Multiple-Access Channel With Partial and Controlled Cribbing Encoders

Himanshu Asnani, *Student Member, IEEE*, and Haim H. Permuter, *Member, IEEE*

***Abstract*—In this paper, we consider a multiple-access channel (MAC) with partial cribbing encoders. This means that each of the two encoders obtains a deterministic function of the output of the other encoder with or without delay. The partial cribbing scheme is especially motivated by the additive noise Gaussian MAC, where perfect cribbing results in the degenerated case of full cooperation between the encoders and requires an infinite entropy link. We derive a single-letter characterization of the capacity of the MAC with partial cribbing for the cases of causal and strictly causal cribbing. Several numerical examples, such as those of quantized cribbing, are presented. We further consider and derive the capacity region where the cribbing depends on actions that are functions of the previous cribbed observations. In particular, we consider a scenario where the action is taken to decide "to crib or not to crib" and show that a naive time-sharing strategy is not optimal.**

***Index Terms*—Backward decoding, block-Markov coding, cribbing encoders, cribbing with actions, Gaussian multiple-access channel (MAC), partial cribbing, quantized cribbing, rate splitting, superposition codes, "To crib or not to crib".**

I. Introduction

In his remarkable dissertation [1], Willems introduced a new problem of the multiple-access channel (MAC) with cribbing encoders and derived its capacity region using a novel decoding technique called "backward decoding." "Cribbing encoder" (or equivalently "perfect cribbing encoder") refers to the encoder which knows the output of the other encoder perfectly, possibly with delay or lookahead. The work by Willems on MACs with cribbing encoders has been extended to the interference channel [2], and to state-dependent MAC [3]. However, for the Gaussian case, where the encoder output takes values in a continuous alphabet, the problem of perfect cribbing is degenerate [4] as it implies full cooperation between the encoders, regardless of the delay in the cribbing. This is due to the fact that in a single epoch, a noiseless continuous signal may transmit an infinite amount of information. Motivated by this fact, we introduce "*partial cribbing*" in this paper, where one encoder only knows a quantized version, or more generally, a deterministic function of the coded output of the other encoder.

In this paper, we consider two kinds of partial cribbing: causal and strictly causal. Causal partial cribbing refers to the setting where at time i, the encoder observes (and uses) the partial cribbing signal without delay, i.e., Z_i. Strictly causal partial cribbing means that at time i, the encoder observes the partial cribbing with a delay, i.e., Z_{i-1}. We derive the capacity region for two different cases according to the causality or the strictly causality of the cribbing.

Case A: The cribbing for both encoders is *strictly causal*.
Case B: The cribbing for one encoder is *causal* and for the other encoder is *strictly causal*.

Fig. 1 depicts the case where one encoder has causal partial cribbing and the other strictly causal partial cribbing, i.e., Case B. To some extent, the partial cribbing problem is related to the semideterministic relay channel [5], which was solved using the "partial decode and forward" technique [6]. When Encoder 2 has no message, the setup is that of a deterministic relay channel and Encoder 2 plays the role of a relay. However, the MAC with partial cribbing is different from the semideterministic relay in the sense that Encoder 2 has its own message to transmit in addition to its role of relaying information from Encoder 1. Another related problem is the semideterministic broadcast channel [7], where one of the receivers obtains a deterministic function of the input channel. In our problem, Encoder 1 "is broadcasting" to Encoder 2 and to the decoder. Thus, this part of the communication resembles the semideterministic broadcast channel. However, in our problem of partial cribbing only the decoder is actually required to decode the message error-free.

The coding scheme presented here for the partial cribbing uses similar techniques to those that were used for the case of perfect cribbing, i.e., block Markov coding, Shannon's strategies, superposition coding, and backward decoding. In addition, we use rate splitting on top of the rate splitting that is inherited in the block-Markov coding. Rate splitting is needed as Encoder 2 can decode only part of the message transmitted by Encoder 1.

Recently, several problems on "action" in information theory have been considered in [8]–[11]. In these problems, the side information is not freely available, but it depends on a cost-constrained action taken by the encoder or the decoder. In this paper, we also consider the case where the cribbing is action dependent. Namely, there is an action that is a function of the previously cribbed observations and this action determines the current cribbing function. These kinds of questions may be raised in cognitive communication systems where sensing other users' signals is a resource with a cost. In particular, we show through a simple example, where the action is "to crib or not to crib," that a naive time-sharing action scheme is not necessarily optimal.

Manuscript received March 20, 2011; revised February 22, 2012; accepted October 24, 2012.Date of publication December 03, 2012; date of current version March 13, 2013. H. Permuter was supported in part by a Marie-Curie fellowship. H. Asnani was supported by The Scott A. and Geraldine D. Macomber Stanford Graduate Fellowship Fund. This paper was presented in part at the 2011 IEEE International Symposium on Information Theory.

H. Asnani is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-3096 USA (e-mail: asnani@stanford.edu).

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

Communicated by E. Erkip, 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.2012.2231462

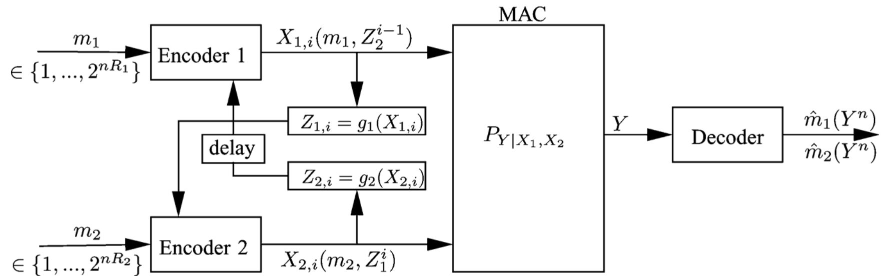


Fig. 1. Partial (deterministic-function) cribbing. Each encoder observes a deterministic function of the other encoder with or without delay. Encoder 1 observes the cribbing in a strictly causal way, i.e., with delay, and Encoder 2 observes the cribbing causally, i.e., without delay. The setting corresponds to Case B in this paper.

The remainder of this paper is organized as follows. In Section II, we introduce the setting of MAC with partial cribbing and state the capacity region for the strictly causal case (Case A), as well as the case of mixed causal and strictly causal cribbing (Case B). In Sections III and IV, respectively, we provide the converse and achievability proofs of the capacity region for each case. In Section V, we consider the case where a common message, known to the encoders, needs to be transmitted to the decoder in addition to the private messages. We show that no additional auxiliary random variable is needed to characterize the capacity region since the partial cribbing is utilized via generating a common message between the users. In Section VI, we consider the case where one of the encoders has no message to send. Hence, it becomes a special case of the semideterministic relay channel with and without delay. We show that, indeed, the region obtained via partial cribbing and the region obtained via a semideterministic relay channel coincide. In Section VII, we consider a Gaussian MAC with quantized cribbing. We provide a simple achievable scheme and show numerically that even with a quantizer of a few bits, we obtain an achievable region that is very close to the perfect cribbing capacity region. In Section VIII, we consider a scenario where a limited-resource action controls the cribbing. In particular, we investigate an example where the action is "to crib or not to crib" and solve it analytically. In Section IX, we conclude this paper and suggest some research directions that have yet been solved, such as noncausal partial cribbing, noisy cribbing, and a few action-related problems.

II. Problem Definition and Main Results

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. The input to the channel from encoder $l \in \{1,2\}$ is denoted by $\{X_{l,1}, X_{l,2}, X_{l,3}, \dots\}$. Encoders 1 and 2 obtain a deterministic function of the output of Encoders 2 and 1, respectively, of the form $Z_{2,i} = g_2(X_{2,i})$, and $Z_{1,i} = g_1(X_{1,i})$. 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 in the following way: $x_l^i = (x_{l,1}, x_{l,2}, \dots, x_{l,i})$, $l \in \{1,2\}$. Since the settings in this paper do not include feedback from the receiver to the transmitters, i.e., $P(x_{1,i}, x_{2,i}|x_1^{i-1}, x_2^{i-1}, y^{i-1}) = P(x_{1,i}, x_{2,i}|x_1^{i-1}, x_2^{i-1})$, (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 $(2^{nR_1}, 2^{nR_2}, n)$ *code* with partial cribbing, as shown in Fig. 1, at time i, consists of an encoding function at Encoder 1

$$\text{Cases A, B,} \quad f_{1,i} : \{1,\dots,2^{nR_1}\} \times \mathcal{Z}_2^{i-1} \mapsto X_{1,i} \tag{3}$$

and an encoding function at Encoder 2 that changes according to the following case settings:

$$\begin{aligned} &\text{Case A} \quad f_{2,i} : \{1,\dots,2^{nR_2}\} \times \mathcal{Z}_1^{i-1} \mapsto X_{1,i}, \\ &\text{Case B} \quad f_{2,i} : \{1,\dots,2^{nR_2}\} \times \mathcal{Z}_1^{i} \mapsto X_{1,i}, \end{aligned} \tag{4}$$

and a decoding function

$$g : \mathcal{Y}^n \mapsto \{1,\dots,2^{nR_1}\} \times \{1,\dots,2^{nR_2}\}. \tag{5}$$

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

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{m_1,m_2} \Pr\{g(Y^n) \neq (m_1, m_2)|(m_1, m_2)\,\text{sent}\}. \tag{6}$$

A rate (R_1, R_2) is said to be *achievable* for the encoders with partial cribbing if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes with $P_e^{(n)} \to 0$. The *capacity region* of the MAC is the closure of all achievable rates.

Let us define the following regions $\mathcal{R}_A$, $\mathcal{R}_B$, which are contained in $\mathbb{R}_+^2$, i.e., the set of nonnegative 2-D real numbers:

$$\mathcal{R}_A = \left\{ \begin{array}{l} R_1 \le H(Z_1|U) + I(X_1;Y|X_2,Z_1,U), \\ R_2 \le H(Z_2|U) + I(X_2;Y|X_1,Z_2,U), \\ R_1+R_2 \le I(X_1,X_2;Y|U,Z_1,Z_2) + H(Z_1,Z_2|U), \\ R_1+R_2 \le I(X_1,X_2;Y), \text{ for some} \\ P(u)P(x_1,z_1|u)P(x_2,z_2|u)P(y|x_1,x_2). \end{array} \right\}. \tag{7}$$

Throughout the paper, the distributions $P(x_1, z_1|u)$, and $P(x_2, z_2|u)$ are restricted to the forms $P(x_1, z_1|u) = P(x_1|u)\mathbf{1}_{\{z_1=g_1(x_1)\}}$ and $P(x_2, z_2|u) = P(x_2|u)\mathbf{1}_{\{z_2=g_2(x_2)\}}$, where $\mathbf{1}_{\{.\}}$ is the indicator function, and this is because as part of the problem settings $z_{1,i} = g_1(x_{1,i})$ and $z_{2,i} = g_2(x_{2,i})$.

The region $\mathcal{R}_B$ is defined with the same set of inequalities as in (7), only that the second inequality in (7) is replaced by

$$R_2 \le H(Z_2|Z_1, U) + I(X_2; Y|X_1, Z_2, U) \tag{8}$$

and the joint distributions are of the form

$$P(u)P(x_1, z_1|u)P(x_2, z_2|z_1, u)P(y|x_1, x_2) \tag{9}$$

where $P(x_1, z_1|u) = P(x_1|u)\mathbf{1}_{\{z_1=g_1(x_1)\}}$ and $P(x_2, z_2|z_1, u) = P(x_2|z_1, u)\mathbf{1}_{\{z_2=g_2(x_2)\}}$.

The following theorem describes the capacity region of a MAC with partial cribbing for two different cases of causality.

Theorem 1 (Capacity Region): The capacity regions of the MAC with strictly causal partial cribbing (Case A), and with mixed causal and strictly causal partial cribbing (Case B), as described in Definition 1, are $\mathcal{R}_A$ and $\mathcal{R}_B$, respectively.

Lemma 2: To exhaust $\mathcal{R}_A$ and $\mathcal{R}_B$, it is enough to restrict the alphabet of U as follows:

$$|\mathcal{U}| \le \min(|\mathcal{Y}| + 3, |\mathcal{X}_1||\mathcal{X}_2| + 2). \tag{10}$$

The proof of Theorem 1 and Lemma 2 is given in the next section.

III. Converse Proof

Here, we provide the converse proof of Theorem 1 for the two cases, A and B.

Converse Proof of Case A: Let us assume that we have a $(2^{nR_1}, 2^{nR_2}, n)$ code as in Definition 1, Case A. We will show the existence of a joint distribution $P(u)P(z_1|u)P(z_2|u)P(x_1|z_1, u)P(x_2|z_2, u)P(y|x_1, x_2)$ that satisfies the inequalities of (7) within some ϵ_n, where ϵ_n goes to zero as $n \to \infty$. Consider

$$\begin{aligned} n(R_1 + R_2) &= H(M_1, M_2) \\ &= H(M_1, M_2) + H(M_1, M_2|Y^n) \\ &\quad - H(M_1, M_2|Y^n) \\ &\overset{(a)}{=} I(M_1, M_2; Y^n) + n\epsilon_n \\ &\overset{(b)}{=} I(X_1^n, X_2^n; Y^n) + n\epsilon_n \\ &= \sum_{i=1}^{n} I(X_1^n, X_2^n; Y_i|Y^{i-1}) + n\epsilon_n \\ &\overset{(c)}{\le} \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i) + n\epsilon_n \end{aligned} \tag{11}$$

where (a) follows from Fano's inequality, (b) from the fact that (X_1^n, X_2^n) is a deterministic function of (M_1, M_2) and the Markov chain $Y^n - (X_1^n, X_2^n) - (M_1, M_2)$, and (c) from the Markov chain $Y_i - (X_{1,i}, X_{2,i}) - (X_1^n, X_2^n, Y^{i-1})$. Now consider

$$\begin{aligned} n(R_1 + R_2) &= H(M_1, M_2) \\ &\overset{(a)}{=} H(M_1, M_2, Z_1^n, Z_2^n) \\ &= H(Z_1^n, Z_2^n) + H(M_1, M_2|Z_1^n, Z_2^n) \\ &\overset{(b)}{=} H(Z_1^n, Z_2^n) + I(M_1, M_2; Y^n|Z_1^n, Z_2^n) + n\epsilon_n \\ &= H(Z_1^n, Z_2^n) + I(X_1^n, X_2^n; Y^n|Z_1^n, Z_2^n) + n\epsilon_n \\ &= \sum_{i=1}^{n} H(Z_{1,i}, Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}) \\ &\quad + I(X_1^n, X_2^n; Y_i|Y^{i-1}, Z_1^n, Z_2^n) + n\epsilon_n \\ &\overset{(c)}{\le} \sum_{i=1}^{n} H(Z_{1,i}, Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}) \\ &\quad + I(X_{1,i}, X_{2,i}; Y_i|Z_1^i, Z_2^i) + n\epsilon_n, \\ &\overset{(d)}{=} \sum_{i=1}^{n} H(Z_{1,i}, Z_{2,i}|U_i) \\ &\quad + I(X_{1,i}, X_{2,i}; Y_i|Z_{1,i}, Z_{2,i}, U_i) + n\epsilon_n \end{aligned} \tag{12}$$

where (a) follows from the fact that (Z_1^n, Z_2^n) are deterministic functions of (M_1, M_2), (b) from Fano's inequality, (c) follows from the Markov Chain $Y_i - (X_{1,i}, X_{2,i}) - (X_1^n, X_2^n, Y^{i-1})$, and (d) from the following definition of the random variable

$$U_i \triangleq (Z_1^{i-1}, Z_2^{i-1}). \tag{13}$$

Furthermore, consider

$$\begin{aligned} nR_1 &= H(M_1) \\ &\overset{(a)}{=} H(M_1|M_2) \\ &\overset{(b)}{=} H(M_1, Z_1^n|M_2) \\ &= H(Z_1^n|M_2) + H(M_1|Z_1^n, M_2) \\ &= H(Z_1^n|M_2) + H(M_1|M_2, Z_1^n) + H(M_1|Y^n, M_2, Z_1^n) \\ &\quad - H(M_1|Y^n, M_2, Z_1^n) \\ &= H(Z_1^n|M_2) + I(Y^n; M_1|M_2, Z_1^n) \\ &\quad + H(M_1|Y^n, M_2, Z_1^n) \\ &\overset{(c)}{\le} \sum_{i=1}^{n} H(Z_{1,i}|Z_1^{i-1}, M_2) + I(Y_i; M_1|Y^{i-1}, M_2, Z_1^n) \\ &\quad + n\epsilon_n \\ &\overset{(d)}{=} \sum_{i=1}^{n} H(Z_{1,i}|Z_1^{i-1}, Z_2^{i-1}, M_2) \\ &\quad + I(Y_i; M_1, X_{1,i}|Y^{i-1}, M_2, X_{2,i}, Z_1^n, Z_2^n) + n\epsilon_n \\ &\overset{(e)}{\le} \sum_{i=1}^{n} H(Z_{1,i}|Z_1^{i-1}, Z_2^{i-1}) + I(Y_i; X_{1,i}|X_{2,i}, Z_1^i, Z_2^i) \\ &\quad + n\epsilon_n \\ &= \sum_{i=1}^{n} H(Z_{1,i}|U_i) + I(Y_i; X_{1,i}|X_{2,i}, U_i, Z_{1,i}) + n\epsilon_n \end{aligned} \tag{14}$$

where (a) follows from the fact that the messages M_1 and M_2 are independent of each other, (b) follows from the fact that Z_1^n is a deterministic function of (M_1, M_2), (c) follows from Fano's inequality, and (d) from the fact that $X_{1,i}$ is a deterministic function of (M_1, Z_2^{i-1}) and $X_{2,i}$ is a deterministic function of (M_2, Z_1^{i-1}) which implies that Z_2^{i-1} is also a deterministic function of (M_2, Z_1^{i-1}). Step (e) follows from the Markov chain $Y_i - (X_{1,i}, X_{2,i}) - (M_1, M_2, Y^{i-1}, Z_1^n, Z_2^n)$ and from the fact that conditioning reduces entropy. Similarly to (14), we obtain

$$nR_2 \le \sum_{i=1}^{n} H(Z_{2,i}|U_i) + I(Y_i; X_{2,i}|X_{1,i}, U_i, Z_{2,i}) + n\epsilon_n. \tag{15}$$

Now let us verify that the three Markov chains

$$\begin{aligned} &Z_{1,i} - U_i - Z_{2,i} \\ &X_{1,i} - (U_i, Z_{1,i}) - X_{2,i} \\ &X_{2,i} - (U_i, Z_{2,i}) - X_{1,i} \end{aligned} \tag{16}$$

hold. The first Markov chain is due to the Markov chain $(M_1, Z_2^{i-1}) - (Z_1^{i-1}, Z_2^{i-1}) - (M_2, Z_1^{i-1})$ or, equivalently, $M_1 - (Z_1^{i-1}, Z_2^{i-1}) - M_2$ and the second Markov chain is due to the Markov chain $(M_1, Z_2^{i-1}) - (Z_1^{i}, Z_2^{i-1}) - (M_2, Z_1^{i-1})$ or, equivalently, $M_1 - (Z_1^i, Z_2^{i-1}) - M_2$. The Markov chain $M_1 - (Z_1^{i-1}, Z_2^{i-1}) - M_2$ follows from the joint distribution $P(m_1, m_2, z_1^{i-1}, z_2^{i-1}) = P(m_1)P(m_2)\prod_{j=1}^{i-1} P(z_{1,j}|z_2^{j-1}, m_1)\prod_{j=1}^{i-1} P(z_{2,j}|z_1^{j-1}, m_2)$ and the observation that $P(m_1|z_1^{i-1}, z_2^{i-1}, m_2)$ can be expressed as (17), given at the bottom of the page, and, therefore, it does not depend on m_2. Similarly, $P(m_1|z_1^i, z_2^{i-1}, m_2)$ also does not depend on m_2 and therefore $M_1 - (Z_1^i, Z_2^{i-1}) - M_2$ holds. The third Markov chain is an exchange between the indices 1 and 2, namely, $M_1, X_{1,i}, Z_{1,i}$ is exchanged with $M_2, X_{2,i}, Z_{2,i}$, respectively. Finally, let Q be a random variable independent of (X_1^n, X_2^n, Y^n), and uniformly distributed over the set $\{1, 2, 3, .., n\}$. We define the random variables $U \triangleq (Q, U_Q)$ and obtain that the region given in (7) is an outer bound to any achievable rate. ■

Once Case A has been proved, Case B follows in a straightforward manner using the following modification.

Converse Proof for Case B: We repeat the same steps as in the converse proof for Case A, except for the bound on R_2 which needs a different treatment that takes into account the causal cribbing. Consider

$$\begin{aligned} nR_2 &\overset{(c)}{\le} \sum_{i=1}^{n} H(Z_{2,i}|Z_2^{i-1}, M_1) + I(Y_i; M_2|Y^{i-1}, M_1, Z_2^n) \\ &\quad + n\epsilon_n \\ &\overset{(d)}{=} \sum_{i=1}^{n} H(Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}, M_1, Z_{1,i}) \\ &\quad + I(Y_i; M_2, X_{2,i}|Y^{i-1}, M_1, X_{1,i}, Z_2^n, Z_1^n) + n\epsilon_n \\ &\overset{(e)}{\le} \sum_{i=1}^{n} H(Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}, Z_{1,i}) \\ &\quad + I(Y_i; X_{2,i}|X_{1,i}, Z_1^i, Z_2^i) + n\epsilon_n \\ &\overset{(f)}{=} \sum_{i=1}^{n} H(Z_{2,i}|U_i, Z_{1,i}) + I(Y_i; X_{2,i}|X_{1,i}, U_i, Z_{2,i}) \\ &\quad + n\epsilon_n \end{aligned} \tag{18}$$

where step (c) follows from steps (a)–(c) in (14), replacing index 1 with index 2. Step (d) follows from the fact that $Z_{1,i}$ and $X_{1,i}$ are deterministic functions of (M_1, Z_2^{i-1}), and $X_{2,i}$ is a deterministic function of (M_2, Z_1^i). Step (e) follows from the Markov chain $Y_i - (X_{1,i}, X_{2,i}) - (M_1, M_2, Y^{i-1}, Z_1^n, Z_2^n)$ and from the fact that conditioning reduces entropy. Step (f) follows from the definition of the auxiliary random variable U_i given in (13). The rest of the inequalities are obtained as in Case A, i.e., (11), (12), and (14). In the final step, we need to show that it suffices to consider a set of distributions of the form given in (9). Hence, we need to show that the Markov chain $X_{2,i} - (U_i, Z_{1,i}, Z_{2,i}) - X_{1,i}$ holds (rather than $X_{2,i} - (U_i, Z_{2,i}) - X_{1,i}$ as in Case A). Since for Case B, the Markov chain $M_2 - (Z_1^i, Z_2^i) - M_1$ holds [the proof is similar to (17)], it follows that $X_{2,i} - (M_2, Z_1^i) - (U_i, Z_{1,i}, Z_{2,i}) - (M_1, Z_2^{i-1}) - X_{1,i}$ holds too. ■

Now we prove Lemma 2 which allows us to bound the cardinality of the auxiliary random variable U without affecting the rate regions $\mathcal{R}_A, \mathcal{R}_B$.

Proof of Lemma 2: We show the proof for Case A; the proof for Case B is similar and hence omitted. We invoke the support lemma [12, p. 310]. The external random variable U must have $|\mathcal{Y}| - 1$ elements to preserve $P(y)$ and four more to preserve the expressions $H(Z_1|U) + I(X_1; Y|X_2, Z_1, U)$, $H(Z_2|U) + I(X_2; Y|X_1, Z_2, U)$, $I(X_1, X_2; Y|U, Z_1, Z_2) + H(Z_1, Z_2|U)$, and $H(Y|X_1, X_2, U)$. Alternatively, the external random variable U must have $|\mathcal{X}_1||\mathcal{X}_2| - 1$ elements to preserve $P(x_1, x_2)$ and three more

$$\begin{aligned} P(m_1|z_1^{i-1}, z_2^{i-1}, m_2) &= \frac{P(m_1)P(m_2)\prod_{j=1}^{i-1} P(z_{1,j}|z_2^{j-1}, m_1)\prod_{j=1}^{i-1} P(z_{2,j}|z_1^{j-1}, m_2)}{\left(P(m_2)\prod_{j=1}^{i-1} P(z_{2,j}|z_1^{j-1}, m_2)\right)\sum_{m_1} P(m_1)\prod_{j=1}^{i-1} P(z_{1,j}|z_2^{j-1}, m_1)} \\ &= \frac{P(m_1)\prod_{j=1}^{i-1} P(z_{1,j}|z_2^{j-1}, m_1)}{\sum_{m_1} P(m_1)\prod_{j=1}^{i-1} P(z_{1,j}|z_2^{j-1}, m_1)} \end{aligned} \tag{17}$$

to preserve the expressions $H(Z_1|U) + I(X_1;Y|X_2,Z_1,U)$, $H(Z_2|U) + I(X_2;Y|X_1,Z_2,U)$, $I(X_1,X_2;Y|U,Z_1,Z_2) + H(Z_1,Z_2|U)$. Hence, the cardinality of U may be bounded by $\min(|\mathcal{Y}|+3, |\mathcal{X}_1||\mathcal{X}_2|+2)$. ■

IV. Achievability Proof of Theorem 1

In this section, we provide the achievability proof of Theorem 1 for the two cases: Case A and Case B. The achievability of Case A can be seen as a special case of the achievable region of MAC with generalized feedback which was derived by Willems *et al.* [1, Th. 7.1], [13] and can be also found in [14, Ch. 11]. We present here the proof of Case A since it will be the basis for the achievability proof of Case B.

Throughout the achievability proofs in the paper, we use the definition of a strong typical set. The set $T_\epsilon^{(n)}(X,Y,Z)$ of ϵ-typical n-sequences is defined by $\{(x^n,y^n,z^n) : \frac{1}{n}N(x,y,z|x^n,y^n,z^n) - p(x,y,z)| \leq \epsilon p(x,y,z) \forall (x,y,z) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\}$, where $N(x,y,z|x^n,y^n,z^n)$ is the number of appearances of (x,y,z) in the n-sequence (x^n,y^n,z^n). Additionally, we will use the following well-known lemma [12], [14]–[16].

Lemma 3 (Joint Typicality Lemma): Consider a joint distribution $P_{X,Y,Z}$ and suppose $(x^n,y^n) \in T_\epsilon^{(n)}(X,Y)$. Let $\tilde{Z}^n$ be distributed according to $\prod_{i=1}^n P_{Z|X}(\tilde{z}_i|x_i)$. Then

$$\Pr\{(x^n,y^n,\tilde{Z}^n) \in T_\epsilon^{(n)}(X,Y,Z)\} \leq 2^{-n(I(Y;Z|X)-\delta(\epsilon))} \tag{19}$$

where $\lim_{\epsilon\to 0}\delta(\epsilon) = 0$.

For the achievability proof, we use the rate-splitting coding technique in addition to the techniques used by Willems [17], i.e., block Markov coding, superposition coding, Shannon's strategies, and backward decoding. The rate-splitting technique introduces additional rate variables which are redundant and we eliminate them using the Fourier–Motzkin elimination [18].

Achievability Proof of Case A: Let us split rate R_1 into two rates R_1' and R_1'' such that $R_1 = R_1' + R_1''$ and, similarly, R_2 into R_2' and R_2'' such that $R_2 = R_2' + R_2''$. Let $m_1' \in [1,\ldots,2^{nR_1'}]$, $m_1'' \in [1,\ldots,2^{nR_1''}]$, $m_2' \in [1,\ldots,2^{nR_2'}]$, and $m_2'' \in [1,\ldots,2^{nR_2''}]$. Note that there is a one-to-one mapping between (m_1',m_1'') and m_1 and between (m_2',m_2'') and m_2.

Code Construction: Divide a block of length Bn into B blocks of length n. We use random coding to generate independently the code for each sub-block b. Construct $2^{n(R_1'+R_2')}$ codewords[1] u^n according to i.i.d. $\sim P(u)$. For every codeword u^n, construct $2^{nR_1'}$ codewords z_1^n according to i.i.d. $\sim P(z_1|u)$ and similarly $2^{nR_2'}$ codewords z_2^n according to i.i.d. $\sim P(z_2|u)$. Furthermore, generate $2^{nR_1''}$ codewords x_1^n according to i.i.d. $\sim P(x_1|z_1,u)$ and similarly $2^{nR_2''}$ codewords x_2^n according to i.i.d. $\sim P(x_2|z_2,u)$. The Markov structure of the code is

$$\begin{aligned} &x_1^n \text{ is determined by } (m_{1,b}', m_{1,b}'') \\ &\quad \text{conditioned on } (m_{1,b-1}', m_{2,b-1}') \\ &x_2^n \text{ is determined by } (m_{2,b}', m_{2,b}'') \\ &\quad \text{conditioned on } (m_{1,b-1}', m_{2,b-1}'). \end{aligned} \tag{20}$$

Encoding at the Transmitters: At block $b \in [1,\ldots,B]$, encode the message $(m_{1,b-1}', m_{2,b-1}') \in [1,..,2^{n(R_1'+R_2')}]$ using $u^n(m_{1,b-1}', m_{2,b-1}')$, encode $m_{1,b}'$ conditioned on $(m_{1,b-1}', m_{2,b-1}')$ using $z_1^n(u^n, m_{1,b}')$, and encode $m_{1,b}''$ conditioned on $(m_{1,b}', m_{1,b-1}', m_{2,b-1}')$ using $x_1^n(z_1^n, u^n, m_{1,b}'')$. Similarly, encode $m_{2,b}'$ conditioned on $(m_{1,b-1}', m_{2,b-1}')$ using $z_2^n(u^n, m_{2,b}')$, and encode $m_{2,b}''$ conditioned on $(m_{2,b}', m_{1,b-1}', m_{2,b-1}')$ using $x_2^n(z_2^n, u^n, m_{2,b}'')$. We assume that $m_{1,0}' = m_{2,0}' = 1$ and $m_{1,B}'' = m_{2,B}'' = 1$ which allow for a backward decoding scheme as explained next.

Decoding at the Transmitters: In the encoding procedure, we assumed that at block b, Transmitter 1 knows $m_{2,b-1}'$ and Transmitter 2 knows $m_{1,b-1}'$. This is possible due to cribbing as follows. At the end of block $b-1$, Transmitter 1 looks for $\hat{m}_{2,b-1}' \in \{1,2,\ldots,2^{nR_2'}\}$ such that

$$z_2^n(u^n, \hat{m}_{2,b-1}') = z_2^n \tag{21}$$

where z_2^n is the cribbed signal received in block $b-1$ and $z_2^n(u^n, \hat{m}_{2,b-1}')$ is the codeword associated with u^n and $\hat{m}_{2,b-1}'$. Note that u^n is known since it is a function of $(m_{1,b-2}', \hat{m}_{2,b-2}')$ that are known at block $b-1$. Similarly, Transmitter 2 looks for $\hat{m}_{1,b-1}' \in \{1,2,\ldots,2^{nR_1'}\}$ such that

$$z_1^n(u^n, \hat{m}_{1,b-1}') = z_1^n \tag{22}$$

where z_1^n is the cribbed signal received in block $b-1$ and $z_2^n(u^n, \hat{m}_{2,b-1}')$ is the codeword associated with u^n and $\hat{m}_{2,b-1}'$. If a pair index $(\hat{m}_{2,b-1}', \hat{m}_{1,b-1}')$ that satisfies (21) and (22) does not exist, an arbitrary message is used (and an error is declared) and if there exists more then one such pair, the smallest calligraphic index is chosen (and an error is also declared).

Decoding at the Receiver: The receiver waits until the end of the block Bn and starts decoding each message in the sub-blocks going backwards $b \in [B, B-1, B-2, \ldots, 1]$. At block b, we assume that $(m_{1,b}', m_{2,b}')$ is already known to the receiver from block $b+1$ and it needs to decode $m_{1,b-1}', m_{2,b-1}', m_{2,b}''$ and $m_{2,b}''$. The decoder uses joint typicality decoding at block b to look for $(\hat{m}_{1,b-1}', \hat{m}_{2,b-1}')$, $\hat{m}_{2,b}''$ and $\hat{m}_{2,b}''$ such that we have (23), given at the bottom of the page. If no such triplet or more than one such triplet is found, an error is declared at block b and, therefore, at the whole superblock nB (we consider $(\hat{m}_{1,b-1}', \hat{m}_{2,b-1}')$ as an index in $[1,\ldots,2^{nR_1'+nR_2'}]$).

[1] It would be more precise to denote the codewords in block b by u_b^n, $x_{1,b}^n$, $x_{2,b}^n$ rather than just u^n, x_1^n, x_2^n. Since it is clear from the context that we are dealing with codewords in block b, we have omitted the subscript b for the sake of brevity.

$$(u^n(\hat{m}_{1,b-1}', \hat{m}_{2,b-1}'), z_1^n(u^n, m_{1,b}'), z_2^n(u^n, m_{2,b}'), x_1^n(z_1^n, u^n, \hat{m}_{1,b}''), x_2^n(z_2^n, u^n, \hat{m}_{2,b}'') \in T_\epsilon^{(n)} \tag{23}$$

Error Analysis: The following lemma will enable us to bound the probability of error of the superblock nB by bounding the probability of error of each block.

Lemma 4: Let $\{A_j\}_{j=1}^{J}$ be a set of events and let A_j^c denotes the complement of the event A_j. Then

$$P(\bigcup_{j=1}^{J} A_j) \leq \sum_{j=1}^{n} P(A_j | \bigcap_{i=1}^{j-1} A_i^c) = \sum_{j=1}^{n} P(A_j | A_1^c, A_2^c, \ldots, A_{j-1}^c). \quad (24)$$

Proof: For simplicity, let us assume that $J = 3$. In a straightforward manner, the proof extends to any number of sets J. For any three sets of events A_1, A_2, A_3, we have

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P\left(A_1 \cup (A_2 \cap A_1^c) \cup (A_3 \cap A_1^c \cap A_2^c)\right) \\ &= P(A_1) + P(A_2 \cap A_1^c) + P(A_3 \cap A_1^c \cap A_2^c) \\ &\leq P(A_1) + \frac{P(A_2 \cap A_1^c)}{P(A_1^c)} + \frac{P(A_3 \cap A_1^c \cap A_2^c)}{P(A_1^c \cap A_2^c)} \\ &= P(A_1) + P(A_2|A_1^c) + P(A_3|A_1^c \cap A_2^c) \\ &= P(A_1) + P(A_2|A_1^c) + P(A_3|A_1^c, A_2^c). \end{aligned} \quad (25)$$

■

Using Lemma 4, we bound the probability of error in the super block Bn by the sum of the probability of having an error in each block b given that in previous blocks, $(b+1, \ldots, B)$, the messages were decoded correctly.

First let us bound the probability that for some b, Transmitter 1 decodes the message $m'_{2,b}$ incorrectly or Transmitter 2 decodes the message $m'_{1,b}$ incorrectly at the end of block b. Using Lemma 4, it suffices to show that the probability of error decoding in each block b goes to zero, assuming that all previous messages in block $(1, 2, \ldots, b-1)$ were decoded correctly.

Let $E_{1,b}$ be the event that Transmitter 1 has an error in decoding $m'_{2,b}$ and let $E_{2,b}$ be the event that Transmitter 2 has an error in decoding $m'_{1,b}$. The term $P(E_{1,b} \cup E_{2,b} | E_{1,b-1}^c, E_{2,b-1}^c)$ is the probability that either Transmitters 1 or 2 incorrectly decoded $m'_{2,b}$ and $m'_{1,b}$, respectively, given that $m'_{1,b-1}$ and $m'_{2,b-1}$ were decoded correctly. Without loss of generality, let us assume that $m'_{1,b} = m'_{2,b} = 1$. An error occurs if and only if there is another message $m'_{1,b} > 1$ that maps to the same codeword as $z_1^n(1, u^n)$ or there is another message $m'_{2,b} > 1$ that maps to the same codeword as $z_2^n(1, u^n)$. The probability that $z_1^n(i, u^n) = z_1^n(1, u^n)$ where $i > 1$ and where $(z_1^n(1, u^n), u^n) \in T_\epsilon^{(n)}(Z_1, U)$ and $z_1^n(i, u^n)$ was generated according to $p(z_1|u)$ is bounded by $2^{-n(H(Z_1|U) - \delta(\epsilon))}$, where $\delta(\epsilon)$ goes to zero as ϵ goes to zero. Hence

$$\begin{aligned} P(E_{1,b} \cup E_{2,b} | E_{1,b-1}^c, E_{2,b-1}^c) &\overset{(a)}{\leq} P(E_{1,b} | E_{1,b-1}^c, E_{2,b-1}^c) \\ &\quad + P(E_{2,b} | E_{1,b-1}^c, E_{2,b-1}^c) \\ &\leq \sum_{i=2}^{2^{nR'_1}} 2^{-n(H(Z_1|U) - \delta(\epsilon))} \\ &\quad + \sum_{i=2}^{2^{nR'_2}} 2^{-n(H(Z_2|U) - \delta(\epsilon))} \\ &\leq 2^{n(R'_1 - H(Z_1|U) + \delta(\epsilon))} \\ &\quad + 2^{n(R'_2 - H(Z_2|U) + \delta(\epsilon))} \end{aligned} \quad (26)$$

where inequality (a) follows from the union bound. Now we bound the probability that the receiver decodes the messages $(m'_{1,b-1}, m'_{2,b-1})$ or $m''_{1,b}$ or $m''_{2,b}$ incorrectly at block b, given that at block $b+1$ the messages $(m'_{1,b}, m'_{2,b})$ were decoded correctly and given that Transmitters 1 and 2 encode the right messages $(m'_{1,b-1}, m'_{2,b-1})$ in block b. Without loss of generality, assume $(m'_{1,b-1}, m'_{2,b-1}) = 1$ (for simplicity, we index both messages by one index), $m''_{1,b} = 1$ and $m''_{2,b} = 1$. Let us define the event (27), given at the bottom of the page. An error occurs if either the correct codewords are not jointly typical with the received sequences, i.e., $E_{1,1,1,b}^c$, or if there exists a different tuple $(i, j, k) \neq (1, 1, 1)$ such that $E_{i,j,k,b}$ occurs. Let $P_{e,b}^{(n)}$ be the error decoding at block b given that in blocks $(b+1, \ldots, B)$, there was no error decoding. From the union bound, we obtain that

$$\begin{aligned} P_{e,b}^{(n)} \leq \Pr(E_{1,1,1,b}^c) &+ \sum_{i=1,j=1,k>1} \Pr(E_{i,j,k,b}) \\ &+ \sum_{i=1,j>1,k=1} \Pr(E_{i,j,k,b}) + \sum_{i=1,j>1,k>1} \Pr(E_{i,j,k,b}) \\ &+ \sum_{i>1,j\geq 1,k\geq 1} \Pr(E_{i,j,k,b}). \end{aligned} \quad (28)$$

Now let us show that each term in (28) goes to zero as the block-length of the code n goes to infinity.

1) Upper-bounding $\Pr(E_{1,1,1}^c)$: Since we assume that the Transmitters 1 and 2 encode the right $(m'_{1,b-1}, m'_{2,b-1})$ and that the receiver decoded the right $(m'_{1,b}, m'_{2,b})$ in block $b+1$, by the law of large numbers (LLN) $\Pr(E_{1,1,1,b}^c) \to 0$.
2) Upper-bounding $\sum_{i=1,j=1,k>1} \Pr(E_{i,j,k})$: The probability that Y^n, which is generated according to $P(y|x_1, x_2) = P(y|x_1, x_2, u, z_1, z_2)$, is jointly typical with x_2^n, which was generated according to $P(x_2|z_2, u) = P(x_2|u, z_1, z_2, x_1)$, where

$$E_{i,j,k,b} \triangleq \left\{ \left(u^n(i), z_1^n(u^n, m'_{1,b}), z_2^n(u^n, m'_{2,b}), x_1^n(u^n, z_1^n, j), x_2^n(u^n, z_2^n, k), y^n\right) \in T_\epsilon^{(n)}(U, Z_1, Z_2, X_1, X_2, Y) \right\} \quad (27)$$

$(x_1^n, z_1^n, z_2^n, u^n) \in T_\epsilon^{(n)}(X_1, Z_1, Z_2, U)$ is bounded by (Lemma 3)

$$\Pr\{(x_1^n, z_1^n, X_2^n, z_2^n, u^n, Y^n) \in T_\epsilon^{(n)} | (x_1^n, z_1^n, z_2^n, u^n) \in T_\epsilon^{(n)}\} \leq 2^{-n(I(X_2;Y|X_1,Z_2,U)-\delta(\epsilon))}. \quad (29)$$

Hence, we obtain

$$\sum_{i=1,j=1,k>1} \Pr(E_{i,j,k,b}) \leq 2^{nR_2''} 2^{-n(I(X_2;Y|X_1,Z_1,Z_2,U)-\delta(\epsilon))}. \quad (30)$$

3) Upper-bounding $\sum_{i=1,j>1,k=1} \Pr(E_{i,j,k,b})$: Similarly to (30), we obtain

$$\sum_{i=1,j>1,k=1} \Pr(E_{i,j,k,b}) \leq 2^{nR_1''} 2^{-n(I(X_1;Y|X_2,Z_1,Z_2,U)-\delta(\epsilon))}. \quad (31)$$

4) Upper-bounding $\sum_{i=1,j>1,k>1} \Pr(E_{i,j,k,b})$ by

$$\sum_{i=1,j>1,k>1} \Pr(E_{i,j,k,b}) \leq 2^{n(R_2''+R_1'')} 2^{-n(I(X_2,X_1;Y|Z_1,Z_2,U)-\delta(\epsilon))}. \quad (32)$$

5) Upper-bounding $\sum_{i>1,j\geq1,k\geq1} \Pr(E_{i,j,k,b})$ by

$$\begin{aligned}\sum_{i>1,j\geq1,k\geq1} \Pr(E_{i,j,k,b}) &\leq 2^{n(R_1''+R_1'+R_2)} 2^{-n(I(X_2,X_1,U,Z_1,Z_2;Y)-\delta(\epsilon))} \\ &= 2^{n(R_1+R_2-I(X_2,X_1;Y)-\delta(\epsilon))}. \end{aligned} \quad (33)$$

To summarize, we obtain that if $R_1' = R_1 - R_1''$, R_1'', $R_2' = R_2 - R_2''$, R_2'' and R_2 satisfy

$$\begin{aligned} R_1 - R_1'' &\leq H(Z_1|U) \\ R_2 - R_2'' &\leq H(Z_2|U) \\ R_1'' &\leq I(X_1;Y|X_2,Z_1,U) \\ R_2'' &\leq I(X_2;Y|X_1,Z_2,U) \\ R_1'' + R_2'' &\leq I(X_1,X_2;Y|Z_1,Z_2,U) \\ R_1 + R_2 &\leq I(X_2,X_1;Y) \end{aligned} \quad (34)$$

then there exists a sequence of codes with a probability of error that goes to zero as the block length goes to infinity. Using the Fourier–Motzkin elimination [18] first for R_1'', we obtain

$$\begin{aligned} R_1 - H(Z_1|U) &\leq I(X_1;Y|X_2,Z_1,U) \\ R_2 - R_2'' &\leq H(Z_2|U) \\ R_2'' &\leq I(X_2;Y|X_1,Z_2,U) \\ R_1 - H(Z_1|U) + R_2'' &\leq I(X_1,X_2;Y|Z_1,Z_2,U) \\ R_1 + R_2 &\leq I(X_2,X_1;Y) \end{aligned} \quad (35)$$

and applying Fourier–Motzkin elimination also for R_2'', we obtain

$$\begin{aligned} R_1 - H(Z_1|U) &\leq I(X_1;Y|X_2,Z_1,U) \\ R_2 - H(Z_2|U) &\leq I(X_2;Y|X_1,Z_2,U) \\ R_1 - H(Z_1|U) + R_2 - H(Z_2|U) &\leq I(X_1,X_2;Y|Z_1,Z_2,U) \\ R_1 + R_2 &\leq I(X_2,X_1;Y) \end{aligned} \quad (36)$$

which is equivalent to the region of Case A in (7). ■

Achievability for Case B: The achievability of Case B is very similar to Case A, only that X_2^n and Z_2^n in each sub-block b need to be generated according to *codetrees* rather than codewords where the branches of the codetree are controlled by $Z_{1,i}$. The codetrees (also called strategies by Shannon in [19]) in Case B give to Transmitter 2 the flexibility to generate codewords that depend causally on the cribbed signal, thereby having a random code generated by the conditional distribution $P(x_2, z_2|z_1, u)$ rather than $P(x_2, z_2|u)$ as in Case A. Here, we explain the parts of the proof that differ from Case A.

Code Design: The division into sub-blocks in Case B is the same as in Case A and the code design of Transmitter 1 in Case B is identical to the one in Case A. The difference is only in the code design of Transmitter 2.

In Case A in sub-block b for a fixed $u^n(m_{1,b-1}', m_{2,b-1}')$, we generated a codeword $z_2^n(u^n, m_{2,b}')$ according to $P(z_2^n|u^n) = \prod_{i=1}^n P(z_{2,i}|u_i)$ for each $m_{2,b}'$. Then, for a fixed $u^n(m_{1,b-1}', m_{2,b-1}')$ and fixed $z_2^n(u^n, m_{2,b}')$, we generated codewords $x_2^n(z_2^n, u^n, m_{2,b}'')$ according to $P(x_2^n|u^n, z_2^n) = \prod_{i=1}^n P(x_{2,i}|u_i, z_{2,i})$ for each $m_{2,b}''$. The codeword is illustrated in the left part of Fig. 2.

However, for Case B, for any $u^n(m_{1,b-1}', m_{2,b-1}')$, we generate codetrees rather then codewords that are associated with $m_{2,b}'$ and $m_{2,b}''$. The branch at time i in the codetree is chosen according to z_1^i. The codetree that is associated with message $m_{2,b}'$, where $u^n(m_{1,b-1}', m_{2,b-1}')$ is fixed, is generated according to the probability $P(z_{2,i}|u_i, z_{1,i})$. The codetree that is associated with message $m_{2,b}''$, where $u^n(m_{1,b-1}', m_{2,b-1}')$ and $m_{2,b}'$ are fixed, is generated according to $P(x_{2,i}|u_i, z_{1,i}, z_{2,i})$. An illustration of the codetree and its comparison with the codeword is given on the right side of Fig. 2.

Encoding at Transmitter 1: Identical to Case A.

Encoding at Transmitter 2: Given a message $m_{2,b}'$ and $u^n(m_{1,b-1}', m_{2,b-1}')$, z_2^n is the sequence that results from the path z_1^n in the codetree that is associated with $m_{2,b}'$. Furthermore, given $m_{2,b}''$, $z_2^n(m_{2,b}', z_1^n)$ and $u^n(m_{1,b-1}', m_{2,b-1}')$ the output of the second transmitter x_2^n is the sequence due to the path determined by z_1^n in the codetree that is associated with $m_{2,b}''$

Decoding at the Transmitters (At the End of Block $b-1$): The decoding scheme of Transmitter 2 is identical to the decoding scheme in Case A, i.e., it looks for $\hat{m}_{1,b-1}' \in \{1, 2, \ldots, 2^{nR_2'}\}$ such that

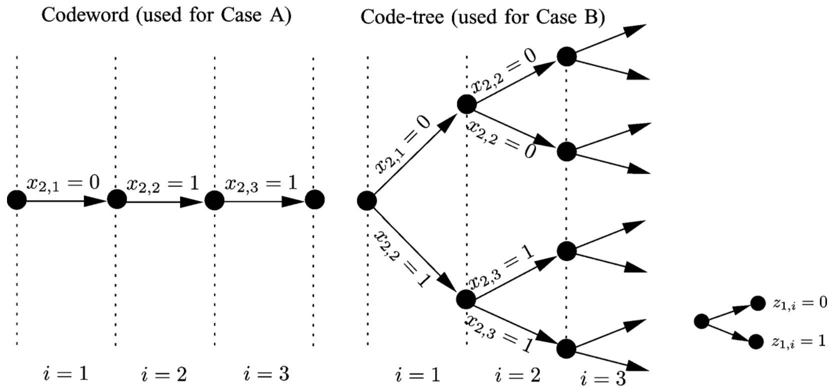
$$z_1^n(u^n, \hat{m}_{1,b-1}') = z_1^n \quad (37)$$

where z_1^n is the cribbed signal received in block $b-1$ and $z_1^n(u^n, \hat{m}_{1,b-1}')$ is the codeword associated with u^n and $\hat{m}_{1,b-1}'$. Note that u^n is known, since it is a function of $(m_{1,b-2}', \hat{m}_{2,b-2}')$ that are known at block $b-1$.

The decoding scheme of Transmitter 1 is slightly different from the scheme in Case A. Transmitter 1 looks for $\hat{m}_{2,b-1}' \in \{1, 2, \ldots, 2^{nR_2'}\}$ such that

$$z_2^n(u^n, \hat{m}_{2,b-1}', z_1^n) = z_2^n \quad (38)$$

where z_2^n is the cribbed signal received in block $b-1$, z_1^n is the output of Transmitter 1 in block $b-1$ and is obviously known



(a) (b)

Fig. 2. Illustration of two coding schemes where (a) a message is mapped to codewords (like in Case A) and (b) with a codetree (like in Case B), where $\mathcal{X}_2$ and $\mathcal{Z}_2$ are binary. The tree branches at every time instant $i = 1, \ldots, n$ within the block b and there exists a codetree for each block b, each cloud center u^n, and each message that is encoded.

to Transmitter 1, and $z_2^n(u^n, \hat{m}'_{2,b-1}, z_1^n)$ is the codeword associated with u^n, $\hat{m}'_{2,b-1}$ and the cribbed signal z_1^n.

Decoding at the Receiver: As in Case A, we use backward decoding and joint typicality decoding. At block b, we assume that $(m'_{1,b}, m'_{2,b})$ is already known to the receiver from block $b+1$ and it needs to decode $m'_{1,b-1}$, $m'_{2,b-1}$, $m''_{2,b}$ and $m''_{2,b}$. At block b, the decoder looks for $(\hat{m}'_{1,b-1}, \hat{m}'_{2,b-1})$, $\hat{m}''_{2,b}$ and $\hat{m}''_{2,b}$ for which we have (39), given at the bottom of the page.

Error analysis: We denote $E_{1,b}$, $E_{2,b}$, and $E_{i,j,k,b}$ as in Case A. The only difference in the analysis is in $P(E_{1,b}|E^c_{1,b-1}, E^c_{2,b-1})$, which is the probability that Transmitter 1 incorrectly decoded $m'_{2,b}$ given that $m'_{1,b-1}$ and $m'_{2,b-1}$ were decoded correctly. Without loss of generality, let us assume that $m'_{2,b} = 1$. An error occurs if and only if there is another message $m'_{2,b} > 1$ that maps to the same codeword as $z_2^n(1, u^n, z_1^n)$. The probability that $z_2^n(i, u^n, z_1^n) = z_2^n(1, u^n, z_1^n)$ where $i > 1$ and where $z_2^n(1, u^n, z_1^n) \in T_\epsilon^{(n)}(Z_1, Z_2, U)$ and $z_2^n(i, u^n, z_1^n)$ was generated according to $p(z_2|u, z_1)$ is bounded by $2^{-n(H(Z_2|U,Z_1)-\delta(\epsilon))}$, where $\delta(\epsilon)$ goes to zero as ϵ goes to zero. Hence

$$P(E_{1,b}|E^c_{1,b-1}, E^c_{2,b-1}) \leq \sum_{i=2}^{2^{nR'_2}} 2^{-n(H(Z_2|Z_1,U)-\delta(\epsilon))} \leq 2^{n(R'_2-H(Z_2|Z_1,U)+\delta(\epsilon))}. \quad (40)$$

Invoking the error analysis of the other events, we obtain a sequence of equations as in (34), except that the second inequality is replaced by

$$R_2 - R''_2 \leq H(Z_2|U, Z_1) \quad (41)$$

and after Fourier–Motzkin elimination, we obtain the region $\mathcal{R}_B$. ■

V. Common Message

Let us now consider the case where a common message, $m_0 \in \{1, 2, \ldots, 2^{nR_0}\}$, is known to Encoders 1 and 2 and needs to be transmitted to the decoder in addition to the private messages m_1, m_2. Hence, Encoder 1 is given by the function

$$\text{Case A, B,}\ f_{1,i}\colon \{1,\ldots,2^{nR_0}\}\times\{1,\ldots,2^{nR_1}\}\times\mathcal{Z}_2^{i-1}\mapsto X_{1,i} \quad (42)$$

and Encoder 2 is given by the functions

$$\begin{aligned}&\text{Case A,}\ f_{2,i}\colon \{1,\ldots,2^{nR_0}\}\times\{1,\ldots,2^{nR_2}\}\times\mathcal{Z}_1^{i-1}\mapsto X_{1,i}\\&\text{Case B,}\ f_{2,i}\colon \{1,\ldots,2^{nR_0}\}\times\{1,\ldots,2^{nR_2}\}\times\mathcal{Z}_1^{i}\mapsto X_{1,i}.\end{aligned} \quad (43)$$

Remarkably, no additional auxiliary random variable is needed to characterize the capacity region, since the partial cribbing is used for generating a common message. Let the rate regions $\mathcal{R}^0_A$ and $\mathcal{R}^0_B$ be defined exactly as $\mathcal{R}_A$ and $\mathcal{R}_B$ except that the last inequality in (7), i.e., $R_1 + R_2 \leq I(X_1, X_2; Y)$, is replaced by

$$R_0 + R_1 + R_2 \leq I(X_1, X_2; Y). \quad (44)$$

Theorem 5 (Capacity Region in the Case of a Common Message): The capacity regions of the MAC with strictly causal (Case A), and mixed causal and strictly causal (Case B) partial cribbing with a common message are $\mathcal{R}^0_A$ and $\mathcal{R}^0_B$, respectively.

Note that if there is no cribbing, i.e., Z_1 and Z_2 are constant, we obtain the capacity region of the MAC with a common message as derived by Slepian and Wolf [20]. We sketch here only the differences between the proof of Theorems 1 and 5.

Proof of Theorem 5: Converse: Similar to the sequence of inequalities in (11), we have

$$n(R_0 + R_1 + R_2) = H(M_0, M_1, M_2) \leq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i) + n\epsilon_n. \quad (45)$$

$$(u^n(\hat{m}'_{1,b-1}, \hat{m}'_{2,b-1}), z_1^n(u^n, m'_{1,b}), z_2^n(u^n, m'_{2,b}, z_1^n), x_1^n(z_1^n, u^n, \hat{m}''_{1,b}), x_2^n(z_1^n, z_2^n, u^n, \hat{m}''_{2,b}) \in T_\epsilon^{(n)} \quad (39)$$

Adding conditioning on M_0 in the sequence of inequalities (12), we obtain

$$\begin{aligned} n(R_1+R_2) &= H(M_1,M_2) \\ &= H(M_1,M_2|M_0) \\ &\le \sum_{i=1}^{n} H(Z_{1,i},Z_{2,i}|Z_1^{i-1},Z_2^{i-1},M_0) \\ &\quad + I(X_{1,i},X_{2,n};Y_i|Z_1^i,Z_2^i,M_0)+n\epsilon_n, \\ &= \sum_{i=1}^{n} H(Z_{1,i},Z_{2,i}|U_i) \\ &\quad + I(X_{1,i},X_{2,n};Y_i|Z_{1,i},Z_{2,i},U_i)+n\epsilon_n, \end{aligned} \tag{46}$$

where the last step is due to the new definition of U_i as

$$U_i \triangleq (M_0, Z_1^{i-1}, Z_2^{i-1}). \tag{47}$$

Similarly, adding conditioning on M_0 in the sequence of inequalities (14), we obtain

$$\begin{aligned} nR_1 &= H(M_1) \\ &= H(M_1|M_2,M_0) \\ &\le \sum_{i=1}^{n} H(Z_{1,i}|Z_1^{i-1},Z_2^{i-1},M_0) \\ &\quad + I(Y_i;X_{1,i}|X_{2,i},Z_1^i,Z_2^i,M_0)+n\epsilon_n \\ &= \sum_{i=1}^{n} H(Z_{1,i}|U_i)+I(Y_i;X_{1,i}|X_{2,i},U_i,Z_{1,i}) \\ &\quad + n\epsilon_n. \end{aligned} \tag{48}$$

In a similar way, we obtain the inequality for R_2 as in (15).

Achievability: The achievability proof is similar to that in Theorem 1, except that we generate $2^{n(R_1'+R_2'+R_0)}$ codewords u^n according to i.i.d. $\sim P(u)$, rather than $2^{n(R_1'+R_2')}$, and wherever we have in the achievability proof of Theorem 1 $(m'_{1,b-1}, m'_{2,b-1})$, we should now have $(m_{0,b}, m'_{1,b-1}, m'_{2,b-1})$. Hence, we obtain the same sequence of inequalities as in (34), except that the last inequality (which corresponds to an error in all messages) is

$$R_0+R_1+R_2 \le I(X_2,X_1;Y). \tag{49}$$

■

VI. Special Case of Partial Cribbing: Semideterministic Relay Channel

As a special case of the partial cribbing encoders, let us consider the case where Encoder 2 has no message to send, i.e., $R_2 = 0$, and only Encoder 2 cribs from Encoder 1, i.e., Z_2 is a constant. We show here that, indeed, the region obtained via partial cribbing when $R_2 = 0$ and the region obtained via semideterministic relay channel coincide.

Case A, Semideterministic Relay With a Delay: This case becomes a special case of the semideterministic relay channel, which was introduced and solved by El-Gamal and Aref [5], where Encoder 2 plays the role of a relay. In this case, the region $\mathcal{R}_A$ becomes

$$\mathcal{R}_A = \left\{ \begin{array}{l} R_1 \le H(Z_1|U)+I(X_1;Y|X_2,Z_1,U), \\ R_1 \le I(X_1,X_2;Y|U,Z_1)+H(Z_1|U), \\ R_1 \le I(X_1,X_2;Y), \text{ for some} \\ P(u)P(z_1|u)P(x_1|z_1,u)P(x_2|u)P(y|x_1,x_2). \end{array} \right\}. \tag{50}$$

Clearly, $H(Z_1|U) + I(X_1;Y|X_2,Z_1,U) \le I(X_1,X_2;Y|U,Z_1)+H(Z_1|U)$. Hence, the region we obtained is $R_1 \le \min(H(Z_1|U)+I(X_1;Y|X_2,Z_1,U), I(X_1,X_2;Y))$ for some $P(u)P(z_1|u)P(x_1|z_1,u)P(x_2|u)$. Now consider

$$\begin{aligned} &H(Z_1|U)+I(X_1;Y|X_2,Z_1,U) \\ &\overset{(a)}{=} H(Z_1|U,X_2)+I(X_1;Y|X_2,Z_1,U) \\ &\overset{(b)}{\le} H(Z_1|X_2)+I(X_1;Y|X_2,Z_1) \end{aligned} \tag{51}$$

where step (a) follows from the Markov chain $X_2 - U - Z_1$ and step (b) from the fact that conditioning reduces entropy and from the Markov chain $Y - (X_1,Z_1,X_2) - U$. By choosing $U = X_2$, we obtain the upper bound of (51) and the expression $I(X_1,X_2;Y)$ does not decrease. Hence, the capacity region is

$$R_1 \le \min(H(Z_1|X_2)+I(X_1;Y|X_2,Z_1), I(X_1,X_2;Y)) \tag{52}$$

for some $P(x_1,x_2)$. Equation (52) coincides with the result in [5].

Case B, Semideterministic Relay Without Delay: In this case, $\mathcal{R}_B$ becomes the set of rates R_1 that satisfy

$$R_1 \le \min(H(Z_1|U)+I(X_1;Y|X_2,Z_1,U), I(X_1,X_2;Y)) \tag{53}$$

for some $P(x_1,z_1,u)P(x_2|u,z_1)$. The case of relays without delay was investigated by El-Gamal *et al.* in [21], where it was shown that the capacity region for the semideterministic relay without delay, which is denoted by $C_{0,\text{semi-det}}$, is

$$C_{0,\text{semi-det}} = \max_{P(u,x_1),x_2=f(u,z_1)} \min(I(X_1;Y,Z_1|U), I(U,X_1;Y)). \tag{54}$$

Furthermore, it was shown by Willems and van der Meulen [22] that the result can be simply obtained using the regular relay with delay and Shannon strategies.

At first glance, the expression in (53) seems to be different from the expression in (54), but with some simple manipulations one can show that the expressions are equivalent. In particular, the first term in (54) can be written as

$$\begin{aligned} I(X_1;Y,Z_1|U) &= I(X_1;Z_1|U)+I(X_1;Y|U,Z_1) \\ &\overset{(a)}{=} H(Z_1|U)+I(X_1;Y|U,Z_1) \\ &\overset{(b)}{=} H(Z_1|U)+I(X_1;Y|U,Z_1,X_2) \end{aligned} \tag{55}$$

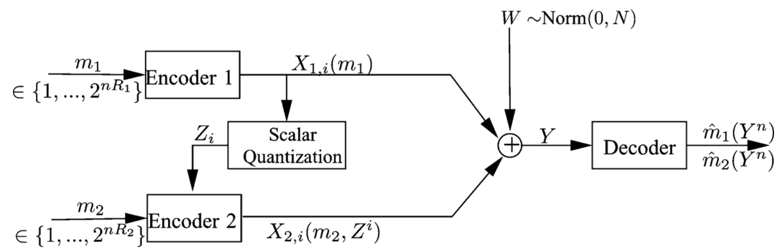


Fig. 3. Gaussian MAC with quantized cribbing. The cribbing that Encoder 2 observes is the quantized signal from Encoder 1. There exist power constraints $\sum_{i=1}^{n} E[X_{1,i}^2] \leq P_1$ and $\sum_{i=1}^{n} E[X_{1,i}^2] \leq P_2$.

where step (a) follows from the fact that Z_1 is a function of X_1 and step (b) follows from the fact that X_2 is a function of (U, Z_1). The second term in (54) may be written as

$$\begin{aligned} I(U, X_1; Y) &\overset{(a)}{=} I(U, X_1, X_2; Y) \\ &\overset{(b)}{=} I(X_1, X_2; Y) \end{aligned} \tag{56}$$

where step (a) follows from the fact that X_2 is a function of (U, X_1) and step (b) follows from the Markov chain $Y - (X_1, X_2) - U$. Now, to conclude that (53) and (54) are equivalent, we need to show that it suffices to consider only distributions where X_2 is a function of (U, Z_1) in (53). It follows from [23, Lemma 1] that there exists a random variable W independent of (U, Z_1) that satisfies $W - (X_2, U, Z_1) - (Y, X_1)$ such that X_2 is a deterministic function of (U, Z_1, W). Therefore

$$\begin{aligned} &H(Z_1|U) + I(X_1; Y|X_2, Z_1, U) \\ &\quad = H(Z_1|U, W) + I(X_1; Y|X_2, Z_1, U, W) \\ &\quad = H(Z_1|\tilde{U}) + I(X_1; Y|X_2, Z_1, \tilde{U}) \end{aligned} \tag{57}$$

where $\tilde{U} = (U, W)$. Hence, it suffices to consider X_2 that is a function of $(\tilde{U}, Z_1)$ and it emerges that (53) is equivalent to (54).

VII. Gaussian MAC With Quantized Cribbing

In this section, we consider the additive Gaussian noise MAC, i.e., $Y = X_1 + X_2 + W$, where W is a memoryless Gaussian noise with variance N, i.e., $W \sim \text{Norm}(0, N)$. We assume power constraints P_1 and P_2 on the inputs from Encoders 1 and 2, respectively. If the encoders do not cooperate, then the capacity is given by

$$\begin{aligned} R_1 &\leq \frac{1}{2} \log\left(1 + \frac{P_1}{N}\right) \\ R_2 &\leq \frac{1}{2} \log\left(1 + \frac{P_2}{N}\right) \\ R_1 + R_2 &\leq \frac{1}{2} \log\left(1 + \frac{P_1 + P_2}{N}\right). \end{aligned} \tag{58}$$

If there is perfect cribbing from Encoders 1 to 2, either with or without delay, the capacity is the same as if Encoder 2 knows the message of Encoder 1, since Encoder 1 can send the message in a single epoch. Hence, the capacity is the union over $0 \leq \rho \leq 1$ of the regions

$$\begin{aligned} R_2 &\leq \frac{1}{2} \log\left(1 + \frac{P_2}{N}(1 - \rho^2)\right) \\ R_1 + R_2 &\leq \frac{1}{2} \log\left(1 + \frac{P_1 + 2\rho\sqrt{P_1 P_2} + P_2}{N}\right). \end{aligned} \tag{59}$$

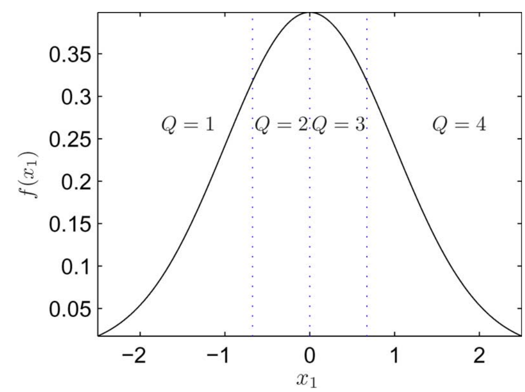


Fig. 4. Two-bit quantizer's boundaries are designed such that if the input signal has a normal distribution with variance $P_1 = 1$, then the output values from the quantizer have equal probability. The input to the 2-bit quantizer is X_1 and the output is $Q \in \{1, 2, 3, 4\}$.

Note that for a Gaussian additive MAC, the gap between the sum-rate of perfect cribbing and no cribbing is at most 1 bit, which might be larger if there are multiple frequencies (like OFDM) or multiple antenna.

Now, let us consider the case where Encoder 2 observes a quantized version of the signal from Encoder 1 without delay. The setting is depicted in Fig. 3. We assume that the quantizer is a scalar quantizer designed such that under a Gaussain input with variance $P_1 = 1$ the discrete values have the same probability (see Fig. 4 for an example of 2-bit quantizer).

Next, we consider a specific choice of distribution for the achievable scheme for the Gaussian MAC with a quantizer cribbing without delay, where the power constraints are $P_1 = P_2 = 1$ and the noise variance is $N = \frac{1}{2}$. We evaluate the region $\mathcal{R}_B$ given by (7) and (9) for the case where U_1, U_2, Z_2 are constants, $X_1 \sim N(0, 1)$, and Z_1 is a quantized version of X_1 such that each value has equal probability. The input distribution is $P_{X_2|Z_1}(x_2|z_1) = \rho P_V(x_2) + (1 - \rho) P_{X_1|Z_1}(x_2|z_1)$, where $V \sim N(0, 1)$ and is independent of X_1 and Z. Note that under these assumptions $X_2 \sim N(0, 1)$ and, therefore, satisfies the power constraint. Fig. 5 depicts a simple achievable scheme for different quantizers. The blue line in Fig. 5 represents the capacity region where there is no cribbing, evaluated according to (58). The red line represents the capacity region where there is perfect cribbing, evaluated according to (59). The lines in between are achievable regions according to the simple scheme we have described above. One can see that the main gain is already due to a 1-bit quantizer and that the difference between the achievable scheme with a 4-bit quantizer and the capacity region where there is perfect cribbing appears negligible.

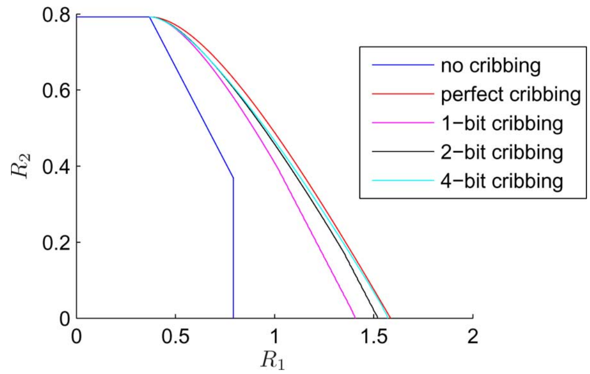


Fig. 5. Achievable regions of Gaussian MAC with a quantizer cribbing.

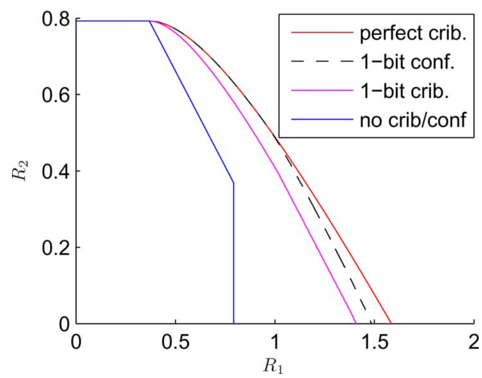


Fig. 6. Achievable regions of Gaussian MAC with quantizer cribbing versus the capacity region of Gaussian MAC with conferencing encoders.

It is also interesting to compare the performance of 1-bit cribbing with the performance of 1-bit conferencing. MAC with 1-bit conferencing is a setting where Encoder 1 may transmit through a noiseless link at a rate of 1-bit to Encoder 2. This setting was introduced and solved by Willems [24] for the general case using auxiliary random variables. Willems showed that it is optimal to generate a common message through the conferencing link and then use a Sleipan–Wolf coding for MAC with common message [20]. Recently Bross *et al.* [25], [26] showed that for the Gaussian MAC its enough to consider only Gaussian joint distribution and derived an analytical expression for the capacity [25, Th. 1].

Fig. 6 depicts the capacity region of the additive MAC with 1-bit conferencing from Encoders 1 to 2 versus the inner bound of the additive Gaussian MAC with 1-bit quantizer cribbing from Encoders 1 to 2. As expected, 1-bit cribbing performs worse than 1-bit conferencing (this can be shown operationally). However, as can be seen from the figure, the difference is small compared to the difference with non cooperative encoders, namely, when there is no conferencing and no cribbing.

VIII. Controlled Cribbing

Here, we consider the case where the cribbing is controlled by an action which depends on previously cribbed signals. In this study, only Encoder 2 cribs causally or strictly causally. More precisely, there is a controller that is an entity which takes action $a_{1,i}$ at time i, which controls the cribbed signal from Encoders 1 to 2. The cribbed signal is given by $z_{1,i} = g_1(a_{1,i}, x_{1,i})$, as shown in Fig. 7. The function $g_1 : \mathcal{A}\times\mathcal{X}_1 \mapsto \mathcal{Z}_1$ is a fixed function given as a part of problem description. The action at time i depends on past cribbed observations, i.e., $a_{1,i}(z_1^{i-1})$, and is a part of the coding scheme design. The action is a limited resource, namely, there is a restriction that $\frac{1}{n}\sum_{i=1}^n E[\Lambda(A_{1,i})] \le \Gamma$, where $\Lambda(a_1)$ is a cost of taking action a_1, and Γ is the action cost constraint.

Let us now formally define a controlled code.

Definition 2: A $(2^{nR_1}, 2^{nR_2}, n)$ *code* with controlled partial cribbing, as shown in Fig. 7, consists at time i of an encoding function at Encoder 1

$$\text{Cases A, B,}\quad f_{1,i} : \{1,\ldots,2^{nR_1}\} \mapsto \mathcal{X}_{1,i} \tag{60}$$

and an encoding function at Encoder 2 that changes according to the following cases:

$$\begin{aligned}&\text{Case A}\quad f_{2,i} : \{1,\ldots,2^{nR_2}\}\times\mathcal{Z}_1^{i-1} \mapsto \mathcal{X}_{1,i}\\ &\text{Case B}\quad f_{2,i} : \{1,\ldots,2^{nR_2}\}\times\mathcal{Z}_1^{i} \mapsto \mathcal{X}_{1,i}\end{aligned} \tag{61}$$

and a controller that takes actions according to the function

$$f_{a,i} : \mathcal{Z}_1^{i-1} \mapsto \mathcal{A}_{1,i} \tag{62}$$

and a decoding function

$$h : \mathcal{Y}^n \mapsto \{1,\ldots,2^{nR_1}\}\times\{1,\ldots,2^{nR_2}\}. \tag{63}$$

The code needs to satisfy the constraint $\frac{1}{n}\sum_{i=1}^n E[\Lambda_1(A_{1,i})] \le \Gamma_1$. The probability of error, achievable rate pairs, and capacity regions are defined in the usual way for MAC as presented in Definition 1.

Let us now define the regions $\mathcal{R}_A^a, \mathcal{R}_B^a$, which are contained in $\mathbb{R}_+^2$, i.e., the set of nonnegative 2-D real numbers:

$$\mathcal{R}_A^a = \bigcup_{\mathcal{P}}\left\{\begin{array}{l}R_1 \le H(Z_1|U,A_1) + I(X_1;Y|X_2,Z_1,U,A_1),\\ R_2 \le I(X_2;Y|X_1,U,A_1),\\ R_1+R_2 \le I(X_1,X_2;Y|U,A_1,Z_1) + H(Z_1|U,A_1),\\ R_1+R_2 \le I(X_1,X_2;Y)\end{array}\right\} \tag{64}$$

where $\mathcal{P}$ is the set of all distributions of the form $P(u,a_1)P(x_1,z_1|u,a_1)P(x_2|u,a_1)P(y|x_1,x_2)$ such that $E[\Lambda_1(A_1)] \le \Gamma_1$. Note that $P(x_1,z_1|u,a_1)$ satisfies $P(x_1,z_1|u,a_1) = P(x_1|u,a_1)\mathbf{1}_{\{z_1=g_1(x_1,a_1)\}}$, where $g_1(x_1,a_1)$ is given as a part of the problem description. The region $\mathcal{R}_B^a$ is defined with the same set of inequalities as in (64), but $\mathcal{P}$ is the set of all distributions of the form $P(u,a_1)P(x_1,z_1|u,a_1)P(x_2|z_1,u,a_1)P(y|x_1,x_2)$ such that $P(x_1,z_1|u,a_1) = P(x_1|u,a_1)\mathbf{1}_{\{z_1=g_1(x_1,a_1)\}}$ and $E[\Lambda(A_1)] \le \Gamma$.

Theorem 6 (Capacity Region): The capacity regions of the MAC with actions and with strictly causal case (Case A), and mixed causal and strictly causal case (Case B), as described in Definition 2, are $\mathcal{R}_A^a$, and $\mathcal{R}_B^a$, respectively.

The proof is based on a minor modification of the proof of the capacity region of the MAC with partial cribbing presented in Theorem 1.

Proof: Achievability: Consider the achievability proof of Theorem 1 and replace U_i with the pair $(U_i, A_{1,i})$. Note that in the proof of Theorem 1, when there is cribbing only from Transmitter 1 to Transmitter 2, U^n in block b is a function of $m'_{1,b-1}$.

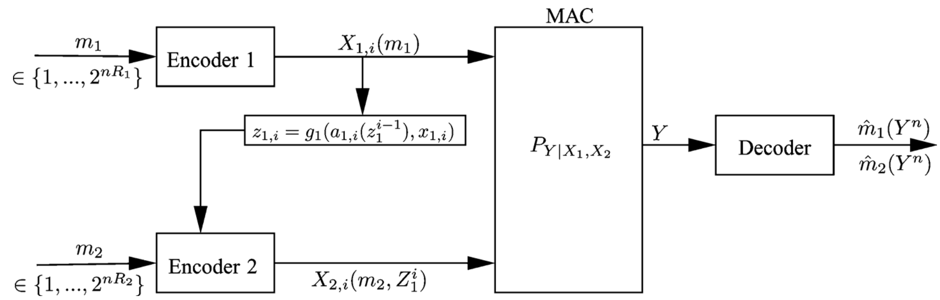


Fig. 7. Partial cribbing with actions. The action at time i is $a_{1,i}$ and is determined by previous cribbed observations, i.e., z_1^{i-1}. The cribbed signal $z_{1,i}$ from Encoders 1 to 2 is given by the deterministic function $z_{1,i} = g_1(a_{1,i}, x_{1,i})$. There exists a constraint on the actions of the form $\frac{1}{n}\sum_{i=1}^n E[\Lambda(A_{1,i})] \leq \Gamma$.

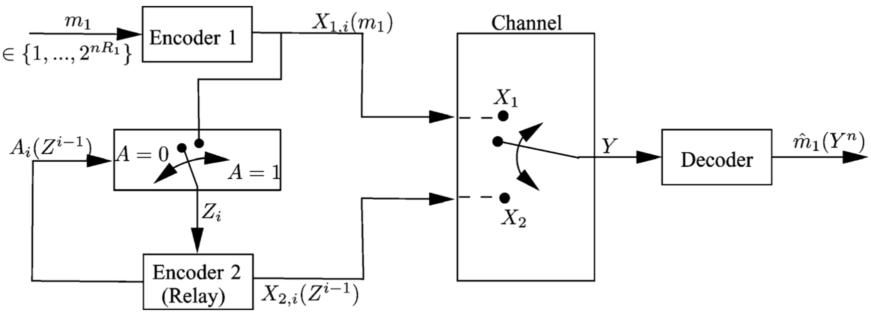


Fig. 8. Example of deterministic cribbing with actions. The relay (Encoder 2) takes an action A_i at time i that depends on previous cribbing, i.e., Z^{i-1}. The cribbing signal Z_i equals to $X_{1,i}$ if $A_i = 1$ and is constant otherwise. The cribbing is a limited resource. Hence, there exists a constraint on the portion of time that Encoder 2 can crib the signal from Encoder 1, namely, $\frac{1}{n}\sum_{i=1}^n E[\Lambda(A_i)] \leq \Gamma$. The output channel Y is randomly chosen with equal probability to be either X_1 or X_2.

Furthermore, in the current setting, the controller is an entity that decides on the actions as a function of previous cribbed signals. From previous cribbed signals, it is able to decode $m'_{1,b-1}$ at the end of block $b-1$, and therefore, one can treat A_i as part of U_i in the achievability part.[2]

Converse: Consider the converse proof of Theorem 1. Following identical steps as in (11)–(15), we obtain

$$\begin{aligned} n(R_1+R_2) &\leq \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i) + n\epsilon_n \\ n(R_1+R_2) &\leq \sum_{i=1}^n H(Z_{1,i}|Z_1^{i-1}) + I(X_{1,i}, X_{2,n}; Y_i|Z_1^i) \\ &\quad + n\epsilon_n, \\ nR_1 &\leq \sum_{i=1}^n H(Z_{1,i}|Z_1^{i-1}) + I(Y_i; X_{1,i}|X_{2,i}, Z_1^i) \\ &\quad + n\epsilon_n \\ nR_2 &\leq \sum_{i=1}^n I(Y_i; X_{2,i}|X_{1,i}, Z_1^i) + n\epsilon_n. \end{aligned} \tag{65}$$

Using the fact that A_i is a function of Z_1^{i-1} and by denoting $U_i \triangleq Z_1^{i-1}$ and, finally, using the auxiliary random variable Q as in the converse proof of Theorem 1, we obtain the converse proof. ■

[2]Note that in (64) the auxiliary r.v. U and the action A_1 appears always jointly and it is tempting to merge $A_{1,i}$ into auxiliary r.v. U. However, in the constraint $E[\Lambda(A_1)] \leq \Gamma$ and in the probability structure $P(x_1, z_1|u, a_1) = P(x_1|u, a_1)\mathbf{1}_{\{z_1 = g_1(x_1, a_1)\}}$, the action variable appears separately from the auxiliary r.v. U. Hence, this merging step is not possible.

Example 1 (Deterministic Relay With Actions): Consider the case where only Encoder 1 has a message to transmit and Encoder 2 has no message of its own to transmit, but helps to increase the rate of Encoder 1. Encoder 2, which plays the role of a relay, takes an action A_i that is a function of the observed signal up to time $i-1$, i.e., Z^{i-1}. If $A_i = 1$, then $Z_i = X_i$, and otherwise, Z_i is a constant. The cribbing signal Z_i is observed at Encoder 2 with a delay. There exists a constraint that $\frac{1}{n}\sum_{i=1}^n E[\Lambda(A_i)] \leq \Gamma$. In addition, Encoder 2 transmits a signal $X_{2,i}$ through the channel at time i, where $X_{2,i}$ is a function of Z^{i-1}. The output channel Y is randomly chosen with equal probability to be either X_1 or X_2. This example is illustrated in Fig. 8 and is a special case of the setting presented in Fig. 7.

The next lemma establishes the capacity region of a deterministic relay with actions, which is a special case of MAC with partial cribbing and actions where $R_2 = 0$.

Lemma 7: The capacity region of partial deterministic cribbing with actions, where only Encoder 1 sends a message, i.e., $R_2 = 0$, and there exists a delay in the cribbing (Case A), is given by (66), shown at the bottom of the page. If there is no delay in the cribbing (Case B), i.e., $X_{2,i}(Z^i)$, then we have (67), given at the bottom of the next page.

$$R_1 = \max_{P_{X_1,X_2,A}: E[\Lambda(A)] \leq \Gamma} \min\{H(Z|X_2, A) + I(X_1; Y|X_2, Z, A), I(Y; X_1, X_2)\} \tag{66}$$

Proof: Since $R_2 = 0$, it follows from (64) by replacing Z_2 with a constant and denoting Z_1 as Z that

$$R_1 \leq \max_{\mathcal{P}} \min\{H(Z|U,A)+I(X_1;Y|X_2,Z,U,A),I(X_1,X_2;Y)\}. \quad (68)$$

For the case where there is a delay in the cribbing (case A), the set of joint distributions $\mathcal{P}$ is of the form $P(u,a,x_1)P(x_2|u,a)P(y|x_1,x_2)$ and Z is a function of A and X. Manipulating the first term in the minimum in (68), we obtain

$$\begin{aligned} R_1 &\stackrel{(a)}{\leq} H(Z|U,A,X_2) + I(X_1;Y|X_2,Z,U,A) \\ &\stackrel{(b)}{\leq} H(Z|A,X_2) + I(X_1;Y|X_2,Z,A) \end{aligned} \quad (69)$$

where step (a) follows from the Markov chain $X_2 - (U,A) - X_1 - Z$ and step (b) from the fact that conditioning reduces entropy and from the Markov chain $Y - (X_1,X_2,Z,A) - U$. By choosing $U = X_2$, the first term of (68) becomes the upper bound in (69); hence, (66) is the capacity region.

When there is no delay in the cribbing, the capacity region is simply (68) where the set of joint distribution $\mathcal{P}$ is of the form $P(u,a,x_1)P(x_2|u,a,z)P(y|x_1,x_2)$ and z is a deterministic function of a and x. ∎

For the case of a delay in the cribbing, the action A_i can be seen as part of the output signal from Encoder 2 to the channel, and indeed, by replacing X_2 in (52) with (X_2, A), we obtain (66). Note that (52) holds for the semideterministic relay and $z_{1,i} = g_1(x_{1,i}, a_i)$ fits the semideterministic relay setting, when $(X_{2,i}, A_i)$ is the output of the relay and $Z_{1,i}$ is the input to the relay. However, in the case of no delay in the cribbing i.e., $X_2(Z^i)$, the replacement of X_2 is not possible since the action must have a delay i.e., $A_i(Z^{i-1})$.

To obtain a numerical solution, when there is a delay in the cribbing, namely, evaluating (66) for the example in Fig. 8, we can assume, without loss of optimality, that

$$\begin{aligned} \Pr(A=1) &= \Gamma \\ \Pr(X_1 = X_2|A=0) &= \alpha_0 \\ \Pr(X_1 = X_2|A=1) &= \alpha_1. \end{aligned} \quad (70)$$

The reason one can assume $\Pr(A=1) = \Gamma$ is as follows: if this is not the case, and one has a code where the portion of $\Pr(A=1)$ is smaller than Γ, one can add actions $A = 1$ for some portion of the time without decreasing the performance of the code. Furthermore, since the channel is symmetric with respect to 0 and 1 (by exchanging 0 and 1 for the inputs to the channel the performance of the code remains the same), only the probability $\Pr(X_1 = X_2)$ is important (a rigorous proof of this claim is given in the Appendix). Furthermore, based on the same reasons, one can also assume that $P(x_1)$ and $P(x_2)$ are $\mathrm{Bernoulli}(\frac{1}{2})$ without loss of optimality. Now we shall compute the terms in (66) as follows:

$$\begin{aligned} I(Y;X_1,X_2) &= H(Y) - H(Y|X_1,X_2) \\ &= 1 - \Gamma + \alpha_1\Gamma - (1-\Gamma)(1-\alpha_0) \\ &= \alpha_1\Gamma + \alpha_0(1-\Gamma) \end{aligned} \quad (71)$$

$$H(Z|X_2,A) = \Gamma H_b(\alpha_1) \quad (72)$$

$$\begin{aligned} I(X_1;Y|X_2,Z,A) &= H(Y|X_2,Z,A) - H(Y|X_1,X_2,A) \\ &\stackrel{(a)}{=} \Gamma(1-\alpha_1) + (1-\Gamma)H_b\left(\frac{1+\alpha_0}{2}\right) \\ &\quad - \Gamma(1-\alpha_1) - (1-\Gamma)(1-\alpha_0) \\ &= (1-\Gamma)\left(H_b\left(\frac{1+\alpha_0}{2}\right) + \alpha_0 - 1\right) \end{aligned} \quad (73)$$

where step (a) in (73) is due to the fact that $\Pr(Y = X_2|X_2, a=0) = \alpha_0 + \frac{1-\alpha_0}{2}$ and, therefore, $H(Y|X_2,Z,A) = \Gamma(1-\alpha_1) + (1-\Gamma)H_b(\frac{1+\alpha_0}{2})$ where $H_b(p)$ is the binary entropy, i.e., $-p\log p - (1-p)\log(1-p)$ for $0 \leq p \leq 1$. Hence, the capacity of the setting in Fig. 8 as a function on the constrain of the action Γ is given by (74), shown at the bottom of the page. The capacity $C(\Gamma)$ is depicted in Fig. 9 and can be found simply using a grid search on $0 \leq \alpha_0, \alpha_1 \leq 1$ or using convex optimization tools. In the case that $\Gamma = 0$, $X_{2,i}$ is independent of the message m_1. Therefore, we obtain that at any time i, the channel from Encoder 1 to the output behaves as a Z-channel if $X_{2,i} = 0$ and as an S channel if $X_{2,i} = 1$ and the capacity of these two channels are $H_b(\frac{1}{5}) - \frac{2}{5}$, and therefore, $C(0) = H_b(\frac{1}{5}) - \frac{2}{5}$. For the case that $\Gamma = 1$, we obtain from (74) that $C(1) = \max_{\alpha_1}\min(\alpha_1, H_b(\alpha_1))$. The α that maximizes the expression of $C(1)$ is the one that solves the equation $\alpha_1 = H_b(\alpha_1)$. From Fig. 9, we note that the capacity curve is strictly concave, and this implies that a naive time sharing between no cribbing ($\Gamma = 0$) and cribbing ($\Gamma = 1$) is strictly suboptimal.

IX. Conclusion and Further Research Directions

We have considered the problem of MACs with partial cribbing encoders, where in a two-encoder MAC, the observed cribbed signal at an encoder is a deterministic function of the output of the other encoder. We have characterized the capacity region for the two cases where the partial cribbing is causal or strictly causal. Rate splitting is the main additional technique used in the achievability proof over the techniques used for perfect cribbing by Willems and van der Meulen [17]. The extension of perfect cribbing to partial cribbing resembles the

$$R_1 = \max_{P_{U,X_1,A}P_{X_2|U,Z,A}: E[\Lambda(A)] \leq \Gamma} \min\{H(Z|U,A) + I(X_1;Y|X_2,Z,U,A), I(Y;X_1,X_2)\} \quad (67)$$

$$C(\Gamma) = \max_{0 \leq \alpha_0,\alpha_1 \leq 1} \min\left(\Gamma H_b(\alpha_1) + (1-\Gamma)\left(H_b\left(\frac{1+\alpha_0}{2}\right) + \alpha_0 - 1\right), \alpha_1\Gamma + \alpha_0(1-\Gamma)\right) \quad (74)$$

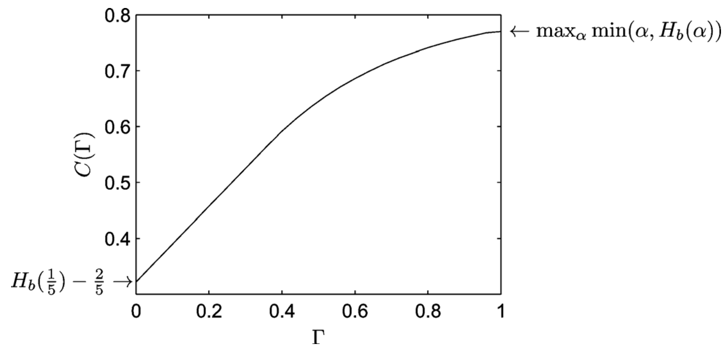


Fig. 9. Capacity of setting in Fig. 8 as a function of the action constraint Γ. For the case where $\Gamma = 0$, the capacity can be solved analytically since it is the capacity of the Z channel. The capacity where $\Gamma = 1$ is the simple expression $\max_{\alpha_1} \min(\alpha_1, H_b(\alpha_1))$, which can be solved numerically by solving $\alpha = H_b(\alpha)$. Note that the capacity curve is strictly concave; hence, time sharing between cribbing and not-cribbing is suboptimal.

extension of the decode-and-forward technique for the relay to the partial-decode-and-forward technique [16]. The method we used for partial cribbing may also be used for noisy cribbing, although in general the capacity region of noisy cribbing is an open question. Another question that has not been solved yet is the noncausal partial cribbing. For the perfect cribbing case, Willems and van der Meulen [17] solved the noncausal case simply by showing that causal and noncausal perfect cribbing results in the same capacity region.

Solving the partial cribbing setting motivated us to solve an action-dependent cribbing problem. In this paper, we considered the case where the action is only a function of the previously observed cribbing. However, the case in Fig. 7, where the action is a function of the previously observed cribbing and the message of the cribbing encoder, i.e., $a_{1,i}(z_1^{i-1}, m_1)$, remains open. Issues of this nature arise naturally in the sphere of cognitive communication systems where sensing other users' signals is a resource with a cost.

APPENDIX
CLAIM OF SYMMETRY IN THE SOLUTION OF EXAMPLE 1

In the process of finding the capacity of the deterministic relay channel with actions, we claim that only the probabilities $\Pr(X_1 = X_2|A = 0)$ and $\Pr(X_1 = X_2|A = 1)$ should be considered due to symmetry. In other words, we claim that it suffices to consider probabilities of $P_{X_1,X_2|A=0}$ where

$$\begin{aligned} P_{X_1,X_2|A=0}(0,0) &= P_{X_1,X_2|A=0}(1,1) \\ P_{X_1,X_2|A=0}(0,1) &= P_{X_1,X_2|A=0}(1,0) \end{aligned} \tag{75}$$

and similar equalities hold for $P_{X_1,X_2|A=1}$.

One can prove such a claim using convexity of the objective or using the converse proof, as we will do here. In the converse proof (Theorem 6), we show that X_1, X_2, and A is actually $X_{1,Q}$, $X_{2,Q}$, and A_Q, respectively, where Q is chosen uniformly over $\{1, 2, \ldots, n\}$.

Consider the operational code as given in Definition 2.

$$\begin{aligned} &\text{Encoder 1} && f_{1,i} : \{1, \ldots, 2^{nR_1}\} \mapsto \mathcal{X}_{1,i} && (76) \\ &\text{Encoder 2} && f_{2,i} : \{1, \ldots, 2^{nR_2}\} \times \mathcal{Z}_1^{i-1} \mapsto \mathcal{X}_{1,i} \\ &\text{Controller} && f_{a,i} : \mathcal{Z}_1^{i-1} \mapsto \mathcal{A}_{1,i} \\ &\text{Decoder 1} && h : \mathcal{Y}^n \mapsto \{1, \ldots, 2^{nR_1}\} \times \{1, \ldots, 2^{nR_2}\}. && (77) \end{aligned}$$

The code needs to satisfy the constraint $\frac{1}{n}\sum_{i=1}^{n} E[\Lambda_1(A_{1,i})] \leq \Gamma_1$, and the probability of error is given by P_e. Now, we design a code that is symmetric in the sense that any input 0 is replaced by 1 and any input 1 is replaced by 0. The symmetric code is given by

$$\begin{aligned} &\text{Encoder 1} && f_{1,i} : \{1, \ldots, 2^{nR_1}\} \mapsto \overline{\mathcal{X}}_{1,i} && (78) \\ &\text{Encoder 2} && f_{2,i} : \{1, \ldots, 2^{nR_2}\} \times \overline{\mathcal{Z}}_1^{i-1} \mapsto \overline{\mathcal{X}}_{1,i} \\ &\text{Contoller} && f_{a,i} : \overline{\mathcal{Z}}_1^{i-1} \mapsto \mathcal{A}_{1,i} \\ &\text{Decoder 1} && h : \overline{\mathcal{Y}}^n \mapsto \{1, \ldots, 2^{nR_1}\} \times \{1, \ldots, 2^{nR_2}\} && (79) \end{aligned}$$

where $\overline{x}$ denotes the complementary value over the binary set $\{0, 1\}$. Note that the symmetric code is a valid code and has the same error performance as the original code because of the symmetry of the channel. Therefore, the concatenation of these two codes is also a valid code, and also note that for the concatenated code, we have

$$P(x_{1,Q}, x_{2,Q}, a_Q) = P(\overline{x}_{1,Q}, \overline{x}_{2,Q}, a_Q). \tag{80}$$

ACKNOWLEDGMENT

The authors are grateful to Tsachy Weissman for very helpful discussions and for supporting the work through NSF grants CCF-1049413, Award-1049413 and 4101–38047, and to Yossef Steinberg for pointing out to us the motivation of partial cribbing for the Gaussian case and for additional helpful discussions. In addition, the authors would like to thank the associate editor and the anonymous referees for their suggestions, which helped them to significantly improve the clarity of the paper.

REFERENCES

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

[2] S. I. Bross, Y. Steinberg, and S. Tinguely, "The causal cognitive interference channel," presented at the Int. Zurich Semin. Commun., Mar. 2010.

[3] S. I. Bross and A. Lapidoth, "The state-dependent multiple-access channel with states available at a cribbing encoder," in *Proc. IEEE 26th Convent. Electr. Electron. Eng. Israel*, 2010, pp. 665–669.

[4] F. M. J. Willems, "The multiple-access channel with cribbing encoders revisited," presented at the Workshop Math. Relay. Cooperat. Commun. Netw., Tutorial Lecture at Math. Sci. Res. Inst., Berkeley, CA, Apr. 10–12, 2006.

[5] A. A. E. Gamal and M. R. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 3, p. 536, May 1982.
[6] T. M. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.
[7] S. I. Gelfand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *IEEE Trans. Inf. Theory*, vol. 16, pp. 24–34, 1980.
[8] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5396–5411, Nov. 2010.
[9] H. H. Permuter and T. Weissman, "Source coding with a side information "vending machine"," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4530–4544, Jul. 2011.
[10] H. Asnani, H. H. Permuter, and T. Weissman, "Probing capacity," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7317–7332, Nov. 2011.
[11] H. Asnani, H. H. Permuter, and T. Weissman, "To feed or not to feed back," 2010 [Online]. Available: arxiv.org/abs/1011.1607
[12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[13] F. Willems, E. C. Van Der Meulen, and J. P. M. Schalkwijk, "An achievable rate region for the multiple access channel with generalized feedback," in *Proc. Commun., Control, Comput. (Allerton)*, Oct. 2010, pp. 284–292.
[14] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 4/5, pp. 265–444, 2007.
[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New-York: Wiley, 2006.
[16] A. E. Gamal and Y.-H. Kim, *Network Information Theory*, 2nd ed. New York: Cambridge Univ. Press, 2011.
[17] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.
[18] N. Lauritzen, Lectures on convex sets 2010, http://home.imf.au.dk/niels/lecconset.pdf.
[19] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devel.*, vol. 2, pp. 289–293, 1958.
[20] D. Slepian and J. K. Wolf, "A coding theorem for multiple-access channel with correlated sources," *Bell Syst. Tech. J.*, vol. 51, pp. 1037–1076, 1973.
[21] A. E. Gamal, N. Hassanpour, and J. P. Mammen, "Relay networks with delays," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3413–3431, Oct. 2007.
[22] F. Willems and E. C. van der Meulen, "Strategy results for the relay channel without delay," in *Proc. 29th Symp. Inf. Theory, Benelux*, May 29–30, 2008, pp. 113–120.
[23] J. Wang, J. Chen, L. Zhao, P. Cuff, and H. H. Permuter, "A random variable substitution lemma with applications to multiple description coding," 2009 [Online]. Available: arxiv.org/abs/0909.3135
[24] F. M. J. Willems, "The discrete memoryless multiple channel with partially cooperating encoders," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 441–445, May 1983.
[25] S. I. Bross, A. Lapidoth, and M. A. Wigger, "The Gaussian MAC with conferencing encoders," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 2702–2706.
[26] S. I. Bross, A. Lapidoth, and M. A. Wigger, The Gaussian MAC with conferencing encoders 2008, Arxiv: abs/0805.0516.

Himanshu Asnani (S'11) is currently a Ph.D. candidate in Information Systems Lab, Electrical Engineering Department at Stanford University. He is advised by Prof. Tsachy Weissman and co-advised by Prof. Balaji Prabhakar. His research interests include information theory, probability theory and statistical learning. He received his B.Tech. from IIT Bombay and M.S. from Stanford University in 2009 and 2011, respectively. He is a Stanford Graduate Fellow (SGF) and recipient of Best Paper Award at MobiHoc 2009. He was also finalist for Student Paper award in ISIT 2011, Saint Petersburg, Russia.

Haim H. Permuter (M'08) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degree in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and Ph.D. degrees 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. He is currently a senior lecturer at Ben-Gurion university.

Dr. Permuter is a recipient of the Fulbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and the 2009 U.S.-Israel Binational Science Foundation Bergmann Memorial Award.