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

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

***Abstract*—We consider a memoryless multiple-access channel (MAC) that is governed by two independent memoryless state sequences, each of which is revealed to a different encoder in a strictly causal or causal way. The special case where one of the state sequences is deterministic (null) corresponds to an MAC governed by a single state that is revealed to only one of the encoders. We show that, even in the strictly causal case, the state information at the encoders can increase the capacity region. It cannot, however, increase the sum-rate capacity. We provide general inner and outer bounds on the capacity region, and we also study a Gaussian example where they coincide. We show that in the causal case, naïve Shannon strategies may be suboptimal.**

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

I. Introduction

We study the capacity region of a memoryless multiple-access channel (MAC) that is governed by two independent memoryless state sequences, each of which is revealed—depending on the scenario—strictly causally or causally to a different encoder. The special *asymmetric case* where one of the state sequences is deterministic (null) corresponds to an MAC governed by a single state sequence, which is revealed to only one of the encoders. Our present work complements [6], which deals with an MAC governed by a *single* state sequence that is revealed to *both* encoders. We shall see that, even in the present case, strictly causal side information (SI) can increase the capacity region. Thus, the gains afforded by SI in [6] and [7] cannot be attributed exclusively to the encoders' ability to cooperate in transmitting a compressed version of the common state: some gains are to be had also when the encoders obtain independent state information and cannot, therefore, cooperate in sending it.

However, when it comes to the sum-rate capacity, the picture changes. Strictly causal SI is beneficial in the common-state scenario [6], [7], but it is useless in the present setting. The increase in the sum-rate capacity when the state is common can thus be attributed to the ability of the encoders to jointly describe the common state and cooperate in sending the description over the MAC.

Why can strictly causal state information be useful on the MAC with independent states? To gain some insight, consider the case where Transmitter 1 is altruistic: although it can send data to the receiver without interfering with Transmitter 2, it is willing to give up all this rate in order to help Transmitter 2. In the absence of SI, the most helpful it can be is by sending a constant symbol (the symbol that interferes least with Transmitter 1). But in the presence of SI, it can do better: it can describe the state sequence it observes to the receiver and in this way help the receiver to decode the message sent by Transmitter 2. The benefits of SI in this scenario are not in allowing the users to cooperate in sending the state but rather in allowing them to trade the rate of one user against the other.

The literature on single-user channels and multiterminal networks that are governed by state sequences is vast. For a recent survey, see [5]. The literature on the causal and strictly causal case is more limited. For single-user memoryless channels, the former was solved by Shannon [9] using "Shannon strategies," and the latter does not increase capacity. The degraded broadcast channel is addressed in [10]. For follow-up work on our problem, see [8] and Theorem 2 ahead.

The rest of 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 in Appendixes A–G and H–I, respectively.

II. Strictly Causal Side Information

A. Basic Definitions

We are given a discrete memoryless state-dependent MAC $P_{Y|S_1,S_2,X_1,X_2}$ with state alphabets $\mathcal{S}_1$ and $\mathcal{S}_2$, state probability mass functions (PMFs) P_{S_1} and P_{S_2}, input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, and output alphabet $\mathcal{Y}$. All the alphabets are finite. We use boldface symbols to denote n-sequences from these alphabets, e.g., $\mathbf{y}$ for an n-sequence over $\mathcal{Y}$ and $\mathbf{s}_1$ for an n-sequence over $\mathcal{S}_1$. The laws governing n-sequences of state and output letters are

$$P^n_{Y|S_1,S_2,X_1,X_2}(\mathbf{y}|\mathbf{s}_1,\mathbf{s}_2,\mathbf{x}_1,\mathbf{x}_2) = \prod_{i=1}^{n} P_{Y|S_1,S_2,X_1,X_2}(y_i|s_{1,i},s_{2,i},x_{1,i},x_{2,i})$$

$$P^n_{S_1,S_2}(\mathbf{s}_1,\mathbf{s}_2) = \prod_{i=1}^{n} P_{S_1}(s_{1,i})\,P_{S_2}(s_{2,i}).$$

Manuscript received July 29, 2011; revised July 04, 2012; accepted October 02, 2012. Date of publication November 27, 2012; date of current version February 12, 2013. Y. Steinberg was supported by the Israel Science Foundation under Grant 280/07. This paper was presented in part at the 2010 IEEE International Symposium on Information Theory.

A. Lapidoth is with the Swiss Federal Institute of Technology (ETH), 8006 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.2012.2230214

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

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

be single-letter cost functions. The cost associated with transmitting the sequence $\mathbf{x}_k$ by encoder k is

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

B. Coding

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

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

$$f_{k,i}\colon \mathcal{S}_k^{i-1} \times \mathcal{M}_k \to \mathcal{X}_k, \quad k=1,2,\ i=1,\ldots,n \tag{1}$$

and a decoding mapping

$$g\colon \mathcal{Y}^n \to \mathcal{M}_1 \times \mathcal{M}_2$$

such that the average input costs are bounded by Γ_k

$$\frac{1}{\nu_k}\sum_{m_k=1}^{\nu_k}\sum_{\mathbf{s}_k\in\mathcal{S}_k^n} P_{\mathbf{S}_k}(\mathbf{s}_k)\ \phi_k\big(\mathbf{f}_k(\mathbf{s}_k,m_k)\big)\le\Gamma_k, \quad k=1,2$$

and the average probability of error P_e is bounded by ϵ. Here

$$\mathbf{f}_k(\mathbf{s}_k,m_k)=\big(f_{k,1}(m_k), f_{k,2}(s_{k,1},m_k),\ldots, f_{k,n}(s_k^{n-1},m_k)\big)$$

where s_k^{n-1} denotes $(s_{k,1},\ldots,s_{k,n-1})$, and

$$P_e = 1 - \frac{1}{\nu_1\nu_2}\sum_{m_1=1}^{\nu_1}\sum_{m_2=1}^{\nu_2}\sum_{\mathbf{s}_1,\mathbf{s}_2} P_{\mathbf{S}_1}(\mathbf{s}_1)\,P_{\mathbf{S}_2}(\mathbf{s}_2) \cdot P\left(g^{-1}(m_1,m_2)|\mathbf{s}_1,\mathbf{s}_2,\mathbf{f}_1(\mathbf{s}_1,m_1),\mathbf{f}_2(\mathbf{s}_2,m_2)\right)$$

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

The rate pair of the code is defined as

$$\left(\frac{1}{n}\log\nu_1, \frac{1}{n}\log\nu_2\right).$$

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 $(n,\nu_1,\nu_2,\Gamma_1,\Gamma_2,\epsilon)$ code with strictly causal SI for the channel $P_{Y|S,X_1,X_2}$ with

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

The capacity-cost region $\mathcal{C}_{\text{s-c}}^{\text{ind}}$ of the channel with strictly causal SI is the closure of the set of all achievable quadruples $(R_1,R_2,\Gamma_1,\Gamma_2)$. The superscript "ind" indicates that the two states are independent, to distinguish the current model from the one treated in [6]. For a given pair (Γ_1,Γ_2) of input costs, $\mathcal{C}_{\text{s-c}}^{\text{ind}}(\Gamma_1,\Gamma_2)$ stands for the section of $\mathcal{C}_{\text{s-c}}^{\text{ind}}$ at (Γ_1,Γ_2). Our interest is in characterizing $\mathcal{C}_{\text{s-c}}^{\text{ind}}(\Gamma_1,\Gamma_2)$.

By the asymmetric case, we shall refer to the case where S_2 is deterministic, in which case the state consists of only one component S_1, which is available to User 1.

C. Outer Bounds

Denote by C_Σ the sum-rate capacity of the MAC without state information

$$C_\Sigma = \max_{P_Q P_{X_1|Q} P_{X_2|Q}} I(X_1,X_2;Y|Q) \tag{2}$$

where the maximum on the right-hand side (RHS) is over all joint PMFs of the form

$$P_{Q,X_1,X_2,Y} = P_Q\, P_{X_1|Q}\, P_{X_2|Q}\, P_{Y|X_1,X_2} \tag{3a}$$

satisfying the constraints

$$\mathsf{E}[\phi_k(X_k)] \le \Gamma_k, \quad k=1,2. \tag{3b}$$

As the next proposition shows, strictly causal SI does not increase the sum-rate capacity. This is in contrast to the case where the channel is governed by a common state, which is revealed strictly causally to both encoders [6].

Proposition 1 (The Sum-Rate Capacity): Strictly causal SI does not increase the sum-rate capacity:

$$\max_{(R_1,R_2)\in\mathcal{C}_{\text{s-c}}^{\text{ind}}(\Gamma_1,\Gamma_2)} (R_1+R_2) = C_\Sigma. \tag{4}$$

Proof: See Appendix A ∎

The following proposition deals with the asymmetric case. It shows that in this case, strictly causal SI can only increase the maximal rate at which the user without the SI can communicate: If User 2 is not provided any SI, then providing User 1 with strictly causal SI cannot increase its rate; it can only increase the rate of User 2. In this sense, the channel from User 1 to the output can be viewed as a single-user channel where strictly causal SI does not increase capacity.

Proposition 2 (Asymmetric Case): If S_2 is deterministic, then strictly causal SI to User 1 does not increase its maximal rate:

$$\max\big\{R_1\colon (R_1,0)\in\mathcal{C}_{\text{s-c}}^{\text{ind}}(\Gamma_1,\Gamma_2)\big\}=\max I(X_1;Y|X_2,Q) \tag{5}$$

where the maximum on the RHS is over joint distributions satisfying (3).

Proof: See Appendix B. ∎

We next present an outer bound on the capacity region. Denote by $\mathcal{O}_{\text{s-c},1}^{\text{ind}}$ the set of all tuples $(R_1,R_2,\Gamma_1,\Gamma_2)$ satisfying

$$R_1 \le I(X_1;Y|X_2,S_2,Q) \tag{6a}$$
$$R_2 \le I(X_2;Y|X_1,S_1,Q) \tag{6b}$$
$$R_1+R_2 \le I(X_1,X_2;Y|Q) \tag{6c}$$
$$\Gamma_k \ge \mathsf{E}[\phi_k(X_k)], \quad k\in\{1,2\} \tag{6d}$$

for some joint distribution of the form

$$P_{Q,X_1,X_2,S_1,S_2,Y}=P_Q\, P_{S_1}\, P_{S_2}\, P_{X_1|Q}\, P_{X_2|Q}\, P_{Y|S_1,S_2,X_1,X_2} \tag{6e}$$

where Q is a time-sharing random variable taking values in an alphabet $\mathcal{Q}$ whose size can be bounded by four.

Proposition 3 (Outer Bound):

$$\mathcal{C}_{\text{s-c}}^{\text{ind}} \subseteq \mathcal{O}_{\text{s-c},1}^{\text{ind}}. \tag{7}$$

Proof: See Appendix C. ■

Note that, because Q is a time-sharing random variable, it follows from (6c) that strictly causal SI does not increase the sum-rate capacity. Consequently, Proposition 1 can be viewed as a corollary of Proposition 3. In fact, Proposition 2 on the asymmetric setting (where S_2 is deterministic) can also be viewed as a corollary by considering (6a).

As the following example shows, the outer bound of Proposition 3 need not be tight.

Remark 1: For some channels the inclusion in (7) is strict.

Example 1: Consider the asymmetric case in which Transmitter 1's input X_1 is binary; Transmitter 2's input $X_2 = (X_2^{(a)}, X_2^{(b)})$ is a binary tuple; and the channel output $Y = (Y_1, Y_2)$ is such that Y_1 is binary, whereas $Y_2 = (Y_2^{(a)}, Y_2^{(b)})$ is a binary tuple. The channel law is

$$Y_1 = X_1 \tag{8a}$$

$$Y_2 = X_2 \oplus S_1 \oplus Z \tag{8b}$$

where $S_1 = (S_1^{(a)}, S_1^{(b)})$ is a random binary tuple whose components are independent and identically distributed (IID) $\text{Bernoulli}(1/2)$, and $Z = (Z^{(a)}, Z^{(b)})$ is a random binary tuple whose components are IID $\text{Bernoulli}(p)$, where $p \approx 0.11$ is the unique constant in the interval $(0, 1/2)$ whose binary entropy function is $1/2$:

$$H_{\text{b}}(p) = \frac{1}{2} \text{ bit}, \quad 0 < p < 0.5. \tag{8c}$$

The state tuple S_1 and the noise tuple Z are independent, and the mod-2 addition $\oplus$ in (8b) is componentwise:

$$Y_2^{(a)} = X_2^{(a)} \oplus S_1^{(a)} \oplus Z^{(a)}$$
$$Y_2^{(b)} = X_2^{(b)} \oplus S_1^{(b)} \oplus Z^{(b)}.$$

For this channel, the rate pair $(0, 1)$ is in $\mathcal{O}_{\text{s-c},1}^{\text{ind}}$ but not in $\mathcal{C}_{\text{s-c}}^{\text{ind}}$.

Proof: See Appendix D. ■

D. Achievable Region

Let $\mathcal{P}_{\text{s-c}}^{\text{ind}}$ be the collection of all PMFs $P_{V_1,V_2,S_1,S_2,X_1,X_2,Y}$ of the form

$$P_{V_1,V_2,S_1,S_2,X_1,X_2,Y} = P_{V_1|S_1} P_{V_2|S_2} P_{S_1} P_{S_2} P_{X_1} P_{X_2} P_{Y|S_1,S_2,X_1,X_2}. \tag{9}$$

Note that (9) implies the Markov relations

$$V_1 \multimap S_1 \multimap (V_2, Y, S_2) \tag{10a}$$

$$V_2 \multimap S_2 \multimap (V_1, Y, S_1) \tag{10b}$$

$$(V_1, V_2) \multimap (S_1, S_2) \multimap (X_1, X_2, Y) \tag{10c}$$

and that X_1, X_2 are independent of each other and of the quadruple (V_1, V_2, S_1, S_2). Let $\mathcal{R}_{\text{s-c}}^{\text{ind}}$ be the convex hull of the collection of all tuples $(R_1, R_2, \Gamma_1, \Gamma_2)$ satisfying

$$0 \le R_1 \le I(X_1; Y|X_2, V_1, V_2) - I(V_1; S_1|Y, V_2) \tag{11a}$$

$$0 \le R_2 \le I(X_2; Y|X_1, V_1, V_2) - I(V_2; S_2|Y, V_1) \tag{11b}$$

$$R_1 + R_2 \le I(X_1, X_2; Y|V_1, V_2) - I(V_1, V_2; S_1, S_2|Y) \tag{11c}$$

$$\Gamma_k \ge \mathsf{E}[\phi_k(X_k)], \quad k = 1, 2 \tag{11d}$$

for some $P_{V_1,V_2,S_1,S_2,X_1,X_2,Y}$ in $\mathcal{P}_{\text{s-c}}^{\text{ind}}$.

Our main achievability result for the strictly causal case is the following.

Theorem 1: The region $\mathcal{R}_{\text{s-c}}^{\text{ind}}$ is achievable:

$$\mathcal{R}_{\text{s-c}}^{\text{ind}} \subseteq \mathcal{C}_{\text{s-c}}^{\text{ind}}. \tag{12}$$

Proof: The proof is in Appendix E. It is based on distributed Wyner–Ziv source coding [4] and on block-Markov channel coding to transmit the Wyner–Ziv codewords and the messages (data). The channel output serves as SI for the reconstructor in the distributed Wyner–Ziv code. Since the two components of the source are independent, there is no direct cooperation between the encoders via a common message as in [6]. Instead, each user spends part of its private rate on the transmission of its Wyner–Ziv codeword. ■

In some cases, the region $\mathcal{R}_{\text{s-c}}^{\text{ind}}$ coincides with $\mathcal{C}_{\text{s-c}}^{\text{ind}}$. The next example is such a case. It is also an example where the outer bound of Proposition 3 is tight.[1] 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 [11].

Example 2: Consider the asymmetric Gaussian MAC with input power constraints $\phi_1(x) = \phi_2(x) = x^2$, where the state S_1 is the channel noise:

$$Y = X_1 + X_2 + S_1, \qquad S_1 \sim \mathcal{N}\big(0, \sigma_{s_1}^2\big). \tag{13}$$

The capacity region of this channel when S_1 is revealed strictly causally to Transmitter 1 comprises all the rate-pairs (R_1, R_2) satisfying

$$R_1 \le \frac{1}{2} \log \left(1 + \frac{\Gamma_1}{\sigma_{s_1}^2}\right) \tag{14a}$$

$$R_1 + R_2 \le \frac{1}{2} \log \left(1 + \frac{\Gamma_1 + \Gamma_2}{\sigma_{s_1}^2}\right). \tag{14b}$$

Proof: That all achievable rate pairs must satisfy (14a) follows from Proposition 2, and that they must also all satisfy (14b) follows from Proposition 1. The achievability is proved in Appendix F. ■

The capacity region in the above example can be strictly larger than the capacity region in the absence of state information [1, Sec. 15.1]. Indeed, in the presence of strictly causal

[1] In fact, in this example, Propositions 1 and 2 already specify the region.

SI to User 1, the network can support positive rates by User 2 even if its allowed average power Γ_2 approaches zero. We thus conclude the following.

Remark 2: Strictly causal state information on the MAC can increase capacity.

In the scheme achieving $\mathcal{R}_{\text{s-c}}^{\text{ind}}$, each transmitter sends in Block $b+1$ a compressed version of the SI information it observed in Block b. But, more generally, it can send a compressed version of the pair of sequences comprising the state sequence that it observed in Block b and, additionally, the sequence comprising its own inputs at Block b. This improvement was proposed by Li *et al.* [8] who obtained the following region.

Theorem 2 [8]: If (R_1, R_2) satisfy

$$R_1 \leq I(X_1, V_1; Y|X_2, V_2) - I(V_1; S_1|X_1) \quad (15)$$

$$R_2 \leq I(X_2, V_2; Y|X_1, V_1) - I(V_2; S_2|X_2) \quad (16)$$

$$R_1 + R_2 \leq I(X_1, X_2, V_1, V_2; Y) - I(V_1; S_1|X_1) - I(V_2; S_2|X_2) \quad (17)$$

for some joint PMF

$$P_{V_1,V_2,S_1,S_2,X_1,X_2,Y} = P_{V_1|S_1,X_1} P_{V_2|S_2,X_2} P_{S_1} P_{S_2} P_{X_1} P_{X_2} P_{Y|S_1,S_2,X_1,X_2} \quad (18)$$

then (R_1, R_2) is achievable.

Li *et al.* showed that their achievable region contains the one of Theorem 1, and they conjectured that for some channels, the inclusion is strict. We next demonstrate that this is indeed the case by providing such a channel.

Example 3: Suppose that the state S_1 is null (deterministic), and the state S_2 is the pair (W_0, W_1), where W_0 and W_1 are IID binary random variables each of entropy $1/2$

$$S_1 = 0 \quad S_2 = (W_0, W_1) \quad (19a)$$

$$H(W_0) = H(W_1) = \frac{1}{2} \text{ bits.} \quad (19b)$$

The MAC has binary inputs $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ and an output $Y = (Y_1, Y_2) \in \{0, 1\} \times \{0, 1\}$ with

$$Y_1 = X_1 \oplus W_{X_2} \quad (20a)$$

$$Y_2 = X_2. \quad (20b)$$

Thus, if X_2 is equal to zero, then Y_1 is the mod-2 sum of X_1 and W_0, and otherwise, it is the mod-2 sum of X_1 and W_1.

The rate pair $(R_1, R_2) = (1, 1/2)$ is in the inner bound of Li *et al.* but not in the inner bound of Theorem 1. In fact, if $(1, R_2)$ is in $\mathcal{R}_{\text{s-c}}^{\text{ind}}$, then R_2 must be zero.

Proof: See Appendix G. ■

III. Causal Side Information

A. Basic Definitions and Coding

The definition of codes and achievable rates remain as in Section II-B, with the only difference being in the definition of encoding mappings: in the causal case, (1) is replaced by

$$f_{k,i}: \mathcal{S}_k^i \times \mathcal{M}_k \to \mathcal{X}_k, \quad k = 1, 2, \quad i = 1, \ldots, n. \quad (21)$$

The capacity region and its section at (Γ_1, Γ_2) are denoted $\mathcal{C}_{\text{cau}}^{\text{ind}}$ and $\mathcal{C}_{\text{cau}}^{\text{ind}}(\Gamma_1, \Gamma_2)$.

In general, causal SI can be more beneficial than strictly causal. Indeed, in the single-user channel, the former can increase capacity [9] and the latter cannot. Since the single-user channel can be viewed as a degenerate MAC, we have the following.

Remark 3: The capacity region of an MAC with causal state information can be larger than with strictly causal state information.

Of course, in some cases, causal SI is no better than strictly causal SI, e.g., when the states are irrelevant, i.e., when $W(y|x_1, x_2, s_1, s_2)$ does not depend on (s_1, s_2).

B. Achievable Region

Let $\mathcal{P}_{\text{cau}}^{\text{ind}}$ be the set of all PMFs $P_{V_1,V_2,U_1,U_2,S_1,S_2,X_1,X_2,Y}$ of the form

$$P_{V_1,V_2,U_1,U_2,S_1,S_2,X_1,X_2,Y} = P_{S_1} P_{S_2} P_{V_1|S_1} P_{V_2|S_2} \cdot P_{U_1} P_{U_2} P_{X_1|U_1,S_1} P_{X_2|U_2,S_2} P_{Y|S_1,S_2,X_1,X_2}. \quad (22)$$

Observe that (22) implies the Markov relations

$$V_1 \multimap S_1 \multimap (V_2, S_2, X_1, X_2, U_1, U_2, Y)$$
$$V_2 \multimap S_2 \multimap (V_1, S_1, X_1, X_2, U_1, U_2, Y)$$
$$(V_1, V_2) \multimap (S_1, S_2) \multimap (X_1, X_2, U_1, U_2, Y). \quad (23)$$

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

$$0 \leq R_1 \leq I(U_1; Y|U_2, V_1, V_2) - I(V_1; S_1|Y, V_2) \quad (24a)$$

$$0 \leq R_2 \leq I(U_2; Y|U_1, V_1, V_2) - I(V_2; S_2|Y, V_1) \quad (24b)$$

$$R_1 + R_2 \leq I(U_1, U_2; Y|V_1, V_2) - I(V_1, V_2; S_1, S_2|Y) \quad (24c)$$

$$\Gamma_k \geq \mathsf{E}[\phi_k(X_k)], \quad k = 1, 2 \quad (24d)$$

for some $P_{V_1,V_2,U_1,U_2,S_1,S_2,X_1,X_2,Y}$ in $\mathcal{P}_{\text{cau}}^{\text{ind}}$.

Theorem 3: In the presence of causal SI, $\mathcal{R}_{\text{cau}}^{\text{ind}}$ is achievable:

$$\mathcal{R}_{\text{cau}}^{\text{ind}} \subseteq \mathcal{C}_{\text{cau}}^{\text{ind}}.$$

Proof: The proof proceeds along the lines of the proof of Theorem 1, except that the input X_k is allowed to depend on the state S_k and that additional external random variables U_1 and U_2 that do not depend on S_1, S_2 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 that is independent of the state. ■

The scheme that Shannon proposed to achieve the capacity of the single-user channel with causal SI does not involve block-Markov coding [9]. In fact, Shannon's scheme ignores the past states, and in his scheme, the present input X_i is a function only of the message and the present state S_i. This allowed Shannon to reduce the channel to one where the inputs are mappings ("strategies") from the state alphabet to the input alphabet and to then reduce the problem to that of coding for a channel without states. This "naïve approach" can also be applied to the MAC.

It leads to the region $\mathcal{R}^{\text{ind}}_{\text{naïve}}$ comprising all rate-pairs (R_1, R_2) satisfying

$$R_1 \le I(T_1; Y | T_2, Q) \tag{25a}$$
$$R_2 \le I(T_2; Y | T_1, Q) \tag{25b}$$
$$R_1 + R_2 \le I(T_1, T_2; Y | Q) \tag{25c}$$

for some PMF of the form

$$P_{Q,T_1,T_2,Y} = P_Q \, P_{T_1|Q} \, P_{T_2|Q} \, P_{Y|T_1,T_2} \tag{25d}$$

where T_k for $k \in \{1, 2\}$ are random Shannon strategies [9] whose realizations are mappings $t_k\colon \mathcal{S}_k \to \mathcal{X}_k$; the random variable Q is for time sharing

$$P_{Y|T_1,T_2}(y|t_1,t_2) = \sum_{s_1 \in \mathcal{S}_1} \sum_{s_2 \in \mathcal{S}_2} P_{S_1}(s_1) \, P_{S_2}(s_2) \, P_{Y|S_1,S_2,X_1,X_2}\big(y | s_1, s_2, t_1(s_1), t_2(s_2)\big); \tag{25e}$$

and

$$\mathsf{E}\big[\phi_k\big(T_k(S_k)\big)\big] \le \Gamma_k, \quad k = 1, 2. \tag{25f}$$

By choosing V_1, V_2 in (24) to be deterministic, we obtain the following.

Remark 4: Every rate pair that is achievable with naïve Shannon strategies must also be in $\mathcal{R}^{\text{ind}}_{\text{cau}}$

$$\mathcal{R}^{\text{ind}}_{\text{naïve}} \subseteq \mathcal{R}^{\text{ind}}_{\text{cau}}. \tag{26}$$

The next example demonstrates that the reverse is not true. Consequently, the naïve approach—while optimal for the single-user channel—need not be optimal on the MAC.

Remark 5: For some channels, there are rate pairs in $\mathcal{R}^{\text{ind}}_{\text{cau}}$ that are not achievable using naïve Shannon strategies.

Example 4: Consider the asymmetric state-dependent MAC consisting of two independent single-user channels, where the state of Channel 2 is available causally at the input of Channel 1. Specifically, let the input and output alphabets be

$$\mathcal{X}_1 = \{0, 1\}, \quad \mathcal{X}_2 = \{0, 1, 2, 3\}, \quad \mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$$

where

$$\mathcal{Y}_1 = \{0, 1\}, \quad \mathcal{Y}_2 = \{0, 1, 2, 3\}.$$

The channel is defined as

$$Y_1 = X_1$$
$$Y_2 = X_2 + S_1 \bmod 4$$

where S_1 takes value in the set $\mathcal{S}_1 = \{0, 1, 2, 3\}$ according the probability vector

$$\big(1 - p, \, p/3, \, p/3, \, p/3\big)$$

and p is small enough so

$$H(S_1) < 1. \tag{27}$$

For this channel, the rate pair $(0, 2)$ is in $\mathcal{R}^{\text{ind}}_{\text{cau}}$, but it cannot be achieved using naïve Shannon strategies

Proof: That the rate pair $(0, 2)$ is in $\mathcal{R}^{\text{ind}}_{\text{cau}}$ is not surprising because $H(S_1)$ is smaller than 1, so in the block-Markov scheme Encoder 1 can losslessly compress the state sequence pertaining to the previous block and transmit it over the channel from X_1 to Y_1 whose capacity is one. The receiver, upon obtaining this state sequence, can subtract it (mod-4) from the Y_2 sequence of the previous block and in this way obtain a clean channel from X_2 to Y_2 of capacity 2. The choice of the auxiliary random variables and the PMF that correspond to this scheme can be found in Appendix A, which also contains a proof that this pair is not achievable with the naïve approach. ■

APPENDIX

A. Proof of Proposition 1

The RHS of (2) is the sum-rate capacity of the same MAC without SI. Consequently, it is also achievable in the presence of (strictly causal) SI, because the SI can always be ignored. It remains to show that the RHS is also an upper bound. Let M_1 and M_2 be the random messages of Users 1 and 2. Starting with Fano's inequality

$$\begin{aligned} n(R_1 + R_2) - n\epsilon_n &\le I(M_1, M_2; Y^n) \\ &= \sum_{i=1}^{n} I(M_1, M_2; Y_i | Y^{i-1}) \\ &\le \sum_{i=1}^{n} I(M_1, M_2, Y^{i-1}; Y_i) \\ &\le \sum_{i=1}^{n} I(M_1, M_2, X_{1,i}, X_{2,i}, Y^{i-1}; Y_i) \\ &= \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i) \end{aligned} \tag{28}$$

where the last equality holds because in the strictly causal case

$$(M_1, M_2, Y^{i-1}) \multimap\!\!- (X_{1,i}, X_{2,i}) \multimap\!\!- Y_i.$$

Since (M_1, S_1^{i-1}) and (M_2, S_2^{i-1}) are independent, so are $X_{1,i}$ and $X_{2,i}$. The claim now follows by the standard time sharing argument. ■

B. Proof of Proposition 2

Since the SI can always be ignored, we only need to prove a converse. Denote by M_1 the random message of User 1, and note that M_1 and X_2^n are independent. By Fano's inequality and this independence

$$\begin{aligned} nR_1 - n\epsilon_n &\le I(M_1;Y^n) \\ &\le I(M_1;Y^n|X_2^n) \\ &= \sum_{i=1}^n I(M_1;Y_i|Y^{i-1},X_2^n) \\ &= \sum_{i=1}^n \Big(H(Y_i|Y^{i-1},X_2^n) - H(Y_i|Y^{i-1},X_2^n,M_1)\Big) \\ &\le \sum_{i=1}^n \Big(H(Y_i|X_{2,i}) - H(Y_i|Y^{i-1},X_2^n,M_1)\Big) \\ &\le \sum_{i=1}^n \Big(H(Y_i|X_{2,i}) - H(Y_i|Y^{i-1},X_2^n,M_1,X_{1,i})\Big) \\ &= \sum_{i=1}^n I(M_1,X_{1,i},Y^{i-1},X_2^{i-1},X_{2,i+1}^n;Y_i|X_{2,i}) \\ &= \sum_{i=1}^n I(X_{1,i};Y_i|X_{2,i}) \end{aligned}$$

where the last equality follows from the Markov relation

$$(M_1,Y^{i-1},X_2^{i-1},X_{2,i+1}^n) \multimap\!- (X_{1,i},X_{2,i}) \multimap\!- Y_i.$$

Since (M_1,S_1^{i-1}) and M_2 are independent, so are $X_{1,i}$ and $X_{2,i}$. The claim now follows by the standard time sharing arguments. ■

C. Proof of Proposition 3

Following the proof of Proposition 1

$$n(R_1+R_2) - n\delta_n \le \sum_{i=1}^n I(X_{1,i},X_{2,i};Y_i) \tag{29}$$

where $X_{1,i}$ and $X_{2,i}$ are independent. For the individual rates

$$\begin{aligned} nR_1 - n\delta_n &\le I(M_1;Y^n,S_2^n|M_2) \\ &= \sum_{i=1}^n I(M_1;Y_i,S_{2,i}|M_2,Y^{i-1},S_2^{i-1}) \\ &= \sum_{i=1}^n I(M_1;S_{2,i}|M_2,Y^{i-1},S_2^{i-1}) \\ &\quad + \sum_{i=1}^n I(M_1;Y_i|M_2,Y^{i-1},S_2^i) \\ &\overset{(a)}{=} \sum_{i=1}^n I(M_1;Y_i|M_2,Y^{i-1},S_2^i) \\ &\overset{(b)}{=} \sum_{i=1}^n I(M_1;Y_i|M_2,X_{2,i},Y^{i-1},S_2^i) \\ &\le \sum_{i=1}^n I(M_1,X_{1,i};Y_i|M_2,X_{2,i},Y^{i-1},S_2^i) \\ &\overset{(c)}{=} \sum_{i=1}^n I(X_{1,i};Y_i|M_2,X_{2,i},Y^{i-1},S_2^i) \\ &\overset{(d)}{\le} \sum_{i=1}^n I(X_{1,i},M_2,Y^{i-1},S_2^{i-1};Y_i|X_{2,i},S_{2,i}) \\ &\overset{(e)}{=} \sum_{i=1}^n I(X_{1,i};Y_i|X_{2,i},S_{2,i}) \end{aligned} \tag{30}$$

where (a) holds due to the Markov relation

$$S_{2,i} \multimap\!- (M_2,Y^{i-1},S_2^{i-1}) \multimap\!- M_1; \tag{31}$$

(b) holds since $X_{k,i}$ is a deterministic function of (M_k,S_k^{i-1}); (c) holds due to the Markov chain

$$Y_i \multimap\!- (M_2,Y^{i-1},S_2^i,X_{1,i}) \multimap\!- M_1; \tag{32}$$

(d) holds since conditioning reduces entropy; and (e) holds due to the Markov chain

$$Y_i \multimap\!- (X_{1,i},X_{2,i},S_{2,i}) \multimap\!- (M_2,Y^{i-1},S_2^{i-1}). \tag{33}$$

In a similar manner, we obtain for R_2

$$nR_2 - n\delta_n \le \sum_{i=1}^n I(X_{2,i};Y_i|X_{1,i},S_{1,i}). \tag{34}$$

The proposition now follows by applying the standard time sharing argument on (30), (34), and (29). ■

D. Analysis of Example 1

To see that the rate pair $(0,1)$ is in $\mathcal{O}_{\text{s-c},1}^{\text{ind}}$, we can consider the distribution on X_1,X_2 according to which they are independent; X_1 is $\text{Bernoulli}(1/2)$; and the components of X_2 are IID $\text{Bernoulli}(1/2)$. With this distribution, the sum-rate constraint is 1 bit and the constraint on R_2 is also 1 bit. It remains to show that the rate pair $(0,1)$ is not achievable, i.e., that it is not in $\mathcal{C}_{\text{s-c}}^{\text{ind}}$.

The key to that is to note that because User 1's input alphabet is binary, and because the entropy of S_1 is 2 bits, Transmitter 1 cannot describe S_1 perfectly. More specifically, consider an n-sequence $\mathbf{X}_1$ produced by Transmitter 1 and an n-tuple of states $\mathbf{S}_1$:

$$\begin{aligned} H(\mathbf{S}_1|\mathbf{X}_1) &= H(\mathbf{S}_1) - I(\mathbf{X}_1;\mathbf{S}_1) \\ &\ge H(\mathbf{S}_1) - H(\mathbf{X}_1) \\ &\ge H(\mathbf{S}_1) - \log|\mathcal{X}_1|^n \\ &= 2n - n \\ &= n \text{ bits}. \end{aligned} \tag{35}$$

As we next show, this inequality yields an upper bound on R_2. Indeed, starting with Fano's inequality

$$\begin{aligned} nR_2 - n\epsilon_n &\le I(M_2; \mathbf{Y}_1, \mathbf{Y}_2) \\ &= I(M_2; \mathbf{X}_1, \mathbf{Y}_2) \\ &\le I(\mathbf{X}_2; \mathbf{Y}_2|\mathbf{X}_1) \\ &= I(\mathbf{X}_2; \mathbf{X}_2 \oplus \mathbf{S}_1 \oplus \mathbf{Z}|\mathbf{X}_1) \\ &= H(\mathbf{X}_2 \oplus \mathbf{S}_1 \oplus \mathbf{Z}|\mathbf{X}_1) - H(\mathbf{X}_2 \oplus \mathbf{S}_1 \oplus \mathbf{Z}|\mathbf{X}_1, \mathbf{X}_2) \\ &= H(\mathbf{X}_2 \oplus \mathbf{S}_1 \oplus \mathbf{Z}|\mathbf{X}_1) - H(\mathbf{S}_1 \oplus \mathbf{Z}|\mathbf{X}_1, \mathbf{X}_2) \\ &= H(\mathbf{X}_2 \oplus \mathbf{S}_1 \oplus \mathbf{Z}|\mathbf{X}_1) - H(\mathbf{S}_1 \oplus \mathbf{Z}|\mathbf{X}_1) \\ &\le 2n - H(\mathbf{S}_1 \oplus \mathbf{Z}|\mathbf{X}_1). \end{aligned} \tag{36}$$

To lower-bound the entropy term on the RHS of (36) we next use (35) and Mrs. Gerber's Lemma [3, Sec. 2.1]:

$$\begin{aligned} \frac{1}{2n} H(\mathbf{S}_1 \oplus \mathbf{Z}|\mathbf{X}_1) &\ge H_{\mathrm{b}}\left(H_{\mathrm{b}}^{-1}\left(\frac{1}{2n} H(\mathbf{S}_1|\mathbf{X}_1)\right) * p\right) \\ &\ge H_{\mathrm{b}}\left(H_{\mathrm{b}}^{-1}\left(\frac{1}{2}\right) * p\right) \\ &= H_{\mathrm{b}}(p * p) \end{aligned} \tag{37}$$

where the second inequality follows from (35); the monotonicity of $H_{\mathrm{b}}^{-1}(\cdot)$; the monotonicity of $\xi \mapsto \xi * p$; and the monotonicity of $H_{\mathrm{b}}(\cdot)$ on the interval $[0, 1/2]$.

Using (37), we now obtain from (36) upon dividing by n and taking the limit as n tends to infinity

$$R_2 \le 2 - 2H_{\mathrm{b}}(p * p) \tag{38}$$

which is strictly smaller than 1.

In fact, it is not difficult to see that the RHS of (38) is achievable: Transmitter 1 uses $n/2$ bits to describe the first component of the n-length state sequence $\mathbf{S}_1$ using a classical rate-distortion codebook, and uses $n/2$ bits to describe the second component. This leads to two parallel binary symmetric channels, each of crossover probability $p * p$.

E. *Proof of Theorem 1*

Lemma 1: If for some PMF of form (9), the rates R_1 and R_2 satisfy inequalities (11), then $R_1 + R_2$ cannot exceed $I(X_1, X_2; Y)$. Consequently, if the latter is zero, then both R_1 and R_2 must be zero.

Proof: From the sum-rate inequality (11c)

$$\begin{aligned} R_1 + R_2 &\le I(X_1, X_2; Y|V_1, V_2) - I(V_1, V_2; S_1, S_2|Y) \\ &= I(X_1, X_2; Y, V_1, V_2) - I(V_1, V_2; S_1, S_2|Y) \\ &= I(X_1, X_2; Y) + I(X_1, X_2; V_1, V_2|Y) \\ &\quad - I(V_1, V_2; S_1, S_2|Y) \\ &= I(X_1, X_2; Y) - H(V_1, V_2|Y, X_1, X_2) \\ &\quad + H(V_1, V_2|Y, S_1, S_2) \\ &= I(X_1, X_2; Y) - H(V_1, V_2|Y, X_1, X_2) \\ &\quad + H(V_1, V_2|Y, X_1, X_2, S_1, S_2) \\ &\le I(X_1, X_2; Y) \end{aligned} \tag{39}$$

where the second line holds because (X_1, X_2) and (V_1, V_2) are independent; the third by the chain rule; the fourth by expressing mutual information in terms of conditional entropies; the fifth by the Markov relation (10c); and the sixth because conditioning cannot increase entropy. ■

Lemma 2: If for some PMF of form (9), the rates R_1 and R_2 satisfy inequalities (11) and $I(X_2; Y, S_1|X_1)$ is zero, then R_2 must be zero and R_1 cannot exceed $I(X_1; Y)$. In this case, (R_1, R_2) is achievable by using the MAC as a single-user channel from X_1 to Y and by ignoring the state information.

Proof: Fix some PMFs of form (9) and rates R_1 and R_2 satisfying inequalities (11). Assume that

$$I(X_2; Y, S_1|X_1) = 0. \tag{40}$$

Starting with (11b)

$$\begin{aligned} R_2 &\le I(X_2; Y|X_1, V_1, V_2) - I(V_2; S_2|Y, V_1) && (41) \\ &\le I(X_2; Y, S_1|X_1, V_1, V_2) - I(V_2; S_2|Y, S_1, V_1) && (42) \\ &= I(X_2; Y, S_1|X_1, V_2) - I(V_2; S_2|Y, S_1) && (43) \end{aligned}$$

with the following justifications.

We first justify (42), where we replaced Y with (Y, S_1). The first term on the RHS of (41) is clearly no larger than the first term on the RHS of (42). And as to the second term

$$\begin{aligned} I(V_2; S_2|Y, V_1) &= H(V_2|Y, V_1) - H(V_2|S_2, Y, V_1) \\ &= H(V_2|Y, V_1) - H(V_2|S_2) \\ &\ge H(V_2|Y, S_1, V_1) - H(V_2|S_2) \\ &= H(V_2|Y, S_1, V_1) - H(V_2|S_2, Y, S_1) \\ &= I(V_2; S_2|Y, S_1, V_1) \end{aligned}$$

where the second line follows from (10b); the third because conditioning cannot increase entropy; and the fourth because (9) implies

$$V_2 -\!\!\circ\!\!- S_2 -\!\!\circ\!\!- (Y, S_1).$$

Having justified (42), we next justify (43), where we have dropped V_1. Starting with the first term

$$\begin{aligned} &I(X_2; Y, S_1|X_1, V_1, V_2) \\ &\quad = H(X_2|X_1, V_1, V_2) - H(X_2|Y, S_1, X_1, V_1, V_2) \\ &\quad = H(X_2|X_1, V_2) - H(X_2|Y, S_1, X_1, V_1, V_2) \\ &\quad = H(X_2|X_1, V_2) - H(X_2|Y, S_1, X_1, V_2) \\ &\quad = I(X_2; Y, S_1|X_1, V_2) \end{aligned}$$

where the second line follows because, under all joint PMFs of form (9), X_2 and (X_1, V_1, V_2) are independent; and the third line because under such PMFs

$$X_2 -\!\!\circ\!\!- (Y, S_1, X_1, V_2) -\!\!\circ\!\!- V_1.$$

To conclude the justification of (43), it remains to consider the second term of (42) and to show that we can drop V_1 from it too:

$$\begin{aligned} I(V_2;S_2|Y,S_1,V_1) &= H(V_2|Y,S_1,V_1) - H(V_2|S_2,Y,S_1,V_1) \\ &= H(V_2|Y,S_1,V_1) - H(V_2|S_2) \\ &= H(V_2|Y,S_1) - H(V_2|S_2) \\ &= I(V_2;S_2|Y,S_1) \end{aligned}$$

where the second equality follows from (10b), and the third equality because for PMFs of form (9)

$$V_2 - \circ - (Y,S_1) - \circ - V_1. \tag{44}$$

Having justified (43), we now use it to conclude that R_2 must be zero. Starting from (43)

$$\begin{aligned} R_2 &\le I(X_2;Y,S_1|X_1,V_2) - I(V_2;S_2|Y,S_1) \\ &= I(X_2;Y,S_1,V_2|X_1) - I(V_2;S_2|Y,S_1) \\ &= I(X_2;Y,S_1|X_1) + I(X_2;V_2|X_1,Y,S_1) - I(V_2;S_2|Y,S_1) \\ &= I(X_2;V_2|X_1,Y,S_1) - I(V_2;S_2|Y,S_1) \\ &= H(V_2|X_1,Y,S_1) - H(V_2|X_1,X_2,Y,S_1) \\ &\quad + H(V_2|S_1,S_2,Y) - H(V_2|S_1,Y) \\ &\le H(V_2|S_1,S_2,Y) - H(V_2|X_1,X_2,Y,S_1) \\ &= H(V_2|S_2) - H(V_2|X_1,X_2,Y,S_1) \\ &\le 0 \end{aligned}$$

where the second line follows from the independence of X_2 and (X_1, V_2); the third by the chain rule; the fourth by (40); the sixth line because conditioning cannot increase entropy; the seventh because for PMFs of form (9), the Markov relation

$$V_2 - \circ - S_2 - \circ - (S_1, Y), \quad P_{V_1,V_2,S_1,S_2,X_1,X_2,Y} \in \mathcal{P}^{\mathrm{ind}}_{\text{s-c}}$$

holds [e.g., by (10b)]; and the final line because conditioning cannot increase entropy. This establishes that the lemma's hypotheses imply that R_2 must be zero.

To conclude the proof, it remains to establish that R_1 cannot exceed $I(X_1;Y)$. Since R_2 is zero, it follows from (39) that

$$\begin{aligned} R_1 &= R_1 + R_2 \\ &\le I(X_1,X_2;Y) \\ &= I(X_1;Y) + I(X_2;Y|X_1) \\ &\le I(X_1;Y) + I(X_2;Y,S_1|X_1) \\ &= I(X_1;Y) \end{aligned}$$

where the last equality follows from the assumption (40). ∎

Proof of Theorem 1: Fix some joint distribution $P_{V_1,V_2,S_1,S_2,X_1,X_2,Y}$ in $\mathcal{P}^{\mathrm{ind}}_{\text{s-c}}$, and let R_1 and R_2 satisfy inequalities (11). We need to show that (R_1, R_2) is achievable. If $I(X_1,X_2;Y)$ is zero, then Lemma 1 implies that both R_1 and R_2 must be zero and hence achievable. It thus remains to prove achievability when

$$I(X_1,X_2;Y) > 0 \tag{45}$$

as we henceforth assume.

As we next argue, (45) implies that $I(X_1;Y|X_2)$ and $I(X_2;Y|X_1)$ cannot be both zero. Indeed, the independence of X_1 and X_2 implies that $I(X_1;Y) \le I(X_1;Y|X_2)$ and $I(X_2;Y) \le I(X_2;Y|X_1)$, so if both $I(X_1;Y|X_2)$ and $I(X_2;Y|X_1)$ are zero, then so are both $I(X_1;Y)$ and $I(X_2;Y)$ and hence so is $I(X_1;Y) + I(X_2;Y|X_1)$, which contradicts (45). In the rest of the proof, we shall treat the case where

$$I(X_1;Y|X_2) > 0; \tag{46}$$

the case where this is violated but $I(X_2;Y|X_1)$ is positive is symmetric.

The significance of (46) is that it guarantees that—by having Encoder 2 transmit a deterministic sequence—Encoder 1 can ignore the state information and still communicate at a positive rate to the uninformed decoder. As we shall see, this will be critical in the termination of the block-Markov scheme. Also critical to the termination of the block-Markov scheme is the assumption

$$I(X_2;Y,S_1|X_1) > 0 \tag{47}$$

which we can make because if this is violated, then achievability is guaranteed by Lemma 2. The significance of (47) is that it guarantees that—by having Encoder 1 transmit a deterministic sequence—Encoder 2 can ignore its state information and communicate at a positive rate to the decoder, provided that the decoder is made cognizant (at some later phase of the coding scheme) of the prevailing S_1-sequence.

Using the Fourier–Motzkin elimination, it can be shown that the fact that the rates R_1, R_2 satisfy inequalities (11) is equivalent to the existence of rates $R_0^{(1)}, R_0^{(2)} \ge 0$ such that

$$0 \le R_1 + R_0^{(1)} \le I(X_1;Y|X_2,V_1,V_2) \tag{48a}$$
$$0 \le R_2 + R_0^{(2)} \le I(X_2;Y|X_1,V_1,V_2) \tag{48b}$$
$$R_1 + R_2 + R_0^{(1)} + R_0^{(2)} \le I(X_1,X_2;Y|V_1,V_2) \tag{48c}$$

and

$$R_0^{(1)} \ge I(V_1;S_1|V_2) - I(V_1;Y|V_2) \tag{49a}$$
$$R_0^{(2)} \ge I(V_2;S_2|V_1) - I(V_2;Y|V_1) \tag{49b}$$
$$R_0^{(1)} + R_0^{(2)} \ge I(V_1,V_2;S_1,S_2) - I(V_1,V_2;Y). \tag{49c}$$

Indeed, form (9) of the joint PMF implies that the sum of (48a) and (49a) yields (11a). Similarly, (48b) and (49b) yield (11b). The sum-rate bound (11c) is obtained by summing (48c) with (49c). Next, observe that

$$\begin{aligned} &I(V_1,V_2;S_1,S_2|Y) \\ &\quad = I(V_1;S_1|Y) + I(V_1;S_2|Y,S_1) \\ &\qquad + I(V_2;S_2|Y,V_1) + I(V_2;S_1|Y,V_1,S_2) \\ &\quad = I(V_1;S_1|Y) + I(V_2;S_2|Y,V_1) \\ &\quad \ge I(V_1;S_1|Y,V_2) + I(V_2;S_2|Y,V_1) \end{aligned} \tag{50}$$

where the second equality and the inequality in (50) hold due to the Markov relations (10a), (10b). By (50), the sum-rate bound (11c) obtained by summing (48c) with (49c) dominates the sum of (48c), (49a), and (49b). Similarly, it can be shown that (11c) dominates all other sum-rate bounds that can be obtained from (48) and (49).

Having established that the condition that the rates R_1, R_2 satisfy inequalities (11) implies (and is, in fact, equivalent to) the existence of rates $R_0^{(1)}, R_0^{(2)} \geq 0$ satisfying (48) and (49), we next interpret the two sets of inequalities.

Inequalities (48) have a channel-coding flavor. They guarantee that the rates

$$\left(R_1 + R_0^{(1)}, R_2 + R_0^{(2)}\right)$$

be achievable on the MAC when the decoder, in addition to the channel output, is also cognizant of V_1 and V_2. Inequalities (49) have a distributed source-coding flavor. They guarantee that in the distributed Wyner–Ziv network [4] where one describing terminal observes S_1, the second observes S_2, and the reconstructor observes Y, the description rates $R_0^{(1)}, R_0^{(2)}$ allow the reconstructor to reconstruct S_1 as V_1 and reconstruct S_2 as V_2.

These interpretations motivate the following block-Markov scheme. The scheme consists of B length-n blocks followed by three terminating blocks of lengths n_1, n_2, and n_3 (to be specified shortly). In Block 1, Encoder 1 transmits fresh information at rate R_1, and Encoder 2 transmits fresh information at rate R_2. In Blocks 2 through B, Encoder 1 transmits nR_1 fresh bits and $nR_0^{(1)}$ bits that describe the length-n state sequence it observed in the previous block. Likewise Encoder 2. The description of the state sequences from the previous block that Encoders 1 and 2 use is based on a binning scheme that was proposed in [4] for the Wyner–Ziv distributed source-coding problem where Encoder 1 describes the state sequence it observed in the previous block, Encoder 2 describes the state sequence it observed in the previous block, and the reconstructor has the n-length channel outputs from the previous block as SI. Constraints (49) guarantee that if the decoder succeeds in decoding the bits sent in the block, it will be able to produce a reconstruction sequence of joint type $P_{V_1,S}$ with the state sequence observed by Encoder 1 in the previous block, and it will also be able to produce a reconstruction sequence of joint type $P_{V_2,S}$ with the state sequence observed by Encoder 2 in the previous block. Constraints (48) guarantee that if (using backward decoding) the decoder is given the description of the states pertaining to the given block, it will be able to decode the bits sent by the two encoders in the block—i.e., both those comprising fresh information and those describing the states pertaining to the previous block. If the decoder were given the description of the states pertaining to Block B, it could then decode all the data using backward decoding: Using the description of the states pertaining to Block B, it would decode the fresh information of Block B as well as the description of the states pertaining to Block $(B-1)$. Using the latter, it would then be able to decode the fresh bits sent in Block $(B-1)$ as well as the description of the states pertaining to Block $(B-2)$, etc.

The purpose of the three terminating block is to provide this information to the receiver, i.e., to convey to the receiver the description of the states pertaining to block B. This is done as follows. In Block $(B+1)$, which is roughly of length $nR_0^{(1)}/I(X_1;Y|X_2)$ [and which is finite by (46)], Encoder 2 sends no data (i.e., sends a deterministic sequence of type P_{X_2}) and Encoder 1 sends the (lossy) description of the state sequence it observed in Block B. After decoding Block $B+1$, the receiver can obtain a description of the state sequence that was observed by Encoder 1 in Block B. The remaining two blocks are used to convey the lossy description of the state sequence that was observed by Encoder 2 in Block B. In Block $(B+2)$, which is roughly of length $nR_0^{(2)}/I(X_2;Y,S_1|X_1)$ (and which is finite by (47)), Encoder 1 sends no data (i.e., sends a deterministic sequence of type P_{X_1}) and Encoder 2 sends the lossy description of the state sequence it observed in Block B. The decoder, however, cannot yet decode this description because the rate $I(X_2;Y,S_1|X_1)$ is not achievable unless the decoder is informed of S_1. This is done in Block $(B+3)$, which is roughly of length $nH(S_1)/I(X_1;Y|X_2)$, where Encoder 2 sends no data and Encoder 1 sends a lossless description of the state sequence it observed in Block $(B+2)$.

Decoding is thus performed as follows: The receiver first decodes Block $(B+3)$ and thus losslessly learns the state sequence that was observed by Encoder 1 in Block $(B+2)$. Using this state sequence, it then decodes Block $(B+2)$ and thus learns the lossy description of the state that was observed by Encoder 2 in Block B. By Decoding Block $(B+1)$, it also learns the description of the state sequence that was observed by Encoder 1 in Block B. Armed with the lossy descriptions of the two state sequences pertaining to Block B, it can now decode Block B and use backward decoding to decode all the remaining blocks $(B-1)$ through 1.

The rates supported by our scheme approach (R_1, R_2) as B tends to infinity. Indeed, Encoder 1 sends a total of BnR_1 data bits, and the total number of channel uses is roughly

$$Bn + \frac{nR_0^{(1)}}{I(X_1;Y|X_2)} + \frac{nR_0^{(2)}}{I(X_2;Y,S_1|X_1)} + \frac{nH(S_1)}{I(X_1;Y|X_2)}$$

so the ratio tends to R_1 as B tends to infinity. Similarly, the rate of Encoder 2 approaches R_2.

We now give a more detailed proof. Recall that since (R_1, R_2) satisfy inequalities (11), there exist $R_1', R_2' > 0$ such that both (48) and (49) holds. Fix such R_1', R_2'. By (46) and (47), there exist $\mu_1, \mu_2 > 0$ such that

$$I(X_1;Y|X_2) > \mu_1 > 0 \tag{51a}$$

and

$$I(X_2;Y,S_1|X_1) > \mu_2 > 0. \tag{51b}$$

As explained earlier, the proof is based on a block-Markov coding scheme with backward decoding. The total transmission time is divided into $B+3$ blocks. Each of the first B blocks is of length n. Block $B+1$ is of length n_1, Block $B+2$ is of length n_2, and Block $B+3$ is of length n_3. In terms of $R_1', R_2', \mu_1,$ and μ_2, we can now specify the lengths of the terminating blocks as

$$n_1 = \frac{nR_1'}{\mu_1} \tag{52a}$$

$$n_2 = \frac{nR_2'}{\mu_2} \tag{52b}$$

$$n_3 = \frac{n_2\left(H(S_1) + \tilde{\delta}\right)}{\mu_1} = \frac{nR_2'\left(H(S_1) + \tilde{\delta}\right)}{\mu_1 \mu_2} \tag{52c}$$

where $\tilde{\delta} > 0$ is some arbitrary fixed constant. Our scheme is thus of rates

$$(R_1, R_2) \cdot \frac{nB}{nB + n_1 + n_2 + n_3} = (R_1,R_2) \cdot \frac{B}{B+R_1'/\mu_1+R_2'/\mu_2+\big(H(S_1)+\tilde{\delta}\big)R_2'/(\mu_1\mu_2)} \quad (53)$$

which approaches (R_1, R_2) for large B.

The blocks are indexed by b, where $b \in \{1, 2, \ldots, B + 3\}$. Denote by $\mathbf{s}_{q,b}$ the state sequence observed by User q during Block b, i.e.,

$$\mathbf{s}_{q,b} = (s_{q,(b-1)n+1}, \ldots, s_{q,bn}),\ q \in \{1, 2\},\ b \in [1 : B]$$
$$\mathbf{s}_{q,B+1} = (s_{q,Bn+1}, \ldots, s_{q,Bn+n_1}),\ q \in \{1, 2\}$$
$$\mathbf{s}_{q,B+2} = (s_{q,Bn+n_1+1}, \ldots, s_{q,Bn+n_1+n_2}),\ q \in \{1, 2\}$$
$$\mathbf{s}_{q,B+3} = (s_{q,Bn+n_1+n_2+1}, \ldots, s_{q,Bn+n_1+n_2+n_3}), q \in \{1, 2\}.$$

In the rest of the proof, we use the definitions of typical sequences and typical sets as in [2]. Thus, $n(a|\mathbf{u})$ is the number of occurrences of the letter a in the n-vector $\mathbf{u} = u^n$. And, for a given PMF P_U over the finite alphabet $\mathcal{U}$, the δ-typical set is

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

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

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

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

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

We further adopt the *Delta-Convention* of [2, Convention 2.11] (see also lemmas 2.12 and 2.13 there). Thus, throughout the proof, the sequences δ_n are fixed, and the dependence of the typical sets on them is omitted.

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

Codebook Generation. Pick real numbers R_{s_1}, R_{s_2} satisfying

$$R_1' \le R_{s_1}, \quad R_2' \le R_{s_2} \quad (54)$$

in a fashion that will be specified later. For each block $b \in [1 : B]$, the codebook is constructed in four steps, as described below.

1) Generate $2^{nR_{s_1}}$ vectors $\mathbf{v}_1(j_1)$, $j_1 = 1, 2, \ldots, 2^{nR_{s_1}}$, IID according to $P_{V_1}(\cdot)$. Randomly partition the indices $\{j_1 : 1 \le j_1 \le 2^{nR_{s_1}}\}$ into $2^{nR_1'}$ bins. Denote by $k_1(j_1)$ the index of the bin to which j_1 belongs, and by $\alpha_1(k_1)$ bin number k_1.
2) Generate $2^{n(R_1+R_1')}$ vectors $\mathbf{x}_1(l, k_1)$, $l = 1, 2 \ldots, 2^{nR_1}$, $k_1 = 1, 2, \ldots, 2^{nR_1'}$, IID, according to $P_{X_1}(\cdot)$.
3) Similarly, generate $2^{nR_{s_2}}$ vectors $\mathbf{v}_2(j_2)$, $j_2 = 1, 2, \ldots 2^{nR_{s_2}}$, IID according to $P_{V_2}(\cdot)$. Randomly partition the indices $\{j_2 : \ 1 \le j_2 \le 2^{nR_{s_2}}\}$ into $2^{nR_2'}$ bins. Denote by $k_2(j_2)$ the index of the bin to which j_2 belongs, and by $\alpha_2(k_2)$ bin number k_2.
4) Generate $2^{n(R_2+R_2')}$ vectors $\mathbf{x}_2(m, k_2)$, $m = 1, 2 \ldots, 2^{nR_2}$, $k_2 = 1, 2, \ldots, 2^{nR_2'}$, IID, according to $P_{X_2}(\cdot)$.

The codebook generation and partition as described above are repeated independently B times, with the same distribution and rates. The last three blocks are devoted to the transmission of the description of the state sequences of block B. They carry no fresh information. The construction of codebooks for the last three blocks is described next.

Block $B + 1$.
1) Generate one length-n_1 codeword $\mathbf{x}_{2,B+1}$ IID P_{X_2}.
2) Generate $2^{nR_1'}$ independent length-n_1 codewords $\mathbf{x}_1(k_1)$ for $k_1 \in \big[1 : 2^{nR_1'}\big]$ each IID P_{X_1}.

Block $B + 2$
1) Generate one length-n_2 codeword $\mathbf{x}_{1,B+2}$ IID P_{X_1}.
2) Generate $2^{nR_2'}$ length-n_2 codewords $\mathbf{x}_2(k_2)$ for $k_2 \in \big[1 : 2^{nR_2'}\big]$ each IID P_{X_2}.

Block $B + 3$.
1) Generate $2^{n_2(H(S_1)+\tilde{\delta})}$ length-n_3 codewords $\mathbf{x}_1(k_3)$ for $k_3 \in \big[1 : 2^{n_2(H(S_1)+\tilde{\delta})}\big]$ each IID P_{X_1}.
2) Generate one length-n_3 codeword $\mathbf{x}_{2,B+3}$ IID P_{X_2}.

For notational convenience, we omit the dependence of the codebooks and codewords lengths on the block number b. It will be clear from the context whether the codewords are of length n, n_1, n_2, or n_3.

Reveal the codebooks to the encoders and decoder.

Encoding. Let $l_b \in \{1, \ldots, 2^{nR_1}\}$ and $m_b \in \{1, \ldots, 2^{nR_2}\}$ be the message indices of the users in Block b. 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, $b \in [1 : B]$. User q is cognizant of $\mathbf{s}_{q,b-1}$ and inspects the $\mathbf{v}_q$ sequences that were generated in Block $b - 1$ to find the first index $j_q \in \{1, \ldots, 2^{nR_{s_q}}\}$ such that

$$\big(\mathbf{v}_q(j_q), \mathbf{s}_{q,b-1}\big) \in \mathcal{T}_{V_q,S_q}. \quad (55)$$

Denote this index by $j_{q,b-1}$. If a vector $\mathbf{v}_q(j_q)$ satisfying (55) does not exist, the user picks a default index, say $j_{q,b-1} = 1$. Denote by $k_{q,b}$ the bin number to which the index $j_{q,b-1}$ belongs. The inputs to the channel are

$$\begin{aligned} &\text{User 1:} \quad \mathbf{x}_1(l_b, k_{1,b}) \\ &\text{User 2:} \quad \mathbf{x}_2(m_b, k_{2,b}). \end{aligned} \quad (56)$$

Block $B+1$. In this block, only User 1 transmits the codeword of the compressed state sequence $\mathbf{s}_{1,B}$. Being cognizant of $\mathbf{s}_{1,B}$, User 1 inspects the $\mathbf{v}_1$ sequences that were generated in Block B and selects the first index $j_1 \in \{1, \ldots, 2^{nR_{s_1}}\}$ such that

$$\big(\mathbf{v}_1(j_1), \mathbf{s}_{1,B}\big) \in \mathcal{T}_{V_1,S_1}. \quad (57)$$

Denote this index by $j_{1,B}$. If a vector $\mathbf{v}_1(j_1)$ satisfying (57) does not exist, the user picks a default index, say $j_{1,B} = 1$. Denote by $k_{1,B+1}$ the bin number to which the index $j_{1,B}$ belongs. The inputs to the channel are

$$\begin{aligned} &\text{User 1:} \quad \mathbf{x}_1(k_{1,B+1}) \\ &\text{User 2:} \quad \mathbf{x}_{2,B+1}. \end{aligned} \tag{58}$$

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

Block $B+2$. In this block, User 2 transmits the codeword of the compressed state sequence $\mathbf{s}_{2,B}$. Being cognizant of $\mathbf{s}_{2,B}$, User 2 inspects the $\mathbf{v}_2$ sequences that were generated in Block B and selects the first index $j_2 \in \{1,\dots,2^{nR_{s_2}}\}$ such that

$$\big(\mathbf{v}_2(j_2), \mathbf{s}_{2,B}\big) \in \mathcal{T}_{V_2,S_2}. \tag{59}$$

Denote this index by $j_{2,B}$. If a vector $\mathbf{v}_2(j_2)$ satisfying (59) does not exist, the user picks a default index, say $j_{2,B}=1$. Denote by $k_{2,B+1}$ the bin number to which the index $j_{2,B}$ belongs. The inputs to the channel are

$$\begin{aligned} &\text{User 1:} \quad \mathbf{x}_{1,B+2} \\ &\text{User 2:} \quad \mathbf{x}_2(k_{2,B+1}). \end{aligned} \tag{60}$$

Block $B+3$. In this block, User 1 transmits an almost-lossless description of $\mathbf{s}_{1,B+2}$ using, for example, the scheme in [1, Sec. 7.13] for transmitting a source over a noisy channel. Thus, if $k_{1,B+3}$ is the index of $\mathbf{s}_{1,B+2}$ in the set of all $\tilde{\delta}$-typical sequences, (with $k_{1,B+3}$ being set to one if $\mathbf{s}_{1,B+2}$ is not typical), it transmits $\mathbf{x}_1(k_{1,B+3})$:

$$\begin{aligned} &\text{User 1:} \quad \mathbf{x}_1(k_{1,B+3}) \\ &\text{User 2:} \quad \mathbf{x}_{2,B+3}. \end{aligned} \tag{61}$$

Decoding. Let $\mathbf{y}_b$ be the channel output at Block b. Decoding begins at Block $B+3$ and proceeds backwards.

Block $B+3$. In this block, the decoder recovers (with high probability) $\mathbf{s}_{2,B+2}$. It looks for an index $\hat{k}_{1,B+3}$ such that

$$\big(\mathbf{x}_1(\hat{k}_{1,B+3}), \mathbf{x}_{2,B+3}, \mathbf{y}_{B+3}\big) \in \mathcal{T}_{X_1,X_2,Y}. \tag{62}$$

If an index $\hat{k}_{1,B+3}$ satisfying (62) does not exist, or is not unique, an error is declared. Otherwise, it sets $\hat{\mathbf{s}}_{1,B+2}$ to be the sequence whose index is $\hat{k}_{1,B+3}$ in the set of $\tilde{\delta}$-typical sequences.

Block $B+2$. The decoder has the output $\mathbf{y}_{B+2}$, and it also has the estimate $\hat{\mathbf{s}}_{1,B+2}$ obtained in the previous decoding step. It looks for an index $\hat{k}_{2,B+2}$ such that

$$\big(\mathbf{x}_{1,B+2}, \mathbf{x}_2(\hat{k}_{2,B+2}), \hat{\mathbf{s}}_{1,B+2}, \mathbf{y}_{B+2}\big) \in \mathcal{T}_{X_1,X_2,S_1,Y}. \tag{63}$$

If an index $\hat{k}_{2,B+2}$ satisfying (63) does not exist, or is not unique, an error is declared.

Block $B+1$. Here, the decoder decodes the compressed state sequence of User 1 in block B. The decoding proceeds as in Block $B+3$: The decoder looks for an index $\hat{k}_{1,B+1}$ such that

$$\big(\mathbf{x}_1(\hat{k}_{1,B+1}), \mathbf{x}_{2,B+1}, \mathbf{y}_{B+1}\big) \in \mathcal{T}_{X_1,X_2,Y}. \tag{64}$$

If an index $\hat{k}_{1,B+1}$ satisfying (64) does not exist, or is not unique, an error is declared.

Block b, $2 \le b \le B$. The decoder has at hand the pair $(\hat{k}_{1,b+1}, \hat{k}_{2,b+1})$, and the channel output $\mathbf{y}_b$. It looks for $\mathbf{v}_1(\hat{j}_{1,b}) \in \alpha_1(\hat{k}_{1,b+1})$ and $\mathbf{v}_2(\hat{j}_{2,b}) \in \alpha_2(\hat{k}_{2,b+1})$, such that

$$(\mathbf{v}_1(\hat{j}_{1,b}), \mathbf{v}_2(\hat{j}_{2,b}), \mathbf{y}_b) \in \mathcal{T}_{V_1,V_2,Y}. \tag{65}$$

If such a pair does not exist, or is not unique, an error is declared.

If decoded correctly, the pair $(\mathbf{v}_1(\hat{j}_{1,b}), \mathbf{v}_2(\hat{j}_{2,b}))$ consists of the compressed state sequences in Block b. This information on the states facilitates the decoding of the messages (l_b, m_b) and the indices $k_{1,b}$, $k_{2,b}$, which are the bin numbers of the states in Block $b-1$. Specifically, the decoder looks for the indices $(\hat{l}_b, \hat{m}_b, \hat{k}_{1,b}, \hat{k}_{2,b})$ such that

$$\begin{aligned} &(\mathbf{x}_1(\hat{l}_b, \hat{k}_{1,b}), \mathbf{x}_2(\hat{m}_b, \hat{k}_{2,b}), \mathbf{v}_1(\hat{j}_{1,b}), \mathbf{v}_2(\hat{j}_{2,b}), \mathbf{y}_b) \in \\ &\mathcal{T}_{X_1,X_2,V_1,V_2,Y}. \end{aligned} \tag{66}$$

If there is no quadruple $(\hat{l}_b, \hat{m}_b, \hat{k}_{1,b}, \hat{k}_{2,b})$ satisfying (66), or there is more than one such quadruple, an error is declared.

Block 1. Since it is the first block, there is no need to decode the bin indices (k_1, k_2). The decoder operates exactly as in Blocks $b \in \{2,\dots,B\}$, except that $\hat{k}_{1,1}$ and $\hat{k}_{2,1}$ are 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. Standard techniques related to single-user channels show that (51) and (52) guarantee that $(k_{1,B+1}, k_{2,B+1})$ are decoded with small probability of error. We thus focus on the decoding of the first B blocks and specifically on the decoding of (l_b, m_b), $b \in [1:B]$ assuming that the tuple $(k_{1,B+1}, k_{2,B+1})$ was decoded correctly. Without loss of generality, we assume that a specific sequence of pairs $(k_{1,2}^{B+3}, k_{2,2}^{B+3})$ is chosen, and that $(l_b, m_b) = (1,1)\ \forall b$.

Fix state sequences $\mathbf{s}_q^B = (\mathbf{s}_{q,1},\dots,\mathbf{s}_{q,B})$, $q=1,2$, and define the events

$$\begin{aligned}
&\mathcal{A}_{1,b}(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) = \big\{(\mathbf{v}_q(j_q), \mathbf{s}_{q,b}) \in \mathcal{T}_{V_q,S_q} \\
&\qquad\qquad \text{for some } j_q \in \{1,\dots,2^{nR_{s_q}}\}, \quad q=1,2\big\} \\
&\mathcal{A}_{2,1}(\mathbf{s}_{1,1}, \mathbf{s}_{2,1}) = \\
&\qquad\qquad \{(\mathbf{x}_1(1,1), \mathbf{x}_2(1,1), \mathbf{s}_{1,1}, \mathbf{s}_{2,1}) \in \mathcal{T}_{X_1,X_2,S_1,S_2}\} \\
&\mathcal{A}_{3,b}(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) = \\
&\qquad\qquad \{(\mathbf{x}_1(1,k_{1,b}), \mathbf{x}_2(1,k_{2,b}), \mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \in \mathcal{T}_{X_1,X_2,S_1,S_2}\} \\
&\mathcal{B}_{2,b}(j_1, j_2) = \{(\mathbf{v}_1(j_1), \mathbf{v}_2(j_2), \mathbf{y}_b) \in \mathcal{T}_{V_1,V_2,Y}\} \\
&\mathcal{B}_{3,b}(l, k_1, m, k_2, j_1, j_2) = \\
&\{(\mathbf{x}_1(l,k_1), \mathbf{x}_2(m,k_2), \mathbf{v}_1(j_1), \mathbf{v}_2(j_2), \mathbf{y}_b) \in \mathcal{T}_{X_1,X_2,V_1,V_2,Y}\}. \\
&\mathcal{A} = \mathcal{A}_{2,1}(\mathbf{s}_{1,1}, \mathbf{s}_{2,1}) \bigcap_{b=1}^{B} \mathcal{A}_{1,b}(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \bigcap_{b=2}^{B+1} \mathcal{A}_{3,b}(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \\
&\beta_b = \mathcal{B}_{2,b}^c(j_{1,b}, j_{2,b}) \bigcup_{\substack{j_1 \in \alpha_1(k_{1,b+1}) \\ j_2 \in \alpha_2(k_{2,b+1}) \\ (j_1,j_2) \neq (j_{1,b},j_{2,b})}} \mathcal{B}_{2,b}(j_1, j_2)
\end{aligned} \tag{67}$$

$$\begin{aligned}
&\gamma_b = \mathcal{B}_{3,b}^c(1, k_{1,b}, 1, k_{2,b}, j_{1,b}, j_{2,b}) \\
&\qquad \bigcup_{\substack{(l,k_1,m,k_2) \neq \\ (1,k_{1,b},1,k_{2,b})}} \mathcal{B}_{3,b}(l, k_1, m, k_2, j_{1,b}, j_{2,b}).
\end{aligned} \tag{68}$$

It is enough to show that

$$\lim_{n\to\infty} P(\mathcal{E}|\mathbf{s}_1^B,\mathbf{s}_2^B)=0 \quad \forall(\mathbf{s}_1^B,\mathbf{s}_2^B)\in\mathcal{T}_{S_1,S_2}^B \tag{69}$$

where $\mathcal{E}=\cup_{b=1}^B\{(\hat{l}_b,\hat{m}_b)\neq(l_b,m_b)\}$ is the error event, and $\mathcal{T}_{S_1,S_2}^B$ is the B product of $\mathcal{T}_{S_1,S_2}$. The probability of error conditioned on the pair $(\mathbf{s}_1^B,\mathbf{s}_2^B)$ can be bounded as

$$\begin{aligned}P(\mathcal{E}|\mathbf{s}_1^B,\mathbf{s}_2^B)&\le P(\mathcal{A}^c|\mathbf{s}_1^B,\mathbf{s}_2^B)+\\&\quad\big[P(\beta_B|\mathbf{s}_1^B,\mathbf{s}_2^B,\mathcal{A})\\&\quad+P(\gamma_B|\mathbf{s}_1^B,\mathbf{s}_2^B,\mathcal{A},\beta_B^c)\big]\\&\quad+\sum_{b=1}^{B-1}\big[P(\beta_b|\mathbf{s}_1^{B+1},\mathbf{s}_2^{B+1},\mathcal{A},\gamma_{b+1}^c)\\&\quad+P(\gamma_b|\mathbf{s}_1^B,\mathbf{s}_2^B,\mathcal{A},\gamma_{b+1}^c,\beta_b^c)\big]\\&=P(\mathcal{A}^c|\mathbf{s}_1^B,\mathbf{s}_2^B)\\&\quad+P(\beta_B|\mathbf{s}_1^B,\mathbf{s}_2^B,\mathcal{A})\\&\quad+P(\gamma_B|\mathbf{s}_1^B,\mathbf{s}_2^B,\mathcal{A},\beta_B^c)\\&\quad+(B-1)[P(\beta_b|\mathbf{s}_1^B,\mathbf{s}_2^B,\mathcal{A},\gamma_{b+1}^c)\\&\quad+P(\gamma_b|\mathbf{s}_1^B,\mathbf{s}_2^B,\mathcal{A},\gamma_{b+1}^c,\beta_b^c)]\end{aligned} \tag{70}$$

for some fixed b. We turn now to bound each of the terms in the RHS of (70). For notational convenience, we drop the conditioning on the state sequences. By classical results on source coding

$$\lim_{n\to\infty}P(\mathcal{A}^c)=0 \tag{71}$$

whenever

$$R_{s_q}>I(V_q;S_q)\quad q=1,2. \tag{72}$$

Conditioned on $\mathcal{A}$, the state sequences and the channel input vectors are all typical. Recall that we assume that the decoder has at hand the correct bin indices $(k_{1,B+1},k_{2,B+1})$. Moreover, conditioned on γ_{b+1}^c, the decoder has at hand the correct bin indices of the last block, i.e., $(\hat{k}_{1,b+1},\hat{k}_{2,b+1})=(k_{1,b+1},k_{2,b+1})$, for $b\in[1:B-1]$. Hence, β_b, $b\in[1:B]$, is the decoding error event in distributed Wyner–Ziv coding. We next evaluate the probability of the union in the RHS of (67):

$$\begin{aligned}P\Big(\bigcup_{\substack{j_1\in\alpha_1(k_{1,B+1})\\ j_1\neq j_{1,B+1}}}\mathcal{B}_{2,B}(j_1,j_{2,B+1})\,|\,\mathcal{A}\Big)\\ \le 2^{n(R_{s_1}-R_1')}2^{-n[I(V_1;V_2,Y)-\delta]}\end{aligned} \tag{73}$$

for some $\delta>0$, $\lim_{n\to\infty}\delta=0$. Similarly

$$\begin{aligned}P\Big(\bigcup_{\substack{j_2\in\alpha_2(k_{2,B+1})\\ j_2\neq j_{2,B+1}}}\mathcal{B}_{2,B}(j_{1,B+1},j_2)\,|\,\mathcal{A}\Big)\\ \le 2^{n(R_{s_2}-R_2')}2^{-n[I(V_2;V_1,Y)-\delta]}\end{aligned} \tag{74}$$

$$\begin{aligned}P\Big(\bigcup_{\substack{j_1\in\alpha_1(k_{1,b+1})\\ j_2\in\alpha_2(k_{2,b+1})\\ j_1\neq j_{1,b}\ j_2\neq j_{2,b}}}\mathcal{B}_{2,B}(j_1,j_2)\,|\,\mathcal{A}\Big)\\ \le 2^{n(R_{s_1}+R_{s_2}-R_1'-R_2')}2^{-n[I(V_1,V_2;Y)-\delta]}.\end{aligned} \tag{75}$$

Therefore, we conclude that

$$\lim_{n\to\infty}P(\beta_B|\mathcal{A})=0$$

if

$$\begin{aligned}R_{s_1}-R_1'&<I(V_1:Y|V_2)\\R_{s_2}-R_2'&<I(V_2;Y|V_1)\\R_{s_1}+R_{s_2}-R_1'-R_2'&<I(V_1,V_2;Y)\end{aligned} \tag{76}$$

where we used the independence of V_1 and V_2.

We proceed to bound the probability of γ_B conditioned on $(\mathcal{A},\beta_B^c)$. Observe that conditioned on β_B^c, the decoder has at hand the compressed version of the state at Block B, $(\mathbf{v}_1(j_{1,B}),\mathbf{v}_2(j_{2,B}))$, and this pair is independent of the inputs in Block B. Decompose the union in (68) as

$$\begin{aligned}&\bigcup_{\substack{(l,k_1,m,k_2)\neq\\(1,k_{1,b},1,k_{2,b})}}\mathcal{B}_{3,b}(l,k_1,m,k_2,j_{1,b},j_{2,b})\\&=\bigcup_{(l,k_1)\neq(1,k_{1,b})}\mathcal{B}_{3,b}(l,k_1,1,k_{2,b},j_{1,b},j_{2,b})\\&\quad\bigcup_{(m,k_2)\neq(1,k_{2,b})}\mathcal{B}_{3,b}(1,k_{1,b},m,k_2,j_{1,b},j_{2,b})\\&\quad\bigcup_{\substack{(l,k_1)\neq(1,k_{1,b})\\(m,k_2)\neq(1,k_{2,b})}}\mathcal{B}_{3,b}(l,k_1,m,k_2,j_{1,b},j_{2,b}).\end{aligned} \tag{77}$$

Therefore, we have

$$\lim_{n\to\infty}P(\gamma_B|\mathcal{A},\beta_B^c)=0$$

provided

$$\begin{aligned}R_1+R_1'&<I(X_1;Y|X_2,V_1,V_2)\\R_2+R_2'&<I(X_2;Y|X_1,V_1,V_2)\\R_1+R_2+R_1'+R_2'&<I(X_1,X_2;Y|V_1,V_2).\end{aligned} \tag{78}$$

The terms $P(\beta_b|\mathcal{A},\gamma_{b+1}^c)$ and $P(\gamma_b|\mathcal{A},\gamma_{b+1}^c,\beta_b^c)$ are treated exactly as $P(\beta_B|\mathcal{A})$ and $P(\gamma_B|\mathcal{A},\beta_B^c)$: the extra conditioning on γ_{b+1}^c only means that the decoder has at hand the bin indices $(k_{1,b+1},k_{2,b+1})$ as we assumed for the last block $b=B$. Hence, they yield the same rate constraints, i.e., (76) and (78). The rate constraints (72) and (76) are equivalent to

$$\begin{aligned}R_1'&>I(V_1;S_1|V_2)-I(V_1;Y|V_2)\\R_2'&>I(V_2;S_2|V_1)-I(V_2;Y|V_1)\\R_1'+R_2'&>I(V_1,V_2;S_1,S_2)-I(V_1,V_2;Y)\end{aligned} \tag{79}$$

where we used the independence of V_1, V_2, S_1, S_2. The rate constraints (78) and (79) are equivalent to (48)–(49). Moreover, (71) guarantees that the input constraints are satisfied. This concludes the proof of Theorem 1. ∎

F. Analysis of Example 2

By ignoring the state, we can achieve the region corresponding to (14) but with the additional constraint $R_2\le\frac{1}{2}\log(1+\frac{\Gamma_2}{\sigma_{s_1}^2})$. Since the convex combination of this region with the rate-point

$$R_1=0,\qquad R_2=\frac{1}{2}\log\left(1+\frac{\Gamma_1+\Gamma_2}{\sigma_{s_1}^2}\right) \tag{80}$$

yields the rate region (14) (without the additional constraint), a time-sharing argument shows that to prove that the rate region (14) is achievable, it suffices to prove that the rate-pair (80) is achievable, which is what we proceed to do.

To this end, set $R_1 = 0$ and make the following substitutions in the definition of the set $\mathcal{R}^{\text{ind}}_{\text{s-c}}$. Set $V_2 = 0$; V_1 zero mean and jointly Gaussian with S_1; and $X_1 \sim \mathcal{N}(0, \Gamma_1)$, $X_2 \sim \mathcal{N}(0, \Gamma_2)$ independent of each other and of (V_1, S_1). Then, (11a)–(11c) reduce to

$$0 \le \frac{1}{2} \log \frac{(\Gamma_1 + \sigma^2_{s_1|v_1})(\Gamma_1 + \Gamma_2 + \sigma^2_{s_1})}{\sigma^2_{s_1}(\Gamma_1 + \Gamma_2 + \sigma^2_{s_1|v_1})} \tag{81}$$

$$R_2 \le \frac{1}{2} \log \left(1 + \frac{\Gamma_2}{\sigma^2_{s_1|v_1}}\right) \tag{82}$$

$$R_2 \le \frac{1}{2} \log \left(1 + \frac{\Gamma_1 + \Gamma_2}{\sigma^2_{s_1}}\right) \tag{83}$$

where $\sigma^2_{s_1|v_1}$ is the variance of S_1 conditioned on V_1. For any positive $\sigma^2_{s_1|v_1}$ satisfying

$$\frac{\Gamma_2\, \sigma^2_{s_1}}{\Gamma_1 + \Gamma_2} - \Gamma_1 \le \sigma^2_{s_1|v_1} \le \frac{\Gamma_2\, \sigma^2_{s_1}}{\Gamma_1 + \Gamma_2} \tag{84}$$

the bound (83) dominates (82) and the RHS of (81) is positive. We thus conclude that (80) is achievable. By the time-sharing argument, this also proves the achievability of the region (14).

We next have to show that no rate-pair outside the region (14) is achievable. This can be shown by recalling the capacity region of the Gaussian MAC without state information and by recalling Propositions 1 and 2.

G. *Analysis of Example 3*

The rate pair $(R_1, R_2) = (1, 1/2)$ is in the inner bound of Li *et al.*. To see this, we set $V_1 = 0$ and $V_2 = W_{X_2}$ with X_1, X_2 IID random bits. However, as we next prove, the pair $(1, 1/2)$ is not in $\mathcal{R}^{\text{ind}}_{\text{s-c}}$.

We prove this by showing that if $(1, R_2)$ is in $\mathcal{R}^{\text{ind}}_{\text{s-c}}$, then R_2 must be zero. Suppose then that $(1, R_2) \in \mathcal{R}^{\text{ind}}_{\text{s-c}}$. Since S_1 is null, it follows from the structure (9) of the joint distribution that V_1 must be independent of all the other random variables. Consequently, we can strike it out from (11a)–(11c). Since $R_1 = 1$, it follows from (11a) that X_1 must be $\text{Bernoulli}(1/2)$ and that $H(X_1|X_2, V_2, Y)$ must be zero. This implies that $H(W_{X_2}|X_2, V_2, Y)$ must also be zero (because $X_1 = Y_1 \oplus W_{X_2}$). Consequently, $H(W_{X_2}|X_2, V_2, Y_1)$ must also be zero (because $Y_2 = X_2$). This implies that

$$H(W_{X_2}|X_2, V_2) = 0 \tag{85a}$$

because X_1 is $\text{Bernoulli}(1/2)$ and independent of (X_2, V_2, W), so Y_1, which is equal to $X_1 \oplus W_{X_2}$, must also be independent of (X_2, V_2, W).

We now distinguish between two cases depending on whether X_2 is deterministic or not. If it is deterministic, then the rate R_2 must be zero by (11b). Consider now the case when it is not. In this case, $\Pr[X_2 = \eta]$ is positive for all $\eta \in \{0, 1\}$. Since V_2 is independent of X_2 [by (9)], and since without changing the inner bound we can assume that $\Pr[V_2 = v_2]$ is positive for all $v_2 \in \mathcal{V}_2$, it follows that in this case

$$\Pr[X_2 = \eta,\, V_2 = v_2] > 0, \quad \eta \in \{0, 1\},\ v_2 \in \mathcal{V}_2. \tag{85b}$$

This combines with (85a) to imply that

$$H(W_\eta|X_2 = \eta, V_2 = v_2) = 0, \quad \eta \in \{0, 1\},\ v_2 \in \mathcal{V}_2. \tag{85c}$$

This implies that

$$H(W_\eta|V_2 = v_2) = 0, \quad \eta \in \{0, 1\},\ v_2 \in \mathcal{V}_2 \tag{85d}$$

because, by (9), X_2 is independent of (V_2, S_2) and hence *a fortiori* of (V_2, W_η). Thus, $H(W_\eta|V_2) = 0$, and since $S_2 = (W_0, W_1)$,

$$H(S_2|V_2) = 0. \tag{85e}$$

Consequently

$$I(V_2; S_2) = H(S_2) = 1. \tag{85f}$$

This implies that also

$$I(V_2; S_2|Y) = 1 \tag{85g}$$

because Y is independent of (V_2, S_2). It now follows from (85g), the fact that V_1 is deterministic, and from (11b) that R_2 must be zero.

H. *Sketch of the Proof of Theorem 3*

Proof of Theorem 3: The region (24) can be written as

$$0 \le R_1 \le I(U_1; Y|U_2, V_1, V_2) - R'_1 \tag{86}$$
$$0 \le R_2 \le I(U_2; Y|U_1, V_1, V_2) - R'_2 \tag{87}$$
$$R_1 + R_2 \le I(U_1, U_2; Y|V_1, V_2) - R'_1 - R'_2 \tag{88}$$
$$R'_1 \ge I(V_1; S_1|V_2) - I(V_1; Y|V_2) \tag{89}$$
$$R'_2 \ge I(V_2; S_2|V_1) - I(V_2; Y|V_1) \tag{90}$$
$$R'_1 + R'_2 \ge I(V_1, V_2; S_1, S_2) - I(V_1, V_2; Y) \tag{91}$$
$$\Gamma_k \ge \mathsf{E}[\phi_k(X_k)]\,, \qquad k = 1, 2 \tag{92}$$

with joint distribution

$$P_{V_1,V_2,U_1,U_2,S_1,S_2,X_1,X_2,Y} = P_{S_1}\, P_{S_2}\, P_{V_1|S_1}\, P_{V_2|S_2} \cdot P_{U_1}\, P_{U_2}\, P_{X_1|U_1,S_1} P_{X_2|U_2,S_2}\, P_{Y|S_1,S_2,X_1,X_2}. \tag{93}$$

The proof proceeds as the proof of Theorem 1, except that additional external random variables U_1 and U_2 are introduced. The codewords $\mathbf{x}_1(l, k_1)$ and $\mathbf{x}_2(m, k_2)$ are replaced by codewords $\mathbf{u}_1(l, k_1)$ and $\mathbf{u}_2(m, k_2)$, independent of the states. The inputs to the channel are allowed now to depend on the states, according to the laws $P_{X_1|U_1,S_1}$ and $P_{X_2|U_2,S_2}$. The proof proceeds exactly along the lines of the proof of Theorem 1, with $(U_1, U_2, \mathbf{u}_1(l, k_1), \mathbf{u}_2(m, k_2))$ replacing $(X_1, X_2, \mathbf{x}_1(l, k_1), \mathbf{x}_2(m, k_2))$ there. The details are omitted. ■

I. Analysis of Example 4

We first show that $(0, 2)$ is in $\mathcal{R}_{\text{cau}}^{\text{ind}}$. This can be shown by a proper choice of the random variables in (24a)–(24c). Since S_2 is deterministic, we set V_2 to be deterministic too, and we set $U_2 = X_2$. For the external random variables of User 1, note that the entropy of S_1 is lower than the capacity of the channel of User 1. Therefore, set $V_1 = S_1$, $U_1 = X_1$, and let X_1 be independent of S_1. With these substitutions and with the choice $R_1 = 0$, (24a)–(24c) reduce to

$$0 \leq I(X_1; Y_1, Y_2 | X_2, S_1) - H(S_1 | Y_1, Y_2) \tag{94a}$$
$$R_2 \leq I(X_2; Y_1, Y_2 | X_1, S_1) \tag{94b}$$
$$R_2 \leq I(X_1, X_2; Y_1, Y_2 | S_1) - H(S_1 | Y_1, Y_2) \tag{94c}$$

with the joint distribution

$$P_{S_1, X_1, X_2, Y} = P_{S_1} \, P_{X_1} \, P_{X_2} \, P_{Y_1 | X_1} \, P_{Y_2 | S_1, X_2}. \tag{95}$$

With the joint distribution (95), we can simplify (94a) to

$$\begin{aligned} 0 &\leq H(Y_1, Y_2 | X_2, S_1) - H(Y_1, Y_2 | X_1, X_2, S_1) - H(S_1 | Y_2) \\ &= H(Y_1 | Y_2, X_2, S_1) + H(Y_2 | X_2, S_1) \\ &\quad - H(Y_1 | X_1, S_1, X_2, Y_2) - H(Y_2 | X_1, X_2, S_1) \\ &\quad - H(S_1 | Y_2) \\ &= H(Y_1) - H(S_1 | Y_2). \end{aligned} \tag{96a}$$

Similarly, (94b) and (94c) simplify to

$$\begin{aligned} R_2 &\leq H(Y_1, Y_2 | X_1, S_1) - H(Y_1, Y_2 | X_1, X_2, S_1) \\ &= H(Y_1 | X_1, S_1, Y_2) + H(Y_2 | X_1, S_1) \\ &\quad - H(Y_1 | X_1, X_2, S_1, Y_2) - H(Y_2 | X_1, X_2, S_1) \\ &= H(Y_2 | X_1, S_1) \\ &= H(X_2) \end{aligned} \tag{96b}$$

$$\begin{aligned} R_2 &\leq H(Y_1, Y_2 | S_1) - H(Y_1, Y_2 | X_1, X_2, S_1) - H(S_1 | Y_1, Y_2) \\ &= H(Y_1 | Y_2, S_1) + H(Y_2 | S_1) - H(Y_1 | X_1, X_2, S_1, Y_2) \\ &\quad - H(Y_2 | X_1, X_2, S_1) - H(S_1 | Y_1, Y_2) \\ &= H(X_1) + H(X_2) - H(S_1 | Y_2). \end{aligned} \tag{96c}$$

Let X_1 be $\text{Bernoulli}(1/2)$, and let X_2 be uniformly distributed over its alphabet $\{0, 1, 2, 3\}$. The bound (27) on the entropy of S_1 implies that (96a) is satisfied and that (96b) is more stringent than (96c). Since $H(X_2) = 2$ bits, we conclude that the tuple $(0, 2)$ is in $\mathcal{R}_{\text{cau}}^{\text{ind}}$.

We next show that if (R_1, R_2) is achievable using naïve Shannon strategies, then R_2 must be strictly smaller than 2 bits. Since S_2 is null, we substitute X_2 for T_2 in (25c). Based on properties of the capacity region of the classical MAC without SI, the maximal rate $R_{2,\max}^{\text{nai}}$ at which User 2 can communicate utilizing the naïve approach is

$$R_{2,\max}^{\text{nai}} = \max_{t_1, P_{X_2}} I(X_2; Y_1, Y_2 | T_1 = t_1). \tag{97}$$

We claim that $R_{2,\max}^{\text{nai}}$ is strictly less than 2 bits. To see this, let us write

$$\begin{aligned} &I(X_2; Y_1, Y_2 | T_1 = t_1) \\ &\quad = H(Y_1, Y_2 | T_1 = t_1) - H(Y_1, Y_2 | T_1 = t_1, X_2) \\ &\quad = H(Y_1 | T_1 = t_1) + H(Y_2 | Y_1, T_1 = t_1) \\ &\qquad - H(Y_1 | T_1 = t_1, X_2) - H(Y_2 | Y_1, T_1 = t_1, X_2) \\ &\quad = I(X_2; Y_2 | Y_1, T_1 = t_1) && (98) \\ &\quad = H(X_2) - H(X_2 | Y_1, Y_2, T_1 = t_1) && (99) \end{aligned}$$

where (98) holds because X_2 is independent of (Y_1, T_1) so $H(Y_1 | T_1 = t_1) = H(Y_1 | T_1 = t_1, X_2)$, and (99) holds because X_2 is independent of (X_1, T_1).

Since X_2 takes value in a set with four elements, it follows from (97) and (99) that R_2^{nai} cannot be 2 if X_2 is not uniform. It thus remains to show that R_2^{nai} cannot be 2 even if X_2 is uniform. By (99), this is equivalent to showing that when X_2 is uniform, the conditional entropy $H(X_2 | Y_1, Y_2, T_1 = t_1)$ is strictly positive for all functions t_1. This can be shown by noting that $t_1(S_1)$ can take on at most two different values and, therefore, cannot determine S_1.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U.K.: Academic, 1981.

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

[4] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2762–2768, Nov. 2004.

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

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

[7] A. Lapidoth and Y. Steinberg, "The multiple access channel with causal and strictly causal side information at the encoders," in *Proc. Int. Zurich Semin.*, Zurich, Switzerland, Mar. 3–5, 2010, pp. 13–16.

[8] M. Li, O. Simeone, and A. Yener, "Multiple access channels with states causally known at transmitters," *Preprint*, Nov 2010 [Online]. Available: arXiv:1011.6639v1

[9] C. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devel.*, vol. 2, pp. 289–293, 1958.

[10] 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.

[11] 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.

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 IEEE 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.