The Benefit of Encoder Cooperation
in the Presence of State Information

Parham Noorzad, Student Member, IEEE, Michelle Effros, Fellow, IEEE,
and Michael Langberg, Senior Member, IEEE

Abstract

In many communication networks, the availability of channel state information at various nodes provides an opportunity for network nodes to work together, or “cooperate.” This work studies the benefit of cooperation in the multiple access channel with a cooperation facilitator, distributed state information at the encoders, and full state information available at the decoder. Under various causality constraints, sufficient conditions are obtained such that encoder cooperation through the facilitator results in a gain in sum-capacity that has infinite slope in the information rate shared with the encoders. This result extends the prior work of the authors on cooperation in networks where none of the nodes have access to state information.

I. INTRODUCTION

In cooperative strategies, various network nodes work together towards a common goal. Previous work [1] shows that under a model of cooperation that incorporates a “cooperation facilitator” (CF)—a node that receives rate-limited information from the encoders of a multiple access channel (MAC) and sends rate-limited information back—even a very low rate cooperation between the MAC encoders can vastly increase the total rate that can be delivered through the MAC. In fact, if we measure cost as the number of bits the CF shares with the encoders and the benefit as the gain in sum-capacity, then for some MACs, the cost-benefit curve has infinite slope in the limit of low cost. This paper extends the exploration of cooperation beyond the networks of [1] to examine the cost-benefit tradeoff of cooperation in networks where state information is present at some nodes.

Networks where state information is available at some nodes appear in many applications, including wireless channels with fading [2], [3], cognitive radios [4], and computer memory with defects [5]. Depending on the application at hand, channel state information may be either fully available at all network nodes or available in a distributed manner; in the latter case, each node has access to a component or a function of the state sequence. Furthermore, the state information may be available non-causally, or alternatively, may be subject to causality constraints. For example, when state information models fading effects in wireless communication [2], the transmitters’ knowledge of state information is strictly causal or causal. On the other hand, when the state sequence...
models a signal that the transmitter sends to another receiver, then the state sequence is available non-causally at the transmitter [6].

In this work, we study the advantage of encoder cooperation in the setting of networks with state information. In this context, network nodes work together to increase transmission rates—not only by sharing message information, but also by sharing state information. (See Figure [1]) As an example of message and state cooperation, Permuter, Shamai, and Somekh-Baruch [7] find the capacity region of the MAC with encoder cooperation under the assumption that distributed, non-causal state information is available at the encoders and full state information is available at the decoder. As their cooperation model, the authors use a special case of the Willems conferencing model [8], originally defined for MACs in the absence of state information.

Indirect forms of cooperation, in the presence of state information, are also considered in the literature. Cemal and Steinberg [9] study a model where a central state-encoder sends rate limited versions of non-causal state information to each encoder, while the decoder has access to full state information.

Here we study cooperation under the CF model. In this model, encoders cooperate indirectly as in [1], rather than directly, as in [8]. The CF enables both message and state cooperation; this proves crucial to the cooperation gain obtained through a CF, which we next describe in more detail.

In earlier work [1], we exhibit single-letter conditions on the channel transition matrix of the MAC which guarantee an infinite slope in sum-capacity as a function of the capacities of the CF output edges; the additive Gaussian MAC [10, p. 544] is an important example of a scenario where the infinite slope phenomenon occurs. In this work, we characterize channels for which the cooperation gain has an infinite slope in the presence of state information (Section IV); interestingly, this includes channels for which the infinite slope phenomenon did not arise in the absence of state information.

For state information at the encoders we consider four cases: (i) no state information, (ii) strictly causal state information, (iii) causal state information, and (iv) non-causal state information. In case (i), the CF is used for sharing message information (a strategy here called “message cooperation”) since no state information is available at the encoders. In cases (ii)-(iv), the CF enables both message and state cooperation. However, here we study message and state cooperation only in case (iv); in this case we show that the use of joint message and state cooperation leads to a weaker sufficient condition for an infinite-slope gain compared to the sole use of message cooperation. Whether in cases (ii) and (iii), the use of joint message and state cooperation likewise leads to a weaker sufficient condition for an infinite-slope gain compared to message cooperation alone, remains an open problem.

Throughout, we assume that any state information available at the encoders is distributed; that is, we assume $S = (S_1, S_2)$, where for $i \in \{1, 2\}$, $S_i$ is available at encoder $i$. As we do not make any assumptions regarding the dependence between $S_1$ and $S_2$, our results apply to the limiting cases of independent states (i.e., independent $S_1$ and $S_2$) and common state (i.e., $S_1 = S_2$).

1As an example, consider the MAC $Y = X_1 + X_2 + S \pmod{3}$, where $S$ is uniform on $\{0, 1, 2\}$, $X_1$ and $X_2$ are binary, and $Y$ is ternary. The infinite slope sum-capacity gain is achievable when the decoder has full knowledge of $S$, but no sum-capacity gain is possible when it does not have access to $S$. 


Figure 1. The network studied in this work consists of a pair of encoders communicating, with the help of a CF, to a decoder through a state-dependent MAC. Full state information is available at the decoder. At time $t \in [n]$, partial state information $\hat{S}_i^t$ is available to encoder $i \in \{1, 2\}$.

Since the decoder starts the decoding process only after receiving all the output symbols in a given transmission block, causality constraints at the decoder do not impose limitations on the availability of state information. Thus we may assume that the decoder either has full state information or no state information. Here we focus on the former scenario. Jafar [11] provides the capacity region of the MAC with distributed independent (causal or non-causal) state information at the encoders and full state information at the decoder. The capacity region is unknown when the encoders have access to state information but the decoder does not [12], [13].

II. MODEL

A. Preliminaries

Let $S_1$, $S_2$, $X_1$, $X_2$, and $Y$ be discrete or continuous alphabets. A MAC with input alphabet $X_1 \times X_2$, output alphabet $Y$, and state alphabet $S := S_1 \times S_2$ is given by the sequence

$$\left\{ p(s^n)p(y^n|s^n, x_1^n, x_2^n) \right\}_{n=1}^{\infty}.$$  

The MAC is said to be memoryless and stationary if for some $p(s)p(y|s, x_1, x_2)$ and all positive integers $n$,

$$p(s^n)p(y^n|s^n, x_1^n, x_2^n) = \prod_{t=1}^{n} p(s_t)p(y_t|s_t, x_{1t}, x_{2t}).$$  

B. Message Cooperation

In this subsection, we define the capacity region of a MAC with a CF that enables message cooperation. We include four scenarios in our definition based on the availability of state information at the encoders: no state, strictly causal, causal, and non-causal. We assume full state information is available at the decoder. In our definition below, for any real number $x \geq 1$, $[x]$ denotes the set $\{1, \ldots, [x]\}$.

We start by defining a $(2^nR_1, 2^nR_2, n)$-code for the MAC with a $(C_{in}, C_{out})$-CF, cost functions $b_i : X_i \to \mathbb{R}_{\geq 0}$ for $i \in \{1, 2\}$, and cost constraints $B_1, B_2 \geq 0$. The pairs $C_{in} = (C_{in}^1, C_{in}^2)$ and $C_{out} = (C_{out}^1, C_{out}^2)$ denote the
CF input and output edge capacities, respectively. Encoder $i$, for $i \in \{1, 2\}$, is represented by $(\varphi^i_{in}, (f_{it})_{t=1}^n)$, the CF is represented by $(\varphi^1_{out}, \varphi^2_{out})$, and the decoder is represented by $g$. These mappings are defined in the order of their use below. For $i \in \{1, 2\}$, the transmission from encoder $i$ to the CF is represented by the mapping 

$$
\varphi^i_{in} : [2^{nR_i}] \rightarrow [2^{nC^i_{in}}]
$$

(1)

and the transmission from the CF to encoder $i$ is represented by 

$$
\varphi^i_{out} : [2^{nC^i_{in}}] \times [2^{nC^i_{out}}] \rightarrow [2^{nC^i_{out}}].
$$

For simplicity, the transmissions to and from the CF occur prior to the transmission of codewords over the channel.

At time $t \in [n]$, for $i \in \{1, 2\}$, the transmission of encoder $i$ over the channel is represented by the mapping 

$$
f_{it} : [2^{nR_i}] \times [2^{nC^i_{out}}] \times \hat{S}^i_t \rightarrow X_i.
$$

(2)

Here $\hat{S}^i_t$ represents any knowledge about the state gathered by encoder $i$ in times $\{1, \ldots, t\}$. Let $*$ be a symbol not in $S_1 \cup S_2$. For $i \in \{1, 2\}$ and $t \in [n]$, we have

$$
\hat{S}^i_t := \begin{cases} 
\ast & \text{no state information} \\
S^i_{it} & \text{causal} \\
S^i_{i(t-1)} & \text{strictly causal} \\
S^i_t & \text{non-causal}.
\end{cases}
$$

For every message pair $(w_1, w_2)$, the codeword of encoder $i$ is required to satisfy the cost constraint 

$$
\sum_{t=1}^n \mathbb{E}b_i \left[ f_{it}(w_i, \varphi^i_{out}(\varphi^1_{in}(w_1), \varphi^2_{in}(w_2)), \hat{S}^i_t) \right] \leq B_i.
$$

(3)

The decoder has full state information and is represented by the mapping 

$$
g : S^n \times Y^n \rightarrow [2^{nR_1}] \times [2^{nR_2}].
$$

The average probability of error is given by

$$
P_e^{(n)} = \Pr \left\{ g(S^n, Y^n) \neq (W_1, W_2) \right\},
$$

where $(W_1, W_2)$ is uniformly distributed over $[2^{nR_1}] \times [2^{nR_2}]$. A rate pair $(R_1, R_2)$ is achievable if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$-codes with $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. We use subscript $\tau \in \{0, T - 1, T, \infty\}$ to specify the dependence of the capacity region and sum-capacity on the availability of state information at the encoders. The following table makes this dependence clear.

| $\tau$ | encoder state information |
|--------|--------------------------|
| 0      | none                     |
| $T - 1$ | strictly causal          |
| $T$    | causal                   |
| $\infty$ | non-causal               |
The capacity region $C_{\tau}(C_{\text{in}}, C_{\text{out}})$ is given by the closure of all achievable rate pairs. The sum-capacity, denoted by $C_{\tau}(C_{\text{in}}, C_{\text{out}})$, is defined as

$$C_{\tau}(C_{\text{in}}, C_{\text{out}}) := \max_{\epsilon} \frac{R_1 + R_2}{C_{\tau}(C_{\text{in}}, C_{\text{out}})}.$$ (4)

For example, $C_{T}(C_{\text{in}}, C_{\text{out}})$ and $C_{T}(C_{\text{in}}, C_{\text{out}})$ denote the capacity region and sum-capacity, respectively, of a MAC with a $(C_{\text{in}}, C_{\text{out}})$-CF and distributed causal state information available at the encoders.

### C. Message and State Cooperation

In the scenario where non-causal state information is available at the encoders, we also study the benefit of joint message and state cooperation. In the definition of a code for the case where non-causal state information is available at the encoders (Subsection II-B), for $i \in \{1, 2\}$, replace (1) and (3) with

$$\phi_{i, \text{in}} : [2^{nR_i}] \times S_{ni}^n \to [2^{nC_{i,\text{in}}}]$$

and

$$\sum_{t=1}^{n} \mathbb{E}b_i \left[f_{it}(w_i, \phi_{1, \text{out}}(\phi_{1, \text{in}}(w_1, S_{1, i}^n)), \phi_{2, \text{in}}(w_2, S_{2, i}^n), S_{ni}^n)\right] \leq B_i.$$ (5)

We denote the capacity region and sum-capacity with $C_{\infty,s}(C_{\text{in}}, C_{\text{out}})$ and $C_{\infty,s}(C_{\text{in}}, C_{\text{out}})$, respectively. The subscript “$s$” indicates the dependence of the cooperation strategy on the channel state information.

### III. Coding Strategy

Here we describe our coding strategies, which are based on random coding arguments. Since our aim is to determine conditions sufficient for an infinite slope cooperation gain, we specifically focus on coding strategies that lead to large gains for small cooperation rates such as the coordination strategy. In particular, in the coding strategies below, the CF does not use its rate for forwarding message or state information [1], since in the cases studied in the literature [7], [8], the gain of such a strategy is at most linear in the cooperation rate. We start with message cooperation and conclude with message and state cooperation.

#### A. Inner Bound for Message Cooperation

For simplicity, we assume the CF has access to both messages by setting $C_{\text{in}} = C_{\text{in}}^{*} = (C_{i,1}^{*}, C_{i,2}^{*})$, where $C_{i,1}^{*}$ and $C_{i,2}^{*}$ are sufficiently large. Despite this assumption, our main result regarding sum-capacity gain, Theorem 6, holds for any $C_{\text{in}} \in \mathbb{R}^2_{>0}$. This is due to the fact that using time-sharing, as stated in the lemma below, we can use the inner bounds for $C_{\text{in}}^{*}$ to obtain inner bounds for any $C_{\text{in}} \in \mathbb{R}^2_{>0}$. The proof appears in Subsection VI-A.

**Lemma 1.** Consider a memoryless stationary MAC. For any $(C_{\text{in}}, C_{\text{out}}) \in \mathbb{R}^2_{>0} \times \mathbb{R}^2_{>0}$, there exists $\mu > 0$, depending only on $C_{\text{in}}$, such that for all $\tau \in \{0, T - 1, T, \infty\}$,

$$C_{\tau}(C_{\text{in}}, C_{\text{out}}) - C_{\tau}(C_{\text{in}}, 0) \geq \mu \left(C_{\tau}(C_{\text{in}}, C_{\text{out}}) - C_{\tau}(C_{\text{in}}, 0)\right).$$

The coordination strategy [1] is the adaptation of Marton’s coding strategy for the broadcast channel [14] to the MAC with encoder cooperation.
We first describe our inner bound for the case where the encoders do not have access to state information. In this case, even though the decoder has access to full state information, we can obtain a suitable inner bound by applying results where state information is absent at both the encoders and the decoder to a modified channel. Specifically, applying \[^{[1]}\text{Theorem 1}\] to the channel
\[
\left( \mathcal{X}_1 \times \mathcal{X}_2, p(y, s|x_1, x_2), \mathcal{Y} \times \mathcal{S} \right),
\]
where
\[
p(y, s|x_1, x_2) = p(s)p(y|s, x_1, x_2),
\]
gives an inner bound for the channel \(p(y|s, x_1, x_2)\) when full state information is available at the decoder. We note that applying Lemma 2 together with the outer bound presented in Subsection VI-E gives the capacity region in the absence of cooperation \((C_{\text{out}} = 0)\) both in the case where no state information is available at the encoders and in the case where the state information available at the encoders is strictly causal.

**Lemma 2.** The set of all rate pairs \((R_1, R_2)\) satisfying
\[
R_1 \leq I(X_1; Y|S_1, S_2, X_2) \\
R_2 \leq I(X_2; Y|S_1, S_2, X_1) \\
R_1 + R_2 \leq I(X_1, X_2; Y|S_1, S_2)
\]
for some distribution \(p(x_1)p(x_2)\) with
\[
I(X_1; X_2) \leq C_1^{\text{out}} + C_2^{\text{out}}
\]
and \(E[b_i(X_i)] \leq B_i\) for \(i \in \{1, 2\}\), is contained in \(\mathcal{C}_0(C_{\text{in}}^*, C_{\text{out}})\).

In the case where the encoders have access to causal state information, the codeword transmitted by an encoder can depend both on its message and the present state information. Lemma 3 provides an inner bound for the capacity region in this scenario. In this inner bound, for \(i \in \{1, 2\}\), \(U_i\) encodes the message of encoder \(i\) in addition to the information it receives from the CF. Note that this inner bound is tight when \(C_{\text{out}} = 0\), even if non-causal state information is available at the encoders. (See Subsection VI-B for the proof of this lemma and Subsection VI-E for the corresponding outer bound in the absence of cooperation.)

**Lemma 3.** The set of all rate pairs satisfying
\[
R_1 \leq I(U_1; Y|S_1, S_2, U_2) \\
R_2 \leq I(U_2; Y|S_1, S_2, U_1) \\
R_1 + R_2 \leq I(U_1, U_2; Y|S_1, S_2)
\]
for some distribution \(p(u_1, u_2)p(x_1|u_1, s_1)p(x_2|u_2, s_2)\) with
\[
I(U_1; U_2) \leq C_1^{\text{out}} + C_2^{\text{out}}
\]
\(^3\)As we show in Subsection VI-B we can choose \(X_1\) and \(X_2\) to be deterministic functions of \((U_1, S_1)\) and \((U_2, S_2)\), respectively.
and \( E[h_i(X_i)] \leq B_i \) for \( i \in \{1, 2\} \), is contained in \( C_T^*(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) \).

**B. Inner Bound for Message and State Cooperation**

As discussed in Subsection [II.C] we only consider message and state cooperation in the scenario where non-causal state information is available at the encoders.

Here we assume that the state alphabet \( S = S_1 \times S_2 \) is discrete and \( H(S_1, S_2) \) is finite. Furthermore, we assume the CF not only has access to both messages, but also knows the state sequences \( \bar{S}_1^n \) and \( \bar{S}_2^n \); equivalently, we set \( \mathbf{C}_{\text{in}} = \bar{C}_{\text{in}} = (\bar{C}_{\text{in}}^1, \bar{C}_{\text{in}}^2) \), where \( \bar{C}_{\text{in}}^1 \) and \( \bar{C}_{\text{in}}^2 \) are sufficiently large. A lemma analogous to Lemma 1 holds in this case.

**Lemma 4.** Fix a memoryless stationary MAC. For any \((\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) \in \mathbb{R}^2_+ \times \mathbb{R}^2_+\), there exists \( \mu > 0 \), depending only on \( \mathbf{C}_{\text{in}} \), such that

\[
C_{(\infty,s)}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) - C_{(\infty,s)}(\mathbf{C}_{\text{in}}, \mathbf{0}) \geq \mu \left( C_{(\infty,s)}(\bar{C}_{\text{in}}, \mathbf{C}_{\text{out}}) - C_{(\infty,s)}(\bar{C}_{\text{in}}, \mathbf{0}) \right),
\]

We next describe our coding strategy for the MAC with message and state cooperation.

**Codebook Generation.** Choose a distribution \( p(x_1, x_2 | s_1, s_2) \). For \( i \in \{1, 2\} \), \( w_i \in [2^n R_i] \), \( z_i \in [2^n C_{\text{out}}^i] \), \( s_i^n \in S_i^n \), generate \( X_i^n(w_i, z_i | s_i^n) \) i.i.d. according to the distribution

\[
\Pr \left\{ X_i^n(w_i, z_i | s_i^n) = x_i^n \middle| s_i^n = s_i^n \right\} = \prod_{t=1}^{n} p(x_i | s_i).
\]

**Encoding.** The CF, having access to \((w_1, w_2)\) and \((S_1^n, S_2^n)\), looks for a pair \((Z_1, Z_2) \in [2^n C_{\text{out}}^1] \times [2^n C_{\text{out}}^2]\) satisfying

\[
(S_1^n, S_2^n, X_1^n(w_1, Z_1 | S_1^n), X_2^n(w_2, Z_2 | S_2^n)) \in A_3^{(n)},
\]

where \( A_3^{(n)} \) is the weakly typical set with respect to the distribution \( p(s_1, s_2) p(x_1, x_2 | s_1, s_2) \). If there is more than one such pair, the CF chooses the smallest pair according to the lexicographical order. If there is no such pair, it sets \((Z_1, Z_2) = (1, 1)\). The CF sends \( Z_i \) to encoder \( i \) for \( i \in \{1, 2\} \). Encoder \( i \) transmits \( X_i^n(w_i, Z_i | S_i^n) \) over \( n \) uses of the channel.

Using [15, p. 130, Lemma A.1.1], it follows that as \( n \) goes to infinity, the probability that a pair \((Z_1, Z_2)\) satisfying (5) exists goes to one if

\[
\begin{align*}
C_{\text{out}}^1 &> H(X_1 | S_1) - H(X_1 | S_1, S_2) + 24 \delta \\
C_{\text{out}}^2 &> H(X_2 | S_2) - H(X_2 | S_1, S_2) + 24 \delta \\
C_{\text{out}}^1 + C_{\text{out}}^2 &> H(X_1 | S_1) + H(X_2 | S_2) - H(X_1, X_2 | S_1, S_2) + 6 \delta.
\end{align*}
\]

**Decoding.** Once the decoder receives \( Y^n \), using \((S_1^n, S_2^n)\), it looks for a pair \((\hat{w}_1, \hat{w}_2)\) that satisfies

\[
(S_1^n, S_2^n, X_1^n(\hat{w}_1, \hat{Z}_1 | S_1^n), X_2^n(\hat{w}_2, \hat{Z}_2 | S_2^n), Y^n) \in A_4^{(n)}.
\]

Here \( A_4^{(n)} \) is the weakly typical set with respect to the distribution \( p(s_1, s_2) p(x_1, x_2 | s_1, s_2) p(y | s_1, s_2, x_1, x_2) \). If there is no such pair, or there is such a pair but it is not unique, the decoder sets \((\hat{w}_1, \hat{w}_2) = (1, 1)\).
The error analysis of the above coding scheme leads to the following lemma, which provides an inner bound for $\mathcal{C}_{\infty,s}(\tilde{C}_{\text{in}}, C_{\text{out}})$.

**Lemma 5.** The set of all rate pairs satisfying
\[
R_1 \leq I(X_1; Y|S_1, S_2, X_2) \\
R_2 \leq I(X_2; Y|S_1, S_2, X_1) \\
R_1 + R_2 \leq I(X_1, X_2; Y|S_1, S_2)
\]
for some distribution $p(x_1, x_2|s_1, s_2)$ with
\[
C_{\text{out}}^1 \geq H(X_1|S_1) - H(X_1|S_1, S_2) \\
C_{\text{out}}^2 \geq H(X_2|S_2) - H(X_2|S_1, S_2) \\
C_{\text{out}}^1 + C_{\text{out}}^2 \geq H(X_1|S_1) + H(X_2|S_2) - H(X_1, X_2|S_1, S_2)
\]
and $\mathbb{E}[b_i(X_i)] \leq B_i$ for $i \in \{1, 2\}$, is contained in $\mathcal{C}_{\infty,s}(\tilde{C}_{\text{in}}, C_{\text{out}})$.

**IV. MAIN RESULT**

Our main result describes conditions on a MAC that, if satisfied, for every fixed $C_{\text{in}} \in \mathbb{R}_{>0}^2$, guarantee an infinite slope in sum-capacity as a function of $C_{\text{out}}$. As sum-capacity depends on the availability of state information at the encoders, so do our conditions. The proof appears in Subsection VI-C.

**Theorem 6.** Let $\mathcal{S}$, $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{Y}$ be finite sets. For any $\tau \in \{0, T-1, T, \infty, (\infty, s)\}$, any MAC in $\mathcal{C}_{\tau}(\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y})$, and any $(C_{\text{in}}, v) \in \mathbb{R}_{>0}^2 \times \mathbb{R}_{>0}^2$,
\[
\lim_{h \to 0^+} \frac{\mathcal{C}_{\tau}(C_{\text{in}}, hv) - \mathcal{C}_{\tau}(C_{\text{in}}, 0)}{h} = \infty.
\]

We next specifically define $\mathcal{C}_{\tau}(\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y})$ for each subscript $\tau \in \{0, T-1, T, \infty, (\infty, s)\}$; as defined in Section II, $\tau$ specifies the availability of state information at the encoders. Note that the definition of $C_{\tau}$ provides sufficient conditions for a large cooperation gain; these conditions may not be necessary.

In our descriptions below, all mentioned distributions satisfy
\[
\mathbb{E}[b_i(X_i)] \leq B_i \quad \text{for } i \in \{1, 2\}.
\]

**A. Message Cooperation**

In this subsection, we define classes of MACs which exhibit a large message cooperation gain as described in Theorem 6.

**No state information.** A MAC is in $\mathcal{C}_0(\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y})$ if
(i) for some $p_0(x_1)p_0(x_2)$ that satisfies
\[
I_0(X_1, X_2; Y|S) = \max_{p(x_1)p(x_2)} I(X_1, X_2; Y|S),
\]
there exist \( p_1(x_1, x_2) \) that satisfies

\[
I_1(X_1, X_2; Y|S) + \mathbb{E}\left[D(p_1(y|S)||p_0(y|S))\right] > I_0(X_1, X_2; Y|S), \quad \text{and}
\]

(ii) \( \text{supp}(p_1(x_1, x_2)) \subseteq \text{supp}(p_0(x_1)p_0(x_2)) \), where “supp” denotes the support.

Intuitively, condition (i) ensures that our channel has the property that the dependence created through message cooperation increases sum-capacity. Condition (ii) allows the CF to use a small rate (i.e., small \( C_{\text{out}} \)) to help the encoders, whose codewords are generated according to \( p_0(x_1)p_0(x_2) \), to transmit codewords whose distribution is sufficiently close to \( p_1(x_1, x_2) \) to achieve a large gain in sum-capacity.

**Strictly causal state information.** As mentioned in Section III the availability of strictly causal state information at the encoders of a MAC without cooperation does not enlarge the capacity region, thus we set \( C_{T-1}(S, X_1, X_2, Y) := C_0(S, X_1, X_2, Y) \).

**Causal state information.** A MAC is in \( C_T(S, X_1, X_2, Y) \) if

(i) for some \( p_0(x_1|s_1)p_0(x_2|s_2) \) that satisfies

\[
I_0(X_1, X_2; Y|S) = \max_{p(x_1|s_1)p(x_2|s_2)} I(X_1, X_2; Y|S),
\]

there exist alphabets \( U_1, U_2 \), distributions \( p_0(u_1)p_0(u_2) \) and \( p_1(u_1, u_2) \), and mappings \( f_i : U_i \times S_i \to X_i \) for \( i \in \{1, 2\} \) such that

\[
p_0(x_1|s_1)p_0(x_2|s_2) = \sum_{u_1, u_2} p_0(u_1)p_0(u_2) \mathbf{1}\{x_1 = f_1(u_1, s_1)\} \mathbf{1}\{x_2 = f_2(u_2, s_2)\},
\]

\[
I_1(U_1, U_2; Y|S) + \mathbb{E}\left[D(p_1(y|S)||p_0(y|S))\right] > I_0(U_1, U_2; Y|S), \quad \text{and}
\]

(ii) \( \text{supp}(p_1(u_1, u_2)) \subseteq \text{supp}(p_0(u_1)p_0(u_2)) \).

In (7), the expressions are calculated using the input distributions

\[
p_0(u_1)p_0(u_2) \mathbf{1}\{x_1 = f_1(u_1, s_1)\} \mathbf{1}\{x_2 = f_2(u_2, s_2)\}, \quad \text{and}
\]

\[
p_1(u_1, u_2) \mathbf{1}\{x_1 = f_1(u_1, s_1)\} \mathbf{1}\{x_2 = f_2(u_2, s_2)\}.
\]

**Non-causal state information.** In the absence of cooperation, the capacity region is the same for regardless of whether the state information at the encoders is causal or non-causal. Thus, similar to the strictly causal case, we set \( C_{\infty}(S, X_1, X_2, Y) := C_T(S, X_1, X_2, Y) \).

**B. Message and State Cooperation**

Here we provide sufficient conditions for a large gain resulting from message and state cooperation.

**Non-causal state information.** A MAC is in \( C_{\infty,s}(S, X_1, X_2, Y) \) if

(i) for some \( p_0(x_1|s_1)p_0(x_2|s_2) \) that satisfies

\[
I_0(X_1, X_2; Y|S) = \max_{p(x_1|s_1)p(x_2|s_2)} I(X_1, X_2; Y|S),
\]

there exists \( p_1(x_1, x_2|s_1, s_2) \) that satisfies

\[
I_1(X_1, X_2; Y|S) + \mathbb{E}\left[D(p_1(y|S)||p_0(y|S))\right] > I_0(X_1, X_2; Y|S), \quad \text{and}
\]

(ii) for all \( (s_1, s_2) \in S \), \( \text{supp}(p_1(x_1, x_2|s_1, s_2)) \subseteq \text{supp}(p_0(x_1|s_1)p_0(x_2|s_2)) \).
V. Example: Gaussian MAC with Binary Fading

While Theorem 6 is stated only for finite alphabet MACs, the result is not limited to such MACs. Specifically, for a given MAC, we can use our inner bounds described in Section III to calculate an inner bound for sum-capacity and verify the result of Theorem 6 directly. We next describe an example of such a MAC.

Consider a MAC that models the wireless communication between two encoders and a decoder in the presence of binary fading. The input-output relationship of this MAC is given by

$$Y = S_1X_1 + S_2X_2 + Z,$$

where $$(S_1, S_2)$$ is uniformly distributed on $$\{0, 1\}^2$$, and $$Z$$ is a Gaussian random variable with mean zero and variance $$N$$. In addition, for $$i \in \{1, 2\}$$ we set the cost function $$b_i(x) = x^2$$ and cost constraint $$B_i = P_i$$, so that the cost constraints correspond to the usual power constraints of the Gaussian MAC.

**Proposition 7.** Consider the Gaussian MAC with binary fading. Fix $$(C_{in}, v) \in \mathbb{R}^2_{>0} \times \mathbb{R}^2_{>0}$$. Then for all $$\tau \in \{0, T - 1, T, \infty, (\infty, s)\}$$,

$$\lim_{h \to 0^+} \frac{C_{\tau}(C_{in}, hv) - C_{\tau}(C_{in}, 0)}{h} = \infty.$$

The proof appears in Subsection VI-D.

VI. Proofs

A. Proof of Lemma 7

Since $$C_{in} \in \mathbb{R}^2_{>0}$$, there exists $$\mu \in (0, 1)$$ such that for $$i \in \{1, 2\},$$

$$C_{in}^i \geq \mu C_{in}^*.$$

Then for each $$\tau \in \{0, T - 1, T, \infty\}$$, a time-sharing argument shows that

$$C_{\tau}(C_{in}, C_{out}) \geq \mu C_{\tau}(C_{in}/\mu, C_{out}) + (1 - \mu)C_{\tau}(0, C_{out})$$

$$\geq \mu C_{\tau}(C_{in}^*, C_{out}) + (1 - \mu)C_{\tau}(0, C_{out}).$$

Thus

$$C_{\tau}(C_{in}, C_{out}) \geq \mu C_{\tau}(C_{in}^*, C_{out}) + (1 - \mu)C_{\tau}(0, C_{out}),$$

which implies

$$C_{\tau}(C_{in}, C_{out}) - C_{\tau}(C_{in}, 0) \geq \mu \left( C_{\tau}(C_{in}^*, C_{out}) - C_{\tau}(C_{in}, 0) \right)$$

since

$$C_{\tau}(0, C_{out}) = C_{\tau}(0, 0) = C_{\tau}(C_{in}, 0) = C_{\tau}(C_{in}^*, 0).$$
B. Proof of Lemma 3

Fix alphabets \( U_1 \) and \( U_2 \), and mappings \( f_i : U_i \times S_i \rightarrow X_i \) for \( i \in \{1, 2\} \).

Applying Lemma 2, where state information is only available at the decoder, to the channel

\[
p(y|s, u_1, u_2) = \sum_{x_1, x_2} p(y|s, x_1, x_2) \mathbf{1}\{x_1 = f_1(u_1, s_1)\} \mathbf{1}\{x_2 = f_2(u_2, s_2)\}
\]

(8)

shows that the set of all rate pairs satisfying

\[
R_1 \leq I(U_1; Y|S, U_2) \\
R_2 \leq I(U_2; Y|S, U_1) \\
R_1 + R_2 \leq I(U_1, U_2; Y|S)
\]

for some distribution \( p(u_1, u_2) \) with

\[
I(U_1; U_2) \leq C_{\text{out}}^1 + C_{\text{out}}^2
\]

is achievable for the channel \( p(y|s, u_1, u_2) \) when no state information is available at the encoders. Note that every code for this channel can be transformed into a code for the channel \( p(y|s, x_1, x_2) \) with causal state information available at the encoders; for all times \( t \in [n] \) and \( i \in \{1, 2\} \), simply apply the mapping \( f_i \) to the pair \( (U_{it}, S_{it}) \), where \( U_{it} \) is the output symbol of encoder \( i \) and \( S_{it} \) is component \( i \) of the state at time \( t \). Note that the new code has the same rate and by (8), the same average error probability as the original code. Thus \( \mathcal{C}_T(C_{\text{in}}^*, C_{\text{out}}) \) contains the set of all rate pairs \( (R_1, R_2) \) satisfying

\[
R_1 \leq I(U_1; Y|S, U_2) \\
R_2 \leq I(U_2; Y|S, U_1) \\
R_1 + R_2 \leq I(U_1, U_2; Y|S)
\]

for some distribution \( p(u_1, u_2) \) with

\[
I(U_1; U_2) \leq C_{\text{out}}^1 + C_{\text{out}}^2
\]

and mappings

\[
f_i : U_i \times S_i \rightarrow X_i \text{ for } i \in \{1, 2\}.
\]

To complete the proof, we show that for every \( \delta \geq 0 \) and every distribution

\[
p(u_1, u_2)p(s_1, s_2)p(x_1|u_1, s_1)p(x_2|u_2, s_2)
\]

satisfying \( I(U_1; U_2) \leq \delta \), there exist alphabets \( U_1' \) and \( U_2' \), mappings

\[
f_i : U_i' \times S_i \rightarrow X_i \text{ for } i \in \{1, 2\},
\]

and distribution \( p(u_1', u_2') \) such that \( I(U_1'; U_2') = I(U_1; U_2) \), and the rate region calculated with respect to

\[
p(u_1', u_2') \mathbf{1}\{x_1 = f_1(u_1', s_1)\} \mathbf{1}\{x_2 = f_2(u_2', s_2)\},
\]
contains the region calculated with respect to \( p(u_1, u_2)p(x_1|u_1, s_1)p(x_2|u_2, s_2) \).

To this end, applying Lemma \(^8\) which appears at the end of this subsection, to \( p(x_i|u_i, s_i) \) demonstrates the existence of a random variable \( V_i \) that is independent of \( (U_i, S_i) \) and a mapping
\[
f_i : V_i \times U_i \times S_i \rightarrow X_i
\]
that satisfies
\[
p(x_i|u_i, s_i) = \sum_{v_i} p(v_i) \mathbf{1}\{x_i = f_i(v_i, u_i, s_i)\}.
\]
Furthermore, without loss of generality, we may assume \( V_1 \) and \( V_2 \) are independent, and \( (V_1, V_2) \) is independent of \( (U_1, U_2, S_1, S_2) \).

Let \( U_i' \coloneqq (U_i, V_i) \) for \( i \in \{1, 2\} \). Then
\[
I(U_1'; U_2') = I(U_1, V_1; U_2, V_2) = H(U_1, V_1) + H(U_2, V_2) - H(U_1, U_2, V_1, V_2) = I(U_1; U_2) + I(V_1; V_2) = I(U_1; U_2).
\]

We now show that the rate region calculated with respect to the distribution
\[
p(s_1, s_2)p(u_1', u_2') \mathbf{1}\{x_1 = f_1(u_1', s_1)\} \mathbf{1}\{x_2 = f_2(u_2', s_2)\}
\]
contains the rate region with respect to
\[
p(s_1, s_2)p(u_1, u_2)p(x_1|u_1, s_1)p(x_2|u_2, s_2).
\]

Recall that \( S = (S_1, S_2) \). We have
\[
I(U_1'; Y|S, U_2') = I(U_1'; Y, U_2'|S) - I(U_1'; U_2'|S) = I(U_1'; Y, U_2'|S) - I(U_1; U_2'|S) = I(U_1; V_1; Y, U_2|S) - I(U_1; U_2|S) \geq I(U_1; Y, U_2|S) - I(U_1; U_2|S) = I(U_1; Y|S, U_2).
\]

Similarly, we show
\[
I(U_2'; Y|S, U_1') \geq I(U_2; Y|S, U_1).
\]

Finally, we have
\[
I(U_1', U_2'; Y|S) = I(U_1, V_1, U_2, V_2; Y|S) \geq I(U_1, U_2; Y|S).
\]

This completes the proof. We next state and prove Lemma \(^8\) which we applied earlier in the proof. In Lemma \(^8\) the scenario where \( \mathcal{X'} \) and \( S \) are finite is a special case of the functional representation lemma \([16] p. 626\).
Lemma 8. Let \( \{F(\cdot|s)\}_{s \in S} \) be a collection of cumulative distribution functions (CDFs) on alphabet \( X \subseteq \mathbb{R} \) and let \( S \) be a random variable with alphabet \( S \). Then there exists a random variable \( U \) independent of \( S \) and a mapping 
\[
g : S \times U \to X
\]
such that the conditional CDF of \( g(S,U) \) given \( S = s \) equals \( F(\cdot|s) \). In the case where \( X \) and \( S \) are finite, we can choose \( U \) such that 
\[
|U| \leq |S|(|X| - 1) + 1.
\]

Proof. We prove the result for general alphabets \( X \subseteq \mathbb{R} \). Let \( U := [0,1] \). Define the mapping \( g : S \times U \to X \) as 
\[
g(s,u) = \inf \left\{ x \in X \mid F(x|s) \geq u \right\}.
\]
Let \( U \) be independent of \( S \) and uniformly distributed on \( U = [0,1] \). From the quantile function theorem [17, Theorem 2], it follows that for all \( s \in S \), \( g(s,U) \) has CDF \( F(\cdot|s) \). Set \( X = g(S,U) \). Then
\[
F_{X|S}(x|s) = \Pr \left\{ X \leq x \mid S = s \right\} = \Pr \left\{ g(S,U) \leq x \mid S = s \right\} = \Pr \left\{ g(s,U) \leq x \right\} = F(x|s).
\]

C. Proof of Theorem 6

From the description of the set \( C_\tau(S,X_1,X_2,Y) \) in Section IV, we see that it suffices to prove Theorem 6 only in the cases \( \tau = 0, \tau = T \), and \( \tau = (\infty, s) \).

The case \( \tau = 0 \). When no state information is available at the encoders, Theorem 6 follows by applying [1, Theorem 3] to the MAC
\[
p(s,y|x_1,x_2) = p(s)p(y|s,x_1,x_2),
\]
with input alphabets \( X_1 \) and \( X_2 \), and output alphabet \( S \times Y \).

The case \( \tau = T \). In this case, it suffices to check that \( p_0(u_1)p_0(u_2) \) satisfies
\[
I_0(U_1,U_2;Y|S) = \max_{p(u_1)p(u_2)} I(U_1,U_2;Y|S)
\]
for the MAC
\[
p(y|s,u_1,u_2) = \sum_{x_1,x_2} 1\{x_1 = f_1(u_1,s_1)\} 1\{x_2 = f_2(u_2,s_2)\} p(y|s,x_1,x_2),
\]
since if (10) holds, then Theorem 6 follows by applying the case \( \tau = 0 \) to the MAC defined by (11).
To prove (10), first note that for all \( p(u_1)p(u_2) \),
\[
I(U_1, U_2; Y|S) = H(Y|S) - H(Y|U_1, U_2, S)
\leq H(Y|S) - H(Y|U_1, U_2, S, X_1, X_2)
\]
\[
= H(Y|S) - H(Y|S, X_1, X_2)
= I(X_1, X_2; Y|S) \leq I_0(X_1, X_2; Y|S).
\]
(13)

For the distribution \( p_0(u_1)p_0(u_2) \), however, the inequalities in (12) and (13) hold with equality due to (6).

The case \( \tau = (\infty, s) \). In this case, we provide a self-contained proof as it is not straightforward to derive it from prior cases. This is due to the fact that in this case, as described in Lemma 5, the family of achievable distributions is constrained by three inequalities rather than one.

Let \( p_0(x_1|s_1)p_0(x_2|s_2) \) be a distribution that satisfies
\[
I_0(X_1, X_2; Y|S) = \max_{p(x_1|s_1)p(x_2|s_2)} I(X_1, X_2; Y|S).
\]

By assumption, there exists a distribution \( p_1(x_1, x_2|s_1, s_2) \) such that
\[
I_1(X_1, X_2; Y|S) + \mathbb{E} \left[ D(p_1(y|S)||p_0(y|S)) \right] > I_0(X_1, X_2; Y|S), \quad (14)
\]
\[
\forall (s_1, s_2) \in S: \text{supp}(p_1(x_1, x_2|s_1, s_2)) \subseteq \text{supp}(p_0(x_1|s_1)p_0(x_2|s_2))
\]
(15)

For every \( \lambda \in (0, 1) \), define
\[
p_\lambda(x_1, x_2|s_1, s_2) := (1 - \lambda)p_0(x_1|s_1)p_0(x_2|s_2) + \lambda p_1(x_1, x_2|s_1, s_2).
\]

Fix \( \epsilon > 0 \) and \( v \in \mathbb{R}^2_{>0} \). Define the mapping \( h : [0, 1] \to \mathbb{R} \) as
\[
h(\lambda) = \frac{1}{v_1} I_\lambda(X_1; S_2|S_1) + \frac{1}{v_2} I_\lambda(X_2; S_1|S_2) + \frac{1}{v_1 + v_2} I_\lambda(X_1; X_2|S_1, S_2) + \epsilon \lambda.
\]

A direct calculation, followed by an application of (15), shows that
\[
\frac{d}{d\lambda} I_\lambda(X_1; S_2|S_1)\bigg|_{\lambda=0^+} = 0
\]
\[
\frac{d}{d\lambda} I_\lambda(X_2; S_1|S_2)\bigg|_{\lambda=0^+} = 0
\]
\[
\frac{d}{d\lambda} I_\lambda(X_1; X_2|S_1, S_2)\bigg|_{\lambda=0^+} = 0.
\]

Note that \( h \) is continuously differentiable and
\[
\frac{dh}{d\lambda}\bigg|_{\lambda=0^+} = \epsilon > 0.
\]

Therefore, by the inverse function theorem, there exists \( h_0 > 0 \) such that \( h \) is invertible on \([0, h_0)\); that is, there exists a mapping \( \lambda^*: [0, h_0) \to [0, 1] \) that satisfies
\[
\lambda = \frac{1}{v_1} I_{\lambda^*(h)}(X_1; S_2|S_1) + \frac{1}{v_2} I_{\lambda^*(h)}(X_2; S_1|S_2) + \frac{1}{v_1 + v_2} I_{\lambda^*(h)}(X_1; X_2|S_1, S_2) + \epsilon \lambda^*(h),
\]
(16)
and
\[
\frac{d\lambda^*}{dh}\bigg|_{h=0^+} = \frac{1}{\epsilon}.
\]
We henceforth write $\lambda^*$ instead of $\lambda^*(h)$ when the value of $h$ is clear from context. By \cite{16}, it now follows that for all $h \in [0, h_0)$,

$$h v_1 \geq I_{\lambda^*}(X_1; S_2|S_1)$$

$$= H_{\lambda^*}(X_1|S_1) - H_{\lambda^*}(X_1|S_1, S_2)$$

$$h v_2 \geq I_{\lambda^*}(X_2; S_1|S_2)$$

$$= H_{\lambda^*}(X_2|S_2) - H_{\lambda^*}(X_2|S_1, S_2)$$

$$h(v_1 + v_2) \geq I_{\lambda^*}(X_1; S_2|S_1) + I_{\lambda^*}(X_2; S_1|S_2) + I_{\lambda^*}(X_1; X_2|S_1, S_2)$$

$$= H_{\lambda^*}(X_1|S_1) + H_{\lambda^*}(X_2|S_2) - H_{\lambda^*}(X_2|S_1, S_2).$$

Thus, by Lemma \cite{5}

$$C_{(\infty, \sigma)}(\bar{C}_{in}, hv) \geq I_{\lambda^*}(X_1, X_2; Y|S) - I_{\lambda^*}(X_1; X_2|S). \quad (17)$$

Since equality holds in \cite{17} at $h = 0$, we have

$$\liminf_{h \to 0^+} \frac{C_{(\infty, \sigma)}(\bar{C}_{in}, hv) - C_{(\infty, \sigma)}(\bar{C}_{in}, 0)}{h} \geq \frac{1}{\epsilon} \frac{d}{d\lambda^*} \left( I_{\lambda^*}(X_1, X_2; Y|S) - I_{\lambda^*}(X_1; X_2|S) \right) \bigg|_{\lambda^* = 0^+}$$

$$= \frac{1}{\epsilon} \frac{d}{d\lambda^*} I_{\lambda^*}(X_1, X_2; Y|S) \bigg|_{\lambda^* = 0^+}$$

$$\geq \frac{1}{\epsilon} \left( I_1(X_1, X_2; Y|S) + \mathbb{E} \left[ D(p_1(y|S)\|p_0(y|S)) \right] - I_0(X_1, X_2; Y|S) \right). \quad (19)$$

The proof of \cite{19} is analogous to \cite{1} Lemma 14 (ii)] and is omitted. Since \cite{19} holds for all $\epsilon > 0$, from \cite{14} it follows that

$$\lim_{h \to 0^+} \frac{C_{(\infty, \sigma)}(\bar{C}_{in}, hv) - C_{(\infty, \sigma)}(\bar{C}_{in}, 0)}{h} = \infty.$$  

\[D. \ Proof \ of \ Proposition \ [7]\]

Since

$$C_{T-1}(C_{\infty}^*, 0) = C_0(C_{\infty}^*, 0) = C_0(0, 0)$$

and

$$C_{(\infty, \sigma)}(C_{\infty}^*, 0) = C_{\infty}(C_{\infty}^*, 0) = C_T(C_{\infty}^*, 0) = C_T(0, 0),$$

it suffices to prove the result only when $\tau = 0$ or $\tau = T$.

When $\tau = 0$, from Lemma \cite{2} it follows that for any distribution $p(x_1)p(x_2)$ satisfying $\mathbb{E}[X_i^2] \leq P_i$ for $i \in \{1, 2\}$ and

$$I(X_1; X_2) \leq C_{out}^1 + C_{out}^2,$$
we have

\[ C_0(C^*_{in}, C_{out}) \geq I(X_1, X_2; Y|S) - I(X_1; X_2) \]
\[ = H(Y|S) - H(Z) \]
\[ = \frac{1}{4} \left( H(X_1 + Z) + H(X_2 + Z) + H(X_1 + X_2 + Z) - 3H(Z) \right) \quad (20) \]

Fix \( h > 0 \). Let \((X_1, X_2)\) be jointly Gaussian with mean zero and covariance matrix

\[ \Sigma := \begin{pmatrix} \sqrt{\mathcal{P}_1} & \rho \sqrt{\mathcal{P}_1 \mathcal{P}_2} \\ \rho \sqrt{\mathcal{P}_1 \mathcal{P}_2} & \sqrt{\mathcal{P}_2} \end{pmatrix}, \]

where \( \rho \in [0, 1] \) is chosen such that

\[ I(X_1; X_2) = \frac{1}{2} \log \frac{1}{1 - \rho^2} := h(v_1 + v_2). \]

Then

\[ \frac{d\rho}{dh} \bigg|_{h=0^+} = \infty. \]

Using (20), it follows that

\[ C_0(C^*_{in}, hv) - C_0(C^*_{in}, 0) \geq \frac{1}{8} \log \left( 1 + \frac{2\rho \sqrt{\mathcal{P}_1 \mathcal{P}_2}}{\mathcal{P}_1 + \mathcal{P}_2 + N} \right) - h(v_1 + v_2), \]

which implies the desired result.

A similar proof follows when \( \tau = T \). In this case, for fixed \( h > 0 \), let \((U_1, U_2)\) be jointly Gaussian with mean zero and covariance matrix

\[ \Sigma := \begin{pmatrix} \sqrt{2\mathcal{P}_1} & 2\rho \sqrt{\mathcal{P}_1 \mathcal{P}_2} \\ 2\rho \sqrt{\mathcal{P}_1 \mathcal{P}_2} & \sqrt{2\mathcal{P}_2} \end{pmatrix}, \]

where \( \rho \in [0, 1] \) satisfies

\[ I(U_1; U_2) = \frac{1}{2} \log \frac{1}{1 - \rho^2} := h(v_1 + v_2). \]

For \( i \in \{1, 2\} \), set \( X_i := S_i U_i \). From Lemma 5, it follows that

\[ C_0(C^*_{in}, C_{out}) \geq I(U_1, U_2; Y|S) - I(U_1; U_2) \]
\[ = H(Y|S) - H(Y|S, U_1, U_2) - I(U_1; U_2) \]
\[ = H(Y|S) - H(Y|S, U_1, U_2, X_1, X_2) - I(U_1; U_2) \]
\[ = H(Y|S) - H(Z) - I(U_1; U_2) \]
\[ = \frac{1}{4} \left( H(X_1 + Z|S_1 = 1) + H(X_2 + Z|S_2 = 1) \\ + H(X_1 + X_2 + Z|S_1 = 1, S_2 = 1) - 3H(Z) \right) - I(U_1; U_2), \quad (21) \]

where (21) follows from the fact that \((X_1, X_2)\) is a deterministic function of \((S, U_1, U_2)\). Simplifying (23) results in

\[ C_T(C^*_{in}, hv) - C_T(C^*_{in}, 0) \geq \frac{1}{8} \log \left( 1 + \frac{4\rho \sqrt{\mathcal{P}_1 \mathcal{P}_2}}{2\mathcal{P}_1 + 2\mathcal{P}_2 + N} \right) - h(v_1 + v_2). \]
E. Outer Bounds in the Absence of Cooperation

We next prove outer bounds for $C_{T-1}(0,0)$ and $C_{\infty}(0,0)$. Together with our inner bounds in Section III, these outer bounds determine the capacity region $C_{\tau}(0,0)$ for all $\tau$, and show

$$C_0(0,0) = C_{T-1}(0,0) \text{ and } C_{T}(0,0) = C_{\infty}(0,0) = C_{(\infty,s)}(0,0).$$

The bounds presented here are well known [16, p. 175] and are included for completeness.

Consider a sequence of $2^{nR_1}, 2^{nR_2}, n$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$ for the MAC with full state information at the decoder. Initially, we do not make any assumptions regarding the presence of state information at the encoders.

We begin with the bound on $R_1$. We have

$$nR_1 = H(W_1)$$
$$= H(W_1|S^n, W_2)$$
$$\simeq I(W_1;Y^n|S^n, W_2)$$
$$= H(Y^n|W_1, W_2, X_2^n) - H(Y^n|W_1, W_2, X_2^n)$$
$$= H(Y^n|S^n, X_2^n) - H(Y^n|S^n, X_2^n)$$
$$= \sum_{t=1}^{n} \left( H(Y_t|Y^{t-1}, S^n, X_2^n) - H(Y_t|Y^{t-1}, S^n, X_1^n, X_2^n) \right). \tag{24}$$

Similarly,

$$nR_2 \simeq \sum_{t=1}^{n} \left( H(Y_t|Y^{t-1}, S^n, X_1^n) - H(Y_t|Y^{t-1}, S^n, X_1^n, X_2^n) \right).$$

Next we bound $R_1 + R_2$. We have

$$n(R_1 + R_2) = H(W_1, W_2)$$
$$= H(W_1, W_2|S^n)$$
$$\simeq I(W_1, W_2; Y^n|S^n)$$
$$= H(Y^n|S^n) - H(Y^n|S^n, W_1, W_2, X^n_2)$$
$$= H(Y^n|S^n) - H(Y^n|S^n, X^n_1, X^n_2)$$
$$= \sum_{t=1}^{n} \left( H(Y_t|Y^{t-1}, S^n) - H(Y_t|Y^{t-1}, S^n, X^n_1, X^n_2) \right).$$

To proceed further, we need to apply the causality constraints of the state information at the encoders.
The case $\tau = T - 1$. In this case, strictly causal state information is available at the encoders; that is, for $i \in \{1, 2\}$ and $t \in [n]$, $X_{it}$ is a deterministic function of $(W_i, S_i^{t-1})$. Continuing from (24), we get

$$nR_1 \lesssim \sum_{t=1}^{n} \left( H(Y_t|S^t, X_{2t}) - H(Y_t|S^t, X_{1t}, X_{2t}) \right)$$

$$= \sum_{t=1}^{n} I(X_{1t}; Y_t|S^t, X_{2t}).$$

Similarly, we get

$$nR_2 \lesssim \sum_{t=1}^{n} I(X_{2t}; Y_t|S^t, X_{1t})$$

$$n(R_1 + R_2) \lesssim \sum_{t=1}^{n} I(X_{1t}, X_{2t}; Y_t|S^t).$$

Setting $Q_t := S_t^{t-1}$ for $t \in [n]$ gives

$$nR_1 \lesssim \sum_{t=1}^{n} I(X_{1t}; Y_t|Q_t, S_t, X_{2t})$$

$$nR_2 \lesssim \sum_{t=1}^{n} I(X_{2t}; Y_t|Q_t, S_t, X_{1t})$$

$$n(R_1 + R_2) \lesssim \sum_{t=1}^{n} I(X_{1t}, X_{2t}; Y_t|Q_t, S_t).$$

Thus $\mathcal{C}_{T-1}(0, 0)$ is contained in the closure of the set of all rate pairs satisfying

$$R_1 \leq I(X_1; Y|Q, S, X_2)$$

$$R_2 \leq I(X_2; Y|Q, S, X_1)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y|Q, S)$$

for some distribution $p(q)p(x_1|q)p(x_2|q)$.

The case $\tau = \infty$. In this case, noncausal state information is available at the encoders, meaning that for $i \in \{1, 2\}$ and $t \in [n]$, $X_{it}$ is a deterministic function of $(W_i, S_i^n)$. From (24), we have

$$nR_1 \lesssim \sum_{t=1}^{n} \left( H(Y_t|S_t^{t}, S_2^{t+n}, X_{2t}) - H(Y_t|S_t^{t}, S_2^{t+n}, X_{1t}, X_{2t}) \right)$$

$$= \sum_{t=1}^{n} I(X_{1t}; Y_t|S_t^{t}, S_2^{t+n}, X_{2t}),$$

where for $t \in [n],

$$S_2^{t+n} = (S_{2t}, S_{2(t+1)}, \ldots, S_{2n}).$$

Similarly, we have

$$nR_2 \lesssim \sum_{t=1}^{n} I(X_{2t}; Y_t|S_t^{t}, S_2^{t+n}, X_{1t})$$

$$n(R_1 + R_2) \lesssim \sum_{t=1}^{n} I(X_{1t}, X_{2t}; Y_t|S_t^{t}, S_2^{t+n}).$$
For $t \in [n]$, following [7], define

$$Q_t := (S_t^{t-1}, S_t^{t+1:n}).$$

By assumption, $(S_1^n, S_2^n) \overset{\text{iid}}{\sim} p(s_1, s_2)$. Thus

$$p(s_1^n, s_2^n | s_1^{t-1}, s_2^{t+1:n}, s_{1t}, s_{2t}) = p(s_1^{t+1:n}, s_2^{t-1} | s_1^t, s_2^n) = p(s_1^{t+1:n} | s_2^{t+1:n}) p(s_2^{t-1} | s_1^t),$$

which implies that $S_1^n$ and $S_2^n$ are independent given $(Q_t, S_1^n, S_2^n)$. Since $(W_1, W_2)$ is independent of $(S_1^n, S_2^n)$, it follows that for $t \in [n]$, $X_1(W_1, S_1^n)$ and $X_2(W_2, S_2^n)$ are independent given $(Q_t, S_1^n, S_2^n)$. Thus $\mathcal{C}_\infty(0, 0)$ is contained in the closure of the set of all rate pairs satisfying

$$R_1 \leq I(X_1; Y | Q, S, X_2)$$
$$R_2 \leq I(X_2; Y | Q, S, X_1)$$
$$R_1 + R_2 \leq I(X_1, X_2; Y | Q, S)$$

for some distribution $p(q)p(x_1|q, s_1)p(x_2|q, s_2)$.

**VII. CONCLUSION**

The presence of distributed state information in a network provides an opportunity for cooperation. In this work, we study encoder cooperation in the MAC under the CF model. When no state information is available at either the encoders or the decoder, [1] provides conditions under which the sum-capacity gain of cooperation has an infinite slope in the limit of small cooperation rate. This work extends these conditions to scenarios where distributed state information is available at the encoders and full state information is available at the decoder.

**REFERENCES**

[1] P. Noorzad, M. Effros, and M. Langberg, “The unbounded benefit of encoder cooperation for the $k$-user MAC,” in *Proc. IEEE Int. Symp. Information Theory*, 2016. [Online]. Available: [http://arxiv.org/abs/1601.06113](http://arxiv.org/abs/1601.06113)

[2] A. J. Goldsmith and P. P. Varaiya, “Capacity of fading channels with channel side information,” *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, 1997.

[3] D. N. C. Tse and S. V. Hanly, “Multiaccess fading channels—Part I: Polymatroid structure, optimal resource allocation and throughput capacities,” *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, 1998.

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

[5] C. Heegard and A. A. El Gamal, “On the capacity of computer memory with defects,” *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 731–739, 1983.

[6] M. H. M. Costa, “Writing on dirty paper,” *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, 1983.

[7] H. H. Permuter, S. Shamai (Shitz), and A. Somekh-Baruch, “Message and state cooperation in multiple access channels,” *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6379–6396, 2011.

[8] F. M. J. Willems, “The discrete memoryless multiple access channel with partially cooperating encoders,” *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 441–445, 1983.

[9] Y. Cemal and Y. Steinberg, “The multiple-access channel with partial state information at the encoders,” *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3992–4003, 2005.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.
[11] S. Jafar, “Capacity with causal and noncausal side information: A unified view,” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5474, 2006.

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

[13] ——, “The multiple-access channel with causal side information: Double state,” *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1379–1393, 2013.

[14] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,” *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, 1979.

[15] P. Noorzad, “Network effects in small networks: A study of cooperation,” Ph.D. dissertation, California Institute of Technology, 2017.

[16] A. El Gamal and Y.-H. Kim, *Network Information Theory*, 2nd ed. Cambridge University Press, 2012.

[17] J. E. Angus, “The probability integral transform and related results,” *SIAM Review*, vol. 36, no. 4, pp. 652–654, 1994.