Feedback, Cribbing, and Causal State Information on the Multiple-Access Channel

Annina Bracher, *Student Member, IEEE*, and Amos Lapidoth, *Fellow, IEEE*

***Abstract*— The benefits afforded by feedback and/or causal state information (SI) on the state-dependent discrete memoryless multiple-access channel (SD-MAC) with cribbing encoder/s are studied. Capacity regions are derived for communication scenarios whose capacities without cribbing are still unknown. It is shown that when the encoders can crib, the SD-MAC behaves less like a MAC and more like a single-user channel: 1) feedback does not help; 2) strictly causal SI does not help; and 3) causal SI to both encoders is best utilized using Shannon strategies. However, in asymmetric settings, the single-user-like behavior may or may not occur. For example, the SD-MAC with only one cribbing encoder is single-user-like when the state is revealed to the cribbing encoder, but not if it is revealed to the noncribbing encoder.**

***Index Terms*— Conferencing, cribbing, feedback, multiple-access channel, state information.**

I. Introduction

A CRIBBING encoder for the multiple-access channel (MAC) is an encoder that, in addition to its own message, also gets to see the past channel inputs that were produced by the other encoder. The symbol it produces at Time i is thus a function of its message and of the symbols that were produced by the other encoder before Time i.[1] As we shall see, when the encoders can crib, the state-dependent, discrete, memoryless MAC (SD-MAC) behaves less like a MAC and more like a single-user channel: feedback does not help, strictly-causal SI does not help, and causal SI at both encoders is optimally utilized using Shannon Strategies.[2] The assumption that the encoders can crib thus enables us to compute the capacity regions for networks whose capacities without cribbing are still unknown. And since cribbing cannot hurt, the cribbing assumption can lead to useful outer bounds.

For example, the capacity of a discrete memoryless multiple-access channel with feedback is, in general, still unknown. However, as we shall see (Theorem 5), if the encoders are allowed to crib then the capacity with feedback is the same as without it. The latter capacity, which was computed by Willems and Van der Meulen [21], is thus an outer bound on the feedback capacity of the MAC without cribbing. In fact, feedback does not increase capacity even if only one encoder is allowed to crib (Theorem 8), thus leading to a tighter outer bound on the feedback capacity (Corollary 9). This bound is tight whenever one of the encoders can compute the symbol produced by the other from the channel output and the symbol it produced itself. And it leads to an operational meaning to the Dependence-Balance bound [8].

The fact that when the encoders can crib feedback does not increase capacity lends credence to the explanation that feedback on the MAC can increase capacity because knowing the output allows each encoder to learn something about the symbols produced by the other and thus learn something about the other's message. But this explanation fails to explain why feedback does not increase capacity also when only one of the encoders can crib.

Another example is the SD-MAC with strictly-causal state information (SI) at the encoders. This channel was studied by Lapidoth and Steinberg, who obtained inner and outer bounds on its capacity [11]. The exact capacity is to date unknown. However, when the encoders can crib, strictly-causal SI does not increase capacity (Theorem 5), so the cribbing capacity is an outer bound.

Things get more interesting when, rather than strictly-causally, the SI is revealed to the encoders *causally*. Here too only inner and outer bounds on the capacity are known [11]. We do know that Shannon Strategies are suboptimal [11]. However, when the encoders can crib, Shannon Strategies are optimal, and we can thus characterize the capacity region (Theorem 11). This region is, of course, an outer bound on the region without cribbing.

The SD-MAC with cribbing encoders need not behave like a single-user channel if information is furnished to the encoders in an asymmetric way. For example, if both encoders crib and the channel state is provided causally to only one encoder, then Shannon Strategies need not be optimal (Example 2). By analyzing various communication scenarios where the SD-MAC with cribbing encoders behaves more like a single-user channel and less like a MAC, and by providing several examples of communication scenarios where it does not,

Manuscript received August 29, 2013; revised September 3, 2014; accepted September 3, 2014. Date of publication October 28, 2014; date of current version November 18, 2014. This paper was presented at the 2012 IEEE Information Theory Workshop.

The authors are with ETH Zurich, Zurich 8092, Switzerland (e-mail: bracher@isi.ee.ethz.ch; lapidoth@isi.ee.ethz.ch).

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

Digital Object Identifier 10.1109/TIT.2014.2365189

[1] In the terminology of Willems and Van der Meulen who introduced cribbing [21], the cribbing encoders we study in this paper are "strictly-causally cribbing encoders".

[2] In this paper, we call every function from the support set of the channel state to the alphabet of a channel input a "Shannon Strategy". We say that Shannon Strategies are optimal if the coding scheme that achieves capacity in the absence of SI also achieves capacity in the presence of causal SI provided that each encoder with causal SI first produces a sequence of Shannon Strategies, and then computes its Time-i channel input by evaluating the i-th Shannon Strategy for the realization of the channel state at Time i. Shannon showed that causal SI to the encoder of the discrete, memoryless single-user channel is optimally utilized using Shannon Strategies [14], [7, Sec. 7.5].

we identify two conditions, which seem to be pivotal to the behavior of the SD-MAC with two cribbing encoders:

Condition 1: The Time-i channel output and each encoder's message are conditionally independent given the information available to the other encoder in forming it's Time-$(i + 1)$ *channel input.*

Condition 2: At every time i, the information that is available strictly-causally only to Encoder 1 (e.g., its message and possibly past channel states) and the information that is available strictly-causally only to Encoder 2 are conditionally independent given the information that is available strictly-causally to both encoders (e.g., cribbed inputs).

As we shall see, Conditions 1 and 2 hold for each communication scenario for which we show that the SD-MAC with two cribbing encoders is single-user-like, and our counterexamples violate Condition 1 and/or Condition 2. For example, Conditions 1 and 2 hold for the SD-MAC with causal SI to both cribbing encoders, for which Shannon Strategies are optimal, and our example of an SD-MAC with causal SI to only one of the cribbing encoders, for which Shannon Strategies are suboptimal, violates Condition 1.

Another setting where allowing the encoders to crib simplifies the analysis is the MAC or SD-MAC with feedback and *conferencing encoders*. Conferencing encoders are allowed to communicate with each other before or during the transmission via noise-free bit pipes of given capacities. In the absence of states and feedback, this network was introduced and solved by Willems [19]. He showed that there is no loss in optimality in requiring that the encoders conduct their conference before transmission begins and in replacing the sequential dialogue by two monologues each of which depends solely on the message of the soliloquising transmitter. Consequently, the resulting capacity is essentially that of the setting considered by Slepian and Wolf [15] where the encoders do not conference but transmit a common message in addition to their private messages.

It is unknown whether requiring the encoders to confer before transmission begins is also optimal in the presence of feedback. But this can be answered when the encoders can crib: we derive the capacity of the MAC with conferencing cribbing encoders that may be furnished with strictly-causal SI and/or feedback and show that it is not increased if the encoders confer during rather than before the transmission phase (Theorem 19). Also in the setting with conferencing encoders, strictly-causal SI and feedback to the cribbing encoders do not increase capacity.

If cribbing is not allowed, then the feedback capacity of the MAC with conferencing encoders is still unknown. We know that it is contained in the cribbing capacity. In fact, we can say more: The feedback capacity of the MAC with conferencing encoders is contained not only in the cribbing capacity of the MAC with conferencing encoders but in any Dependence-Balance outer bound on the feedback capacity of the MAC with a common message (Theorem 22).[3] This implies that, also in the presence of feedback, conferencing before transmission begins is optimal whenever a version of the Dependence-Balance bound is tight.

Ours is, of course, not the only work that builds on the original work of Willems and Van der Meulen [21], which introduced the (stateless) MAC with cribbing encoders. Recent work includes the work of Permuter and Asnani [1] on the (stateless) MAC with "partial cribbing encoders" that observe a deterministic function of each other's channel input. Limiting the cardinality of the co-domains of these functions leads to models that are more conservative than the cribbing model, especially when the cardinality of the input alphabets is very large.

Cooperative encoding on the SD-MAC was studied by Somekh-Baruch, Shamai, and Verdú [16] who computed the capacity of the MAC under the assumption that both encoders observe a common message and that Encoder 2 additionally observes a private message and non-causal or causal SI. Bross and Lapidoth [3] established the capacity of the SD-MAC under the assumption that the encoders observe private messages and that Encoder 2 cribs and observes the state sequence in a non-causal fashion [3].[4] For a comprehensive survey of the literature on state-dependent single-user and multi-terminal networks, see [9].

The remainder of this paper is structured as follows. First, we briefly describe our notation. The second section is dedicated to the SD-MAC with cribbing encoders that are furnished with strictly-causal SI and feedback. In the third section, we study the SD-MAC with causal SI at the cribbing encoders. The results on the MAC with conferencing encoders are presented in the fourth section. We conclude the paper with a brief summary.

A. Notation

We denote by $\mathcal{X}_1$ and $\mathcal{X}_2$ the support sets of the channel inputs produced by Encoder 1 and Encoder 2. The finite support set of the channel output is denoted by $\mathcal{Y}$ and that of the channel state by $\mathcal{S}$. We use $|\cdot|$ for the cardinality of a set, e.g., $|\mathcal{Y}|$ is the cardinality of the output alphabet $\mathcal{Y}$. We sometimes write $(\mathcal{X}_1 \times \mathcal{X}_2, W(y|x_1, x_2), \mathcal{Y})$ for a MAC of transition law $W(y|x_1, x_2)$ and $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{S}, W(y|x_1, x_2, s), \mathcal{Y})$ for an SD-MAC of transition law $W(y|x_1, x_2, s)$. We denote by $\mathcal{X}_1^{\mathcal{S}}$ the set of all functions from $\mathcal{S}$ to $\mathcal{X}_1$ and similarly for $\mathcal{X}_2^{\mathcal{S}}$.

Random variables are denoted by upper-case letters and their realizations or the elements of their support sets by lower-case letters, e.g., Y denotes the random channel output and $y \in \mathcal{Y}$ a value it may take.

Codewords, information exchanged during a conference, and sequences of Shannon Strategies are denoted by bold face lower or upper case letters depending on whether they are deterministic or random, e.g., $\mathbf{x}_1(m_1)$ is the codeword corresponding to Message $m_1 \in [1 : 2^{nR_1}] \triangleq \{1, 2, \ldots, 2^{nR_1}\}$. The integer n stands for the block-length and unless otherwise specified, sequences are assumed to be of length n.

Variables that occur at Time i are characterized by a subscript i, and for $k \in \{1, 2\}$ we define $\underline{k}$ to be the element

[3]Note that the outer bounds in [8] apply to the case with two private messages. The extension to the setting considered by Slepian and Wolf [15] where the encoders additionally observe a common message is however straightforward.

[4]If the input alphabet of the less-informed encoder is continuous, then one-sided cribbing is as good as (and no better than) a common message.

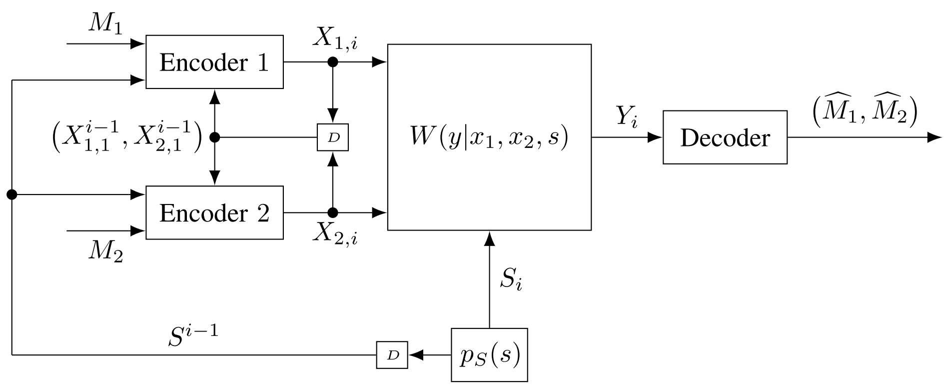

Fig. 1. SD-MAC with strictly-causal SI to the cribbing encoders.

of $\{1, 2\}$ that is not k, e.g., for $k = 1$ we denote by $X_{k,i}$ the channel input of Encoder 1 at Time i and by $X_{\underline{k},i}$ that of Encoder 2. Sequences of variables that occur in the time-range j to i bear a subscript j and a superscript i where the subscript $j = 1$ may be dropped, e.g., $X_{1,4}^5$ denotes the fourth and fifth channel input produced by Encoder 1, and Y^n denotes the entire output sequence.

A joint probability mass function (PMF), its marginal PMF, and its conditional PMF are all denoted by the same function $p(\cdot)$, with the exact meaning specified by the subscripts or arguments, e.g., $p_{X_1,X_2}(0, 1)$ denotes the probability of the event $(X_1, X_2) = (0, 1)$ and $p(x_1|x_2)$ the probability that $X_1 = x_1$ given $X_2 = x_2$.

We denote the set of ϵ-weakly-typical sequences of length n by $\mathcal{A}_\epsilon^{(n)}$, e.g., $\mathcal{A}_\epsilon^{(n)}(X, Y)$ is the set of length-n sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ that are jointly ϵ-weakly-typical w.r.t. $p_{X,Y}(\cdot, \cdot)$.

B. Functional Representation Lemma

In some scenarios—particulary if an encoder cribs—it can be shown that feedback is no better than strictly-causal SI, so if the latter does not increase capacity, then nor does the former. This is typically shown using the Functional Representation Lemma:

Lemma 1 [21, Functional Representation Lemma]: Given two discrete random variables X and Y, there exists a discrete random variable S, which is independent of X, and a function $g : \mathcal{X} \times \mathcal{S} \to \mathcal{Y}$ such that $Y = g(X, S)$.

The lemma allows us to view a MAC $W(y|x_1, x_2)$ as an SD-MAC $W(y|x_1, x_2, s)$ whose output can be computed from its inputs and state. If the state is revealed strictly-causally to a cribbing encoder, then this encoder can compute the past channel outputs from the past states, the past symbols it produced, and the past symbols that the other encoder produced (which it has learned by cribbing).

II. Cribbing and Strictly-Causal State Information

The SI discussed in this section is strictly-causal, and either both encoders crib (Section II-A) or only one (Section II-B).

A. Both Encoders Crib

We consider an SD-MAC where both encoders crib and both obtain the SI strictly-causally (see Figure 1). There is no need to address the case where the SI is revealed to only one, because even revealing it to both does not increase capacity (Proposition 3).

Recall that for $k \in \{1, 2\}$ we defined $\underline{k}$ to be the element of $\{1, 2\}$ that is not k.

Definition 1: For any two sets $\mathcal{M}_1$, $\mathcal{M}_2$ and positive integer $n \in \mathbb{N}$, an $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code for the SD-MAC $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{S}, W(y|x_1, x_2, s), \mathcal{Y})$ with strictly-causal SI to the cribbing encoders consists of two sequences of encoder mappings

$$f_{k,i} : \mathcal{M}_k \times \mathcal{X}_{\underline{k},1}^{i-1} \times \mathcal{S}^{i-1} \to \mathcal{X}_k, \tag{1}$$

where $k \in \{1, 2\}$, $i \in [1 : n]$, and a decoding mapping

$$\phi : \mathcal{Y}^n \to \mathcal{M}_1 \times \mathcal{M}_2 \tag{2}$$

such that the average probability of error P_e does not exceed ϵ, where

$$P_e = \sum_{\substack{(m_1,m_2)\in\mathcal{M}_1\times\mathcal{M}_2,\\ s^n\in\mathcal{S}^n,\ y^n\notin\phi^{-1}(m_1,m_2)}} \frac{\prod_{i=1}^{n} p(s_i)\, W(y_i|x_{1,i}, x_{2,i}, s_i)}{|\mathcal{M}_1|\,|\mathcal{M}_2|}, \tag{3}$$

$\phi^{-1}(m_1, m_2) \subset \mathcal{Y}^n$ is the decoding set of the message pair (m_1, m_2), and

$$x_{k,i} = f_{k,i}(m_k, x_{\underline{k},1}^{i-1}, s^{i-1}). \tag{4}$$

The rate pair (R_1, R_2) of the code is

$$R_1 = \frac{1}{n} \log |\mathcal{M}_1|, \quad R_2 = \frac{1}{n} \log |\mathcal{M}_2|. \tag{5}$$

A rate pair (R_1, R_2) is achievable if for every $\epsilon > 0$ and sufficiently large n there exists an $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code with rate pair $(\tilde{R}_1, \tilde{R}_2)$ satisfying $\tilde{R}_1 \geq R_1$ and $\tilde{R}_2 \geq R_2$. The capacity region is the closure of all achievable rate pairs.

We refer to this network as an SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^{i-1}).$$

If the encoders are also furnished with feedback, then we refer to the network as an SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s^{i-1}, y^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^{i-1}, y^{i-1}).$$

The capacity of the stateless MAC with cribbing encoders was found by Willems and Van der Meulen:

Theorem 2 [21, Th. 5]: The capacity region of the MAC $W(y|x_1, x_2)$ with cribbing encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1})$$

is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \leq H(X_1|U) \tag{6a}$$

$$R_2 \leq H(X_2|U) \tag{6b}$$

$$R_1 + R_2 \leq I(X_1, X_2; Y) \tag{6c}$$

for some joint PMF of the form

$$p(u, x_1, x_2, y) = p(u)\, p(x_1|u)\, p(x_2|u)\, W(y|x_1, x_2). \tag{7}$$

The capacity is achieved by the Block-Markov coding scheme of [21, Sec. V, Situation 5], which can be roughly described as follows: The encoders transmit their messages in blocks and—using the cribbed information—decode each other's previous message blocks in order to establish cooperation. During every transmission block, the encoders cooperatively send resolution information on the previous message blocks, which each encoder individually superimposes with its next message block. The resolution information can be associated with the random variable U in Theorem 2. Since Encoder 1 selects its channel input X_1 based on the resolution information and on its fresh information, and similarly Encoder 2, the channel inputs X_1 and X_2 are conditionally independent given U.

We offer the following intuition why the above Block-Markov coding scheme is optimal: One can check that the Time-i channel output and each encoder's message are conditionally independent given the information available to the other encoder in forming its Time-$(i+1)$ channel input, i.e., Condition 1 holds. As a consequence, there is no loss in optimality in requiring the encoders to decode each other's message (because if the decoder can decode a message, then so can the other encoder). Since also Condition 2 holds, i.e., the messages of the encoders are conditionally independent given the information that is available to both encoders (namely, all previous channel inputs), the decoded information is optimally utilized by the above Block-Markov coding scheme, where the channel inputs are independent conditional on the resolution information.

The capacity of the SD-MAC generally increases if the state is revealed to the encoders in a strictly-causal fashion [11]. But this is not the case when the encoders are allowed to crib. We state this as a proposition because this result will be strengthened in Theorem 5, which also allows feedback.

Proposition 3: Revealing the channel state in a strictly-causal fashion does not increase the capacity region of the SD-MAC with cribbing encoders.

Proof: See Section A-A. □

The following may offer some intuition for this result: It can be shown that Conditions 1 and 2 continue to hold when strictly-causal SI is provided to the cribbing encoders. As before, this implies that Block-Markov coding is optimal, and we can thus assume that it is used. In the presence of strictly-causal SI, we can view the cribbing link from Encoder 1 to Encoder 2 as a point-to-point channel with input X_1 and output (X_1, S). We know that the capacity of a point-to-point channel does not increase if the encoder (in this case Encoder 1) learns S strictly-causally. It is also clear that the described channel is no better than a channel whose input and output are both X_1. Hence, strictly-causal SI does not enhance the cribbing link from Encoder 1 to Encoder 2, and similarly for the cribbing link from Encoder 2 to Encoder 1. This implies that strictly-causal SI does not increase the amount of resolution information learned by cribbing and thus does not allow for additional cooperation between the encoders. We expect, moreover, that strictly-causal SI does not help the (fully-cooperative) transmission of resolution information because it does not help on the point-to-point channel.

Allowing the encoders to crib cannot hurt. And once they can crib the proposition shows that strictly-causal SI does not help. Thus:[5]

Corollary 4: The capacity region of the SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, s^{i-1}), \quad x_{2,i}(m_2, s^{i-1})$$

is contained in that of the same SD-MAC with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}).$$

Using the Functional Representation Lemma (Section I-B), we can strengthen Proposition 3 and show that when the encoders can crib the capacity is not increased even when we allow both strictly-causal SI and feedback:

Theorem 5: The capacity region of the SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s^{i-1}, y^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^{i-1}, y^{i-1})$$

is that of the same SD-MAC with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}).$$

Proof: See Section A-B. □

Allowing the encoders to crib cannot hurt. And once they can crib the theorem shows that feedback does not help. Thus:[6]

Corollary 6: The capacity region of the MAC $W(y|x_1, x_2)$ with encoders

$$x_{1,i}(m_1, y^{i-1}), \quad x_{2,i}(m_2, y^{i-1})$$

is contained in that of the same MAC with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}).$$

This outer bound is tight if each encoder can compute the other encoder's output from its own output and the channel output.

[5] The outer bound of Corollary 4 is looser than that of [11, Proposition. 3], but it is easier to evaluate. It is tighter than the full-cooperation bound [11, Proposition. 1].

[6] This result is also obtained if we view the channel inputs as the Parallel Channel Extension to the Dependence-Balance bound and set $Z = (X_1, X_2)$ in [8, Th. 3]. We discuss the Dependence-Balance bound for the MAC with conferencing encoders in Section IV-B ahead. If the encoders do not conference, then the Dependence-Balance bound [8, Th. 3] is that of Theorem 22 ahead with $C_{1,2} = C_{2,1} = 0$.

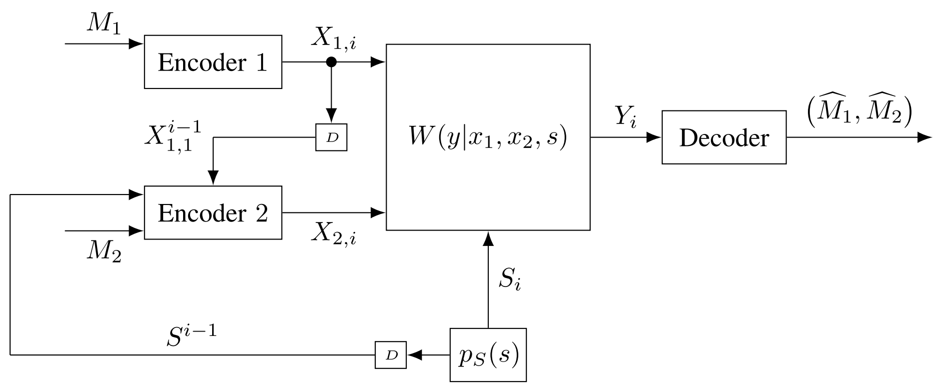

Fig. 2. SD-MAC with strictly-causal SI to the cribbing Encoder 2.

Proof: Theorem 5 implies that the feedback capacity region is contained in the cribbing capacity region. To see that this outer bound is tight if each encoder can compute the other encoder's output based on its own output and the channel output, observe that for such channels feedback is as good as the combination of feedback and cribbing. □

The outer bound of Corollary 6 will be strengthened in Corollary 9 ahead, where we show that even one cribbing encoder is better than feedback.

B. One Encoder Cribs

Suppose now that only one encoder cribs. One could then consider a scenario as in Figure 2 where the strictly-causal SI is furnished to the cribbing encoder or one where it is furnished to the non-cribbing encoder. Likewise for feedback. We shall see that in the former case the SI is useless but in the latter case it can be beneficial. Feedback does not increase capacity in either case.

An $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code with strictly-causal SI to the cribbing Encoder 2 consists of two sequences of encoder mappings

$$f_{1,i} : \mathcal{M}_1 \to \mathcal{X}_1, \tag{8a}$$
$$f_{2,i} : \mathcal{M}_2 \times \mathcal{X}_{1,1}^{i-1} \times \mathcal{S}^{i-1} \to \mathcal{X}_2, \tag{8b}$$

where $i \in [1 : n]$, such that the average probability of error does not exceed ϵ and

$$x_{1,i} = f_{1,i}(m_1), \tag{9a}$$
$$x_{2,i} = f_{2,i}(m_2, x_{1,1}^{i-1}, s^{i-1}). \tag{9b}$$

We refer to this network as an SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^{i-1}).$$

The capacity of this network when the state is null was found by Willems and Van der Meulen:

Theorem 7 [21, Th. 2]: The capacity region of the MAC $W(y|x_1, x_2)$ with encoders

$$x_{1,i}(m_1), \quad x_{2,i}(m_2, x_{1,1}^{i-1})$$

is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \le H(X_1|U) \tag{10a}$$
$$R_2 \le I(X_2; Y|X_1, U) \tag{10b}$$
$$R_1 + R_2 \le I(X_1, X_2; Y) \tag{10c}$$

for some joint PMF of the form

$$p(u, x_1, x_2, y) = p(u)\, p(x_1|u)\, p(x_2|u)\, W(y|x_1, x_2). \tag{11}$$

The capacity is achieved by the Block-Markov coding scheme of [21, Sec. V, Situation 2], which is similar to the one for the case where both encoders crib: The encoders divide their messages into blocks, and during every transmission block, the encoders cooperatively send resolution information, which each encoder individually superimposes with its next message block. Here, only the cribbing Encoder 2 decodes the previous message blocks of Encoder 1, and the encoders thus only send resolution information on the previous message blocks of Encoder 1. Since the decoder obtains no resolution information about the message blocks of Encoder 2, the rate at which Encoder 2 can transmit its message blocks is upper-bounded by (10b).

To explain why this scheme is optimal, we offer a similar intuition as in the case where both encoders crib: The Time-i channel output and the message of Encoder 1 are conditionally independent given the information available to Encoder 2 in forming its Time-$(i+1)$ channel input. As a consequence, there is no loss in optimality in requiring Encoder 2 to decode the message of Encoder 1. Since the messages of the encoders are conditionally independent given the information that is available to both encoders (namely, all previous channel inputs of Encoder 1), the decoded information is optimally utilized by the above Block-Markov scheme, where the channel inputs are independent conditional on the resolution information.

Suppose now that the state is not null. If none of the encoders cribs, then strictly-causal SI to even just one of the encoders can increase capacity [10], [12]. This is still true when one of the encoders cribs provided that it is the non-cribbing encoder to whom the SI is furnished:

Remark 1: The capacity region of the SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, s^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1})$$

can be strictly larger than that of the same SD-MAC with encoders

$$x_{1,i}(m_1), \quad x_{2,i}(m_2, x_{1,1}^{i-1}).$$

This is illustrated by the following example:

Example 1: Consider an SD-MAC whose inputs X_1, X_2 and state $S \sim \mathrm{Ber}(1/2)$ are binary, and whose output Y is

$$Y = (X_1, X_2 \oplus S). \tag{12}$$

Suppose that Encoder 2 cribs. In the absence of SI, the rate pair $(R_1, R_2) = (0, 1)$ is not achievable. But it is achievable if SI is provided to the non-cribbing encoder.

The example is formally analyzed in Appendix A-C. The intuition is that the non-cribbing encoder can transmit the state sequence to the decoder and thus help it decode the cribbing encoder's message.

The scenario where strictly-causal SI is provided only to the cribbing encoder is different: in this case the SI is useless. We present this result in a stronger form by also allowing feedback:

Theorem 8: The capacity region of the SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, y^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^{i-1}, y^{i-1})$$

is that of the same SD-MAC with encoders

$$x_{1,i}(m_1), \quad x_{2,i}(m_2, x_{1,1}^{i-1}).$$

Proof: See Section A-D. □

It is perhaps surprising that feedback to the non-cribbing encoder does not help. After all, it enables Encoder 1 to decode the message blocks of Encoder 2 and therefore allows for additional cooperation by allowing the encoders to also send resolution information on the previous message blocks of Encoder 2. It turns out that this additional cooperation does not increase capacity. Perhaps this is because once the decoder has retrieved the message of Encoder 1 it knows just as much about the message of Encoder 2 as Encoder 1, and therefore the resolution information that the encoders can offer on the previous message blocks of Encoder 2 cannot help the decoder.

Allowing one encoder to crib cannot hurt. And once one encoder cribs the theorem shows that strictly-causal SI to the cribbing encoder and feedback to both encoders do not help. Thus:[7,8]

Corollary 9: The capacity region of the SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, y^{i-1}), \quad x_{2,i}(m_2, s^{i-1}, y^{i-1})$$

is contained in that of the same SD-MAC with encoders

$$x_{1,i}(m_1), \quad x_{2,i}(m_2, x_{1,1}^{i-1}).$$

Since a null state can be viewed as being known to either encoder, the capacity region of the stateless MAC $W(y|x_1, x_2)$ with encoders

$$x_{1,i}(m_1, y^{i-1}), \quad x_{2,i}(m_2, y^{i-1})$$

is contained in the intersection of the capacity region of the same MAC with encoders

$$x_{1,i}(m_1), \quad x_{2,i}(m_2, x_{1,1}^{i-1})$$

and that of the same MAC with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}), \quad x_{2,i}(m_2).$$

Remark 2: The outer bound in Corollary 9 is expressed in terms of the capacity region of an enhanced network where one of the encoders can crib and where capacity is known and is not increased by feedback. It is reminiscent of the outer bound on the feedback capacity of a broadcast channel (BC) in terms of the capacity region of an enhanced network where one of the receivers also sees the signal received by the other. The enhanced network is a physically degraded BC whose capacity is known and is not increased by feedback [6].

The outer bound of Corollary 9 on the capacity of the MAC with feedback need not be tight. But for some channels it is:

Corollary 10: If (at least) one encoder can compute the other encoder's output from its own output and the channel output, then the outer bound of Corollary 9 on the feedback capacity of the stateless MAC is tight.[9] Moreover, one-sided feedback to the encoder that can perform this computation is as good as feedback to both.

Proof: Feedback to the encoder that can compute the other encoder's output based on its own output and the channel output is at least as beneficial as allowing it to crib. Hence the outer bound of Corollary 9 is achievable. □

Remark 3 [17]: If X_1 is computable from (X_2, Y), then (10) and (11) are equivalent to the Cover-Leung constraints

$$R_1 \leq I(X_1; Y|X_2, U) \tag{13a}$$
$$R_2 \leq I(X_2; Y|X_1, U) \tag{13b}$$
$$R_1 + R_2 \leq I(X_1, X_2; Y), \tag{13c}$$

where the joint PMF is of the form

$$p(u, x_1, x_2, y) = p(u)\, p(x_1|u)\, p(x_2|u)\, W(y|x_1, x_2). \tag{14}$$

Since the Cover-Leung region is achievable by one-sided feedback (no matter to which encoder) [4], [18], Remark 3 yields an alternative proof of Corollary 10 and allows us to even strengthen it:

Remark 4: If (at least) one encoder can compute the other encoder's output from its own output and the channel output, then one-sided feedback (no matter to which encoder) is as good as feedback to both encoders.

[7] The outer bound of Corollary 9 on the feedback capacity of the MAC is also obtained if we view the cribbed channel input as the Parallel Channel Extension to the Dependence-Balance bound: setting $Z = X_1$ in [8, Th. 3] results in the capacity region with a cribbing Encoder 2 and likewise $Z = X_2$ in that with a cribbing Encoder 1 (see Footnote 6).

[8] The network without feedback is a special case of the double-state MAC that was studied in [10] and [12].

[9] Note the analogy to Corollary 6, which states that the feedback capacity of the MAC is contained in the capacity without feedback but with two cribbing encoders and that the bound is tight if each encoder can compute the other encoder's output based on its own output and the channel output.

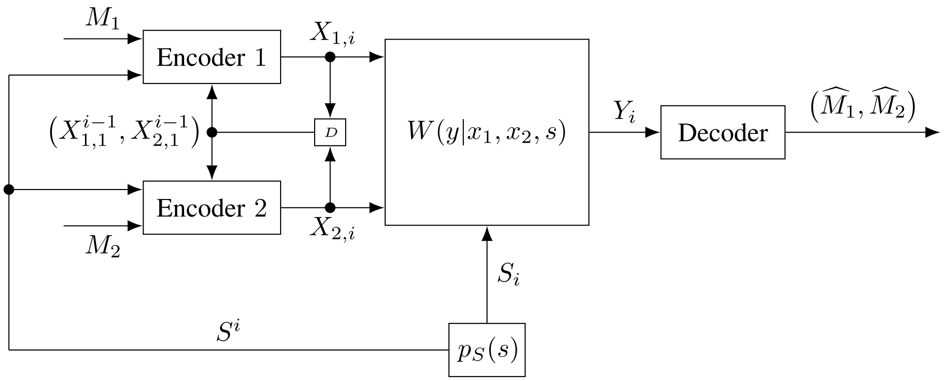

Fig. 3. SD-MAC with causal SI to the cribbing encoders.

III. Cribbing and Causal State Information

The SI in this section is causal. In Section III-A both encoders crib and both are cognizant of the SI. In Section III-B both encoders crib but the SI presented to them differ. Finally, in Section III-C only one encoder cribs and only the cribbing encoder is presented with the SI.

A. Both Encoders Crib and Observe the State

Consider an SD-MAC where both encoders crib and both obtain the SI causally (see Figure 3). For this setting an $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code is defined as in Definition 1 except that the inputs are allowed to depend also on the current channel state. The cribbing, however, is still strictly-causal. The code thus consists of two sequences of encoder mappings

$$f_{k,i} \colon \mathcal{M}_k \times \mathcal{X}_{\underline{k},1}^{i-1} \times \mathcal{S}^i \to \mathcal{X}_k, \tag{15}$$

where $k \in \{1, 2\}$, $i \in [1 : n]$, such that the average probability of error does not exceed ϵ and

$$x_{k,i} = f_{k,i}(m_k, x_{\underline{k},1}^{i-1}, s^i). \tag{16}$$

(Recall that for $k \in \{1, 2\}$ we defined $\underline{k}$ to be the element of $\{1, 2\}$ that is not k.) We refer to this network as an SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s^i), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^i).$$

On the single-user channel Shannon showed how to optimally use causal state information [14]: Each message m is mapped to a length-n sequence of functions $h_{m,1}, \ldots, h_{m,n} \in \mathcal{X}^{\mathcal{S}}$, and the Time-$i$ channel input that the encoder produces to convey Message m is $h_{m,i}(S_i)$. The elements of $\mathcal{X}^{\mathcal{S}}$ are called Shannon Strategies, and the mapping of messages to n-tuples of Shannon Strategies is obtained using random coding by choosing a capacity-achieving distribution on $\mathcal{X}^{\mathcal{S}}$. Once this mapping has been chosen, the Time-i encoder output depends only on the message and the Time-i state S_i. Knowledge of the past states is unnecessary.

In the absence of cribbing, the capacity of the SD-MAC with causal SI to the encoders is to date unknown. A naive approach is to extend Shannon's approach by choosing a product distribution on $\mathcal{X}_1^{\mathcal{S}} \times \mathcal{X}_2^{\mathcal{S}}$, by using random coding (and time-sharing) to produce n-tuples of Shannon Strategies $h_{m,i,k} \in \mathcal{X}_k^{\mathcal{S}}$, and by having the Time-i output of Encoder k be $h_{m,i,k}(S_i)$ whenever it wishes to convey Message m. Unlike the single-user channel, on the MAC this naive approach is not optimal [11].

When the encoders can crib the MAC behaves more like a single-user channel in the sense that combining Shannon Strategies with the scheme that achieves capacity in the absence of SI is optimal. The capacity of the SD-MAC with causal SI to the cribbing encoders is achieved by a Block-Markov coding scheme like that of [21, Sec. V, Situation 5]: At the beginning of each transmission block, the encoders use the past state information and the past cribbed information pertaining to the previous block to determine the resolution message to be transmitted in the present block. They then transmit this common message and their fresh information in the present block by applying the Slepian-Wolf coding scheme for the MAC with a common message but with input alphabets comprising the Shannon Strategies $\mathcal{X}_1^{\mathcal{S}}$ and $\mathcal{X}_2^{\mathcal{S}}$. In the following theorem the elements of $\mathcal{X}_k^{\mathcal{S}}$, $k \in \{1, 2\}$ are indexed by $t_k \in \left[1 : |\mathcal{X}_k^{\mathcal{S}}|\right]$, and the Shannon Strategy indexed by t_k is denoted $g_k(t_k, \cdot)$.

It is important to note that cribbing allows an encoder to observe the previous channel inputs that were produced by the other encoder but not the previous strategies that were used to produce these inputs. If at Time $i - 1$ this other encoder uses the Shannon Strategy t_k, then at Time i the cribbing encoder will not learn t_k but only the result of applying $g_k(t_k, \cdot)$ to the state S_{i-1}.

Theorem 11: The capacity region of the SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s^i), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^i)$$

is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \le H(X_1|S, U) \tag{17a}$$
$$R_2 \le H(X_2|S, U) \tag{17b}$$
$$R_1 + R_2 \le I(T_1, T_2; Y) \tag{17c}$$

for some random variables U, T_1, and T_2; functions $g_1 \colon \mathcal{T}_1 \times \mathcal{S} \to \mathcal{X}_1$ and $g_2 \colon \mathcal{T}_2 \times \mathcal{S} \to \mathcal{X}_2$; and a joint PMF of the form

$$\begin{aligned} &p(u, t_1, t_2, x_1, x_2, s, y) \\ &\quad = p(u)\, p(t_1|u)\, p(t_2|u)\, p(s) \\ &\qquad \times \mathbb{1}_{\{x_1 = g_1(t_1, s)\}} \mathbb{1}_{\{x_2 = g_2(t_2, s)\}} W(y|x_1, x_2, s). \end{aligned} \tag{18}$$

Proof: See Section A-E. $\square$

One can show that Conditions 1 and 2, which we believe to be key to the single-user-like behavior of the MAC with two cribbing encoders, also hold for the network of Theorem 11.

As we next show, in the setting of Theorem 11 providing additional feedback and/or strictly-causal SI $\tilde{S}^{i-1}$ does not increase capacity. Consider the SD-MAC $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{S} \times \tilde{\mathcal{S}}, W(y|x_1, x_2, (s, \tilde{s})), \mathcal{Y})$, which is governed by two (possibly correlated) channel states S and $\tilde{S}$, with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s^i, \tilde{s}^{i-1}, y^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^i, \tilde{s}^{i-1}, y^{i-1}).$$

We know that in the absence of causal SI, i.e., if S is null, the capacity region does not increase if the encoders are furnished with strictly-causal SI and feedback (Theorem 5). A similar result holds if S is not null:

Theorem 12: The capacity region of the SD-MAC $W(y|x_1, x_2, (s, \tilde{s}))$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s^i, \tilde{s}^{i-1}, y^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^i, \tilde{s}^{i-1}, y^{i-1})$$

is that (of Theorem 11) of the same SD-MAC with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s^i), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^i).$$

Proof: See Section A-F. □

Allowing the encoders to crib cannot hurt. And once they can crib the theorem shows that strictly-causal SI and feedback do not help. Thus:

Corollary 13: The capacity region of the SD-MAC $W(y|x_1, x_2, (s, \tilde{s}))$ with encoders

$$x_{1,i}(m_1, s^i, \tilde{s}^{i-1}, y^{i-1}), \quad x_{2,i}(m_2, s^i, \tilde{s}^{i-1}, y^{i-1})$$

is contained in that of the same SD-MAC with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s^i), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^i).$$

B. Both Encoders Crib and Observe Distinct States

Consider the SD-MAC $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}, W(y|x_1, x_2, (s_1, s_2, s)), \mathcal{Y})$, which is governed by three (possibly correlated) state sequences $\{S_{1,i}\}$, $\{S_{2,i}\}$, and $\{S_i\}$, and assume that both encoders crib. The state S_1 is revealed to Encoder 1 causally and likewise S_2 to Encoder 2. The state S is revealed to both encoders but strictly-causally. For this setting an $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code consists of two sequences of encoder mappings

$$f_{k,i} : \mathcal{M}_k \times \mathcal{X}_{\underline{k},1}^{i-1} \times \mathcal{S}_{k,1}^{i} \times \mathcal{S}^{i-1} \to \mathcal{X}_k, \tag{19}$$

where $k \in \{1, 2\}$, $i \in [1 : n]$, such that the average probability of error does not exceed ϵ and

$$x_{k,i} = f_{k,i}(m_k, x_{\underline{k},1}^{i-1}, s_{k,1}^{i}, s^{i-1}).$$

We refer to this network as an SD-MAC $W(y|x_1, x_2, (s_1, s_2, s))$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s_{1,1}^{i}, s^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s_{2,1}^{i}, s^{i-1}).$$

An inner bound on the capacity region can be obtained by considering a coding scheme that combines Shannon Strategies with Block-Markov coding. The resulting scheme is like the one of Theorem 11 except that the resolution information is now computed by Encoder k not only based on its past cribbing and the past realizations of S but also based on the past realizations of S_k.

Theorem 14: For the SD-MAC $W(y|x_1, x_2, (s_1, s_2, s))$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s_{1,1}^{i}, s^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s_{2,1}^{i}, s^{i-1}),$$

all rate pairs (R_1, R_2) satisfying

$$R_1 \le I(T_1; X_1|S, S_2, U) \tag{20a}$$

$$R_2 \le I(T_2; X_2|S, S_1, U) \tag{20b}$$

$$R_1 + R_2 \le I(T_1, T_2; Y) \tag{20c}$$

for some random variables U, T_1, and T_2; functions $g_1 : \mathcal{T}_1 \times \mathcal{S}_1 \to \mathcal{X}_1$ and $g_2 : \mathcal{T}_2 \times \mathcal{S}_2 \to \mathcal{X}_2$; and a joint PMF of the form

$$\begin{aligned} &p(u, t_1, t_2, x_1, x_2, s_1, s_2, s, y) \\ &\quad = p(u)\, p(t_1|u)\, p(t_2|u)\, p(s_1, s_2, s)\, \mathbb{1}_{\{x_1 = g_1(t_1, s_1)\}} \\ &\qquad \times \mathbb{1}_{\{x_2 = g_2(t_2, s_2)\}} W(y|x_1, x_2, (s_1, s_2, s)) \end{aligned} \tag{21}$$

are achievable.

Proof: See Section A-G. □

When S_1 and S_2 are the same and S is denoted $\tilde{S}$ the setting reduces to that of Theorem 12 and the bound is tight. But it need not be tight when S_1 and S_2 differ:

Remark 5: The inner bound of Theorem 14 need not be tight.

An example where the inner bound of Theorem 14 is loose is the following:

Example 2: For the SD-MAC $W(y|x_1, x_2, s_2)$ with cribbing encoders, binary inputs, a binary channel state $S_2 \sim$ Ber$(1/2)$, and a binary output $Y = X_2 \oplus S_2$, for which S and S_1 are null, any rate pair in the set

$$\left\{(R_1, R_2) \in (R_0^+)^2 : R_1 + R_2 \le 1\right\} \tag{22}$$

is achievable. The inner bound of Theorem 14 does not contain the rate pair $(R_1, R_2) = (0, 1)$ and is thus loose.[10]

The example is formally analyzed in Appendix A-H. Here, we provide an informal description of the proof idea: The rate pair (0, 1) is achievable by having Encoder 2 produce the XOR of its data bit with the state S_2 so as to have the output be the data bit. If, however, the encoders use the coding scheme of Theorem 14, then $R_2 = 1$ is not achievable. The reason is that this coding scheme requires the encoders to estimate each other's private message to establish the resolution message. And it is impossible for Encoder 1 to decode the data bits that are fed to Encoder 2 because the cribbing only allows it to see the outputs of Encoder 2, which are the result of the XOR of the data bits with S_2.

Our next result is an outer bound on the capacity region of the network of Theorem 14.

Theorem 15: The capacity region of the SD-MAC $W(y|x_1, x_2, (s_1, s_2, s))$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s_{1,1}^{i}, s^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s_{2,1}^{i}, s^{i-1})$$

[10] The capacity region of this SD-MAC can be obtained from Theorem 18 ahead: since the output $Y = X_2 \oplus S_2$ is independent of the input X_1 we can w.l.g. assume that only Encoder 2 cribs.

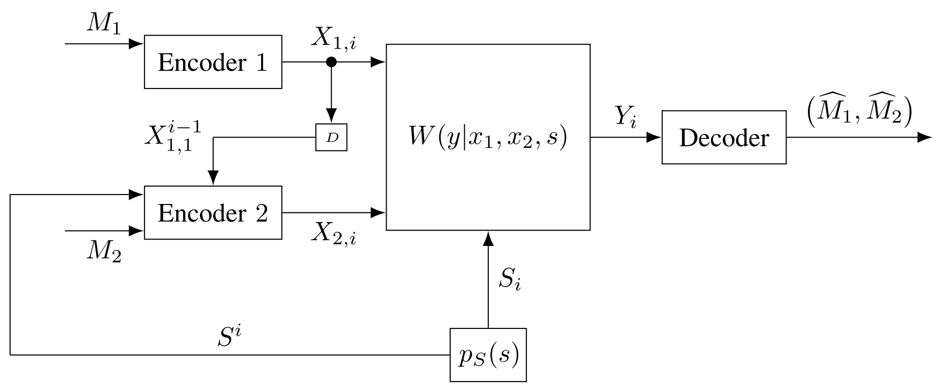

Fig. 4. SD-MAC with causal SI to the cribbing Encoder 2.

is contained in the set of rate pairs (R_1, R_2) *satisfying*

$$R_1 \le H(X_1|S_1, U) \quad (23a)$$
$$R_2 \le H(X_2|S_2, U) \quad (23b)$$
$$R_1 + R_2 \le I(T_1, T_2; Y) \quad (23c)$$

for some random variables U, T_1, *and* T_2*; functions* $g_1 \colon \mathcal{T}_1 \times \mathcal{S}_1 \to \mathcal{X}_1$ *and* $g_2 \colon \mathcal{T}_2 \times \mathcal{S}_2 \to \mathcal{X}_2$*; and a joint PMF of the form*

$$\begin{aligned} &p(u, t_1, t_2, x_1, x_2, s_1, s_2, s, y) \\ &\quad = p(u)\, p(t_1|u)\, p(t_2|u)\, p(s_1, s_2, s)\, \mathbb{1}_{\{x_1 = g_1(t_1, s_1)\}} \\ &\qquad \times \mathbb{1}_{\{x_2 = g_2(t_2, s_2)\}} W(y|x_1, x_2, (s_1, s_2, s)). \end{aligned} \quad (24)$$

Proof: See Section A-I. □

The inner and outer bounds of Theorems 14 and 15 coincide when there exists a one-to-one relationship between (S, S_1) and (S, S_2).[11] In this case Shannon Strategies and Block-Markov coding is thus optimal.

Corollary 16: For an SD-MAC $W(y|x_1, x_2, (s_1, s_2, s))$ *with encoders*

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s_{1,1}^{i}, s^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s_{2,1}^{i}, s^{i-1})$$

and a one-to-one relationship between (S, S_1) *and* (S, S_2) *the capacity region coincides with the outer bound of Theorem 15.*

Proof: We show that the achievable region of Theorem 14 coincides with the outer bound of Theorem 15. Fix sets $\mathcal{T}_1$ and $\mathcal{T}_2$; functions $g_1 \colon \mathcal{T}_1 \times \mathcal{S}_1 \to \mathcal{X}_1$ and $g_2 \colon \mathcal{T}_2 \times \mathcal{S}_2 \to \mathcal{X}_2$; and a joint PMF of the form (21). Theorem 15 implies that a rate pair (R_1, R_2) is achievable if it satisfies (20). Note that

$$I(T_1; X_1|S, S_2, U) \overset{a)}{=} I(T_1; X_1|S, S_1, U) \quad (25)$$
$$\overset{b)}{=} H(X_1|S, S_1, U) \quad (26)$$
$$\overset{c)}{=} H(X_1|S_1, U), \quad (27)$$

where a) holds because (S, S_1) and (S, S_2) are in a one-to-one relationship, b) is true because $X_1 = g_1(T_1, S_1)$, and c) holds because (21) implies that X_1, (S_1, U), and S form a Markov chain in that order. Similarly, one can show that

$$I(T_2; X_2|S, S_1, U) = H(X_2|S_2, U). \quad (28)$$

Hence, the rate constraints in (20) are equivalent to those in (23). Since (21) and (24) are identical, we conclude that the regions of Theorems 14 and 15 coincide. □

The corollary implies that Shannon Strategies and Block-Markov coding is optimal if the state S_1, which is available causally to Encoder 1, is available strictly-causally to Encoder 2, and similarly for S_2. We next show that this also holds if Encoder 2 observes S_1 strictly-causally and, rather than observing S_2 strictly-causally, Encoder 1 is presented with feedback.

Theorem 17: For an SD-MAC $W(y|x_1, x_2, (s_1, s_2, s))$ *with encoders*

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s_{1,1}^{i}, s^{i-1}, y^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s_{2,1}^{i}, s^{i-1}),$$

and where S_1 *is computable from* (S, S_2)*, the capacity region is the set of rate pairs* (R_1, R_2) *satisfying*

$$R_1 \le H(X_1|S_1, U) \quad (29a)$$
$$R_2 \le I(T_2; X_2, Y|X_1, S, S_1, U) \quad (29b)$$
$$R_1 + R_2 \le I(T_1, T_2; Y) \quad (29c)$$

for some random variables U, T_1, *and* T_2*; functions* $g_1 \colon \mathcal{T}_1 \times \mathcal{S}_1 \to \mathcal{X}_1$ *and* $g_2 \colon \mathcal{T}_2 \times \mathcal{S}_2 \to \mathcal{X}_2$*; and a joint PMF of the form*

$$\begin{aligned} &p(u, t_1, t_2, x_1, x_2, s_1, s_2, s, y) \\ &\quad = p(u)\, p(t_1|u)\, p(t_2|u)\, p(s_1, s_2, s)\, \mathbb{1}_{\{x_1 = g_1(t_1, s_1)\}} \\ &\qquad \times \mathbb{1}_{\{x_2 = g_2(t_2, s_2)\}} W(y|x_1, x_2, (s_1, s_2, s)). \end{aligned} \quad (30)$$

The capacity region does not increase if additional feedback is furnished to Encoder 2, i.e., if the encoders are of the form

$$x_{1,i}(m_1, x_{2,1}^{i-1}, s_{1,1}^{i}, s^{i-1}, y^{i-1}),\ x_{2,i}(m_2, x_{1,1}^{i-1}, s_{2,1}^{i}, s^{i-1}, y^{i-1}).$$

Proof: See Section A-J. □

The networks in Corollary 16 and Theorem 17 behave single-user-like in the sense that combining Shannon Strategies with the coding scheme that achieves capacity in the absence of SI is optimal, and one can check that they satisfy Conditions 1 and 2.

C. One Encoder Cribs

Suppose now that only Encoder 2 cribs and that it is the only encoder furnished with SI (causally) (see Figure 4).

[11] This scenario is not equivalent to a common causal SI because S_1 and S_2 are revealed here causally while S is only revealed strictly-causally.

For this setting an $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code consists of two sequences of encoder mappings

$$f_{1,i} \;:\; \mathcal{M}_1 \to \mathcal{X}_1, \tag{31a}$$
$$f_{2,i} \;:\; \mathcal{M}_2 \times \mathcal{X}_{1,1}^{i-1} \times \mathcal{S}^i \to \mathcal{X}_2, \tag{31b}$$

where $i \in [1:n]$, such that the average probability of error does not exceed ϵ and

$$x_{1,i} = f_{1,i}(m_1), \tag{32a}$$
$$x_{2,i} = f_{2,i}(m_2, x_{1,1}^{i-1}, s^i). \tag{32b}$$

We refer to this network as an SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^i).$$

We shall see that—as when both encoders crib and both obtain SI—this network too is single-user-like. In fact, the parallel to the single-user channel is even stronger: Define $\mathcal{T} = \left[1 : |\mathcal{X}_2^{\mathcal{S}}|\right]$, index the elements of $\mathcal{X}_2^{\mathcal{S}}$ by $t \in \mathcal{T}$, and let $g(t, \cdot)$ denote the Shannon Strategy indexed by t. The capacity of this network is achieved by applying the Block-Markov coding scheme of [21, Sec. V, Situation 2] to the (stateless) MAC with input alphabets $\mathcal{X}_1$ and $\mathcal{T}$, output alphabet $\mathcal{Y}$, and transition law

$$\widehat{W}(y|x_1, t) = \sum_{s \in \mathcal{S}} p_S(s)\, W(y|x_1, g(t, s), s).$$

Such a simple scheme is optimal for this network because in this network the encoder whose output is cribbed is not furnished with SI and hence need not use Shannon Strategies. Consequently, the situation where an encoder uses a Shannon Strategy $h(\cdot)$ but only $h(S)$ is cribbed does not arise.

Theorem 18: The capacity region of the SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, s^i)$$

is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \le H(X_1|U) \tag{33a}$$
$$R_2 \le I(T; Y|X_1, U) \tag{33b}$$
$$R_1 + R_2 \le I(X_1, T; Y) \tag{33c}$$

for some random variables U and T; a function $g\colon \mathcal{T} \times \mathcal{S} \to \mathcal{X}_2$; and a joint PMF of the form

$$\begin{aligned} p(u, t, x_1, x_2, s, y) &= p(u)\, p(x_1|u)\, p(t|u)\, p(s) \\ &\quad \times \mathbb{1}_{\{x_2 = g(t,s)\}} W(y|x_1, x_2, s). \end{aligned} \tag{34}$$

Proof: See Section A-K. □

IV. Conferencing Encoders

This section discusses the MAC with conferencing encoders, which sequentially exchange information via noise-free bit pipes of given capacities. Throughout, we denote by $C_{1,2}$ the capacity of the bit pipe from Encoder 1 to Encoder 2 and by $C_{2,1}$ that of the bit pipe from Encoder 2 to Encoder 1. Section IV-A studies the MAC with conferencing encoders under the assumption that at least one encoder cribs. It discusses strictly-causal SI to the cribbing encoder(s) and feedback to both encoders. In Section IV-B cribbing is not allowed and we discuss feedback on the MAC with conferencing encoders.

A. Cribbing

We assume that either both encoders crib (Section IV-A1) or only one (Section IV-A2).

1) Both Encoders Crib: Consider an SD-MAC where both encoders conference and crib. Recall that for $k \in \{1, 2\}$ we defined $\underline{k}$ to be the element of $\{1, 2\}$ that is not k.

Definition 2: For any two sets $\mathcal{M}_1$, $\mathcal{M}_2$ and any positive integer $n \in \mathbb{N}$, an $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code for the MAC with a cribbing Encoder 1 that conferences at rate $C_{1,2}$ and a cribbing Encoder 2 that conferences at rate $C_{2,1}$ consists of four sequences of encoder mappings

$$h_{k,i} \;:\; \mathcal{M}_k \times \mathcal{X}_{\underline{k},1}^{i-1} \times \mathcal{G}_{\underline{k},1}^{i-1} \to \mathcal{G}_{k,i}, \tag{35a}$$
$$f_{k,i} \;:\; \mathcal{M}_k \times \mathcal{X}_{\underline{k},1}^{i-1} \times \mathcal{G}_{\underline{k},1}^{i} \to \mathcal{X}_k, \tag{35b}$$

where $k \in \{1, 2\}$, $i \in [1:n]$, and a decoding mapping

$$\phi\colon \mathcal{Y}^n \to \mathcal{M}_1 \times \mathcal{M}_2 \tag{36}$$

such that the alphabets of the conferred information packets satisfy

$$\sum_{i=1}^{n} \log_2 |\mathcal{G}_{k,i}| \le n C_{k,\underline{k}}, \tag{37}$$

and such that the average probability of error P_e does not exceed ϵ, where

$$P_e = \sum_{\substack{(m_1,m_2) \in \mathcal{M}_1 \times \mathcal{M}_1, \\ y^n \notin \phi^{-1}(m_1,m_2)}} \frac{\prod_{i=1}^{n} W(y_i|x_{1,i}, x_{2,i})}{|\mathcal{M}_1|\, |\mathcal{M}_2|}, \tag{38}$$

$\phi^{-1}(m_1, m_2) \subset \mathcal{Y}^n$ is the decoding set of the message pair (m_1, m_2), and

$$g_{k,i} = h_{k,i}(m_k, x_{\underline{k},1}^{i-1}, g_{\underline{k},1}^{i-1}), \tag{39a}$$
$$x_{k,i} = f_{k,i}(m_k, x_{\underline{k},1}^{i-1}, g_{\underline{k},1}^{i}). \tag{39b}$$

The rate pair (R_1, R_2) of the code is

$$R_1 = \frac{1}{n} \log |\mathcal{M}_1|, \quad R_2 = \frac{1}{n} \log |\mathcal{M}_2|. \tag{40}$$

We refer to this network as a MAC $W(y|x_1, x_2)$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, g_{2,1}^{i}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i}).$$

If, in addition, strictly-causal SI and feedback are furnished to the cribbing encoders, then an $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code consists of four sequences of encoder mappings

$$h_{k,i} \;:\; \mathcal{M}_k \times \mathcal{X}_{\underline{k},1}^{i-1} \times \mathcal{G}_{\underline{k},1}^{i-1} \times \mathcal{S}^{i-1} \times \mathcal{Y}^{i-1} \to \mathcal{G}_{k,i}, \tag{41a}$$
$$f_{k,i} \;:\; \mathcal{M}_k \times \mathcal{X}_{\underline{k},1}^{i-1} \times \mathcal{G}_{\underline{k},1}^{i} \times \mathcal{S}^{i-1} \times \mathcal{Y}^{i-1} \to \mathcal{X}_k, \tag{41b}$$

where $k \in \{1, 2\}$, $i \in [1 : n]$, such that the average probability of error P_e does not exceed ϵ, where

$$P_e = \sum_{\substack{(m_1,m_2)\in\mathcal{M}_1\times\mathcal{M}_2,\\ s^n\in\mathcal{S}^n,\, y^n\notin\phi^{-1}(m_1,m_2)}} \frac{\prod_{i=1}^n p(s_i)\, W(y_i|x_{1,i}, x_{2,i}, s_i)}{|\mathcal{M}_1|\,|\mathcal{M}_2|}, \quad (42)$$

$\phi^{-1}(m_1, m_2) \subset \mathcal{Y}^n$ is the decoding set of the message pair (m_1, m_2), and

$$g_{k,i} = h_{k,i}(m_k, x_{\underline{k},1}^{i-1}, g_{\underline{k},1}^{i-1}, s^{i-1}, y^{i-1}), \quad (43a)$$
$$x_{k,i} = f_{k,i}(m_k, x_{\underline{k},1}^{i-1}, g_{\underline{k},1}^{i}, s^{i-1}, y^{i-1}). \quad (43b)$$

We refer to this network as an SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, g_{2,1}^{i}, s^{i-1}, y^{i-1}),\ x_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i}, s^{i-1}, y^{i-1}).$$

Theorem 19: The capacity region of the MAC $W(y|x_1, x_2)$ with encoders

$$x_{1,i} = x_{1,i}(m_1, x_{2,1}^{i-1}, g_{2,1}^{i}), \quad x_{2,i} = x_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i})$$

is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \le H(X_1|U) + C_{1,2} \quad (44a)$$
$$R_2 \le H(X_2|U) + C_{2,1} \quad (44b)$$
$$R_1 + R_2 \le I(X_1, X_2; Y) \quad (44c)$$

for some joint PMF of the form

$$p(u, x_1, x_2, y) = p(u)\, p(x_1|u)\, p(x_2|u)\, W(y|x_1, x_2). \quad (45)$$

It is achievable also if the conference is restricted to take place before transmission begins. Moreover, strictly-causal SI and/or feedback do not increase the capacity region, i.e., the capacity region of the SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, x_{2,1}^{i-1}, g_{2,1}^{i}, s^{i-1}, y^{i-1}),\ x_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i}, s^{i-1}, y^{i-1})$$

is the set of rate pairs (R_1, R_2) satisfying (44) *for some joint PMF of the form* (45).

Proof: See Section B-A. □

That conferencing prior to transmission is optimal in the absence of cribbing, SI, and feedback was shown in [19]. It is perhaps surprising that this is also optimal for the network of Theorem 19, where during the transmission the encoders are presented with additional information (such as cribbed inputs, SI, and feedback).

2) One Encoder Cribs: Consider an SD-MAC with conferencing encoders where only Encoder 2 cribs. An $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code for this network consists of four sequences of encoder mappings

$$h_{1,i} : \mathcal{M}_1 \times \mathcal{G}_{2,1}^{i-1} \to \mathcal{G}_{1,i}, \quad (46a)$$
$$h_{2,i} : \mathcal{M}_2 \times \mathcal{X}_{1,1}^{i-1} \times \mathcal{G}_{1,1}^{i-1} \to \mathcal{G}_{2,i}, \quad (46b)$$
$$f_{1,i} : \mathcal{M}_1 \times \mathcal{G}_{2,1}^{i} \to \mathcal{X}_1, \quad (46c)$$
$$f_{2,i} : \mathcal{M}_2 \times \mathcal{X}_{1,1}^{i-1} \times \mathcal{G}_{1,1}^{i} \to \mathcal{X}_2, \quad (46d)$$

where $i \in [1 : n]$, such that the average probability of error does not exceed ϵ and

$$g_{1,i} = h_{1,i}(m_1, g_{2,1}^{i-1}), \quad (47a)$$
$$g_{2,i} = h_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i-1}), \quad (47b)$$
$$x_{1,i} = f_{1,i}(m_1, g_{2,1}^{i}), \quad (47c)$$
$$x_{2,i} = f_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i}). \quad (47d)$$

We refer to this network as a MAC $W(y|x_1, x_2)$ with encoders

$$x_{1,i}(m_1, g_{2,1}^{i}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i}).$$

If, in addition, Encoder 2 observes the state sequence strictly-causally and feedback is furnished to both encoders, then an $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code consists of four sequences of encoder mappings

$$h_{1,i} : \mathcal{M}_1 \times \mathcal{G}_{2,1}^{i-1} \times \mathcal{Y}^{i-1} \to \mathcal{G}_{1,i}, \quad (48a)$$
$$h_{2,i} : \mathcal{M}_2 \times \mathcal{X}_{1,1}^{i-1} \times \mathcal{G}_{1,1}^{i-1} \times \mathcal{S}^{i-1} \times \mathcal{Y}^{i-1} \to \mathcal{G}_{2,i}, \quad (48b)$$
$$f_{1,i} : \mathcal{M}_1 \times \mathcal{G}_{2,1}^{i} \times \mathcal{Y}^{i-1} \to \mathcal{X}_1, \quad (48c)$$
$$f_{2,i} : \mathcal{M}_2 \times \mathcal{X}_{1,1}^{i-1} \times \mathcal{G}_{1,1}^{i} \times \mathcal{S}^{i-1} \times \mathcal{Y}^{i-1} \to \mathcal{X}_2, \quad (48d)$$

where $i \in [1 : n]$, such that the average probability of error does not exceed ϵ and

$$g_{1,i} = h_{1,i}(m_1, g_{2,1}^{i-1}, y^{i-1}), \quad (49a)$$
$$g_{2,i} = h_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i-1}, s^{i-1}, y^{i-1}), \quad (49b)$$
$$x_{1,i} = f_{1,i}(m_1, g_{2,1}^{i}, y^{i-1}), \quad (49c)$$
$$x_{2,i} = f_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i}, s^{i-1}, y^{i-1}). \quad (49d)$$

We refer to this network as an SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, g_{2,1}^{i}, y^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i}, s^{i-1}, y^{i-1}).$$

Theorem 20: The capacity region of the MAC $W(y|x_1, x_2)$ with encoders

$$x_{1,i}(m_1, g_{2,1}^{i}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i})$$

is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \le H(X_1|U) + C_{1,2} \quad (50a)$$
$$R_2 \le I(X_2; Y|X_1, U) + C_{2,1} \quad (50b)$$
$$R_1 + R_2 \le I(X_1, X_2; Y) \quad (50c)$$

for some joint PMF of the form

$$p(u, x_1, x_2, y) = p(u)\, p(x_1|u)\, p(x_2|u)\, W(y|x_1, x_2). \quad (51)$$

It is achievable also if the conference is restricted to take place before transmission begins. Moreover, strictly-causal SI to Encoder 2 and feedback to both encoders do not increase the capacity region, i.e., the capacity region of the SD-MAC $W(y|x_1, x_2, s)$ with encoders

$$x_{1,i}(m_1, g_{2,1}^{i}, y^{i-1}), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^{i}, s^{i-1}, y^{i-1})$$

is the set of rate pairs (R_1, R_2) satisfying (50) *for some joint PMF of the form* (51).

Proof: See Section B-B. □

Suppose now that cribbing is not allowed but feedback is available to both encoders. An $(n, \mathcal{M}_1, \mathcal{M}_2, \epsilon)$ code for the MAC with conferencing encoders and feedback is defined as in Definition 2 except that it consists of four sequences of encoder mappings

$$h_{k,i} \ : \ \mathcal{M}_k \times \mathcal{G}_{\underline{k},1}^{i-1} \times \mathcal{Y}^{i-1} \to \mathcal{G}_{k,i}, \tag{52}$$

$$f_{k,i} \ : \ \mathcal{M}_k \times \mathcal{G}_{\underline{k},1}^{i} \times \mathcal{Y}^{i-1} \to \mathcal{X}_k, \tag{53}$$

where $k \in \{1, 2\}$, $i \in [1 : n]$, such that the average probability of error does not exceed ϵ and

$$g_{k,i} = h_{k,i}(m_k, g_{\underline{k},1}^{i-1}, y^{i-1}), \tag{54a}$$

$$x_{1,i} = f_{1,i}(m_k, g_{\underline{k},1}^{i}, y^{i-1}). \tag{54b}$$

We refer to this network as a MAC $W(y|x_1, x_2)$ with encoders

$$x_{1,i}(m_1, g_{2,1}^i, y^{i-1}), \quad x_{2,i}(m_2, g_{1,1}^i, y^{i-1}).$$

Since cribbing cannot hurt we can use Theorem 20 to obtain the following outer bound:

Corollary 21: The capacity region of the MAC $W(y|x_1, x_2)$ with encoders

$$x_{1,i}(m_1, g_{2,1}^i, y^{i-1}), \quad x_{2,i}(m_2, g_{1,1}^i, y^{i-1})$$

is contained in that of the same MAC with encoders

$$x_{1,i}(m_1, g_{2,1}^i), \quad x_{2,i}(m_2, x_{1,1}^{i-1}, g_{1,1}^i).$$

This outer bound is tight and conferencing prior to transmission is optimal if Encoder 2 can compute the output of Encoder 1 based on its own output and the channel output.

Proof: The outer bound follows from Theorem 20 and the fact that cribbing cannot hurt. If Encoder 2 can compute the output of Encoder 1 based on its own output and the channel output, then feedback allows Encoder 2 to crib. But since feedback does not increase capacity if one encoder cribs (Theorem 20), the feedback capacity of such a MAC is that of Theorem 20. □

B. Feedback and Conferencing on MACs in the Class $\mathcal{D}_2$

Consider the MAC $W(y|x_1, x_2)$ with feedback and conferencing, i.e., with encoders

$$x_{1,i}(m_1, g_{2,1}^i, y^{i-1}), \quad x_{2,i}(m_2, g_{1,1}^i, y^{i-1}).$$

In this section, we strengthen the result of Corollary 21: we derive tighter outer bounds on the capacity region (Theorem 22) and exhibit a larger class of MACs for which conferencing prior to transmission is optimal (Corollary 23).

By choosing to hold their conference before transmission begins, conferencing encoders can establish a common message by exchanging fixed bit-portions of their private messages. (The common message can comprise $nC_{1,2}$ bits of Encoder 1's private message M_1 and $nC_{2,1}$ bits of M_2.) This is also true in the presence of feedback. Hence, the feedback capacity of a MAC with conferencing encoders contains the feedback capacity of the MAC with a common message of rate $C_{1,2}+C_{2,1}$.[12] Since it is in general not known whether conferencing before transmission begins is optimal, it is unknown whether this inclusion can be strict. Moreover, this bound is not explicit because the feedback capacity of a MAC with a common message is unknown. But any inner bound to it would also yield an inner bound on the feedback capacity of the MAC with conferencing.

An interesting question is whether this also holds for outer bouds: Is every (known) outer bound on the feedback capacity of the MAC with a common message also an outer bound on the feedback capacity of the MAC with conferencing? Here, we answer this question in the affirmative for the different versions of the Dependence-Balance outer bound. In [8, Th. 3] Hekstra and Willems established Dependence-Balance outer bounds on the feedback capacity of the MAC (without common message or conferencing). The different versions can be readily extended to the setting with a common message (in [13] this is done for the vanilla-version without Adaptive Parallel Channel Extension). If we assume that the common message comprises $nC_{1,2}$ bits of Encoder 1's private message M_1 and $nC_{2,1}$ bits of M_2, then—as we next show—each of these versions also outer-bounds the feedback capacity of the MAC with conferencing:

Theorem 22: The capacity region of the MAC $W(y|x_1, x_2)$ with encoders

$$x_{1,i}(m_1, g_{2,1}^i, y^{i-1}), \quad x_{2,i}(m_2, g_{1,1}^i, y^{i-1})$$

is contained in the set of rate pairs (R_1, R_2) satisfying

$$R_1 \le I(X_1; Y, Z|X_2, U) + C_{1,2} \tag{55a}$$

$$R_1 \le I(X_1; Y|X_2) + C_{1,2} \tag{55b}$$

$$R_2 \le I(X_2; Y, Z|X_1, U) + C_{2,1} \tag{55c}$$

$$R_2 \le I(X_2; Y|X_1) + C_{2,1} \tag{55d}$$

$$R_1 + R_2 \le I(X_1, X_2; Y, Z|U) + C_{1,2} + C_{2,1} \tag{55e}$$

$$R_1 + R_2 \le I(X_1, X_2; Y) \tag{55f}$$

for some joint PMF having the form

$$p(u, x_1, x_2, y, z) = p(u, x_1, x_2)\, W(y|x_1, x_2)\, p(z|x_1, x_2, y, u) \tag{56}$$

and satisfying the Dependence-Balance constraint

$$0 \le I(X_1; X_2|Y, Z, U) - I(X_1; X_2|U). \tag{57}$$

Proof: See Section B-C. □

Remark 6: Hekstra and Willems refer to the random variable Z in Theorem 22 as an Adaptive Parallel Channel Extension *[8]. The term alludes to the fact that* (55a), (55c), (55e)*, and* (55f) *would still hold if the encoders were additionally furnished with feedback from a channel with output Z, i.e., if they were to observe Z strictly-causally. (A channel with output Z is adaptive in the sense that its transition law may depend also on the auxiliary random variable U.)*

One possibility is to choose (nonadaptive) Parallel Channel Extensions that allow for cribbing: If $Z = (X_1, X_2)$, then observing Z strictly-causally allows each encoder to crib, and (55a) *amounts to* (44a) *while* (55c) *amounts to* (44b)*. Similarly, if $Z = X_1$, then observing Z strictly-causally allows*

[12] More precisely, the common message is of rate $\min\{R_1, C_{1,2}\} + \min\{R_2, C_{2,1}\}$.

Encoder 2 to crib, and (55a) *amounts to* (50a) *while* (55c) *amounts to* (50b).

We have seen that (in the presence of feedback) conferencing is at least as good as a common message. Suppose now the MAC is such that a version of the Dependence-Balance bound is tight for its feedback capacity with a common message. Then the theorem implies that also the converse holds, i.e., conferencing is no better than a common message. Thus:

Corollary 23: Suppose the MAC $W(y|x_1, x_2)$ is such that a version of the Dependence-Balance bound is tight for its feedback capacity with a common message. Then this version is also tight for the MAC with encoders

$$x_{1,i}(m_1, g^i_{2,1}, y^{i-1}), \quad x_{2,i}(m_2, g^i_{1,1}, y^{i-1}),$$

and conferencing prior to transmission is optimal.

We next exhibit a class of such MACs: the class $\mathcal{D}_2$ for which Hekstra and Willems [8] exhibited the existence of an Adaptive Parallel Channel Extension for which the Dependence-Balance outer bound coincides with the Cover-Leung inner bound.

Definition 3: A MAC is said to be in the class $\mathcal{D}_2$ if there exist a finite alphabet $\mathcal{A}$ and two functions

$$f_k : \mathcal{Y} \times \mathcal{X}_k \to \mathcal{A},\ k \in \{1, 2\},$$

such that $W(y|x_1, x_2) = 0$ whenever $f_1(y, x_1) \neq f_2(y, x_2)$ and such that for $A \triangleq f_1(Y, X_1)$ (or, equivalently, $A \triangleq f_2(Y, X_2)$) $I(X_1; X_2|Y, A) = 0$ whenever $I(X_1; X_2) = 0$.

The following may offer some intuition on why for every MAC in the class $\mathcal{D}_2$ there exists a version of the Dependence-Balance bound that coincides with the Cover-Leung inner bound: If the MAC is in the class $\mathcal{D}_2$ and feedback is presented to the encoders, then the Parallel Channel Extension $Z = A$ is physically present because each encoder can compute A from its own output and the channel output. Consequently, the assumption that each encoder observes Z strictly-causally does not make the Dependence-Balance bound loose. If, moreover, $I(X_1; X_2|Y, A) = 0$ whenever $I(X_1; X_2) = 0$, then one can show that Condition 2 holds, i.e., the encoder's messages are conditionally independent given the past channel outputs and the past outputs of the Parallel Channel Extension, which are available to both encoders. Since the encoders are presented with feedback, Condition 1 trivially holds. This suggests that the Cover-Leung inner bound, which is based on Block-Markov coding, is tight.

As the next remark shows, the MACs that we encountered in Corollaries 6, 10, and 21 are in the class $\mathcal{D}_2$:

Remark 7 [8]: If at least one encoder, say Encoder 2, can compute the other encoder's output from its own output and the channel output, then the MAC is in the class $\mathcal{D}_2$ (choose $A = X_1$).

Suppose now that the MAC is in the class $\mathcal{D}_2$. One can readily extend the argument of Hekstra and Willems to the setting with a common message and show that also in this setting there exists a tight version of the Dependence-Balance bound that coincides with the Cover-Leung inner bound. Corollary 23 thus implies:

Theorem 24: Suppose the MAC $W(y|x_1, x_2)$ is in the class $\mathcal{D}_2$. Then its capacity region with encoders

$$x_{1,i}(m_1, g^i_{2,1}, y^{i-1}), \quad x_{2,i}(m_2, g^i_{1,1}, y^{i-1})$$

is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(X_1; Y|X_2, U) + C_{1,2} \tag{58}$$

$$R_2 \leq I(X_2; Y|X_1, U) + C_{2,1} \tag{59}$$

$$R_1 + R_2 \leq I(X_1, X_2; Y) \tag{60}$$

for some joint PMF of the form

$$p(u, x_1, x_2, y) = p(u)\, p(x_1|u)\, p(x_2|u)\, W(y|x_1, x_2). \tag{61}$$

Moreover, conferencing prior to transmission is optimal.

Proof: See Section B-D. □

Remark 8: If Encoder 2 can compute Encoder 1's output from its own output and the channel output, then feedback allows it to crib, and the feedback capacity of Theorem 24 is that of Theorem 20 with a cribbing Encoder 2. If both encoders can compute the other encoder's output from their own output and the channel output, then feedback allows them both to crib and the feedback capacity of Theorem 24 is that of Theorem 19 with two cribbing encoders.[13]

In the presence of feedback and conferencing Conditions 1 and 2 are again key: all the networks with conferencing and feedback that we solved satisfy these conditions.

V. Conclusion

If at least one encoder cribs, then the SD-MAC behaves less like a MAC and more like a single-user channel: providing output feedback to both encoders and strictly-causal SI to the cribbing encoder(s) does not increase capacity, and causal SI to the cribbing encoder(s) is optimally utilized using Shannon Strategies. In asymmetric communication scenarios the SD-MAC with cribbing need not behave single-user-like. For example, the capacity of the SD-MAC with one cribbing encoder typically increases if strictly-causal SI is presented to the encoder that does not crib. Moreover, Shannon Strategies need not be optimal if both encoders crib but the causal SI is provided to only one encoder. It remains an open problem to characterize the capacity of the SD-MAC for scenarios where it does not behave like a single-user channel. For the SD-MAC with two cribbing encoders, we have argued that the following two conditions are pivotal to the single-user-like behavior:

Condition 1: The Time-i channel output and each encoder's message are conditionally independent given the information available to the other encoder in forming it's Time-$(i+1)$ channel input.

Condition 2: At every time i, the information that is available strictly-causally only to Encoder 1 and the information that is available strictly-causally only to Encoder 2 are conditionally independent given the information that is available strictly-causally to both encoders.

[13] In the absence of conferencing, i.e., when $C_{1,2} = C_{2,1} = 0$, we made similar observations in Footnotes 6 and 7.

As to the feedback capacity when the encoders can conference, we characterized the capacity in two cases: when the channel is in the class $\mathcal{D}_2$ and when at least one encoder can crib. In both cases there is no loss of optimality in having the encoders hold the conference prior to transmission. Incidentally, in both cases Conditions 1 and 2 hold.

Appendix A
Proofs Related to Cribbing

The following lemma is well-known. We prove it for the sake of completeness.

Lemma 25: Two random variables X and Y are conditionally independent given Z if and only if the joint PMF of (X, Y, Z) is of the form

$$p(x, y, z) = g_1(x, z) g_2(y, z). \tag{62}$$

Proof: To see that (62) is necessary, assume the random variables X, Z, and Y form a Markov chain in that order. Then, their joint PMF can be written as

$$p(x, y, z) = p(x, z)\, p(y|x, z) = p(x, z)\, p(y|z) \tag{63}$$

and is therefore of the form (62).

We next argue that (62) is sufficient. Indeed, if the joint PMF of the random variables X, Y, and Z is of the form (62), then the conditional PMF of X given Y and Z is

$$\begin{aligned} p(x|y, z) &= \frac{p(x, y, z)}{p(y, z)} &(64)\\ &= \frac{g_1(x, z) g_2(y, z)}{\sum_{\tilde{x}} g_1(\tilde{x}, z) g_2(y, z)} &(65)\\ &= \frac{g_1(x, z)}{\sum_{\tilde{x}} g_1(\tilde{x}, z)} &(66)\\ &= \frac{\sum_y g_1(x, z) g_2(y, z)}{\sum_{\tilde{x}, y} g_1(\tilde{x}, z) g_2(y, z)} &(67)\\ &= p(x|z). &(68)\end{aligned}$$

Hence, X and Y are conditionally independent given Z. □

A. Proof of Proposition 3

Proof: The claim is established by means of a converse, which is similar to the one in [21, Sec. V, Situation 5] but accounts for the SI. Let $Q \sim \text{Unif}[1:n]$ be independent of (M_1, M_2, S^n), and denote

$$U_i \triangleq (S^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}), \tag{69a}$$

$$U \triangleq (U_Q, Q),\ X_1 \triangleq X_{1,Q},\ X_2 \triangleq X_{2,Q}, Y \triangleq Y_Q. \tag{69b}$$

The rate of Encoder 1 satisfies

$$\begin{aligned} n(R_1 - \epsilon_n) &\overset{a)}{\leq} I(M_1; Y^n, S^n, M_2) &(70)\\ &\overset{b)}{=} I(M_1; Y^n|S^n, M_2) &(71)\\ &\overset{c)}{\leq} I(X_{1,1}^n, M_1; Y^n|S^n, M_2) &(72)\\ &\overset{d)}{=} I(X_{1,1}^n; Y^n|S^n, M_2) \\ &\quad + I(M_1; Y^n|S^n, X_{1,1}^n, M_2) &(73)\\ &\overset{e)}{=} I(X_{1,1}^n; Y^n|S^n, M_2) &(74)\\ &\overset{f)}{\leq} \sum_{i=1}^n H(X_{1,i}|S^n, X_{1,1}^{i-1}, M_2) &(75)\\ &\overset{g)}{=} \sum_{i=1}^n H(X_{1,i}|S^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, M_2, S_i^n) &(76)\\ &\overset{h)}{\leq} \sum_{i=1}^n H(X_{1,i}|U_i) &(77)\\ &\overset{i)}{=} nH(X_1|U), &(78)\end{aligned}$$

where $a)$ follows from Fano's inequality, $b)$ holds since M_1, S^n, and M_2 are independent, $c)$ is true because conditioning cannot increase entropy, $d)$ is due to the chain-rule, $e)$ holds since $X_{2,1}^n = f_{2,1}^n(M_2, X_{1,1}^{n-1}, S^{n-1})$ and since (M_1, M_2) and Y^n are conditionally independent given $(X_{1,1}^n, X_{2,1}^n, S^n)$, $f)$ is due to the non-negativity of entropy and to the chain-rule, $g)$ is true because $X_{2,1}^{i-1} = f_{2,1}^{i-1}(M_2, X_{1,1}^{i-2}, S^{i-2})$, $h)$ is a consequence of (69a) and of the fact that conditioning cannot increase entropy, and $i)$ is due to (69b). By symmetry,

$$n(R_2 - \epsilon_n) \leq nH(X_2|U). \tag{79}$$

The sum-rate satisfies

$$\begin{aligned} n(R_1 + R_2 - \epsilon_n) &\overset{a)}{\leq} I(M_1, M_2; Y^n) &(80)\\ &\overset{b)}{=} \sum_{i=1}^n I(M_1, M_2; Y_i|Y^{i-1}) &(81)\\ &\overset{c)}{=} nI(M_1, M_2; Y|Y^{Q-1}, Q) &(82)\\ &\overset{d)}{\leq} nI(X_1, X_2, M_1, M_2, Y^{Q-1}, Q; Y) &(83)\\ &\overset{e)}{=} nI(X_1, X_2; Y), &(84)\end{aligned}$$

where $a)$ follows from Fano's inequality, $b)$ is due to the chain-rule, $c)$ is a consequence of (69b), $d)$ holds since conditioning cannot increase entropy, and $e)$ is true because $\big(M_1, M_2, Y^{Q-1}, Q\big)$ and Y are conditionally independent given (X_1, X_2).

As we argue next, the joint PMF satisfies (7). Clearly, Y, (X_1, X_2), and U form a Markov chain in that order. It thus remains to verify that X_1 and X_2 are conditionally independent given U. Similarly as in [21, Eqs. (58)–(60)] we can write

$$\begin{aligned} &p(u_i, m_1, m_2) \\ &\quad = p(m_1) p(m_2) p(s^{i-1}) \\ &\qquad \times \prod_{j=1}^{i-1} p(x_{1,j}|m_1, x_{2,1}^{j-1}, s^{j-1}) p(x_{2,j}|m_2, x_{1,1}^{j-1}, s^{j-1}).\end{aligned}$$

Since the joint PMF is of the form (62), Lemma 25 implies that M_1 and M_2 are conditionally independent given U_i. As a consequence, also $X_{1,i}$ and $X_{2,i}$, which are obtained via the encoding functions (1) and hence computable from (M_1, U_i) and (M_2, U_i), are conditionally independent given U_i. Since Q is deterministic given U, it follows that X_1, U, and X_2 form a Markov chain in that order. □

B. A Proof of Theorem 5

Proof: If we denote $X = (X_1, X_2, S)$, then Lemma 1 implies that any SD-MAC of transition law $W(y|x_1, x_2, s)$ can also be viewed as an SD-MAC whose output is a deterministic function of the channel inputs, the state S, and a second channel state V, which is independent of X_1, X_2, and S. Put differently, the Time-i channel output is

$$Y_i = g(x_{1,i}, x_{2,i}, S_i, V_i), \tag{85}$$

where $g\colon \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{S} \times \mathcal{V} \to \mathcal{Y}$ is a function and $\{V_i\}$ is an IID state sequence. For i in $[1:n]$, define the super-state

$$\Xi_i \triangleq (S_i, V_i), \tag{86}$$

and note that $\{\Xi_i\}$ is also an IID state sequence. Providing feedback and revealing the state S_i strictly-causally to the cribbing encoders is no better than revealing the super-state Ξ_i strictly-causally (since S_i is computable from Ξ_i and Y_i is deterministic given $X_{1,i}$, $X_{2,i}$, and Ξ_i). According to Proposition 3, revealing the super-state strictly-causally to the cribbing encoders does not increase the capacity region. □

C. Analysis of Example 1

To see that in the absence of SI the rate pair $(R_1, R_2) = (0, 1)$ lies outside the capacity region of the MAC with a cribbing Encoder 2, note that $Y_2 = X_2 \oplus S$ is independent of (X_1, X_2). Since U, (X_1, X_2), and Y form a Markov chain in that order, Y_2 is also independent of the tuple (X_1, X_2, U). The claim now follows from (10b) of Theorem 7:

$$R_2 \le I(X_2; Y|X_1, U) \overset{a)}{=} I(X_2; X_1|X_1, U) = 0, \tag{87}$$

where $a)$ holds since Y_2 is independent of (X_1, X_2, U).

If the state S is revealed to Encoder 1 in a strictly-causal fashion, then the rate pair $(R_1, R_2) = (0, 1)$ is achievable. To see this, assume that Encoder 2 sends the uncoded binary representation of its message (and a zero at Time n) and that Encoder 1 transmits the random state sequence by choosing $X_{1,i} = S_{i-1}$. Then, the receiver can decode the message of Encoder 2 since it can recover $X_{2,i} = Y_{2,i} \oplus Y_{1,i+1}$.

D. A Proof of Theorem 8

Proof: The claim is established by means of a converse, which is similar to the one in [21, Sec. V, Situation 2] but accounts for the SI and the feedback. If we denote $X = (X_1, X_2, S)$, then Lemma 1 implies that any SD-MAC of transition law $W(y|x_1, x_2, s)$ can also be viewed as an SD-MAC whose output is a deterministic function of the channel inputs, the state S, and a second channel state V, which is independent of X_1, X_2, and S. Hence, the Time-i channel output is

$$Y_i = g(x_{1,i}, x_{2,i}, S_i, V_i), \tag{88}$$

where $g\colon \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{S} \times \mathcal{V} \to \mathcal{Y}$ is a function and $\{V_i\}$ is an IID state sequence. Let $Q \sim \mathrm{Unif}[1:n]$ and denote

$$U_i \triangleq (X_{1,1}^{i-1}, Y^{i-1}), \tag{89a}$$

$$U \triangleq (U_Q, Q),\ X_1 \triangleq X_{1,Q},\ X_2 \triangleq X_{2,Q}, \tag{89b}$$

$$Y \triangleq Y_Q,\ S \triangleq S_Q,\ V \triangleq V_Q. \tag{89c}$$

The rate of Encoder 1 satisfies

$$\begin{aligned}
n(R_1 - \epsilon_n) &\overset{a)}{\le} I(X_{1,1}^n; Y^n|S^n, V^n, M_2) \\
&\quad + I(M_1; Y^n|S^n, V^n, X_{1,1}^n, M_2) &(90)\\
&\overset{b)}{=} I(X_{1,1}^n; Y^n|S^n, V^n, M_2) &(91)\\
&\overset{c)}{\le} H(X_{1,1}^n|S^n, V^n, M_2) &(92)\\
&\overset{d)}{=} \sum_{i=1}^n H(X_{1,i}|S^n, V^n, X_{1,1}^{i-1}, M_2) &(93)\\
&\overset{e)}{=} \sum_{i=1}^n H(X_{1,i}|S^{i-1}, V^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, M_2, S_i^n, V_i^n)\\
&\overset{f)}{=} \sum_{i=1}^n H(X_{1,i}|S^{i-1}, V^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, Y^{i-1}, M_2, S_i^n, V_i^n)\\
&\overset{g)}{\le} \sum_{i=1}^n H(X_{1,i}|X_{1,1}^{i-1}, Y^{i-1}) &(94)\\
&\overset{h)}{=} nH(X_1|U), &(95)
\end{aligned}$$

where $a)$ follows from (73), $b)$ is true because $X_{2,i} = f_{2,i}(M_2, X_{1,1}^{i-1}, Y^{i-1}, S^{i-1})$ and (88) imply that $X_{2,1}^n = \tilde{f}_{2,1}^n(M_2, X_{1,1}^{n-1}, S^{n-1}, V^{n-1})$ and because Y^n is deterministic given $(X_{1,1}^n, X_{2,1}^n, S^n, V^n)$, $c)$ is due to the non-negativity of entropy, $d)$ is an application of the chain-rule, $e)$ holds since $X_{2,1}^{i-1} = \tilde{f}_{2,1}^{i-1}(M_2, X_{1,1}^{i-2}, S^{i-2}, V^{i-2})$ (by the arguments that led to $b)$), $f)$ is true because (88) guarantees that Y^{i-1} can be computed from $(X_{1,1}^{i-1}, X_{2,1}^{i-1}, S^{i-1}, V^{i-1})$ via $g(\cdot)$, $g)$ holds since conditioning cannot increase entropy, and $h)$ follows from (89).

The rate of the Encoder 2 satisfies

$$\begin{aligned}
n(R_2 - \epsilon_n) &\overset{a)}{\le} I(M_2; Y^n, M_1) &(96)\\
&\overset{b)}{=} \sum_{i=1}^n I(M_2; Y_i|Y^{i-1}, M_1) &(97)\\
&\overset{c)}{=} \sum_{i=1}^n I(M_2; Y_i|X_{1,i}, X_{1,1}^{i-1}, Y^{i-1}, M_1) &(98)\\
&\overset{d)}{\le} \sum_{i=1}^n I(X_{2,i}, M_2; Y_i|X_{1,i}, X_{1,1}^{i-1}, Y^{i-1}, M_1) &(99)\\
&\overset{e)}{\le} \sum_{i=1}^n I(X_{2,i}; Y_i|X_{1,i}, X_{1,1}^{i-1}, Y^{i-1}) &(100)\\
&\overset{f)}{=} nI(X_2; Y|X_1, U), &(101)
\end{aligned}$$

where $a)$ follows from Fano's inequality, $b)$ is due to the chain-rule and to the independence of M_1 and M_2, $c)$ is true because $X_{1,1}^i = f_{1,1}^i(M_1, Y^{i-1})$, $d)$ follows from the fact that conditioning cannot increase entropy, $e)$ holds since $(X_{1,1}^{i-1}, Y^{i-1}, M_1, M_2)$ and Y_i are conditionally independent given $(X_{1,i}, X_{2,i})$ and since conditioning cannot increase entropy, and $f)$ follows from (89).

The bound on the sum-rate follows from (84).

As we argue next, the joint PMF satisfies (11). Clearly, Y, (X_1, X_2), and U form a Markov chain in that order. It thus remains to verify that X_1 and X_2 are conditionally independent given U. Similarly as in [21, Eqs. (58)–(60)] we can write

$$\begin{aligned} p(u_i, m_1, m_2, s^{i-1}) &= \sum_{x_{2,1}^{i-1}} p(m_1)p(m_2)p(s^{i-1}) \\ &\quad \times \prod_{j=1}^{i-1}\Big[p(x_{1,j}|m_1, y^{j-1})p(x_{2,j}|m_2, x_{1,1}^{j-1}, y^{j-1}, s^{j-1}) \\ &\qquad\qquad \times W(y_j|x_{1,j}, x_{2,j}, s_j)\Big] \\ &= p(m_1)p(m_2)p(s^{i-1}) \\ &\quad \times \prod_{j=1}^{i-1}\Big[p(x_{1,j}|m_1, y^{j-1})\sum_{x_{2,j}} p(x_{2,j}|m_2, x_{1,1}^{j-1}, y^{j-1}, s^{j-1}) \\ &\qquad\qquad \times W(y_j|x_{1,j}, x_{2,j}, s_j)\Big]. \end{aligned}$$

Since the joint PMF is of the form (62), Lemma 25 implies that M_1 and (M_2, S^{i-1}) are conditionally independent given U_i. As a consequence, also $X_{1,i}$ and $X_{2,i}$, which are computable from (M_1, U_i) and (M_2, U_i, S^{i-1}) via the encoding functions, are conditionally independent given U_i. Since Q is deterministic given U, it follows that X_1, U, and X_2 form a Markov chain in that order. □

E. A Proof of Theorem 11

Proof: The proof is accomplished by means of a converse and a direct part.

1) Converse: The converse is similar to that of Section A-A. Let $Q \sim \text{Unif}[1:n]$ be independent of (M_1, M_2) and denote

$$U_i \triangleq \big(S^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}\big), \tag{102a}$$

$$T_{1,i} \triangleq (M_1, U_i), \quad T_{2,i} \triangleq (M_2, U_i), \tag{102b}$$

$$U \triangleq (U_Q, Q),\ T_1 \triangleq (T_{1,Q}, Q),\ T_2 \triangleq (T_{2,Q}, Q), \tag{102c}$$

$$X_1 \triangleq X_{1,Q},\ X_2 \triangleq X_{2,Q},\ Y \triangleq Y_Q,\ S \triangleq S_Q. \tag{102d}$$

The rate of Encoder 1 satisfies

$$\begin{aligned} n(R_1 - \epsilon_n) &\overset{a)}{\leq} \sum_{i=1}^{n} H(X_{1,i}|S_i, S^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, S_{i+1}^{n}, M_2) && (103) \\ &\overset{b)}{=} \sum_{i=1}^{n} H(X_{1,i}|S_i, U_i) && (104) \\ &\overset{c)}{=} nH(X_1|S, U), && (105) \end{aligned}$$

where $a)$ follows from (76) (note that $e)$ and $g)$ also hold if causal SI is available and $X_{2,1}^{i} = f_{2,1}^{i}(M_2, X_{1,1}^{i-1}, S^{i})$), $b)$ holds because of (102a) and since conditioning cannot increase entropy, and $c)$ is due to (102). By symmetry,

$$n(R_2 - \epsilon_n) \leq nH(X_2|S, U). \tag{106}$$

The sum-rate satisfies

$$\begin{aligned} &n(R_1 + R_2 - \epsilon_n) \\ &\quad\overset{a)}{\leq} nI\big(M_1, M_2, Y^{Q-1}, S^{Q-1}, X_{1,1}^{Q-1}, X_{2,1}^{Q-1}, Q; Y\big) \\ &\quad\overset{b)}{=} nI\big(M_1, M_2, S^{Q-1}, X_{1,1}^{Q-1}, X_{2,1}^{Q-1}, Q; Y\big) && (107) \\ &\quad\overset{c)}{=} nI(T_1, T_2; Y), && (108) \end{aligned}$$

where $a)$ follows from (82) and the fact that conditioning cannot increase entropy, $b)$ holds since Y, $\big(Q, X_{1,1}^{Q-1}, X_{2,1}^{Q-1}, S^{Q-1}, M_1, M_2\big)$, and Y^{Q-1} form a Markov chain in that order, and $c)$ is due to (102).

As we argue next, the joint PMF satisfies (18). Clearly, S and (U, T_1, T_2) are independent, and (U, T_1, T_2) and Y are (by Lemma 25) conditionally independent given (X_1, X_2, S). Furthermore, the encoding scheme (15) guarantees that there exist functions $g_1\colon \mathcal{T}_1 \times \mathcal{S} \to \mathcal{X}_1$ and $g_2\colon \mathcal{T}_2 \times \mathcal{S} \to \mathcal{X}_2$ such that

$$X_1 = g_1(T_1, S), \quad X_2 = g_2(T_2, S). \tag{109}$$

The joint distribution of (U_i, M_1, M_2) can be written as

$$\begin{aligned} p(u_i, m_1, m_2) &= p(m_1)p(m_2)p(s^{i-1}) \\ &\quad \times \prod_{j=1}^{i-1} p(x_{1,j}|m_1, x_{2,1}^{j-1}, s^{j})p(x_{2,j}|m_2, x_{1,1}^{j-1}, s^{j}). \end{aligned}$$

Since the PMF is of the form (62), Lemma 25 implies that M_1 and M_2 and therefore also the auxiliary random variables $T_{1,i}$ and $T_{2,i}$ are conditionally independent given U_i. Since Q is deterministic given U, it follows that T_1, U, and T_2 form a Markov chain in that order.

2) Direct Part: We prove that combining Shannon Strategies [14] with the Block-Markov coding scheme of [21, Sec. V, Situation 5] is optimal.[14] Assume that the encoding is done in B blocks. Define $\mathcal{T}_1 = \big[1 : |\mathcal{X}_1^{\mathcal{S}}|\big]$ and $\mathcal{T}_2 = \big[1 : |\mathcal{X}_2^{\mathcal{S}}|\big]$, index the elements of $\mathcal{X}_1^{\mathcal{S}}$ by $t_1 \in \mathcal{T}_1$ and those of $\mathcal{X}_2^{\mathcal{S}}$ by $t_2 \in \mathcal{T}_2$, and define the functions $g_1\colon \mathcal{T}_1 \times \mathcal{S} \to \mathcal{X}_1$ and $g_2\colon \mathcal{T}_2 \times \mathcal{S} \to \mathcal{X}_2$ so that $g_1(t_1, \cdot)$ denotes the element of $\mathcal{X}_1^{\mathcal{S}}$ indexed by t_1 and $g_2(t_2, \cdot)$ denotes the element of $\mathcal{X}_2^{\mathcal{S}}$ indexed by t_2. Fix a sufficiently small $\epsilon > 0$ and a joint PMF

$$p(u, t_1, t_2) = p(u)\, p(t_1\,|u)\, p(t_2\,|u). \tag{110}$$

a) Codebook generation: For b in $[1 : B]$, draw $2^{n(R_1+R_2)}$ length-n sequences $\mathbf{u}$ from the PMF $\prod_{i=1}^{n} p(u_i)$. Index them $m_0 = \big(m_0^{(1)}, m_0^{(2)}\big)$, $m_0^{(1)}$ in $[1 : 2^{nR_1}]$, $m_0^{(2)}$ in $[1 : 2^{nR_2}]$. For k in $\{1, 2\}$, m_0 in $\big\{(1, 1), \ldots, (2^{nR_1}, 2^{nR_2})\big\}$, and m_k in $[1 : 2^{nR_k}]$ draw a length-n sequences $\mathbf{t}_k$ from the PMF $\prod_{i=1}^{n} p\big(t_{k,i}|u_i(m_0)\big)$ and label it $\mathbf{t}_k(m_0, m_k)$.

[14]Since the encoders crib each other's channel inputs and not the Shannon Strategies, we cannot prove the direct part by simply evaluating the capacity region of Theorem 2 for the SD-MAC into which the original channel is transformed when the Shannon Strategies are viewed as channel inputs.

b) Encoding: Split the messages m_1 of Encoder 1 and m_2 of Encoder 2 into $B-1$ blocks b in $[1:B-1]$ of equal length: $m_1 = m_{1,1}, \ldots, m_{1,B-1}$ and $m_2 = m_{2,1}, \ldots, m_{2,B-1}$. In the first block, Encoder 1 chooses the sequences $\mathbf{t}_{1,1} = \mathbf{t}_1\big((1,1), m_{1,1}\big)$, and Encoder 2 chooses $\mathbf{t}_{2,1} = \mathbf{t}_2((1,1), m_{2,1})$. For b in $[2:B]$, cribbing and strictly-causal SI allow Encoder 1 to form the estimate $\hat{m}_{2,b-1}$ of $m_{2,b-1}$ and Encoder 2 to form the estimate $\tilde{m}_{1,b-1}$ of $m_{1,b-1}$. The estimates are used to send resolution information, which each encoder individually superimposes with its next message block. Denote $\hat{m}_{0,b} = (m_{1,b-1}, \hat{m}_{2,b-1})$ and $\tilde{m}_{0,b} = (\tilde{m}_{1,b-1}, m_{2,b-1})$. Let b be an element of $[2:B-1]$. In Block b, Encoder 1 chooses the sequence $\mathbf{t}_{1,b} = \mathbf{t}_1(\hat{m}_{0,b}, m_{1,b})$, and Encoder 2 chooses $\mathbf{t}_{2,b} = \mathbf{t}_2(\tilde{m}_{0,b}, m_{2,b})$. In the last block B, only resolution information is sent, i.e., $\mathbf{t}_{1,B} = \mathbf{t}_1(\hat{m}_{0,B}, 1)$, $\mathbf{t}_{2,B} = \mathbf{t}_2(\tilde{m}_{0,B}, 1)$. Let b be an element of $[1:B]$ and i an element of $[1:n]$. At Time $(b-1)n+i$, Encoder 1 forms the channel input $x_{1,(b-1)n+i}$ by evaluating the function $g_1(\cdot,\cdot)$ for the index $[\mathbf{t}_{1,b}]_i$ and the realization $s_{(b-1)n+i}$ of the channel state $S_{(b-1)n+i}$ and likewise Encoder 2 forms $x_{2,(b-1)n+i}$ by evaluating $g_2(\cdot,\cdot)$ for $[\mathbf{t}_{2,b}]_i$ and $s_{(b-1)n+i}$, i.e., $x_{1,(b-1)n+i} = g_1\big([\mathbf{t}_{1,b}]_i, s_{(b-1)n+i}\big)$, $x_{2,(b-1)n+i} = g_2\big([\mathbf{t}_{2,b}]_i, s_{(b-1)n+i}\big)$.

c) Handling cribbed information: To enable cooperation after Block b in $[1:B-1]$, Encoder 1 chooses $\hat{m}_{2,b}$ such that

$$\begin{aligned}&\big(\mathbf{u}(\hat{m}_{0,b}), \mathbf{t}_2(\hat{m}_{0,b}, \hat{m}_{2,b}), s_{(b-1)n+1}^{bn}, x_{2,(b-1)n+1}^{bn}\big)\\ &\quad \in \mathcal{A}_\epsilon^{(n)}(U, T_2, S, X_2).\end{aligned}$$

Likewise, Encoder 2 chooses $\tilde{m}_{1,b}$ such that

$$\begin{aligned}&\big(\mathbf{u}(\tilde{m}_{0,b}), \mathbf{t}_1(\tilde{m}_{0,b}, \tilde{m}_{1,b}), s_{(b-1)n+1}^{bn}, x_{1,(b-1)n+1}^{bn}\big)\\ &\quad \in \mathcal{A}_\epsilon^{(n)}(U, T_1, S, X_1).\end{aligned}$$

Note that the estimates $\hat{m}_{0,b}$ and $\tilde{m}_{0,b}$ are formed after Block $b-1$ and that the previous channel inputs $x_{2,(b-1)n+1}^{bn}$ and $x_{1,(b-1)n+1}^{bn}$ are cribbed. Naturally, $\hat{m}_{0,1} = \tilde{m}_{0,1} = m_{0,1} = (1,1)$.

d) Decoding: The receiver retrieves the transmitted information through backward decoding, i.e., it waits until the last block B was transmitted and then looks for $\check{m}_{0,B}$ such that

$$\begin{aligned}&\big(\mathbf{u}(\check{m}_{0,B}), \mathbf{t}_1(\check{m}_{0,B}, 1), \mathbf{t}_2(\check{m}_{0,B}, 1), y_{(B-1)n+1}^{Bn}\big)\\ &\quad \in \mathcal{A}_\epsilon^{(n)}(U, T_1, T_2, Y).\end{aligned}$$

Fix b in $[2:B-1]$ and assume that the decoder has already found

$$\check{m}_{0,B}, (\check{m}_{0,B-1}, \check{m}_{1,B-1}, \check{m}_{2,B-1}), \ldots, (\check{m}_{0,b+1}, \check{m}_{1,b+1}, \check{m}_{2,b+1}).$$

To decode Block b, the receiver first sets $\check{m}_{1,b} = \check{m}_{0,b+1}^{(1)}$, $\check{m}_{2,b} = \check{m}_{0,b+1}^{(2)}$ and then looks for $\check{m}_{0,b}$ such that

$$\begin{aligned}&\big(\mathbf{u}(\check{m}_{0,b}), \mathbf{t}_1(\check{m}_{0,b}, \check{m}_{1,b}), \mathbf{t}_2(\check{m}_{0,b}, \check{m}_{2,b}), y_{(b-1)n+1}^{bn}\big)\\ &\quad \in \mathcal{A}_\epsilon^{(n)}(U, T_1, T_2, Y).\end{aligned}$$

With the knowledge of $\check{m}_{0,b}$, the information in Block $b-1$ can be decoded next. The procedure stops after Block 2 since $m_{0,1} = (1,1)$.

e) Analysis of the error probability: The error event is

$$\mathcal{E} = \bigcup_{b=1}^{B-1} \big(\{\check{M}_{1,b} \neq M_{1,b}\} \cup \{\check{M}_{2,b} \neq M_{2,b}\}\big). \tag{111}$$

Define the critical events

$$\begin{aligned}\mathcal{E}_{b,m_1}^1 &= \Big\{\big(\mathbf{u}(M_{0,b}), \mathbf{t}_1(M_{0,b}, m_1), S_{(b-1)n+1}^{bn}, X_{1,(b-1)n+1}^{bn}\big)\\ &\qquad \in \mathcal{A}_\epsilon^{(n)}(U, T_1, S, X_1)\Big\},\\ \mathcal{E}_{b,m_2}^2 &= \Big\{\big(\mathbf{u}(M_{0,b}), \mathbf{t}_2(M_{0,b}, m_2), S_{(b-1)n+1}^{bn}, X_{2,(b-1)n+1}^{bn}\big)\\ &\qquad \in \mathcal{A}_\epsilon^{(n)}(U, T_2, S, X_2)\Big\},\\ \mathcal{E}_{b,m_0}^0 &= \Big\{\big(\mathbf{u}(m_0), \mathbf{t}_1(m_0, M_{1,b}), \mathbf{t}_2(m_0, M_{2,b}), Y_{(b-1)n+1}^{bn}\big)\\ &\qquad \in \mathcal{A}_\epsilon^{(n)}(U, T_1, T_2, Y)\Big\},\end{aligned}$$

and note that

$$\begin{aligned}\mathcal{E} \subseteq \bigcup_{b=1}^{B-1} &\Bigg(\mathcal{E}_{b,M_{1,b}}^{1c} \cup \mathcal{E}_{b,M_{2,b}}^{2c} \cup \bigcup_{m_1 \neq M_{1,b}} \mathcal{E}_{b,m_1}^1\\ &\cup \bigcup_{m_2 \neq M_{2,b}} \mathcal{E}_{b,m_2}^2\Bigg) \cup \bigcup_{b=2}^{B}\Bigg(\mathcal{E}_{b,M_{0,b}}^{0c} \cup \bigcup_{m_0 \neq M_{0,b}} \mathcal{E}_{b,m_0}^0\Bigg).\end{aligned} \tag{112}$$

Because of the union bound, since on average over the realization of the message pair (M_1, M_2) the probability of each critical event is independent of the transmission block, and since the codebook generation is symmetrical in the index sequences, the error probability averaged over the ensemble of codes satisfies

$$\begin{aligned}P_e \leq (B-1)\Bigg(&P\big(\mathcal{E}_{2,1}^{1c}\big) + P\big(\mathcal{E}_{2,1}^{2c}\big) + \sum_{m_1 \neq 1} P(\mathcal{E}_{2,m_1}^1)\\ &+ \sum_{m_2 \neq 1} P(\mathcal{E}_{2,m_2}^2) + P\big(\mathcal{E}_{2,(1,1)}^{0c}\big)\\ &+ \sum_{m_0 \neq (1,1)} P(\mathcal{E}_{2,m_0}^0)\Bigg),\end{aligned} \tag{113}$$

where $(M_{0,2}, M_{1,2}, M_{2,2}) = ((1,1), 1, 1)$. By the weak-typicality Lemma

$$P\big(\mathcal{E}_{2,1}^{1c}\big), P\big(\mathcal{E}_{2,1}^{2c}\big), P\big(\mathcal{E}_{2,(1,1)}^{0c}\big) \to 0 \ (n \to \infty). \tag{114}$$

The properties of weakly-typical sequences also imply

$$\begin{aligned}&\sum_{m_1 \neq 1} P\big(\mathcal{E}_{2,m_1}^1\big)\\ &\quad \leq 2^{nR_1} \sum_{\mathcal{A}_\epsilon^{(n)}(U,T_1,S,X_1)} p\big(u^n, s^n, x_{1,1}^n\big)\, p\big(t_{1,1}^n | u^n\big) &(115)\\ &\quad \leq 2^{nR_1} 2^{n(H(U,T_1,S,X_1)+\epsilon)}\\ &\qquad \times 2^{-n(H(U,S,X_1)-\epsilon)} 2^{-n(H(T_1|U)-2\epsilon)} &(116)\\ &\quad \leq 2^{-n(I(T_1;S,X_1|U)-4\epsilon-R_1)} &(117)\\ &\quad \overset{a)}{=} 2^{-n(H(X_1|S,U)-4\epsilon-R_1)}, &(118)\end{aligned}$$

$$\sum_{m_2\neq 1} P\big(\mathcal{E}^2_{2,m_2}\big)$$
$$\overset{b)}{\leq} 2^{-n(H(X_2|S,U)-4\epsilon-R_2)}, \sum_{m_0\neq(1,1)} P\big(\mathcal{E}^0_{2,m_0}\big)$$
$$\leq 2^{n(R_1+R_2)} \sum_{\mathcal{A}^{(n)}_\epsilon(U,T_1,T_2,Y)} p\big(u^n, t^n_{1,1}, t^n_{2,1}\big)\, p\big(y^n\big)$$
$$\leq 2^{n(R_1+R_2)} 2^{n(H(U,T_1,T_2,Y)+\epsilon)} 2^{-n(H(U,T_1,T_2)-\epsilon)} 2^{-n(H(Y)-\epsilon)} \tag{119}$$
$$\leq 2^{-n(I(U,T_1,T_2;Y)-3\epsilon-(R_1+R_2))} \tag{120}$$
$$\overset{c)}{=} 2^{-n(I(T_1,T_2;Y)-3\epsilon-(R_1+R_2))}, \tag{121}$$

where a) holds since (U, T_1) and S are independent and since X_1 is deterministic given (T_1, S), b) is due to symmetry, and c) holds since U, (T_1, T_2), and Y form a Markov chain in that order.

Equations (113)–(121) imply that $P_e \to 0\,(n \to \infty)$ provided that B grows sufficiently slowly with n and that the rate pair satisfies

$$R_1 < H(X_1|S, U) - 4\epsilon \tag{122}$$
$$R_2 < H(X_2|S, U) - 4\epsilon \tag{123}$$
$$R_1 + R_2 < I(T_1, T_2; Y) - 3\epsilon. \tag{124}$$

To conclude the proof, note that for k in $\{1, 2\}$

$$\frac{1}{nB}\log_2|\mathcal{M}_k| = \tfrac{B-1}{B} R_k \to R_k\ (B \to \infty). \tag{125}$$

□

F. A Proof of Theorem 12

Proof: To establish the claim, we slightly adapt the converse in Section A-E. Assume the encoders observe the channel state S causally and the channel state $\tilde{S}$ strictly-causally. For $X = (X_1, X_2, S, \tilde{S})$, Lemma 1 implies that there exists a random variable L, which is independent of X, and a function $g\colon \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{S} \times \tilde{\mathcal{S}} \times \mathcal{L} \to \mathcal{Y}$ such that

$$Y = g(X_1, X_2, S, \tilde{S}, L). \tag{126}$$

If we replace (102a) by

$$U_i \triangleq (S^{i-1}, \tilde{S}^{i-1}, L^{i-1}, X^{i-1}_{1,1}, X^{i-1}_{2,1}) \tag{127}$$

and S^{i-1}, s^{i-1}, and s^{j-1} by $(S^{i-1}, \tilde{S}^{i-1}, L^{i-1})$, $(s^{i-1}, \tilde{s}^{i-1}, l^{i-1})$, and $(s^{j-i}, \tilde{s}^{j-1}, l^{j-1})$, then the converse in Section A-E still applies. (Note that S_i, s_i, and s_j remain unchanged.) □

G. A Proof of Theorem 14

Proof: Since the proof is essentially that of the direct part in Section A-E, we only highlight the differences. Fix a sufficiently small $\epsilon > 0$, functions $g_1\colon \mathcal{T}_1 \times \mathcal{S}_1 \to \mathcal{X}_1$, $g_2\colon \mathcal{T}_2 \times \mathcal{S}_2 \to \mathcal{X}_2$, and a joint PMF

$$p(u, t_1, t_2) = p(u)\, p(t_1|u)\, p(t_2|u)\,. \tag{128}$$

1) Codebook Generation: Proceed as in Section A-E.

2) Encoding: Let b be an element of $[1 : B]$ and i an element of $[1 : n]$. At Time $(b-1)n+i$, Encoder 1 forms the channel input $x_{1,(b-1)n+i}$ by evaluating the function $g_1(\cdot,\cdot)$ for the index $[\mathbf{t}_{1,b}]_i$ and the realization $s_{1,(b-1)n+i}$ of the channel state $S_{1,(b-1)n+i}$, and likewise Encoder 2 forms $x_{2,(b-1)n+i}$ by evaluating $g_2(\cdot,\cdot)$ for the realization $s_{2,(b-1)n+i}$ of $S_{2,(b-1)n+i}$, i.e., $x_{1,(b-1)n+i} = g_1\big([\mathbf{t}_{1,b}]_i, s_{1,(b-1)n+i}\big)$, $x_{2,(b-1)n+i} = g_2\big([\mathbf{t}_{2,b}]_i, s_{2,(b-1)n+i}\big)$.

3) Handling of Cribbed Information: To enable cooperation after block b in $[1 : B-1]$, Encoder 1 chooses $\hat{m}_{2,b}$ such that

$$\big(\mathbf{u}(\hat{m}_{0,b}), \mathbf{t}_2(\hat{m}_{0,b}, \hat{m}_{2,b}), s^{bn}_{1,(b-1)n+1}, s^{bn}_{(b-1)n+1}, x^{bn}_{2,(b-1)n+1}\big) \in \mathcal{A}^{(n)}_\epsilon(U, T_2, S_1, S, X_2). \tag{129}$$

Likewise, Encoder 2 choose $\tilde{m}_{1,b}$ such that

$$\big(\mathbf{u}(\tilde{m}_{0,b}), \mathbf{t}_1(\tilde{m}_{0,b}, \tilde{m}_{1,b}), s^{bn}_{2,(b-1)n+1}, s^{bn}_{(b-1)n+1}, x^{bn}_{1,(b-1)n+1}\big) \in \mathcal{A}^{(n)}_\epsilon(U, T_1, S_2, S, X_1). \tag{130}$$

4) Decoding: Proceed as in Section A-E.

5) Analysis of the Error Probability: Let $\mathcal{E}^1_{b,m_1}$ and $\mathcal{E}^2_{b,m_2}$ denote the events

$$\mathcal{E}^1_{b,m_1} = \Big\{\big(\mathbf{u}(M_{0,b}), \mathbf{t}_1(M_{0,b}, m_1), S^{bn}_{2,(b-1)n+1}, S^{bn}_{(b-1)n+1}, X^{bn}_{1,(b-1)n+1}\big) \in \mathcal{A}^{(n)}_\epsilon(U, T_1, S_2, S, X_1)\Big\},$$
$$\mathcal{E}^2_{b,m_2} = \Big\{\big(\mathbf{u}(M_{0,b}), \mathbf{t}_2(M_{0,b}, m_2), S^{bn}_{1,(b-1)n+1}, S^{bn}_{(b-1)n+1}, X^{bn}_{2,(b-1)n+1}\big) \in \mathcal{A}^{(n)}_\epsilon(U, T_2, S_1, S, X_2)\Big\}.$$

The properties of weakly-typical sequences imply

$$\sum_{m_1\neq 1} P(\mathcal{E}^1_{2,m_1})$$
$$\leq 2^{nR_1} \sum_{\mathcal{A}^{(n)}_\epsilon(U,T_1,S_2,S,X_1)} p\big(u^n, s^n_{2,1}, s^n, x^n_{1,1}\big)\, p\big(t^n_{1,1}\,\big|u^n\big)$$
$$\leq 2^{nR_1} 2^{n(H(U,T_1,S_2,S,X_1)+\epsilon)} \times 2^{-n(H(U,S_2,S,X_1)-\epsilon)} 2^{-n(H(T_1|U)-2\epsilon)} \tag{131}$$
$$\leq 2^{-n(I(T_1;S,S_2,X_1|U)-4\epsilon-R_1)} \tag{132}$$
$$\overset{a)}{=} 2^{-n(I(T_1;X_1|S,S_2,U)-4\epsilon-R_1)}, \tag{133}$$
$$\sum_{m_2\neq 1} P(\mathcal{E}^2_{2,m_2}) \overset{b)}{\leq} 2^{-n(I(T_2;X_2|S,S_1,U)-4\epsilon-R_2)}, \tag{134}$$

where a) holds since (U, T_1) and (S, S_2) are independent, and b) is due to symmetry. With (133) and (134) at hand, the claim follows as in Section A-E. □

H. Analysis of Example 2

To see that any rate pair

$$\big\{(R_1, R_2) \in (R^+_0)^2 : R_1 + R_2 \leq 1\big\} \tag{135}$$

is achievable, let $X \sim \mathrm{Ber}(1/2)$ and $X_2 = X \oplus S_2$. Then, $Y = X$. Since Encoder 2 learns the input symbols of Encoder 1 through cribbing, X can carry one bit of information about M_1 or M_2.

In contrast, Theorem 14 allows rate pairs $(R_1, R_2) \in (\mathbb{R}_0^+)^2$ satisfying

$$R_1 \leq I(T_1; X_1|S_2, U) = H(X_1|S_2, U) \quad (136)$$

$$R_2 \leq I(T_2; X_2|U) \quad (137)$$

$$R_1 + R_2 \leq I(T_1, T_2; Y) \quad (138)$$

for some random variables U, T_1, and T_2; functions g_1: $\mathcal{T}_1 \to \mathcal{X}_1$ and g_2: $\mathcal{T}_2 \times \mathcal{S}_2 \to \mathcal{X}_2$; and a joint PMF of the form

$$\begin{aligned} &p(u, t_1, t_2, x_1, x_2, s_2, y) \\ &\quad = p(u)\, p(t_1|u)\, p(t_2|u)\, p(s_2) \\ &\qquad \times \mathbb{1}_{\{x_1 = g_1(t_1)\}} \mathbb{1}_{\{x_2 = g_2(t_2, s_2)\}} W(y|x_1, x_2, s_2). \end{aligned} \quad (139)$$

The rate of Encoder 2 can only be one if $I(T_2; X_2|U) = 1$, which in turn implies that X_2 is deterministic given T_2. But in this case

$$\begin{aligned} R_1 + R_2 &\leq I(T_1, T_2; Y) && (140)\\ &= H(X_2 \oplus S_2) - H(S_2|X_1, X_2, T_1, T_2) && (141)\\ &= 0, && (142) \end{aligned}$$

where the last equality holds since the tuple (X_1, X_2, T_1, T_2) is independent of S_2 if $X_2 = g_2(T_2)$. This proves that the inner bound of Theorem 14 is strictly smaller than the capacity region.

I. A Proof of Theorem 15

Proof: To prove the claim, we adapt the converse of Section A-E to the new setting. In definition (102) replace (102a) by

$$U_i \triangleq \big(S^{i-1}, S_{1,1}^{i-1}, S_{2,1}^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}\big). \quad (143)$$

If we replace S_i by $(S_i, S_{1,i}, S_{2,i})$ for every i in $[1:n]$, then the converse of Section A-E implies

$$\begin{aligned} n(R_1 - \epsilon_n) &\leq nH(X_1|S, S_1, S_2, U) && (144)\\ &\overset{a)}{=} nH(X_1|S_1, U), && (145)\\ n(R_2 - \epsilon_n) &\leq nH(X_2|S, S_1, S_2, U) && (146)\\ &\overset{b)}{=} nH(X_2|S_2, U), && (147)\\ n(R_1 + R_2) &\leq nI(T_1, T_2; Y), && (148) \end{aligned}$$

where $a)$ holds since X_1, (S_1, U), and (S, S_2) form a Markov chain in that order, and $b)$ follows similarly.

As we argue next, the joint PMF satisfies (24). Indeed, (S_1, S_2, S) and (U, T_1, T_2) are independent, and (U, T_1, T_2) and Y are (by Lemma 25) conditionally independent given (X_1, X_2, S_1, S_2, S). Since the encoding mappings are of the form (19), there exist functions g_1: $\mathcal{T}_1 \times \mathcal{S}_1 \to \mathcal{X}_1$ and g_2: $\mathcal{T}_2 \times \mathcal{S}_2 \to \mathcal{X}_2$ such that

$$X_1 = g_1(T_1, S_1), \quad X_2 = g_2(T_2, S_2). \quad (149)$$

The joint PMF of (U_i, M_1, M_2) can be written as

$$\begin{aligned} p(u_i, m_1, m_2) = p(m_1)p(m_2)p(s_{1,1}^{i-1}, s_{2,1}^{i-1}, s^{i-1}) \\ \times \prod_{j=1}^{i-1}\big[p(x_{1,j}|m_1, x_{2,1}^{j-1}, s_{1,1}^{j}, s^{j-1}) \\ \times p(x_{2,j}|m_2, x_{1,1}^{j-1}, s_{2,1}^{j}, s^{j-1})\big], \end{aligned} \quad (150)$$

and is therefore of the form (62). Hence, Lemma 25 implies that M_1 and M_2 and hence also the auxiliary random variables $T_{1,i}$ and $T_{2,i}$ are conditionally independent given U_i. Because Q is deterministic given U, we conclude that T_1, U, and T_2 form a Markov chain in that order. □

J. A Proof of Theorem 17

Proof: The proof has a converse and a direct part. We establish the converse for the case where feedback is available to both encoders and the direct part for the case where feedback is only available to Encoder 1.

1) Converse: Let $Q \sim$ Unif$[1:n]$ and denote

$$U_i \triangleq \big(S^{i-1}, S_{1,1}^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, Y^{i-1}\big), \quad (151a)$$

$$T_{1,i} \triangleq (M_1, U_i), \; T_{2,i} \triangleq \big(M_2, S_{2,1}^{i-1}, U_i\big), \quad (151b)$$

$$U \triangleq (U_Q, Q), \; T_1 \triangleq (T_{1,Q}, Q), \; T_2 \triangleq (T_{2,Q}, Q), \quad (151c)$$

$$X_1 \triangleq X_{1,Q}, \; X_2 \triangleq X_{2,Q}, \; Y \triangleq Y_Q, \quad (151d)$$

$$S_1 \triangleq S_{1,Q}, \; S_2 \triangleq S_{2,Q}. \quad (151e)$$

As we argue next, the joint PMF satisfies (30). It is clear that (S_1, S_2, S) and (U, T_1, T_2) are independent, and that (U, T_1, T_2) and Y are conditionally independent given (X_1, X_2, S_1, S_2, S). Moreover, the encoding functions

$$X_{1,i} = f_{1,i}(M_1, X_{2,1}^{i-1}, S_{1,1}^{i}, S^{i-1}, Y^{i-1}), \quad (152)$$

$$X_{2,i} = f_{2,i}(M_2, X_{1,1}^{i-1}, S_{2,1}^{i}, S^{i-1}, Y^{i-1}) \quad (153)$$

and definition (151) guarantee that there exist functions g_1: $\mathcal{T}_1 \times \mathcal{S}_1 \to \mathcal{X}_1$ and g_2: $\mathcal{T}_2 \times \mathcal{S}_2 \to \mathcal{X}_2$ such that

$$X_1 = g_1(T_1, S_1), \quad X_2 = g_2(T_2, S_2). \quad (154)$$

The joint PMF of $\big(U_i, M_1, M_2, S_{2,1}^{i-1}\big)$ satisfies

$$\begin{aligned} &p(u_i, m_1, m_2, s_{2,1}^{i-1}) \\ &\quad = p(m_1)p(m_2)p(s_{1,1}^{i-1}, s_{2,1}^{i-1}, s^{i-1}) \\ &\qquad \times \prod_{j=1}^{i-1}\big[p(x_{1,j}|m_1, x_{2,1}^{j-1}, s_{1,1}^{j}, s^{j-1}, y^{j-1}) \\ &\qquad\qquad \times p(x_{2,j}|m_2, x_{1,1}^{j-1}, s_{2,1}^{j}, s^{j-1}, y^{j-1}) \\ &\qquad\qquad \times W(y_j|x_{1,j}, x_{2,j}, (s_{1,j}, s_{2,j}, s_j))\big]. \end{aligned} \quad (155)$$

Since the PMF is of the form (62), Lemma 25 implies that M_1 and $\big(M_2, S_2^{i-1}\big)$ and therefore also the auxiliary random variables $T_{1,i}$ and $T_{2,i}$ are conditionally independent given U_i. Because Q is deterministic given U, we conclude that T_1, U, and T_2 form a Markov chain in that order.

The rate of Encoder 1 satisfies

$$\begin{aligned} &n(R_1 - \epsilon_n) \\ &\quad \overset{a)}{=} I\big(M_1; Y^n, X_{1,1}^n, X_{2,1}^n, M_2, S^n, S_{1,1}^n, S_{2,1}^n\big) && (156)\\ &\quad \overset{b)}{=} nI(T_1; Y, X_1, X_2, S, S_1, S_2|U, T_2) && (157)\\ &\quad \overset{c)}{=} nI(T_1; Y, X_1|X_2, S, S_1, S_2, U, T_2) && (158)\\ &\quad \overset{d)}{=} nH(X_1|X_2, S, S_1, S_2, U, T_2) && (159)\\ &\quad \overset{e)}{=} nH(X_1|S_1, U), && (160) \end{aligned}$$

where a) follows from Fano's inequality, b) is true because of the chain-rule, the independence of M_1 and M_2, and (151), (c) holds since (U, T_1, T_2) and (S, S_1, S_2) are independent and since $X_2 = g_2(T_2, S_2)$, (d) is true because (U, T_1, T_2) and Y are conditionally independent given (X_1, X_2, S, S_1, S_2) and because $X_1 = g_1(T_1, S_1)$, and (e) holds since X_1 and (X_2, S, S_2, T_2) are conditionally independent given (S_1, U).
The rate of Encoder 2 satisfies

$$
\begin{aligned}
&n(R_2 - \epsilon_n) \\
&\overset{a)}{=} I(M_2; Y^n, X_{1,1}^n, X_{2,1}^n, M_1, S^n, S_{1,1}^n) && (161)\\
&\overset{b)}{=} \sum_{i=1}^{n} I(M_2; Y_i, X_{1,i}, X_{2,i}, S_i, S_{1,i}|U_i, T_{1,i}) && (162)\\
&\overset{c)}{\leq} \sum_{i=1}^{n} I(M_2, S_{2,1}^{i-1}; Y_i, X_{1,i}, X_{2,i}, S_i, S_{1,i}|U_i, T_{1,i})\\
&\overset{d)}{=} nI(T_2; Y, X_1, X_2, S, S_1|U, T_1) && (163)\\
&\overset{e)}{=} nI(T_2; Y, X_2|X_1, S, S_1, U, T_1) && (164)\\
&\overset{f)}{=} nI(T_2; Y, X_2|X_1, S, S_1, U), && (165)
\end{aligned}
$$

where a) follows from Fano's inequality, b) is true because of the chain-rule, the independence of M_1 and M_2, and (151a)–(151b), (c) holds since conditioning cannot increase entropy, (d) is due to (151), (e) is true because (U, T_1, T_2) and (S, S_1) are independent and since $X_1 = g_1(T_1, S_1)$, and (f) holds since T_1 and (Y, X_2, T_2) are conditionally independent given (X_1, S, S_1, U). The sum-rate satisfies

$$n(R_1 + R_2 - \epsilon_n) \leq nI(T_1, T_2; Y), \tag{166}$$

where we used (82), the fact that conditioning cannot increase entropy, and (151).

2) Direct Part: Since the proof is essentially that of the direct part in Section A-E, we only highlight the differences. Fix a sufficiently small $\epsilon > 0$, functions $g_1 : \mathcal{T}_1 \times \mathcal{S}_1 \to \mathcal{X}_1$, $g_2 : \mathcal{T}_2 \times \mathcal{S}_2 \to \mathcal{X}_2$, and a joint PMF

$$p(u, t_1, t_2) = p(u)\, p(t_1|u)\, p(t_2|u)\,. \tag{167}$$

a) Codebook generation: Proceed as in Section A-E.

Encoding: Let b be an element of $[1 : B]$ and i an element of $[1 : n]$. At Time $(b-1)n + i$, Encoder 1 forms the channel input $x_{1,(b-1)n+i}$ by evaluating the function $g_1(\cdot, \cdot)$ for the index $[\mathbf{t}_{1,b}]_i$ and the realization $s_{1,(b-1)n+i}$ of the channel state $S_{1,(b-1)n+i}$, and likewise Encoder 2 forms $x_{2,(b-1)n+i}$ by evaluating the function $g_2(\cdot, \cdot)$ for $[\mathbf{t}_{2,b}]_i$ and the realization $s_{2,(b-1)n+i}$ of $S_{2,(b-1)n+i}$, i.e., $x_{1,(b-1)n+i} = g_1\big([\mathbf{t}_{1,b}]_i, s_{1,(b-1)n+i}\big)$, $x_{2,(b-1)n+i} = g_2\big([\mathbf{t}_{2,b}]_i, s_{2,(b-1)n+i}\big)$.

b) Handling cribbed information: To enable cooperation after block b in $[1 : B-1]$, Encoder 1 chooses $\hat{m}_{2,b}$ such that

$$
\begin{aligned}
&\big(\mathbf{u}(\hat{m}_{0,b}), \mathbf{t}_2(\hat{m}_{0,b}, \hat{m}_{2,b}), s_{(b-1)n+1}^{bn}, s_{1,(b-1)n+1}^{bn}, y_{(b-1)n+1}^{bn},\\
&\quad x_{1,(b-1)n+1}^{bn}, x_{2,(b-1)n+1}^{bn}\big) \in \mathcal{A}_\epsilon^{(n)}(U, T_2, S, S_1, Y, X_1, X_2).
\end{aligned}
$$

By assumption, Encoder 2 can compute $s_{1,(b-1)n+1}^{bn}$. It thus chooses $\tilde{m}_{1,b}$ such that

$$
\begin{aligned}
&\big(\mathbf{u}(\tilde{m}_{0,b}), \mathbf{t}_1(\tilde{m}_{0,b}, \tilde{m}_{1,b}), s_{1,(b-1)n+1}^{bn}, x_{1,(b-1)n+1}^{bn}\big)\\
&\quad \in \mathcal{A}_\epsilon^{(n)}(U, T_1, S_1, X_1)\,. && (168)
\end{aligned}
$$

c) Decoding: Proceed as in Section A-E.

d) Analysis of the error probability: Let $\mathcal{E}_{b,m_1}^1$ and $\mathcal{E}_{b,m_2}^2$ denote the events

$$
\begin{aligned}
\mathcal{E}_{b,m_1}^1 &= \Big\{\big(\mathbf{u}(M_{0,b}), \mathbf{t}_1(M_{0,b}, m_1), S_{1,(b-1)n+1}^{bn}, X_{1,(b-1)n+1}^{bn}\big)\\
&\qquad \in \mathcal{A}_\epsilon^{(n)}(U, T_1, S_1, X_1)\Big\}, && (169)\\
\mathcal{E}_{b,m_2}^2 &= \Big\{\big(\mathbf{u}(M_{0,b}), \mathbf{t}_2(M_{0,b}, m_2), S_{(b-1)n+1}^{bn}, S_{1,(b-1)n+1}^{bn},\\
&\qquad Y_{(b-1)n+1}^{bn}, X_{1,(b-1)n+1}^{bn}, X_{2,(b-1)n+1}^{bn}\big)\\
&\qquad \in \mathcal{A}_\epsilon^{(n)}(U, T_2, S, S_1, Y, X_1, X_2)\Big\}. && (170)
\end{aligned}
$$

The properties of weakly-typical sequences imply

$$
\begin{aligned}
&\sum_{m_1 \neq 1} P(\mathcal{E}_{2,m_1}^1)\\
&\quad \leq 2^{nR_1} \sum_{\mathcal{A}_\epsilon^{(n)}(U,T_1,S_1,X_1)} p(u^n, s_{1,1}^n, x_{1,1}^n)\, p(t_{1,1}^n|u^n) && (171)\\
&\quad \leq 2^{nR_1} 2^{n(H(U,T_1,S_1,X_1)+\epsilon)}\\
&\qquad \times 2^{-n(H(U,S_1,X_1)-\epsilon)} 2^{-n(H(T_1|U)-2\epsilon)} && (172)\\
&\quad \leq 2^{-n(I(T_1;X_1,S_1|U)-4\epsilon-R_1)} && (173)\\
&\quad \overset{(a)}{\leq} 2^{-n(H(X_1|S_1,U)-4\epsilon-R_1)}, && (174)
\end{aligned}
$$

where a) holds since S_1 is independent of (U, T_1) and since X_1 is deterministic given (T_1, S_1). Similarly, we find that

$$
\begin{aligned}
&\sum_{m_2 \neq 1} P(\mathcal{E}_{2,m_2}^1)\\
&\leq 2^{nR_2} \sum_{\mathcal{A}_\epsilon^{(n)}(U,T_2,S,S_1,Y,X_1,X_2)} \big[p(u^n, s^n, s_{1,1}^n, y^n, x_{1,1}^n, x_{2,1}^n)\, p(t_{2,1}^n|u^n)\big]\\
&\leq 2^{nR_2} 2^{n(H(U,T_2,S,S_1,Y,X_1,X_2)+\epsilon)}\\
&\quad \times 2^{-n(H(U,S,S_1,Y,X_1,X_2)-\epsilon)} 2^{-n(H(T_2|U)-2\epsilon)} && (175)\\
&\leq 2^{-n(I(T_2;Y,X_1,X_2,S,S_1|U)-4\epsilon-R_2)} && (176)\\
&\overset{a)}{=} 2^{-n(I(T_2;Y,X_2|X_1,S,S_1,U)-4\epsilon-R_2)}, && (177)
\end{aligned}
$$

where a) holds since T_2 and (X_1, S, S_1) are conditionally independent given U. With (174) and (177) at hand, we conclude as in Section A-E. □

K. A Proof of Theorem 18

Proof: The proof has a converse and a direct part.

1) Converse: The converse is similar as the one in Section A-D. Let $Q \sim \text{Unif}[1 : n]$ and denote

$$
\begin{aligned}
U_i &\triangleq \big(X_{1,1}^{i-1}, Y^{i-1}\big), && (178a)\\
T_i &\triangleq \big(M_2, S^{i-1}, U_i\big), && (178b)\\
U &\triangleq (U_Q, Q),\ \ T \triangleq (T_Q, Q), && (178c)\\
X_1 &\triangleq X_{1,Q},\ \ X_2 \triangleq X_{2,Q},\ \ Y \triangleq Y_Q,\ \ S \triangleq S_Q. && (178d)
\end{aligned}
$$

The constraint on the rate of Encoder 1 is due to (95) (note that b) and e) also hold if causal SI is available and

$X_{2,1}^i = f_{2,1}^i(M_2, X_{1,1}^{i-1}, S^i))$. The rate of Encoder 2 satisfies

$$
\begin{aligned}
&n(R_2 - \epsilon_n) \\
&\overset{a)}{\leq} \sum_{i=1}^{n} I(M_2; Y_i | X_{1,i}, X_{1,1}^{i-1}, Y^{i-1}, M_1) && (179)\\
&\overset{b)}{\leq} \sum_{i=1}^{n} I(M_2, S^{i-1}; Y_i | X_{1,i}, X_{1,1}^{i-1}, Y^{i-1}, M_1) && (180)\\
&\overset{c)}{\leq} \sum_{i=1}^{n} I(M_2, S^{i-1}; Y_i | X_{1,i}, X_{1,1}^{i-1}, Y^{i-1}) && (181)\\
&\overset{d)}{=} nI(T; Y | X_1, U), && (182)
\end{aligned}
$$

where a) follows from (98), b) holds since conditioning cannot increase entropy, c) is true because M_1, $\bigl(X_{1,i}, M_2, X_{1,1}^{i-1}, Y^{i-1}, S^{i-1}\bigr)$, and Y_i form a Markov chain in that order and since conditioning cannot increase entropy, and d) is due to (178).

The sum-rate satisfies

$$
\begin{aligned}
&n(R_1 + R_2 - \epsilon_n) \\
&\overset{a)}{\leq} nI\bigl(X_1, M_1, M_2, S^{Q-1}, X_{1,1}^{Q-1}, Y^{Q-1}, Q; Y\bigr) && (183)\\
&\overset{b)}{=} nI\bigl(X_1, M_2, S^{Q-1}, X_{1,1}^{Q-1}, Y^{Q-1}, Q; Y\bigr) && (184)\\
&\overset{c)}{=} nI(X_1, T; Y), && (185)
\end{aligned}
$$

where a) follows from (82) and the fact that conditioning cannot increase entropy, b) holds since Y, $\bigl(X_1, M_2, S^{Q-1}, X_{1,1}^{Q-1}, Y^{Q-1}\bigr)$, and M_1 form a Markov chain in that order, and c) is due to (178).

As we argue next, the joint PMF is of the form (34). It is clear that S and (U, X_1, T) are independent and that U and Y are conditionally independent given (X_1, T, S). Moreover, the encoding function (31a) and the independence of M_1, M_2, and S^{i-1} imply that T_i and $X_{1,i}$ are conditionally independent given U_i. Since Q is deterministic given U, it follows that X_1, U, and T form a Markov chain in that order. Because of (178) and (31b) there exists a function $g\colon \mathcal{T} \times \mathcal{S} \to \mathcal{X}_2$ such that

$$X_2 = g(T, S). \tag{186}$$

2) Direct Part: To prove that the rate region is also achievable, we use Shannon Strategies [14] to transform the given channel with causal SI to the cribbing encoder into a (stateless) MAC with a cribbing encoder. For the latter channel we can invoke Theorem 7.

Define $\mathcal{T} = \bigl[1 : |\mathcal{X}_2^{\mathcal{S}}|\bigr]$, index the elements of $\mathcal{X}_2^{\mathcal{S}}$ by $t \in \mathcal{T}$, and let $g(t, \cdot)$ denote the Shannon Strategy indexed by t. If Encoder 2 computes its Time-i input as $X_{2,i} = g(T_i, S_i)$, where $T_i = f_{2,i}\bigl(M_2, X_{1,1}^{i-1}\bigr)$, then we can view the SD-MAC $W(y|x_1, x_2, s)$ with causal SI to the cribbing Encoder 2 as a (stateless) MAC with a cribbing Encoder 2, input alphabets $\mathcal{X}_1$ and $\mathcal{T}$, output alphabet $\mathcal{Y}$, and transition law

$$\widehat{W}(y|x_1, t) = \sum_{s \in \mathcal{S}} p_S(s)\, W(y|x_1, g(t, s), s). \tag{187}$$

For such a network Theorem 7 implies that a rate pair (R_1, R_2) is achievable if

$$
\begin{aligned}
R_1 &\leq H(X_1|U) && (188)\\
R_2 &\leq I(T; Y|X_1, U) && (189)\\
R_1 + R_2 &\leq I(X_1, T; Y) && (190)
\end{aligned}
$$

hold for some joint PMF of the form

$$p(u, x_1, t, y) = p(u)\, p(x_1|u)\, p(t|u)\, \widehat{W}(y|x_1, t). \tag{191}$$

To conclude the proof, note that the relations $X_2 = g(T, S)$ and (187) imply (34). □

Appendix B
Proofs Related to Conferencing

A. A Proof of Theorem 19

Proof: The proof has a converse and a direct part.

1) Converse: The claim can be established with similar arguments as in Sections A-A and A-B (see [2, Proof of Th. 4.1] that applies to the case where only Encoder 2 cribs). Here, we give a proof based on the proof of Theorem 22. To this end, we view the cribbed input symbols and the SI as a Parallel Channel Extension and let $Z_i = \bigl(X_{1,i}, X_{2,i}, S_i\bigr)$. Then (272), (279), and (293) imply

$$
\begin{aligned}
R_1 &\leq I(X_1; Y, X_1, X_2, S|X_2, U) + C_{1,2} && (192)\\
&= H(X_1|X_2, U) + C_{1,2}, && (193)\\
R_2 &\leq I(X_2; Y, X_1, X_2, S|X_1, U) + C_{2,1} && (194)\\
&= H(X_2|X_1, U) + C_{2,1}, && (195)\\
R_1 + R_2 &\leq I(X_1, X_2; Y). && (196)
\end{aligned}
$$

Moreover, the Dependence-Balance constraint (300) reads

$$
\begin{aligned}
0 &\leq I(X_1; X_2|Y, X_1, X_2, S, U) - I(X_1; X_2|U) && (197)\\
&= -I(X_1; X_2|U). && (198)
\end{aligned}
$$

It is satisfied iff X_1, U, and X_2 form a Markov-chain in that order. But in this case (193) and (195) simplify to

$$
\begin{aligned}
R_1 &\leq H(X_1|U) + C_{1,2}, && (199)\\
R_2 &\leq H(X_2|U) + C_{2,1}. && (200)
\end{aligned}
$$

To conclude the proof, note that U, (X_1, X_2), and Y form a Markov chain in that order.

2) Direct Part: To see that the rate region is also achievable, assume that the encoders conference prior to transmission. If Encoder 1 sends the last $nC_{1,2}$ bits of its message M_1 during the conference and likewise Encoder 2 the last $nC_{2,1}$ bits of M_2, then all rate pairs in the capacity region of the MAC with cribbing encoders that observe private messages of rates

$$
\begin{aligned}
\tilde{R}_1 &= \max\{R_1 - C_{1,2}, 0\}, && (201a)\\
\tilde{R}_2 &= \max\{R_2 - C_{2,1}, 0\} && (201b)
\end{aligned}
$$

and a common message of rate $R_0 = \min\{R_1, C_{1,2}\} + \min\{R_2, C_{2,1}\}$ are achievable. Suppose the capacity region of the MAC with a common message and cribbing encoders contains the set of rate triples $(R_0, \tilde{R}_1, \tilde{R}_2)$ satisfying

$$\tilde{R}_1 \le H(X_1|U) \tag{202a}$$
$$\tilde{R}_2 \le H(X_2|U) \tag{202b}$$
$$R_0 + \tilde{R}_1 + \tilde{R}_2 \le I(X_1, X_2; Y) \tag{202c}$$

for some joint PMF of the form

$$p(u, x_1, x_2, y) = p(u)\, p(x_1|u)\, p(x_2|u)\, p(y|x_1, x_2). \tag{203}$$

Then, (201) implies that the set of rate pairs (R_1, R_2) satisfying

$$R_1 \le H(X_1|U) + C_{1,2} \tag{204}$$
$$R_2 \le H(X_2|U) + C_{2,1} \tag{205}$$
$$R_1 + R_2 \le I(X_1, X_2; Y) \tag{206}$$

for some joint PMF of the form

$$p(u, x_1, x_2, y) = p(u)\, p(x_1|u)\, p(x_2|u)\, p(y|x_1, x_2) \tag{207}$$

is achievable. To conclude the proof, we show that the capacity region of the MAC with a common message and cribbing encoders contains any rate triple that satisfies (202) for some joint PMF of the form (203). To this end, we use a random coding argument that combines the ideas of [15] and [21]: As in [21, Sec. V, Situation 2], the encoding is done in B blocks during each of which both encoders superimpose private message blocks over a common part. In contrast to the scheme of [21], the common part not only comprises resolution information on the previous transmission block but also information on the common message (see [15]).

Fix a sufficiently small $\epsilon > 0$ and a joint PMF

$$p(u, x_1, x_2) = p(u)\, p(x_1|u)\, p(x_2|u)\,. \tag{208}$$

a) Codebook generation: For b in $[1 : B]$, draw $2^{n(R_0+R_1+R_2)}$ length-n sequences $\mathbf{u}$ from the PMF $\prod_{i=1}^n p(u_i)$. Index them $m_{-1} = (m_{-1}^{(0)}, m_{-1}^{(1)}, m_{-1}^{(2)})$, $m_{-1}^{(0)}$ in $[1 : 2^{nR_0}]$, $m_{-1}^{(1)}$ in $[1 : 2^{nR_1}]$, $m_{-1}^{(2)}$ in $[1 : 2^{nR_2}]$. For k in $\{1, 2\}$, m_{-1} in $\{(1, 1, 1), \ldots, (2^{nR_0}, 2^{nR_1}, 2^{nR_2})\}$, and m_k in $[1 : 2^{nR_k}]$ draw a length-n codeword $\mathbf{x}_k$ from the PMF $\prod_{i=1}^n p\big(x_{k,i}|u_i(m_{-1})\big)$ and label it $\mathbf{x}_k(m_{-1}, m_k)$.

b) Encoding: Split the common message m_0, the private message m_1 of Encoder 1, and the private message m_2 of Encoder 2 into $B-1$ blocks of equal length: $m_0 = m_{0,2}, \ldots, m_{0,B}$, $m_1 = m_{1,1}, \ldots, m_{1,B-1}$, and $m_2 = m_{2,1}, \ldots, m_{2,B-1}$. In the first block, Encoder 1 chooses the codeword $\mathbf{x}_{1,1} = \mathbf{x}_1\big((1, 1, 1), m_{1,1}\big)$, and Encoder 2 chooses $\mathbf{x}_{2,1} = \mathbf{x}_2\big((1, 1, 1), m_{2,1}\big)$. For b in $[2 : B]$, cribbing allows Encoder 1 to form the estimate $\hat{m}_{2,b-1}$ of $m_{2,b-1}$ and Encoder 2 to form the estimate $\tilde{m}_{1,b-1}$ of $m_{1,b-1}$. The estimates are used to send resolution information, which each encoder individually superimposes with its next private message block. Denote $m_{-1,b} = (m_{0,b}, m_{1,b-1}, m_{2,b-1})$, $\hat{m}_{-1,b} = (m_{0,b}, m_{1,b-1}, \hat{m}_{2,b-1})$, and $\tilde{m}_{-1,b} = (m_{0,b}, \tilde{m}_{1,b-1}, m_{2,b-1})$. Let b be an element of $[2 : B-1]$. In Block b, Encoder 1 chooses the codeword $\mathbf{x}_{1,b} = \mathbf{x}_1(\hat{m}_{-1,b}, m_{1,b})$, and Encoder 2 chooses $\mathbf{x}_{2,b} = \mathbf{x}_2(\tilde{m}_{-1,b}, m_{2,b})$. In the last block B, only resolution and common information are sent, i.e., $\mathbf{x}_{1,B} = \mathbf{x}_1(\hat{m}_{-1,B}, 1)$, $\mathbf{x}_{2,B} = \mathbf{x}_2(\tilde{m}_{-1,B}, 1)$.

c) Handling cribbed information: To enable cooperation after block b in $[1 : B-1]$, Encoder 1 chooses $\hat{m}_{2,b}$ such that

$$\big(\mathbf{u}(\hat{m}_{-1,b}), \mathbf{x}_2(\hat{m}_{-1,b}, \hat{m}_{2,b}), x_{2,(b-1)n+1}^{bn}\big) \in \mathcal{A}_\epsilon^{(n)}(U, X_2, X_2). \tag{209}$$

Likewise, Encoder 2 chooses $\tilde{m}_{1,b}$ such that

$$\big(\mathbf{u}(\tilde{m}_{-1,b}), \mathbf{x}_1(\tilde{m}_{-1,b}, \tilde{m}_{1,b}), x_{1,(b-1)n+1}^{bn}\big) \in \mathcal{A}_\epsilon^{(n)}(U, X_1, X_1). \tag{210}$$

Note that the estimates $\hat{m}_{-1,b}$ and $\tilde{m}_{-1,b}$ are formed after Block $b-1$ and that the previous channel inputs $x_{2,(b-1)n+1}^{bn}$ and $x_{1,(b-1)n+1}^{bn}$ are cribbed. Naturally, $\hat{m}_{-1,1} = \tilde{m}_{-1,1} = m_{-1,1} = (1, 1, 1)$.

d) Decoding: The receiver retrieves the transmitted information through backward decoding, i.e., it waits until the last block B was transmitted and then looks for $\check{m}_{-1,B}$ such that

$$\big(\mathbf{u}(\check{m}_{-1,B}), \mathbf{x}_1(\check{m}_{-1,B}, 1), \mathbf{x}_2(\check{m}_{-1,B}, 1), y_{(B-1)n+1}^{Bn}\big) \in \mathcal{A}_\epsilon^{(n)}(U, X_1, X_2, Y). \tag{211}$$

Fix b in $[2 : B-1]$ and assume the decoder has already found

$$\check{m}_{-1,B}, (\check{m}_{-1,B-1}, \check{m}_{1,B-1}, \check{m}_{2,B-1}), \ldots, (\check{m}_{-1,b+1}, \check{m}_{1,b+1}, \check{m}_{2,b+1}).$$

To decode Block b, the receiver first sets $\check{m}_{1,b} = \check{m}_{-1,b+1}^{(1)}$, $\check{m}_{2,b} = \check{m}_{-1,b+1}^{(2)}$ and then looks for $\check{m}_{-1,b}$ such that

$$\big(\mathbf{u}(\check{m}_{-1,b}), \mathbf{x}_1(\check{m}_{-1,b}, \check{m}_{1,b}), \mathbf{x}_2(\check{m}_{-1,b}, \check{m}_{2,b}), y_{(b-1)n+1}^{bn}\big) \in \mathcal{A}_\epsilon^{(n)}(U, X_1, X_2, Y). \tag{212}$$

With the knowledge of $\check{m}_{-1,b}$, the information in Block $b-1$ can be decoded next. The procedure stops after Block 2 since $m_{-1,1} = (1, 1, 1)$. For b in $[2 : B]$ the receiver sets $\check{m}_{0,b} = \check{m}_{-1,b}^{(0)}$.

e) Analysis of the error probability: The error event is

$$\mathcal{E} = \bigcup_{b=2}^{B} \{\check{M}_{0,b} \ne M_{0,b}\} \cup \bigcup_{b=1}^{B-1} \big(\{\check{M}_{1,b} \ne M_{1,b}\} \cup \{\check{M}_{2,b} \ne M_{2,b}\}\big). \tag{213}$$

Define the critical events

$$\mathcal{E}_{b,m_1}^1 = \Big\{\big(\mathbf{u}(M_{-1,b}), \mathbf{x}_1(M_{-1,b}, m_1), X_{1,(b-1)n+1}^{bn}\big) \in \mathcal{A}_\epsilon^{(n)}(U, X_1, X_1)\Big\}, \tag{214}$$

$$\mathcal{E}_{b,m_2}^2 = \Big\{\big(\mathbf{u}(M_{-1,b}), \mathbf{x}_2(M_{-1,b}, m_2), X_{2,(b-1)n+1}^{bn}\big) \in \mathcal{A}_\epsilon^{(n)}(U, X_2, X_2)\Big\}, \tag{215}$$

$$\mathcal{E}_{b,m_{-1}}^{-1} = \Big\{\big(\mathbf{u}(m_{-1}), \mathbf{x}_1(m_{-1}, M_{1,b}), \mathbf{x}_2(m_{-1}, M_{2,b}), Y_{(b-1)n+1}^{bn}\big) \in \mathcal{A}_\epsilon^{(n)}(U, X_1, X_2, Y)\Big\}, \tag{216}$$

and note that

$$\mathcal{E} \subseteq \bigcup_{b=1}^{B-1} \Bigg(\mathcal{E}_{b,M_{1,b}}^{1c} \cup \mathcal{E}_{b,M_{2,b}}^{2c} \cup \bigcup_{m_1 \neq M_{1,b}} \mathcal{E}_{b,m_1}^{1} \cup \bigcup_{m_2 \neq M_{2,b}} \mathcal{E}_{b,m_2}^{2} \Bigg) \cup \bigcup_{b=2}^{B} \Bigg(\mathcal{E}_{b,M_{-1,b}}^{-1c} \cup \bigcup_{m_{-1} \neq M_{-1,b}} \mathcal{E}_{b,m_{-1}}^{-1} \Bigg).$$

Because of the union bound, since on average over the realization of the message triple (M_0, M_1, M_2) the probability of each critical event is independent of the transmission block, and since the codebook generation is symmetrical in the index sequences, the error probability averaged over the ensemble of codes satisfies

$$P_e \leq (B-1)\Bigg(P\big(\mathcal{E}_{2,1}^{1c}\big) + P\big(\mathcal{E}_{2,1}^{2c}\big) + \sum_{m_1 \neq 1} P(\mathcal{E}_{2,m_1}^{1}) + \sum_{m_2 \neq 1} P(\mathcal{E}_{2,m_2}^{2}) + P\big(\mathcal{E}_{2,(1,1,1)}^{-1c}\big) + \sum_{m_{-1} \neq (1,1,1)} P(\mathcal{E}_{2,m_{-1}}^{-1}) \Bigg), \tag{217}$$

where $(M_{-1,2}, M_{1,2}, M_{2,2}) = \big((1,1,1),1,1\big)$. By the weak-typicality Lemma

$$P(\mathcal{E}_{2,1}^{1c}),\, P(\mathcal{E}_{2,1}^{2c}),\, P(\mathcal{E}_{2,(1,1,1)}^{-1c}) \to 0 \ (n \to \infty). \tag{218}$$

The properties of weakly-typical sequences also imply

$$\begin{aligned}
&\sum_{m_1 \neq 1} P(\mathcal{E}_{2,m_1}^{1}) \\
&\leq 2^{nR_1} \sum_{(u^n, x_{1,1}^n, \tilde{x}_{1,1}^n) \in \mathcal{A}_\epsilon^{(n)}(U, X_1, X_1)} p\big(u^n, x_{1,1}^n\big)\, p\big(\tilde{x}_{1,1}^n | u^n\big) \\
&\leq 2^{nR_1} 2^{n(H(U,X_1,X_1)+\epsilon)} \times 2^{-n(H(U,X_1)-\epsilon)} 2^{-n(H(X_1|U)-2\epsilon)} &(219)\\
&\leq 2^{-n(H(X_1|U)-4\epsilon-R_1)}, &(220)\\
&\sum_{m_2 \neq 1} P(\mathcal{E}_{2,m_2}^{2}) \overset{a)}{\leq} 2^{-n(H(X_2|U)-4\epsilon-R_2)}, &(221)\\
&\sum_{m_{-1} \neq (1,1,1)} P(\mathcal{E}_{2,m_0}^{-1}) \\
&\leq 2^{n(R_0+R_1+R_2)} \sum_{\big(u^n, x_{1,1}^n, x_{2,1}^n, y^n\big) \in \mathcal{A}_\epsilon^{(n)}(U, X_1, X_2, Y)} p\big(u^n, x_{1,1}^n, x_{2,1}^n\big)\, p\big(y^n\big) \\
&\leq 2^{n(R_0+R_1+R_2)} 2^{n(H(U,X_1,X_2,Y)+\epsilon)} \times 2^{-n(H(U,X_1,X_2)-\epsilon)} 2^{-n(H(Y)-\epsilon)} &(222)\\
&\leq 2^{-n(I(U,X_1,X_2;Y)-3\epsilon-(R_0+R_1+R_2))} &(223)\\
&\overset{b)}{=} 2^{-n(I(X_1,X_2;Y)-3\epsilon-(R_0+R_1+R_2))}, &(224)
\end{aligned}$$

where $a)$ is due to symmetry, and $b)$ is true because U, (X_1, X_2), and Y form a Markov chain in that order. Equations (217)–(224) imply that $P_e \to 0\, (n \to \infty)$ provided that B grows sufficiently slowly with n and that the rate triple satisfies

$$\begin{aligned}
R_1 &< H(X_1|U) - 4\epsilon &(225)\\
R_2 &< H(X_2|U) - 4\epsilon &(226)\\
R_0 + R_1 + R_2 &< I(X_1, X_2; Y) - 3\epsilon. &(227)
\end{aligned}$$

To conclude the proof, note that for k in $\{0, 1, 2\}$

$$\frac{1}{nB} \log_2 |\mathcal{M}_k| = \frac{B-1}{B} R_k \to R_k \ (B \to \infty). \tag{228}$$

□

B. A Proof of Theorem 20

Proof: The proof has a converse and a direct part.

1) Converse: The claim can be established with similar arguments as in Section A-D (see the proof of [2, Th. 4.1]). For the case where no SI is available to Encoder 2, we give a proof based on the proof of Theorem 22. To this end, we view the cribbed input symbols as a Parallel Channel Extension and let $Z_i = X_{1,i}$. Then (272), (279), and (293) imply

$$\begin{aligned}
R_1 &\leq I(X_1; Y, X_1 | X_2, U) + C_{1,2} &(229)\\
&= H(X_1|X_2, U) + C_{1,2}, &(230)\\
R_2 &\leq I(X_2; Y, X_1 | X_1, U) + C_{2,1} &(231)\\
&= I(X_2; Y | X_1, U) + C_{2,1}, &(232)\\
R_1 + R_2 &\leq I(X_1, X_2; Y). &(233)
\end{aligned}$$

Moreover, the Dependence-Balance constraint (300) reads

$$\begin{aligned}
0 &\leq I(X_1; X_2 | Y, X_1, U) - I(X_1; X_2 | U) \\
&= -I(X_1; X_2 | U). &(234)
\end{aligned}$$

It is satisfied iff X_1, U, and X_2 form a Markov chain in that order. But in this case (230) simplifies to

$$R_1 \leq H(X_1|U) + C_{1,2}. \tag{235}$$

To conclude the proof, note that U, (X_1, X_2), and Y form a Markov chain in that order.

If strictly-causal SI is available to Encoder 2, then the converse no longer is a direct consequence of the arguments in Section B-C. If we let $Z_i = X_{1,i}$, replace Z_i by $\tilde{Z}_i \triangleq (Z_i, S_i)$ in the derivation of (272), and define $V_i = \tilde{Z}^{i-1}$ as well as $V = V_Q$, then (272), (279), and (293) imply

$$\begin{aligned}
R_1 &\leq I(X_1; Y, X_1, S | X_2, U, V) + C_{1,2} &(236)\\
&\leq H(X_1|X_2, U) + C_{1,2}, &(237)\\
R_2 &\leq I(X_2; Y, X_1 | X_1, U) + C_{2,1} &(238)\\
&= I(X_2; Y | X_1, U) + C_{2,1}, &(239)\\
R_1 + R_2 &\leq I(X_1, X_2; Y). &(240)
\end{aligned}$$

As we argue next, the joint PMF satisfies (51). Indeed, U, (X_1, X_2), and Y form a Markov chain in that order. It thus remains to verify that X_1 and X_2 are conditionally independent given U. Observe that

$$
\begin{aligned}
&p(u_i, s^{i-1}, m_1, m_2)\\
&= \sum_{x_{2,1}^{i-1}} p(s^{i-1}) p(m_1) p(m_2) \prod_{j=1}^{i-1} \big[p(g_{1,j}|m_1, g_{2,1}^{j-1}, y^{j-1})\\
&\quad \times p(g_{2,j}|m_2, x_{1,1}^{j-1}, g_{1,1}^{j-1}, s^{j-1}, y^{j-1}) p(x_{1,j}|m_1, g_{2,1}^{j}, y^{j-1})\\
&\quad \times p(x_{2,j}|m_2, x_{1,1}^{j-1}, g_{1,1}^{j}, s^{j-1}, y^{j-1}) p(y_j|x_{1,j}, x_{2,j}, s_j)\big]\\
&\quad \times p(g_{1,i}|m_1, g_{2,1}^{i-1}, y^{i-1}) p(g_{2,i}|m_2, x_{1,1}^{i-1}, g_{1,1}^{i-1}, s^{i-1}, y^{i-1})\\
&= p(s^{i-1}) p(m_1) p(m_2) \prod_{j=1}^{i-1} \big[p(g_{1,j}|m_1, g_{2,1}^{j-1}, y^{j-1})\\
&\quad \times p(g_{2,j}|m_2, x_{1,1}^{j-1}, g_{1,1}^{j-1}, s^{j-1}, y^{j-1}) p(x_{1,j}|m_1, g_{2,1}^{j}, y^{j-1})\\
&\quad \times \sum_{x_{2,j}} p(x_{2,j}|m_2, x_{1,1}^{j-1}, g_{1,1}^{j}, s^{j-1}, y^{j-1}) p(y_j|x_{1,j}, x_{2,j}, s_j)\big]\\
&\quad \times\ p(g_{1,i}|m_1, g_{2,1}^{i-1}, y^{i-1}) p(g_{2,i}|m_2, x_{1,1}^{i-1}, g_{1,1}^{i-1}, s^{i-1}, y^{i-1})\ .
\end{aligned}
$$

Since the joint PMF is of the form (62), Lemma 25 implies that M_1 and (M_2, S^{i-1}) are conditionally independent given U_i. The encoding functions (48c) and (48d) and the definition (266a) of the auxiliary random variable U_i imply that $X_{1,i}$ is computable from (M_1, U_i) and similarly $X_{2,i}$ from (M_2, U_i, S^{i-1}). Hence, $X_{1,i}$ and $X_{2,i}$ are conditionally independent given U_i. Since the auxiliary random variable U comprises Q, we conclude that X_1, U, and X_2 form a Markov chain in that order.

2) Direct Part: Achievability is established similarly as in Section B-A. Observe that all rate pairs in the capacity region of the MAC with a cribbing Encoder 2 and where the senders observe private messages of rates

$$\tilde{R}_1 = \max\{R_1 - C_{1,2}, 0\}, \tag{241a}$$
$$\tilde{R}_2 = \max\{R_2 - C_{2,1}, 0\} \tag{241b}$$

and a common message of rate $R_0 = \min\{R_1, C_{1,2}\} + \min\{R_2, C_{2,1}\}$ are achievable. Hence, it is enough to show that the capacity region of the MAC with a common message and a cribbing Encoder 2 contains the set of rate triples $(R_0, \tilde{R}_1, \tilde{R}_2)$ satisfying

$$\tilde{R}_1 \le H(X_1|U) \tag{242a}$$
$$\tilde{R}_2 \le I(X_2; Y|X_1, U) \tag{242b}$$
$$R_0 + \tilde{R}_1 + \tilde{R}_2 \le I(X_1, X_2; Y) \tag{242c}$$

for some joint PMF of the form

$$p(u, x_1, x_2, y) = p(u)\, p(x_1|u)\, p(x_2|u)\, p(y|x_1, x_2). \tag{243}$$

To this end, we employ a random coding argument. The encoding is done in B blocks during each of which both encoders superimpose private message blocks over a common part, which comprises resolution information for the previous transmission block and information about the common message.

Fix a sufficiently small $\epsilon > 0$ and a joint PMF

$$p(u, x_1, x_2) = p(u)\, p(x_1|u)\, p(x_2|u)\,. \tag{244}$$

a) Codebook generation: For b in $[1 : B]$, draw $2^{n(R_0+R_1)}$ length-n sequences $\mathbf{u}$ from the PMF $\prod_{i=1}^n p(u_i)$. Index them $m_{-1} = (m_{-1}^{(0)}, m_{-1}^{(1)})$, $m_{-1}^{(0)}$ in $[1 : 2^{nR_0}]$, $m_{-1}^{(1)}$ in $[1 : 2^{nR_1}]$. For k in $\{1, 2\}$, m_{-1} in $\{(1,1), \ldots, (2^{nR_0}, 2^{nR_1})\}$, and $m_k \in [1 : 2^{nR_k}]$ draw a length-n codeword $\mathbf{x}_k$ from the PMF $\prod_{i=1}^n p\big(x_{k,i}|u_i(m_{-1})\big)$ and label it $\mathbf{x}_k(m_{-1}, m_k)$.

b) Encoding: Split the common message m_0, the private message m_1 of Encoder 1, and the private message m_2 of Encoder 2 into $B-1$ blocks b in $[1 : B-1]$ of equal length: $m_0 = m_{0,2}, \ldots, m_{0,B}$, $m_1 = m_{1,1}, \ldots, m_{1,B-1}$, and $m_2 = m_{2,2}, \ldots, m_{2,B}$. In the first block, Encoder 1 chooses the codeword $\mathbf{x}_{1,1} = \mathbf{x}_1\big((1,1), m_{1,1}\big)$, and Encoder 2 chooses $\mathbf{x}_{2,1} = \mathbf{x}_2\big((1,1), 1\big)$. For b in $[2 : B]$, cribbing allows Encoder 2 to form the estimate $\tilde{m}_{1,b-1}$ of $m_{1,b-1}$. The estimate is used to send resolution information, which each encoder individually superimposes with its next private message block. Denote $m_{-1,b} = (m_{0,b}, m_{1,b-1})$ and $\tilde{m}_{-1,b} = (m_{0,b}, \tilde{m}_{1,b-1})$. Let b be an element of $[2 : B-1]$. In Block b, Encoder 1 chooses the codeword $\mathbf{x}_{1,b} = \mathbf{x}_1(m_{-1,b}, m_{1,b})$, and Encoder 2 chooses $\mathbf{x}_{2,b} = \mathbf{x}_2(\tilde{m}_{-1,b}, m_{2,b})$. In the last block B, Encoder 1 does not transmit any private information, i.e., $\mathbf{x}_{1,B} = \mathbf{x}_1(m_{-1,B}, 1)$, $\mathbf{x}_{2,B} = \mathbf{x}_2(\tilde{m}_{-1,B}, m_{2,B})$.

c) Handling cribbed information: To enable cooperation after block b in $[1 : B-1]$, Encoder 2 chooses $\tilde{m}_{1,b}$ such that

$$\big(\mathbf{u}(\tilde{m}_{-1,b}), \mathbf{x}_1(\tilde{m}_{-1,b}, \tilde{m}_{1,b}), x_{1,(b-1)n+1}^{bn}\big) \in \mathcal{A}_\epsilon^{(n)}(U, X_1, X_1).$$

Note that the estimate $\tilde{m}_{-1,b}$ is formed after Block $b-1$ and that the previous channel inputs $x_{1,(b-1)n+1}^{bn}$ are cribbed. Naturally, $\tilde{m}_{-1,1} = m_{-1,1} = (1, 1)$.

d) Decoding: The receiver retrieves the transmitted information through backward decoding, i.e., it waits until the last block B was transmitted and then looks for $\check{m}_{-1,B}$, $\check{m}_{2,B}$ such that

$$
\begin{aligned}
&\big(\mathbf{u}(\check{m}_{-1,B}), \mathbf{x}_1(\check{m}_{-1,B}, 1), \mathbf{x}_2(\check{m}_{-1,B}, \check{m}_{2,B}), y_{(B-1)n+1}^{Bn}\big)\\
&\quad \in \mathcal{A}_\epsilon^{(n)}(U, X_1, X_2, Y).
\end{aligned} \tag{245}
$$

Fix b in $[2 : B-1]$ and assume the decoder has already found

$$
\begin{aligned}
&(\check{m}_{-1,B}, \check{m}_{2,B}), (\check{m}_{-1,B-1}, \check{m}_{1,B-1}, \check{m}_{2,B-1}), \ldots,\\
&\quad (\check{m}_{-1,b+1}, \check{m}_{1,b+1}, \check{m}_{2,b+1}).
\end{aligned} \tag{246}
$$

To decode Block b, the receiver first sets $\check{m}_{1,b} = \check{m}_{-1,b+1}^{(1)}$ and then looks for $\check{m}_{-1,b}$, $\check{m}_{2,b}$ such that

$$
\begin{aligned}
&\big(\mathbf{u}(\check{m}_{-1,b}), \mathbf{x}_1(\check{m}_{-1,b}, \check{m}_{1,b}), \mathbf{x}_2(\check{m}_{-1,b}, \check{m}_{2,b}), y_{(b-1)n+1}^{bn}\big)\\
&\quad \in \mathcal{A}_\epsilon^{(n)}(U, X_1, X_2, Y).
\end{aligned} \tag{247}
$$

With the knowledge of $\check{m}_{-1,b}$, the information in Block $b-1$ can be decoded next. The procedure stops after Block 2 since $m_{-1,1} = (1, 1)$, $m_{2,1} = 1$. For b in $[2 : B]$ the receiver sets $\check{m}_{0,b} = \check{m}_{-1,b}^{(0)}$.

e) Analysis of the error probability: The error event is

$$
\begin{aligned}
\mathcal{E} &= \bigcup_{b=2}^{B} \big(\{\check{M}_{0,b} \neq M_{0,b}\} \cup \{\check{M}_{2,b} \neq M_{2,b}\}\big)\\
&\quad \cup \bigcup_{b=1}^{B-1} \{\check{M}_{1,b} \neq M_{1,b}\}.
\end{aligned} \tag{248}
$$

Define the critical events

$$\mathcal{E}^1_{b,m_1} = \Big\{ \big(\mathbf{u}(M_{-1,b}), \mathbf{x}_1(M_{-1,b}, m_1), X^{bn}_{1,(b-1)n+1}\big) \in \mathcal{A}^{(n)}_\epsilon(U, X_1, X_1) \Big\}, \tag{249}$$

$$\mathcal{E}^2_{b,m_2} = \Big\{ \big(\mathbf{u}(M_{-1,b}), \mathbf{x}_1(M_{-1,b}, M_1), \mathbf{x}_2(M_{-1,b}, m_2), Y^{bn}_{(b-1)n+1}\big) \in \mathcal{A}^{(n)}_\epsilon(U, X_1, X_2, Y) \Big\}, \tag{250}$$

$$\mathcal{E}^{-1}_{b,m_{-1},m_2} = \Big\{ \big(\mathbf{u}(m_{-1}), \mathbf{x}_1(m_{-1}, M_{1,b}), \mathbf{x}_2(m_{-1}, m_2), Y^{bn}_{(b-1)n+1}\big) \in \mathcal{A}^{(n)}_\epsilon(U, X_1, X_2, Y) \Big\}, \tag{251}$$

and note that

$$\mathcal{E} \subseteq \bigcup_{b=1}^{B-1} \Bigg(\mathcal{E}^{1c}_{b,M_{1,b}} \cup \bigcup_{m_1 \neq M_{1,b}} \mathcal{E}^1_{b,m_1} \Bigg) \cup \bigcup_{b=2}^{B} \Bigg(\mathcal{E}^{-1c}_{b,M_{-1,b},M_{2,b}} \cup \bigcup_{m_{-1} \neq M_{-1,b}, m_2} \mathcal{E}^{-1}_{b,m_{-1},m_2} \cup \bigcup_{m_2 \neq M_{2,b}} \mathcal{E}^2_{b,m_2} \Bigg). \tag{252}$$

Because of the union bound, since on average over the realization of the message triple (M_0, M_1, M_2) the probability of each critical event is independent of the transmission block, and since the codebook generation is symmetrical in the index sequences, the error probability averaged over the ensemble of codes satisfies

$$P_e \leq (B-1) \Bigg(P(\mathcal{E}^{-1c}_{2,(1,1),1}) + \sum_{m_{-1} \neq (1,1), m_2} P(\mathcal{E}^{-1}_{2,m_{-1},m_2}) + \sum_{m_2 \neq 1} P(\mathcal{E}^2_{2,m_2}) + P(\mathcal{E}^{1c}_{2,1}) + \sum_{m_1 \neq 1} P(\mathcal{E}^1_{2,m_1}) \Bigg), \tag{253}$$

where $(M_{-1,2}, M_{1,2}, M_{2,2}) = \big((1,1),1,1\big)$. By the weak-typicality Lemma

$$P(\mathcal{E}^{1c}_{2,1}),\ P(\mathcal{E}^{-1c}_{2,(1,1),1}) \to 0 \ (n \to \infty). \tag{254}$$

The properties of weakly-typical sequences also imply

$$\sum_{m_1 \neq 1} P(\mathcal{E}^1_{2,m_1}) \overset{a)}{\leq} 2^{-n(H(X_1|U) - 4\epsilon - R_1)}, \tag{255}$$

$$\begin{aligned} \sum_{m_2 \neq 1} P(\mathcal{E}^2_{2,m_2}) &\leq 2^{nR_2} \sum_{(u^n, x^n_{1,1}, x^n_{2,1}, y^n) \in \mathcal{A}^{(n)}_\epsilon(U, X_1, X_2, Y)} \times p\big(u^n, x^n_{1,1}, y^n\big)\, p\big(x^n_{2,1}|u^n\big) \\ &\leq 2^{nR_2} 2^{n(H(U,X_1,X_2,Y)+\epsilon)} \times 2^{-n(H(U,X_1,Y)-\epsilon)} 2^{-n(H(X_2|U)-2\epsilon)} &(256) \\ &\leq 2^{-n(I(X_2;X_1,Y|U)-4\epsilon-R_2)} &(257) \\ &\overset{b)}{=} 2^{-n(I(X_2;Y|X_1,U)-4\epsilon-R_2)}, &(258) \end{aligned}$$

$$\begin{aligned} \sum_{m_{-1} \neq (1,1), m_2} P(\mathcal{E}^{-1}_{2,m_0}) &\leq 2^{n(R_0+R_1+R_2)} \sum_{\big(u^n, x^n_{1,1}, x^n_{2,1}, y^n\big) \in \mathcal{A}^{(n)}_\epsilon(U, X_1, X_2, Y)} p\big(u^n, x^n_{1,1}, x^n_{2,1}\big)\, p\big(y^n\big) \\ &\leq 2^{n(R_0+R_1+R_2)} 2^{n(H(U,X_1,X_2,Y)+\epsilon)} \times 2^{-n(H(U,X_1,X_2)-\epsilon)} 2^{-n(H(Y)-\epsilon)} &(259) \\ &\leq 2^{-n(I(U,X_1,X_2;Y)-3\epsilon-(R_0+R_1+R_2))} &(260) \\ &\overset{c)}{=} 2^{-n(I(X_1,X_2;Y)-3\epsilon-(R_0+R_1+R_2))}, &(261) \end{aligned}$$

where $a)$ is (220), $b)$ holds since X_1 and X_2 are conditionally independent given U, and $c)$ is true because U, (X_1, X_2), and Y form a Markov chain in that order. Equations (253)–(261) imply that $P_e \to 0$ $(n \to \infty)$ provided that B grows sufficiently slowly with n and that the rate triple satisfies

$$\begin{aligned} R_1 &< H(X_1|U) - 4\epsilon &(262) \\ R_2 &< I(X_2; Y|X_1, U) - 4\epsilon &(263) \\ R_0 + R_1 + R_2 &< I(X_1, X_2; Y) - 3\epsilon. &(264) \end{aligned}$$

To conclude the proof, note that for k in $\{0, 1, 2\}$

$$\frac{1}{nB} \log_2 |\mathcal{M}_k| = \frac{B-1}{B} R_k \to R_k \ (B \to \infty). \tag{265}$$

$\square$

C. A Proof of Theorem 22

Proof: For i in $[1:n]$, let Z_i denote the Time-i output of the parallel channel. We assume that it only depends on $\big(X_{1,i}, X_{2,i}, Y^i, Z^{i-1}, G^i_{1,1}, G^i_{2,1}\big)$. Let $Q \sim \text{Unif}[1:n]$ and denote

$$\begin{aligned} U_i &\triangleq \big(Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1}\big), &(266a) \\ U &\triangleq (U_Q, Q),\ X_1 \triangleq X_{1,Q}, \quad X_2 \triangleq X_{2,Q}, &(266b) \\ Y &\triangleq Y_Q,\ Z \triangleq Z_Q. &(266c) \end{aligned}$$

The rate of Encoder 1 satisfies

$$\begin{aligned} n(R_1 - \epsilon_n) &\overset{a)}{\leq} I(M_1; Y^n, Z^n, G^n_{1,1}, G^n_{2,1}, M_2) &(267) \\ &\overset{b)}{=} I(M_1; Y^n, Z^n, G^n_{1,1}, G^n_{2,1}|M_2) &(268) \\ &\overset{c)}{=} \sum_{i=1}^n \big[I(M_1; G_{1,i}, G_{2,i}|Y^{i-1}, Z^{i-1}, G^{i-1}_{1,1}, G^{i-1}_{2,1}, M_2) \\ &\quad + I(M_1; Y_i, Z_i|X_{2,i}, Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1}, M_2) \big] \\ &\overset{d)}{\leq} \sum_{i=1}^n I(M_1; Y_i, Z_i|X_{2,i}, Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1}, M_2) + nC_{1,2} &(269) \\ &\overset{e)}{\leq} \sum_{i=1}^n I(X_{1,i}, M_1; Y_i, Z_i|X_{2,i}, Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1}, M_2) + nC_{1,2} &(270) \\ &\overset{f)}{\leq} \sum_{i=1}^n I(X_{1,i}; Y_i, Z_i|X_{2,i}, Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1}) \end{aligned}$$

$$+ nC_{1,2} \tag{271}$$
$$\overset{g)}{=} nI(X_1; Y, Z|X_2, U) + nC_{1,2}, \tag{272}$$

where a) follows from Fano's inequality, b) holds since M_1 and M_2 are independent, c) is true because of the chain-rule and since $X_{2,i}$ is deterministic given $(M_2, G^i_{1,1}, Y^{i-1})$, d) is a consequence of (276) below, e) is true because conditioning cannot increase entropy, f) holds since (M_1, M_2), $\left(X_{1,i}, X_{2,i}, Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1}\right)$, and (Y_i, Z_i) form a Markov chain in that order and since conditioning cannot increase entropy (recall that Z_i solely depends on $\left(X_{1,i}, X_{2,i}, Y^i, Z^{i-1}, G^i_{1,1}, G^i_{2,1}\right)$), and g) is due to (266). The following computation proves inequality d)

$$\sum_{i=1}^n I(M_1; G_{1,i}, G_{2,i}|Y^{i-1}, Z^{i-1}, G^{i-1}_{1,1}, G^{i-1}_{2,1}, M_2)$$
$$\overset{a)}{\leq} \sum_{i=1}^n H(G_{1,i}, G_{2,i}|Y^{i-1}, Z^{i-1}, G^{i-1}_{1,1}, G^{i-1}_{2,1}, M_2)$$
$$\overset{b)}{=} \sum_{i=1}^n H(G_{1,i}|Y^{i-1}, Z^{i-1}, G^{i-1}_{1,1}, G^i_{2,1}, M_2) \tag{273}$$
$$\overset{c)}{\leq} \sum_{i=1}^n H(G_{1,i}) \tag{274}$$
$$\overset{d)}{\leq} \sum_{i=1}^n \log_2|\mathcal{G}_{1,i}| \tag{275}$$
$$\overset{e)}{\leq} nC_{1,2}, \tag{276}$$

where a) is due to the non-negativity of entropy, b) follows from the chain-rule and the fact that $G_{2,i}$ is deterministic given $\left(M_2, G^{i-1}_{1,1}, Y^{i-1}\right)$, c) is true because conditioning cannot increase entropy, d) holds since entropy is maximized by the uniform distribution, and e) is due to (37).

The rate of Encoder 1 also satisfies

$$n(R_1 - \epsilon_n)$$
$$\overset{a)}{\leq} \sum_{i=1}^n I(X_{1,i}, M_1; Y_i|X_{2,i}, Y^{i-1}, G^i_{1,1}, G^i_{2,1}, M_2) + nC_{1,2}$$
$$\overset{b)}{\leq} \sum_{i=1}^n I(X_{1,i}; Y_i|X_{2,i}) + nC_{1,2} \tag{277}$$
$$\overset{c)}{=} nI(X_1; Y|X_2) + nC_{1,2}, \tag{278}$$

where a) is (270) evaluated for a deterministic Z_i, b) holds since conditioning cannot increase entropy and since Y_i, $(X_{1,i}, X_{2,i})$, and $\left(M_1, M_2, Y^{i-1}, G^i_{1,1}, G^i_{2,1}\right)$ form a Markov chain in that order, and c) is due to (266).

By symmetry, the rate of Encoder 2 satisfies the constraints,

$$n(R_2 - \epsilon_n) \leq nI(X_2; Y, Z|X_1, U) + nC_{2,1}, \tag{279}$$
$$n(R_2 - \epsilon_n) \leq nI(X_2; Y|X_1) + nC_{2,1}. \tag{280}$$

The sum-rate satisfies

$$n(R_1 + R_2 - \epsilon_n)$$
$$\overset{a)}{\leq} I(M_1, M_2; Y^n, Z^n, G^n_{1,1}, G^n_{2,1}) \tag{281}$$
$$\overset{b)}{=} \sum_{i=1}^n \big[I(M_1, M_2; G_{1,i}, G_{2,i}|Y^{i-1}, Z^{i-1}, G^{i-1}_{1,1}, G^{i-1}_{2,1})$$
$$+ I(M_1, M_2; Y_i, Z_i|Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1})\big] \tag{282}$$
$$\overset{c)}{\leq} \sum_{i=1}^n I(M_1, M_2; Y_i, Z_i|Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1})$$
$$+ n(C_{1,2} + C_{2,1}) \tag{283}$$
$$\overset{d)}{\leq} \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i, Z_i|Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1})$$
$$+ n(C_{1,2} + C_{2,1}) \tag{284}$$
$$\overset{e)}{=} nI(X_1, X_2; Y, Z|U) + n\left(C_{1,2} + C_{2,1}\right), \tag{285}$$

where a) follows from Fano's inequality, b) is due to the chain-rule, c) is a consequence of (288) below, d) is true because conditioning cannot increase entropy and because (M_1, M_2), $\left(X_{1,i}, X_{2,i}, Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1}\right)$, and (Y_i, Z_i) form a Markov chain in that order (recall that Z_i solely depends on $\left(X_{1,i}, X_{2,i}, Y^{i-1}, Z^{i-1}, G^i_{1,1}, G^i_{2,1}\right)$), and e) is due to (266).

The following computation proves inequality c)

$$\sum_{i=1}^n I(M_1, M_2; G_{1,i}, G_{2,i}|Y^{i-1}, Z^{i-1}, G^{i-1}_{1,1}, G^{i-1}_{2,1})$$
$$\overset{a)}{\leq} \sum_{i=1}^n H(G_{1,i}, G_{2,i}) \tag{286}$$
$$\overset{b)}{\leq} \sum_{i=1}^n \log_2\left(|\mathcal{G}_{1,i}||\mathcal{G}_{2,i}|\right) \tag{287}$$
$$\overset{c)}{\leq} n\left(C_{1,2} + C_{2,1}\right), \tag{288}$$

where a) is due to the non-negativity of entropy and the fact that conditioning cannot increase entropy, b) holds since entropy is maximized by the uniform distribution, and c) is due to (37).

The sum-rate also satisfies

$$n(R_1 + R_2 - \epsilon_n)$$
$$\overset{a)}{\leq} \sum_{i=1}^n I(M_1, M_2; Y^n) \tag{289}$$
$$\overset{b)}{=} \sum_{i=1}^n I(M_1, M_2; Y_i|Y^{i-1}) \tag{290}$$
$$\overset{c)}{=} nI(M_1, M_2; Y|Y^{Q-1}, Q) \tag{291}$$
$$\overset{d)}{\leq} nI(X_1, X_2, M_1, M_2, Y^{Q-1}, Q; Y) \tag{292}$$
$$\overset{e)}{=} nI(X_1, X_2; Y), \tag{293}$$

where a) follows from Fano's inequality, b) is obtained when applying the chain-rule, c) is due to (266) and the uniform distribution of Q, d) holds since conditioning cannot increase entropy, and e) is true because $\bigl(M_1, M_2, Y^{Q-1}, Q\bigr)$ and Y are conditionally independent given (X_1, X_2).

The Dependence-Balance constraint is due to

$$
\begin{aligned}
0 &\overset{a)}{\leq} I(M_1; M_2|Y^n, Z^n, G_{1,1}^n, G_{2,1}^n) - I(M_1; M_2) \\
&\overset{b)}{=} I(M_1; Y^n, Z^n, G_{1,1}^n, G_{2,1}^n|M_2) \\
&\quad - I(M_1; Y^n, Z^n, G_{1,1}^n, G_{2,1}^n) && (294) \\
&\overset{c)}{=} \sum_{i=1}^n \Bigl[I(M_1; G_{1,i}, G_{2,i}|Y^{i-1}, Z^{i-1}, G_{1,1}^{i-1}, G_{2,1}^{i-1}, M_2) \\
&\qquad + I(M_1; Y_i, Z_i|Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i, M_2) \\
&\qquad - I(M_1; G_{1,i}, G_{2,i}|Y^{i-1}, Z^{i-1}, G_{1,1}^{i-1}, G_{2,1}^{i-1}) \\
&\qquad - I(M_1; Y_i, Z_i|Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i) \Bigr] && (295) \\
&\overset{d)}{\leq} \sum_{i=1}^n \Bigl[I(M_1; Y_i, Z_i|Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i, M_2) \\
&\qquad - I(M_1; Y_i, Z_i|Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i) \Bigr] && (296) \\
&\overset{e)}{=} \sum_{i=1}^n \bigl[I(X_{1,i}, M_1; Y_i, Z_i|X_{2,i}, Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i, M_2) \\
&\qquad - I(X_{1,i}, M_1; Y_i, Z_i|Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i) \bigr] && (297) \\
&\overset{f)}{\leq} \sum_{i=1}^n \bigl[I(X_{1,i}; Y_i, Z_i|X_{2,i}, Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i) \\
&\qquad - I(X_{1,i}; Y_i, Z_i|Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i) \bigr] && (298) \\
&\overset{g)}{=} \sum_{i=1}^n \bigl[I(X_{1,i}; X_{2,i}|Y_i, Z_i, Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i) \\
&\qquad - I(X_{1,i}; X_{2,i}|Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i) \bigr] && (299) \\
&\overset{h)}{=} n\bigl[I(X_1; X_2|Y, Z, U) - I(X_1; X_2|U) \bigr], && (300)
\end{aligned}
$$

where a) holds since mutual information is non-negative and since M_1 and M_2 are independent, b) is obtained when we regroup the entropy terms, c) is a consequence of the chain-rule, d) follows from (302) below, e) is true because $X_{1,i}$ is deterministic given $\bigl(M_1, G_{2,1}^i, Y^{i-1}\bigr)$ and likewise $X_{2,i}$ given $\bigl(M_2, G_{1,1}^i, Y^{i-1}\bigr)$, f) holds since (M_1, M_2) and (Y_i, Z_i) are conditionally independent given $\bigl(X_{1,i}, X_{2,i}, Y^{i-1}, Z^{i-1}, G_{1,1}^i, G_{2,1}^i\bigr)$ and since conditioning cannot increase entropy, g) is obtained when we regroup the entropy terms, and h) is due to (266).

The following computation proves inequality d)

$$
\begin{aligned}
&I(M_1; G_{1,i}, G_{2,i}|Y^{i-1}, Z^{i-1}, G_{1,1}^{i-1}, G_{2,1}^{i-1}) \\
&\quad - I(M_1; G_{1,i}, G_{2,i}|Y^{i-1}, Z^{i-1}, G_{1,1}^{i-1}, G_{2,1}^{i-1}, M_2) \\
&\quad \overset{a)}{=} I(M_1; G_{2,i}|Y^{i-1}, Z^{i-1}, G_{1,1}^{i-1}, G_{2,1}^{i-1}) \\
&\qquad + H(G_{1,i}|Y^{i-1}, Z^{i-1}, G_{1,1}^{i-1}, G_{2,1}^i) \\
&\qquad - H(G_{1,i}|Y^{i-1}, Z^{i-1}, G_{1,1}^{i-1}, G_{2,1}^i, M_2) && (301) \\
&\quad \overset{b)}{\geq} 0, && (302)
\end{aligned}
$$

where a) holds because of the chain-rule and since $G_{1,i}$ is deterministic given $\bigl(M_1, G_{2,1}^{i-1}, Y^{i-1}\bigr)$ and likewise $G_{2,i}$ given $\bigl(M_2, G_{1,1}^{i-1}, Y^{i-1}\bigr)$, and b) is true because conditioning cannot increase entropy. □

D. A Proof of Theorem 24

Proof: Let $A = f_1(Y, X_1)$ be the random variable of Definition 3 and recall that $A = f_2(Y, X_2)$ almost surely. As in [8, Sec. VI, e)], there exists a random variable C with $H(C|A, Y, U) = 0$ such that

$$Z \triangleq (A, C) \tag{303}$$

is an admissible Parallel Channel Extension (in the sense that (M_1, M_2), (X_1, X_2, U), and (Y, Z) form a Markov chain in that order) and such that $I(X_1; X_2|Y, Z, U) = 0$. Because of (57), this implies that $I(X_1; X_2|U) = 0$, and thus

$$
\begin{aligned}
&p(u, x_1, x_2, y, z) \\
&\quad = p(u)\, p(x_1|u)\, p(x_2|u) \\
&\qquad \times W(y|x_1, x_2) p(z|x_1, x_2, y, u)\,. && (304)
\end{aligned}
$$

For the above Parallel Channel Extension Z, the rate of Encoder 1 satisfies

$$
\begin{aligned}
R_1 &\overset{a)}{\leq} I(X_1; Y, Z|X_2, U) + C_{1,2} && (305) \\
&\overset{b)}{=} I(X_1; Y|X_2, U) + I(X_1; A|Y, X_2, U) \\
&\quad + I(X_1; C|A, Y, X_2, U) + C_{1,2} && (306) \\
&\overset{c)}{=} I(X_1; Y|X_2, U) + C_{1,2}, && (307)
\end{aligned}
$$

where a) is (55a), b) is a consequence of (303) and the chain-rule, and c) is true because A is deterministic given (Y, X_2) and $H(C|A, Y, U) = 0$.

Similarly, (55c) implies

$$R_2 \leq I(X_2; Y|X_1, U) + C_{2,1}. \tag{308}$$

Consequently, the capacity region is contained in the set of rate pairs (R_1, R_2) satisfying (307), (308), and (55f) for some joint PMF of the form (304). To conclude, observe that the derived outer bound coincides with the Cover-Leung inner bound [4] for the setting with a common message, which can comprise $nC_{1,2}$ bits of Message 1 and $nC_{2,1}$ bits of Message 2.[15] □

Acknowledgments

Discussions with Y. Steinberg are gratefully acknowledged. The authors would also like to thank the anonymous reviewers and the Associate Editor whose valuable comments helped substantially improve the quality of the paper.

[15]The Cover-Leung inner bound on the feedback capacity of the MAC applies to the setting without common message [4]. We can readily extend it to the setting with a common message by letting the common information comprise not only resolution information but also a fresh common message block. Put differently, one can establish the Cover-Leung inner bound on the feedback capacity of the MAC with a common message by using the random coding argument of Section B-A but by letting each encoder estimate the other encoder's private message based on feedback (instead of cribbing), i.e., Encoder 1 chooses $\hat{m}_{2,b}$ such that (instead of (209)) $\bigl(\mathbf{u}(\hat{m}_{-1,b}), \mathbf{x}_2(\hat{m}_{-1,b}, \hat{m}_{2,b}), y_{(b-1)n+1}^{bn}\bigr) \in \mathcal{A}_\epsilon^{(n)}(U, X_2, Y)$ holds, and Encoder 2 chooses $\tilde{m}_{1,b}$ such that (instead of (210)) $\bigl(\mathbf{u}(\tilde{m}_{-1,b}), \mathbf{x}_1(\tilde{m}_{-1,b}, \tilde{m}_{1,b}), y_{(b-1)n+1}^{bn}\bigr) \in \mathcal{A}_\epsilon^{(n)}(U, X_1, Y)$ holds.

REFERENCES

[1] 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.

[2] A. Bracher, "Cribbing, conferencing, feedback, and state-information on the multiple-access channel," M.S. thesis, Dept. Inf. Technol. Elect. Eng., Swiss Federal Inst. Technol., Zürich, Switzerland, Jul. 2012.

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

[4] T. M. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 27, no. 3, pp. 292–298, May 1981.

[5] N. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 100–102, Jan. 1975.

[6] A. El Gamal, "The feedback capacity of degraded broadcast channels (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 379–381, May 1978.

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

[8] A. P. Hekstra and F. M. J. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 44–53, Jan. 1989.

[9] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 6, pp. 445–586, Jan. 2008.

[10] A. Lapidoth and Y. Steinberg, "The multiple access channel with two independent states each known causally to one encoder," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 480–484.

[11] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal side information: Common state," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 32–50, Jan. 2013.

[12] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal side information: Double state," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1379–1393, Mar. 2013.

[13] C. Schürch, "The multiple-access channel with feedback and a common message," M.S. thesis, Dept. Inf. Technol. Elect. Eng., Swiss Federal Inst. Technol., Zürich, Switzerland, Aug. 2009.

[14] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, no. 4, pp. 289–293, Oct. 1958.

[15] 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, Sep. 1973.

[16] A. Somekh-Baruch, S. Shamai, 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.

[17] F. M. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 93–95, Jan. 1982.

[18] F. M. J. Willems and E. C. van der Meulen, "Partial feedback for the discrete memoryless multiple access channel (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 2, pp. 287–290, Mar. 1983.

[19] 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.

[20] F. M. J. Willems, "On multiple access channels with feedback (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 30, no. 6, pp. 842–845, Nov. 1984.

[21] F. M. J. Willems and E. C. 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.

Annina Bracher (S'14) received the B.Sc. and M.Sc. degrees in electrical engineering (both with distinction) from the Swiss Federal Institute of Technology in Zurich in 2010 and 2012 and an additional M.Sc. degree in engineering from Princeton University in 2014. She is currently pursuing a Ph.D. degree in electrical engineering at the Swiss Federal Institute of Technology in Zurich. Her research interests are in information theory and signal processing.

Amos Lapidoth (S'89–M'95–SM'00–F'04) received the B.A. degree in mathematics (summa cum laude, 1986), the B.Sc. degree in electrical engineering (summa cum laude, 1986), and the M.Sc. degree in electrical engineering (1990) all from the Technion-Israel Institute of Technology. He received the Ph.D. degree in electrical engineering from Stanford University in 1995.

In the years 1995-1999 he was an Assistant and Associate Professor at the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, and was the KDD Career Development Associate Professor in Communications and Technology. He is now Professor of Information Theory at the Swiss Federal Institute of Technology in Zurich. His research interests are in digital communications and information theory.

Dr. Lapidoth served in the years 2003-2004 and 2009 as Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.