On the Power of Cooperation: Can a Little Help a Lot? (Extended Version)

Parham Noorzad*, Michelle Effros*, Michael Langberg†, Tracey Ho*
*California Institute of Technology
Emails: {parham, effros, tho}@caltech.edu
†State University of New York at Buffalo
Email: mikel@buffalo.edu

Abstract—In this paper, we propose a new cooperation model for discrete memoryless multiple access channels (DM-MACs). While in prior cooperation models (e.g., DM-MAC with conferencing encoders) the increase in a network’s sum-capacity that results from cooperation is no more than the rate introduced by the cooperation mechanism, we show that in our model for a particular sequence of channels, the increase in sum-capacity is significantly larger than the cooperation rate. This result contrasts with the network coding literature for networks of point to point links, where attempts to find examples in which similar small network modifications yield large capacity benefits have to date been unsuccessful.

I. INTRODUCTION

Cooperation is a potentially powerful strategy in distributed communication systems. It can both increase the possible transmission rates of source messages and improve the reliability of a network [8]. To date, cooperation is not completely understood. In this paper, we focus on the effect of cooperation on the capacity region and discuss situations where a small amount of cooperation results in a large increase in sum-capacity (a formal definition follows in Section II).

One model of cooperation, proposed by Willems in [15], is the conferencing encoders (CE) model for the DM-MAC. In the CE model, there is a noiseless link of capacity \( C_{12} \) from the first encoder to the second and a corresponding link of capacity \( C_{21} \) back. These links allow a finite number of rounds of communication between the two encoders; the total number of bits sent by each encoder is bounded by the capacity of its outgoing link. A similar type of cooperation is applied in the broadcast channel with conferencing decoders [3] and the interference channel with conferencing encoders [15]. More recently, the authors of [14] investigate the case where each encoder has partial state information and conferencing enables information exchange about both the state and the messages.

In this paper, we introduce the cooperation facilitator (CF) model for the DM-MAC. The cooperation facilitator is a node that has complete access to both source messages and sends the same information to both encoders through a noiseless bottleneck link of finite capacity. We define the cooperation rate as the capacity of this link. If we remove this bottleneck link from the network, then the two transmitters are not able to cooperate, and their transmitted codewords must be independent. This is the case of the DM-MAC with independent encoders (IE), whose capacity region was first derived by Ahlswede [1] and Liao [12].

Since removing the bottleneck link transforms the CF network into the IE network, the proposed cooperation model is related to the edge removal problem in network coding [5, 7, 9–11]. For networks of noiseless links, there are no known examples of networks for which removing a single edge of capacity \( \delta \) changes the capacity region by more than \( \delta \) in each dimension, and in some cases it is known that an impact of more than \( \delta \) per dimension is not possible [5, 7]. Therefore, at least in the situations investigated in [5, 7], inserting a cooperation facilitator in a network cannot increase the sum-capacity by more than a constant times the cooperation rate.

How much can cooperation help in a DM-MAC? In the CE model, the increase in sum-capacity is at most the sum of the capacities of the noiseless links between the two encoders (Section II). Given the previous discussion, one may wonder whether a similar result holds for the CF model, that is, whether the increase in sum-capacity is limited to a constant times the cooperation rate. In what follows, we see that the benefit of cooperation can far exceed what might be expected based on the CE and edge removal examples. Specifically, we describe a sequence of DM-MACs with increasing alphabet sizes and set the cooperation rate for each channel as a function of its alphabet size. We then show that the increase in sum-capacity that results from cooperation grows more quickly than any polynomial function of the cooperation rate.

In the next section, we review the CE model and its capacity region as presented by Willems [15]. We give a formal introduction to the CF model in Section III.

II. THE CONFERENCING ENCODERS MODEL

In the CE model, each encoder shares some information regarding its message with the other encoder prior to transmission over the channel. This sharing of information is achieved through a \( K \)-step conference over noiseless links of capacities \( C_{12} \) and \( C_{21} \). A \( K \)-step conference consists of two sets of functions, \( \{h_{11}, \ldots, h_{1K}\} \) and \( \{h_{21}, \ldots, h_{2K}\} \), which
where \( X \) sum-capacity is given by

\[
C = \sum_{k=1}^{K} \log |\mathcal{V}_k| \leq nC_{12},
\]

for \( k = 1, \ldots, K \). In step \( k \), encoder 1 (encoder 2) computes \( V_{1k} (V_{2k}) \) and sends it to encoder 2 (encoder 1). Since the noiseless links between the two encoders are of capacity \( C_{12} \) and \( C_{21} \), respectively, we require

\[
\sum_{k=1}^{K} \log |\mathcal{V}_k| \leq nC_{12},
\]

\[
\sum_{k=1}^{K} \log |\mathcal{V}_k| \leq nC_{21}
\]

where \( \mathcal{V}_k \) is the alphabet of the random variable \( V_k \) for \( i = 1, 2 \) and \( k = 1, \ldots, K \). The outputs of the encoders, \( X_1^n \) and \( X_2^n \), are given by

\[
X_1^n = f_{1n} (W_1, V_1^n),
\]

\[
X_2^n = f_{2n} (W_2, V_2^n)
\]

where \( f_{1n} \) and \( f_{2n} \) are deterministic functions. The capacity region of the DM-MAC with conferencing encoders \([15]\) is given by the set of all rate pairs \((R_1, R_2)\) such that

\[
R_1 \leq I (X_1; Y | X_2, U) + C_{12},
\]

\[
R_2 \leq I (X_2; Y | X_1, U) + C_{21},
\]

\[
R_1 + R_2 \leq I (X_1, X_2; Y | U) + C_{12} + C_{21},
\]

\[
R_1 + R_2 \leq I (X_1, X_2; Y)
\]

for some distribution \( p(u)p(x_1|u)p(x_2|u)p(y|x_1, x_2) \).

For a given capacity region \( C \subseteq \mathbb{R}_{+}^{2} \), the sum-capacity \( C_S \) is defined as

\[
C_S = \max \left\{ R_1 + R_2 \mid (R_1, R_2) \in C \right\}.
\]

In the IE model, following Ahlswede [12] and Liao [12], the sum-capacity is given by

\[
C_{S-IE} = \max_{p(x_1)p(x_2)} I (X_1, X_2; Y).
\]

By studying the capacity region of the CE model, we deduce

\[
C_{S-CE} \leq C_{S-IE} + C_{12} + C_{21}.
\]

Thus, with conferencing, the sum-capacity increases at most linearly in \((C_{12}, C_{21})\) over the sum-capacity of the IE model.

III. THE COOPERATION FACILITATOR MODEL

Consider the DM-MAC

\[
(X_1 \times X_2, p_{Y|X_1,X_2}(y|x_1, x_2), Y),
\]

where \( X_1, X_2 \), and \( Y \) are finite sets and \( p_{Y|X_1,X_2}(y|x_1, x_2) \) denotes the conditional distribution of the output, \( Y \), given the inputs, \( X_1 \) and \( X_2 \). To simplify notation, we suppress the subscript of the probability distributions when the corresponding random variables are clear from context. For example, we write \( p(x) \) instead of \( p_X(x) \).

In this network there are two sources, source 1 and source 2, whose outputs are the messages \( W_1 \in \mathcal{W}_1 = \{1, \ldots, 2^{nR_1}\} \) and \( W_2 \in \mathcal{W}_2 = \{1, \ldots, 2^{nR_2}\} \), respectively. The random variables \( W_1 \) and \( W_2 \) are independent and uniformly distributed over their corresponding alphabets. The real numbers \( R_1 \) and \( R_2 \) are nonnegative and are called the message rates. The cooperation facilitator is represented by the function

\[
\phi_n : \mathcal{W}_1 \times \mathcal{W}_2 \to Z,
\]

where \( Z = \{1, \ldots, 2^{nC}\} \) is determined by the cooperation rate \( C \). The output of the cooperation facilitator, \( Z = \phi_n(W_1, W_2) \), is available to both encoders. Each encoder chooses a blocklength-\( n \) codeword as a function of its own source and \( Z \) and sends that codeword to the receiver using \( n \) transmissions. Hence the two encoders can be represented by the functions

\[
f_{1n} : W_1 \times Z \to X_1^n,
\]

\[
f_{2n} : W_2 \times Z \to X_2^n.
\]

We denote the output of the encoders by \( X_1^n = f_{1n} (W_1, Z) \) and \( X_2^n = f_{2n}(W_2, Z) \). Let \( Y^n \) be the output of the channel when the pair \((X_1^n, X_2^n)\) is transmitted. Using \( Y^n \), the decoder estimates the original messages via a decoding function \( g_n : Y^n \to W_1 \times W_2 \). See Figure 1 for a schematic of the network model.

A \((2^{nR_1}, 2^{nR_2}, n)\) code for the multiple access channel with a cooperation facilitator with output rate \( C \) is defined as the quadruple

\[
(\phi_n, f_{1n}, f_{2n}, g_n).
\]

The average probability of error for this code is given by

\[
P_e^{(n)} = \Pr \left( g_n (Y^n) \neq (W_1, W_2) \right).
\]

We say the rate pair \((R_1, R_2)\) is achievable if there exists a sequence of \((2^{nR_1}, 2^{nR_2}, n)\) codes such that \( P_e^{(n)} \) tends to zero as the blocklength, \( n \), tends to infinity. The capacity region, \( C \), is the closure of the set of all achievable rate pairs.

Given a pair of functions \( f, g : Z^+ \to Z^+ \), we say \( f = o(g) \) if

\[
\lim_{m \to \infty} \frac{f(m)}{g(m)} = 0.
\]

We say \( f = \omega(g) \) if \( g = o(f) \).
Consider a sequence of DM-MACs with cooperation rate but also asymptotically larger than any polynomial function of a sequence of DM-MACs with properties that are useful rate.

(Note that in the above theorem, other than the conditions \( \mathcal{X}_1(m) = \mathcal{X}_2(m) = \{1, \ldots, 2^m\} \) and output alphabet \( \mathcal{Y}(m) = (\mathcal{X}_1(m) \times \mathcal{X}_2(m)) \cup \{ (E, E) \} \), where “E” denotes an erasure symbol. For each \( (x_1, x_2, y) \in \mathcal{X}_1(m) \times \mathcal{X}_2(m) \times \mathcal{Y}(m) \), \( p(m)(y|x_1, x_2) \) is defined as a function of the corresponding entry \( b_{x_1, x_2} \) of a binary matrix \( B(m) = (b_{ij})_{i,j=1} \). Precisely,

\[
p(m)(y|x_1, x_2) = \begin{cases} 1 - b_{x_1, x_2}, & y = (x_1, x_2) \\ b_{x_1, x_2}, & y = (E, E). \end{cases}
\]

That is, when \( (x_1, x_2) \) is transmitted, \( y = (x_1, x_2) \) is received if \( b_{x_1, x_2} = 0 \), and \( y = (E, E) \) is received if \( b_{x_1, x_2} = 1 \). Thus, we can interpret the 0 and 1 entries of \( B(m) \) as “good” and “bad” entries, respectively. We define the sets

\[
0_{B(m)} = \{(i, j) : b_{ij} = 0\}, \quad 1_{B(m)} = \{(i, j) : b_{ij} = 1\}
\]

to be the set of good and bad entries of \( \mathcal{X}(m) \times \mathcal{X}(m) \), respectively. To simplify notation, we drop \( m \) as a superscript when it is fixed.

For every \( S, T \subseteq \mathcal{X} \), let \( B_{S,T} \) be the submatrix obtained from \( B \) by keeping the rows with indices in \( S \) and columns with indices in \( T \). For every \( x \in \mathcal{X} \), let \( B_{x,T} = B_{\{x\},T} \) and \( B_{S,x} = B_{S,\{x\}} \).

The proof of Theorem \ref{main} requires that \( B \) satisfies two properties. One is that every sufficiently large submatrix of \( B \) should have a large fraction of bad entries. This property ensures that the sum-capacity of our channel without cooperation is small (Section \ref{sec:main}). The second property is that every submatrix of a specific type should have at least one good entry. This property enables a significantly higher sum-capacity under low-rate cooperation using the cooperation facilitator model (Section \ref{sec:main}). Lemma 2 demonstrates that these two properties can be simultaneously achieved. A proof of this lemma, based on the probabilistic method, is given in the appendix.

Let \( f, g : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+ \) be two functions such that \( f(m) = \omega(m) \) and \( g(m) = \log m + \omega(1) \). Then for every \( \epsilon > 0 \), there exists a sequence of \( (0, 1) \)-matrices \( \{B(m) = (b_{ij})_{i,j=1} \}_m \) such that

1. for every \( S, T \subseteq \mathcal{X}(m) \) such that \( |S|, |T| \geq f(m) \),

\[
\frac{|(S \times T) \cap 1_{B(m)}|}{|S||T|} \geq 1 - \epsilon,
\]

In the next section (Section \ref{sec:main}), we prove the existence of a sequence of DM-MACs with properties that are useful for proving Theorem \ref{main}. In Section \ref{sec:main}, we show that for the sequence of channels of Section \ref{sec:main}

\[
2m - C(m) \leq C_{S-\text{IE}}(m) \leq 2m.
\]

IV. CHANNEL CONSTRUCTION

For a fixed positive integer \( m \), the channel

\[
\left( \mathcal{X}(m) \times \mathcal{X}(m), p(m)(y|x_1, x_2), \mathcal{Y}(m) \right)
\]

used in the proof of Theorem \ref{main} has input alphabets \( \mathcal{X}_1(m) = \mathcal{X}_2(m) = \{1, \ldots, 2^m\} \) and output alphabet \( \mathcal{Y}(m) = (\mathcal{X}_1(m) \times \mathcal{X}_2(m)) \cup \{ ((E, E)) \} \), where “E” denotes an erasure symbol.
that is, in every sufficiently large submatrix of $B^{(m)}$, the fraction of bad entries is larger than $1 - \epsilon$, and

(2) for every $x \in \mathcal{X}^{(m)}$, and $k \in \{0, 1, \ldots, 2^{m-g(m)} - 1\}$, both $B_{x_k(x_k)}^{(m)}$ and $B_{x_k^{(m)}}$ each contain at least one good entry, where

$$\mathcal{X}^{(m)}_k = \left\{ k2^g(m) + l | l = 1, \ldots, 2^g(m) \right\}.$$

Let $f, g$ be two functions that satisfy the properties of Lemma 2 and the additional constraint $\log f(m) = o(m)$. For the next two sections, fix a sequence of channels as defined by (2) using matrices $B^{(m)}$ such that the $m$th matrix in the sequence satisfies the properties proved possible in Lemma 2 with respect to the given functions $f$ and $g$.

V. INNER AND OUTER BOUNDS FOR THE CF CAPACITY REGION

For the $m$th channel, we show the achievability of the rate pairs $(m, m - g(m))$ and $(m - g(m), m)$, with cooperation rate $C^{(m)} = g(m)$, by employing blocklength-1 codes ($n = 1$). Employing time sharing between these codes results in an inner bound for the capacity region given by

$$R_1, R_2 \leq m,$$

$$R_1 + R_2 \leq 2m - g(m).$$

If $R_1 = m$, $R_2 = m - g(m)$, and $n = 1$, then the independent, uniformly distributed messages $W_1$ and $W_2$ have alphabets $\mathcal{W}_1 = \{1, \ldots, 2^m\}$ and $\mathcal{W}_2 = \{1, \ldots, 2^{n-g(m)}\}$, respectively. By the second property of our channel in Lemma 2, for every $w_1 \in \mathcal{W}_1$ and $w_2 \in \mathcal{W}_2$, the submatrix $B_{w_1, w_2-1}$ contains at least one good entry. Let $z = \phi(w_1, w_2)$, the output of the cooperation facilitator, be an element of $Z = \{1, \ldots, 2^g(m)\}$ such that $(w_1, (w_2 - 1)2^g(m) + z)$ is a good entry of $B_{w_1, w_2-1}$. If there’s more than one good entry, we pick the one that results in the smallest $z$.

Suppose encoder 1 sends $x_1 = w_1$ and encoder 2 sends $x_2 = (w_2 - 1)2^g(m) + z$. Then by the definition of our channel (2), the channel output is $y = (x_1, x_2)$ with probability one, and hence zero error decoding is possible. Thus the rate pair $(m, m - g(m))$ is achievable. Note that for this achievable scheme to work, only the second encoder needs to know the value of $z$. A similar argument proves the achievability of $(m - g(m), m)$.

To find an outer bound for the capacity region, we use the capacity region of the CE model in a special case. Consider the situation that encoder 1 has access to both messages and can transmit information to encoder 2 on a noiseless link of capacity $C^{(m)}$. Then it is easy to see that the capacity region of this network contains the capacity region of the CF model. This situation, however, is the same as the CE model for $C_{12} = C^{(m)}$ and $C_{21} = \infty$. Hence an outer bound for the capacity region is given by the set of all rate pairs $(R_1, R_2)$ such that

$$R_1 \leq I(X_1; Y|X_2, U) + C(m),$$

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

for some distribution $p(u)p(x_1|u)p(x_2|u)$. Note that

$$I(X_1; Y|X_2, U) \leq H(X_1) \leq m,$$

$$I(X_1, X_2; Y) \leq H(X_1, X_2) \leq 2m,$$

and $C^{(m)} = g(m)$, so the region

$$R_1 \leq m + g(m),$$

$$R_1 + R_2 \leq 2m$$

is an outer bound for the CF model. Note that if we switch the roles of encoders 1 and 2, we get the outer bound

$$R_2 \leq m + g(m),$$

$$R_1 + R_2 \leq 2m.$$

Since the intersection of two outer bounds is also an outer bound, the set of all rate pairs $(R_1, R_2)$ such that

$$R_1, R_2 \leq m + g(m),$$

$$R_1 + R_2 \leq 2m$$

is an outer bound for the CF model as well.

VI. INNER AND OUTER BOUNDS FOR THE IE CAPACITY REGION

Consider the $m$th channel of the construction in Section IV. In the case where there is no cooperation, we show that the set of all rate pairs $(R_1, R_2)$ satisfying

$$R_1 + R_2 \leq m - g(m)$$

is an inner bound for the capacity region. To this end, we show the achievability of the rate pairs $(m - g(m), 0)$ and $(0, m - g(m))$. The achievability of all other rate pairs in the inner bound follow by time-sharing between the encoders. Similar to the previous section, let $n = 1$. Then $W_1 \in \{1, \ldots, 2^m - g(m)\}$ and $W_2 = 1$ with probability one. By our channel construction, for every $w \in \mathcal{W}_1$, $B_{w, w-1}$ contains at least one good entry. This means that the first column of $B^{(m)}$ contains at least $|\mathcal{W}_1| = 2^{m-g(m)}$ good entries. Suppose encoder 1 transmits uniformly on these $2^{m-g(m)}$ good entries and encoder 2 transmits $x_2 = 1$. Then the input is always on a good entry and the channel output is the same as the channel input. Thus the pair $(m - g(m), 0)$ is achievable. A similar argument shows that the pair $(0, m - g(m))$ is achievable and the inner bound follows. We next find an outer bound for the IE capacity region.

Let $Y_1$ and $Y_2$ be the components of $Y$; that is, if $Y = (x_1, x_2)$, then $Y_1 = x_1$ and $Y_2 = x_2$, and if $Y = (E, E)$, then $Y_1 = Y_2 = E$. Note that $Y_1, Y_2 \in \mathcal{X} \cup \{E\}$. In the case of independent encoders, $X_1$ and $X_2$ are independent, and the distribution of $Y_1$ is given by

$$p(y_1) = \begin{cases} \gamma y_1 & y_1 \in \mathcal{X}, \\ 1 - \gamma & y_1 = E, \end{cases}$$

where

$$\gamma x_1 = p(x_1) \sum_{x_2 : b_{x_1, x_2} = 0} p(x_2),$$

Note that
for every $x_1 \in X$, and $\gamma = \sum_{x_1} \gamma_{x_1}$. The capacity region for the IE model (no cooperation) is due to Ahlswede [1,2] and Liao [12]. If $\mathcal{A}^{(m)}$ is the set of all pairs $(R_1, R_2)$ such that

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

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

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

(4)

for some distribution $p(x_1)p(x_2)p(y|x_1, x_2)$, then the capacity region is given by $\text{conv}(\mathcal{A}^{(m)})$, where $\text{conv}(\mathcal{A}^{(m)})$ denotes the convex hull of $\mathcal{A}^{(m)}$. For our channel, $Y$, $Y_1$, and $Y_2$ are deterministic functions of $(X_1, X_2)$, $(X_1, Y_2)$ and $(Y_1, X_2)$, respectively, and the bounds simplify as

$$
R_1 \leq I (X_1; Y | X_2) = H (Y_1 | X_2) \leq H (Y_1),
$$

$$
R_2 \leq I (X_2; Y | X_1) = H (Y_2 | X_1) \leq H (Y_2).
$$

(5)

To bound $H (Y_1)$, we apply the following lemma, proved in the appendix. This lemma bounds the probability that a random variable $X$ falls in a specific set $T$; the bound is given as a function of the entropy of $X$ and the cardinality of $T$. For any set $T$, we denote its characteristic function by $\chi_T$.

**Lemma 3.** Let $X$ be a discrete random variable with values in $X$, and let $T$ be a subset of $X$. If $q : T \to \mathbb{R}_{\geq 0}$ is a function and $\alpha = \sum_{x \in T} q(x)$, then

$$
-\sum_{x \in T} q(x) \log q(x) \leq \alpha \log |T| - \alpha \log \alpha.
$$

(6)

In the case where $q(x) = p(x)\chi_T(x)$, the above inequality implies

$$
\Pr (X \in T) \leq K \left( 1 - \frac{H (X) - 1}{\log |X|} \right),
$$

(7)

where $K = \left( 1 - \frac{\log |T|}{\log |X|} \right)^{-1}$.

By (3),

$$
H (Y_1) = -\sum_{x_1} \gamma_{x_1} \log \gamma_{x_1} - (1 - \gamma) \log (1 - \gamma).
$$

Applying (6) from Lemma 3

$$
H (Y_1) \leq \gamma m + H (\gamma) \leq \gamma m + 1.
$$

(8)

We next bound $H_1$. To this end, we write each distribution $(p(x_1)$ and $p(x_2))$ as a particular convex combination of uniform distributions as stated in the next lemma.

**Lemma 4.** If $X$ is a discrete random variable with a finite alphabet $X$, then there exists a positive integer $k$, a sequence of positive numbers $\{\alpha_j\}_{j=1}^k$, and a sequence of non-empty subsets of $X$, $\{S_j\}_{j=1}^k$ such that the following properties are satisfied.

(a) For every $j$, $1 \leq j \leq k$, $S_{j+1} \subseteq S_j$.

(b) $\sum_{j=1}^k \alpha_j = 1$.

(c) For all $x \in X$,

$$
p (x) = \sum_{j=1}^k \alpha_j \frac{\chi_{S_j} (x)}{|S_j|}.
$$

(d) For every $C, 0 < C < |X|$,.

$$
\sum_{j: |S_j| \leq C} \alpha_j \leq K \left( 1 - \frac{H (X) - 1}{\log |X|} \right),
$$

where $K = \left( 1 - \frac{\log C}{\log |X|} \right)^{-1}$.

The proof of this lemma is given in the appendix. Let

$$
p (x_1) = \sum_{i=1}^k \alpha_{i(1)} \frac{\chi_{S_{i(1)}^{(1)}} (x_1)}{|S_{i(1)}^{(1)}|},
$$

$$
p (x_2) = \sum_{j=1}^l \alpha_{j(2)} \frac{\chi_{S_{j(2)}^{(2)}} (x_2)}{|S_{j(2)}^{(2)}|}
$$

be the decompositions of $p (x_1)$ and $p (x_2)$ according to Lemma 4. Then

$$
\gamma = \sum_{x_1, x_2 : d_{x_1, x_2} = 0} p (x_1)p (x_2)
$$

$$
= \sum_{i=1}^k \sum_{j=1}^l \alpha_{i(1)} \alpha_{j(2)} \beta_{ij},
$$

where

$$
\beta_{ij} = \sum_{x_1, x_2 : d_{x_1, x_2} = 0} \frac{\chi_{S_{i(1)}^{(1)}} (x_1) \chi_{S_{j(2)}^{(2)}} (x_2)}{|S_{i(1)}^{(1)}||S_{j(2)}^{(2)}|}
$$

$$
= \left| \left( S_{i(1)}^{(1)} \times S_{j(2)}^{(2)} \right) \cap 0_B \right| 
$$

$$
\leq \frac{|X|}{|S_{i(1)}^{(1)}||S_{j(2)}^{(2)}|}
$$

(9)

For every $i$ and $j$, $\beta_{ij} \leq 1$. If, however, $\min \{|S_{i(1)}^{(1)}|, |S_{j(2)}^{(2)}|\} \geq f (m)$, then by the first property of our channel (Lemma 2), $\beta_{ij} \leq \epsilon$. Thus by part (d) of Lemma 4

$$
\gamma < \epsilon + \sum_{i, j: \min \{|S_{i(1)}^{(1)}|, |S_{j(2)}^{(2)}|\} < f (m)} \alpha_{i(1)} \alpha_{j(2)}
$$

$$
= \epsilon + 1 - \sum_{i, j: \min \{|S_{i(1)}^{(1)}|, |S_{j(2)}^{(2)}|\} \geq f (m)} \alpha_{i(1)} \alpha_{j(2)}
$$

$$
= \epsilon + 1 - \left( 1 - \sum_{i: |S_{i(1)}^{(1)}| < f (m)} \alpha_{i(1)} \right) \left( 1 - \sum_{j: |S_{j(2)}^{(2)}| < f (m)} \alpha_{j(2)} \right)
$$

$$
\leq \epsilon + 1 - \left( 1 - K_m \left( 1 - \frac{H (X) - 1}{m} \right) \right) 
$$

$$
\times \left( 1 - K_m \left( 1 - \frac{H (X) - 1}{m} \right) \right),
$$

where

$$
K_m = \left( 1 - \frac{\log f (m)}{m} \right)^{-1}.
$$

Note that $K_m \to 1$ as $m \to \infty$ since $\log f (m) = o (m)$ by assumption. Since by (4), $R_i \leq$
Combining the previous inequality with (5) and (8) gives
\[
\frac{R_1}{m} \leq \epsilon + \frac{1}{m} + K_m \left( \frac{2 - \frac{R_1 + R_2 - 2}{m}}{m} \right) - K^2_m \left( \frac{1 - \frac{R_1 - 1}{m}}{m} \right) \left( \frac{1 - \frac{R_1 - 1}{m}}{m} \right). 
\]
If we let \( x = \frac{R_1}{m} \) and \( y = \frac{R_2}{m} \), then
\[
x \leq \epsilon + \frac{1}{m} + K_m \left( \frac{2 + \frac{2}{m} - x - y}{m} \right) - K^2_m \left( \frac{1 + \frac{2}{m} - x - y + \frac{x - 1}{m} \left( y - \frac{1}{m} \right) }{m} \right),
\]
or
\[
(x - a_m)(y + b_m) \leq c_m, \tag{9}
\]
where
\[
a_m = 1 + \frac{1}{m} - \frac{1}{K_m},
\]
\[
b_m = -1 - \frac{1}{m} + \frac{1}{K_m} + \frac{1}{K^2_m},
\]
\[
c_m = -1 - \frac{2}{m} - \frac{1}{m^2} + \left( \frac{2 + \frac{2}{m}}{K_m} \right) \frac{1}{m} + \left( \epsilon + \frac{1}{m} \right) \frac{1}{K^2_m} - a_m b_m.
\]
By symmetry, we can also show
\[
(x + b_m)(y - a_m) \leq c_m. \tag{10}
\]
As the capacity region is given by \( \text{conv}(\mathcal{S}^{(m)}) \), the definition of sum-capacity (1) implies
\[
\frac{1}{m} C_{S-\text{IE}}^{(m)} \leq \frac{1}{m} \max_{(R_1, R_2) \in \text{conv}(\mathcal{S}^{(m)})} (R_1 + R_2) = \max_{(x, y) \in \text{conv}(S^{(m)})} (x + y).
\]
Thus
\[
\lim_{m \to \infty} \frac{C_{S-\text{IE}}^{(m)}}{m} \leq \lim_{m \to \infty} \max_{(x, y) \in \text{conv}(S^{(m)})} (x + y). \tag{11}
\]
To find the limit on the right side, we make use of the following lemma proved in the appendix.

**Lemma 5.** Suppose \( \{a_m\}_{m=1}^{\infty}, \{b_m\}_{m=1}^{\infty} \) and \( \{c_m\}_{m=1}^{\infty} \) are three sequences of real numbers such that \( \lim_{m \to \infty} a_m = a, \lim_{m \to \infty} b_m = b, \lim_{m \to \infty} c_m = c \), where
\[
b, c, a + b, ab + c > 0,
\]
and
\[
\sqrt{(a + b)^2 + 4c} > b + \frac{c}{b}.
\]
For every positive integer \( m \), let \( S^{(m)} \) be defined as above. Then
\[
\lim_{m \to \infty} \max_{(x, y) \in \text{conv}(S^{(m)})} (x + y) = a + \sqrt{(a + b)^2 + 4c}.
\]
It is easy to see that the sequences above satisfy the assumptions of Lemma 5. Thus
\[
\lim_{m \to \infty} \frac{C_{S-\text{IE}}^{(m)}}{m} \leq \sqrt{b^2 + 4c} - 1 < \frac{2}{\phi} + \epsilon,
\]
where \( \phi = \frac{1 + \sqrt{5}}{2} \). Therefore for all but finitely many \( m \),
\[
C_{S-\text{IE}}^{(m)} \leq \left( \frac{2}{\phi} + \epsilon \right) m.
\]

**VII. Conclusion**

In this paper, we present a new model for cooperation and study its benefits in the case of the encoders of a DM-MAC. Specifically, we present channels for which the gain in sum-capacity is “nearly” exponential in the cooperation rate. The CF model can be generalized to other network settings, and its study is subject to future work.

**Appendix A**

**Proof of Lemma 2**

We use the probabilistic method. We assign a probability to every \( 2^m \times 2^m \) (0,1)-matrix and show that the probability of a matrix having both properties is positive for sufficiently large \( m \); hence, there exists at least one such matrix. Fix \( \epsilon \geq 0 \), and let \( B = (b_{ij})_{i,j=1}^{2^m} \) be a random matrix with \( b_{ij} \overset{iid}{\sim} \text{Bernoulli}(p) \), where \( 1 - \epsilon < p < 1 \). Let
\[
\Gamma = \{ S : S \subseteq X, |S| \geq f(m) \}.
\]
For every $S, T \in \Gamma$, define the event
\[ E_{S,T} = \left\{ \frac{|(S \times T) \cap 1_B|}{|S||T|} \leq 1 - \epsilon \right\} . \]

It follows
\[
\Pr \left( \bigcup_{S,T \in \Gamma} E_{S,T} \right) 
\leq \sum_{S,T \in \Gamma} \Pr \left( E_{S,T} \right) 
= \sum_{S,T \in \Gamma} \Pr \left( |(S \times T) \cap 1_B| \leq (1 - \epsilon) |S||T| \right) 
= \sum_{S,T \in \Gamma} \sum_{k=0}^{[1-(\epsilon)|S||T|]} \binom{|S||T|}{k} p^k (1-p)^{|S||T|-k} 
= \sum_{i,j=1}^{2^{m}} e^{(i+j)m \ln 2 - 2(p-1+\epsilon)^2 i j} 
\leq (2^m - f(m) + 1)^2 e^{2m f(m) \ln 2 - 2(p-1+\epsilon)^2 (f(m))^2} 
< e^{2m (1+f(m)) \ln 2 - 2(p-1+\epsilon)^2 (f(m))^2} 
= e^{2(f(m))^2 ((1+\frac{1}{m}) \frac{m}{m} \ln 2 - 2(p-1+\epsilon)^2)} .
\]

The exponent of the right hand side of the previous inequality
\[ 2 (f(m))^2 \left( 1 + \frac{1}{f(m)} \right) \frac{m}{f(m)} \ln 2 - 2(p-1+\epsilon)^2 \]
go to $-\infty$ as $m \to \infty$, so as $m$ tends to infinity
\[
\Pr \left( \bigcup_{S,T \in \Gamma} E_{S,T} \right) \to 0 .
\]

That is, the probability that the fraction of bad entries in a sufficiently large submatrix is less than $1-\epsilon$ is going to zero.

We next calculate the probability that $B$ doesn’t satisfy the second property,
\[
\Pr \left( \exists x \in X, k \in [1, 2m-g(m)] : (B_{x,x_k} \cup B_{x_k,x}) \cap 0_B = \emptyset \right) 
\leq \sum_{x \in X} \sum_{k=1}^{2m-g(m)} \Pr \left( B_{x,x_k} \cap 0_B = \emptyset \right) 
+ \sum_{x \in X} \sum_{k=1}^{2m-g(m)} \Pr \left( B_{x_k,x} \cap 0_B = \emptyset \right) 
= 2^{2m-g(m)+1} p^{2g(m)} 
= 2^{2g(m)} \left( \frac{2m-g(m)+1}{2g(m)} + \log p \right) .
\]

Since $m = o(2g(m))$, the exponent of the right hand side of the previous inequality
\[ 2g(m) \left( \frac{2m-g(m)+1}{2g(m)} + \log p \right) \]
go to $-\infty$ as $m \to \infty$, which implies that the left hand side
\[
\Pr \left( \exists x \in X, k \in [1, 2m-g(m)] : (B_{x,x_k} \cup B_{x_k,x}) \cap 0_B = \emptyset \right) 
\]
go to zero as $m \to \infty$. Thus, by the union bound the probability that the matrix doesn’t satisfy either of these properties is going to zero. Therefore, for large enough $m$, almost every $(0,1)$-matrix satisfies both of these properties, though we only need one such matrix.
APPENDIX B

PROOF OF LEMMA 3

For the first part, if \( \alpha = 0 \), then \( \log \frac{q(x)}{\alpha} = 0 \) for every \( x \in T \) and both sides equal zero. Otherwise,

\[
- \sum_{x \in T} q(x) \log q(x) = - \alpha \sum_{x \in T} \frac{q(x)}{\alpha} \log \frac{q(x)}{\alpha} - \alpha \log \alpha
\]

\[
\leq \alpha \log |T| - \alpha \log \alpha,
\]

since \( \frac{q(x)}{\alpha} \) is a probability mass function with entropy \( \sum_{x \in T} \frac{q(x)}{\alpha} \log \frac{1}{\alpha} \) and alphabet size \( |T| \).

For the second part, if

\[
q(x) = p(x) \chi_T(x),
\]

then by the previous inequality,

\[
- \sum_{x \in T} p(x) \log p(x) = - \sum_{x \in T} q(x) \log q(x)
\]

\[
\leq \alpha \log |T| - \alpha \log \alpha,
\]

where

\[
\alpha = \sum_{x \in T} q(x) = \Pr(x \in T).
\]

Similarly, replacing \( \mathcal{X} \setminus T \) with \( T \) results in

\[
- \sum_{x \in \mathcal{X} \setminus T} p(x) \log p(x)
\]

\[
\leq (1 - \alpha) \log |\mathcal{X} \setminus T| - (1 - \alpha) \log (1 - \alpha).
\]

Adding the previous two inequalities gives

\[
H(X) \leq \alpha \log |T| + (1 - \alpha) \log |\mathcal{X} \setminus T| + H(\alpha)
\]

\[
\leq \alpha \log |T| + (1 - \alpha) \log |\mathcal{X}| + 1.
\]

Therefore,

\[
\frac{H(X)}{\log |\mathcal{X}|} \leq 1 + \frac{1}{\log |\mathcal{X}|} \left( 1 - \frac{\log |T|}{\log |\mathcal{X}|} \right) \alpha,
\]

and

\[
\alpha \leq \frac{1 - H(X) - 1}{1 - \frac{\log |T|}{\log |\mathcal{X}|}}.
\]

APPENDIX C

PROOF OF LEMMA 4

Let \( k \) be the cardinality of the range of \( p : \mathcal{X} \to \mathbb{R} \). Then there exists a sequence \( \{x_j\}_{j=1}^{k} \) such that

\[
\{p(x)|x \in \mathcal{X}\} = \{p(x_j) | 1 \leq j \leq k\},
\]

and

\[
0 < p(x_1) < \cdots < p(x_k) \leq 1.
\]

For \( j, 1 \leq j \leq k \), define

\[
S_j = \{x \in \mathcal{X} | p(x) \geq p(x_j)\},
\]

and let \( S_{k+1} = \emptyset \). Then for \( j, 1 \leq j \leq k \), \( S_{j+1} \subseteq S_j \) (part a) and

\[
S_j \setminus S_{j+1} = \{x \in \mathcal{X} | p(x) = p(x_j)\}.
\]

Thus the number of \( x \in \mathcal{X} \) such that \( p(x) = p(x_j) \) equals \( |S_j \setminus S_{j+1}| \). For \( j \in \{2, \ldots, k\} \), define

\[
\alpha_j = |S_j| (p(x_j) - p(x_{j-1})),
\]

and let \( \alpha_1 = |S_1| p(x_1) \). To show (b), note that

\[
\sum_{j=1}^{k} \alpha_j = \sum_{j=1}^{k} |S_j \setminus S_{j+1}| p(x_j)
\]

\[
= \sum_{x \in \mathcal{X}} p(x) = 1.
\]

In part (c), the left hand side simplifies as

\[
\sum_{j=1}^{k} \frac{\alpha_j \chi_{S_j}(x)}{|S_j|} = \sum_{j=1}^{k} (p(x_j) - p(x_{j-1})) \chi_{S_j}(x)
\]

\[
= \sum_{j=1}^{k} p(x_j) \chi_{S_j \setminus S_{j+1}}(x)
\]

\[
= p(x).
\]

In (d), if the set \( \{j|1 \leq j \leq k, |S_j| \leq C\} \) is empty, then there’s nothing to prove. Otherwise, it’s a nonempty subset of \( \{1, \ldots, k\} \) and thus has a minimum, which we call \( j^* \). Then

\[
\sum_{j:|S_j| \leq C} \alpha_j = \sum_{j=j^*}^{k} \alpha_j
\]

\[
= \sum_{j=j^*}^{k} |S_j| (p(x_j) - p(x_{j-1}))
\]

\[
= \sum_{j=j^*}^{k} |S_j \setminus S_{j+1}| p(x_j) - |S_{j^*}| p(x_{j^*+1})
\]

\[
= \sum_{x \in S_{j^*}} p(x) - |S_{j^*}| p(x_{j^*+1})
\]

\[
\leq \sum_{x \in S_{j^*}} p(x).
\]

By (4) of Lemma 3

\[
\sum_{x \in S_{j^*}} p(x) \leq \frac{1}{1 - \frac{\log |S_{j^*}|}{\log |\mathcal{X}|}} \left( 1 - \frac{H(X) - 1}{\log |\mathcal{X}|} \right)
\]

\[
\leq \frac{1}{1 - \frac{\log C}{\log |\mathcal{X}|}} \left( 1 - \frac{H(X) - 1}{\log |\mathcal{X}|} \right),
\]

since \( |S_{j^*}| \leq C \).

APPENDIX D

PROOF OF LEMMA 5

Prior to proving Lemma 5, we state and prove the following lemma.

**Lemma 6.** Suppose \( a, b, \) and \( c \) are three real numbers such that \( b, c, a + b, \) and \( ab + c \) are positive, and

\[
\sqrt{(a+b)^2 + 4c} > b + \frac{c}{b}.
\]
Let $S$ be the set of all pairs $(x, y)$ such that $x, y \geq 0$, and
\[
\begin{cases}
(x - a)(y + b) \leq c, \\
(x + b)(y - a) \leq c.
\end{cases}
\]
If $x_0$ is the unique positive solution to the equation

\[(x - a)(x + b) = c,
\]
then

\[\max_{(x,y) \in \text{conv}(S)} (x + y) = 2x_0.\]

**Proof:** Since
\[(x - a)(y + b) - (x + b)(y - a) = (a + b)(x - y),\]
and $a + b > 0$, the set $S$ can be written as $S = S_1 \cup S_2$ (see Figure 3), where $S_1$ is the set of all pairs $(x, y)$ such that $0 \leq x \leq y$ and

\[(x + b)(y - a) \leq c,
\]
and $S_2$ is the set of all pairs $(x, y)$ such that $0 \leq y \leq x$ and

\[(x - a)(y + b) \leq c.
\]
The intersection of $S_1$ and $S_2$ consists of all pairs $(x, x)$ such that $0 \leq x \leq x_0$ where

\[x_0 = \frac{a - b + \sqrt{(a + b)^2 + 4c}}{2}.
\]

Note that since $c$ and $ab + c$ are positive, $0 < x_0 < a + \frac{c}{b}$. The convex hull of $S_1$ consists of all pairs $(x, y)$ such that $0 \leq x \leq y$ and

\[\left(a + \frac{c}{b} - x_0\right)x + x_0y \leq \left(a + \frac{c}{b}\right)x_0,
\]
and the convex hull of $S_2$ consists of all pairs $(x, y)$ such that $0 \leq y \leq x$ and

\[x_0x + \left(a + \frac{c}{b} - x_0\right)y \leq \left(a + \frac{c}{b}\right)x_0.
\]

Note that $\text{conv}(S_1) \cup \text{conv}(S_2)$ is the region bounded by the axes $y = 0, x = 0$, and the piecewise linear function

\[h(x) = \begin{cases}
x_0 - a + \frac{c}{b}x + a + \frac{c}{b} & 0 \leq x \leq x_0, \\
x_0 - \frac{x_0}{a + \frac{c}{b}}x - x_0 - a + \frac{c}{b} & x_0 < x \leq a + \frac{c}{b}.
\end{cases}\]

Since $2x_0 \geq a + \frac{c}{b}$ by assumption,
\[\frac{x_0 - a - \frac{c}{b}}{x_0} \geq \frac{x_0}{x_0 - a - \frac{c}{b}}.
\]
This means the slope of $h$ is decreasing, or equivalently, $h$ is a concave function. Thus $\text{conv}(S_1) \cup \text{conv}(S_2)$ is convex. But

\[S \subseteq \text{conv}(S_1) \cup \text{conv}(S_2) \subseteq \text{conv}(S),
\]
so

\[\text{conv}(S) = \text{conv}(S_1) \cup \text{conv}(S_2).
\]

This implies

\[\max_{(x,y) \in \text{conv}(S)} (x + y) = 2x_0.\]

Using this lemma, we may prove Lemma 5. There exists an $M > 0$ such that for every $m \geq M$,

\[b_m, c_m, a_m + b_m, a_m b_m + c_m > 0,
\]
and

\[\sqrt{(a_m + b_m)^2 + 4c_m} - b_m - \frac{c_m}{b_m} > 0.
\]

Let $x_0^{(m)}$ and $x_0$ be the unique positive solutions to the equations

\[(x_0^{(m)} - a_m)(x_0^{(m)} + b_m) = c_m,
\]
and

\[(x_0 - a)(x_0 + b) = c.
\]

Since $x_0^{(m)}$ and $x_0$ are continuous functions of $(a_m, b_m, c_m)$ and $(a, b, c)$, respectively, we have

\[\lim_{m \to \infty} x_0^{(m)} = x_0.
\]

Thus by Lemma 6

\[\lim_{m \to \infty} \max_{(x,y) \in \text{conv}(S^{(m)})} (x + y) = \lim_{m \to \infty} 2x_0^{(m)} = 2x_0 = a - b + \sqrt{(a + b)^2 + 4c}.
\]

**REFERENCES**

[1] Rudolf Ahlswede, *Multi-way communication channels* (1973).

[2] ---, *The capacity region of a channel with two senders and two receivers*, The Annals of Probability (1974), 805–814.

[3] Ron Dabora and Sergio D Servetto, *Broadcast channels with cooperating decoders*, Information Theory, IEEE Transactions on 52 (2006), no. 12, 5438–5454.

[4] Abbas El Gamal and Young-Han Kim, *Network information theory*, Cambridge University Press, New York, NY, USA, 2012.

[5] Tracey Ho, Michelle Effros, and Shiri Jalali, *On equivalence between network topologies*, Communication, control, and computing (allerton), 2010 48th annual allerton conference on, 2010, pp. 391–398.

[6] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of American Statistical Association 58 (1963), no. 301, 13–30.

[7] Shiri Jalali, Michelle Effros, and Tracey Ho, *On the impact of a single edge on the network coding capacity*, Information theory and applications workshop (ita), 2011, 2011, pp. 1–5.

[8] Gerhard Kramer, Ivana Marić, and Roy D. Yates, *Cooperative communications*, Found. Trends Netw. 1 (August 2006), no. 3, 271–425.
[9] M Langberg and M Effros, *Source coding for dependent sources*, Information theory workshop (itw), 2012 ieee, 2012, pp. 70–74.

[10] Michael Langberg and Michelle Effros, *Network coding: Is zero error always possible?*, Communication, control, and computing (allerton), 2011 49th annual allerton conference on, 2011, pp. 1478–1485.

[11] Eun Jee Lee, Michael Langberg, and Michelle Effros, *Outer bounds and a functional study of the edge removal problem*, Information theory workshop (itw), 2013 ieee, 2013, pp. 1–5.

[12] H. Liao, *Multiple access channels*, Ph.D. Thesis, 1972.

[13] Ivana Maric, Roy D Yates, and Gerhard Kramer, *Capacity of interference channels with partial transmitter cooperation*, Information Theory, IEEE Transactions on 53 (2007), no. 10, 3536–3548.

[14] Haim H Permuter, Shlomo Shamai, and Anelia Somekh-Baruch, *Message and state cooperation in multiple access channels*, Information Theory, IEEE Transactions on 57 (2011), no. 10, 6379–6396.

[15] F Willems, *The discrete memoryless multiple access channel with partially cooperating encoders (corresp.*)*, Information Theory, IEEE Transactions on 29 (1983), no. 3, 441–445.