

The Multiple-Access Channel With Causal Side Information: Common State

Amos Lapidoth, *Fellow, IEEE*, and Yossef Steinberg, *Fellow, IEEE*

Abstract—We show that if a memoryless multiple-access channel (MAC) is governed by an independent and identically distributed state sequence, then—unlike the single-user case—the capacity region is typically increased if the state is revealed to the encoders in a strictly causal way. For this scenario, we derive inner and outer bounds on the capacity region. For the Gaussian MAC whose state sequence comprises the channel noise, we compute the capacity region and propose a variation on the Schalkwijk–Kailath scheme that achieves capacity with a double-exponential decay of the maximal probability of error. We also study the causal case for which we derive an achievable region, which is typically strictly larger than the region achievable with naïve Shannon strategies.

Index Terms—Causal state information, feedback, multiple-access channel (MAC), Shannon strategies, side information (SI), state, strictly causal state information.

I. INTRODUCTION

IN this paper, we study the capacity of a memoryless multiple-access channel (MAC) that is governed by an independent and identically distributed (i.i.d.) state sequence, which is revealed to the two encoders—depending on the scenario—strictly causally or causally. These two scenarios have not received much attention, perhaps because of the nature of their single-user counterparts. For memoryless single-user channels, revealing the state strictly causally to the encoder does not increase capacity. And if causally, then the problem is solved elegantly using Shannon strategies [13]. (Shannon strategies are also optimal for the degraded broadcast channel [15].) As we shall see, the situation is quite different on the MAC. Strictly causal side information (SI) *can* increase the capacity region. And in the causal case, the naïve use of Shannon strategies is suboptimal.

That, like feedback, strictly causal SI can be beneficial in some memoryless networks can be learned from the work of Dueck [4] on the additive-noise broadcast channel (BC) with feedback. In additive-noise BCs, if we think about the noise as channel state, then providing the encoder with a feedback link

from the channel outputs is equivalent to revealing the state to the encoder strictly causally. And Dueck provided an example of such channels where feedback increases the capacity region. In Dueck’s BC, the noise is common to the two receivers. The encoder learns the channel noise via the feedback and transmits it to the two users. Although valuable rate—that otherwise could be used to transmit user messages—is spent on the transmission of the noise, the net effect is an increase in channel capacity because the noise is common to both users.

The first model we study is the MAC with strictly causal SI. Since only past (or delayed) samples of the state are known to the encoders, neither binning (as in Gel’fand and Pinsker’s channel [8]) nor Shannon strategies [13] can be employed. Instead, we derive an achievable region based on block Markov coding. The encoders, having access to a common state sequence (from the previous block), compress and transmit it to the decoder. The users cannot establish cooperation in the transmission of their private messages, but they do cooperate in the transmission of the compressed state, and they thus increase throughput.

Some of the gains from strictly causal state information can be attributed to the ability of the two encoders to compress the state information and to send the compressed version in cooperation. But strictly causal SI is beneficial even in the “double-state” scenario, where the channel is governed by two independent state sequences, each of which is revealed to a different encoder strictly causally. The double-state scenario is studied in [9] and [10]. Here, we only study the case of a single state, which is revealed (causally or strictly causally) to *both* encoders.

It is instructive to consider the additive Gaussian noise MAC, where the state is the channel noise. For this MAC, our proposed inner bound is tight, and we have a complete characterization of the capacity region. This region turns out to be identical to the capacity region of the additive Gaussian noise MAC without SI but with full cooperation between the encoders. The full-cooperation capacity region is always an outer bound on the capacity of the MAC with strictly causal SI (Proposition 1), and it is interesting that on the Gaussian MAC it is achievable. In fact, the achievability can be demonstrated not only using our inner bound but also by an explicit scheme that is reminiscent of the Schalkwijk–Kailath coding scheme for the single-user Gaussian channel with feedback [12]. Although the users do not have access to each other’s message and no feedback is available, a Schalkwijk–Kailath type algorithm can be devised for this channel, yielding the full-cooperation region with a double-exponential decay in the maximal probability of error.

Our block Markov scheme and the Schalkwijk–Kailath algorithm make use of all the past samples of the channel noise. It turns out, however, that much less is needed to achieve the

Manuscript received July 29, 2011; accepted December 12, 2011. Date of current version September 19, 2012; date of current version December 19, 2012. Y. Steinberg was supported by the Israel Science Foundation under Grant 280/07. This paper was presented in part at the International Zurich Seminar, Zurich, Switzerland, March 3–5, 2010 and in part at Wireless Advanced, London, U.K., June 20–22, 2011.

A. Lapidoth is with ETH Zurich (Swiss Federal Institute of Technology), CH-8092 Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch).

Y. Steinberg is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: ysteinbe@ee.technion.ac.il).

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

Digital Object Identifier 10.1109/TIT.2012.2216096

full-cooperation region: it suffices that only the first sample of the channel noise be known (strictly causally) to the encoders (see Section H in the Appendix).¹

The second model we study is the MAC with a common state that is revealed to the encoders *causally*, so the time- i inputs can depend on the time- i state S_i . Here again our achievability result is based on block Markov coding, but in conjunction with Shannon strategies. Our achievable region contains the naïve region, which uses Shannon strategies for the MAC without block Markov coding. And in some cases, the inclusion is strict.

This paper is organized as follows. The definitions and main results pertaining to the first scenario, in which the encoders learn the state sequence *strictly causally*, can be found in Section II. Those pertaining to the second scenario, where the encoders learn the state sequence *causally*, are in Section III. These sections also contain the key examples. The proofs are given in Sections A–K and Sections L and M in the Appendix, respectively.

II. STRICTLY CAUSAL SIDE INFORMATION

A. Basic Definitions

We are given a discrete memoryless state-dependent MAC $P_{Y|S, X_1, X_2}$ with state alphabet \mathcal{S} , state probability mass function (PMF) P_S , input alphabets \mathcal{X}_1 and \mathcal{X}_2 , and output alphabet \mathcal{Y} . The alphabets \mathcal{S} , \mathcal{X}_1 , \mathcal{X}_2 , and \mathcal{Y} are all finite. Sequences of letters from \mathcal{S} are denoted $s^n = (s_1, s_2, \dots, s_n)$ and $s_i^j = (s_i, s_{i+1}, \dots, s_j)$ with similar notation for all alphabets, e.g., $x_1^n = (x_{1,1}, x_{1,2}, \dots, x_{1,n})$, $x_{2,i}^j = (x_{2,i}, x_{2,i+1}, \dots, x_{2,j})$. Sometimes, we denote n -sequences by boldface letters, e.g., \mathbf{y} , \mathbf{x}_1 , \mathbf{s} , etc. The laws governing n -sequences of state and output letters are

$$P_{Y|S, X_1, X_2}^n(\mathbf{y}|\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2) = \prod_{i=1}^n P_{Y|S, X_1, X_2}(y_i|s_i, x_{1,i}, x_{2,i})$$

$$P_S^n(\mathbf{s}) = \prod_{i=1}^n P_S(s_i).$$

For notational convenience, we henceforth omit the superscript n , and we denote the channel by P . Let

$$\phi_k: \mathcal{X}_k \rightarrow [0, \infty), \quad k = 1, 2$$

be single-letter cost functions. The cost associated with the transmission of the sequence \mathbf{x}_k by Encoder k is

$$\phi_k(\mathbf{x}_k) = \frac{1}{n} \sum_{i=1}^n \phi_k(x_{k,i}), \quad k \in \{1, 2\}.$$

B. Coding

Given positive integers ν_1 and ν_2 , let \mathcal{M}_1 denote the set $\{1, 2, \dots, \nu_1\}$, and let \mathcal{M}_2 denote the set $\{1, 2, \dots, \nu_2\}$.

¹In our original submission, we proposed a scheme that requires knowledge of the first *two* samples of the state. We are thankful to the reviewer for pointing out that *one* sample suffices.

Definition 1 (A Code With Strictly Causal SI): An $(n, \nu_1, \nu_2, \Gamma_1, \Gamma_2, \epsilon)$ code with strictly causal SI at the encoders is a pair of sequences of encoder mappings

$$f_{k,i}: \mathcal{S}^{i-1} \times \mathcal{M}_k \rightarrow \mathcal{X}_k, \quad k = 1, 2, \quad i = 1, \dots, n \quad (1)$$

and a decoding map

$$g: \mathcal{Y}^n \rightarrow \mathcal{M}_1 \times \mathcal{M}_2$$

such that the input costs are bounded by Γ_k

$$\frac{1}{\nu_1 \nu_2} \sum_{m_1=1}^{\nu_1} \sum_{m_2=1}^{\nu_2} \sum_{\mathbf{s} \in \mathcal{S}^n} P_S(\mathbf{s}) \phi_k(\mathbf{f}_k(\mathbf{s}, m_k)) \leq \Gamma_k, \quad k = 1, 2$$

and the average probability of error P_e does not exceed ϵ . Here

$$P_e = 1 - \frac{1}{\nu_1 \nu_2} \sum_{m_1=1}^{\nu_1} \sum_{m_2=1}^{\nu_2} \sum_{\mathbf{s} \in \mathcal{S}^n} P_S(\mathbf{s}) P(g^{-1}(m_1, m_2) | \mathbf{s}, \mathbf{f}_1(\mathbf{s}, m_1), \mathbf{f}_2(\mathbf{s}, m_2)) \quad (2)$$

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

$$\mathbf{f}_k(\mathbf{s}, m_k) = (f_{k,1}(m_k), f_{k,2}(s_1, m_k), \dots, f_{k,n}(s^{n-1}, m_k)).$$

The rate pair (R_1, R_2) of the code is²

$$R_1 = \frac{1}{n} \log \nu_1, \quad R_2 = \frac{1}{n} \log \nu_2.$$

A rate-cost quadruple $(R_1, R_2, \Gamma_1, \Gamma_2)$ is said to be achievable if for every $\epsilon > 0$ and sufficiently large n there exists an $(n, 2^{nR_1}, 2^{nR_2}, \Gamma_1, \Gamma_2, \epsilon)$ code with strictly causal SI for the channel $P_{Y|S, X_1, X_2}$. The capacity-cost region of the channel with strictly causal SI is the closure of the set of all achievable quadruples $(R_1, R_2, \Gamma_1, \Gamma_2)$, and is denoted by $\mathcal{C}_{s-c}^{\text{com}}$. The subscript “s-c” stands for “strictly causal,” and the superscript “com” stands for “common” to remind us that the state is revealed to both encoders. For a given pair (Γ_1, Γ_2) of input costs, $\mathcal{C}_{s-c}^{\text{com}}(\Gamma_1, \Gamma_2)$ stands for the section of $\mathcal{C}_{s-c}^{\text{com}}$ at (Γ_1, Γ_2) . Our interest is in characterizing $\mathcal{C}_{s-c}^{\text{com}}(\Gamma_1, \Gamma_2)$.

C. Outer Bounds

We present three outer bounds. The first and simplest states that the capacity of the MAC with strictly causal SI available at both encoders cannot be larger than the capacity of the same MAC without SI but with fully cooperating encoders.³ Let $\mathcal{C}_{s-c,1}^{\text{com}}$ stand for the collection of all $(R_1, R_2, \Gamma_1, \Gamma_2)$ satisfying

$$R_1 + R_2 \leq I(X_1, X_2; Y) \quad (3a)$$

$$\Gamma_k \geq \mathbb{E}[\phi_k(X_k)], \quad k \in \{1, 2\} \quad (3b)$$

for some joint distribution of the form

$$P_{X_1, X_2, S, Y} = P_S P_{X_1, X_2} P_{Y|S, X_1, X_2}. \quad (4)$$

²We use base-two logarithms throughout, and all rates are in bits per channel use.

³An improved bound that is based on cribbing can be found in [2, Corollary 2] and [1, Corollary 2.4 and Remark 2.1].

Proposition 1 (The Full-Cooperation Outer Bound):

$$\mathcal{C}_{s-c}^{\text{com}} \subseteq \mathcal{O}_{s-c,1}^{\text{com}}. \quad (5)$$

Proof: For a proof, see Section A in the Appendix. The intuition is that cooperation cannot hurt, and once we allow cooperation the channel behaves like a single-user channel for which strictly causal state information does not increase capacity. ■

The next outer bound states that the capacity region of the MAC with common strictly causal SI at the encoders is contained in the capacity region of the same MAC with an informed decoder and uninformed encoders. Let $\mathcal{O}_{s-c,2}^{\text{com}}$ stand for the collection of all $(R_1, R_2, \Gamma_1, \Gamma_2)$ satisfying

$$R_1 \leq I(X_1; Y|X_2, S, Q) \quad (6a)$$

$$R_2 \leq I(X_2; Y|X_1, S, Q) \quad (6b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y|S, Q) \quad (6c)$$

$$\Gamma_k \geq \mathbb{E}[\phi_k(X_k)], \quad k \in \{1, 2\} \quad (6d)$$

for some joint distribution of the form

$$P_{Q, X_1, X_2, S, Y} = P_S P_Q P_{X_1|Q} P_{X_2|Q} P_{Y|S, X_1, X_2} \quad (7)$$

where Q is a time-sharing random variable taking values in a set \mathcal{Q} , whose size can be bounded as

$$|\mathcal{Q}| \leq 4. \quad (8)$$

Proposition 2 (The Informed-Decoder Outer Bound):

$$\mathcal{C}_{s-c}^{\text{com}} \subseteq \mathcal{O}_{s-c,2}^{\text{com}}. \quad (9)$$

Proof: See Section B in the Appendix. ■

As to the third outer bound, let $\mathcal{O}_{s-c,3}^{\text{com}}$ denote the set of all quadruples $(R_1, R_2, \Gamma_1, \Gamma_2)$ satisfying

$$R_1 \leq I(U_1; Y|U_2, V, F) - I(U_1; S|U_2, V, F) \quad (10a)$$

$$R_2 \leq I(U_2; Y|U_1, V, F) - I(U_2; S|U_1, V, F) \quad (10b)$$

$$R_1 + R_2 \leq I(U_1, U_2; Y|V, F) - I(U_1, U_2; S|V, F) \quad (10c)$$

$$0 \leq I(V, F; Y) - I(V, F; S) \quad (10d)$$

$$\Gamma_k \geq \mathbb{E}[\phi_k(X_k)], \quad k \in \{1, 2\} \quad (10e)$$

for some joint distribution of the form

$$P_{U_1, U_2, V, F, S, X_1, X_2, Y} = P_{U_1} P_{U_2} P_V P_S P_{F|U_1, U_2, V, S} P_{X_1|U_1, V} P_{X_2|U_2, V} P_{Y|S, X_1, X_2}. \quad (10f)$$

Proposition 3:

$$\mathcal{C}_{s-c}^{\text{com}} \subseteq \mathcal{O}_{s-c,3}^{\text{com}}. \quad (11)$$

Proof: See Section C in the Appendix. ■

The latter bound is more difficult to evaluate than the first two but, as we next show, it is also tighter.

Proposition 4: The outer bound $\mathcal{O}_{s-c,3}^{\text{com}}$ is at least as tight as the full-cooperation outer bound $\mathcal{O}_{s-c,1}^{\text{com}}$

$$\mathcal{O}_{s-c,3}^{\text{com}} \subseteq \mathcal{O}_{s-c,1}^{\text{com}}. \quad (12)$$

Proof: See Section D in the Appendix. For a stronger statement, see [1, Remark 2.4]: the outer bound $\mathcal{O}_{s-c,3}^{\text{com}}$ is at least as tight as the cribbing bound [2, Corollary 2], [1, Corollary 2.4], and the latter is at least as tight (and sometimes tighter) than the full-cooperation bound. ■

Proposition 5: The outer bound $\mathcal{O}_{s-c,3}^{\text{com}}$ is at least as tight as the informed-decoder outer bound $\mathcal{O}_{s-c,2}^{\text{com}}$

$$\mathcal{O}_{s-c,3}^{\text{com}} \subseteq \mathcal{O}_{s-c,2}^{\text{com}}. \quad (13)$$

Proof: See Section E in the Appendix. ■

Although the outer bound $\mathcal{O}_{s-c,3}^{\text{com}}$ is at least as tight as the full-cooperation and informed-decoder outer bounds, the latter two may still be useful because they are much easier to evaluate: they do not require any optimization over auxiliary random variables. But, as the following example demonstrates, this simplicity comes with a price.

Remark 1: For some channels, the outer bound of Proposition 3 is strictly contained in the intersection of the full-cooperation outer bound (Proposition 1) and the informed-decoder outer bound (Proposition 2).⁴

Example 1: Consider the channel where $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \{0, 1\}$, and where the common state S acts as a random switch that connects a randomly chosen transmitter to the output: $Y = x_S$, where S is uniform over $S = \{1, 2\}$.

Both the full-cooperation and the informed-decoder outer bounds contain the rate pair $(1/2, 1/2)$

$$\begin{aligned} \mathcal{O}_{s-c,1}^{\text{com}} &= \{(R_1, R_2) : R_1 + R_2 \leq 1\} \\ \mathcal{O}_{s-c,2}^{\text{com}} &= \{(R_1, R_2) : 0 \leq R_1, R_2 \leq 1/2\} \end{aligned}$$

but the outer bound $\mathcal{O}_{s-c,3}^{\text{com}}$ of Proposition 3 does not contain any rate pairs summing to one.

Proof: See Section F in the Appendix. ■

D. Achievable Region

Let $\mathcal{P}_{s-c}^{\text{com}}$ be the collection of all random variables (U, V, X_1, X_2, S, Y) whose joint distribution satisfies

$$P_{U, V, X_1, X_2, S, Y} = P_S P_U P_{X_1|U} P_{X_2|U} P_{V|S} P_{Y|S, X_1, X_2}. \quad (14)$$

Note that (14) implies the Markov relations $X_1 \circ - U \circ - X_2$ and $V \circ - S \circ - Y$, and it also implies that the triplet (X_1, U, X_2) is independent of (V, S) . Let $\mathcal{R}_{s-c}^{\text{com}}$ be the convex hull of the collection of all $(R_1, R_2, \Gamma_1, \Gamma_2)$ satisfying

$$R_1 \leq I(X_1; Y|X_2, U, V) \quad (15a)$$

$$R_2 \leq I(X_2; Y|X_1, U, V) \quad (15b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y|U, V) \quad (15c)$$

$$R_1 + R_2 \leq I(X_1, X_2, V; Y) - I(V; S) \quad (15d)$$

$$\Gamma_k \geq \mathbb{E}[\phi_k(X_k)], \quad k = 1, 2 \quad (15e)$$

for some $(U, V, X_1, X_2, S, Y) \in \mathcal{P}_{s-c}^{\text{com}}$. Our main achievability result for the strictly causal case is given in the following.

⁴This remark is strengthened in [1, Remark 5.3].

Theorem 1 (An Achievable Region With Strictly Causal SI): Every rate pair in $\mathcal{R}_{s-c}^{\text{com}}$ is achievable on the MAC with strictly causal SI

$$\mathcal{R}_{s-c}^{\text{com}} \subseteq \mathcal{C}_{s-c}^{\text{com}}. \quad (16)$$

Proof: The proof is given in Section G in the Appendix. It is based on a scheme where a lossy version of the state is conveyed to the decoder using Wyner–Ziv compression [17] and block Markov encoding for the MAC with common message [14]. ■

In some cases, the region $\mathcal{R}_{s-c}^{\text{com}}$ coincides with the capacity region $\mathcal{C}_{s-c}^{\text{com}}$. The next example is such a case. Although Theorem 1 is proved for the discrete memoryless case, we apply it here for the Gaussian model. Extension to continuous alphabets can be done as in [16].

Example 2: Consider the Gaussian MAC where the state is the channel noise

$$Y = x_1 + x_2 + S, \quad S \sim \mathcal{N}(0, \sigma_s^2) \quad (17)$$

and where we impose average power constraints on the inputs

$$\phi_1(x) = \phi_2(x) = x^2. \quad (18)$$

The capacity region $\mathcal{C}_{s-c}^{\text{com}}(\Gamma_1, \Gamma_2)$ of this channel when S is revealed strictly causally to the two encoders comprises all rate pairs (R_1, R_2) satisfying

$$R_1 + R_2 \leq \frac{1}{2} \log \left(1 + \frac{(\sqrt{\Gamma_1} + \sqrt{\Gamma_2})^2}{\sigma_s^2} \right). \quad (19)$$

Proof: Since the set of rate pairs satisfying (19) coincides with the full-cooperation outer bound (Proposition 1), it only remains to prove achievability. Two achievability proofs are presented in Section H in the Appendix. The first is based on a judicious choice of the auxiliary random variables defining $\mathcal{R}_{s-c}^{\text{com}}$. The second is based on a Schalkwijk–Kailath type algorithm. The latter approach provides a scheme whose maximal probability of error decays double exponentially to zero. A small variation demonstrates that to achieve the full-cooperation region, not all states must be revealed strictly causally to the encoders: it suffices to reveal (strictly causally) just the state corresponding to the first time instant. This variation extends to the case where the state sequence is a general stationary Gaussian noise, i.e., not necessarily white: in this case too, revealing the time-one state suffices to achieve the full-cooperation region (which now no longer takes the form (19)). ■

This example demonstrates:

Note 1: Revealing the state governing a MAC strictly causally to both encoders can increase the sum-rate capacity.

Such an increase in the sum-rate capacity is impossible if the MAC is governed by two independent states that are each known strictly causally to a different encoder [9], [10].

E. Improved Achievable Region

We next present a tighter inner bound. For simplicity, we assume that no cost constraints are imposed. The improved region

is based on the following scheme. At Block $\nu + 1$, we still use the MAC by sending private messages and a common message. The common message is still a compressed version of the state information from the previous block. The improvement comes from the fact that the private messages need not be entirely composed of fresh information. The private message of Transmitter 1 has two parts. The first, of rate R_1 , is indeed fresh information. But the second, of rate $R_0^{(1)}$, is a compressed version of the pair of sequences (x_1, s) from Block ν (again with the SI being the channel outputs of Block ν). Since Transmitter 1 knows which symbols it sent in the previous block, and since it knows the state of the channel in the previous block, it can compress the pair (x_1, s) . Likewise Transmitter 2. Using Gastpar’s results on the compression of correlated sources with SI [6], we obtain the improved inner bound, which is described in Theorem 2. In this bound, the constraints (25) guarantee that the rates $R_0^{(1)}$, $R_0^{(2)}$, and R_0 be large enough so that they suffice to describe the sequences from Block ν corresponding to V , V_1 , and V_2 . Constraints (21)–(24) guarantee that once Block $\nu + 1$ is decoded and the sequences from Block ν corresponding to V , V_1 , and V_2 become available, the individual rates $R_1 + R_0^{(1)}$, $R_2 + R_0^{(2)}$, and the common rate R_0 become achievable in Block ν .

Theorem 2 (Improved Achievable Region): Assume that the alphabets \mathcal{X}_1 , \mathcal{X}_2 , \mathcal{Y} , and \mathcal{S} are all finite, and assume that no cost constraints are imposed. The rate pair (R_1, R_2) is achievable if for some joint distribution of the form

$$P_{U,V,V_1,V_2,X_1,X_2,S,Y} = P_S P_U P_{X_1|U} P_{X_2|U} P_{V|S} P_{V_1|S,X_1} P_{V_2|S,X_2} P_{Y|S,X_1,X_2} \quad (20)$$

there exist nonnegative numbers $R_0^{(1)}$ and $R_0^{(2)}$ such that

$$R_1 + R_0^{(1)} \leq I(X_1; Y, V_1, V_2, V | X_2, U) \quad (21)$$

$$R_2 + R_0^{(2)} \leq I(X_2; Y, V_1, V_2, V | X_1, U) \quad (22)$$

$$R_1 + R_2 + R_0^{(1)} + R_0^{(2)} \leq I(X_1, X_2; Y, V_1, V_2, V | U) \quad (23)$$

$$R_0 + R_1 + R_2 + R_0^{(1)} + R_0^{(2)} \leq I(X_1, X_2; Y, V_1, V_2, V) \quad (24)$$

and

$$R_0^{(1)} \geq I(X_1, S; V_1 | V, V_2, Y) \quad (25a)$$

$$R_0^{(2)} \geq I(X_2, S; V_2 | V, V_1, Y) \quad (25b)$$

$$R_0 \geq I(S; V | V_1, V_2, Y) \quad (25c)$$

$$R_0^{(1)} + R_0^{(2)} \geq I(X_1, X_2, S; V_1, V_2 | V, Y) \quad (25d)$$

$$R_0^{(1)} + R_0 \geq I(X_1, S; V_1, V | V_2, Y) \quad (25e)$$

$$R_0^{(2)} + R_0 \geq I(X_2, S; V_2, V | V_1, Y) \quad (25f)$$

$$R_0^{(1)} + R_0^{(2)} + R_0 \geq I(X_1, X_2, S; V_1, V_2, V | Y). \quad (25g)$$

The set of rate pairs satisfying these constraints for some such joint distribution is denoted $\mathcal{R}_{s-c}^{\text{com},*}$.

Proof: The remaining part of the proof pertaining to the termination of the block Markov scheme is sketched in Section I in the Appendix. ■

If we only consider joint distributions where V_1 and V_2 are deterministic, and if we set $R_0^{(1)}, R_0^{(2)}$ to zero, we obtain the inner bound of Theorem 1. Thus:

Remark 2: The improved inner bound contains the inner bound $\mathcal{R}_{s-c}^{\text{com}}$ of Theorem 1.

As the following example shows:

Remark 3: For some channels, the improved inner bound is strictly larger than the bound of Theorem 1.

Example 3: Consider a MAC with two binary inputs $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$; a common state $S = (S_0, S_1) \in \{0, 1\}^2$, where S_0, S_1 are i.i.d. with entropy

$$H(S_0) = H(S_1) = 1/2 \quad (26)$$

and an output $Y = (Y_1, Y_2) \in \{0, 1\}^2$ with

$$Y_1 = X_1 \oplus S_{X_2} \quad (27a)$$

$$Y_2 = X_2. \quad (27b)$$

Thus, if X_2 is equal to zero, then Y_1 is the mod-2 sum of X_1 and S_0 , and otherwise, it is the mod-2 sum of X_1 and S_1 . We study the highest rate at which User 2 can communicate when User 1 transmits at rate 1. For this channel

$$\max\{R_2 : (1, R_2) \in \mathcal{R}_{s-c}^{\text{com}}\} = 0 \quad (28)$$

but

$$\max\{R_2 : (1, R_2) \in \mathcal{C}_{s-c}^{\text{com}}\} = 1/2 \quad (29)$$

and the rate-pair $(1, 1/2)$ is in the improved inner bound $\mathcal{R}_{s-c}^{\text{com},*}$.

Proof: The proof is given in Section J in the Appendix; here, we only provide some intuition. Since the decoder observes X_1 only via (27a), in order for R_1 to equal 1, the receiver needs to learn S_{X_2} . Since in the scheme yielding $\mathcal{R}_{s-c}^{\text{com}}$ the encoders only transmit a (possibly compressed) description of the state (in addition to their private messages), and since for R_1 to equal 1 the receiver must be able to compute S_{X_2} from this description, this description must actually convey the full state $S = (S_0, S_1)$, which is of rate 1. And since the channel (27b) from X_2 to the receiver is only of capacity 1, it cannot support any private rate in addition to the full state, thus leading to (28).

However, one can do better using a block Markov scheme where User 2 does not convey the full state S of the previous block but only the values of S_{X_2} from the previous blocks. (It is cognizant of these values thanks to the strictly causal state information and because it knows the values of X_2 from the previous block, as they were sent by it). The rate of S_{X_2} is only 1/2, so it can use the remaining capacity on the channel (27b) to send a private message at rate 1/2, thus demonstrating the achievability of (29).

The rate pair $(1, 1/2)$ is achievable by the improved scheme because it allows User 2 to describe/compress the pair (S, X_2) and not just S . The improved scheme thus allows User 2 to describe (S, X_2) (in a lossy manner) as S_{X_2} . ■

In deriving the improved bound, we used the results of Gastpar [6] on general Wyner–Ziv networks; we did not exploit the additional structure in our problem. In our problem, the three terminals observe (X_1, S) , (X_2, S) , and S . For this scenario, the results of Gastpar can be improved by utilizing not only binning but also superposition coding [1, Ch. 7]. Indeed, for this setting, it has been shown in [1, Th. 7.1] that we can replace (25) with

$$R_0^{(1)} \geq I(V_1; X_1, X_2, S|V_2, V, Y) \quad (30a)$$

$$R_0^{(2)} \geq I(V_2; X_1, X_2, S|V_1, V, Y) \quad (30b)$$

$$R_0^{(1)} + R_0^{(2)} \geq I(V_1, V_2; X_1, X_2, S|V, Y) \quad (30c)$$

$$R_0^{(1)} + R_0^{(2)} + R_0 \geq I(V_1, V_2, V; X_1, X_2, S|Y) \quad (30d)$$

where the joint PMF is now of the form

$$P_{U,V,V_1,V_2,X_1,X_2,S,Y} = P_S P_U P_{X_1|U} P_{X_2|U} P_{V|S} P_{V_1|S,X_1,V} P_{V_2|S,X_2,V} P_{Y|S,X_1,X_2}. \quad (31)$$

We denote the resulting inner bound by $\mathcal{R}_{s-c}^{\text{com},**}$.

Note that if—rather than allowing any PMF of the form (31)—we restrict ourselves to PMFs of the form (20), then the constraints in (30) are a subset of those in (25). Hence

$$\mathcal{R}_{s-c}^{\text{com},**} \supseteq \mathcal{R}_{s-c}^{\text{com},*}. \quad (32)$$

The analysis of the termination of the block Markov scheme is given in Section K in the Appendix.

III. CAUSAL SIDE INFORMATION

A. Basic Definitions and Coding

When the state information is revealed to the encoders *causally*, the encoding functions have the form

$$f_{k,i}: \mathcal{S}^i \times \mathcal{M}_k \rightarrow \mathcal{X}_k, \quad k = 1, 2, \quad i = 1, \dots, n. \quad (33)$$

The definition of codes and achievable rates remain otherwise as in Section II-A and Section II-B. The capacity region and its section at (Γ_1, Γ_2) are now denoted $\mathcal{C}_{\text{cau}}^{\text{com}}$ and $\mathcal{C}_{\text{cau}}^{\text{com}}(\Gamma_1, \Gamma_2)$.

B. Achievable Region

Let $\mathcal{P}_{\text{cau}}^{\text{com}}$ be the collection of all random variables $(U, U_1, U_2, V, X_1, X_2, S, Y)$ whose joint distribution can be written as

$$P_{U,U_1,U_2,V,X_1,X_2,S,Y} = P_S P_U P_{U_1|U} P_{U_2|U} P_{V|S} P_{X_1|U,U_1,S} P_{X_2|U,U_2,S} P_{Y|S,X_1,X_2}. \quad (34)$$

Observe that (34) implies the Markov relations

$$U_1 \circ - U \circ - U_2 \quad (35a)$$

$$V \circ - S \circ - Y \quad (35b)$$

and that

$$(U_1, U, U_2) \text{ is independent of } (V, S). \quad (35c)$$

Let $\mathcal{R}_{\text{cau}}^{\text{com}}$ be the convex hull of the collection of all $(R_1, R_2, \Gamma_1, \Gamma_2)$ satisfying

$$R_1 \leq I(U_1; Y|U_2, U, V) \quad (36a)$$

$$R_2 \leq I(U_2; Y|U_1, U, V) \quad (36b)$$

$$R_1 + R_2 \leq I(U_1, U_2; Y|U, V) \quad (36c)$$

$$R_1 + R_2 \leq I(U_1, U, U_2, V; Y) - I(V; S) \quad (36d)$$

$$\Gamma_k \geq \mathbb{E}[\phi_k(X_k)], \quad k = 1, 2 \quad (36e)$$

for some $(U, U_1, U_2, V, X_1, X_2, S, Y) \in \mathcal{P}_{\text{cau}}^{\text{com}}$. Our main result for the causal case is the following.

Theorem 3 (Achievability Result for Causal Common SI):

$$\mathcal{R}_{\text{cau}}^{\text{com}} \subseteq \mathcal{C}_{\text{cau}}^{\text{com}}. \quad (37)$$

Proof: A sketch of the proof is given in Section L in the Appendix. It proceeds along the lines of the proof of Theorem 1, except that the inputs X_k , $k = 1, 2$, are allowed to depend on the state S , and that additional external random variables U_1 and U_2 that do not depend on S are introduced. This resembles the situation in coding for the single-user channel with causal SI, where a random Shannon strategy can be represented by an external random variable independent of the state [5, Remark 7.6]. ■

Remark 4: In a similar fashion, one can obtain an improved achievable region by building on the coding scheme for Theorem 2.

The proposed scheme outperforms the “naïve approach” of using Shannon strategies without block Markov coding. The naïve approach leads to the rate region comprising all rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(T_1; Y|T_2, Q) \quad (38a)$$

$$R_2 \leq I(T_2; Y|T_1, Q) \quad (38b)$$

$$R_1 + R_2 \leq I(T_1, T_2; Y|Q) \quad (38c)$$

where T_1 is a random Shannon strategy [13], i.e., a chance variable taking values in the set of mapping from \mathcal{S} to \mathcal{X}_1 ; the chance variable T_2 takes values in the set of mapping from \mathcal{S} to \mathcal{X}_2 ; the conditional law of Y given $T_1 = t_1$ and $T_2 = t_2$ is

$$P_{Y|T_1, T_2}(y|t_1, t_2) = \sum_{s \in \mathcal{S}} P_S(s) P_{Y|S, X_1, X_2}(y|s, t_1(s), t_2(s)) \quad (38d)$$

the random variable Q is for time sharing and the allowed joint distribution are of the form

$$P_Q P_{T_1|Q} P_{T_2|Q} P_{Y|T_1, T_2}. \quad (38e)$$

Restricting V in (36a)–(36d) to be deterministic reduces our achievable region $\mathcal{R}_{\text{cau}}^{\text{com}}$ to the set of rates achievable with the naïve approach. Thus:

Remark 5: Every rate pair that is achievable using the naïve approach is also in $\mathcal{R}_{\text{cau}}^{\text{com}}$.

The next example shows that the reverse statement is not true.

Remark 6: For some channels, there exist rate pairs that are in $\mathcal{R}_{\text{cau}}^{\text{com}}$ but that are not achievable using the naïve approach.

Example 4: Consider the noiseless binary MAC where

$$\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \{0, 1\}, \quad \mathcal{S} = \{1, 2\}$$

and $P(S = 2) = 1 - P(S = 1) = p$ for some $1 \geq p > 0.5$. The state S determines which of the two inputs is connected to the output

$$Y = X_S.$$

For this channel

$$\max_{(R_1, R_2) \in \mathcal{R}_{\text{cau}}^{\text{com}}} R_1 \geq \min\{1 - p, 1 - H_b(p)\} \quad \text{bits} \quad (39)$$

(where $H_b(\cdot)$ denotes the binary entropy function $\xi \log \xi^{-1} + (1 - \xi) \log(1 - \xi)^{-1}$) whereas with the naïve approach, the highest achievable rate for Transmitter 1 is

$$\log_2 \left(1 + (1 - p)p^{\frac{p}{1-p}} \right) \quad \text{bits}. \quad (40)$$

For sufficiently large values of p , this is strictly smaller than the right-hand side (RHS) of (39).

If p is large enough so that $p > H_b(p)$, then the RHS of (39) is equal to $1 - p$, which is the highest rate at which Transmitter 1 could have communicated even if the state had been known to the receiver.

Proof: See Section M in the Appendix. ■

APPENDIX

A) Proof of the Full-Cooperation Outer Bound:

Proof of Proposition 1: Let $(R_1, R_2) \in \mathcal{C}_{\text{s-c}}^{\text{com}}(\Gamma_1, \Gamma_2)$. Then, there exists a sequence of $(n, 2^{nR_1}, 2^{nR_2}, \Gamma_1, \Gamma_2, \epsilon_n)$ codes with strictly causal SI, such that $\lim_{n \rightarrow \infty} \epsilon_n = 0$. Denote by M_k the random message of user k , $k = 1, 2$. Starting with Fano’s inequality

$$\begin{aligned} n(R_1 + R_2 - \delta_n) &\leq I(M_1, M_2; Y^n) \\ &= \sum_{i=1}^n I(M_1, M_2; Y_i | Y^{i-1}) \\ &\leq \sum_{i=1}^n I(M_1, M_2, Y^{i-1}; Y_i) \\ &= \sum_{i=1}^n I(U_{1,i}, U_{2,i}; Y_i) \\ &\stackrel{(a)}{\leq} \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i) \end{aligned} \quad (41)$$

where $\lim_{n \rightarrow \infty} \delta_n = 0$

$$U_{k,i} = (M_k, Y^{i-1}), \quad k \in 1, 2 \quad (42)$$

and the input constraints are satisfied

$$\frac{1}{n} \sum_{i=1}^n \phi_k(X_{k,i}) \leq \Gamma_k, \quad k \in 1, 2. \quad (43)$$

Inequality (a) in (41) holds by the data processing inequality: due to the coding scheme, the quadruple $(U_{1,i}, U_{2,i}, X_{1,i}, X_{2,i})$ is independent of S_i , so

$$(U_{1,i}, U_{2,i}) \circ - (X_{1,i}, X_{2,i}) \circ - Y_i. \quad (44)$$

Since the joint distribution of X_1 and X_2 is arbitrary, we do not have to employ time-sharing considerations on (41) and (43), and any achievable quadruple $(R_1, R_2, \Gamma_1, \Gamma_2)$ satisfies

$$R_1 + R_2 \leq I(X_1, X_2; Y) \quad (45a)$$

$$\Gamma_k \geq \mathbb{E}[\phi_k(X_k)], \quad k = 1, 2 \quad (45b)$$

for some joint distribution of the form

$$P_{S, X_1, X_2, Y} = P_S P_{X_1, X_2} P_{Y|S, X_1, X_2}. \quad (46)$$

■

B) Proof of the Informed-Decoder Outer Bound:

Proof of Proposition 2: Starting with Fano's inequality,

$$\begin{aligned} nR_1 - n\delta_n &\leq I(M_1; Y^n S^n | M_2) \\ &= \sum_{i=1}^n I(M_1; Y_i, S_i | M_2, Y^{i-1}, S^{i-1}) \\ &= \sum_{i=1}^n I(M_1; S_i | M_2, Y^{i-1}, S^{i-1}) \\ &\quad + \sum_{i=1}^n I(M_1; Y_i | M_2, Y^{i-1}, S^i) \\ &\stackrel{(a)}{=} \sum_{i=1}^n I(M_1; Y_i | M_2, Y^{i-1}, S^i) \\ &\stackrel{(b)}{=} \sum_{i=1}^n I(M_1, X_{1,i}; Y_i | M_2, X_{2,i}, Y^{i-1}, S^i) \\ &\stackrel{(c)}{\leq} \sum_{i=1}^n I(M_1, M_2, X_{1,i}, Y^{i-1}; Y_i | X_{2,i}, S^i) \\ &\stackrel{(d)}{=} \sum_{i=1}^n I(X_{1,i}; Y_i | X_{2,i}, S^i) \end{aligned} \quad (47)$$

where (a) holds due to the Markov relation

$$S_i \circ - (M_2, Y^{i-1}, S^{i-1}) \circ - M_1 \quad (48)$$

(b) holds because $X_{k,i}$ is a deterministic function of (M_k, S^{i-1}) , $k \in \{1, 2\}$; (c) holds because conditioning cannot increase entropy; and (d) holds due to the Markov chain

$$Y_i \circ - (X_{1,i}, X_{2,i}, S_i) \circ - (M_1, M_2, Y^{i-1}, S^{i-1}). \quad (49)$$

In a similar manner, we obtain for R_2 and the sum rate

$$nR_2 - n\delta_n \leq \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}, S^i) \quad (50)$$

$$n(R_1 + R_2 - \delta_n) \leq \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i | S^i). \quad (51)$$

Defining

$$Q_i = S^{i-1} \quad (52)$$

we can rewrite (47), (50), and (51) as

$$R_1 - \delta_n \leq \frac{1}{n} \sum_{i=1}^n I(X_{1,i}; Y_i | X_{2,i}, S_i, Q_i) \quad (53a)$$

$$R_2 - \delta_n \leq \frac{1}{n} \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}, S_i, Q_i) \quad (53b)$$

$$R_1 + R_2 - \delta_n \leq \frac{1}{n} \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i | S_i, Q_i) \quad (53c)$$

$$\Gamma_k \geq \frac{1}{n} \sum_{i=1}^n \phi_k(X_{k,i}). \quad (53d)$$

Recalling that

$$X_{k,i} = f_{k,i}(M_k, S^{i-1}), \quad k \in \{1, 2\}$$

and that M_1, M_2, S^{i-1} are independent, we conclude that

$$X_{1,i} \circ - S^{i-1} \circ - X_{2,i}$$

and the joint PMF thus factorizes as

$$P_{Q_i, X_{1,i}, X_{2,i}, S_i, Y_i} = P_S P_{Q_i} P_{X_{1,i}|Q_i} P_{X_{2,i}|Q_i} P_{Y_i|S, X_1, X_2}. \quad (53e)$$

Taking the limit $n \rightarrow \infty$ in (53) and using the fact that $\mathcal{O}_{s-c,2}^{\text{com}}$ is convex yields the proposition. The proof that the alphabet size of Q need not be larger than 4 is standard and follows from Carathéodory's theorem. ■

C) Proof of Proposition 3:

Let $(R_1, R_2, \Gamma_1, \Gamma_2)$ be in $\mathcal{C}_{s-c}^{\text{com}}$. Then, we have a sequence of $(n, 2^{nR_1}, 2^{nR_2}, \Gamma_1, \Gamma_2, \epsilon_n)$ codes with strictly causal SI, such that $\lim_{n \rightarrow \infty} \epsilon_n = 0$. Denote by M_k the random message of user k , $k = 1, 2$. Starting with Fano's inequality

$$\begin{aligned} nR_1 - n\delta_n &\leq I(M_1; Y^n | M_2) \\ &= I(M_1; Y^n | M_2) - I(M_1; S^n | M_2) \\ &= \sum_{i=1}^n [I(M_1; Y_i | M_2, Y_{i+1}^n) - I(M_1; S_i | M_2, S^{i-1})] \\ &= \sum_{i=1}^n \left[I(M_1, S^{i-1}; Y_i | M_2, Y_{i+1}^n) \right. \\ &\quad - I(S^{i-1}; Y_i | M_1, M_2, Y_{i+1}^n) \\ &\quad - I(M_1, Y_{i+1}^n; S_i | M_2, S^{i-1}) \\ &\quad \left. + I(Y_{i+1}^n; S_i | M_1, M_2, S^{i-1}) \right] \\ &\stackrel{(a)}{=} \sum_{i=1}^n \left[I(M_1, S^{i-1}; Y_i | M_2, Y_{i+1}^n) \right. \\ &\quad \left. - I(M_1, Y_{i+1}^n; S_i | M_2, S^{i-1}) \right] \\ &= \sum_{i=1}^n \left[I(S^{i-1}; Y_i | M_2, Y_{i+1}^n) + I(M_1; Y_i | M_2, Y_{i+1}^n, S^{i-1}) \right] \end{aligned}$$

$$\begin{aligned}
& - I(Y_{i+1}^n; S_i | M_2, S^{i-1}) - I(M_1; S_i | M_2, Y_{i+1}^n, S^{i-1}) \\
& \stackrel{(b)}{=} \sum_{i=1}^n \left[I(M_1; Y_i | M_2, Y_{i+1}^n, S^{i-1}) \right. \\
& \quad \left. - I(M_1; S_i | M_2, Y_{i+1}^n, S^{i-1}) \right] \\
& = \sum_{i=1}^n \left[I(U_{1,i}; Y_i | U_{2,i}, V_i, F_i) \right. \\
& \quad \left. - I(U_{1,i}; S_i | U_{2,i}, V_i, F_i) \right] \tag{54}
\end{aligned}$$

where (a) and (b) hold due to Csiszar's lemma [5, Sec. 2.3], which states that for any pair of random vectors (A^n, B^n)

$$\sum_{i=1}^n [I(A_{i+1}^n; B_i | B^{i-1}) - I(B^{i-1}; A_i | A_{i+1}^n)] = 0 \tag{55}$$

and in the last equality of (54) we defined

$$U_{k,i} = M_k, \quad k \in \{1, 2\} \tag{56a}$$

$$V_i = S^{i-1} \tag{56b}$$

$$F_i = Y_{i+1}^n. \tag{56c}$$

In a similar manner, we obtain

$$\begin{aligned}
n(R_2 - \delta_n) & \leq \sum_{i=1}^n \left[I(U_{2,i}; Y_i | U_{1,i}, V_i, F_i) \right. \\
& \quad \left. - I(U_{2,i}; S_i | U_{1,i}, V_i, F_i) \right]. \tag{57}
\end{aligned}$$

As to the sum-rate constraint

$$\begin{aligned}
& n(R_1 + R_2 - \delta_n) \\
& \leq I(M_1, M_2; Y^n) \\
& = I(M_1, M_2; Y^n) - I(M_1, M_2; S^n) \\
& = \sum_{i=1}^n \left[I(M_1, M_2; Y_i | Y_{i+1}^n) - I(M_1, M_2; S_i | S^{i-1}) \right] \\
& = \sum_{i=1}^n \left[I(M_1, M_2, S^{i-1}; Y_i | Y_{i+1}^n) \right. \\
& \quad - I(S^{i-1}; Y_i | M_1, M_2, Y_{i+1}^n) \\
& \quad - I(M_1, M_2, Y_{i+1}^n; S_i | S^{i-1}) \\
& \quad \left. + I(Y_{i+1}^n; S_i | M_1, M_2, S^{i-1}) \right] \\
& = \sum_{i=1}^n \left[I(M_1, M_2, S^{i-1}; Y_i | Y_{i+1}^n) \right. \\
& \quad \left. - I(M_1, M_2, Y_{i+1}^n; S_i | S^{i-1}) \right] \\
& = \sum_{i=1}^n \left[I(S^{i-1}; Y_i | Y_{i+1}^n) + I(M_1, M_2; Y_i | Y_{i+1}^n, S^{i-1}) \right. \\
& \quad \left. - I(Y_{i+1}^n; S_i | S^{i-1}) - I(M_1, M_2; S_i | Y_{i+1}^n, S^{i-1}) \right] \\
& = \sum_{i=1}^n \left[I(M_1, M_2; Y_i | Y_{i+1}^n, S^{i-1}) \right. \\
& \quad \left. - I(M_1, M_2; S_i | Y_{i+1}^n, S^{i-1}) \right] \\
& = \sum_{i=1}^n \left[I(U_{1,i}, U_{2,i}; Y_i | V_i, F_i) \right. \\
& \quad \left. - I(U_{1,i}, U_{2,i}; S_i | V_i, F_i) \right]. \tag{58}
\end{aligned}$$

As to the positivity constraint (10d)

$$\begin{aligned}
& \sum_{i=1}^n [I(F_i, V_i; Y_i) - I(F_i, V_i; S_i)] \\
& = \sum_{i=1}^n [H(Y_{i+1}^n, S^{i-1} | S_i) - H(Y_{i+1}^n, S^{i-1} | Y_i)] \\
& = \sum_{i=1}^n [H(Y_{i+1}^n, S^i) - H(S_i) - H(Y_i^n, S^{i-1}) + H(Y_i)] \\
& = -H(Y^n) + H(S^n) + \sum_{i=1}^n [H(Y_i) - H(S_i)] \\
& = -H(Y^n) + \sum_{i=1}^n H(Y_i) \\
& \geq 0. \tag{59}
\end{aligned}$$

Finally, note that $U_{1,i}, U_{2,i}$, and S_i are independent for each i and, due to the coding scheme, $X_{k,i}$ is a deterministic function of $(U_{k,i}, V_i)$, for $k \in \{1, 2\}$. Thus, the joint PMF factorizes as

$$\begin{aligned}
& P_{S_i} P_{U_{1,i}} P_{U_{2,i}} P_{V_i | U_{1,i}, U_{2,i}, S_i} \\
& \quad P_{X_{1,i} | U_{1,i}, V_i} P_{X_{2,i} | U_{2,i}, V_i} P_{Y_i | S_i, X_{1,i}, X_{2,i}}. \tag{60}
\end{aligned}$$

The outer bound (10) now follows by letting n tend to infinity in (54), (57), (58), and in (59), and by using the fact that the region (10) is convex. \blacksquare

D) Proof of Proposition 4:

Fix some joint PMF of the form (10f). We need to show that the polytope defined by the inequalities in (10) is contained in $\mathcal{O}_{s-c,1}^{\text{com}}$. Starting with the sum-rate constraint (10c) and using the positivity constraint (10d), we have

$$\begin{aligned}
R_1 + R_2 & \leq I(U_1, U_2; Y | V, F) - I(U_1, U_2; S | V, F) \\
& = I(U_1, U_2, V, F; Y) - I(V, F; Y) \\
& \quad - (I(U_1, U_2, V, F; S) - I(V, F; S)) \\
& \leq I(U_1, U_2, V, F; Y) - I(U_1, U_2, V, F; S) \\
& = I(U_1, U_2, V; Y) + I(F; Y | U_1, U_2, V) \\
& \quad - (I(U_1, U_2, V; S) + I(F; S | U_1, U_2, V)) \\
& = I(U_1, U_2, V; Y) \\
& \quad + I(F; Y | U_1, U_2, V) - I(F; S | U_1, U_2, V) \\
& \leq I(U_1, U_2, V; Y) \\
& \leq I(X_1, X_2; Y)
\end{aligned}$$

where the first line follows from the sum-rate constraint (10c); the second from the chain rule; the third from the positivity constraint (10d); the fourth from the chain rule; the fifth from (10f); the sixth from the (conditional) data processing inequality because by (10f)

$$F \text{--} \circ \text{--} (U_1, U_2, V, S) \text{--} \circ \text{--} Y \tag{61}$$

and the final inequality holds because by (10f)

$$(U_1, U_2, V) \text{--} \circ \text{--} (X_1, X_2) \text{--} \circ \text{--} Y. \tag{62}$$

\blacksquare

E) Proof of Proposition 5:

Inspecting Inequalities (10), we see that $\mathcal{O}_{s-c,3}^{\text{com}}$ is contained in the result of applying Proposition 3 to a new channel where the output Y is replaced by the augmented output (Y, S) . We will show that the result of applying Proposition 3 to the augmented channel is contained in $\mathcal{O}_{s-c,2}^{\text{com}}$. Substituting (Y, S) for Y in (10) leads to the inequalities

$$R_1 \leq I(U_1; Y|U_2, V, F, S) \quad (63a)$$

$$R_2 \leq I(U_2; Y|U_1, V, F, S) \quad (63b)$$

$$R_1 + R_2 \leq I(U_1, U_2; Y|V, F, S) \quad (63c)$$

where the positivity constraint is omitted because in the augmented channel it is always satisfied. Fix some joint distribution of the form (10f).

Addressing the sum-rate constraint, we have from (63c)

$$\begin{aligned} R_1 + R_2 &\leq I(U_1, U_2; Y|V, F, S) \\ &= H(Y|V, F, S) - H(Y|V, F, S, U_1, U_2) \\ &\leq H(Y|V, S) - H(Y|V, F, S, U_1, U_2, X_1, X_2) \\ &= H(Y|V, S) - H(Y|V, S, X_1, X_2) \\ &= I(X_1, X_2; Y|S, V) \end{aligned} \quad (64)$$

where the second inequality holds because conditioning cannot increase entropy; the subsequent equality because conditional on (S, X_1, X_2) the output Y is independent of all the other variables; and the final equality follows from the definition of conditional mutual information. Notice that by (10f)

$$X_1 \text{---} V \text{---} X_2 \quad (65)$$

and

$$(X_1, X_2, V) \text{ is independent of } S. \quad (66)$$

Consider now the constraint (63a) on R_1

$$\begin{aligned} R_1 &\leq I(U_1; Y|U_2, V, F, S) \\ &\leq H(Y|U_2, V, S) - H(Y|U_1, U_2, V, F, S) \\ &= H(Y|U_2, V, S) - H(Y|U_1, U_2, V, S) \\ &= I(U_1; Y|U_2, V, S) \\ &\leq I(X_1; Y|U_2, V, S) \end{aligned} \quad (67)$$

where the second line follows because conditioning reduces entropy; the third line from

$$(U_1, Y) \text{---} (U_1, U_2, V, S) \text{---} F \quad (68)$$

and the last line follows from the (conditional) data processing inequality, because for every joint law of the form (10f)

$$U_1 \text{---} (U_2, V, S, X_1) \text{---} Y. \quad (69)$$

Continuing from (67)

$$\begin{aligned} R_1 &\leq I(X_1; Y|U_2, V, S) \\ &= H(X_1|V, S, U_2, X_2) - H(X_1|U_2, V, S, Y) \\ &= H(X_1|V, S, X_2) - H(X_1|U_2, V, S, Y) \\ &\leq H(X_1|X_2, V, S) - H(X_1|U_2, V, S, Y, X_2) \\ &= H(X_1|X_2, V, S) - H(X_1|V, S, Y, X_2) \\ &= I(X_1; Y|X_2, V, S). \end{aligned} \quad (70)$$

A symmetric argument yields

$$R_2 \leq I(X_2; Y|X_1, V, S). \quad (71)$$

The inequalities (71), (70), and (64) combine with (65) and (66) to yield the informed-decoder constraints (with the role of Q played by V). ■

F) Analysis of Example 1:

Proof: The proof is by contradiction. We assume that the rate pair (R_1, R_2) sums to one

$$R_1 + R_2 = 1 \quad (72a)$$

and that for some joint PMF of the form (10f) all the constraints (10a)–(10d) are satisfied. We then show that this leads to a contradiction.

From (72a), (10c), and the fact that $H(Y) \leq 1$ (because \mathcal{Y} has only two elements), we conclude that

$$Y \text{ is independent of } (V, F) \quad (72b)$$

$$(U_1, U_2) \text{---} (V, F) \text{---} S \quad (72c)$$

and

$$H(Y|U_1, U_2, V, F) = 0. \quad (72d)$$

From (72b) and the positivity constraint (10d)

$$S \text{ is independent of } (V, F). \quad (72e)$$

It now follows from (72c) and (72f) that

$$S \text{ is independent of } (U_1, U_2, V, F). \quad (72f)$$

From (72d), (72f), and the fact that $Y = X_S$, we conclude that

$$\Pr[X_1 = X_2 | U_1, U_2, V, F] = 1$$

(with probability one) and hence also unconditionally

$$\Pr[X_1 = X_2] = 1. \quad (72g)$$

From (10f)

$$X_1 \text{---} (U_1, V) \text{---} (S, U_2, F, X_2)$$

and hence also

$$X_1 \text{---} (U_1, V, F) \text{---} X_2. \quad (72h)$$

From (72g) and (72h) we see that, conditionally on (U_1, V, F) , the chance variables X_1 and X_2 are independent but also deterministically equal to each other. This implies that $X_1 = X_2$ and that both are deterministic functions of (U_1, V, F) . Because $Y = X_S$, this implies that $H(Y|U_1, V, F) = 0$ and that hence by (10b)

$$R_2 = 0. \quad (72i)$$

By symmetry, we also have

$$R_1 = 0. \quad (72j)$$

We have thus reached a contradiction because (72j), (72i), and (72a) cannot all hold. ■

G) Proof of Theorem 1: Throughout the achievability proof, we use the definition of typical sequences and typical sets as in [3]. Thus, denote by $n(a|\mathbf{u})$ the number of occurrences of the letter a in the n -vector $\mathbf{u} = u^n$. Then, for a given PMF P_U over the finite alphabet \mathcal{U} , the δ -typical set is defined as

$$\mathcal{T}_U^\delta = \{ \mathbf{u} \in \mathcal{U}^n \mid |n^{-1}n(a|\mathbf{u}) - P_U(a)| < \delta, \forall a \in \mathcal{U} \\ n(a|\mathbf{u}) = 0 \text{ when } P_U(a) = 0 \}.$$

For a joint PMF $P_{U,V}$ and a given n -vector \mathbf{v} , the conditional δ -typical set is

$$\mathcal{T}_{U|V}^\delta(\mathbf{v}) = \{ \mathbf{u} \in \mathcal{U}^n : |n^{-1}n(a, b|\mathbf{u}, \mathbf{v}) - n^{-1}n(b|\mathbf{v})P_{U|V}(a|b)| < \delta \\ \forall (a, b) \in \mathcal{U} \times \mathcal{V}, n(a, b|\mathbf{u}, \mathbf{v}) = 0 \text{ for } P_{U|V}(a|b) = 0 \}.$$

Typical sets will be used with δ depending on n , such that

$$\delta_n \rightarrow 0, \quad \sqrt{n} \cdot \delta_n \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

We further adopt the *Delta-Convention* of [3, Convention 2.11] (see also lemmas 2.12 and 2.13 there). Thus, throughout the paper, the sequences δ_n are fixed, and the dependence of the typical sets on them is omitted.

Proof of Theorem 1: We first claim that the region $\mathcal{R}_{s-c}^{\text{com}}$ is contained in the full cooperation region $\mathcal{O}_{s-c,1}^{\text{com}}$, as defined in (3) and (4). To this end, observe that we can write the bound on the sum rate (15d) as

$$\begin{aligned} R_1 + R_2 &\leq I(X_1, X_2; Y) + I(V; Y|X_1, X_2) - I(V; S) \\ &= I(X_1, X_2; Y) + I(V; X_1, X_2, Y) - I(V; S) \quad (73a) \\ &\leq I(X_1, X_2; Y) \quad (73b) \end{aligned}$$

where the equality in (73a) holds since by (14), V is independent of (X_1, X_2) , and the inequality in (73b) holds because of the Markov chain

$$V \text{---} S \text{---} (X_1, X_2, Y).$$

Fix a distribution as in (14). In view of (73b), we can assume that for the chosen distribution

$$I(X_1, X_2; Y) > \mu > 0 \quad (74)$$

for some real number μ , as otherwise there is nothing to prove.

The rate pairs (R_1, R_2) that satisfy the constraints (15) can also be characterized as those for which there exists some $R_0 \geq 0$ such that the triple (R_0, R_1, R_2) satisfies the constraints

$$R_1 \leq I(X_1; Y|X_2, U, V) \quad (75a)$$

$$R_2 \leq I(X_2; Y|X_1, U, V) \quad (75b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y|U, V) \quad (75c)$$

$$R_0 + R_1 + R_2 \leq I(X_1, X_2; Y|V) \quad (75d)$$

$$R_0 \geq I(V; S) - I(V; Y) \quad (75e)$$

$$\Gamma_k \geq \mathbb{E}[\phi_k(X_k)], \quad k = 1, 2 \quad (75f)$$

where as in (14)

$$P_{U,V,X_1,X_2,S,Y} = P_S P_{X_1|U} P_{X_2|U} P_U P_{V|S} P_{Y|S,X_1,X_2} \quad (76)$$

so

$$X_1 \text{---} U \text{---} X_2 \quad (77)$$

$$V \text{---} S \text{---} Y. \quad (78)$$

Indeed, by (78) and the data processing inequality, the RHS of (75e) is nonnegative, so we can choose R_0 so that it hold with equality. We can then substitute this choice of R_0 in (75d) to obtain (15d). To prove Theorem 1, it thus suffices to show that any pair (R_1, R_2) satisfying (75a)–(75e) for some $R_0 \geq 0$ is achievable. This is what we proceed to do next. Before giving a formal proof, we provide some interpretation for (75a)–(75e). Ignoring V , Constraints (75a)–(75d) and (77) are reminiscent of the capacity region of the MAC with a common message of rate R_0 and without SI. The random variable V is independent of (X_1, U, X_2) , but it does depend on S . Constraint (75e) can be interpreted as guaranteeing that the rate corresponding to Wyner–Ziv coding of S with SI Y does not exceed the common rate of the MAC.

To prove this result, we consider a block Markov coding scheme with backward decoding. The total transmission time is divided into $B + 1$ blocks. Each of the first B blocks is of length n . The length of block $B + 1$ is denoted by \tilde{n} and is set to

$$\tilde{n} = n \cdot \frac{R_0}{\mu}. \quad (79)$$

The blocks are indexed by b , so $b \in [1 : B + 1]$. In the first block, User 1 and User 2 transmit messages at rates R_1 and R_2 without a common message. In Block b , for $b \in [2 : B]$, the users cooperatively transmit a common message at rate R_0 , and superimpose on it their fresh messages at rates R_1 and R_2 . The common message consists of the Wyner–Ziv codeword for describing the channel state sequence in the previous block, Block $b - 1$, with the channel outputs at Block $b - 1$ serving as the decoder's SI. This Wyner–Ziv codeword is independent of the channel state sequence during its transmission. In Block $B + 1$, the users do not transmit any fresh information: they only transmit common information on the state sequence in Block B . The blocklength given by (79) guarantees that this information can be transmitted in block $B + 1$. The decoding proceeds backward: based on the output sequence in block $B + 1$, the decoder decodes the

common message, which consists of a Wyner–Ziv codeword of the state sequence during Block B . This codeword and the channel output sequence of Block B give it some information about the state sequence at Block B . It then uses this information to decode the private messages and the common message of Block B . This common message of Block B and the channel output sequence of Block $B - 1$ give the decoder some information about the state sequence in Block $B - 1$. This process continues until Block 1 is reached. Overall, the users transmit messages at rates

$$(R_1, R_2) \cdot \frac{nB}{nB + \tilde{n}} = (R_1, R_2) \cdot \frac{B}{B + R_0/\mu} \quad (80)$$

which approach (R_1, R_2) as B increases.

We proceed now to a more detailed description of the codebooks and coding scheme.

Codebook Generation: Choose real numbers R_p, R_s, R_0 satisfying

$$0 \leq R_p \leq R_s \quad (81)$$

$$R_p \leq R_0. \quad (82)$$

For each block b , the codebook is constructed in three steps, as described in the following.

- 1) Generate 2^{nR_s} independent length- n codewords

$$\mathbf{v}(j) \in \mathcal{V}^n, \quad j = 1, \dots, 2^{nR_s}$$

each i.i.d. according to the distribution $P_V(\cdot)$ on the finite set \mathcal{V} . Randomly partition the indices $[1 : 2^{nR_s}]$ into 2^{nR_p} bins. Denote by

$$k(j) \in [1 : 2^{nR_p}], \quad j = 1, \dots, 2^{nR_s}$$

the index of the bin to which j belongs, and by $\alpha(k)$ the content of Bin k

$$\alpha(k) = \left\{ j \in [1 : 2^{nR_s}] : k(j) = k \right\}, \quad k = 1, \dots, 2^{nR_p}.$$

- 2) Generate 2^{nR_0} independent length- n codewords

$$\mathbf{u}(k') \in \mathcal{U}^n, \quad k' = 1, 2, \dots, 2^{nR_0}$$

each i.i.d. according to $P_U(\cdot)$.

- 3) For every $k' \in [1 : 2^{nR_0}]$, generate

- a) 2^{nR_1} vectors $\mathbf{x}_1(l|k')$, $l = 1, 2, \dots, 2^{nR_1}$, independent of each other, according to $\prod_{i=1}^n P_{X_1|U}(\cdot|u_i(k'))$.
- b) 2^{nR_2} vectors $\mathbf{x}_2(m|k')$, $m = 1, 2, \dots, 2^{nR_2}$, independent of each other, according to $\prod_{i=1}^n P_{X_2|U}(\cdot|u_i(k'))$.

The codebook generation and partition as described previously are repeated independently B times, with the same distribution and rates. For the last block $B+1$, the codebook that compresses the state is not needed, since the state \mathbf{s}_{B+1} is not sent to the receiver. Therefore, at block $B+1$, Step 1 above can be omitted. Steps 2 and 3 are repeated for the last block, with blocklength \tilde{n}

(as given in (79)) instead of n . For notational convenience, we drop the dependence of the codebooks and codewords length on the block number b . It will be clear from the context whether the length is n (blocks $[1 : B]$) and when it is \tilde{n} (block $B + 1$).

Reveal the codebooks to the encoders and decoder.

Encoding: Let $l_b \in \{1, \dots, 2^{nR_1}\}$ and $m_b \in \{1, \dots, 2^{nR_2}\}$ be the message indices of the users in Block b , and let \mathbf{s}_b be the common state sequence in that block, i.e., $\mathbf{s}_b = (s_{(b-1)n+1}, \dots, s_{bn})$. The operation of the two encoders depend on the block number as follows.

Block 1. The users send $\mathbf{x}_1(l_1|1)$ and $\mathbf{x}_2(m_1|1)$.

Block b , for $b \in [2 : B + 1]$. Both encoders are cognizant of \mathbf{s}_{b-1} . They select the first index $j \in \{1, \dots, 2^{nR_s}\}$ such that

$$(\mathbf{v}(j), \mathbf{s}_{b-1}) \in \mathcal{T}_{V,S} \quad (83)$$

where $\mathcal{T}_{V,S}$ denotes the set of n -length typical pairs (\mathbf{v}, \mathbf{s}) . Denote this index by j_{b-1} . If a vector $\mathbf{v}(j)$ satisfying (83) does not exist, the users pick a default vector, say $j_{b-1} = 1$. Denote by k_b the bin number to which j_{b-1} belongs, i.e., $k(j_{b-1})$. The inputs to the channel now depend on whether we are in the last block or not. For $b \in \{2, \dots, B\}$, the inputs are

$$\begin{aligned} \text{User 1: } & \mathbf{x}_1(l_b|k_b) \\ \text{User 2: } & \mathbf{x}_2(m_b|k_b) \end{aligned} \quad (84)$$

whereas in Block $B + 1$ the inputs are

$$\begin{aligned} \text{User 1: } & \mathbf{x}_1(1|k_{B+1}) \\ \text{User 2: } & \mathbf{x}_2(1|k_{B+1}). \end{aligned} \quad (85)$$

That is, no user messages are sent in the last block.

Decoding: Denote the channel outputs in Block b by

$$\mathbf{y}_b = (y_{(b-1)n+1}, \dots, y_{bn}).$$

Decoding starts at Block $B + 1$ and proceeds backward.

Block $B + 1$. Having observed \mathbf{y}_{B+1} , the decoder looks for an index $\hat{k}_{B+1} \in [1 : 2^{nR_p}]$ such that

$$(\mathbf{x}_1(1|\hat{k}_{B+1}), \mathbf{x}_2(1|\hat{k}_{B+1}), \mathbf{y}_{B+1}) \in \mathcal{T}_{X_1, X_2, Y}. \quad (86)$$

If an index $\hat{k}_{B+1} \in [1 : 2^{nR_p}]$ satisfying (86) does not exist, or is not unique, an error is declared.

Block b , $2 \leq b \leq B$. The decoder has at hand an estimate \hat{k}_{b+1} . It looks in Bin \hat{k}_{b+1} for an index \hat{j}_b such that

$$(\mathbf{v}(\hat{j}_b), \mathbf{y}_b) \in \mathcal{T}_{V,Y}. \quad (87)$$

If there is no such vector in Bin \hat{k}_{b+1} , or there is more than one, an error is declared.

Observe that if decoded correctly, the vector $\mathbf{v}(\hat{j}_b)$ is the compressed state in Block b . This information on the state facilitates the decoding of the messages (l_b, m_b) , and the index k_b , which is the bin number of the state in Block $b - 1$. Specifically, the decoder looks for the indices $(\hat{l}_b, \hat{m}_b, \hat{k}_b)$, $\hat{l}_b \in [1 : 2^{nR_1}]$, $\hat{m}_b \in [1 : 2^{nR_2}]$, $\hat{k}_b \in [1 : 2^{nR_p}]$, such that

$$(\mathbf{x}_1(\hat{l}_b|\hat{k}_b), \mathbf{u}(\hat{k}_b), \mathbf{x}_2(\hat{m}_b|\hat{k}_b), \mathbf{v}(\hat{j}_b), \mathbf{y}_b) \in \mathcal{T}_{X_1, U, X_2, V, Y}. \quad (88)$$

If there is no triple $(\hat{l}_b, \hat{m}_b, \hat{k}_b)$ satisfying (88), or there is more than one such triple, an error is declared.

Block 1. Since it is the first block, there is no need to decode the bin index k . The decoder operates exactly as in Blocks $b \in \{2, \dots, B\}$, except that \hat{k}_1 is set to 1.

The decoder output is the sequence of pairs (\hat{l}_b, \hat{m}_b) , $b = 1, \dots, B$.

Probability of error analysis: Define the error event

$$\mathcal{E} = \bigcup_{b=1}^B \left\{ (\hat{l}_b, \hat{m}_b) \neq (l_b, m_b) \right\}.$$

Denote by K_2^{B+1} the random sequence of bin numbers chosen by the encoder in (84) and (85). We can write

$$P(\mathcal{E}) = \sum_{\tilde{k}_2^{B+1}} P(\mathcal{E} | K_2^{B+1} = \tilde{k}_2^{B+1}) P(K_2^{B+1} = \tilde{k}_2^{B+1}). \quad (89)$$

Since the partitioning is done for every block b independently of other blocks and of the data, we have

$$P(K_2^{B+1} = \tilde{k}_2^{B+1}) = 2^{-nBR_p} \quad \forall \tilde{k}_2^{B+1}. \quad (90)$$

Moreover, by symmetry, the probability of error does not depend on the sequence \tilde{k}_2^{B+1} . Hence, we can assume throughout that a specific sequence k_2^{B+1} is chosen. Similarly, we can assume that the message pair $(l, m) = (1, 1)$ was sent in all blocks. Fix a state sequence $\mathbf{s}^{B+1} = (\mathbf{s}_1, \dots, \mathbf{s}_{B+1})$ and define the events

$$\begin{aligned} \mathcal{A}_{1,b}(\mathbf{s}_b) &= \{(\mathbf{v}(j), \mathbf{s}_b) \in \mathcal{T}_{V,S} \text{ for some } j \in \{1, \dots, 2^{nR_s}\}\} \\ \mathcal{A}_{2,1}(\mathbf{s}_1) &= \{(\mathbf{x}_1(1|1), \mathbf{x}_2(1|1), \mathbf{s}_1) \in \mathcal{T}_{X_1, X_2, S}\} \\ \mathcal{A}_{3,b}(\mathbf{s}_b) &= \{(\mathbf{x}_1(1|k_b), \mathbf{u}(k_b), \mathbf{x}_2(1|k_b), \mathbf{s}_b) \in \mathcal{T}_{X_1, U, X_2, S}\} \\ \mathcal{B}_{1,B+1}(k) &= \{(\mathbf{x}_1(1|k), \mathbf{x}_2(1|k), \mathbf{y}_{B+1}) \in \mathcal{T}_{X_1, X_2, Y}\} \\ \mathcal{B}_{2,b}(j) &= \{(\mathbf{v}(j), \mathbf{y}_b) \in \mathcal{T}_{V,Y}\} \\ \mathcal{B}_{3,b}(l, k, m, j) &= \{(\mathbf{x}_1(l|k), \mathbf{u}(k), \mathbf{x}_2(m|k), \mathbf{v}(j), \mathbf{y}_b) \in \mathcal{T}_{X_1, U, X_2, V, Y}\} \end{aligned}$$

where $b \in [1 : B]$. Further define

$$\begin{aligned} \mathcal{A} &= \mathcal{A}_{2,1}(\mathbf{s}_1) \prod_{b=1}^B \mathcal{A}_{1,b}(\mathbf{s}_b) \prod_{b=2}^{B+1} \mathcal{A}_{3,b}(\mathbf{s}_b) \\ \gamma_{B+1} &= \mathcal{B}_{1,B+1}^c(k_{B+1}) \bigcup_{k \neq k_{B+1}} \mathcal{B}_{1,B+1}(k) \\ \beta_b &= \mathcal{B}_{2,b}^c(j_b) \bigcup_{\substack{j \in \alpha(k_{b+1}) \\ j \neq j_b}} \mathcal{B}_{2,b}(j) \\ \gamma_b &= \mathcal{B}_{3,b}^c(1, k_b, 1, j_b) \bigcup_{(l,k,m) \neq (1,k_b,1)} \mathcal{B}_{3,b}(l, k, m, j_b) \quad (91) \end{aligned}$$

where, again, $b \in [1 : B]$ and with the understanding that for $b = 1$, the value of k_b in the definition of γ_b is set to 1 (see the decoding process at block 1). We can write

$$P(\mathcal{E}) \leq \sum_{\mathbf{s}^{B+1} \in \mathcal{T}_S^{B+1}} P(\mathcal{E} | \mathbf{s}^{B+1}) P_S(\mathbf{s}^{B+1}) + P_S(\mathcal{T}_S^{B+1,c}) \quad (92)$$

where \mathcal{T}_S^{B+1} is the $B + 1$ product of \mathcal{T}_S . It is enough to show that

$$\lim_{n \rightarrow \infty} P(\mathcal{E} | \mathbf{s}^{B+1}) = 0 \quad \forall \mathbf{s}^{B+1} \in \mathcal{T}_S^{B+1}. \quad (93)$$

The probability of error conditioned on a given state sequence can be bounded as follows:

$$\begin{aligned} P(\mathcal{E} | \mathbf{s}^{B+1}) &\leq P(\mathcal{A}^c | \mathbf{s}^{B+1}) + P(\gamma_{B+1} | \mathbf{s}^{B+1}, \mathcal{A}) + \\ &\quad \sum_{b=1}^B [P(\beta_b | \mathbf{s}^{B+1}, \mathcal{A}, \gamma_{b+1}^c) + P(\gamma_b | \mathbf{s}^{B+1}, \mathcal{A}, \gamma_{b+1}^c, \beta_b^c)] \\ &= P(\mathcal{A}^c | \mathbf{s}^{B+1}) + P(\gamma_{B+1} | \mathbf{s}^{B+1}, \mathcal{A}) \\ &\quad + P(\beta_B | \mathbf{s}^{B+1}, \mathcal{A}, \gamma_{B+1}^c) + P(\gamma_B | \mathbf{s}^{B+1}, \mathcal{A}, \gamma_{B+1}^c, \beta_B^c) \\ &\quad + (B-1)[P(\beta_b | \mathbf{s}^{B+1}, \mathcal{A}, \gamma_{b+1}^c) \\ &\quad + P(\gamma_b | \mathbf{s}^{B+1}, \mathcal{A}, \gamma_{b+1}^c, \beta_b^c)] \quad (94) \end{aligned}$$

for some fixed b in the range $[1 : B]$. We proceed to bound each of the terms in (94). We assume throughout that $\mathbf{s}^{B+1} \in \mathcal{T}_S^{B+1}$. For notational convenience, we drop the dependence on \mathbf{s}^{B+1} .

$$\begin{aligned} P(\mathcal{A}^c) &\leq P(\mathcal{T}_{X_1, X_2 | S}^c(\mathbf{s}_1)) + BP(\mathcal{T}_{X_1, U, X_2 | S}^c(\mathbf{s}_1)) \\ &\quad + BP(\mathcal{A}_{1,1}^c(\mathbf{s}_1)). \quad (95) \end{aligned}$$

The third term in the RHS of (95) is the probability of error in lossy coding of the state \mathbf{s}_1 with the vectors $\{\mathbf{v}(j)\}$. By classical results

$$\lim_{n \rightarrow \infty} P(\mathcal{A}^c) = 0 \quad (96)$$

provided

$$R_s > I(V; S). \quad (97)$$

Conditioned on \mathcal{A} , the codewords at the channel input are jointly typical with the state. By classical arguments

$$\lim_{n \rightarrow \infty} P(\gamma_{B+1} | \mathcal{A}) = 0 \quad (98)$$

provided

$$R_p \leq R_0 < \frac{\tilde{n}}{n} I(X_1, X_2; Y). \quad (99)$$

Conditioned on \mathcal{A} and γ_{B+1}^c , $(\mathbf{v}(j_B), \mathbf{s}_B) \in \mathcal{T}_{V,S}$ for some $j_B \in \alpha(k_{B+1})$, and the decoder knows the bin number k_{B+1} . Thus, using standard arguments as in Wyner–Ziv coding [17]

$$\lim_{n \rightarrow \infty} P(\beta_B | \mathcal{A}, \gamma_{B+1}^c) = 0 \quad (100)$$

whenever

$$R_s - R_p < I(V; Y). \quad (101)$$

We proceed to bound the probability of the event γ_B , conditioned on the events $(\mathcal{A}, \gamma_{B+1}^c, \beta_B^c)$. The conditioning on the pair $(\gamma_{B+1}^c, \beta_B^c)$ guarantees that the decoder knows the correct

value of $\mathbf{v}(j_B)$, and the index j_B is held fixed in (91). Decompose the union in (91) as

$$\begin{aligned} & \bigcup_{(l,k,m) \neq (1,k_b,1)} \mathcal{B}_{3,b}(l,k,m,j_b) = \\ & \bigcup_{l \neq 1} \mathcal{B}_{3,b}(l,k_b,1,j_b) \bigcup_{k \neq k_b} \mathcal{B}_{3,b}(1,k,1,j_b) \\ & \bigcup_{m \neq 1} \mathcal{B}_{3,b}(1,k_b,m,j_b) \bigcup_{l \neq 1, m \neq 1} \mathcal{B}_{3,b}(l,k_b,m,j_b) \\ & \bigcup_{l \neq 1, k \neq k_b} \mathcal{B}_{3,b}(l,k,1,j_b) \bigcup_{k \neq k_b, m \neq 1} \mathcal{B}_{3,b}(1,k,m,j_b) \\ & \bigcup_{l \neq 1, k \neq k_b, m \neq 1} \mathcal{B}_{3,b}(l,k,m,j_b). \end{aligned}$$

By standard methods

$$\lim_{n \rightarrow \infty} P(\gamma_B | \mathcal{A}, \gamma_{B+1}^c, \beta_B^c) = 0 \quad (102)$$

whenever the following inequalities are satisfied:

$$R_1 < I(X_1; Y | X_2, U, V) \quad (103)$$

$$R_0 < I(U, X_1, X_2; Y | V)$$

$$R_2 < I(X_2; Y | X_1, U, V) \quad (104)$$

$$R_1 + R_2 < I(X_1, X_2; Y | U, V) \quad (105)$$

$$R_0 + R_1 < I(X_1, X_2, U; Y | V)$$

$$R_0 + R_2 < I(X_1, X_2, U; Y | V)$$

$$R_0 + R_1 + R_2 < I(X_1, U, X_2; Y | V). \quad (106)$$

In view of (76), inequalities (103)–(106) dominate the other aforementioned inequalities, and (106) is equivalent to

$$R_0 + R_1 + R_2 < I(X_1, X_2; Y | V). \quad (107)$$

The last two terms in (94), $P(\beta_b | \mathbf{s}^{B+1}, \mathcal{A}, \gamma_{b+1}^c)$ and $P(\gamma_b | \mathbf{s}^{B+1}, \mathcal{A}, \gamma_{b+1}^c, \beta_b^c)$, are treated exactly as $P(\beta_B | \mathbf{s}^{B+1}, \mathcal{A}, \gamma_{B+1}^c)$ and $P(\gamma_B | \mathbf{s}^{B+1}, \mathcal{A}, \gamma_{B+1}^c, \beta_b^c)$. The only difference between these terms is in the definitions of γ_{B+1} and γ_b , but this does not add any difficulty.

We now collect terms. First, note that (99) is always satisfied due to (79) and the choice of μ (74). Choose

$$R_0 > I(V; S) - I(V; Y).$$

Then, (97) and (101) are satisfied for some $R_p < \min\{R_0, R_s\}$. Therefore, the following constraints suffice to guarantee (93):

$$R_1 < I(X_1; Y | X_2, U, V) \quad (108)$$

$$R_2 < I(X_2; Y | X_1, U, V) \quad (109)$$

$$R_1 + R_2 < I(X_1, X_2; Y | U, V) \quad (110)$$

$$R_0 + R_1 + R_2 < I(X_1, X_2; Y | V) \quad (111)$$

$$R_0 > I(V; S) - I(V; Y). \quad (112)$$

In addition, (96) guarantees that the input constraints are satisfied. This completes the proof of Theorem 1. ■

H) Analysis of Example 2: The collection of rate pairs that satisfy (19) is the capacity region of the same MAC when the

encoders are uninformed of S but can fully cooperate. Consequently, by Proposition 1, this is an outer bound on the capacity of our MAC with strictly causal SI and no cooperation. To conclude the analysis, it thus only remains to prove that every rate pair satisfying (19) is achievable. We show this in two different ways.

The first is by judiciously choosing the random variables in (15). We first examine the maximal rate for Encoder 1. To this end, we choose $U = X_2$, and we choose (X_1, X_2, V, S) to be zero-mean jointly Gaussian with (X_1, X_2) independent of (V, S) and with X_1 and X_2 of variance Γ_1 and Γ_2 . The correlation between X_1 and X_2 and between V and S will be addressed shortly. With these choices, (15b) simplifies to $R_2 = 0$; (15a) and (15c) simplify to

$$R_1 \leq \frac{1}{2} \log \left(\frac{\sigma_{x_1|x_2}^2 + \sigma_{s|v}^2}{\sigma_{s|v}^2} \right) \quad (113)$$

and (15d) simplifies to

$$R_1 \leq \frac{1}{2} \log \left(\frac{\Gamma_\Sigma + \sigma_s^2}{\sigma_s^2} \right) \quad (114)$$

where $\sigma_{x_1|x_2}^2$ is the variance of X_1 conditioned on X_2 , where $\sigma_{s|v}^2$ is the variance of S given V , and where Γ_Σ is the power in the sum $X_1 + X_2$

$$\Gamma_\Sigma = \mathbb{E}[(X_1 + X_2)^2]. \quad (115)$$

Given an arbitrary $\delta > 0$, we can choose $\sigma_{x_1|x_2}^2$ to be strictly positive but small enough so that the RHS of (114) will be within δ of the RHS of (19). Since we have chosen $\sigma_{x_1|x_2}^2$ strictly positive, we can now choose $\sigma_{s|v}^2$ small enough (but positive) so that the RHS of (113) will exceed the RHS of (19). With this choice of $\sigma_{s|v}^2$, the inequality (113) is inactive, and the RHS of (114) is achievable for R_1 . Since this RHS is within δ of the RHS of (19), this demonstrates that the rate pair $(R_1, 0)$ where R_1 is the RHS of (19) is achievable. By symmetry, we can also achieve the rate pair $(0, R_2)$, where R_2 is the RHS of (19). A time-sharing argument now concludes the proof.

The second way of establishing that rate pairs satisfying (19) are achievable is based on a Schalkwijk–Kailath type algorithm [12] that achieves the rate region (19) with a double-exponential decay of the maximal probability of error. As earlier, we first prove the achievability of the rate pair

$$\left(\frac{1}{2} \log \left(1 + \frac{\Gamma_1 + \Gamma_2 + 2\sqrt{\Gamma_1\Gamma_2}}{\sigma_s^2} \right), 0 \right). \quad (116)$$

The achievability of the region (19) then follows by symmetry and time sharing. Split the interval $[0, 1]$ into ν_1 equally spaced subintervals. Let θ_1 be the center of one of these subintervals, representing the message of User 1, as in [12]. At the first time instance, User 1 transmits θ_1 and User 2 is silent so

$$X_{1,1} = \theta_1, \quad X_{2,1} = 0 \quad (117)$$

and the channel output Y_1 is

$$Y_1 = \theta_1 + S_1. \quad (118)$$

As of time instance $i = 2$, the first noise sample S_1 is known to both encoders, who proceed to transmit it to the decoder with full cooperation. Since as of this point the encoders are transmitting a common message, knowing the states in a strictly causal manner is equivalent to having causal feedback. The users can thus send S_1 by mimicking the Schalkwijk–Kailath algorithm for the single-user Gaussian with feedback and power $\Gamma_1 + \Gamma_2 + 2\sqrt{\Gamma_1\Gamma_2}$.

Applying this algorithm, it follows from [12] that after n channel uses the receiver can construct a maximum likelihood estimate \hat{S}_1 for S_1 , whose error satisfies

$$\mathbb{E} \left[(S_1 - \hat{S}_1^{(n)})^2 | S_1 \right] = \frac{\sigma_s^2}{(\alpha^2)^n} \quad (119)$$

where

$$\alpha^2 = 1 + \gamma^2, \quad \gamma^2 = \frac{(\Gamma_1^{\frac{1}{2}} + \Gamma_2^{\frac{1}{2}})^2}{\sigma_s^2}. \quad (120)$$

Based on this estimate of S and on Y_1 , the decoder can now estimate θ_1 by

$$\hat{\theta}_1 = Y_1 - \hat{S}_1^{(n)} \quad (121)$$

with error

$$\mathbb{E} \left[(\theta_1 - \hat{\theta}_1^{(n)})^2 | S_1 \right] = \frac{\sigma_s^2}{(\alpha^2)^n}. \quad (122)$$

Thus, if we choose $\nu_1 = nr_1$ with $r_1 \leq \frac{1}{2} \log(1 + \gamma^2)$, then the probability of error will vanish doubly exponentially as $n \rightarrow \infty$. This proves the achievability of the rate pair (116). Using symmetry and a time-sharing argument proves that this algorithm achieves the entire region (19).

We next show that the region (19) is achievable also under the weaker assumption that only the time-one state S_1 is revealed to the encoders strictly causally, i.e., that S_1 is known to the encoders as of time $i = 2$. We begin by showing that even under this weaker assumption, we can achieve the rate pair (116). For typographical reasons, we now denote the blocklength by $n + 1$, so after the first channel use we still have n remaining channel uses. As earlier, User 1 maps its message to some number θ_1 in a manner that will be explained later. At time $i = 1$, the transmitted symbols and the corresponding output are as in (117) and (118). As of time $i = 2$, the time-one state S_1 is known to both encoders, who cooperate in transmitting it to the receiver as follows. Given any $\epsilon > 0$, we first pick some $\beta > 0$ sufficiently large so that

$$\Pr[|S_1| \geq \beta] \leq \epsilon. \quad (123)$$

Note that

$$(|S_1| < \beta) \Rightarrow \left(\frac{1}{2} \left(\frac{S_1}{\beta} + 1 \right) \in (0, 1) \right). \quad (124)$$

As of time $i = 2$, both encoders know S_1 and they cooperate in sending a quantized version of it. More specifically, they use the remaining channel inputs to send the first nr_1 bits to the right of the period in the (nonterminating) binary representation of

$$\frac{1}{2} \left(\frac{S_1}{\beta} + 1 \right). \quad (125)$$

(This corresponds to uniformly quantizing (125) using nr_1 bits.) To send these bits, they use a Gaussian codebook that achieves the capacity of the single-user Gaussian channel under the maximal error probability criterion. Thus, for sufficiently large n , the number of bits that can be sent reliably is nr_1 , where r_1 is any rate satisfying

$$r_1 < \frac{1}{2} \log(1 + \gamma^2). \quad (126)$$

Based on these bits, the decoder can determine $(\beta^{-1}S_1 + 1)/2$ to within an error of 2^{-nr_1-1} whenever $|S_1| < \beta$, i.e., it can form an estimate \hat{S}_1 such that

$$(|S_1| < \beta) \Rightarrow \left(|\hat{S}_1 - S_1| < \beta 2^{-nr_1} \right). \quad (127)$$

The decoder then guesses θ_1 based on $Y_1 - \hat{S}_1$. To make sure that this guess is correct whenever $|S_1|$ is smaller than β and the nr_1 bits describing (125) are correctly decoded, we allow θ_1 to take on only the values

$$\left(\kappa + \frac{1}{2} \right) 2^{-\nu}, \quad \kappa = 0, \dots, 2^\nu - 1 \quad (128)$$

where ν is chosen as

$$\nu = \lceil nr_1 + \log_2(2\beta)^{-1} \rceil - 1 \quad (129)$$

so as to guarantee that the distance between the nearest values of θ_1 , i.e., $2^{-\nu}$, is greater than $2\beta 2^{-nr_1}$. With this constellation for θ_1 , we can send ν bits corresponding to the first ν bits in the binary expansion of the transmitted θ_1 . The rate is thus

$$\frac{\lceil nr_1 + \log_2(2\beta)^{-1} \rceil - 1}{n + 1} \quad (130)$$

which converges to r_1 , as n tends to infinity. Thus, the pair (116) is achievable.

We next explain how to achieve other rate pairs of the same sum as the rate pair (116). Recall that, in our scheme, θ_1 conveys ν bits corresponding to the first ν bits in the binary expansion of θ_1 . Suppose now that we distribute these ν bits among the two users by expressing ν as $\nu_1 + \nu_2$ for nonnegative integers ν_1 and ν_2 , and by allowing User 1 to send ν_1 bits and User 2 to send ν_2 bits as follows. The first symbol sent by User 1 is

$$0. \overbrace{d_1^{(1)} d_1^{(2)} \dots d_1^{(\nu_1)}}^{\nu} \underbrace{00 \dots 0}_{\nu_2} 1$$

where $d_1^{(1)}, \dots, d_1^{(\nu_1)}$ are the ν_1 bits it wishes to convey. The first symbol sent by User 2 is

$$0. \underbrace{00 \dots 0}_{\nu_1} \overbrace{d_2^{(1)} d_2^{(2)} \dots d_2^{(\nu_2)}}^{\nu} 0$$

where $d_2^{(1)}, \dots, d_2^{(\nu_2)}$ are the ν_2 bits it wishes to convey. Since the channel adds the inputs, the time-one received symbol Y_1 is

$$Y_1 = 0. d_1^{(1)} d_1^{(2)} \dots d_1^{(\nu_1)} d_2^{(1)} d_2^{(2)} \dots d_2^{(\nu_2)} 1 + S_1.$$

As earlier, the transmitters use the next n channel uses to describe the quantization of S_1 . The decoder can then estimate S_1 and with high probability recover

$$0.d_1^{(1)}d_1^{(2)}\dots d_1^{(\nu_1)}d_2^{(1)}d_2^{(2)}\dots d_2^{(\nu_2)}1$$

from which it can recover the data bits. With this scheme, all rates summing to $0.5 \log(1 + \gamma^2)$ are achievable, and only the first state S_1 is needed.

1) Sketch of the Proof of Theorem 2: To conclude the sketch of the achievability of the improved inner bound, we still need to describe how the block Markov scheme is terminated. We thus assume that B blocks have been transmitted, and we proceed to describe Blocks $B + 1$, $B + 2$, and $B + 3$. We think about these blocks as “overhead,” because they contain no fresh information. Fortunately, this overhead does not affect the throughput because we can choose B very large.

The next lemma shows that if the full-cooperation capacity of the MAC without SI is zero, then the improved inner bound contains only the rate pair $(0, 0)$ and is thus trivially an inner bound.

Lemma 1: If the capacity of the MAC without any SI but with full cooperation is zero, i.e., if

$$\max_{P_{X_1, X_2}} I(X_1, X_2; Y) = 0 \quad (131)$$

then the improved inner bound $\mathcal{R}_{s-c}^{\text{com},*}$ contains only the all-zero rate tuple.

Proof: It follows from (25g) and (24) that if (R_1, R_2) is in $\mathcal{R}_{s-c}^{\text{com},*}$, then for some joint distribution of the form (20)

$$\begin{aligned} R_1 + R_2 &\leq I(X_1, X_2; Y, V_1, V_2, V) - I(X_1, X_2, S; V_1, V_2, V|Y) \\ &= I(X_1, X_2; Y) + I(X_1, X_2; V_1, V_2, V|Y) \\ &\quad - I(X_1, X_2, S; V_1, V_2, V|Y). \end{aligned}$$

Consequently, if (131) holds and $I(X_1, X_2; Y)$ is hence zero, then $R_1 + R_2$ must be upper bounded by $I(X_1, X_2; V_1, V_2, V|Y) - I(X_1, X_2, S; V_1, V_2, V|Y)$, which is nonpositive (irrespective of the joint PMF). ■

In view of Lemma 1, it only remains to prove the achievability of the improved inner bound when the full-cooperation capacity without SI is positive. We henceforth therefore assume

$$\max_{P_{X_1, X_2}} I(X_1, X_2; Y) > 0. \quad (132)$$

The next lemma shows that we can also assume that the channel between Transmitter 1 (uninformed) and the receiver (informed) is of positive capacity and likewise from Transmitter 2.

Lemma 2: If the channel between Transmitter 1 (uninformed) to the receiver (informed) is of zero capacity, i.e., if

$$\max_{x_2 \in \mathcal{X}_2} \max_{P_{X_1}} I(X_1; Y, S|X_2 = x_2) = 0 \quad (133)$$

then the improved inner bound $\mathcal{R}_{s-c}^{\text{com},*}$ contains only rate pairs (R_1, R_2) with $R_1 = 0$ and $R_2 \leq \max I(X_2; Y)$. An analogous result holds if

$$\max_{x_1 \in \mathcal{X}_1} \max_{P_{X_2}} I(X_2; Y, S|X_1 = x_1) = 0. \quad (134)$$

Proof: We first prove that if a rate pair (R_1, R_2) is in $\mathcal{R}_{s-c}^{\text{com},*}$, and if (133) holds, then R_1 must be zero. Fix some joint distribution of the form (20) and let (R_1, R_2) satisfy the inequalities of Theorem 2. We next argue that Hypothesis (133) implies

$$I(X_1; Y, V_2, V|X_2, U) = 0. \quad (135)$$

Indeed

$$I(X_1; Y, V_2, V|X_2, U) \leq I(X_1; Y, V_2, V|X_2, U, S) \quad (136a)$$

$$= I(X_1; Y|X_2, U, S, V_2, V) \quad (136b)$$

$$= I(X_1; Y|X_2, U, S) \quad (136c)$$

$$= I(X_1; Y, S|X_2, U) \quad (136d)$$

$$\leq \max_{u \in \mathcal{U}} \max_{x_2 \in \mathcal{X}_2} I(X_1; Y, S|X_2 = x_2, U = u) \quad (136e)$$

$$\leq \max_{u \in \mathcal{U}} \max_{x_2 \in \mathcal{X}_2} \max_{P_{X_1|U=u}} I(X_1; Y, S|X_2 = x_2, U = u) \quad (136f)$$

$$= \max_{x_2 \in \mathcal{X}_2} \max_{P_{X_1}} I(X_1; Y, S|X_2 = x_2) \quad (136g)$$

where the first line follows from

$$X_1 - \circ - (X_2, U) - \circ - S \quad (137)$$

the second from the chain rule and because

$$X_1 - \circ - (X_2, U, S) - \circ - (V_2, V) \quad (138)$$

so $I(X_1; V_2, V|X_2, U, S)$ is zero; the third from

$$(X_1, Y) - \circ - (X_2, U, S) - \circ - (V_2, V) \quad (139)$$

the fourth again by (137); the fifth by upper bounding the average by the maximal value; the sixth by maximizing over the conditional distribution of X_1 given $U = u$; and the last because the maximization over u on the RHS of (136f) is unnecessary.

Continuing our proof that R_1 must be zero, we note that (21) and (25a) imply

$$\begin{aligned} R_1 &\leq I(X_1; Y, V_1, V_2, V|X_2, U) - I(X_1, S; V_1|V, V_2, Y) \\ &= I(X_1; Y, V_2, V|X_2, U) + I(X_1; V_1|X_2, U, Y, V_2, V) \\ &\quad - I(X_1, S; V_1|V, V_2, Y) \end{aligned} \quad (140)$$

$$\begin{aligned} &= I(X_1; V_1|X_2, U, Y, V_2, V) - I(X_1, S; V_1|V, V_2, Y) \\ &= H(V_1|X_2, U, Y, V_2, V) - H(V_1|X_1, X_2, U, Y, V_2, V) \\ &\quad + H(V_1|X_1, S, V, V_2, Y) - H(V_1|V, V_2, Y) \\ &= H(V_1|X_2, U, Y, V_2, V) - H(V_1|X_1, X_2, U, Y, V_2, V) \\ &\quad + H(V_1|X_1, S) - H(V_1|V, V_2, Y) \\ &\leq 0 \end{aligned} \quad (141)$$

where the second equality (third line) follows from (135); the last equality because (20) implies that, conditional on (X_1, S) , the auxiliary random variable V_1 is independent of (X_2, U, Y, V_2, V) ; and where in the last inequality we have used

$$H(V_1|X_2, U, Y, V_2, V) \leq H(V_1|V, V_2, Y)$$

(conditioning reduces entropy) and

$$H(V_1|X_1, X_2, U, Y, V_2, V) \geq H(V_1|X_1, S)$$

which can be argued as follows:

$$\begin{aligned} H(V_1|X_1, X_2, U, Y, V_2, V) &\geq H(V_1|X_1, S, X_2, U, Y, V_2, V) \\ &= H(V_1|X_1, S) \end{aligned}$$

where the first inequality is because conditioning cannot increase entropy, and the second by (20), which implies that, conditional on (X_1, S) , the auxiliary random variable V_1 is independent of (X_2, U, Y, V_2, V) .

Having established that R_1 is zero, we now conclude from (25g) and (24)

$$\begin{aligned} R_2 &= R_1 + R_2 \\ &\leq I(X_1, X_2; Y, V_1, V_2, V) - I(X_1, X_2, S; V_1, V_2, V|Y) \\ &= I(X_1, X_2; Y) \\ &\quad + I(X_1, X_2; V_1, V_2, V|Y) - I(X_1, X_2, S; V_1, V_2, V|Y) \\ &\leq I(X_1, X_2; Y) \\ &= I(X_2; Y) + I(X_1; Y|X_2) \\ &= I(X_2; Y) \end{aligned} \tag{142} \tag{143} \tag{144}$$

where the last equality follows from (133). ■

Lemma 2 shows that if either (133) or (134) holds, then the improved inner bound is achievable. It thus only remains to prove its achievability when

$$\max_{x_2 \in \mathcal{X}_2} \max_{P_{X_1}} I(X_1; Y, S|X_2 = x_2) > 0 \tag{145}$$

and

$$\max_{x_1 \in \mathcal{X}_1} \max_{P_{X_2}} I(X_2; Y, S|X_1 = x_1) > 0 \tag{146}$$

both of which we now assume.

With the aid of (132), (145), and (146), we can now describe the termination of the block Markov scheme. This is done in three blocks $B+1$, $B+2$, and $B+3$ whose length is some finite multiple of n . Block $B+1$ is split into two parts. In the first, Transmitter 1 sends the \mathbf{v}_1 -sequence of Block B assuming that the receiver knows the state sequence \mathbf{s} of Block $B+1$. This can be done (under this assumption) by (145). In the second, Transmitter 2 sends the \mathbf{v}_2 -sequence of Block B assuming that the receiver knows the state of Block $B+1$. This is possible by (146). In Block $B+2$, the transmitters cooperate to send the sequence \mathbf{s} of Block $B+1$ (this is possible by (132)), and in

Block $B+3$ they cooperate to send the \mathbf{v} sequence of Block B (again possible by (132)).

Decoding is performed as follows. The decoder first decodes Block $B+3$ without any SI and thus learns the sequence \mathbf{v} of Block B . It then decodes Block $B+2$ (again without any SI) and learns the state sequence \mathbf{s} of Block $B+1$. Now that it knows the state sequence of Block- $B+1$, it can decode that block and learn the \mathbf{v}_1 -sequence and the \mathbf{v}_2 -sequence of Block B . From here on, it can proceed with the regular backward decoding: in decoding Block b , it knows the sequences \mathbf{v} , \mathbf{v}_1 , and \mathbf{v}_2 of Block b and it can therefore decode the common message and the messages transmitted by each of the transmitters in Block b . From this decoding, it learns the private messages of Block b , and the sequences \mathbf{v} , \mathbf{v}_1 , and \mathbf{v}_2 of Block $b-1$.

J) Analysis of Example 3: We first prove (29). To this end, we note that if $(1, R_2)$ is achievable, then R_2 cannot exceed $1/2$. This can be shown using the full-cooperation outer bound (Proposition 1), which implies that (R_1, R_2) can only be achievable if $R_1 + R_2 \leq 3/2$. Of more interest to us is the fact that the rate pair $(1, 1/2)$ is achievable. We demonstrate this using the improved inner bound. Indeed, it is straightforward to verify that setting

$$R_0^{(1)} = R_0 = 0, \quad R_0^{(2)} = 1/2 \tag{147a}$$

$$V = V_1 = 0, \quad V_2 = S_{X_2} \tag{147b}$$

$$U = 0 \tag{147c}$$

$$X_1, X_2 \sim \text{i.i.d. Bernoulli } 1/2 \tag{147d}$$

and

$$(R_1, R_2) = (1, 1/2) \tag{147e}$$

satisfies all the required inequalities. This choice corresponds to the following block Markov scheme: in the block Markov scheme, Transmitter 1 sends its data uncoded. At Block $b+1$, Transmitter 2 sends n bits, half of which are fresh data bits and half of which are used to describe the n -length sequence \mathbf{s}_{x_2} of the previous block. Note that Transmitter 2 does not describe the entire state sequence \mathbf{s} of the previous block but only \mathbf{s}_{x_2} . This latter sequence is known to Transmitter 2 at the beginning of Block $b+1$ thanks to the strictly causal state information and because it knows the sequence \mathbf{x}_2 it transmitted in the previous block. And $n/2$ bits suffice to describe this sequence because S_{X_2} is of entropy $1/2$.

We now turn to proving (28). We fix some distribution $P_{U, V, X_1, X_2, S, Y}$ of the form (14), we assume that $(R_1 = 1, R_2)$ satisfy inequalities (15), and we then prove that R_2 must be zero. Since $R_1 = 1$ and since \mathcal{X}_1 is binary, inequality (15a) must hold with equality, and X_1 must be independent of (X_2, U, V) . By (14), this implies that

$$X_1 \text{ is independent of } (X_2, U, V, S). \tag{148a}$$

From (15a) (that we know holds with equality) and the fact that $R_1 = 1$, we also infer that

$$\begin{aligned} 1 &= H(Y|X_2, U, V) - H(Y|X_1, X_2, U, V) \\ &= H(Y_1|X_2, U, V) - H(Y_1|X_1, X_2, U, V) \end{aligned} \tag{148b}$$

where the second equality holds because Y_2 is a deterministic function of X_2 . Since Y_1 is binary, $H(Y_1|X_2, U, V)$ is upper bounded by 1, and we conclude from (148b) that

$$\begin{aligned} 0 &= H(Y_1|X_1, X_2, U, V) \\ &= H(Y_1 \oplus X_1|X_1, X_2, U, V) \\ &= H(S_{X_2}|X_1, X_2, U, V) \\ &= H(S_{X_2}|X_2, U, V) \end{aligned} \quad (148c)$$

where the last equality follows from (148a).

We next show that

$$U \circ - (X_2, V) \circ - S_{X_2}. \quad (148d)$$

To this end, we note that, by (14), the pair (V, S) is independent of (U, X_2) and hence

$$U \circ - (X_2, V) \circ - S. \quad (148e)$$

Since S_{X_2} is a deterministic function of (X_2, V, S) , this implies (148d), because if $A \circ - B \circ - C$ forms a Markov chain, then $A \circ - B \circ - f(B, C)$. Having established (148d), we now obtain from (148c)

$$H(S_{X_2}|X_2, V) = 0. \quad (148f)$$

We now focus on the case where X_2 is not deterministic

$$\Pr[X_2 = \eta] > 0, \quad \eta \in \{0, 1\} \quad (148g)$$

because if X_2 is deterministic then R_2 must be zero by (15b). We also assume that the PMF of V is strictly positive

$$\Pr[V = v] > 0, \quad v \in \mathcal{V} \quad (148h)$$

because outcomes of the auxiliary random variable that have zero probability can be removed from \mathcal{V} without affecting the inner bound. Since, by (14), V is independent of X_2 , it follows from (148g) and (148h) that

$$\Pr[X_2 = \eta, V = v] > 0, \quad \eta \in \{0, 1\}, v \in \mathcal{V}. \quad (148i)$$

This and (148f) imply that

$$H(S_\eta|X_2 = \eta, V = v) = 0, \quad \eta \in \{0, 1\}, v \in \mathcal{V}. \quad (148j)$$

Since, by (14), X_2 is independent of (V, S) and, *a fortiori*, of (V, S_η) , it follows from (148j) that

$$H(S_\eta|V = v) = 0, \quad \eta \in \{0, 1\}, v \in \mathcal{V}. \quad (148k)$$

Thus, $H(S_\eta|V) = 0$, and since $S = (S_0, S_1)$

$$H(S|V) = 0. \quad (148l)$$

Consequently

$$\begin{aligned} I(V; S) &= H(S) \\ &= 1 \end{aligned} \quad (148m)$$

where the second equality follows from (26) and the independence of S_0 and S_1 . From (148m), (15d), and the fact that \mathcal{Y} has four elements, we then conclude that $R_1 + R_2 \leq 1$. This combines with $R_1 = 1$ to establish that R_2 must be zero.

*K) Termination of the Block Markov Scheme for $\mathcal{R}_{s-c}^{\text{Com},**}$.* The termination of the block Markov scheme for $\mathcal{R}_{s-c}^{\text{Com},**}$ is identical to that for $\mathcal{R}_{s-c}^{\text{Com},*}$. Since the proof of Lemma 1 only relies on (24), (25g), and on the properties of mutual information, and since (30d) is identical to (25g), Lemma 1 also holds when we replace $\mathcal{R}_{s-c}^{\text{Com},*}$ with $\mathcal{R}_{s-c}^{\text{Com},**}$. Consequently, we only have to consider the case where (132) holds.

We next show that Lemma 2 too continues to hold when we replace $\mathcal{R}_{s-c}^{\text{Com},*}$ with $\mathcal{R}_{s-c}^{\text{Com},**}$. To see this, we first note that (133) implies (135) also for all PMFs of the form (31). Indeed, the proof in Section I in the Appendix that (133) implies (135) relies only on the basic properties of mutual information and on the Markov relations (137)–(139), which also hold for PMFs of the form (31).

The proof that R_1 must be zero also if we replace $\mathcal{R}_{s-c}^{\text{Com},*}$ with $\mathcal{R}_{s-c}^{\text{Com},**}$ in the hypothesis of Lemma 2 (cf., (141)) requires some slight alteration: rather than using (21) and (25a) to infer (140), we now use (21) and (30a) to replace (140) with

$$\begin{aligned} R_1 &\leq I(X_1; Y, V_1, V_2, V|X_2, U) - I(X_1, X_2, S; V_1|V, V_2, Y) \\ &= I(X_1; Y, V_2, V|X_2, U) + I(X_1; V_1|X_2, U, Y, V_2, V) \\ &\quad - I(X_1, X_2, S; V_1|V, V_2, Y) \\ &= I(X_1; V_1|X_2, U, Y, V_2, V) - I(X_1, X_2, S; V_1|V, V_2, Y) \\ &= H(V_1|X_2, U, Y, V_2, V) - H(V_1|X_1, X_2, U, Y, V_2, V) \\ &\quad + H(V_1|X_1, X_2, S, V, V_2, Y) - H(V_1|V, V_2, Y) \\ &= H(V_1|X_2, U, Y, V_2, V) - H(V_1|V, V_2, Y) \\ &\quad + H(V_1|X_1, X_2, S, V, V_2, Y) \\ &\quad - H(V_1|X_1, X_2, U, Y, V_2, V) \\ &\leq H(V_1|X_1, X_2, S, V, V_2, Y) - H(V_1|X_1, X_2, U, Y, V_2, V) \\ &\leq 0 \end{aligned}$$

where the third line follows from (135); the sixth because conditioning cannot increase entropy; and the seventh can be justified as follows:

$$\begin{aligned} &H(V_1|X_1, X_2, U, Y, V_2, V) \\ &\geq H(V_1|X_1, X_2, U, Y, V_2, V, S) \\ &= H(V_1|X_1, V, S) \\ &\geq H(V_1|X_1, X_2, S, V, V_2, Y) \end{aligned} \quad (149)$$

where the first line holds because conditioning (on S) cannot increase entropy; the second because the PMFs of the form (31) satisfy the Markov condition

$$V_1 \circ - (S, X_1, V) \circ - (X_2, U, Y, V_2, V)$$

and the third because conditioning cannot increase entropy.

To conclude the proof that Lemma 2 continues to hold when we replace $\mathcal{R}_{s-c}^{\text{Com},*}$ with $\mathcal{R}_{s-c}^{\text{Com},**}$, it remains to show that, under the lemma's hypothesis, R_2 must satisfy (144). The derivation is identical to the one in the proof of the lemma in Section I in

the Appendix, because in deriving (144) we only used the fact that R_1 must be zero, (24), (25g) (which is identical to (30d)), the properties of mutual information, and (133).

L) *Proof of Theorem 3:*

As in the proof of Theorem 1, we write the region (36a)–(36e) as

$$R_1 \leq I(U_1; Y|U_2, U, V) \quad (150)$$

$$R_2 \leq I(U_2; Y|U_1, U, V) \quad (151)$$

$$R_1 + R_2 \leq I(U_1, U_2; Y|U, V) \quad (152)$$

$$R_0 + R_1 + R_2 \leq I(U_1, U, U_2; Y|V) \quad (153)$$

$$R_0 \geq I(V; S) - I(V; Y) \quad (154)$$

$$\Gamma_k \geq \mathbb{E}[\phi_k(X_k)], \quad k = 1, 2. \quad (155)$$

To prove Theorem 3, it is enough to show that any pair (R_1, R_2) satisfying (150)–(154) for some $R_0 \geq 0$ is achievable. We proceed along the lines of the proof of the strictly causal case, with minor modifications in the construction of the codewords. Specifically, new external random variables U_1 and U_2 are introduced, with joint law (34). The codewords $\mathbf{x}_1(l|k)$ and $\mathbf{x}_2(m|k)$ are replaced by $\mathbf{u}_1(l|k)$ and $\mathbf{u}_2(m|k)$, generated by $\prod_{i=1}^n P_{U_1|U}(\cdot|u_i(k))$ and $\prod_{i=1}^n P_{U_2|U}(\cdot|u_i(k))$, respectively. The input word of User 1 is now generated by $\prod_{i=1}^n P_{X_1|U, U_1, S}(\cdot|u_i(l|k), u_{1,i}(k), s_i)$, and similarly for User 2. The advantage of this construction over the code construction for the strictly causal case is that the input at time i can depend on s_i . From this point, we follow the arguments leading to (106), with U_1 and U_2 replacing X_1 and X_2 . Note that now

$$I(U_1, U, U_2; Y|V) \geq I(U_1, U_2; Y|V)$$

and therefore we do not drop the random variable U in (153), as it will lead to lower achievable rates. Regarding the termination: as in the strictly causal case, it can be easily shown that the region defined via (34) and (36) is contained in the full-cooperation region of the causal model

$$R_1 + R_2 \leq \max I(F; Y) \quad (156)$$

where the maximization is over all joint distributions

$$P_S P_F P_{X_1|S, F} P_{X_2|S, F}.$$

Hence, we can assume that the RHS of (156) is strictly positive, as otherwise $\mathcal{R}_{\text{cau}}^{\text{com}}$ is trivially achievable. If the RHS of (156) is strictly positive, then the users can cooperate in Block $B + 1$ to transmit the compressed state V of Block B without any fresh information. The details are omitted. ■

M) *Analysis of Example 4:* We begin by proving (39), i.e., that $\mathcal{R}_{\text{cau}}^{\text{com}}$ contains the rate pair

$$\left(\min\{1 - p, 1 - H_b(p)\}, 0 \right). \quad (157)$$

To see this, consider the choice

$$V = S, \quad U_2 = 0, \quad X_1 = U_1, \quad X_2 = U$$

with U_1, U , and S being independent and with U_1 and U taking on the values 0 and 1 equiprobably. With this choice, (36b) reduces to $R_2 = 0$ and consequently (36a) and (36c) reduce to

$$\begin{aligned} R_1 &\leq H(Y|U, S) \\ &= \Pr[S = 1]H(Y|U, S = 1) \\ &= (1 - p) \end{aligned}$$

and (36d) to

$$\begin{aligned} R_1 &\leq H(Y) - H(S) \\ &= 1 - H_b(p). \end{aligned}$$

This establishes that the rate pair in (157) is in $\mathcal{R}_{\text{cau}}^{\text{com}}$. It is interesting to note that when $p > H_b(p)$, the rate pair $(1 - p, 0)$ is achievable. This shows that the highest rate for User 1 in this case is $(1 - p)$ because this rate cannot be exceeded by User 1 even when the receiver knows the state S .

The Naïve Approach: From the region (38) and the characterization of the extreme points of the capacity region of the classical MAC, the maximal rate that User 1 can transmit is

$$\max I(T_1; Y|T_2 = t_2) \quad (158)$$

where the maximum is over the distribution of T_1 and over all the strategies (mappings) $t_2: \mathcal{S} \rightarrow \mathcal{X}_2$. For this MAC, the law $P_{Y|T_1, T_2}(y|t_1, t_2)$ does not depend on the value to which t_2 maps the state $s = 1$, i.e., the strategy t_2 influences the output only when $S = 2$, in which case it gives a certain input X_2 , connected directly to the output. User 1 is then disconnected. Therefore, the exact value of t_2 is immaterial. Without loss of generality, we thus assume that $t_2(s)|_{s=2} = 0$.

Similarly, t_1 influences the output only when $S = 1$, in which case it gives a certain input X_1 directly connected to the output. Since the strategies are chosen independently of S , the MAC reduces to a Z -channel from User 1

$$P(Y = 0|X = 0) = 1, \quad P(Y = 0|X = 1) = p. \quad (159)$$

The capacity of the Z -channel is

$$\log_2 \left(1 + (1 - p)p^{\frac{p}{1-p}} \right) \quad \text{bits}. \quad (160)$$

For sufficiently large values of p , this is strictly smaller than $\min\{1 - p, H_b(p)\}$, thus demonstrating that for this channel, our rate region $\mathcal{R}_{\text{cau}}^{\text{com}}$ contains rate pairs that are not achievable using the naïve approach.

ACKNOWLEDGMENT

Comments and suggestions by A. Bracher, C. Bunte, the Associate Editor, and the anonymous reviewers are gratefully acknowledged.

REFERENCES

- [1] A. Bracher, Cribbing, conferencing, feedback, and state-information on the multiple-access channel M.S. thesis, Swiss Federal Institute of Technology, Signal and Information Processing Laboratory. Zurich, Switzerland, 2012.
- [2] A. Bracher, A. Lapidoth, and Y. Steinberg, "On feedback, cribbing, and causal state-information on the multiple-access channel," in *Proc. Information Theory Workshop (ITW'12)*, Lausanne, Switzerland, Sep. 2012, pp. 134–138.
- [3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U.K.: Academic Press, 1981.
- [4] G. Dueck, "Partial feedback for two-way and broadcast channels," *Inf. Control*, vol. 46, pp. 1–15, 1980.
- [5] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [6] M. Gastpar, "On Wyner-Ziv networks," in *Proc. 37th Asilomar Conf. Signals, Syst., Comput.*, Asilomar, CA, Nov. 2003, pp. 855–859.
- [7] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2762–2768, Nov. 2004.
- [8] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Inf. Control*, vol. 9, no. 1, pp. 19–31, 1980.
- [9] 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.
- [10] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal side information: Independent states," *IEEE Trans. Inf. Theory*, to appear.
- [11] M. Li, O. Simeone, and A. Yener, Multiple access channels with states causally known at transmitters 2010, ArXiv:1011.6639v1 30.
- [12] J. P. M. Schalkwijk, "A coding scheme for additive noise channels with feedback—Part II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 183–189, Apr. 1966.
- [13] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, 1958.
- [14] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, Sep. 1973.
- [15] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inf. Theory*, vol. IT-51, no. 8, pp. 2867–2877, Aug. 2005.
- [16] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: General sources," *Inf. Control*, vol. 38, pp. 60–80, 1978.
- [17] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

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 (ETH) in Zurich, Switzerland. He served in the years 2003–2004 and 2009 as Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.

Dr. Lapidoth's research interests are in Digital Communications and Information Theory. He is the author of the textbook *A Foundation in Digital Communication*, published by Cambridge University Press in 2009.

Yossef Steinberg (M'96–SM'09–F'11) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering in 1983, 1986, and 1990, respectively, all from Tel-Aviv University, Tel-Aviv, Israel. He was a Lady Davis Fellow in the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel, and held visiting appointments in the Department of Electrical Engineering at Princeton University, Princeton, NJ, and at the C I Center, George Mason University, Fairfax, VA. From 1995 to 1999, he was with the Department of Electrical Engineering, Ben Gurion University, Beer-Sheva, Israel. In 1999, he joined the Department of Electrical Engineering at the Technion. Dr. Steinberg served in the years 2004–2007 as Associate Editor for Shannon Theory, and currently serves as Associate Editor at large, for the TRANSACTIONS ON INFORMATION THEORY. Dr. Steinberg's research interests are in Digital Communications, Information Theory, and Estimation. He won the 2007 best paper award, jointly with Hanan Weingarten and Shlomo Shamai.