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

Annina Bracher
ETH Zurich
Zurich, Switzerland
Email: brachera@student.ethz.ch

Amos Lapidoth
ETH Zurich
Zurich, Switzerland
Email: lapidoth@isi.ee.ethz.ch

Yossef Steinberg
Technion—Israel Institute of Technology
Haifa, Israel
Email: ysteinbe@ee.technion.ac.il

***Abstract*—We show that the capacity region of the state-dependent multiple-access channel (SD-MAC) with strictly-causally cribbing encoders is not enlarged if strictly-causal state-information (SI) and feedback are furnished to the encoders. We also derive the capacity region of the SD-MAC with causal SI at the cribbing encoders and show that Shannon strategies are optimal. Such strategies are generally suboptimal if the encoders access distinct SI. However, Shannon strategies are optimal and we have a characterization of the capacity region for the case where both encoders crib, causal SI is revealed to one encoder, and feedback is available to the other encoder.**

I. Introduction

Unlike the memoryless single-user channel, the capacity of the memoryless multiple-access channel (MAC) typically increases when feedback from the channel output to the encoders is introduced [1]. The intuitive explanation for this phenomenon is that, via the feedback link, each encoder learns something about the symbol transmitted by the other encoder and hence also something about the other encoder's message, thus facilitating some cooperation. If this is the right explanation, then one would expect that feedback could not increase the capacity region of the MAC with cribbing encoders [2], where, before transmitting its time-i symbol, each encoder learns the symbol transmitted by the other encoder at time $i-1$. We shall see that this is indeed the case.

Like feedback on the MAC, strictly-causal state-information (SI) can increase the sum-rate capacity of the state-dependent multiple-access channel (SD-MAC) [3]. In analogy to the aforementioned result, we shall see that once the encoders are allowed to crib, strictly-causal SI becomes useless: the capacity region of the SD-MAC with cribbing encoders is not enlarged if SI is revealed to the encoders in a strictly-causal fashion. In fact, a stronger result holds: the capacity region of the SD-MAC with cribbing encoders is not increased even if we furnish the encoders with *both* feedback *and* strictly-causal SI (Theorem 3).

This result suggests that—unlike the SD-MAC with causal SI and no cribbing for which Shannon strategies are sub-optimal [3]—for the SD-MAC with causal SI and cribbing encoders Shannon strategies are optimal. This is indeed the case (Theorem 5). It is not, however, the case when the channel is governed by two (correlated) memoryless state sequences each of which is revealed causally to only one encoder.

We conclude by computing the capacity of the SD-MAC with two cribbing encoders, causal SI at one encoder, and one-sided feedback at the other encoder.

II. The SD-MAC with Cribbing Encoders

We consider a MAC with cribbing encoders, a state, and a single output. The encoders produce the symbols x_1 and x_2 in the finite sets $\mathcal{X}_1$ and $\mathcal{X}_2$. The state S takes value in the finite set $\mathcal{S}$, and the output Y takes value in the finite set $\mathcal{Y}$. The state sequence $S_1, \ldots, S_n$ is independent and identically distributed (IID) according to some given probability mass function (PMF) $P_S(\cdot)$. Here n denotes the transmission's blocklength. When the inputs are x_1, x_2 and the state is s, the channel output Y is of PMF

$$W(y|x_1, x_2, s).$$

In this paper we only consider strictly-causal cribbing as in [2, Sit. 5]. Denoting the set of messages of Encoder 1 and Encoder 2 by $\mathcal{M}_1$ and $\mathcal{M}_2$, we can now describe the encoders as follows. At every time instant i in $[1:n] = \{1, \ldots, n\}$, Encoder 1 produces the symbol $X_{1,i}$ based on its message M_1 and on the symbols previously transmitted by Encoder 2, namely, $X_{2,1}^{i-1} = X_{2,1}, \ldots, X_{2,i-1}$. Likewise Encoder 2. We study the capacity region of this channel with and without strictly-causal or causal SI under the average probability of error criterion.

In the absence of SI, the capacity region of this channel was found in [2] to be the set of rate-pairs (R_1, R_2) satisfying

$$R_1 \leq H(X_1|U) \tag{1a}$$

$$R_2 \leq H(X_2|U) \tag{1b}$$

$$R_1 + R_2 \leq I(X_1, X_2; Y) \tag{1c}$$

for some joint distribution $P_{S,U,X_1,X_2,Y}$ of the form

$$P_S P_U P_{X_1|U} P_{X_2|Y} W_{Y|X_1,X_2,S}. \tag{2}$$

Key here is that X_1 and X_2 are conditionally independent given U, and that the output Y is independent of U given (X_1, X_2). The following lemma is useful to establish conditional independence.

Lemma 1. *X and Y are conditionally independent given Z if, and only if, the joint distribution $P_{X,Y,Z}$ is of the form*

$$P(x, y, z) = g_1(x, z)\, g_2(y, z)\,. \tag{3}$$

978-1-4673-0223-4/12/$31.00 ©2012 IEEE

STRICTLY-CAUSAL CASE

If SI is available strictly-causally to the encoders, then $X_{1,i}$ is additionally allowed to depend on $S^{i-1} = S_1, \ldots, S_{i-1}$ and likewise $X_{2,i}$. The main result of this section is that the capacity region is not enlarged if SI is revealed to the cribbing encoders in a strictly-causal fashion. It is interesting to note that an analogous result does not hold in the absence of cribbing [3]: in the absence of cribbing, strictly-causal SI may increase capacity.

Theorem 1. *The capacity region of the SD-MAC with cribbing encoders does not increase if the state is revealed strictly-causally to the cribbing encoders.*

Proof: The claim is established by means of a converse, which is similar to the one in [2, Sec. V., Sit. 5] but that accounts for the SI: instead of $\left(X_{1,1}^{i-1}, X_{2,1}^{i-1}\right)$, we define the auxiliary random variable

$$U_i \triangleq \left(S^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}\right). \tag{4}$$

The rate of Encoder 1 satisfies

$$\begin{aligned} n\left(R_1 - \epsilon_n\right) & \qquad (5)\\ &\overset{a)}{\leq} I(M_1; Y^n, S^n, M_2) \\ &\overset{b)}{\leq} I(M_1; Y^n \,|S^n, M_2) \\ &\overset{c)}{\leq} I\left(X_{1,1}^n, M_1; Y^n \,|S^n, M_2\right) \\ &\overset{d)}{=} I\left(X_{1,1}^n; Y^n \,|S^n, M_2\right) \\ &\quad + I\left(M_1; Y^n \,|S^n, X_{1,1}^n, M_2\right) \\ &\overset{e)}{=} I\left(X_{1,1}^n; Y^n \,|S^n, M_2\right) \\ &\overset{f)}{=} \sum_{i=1}^n H\left(X_{1,i} \,|S^n, X_{1,1}^{i-1}, M_2\right) \\ &\overset{g)}{=} \sum_{i=1}^n H\left(X_{1,i} \,|S^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, M_2, S_i^n\right) \qquad (6) \\ &\overset{h)}{=} \sum_{i=1}^n H(X_{1,i} \,|U_i), \qquad (7) \end{aligned}$$

where $a)$ follows from Fano's inequality, $b)$ holds since M_1, S^n, and M_2 are independent, $c)$ is true because conditioning cannot increase entropy, $d)$ is due to the chain-rule, $e)$ holds since $X_{2,1}^n = f_{2,1}^n\left(M_2, X_{1,1}^{n-1}, S^{n-1}\right)$ and since M_1 and Y^n are conditionally independent given $\left(X_{1,1}^n, X_{2,1}^n, S^n\right)$, $f)$ is due to the non-negativity of entropy and the chain-rule, $g)$ is true because $X_{2,1}^{i-1} = f_{2,1}^{i-1}\left(M_2, X_{1,1}^{i-2}, S^{i-2}\right)$, and $h)$ is a consequence of (4) and the fact that conditioning cannot increase entropy. By symmetry,

$$n\left(R_2 - \epsilon_n\right) \leq \sum_{i=1}^n H(X_{2,i} \,|U_i).$$

The sum-rate satisfies

$$\begin{aligned} n\left(R_1 + R_2 - \epsilon_n\right) &\overset{a)}{\leq} I(M_1, M_2; Y^n) \\ &\overset{b)}{=} \sum_{i=1}^n I\left(M_1, M_2; Y_i \,\middle|Y^{i-1}\right) \qquad (8) \\ &\overset{c)}{\leq} \sum_{i=1}^n I\left(X_{1,i}, X_{2,i}, M_1, M_2, Y^{i-1}; Y_i\right) \\ &\overset{d)}{\leq} \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i), \end{aligned}$$

where $a)$ follows from Fano's inequality, $b)$ is due to the chain-rule, $c)$ holds since conditioning cannot increase entropy, and $d)$ is true because $\left(M_1, M_2, Y^{i-1}\right)$ and Y_i are conditionally independent given $(X_{1,i}, X_{2,i})$.

As we argue next, the joint PMF satisfies (2). Indeed, U_i and Y_i are conditionally independent given $(X_{1,i}, X_{2,i})$. As in [2, (58) - (60)], we can write

$$\begin{aligned} &P(u_i, m_1, m_2) \\ &\quad = P(m_1)\, P(m_2)\, P\left(s^{i-1}\right) \prod_{j=1}^{i-1} \Big[P\Big(x_{1,j} \,\Big|m_1, x_{2,1}^{j-1}, s^{j-1}\Big) \\ &\qquad P\Big(x_{2,j} \,\Big|m_2, x_{1,1}^{j-1}, s^{j-1}\Big)\Big]. \qquad (9) \end{aligned}$$

Since the PMF is of the form (3), Lemma 1 implies that M_1 and M_2 are conditionally independent given U_i. As a consequence, also $X_{1,i}$ and $X_{2,i}$, which only depend on (M_1, U_i) and (M_2, U_i), are conditionally independent given U_i. Since the capacity region of the MAC with cribbing encoders is convex [2, App. A], the claim follows. ■

The capacity region of the SD-MAC with strictly-causal SI cannot decrease if we also allow the encoders to crib. Since in the latter case strictly-causal SI does not increase capacity, we obtain the following results, which is tighter than the full-cooperation bound of [3]:

Corollary 2. *The capacity region of the SD-MAC with strictly-causal SI is contained in the capacity region of the MAC without SI but with cribbing encoders.*

Theorem 1 can be strengthened by also allowing feedback:

Theorem 3. *The capacity region of the SD-MAC with cribbing encoders is not enlarged even if both strictly-causal SI and feedback are furnished to the encoders.*

Proof: As in [2, Functional Representation Lemma], any SD-MAC of law $W(y|x_1, x_2, s)$ can be described as an SD-MAC whose time-i output Y_i is

$$Y_i = g\left(x_{1,i}, x_{2,i}, S_i, \Theta_i\right), \tag{10}$$

where $\{\Theta_i\}$ are IID and take value in a finite set. Define

$$\Xi_i = (S_i, \Theta_i), \quad i \in [1:n]. \tag{11}$$

We can conclude from Theorem 1 that revealing the super-state Ξ to the encoders in a strictly-causal fashion does not increase the capacity region of the SD-MAC with cribbing encoders. That is, allowing $X_{1,i}$ to depend not only on $(M_1, X_{2,1}^{i-1})$ but also on Ξ^{i-1} and likewise for $X_{2,i}$ does not increase the capacity region. But providing feedback and revealing the state S_i strictly-causally to the cribbing encoders is no better than the case where $X_{1,i}$ may depend on $(M_1, X_{2,1}^{i-1}, \Xi^{i-1})$ (and likewise for $X_{2,i}$), since (10) guarantees that any function of $(M_1, X_{2,1}^{i-1}, Y^{i-1}, S^{i-1})$ can be represented as a function of $(M_1, X_{2,1}^{i-1}, \Xi^{i-1})$. ■

Since cribbed information may be ignored and since feedback cannot increase the capacity of the MAC with cribbing encoders, we find:

Corollary 4. *The feedback capacity region of the MAC is contained in the capacity region of the MAC without feedback but with cribbing encoders.*

This result can also be obtained from [4, Thm. 2] if we use $Z = (X_1, X_2)$. Note that the outer bound in [7] is tighter.

III. Causal Case

If common causal SI is available to the cribbing encoders, then $X_{1,i}$ and $X_{2,i}$ are additionally allowed to depend on S^i. As our next result shows, Shannon strategies are optimal.

Theorem 5. *The capacity-region of the SD-MAC with common causal SI at the cribbing encoders is the set of rate-pairs (R_1, R_2) satisfying*

$$R_1 \leq H(X_1 | S, U) \tag{12a}$$

$$R_2 \leq H(X_2 | S, U) \tag{12b}$$

$$R_1 + R_2 \leq I(T_1, T_2, U; Y) \tag{12c}$$

for some joint distribution $P_{S,U,T_1,T_2,X_1,X_2,Y}$ of the form

$$\begin{aligned} &P(s, u, t_1, t_2, x_1, x_2, y) \\ &\quad = P(s)\, P(u)\, P(t_1 | u)\, P(t_2 | u)\, \mathbb{1}\{x_1 = g_1(t_1, u, s)\} \\ &\qquad \mathbb{1}\{x_2 = g_2(t_2, u, s)\}\, W(y | x_1, x_2, s)\,. \end{aligned} \tag{13}$$

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

Converse: Define the auxiliary random variables

$$U_i \triangleq \left(S^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}\right), T_{1,i} \triangleq M_1, T_{2,i} \triangleq M_2. \tag{14}$$

The rate of Encoder 1 satisfies

$$\begin{aligned} &n\,(R_1 - \epsilon_n) \\ &\quad \overset{a)}{\leq} \sum_{i=1}^{n} H\big(X_{1,i} | S_i, S^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, S_{i+1}^{n}, M_2\big) \\ &\quad \overset{b)}{\leq} \sum_{i=1}^{n} H(X_{1,i} | S_i, U_i)\,, \end{aligned}$$

where $a)$ follows from (6) (note that $e)$ and $g)$ also apply if $X_{2,i} = f_{2,i}\big(M_2, X_{1,1}^{i-1}, S^i\big)$) and $b)$ holds because of (14) and since conditioning cannot increase entropy. Likewise

$$n\,(R_2 - \epsilon_n) \leq \sum_{i=1}^{n} H(X_{2,i} | S_i, U_i)\,.$$

The sum-rate satisfies

$$\begin{aligned} &n\,(R_1 + R_2 - \epsilon_n) \\ &\quad \overset{a)}{\leq} \sum_{i=1}^{n} I\big(M_1, M_2, S^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, Y^{i-1}; Y_i\big) \\ &\quad \overset{b)}{=} \sum_{i=1}^{n} I(T_{1,i}, T_{2,i}, U_i; Y_i)\,, \end{aligned} \tag{15}$$

where $a)$ holds because of (8) and since conditioning cannot increase entropy and $b)$ is due to (14) and the Markov chain $Y^{i-1} - (U_i, M_1, M_2) - Y_i$.

The joint PMF satisfies (13) since Y_i and $(U_i, T_{1,i}, T_{2,i})$ are conditionally independent given $(X_{1,i}, X_{2,i}, S_i)$, $X_{1,i}$ and $X_{2,i}$ are deterministic given $(T_{1,i}, U_i, S_i)$ and $(T_{2,i}, U_i, S_i)$, and, if s^{j-1} is replaced by s^j, (9) implies that $M_1 = T_{1,i}$ and $M_2 = T_{2,i}$ are conditionally independent given U_i. To conclude the converse, note that the region (12) is convex (time-sharing may be achieved via U_i).

Direct Part: To prove achievability, we use *backward decoding* [2] but instead of directly coding over input-symbols, the messages are mapped to *Shannon strategy* sequences and the channel inputs are generated by evaluating each Shannon strategy for the observed realization of the channel state [6].

Along the line of [2, Sec. V., Sit. 5], fix $\epsilon > 0$, functions $g_1 \colon \mathcal{T}_1 \times \mathcal{U} \times \mathcal{S} \mapsto \mathcal{X}_1$, $g_2 \colon \mathcal{T}_2 \times \mathcal{U} \times \mathcal{S} \mapsto \mathcal{X}_2$, and a PMF

$$P(u, t_1, t_2) = P(u)\, P(t_1 | u)\, P(t_2 | u)\,.$$

Codebook Generation: For b in $[1 : B]$, draw $2^{n(R_1+R_2)}$ length-n sequences u from the PMF $\prod_{i=1}^{n} P(u_i)$. Index them $m_0 = (m_{0,1}, m_{0,2})$, $m_{0,1}$ in $\left[1 : 2^{nR_1}\right]$, $m_{0,2}$ in $\left[1 : 2^{nR_2}\right]$. For l in $\{1, 2\}$ and for every m_0 in $\left\{(1,1), \ldots, \left(2^{nR_1}, 2^{nR_2}\right)\right\}$, draw 2^{nR_l} length-n Shannon strategy sequences t_l from the PMF $\prod_{i=1}^{n} P(t_{l,i} | u_i(m_0))$.

Encoding: Split the messages m_1 and m_2 into $B-1$ blocks b in $[1 : B-1]$ of equal length, i.e., $m_1 = m_{1,1}, \ldots, m_{1,B-1}$ and $m_2 = m_{2,1}, \ldots, m_{2,B-1}$. In the first block, the encoders choose the sequences $\mathrm{t}_{1,1} = \mathrm{t}_1((1,1), m_{1,1})$, $\mathrm{t}_{2,1} = \mathrm{t}_2((1,1), m_{2,1})$. For b in $[2 : B]$, the encoders form estimates $\hat{m}_{2,b-1}$ and $\tilde{m}_{1,b-1}$ of $m_{2,b-1}$ and $m_{1,b-1}$. Denote $\hat{m}_{0,b} = (m_{1,b-1}, \hat{m}_{2,b-1})$ and $\tilde{m}_{0,b} = (\tilde{m}_{1,b-1}, m_{2,b-1})$. Then, the sequences chosen in block b are $\mathrm{t}_{1,b} = \mathrm{t}_1(\hat{m}_{0,b}, m_{1,b})$, $\mathrm{t}_{2,b} = \mathrm{t}_2(\tilde{m}_{0,b}, m_{2,b})$. In the last block B, only resolution information is sent, i.e., $\mathrm{t}_{1,B} = \mathrm{t}_1(\hat{m}_{0,B}, 1)$, $\mathrm{t}_{2,B} = \mathrm{t}_2(\tilde{m}_{0,B}, 1)$. For b in $[1 : B]$ and at time i in $[1 : n]$, the channel inputs are $x_{1,(b-1)n+i} = g_1\big([\mathrm{t}_{1,b}]_i, s_{(b-1)n+i}\big)$, $x_{2,(b-1)n+i} = g_2\left([\mathrm{t}_{2,b}]_i, s_{(b-1)n+i}\right)$.

Handling Cribbed Information: For b in $[1 : B-1]$, the encoders choose $\hat{m}_{2,b}$ and $\tilde{m}_{1,b}$ s.t.

$$\begin{aligned} &\left(\mathrm{u}(\hat{m}_{0,b}), \mathrm{t}_2(\hat{m}_{0,b}, \hat{m}_{2,b}), s_{(b-1)n+1}^{bn}, x_{2,(b-1)n+1}^{bn}\right) \\ &\quad \in \mathcal{T}_\epsilon^{(n)}(U, T_2, S, X_2) \\ &\left(\mathrm{u}(\tilde{m}_{0,b}), \mathrm{t}_1(\tilde{m}_{0,b}, \tilde{m}_{1,b}), s_{(b-1)n+1}^{bn}, x_{1,(b-1)n+1}^{bn}\right) \\ &\quad \in \mathcal{T}_\epsilon^{(n)}(U, T_1, S, X_1)\,. \end{aligned}$$

($\check{m}_{0,1} = \tilde{m}_{0,1} = m_{0,1} = (1,1)$, $\check{m}_{0,b}$, $\tilde{m}_{0,b}$ are formed after block $b-1$, and $x^{bn}_{2,(b-1)n+1}$, $x^{bn}_{1,(b-1)n+1}$ are cribbed.)

Decoding: The receiver looks for $\check{m}_{0,B}$ s.t.

$$\Big(\mathrm{u}(\check{m}_{0,B})\,, \mathrm{t}_1(\check{m}_{0,B},1)\,, \mathrm{t}_2(\check{m}_{0,B},1)\,, y^{Bn}_{(B-1)n+1}\Big) \in \mathcal{T}^{(n)}_{\epsilon}(U,T_1,T_2,Y)\,.$$

Fix b in $[2:B-1]$ and assume the decoder has already found $(\check{m}_{0,b+1}, \check{m}_{1,b+1}, \check{m}_{2,b+1})$. Upon setting $\check{m}_{1,b} = [\check{m}_{0,b+1}]_1$ and $\check{m}_{2,b} = [\check{m}_{0,b+1}]_2$, it looks for $\check{m}_{0,b}$ s.t.

$$\Big(\mathrm{u}(\check{m}_{0,b})\,, \mathrm{t}_1(\check{m}_{0,b},\check{m}_{1,b})\,, \mathrm{t}_2(\check{m}_{0,b},\check{m}_{2,b})\,, y^{bn}_{(b-1)n+1}\Big) \in \mathcal{T}^{(n)}_{\epsilon}(U,T_1,T_2,Y)\,.$$

Probability of Error: The average error probability is

$$P_e = \Pr\left[\bigcup_{b=1}^{B-1} \left(\check{M}_{1,b} \neq M_{1,b} \cup \check{M}_{2,b} \neq M_{2,b}\right)\right].$$

Define the events

$$\mathcal{E}^1_{b,m_1} = \Big\{\Big(\mathrm{u}(M_{0,b})\,, \mathrm{t}_1(M_{0,b},m_1)\,, S^{bn}_{(b-1)n+1}, X^{bn}_{1,(b-1)n+1}\Big) \in \mathcal{T}^{(n)}_{\epsilon}(U,T_1,S,X_1)\Big\}$$

$$\mathcal{E}^2_{b,m_2} = \Big\{\Big(\mathrm{u}(M_{0,b})\,, \mathrm{t}_2(M_{0,b},m_2)\,, S^{bn}_{(b-1)n+1}, X^{bn}_{2,(b-1)n+1}\Big) \in \mathcal{T}^{(n)}_{\epsilon}(U,T_2,S,X_2)\Big\}$$

$$\mathcal{E}^0_{b,m_\bullet} = \Big\{\Big(\mathrm{u}(m_0)\,, \mathrm{t}_1(m_0,M_{1,b})\,, \mathrm{t}_2(m_0,M_{2,b})\,, Y^{bn}_{(b-1)n+1}\Big) \in \mathcal{T}^{(n)}_{\epsilon}(U,T_1,T_2,Y)\Big\}\,.$$

For b in $[1:B]$, we can w.l.g. assume $(m_{0,b}, m_{1,b}, m_{2,b}) = ((1,1)\,,1,1)$. In particular,

$$P_e \leq \Pr\left\{\bigcup_{b=1}^{B-1}\left(\mathcal{E}^{1c}_{b,1} \cup \mathcal{E}^{2c}_{b,1} \cup \bigcup_{m_1\neq 1}\mathcal{E}^1_{b,m_1} \cup \bigcup_{m_2\neq 1}\mathcal{E}^2_{b,m_2}\right) \cup \bigcup_{b=2}^{B}\left(\mathcal{E}^{0c}_{b,(1,1)} \cup \bigcup_{m_\bullet\neq(1,1)}\mathcal{E}^0_{b,m_\bullet}\right)\right\}.$$

Since the codebook generation is independent of the block b

$$P_e \leq (B-1)\left(\Pr\big[\mathcal{E}^{1c}_{2,1}\big] + \Pr\big[\mathcal{E}^{2c}_{2,1}\big] + \sum_{m_1\neq 1}\Pr\big[\mathcal{E}^1_{2,m_1}\big] + \sum_{m_2\neq 1}\Pr\big[\mathcal{E}^2_{2,m_2}\big] + \Pr\Big[\mathcal{E}^{0c}_{2,(1,1)}\Big] + \sum_{m_\bullet\neq(1,1)}\Pr\big[\mathcal{E}^0_{2,m_\bullet}\big]\right). \tag{16}$$

Because of the properties of weakly typical sequences

$$\Pr\big[\mathcal{E}^{1c}_{2,1}\big]\,, \Pr\big[\mathcal{E}^{2c}_{2,1}\big]\,, \Pr\Big[\mathcal{E}^{0c}_{2,(1,1)}\Big] \to 0\,(n\to\infty) \tag{17}$$

$$\begin{aligned}
\sum_{m_1\neq 1}\Pr\big[\mathcal{E}^1_{2,m_1}\big] &\leq 2^{nR_1}\sum_{\mathcal{T}^{(n)}_{\epsilon}(U,T_1,S,X_1)} P\big(u^n,s^n,x^n_{1,1}\big)\,P\big(t^n_{1,1}\,|u^n\big)\\
&\leq 2^{nR_1}2^{-n(-H(U,T_1,S,X_1)+H(U,S,X_1)+H(T_1|U)-4\epsilon)}\\
&\overset{a)}{=} 2^{-n(H(X_1|S,U)-4\epsilon-R_1)} \qquad (18)
\end{aligned}$$

$$\sum_{m_2\neq 1}\Pr\big[\mathcal{E}^2_{2,m_2}\big] \leq 2^{-n(H(X_2|S,U)-4\epsilon-R_2)} \tag{19}$$

$$\begin{aligned}
\sum_{m_\bullet\neq(1,1)}\Pr\big[\mathcal{E}^0_{2,m_\bullet}\big] &\leq 2^{n(R_1+R_2)}\sum_{\mathcal{T}^{(n)}_{\epsilon}(U,T_1,T_2,Y)} P\big(u^n,t^n_{1,1},t^n_{2,1}\big)\,P(y^n)\\
&\leq 2^{n(R_1+R_2)}2^{-n(-H(U,T_1,T_2,Y)+H(U,T_1,T_2)+H(Y)-3\epsilon)}\\
&\leq 2^{-n(I(T_1,T_2,U;Y)-3\epsilon-(R_1+R_2))}, \qquad (20)
\end{aligned}$$

where $a)$ holds since (T_1,U) and S are independent and X_1 is deterministic given (T_1,U,S). Equations (16) - (20) imply $P_e \to 0\,(n\to\infty)$ provided that B is sufficiently small and

$$\begin{aligned}
R_1 &< H(X_1\,|S,U) - 4\epsilon\\
R_2 &< H(X_2\,|S,U) - 4\epsilon\\
R_1 + R_2 &< I(T_1,T_2,U;Y) - 3\epsilon.
\end{aligned}$$

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

$$\frac{1}{nB}\log|\mathcal{M}_l| = \frac{B-1}{B}R_l \to R_l\,(B\to\infty)\,.$$

∎

Note 6. *The capacity region of the SD-MAC is not increased if, in addition to common causal SI, feedback is furnished to each of the cribbing encoders [8, Thm. 2.20].*

The random coding argument of Theorem 5 (but where each encoder uses only the state sequence it observes to estimate the common message) is also applicable if the MAC is governed by two arbitrarily correlated state sequences each of which is revealed causally to only one of the cribbing encoders. It does not, however, result in a tight inner bound. This is seen in the following example where one of the state sequences is null.

Consider the SD-MAC $Y = X_2 \oplus S$ with binary inputs and a Bernoulli state with success probability $1/2$. We assume that both encoders crib but the state is available causally only to Encoder 2. Let Message 2 take value in $[1:2^n]$ and denote its binary representation $w^n_{2,1}$. If for i in $[1:n]$ Encoder 2 transmits $X_{2,i} = w_{2,i} \oplus S_i$, the receiver observes $Y_i = X_{2,i} \oplus S_i = w_{2,i}$ and can perfectly decode Message 2 after n transmissions. In particular, $R_2 = 1$ is achievable. If the encoders use Shannon strategies and backward decoding as in Theorem 5, $R_2 = 1$ is not achievable [8, Ex. 2.2]. For the described coding-scheme this is intuitively clear: to establish cooperation in backward decoding Encoder 1 must decode Message 2. This is, however, impossible since the state sequence assumes any value in $[1:2^n]$ with equal probability and $X^n_{2,1} = w^n_{2,1} \oplus S^n$ is therefore independent of $w^n_{2,1}$.

The following conditions are sufficient to prove that encoding as in Theorem 5 is optimal:

1) Message 2 and the output sequence are conditionally independent given the information available to Encoder 1 and vice versa.
2) The state sequences are conditionally independent given the common information (cf. (13)).

Let S_1 and S_2 denote the states that are available causally to Encoders 1 and 2. The above conditions are satisfied and Shannon strategies are optimal if S_1 and S_2 are furnished strictly causally to both encoders, or feedback is available to both encoders and S_1 is deterministic given S_2 and Y, or S_1 is deterministic given S_2 and feedback is provided only to Encoder 1 [8, Cor. 2.23 and Thm. 2.25]. For example, if one state sequence is null, the following result holds:

Theorem 7. *The capacity region of the SD-MAC with cribbing encoders, causal SI at Encoder 2, and feedback at Encoder 1 is the set of rate-pairs* (R_1, R_2) *satisfying*

$$R_1 \leq H(X_1 | U) \tag{21a}$$

$$R_2 \leq I(T_2; Y, X_2 | X_1, U) \tag{21b}$$

$$R_1 + R_2 \leq I(X_1, T_2, U; Y) \tag{21c}$$

for some joint distribution $P_{S,U,X_1,T_2,X_2,Y}$ *of the form*

$$\begin{aligned} &P(s, u, x_1, t_2, x_2, y) \\ &\quad = P(s)\, P(u)\, P(x_1 | u)\, P(t_2 | u) \\ &\quad\quad \mathbb{1}\{x_2 = g_2(t_2, u, s)\}\, W(y | x_1, x_2, s)\,. \end{aligned} \tag{22}$$

Proof: We briefly highlight the difference to Theorem 5.

Converse: Define the auxiliary random variables

$$U_i \triangleq \left(Y^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}\right), T_{2,i} \triangleq \left(M_2, S^{i-1}\right). \tag{23}$$

If we argue as in Theorem 3 and recall that conditioning cannot increase entropy, the bound on R_1 follows from (7). The sum-rate bound is due to (15), $X_{1,i} = f_{1,i}\left(M_1, X_{2,1}^{i-1}, Y^{i-1}\right)$, and the Markov chain $M_1 - (X_{1,i}, T_{2,i}, U_i) - Y_i$ (cf. (24)). The rate of Encoder 2 satisfies

$$\begin{aligned} &n\,(R_2 - \epsilon_n) \\ &\overset{a)}{\leq} I\left(M_2; Y^n, X_{1,1}^n, X_{2,1}^n, M_1\right) \\ &\overset{b)}{=} \sum_{i=1}^n I\left(M_2; Y_i, X_{1,i}, X_{2,i} \,|\, Y^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, M_1\right) \\ &\overset{c)}{=} \sum_{i=1}^n I\left(M_2; Y_i, X_{2,i} \,|\, X_{1,i}, Y^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, M_1\right) \\ &\overset{d)}{=} \sum_{i=1}^n I\left(M_2, S^{i-1}; Y_i, X_{2,i} \,|\, X_{1,i}, Y^{i-1}, X_{1,1}^{i-1}, X_{2,1}^{i-1}, M_1\right) \\ &\overset{e)}{\leq} \sum_{i=1}^n I(T_{2,i}; Y_i, X_{2,i} \,|\, X_{1,i}, U_i) \end{aligned}$$

where $a)$ follows from Fano's inequality, $b)$ is due to the chain-rule and the independence of M_1 and M_2, $c)$ is true because $X_{1,i} = f_{1,i}\left(M_1, X_{2,1}^{i-1}, Y^{i-1}\right)$, $d)$ holds since conditioning cannot increase entropy, and $e)$ is due to (23) and the Markov chain $M_1 - (X_{1,i}, U_i) - (Y_i, X_{2,i}, T_{2,i})$ (cf. (24)).

As we argue next, the joint PMF satisfies (22). Indeed, $(T_{2,i}, U_i)$ and Y_i are conditionally independent given $(X_{1,i}, X_{2,i}, S_i)$ and $X_{2,i}$ is deterministic given $(T_{2,i}, U_i, S_i)$. Since the joint PMF of $\left(U_i, M_1, M_2, S^{i-1}, X_{1,i}, X_{2,i}, Y_i\right)$ is

$$\begin{aligned} &P\left(u_i, m_1, m_2, s^{i-1}, x_{1,i}, x_{2,i}, y_i\right) = P(m_1)\, P(m_2)\, P\left(s^{i-1}\right) \\ &\prod_{j=1}^{i-1} \Big[P\left(x_{1,j} \,\middle|\, m_1, x_{2,1}^{j-1}, y^{j-1}\right) P\left(x_{2,j} \,\middle|\, m_2, x_{1,1}^{j-1}, s^{j}\right) \\ &\quad W(y_j | x_{1,j}, x_{2,j}, s_j) \Big] P\left(x_{1,i} \,\middle|\, m_1, x_{2,1}^{i-1}, y^{i-1}\right) \\ &\quad P\left(y_i, x_{2,i} \,\middle|\, x_{1,i}, m_2, x_{1,1}^{i-1}, s^{i-1}\right), \end{aligned} \tag{24}$$

$M_1 - U_i - T_{2,i}$ and $M_1 - (X_{1,i}, U_i) - (Y_i, X_{2,i}, T_{2,i})$ are Markov chains (cf. Lemma 1). To conclude the converse, note that the rate-region (21) is convex.

Direct Part: Fix $\epsilon > 0$, $g_2 \colon \mathcal{T}_2 \times \mathcal{U} \times \mathcal{S} \mapsto \mathcal{X}_2$, and

$$P(u, x_1, t_2) = P(u)\, P(x_1 | u)\, P(t_2 | u)\,.$$

Codebook Generation: Encoder 1 directly codes over input-symbols: for m_0 in $\left\{(1,1), \ldots, \left(2^{nR_1}, 2^{nR_2}\right)\right\}$ draw 2^{nR_1} length-n codewords x_1 from the PMF $\prod_{i=1}^n P(x_{1,i} | u_i(m_0))$.

Handling Cribbed Information: Choose $\hat{m}_{2,b}$ and $\tilde{m}_{1,b}$ s.t.

$$\begin{aligned} &\Big(\mathrm{u}(\hat{m}_{0,b}), \mathrm{t}_2(\hat{m}_{0,b}, \hat{m}_{2,b}), y_{(b-1)n+1}^{bn}, x_{1,(b-1)n+1}^{bn}, \\ &\quad x_{2,(b-1)n+1}^{bn}\Big) \in \mathcal{T}_\epsilon^{(n)}(U, T_2, Y, X_1, X_2) \\ &\Big(\mathrm{u}(\tilde{m}_{0,b}), \mathrm{x}_1(\tilde{m}_{0,b}, \tilde{m}_{1,b}), x_{1,(b-1)n+1}^{bn}\Big) \in \mathcal{T}_\epsilon^{(n)}(U, X_1, X_1)\,. \end{aligned}$$

Decoding: For $b \in [1:n]$, the decoder looks for $\check{m}_{0,b}$ s.t.

$$\begin{aligned} &\Big(\mathrm{u}(\check{m}_{0,b}), \mathrm{x}_1(\check{m}_{0,b}, \check{m}_{1,b}), \mathrm{t}_2(\check{m}_{0,b}, \check{m}_{2,b}), y_{(b-1)n+1}^{bn}\Big) \\ &\quad \in \mathcal{T}_\epsilon^{(n)}(U, X_1, T_2, Y)\,. \end{aligned}$$

Probability of Error: The analysis parallels Theorem 5 but accounts for the differences in the handling of cribbed information and in the decoding (cf. [8, Thm. 2.25]). ■

REFERENCES

[1] N.T. Gaarder and J.K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. on Inform. Theory*, IT-21, No. 1, pp. 100–102, Jan. 1975.

[2] F.M.J. Willems and E. Van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. on Inform. Theory*, IT-31, No. 3, pp. 313–327, May 1985.

[3] A. Lapidoth and Y. Steinberg, "The Multiple-Access Channel with Causal Side Information: Common State," *IEEE Trans. on Inform. Theory*, to appear.

[4] A.P. Hekstra and F.M.J. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. on Inform. Theory*, IT-35, No. 1, pp. 44–53, Jan. 1989.

[5] T.M. Cover and C.S.K. Leung, "An achievable rate-region for the multiple-access channel with feedback," *IEEE Trans. on Inform. Theory*, IT-27, No. 3, pp. 292–298, May 1981.

[6] C.E. Shannon, "Channels with side-information at the transmitter," *IBM Journal of Research and Development*, Vol. 2, pp. 289–293, Oct. 1958.

[7] R. Tandon and S. Ulukus, "Outer Bounds for Multiple-Access Channels with Feedback Using Dependence Balance," *IEEE Trans. on Inform. Theory*, IT-55, No. 10, pp. 4494–4507, Oct. 2009.

[8] A. Bracher, "Cribbing, conferencing, feedback, and state-information on the Multiple-Access Channel," Master's thesis under the supervision of A. Lapidoth, *Swiss Federal Institute of Technology, Zürich*, July 2012.