The State-Dependent Multiple-Access Channel with States Available at a Cribbing Encoder

Shraga I. Bross, and Amos Lapidoth

Abstract—The two-user discrete memoryless state-dependent multiple-access channel (MAC) models a scenario in which two encoders transmit independent messages to a single receiver via a MAC whose channel law is governed by the pair of encoders’ inputs and by an i.i.d. state random variable. In the cooperative state-dependent MAC model it is further assumed that Message 1 is shared by both encoders whereas Message 2 is known only to Encoder 2 – the cognitive transmitter. The capacity of the cooperative state-dependent MAC where the realization of the state sequence is known non-causally to the cognitive encoder has been derived by Somekh-Baruch et al.

In this work we dispense of the assumption that Message 1 is shared a-priori by both encoders. Instead, we study the case in which Encoder 2 cribbs causally from Encoder 1. We determine the capacity region for both, the case where Encoder 2 cribbs strictly causal – i.e. its current channel input depends on its message as well as all past inputs and the case where Encoder 2 cribbs causally from Encoder 1.

Index Terms—State-dependent MAC, Gel’fand-Pinsker channel, cribbing encoder.

I. INTRODUCTION

The two-user discrete memoryless state-dependent multiple-access channel (MAC) models a scenario in which two encoders transmit independent messages to a single receiver via a MAC whose channel law is governed by the pair of encoders’ inputs and by an i.i.d. state random variable $S$. In the cooperative state-dependent MAC model it is further assumed that Message 1 is shared by both encoders whereas Message 2 is known only to Encoder 2 – the cognitive transmitter. The capacity of the cooperative state-dependent MAC where the realization of the state sequence is known non-causally to the cognitive encoder has been derived by Somekh-Baruch et al. in [1].

In this work we dispense of the assumption that Message 1 is shared a-priori by both encoders. Instead, we study a “more realistic” model in which Encoder 2 “cribs” and learns the sequence of channel inputs emitted by Encoder 1 before generating its next channel input. Specifically, we study both, the case where Encoder 2 cribbs strictly causal – i.e. its current channel input depends on its message as well as the past inputs of Encoder 1 (in the sense of [2] Situation 3)), and the case where Encoder 2 cribbs causally – i.e. its current channel input depends on its message as well as all past including the current inputs of Encoder 1 (in the sense of [2] Situation 3)). The model is depicted in Figure 1.

For both cases – strictly causal cribbing as well as causal cribbing – we provide a complete characterization of the capacity region.

The paper is organized as follows. In Section II we provide a formal definition for the state-dependent MAC with a cribbing encoder. In Section III we present our main results, while Section IV is devoted for the proofs.

II. CHANNEL MODEL

A discrete memoryless state-dependent multiple-access channel is a triple $(X_1 \times X_2 \times S, p(y|x_1, x_2, s), Y)$ where $X_1$ and $X_2$ are finite sets corresponding to the input alphabets of Encoder 1 and Encoder 2 respectively, $S$ is a finite set corresponding to the alphabet of the state governing the channel law, the finite set $Y$ is the output alphabet at the receiver, and $p(y|x_1, x_2, s)$ is a collection of probability laws on $Y$ indexed by the input symbols $x_1 \in X_1$ and $x_2 \in X_2$ and $s \in S$. The channel’s law extends to $n$-tuples according to the memoryless law

$$\Pr(y^n|x^n_1, x^n_2, s^n) = \prod_{k=1}^{n} p(y_k|x_{1,k}, x_{2,k}, s_k),$$

where $x_{1,k}, x_{2,k}, s_k$ and $y_k$ denote the inputs, state and output of the channel at time $k$, and $x_{1,k} \triangleq (x_{1,1}, \ldots, x_{1,k})$.

Encoder 1 sends a message $W_1$, which is drawn uniformly over the set $\{1, \ldots, e^{nR_1}\} \triangleq \mathcal{W}_1$ to the receiver, while Encoder 2 sends to the receiver a message $W_2$ which is independent of $W_1$ and is drawn uniformly over the set $\{1, \ldots, e^{nR_2}\} \triangleq \mathcal{W}_2$. The channel state sequence $S^n$, which is drawn i.i.d. according to the law $p_S$, is available non-causally to Encoder 2. It is further assumed that Encoder 2 “cribs” causally
and learns the sequence of channel inputs emitted by Encoder 1 in all past transmissions (in the sense of [2] Situation 2) before generating its next channel input. The model is depicted in Figure 1.

An \((e^{nR_1}, e^{nR_2}, n)\) code for the state-dependent multiple-access channel with a strictly causal cribbing encoder consists of:

1) Encoder 1 defined by a deterministic mapping
\[ f_1: W_1 \rightarrow X_1^n \] (1)
which maps a message \(w_1 \in W_1\) to a codeword \(x_1^n \in X_1^n\).

2) Encoder 2 defined by a collection of encoding functions
\[ f_{2,k}^{(sc)}: W_2 \times S^n \times X_1^{k-1} \rightarrow X_2 \quad k = 1, 2, \ldots, n \] (2)
which, based on the message \(w_2 \in W_2\), the state sequence \(s^n \in S^n\) and what was learned from the other encoder by cribbing \(x_1^{k-1} \in X_1^{k-1}\), map into the next channel input \(x_{2,k} \in X_2\).

3) The receiver decoder defined by a mapping
\[ g: Y^n \rightarrow W_1 \times W_2 \]
which maps a received sequence \(y^n\) to a message pair \((\hat{w}_1, \hat{w}_2)\) in \(W_1 \times W_2\).

An \((e^{nR_1}, e^{nR_2}, n)\) code for the state-dependent multiple-access channel with a causal cribbing encoder differs from that for a strictly causal encoder just in the encoding rule at Encoder 2 which is defined by a collection of encoding functions
\[ f_{2,k}^{(c)}: W_2 \times S^n \times X_1^k \rightarrow X_2 \quad k = 1, 2, \ldots, n \] (3)
which, based on the message \(w_2 \in W_2\), the state sequence \(s^n \in S^n\) and what was learned from the other encoder by cribbing \(x_1^k \in X_1^k\), map into the current channel input \(x_{2,k} \in X_2\).

For a given code, the block average probability of error is
\[ P_e^{(n)} = \frac{1}{e^{n(R_1 + R_2)}} \sum_{w_1=1}^{e^{nR_1}} \sum_{w_2=1}^{e^{nR_2}} P_e^{(n)}(w_1, w_2) \]
where
\[ P_e^{(n)}(w_1, w_2) = \Pr \{ (\hat{w}_1, \hat{w}_2) \neq (w_1, w_2) | (w_1, w_2) \text{ sent} \} . \]
A rate-pair \((R_1, R_2)\) is said to be achievable if there exists a sequence of \((e^{nR_1}, e^{nR_2}, n)\) codes with \(\lim_{n \rightarrow \infty} P_e^{(n)} = 0\). The capacity region of the state-dependent MAC with a cribbing encoder is the closure of the set of achievable rate-pairs.

III. MAIN RESULTS

Our first result is a characterization of the capacity region for the two-user discrete memoryless state-dependent MAC with state-sequence available non-causally at a strictly causal cribbing encoder. By combining the coding strategies from [1] and [2] we prove the following.

\[ \text{Theorem 1:} \]
Consider the discrete memoryless state-dependent MAC \((X_1 \times X_2 \times S, p(y|x_1, x_2, s), \mathcal{Y})\) with state-sequence available non-causally at a strictly causal cribbing encoder and finite alphabets \(S, X_1, X_2\). The capacity region of this channel is
\[ \mathcal{C} = \bigcup_{PVUX_1X_2Y} \left\{ (R_1, R_2) : \begin{align*}
0 &\leq R_1 \leq H(X_1|V) \\
0 &\leq R_2 \leq I(U; Y|VX_1) - I(U; S|V) \\
0 &\leq R_1 + R_2 \leq I(VUX_1; Y) - I(U; S|V) \end{align*} \right\} , \] (4)
where the union in (4) is over all laws on \(V \in \mathcal{V}, S \in \mathcal{S}, U \in \mathcal{U}, X_1 \in \mathcal{X}_1, X_2 \in \mathcal{X}_2, Y \in \mathcal{Y}\) of the form
\[ p_{VUSX_1X_2Y}(v, s, u, x_1, x_2, y) = p_V(v)p_S(s)p_{X_1|V}(x_1|v)p_{UX_2|SV}(u, x_2|s, v)p(y|x_1, x_2, s). \] (5)

The cardinalities of the auxiliary random variables \(V, U\) are bounded by
\[ |\mathcal{V}| \leq |\mathcal{X}_1||\mathcal{X}_2||\mathcal{S}| + 5 \]
\[ |\mathcal{U}| \leq |\mathcal{X}_1||\mathcal{X}_2||\mathcal{S}||\mathcal{V}| + 2. \]

Our second result is a characterization of the capacity region for the two-user discrete memoryless state-dependent MAC with state-sequence available non-causally at a causal cribbing encoder.

\[ \text{Theorem 2:} \]
Consider the discrete memoryless state-dependent MAC \((X_1 \times X_2 \times S, p(y|x_1, x_2, s), \mathcal{Y})\) with state-sequence available non-causally at a causal cribbing encoder and finite alphabets \(S, X_1, X_2\). The capacity region of this channel is the set of rate pairs satisfying (4) except that the union is taken over all laws on \(V \in \mathcal{V}, S \in \mathcal{S}, U \in \mathcal{U}, X_1 \in \mathcal{X}_1, X_2 \in \mathcal{X}_2, Y \in \mathcal{Y}\) of the form
\[ p_{VUSX_1X_2Y}(v, s, u, x_1, x_2, y) = p_V(v)p_S(s)p_{X_1|V}(x_1|v)p_{UX_2|SV}(u, x_2|s, v)p(y|x_1, x_2, s). \] (6)

IV. PROOFS

A. Proof of the achievability part in Theorem 1

We propose a coding scheme that is based on Block-Markov superposition encoding and which combines the coding technique of [1] with that of [2], while the decoder uses backward decoding.
1) Coding Scheme: We consider $B$ blocks, each of $n$ symbols. A sequence of $B-1$ message pairs $(W_1^{(b)}, W_2^{(b)})$, for $b = 1, \ldots, B-1$, will be transmitted during $B$ transmission blocks. Here the sequence $(W_1^{(b)})$ is an i.i.d. sequence of uniform random variables over $\{1, \ldots, e^{nR_1}\}$ and independent thereof $(W_2^{(b)})$ is an i.i.d. sequence of uniform random variables over $\{1, \ldots, e^{nR_2}\}$. As $B \to \infty$, for fixed $n$, the rate pair of the message $(W_1, W_2)$, $(R_1, R_2) = (R_1(B-1)/B, R_2(B-1)/B)$, is arbitrarily close to $(R_1, R_2)$.

We assume a tuple of random variables $V \in \mathcal{V}, S \in \mathcal{S}, U \in \mathcal{U}, X_1 \in \mathcal{X}_1, X_2 \in \mathcal{X}_2, Y \in \mathcal{Y}$, of joint law $[5]$.

Random coding and partitioning: In each block $b, b = 1,2, \ldots, B$, we shall use the following code:

- Generate $e^{nR_1}$ sequences $v = (v_1, \ldots, v_n)$, each with probability $\Pr(v) = \prod_{i=1}^{n} p_i(v_i)$. Label them $v(\omega_0)$ where $\omega_0 \in \{1, \ldots, e^{nR_1}\}$.

- For each $v(\omega_0)$ generate $e^{nR_2}$ sequences $x_1 = (x_{1,1}, x_{1,2}, \ldots, x_{1,n})$, each with probability $\Pr(x_1|v(\omega_0)) = \prod_{k=1}^{n} p_{X_1}(x_{1,k}|v_k(\omega_0))$. Label them $x_1(i, \omega_0), i \in \{1, \ldots, e^{nR_2}\}$.

- For each $v(\omega_0)$ generate $e^{nR_2+R'}$ sequences $u = (u_1, \ldots, u_n)$, each with probability $\Pr(u|v(\omega_0)) = \prod_{k=1}^{n} p_{U|V}(u_k|v_k(\omega_0))$. Randomly partition the set $\{u\}$ into $e^{nR_2}$ bins, each consisting of $e^{nR'}$ codewords. Now label the codewords by $u(j, \omega_0), j \in \{1, \ldots, e^{nR_2}\}$, $j \in \{1, \ldots, e^{nR'}\}$ where $j$ identifies the bin and $j$ the index within the bin.

Encoding: We denote the realizations of the sequences $(W_1^{(b)})$ and $(W_2^{(b)})$ by $(w_1^{(b)})$ and $(w_2^{(b)})$, and the realization of the state sequence $(S_1^{(b)}, S_2^{(b)}, \ldots, S_n^{(b)})$ by $s^{(b)}$. The code builds upon a Block-Markov structure in which the message $(w_1^{(b)}, w_2^{(b)})$ is encoded over the successive blocks $b$ and $(b+1)$ such that, $\omega_0^{(b+1)} = w_1^{(b)}$, for $b = 1, \ldots, B-1$.

The messages $(w_1^{(b)})$ and $(w_2^{(b)})$, $b = 1,2, \ldots, B-1$ are encoded as follows:

In block 1 the encoders send
\[
\begin{align*}
x_1^{(1)} &= x_1(w_1^{(1)}, 1) \\
x_2^{(1)} &= x_2(s^{(1)}, w_2^{(1)}, 1).
\end{align*}
\]

Here, the encoding $x_2(s^{(b)}, w_2^{(b)}, \omega_0^{(b)})$ is defined as follows:

1) Find the typical $u(w_2^{(b)}, j_0, \omega_0^{(b)})$: Search within the bin $u(w_2^{(b)}, \cdot, \omega_0^{(b)})$ for the lowest $j_0 \in \{1, \ldots, e^{nR_2}\}$ such that $u(w_2^{(b)}, j_0, \omega_0^{(b)})$ is jointly typical with the pair $(v(\omega_0^{(b)}), s^{(b)}))$; denote this $j_0$ as $j_0(s^{(b)}, w_2^{(b)}, \omega_0^{(b)})$. If such $j_0$ is not found or if the state sequence $s^{(b)}$ is non-typical an error is declared and $j_0(s^{(b)}, w_2^{(b)}, \omega_0^{(b)}) = 1$.

2) Generate the codeword $x_2(s^{(b)}, w_2^{(b)}, \omega_0^{(b)})$ by drawing its components i.i.d. conditionally on the triple $(s^{(b)}, u(w_2^{(b)}, j_0, \omega_0^{(b)}), v(\omega_0^{(b)}))$, where the conditional law is induced by $[4]$.

Suppose that, as a result of cribbing from Encoder 1, before the beginning of block $b = 2, 3, \ldots, B$, Encoder 2 has an estimate $\hat{w}_1^{(b-1)}$ for $w_1^{(b-1)}$. Then, in block $b = 2, 3, \ldots, B-1$, the encoders send
\[
\begin{align*}
x_1^{(b)} &= x_1(w_1^{(b)}, \hat{w}_1^{(b-1)}) \\
x_2^{(b)} &= x_2(s^{(b)}, w_2^{(b)}, \hat{w}_1^{(b-1)}),
\end{align*}
\]
and in block $B$
\[
\begin{align*}
x_1^{(B)} &= x_1(1, \hat{w}_1^{(B-1)}) \\
x_2^{(B)} &= x_2(s^{(B)}, 1, \hat{w}_1^{(B-1)}).
\end{align*}
\]

Decoding at the receiver: After the reception of block-$B$ the receiver uses backward decoding starting from block $B$ downward to block 1 and decodes the messages as follows. In block $B$ the receiver looks for $\hat{w}_1^{(B-1)}$ such that
\[
\left(v(\hat{w}_1^{(B-1)}), x_1(1, \hat{w}_1^{(B-1)}), u(1, j_0, \hat{w}_1^{(B-1)}), \right) \\
x_2(s^{(B)}, 1, \hat{w}_1^{(B-1)}, 1) \in A(v, X_1, U, X_2, Y),
\]
where $j_0 = j_0(s^{(B)}, 1, \hat{w}_1^{(B-1)})$.

Next, assume that, decoding backwards up to (and including) block $b = B-1$, the receiver decoded $w_1^{(B-1)}$, $w_2^{(B-2)}$, $\ldots$, $w_2^{(b+1)}$, $\hat{w}_1^{(b+1)}$. To decode block $b$, the receiver looks for $(w_2^{(b)}, \hat{w}_1^{(b+1)})$ such that
\[
\left(v(\hat{w}_1^{(b+1)}), x_1(\hat{w}_1^{(b+1)}, 1), u(\hat{w}_2^{(b)}, j_0, w_1^{(b)}), \right) \\
x_2(s^{(b)}, w_2^{(b)}, w_1^{(b)}, 1) \in A(v, X_1, U, X_2, Y),
\]
where $j_0 = j_0(s^{(b)}, \hat{w}_2^{(b)}, \hat{w}_1^{(b+1)})$.

Decoding at Encoder 2: To obtain cooperation, after block $b = 2, 1, \ldots, B-1$, Encoder 2 chooses $\hat{w}_1^{(b)}$ such that
\[
\left(v(\omega_0^{(b)}), x_1(\hat{w}_1^{(b)}, \omega_0^{(b)}), \right) \in A(v, X_1, X_1),
\]
where $\omega_0^{(b)} = \hat{w}_1^{(b-1)}$ was determined at the end of block $b-1$ and $\hat{w}_0^{(1)} = 1$.

When a decoding step either fails to recover a unique index (or index pair) which satisfies the decoding rule, or there is more than one index (or index pair), then an index (or an index pair) is chosen at random.

2) Bounding the Probability of Error: Genie-aided arguments as in [3] and [4] can be used to show that the probability that either Endor2 makes an encoding error or the receiver makes a decoding error after block $b$ in the above scheme is upper bounded by the probability that at least one of the following events $E_0^{(b)} - E_0^{(b)}$ happens.

Error events:

- $E_0^{(b)}$:
\[
\left(v(\omega_0^{(b)}), u(w_2^{(b)}, j_0, \omega_0^{(b)}), x_1(w_1^{(b)}, \omega_0^{(b)})) \not\in A(v, U, X_1).
\]
• $E_1^{(b)}$: There exists $\tilde{w}_1 \neq w_1^{(b)}$ such that
\[
\left( v(\omega_0^{(b)}), x_1(\tilde{w}_1, \omega_0^{(b)}), x_1^{(b)} \right) \in \mathcal{A}_r(V, X_1, X_1).
\]

• $E_2^{(b)}$: There doesn’t exist $j_0 \in \{1, \ldots, e^{nR'}\}$ such that
\[
\left( v(\omega_0^{(b)}), u(w_2^{(b)}, j_0, \omega_0^{(b)}), s^{(b)} \right) \in \mathcal{A}_r(V, U, S).
\]

• $E_3^{(b)}$:
\[
\left( v(\omega_0^{(b)}), u(w_2^{(b)}, j_0, \omega_0^{(b)}), x_1^{(b)}, \omega_0^{(b)} \right),
\]
\[
x_2(s^{(b)}, w_3^{(b)}, w_4^{(b)}), y^{(b)} \right)
\]
\[
\not\in \mathcal{A}_r(V, U, X_1, X_2, Y).
\]

• $E_4^{(b)}$: There exists $\tilde{\omega}_0 \neq \omega_0^{(b)}$ such that
\[
\left( v(\omega_0^{(b)}), x_1(\omega_1^{(b)}, \omega_0^{(b)}), u(j, j_0, \tilde{\omega}_0) \right),
\]
\[
x_2(s^{(b)}, j, \tilde{\omega}_0^{(b)}), y^{(b)} \right)
\]
\[
\in \mathcal{A}_r(V, U, X_1, X_2, Y),
\]
for some pair $(j, j_0), j \in \mathcal{W}_2, j_0 \in \{1, \ldots, e^{nR'}\}$.

• $E_5^{(b)}$: There exists $\tilde{w}_2 \neq w_2^{(b)}$ such that
\[
\left( v(\omega_0^{(b)}), x_1^{(b)}, \omega_0^{(b)} \right),
\]
\[
x_2(\tilde{s}(b), \tilde{w}_2, \omega_0^{(b)}), y^{(b)} \right)
\]
\[
\in \mathcal{A}_r(V, U, X_1, X_2, Y),
\]
for some index $j_0 \in \{1, \ldots, e^{nR'}\}$.

We define the event
\[
F_1^{(b)} \triangleq \bigcup_{j=1}^{5} E_j^{(b)}, \quad b = 1, \ldots, B,
\]
the event
\[
F_2 \triangleq \bigcup_{j=1}^{B} E_0^{(b)},
\]
the event
\[
F_3 \triangleq \bigcup_{j=1}^{B} \left( E_0^{(b)} \cup E_1^{(b)} \right),
\]
the event
\[
F_4 \triangleq \bigcup_{j=1}^{B} \left( E_0^{(b)} \cup E_1^{(b)} \cup E_2^{(b)} \right),
\]
and the event
\[
F_4 \triangleq \bigcup_{j=1}^{B} \left( E_0^{(b)} \cup E_1^{(b)} \cup E_2^{(b)} \cup E_3^{(b)} \right).
\]

We can upper bound the average probability of error $\bar{P}_e$ averaged over all codebooks and all random partitions by
\[
\bar{P}_e \leq \sum_{b=1}^{B} \left\{ \Pr \left[ E_0^{(b)} \right] + \Pr \left[ E_1^{(b)} | F_2^{(c)} , E_1^{(1 \ldots b-1)^c} \right] \right\}
\]
\[
+ \sum_{b=1}^{B} \left\{ \Pr \left[ E_2^{(b)} | F_3 \right] + \Pr \left[ E_3^{(b)} | F_4^{(c)} , E_3^{(1 \ldots b-1)^c} \right] \right\}
\]
\[
+ \sum_{b=1}^{B} \Pr \left[ F_1^{(b)} | F_4^{(c)} , F_1^{(b+1 \ldots B)^c} \right],
\]
where $F_1^{(1 \ldots b-1)^c}$ denotes the complement of the event $F_1^{(1 \ldots b-1)}$.

Furthermore, we can upper bound each of the summands in the last component as
\[
\Pr \left( F_1^{(b)} | F_4^{(c)} , F_1^{(b+1 \ldots B)^c} \right)
\]
\[
= \Pr \left( \bigcup_{j=4}^{B} E_{j+1}^{(b)} | F_4^{(c)} , F_1^{(b+1 \ldots B)^c} \right)
\]
\[
\leq \Pr \left( E_{4}^{(b)} | F_4^{(c)} , F_1^{(b+1 \ldots B)^c} \right)
\]
\[
+ \Pr \left( F_5^{(b)} | F_4^{(c)} , F_1^{(b+1 \ldots B)^c} \right).
\]

In the following we separately examine each of the above summands.

By Lemma [1] (Appendix A) $\Pr \left[ E_3^{(b)} | F_4^{(c)} , E_3^{(1 \ldots b-1)^c} \right]$ and $\Pr \left[ E_0^{(b)} \right]$ can be made arbitrarily small for sufficiently large $n$. Also, by Lemma [2]
• If
\[
R_1 < H(X_1 | V),
\]
then $\Pr \left[ E_1^{(b)} | F_2^{(c)} , E_1^{(1 \ldots b-1)^c} \right]$ can be made arbitrarily small, provided that $n$ is sufficiently large;
• If
\[
R_1 + R_2 + R' < I(V U X_1; Y),
\]
then $\Pr \left( E_4^{(b)} | F_4^{(c)} , F_1^{(b+1 \ldots B)^c} \right)$ can be made arbitrarily small, provided that $n$ is sufficiently large;
• If
\[
R_2 + R' < I(U; Y | V X_1)
\]
then $\Pr \left( E_5^{(b)} | F_4^{(c)} , F_1^{(b+1 \ldots B)^c} \right)$ can be made arbitrarily small, provided that $n$ is sufficiently large;

Finally, by the covering lemma (See [5], [6], [7] or [9] Chapter 13), if
\[
R' > I(U; S | V)
\]
then $\Pr \left[ E_2^{(b)} | F_3^{(c)} \right]$ can be made arbitrarily small, provided that $n$ is sufficiently large.

The combination of (7), (8), (9), and (10) establishes the achievability of the rate region [4] for a law of the form (5).
B. Proof of the converse in Theorem 2

Consider an \((e^{R_1}, e^{R_2}, n)\) code with average block error probability \(P_e(n)\), and a law on \(W_1 \times W_2 \times X_1^n \times X_2^n \times Y^n \times S^n\) given by

\[
P(w_1, w_2, x_1^n, x_2^n, y^n, s^n) = p_{w_1, w_2} p_{x_1} \prod_{k=1}^n p_{s_k} x_{1,k} x_{2,k} s_k.
\]

(11)

Let \(V_k\) be the random variable defined by

\[
V_k \triangleq X_k^{k-1},
\]

(12)

and let \(U_k\) be the random variable defined by

\[
U_k \triangleq W_2 Y^{k-1} S_{k+1}^n.
\]

(13)

We start with an upper bound on \(R_1\) by following similar steps as in [2] Section V—Converse for situation 2.

\[
n R_1 = H(W_1|W_2) = I(W_1; Y^n|W_2) + H(W_1|W_2 Y^n) \leq I(W_1; Y^n|W_2) + n \delta(P_e) = I(X_1^n; Y^n|W_2) + n \delta(P_e) \leq \sum_{k=1}^n H(X_1|X_1^{k-1}) + n \delta(P_e) = \sum_{k=1}^n H(X_1|V_k) + n \delta(P_e).
\]

(14)

where (a) follows from the encoding relation (1).

Next, consider \(R_2\)

\[
n R_2 = H(W_2|W_1) \leq I(W_2; Y^n|W_1) + n \delta(P_e) = \sum_{k=1}^n I(W_2; Y_k|W_1 Y^{k-1}) + n \delta(P_e) \leq \sum_{k=1}^n I(W_2 Y^{k-1}; Y_k|W_1) + n \delta(P_e) = \sum_{k=1}^n [I(W_2 Y^{k-1} S_{k+1}^n; Y_k|W_1) - I(S_{k+1}^n; Y_k|W_1 W_2 Y^{k-1})] + n \delta(P_e)
\]

(15)

(b) follows by the Csiszár-Körner’s identity [5, Lemma 7];
(c) follows since \((W_2 S_{k+1}^n)\) is independent of \(S_k\);
(d) follows by the encoding relation (1);
(e) follows since \(W_1 \mapsto X_1 X_1^{k-1} \mapsto W_2 Y_k Y^{k-1} S_{k+1}^n\) and \(W_1 \mapsto X_1 X_1^{k-1} Y_k\) are Markov strings; and
(f) follows since \(W_1 \mapsto X_1^{k-1} \mapsto W_2 Y_k Y^{k-1} S_{k+1}^n\) is a Markov string.

Finally, we consider the sum-rate \(R_1 + R_2\)

\[
n(R_1 + R_2) = H(W_1 W_2) \leq I(W_1 W_2; Y^n) + n \delta(P_e) = \sum_{k=1}^n I(W_1 W_2; Y_k|Y^{k-1}) + n \delta(P_e)
\]

(16)
Here, (g) follows by the same procedure as (b) and (c); (h) follows by the encoding relation (11) and since $W_1 \oplus X_{1,k}X_1' \oplus W_2Y_{k+1} \oplus Y_k$ is a Markov string; and (i) follows since $W_1$ is independent of $S_k$ and since $W_1 \oplus X_1'Y_{k+1} \oplus S_k$ and $W_1 \oplus X_{1,k}' \oplus S_k$ are Markov strings.

Next we verify the joint law of the auxiliary random variables. By (11) and the encoding rule (2) we may write

$$p_{W_1W_2X_1^k-1X_1,k} = p_{W_1X_1^k-1|W_1}p_{X_1|W_1}p_{S_{k+1}|P_{S_k}P_{S_{k+1}}}$$

Summing this joint law over $w_1$ we obtain

$$\sum_{w_1} p_{W_1W_2X_1^k-1X_1,k} = p_{W_2X_1^k-1X_1,k}p_{S_{k+1}|P_{S_k}X_1^k}$$

Finally, we may express (16) as follows

$$R_1 + R_2 \leq \frac{1}{n} \sum_{k=1}^{n} \left[ I(V_kU_kX_1,k;Y_k) - I(U_k;S_k|V_k) \right]$$

$$= I(V,U,X_1,Y) - I(U;S|V,J)$$

$$\leq I(V,J,U,X_1;Y) - I(J;Y) - I(U;S|V,J)$$

This establishes the single letter expression for the achievable rate region (4). The convexity of the rate region (4) can be shown in a similar way.

The inequalities (14), (15), (16) combined with their respective single-letter expressions and the Markov relation (17) establish the converse part of Theorem 1.

C. Bounds on alphabets sizes in Theorem 7

We consider the alphabet sizes of $U$ and $V$. Specifically, let $P_{X_1,x_2,S,V,U}$ be a distribution satisfying the Markov conditions required in (5). For convenience, $P_{X_1,x_2,S,V,U|V(x_1,x_2,s,u|v)}$ will be denoted in the sequel as $P(\cdot|v)$. We would like to bound the sizes of the alphabets $V$ and $U$, while preserving the region given in (4). For a generic distribution $Q$ on $X_1 \times X_2 \times S \times U$, define the functionals

$$q_{x_1,x_2,s}(Q) = \sum_u Q(x_1,x_2,s,u), \quad x_1,x_2,s \in X_1 \times X_2 \times S$$

$$J_1(Q) = \sum_{x_1,x_2,s,u} Q(x_1,x_2,s,u) \log \frac{1}{\sum_{x_1',x_2',s',u'} Q(x_1',x_2',s',u')}$$

$$J_2(Q) = \sum_{x_1,x_2,s,u} Q(x_1,x_2,s,u) \log \frac{1}{\sum_{x_1',x_2',s',u'} Q(x_1',x_2',s',u')}$$

$$J_3(Q) = \sum_{x_1,x_2,s,u} Q(x_1,x_2,s,u) \log \frac{1}{\sum_{x_1',x_2',s'} Q(x_1',x_2',s',u')}$$

$$J_4(Q) = \sum_{x_1,x_2,s,u} Q(x_1,x_2,s,u) \log \frac{1}{\sum_{x_1',x_2',s',u'} Q(x_1',x_2',s',u')}$$

We may express (14) as follows

$$R_1 \leq \frac{1}{n} \sum_{k=1}^{n} H(X_1,k|V_k) = H(X_1|V),$$

where in the last step we’ve defined $\bar{V} \triangleq (V,J)$. Similarly, we may express (15) as follows

$$R_2 \leq \frac{1}{n} \sum_{k=1}^{n} [I(U_k;Y_k|V_kX_1,k) - I(U_k;S_k|V_k)]$$

$$= I(U;Y|V,X_1,J) - I(U;S|V,J)$$

$$= I(U;Y|\bar{V},X_1) - I(U;S|\bar{V}),$$

(19)
Substituting the distribution \( P_{X_1, X_2, S, U} | V (\cdot | v) \) in the functionals, and averaging them with respect to \( v \), we obtain
\[
\sum_u P_V (v) q_{x_1, x_2, s, v} (P (\cdot | v)) = P_{X_1, X_2, s} (x_1, x_2, s)
\] (22a)
\[
\sum_u P_V (v) J_1 (P (\cdot | v)) = H (X_1 | V)
\] (22b)
\[
\sum_u P_V (v) J_2 (P (\cdot | v)) = H (S | V)
\] (22c)
\[
\sum_u P_V (v) J_3 (P (\cdot | v)) = H (S | U, V)
\] (22d)
\[
\sum_u P_V (v) J_4 (P (\cdot | v)) = H (Y | V, X_1)
\] (22e)
\[
\sum_u P_V (v) J_5 (P (\cdot | v)) = H (Y | V, U, X_1).
\] (22f)

Observe that preserving the values of the right hand sides of (22a)-(22f), guarantees that we also preserve the region (4). We used here the Markov structure \( Y \rightarrow (X_1, X_2, S) \rightarrow (V, U) \), and the fact that if we preserve the joint distribution of \( X_1, X_2, S \), the distribution of \( Y \) is also preserved. By the Support Lemma we can restrict the alphabet of \( V \) to:
\[
|V| \leq |X_1| |X_2| |S| \geq 5.
\] (23)

Note that this bound is independent of the alphabet of \( U \).

We now fix some \( V \) with bounded alphabet as above, and proceed to bound the alphabet of \( U \). Let \( \tilde{Q} \) be a generic distribution on \( X_1 \times X_2 \times S \times V \), and define the functionals
\[
\tilde{q}_{x_1, x_2, s, v} (\tilde{Q}) = \tilde{Q} (x_1, x_2, s, v)
\] (24a)
\[
\tilde{J}_1 (\tilde{Q}) = \sum_{x_1, x_2, s, v} \tilde{Q} (x_1, x_2, s, v)
\cdot \log \sum_{x_1', x_2', s', v} \tilde{Q} (x_1', x_2', s', v)
\sum_{x_1, x_2, s, v} \tilde{Q} (x_1, x_2, s, v)
\] (24b)
\[
\tilde{J}_2 (\tilde{Q}) = \sum_{x_1, x_2, s, v} \tilde{Q} (x_1, x_2, s, v)
\cdot \log \frac{1}{\sum_{x_1', s'} \tilde{Q} (x_1', x_2', s', v) P_Y | X_1, X_2, S (y | x_1, x_2', s')}
\] (24c)

Since
\[
P_{Y | V, X_1} (y | v, x_1) = \sum_{x_2, s} P_{Y | X_1, X_2, S} (y | x_1, x_2, s)
\cdot \frac{P_{X_1, X_2, S, V} (x_1, x_2, s, v)}{\sum_{x_2', s'} P_{X_1, X_2, S, V} (x_1, x_2', s', v)},
\]
in order to preserve the value of \( H (Y | V, X_1) \), it suffices to preserve the joint distribution of \( X_1, X_2, S, V \).

For convenience, we use in the sequel the shorthand notation \( P (\cdot | u) = P_{X_1, X_2, S, V} | U (\cdot | u) \). Substituting the distribution \( P_{X_1, X_2, S, V} | U (\cdot | u) \) in the functionals (24) and averaging over \( u \), we obtain
\[
\sum_u P_U (u) q_{x_1, x_2, s, v} (P (\cdot | u)) = P_{X_1, X_2, S, V} (x_1, x_2, s, v)
\] (25a)
\[
\sum_u P_U (u) J_1 (P (\cdot | u)) = H (S | V, U)
\] (25b)
\[
\sum_u P_U (u) J_2 (P (\cdot | u)) = H (Y | V, U, X_1)
\] (25c)

Applying again the Support Lemma, we see that it suffices to bound the alphabet size of \( U \) as
\[
|U| \leq |X_1| |X_2| |S| |V| + 2.
\] (26)

This completes the proof of the bounds on the alphabet sizes.

D. Proof of Theorem 2

The achievability part follows similarly to that of Theorem 1, the only difference being in the way the code word \( x_2 (s (b), w_2 (b), \omega (b)) \) is generated. Here the second encoder generates the code word \( x_2 (s (b), w_2 (b), \omega (b)) \) by drawing its elements i.i.d. conditionally on the quadruple \( (s (b), u (w_2 (b), j_0 (w_2 (b), \omega (b)), v (\omega (b), x_1 (b)), (22e)

For the converse, consider an \( (e^{nR_1}, e^{nR_2}, n) \) code with average block error probability \( P_e (n) \), and a law on \( W_1 \times W_2 \times X_1^n \times X_2^n \times S^n \) given by
\[
p_{W_1} p_{W_2} I (x_1^n = f_1 (w_1)) p_{X_2^n | W_1} p_{W_2} p_{Y^n | X_1^n, X_2^n, S^n} \prod_{k=1}^{n} p_{Y_k | X_{1,k} X_{2,k} S_k}.
\] (27)

The Fano inequalities for the causal cribbing case yield the same inequalities (14), (15), and (16).

It remains to verify the joint law of the auxiliary random variables.

By (27) and the encoding rule (3) we may write
\[
p_{W_1} p_{W_2} X_1^{k-1} X_1 k S^{k-1} S_k S_{k+1} X_2^k Y^{k-1} =
\]
\[
p_{W_1} p_{X_1^{k-1} | W_1} p_{X_1 k} p_{W_1} p_{X_2^{k-1} | W_2} p_{S_k} p_{S_{k+1}} p_{W_2} P_{X_2^{k-1} | W_2} p_{X_2^{k-1} | X_1 k} p_{X_2^{k-1} S^n} p_{X_1^{k-1} X_2^{k-1} S_k}
\]
\[
p_{Y^{k-1} | X_1^{k-1} X_2^{k-1} S_k}.
\]

Summing this joint law over \( w_1 \) we obtain
\[
\sum_{w_1} p_{W_1} p_{W_2} X_1^{k-1} X_1 k S^{k-1} S_k S_{k+1} X_2^k Y^{k-1} =
\]
\[
p_{W_2} P_{X_1^{k-1} | X_1 k} p_{W_2} p_{X_2^{k-1} | W_2} p_{S_k} p_{S_{k+1}} p_{W_2} P_{X_2^{k-1} | W_2} p_{X_2^{k-1} | X_1 k} p_{X_2^{k-1} S^n} p_{X_1^{k-1} X_2^{k-1} S_k}
\]
\[
p_{Y^{k-1} | X_1^{k-1} X_2^{k-1} S_k}.
\]
Summing this joint law over all possible sub-sequences $(s_1, s_2, \ldots, s_{k-1})$ we obtain
\[
\sum_{(s_1, s_2, \ldots, s_{k-1})} P_{W_2X_1}^{k-1} X_1, X_k S_k \equiv S_{k+1} X_2 \equiv Y^{k-1} = P_{X_1}^{k-1} P_{X_1} S_{k+1} X_2 \equiv Y^{k-1}
\]
This establishes the Markov relation
\[
U_k \not\equiv S_k V_k \not\equiv X_{1,k},
\]
as well as the fact that conditionally on $V_k U_k S_{1,k}$ the r.v. $X_{2,k}$ is independent of the rest.

**APPENDIX**

**A. Strong Typicality**

Let $\{X^{(1)}, X^{(2)}, \ldots, X^{(k)}\}$ denote a finite collection of discrete random variables with some joint distribution $P(X^{(1)}, X^{(2)}, \ldots, X^{(k)})$ with \(\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \cdots \times \mathcal{X}^{(k)}\). Let $S$ be the number of indices $(1, 2, \ldots, k)$ with $s_j = s$. By the law of large numbers, for any subset $S$ of these random variables and consider $n$ independent copies of $S$. Thus, with $S \triangleq (S_1, S_2, \ldots, S_n)$,
\[
\Pr\{S = s\} = \prod_{j=1}^n \Pr\{S_j = s_j\}. 
\]

Let $N(s; s)$ be the number of indices $j \in \{1, 2, \ldots, n\}$ such that $S_j = s$. By the law of large numbers, for any subset $S$ of random variables and for all $s \in S$,
\[
\frac{1}{n} N(s; s) \to P(s), \tag{29}
\]
as well as
\[
-\frac{1}{n} \ln P(s_1, s_2, \ldots, s_n) = -\frac{1}{n} \sum_{j=1}^n \ln P(s_j) \to H(S). \tag{30}
\]

The convergence in (29) and (30) takes place simultaneously with probability one for all nonempty subsets $S$ [9].

**Definition 1:** The set $\mathcal{A}_e$ of $\epsilon$-strongly typical $n$-sequences is defined by (see [9, Chapter 3.12,13])
\[
\mathcal{A}_e \triangleq \mathcal{A}_e \left( X^{(1)}, X^{(2)}, \ldots, X^{(k)} \right) \triangleq \left\{ (x^{(1)}, x^{(2)}, \ldots, x^{(k)}) : \right. \\
\frac{1}{n} N \left( (x^{(1)}, x^{(2)}, \ldots, x^{(k)}); x^{(1)}, x^{(2)}, \ldots, x^{(k)} \right) \\
- P \left( x^{(1)}, x^{(2)}, \ldots, x^{(k)} \right) \epsilon \left\langle \| \mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \cdots \times \mathcal{X}^{(k)} \| \right. \\
\forall \left( x^{(1)}, x^{(2)}, \ldots, x^{(k)} \right) \in \mathcal{X}^{(1)} \times \cdots \times \mathcal{X}^{(k)}, \right\}
\]
where $|X|$ is the cardinality of the set $X$.

Let $\mathcal{A}_e(S)$ be defined similar to $\mathcal{A}_e$ but now with constraints corresponding to all nonempty subsets $S$. We recall now two basic lemmas (for the proofs we refer to [9]).

**Lemma 1:** For any $\epsilon > 0$ the following statements hold for every integer $n \geq 1$:

1. If $s \in \mathcal{A}_e(S)$, then $\exp(-n(H(S) + \epsilon)) \leq \Pr\{S = s\} \leq \exp(-n(H(S) - \epsilon))$.
2. If $S_1, S_2 \subseteq \{X_1, X_2, \ldots, X_k\}$ and $(s_1, s_2) \in \mathcal{A}_e(S_1 \cup S_2)$, then
\[
\exp(-n(H(S_1 | S_2) + 2\epsilon)) \leq \Pr\{S_1 = s_1 | S_2 = s_2\} \leq \exp(-n(H(S_1 | S_2) - 2\epsilon)).
\]

Moreover, the following statements hold for every sufficiently large $n$:
3. $\Pr\{\mathcal{A}_e(S)\} \geq 1 - \epsilon$,
4. $(1 - \epsilon)\exp(n(H(S) - \epsilon)) \leq \mathcal{A}_e(S) \leq \exp(n(H(S) + \epsilon))$.

**Lemma 2:** Let the discrete random variables $X, Y, Z$ have joint distribution $P_{X,Y,Z}(x, y, z)$. Let $X'$ and $Y'$ be conditionally independent given $Z$, with the marginal laws
\[
P_{X' | Z}(x | z) = \sum_y P_{X,Y,Z}(x, y, z) / P_Z(z),
\]
\[
P_{Y' | Z}(y | z) = \sum_x P_{X,Y,Z}(x, y, z) / P_Z(z).
\]

Let $(X, Y, Z) \sim \prod_{x=1}^{k} P_{X,Y,Z}(x_k, y_k, z_k)$ and $(X', Y', Z) \sim \prod_{x=1}^{k} P_{X' | Z}(x_k | z_k) P_{Y' | Z}(y_k | z_k) P_Z(z_k)$. Then
\[
\Pr\{X', Y', Z \in \mathcal{A}_e(X, Y, Z)\} \leq \exp(-n[I(X; Y | Z) - \epsilon]).
\]

**REFERENCES**

[1] A. Somekh-Baruch, S. Shamai (Shitz) and S. Verdú, “Cooperative multiple-access encoding with states available at one transmitter,” *IEEE Trans. Inform. Theory*, vol. IT-54, no. 10, pp. 4448-4469, Oct. 2008.

[2] F.M.J. Willems and E.C. van der Meulen, “The discrete memoryless multiple-access channel with cribbing encoders”, *IEEE Trans. Inform. Theory*, vol. IT-31, no. 3, pp. 313-327, May 1985.

[3] B. Rimoldi and R. Urbanke, “A rate-splitting approach to the Gaussian multiple-access channel”, *IEEE Trans. Inform. Theory*, vol. IT-42, no. 2, pp. 364 – 375, Mar 1996.

[4] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, John Wiley & Sons, 1965.

[5] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” *IEEE Trans. Inform. Theory*, vol. IT-24, No. 3 pp. 339-348, May 1978.

[6] R. Ahlswede and J. Körner, “Source coding with side information and a converse for degraded broadcast channels,” *IEEE Trans. Inform. Theory*, vol. IT-21, No. 6, pp. 629-637, Nov. 1975.

[7] A. D. Wyner, “On source coding with side information at the decoder,” *IEEE Trans. Inform. Theory*, vol. IT-21, No. 6, pp. 294-300, May 1975.

[8] T. Berger, “Multiterminal source coding,” Lecture notes presented at the 1977 CISM Summer School, Udine, Italy, July 18-20, 1977, Springer-Verlag.

[9] T. M. Cover and J. A. Thomas, “Elements of Information Theory,” Wiley, 1991.

[10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.