The multiple access channel with causal and strictly causal side information at the encoders

Author(s):
Lapidoth, Amos; Steinberg, Yossef

The Multiple Access Channel with Causal and Strictly Causal Side Information at the Encoders

Amos Lapidoth

Yossef Steinberg

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

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. We consider two general models. In the first model, the state sequence is available at the two encoders in a strictly causal manner. We derive an achievable region, which is tight for the special case of a Gaussian MAC where the state sequence comprises the channel noise. Although the two senders do not have access to each other's massage and no feedback is present, the capacity for this case coincides with the capacity of the same MAC without side information, but with full cooperation between the users. A Schalkwijk-Kailath type algorithm is developed, which achieves this capacity with a double exponential decay of the maximal probability of error. In the second model we consider, the state sequence is available, as in Shannon's model, to the two encoders in a causal manner. A simple extension of the previous result, with the inclusion of Shannon strategies, yields an achievability result for this problem.

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 at the encoder has been studied extensively in two main scenarios: causal state information and noncausal state information. 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 state-information (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 [1]. Specifically, Deuck 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 user messages–is spent on the transmission of the noise, the net effect is an increase in channel capacity, due to the noise being common to both users. In Dueck's example, the noise is transmitted losslessly to the two users. However, based on his observations, it is straightforward to construct examples where only lossy transmission of the noise is possible, and yet the capacity region is increased by this use of feedback. There is only one

encoder in the BC and thus, identifying the additive noise as channel state, feedback in Dueck's example is equivalent to knowledge of the state in a strictly causal manner.

In this paper we study the state-dependent multiple access channel (MAC) with common state information at the encoders. Two main models are 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 delay), and the causal model, in the spirit of Shannon [4], where at time i both encoders have access to a common state sequence up to (and including) time i.

As in the case of broadcast channels, strictly causal knowledge of the state increases the MAC's capacity. Since only past (or delayed) samples of the state are known, neither binning (as in Gel'fand and Pinsker's channel [2]) nor strategies [4] can be employed. Instead, we derive a general achievable region based on a block-Markov coding scheme. The encoders, having access to a common state sequence, compress and transmit it to the decoder. The users cannot establish cooperation in the transmission of the messages, but they do cooperate in the transmission of the compressed state, thus increasing the achievable rates. The resulting region is tight for the Gaussian MAC where the state comprises the channel noise. Specifically, it is shown that for this channel, a proper choice of the random variables in our achievable region yields the capacity region of the same MAC without side information but with full cooperation between the encoders. Since strictly causal state information does not increase the capacity of single user channels, it also cannot increase the capacity of the MAC with full cooperation. Consequently, full cooperation is the best that one can hope for, and thus the region must be tight. Although the users do not have access to each other's message and no feedback is available, a Schalkwijk-Kailath type algorithm can be devised for this channel, yielding the full cooperation region with a double exponential decay in the probability of error. The general achievability result, and the Schalkwijk-Kailath algorithm, make use of all the past samples of the channel noise. It turns out, however, that much less is needed to achieve the full cooperation region. Assume that, instead of having all the past noise samples, only S_1 and S_2 are known to the encoders, in a strictly causal

manner (i.e., available at times 3,4,....). As demonstrated in Section II, although only two noise samples are known, the full cooperation region can still be achieved.

The causal model is also treated with block Markov coding, but the transmission at time i can depend on the state S_i. All other ingredients of the coding scheme remain intact. The resulting achievable region contains the naive region, which uses Shannon strategies for the MAC without block Markov coding, with the inclusion being in some cases strict.

II. PROBLEM FORMULATION AND MAIN RESULTS

A. Basic definitions

We are given a discrete memoryless state-dependent MAC $P_{Y|S,X_1,X_2}$ with state alphabet $\mathcal{S}$, state probability mass function (PMF) P_S, input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, and output alphabet $\mathcal{Y}$. Sequences of letters from $\mathcal{S}$ are denoted by $s^n = (s_1, s_2, \ldots, s_n)$ and $s_i^j = (s_i, s_{i+1} \ldots, s_j)$. Similar notation holds for all alphabets, e.g. $x_1^n = (x_{1,1}, x_{1,2}, \ldots, x_{1,n})$, $x_{2,i}^j = (x_{2,i}, x_{2,i+1}, \ldots, x_{2,j})$. When there is no risk of ambiguity, n-sequences will sometimes be denoted by boldface letters, $\boldsymbol{y}$, $\boldsymbol{x}_1$, etc. The laws governing n sequences of state and output letters are given by

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

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

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

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

B. The strictly causal model

Definition 1: Given positive integers M_1, M_2, let $\mathcal{M}_1$ be the set $\{1, 2, \ldots, M_1\}$ and similarly, $\mathcal{M}_2$ the set $\{1, 2, \ldots, M_2\}$. An $(n, M_1, M_2, \Gamma_1, \Gamma_2, \epsilon)$ code with strictly causal side information at the encoders is a pair of sequences of encoder mappings

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

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 Γ_k

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

and the average probability of error is bounded by ϵ

$$\begin{aligned} P_e &= 1 - \frac{1}{M_1 M_2} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \sum_{\boldsymbol{s}} P_S(\boldsymbol{s}) \cdot \\ &\quad P\left(g^{-1}(m_1, m_2)|\boldsymbol{s}, \boldsymbol{f}_1(\boldsymbol{s}, m_1), \boldsymbol{f}_2(\boldsymbol{s}, m_2)\right) \leq \epsilon, \end{aligned}$$

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

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

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

$$R_1 = \frac{1}{n} \log M_1, \quad R_2 = \frac{1}{n} \log M_2.$$

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 SI is the closure of the set of all achievable quadruples $(R_1, R_2, \Gamma_1, \Gamma_2)$, and is denoted by $\mathcal{C}_{sc}$. For a given pair (Γ_1, Γ_2) of input costs, $\mathcal{C}_{sc}(\Gamma_1, \Gamma_2)$ stands for the section of $\mathcal{C}_{sc}$ at (Γ_1, Γ_2). Our interest is in characterizing $\mathcal{C}_{sc}(\Gamma_1, \Gamma_2)$.

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

$$P_{U,V,X_1,X_2,S,Y} = P_S P_{X_1|U} P_{X_2|U} P_U P_{V|S} P_{Y|S,X_1,X_2}. \quad (2)$$

Note that (2) implies the Markov relations $X_1 \leftrightarrow U \leftrightarrow X_2$ and $V \leftrightarrow S \leftrightarrow Y$, and that the triplet (X_1, U, X_2) is independent of (V, S). Let $\mathcal{R}_{sc}$ be the convex hull of the collection of all $(R_1, R_2, \Gamma_1, \Gamma_2)$ satisfying

$$R_1 \leq I(X_1; Y|X_2, U, V) \quad (3)$$
$$R_2 \leq I(X_2; Y|X_1, U, V) \quad (4)$$
$$R_1 + R_2 \leq I(X_1, X_2; Y|U, V) \quad (5)$$
$$R_1 + R_2 \leq I(X_1, X_2, V; Y) - I(V; S) \quad (6)$$
$$\Gamma_k \geq \mathbb{E}\phi_k(X_k), \quad k = 1, 2$$

for some $(U, V, X_1, X_2, S, Y) \in \mathcal{P}_{sc}$. Our main result for the strictly causal case is the following.

Theorem 1: $\mathcal{R}_{sc} \subseteq \mathcal{C}_{sc}$.

The proof is based on a scheme where a lossy version of the state is conveyed to the decoder using Wyner-Ziv compression [7] and block-Markov encoding for the MAC with common message [5]. The proof is omitted. In some cases, the region $\mathcal{R}_{cs}$ coincides with $\mathcal{C}_{cs}$. 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 [6].

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

$$Y = X_1 + X_2 + S, \quad S \sim N(0, \sigma_s^2). \quad (7)$$

The capacity region of this channel, when S is known strictly causally at the two encoders, is the collection of all pairs (R_1, R_2) satisfying

$$R_1 + R_2 \leq \frac{1}{2} \log\left(1 + \frac{(\Gamma_1^{\frac{1}{2}} + \Gamma_2^{\frac{1}{2}})^2}{\sigma_s^2}\right). \quad (8)$$

The region (8) is the capacity region of the same MAC when S is not known to any of the encoders, but with full cooperation

between the users—a situation equivalent to a single user channel with a vector input constraint. Since strictly causal SI does not increase the capacity of a single user channel, we only have to show achievability in (8). We show it by properly choosing the random variables in (3)–(6). Let us first examine the maximal R_1. Set $U = X_2$, and let X_1, X_2, V, S be zero mean, jointly Gaussian, with X_1, X_2 independent of V, S. Then (3)-(6) reduce to the two bounds on R_1:

$$R_1 \leq \frac{1}{2} \log \left(\frac{\sigma^2_{x_1|x_2} + \sigma^2_{s|v}}{\sigma^2_{s|v}} \right) \tag{9}$$

$$R_1 \leq \frac{1}{2} \log \left(\frac{\Gamma_\Sigma + \sigma^2_s}{\sigma^2_s} \right) \tag{10}$$

where $\sigma^2_{x_1|x_2}$ is the variance of X_1 conditioned on X_2; $\sigma^2_{s|v}$ is analogously defined; and Γ_Σ is the power of the sum $X_1 + X_2$. In full cooperation, $\Gamma_\Sigma = (\Gamma_1^{\frac{1}{2}} + \Gamma_2^{\frac{1}{2}})^2$, but then $\sigma^2_{x_1|x_2} = 0$, which nullifies the right hand side of (9). Note, however, that we can take the limit $\sigma^2_{s|v} \to 0$ without effecting (10). Thus we can approach the full cooperation rate as closely as desired, by first reducing $\sigma^2_{s|v}$, so that the right hand side of (9) is kept high, and then reducing $\sigma^2_{x_1|x_2}$. This proves that with $R_2 = 0$, the rate

$$R_1 = \frac{1}{2} \log \left(1 + \frac{(\Gamma_1^{\frac{1}{2}} + \Gamma_2^{\frac{1}{2}})^2}{\sigma^2_s} \right) \tag{11}$$

is achievable. By symmetry and time sharing, (8) is achievable.

We next describe a Schalkwijk-Kailath type algorithm [3], that achieves the same rate, with double exponential decay in the maximal probability of error. As with the proof of (8), first the achievability of (11) is shown. The rest will follow by symmetry and time sharing. Split the interval $[0, 1]$ into M_1 equally spaced sub-intervals. Let θ_1 be the center of one of these sub-intervals, representing the message of user 1, as in [3]. At the first time instance, the users transmit

$$X_{1,1} = \theta_1, \quad X_{2,1} = 0. \tag{12}$$

The corresponding channel output is

$$Y_1 = \theta_1 + S_1. \tag{13}$$

Starting from time instance $i = 2$ and on, the noise sample S_1 is known at both encoders. Thus the two encoders now cooperate to transmit S_1 to the decoder. Since they now have a common message to transmit, knowing the states in a strictly causal manner is equivalent to feedback. Applying the same algorithm as in [3], after n iterations the receiver constructs a Maximum Likelihood estimate of S_1, denoted $\hat{S}_1^{(n)}$, whose error satisfies

$$\mathbb{E}((S_1 - \hat{S}_1^{(n)})^2|S_1) = \frac{\sigma^2_s}{(\alpha^2)^n} \tag{14}$$

with

$$\alpha^2 = 1 + \gamma^2, \quad \gamma^2 = \frac{(\Gamma_1^{\frac{1}{2}} + \Gamma_2^{\frac{1}{2}})^2}{\sigma^2_s}. \tag{15}$$

Based on this estimate and on Y_1, the decoder can now construct an estimate $\hat{\theta}_1^{(n)}$ of θ_1

$$\hat{\theta}_1 = Y_1 - \hat{S}_1^{(n)} \tag{16}$$

whose error satisfies

$$\mathbb{E}((\theta_1 - \hat{\theta}_1^{(n)})^2|S_1) = \frac{\sigma^2_s}{(\alpha^2)^n}. \tag{17}$$

Therefore, choosing $M_1 = nr_1$, with $r_1 \leq \frac{1}{2} \log(1 + \gamma^2)$, the probability of error vanishes doubly exponentially as $n \to \infty$. This proves that (11) is achievable. By symmetry, and applying time sharing, this algorithm achieves the region (8).

We next show that the region (8) is achievable also when the states at only two time instances, say S_1 and S_2, are known from time $i = 3$ and on. It suffices to show achievability of (11) with only S_1 known, from time $i = 2$. First transmissions and output are given by (12), (13). At times $i = 2$ and on, both users know S_1. They cooperate in transmitting to the receiver a quantized version of S_1 via a regular code for the single user Gaussian channel. Specifically, fix $\epsilon > 0$ and choose β such that $P_S(|S_1| > \beta) \leq \epsilon$. Define $r_1 = \frac{1}{2} \log(1 + \gamma^2)$. We employ two partitions. First, partition the interval $[0, 1]$ into $M_1 = 2^{nr_1}/(4\beta)$ sub intervals, where the centers represent the messages of user 1. Let θ_1 be the center of one of these sub intervals. Partition the interval $[-\beta, \beta]$ into 2^{nr_1} sub intervals, and denote by $m_q, q = 1, 2, \ldots, 2^{nr_1}$ their center points. Define

$$S_q = \arg \min_{m_q} |S_1 - m_q| \tag{18}$$

The two users transmit S_q to the receiver, via a single user code. Denote by $\hat{S}_q$ the receiver's estimate of S_q. Clearly, for n large enough,

$$P \left(\left| S_1 - \hat{S}_q \right| \geq \frac{2\beta}{(1 + \gamma^2)^{n/2}} \right) \leq 2\epsilon \tag{19}$$

implying that the receiver can detect θ_1 with probability of error not exceeding 2ϵ. Note that M_1 provides the claimed rate.

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

$$f_{k,i} \colon \mathcal{S}^i \times \mathcal{M}_k \to \mathcal{X}_k, \quad k = 1, 2, \quad i = 1, \ldots, n. \tag{20}$$

The capacity region and its section at (Γ_1, Γ_2) are denoted by $\mathcal{C}_c$ and $\mathcal{C}_c(\Gamma_1, \Gamma_2)$, respectively. Let $\mathcal{P}_c$ be the collection of all random variables $(U, U_1, U_2, V, X_1, X_2, S, Y)$ whose joint distribution can be written as

$$P_U P_{U_1|U} P_{U_2|U} P_{V|S} P_S P_{X_1|U,U_1,S} P_{X_2|U,U_2,S} P_{Y|S,X_1,X_2}. \tag{21}$$

Observe that (21) implies the Markov relations $U_1 \oplus U \oplus U_2$ and $V \oplus S \oplus Y$, and that the triple (U_1, U, U_2) is independent

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

$$
\begin{align}
R_1 &\leq I(U_1; Y | U_2, U, V) \tag{22}\\
R_2 &\leq I(U_2; Y | U_1, U, V) \tag{23}\\
R_1 + R_2 &\leq I(U_1, U_2; Y | U, V) \tag{24}\\
R_1 + R_2 &\leq I(U_1, U_2, V; Y) - I(V; S) \tag{25}\\
\Gamma_k &\geq \mathbb{E}\phi_k(X_k), \qquad k = 1, 2 \tag{}
\end{align}
$$

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

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

The proof proceeds along the lines of the proof of Theorem 1, except that the inputs X_k, $k = 1, 2$, are allowed to depend on the state S, and that additional external random variables U_1 and U_2 that do not depend on S are introduced. This resembles the situation in coding for the single user channel with causal side information, where a random Shannon strategy can be represented by an external random variable independent of the state. The proposed scheme outperforms the naive approach of using strategies without block Markov encoding of the state. This latter naive approach leads to the region comprising all (R_1, R_2) satisfying

$$
\begin{align}
R_1 &\leq I(T_1; Y | T_2, Q) \tag{}\\
R_2 &\leq I(T_2; Y | T_1, Q) \tag{}\\
R_1 + R_2 &\leq I(T_1, T_2; Y | Q) \tag{26}
\end{align}
$$

for some $P_Q P_{T_1|Q} P_{T_2|Q}$, where T_k are random Shannon strategies [4], whose realizations are mappings $t_k \colon \mathcal{S} \to \mathcal{X}_k$, $k = 1, 2$; Q is a time sharing random variable; and

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

Clearly $\mathcal{R}_c$ contains the region of the naive approach as we can choose V in (22)–(25) to be a null random variable. The next example demonstrates that the inclusion can be strict.

Example 2: Noiseless binary MAC, with input selector. Consider the noiseless binary MAC where $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \{0, 1\}$, $\mathcal{S} = \{1, 2\}$ and $P_S(S = 2) = p$ for some $p > 0.5$. The state S determines which of the two inputs is connected to the output:

$$
Y = X_S.
$$

Block Markov Coding. Both users know the state and hence know, at each time, which user is connected to the output. Thus, they can compress the state using $H(S) = H_b(p)$ bits per channel use and transmit the state sequence to the decoder, via block Markov coding. If they do so, the decoder knows S, and the users can now share between them a clean channel. Since they already spent $H_b(p)$ bits in transmitting the state, the net rate remaining to share between them is

$$
R_1 + R_2 = 1 - H_b(p). \tag{27}
$$

Note, however, that not all the line (27) is achievable. The users do not know each other's message. Thus, user 1 can

transmit its own message only $(1 - p)$ fraction of the time. We conclude that the following rate is achievable for user 1:

$$
R_1 = [1 - H_b(p)](1 - p) \quad \text{[bits]}. \tag{28}
$$

The Naive Approach. From the region (26) and the extreme points of the capacity region of the classical MAC, the maximal rate that user 1 can transmit is:

$$
R_1 = \max I(T_1; Y | T_2 = t_2), \tag{29}
$$

where the maximum is over the distribution of T_1 and over all mappings $t_2 \colon \mathcal{S} \to \mathcal{X}_2$. The strategy t_2 influences the output only when $S = 2$, in which case it gives a certain input X_2, connected directly to the output. User 1 is then disconnected. Therefore, the exact value of t_2 is immaterial. Assume that $t_2(s = 2) = 0$.

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

$$
P(Y = 0 | X = 0) = 1, \quad P(Y = 0 | X = 1) = p. \tag{30}
$$

The capacity of this channel is given by

$$
C(p) = \log_2\left(1 + (1 - p)p^{\frac{p}{1-p}}\right) \quad \text{[bits]}. \tag{31}
$$

At the limit where p approaches 1, we have

$$
C(p) \approx (1 - p)e^{-1}\log_2 e \approx 0.53(1 - p), \tag{32}
$$

which, at the limit $p \to 1$, is strictly less than (28).

REFERENCES

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

[2] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Inform. & Control*, vol. 9, no. 1, pp. 19–31, 1980.

[3] J. P. M. Schalkwijk, "A coding scheme for additive noise channels with feedback—Part II: band-limited signals," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 183–189, April 1966

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

[5] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell System technical Journal*, vol. 52, pp. 1037–1076, Sept. 1973.

[6] 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.

[7] 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.