# The Multiple Access Channel with Two Independent States Each Known Causally to One Encoder

Amos Lapidoth
Signal and Information Processing Laboratory
ETH Zürich
CH-8092 Zürich, Switzerland
amos.lapidoth@isi.ee.ethz.ch

Yossef Steinberg
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel
ysteinbe@ee.technion.ac.il

***Abstract*—We study the state-dependent multiple access channel (MAC) with causal side information at the encoders. The channel state consists of two independent components, $S_1$ and $S_2$, available at Encoder 1 and Encoder 2, respectively. The problem where the state is available at only one of the encoders is a special case . We consider two scenarios. In the first, the states are available at the encoders in a strictly causal manner. We derive an achievable region, which is tight for a Gaussian MAC where the state sequence comprises the channel noise and is available at one of the encoders only. In the second scenario the state sequence is available to the encoders in a causal manner, as in Shannon's model. A simple extension of the previous result to Shannon strategies yields an achievability result. Our region contains as a special case the naïve rate region obtained when each of the users applies Shannon strategies. In some cases the inclusion is strict.**

***Index Terms*—Causal state information, feedback, multiple access channel, strictly-causal state-information.**

## I. Introduction

The problem of coding for state-dependent channels with state information (SI) at the encoder has been studied extensively in two main scenarios: causal SI, and noncausal SI. The case where the state is available in a strictly-causal manner or with a given fixed delay has not attracted much attention, possibly because in single-user channels strictly-causal SI does not increase capacity. However, like feedback, strictly-causal SI can be beneficial in multiple user channels. This can be seen using the examples of Dueck [3]. Specifically, Dueck constructs an additive noise broadcast channel (BC), where the noise is common to the two users. The input and additive noise are defined in a way that the resulting BC is not degraded. 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 data—is spent on the transmission of the noise, the net effect is an increase in channel capacity (perhaps because the noise is common to both users). In Dueck's example the noise is transmitted to the two users losslessly. But it is straightforward to construct examples where only lossy transmission of the noise is possible, and yet this use of feedback increases capacity. There is only one encoder in the BC, so we can identify the additive noise as channel state and think about the availability of feedback in Dueck's example as knowledge of the state in a strictly causal manner.

Of most relevance to this contribution is [7], where the state-dependent multiple access channel (MAC) with *common* SI at the encoders was studied. Two main models were considered: the strictly causal model, where at time $i$ both encoders have access to a common state sequence up to time $i-1$ (or possibly with larger fixed delay), and the causal model, in the spirit of Shannon [8], where at time $i$ both encoders have access to a common state sequence up to (and including) time $i$. In accordance with the insight gained from Dueck's example, it was shown that knowledge of the state in a strictly causal manner increases the MAC's capacity. The main idea in the achievability results of [7] is to use a block Markov coding scheme in which the two users cooperatively transmit a compressed version of the state to the decoder. Although rate is spent on the transmission of the compressed state, the net effect can be an increase in the capacity region, because knowledge of the state (or a compressed version thereof) at the decoder helps both users. The same approach is also beneficial in the causal case, where at time $i$ the state $S_i$ is known to both encoders. A block Markov coding scheme is constructed exactly as in the strictly causal case, with the only exceptions that the inputs at each of the entries are allowed to depend on the state and that additional external random variables that do not depend on the state are introduced. As in the single-user channel with causal side information, the external random variables are viewed as representing Shannon strategies. The naïve approach of using strategies without block Markov coding is a special case of this scheme, since the rate at which the state is described can be chosen as zero. An example is constructed in [7] where the block Markov scheme outperforms the naïve scheme.

In this paper we study the state-dependent MAC where the state $(S_1, S_2)$ consists of two independent components $S_1$ and $S_2$ that are available to Encoder 1 and Encoder 2, respectively. Unlike the scenario studied in [7], here the SI at the two encoders are independent, so the users cannot cooperate in transmitting the state sequences. The special case when $S_2$ is determinstic corresponds to a state-dependent MAC with state available at Encoder 1 only. We refer to this special case as the *asymmetric* state-dependent MAC.

For the strictly causal case we suggest a block Markov coding scheme, where each user transmits a compressed version of its SI. Instead of full cooperation in the state transmissions, the users employ a distributed Wyner-Ziv coding scheme [10], [4], where the channel output serves as the decoder's side information about the two compressed state sequences. Note that although the two state sequences are independent, they do depend on each other given the channel output, so a distributed Wyner-Ziv scheme is superior to a scheme that employs two independent Wyner-Ziv codes. The resulting region is tight for the Gaussian asymmetric state-dependent MAC where the state comprises the channel noise. As in [7], the block Markov scheme with distributed Wyner-Ziv coding is applied also to the causal model. Here we use Shannon strategies to allow dependence between the inputs and channel states. The resulting rate region includes, as a special case, the naïve rate region which uses Shannon strategies for the MAC without block Markov coding. We show via an example that in some cases the inclusion is strict. The combination of Shannon strategies and block-Markov coding was suggested by El Gamal *et. al.* [1] in the context of the relay channel.

The state-dependent MAC with causal side information was studied in the past by Das and Narayan [2] and by Jafar [6]. Das and Narayan developed a non single-letter characterization of the capacity region of the state-dependent MAC with causal SI. The achievability results in [2, Theorem 2] hinge on the techniques developed for the multiple-access channel by Han [5] and do not explicitly include block-Markov coding schemes. However, note that optimization of limits of information functions over input distributions include implicitly also block-Markov coding schemes, since the input distributions can have arbitrary structures. Jafar suggested in [6] an achievable region for the MAC with causal SI, which is in fact identical to the naïve region, but without the convex hull operation. We will comment on Jafar's region after Theorem 2 in Section II-C.

## II. Problem formulation and main results

### 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 $\mathcal{Y}$. Sequences of letters from $\mathcal{S}_k$ are denoted by $s_k^n = (s_{k,1}, s_{k,2}, \ldots, s_{k,n})$ and $s_{k,i}^j = (s_{k,i}, s_{k,i+1} \ldots, s_{k,j})$, for $k = 1, 2$. Similar notation holds for all alphabets, e.g. $x_1^n = (x_{1,1}, x_{1,2}, \ldots, x_{1,n})$, and $x_{2,i}^j = (x_{2,i}, x_{2,i+1}, \ldots, x_{2,j})$. Sometimes we denote $n$-sequences by boldface letters, e.g., $\boldsymbol{x}_1$, $\boldsymbol{s}_1$, $\boldsymbol{y}$, etc. The laws governing $n$-sequences of state and output letters are given by

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

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

For notational convenience, we henceforth omit the superscript $n$, and we denote the channel by $P$. Let $\phi_k \colon \mathcal{X}_k \to [0, \infty)$, $k = 1, 2$, be single-letter cost functions. The cost associated with the transmission of the sequence $\boldsymbol{x}_k$ by Transmitter $k$ is defined as

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

### B. The strictly causal model

*Definition 1:* Given positive integers $\mu_1$ and $\mu_2$, let $\mathcal{M}_1$ be the set $\{1, 2, \ldots, \mu_1\}$ and let $\mathcal{M}_2$ be the set $\{1, 2, \ldots, \mu_2\}$. An $(n, \mu_1, \mu_2, \Gamma_1, \Gamma_2, \epsilon)$ code with strictly causal independent side information at the encoders is a pair of sequences of encoder mappings

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

and a decoding map

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

such that the input cost costs are bounded by $\Gamma_k$

$$\phi_k(\boldsymbol{x}_k) \le \Gamma_k, \quad k = 1, 2,$$

and the average probability of error $P_e$is bounded by $\epsilon$

$$P_e = 1 - \frac{1}{\mu_1 \mu_2} \sum_{m_1=1}^{\mu_1} \sum_{m_2=1}^{\mu_2} \sum_{\boldsymbol{s}_1, \boldsymbol{s}_2} P_{S_1}(\boldsymbol{s}_1) P_{S_2}(\boldsymbol{s}_2) \cdot P\left(g^{-1}(m_1, m_2) | \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{f}_1(\boldsymbol{s}_1, m_1), \boldsymbol{f}_2(\boldsymbol{s}_2, m_2)\right) \le \epsilon,$$

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

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

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

$$R_1 = \frac{1}{n} \log \mu_1, \qquad R_2 = \frac{1}{n} \log \mu_2.$$

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 side information for the channel $P_{Y|S,X_1,X_2}$. The capacity-cost region of the channel with strictly causal independent SI is the closure of the set of all achievable quadruples $(R_1, R_2, \Gamma_1, \Gamma_2)$ and is denoted by $\mathcal{C}_{\mathrm{sc}}^{\mathrm{i}}$. The superscript i stands for independent, to distinguish the current model from the one treated in [7]. For a given pair $(\Gamma_1, \Gamma_2)$ of input costs, $\mathcal{C}_{\mathrm{sc}}^{\mathrm{i}}(\Gamma_1, \Gamma_2)$ stands for the section of $\mathcal{C}_{\mathrm{sc}}^{\mathrm{i}}$ at $(\Gamma_1, \Gamma_2)$. Our interest is in characterizing $\mathcal{C}_{\mathrm{sc}}^{\mathrm{i}}(\Gamma_1, \Gamma_2)$.

We refer to the situation where $S_2$ is a deterministic random variable, i.e., where the state consists of only one component $S_1$, which is available to User 1, as the *asymmetric case*.

The sum capacity of the channel, denoted by $C_{\Sigma,\mathrm{sc}}^{\mathrm{i}}$, is defined as

$$C_{\Sigma,\mathrm{sc}}^{\mathrm{i}}(\Gamma_1, \Gamma_2) = \max_{(R_1, R_2) \in \mathcal{C}_{\mathrm{sc}}^{\mathrm{i}}(\Gamma_1, \Gamma_2)} (R_1 + R_2). \tag{2}$$

A few simple properties of $\mathcal{C}^{\rm i}_{\rm cs}$ can be readily obtained.

*Proposition 1:* Strictly-causal independent SI does not increase the sum-rate capacity:

$$C^{\rm i}_{\Sigma,{\rm sc}}(\Gamma_1,\Gamma_2)=\max I(X_1,X_2;Y), \tag{3}$$

where the maximum is over all product distributions $P_{X_1}P_{X_2}$ satisfying the input constraints

$$\mathbb{E}\phi_k(X_k)\le\Gamma_k,\quad k=1,2. \tag{4}$$

The proof of Proposition 1 is deferred to Section III. A simple property of the capacity region for the asymmetric case is stated next.

*Proposition 2:* Let $S_2$ be deterministic. Then the maximal rate of User 1 with strictly causal SI is equal to its single user capacity without SI

$$\max\left\{R_1:\ (R_1,0)\in\mathcal{C}^{\rm i}_{\rm sc}(\Gamma_1,\Gamma_2)\right\}=\max I(X_1;Y|X_2),$$

where the maximum in the right hand side is over all $P_{X_1}P_{X_2}$ satisfying the input constraints (4).
The proof is deferred to Section III.

Let $\mathcal{P}^{\rm i}_{\rm sc}$ be the collection of all random variables $(V_1,V_2,S_1,S_2,X_1,X_2,Y)$ whose joint distribution satisfies

$$P_{V_1,V_2,S_1,S_2,X_1,X_2,Y}=\\ P_{V_1|S_1}P_{V_2|S_2}P_{S_1}P_{S_2}P_{X_1}P_{X_2}P_{Y|S_1,S_2,X_1,X_2}. \tag{5}$$

Note that (5) implies the Markov relations

$$\begin{aligned}&V_1 \multimap S_1 \multimap (V_2,Y,S_2)\\&V_2 \multimap S_2 \multimap (V_1,Y,S_1)\\&(V_1,V_2)\multimap(S_1,S_2)\multimap Y\end{aligned} \tag{6}$$

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

$$0\le R_1\le I(X_1;Y|X_2,V_1,V_2)-I(V_1;S_1|Y,V_2) \tag{7}$$
$$0\le R_2\le I(X_2;Y|X_1,V_1,V_2)-I(V_2;S_2|Y,V_1) \tag{8}$$
$$R_1+R_2\le I(X_1,X_2;Y|V_1,V_2)-I(V_1,V_2;S_1,S_2|Y) \tag{9}$$
$$\Gamma_k\ge\mathbb{E}\phi_k(X_k),\qquad k=1,2$$

for some $(V_1,V_2,S_1,S_2,X_1,X_2,Y)\in\mathcal{P}^{\rm i}_{\rm sc}$. Our main result for the strictly causal case is the following.

*Theorem 1:* $\mathcal{R}^{\rm i}_{\rm sc}\subseteq\mathcal{C}^{\rm i}_{\rm sc}$.
The proof is based on a scheme where lossy versions of the state sequences are conveyed to the decoder using distributed Wyner-Ziv compression [4] followed by block-Markov encoding to transmit the messages and the Wyner-Ziv codewords. The channel output serves as the decoder's side information 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 [7]. Instead, each user spends part of its private rate on the transmission of its Wyner-Ziv codeword. The details of the proof are omitted. In some cases, the region $\mathcal{R}^{\rm i}_{cs}$ coincides with $\mathcal{C}^{\rm i}_{cs}$. 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 [9].

*Example 1:* Consider the asymmetric Gaussian MAC with input power constraints $\mathbb{E}X_k^2\le\Gamma_k$, $k=1,2$, where the state $S_1$ comprises the channel noise:

$$Y=X_1+X_2+S_1,\qquad S_1\sim N(0,\sigma^2_{s_1}). \tag{10}$$

The capacity region of this channel when $S_1$ is known strictly causally to Encoder 1 is the collection of all pairs $(R_1,R_2)$ satisfying

$$R_1\le\frac{1}{2}\log\left(1+\frac{\Gamma_1}{\sigma^2_{s_1}}\right) \tag{11}$$
$$R_1+R_2\le\frac{1}{2}\log\left(1+\frac{\Gamma_1+\Gamma_2}{\sigma^2_{s_1}}\right). \tag{12}$$

Note that this collection is the convex hull of the union of the capacity region when $S_1$ is unknown and the set comprising the rate pair

$$R_1=0,\qquad R_2=\frac{1}{2}\log\left(1+\frac{\Gamma_1+\Gamma_2}{\sigma^2_{s_1}}\right). \tag{13}$$

A time-sharing argument thus demonstrates that, to prove the achievability of (11)–(12) in the asymmetric case, it suffices to show that the rate pair in (13) is achievable. To this end, make the following substitutions in the definition of the set $\mathcal{R}^{\rm i}_{\rm sc}$: $V_2=0$; $V_1$ is zero mean and jointly Gaussian with $S_1$; and $X_1$, $X_2$ are zero mean independent Gaussians independent of $(V_1,S_1)$. Then (7)–(9) reduce to

$$0\le\frac{1}{2}\log\frac{(\Gamma_1+\sigma^2_{s_1|v_1})(\Gamma_1+\Gamma_2+\sigma^2_{s_1})}{\sigma^2_{s_1}(\Gamma_1+\Gamma_2+\sigma^2_{s_1|v_1})} \tag{14}$$
$$R_2\le\frac{1}{2}\log\left(1+\frac{\Gamma_2}{\sigma^2_{s_1|v_1}}\right) \tag{15}$$
$$R_2\le\frac{1}{2}\log\left(1+\frac{\Gamma_1+\Gamma_2}{\sigma^2_{s_1}}\right) \tag{16}$$

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

$$\frac{\Gamma_2\,\sigma^2_{s_1}}{\Gamma_1+\Gamma_2}-\Gamma_1\le\sigma^2_{s_1|v_1}\le\frac{\Gamma_2\,\sigma^2_{s_1}}{\Gamma_1+\Gamma_2} \tag{17}$$

the bound (16) dominates (15) and the right hand side of (14) is positive. We thus conclude that (13) is achievable.

We next have to show that the achievable region (11)–(12) is tight. This follows from the capacity region of the Gaussian MAC without SI, and Propositions 1 and 2.

### C. The causal model

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

$$f_{k,i}\colon\mathcal{S}^i_k\times\mathcal{M}_k\to\mathcal{X}_k,\quad k=1,2,\quad i=1,\ldots,n. \tag{18}$$

The capacity region and its section at $(\Gamma_1, \Gamma_2)$ are denoted by $\mathcal{C}_c^i$ and $\mathcal{C}_c^i(\Gamma_1, \Gamma_2)$, respectively. Let $\mathcal{P}_c^i$ be the collection of all random variables $(V_1, V_2, U_1, U_2, S_1, S_2, X_1, X_2, Y)$ whose joint distribution can be written as

$$\begin{aligned} P_{V_1,V_2,U_1,U_2,S_1,S_2,X_1,X_2,Y} &= P_{V_1|S_1} P_{V_2|S_2} P_{U_1} P_{U_2} P_{S_1} P_{S_2} \cdot \\ &P_{X_1|U_1,S_1} P_{X_2|U_2,S_2} P_{Y|S_1,S_2,X_1,X_2}. \end{aligned} \tag{19}$$

Observe that (19) implies the Markov relations

$$\begin{aligned} &V_1 \multimap S_1 \multimap (V_2, S_2, X_1, X_2, U_1, U_2, Y) \\ &V_2 \multimap S_2 \multimap (V_1, S_1, X_1, X_2, U_1, U_2, Y) \\ &(V_1, V_2) \multimap (S_1, S_2) \multimap (X_1, X_2, U_1, U_2, Y). \end{aligned} \tag{20}$$

Let $\mathcal{R}_c^i$ be the convex hull of the collection of all $(R_1, R_2, \Gamma_1, \Gamma_2)$ satisfying

$$0 \le R_1 \le I(U_1; Y|U_2, V_1, V_2) - I(V_1; S_1|Y, V_2) \tag{21}$$
$$0 \le R_2 \le I(U_2; Y|U_1, V_1, V_2) - I(V_2; S_2|Y, V_1) \tag{22}$$
$$R_1 + R_2 \le I(U_1, U_2; Y|V_1, V_2) - I(V_1, V_2; S_1, S_2|Y) \tag{23}$$
$$\Gamma_k \ge \mathbb{E}\phi_k(X_k), \qquad k = 1, 2$$

for some $(V_1, V_2, U_1, U_2, S_1, S_2, X_1, X_2, Y) \in \mathcal{P}_c^i$. Our result for the causal case is stated next.

*Theorem 2:* $\mathcal{R}_c^i \subseteq \mathcal{C}_c^i$.

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 side information, where a random Shannon strategy can be represented by an external random variable independent of the state. The proposed scheme outperforms the naïve approach of using strategies without block Markov encoding of the state. This latter naïve approach leads to the region comprising all rate pairs $(R_1, R_2)$ satisfying

$$\begin{aligned} R_1 &\le I(T_1; Y|T_2, Q) \\ R_2 &\le I(T_2; Y|T_1, Q) \\ R_1 + R_2 &\le I(T_1, T_2; Y|Q) \end{aligned} \tag{24}$$

for some $P_Q P_{T_1|Q} P_{T_2|Q}$, where $T_k$ are random Shannon strategies [8] whose realizations are mappings $t_k \colon \mathcal{S}_k \to \mathcal{X}_k$, $k = 1, 2$; $Q$ is a time sharing random variable, and

$$\begin{aligned} P_{Y|T_1,T_2}(y|t_1, t_2) = \sum_{s_1 \in \mathcal{S}_1} \sum_{s_2 \in \mathcal{S}_2} & P_{S_1}(s_1) P_{S_2}(s_2) \cdot \\ & P_{Y|S_1,S_2,X_1,X_2}(y|s_1, s_2, t_1(s_1), t_2(s_2)). \end{aligned}$$

The naïve region (24) contains the region suggested by S. A. Jafar in [6, Section VI], as the latter does not include the time sharing random variable, or convex hull operation. Note that the time sharing random variable cannot be included in the coding random variables $(U_1, U_2)$ of [6, Section VI], since they should be kept independent of each other. Clearly $\mathcal{R}_c$ contains the region of the naïve approach as we can choose $V_1, V_2$ in (21)–(23) to be deterministic. The next example demonstrates that the inclusion can be strict.

*Example 2:* 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

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_2 \oplus S_1, \end{aligned}$$

where $\oplus$ stands for modulo 4 addition, $S_1$ is additive noise given by

$$\mathcal{S}_1 = \{0, 1, 2, 3\}, \quad P_S = (1 - p, p/3, p/3, p/3),$$

and $p \in (0, 1)$ is small enough to guarantee that

$$H(S_1) < 1. \tag{25}$$

We now characterize the maximal rate of User 2.

**Block Markov Coding.** For this channel the rate pair $(0, 2)$ is in $\mathcal{R}_c^i$, and every rate pair $(R_1, R_2)$ in $\mathcal{C}_c^i$ must satisfy $R_2 \le 2$. Thus,

$$R_{2,\max}^{(\text{bm})} = 2 \quad \text{[bits]}. \tag{26}$$

That no achievable rate $R_2$ can exceed 2 bits is obvious, because this is even true if the state is known to all parties. That $(0, 2)$ is in $\mathcal{R}_c^i$ can be shown by a proper choice of the random variables in (21)–(23) as follows. Since $S_2$ is null, let $V_2$ also be a null random variable, and 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 choose $V_1 = S_1$, $U_1 = X_1$, and let $X_1$ be independent of $S_1$. With these substitutions, (21)–(23) reduce to

$$0 \le I(X_1; Y_1, Y_2|X_2, S_1) - H(S_1|Y_1, Y_2) \tag{27}$$
$$R_2 \le I(X_2; Y_1, Y_2|X_1, S_1) \tag{28}$$
$$R_2 \le I(X_1, X_2; Y_1, Y_2|S_1) - H(S_1|Y_1, Y_2) \tag{29}$$

with the joint distribution

$$P_{S_1,X_1,X_2,Y} = P_{S_1} P_{X_1} P_{X_2} P_{Y_1|X_1} P_{Y_2|S_1,X_2}. \tag{30}$$

With the joint distribution (30), we obtain from (27)

$$\begin{aligned} 0 &\le H(Y_1, Y_2|X_2, S_1) - H(Y_1, Y_2|X_1, X_2, S_1) - H(S_1|Y_2) \\ &= H(Y_1|Y_2, X_2, S_1) + H(Y_2|X_2, S_1) \\ &\quad - H(Y_1|X_1, S_1, X_2, Y_2) - H(Y_2|X_1, X_2, S_1) \\ &\quad - H(S_1|Y_2) \\ &= H(Y_1) - H(S_1|Y_2). \end{aligned} \tag{31}$$

Similarly, from (28), (29) we obtain the bounds

$$\begin{aligned} R_2 &\le H(Y_1, Y_2|X_1, S_1) - H(Y_1, Y_2|X_1, X_2, S_1) \\ &= H(Y_1|X_1) + H(Y_2|X_1, S_1) - H(Y_1|X_1) \\ &\quad - H(Y_2|X_1, X_2, S_1) \\ &= H(Y_2|X_1, S_1) = H(X_2) \end{aligned} \tag{32}$$

$$
\begin{aligned}
R_2 &\le H(Y_1,Y_2|S_1) - H(Y_1,Y_2|X_1,X_2,S_1) - H(S_1|Y_1,Y_2) \\
&= H(Y_1) + H(Y_2|S_1) - H(Y_1|X_1) \\
&\quad - H(Y_2|X_2,S_1) - H(S_1|Y_2) \\
&= H(X_2) + H(Y_1) - H(S_1|Y_2). \qquad (33)
\end{aligned}
$$

Let $X_1$ be Bernoulli(1/2), and let $X_2$ be uniformly distributed over its alphabet $\{0,1,2,3\}$. Due to the bound (25) on the entropy of $S_1$, (31) is satisfied. Consequently (32) dominates (33), and since $H(X_2) = 2$ bits, we conclude that (26) is achievable.

**The Naïve Approach.** Since the state $S_2$ is null, we substitute $X_2$ instead of $T_2$ in (24). Based on properties of the capacity region of the classical MAC without side information, the maximal rate $R_{2,\max}^{(\text{naïve})}$ at which User 2 can communicate utilizing the naïve approach is given by

$$
R_{2,\max}^{(\text{naïve})} = \max_{t_1, P_{X_2}} I(X_2;Y_1,Y_2|T_1=t_1). \qquad (34)
$$

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

$$
\begin{aligned}
&I(X_2;Y_1,Y_2|T_1=t_1) \\
&\quad = H(Y_1,Y_2|T_1=t_1) - H(Y_1,Y_2|T_1=t_1,X_2) \\
&\quad = H(Y_1|T_1=t_1) + H(Y_2|Y_1,T_1=t_1) \\
&\qquad - H(Y_1|T_1=t_1,X_2) - H(Y_2|Y_1,T_1=t_1,X_2) \\
&\quad = I(X_2;Y_2|Y_1,T_1=t_1) \qquad (35) \\
&\quad = H(X_2) - H(X_2|Y_1,Y_2,T_1=t_1), \qquad (36)
\end{aligned}
$$

where (35) 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 (36) 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 (34) and (36) that $R_2^{(\text{naïve})}$ cannot be 2 if $X_2$ is not uniform. It thus remains to show that $R_2^{(\text{naïve})}$ cannot be 2 even if $X_2$ is uniform. By (36), 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 $S_1$.

## III. Proofs

*Proof of Proposition 1:* First note that the right hand side of (3) is the sum-rate capacity of the same MAC without SI and hence is achievable in the presence of strictly causal SI. It remains to show that this is also an upper bound. Consider the result of applying any coding scheme to messages $M_1$ and $M_2$ that are chosen uniformly and independently at random. Starting with the Fano inequality, we have

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

where the last equality holds because in the strictly causal case the Markov relation $(M_1,M_2,Y^{i-1}) \multimap (X_{1,i},X_{2,i}) \multimap Y_i$ is satisfied. Since $(M_1,S_1^{i-1})$ and $(M_2,X_2^{i-1})$ are independent of each other, so are $X_{1,i}$ and $X_{2,i}$. The claim now follows by the standard time sharing argument. □

*Proof of Proposition 2:* We need to show only the converse part. Denote by $M_1$ the random message of User 1, and note that $M_1$ and $X_2^n$ are independent of each other. By Fano's inequality

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

where the last equality is due to the Markov relation $(M_1,Y^{i-1},X_2^{i-1},X_{2,i+1}^n) \multimap (X_{1,i},X_{2,i}) \multimap Y_i$. The claim now follows by the standard time sharing arguments. □

## References


[1] A. El Gamal, N. Hassanpour, J. Mammen, "Relay networks with delays," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3413-3431, October 2007.

[2] A. Das and P. Narayan, "Capacities of time-varying multiple-access channels with side information," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 4-25, January 2002.

[3] G. Dueck, "Partial feedback for two-way and broadcast channels," *Inf. Contr.*, vol. 46, pp. 1–15, 1980.

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

[5] T. S. Han, "An information-spectrum approach to capacity theorems for the general multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2773-2795, Nov. 1998.

[6] S. A. Jafar, "Capacity with causal and non-causal side information - A unified view," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468-5475, December 2006.

[7] A. Lapidoth and Y. Steinberg, "The multiple access channel with causal and strictly causal side information at the encoders," to appear in *Proc. of the 21st International Zurich Seminar on Communications*, March 3–5 2010, Zurich, Switzerland.

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

[9] A. D. Wyner, "The rate-distortion function for source coding with side infomation at the decoder—II: general sources," *Inf. Contr.*, vol. 38, pp. 60–80, 1978.

[10] 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.