MAC with quantized cribbing has been described (This region was not proven to be optimal), whereas in [11], an achievable region for the Gaussian MAC with a common message was provided (This region was proven to be optimal). In our achievable region, we choose the distributions of the RVs as a combination of the choices made in [4] and [11]. We set the following distributions:

$$X_1 = \lambda U + X_1', \tag{61}$$

$$X_2 = \bar{\lambda} U + X_2', \tag{62}$$

where

$$\begin{aligned} &U \sim N(0, P_0), \quad P_0 = \left(\sqrt{\bar{\beta}_1 P_1} + \sqrt{\bar{\beta}_2 P_2}\right)^2, \\ &P_{X_2'|Z,U}(x_2'|z, u) = \bar{\rho} P_{X_2''}(x_2') + \rho P_{X_1'|Z,U}(x_2'|z, u), \\ &X_1' \sim N(0, \beta_1 P_1), \quad X_2'' \sim N(0, \beta_2 P_2), \\ &\lambda = \sqrt{\frac{\bar{\beta}_1 P_1}{P_0}}, \quad \bar{\lambda} = 1 - \lambda, \\ &\beta_1, \beta_2, \rho \in [0, 1]. \end{aligned} \tag{63}$$

The intuition behind the choice of these distributions is as follows. The common message, signified as U, is obtained via the rate-limited link and the two encoders cooperate to send that common message. Since the cooperation and cribbing are

---

$$\begin{aligned} R_1 - C_{12} &\leq H(Z_1|U) + I(X_1; Y|X_2, Z_1, U), \\ R_2 - C_{21} &\leq H(Z_2|U) + I(X_2; Y|X_1, Z_2, U), \\ (R_1 - C_{12}) + (R_2 - C_{21}) &\leq I(X_1, X_2; Y|U, Z_1, Z_2) + H(Z_1, Z_2|U), \\ (C_{12} + C_{21}) + (R_1 - C_{21}) + (R_2 - C_{21}) &\leq I(X_1, X_2; Y), \end{aligned} \tag{57}$$

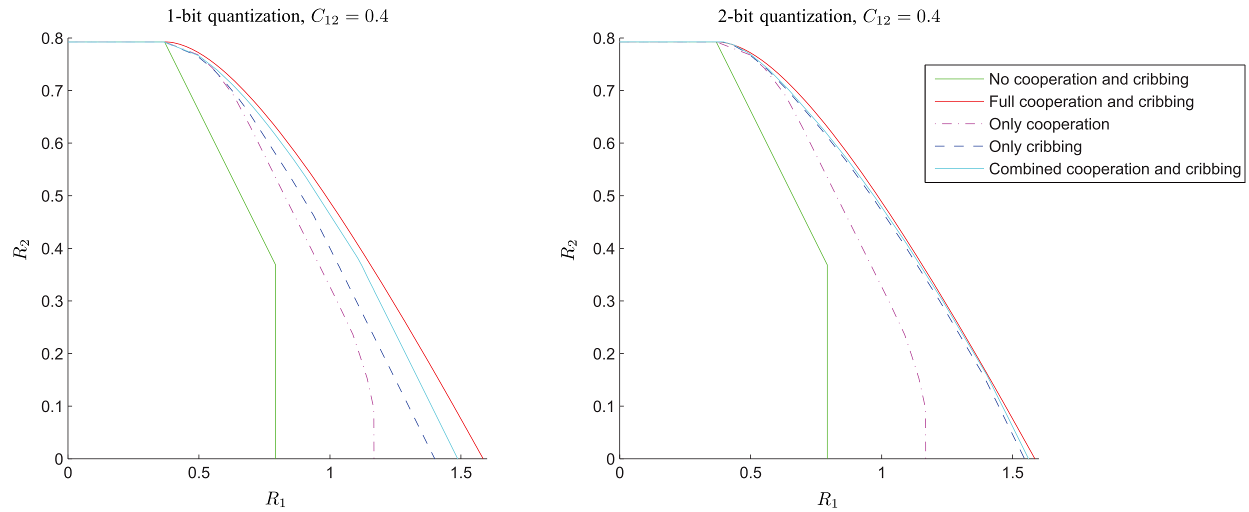

Fig. 4. Achievable regions for the Gaussian MAC with combined cooperation and quantized cribbing.

one-sided, only Encoder 2 can help Encoder 1 send its private message. The idea behind the choice of $P_{X_2'|Z,U}(x_2'|z,u)$ is that Encoder 2 will send $\bar{\rho}$ of the time its private message and ρ of the time the estimation of Encoder 1's private message, X_1', conditioned on the cribbing Z and the cooperation U. Notice that under these definitions, by setting the power constraints as $P_1 = P_2 = 1$, the power constraints on both encoders hold. Evaluation of region $\mathcal{R}_B$ with Z_2 constant and $N = \frac{1}{2}$ is depicted in Fig. 4; achievable regions for 1-bit and 2-bit quantizations are illustrated where $C_{12} = 0.4$. When only one bit of quantization is available (LHS of Fig. 4), the region of combined cooperation and cribbing encloses special cases of cribbing [4] and cooperation [11]. However, when two bits of quantization are available (RHS of Fig. 4), combining cooperation and cribbing does not significantly increase the region. This is because the difference between the achievable region with a 2-bit quantizer ($C_{12} = 0$) and full cooperation is negligible. As a result, for 2-bit quantization, this achievable region practically achieves the capacity region of the Gaussian MAC with full message cooperation or perfect cribbing.

IV. Dual Rate Distortion Setting

In this section we show how our methods for combined cooperation and cribbing can be implemented in the rate-distortion dual. The information-theoretic duality between rate distortion and channel coding was first introduced by Shannon in [12]. An important duality between the Wyner-Ziv rate distortion problem [17] and the Gelfand-Pinsker channel coding problem [18] was pointed out by Cover and Chiang in [19] (see [20] and [21] for further reading). In some cases, the corner points of a rate distortion region and its dual channel coding capacity region are the same. This property can help one find a region based on its dual region.

In general, there is no solution for the dual setting of the MAC. However, the rate distortion dual of the MAC with degraded message set has been solved. In [13], Asnani et al. considered the SR problem with decoder cooperation and its channel coding dual. We establish the duality between the MAC with a degraded message set and combined cooperation and partial cribbing and the SR problem with combined cooperation and partial cribbing at the decoders. As expected, the rate region for the rate-distortion dual is established using only one auxiliary RV. Table I, at the top of the next page, describes the principles of duality between channel coding and source coding.

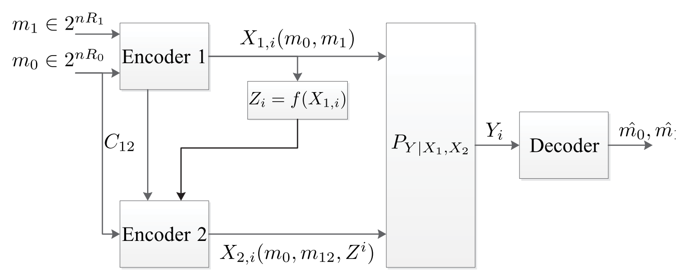

Fig. 5. MAC with common message, private message, and combined cooperation and cribbing. Encoder 2 obtains message M_{12} prior to transmission. The cribbing is done causally.

We start with the channel coding problem and consider the setting depicted in Fig. 5.

Following Theorem 2, the capacity region for this setting is

$$\mathcal{R}_{MAC} = \left\{ \begin{array}{c} R_1 \leq I(X_1; Y|Z, U) + H(Z|U) + C_{12}, \\ R_0 + R_1 \leq I(X_1, U; Y), \text{ for} \\ P(u)P(x_1|u)\mathbb{1}_{z=f(x_1)}P(x_2|u,z)P(y|x_1,x_2). \end{array} \right\}. \tag{64}$$

We go on to define the SR setting with combined cooperation and partial cribbing at the decoders.

A. Successive Refinement With Combined Cooperation and Partial Cribbing at the Decoders

We address the rate distortion setting depicted in Fig. 6.

The source sequence $X_i \in \mathcal{X}, i = 1, 2, \ldots$ is drawn i.i.d. $\sim p(x)$. Let $\hat{\mathcal{X}}_1$ and $\hat{\mathcal{X}}_2$ denote the reconstruction alphabets, and $d_i : \mathcal{X} \times \hat{\mathcal{X}}_i \mapsto [0, \infty)$, for $i = 1, 2$ denote single letter distortion measures. Distortion between sequences

TABLE I

PRINCIPLES OF DUALITY BETWEEN CHANNEL CODING AND SOURCE CODING

Channel coding	Source coding
Channel decoder	Source encoder
Encoder 1 input $(M_0, M_1) \in \{1, \ldots, 2^{n(R_0+R_1)}\}$	Decoder 1 input $(T_0, T_1) \in \{1, \ldots, 2^{n(R_0+R_1)}\}$
Encoder 1 output $X_1 \in \mathcal{X}_1$	Decoder 1 output $\hat{X}_1 \in \hat{\mathcal{X}}_1$
Encoder 2 input $M_0 \in \{1, \ldots, 2^{nR_0}\}$, $Z_i(X_{1,i}), M_{12}(M_0, M_1)$	Decoder 2 input $T_0 \in \{1, \ldots, 2^{nR_0}\}$, $Z_i(\hat{X}_{1,i}), T_{12}(T_0, T_1)$
Encoder 2 output $X_2 \in \mathcal{X}_2$	Decoder 2 output $\hat{X}_2 \in \hat{\mathcal{X}}_2$
Decoder input $Y \in \mathcal{Y}$	Encoder input $X \in \mathcal{X}$
Decoder output $(\hat{M}_0, \hat{M}_1) \in \{1, \ldots, 2^{n(R_0+R_1)}\}$	Encoder output $(T_0, T_1) \in \{1, \ldots, 2^{n(R_0+R_1)}\}$
Encoding function $f_1 : \mathcal{M}_0 \times \mathcal{M}_1 \mapsto \mathcal{X}_1^n$	Decoding function $g_1 : \mathcal{T}_0 \times \mathcal{T}_1 \mapsto \hat{\mathcal{X}}^n$
Causal cribbing encoding function $f_2 : \mathcal{M}_0 \times \mathcal{M}_{12} \times \mathcal{Z}^i \mapsto \mathcal{X}_{2,i}$	Causal cribbing decoding function $g : \mathcal{T}_0 \times \mathcal{T}_{12} \times \mathcal{Z}^i \mapsto \hat{\mathcal{X}}_{2,i}$
Decoding function $g : \mathcal{Y}^n \mapsto \mathcal{M}_0 \times \mathcal{M}_1$	Encoding function $f_0 : \mathcal{X}^n \mapsto \mathcal{T}_0$, $f_1 : \mathcal{X}^n \mapsto \mathcal{T}_1$
Auxiliary RV U	Auxiliary RV U
Joint distribution $p(u, x_1, x_2, y)$	Joint distribution $p(u, \hat{x}_1, \hat{x}_2, x)$
Constraint: $p(y\|x_1, x_2)$ is fixed	Constraint: $p(x)$ is fixed

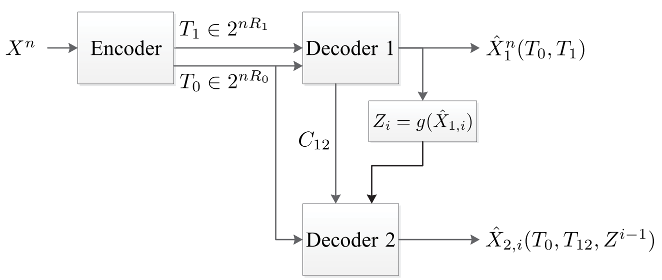

Fig. 6. SR with combined cooperation and partial cribbing at the decoders. The cribbing is done causally.

is defined in the usual way;

$$d_l(x^n, \hat{x}_l^n) = \frac{1}{n}\sum_{j=1}^{n} d_l(x_j, \hat{x}_{l,j}), \text{ for } l = 1, 2. \tag{65}$$

Definition 2: A $(2^{nR_0}, 2^{nR_1}, 2^{nC_{12}}, n)$ *rate-distortion code* for the SR with combined cooperation and partial cribbing at the decoders, as shown in Fig. 6, consists of encoding functions

$$f_0 : \mathcal{X}^n \mapsto \{1, \ldots, 2^{nR_0}\}, \tag{66}$$
$$f_1 : \mathcal{X}^n \mapsto \{1, \ldots, 2^{nR_1}\}, \tag{67}$$
$$f_{12} : \{1, \ldots, 2^{nR_0}\} \times \{1, \ldots, 2^{nR_1}\} \mapsto \{1, \ldots, 2^{nC_{12}}\}, \tag{68}$$

and decoding functions at Decoder 1 and Decoder 2

$$g_1 : \{1, \ldots, 2^{nR_0}\} \times \{1, \ldots, 2^{nR_1}\} \mapsto \hat{\mathcal{X}}_1^n, \tag{69}$$
$$g_{2,i} : \{1, \ldots, 2^{nR_0}\} \times \{1, \ldots, 2^{nC_{12}}\} \times \mathcal{Z}^i \mapsto \mathcal{X}_{2,i}. \tag{70}$$

where $i \in \{1, \ldots, n\}$.

A rate (R_0, R_1, D_1, D_2) is said to be *achievable* for the SR with combined cooperation and partial cribbing at the decoders if $\forall \epsilon > 0$ and a $(2^{nR_0}, 2^{nR_1}, 2^{nC_{12}}, n)$ rate-distortion code, the expected distortion for the decoders, is bounded as

$$E\left[d_l(X^n, \hat{X}_l^n)\right] \leq D_l + \epsilon, \text{ for } l = 1, 2. \tag{71}$$

The *rate-distortion region* $\mathcal{R}(D_1, D_2)$ is defined as the closure of the set of all achievable rate-distortion tuples (R_0, R_1, D_1, D_2).

Let us define the following region $\mathcal{R}_{SR}(D_1, D_2)$ that is contained in $\mathbb{R}_+^2$.

$$\mathcal{R}_{SR}(D_1, D_2) = \left\{ \begin{array}{c} R_0 \geq I(X; Z, U) - H(Z|U) - C_{12}, \\ R_0 + R_1 \geq I(\hat{X}_1, U; X), \text{ for} \\ P(x, x_1, u)\mathbb{1}_{z=g(x_1), x_2 = f(u, z_1)} \text{ s.t.} \\ E\left[d_i(X^n, \hat{X}_i^n)\right] \leq D_i + \epsilon, \text{ for } i = 1, 2. \end{array} \right\}. \tag{72}$$

Theorem 3 (Rate Distortion Region of the Successive Refinement With Combined Cooperation and Partial Cribbing Decoders): The rate-distortion region for the SR with combined cooperation and partial cribbing, as defined in Def. 2, is $\mathcal{R}_{SR}(D_1, D_2)$.

Proof (Achievability): The achievability for this model is the same as in [13] where the achievable region was

$$\begin{aligned} \tilde{R}_0 &\geq I(X; Z, U) - H(Z|U), \\ \tilde{R}_0 + \tilde{R}_1 &\geq I(\hat{X}_1, U; X). \end{aligned} \tag{73}$$

In our case, we use rate splitting and set the following rates

$$\tilde{R}_0 = R_0 + C_{12}, \tag{74}$$
$$\tilde{R}_1 = R_0 - C_{12}. \tag{75}$$

By setting these rates we obtain the region in (72).

Converse: Assume we have a $(2^{nR_0}, 2^{nR_1}, 2^{nC_{12}}, n)$ rate distortion code s.t. a (R_0, R_1, D_1, D_2) tuple is feasible. For the first inequality

$$nR_0 \geq H(T_0) \tag{76}$$
$$\stackrel{(a)}{=} H(Z^n, T_0, T_{12}) - H(Z^n|T_{12}, T_0) - H(T_{12}|T_0) \tag{77}$$

TABLE II
CORNER POINTS OF MAC AND SR

	(R_0, R_1)
MAC (Theorem 2)	$(I(Y;Z,U) - H(Z\|U) - C_{12}, I(Y;X_1\|Z,U) + H(Z\|U) + C_{12})$ $(I(Y;X_1,U), 0)$
SR (Theorem 3)	$(I(X;Z,U) - H(Z\|U) - C_{12}, I(X;\hat{X}_1\|Z,U) + H(Z\|U) + C_{12})$ $(I(X;\hat{X}_1,U), 0)$

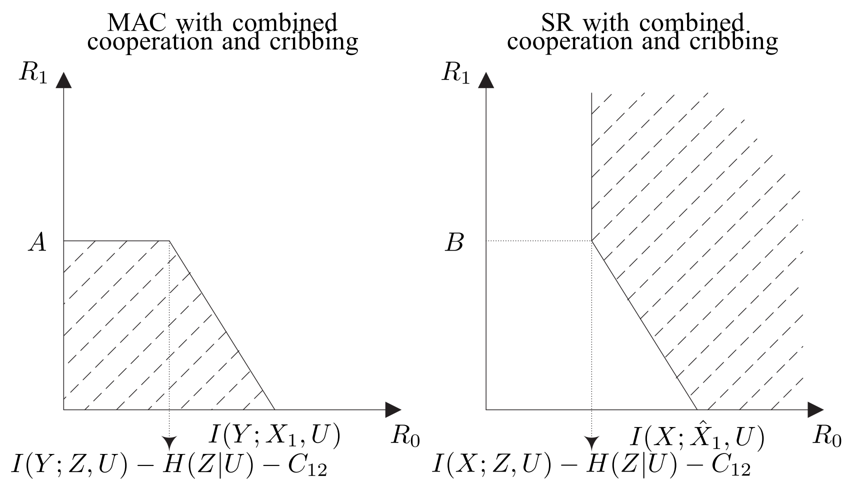

Fig. 7. Capacity region of the MAC and rate-distortion region of SR with combined cooperation and cribbing where A is $I(Y;X_1|Z,U) + H(Z|U) + C_{12}$ and B is $I(X;\hat{X}_1|Z,U) + H(Z|U) + C_{12}$.

$$
\begin{aligned}
&\overset{(b)}{\geq} I(X^n;Z^n,T_0,T_{12}) - H(Z^n|T_{12},T_0) - H(T_{12}) && (78)\\
&\overset{(c)}{\geq} \sum_{i=1}^n [I(X_i;Z^n,T_0,T_{12}|X^{i-1}) \\
&\qquad - H(Z_i|T_{12},T_0,Z^{i-1})] - nC_{12} && (79)\\
&= \sum_{i=1}^n [I(X_i;Z^n,T_0,T_{12},X^{i-1}) \\
&\qquad - H(Z_i|T_{12},T_0,Z^{i-1})] - nC_{12} && (80)\\
&\overset{(d)}{\geq} \sum_{i=1}^n [I(X_i;Z^i,T_0,T_{12}) - H(Z_i|T_{12},T_0,Z^{i-1})] \\
&\quad - nC_{12} && (81)\\
&\overset{(e)}{=} \sum_{i=1}^n [I(X_i;Z_i,U_i) - H(Z_i|U_i)] - nC_{12} && (82)\\
&= n\sum_{i=1}^n \frac{1}{n}[I(X_i;Z_i,U_i) - H(Z_i|U_i)] - nC_{12} && (83)\\
&\overset{(f)}{=} n[I(X_Q;Z_Q,U_Q|Q) - H(Z_Q,U_Q|Q) - C_{12}] \\
& && (84)\\
&= n[I(X_Q;Z_Q,U_Q,Q) - H(Z_Q,U_Q|Q) - C_{12}] \\
& && (85)\\
&\geq n[I(X_Q;Z_Q,U_Q) - H(Z_Q,U_Q) - C_{12}], && (86)
\end{aligned}
$$

where (a) and (c) follow from the chain rule, (b) follows since conditionality reduces entropy, (d) follows since X_i is independent of X^{i-1}, (e) follows by setting the RV $U_i = (Z^{i-1}, T_0, T_{12})$, and (f) follows by defining the RV Q independent of X^n and uniformly distributed over the set $\{1, 2, 3, \ldots, n\}$. For the second inequality

$$
\begin{aligned}
n(R_0 + R_1) &\geq H(T_0, T_1) && (87)\\
&\overset{(a)}{=} I(X^n; T_0, T_1) && (88)\\
&= \sum_{i=1}^n I(X_i; T_0, T_1|X^{i-1}) && (89)\\
&\overset{(b)}{=} \sum_{i=1}^n I(X_i; T_0, T_1, X^{i-1}) && (90)\\
&\overset{(c)}{=} \sum_{i=1}^n I(X_i; T_0, T_1, \hat{X}_{1,i}, Z^{i-1}, T_{12}, X^{i-1}) && (91)\\
&\geq \sum_{i=1}^n I(X_i; \hat{X}_{1,i}, Z^{i-1}, T_0, T_{12}) && (92)\\
&= \sum_{i=1}^n I(X_i; \hat{X}_{1,i}, U_i) && (93)\\
&= nI(X_Q; \hat{X}_{1,Q}, U_Q), && (94)
\end{aligned}
$$

where (a) follows since (T_0, T_1) is a function of X^n, (b) follows since $X_{1,i}$ is independent of X_1^{i-1}, and (c) follows since $(\hat{X}_{1,i}, Z^{i-1}, T_{12})$ is a function of (T_0, T_1). We complete the proof by noting that the joint distribution of $(X_Q, \hat{X}_{1,Q}, Z_Q, U_Q)$ is the same as that of $(X, \hat{X}_1, Z, U)$. ■

B. Duality Results Between the MAC and the Successive Refinement Settings With Combined Cooperation and Partial Cribbing

After establishing regions $\mathcal{R}_{MAC}$ and $\mathcal{R}_{SR}(D_1, D_2)$, we now point out the dualities between the two settings. The similarity between the rate regions of the two settings is evident. Let us consider the corner points depicted in Table II and Fig. 7.

One can see that the corner points are the same if we apply the duality rules $\hat{X}_1 \leftrightarrow X_1$, $\hat{X}_2 \leftrightarrow X_2$, $X \leftrightarrow Y$ and $\geq \leftrightarrow \leq$. We notice that only one RV was used to describe the common

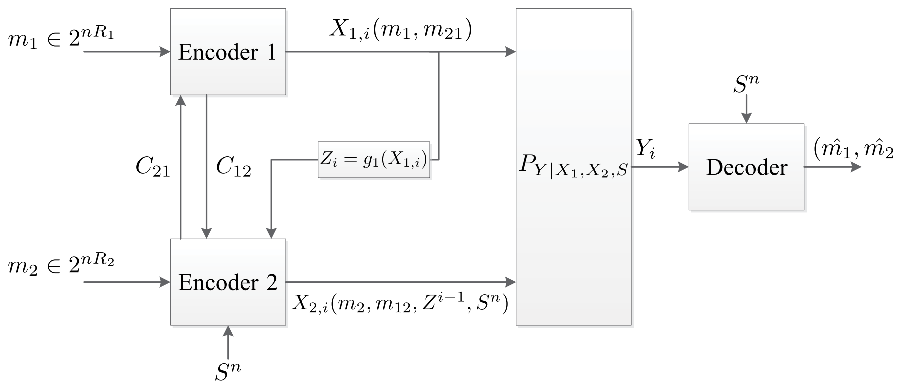

Fig. 8. The MAC with cooperation and state known at a partially cribbing encoder and at the decoder. Encoder 1 and Encoder 2 obtain messages M_{21} and M_{12} prior to transmission. Corresponding to the strictly causal case, the partial cribbing is done strictly causally only by Encoder 2.

message in both settings. This means that our methods of combining cooperation and cribbing can also be implemented in source coding problems. In the next section we address another case, in which only one RV is needed to describe both cooperation and cribbing.

V. State-Dependent MAC With Combined Cooperation and Partial Cribbing

Following our results from Section II, we now show that our methods can also be implemented for a state-dependent channel where also only one auxiliary RV is needed. Let us consider the MAC with cooperation and non-causal state known at a partially cribbing encoder and at the decoder, depicted in Fig. 8.

We note that message M_{12} is sent prior to message M_{21}. For this model we address two different cases:

- The strictly causal case (sc): Encoder 2 obtains Z_i with unit delay.
- The causal case (c): Encoder 2 obtains Z_i without delay.

The channel probability does not depend on the time index i and is memoryless, i.e.,

$$P(y_i|x_1^i, x_2^i, s^i, y^{i-1}) = P(y_i|x_{1,i}, x_{2,i}, s_i) \tag{95}$$

Definition 3: A $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, 2^{nC_{21}}, n)$ *code* for the MAC with cooperation and non-causal state known at a partially cribbing encoder and at the decoder, as shown in Fig. 8, consists at time i of encoding functions at Encoder 1 and Encoder 2,

$$f_{12} : \{1, \dots, 2^{nR_1}\} \mapsto \{1, \dots, 2^{nC_{12}}\}, \tag{96}$$

$$f_{21} : \{1, \dots, 2^{nR_2}\} \times \mathcal{S}^n \times \{1, \dots, 2^{nC_{12}}\} \mapsto \{1, \dots, 2^{nC_{21}}\}, \tag{97}$$

$$f_1 : \{1, \dots, 2^{nC_{21}}\} \times \{1, \dots, 2^{nR_1}\} \mapsto \mathcal{X}_1^n, \tag{98}$$

$$f_{2,i}^{sc} : \{1, \dots, 2^{nC_{12}}\} \times \{1, \dots, 2^{nR_2}\} \times \mathcal{S}^n \times \mathcal{Z}^{i-1} \mapsto \mathcal{X}_{2,i}, \tag{99}$$

$$f_{2,i}^{c} : \{1, \dots, 2^{nC_{12}}\} \times \{1, \dots, 2^{nR_2}\} \times \mathcal{S}^n \times \mathcal{Z}^{i} \mapsto \mathcal{X}_{2,i}, \tag{100}$$

and a decoding function

$$g : \mathcal{S}^n \times \mathcal{Y}^n \mapsto \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\}. \tag{101}$$

The average probability of error for a $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, 2^{nC_{21}}, n)$ code is defined in (102) at the bottom of the page.

Let us define the following regions, $\mathcal{R}_{State}^{sc}$ and $\mathcal{R}_{State}^{c}$, that are contained in $\mathbb{R}_+^2$. The region $\mathcal{R}_{State}^{sc}$ is defined in (103) at the bottom of the page. The region $\mathcal{R}_{State}^{c}$ is defined with the same set of inequalities as in (103), but the joint distribution is of the form

$$P(s)P(u|s)P(x_1|u)\mathbb{1}_{z=g_1(x_1)}P(x_2|s,u,z)P(y|x_1,x_2,s). \tag{104}$$

Theorem 4 (Capacity Region of the MAC With Cooperation and State Known at a Partial Cribbing Encoder): The capacity regions of the MAC with cooperation and non-causal state known at a partially cribbing encoder and at the decoder for the strictly causal case and the causal case, as described in Def. 3, are $\mathcal{R}_{State}^{sc}$ and $\mathcal{R}_{State}^{c}$, respectively.

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{m_1,m_2} \Pr\{g(Y^n, S^n) \neq (m_1, m_2) | (m_1, m_2) \text{ sent}\}. \tag{102}$$

$$\mathcal{R}_{State}^{sc} = \left\{ \begin{array}{rl} C_{21} \geq & I(U;S), \\ R_1 \leq & H(Z|U) + I(X_1;Y|S,U,X_2,Z) + C_{12}, \\ R_2 \leq & I(X_2;Y|X_1,S,U) + C_{21} - I(U;S), \\ R_1 + R_2 \leq & I(X_1,X_2;Y|S), \\ R_1 + R_2 \leq & I(X_1,X_2;Y|U,Z,S) + H(Z|U) + C_{12} + C_{21} - I(U;S), \text{ for} \\ & P(s)P(u|s)P(x_1|u)\mathbb{1}_{z=g_1(x_1)}P(x_2|s,u)P(y|x_1,x_2,s). \end{array} \right\} \tag{103}$$

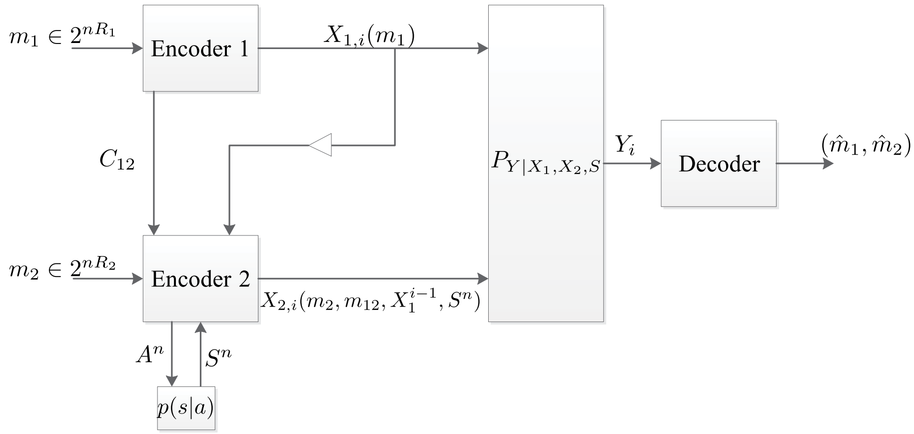

Fig. 9. The MAC with one-way cooperation and action-dependent state known at a cribbing encoder. Encoder 2 obtains message M_{12} prior to transmission. The cribbing is done strictly causally only by Encoder 2. This setting corresponds to the strictly causal case.

The role of the RV U is to generate an empirical coordination between the two encoders regarding the state channel and to generate a common message between the two encoders by combining the cooperation links and the partial cribbing. We now examine two special cases of this capacity region.

Case 1: The One-Sided Cooperation and No Cribbing Case, i.e., $|\mathcal{Z}| = 1$ and $C_{12} = 0$: In this case $H(Z|U) = 0$ and hence, the region $\mathcal{R}^{sc}_{State}$ coincides with the region in [22, Th. 1].

Case 2: $|\mathcal{S}| = 1$, MAC with a Constant State: Notice that in this case, $I(U;S) = 0$ and the region $\mathcal{R}^{sc}_{State}$ reduces to region $\mathcal{R}^2_{State}$ given in (105) at the bottom of the page. which is the region in Theorem 1 where $Z_1 = Z$ and only Encoder 2 cribs from Encoder 1, i.e., $|\mathcal{Z}_2| = 1$.

The proof of Theorem 4 is given in Appendix C.

VI. MAC With Cooperation and Action-Dependent State Known at a Cribbing Encoder

Although we have shown that for combined cooperation and cribbing only one auxiliary RV is needed to describe the capacity region, in some cases this is not possible. For instance, if the roles of the cribbing and cooperation in the communication setting are different, then more than one auxiliary RV is needed. In this section, we introduce a MAC with cooperation and action-dependent state known at a cribbing encoder. Because of the nature of actions and of non-causal states, the actions depend only on the cooperation and, therefore, two auxiliary RVs are needed, one for the cooperation and one for the cribbing.

Consider the MAC with one-way cooperation and action-dependent state known at a cribbing encoder, depicted in Fig. 9. Notice that the action A^n is taken from (m_2, m_{12}).

We address two cases for this setting:

- The strictly causal case (sc): Encoder 2 obtains $X_{1,i}$ with unit delay.
- The causal case (c): Encoder 2 obtains $X_{1,i}$ without delay.

The channel probability is defined as in (95).

Definition 4: A $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, n)$ *code* for the MAC with one-way cooperation and action-dependent state known at a cribbing encoder, as shown in Fig. 9, consists at time i of encoding functions at Encoder 1 and Encoder 2

$$f_{12} : \{1, \dots, 2^{nR_1}\} \mapsto \{1, \dots, 2^{nC_{12}}\}, \tag{106}$$

$$f_1 : \{1, \dots, 2^{nR_1}\} \mapsto \mathcal{X}_1^n, \tag{107}$$

$$f_A : \{1, \dots, 2^{nR_2}\} \times \{1, \dots, 2^{nC_{12}}\} \mapsto \mathcal{A}^n, \tag{108}$$

$$f^{sc}_{2,i} : \{1, \dots, 2^{nR_2}\} \times \{1, \dots, 2^{nC_{12}}\} \times \mathcal{S}^n \times \mathcal{X}_1^{i-1} \mapsto \mathcal{X}_{2,i}, \tag{109}$$

$$f^{c}_{2,i} : \{1, \dots, 2^{nR_2}\} \times \{1, \dots, 2^{nC_{12}}\} \times \mathcal{S}^n \times \mathcal{X}_1^{i} \mapsto \mathcal{X}_{2,i}, \tag{110}$$

and a decoding function

$$g : \mathcal{Y}^n \mapsto \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\}. \tag{111}$$

The average probability of error for a $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, n)$ code is defined as

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \cdot \sum_{m_1,m_2} \Pr\{g(Y^n) \neq (m_1, m_2) | (m_1, m_2) \text{ sent}\}. \tag{112}$$

$$\mathcal{R}^2_{State} = \left\{ \begin{array}{l} R_1 \leq H(Z|U) + I(X_1; Y|U, X_2, Z) + C_{12}, \\ R_2 \leq I(X_2; Y|X_1, U) + C_{21}, \\ R_1 + R_2 \leq I(X_1, X_2; Y), \\ R_1 + R_2 \leq I(X_1, X_2; Y|U, Z) + H(Z|U) + C_{12} + C_{21}, \text{ for} \\ P(u)P(x_1|u)\mathbb{1}_{z=g_1(x_1)} P(x_2|u) P(y|x_1, x_2). \end{array} \right\} \tag{105}$$

$$\mathcal{R}^{sc}_{Action} = \left\{ \begin{array}{c} R_1 \le \min\{H(X_1|V,W), I(Y;V,X_1,U|W,A) - I(S;U|W,V,A)\} + C_{12}, \\ R_2 \le I(U,A;Y|X_1,V,W) - I(U;S|W,V,A), \\ R_1 + R_2 \le I(X_1,V,U,A;Y|W) - I(U;S|W,V,A) + C_{12}, \\ R_1 + R_2 \le I(X_1,V,U,A,W;Y) - I(U;S|W,V,A), \text{ for} \\ P(w)P(v|w)p(a|w)P(s|a)P(x_1|v,w)P(u,x_2|s,v,a,w)P(y|x_1,x_2,s). \end{array} \right\} \quad (113)$$

$$P(w)P(v|w)p(a|w)P(s|a)P(x_1|v,w)P(u|s,v,a,w)P(x_2|v,u,s,a,w,x_1)P(y|x_1,x_2,s) \quad (114)$$

$$\mathcal{R}^{1}_{Action} = \left\{ \begin{array}{c} R_2 \le I(U,A;Y|X_1,V,W) - I(U;S|W,V,A), \\ R_1 + R_2 \le I(X_1,V,U,A,W;Y) - I(U;S|W,V,A), \text{ for} \\ P(w)P(v|w)p(a|w)P(s|a)P(x_1|v,w)P(u,x_2|s,v,a,w)P(y|x_1,x_2,s). \end{array} \right\} \quad (115)$$

$$\mathcal{R}^{2}_{Action} = \left\{ \begin{array}{c} R_1 \le H(X_1|V,W), \\ R_2 \le I(U;Y|X_1,V,W) - I(U;S|W,V), \\ R_1 + R_2 \le I(X_1,V,U;Y|W) - I(U;S|W,V), \\ R_1 + R_2 \le I(X_1,V,U,W;Y) - I(U;S|W,V), \text{ for} \\ P(w)P(v|w)P(s)P(x_1|v,w)P(u,x_2|s,v,w)P(y|x_1,x_2,s). \end{array} \right\} \quad (116)$$

Let us define the following regions $\mathcal{R}^{sc}_{Action}$ and $\mathcal{R}^{c}_{Action}$ that are contained in $\mathbb{R}^2_+$. The region $\mathcal{R}^{sc}_{Action}$ is defined in (113) at the top of the page. The region $\mathcal{R}^{c}_{Action}$ is defined with the same set of inequalities as in (113), but the joint distribution is of the form given in (114) at the top of the page.

Theorem 5 (Capacity Region of the MAC With Cooperation and Action-Dependent State Known at a Cribbing Encoder): The capacity regions of the MAC with one-way cooperation and action-dependent state known at a strictly causal and causal cribbing encoder, as described in Def. 3, are $\mathcal{R}^{sc}_{Action}$ and $\mathcal{R}^{c}_{Action}$, respectively.

In this case, U is a Gelfand-Pinsker coding RV [18]. The role of the RV W is to generate a common message based on the cooperation link, whereas the RV V generates a common message based on the cribbing. We cannot combine the cooperation and cribbing in this case because only part of the common information of both encoders is being used to generate the action sequence A^n. This example shows that in cases where only part of the common information that the encoders share is being used for certain purposes, cooperation and cribbing cannot be combined into one RV. We now address two previous results in this field and show that they are special cases of our result.

Case 1: The Action-Dependent MAC Where $C_{12} = R_1$: In this case the region reduces to region $\mathcal{R}^{1}_{Action}$ given in (115) at the top of the page. First, we notice that the cribbing in this case is redundant. Second, since the action is now taken from (M_1, M_2), we can set the RV $W = X_1$ and V as a constant and the region coincides with the capacity region in [23].

Case 2: The State-Dependent MAC With State Known at a Cribbing Encoder, i.e., $|\mathcal{A}| = 1$ and $C_{12} = 0$: Notice that in this case, the state is not action-dependent and the region reduces to region $\mathcal{R}^{2}_{Action}$ given in (116) at the top of the page. If we set W as constant, the region coincides with the capacity region in [24]. Since these regions are equal, this shows that the capacity region in [24] is a special case of the region in Theorem 5.

The proof of Theorem 5 is given in Appendix D.

VII. Conclusions and Future Work

In this paper, we have presented the capacity region for the MAC with combined cooperation and partial cribbing. Remarkably, the solution necessitates the use of only one auxiliary RV, which represents the common information obtained at the two encoders via cooperation and cribbing. Additionally, we have shown an achievability scheme for the Gaussian MAC with combined one-sided cooperation and causal partial cribbing. In this case, partial cribbing is a scalar quantization of Encoder 1's output obtained by Encoder 2. Numerical plots of achievability regions were presented for different numbers of quantization bits and capacity links. The plots indicated under which certain conditions (full message cooperation or perfect cribbing) the outer bound is achieved. To obtain further support for our results, we considered a dual setting for the MAC with a degraded message set and combined cooperation and cribbing. By converting our channel coding methods to rate-distortion, we successfully characterized the rate-distortion region for the dual model using a single auxiliary RV. Again, this RV represents the common information obtained at the two decoders via cooperation and cribbing. Continuing with an assessment of whether using only one auxiliary RV is possible for the state-dependent MAC with combined cooperation and partial cribbing, we applied our methods to find the capacity region for a MAC with cooperation and state known non-causally at a cribbing encoder and at the decoder. The capacity region consisted of only one auxiliary RV. Additionally, we addressed a MAC with one-way cooperation and cribbing and action-dependent state, where the action was based on the cooperation between the encoders. In this case two auxiliary RVs were needed. We deduced a rule that if only part of the common information that the encoders share is being used for arbitrary purposes, then cooperation and cribbing cannot be combined into one auxiliary RV. For future work, we suggest that the non-causal partial cribbing case and the interference channel with combined cooperation and cribbing be considered. An additional case to consider is that where the state or action is known at the weak encoder (the non-cognitive encoder).

Appendix A
Cooperation Prior to and During Transmission

For this model, we generalize Willems' definition from [1] to the case where partial cribbing is present. We consider the transmissions of sequences $(V_{1,-K+1}^n, V_{2,-K+1}^n)$ between the two encoders, where K rounds occur prior to transmission over the channel and n rounds occur during transmission over the channel. The number of information bits transmitted after $n+K$ rounds is nC_{12} and nC_{21} at Encoder 2 and Encoder 1, respectively, i.e.,

$$\sum_{i=-K+1}^{n} \log(|\mathcal{V}_{1,i}|) \leq nC_{12},$$
$$\sum_{i=-K+1}^{n} \log(|\mathcal{V}_{2,i}|) \leq nC_{21}. \tag{117}$$

Accordingly, we redefine the encoding functions for $i \in \{-K+1, \ldots, 0\}$ as

$$h_{1,i} : \{1, \ldots, 2^{nR_1}\} \times \mathcal{V}_{2,-K+1}^{i-1} \mapsto \mathcal{V}_{1,i}, \tag{118}$$
$$h_{2,i} : \{1, \ldots, 2^{nR_2}\} \times \mathcal{V}_{1,-K+1}^{i-1} \mapsto \mathcal{V}_{2,i}, \tag{119}$$

and for $i \in \{1, \ldots, n\}$ as

$$h_{1,i} : \{1, \ldots, 2^{nR_1}\} \times \mathcal{V}_{2,-K+1}^{i-1} \times \mathcal{Z}_2^{i-1} \mapsto \mathcal{V}_{1,i}, \tag{120}$$
$$h_{2,i} : \{1, \ldots, 2^{nR_2}\} \times \mathcal{V}_{1,-K+1}^{i-1} \times \mathcal{Z}_1^{i-1} \mapsto \mathcal{V}_{2,i}, \tag{121}$$
$$f_{1,i} : \{1, \ldots, 2^{nR_1}\} \times \mathcal{V}_{2,-K+1}^{i-1} \times \mathcal{Z}_2^{i-1} \mapsto \mathcal{X}_{1,i}, \tag{122}$$
$$f_{2,i} : \{1, \ldots, 2^{nR_2}\} \times \mathcal{V}_{1,-K+1}^{i-1} \times \mathcal{Z}_1^{i-1} \mapsto \mathcal{X}_{2,i}. \tag{123}$$

We note that for each transmission of the set (m_1, m_2) over the channel, we use $n+K$ cooperation transmissions. We allow cooperation to occur in K rounds prior to transmission, as in Willems' model. Additionally, we allow the cooperation to occur during transmission over the channel where cribbing takes place. Indeed, this model generalizes Willems' model to use cooperation before and during transmission. In the proof we show only the case of strictly causal partial cribbing. The proof for causal partial cribbing is omitted for brevity.

A. Achievability

The scheme is the same as in [4] with minor modifications. Since cooperation occurs before and during transmission, the common message obtained by the two encoders prior to transmission is that of the previous block, i.e., $m_{0,b-1}$. Thus, in the first block, no common information is known. From the second block, the common messages known prior to transmission are the cooperation messages obtained during transmission in the previous block. Hence, we encode the common message u^n using the message $m_{0,b-1}$. The rest of the proof is the same as in Subsection II-B2.

B. Converse

Given an achievable rate (R_1, R_2), we need to show that there exists a joint distribution of the form $P(u)P(x_1|u)\mathbb{1}_{z_1=g_1(x_1)}P(x_2|u)\mathbb{1}_{z_2=g_2(x_2)}P(y|x_1,x_2)$, such that the inequalities (10) are satisfied. Since (R_1, R_2) is an achievable rate-pair, there exists a $(2^{nR_1}, 2^{nR_2}, n)$ code with an arbitrarily small error probability $P_e^{(n)}$. By Fano's inequality,

$$H(M_1, M_2|Y^n) \leq n(R_1+R_2)P_e^{(n)} + H(P_e^{(n)}). \tag{124}$$

We set

$$(R_1+R_2)P_e^{(n)} + \frac{1}{n}H(P_e^{(n)}) \triangleq \epsilon_n, \tag{125}$$

where $\epsilon_n \to 0$ as $P_e^{(n)} \to 0$. Hence,

$$\begin{aligned} &H(M_1, M_2|Y^n, V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n, Z_2^n) \\ &\quad \leq H(M_1, M_2|Y^n) \leq n\epsilon_n, \end{aligned} \tag{126}$$
$$\begin{aligned} &H(M_1|Y^n, M_2, V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n) \\ &\quad \leq H(M_1, M_2|Y^n) \leq n\epsilon_n, \end{aligned} \tag{127}$$
$$\begin{aligned} &H(M_2|Y^n, M_1, V_{2,-K+1}^n, V_{2,-K+1}^n, Z_2^n) \\ &\quad \leq H(M_1, M_2|Y^n) \leq n\epsilon_n. \end{aligned} \tag{128}$$

For R_1, we have the following:

$$\begin{aligned} nR_1 &= H(M_1) && (129)\\ &= H(M_1|M_2) && (130)\\ &\overset{(a)}{=} H(M_1, V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n|M_2) && (131)\\ &= H(V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n|M_2) \\ &\quad + H(M_1|V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n, M_2), && (132) \end{aligned}$$

where (a) follows since $(V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n)$ is a function of (M_1, M_2). For the first term in (132) we have

$$\begin{aligned} &H(V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n|M_2) \\ &= H(V_{1,-K+1}^0, V_{2,-K+1}^0|M_2) \\ &\quad + H(V_1^n, V_2^n, Z_1^n|M_2, V_{1,-K+1}^0, V_{2,-K+1}^0) && (133)\\ &\overset{(b)}{=} \sum_{i=-K+1}^{0} H(V_{1,i}, V_{2,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, M_2) \\ &\quad + \sum_{i=1}^{n} H(V_{1,i}, V_{2,i}, Z_{1,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, Z_1^{i-1}, M_2) && (134)\\ &\overset{(c)}{=} \sum_{i=-K+1}^{0} H(V_{1,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, M_2) \\ &\quad + \sum_{i=1}^{n} H(V_{1,i}, Z_{1,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, Z_1^{i-1}, M_2) && (135)\\ &\overset{(d)}{=} \sum_{i=-K+1}^{0} H(V_{1,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, M_2) \\ &\quad + \sum_{i=1}^{n} H(V_{1,i}, Z_{1,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, Z_1^{i-1}, Z_2^{i-1}, M_2) && (136)\\ &\leq \sum_{i=1}^{n} H(Z_{1,i}|V_{1,-K+1}^{i}, V_{2,-K+1}^{i-1}, Z_1^{i-1}, Z_2^{i-1}, M_2) \\ &\quad + \sum_{i=-K+1}^{n} H(V_{1,i}) && (137)\\ &\overset{(e)}{\leq} nC_{12} + \sum_{i=1}^{n} H(Z_{1,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, Z_1^{i-1}, Z_2^{i-1}). && (138) \end{aligned}$$

where (b) follows from the chain rule, (c) follows since $V_{2,i}$ is a function of $(M_2, V_{1,-K+1}^{i-1}, Z_1^{i-1})$, (d) follows since Z_2^{i-1} is a function of $(M_2, V_{1,-K+1}^{i-2}, Z_1^{i-2})$, and (e) follows from the definition in (117). For the second term in (132) we have

$$\begin{aligned}
&H(M_1|V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n, M_2)\\
&\overset{(f)}{=} I(M_1; Y^n|V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n, M_2) + n\epsilon_n \qquad (139)\\
&\overset{(g)}{=} \sum_{i=1}^n I(M_1, X_{1,i}; Y_i|Y^{i-1}, Z_1^n, Z_2^n, V_{1,-K+1}^n,\\
&\qquad V_{2,-K+1}^n, M_2, X_{2,i}) + n\epsilon_n \qquad (140)\\
&\overset{(h)}{\leq} \sum_{i=1}^n I(X_{1,i}; Y_i|Z_1^i, Z_2^{i-1}, V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, X_{2,i}) + n\epsilon_n, \qquad (141)
\end{aligned}$$

where (f) follows from Fano's inequality, (g) follows since $Z_2^n = f(M_2, V_{1,-K+1}^n, Z_1^n)$, $X_{1,i} = f(M_1, Z_2^{i-1}, V_{2,-K+1}^{i-1})$, and $X_{2,i} = f(M_2, Z_1^{i-1}, V_{1,-K+1}^{i-1})$, and step (h) follows since conditioning reduces entropy and from the Markov chain $Y_i-(X_{1,i}, X_{2,i}, V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, Z_1^i, Z_2^{i-1})-(M_1, M_2, Y^{i-1}, V_{1,i}^n, V_{2,i}^n, Z_{1,i+1}^n, Z_{2,i}^n)$. We set the following RV

$$U_i \triangleq (V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, Z_1^{i-1}, Z_2^{i-1}), \qquad (142)$$

and obtain

$$\begin{aligned}
R_1 \leq \frac{1}{n}\sum_{i=1}^n [H(Z_{1,i}|U_i) + I(X_{1,i}; Y_i|X_{2,i}, Z_{1,i}, U_i)]\\
+ C_{12} + \epsilon_n. \qquad (143)
\end{aligned}$$

Similar to (143), we obtain

$$\begin{aligned}
R_2 \leq \frac{1}{n}\sum_{i=1}^n [H(Z_{2,i}|U_i) + I(X_{2,i}; Y_i|X_{1,i}, Z_{2,i}, U_i)]\\
+ C_{21} + \epsilon_n. \qquad (144)
\end{aligned}$$

Now, consider

$$\begin{aligned}
n(R_1 + R_2) &= H(M_1, M_2) \qquad (145)\\
&\overset{(a)}{=} H(M_1, M_2, V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n, Z_2^n) \qquad (146)\\
&= H(V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n, Z_2^n)\\
&\quad + H(M_1, M_2|V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n, Z_2^n), \qquad (147)
\end{aligned}$$

where (a) follows since $(V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n)$ is a function of (M_1, M_2). For the first term in (147) we have

$$\begin{aligned}
&H(V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n, Z_2^n)\\
&= H(V_{1,-K+1}^0, V_{2,-K+1}^0)\\
&\quad + H(V_1^n, V_2^n, Z_1^n, Z_1^n|V_{1,-K+1}^0, V_{2,-K+1}^0) \qquad (148)\\
&\overset{(b)}{=} \sum_{i=1}^n H(V_{1,i}, V_{2,i}, Z_{1,i}, Z_{2,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, Z_1^{i-1}, Z_2^{i-1})\\
&\quad + \sum_{i=-K+1}^0 H(V_{1,i}, V_{2,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}) \qquad (149)\\
&\leq \sum_{i=-K+1}^n H(V_{1,i}) + H(V_{2,i})\\
&\quad + \sum_{i=1}^n H(Z_{1,i}, Z_{2,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, Z_1^{i-1}, Z_2^{i-1}) \qquad (150)\\
&\overset{(c)}{\leq} \sum_{i=1}^n H(Z_{1,i}|V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, Z_1^{i-1}, Z_2^{i-1})\\
&\quad + nC_{12} + nC_{21}. \qquad (151)
\end{aligned}$$

where (b) follows from the chain rule and (c) follows from the definition in (117). For the second term in (147) we have

$$\begin{aligned}
&H(M_1, M_2|V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n, Z_2^n)\\
&\overset{(d)}{=} I(M_1, M_2; Y^n|V_{1,-K+1}^n, V_{2,-K+1}^n, Z_1^n, Z_2^n) + n\epsilon_n \qquad (152)\\
&\overset{(e)}{=} \sum_{i=1}^n I(M_1, X_{1,i}, M_2, X_{2,i}; Y_i|Y^{i-1}, Z_1^n, Z_2^n,\\
&\qquad V_{1,-K+1}^n, V_{2,-K+1}^n) + n\epsilon_n \qquad (153)\\
&\overset{(f)}{\leq} \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i|Z_1^i, Z_2^i, V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}) + n\epsilon_n, \qquad (154)
\end{aligned}$$

where (d) follows from Fano's inequality, (e) follows since $(X_{1,i}, X_{2,i}) = f(M_1, M_2)$, and step (f) follows since conditioning reduces entropy and from the Markov chain $Y_i - (X_{1,i}, X_{2,i}, V_{1,-K+1}^{i-1}, V_{2,-K+1}^{i-1}, Z_1^i, Z_2^i) - (M_1, M_2, Y^{i-1}, V_{1,i}^n, V_{2,i}^n, Z_{1,i+1}^n, Z_{2,i+1}^n)$. From the definition of the RV U, we obtain

$$\begin{aligned}
R_1 + R_2 \leq C_{12} + C_{21} + \frac{1}{n}\sum_{i=1}^n [H(Z_{1,i}, Z_{2,i}|U_i)\\
+ I(X_{1,i}, X_{2,i}; Y_i|Z_{1,i}, Z_{2,i}, U_i)] + \epsilon_n. \qquad (155)
\end{aligned}$$

For the fourth inequality and the proof of the Markov chains, the derivations are similar to the proof for Theorem 1. ■

Appendix B
Proof of Theorem 2

The proof for Theorem 2 follows directly from the proof of Theorem 1 with minor modifications. We use the achievability for the case of the MAC with partial cribbing and a common message, given in [4], but now we set $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2)$ to be

$$\tilde{R}_0 = R_0 + C_{12} + C_{21}, \qquad (156)$$
$$\tilde{R}_1 = R_1 - C_{12}, \qquad (157)$$
$$\tilde{R}_2 = R_2 - C_{21}. \qquad (158)$$

Given these modifications, we obtain the region $\mathcal{R}_0^A$. The achievability for Case B is the same as that for Case A, but now we use code-trees since Z_1 is known causally at Encoder 2. In the converse, we follow the same derivations as in the converse for Theorem 2. For the first three inequalities, we condition all terms upon M_0 and set

$$U_i = (M_0, M_{12}, M_{21}, Z_1^{i-1}, Z_2^{i-1}). \qquad (159)$$

For the forth inequality, we obtain

$$R_0 + R_1 + R_2 = H(M_0, M_1, M_2) \tag{160}$$
$$\leq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i) + n\epsilon_n. \tag{161}$$

This completes the proof. ∎

Appendix C
Proof of Theorem 4

A. Converse

Converse for the strictly causal case: Given an achievable rate-pair (R_1, R_2), we need to show that there exists a joint distribution of the form $P(s)P(u|s)P(z, x_1|u)$ $P(x_2|s, u)P(y|x_1, x_2, s)$ such that the inequalities in (103) are satisfied. Since (R_1, R_2) is an achievable rate-pair, there exists a $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, 2^{nC_{21}}, n)$ code with an arbitrarily small error probability $P_e^{(n)}$. By Fano's inequality,

$$H(M_1, M_2|Y^n, S^n) \leq n(R_1 + R_2)P_e^{(n)} + H(P_e^{(n)}). \tag{162}$$

We set

$$(R_1 + R_2)P_e^{(n)} + \frac{1}{n}H(P_e^{(n)}) \triangleq \epsilon_n, \tag{163}$$

where $\epsilon_n \to 0$ as $P_e^{(n)} \to 0$. Hence,

$$H(M_1|Y^n, M_2, S^n) \leq H(M_1, M_2|Y^n, S^n) \leq n\epsilon_n, \tag{164}$$
$$H(M_2|Y^n, M_1, S^n) \leq H(M_1, M_2|Y^n, S^n) \leq n\epsilon_n. \tag{165}$$

For R_1, we have the following:

$$
\begin{aligned}
nR_1 &= H(M_1) && (166)\\
&= H(M_1|M_{12}) + H(M_{12}) && (167)\\
&\overset{(a)}{=} H(M_1|M_{12}, M_2, S^n) + H(M_{12}) && (168)\\
&= I(M_1; Y^n|M_{12}, M_2, S^n)\\
&\quad + H(M_1|Y^n, M_{12}, M_2, S^n) + H(M_{12}) && (169)\\
&\overset{(b)}{\leq} I(M_1; Y^n|M_{12}, M_2, S^n) + nC_{12} + n\epsilon_n && (170)\\
&\overset{(c)}{=} I(X_1^n, Z^n; Y^n|M_{12}, M_2, S^n) + nC_{12} + n\epsilon_n && (171)\\
&\overset{(d)}{=} I(Z^n; Y^n|M_{12}, M_2, S^n) + I(X_1^n; Y^n|M_{12}, M_2, S^n, Z^n)\\
&\quad + nC_{12} + n\epsilon_n && (172)\\
&\overset{(e)}{=} \sum_{i=1}^{n}[I(Z_i; Y^n|M_{12}, M_{21}, M_2, Z^{i-1}, S^n)\\
&\quad + I(X_1^n; Y_i|Y^{i-1}, M_{12}, M_{21}, M_2, S^n, Z^n)]\\
&\quad + nC_{12} + n\epsilon_n && (173)\\
&\overset{(f)}{\leq} \sum_{i=1}^{n}[H(Z_i|M_{21}, Z^{i-1}, M_{12}, S^{i-1})\\
&\quad + I(X_1^n; Y_i|Y^{i-1}, M_{12}, M_{21}, M_2, S^n, Z^n, X_{2,i})]\\
&\quad + nC_{12} + n\epsilon_n && (174)\\
&\overset{(g)}{\leq} \sum_{i=1}^{n}[H(Z_i|M_{21}, Z^{i-1}, M_{12}, S^{i-1})\\
&\quad + I(X_{1,i}; Y_i|M_{21}, S^i, Z^{i-1}, M_{12}, X_{2,i}, Z_i)]\\
&\quad + nC_{12} + n\epsilon_n && (175)\\
&\overset{(h)}{=} \sum_{i=1}^{n}[H(Z_i|U_i) + I(X_{1,i}; Y_i|U_i, X_{2,i}, S_i, Z_i)]\\
&\quad + nC_{12} + n\epsilon_n, && (176)
\end{aligned}
$$

where (a) follows from the fact that the messages M_1 and (M_2, S^n) are independent, (b) follows from Fano's inequality, (c) follows from the Markov chain $M_1 - (X_1^n, Z^n, M_{12}, M_2, S^n) - Y^n$, (d) and (e) follow from the chain rule and since $M_{21} = f(S^n, M_2, M_{12})$, (f) follows since conditioning reduces entropy and since $X_{2,i} = f(S^n, Z^{i-1}, M_{12}, M_2)$, (g) follows from the Markov Chain $Y_i - (X_{1,i}, X_{2,i}, S^i, M_{12}, M_{21}, Z^i) - (Y^{i-1}, M_2, S_{i+1}^n, Z_{i+1}^n)$, and (h) follows by setting the RV

$$U_i \triangleq (M_{12}, M_{21}, Z^{i-1}, S^{i-1}). \tag{177}$$

Thus, we obtain

$$
\begin{aligned}
R_1 &\leq \frac{1}{n}\sum_{i=1}^{n}[H(Z_i|U_i) + I(X_{1,i}; Y_i|U_i, X_{2,i}, S_i, Z_i)]\\
&\quad + C_{12} + \epsilon_n. && (178)
\end{aligned}
$$

Next, we consider R_2;

$$
\begin{aligned}
nR_2 &= H(M_2) && (179)\\
&\overset{(a)}{=} H(M_2|S^n, M_1) && (180)\\
&\overset{(b)}{=} H(M_{21}, M_2|S^n, M_1) && (181)\\
&= H(M_2|S^n, M_{21}, M_1) + H(M_{21}|S^n, M_1) && (182)\\
&\overset{(c)}{\leq} I(M_2; Y^n|S^n, M_1, M_{21}) +\\
&\quad - I(M_{21}; S^n|M_1) + H(M_{21}|M_1)n\epsilon_n && (183)\\
&\overset{(d)}{\leq} \sum_{i=1}^{n}[I(M_2; Y_i|Y^{i-1}, S^n, M_1, M_{21})\\
&\quad - I(S_i; M_{21}|S^{i-1}, M_1)] + nC_{21} + n\epsilon_n && (184)\\
&\overset{(e)}{=} \sum_{i=1}^{n}[I(M_2, X_{2,i}; Y_i|Y^{i-1}, M_1, M_{12}, M_{21}, S^n, X_{1,i}, Z^{i-1})\\
&\quad - I(S_i; M_{21}, S^{i-1}, M_1, M_{12}, Z^{i-1})] + nC_{21} + n\epsilon_n && (185)\\
&\overset{(f)}{\leq} \sum_{i=1}^{n}[I(X_{2,i}; Y_i|M_{21}, M_{12}, S^i, Z^{i-1}, X_{1,i})\\
&\quad - I(S_i; M_{21}, S^{i-1}, M_{12}, Z^{i-1})] + nC_{21} + n\epsilon_n && (186)\\
&= \sum_{i=1}^{n}[I(X_{2,i}; Y_i|U_i, S_i, X_{1,i}) - I(S_i; U_i)] + nC_{21} + n\epsilon_n, && (187)
\end{aligned}
$$

where (a) follows since M_2 is independent of S^n and M_1, (b) follows since $M_{21} = f(S^n, M_2, M_1)$, (c) follows from Fano's inequality, (d) follows from the chain rule, (e) follows since S_i is independent of (S^{i-1}, M_1) and since $(M_{12}, Z^{i-1}, X_{1,i}) = f(M_{21}, M_1)$, and (f) follows from the same argument as in (175) and since conditioning reduces entropy. Thus, we obtain

$$
\begin{aligned}
R_2 &\leq \frac{1}{n}\sum_{i=1}^{n}[I(X_{2,i}; Y_i|U_i, S_i, X_{1,i}) - I(S_i; U_i)]\\
&\quad + C_{21} + \epsilon_n. && (188)
\end{aligned}
$$

Now, consider

$$
\begin{aligned}
&n(R_1 + R_2)\\
&\quad = H(M_1, M_2, M_{12}) && (189)\\
&\quad \overset{(a)}{=} H(M_1, M_2|S^n, M_{12}) + H(M_{12}) && (190)
\end{aligned}
$$

$$\begin{aligned}
&\leq H(M_1, M_2|M_{21}, S^n, M_{12}) \\
&\quad + H(M_{21}|S^n, M_{12}) + nC_{12} && (191)\\
&\overset{(b)}{\leq} I(M_1, M_2, Z^n; Y^n|S^n, M_{12}, M_{21}) \\
&\quad + H(M_{21}|S^n, M_1) + nC_{12} + n\epsilon_n && (192)\\
&\leq I(M_1, M_2; Y^n|S^n, M_{12}, M_{21}, Z^n) \\
&\quad + I(Z^n; Y^n|S^n, M_{12}, M_{21}) + H(M_{21}|S^n, M_1) \\
&\quad + nC_{12} + n\epsilon_n && (193)\\
&\overset{(c)}{\leq} I(X_1^n, X_2^n; Y^n|S^n, M_{12}, M_{21}, Z^n) + H(M_{21}|S^n, M_1) \\
&\quad + \sum_{i=1}^n H(Z_i|U_i) + nC_{12} + n\epsilon_n && (194)\\
&\overset{(d)}{\leq} I(X_1^n, X_2^n; Y^n|S^n, M_{12}, M_{21}, Z^n) \\
&\quad + \sum_{i=1}^n [H(Z_i|U_i) - I(S_i; U_i)] + nC_{12} + nC_{21} \\
&\quad + n\epsilon_n && (195)\\
&\overset{(e)}{=} \sum_{i=1}^n [I(X_1^n, X_2^n; Y_i|S^n, Y^{i-1}, M_{12}, M_{21}, Z^n) \\
&\quad + H(Z_i|U_i) - I(S_i; U_i)] + nC_{12} + nC_{21} + n\epsilon_n && (196)\\
&\overset{(f)}{\leq} \sum_{i=1}^n [I(X_{1,i}, X_{2,i}; Y_i|S_i, S^{i-1}, M_{21}, M_{12}, Z^i) \\
&\quad + H(Z_i|U_i) - I(S_i; U_i)] + nC_{12} + nC_{21} + n\epsilon_n && (197)\\
&\leq \sum_{i=1}^n [I(X_{1,i}, X_{2,i}; Y_i|S_i, U_i, Z_i) + H(Z_i|U_i) \\
&\quad - I(S_i; U_i)] + nC_{12} + nC_{21} + n\epsilon_n, && (198)
\end{aligned}$$

where (a) follows since (M_1, M_2) is independent of S^n, (b) follows since $Z^n = f(M_1, M_{21})$, (c) follows from the same arguments as those given in (172)-(176) and from the Markov chain $(M_1, M_2) - (X_1^n, X_2^n, M_{12}, M_{21}, Z^n, S^n) - Y^n$, (d) follows from the same arguments as those given in (183)-(186), (e) follows from the chain rule, and (f) follows from the same argument as those given in (175). Thus, we obtain

$$\begin{aligned}
R_1 + R_2 \leq C_{12} + C_{21} + \frac{1}{n}\sum_{i=1}^n [I(X_{1,i}, X_{2,i}; Y_i|S_i, U_i, Z_i) \\
+ H(Z_i|U_i) - I(S_i; U_i)] + \epsilon_n. && (199)
\end{aligned}$$

Additionally,

$$\begin{aligned}
n(R_1 + R_2) &\leq H(M_1, M_2) && (200)\\
&\overset{(a)}{=} H(M_1, M_2|S^n) && (201)\\
&\overset{(b)}{\leq} I(M_1, M_2; Y^n|S^n) + n\epsilon_n && (202)\\
&\overset{(c)}{\leq} I(X_1^n, X_2^n; Y^n|S^n) + n\epsilon_n && (203)\\
&\overset{(d)}{=} \sum_{i=1}^n I(X_1^n, X_2^n; Y_i|S^n, Y^{i-1}) + n\epsilon_n && (204)\\
&\overset{(e)}{\leq} \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i|S_i) + n\epsilon_n, && (205)
\end{aligned}$$

where (a) follows since (M_1, M_2) is independent of S^n, (b) follows from Fano's inequality, (c) follows from encoding relations (96)-(100), (d) follows from the chain rule, and step (e) follows from the Markov Chain $Y_i - (X_{1,i}, X_{2,i}, S_i) - Y^{i-1}$ and since conditioning reduces entropy. Thus, we obtain

$$R_1 + R_2 \leq \frac{1}{n}\sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i|S_i) + \epsilon_n. \quad (206)$$

Finally,

$$\begin{aligned}
nC_{21} &\geq H(M_{21}) && (207)\\
&\geq H(M_{21}|M_1) && (208)\\
&\geq I(M_{21}; S^n|M_1) && (209)\\
&= \sum_{i=1}^n I(S_i; M_{21}|S^{i-1}, M_1) && (210)\\
&\overset{(a)}{=} \sum_{i=1}^n I(S_i; M_{21}, S^{i-1}, M_1) && (211)\\
&\geq \sum_{i=1}^n I(S_i; M_{21}, S^{i-1}, Z^{i-1}, M_{12}) && (212)\\
&= \sum_{i=1}^n I(S_i; U_i), && (213)
\end{aligned}$$

where (a) follows since S_i is independent of (S^{i-1}, M_1). Finally, let Q be an RV independent of (X_1^n, X_2^n, Y^n) and uniformly distributed over the set $\{1, 2, 3, \ldots, n\}$. We define the RV $U \triangleq (Q, U_Q)$ and obtain the region given in (103).

To complete the converse, we need to show the following Markov relations:

- $Z_i - U_i - S_i$, $X_{1,i} - (U_i, Z_i) - S_i$, and $X_{2,i} - (M_{12}, M_{21}, Z^{i-1}, S^{i-1}) - X_{1,i}$ - These Markov relations can be proven by using the undirected graph method in Fig. 11. For the first Markov chain, see that it is impossible to get from node Z_i to node S_i without going through nodes $(S^{i-1}, Z^{i-1}, M_{12}, M_{21})$. For the second Markov chain, it is impossible to get from node $X_{1,i}$ to node S_i without going through nodes $(S^{i-1}, Z^i, M_{12}, M_{21})$. Finally, for the third Markov chain, we can see that it is impossible to get from node $X_{1,i}$ to node $X_{2,i}$ without going through nodes $(S^i, Z^{i-1}, M_{12}, M_{21})$.
- $Y_i - (X_{1,i}, X_{2,i}) - (Z_{1,i}, U_i)$ - Follows from the fact that the channel output at any time i is assumed to depend only on the channel inputs and state at time i.

This completes the converse part. ■

Converse for the Causal Case: For the causal case, we repeat the same converse as for the strictly causal case, except that in the final step we need to show the Markov chain $X_{2,i} - (U_i, Z_i, S_i) - X_{1,i}$, rather than $X_{2,i} - (U_i, S_i) - X_{1,i}$, as in the strictly causal case. If we change node Z^{i-1} to Z^i in Fig. 11, we can see that the Markov chain $X_{2,i} - (M_{12}, M_{21}, Z^i, S^i) - X_{1,i}$ holds since we cannot get from node $X_{2,i}$ to node $X_{1,i}$ without going through nodes $(M_{12}, M_{21}, Z^i, S^i)$. ■

B. Achievability

To prove the achievability, we will consider a similar setting and then, by incorporating a minor modification, we will prove

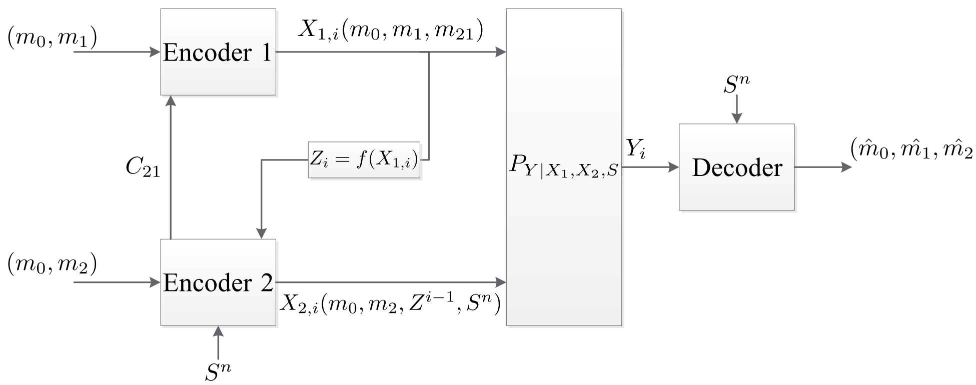

Fig. 10. MAC with a common message and state known at a partially cribbing Encoder.

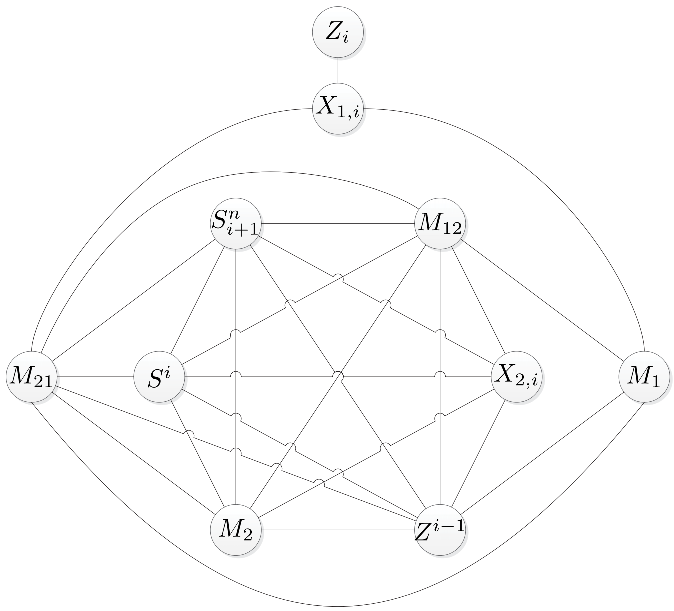

Fig. 11. Proof of the Markov chains $Z_i - U_i - S_i$, $X_{1,i} - (U_i, Z_i) - S_i$, and $X_{2,i} - (M_{12}, M_{21}, Z^{i-1}, S^{i-1}) - X_{1,i}$ using the undirected graphical technique [16, Sec. II]. This graph corresponds to the joint distribution $P(s^n)P(m_1)P(m_2)P(m_{12}|m_1)P(m_{21}|m_2, s^n, m_{21})P(z^{i-1}|m_1, m_{21})$ $P(x_{1,i}|m_{21}, m_1)P(z_i|x_{1,i})P(x_{2,i}|m_{12}, m_2, s^n, z^{i-1})$.

our setting. We first prove the achievability for the strictly causal case.

Achievability for the strictly causal case: Let us look at a similar model depicted in Fig. 10.

First, we will solve the achievability for this model. Fix a joint distribution $P(s)P(u|s)P(z, x_1|u)$ $P(x_2|s, u)P(y|x_1, x_2, s)$ where $P(s)$ and $P(y|x_1, x_2, s)$ are given by the channel. In the following achievability scheme, we use block Markov coding, rate splitting, and double binning.

Coding Scheme: We consider B blocks, each consisting of n symbols; thus we transmit nB symbols. We transmit $B-1$ messages M_1 in B blocks of information. Here, $M_1 \in \{1, \dots, 2^{nR_1}\}$; thus asymptotically, for a large enough n, our transmission rate would be $\frac{nR_1(B-1)}{nB} \overset{n\to\infty}{\longrightarrow} R_1$. At each block, we split messages M_1 and M_2 into (M_1', M_1'') and (M_2', M_2'') at rates (R_1', R_1'') and (R_2', R_2''), respectively. We note that $R_1' + R_1'' = R_1$ and $R_2' + R_2'' = R_2$.

Code Design: The following binning process is depicted in Fig. 12. Generate $2^{n(R_0+R_1'+C_{21})}$ codewords u^n i.i.d. using $P(u^n) = \Pi_{i=1}^n P(u_i)$. Bin all u^ns into $2^{n(R_0+R_1')}$ super-bins. In each super-bin, bin all u^ns into $2^{nR_2'}$ bins. Thus, we have $2^{n(R_0+R_1')}$ super-bins, each consisting of $2^{nR_2'}$ bins, where in each bin we have $2^{n(C_{21}-R_2')}$ u^n codewords. For each u^n, generate $2^{nR_1'}$ codewords z^n i.i.d. using $P(z^n|u^n) = \Pi_{i=1}^n P(z_i|u_i)$. For each pair (u^n, z^n), generate $2^{nR_1''}$ codewords x_1^n i.i.d. using $P(x_1^n|u^n, z^n) = \Pi_{i=1}^n P(x_{1,i}|u_i, z_i)$. Additionally, for each pair (u^n, s^n), generate $2^{nR_2''}$ codewords x_2^n i.i.d. using $P(x_2^n|u^n, s^n) = \Pi_{i=1}^n P(x_{2,i}|u_i, s_i)$.

Encoding: We denote the realizations of the sequences $(M_0, M_1', M_1'', M_2', M_2'')$ at block b as $(m_{0,b}, m_{1,b}', m_{1,b}'', m_{2,b}', m_{2,b}'')$. Since we use block Markov coding, we set $m_{1,B}' = 1$. In block $b \in \{1, \dots, B\}$, Encoder 2 looks in super-bin $(m_{0,b}, m_{1,b-1}')$ and bin $m_{2,b}'$ for u^n such that $(u^n, s^n) \in T_\epsilon^{(n)}(U, S)$ and sends its index l inside the super-bin over the rate-limited cooperation link to Encoder 1, where $l \in \{1, \dots, 2^{nC_{21}}\}$. If such a codeword u^n does not exist, namely, among the codewords in the bin none is jointly typical with s^n, choose an arbitrary u^n from the bin $m_{2,b}'$ (in such a case the decoder will declare an error). Encoder 1 looks in super-bin $(m_{0,b}, m_{1,b}')$ for the bin that $u^n(l)$ lies in. That bin's index is $m_{2,b}'$. Encoder 1 then encodes message $m_{1,b}'$ conditioned on $(m_{0,b}, m_{1,b-1}', m_{2,b}')$ using $z^n(m_{1,b}', u^n)$ and encodes message $m_{1,b}''$ conditioned on $(m_{0,b}, m_{1,b-1}', m_{2,b}', m_{1,b}')$ using $x_1^n(m_{1,b}'', u^n, z^n)$. Encoder 2 encodes message $m_{2,b}''$ conditioned on $(m_{0,b}, m_{1,b-1}', m_{2,b}')$ and s^n using $x_2^n(m_{2,b}, u^n, s^n)$. Send $x_1^n(m_{1,b}'', u^n, z^n)$ and $x_2^n(m_{2,b}'', u^n, s^n)$ over the channel.

Decoding at Encoder 2: At the end of block b, Encoder 2 tries to decode message $m_{1,b}'$. Given $(m_{0,b}, m_{2,b}')$ and assuming that message $m_{1,b-1}'$ was decoded correctly at the end of block $b-1$, Encoder 2 looks for $\hat{m}_{1,b}'$ s.t.

$$(u^n(m_{0,b}, m_{1,b-1}', m_{2,b}'), z^n(\hat{m}_{1,b}', u^n)) \in T_\epsilon^{(n)}(U, Z). \quad (214)$$

If no such $\hat{m}_{1,b}'$, or more than one such $\hat{m}_{1,b}'$, is found, an error is declared at block b and therefore, in the whole superblock nB.

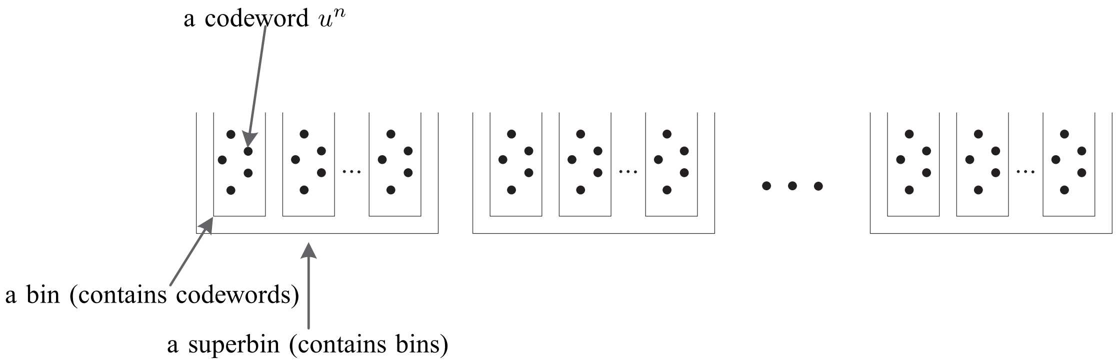

Fig. 12. The binning process as explained in the code design. There are $2^{n(R_0+R_1')}$ super-bins and $2^{nR_2'}$ bins in each super-bin. The number of codewords in each bin must be greater than $I(U;S)$ in order to find u^n such that $(u^n,s^n)\in T_\epsilon^{(n)}(U,S)$.

Decoding at the Receiver: At the end of block B, the decoding is done backwards. At block b, assuming that $(m_{0,b+1},m_{1,b},m'_{2,b+1})$ was decoded correctly in block $b+1$, the decoder looks for the set $(\hat{m}_{0,b},\hat{m}'_{1,b-1},\hat{m}''_{1,b},\hat{m}'_{2,b},\hat{m}''_{2,b})$ s.t.

$$(u^n(\hat{m}_{0,b},\hat{m}'_{1,b-1},\hat{m}'_{2,b},s^n),z^n(\hat{m}'_{1,b},u^n),x_1^n(m''_{1,b},u^n,z^n),$$
$$x_2^n(\hat{m}''_{2,b},u^n,s^n),s^n,y^n)\in T_\epsilon^{(n)}(U,Z,X_1,X_2,S,Y).$$

If no such tuple, or more than one such tuple, is found, an error is declared in block b and therefore, at the whole super-block nB.

Error Analysis: The probability that $z^n(1,u^n)=z^n(i,u^n)$, where $i>1$ and where $(u^n,z^n(1,u^n))\in T_\epsilon^{(n)}(U,Z)$ is bounded by $2^{-n(H(Z|U)-\delta(\epsilon))}$, where $\delta(\epsilon)$ goes to zero as ϵ goes to zero. Hence, if

$$R_1' < H(Z|U), \tag{215}$$

then the probability that an incorrect message $m'_{1,b}$ was decoded goes to zero for a large enough n. To find in super-bin $(\hat{m}_{0,b},\hat{m}'_{1,b-1})$ and in bin $m'_{2,b}$ a codeword u^n that is jointly typical with s^n, we need to have more than $I(U;S)$ codewords in each bin; thus, if

$$C_{21}-R_2' \geq I(U;S), \tag{216}$$

then the probability of finding a codeword u^n such that $(u^n,s^n)\in T_\epsilon^{(n)}(U,S)$ goes to 1 for a large enough n. We define the following event at block b:

$$E_{i,j,k,b}\triangleq(u^n(i,s^n),z^n(\hat{m}'_{1,b},u^n),x_1^n(j,u^n,z^n),$$
$$x_2^n(k,s^n),s^n,y^n)\in T_\epsilon^{(n)}(U,Z,X_1,X_2,S,Y). \tag{217}$$

We can bound the probability of error as follows:

$$\begin{aligned}P_{e,b}^{(n)} &\leq \Pr(E^c_{1,1,1,b})+\sum_{i=1,j=1,k>1}\Pr(E_{1,1,k,b})\\ &\quad+\sum_{i=1,j>1,k=1}\Pr(E_{1,j,1,b})+\sum_{i=1,j>1,k>1}\Pr(E_{1,j,k,b})\\ &\quad+\sum_{i>1,j>1,k>1}\Pr(E_{i,j,k,b}).\end{aligned} \tag{218}$$

We now show that each term in (218) goes to zero for a large enough n.

- Upper-bounding $\Pr(E^c_{1,1,1,b})$: Since we assume that Encoders 1 and 2 encode the correct message-tuple $(m_{0,b},m'_{1,b-1},m''_{1,b},m'_{2,b},m''_{2,b})$ at block b and that the decoder decoded the right $(m_{0,b+1},m'_{1,b},m''_{1,b+1},m'_{2,b+1},m''_{2,b+1})$ at block $b+1$, by the L.L.N., $\Pr(E^c_{1,1,1,b})\to 0$.
- Upper-bounding $\sum_{i=1,j=1,k>1}\Pr(E_{1,1,k,b})$: Assuming that $m'_{1,b}$ was decoded correctly at block $b+1$, the probability for this event is bounded by

$$\sum_{i=1,j=1,k>1}\Pr(E_{1,1,k,b})\leq 2^{nR_2''}2^{-n(I(X_2;Y|S,U,Z,X_1)-\delta(\epsilon)} \tag{219}$$
$$=2^{nR_2''}2^{-n(I(X_2;Y|S,U,X_1)-\delta(\epsilon)}. \tag{220}$$

- Upper-bounding $\sum_{i=1,j>1,k=1}\Pr(E_{1,j,1,b})$: Assuming that $m'_{1,b}$ was decoded correctly at block $b+1$, the probability for this event is bounded by

$$\sum_{i=1,j>1,k=1}\Pr(E_{1,j,1,b})\leq 2^{n(R_1'')}2^{-n(I(X_1;Y|S,U,Z,X_2)-\delta(\epsilon)}. \tag{221}$$

- Upper-bounding $\sum_{i=1,j>1,k>1}\Pr(E_{1,j,k,b})$: Assuming that $m'_{1,b}$ was decoded correctly at block $b+1$, the probability for this event is bounded by

$$\sum_{i=1,j>1,k>1}\Pr(E_{1,j,k,b})\leq 2^{n(R_1''+R_2'')}2^{-n(I(X_1,X_2;Y|S,U,Z)-\delta(\epsilon)}. \tag{222}$$

- Upper-bounding $\sum_{i>1,j>1,k>1}\Pr(E_{i,j,k,b})$: Assuming that $m'_{1,b}$ was decoded correctly at block $b+1$, the probability for this event is bounded by

$$\sum_{i>1,j>1,k>1}\Pr(E_{1,j,k,b})\leq 2^{n(R_0+R_1'+R_1''+R_2'+R_2'')}\cdot 2^{-n(I(U,V,Z,X_1,X_2;Y|S)-\delta(\epsilon)} \tag{223}$$
$$\leq 2^{n(R_0+R_1'+R_1''+R_2'+R_2'')}\cdot 2^{-n(I(X_1,X_2;Y|S)-\delta(\epsilon)}. \tag{224}$$

To summarize, we note that $R_1' = R_1 - R_1''$ and $R_2' = R_2 - R_2''$, and thus, we obtained that if (R_1'', R_2'', R_1, R_2) satisfy

$$R_1 - R_1'' \le H(Z|U), \tag{225}$$
$$R_2 - R_2'' \le C_{21} - I(U;S), \tag{226}$$
$$R_2'' \le I(X_2;Y|S,U,X_1), \tag{227}$$
$$R_1'' \le I(X_1;Y|S,U,Z,X_2), \tag{228}$$
$$R_1'' + R_2'' \le I(X_1,X_2;Y|S,U,Z), \tag{229}$$
$$R_0 + R_1'' + R_2'' \le I(X_1,X_2;Y|S), \tag{230}$$

then there exists a code with a probability of error that goes to zero as the block length goes to infinity. Using the Fourier-Motzkin elimination and by setting $R_1 = \tilde{R}_1, R_0 = \tilde{R}_0$, we obtain the following region

$$\begin{aligned}\tilde{R}_1 &\le H(Z|V,U) + I(X_1;Y|S,U,X_2,Z),\\ R_2 &\le I(X_2;Y|X_1,S,U) + C_{21} - I(U;S),\\ \tilde{R}_1 + R_2 &\le I(X_1,X_2;Y|U,Z,S) + H(Z|U)\\ &\quad + C_{21} - I(U;S),\\ \tilde{R}_0 + \tilde{R}_1 + R_2 &\le I(X_1,X_2;Y|S).\end{aligned} \tag{231}$$

After establishing the capacity region for the MAC with common message, one-sided cooperation and non-causal state known at a partially cribbing encoder and at the decoder, we apply the following modification. If we set

$$\tilde{R}_0 = C_{12}, \tag{232}$$
$$\tilde{R}_1 = R_1 - C_{12}, \tag{233}$$

then the inequalities can be rewritten as

$$\begin{aligned}R_1 - C_{12} &\le H(Z|U) + I(X_1;Y|S,U,X_2,Z),\\ R_2 &\le I(X_2;Y|X_1,V,S,U)\\ &\quad + C_{21} - I(U;S),\\ R_1 - C_{12} + R_2 &\le I(X_1,X_2;Y|U,Z,S)\\ &\quad + H(Z|U) + C_{21} - I(U;S),\\ C_{12} + (R_1 - C_{12}) + R_2 &\le I(X_1,X_2;Y|S),\end{aligned} \tag{234}$$

and thus, we obtain the region in (103). ■

Achievability for the Causal Case: The achievability part follows similar to that of the strictly causal case, but now the generation of X_2^n is done i.i.d. according to the conditional distribution of $p(x_2|u,s,z)$ induced by (104). ■

Appendix D
Proof of Theorem 5

A. Converse

Converse for the Strictly Causal Case: Given an achievable rate-pair (R_1, R_2), we need to show that there exists a joint distribution of the form $P(w)P(v|w)p(a|w)$ $P(s|a)P(x_1|v,w)P(u,x_2|s,v,a,w)$ $P(y|x_1,x_2,s)$ such that the inequalities in (113) are satisfied. Since (R_1, R_2) is an achievable rate-pair, there exists a $(2^{nR_1}, 2^{nR_2}, 2^{nC_{12}}, n)$ code with an arbitrarily small error probability $P_e^{(n)}$. By Fano's inequality,

$$H(M_1, M_2|Y^n) \le n(R_1+R_2)P_e^{(n)} + H(P_e^{(n)}). \tag{235}$$

We set

$$(R_1+R_2)P_e^{(n)} + \frac{1}{n}H(P_e^{(n)}) \triangleq \epsilon_n, \tag{236}$$

where $\epsilon_n \to 0$ as $P_e^{(n)} \to 0$. Hence,

$$H(M_1|Y^n, M_2) \le H(M_1, M_2|Y^n) \le n\epsilon_n, \tag{237}$$
$$H(M_2|Y^n, M_1) \le H(M_1, M_2|Y^n) \le n\epsilon_n. \tag{238}$$

For R_1, we have the following:

$$\begin{aligned}nR_1 &= H(M_1) &(239)\\ &= H(M_1, M_{12}) &(240)\\ &\overset{(a)}{=} H(M_1|M_2,M_{12}) + H(M_{12}) &(241)\\ &\le nC_{12} + I(M_1;Y^n|M_2,M_{12})\\ &\quad + H(M_1|Y^n,M_2,M_{12}) &(242)\\ &\overset{(b)}{\le} nC_{12} + I(M_1;Y^n|M_2,M_{12}) + n\epsilon_n &(243)\\ &\overset{(c)}{=} nC_{12} + I(X_1^n;Y^n|M_2,M_{12}) + n\epsilon_n &(244)\\ &\overset{(d)}{=} nC_{12} + \sum_{i=1}^n I(X_{1,i};Y^n|M_2,X_1^{i-1},M_{12}) + n\epsilon_n &(245)\\ &\le nC_{12} + \sum_{i=1}^n H(X_{1,i}|M_2,X_1^{i-1},M_{12}) + n\epsilon_n &(246)\\ &\overset{(e)}{\le} nC_{12} + \sum_{i=1}^n H(X_{1,i}|V_i,W_i) + n\epsilon_n, &(247)\end{aligned}$$

where (a) follows from the fact that the messages M_1 and M_2 are independent, (b) follows from Fano's inequality, (c) follows from the encoding relation in (107), (d) follows from the chain rule, and step (e) follows since conditioning reduces entropy and by setting the RVs

$$V_i \triangleq X_1^{i-1}, \tag{248}$$
$$W_i \triangleq M_{12}. \tag{249}$$

Thus, we obtain

$$R_1 \le C_{12} + \frac{1}{n}\sum_{i=1}^n H(X_{1,i}|V_i,W_i) + \epsilon_n. \tag{250}$$

Additionally,

$$\begin{aligned}nR_1 &= H(M_1) &(251)\\ &= H(M_1|M_2,M_{12}) + H(M_{12}) &(252)\\ &\overset{(a)}{\le} I(M_1;Y^n|M_2,M_{12}) + nC_{12} + n\epsilon_n &(253)\\ &\overset{(b)}{=} \sum_{i=1}^n I(M_1;Y_i|Y^{i-1},M_2,M_{12}) + nC_{12} + n\epsilon_n &(254)\\ &\le \sum_{i=1}^n I(Y^{i-1},M_1,M_2;Y_i|M_{12}) + nC_{12} + n\epsilon_n &(255)\\ &= \sum_{i=1}^n [I(Y^{i-1},M_1,M_2,S_{i+1}^n;Y_i|M_{12})\\ &\quad - I(S_{i+1}^n;Y_i|M_1,M_2,Y^{i-1},M_{12})]\\ &\quad + nC_{12} + n\epsilon_n &(256)\\ &\overset{(c)}{=} \sum_{i=1}^n [I(Y^{i-1},M_1,M_2,S_{i+1}^n,X_1^{i-1},X_{1,i};Y_i|M_{12})\\ &\quad - I(S_i;Y^{i-1}|M_1,M_2,S_{i+1}^n,M_{12})]\\ &\quad + nC_{12} + n\epsilon_n &(257)\end{aligned}$$

$$\overset{(d)}{=} \sum_{i=1}^{n}[I(Y^{i-1}, M_2, S_{i+1}^n, X_1^{i-1}, X_{1,i}; Y_i|A_i, M_{12}) - I(S_i; Y^{i-1}|M_1, M_2, A_i, S_{i+1}^n, M_{12})] + nC_{12} + n\epsilon_n \tag{258}$$

$$\overset{(e)}{\leq} \sum_{i=1}^{n}[I(Y^{i-1}, M_2, S_{i+1}^n, X_1^{i-1}, X_{1,i}; Y_i|A_i, M_{12}) - I(S_i; Y^{i-1}, M_2, S_{i+1}^n|M_1, A_i, M_{12})] + nC_{12} + n\epsilon_n \tag{259}$$

$$= \sum_{i=1}^{n}[I(Y^{i-1}, M_2, S_{i+1}^n, X_1^{i-1}, X_{1,i}; Y_i|A_i, M_{12}) - I(S_i; Y^{i-1}, M_2, S_{i+1}^n|M_1, A_i, X_1^{i-1}, M_{12})] + nC_{12} + n\epsilon_n \tag{260}$$

$$\overset{(f)}{=} \sum_{i=1}^{n}[I(Y^{i-1}, M_2, S_{i+1}^n, X_1^{i-1}, X_{1,i}; Y_i|A_i, M_{12}) - I(S_i; Y^{i-1}, M_2, S_{i+1}^n|A_i, X_1^{i-1}, M_{12})] + nC_{12} + n\epsilon_n \tag{261}$$

$$\overset{(g)}{=} \sum_{i=1}^{n}[I(V_i, U_i, X_{1,i}; Y_i|A_i, W_i) - I(S_i; U_i|V_i, A_i, W_i)] + nC_{12} + n\epsilon_n, \tag{262}$$

where (a) follows from Fano's inequality, (b) follows from the chain rule, (c) follows since $X_1^i = f(M_1)$ and by using the Csiszar Sum Equality, (d) follows since $A_i = f(M_{12}, M_2)$ and from the Markov Chain $M_1 - (M_{12}, X_{1,i}, X_1^{i-1}, Y^{i-1}, M_2, S_{i+1}^n, A_i, M_{12}) - Y_i$, (e) follows since S_i is independent of (M_2, S_{i+1}^n) given (M_1, A_i), (f) follows from the Markov Chain $M_1 - (M_{12}, X_1^{i-1}, A_i) - (Y^{i-1}, M_2, S_{i+1}^n)$, and (g) follows by setting the RVs W,V and

$$U_i \triangleq (Y^{i-1}, M_2, S_{i+1}^n). \tag{263}$$

Thus, we obtain

$$R_1 \leq \frac{1}{n}\sum_{i=1}^{n}[I(V_i, U_i, X_{1,i}; Y_i|A_i, W_i) - I(S_i; U_i|V_i, A_i, W_i)] + C_{12} + \epsilon_n. \tag{264}$$

Next, we consider R_2

$$nR_2 = H(M_2) \tag{265}$$

$$= H(M_2|M_1) \tag{266}$$

$$\overset{(a)}{\leq} I(M_2; Y^n|M_1) + n\epsilon_n \tag{267}$$

$$= \sum_{i=1}^{n} I(M_2; Y_i|Y^{i-1}, M_1) + n\epsilon_n \tag{268}$$

$$\leq \sum_{i=1}^{n} I(Y^{i-1}, M_2; Y_i|M_1) + n\epsilon_n \tag{269}$$

$$\overset{(b)}{=} \sum_{i=1}^{n}[I(Y^{i-1}, M_2, S_{i+1}^n; Y_i|M_1) - I(S_{i+1}^n; Y_i|M_1, M_2, Y^{i-1})] + n\epsilon_n \tag{270}$$

$$\overset{(c)}{=} \sum_{i=1}^{n}[I(Y^{i-1}, M_2, S_{i+1}^n; Y_i|M_1, M_{12}, X_{1,i}, X_1^{i-1}) - I(S_i;Y^{i-1},M_2, S_{i+1}^n|A_i,M_{12},X_1^{i-1})] + n\epsilon_n \tag{271}$$

$$\overset{(d)}{=} \sum_{i=1}^{n}[I(Y^{i-1}, M_2, S_{i+1}^n, A_i; Y_i|M_1, M_{12}, X_{1,i}, X_1^{i-1}) - I(S_i;Y^{i-1},M_2, S_{i+1}^n|A_i,M_{12},X_1^{i-1})] + n\epsilon_n \tag{272}$$

$$\overset{(e)}{\leq} \sum_{i=1}^{n}[I(Y^{i-1}, M_2, S_{i+1}^n, A_i; Y_i|M_{12}, X_{1,i}, X_1^{i-1}) - I(S_i;Y^{i-1},M_2, S_{i+1}^n|A_i,M_{12},X_1^{i-1})] + n\epsilon_n \tag{273}$$

$$\overset{(f)}{=} \sum_{i=1}^{n}[I(U_i, A_i; Y_i|W_i, X_{1,i}, V_i) - I(S_i; U_i|W_i, V_i, A_i)] + n\epsilon_n, \tag{274}$$

where (a) follows from Fano's inequality, (b) follows from the chain rule, (c) follows since $(M_{12}, X_1^i) = f(M_1)$ and from the same arguments as given in (257) - (262), (d) follows since $A_i = f(M_{12}, M_2)$, (e) follows from the same arguments as given in (258), and (f) follows by setting the RVs U, V and W. Thus, we obtain

$$R_2 \leq \frac{1}{n}\sum_{i=1}^{n}[I(U_i, A_i; Y_i|W_i, X_{1,i}, V_i) - I(S_i; U_i|W_i, V_i, A_i)] + \epsilon_n. \tag{275}$$

Now, consider

$$n(R_1 + R_2) = H(M_1, M_2) \tag{276}$$

$$= H(M_1, M_2|M_{12}) + H(M_{12}) \tag{277}$$

$$\overset{(a)}{\leq} I(M_1, M_2; Y^n|M_{12}) + nC_{12} + n\epsilon_n \tag{278}$$

$$\overset{(b)}{=} \sum_{i=1}^{n} I(M_1, M_2;Y_i|Y^{i-1}, M_{12}) + nC_{12} + n\epsilon_n \tag{279}$$

$$\overset{(c)}{\leq} \sum_{i=1}^{n}[I(M_1, Y^{i-1}, M_2, S_{i+1}^n; Y_i|M_{12}) - I(Y^{i-1}, M_2, S_{i+1}^n; S_i|M_{12}, A_i, X_1^{i-1})] + nC_{12} + n\epsilon_n \tag{280}$$

$$\overset{(d)}{=} \sum_{i=1}^{n}[I(M_1,X_{1,i},X_1^{i-1},Y^{i-1},M_2,S_{i+1}^n;Y_i|M_{12}) - I(Y^{i-1}, M_2, S_{i+1}^n; S_i|M_{12}, A_i, X_1^{i-1})] + nC_{12} + n\epsilon_n \tag{281}$$

$$\overset{(e)}{=} \sum_{i=1}^{n}[I(X_{1,i},X_1^{i-1},Y^{i-1},M_2,S_{i+1}^n,A_i;Y_i|M_{12}) - I(Y^{i-1}, M_2, S_{i+1}^n; S_i|M_{12}, A_i, X_1^{i-1})] + nC_{12} + n\epsilon_n \tag{282}$$

$$= \sum_{i=1}^{n}[I(U_i, V_i, X_{1,i}, A_i; Y_i|W_i) - I(U_i; S_i|V_i, A_i, W_i)] + nC_{12} + n\epsilon_n, \tag{283}$$

where (a) follows from Fano's inequality, (b) follows from the chain rule, (c) follows from the same arguments as given in (257)-(262), (d) follows since $X_1^i = f(M_1)$ and $A_i = f(M_{12}, M_2)$, and (e) follows from the same arguments as given in (258). Thus we obtain

$$R_1 + R_2 \leq \frac{1}{n}\sum_{i=1}^{n}[I(U_i, V_i, X_{1,i}, A_i; Y_i|W_i) - I(U_i; S_i|V_i, A_i|W_i)] + C_{12} + \epsilon_n. \tag{284}$$

Again,

$$\begin{aligned}
&n(R_1+R_2)\\
&\quad = H(M_1,M_2) &(285)\\
&\quad \overset{(a)}{\leq} I(M_1,M_2;Y^n)+n\epsilon_n &(286)\\
&\quad \overset{(b)}{=} \sum_{i=1}^{n} I(M_1,M_2;Y_i|Y^{i-1})+n\epsilon_n &(287)\\
&\quad \overset{(c)}{\leq} \sum_{i=1}^{n}[I(M_1,Y^{i-1},M_2,S_{i+1}^n;Y_i)\\
&\qquad - I(Y^{i-1},M_2,S_{i+1}^n;S_i|M_{12},A_i,X_1^{i-1})]+n\epsilon_n &(288)\\
&\quad \overset{(d)}{=} \sum_{i=1}^{n}[I(M_1,M_{12},X_{1,i},X_1^{i-1},Y^{i-1},M_2,S_{i+1}^n;Y_i)\\
&\qquad - I(Y^{i-1},M_2,S_{i+1}^n;S_i|M_{12},A_i,X_1^{i-1})]+n\epsilon_n &(289)\\
&\quad \overset{(e)}{=} \sum_{i=1}^{n}[I(M_{12},X_{1,i},X_1^{i-1},Y^{i-1},M_2,S_{i+1}^n,A_i;Y_i)\\
&\qquad - I(Y^{i-1},M_2,S_{i+1}^n;S_i|M_{12},A_i,X_1^{i-1})]+n\epsilon_n &(290)\\
&\quad \leq \sum_{i=1}^{n}[I(W_i,U_i,V_i,X_{1,i},A_i;Y_i)\\
&\qquad - I(U_i;S_i|W_i,V_i,A_i)]+n\epsilon_n, &(291)
\end{aligned}$$

d where (a) follows from Fano's inequality, (b) follows from the chain rule, (c) follows from the same arguments as given in (257)-(262), (d) follows since $X_1^i = f(M_1)$ and $A_i = f(M_{12},M_2)$, and (e) follows from the same arguments as given in (258). Thus, we obtain

$$\begin{aligned}
R_1+R_2 \leq \frac{1}{n}\sum_{i=1}^{n}&[I(W_i,U_i,V_i,X_{1,i},A_i;Y_i)\\
&-I(U_i;S_i|W_i,V_i,A_i)]+\epsilon_n. \qquad (292)
\end{aligned}$$

Finally, we need to prove the following Markov chains:

- $A_i - W_i - V_i$ -

$$\begin{aligned}
p(a_i|m_{12},x_1^{i-1}) &= \sum_{m_2\in\mathcal{M}_2} p(m_2|m_{12},x_1^{i-1})\\
&\qquad \cdot p(a_i|m_{12},m_2,x_1^{i-1})\\
&\overset{(a)}{=} \sum_{m_2\in\mathcal{M}_2} p(m_2|m_{12})p(a_i|m_{12},m_2)\\
&= p(a_i|m_{12}), \qquad (293)
\end{aligned}$$

  where (a) follows since m_2 is independent of m_1 and since $a_i = f(m_2,m_{12})$.
- $S_i - A_i - (W_i,V_i)$ - Follows from the fact that the channel state at any time i is assumed to depend only on the action at time i.
- $X_{1,i} - (V_i,W_i) - (A_i,S_i)$ -

$$\begin{aligned}
p(x_{1,i}|m_{12},x_1^{i-1},a_i,s_i) &= \sum_{m_1\in\mathcal{M}_1} p(m_1|m_{12},x_1^{i-1},a_i,s_i)\\
&\qquad \cdot p(x_{1,i}|m_{12},m_1,x_1^{i-1},a_i,s_i)\\
&\overset{(a)}{=} \sum_{m_1\in\mathcal{M}_1} p(m_1|m_{12},x_1^{i-1})\\
&\qquad \cdot p(x_{1,i}|m_1,m_{12},x_1^{i-1})\\
&= p(x_{1,i}|m_{12},x_1^{i-1}), \qquad (294)
\end{aligned}$$

  where (a) follows since m_1 is independent of (a_i,s_i) given (m_{12},x_1^{i-1}) and since $x_{1,i} = f(m_1)$.
- $(U_i,X_{2,i}) - S_i,A_i,W_i,V_i - X_{1,i}$ -

$$\begin{aligned}
&p(x_{1,i}|m_{12},x_1^{i-1},a_i,s_i^n,y^{i-1},m_2,x_{2,i})\\
&\quad = \sum_{m_1\in\mathcal{M}_1} p(m_1|m_{12},x_1^{i-1},a_i,s_i^n,y^{i-1},m_2,x_{2,i})\\
&\qquad \cdot p(x_{1,i}|m_1,m_{12},x_1^{i-1},a_i,s_i^n,y^{i-1},m_2,x_{2,i})\\
&\quad \overset{(a)}{=} \sum_{m_1\in\mathcal{M}_1} p(m_1|m_{12},x_1^{i-1},a_i,s_i)\\
&\qquad \cdot p(x_{1,i}|m_1,m_{12},x_1^{i-1},a_i,s_i)\\
&\quad = p(x_{1,i}|m_{12},x_1^{i-1},a_i,s_i), \qquad (295)
\end{aligned}$$

  where (a) follows since m_1 is independent of $(s_{i+1}^n,y^{i-1},m_2,x_{2,i})$ given $(m_{12},x_1^{i-1},a_i,s_i)$ and since $x_{1,i} = f(m_1)$.
- $Y_i - (X_{1,i},X_{2,i},S_i) - (W_i,V_i,U_i,A_i)$ - Follows from the fact that the channel output at any time i is assumed to depend only on the channel inputs and state at time i.

Finally, let Q be an RV independent of (X_1^n,X_2^n,Y^n) and uniformly distributed over the set $\{1,2,3,\ldots,n\}$. We define the RV $W \triangleq (Q,W_Q)$ and obtain the region given in (113). ■

Converse for the Causal Case: For the causal case we repeat the same approach as for the strictly causal case, except that in the final step we need to show the Markov chain $U_i - (S_i,A_i,W_i,V_i) - X_{1,i}$. We can see from the following derivations that this Markov chain holds

$$\begin{aligned}
&p(x_{1,i}|m_{12},x_1^{i-1},a_i,s_i^n,y^{i-1},m_2)\\
&\quad = \sum_{m_1\in\mathcal{M}_1} p(m_1|m_{12},x_1^{i-1},a_i,s_i^n,y^{i-1},m_2)\\
&\qquad \cdot p(x_{1,i}|m_1,m_{12},x_1^{i-1},a_i,s_i^n,y^{i-1},m_2)\\
&\quad \overset{(a)}{=} \sum_{m_1\in\mathcal{M}_1} p(m_1|m_{12},x_1^{i-1},a_i,s_i)\\
&\qquad \cdot p(x_{1,i}|m_1,m_{12},x_1^{i-1},a_i,s_i)\\
&\quad = p(x_{1,i}|m_{12},x_1^{i-1},a_i,s_i), \qquad (296)
\end{aligned}$$

where (a) follows since m_1 is independent of (s_{i+1}^n,y^{i-1},m_2) given $(m_{12},x_1^{i-1},a_i,s_i)$ and since $x_{1,i} = f(m_1)$. ■

B. Achievability

Achievability for the Strictly Causal Case: Fix a joint distribution $P(w)P(v|w)P(a|w)P(s|a)P(x_1|v,w)P(u|s,w,v,a)$ $p(x_2|w,a,v,u,s)P(y|x_1,x_2,s)$ where $P(s|a)$ and $P(y|x_1,x_2,s)$ are given by the channel. In the following achievability scheme, we use block Markov coding, rate splitting, and Gelfand-Pinsker coding.

Coding Scheme: We consider B blocks, each consisting of n symbols; thus, we transmit nB symbols. We transmit $B-1$ messages M_1 in B blocks of information. Here, $M_1 \in \{1,\ldots,2^{nR_1}\}$; thus, asymptotically, for a large enough n, our transmission rate would be $\frac{nR_1(B-1)}{nB} \xrightarrow{n\to\infty} R_1$. We also split message M_1 into (M_1', M_1'') such that $(R_1', R_1'') = (C_{12}, R_1 - C_{12})$.

Code Design: Generate $2^{nR_1'}$ codewords w^n i.i.d. using $P(w^n) = \Pi_{i=1}^n P(w_i)$. For each w^n, generate $2^{nR_1''}$ codewords v^n i.i.d. using $P(v^n|w^n) = \Pi_{i=1}^n P(v_i|w_i)$. For each w^n, generate 2^{nR_2} codewords a^n i.i.d. using $P(a^n|w^n) = \Pi_{i=1}^n P(a_i|w_i)$. For each pair (w^n, v^n), generate $2^{nR_1''}$ codewords x_1^n i.i.d. using $P(x_1^n|v^n, w^n) = \Pi_{i=1}^n P(x_{1,i}|v_i, w_i)$. Additionally, for each triplet (w^n, v^n, a^n), generate $2^{n(R_2+\tilde{R})}$ codewords u^n i.i.d. using $P(u^n|a^n, v^n, w^n) = \Pi_{i=1}^n P(u_i|a_i, v_i, w_i)$. Randomly bin all u^n codewords into 2^{nR_2} bins where each bin contains $2^{n\tilde{R}}$ codewords.

Encoding: We denote the realizations of the messages (M_1', M_1'', M_2) at block b as $(m_{1,b}', m_{1,b}'', m_{2,b})$. Since we use block Markov coding, we set $m_{1,B} = 1$. In block $b \in \{1, \ldots, B\}$, send $m_{1,b}'$ from Encoder 1 to Encoder 2 via the rate-limited cooperation link. Encode message $m_{1,b}'$ using $w^n(m_{1,b}')$. Encode message $m_{1,b-1}''$ conditioned on $m_{1,b}'$ using $v^n(m_{1,b-1}'', w^n)$ and encode message $m_{1,b}''$ conditioned on $(m_{1,b-1}'', m_{1,b}')$ using $x_1^n(m_{1,b}'', v^n, w^n)$. Given $(m_{1,b}', m_{2,b})$, Encoder 2 chooses an action sequence a^n. Given (s^n, w^n, v^n, a^n), look in bin $m_{2,b}$ for a codeword $u^n(w^n,\ v^n,\ a^n,\ m_{2,b},\ l)$ that is jointly typical with $(w^n(m_{1,b}'), v^n(m_{1,b-1}''), s^n, a^n(m_{2,b}))$, where $l \in \{1, \ldots, 2^{n\tilde{R}}\}$. Send $x_1^n(m_{1,b}'', w^n, v^n)$ and x_2^n according to $p(x_2|w, v, u, s)$ i.i.d. over the channel.

Decoding at Encoder 2: At the end of block b, Encoder 2 tries to decode message $m_{1,b}''$. Given $m_{1,b}'$ and assuming that message $m_{1,b-1}''$ was decoded correctly at the end of block $b-1$, Encoder 2 looks for $\hat{m}_{1,b}''$ s.t.

$$(w^n(m_{1,b}'), v^n(m_{1,b-1}'', w^n), x_1^n(\hat{m}_{1,b}'', w^n, v^n)) \in T_\epsilon^{(n)}(W, V, X_1). \tag{297}$$

If no such $\hat{m}_{1,b}'$, or more than one such $\hat{m}_{1,b}'$, is found, an error is declared at block b and therefore, in the whole super-block nB.

Decoding at the Receiver: At the end of block B, the decoding is done backwards. At block b, assuming that $m_{1,b}$ was decoded correctly in block $b+1$, the decoder looks for the triplet $(m_{1,b}', m_{1,b-1}'', \hat{m}_{2,b})$ s.t.

$$\begin{aligned}&(w^n(\hat{m}_{1,b}'), v^n(\hat{m}_{1,b-1}'', w^n), x_1^n(m_{1,b}'', w^n, v^n),\\ &\quad a^n(\hat{m}_{2,b}, w^n), u^n(\hat{m}_{2,b}, w^n, v^n, s^n, a^n, l), y^n)\\ &\quad \in T_\epsilon^{(n)}(W, V, X_1, A, U, Y).\end{aligned} \tag{298}$$

If no such pair, or more than one such pair, is found, an error is declared at block b and therefore, in the whole super-block nB.

Error Analysis: Without loss of generality, we assume that $(m_{1,b}', m_{1,b-1}'', m_{2,b}) = (1, 1, 1)$. The probability that $x_1^n(1, w^n, v^n) = x_1^n(i, w^n, v^n)$ where $i > 1$ and where $(w^n(1), v^n(1, w^n), x_1^n(1, w^n, v^n)) \in T_\epsilon^{(n)}(W, V, X_1)$ is bounded by $2^{-n(H(X_1|V,W)-\delta(\epsilon))}$, where $\delta(\epsilon)$ goes to zero as ϵ goes to zero. Hence, if

$$R_1 - C_{12} < H(X_1|V, W), \tag{299}$$

then the probability that an incorrect message $m_{1,b}$ was decoded goes to zero for a large enough n. We define the following event at block b:

$$\begin{aligned}E_{i,j,k,l,b} \triangleq\ &(w^n(i), v^n(j, w^n), x_1^n(\hat{m}_{1,b}'', v^n, w^n),\\ &a^n(k, w^n), u^n(k, v^n, s^n, a^n, w^n, l), y^n)\\ &\in T_\epsilon^{(n)}(W, V, X_1, A, U, Y).\end{aligned} \tag{300}$$

We can bound the probability of error as follows:

$$\begin{aligned}P_{e,b}^{(n)} \le\ &\Pr(E_{1,1,1,1,b}^c) + \sum_{\substack{i=1,j=1\\k>1,l>1}} \Pr(E_{1,1,k,l,b})\\ &+ \sum_{\substack{i=1,j>1\\k=1,l>1}} \Pr(E_{1,j,1,l,b}) + \sum_{\substack{i=1,j>1\\k>1,l>1}} \Pr(E_{1,j,k,l,b})\\ &+ \sum_{\substack{i>1,j>1\\k>1,l>1}} \Pr(E_{i,j,k,l,b}).\end{aligned} \tag{301}$$

We now show that each term in (301) goes to zero for a large enough n.

- Upper-bounding $\Pr(E_{1,1,1,1,b}^c)$: Since we assume that Transmitters 1 and 2 encode the correct message triplet $(m_{1,b}', m_{1,b-1}'', m_{2,b})$ at block b and that the receiver decoded the right $(m_{1,b+1}', m_{1,b}'', m_{2,b+1})$ at block $b+1$, by the LLN, $\Pr(E_{1,1,1,b}^c) \to 0$.
- Upper-bounding $\sum_{\substack{i=1,j=1\\k>1,l>1}} \Pr(E_{1,1,k,l,b})$: Assuming that $m_{1,b}''$ was decoded correctly at block $b+1$, the probability for this event is bounded by
$$\sum_{\substack{i=1,j=1\\k>1,l>1}} \Pr(E_{1,1,k,l,b}) \le 2^{n(R_2+\tilde{R})} 2^{-n(I(U,A;Y|W,V,X_1)-\delta(\epsilon)}. \tag{302}$$
- Upper-bounding $\sum_{\substack{i=1,j>1\\k=1,l>1}} \Pr(E_{1,j,1,l,b})$: Similar to (302) we obtain
$$\sum_{i=1,j>1,k=1,l>1} \Pr(E_{1,j,1,l,b}) \le 2^{n(R_1-C_{12}+\tilde{R})} \cdot 2^{-n(I(V,X_1,U;Y|W,A)-\delta(\epsilon)}. \tag{303}$$
- Upper-bounding $\sum_{\substack{i=1,j>1\\k>1,l>1}} \Pr(E_{1,j,k,l,b})$: Similar to (302) we obtain
$$\sum_{i=1,j>1,k>1,l>1} \Pr(E_{1,j,k,l,b}) \le 2^{n(R_1-C_{12}+R_2+\tilde{R})} \cdot 2^{-n(I(U,A,V,X_1;Y|W)-\delta(\epsilon)}. \tag{304}$$
- Upper-bounding $\sum_{\substack{i>1,j>1\\k>1,l>1}} \Pr(E_{1,j,k,l,b})$: Similar to (302) we obtain
$$\sum_{i>1,j>1,k>1,l>1} \Pr(E_{1,j,k,l,b}) \le 2^{n(R_1+R_2+\tilde{R})} \cdot 2^{-n(I(U,A,W,V,X_1;Y)-\delta(\epsilon)}. \tag{305}$$

Finally, we analyze the probability of error for finding u^n at Encoder 2. By the covering lemma, if

$$\tilde{R} > I(U; S|W, V, A) \tag{306}$$

then with high probability, in block b we can find a codeword u^n that is jointly typical with s^n in bin number $m_{2,b}$.

The combination of (299), (302), (303), (304), (305), and (306) yields the capacity region in (113), thus completing the proof. ■

Achievability for the Causal Case: The achievability part follows similar to that of the strictly causal case, but now the generation of X_2^n is done i.i.d. according to the conditional distribution of $p(x_2|w, v, u, s, x_1)$ induced by (114). ■

ACKNOWLEDGEMENT

The authors would like to thank the Associate Editor and the reviewers for their valuable comments and suggestions that helped improving this paper.

REFERENCES

[1] F. M. J. Willems, "Information-theoretical results for the discrete memoryless multiple access channel," Ph.D. dissertation, Dept. Wiskunde, Katholieke Univ. Leuven, Leuven, Belgium, 1982.

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

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

[4] 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.

[5] 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.

[6] X. Lin, J. Andrews, A. Ghosh, and R. Ratasuk, "An overview of 3GPP device-to-device proximity services," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 40–48, Apr. 2014.

[7] Y. Cao and B. Chen, "The cognitive radio channel: From spectrum sensing to message cribbing," *Ad Hoc Netw.*, vol. 28, pp. 868–883, Sep. 2009.

[8] J. Shimonovich, A. S. Baruch, and S. Shamai (Shitz), "Cognitive cooperative communications on the multiple access channel," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2013, pp. 1–5.

[9] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1037–1076, 1973.

[10] A. Bracher and A. Lapidoth, "Feedback, cribbing, and causal state information on the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7627–7654, Dec. 2014.

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

[12] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, vol. 4, no. 1, pp. 142–163, 1959.

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

[14] Y. Steinberg, "Channels with cooperation links that may be absent," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, pp. 1947–1951.

[15] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5396–5411, Nov. 2010.

[16] 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.

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

[18] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

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

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

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

[22] 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.

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

[24] S. I. Bross and A. Lapidoth, "The state-dependent multiple-access channel with states available at a cribbing encoder," in *Proc. IEEE 26th Conv. Elect. Electron. Eng. Israel (IEEEI)*, Nov. 2010, pp. 665–669. [Online]. Available: http://arxiv.org/abs/1007.3896

[25] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú, "Cooperative multiple-access encoding with states available at one transmitter," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4448–4469, Oct. 2008.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2012.

[27] Z. Goldfeld, H. H. Permuter, and B. M. Zaidel, "The finite state MAC with cooperative encoders and delayed CSI," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6181–6203, Oct. 2014.

Tal Kopetz (S'14) received his B.Sc. (magna cum laude) degree in Electrical and Computer Engineering from Ben-Gurion University of the Negev, Israel, in 2013. He is studying his M.Sc. degree on the fast track program for honor students in Electrical and Computer Engineering at that same institution.

Tal is a recipient of the Deans List Award, and excellence scholarships from the Electrical and Computer Engineering department, Ben-Gurion University of the Negev, Israel.

Tal is currently working as an Imaging and Computer Vision Algorithms Developer at CEVA, Inc.

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

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

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

Shlomo (Shitz) Shamai (F'94) 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 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 and the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering.

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*), and the 2015 IEEE Communications Society Best Tutorial Paper Award. 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 served on the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY.