Every Bit Counts: Second-Order Analysis of Cooperation in the Multiple-Access Channel

Oliver Kosut, Michelle Effros, Michael Langberg

***Abstract*—The work at hand presents a finite-blocklength analysis of the multiple access channel (MAC) sum-rate under the cooperation facilitator (CF) model. The CF model, in which independent encoders coordinate through an intermediary node, is known to show significant rate benefits, even when the rate of cooperation is limited. We continue this line of study for cooperation rates which are sub-linear in the blocklength n. Roughly speaking, our results show that if the facilitator transmits $\log K$ bits, then there is a sum-rate benefit of order $\sqrt{\log K/n}$ compared to the best-known achievable rate. This result extends across a wide range of K: even a single bit of cooperation is shown to provide a sum-rate benefit of order $1/\sqrt{n}$.**

I. Introduction

The multiple access channel (MAC) model lies at an interesting conceptual intersection between the notions of cooperation and interference in wireless communications. When viewed from the perspective of any single transmitter, codewords transmitted by other transmitters can only inhibit the first transmitter's *individual* communication rate; thus each transmitter sees the others as a source of interference. When viewed from the perspective of the receiver, however, maximizing the *total* rate delivered to the receiver often requires all transmitters to communicate simultaneously; from the receiver's perspective, then, the transmitters must cooperate through their simultaneous transmissions to maximize the sum-rate delivered to the receiver.

Simultaneous transmission is, perhaps, the weakest form of cooperation imaginable in a wireless communication model. Nonetheless, the fact that even simultaneous transmission of independent codewords from interfering transmitters can increase the sum-rate deliverable to the MAC receiver begs the question of how much more could be achieved through more significant forms of MAC transmitter cooperation.

The information theory literature devotes considerable effort to studying the impact of encoder cooperation in the MAC. A variety of cooperation models are considered. Examples include the "conferencing" cooperation model [2], in which encoders share information directly in order to coordinate their channel inputs, the "cribbing" cooperation model [3], in which transmitters cooperate by sharing their codeword information (at times causally), and the "cooperation facilitator" (CF) cooperation model [4] in which users coordinate their channel inputs with the help of an intermediary called the CF. The CF distinguishes the amount of information that must be understood to facilitate cooperation (i.e., the rate $R_{\rm IN}$ *to* the CF) from the amount of information employed in the coordination (i.e., the rate $R_{\rm OUT}$ *from* the CF). Key results using the CF model show that for many MACs, no matter what the (non-zero) fixed rate $C_{\rm IN}$, the curve describing the maximal sum-rate as a function of $R_{\rm OUT}$ has infinite slope at $R_{\rm OUT} = 0$ [5]. That is, *very little coordination through a CF can change the MAC capacity considerably.* This phenomenon holds for both average and maximum error sum-rates; it is most extreme in the latter case, where even a finite number of bits (independent of the blocklength) — that is, $R_{\rm OUT} = 0$ — can suffice to change the MAC capacity region [6]–[8].

We study the CF model for 2-user MACs under the average error criterion. In this setting, the maximal sum-rate is a *continuous* function of $R_{\rm OUT}$ at $R_{\rm OUT} = 0$ [7], [8], implying that any sub-linear cooperation rate cannot increase the achievable *first-order* rates. However, sub-linear CF cooperation may still increase sum-rate, albeit through second or higher-order terms. In this work, we seek to understand the impact of the CF over a wide range of cooperation rates. Specifically, we consider a CF that, after viewing both messages, can transmit one of K signals to both transmitters. We prove achievable bounds that express the benefit of this cooperation as a function of K. These bounds extend all the way from constant K to exponential K. Interestingly, we find that even for $K = 2$ (i.e., one bit of cooperation), there is a benefit in the second-order (i.e., dispersion) term, corresponding to an improvement of $O(\sqrt{n})$ message bits compared to the best-known achievability bound. We prove two main achievable bounds, each of which is better for a different range of K values. The proof of the first bound is based on refined asymptotic analysis similar to typical second-order bounds. The proof of the second bound is based on the method of types. For a wide range of K values, we find that the benefit is $O(\sqrt{n \log K})$ message bits. We have no matching converse — indeed, even without cooperation the second-order rate of the MAC is open — so in all cases we compare against the best known achievability bounds.

II. Problem Setup

An (M_1, M_2, K) facilitated multiple access code for multiple access channel (MAC)

$$(\mathcal{X}_1 \times \mathcal{X}_2, p_{Y|X_1,X_2}(y|x_1,x_2), \mathcal{Y})$$

O. Kosut is with the School of Electrical, Computer and Energy Engineering at Arizona State University. Email: okosut@asu.edu

M. Effros is with the Department of Electrical Engineering at the California Institute of Technology. Email: effros@caltech.edu

M. Langberg is with the Department of Electrical Engineering at the University at Buffalo (State University of New York). Email: mikel@buffalo.edu

This work is supported in part by NSF grants CCF-1817241, CCF-1908725, and CCF-1909451. The full version of this work appears in [1].

is defined by a facilitator code

$$e: \quad [M_1] \times [M_2] \to [K]$$

a pair of encoders

$$f_1: \quad [M_1] \times [K] \to \mathcal{X}_1$$
$$f_2: \quad [M_2] \times [K] \to \mathcal{X}_2$$

and a decoder

$$g: \mathcal{Y} \to [M_1] \times [M_2].$$

We assume that all alphabets are finite. The encoder's output is sometimes described using the abbreviated notation

$$\begin{aligned} X_1(m_1, m_2) &= f_1(m_1, e(m_1, m_2)) \\ X_2(m_1, m_2) &= f_2(m_2, e(m_1, m_2)). \end{aligned}$$

The average error probability for the given code is

$$P_e = \frac{1}{M_1 M_2} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \Pr\big(g(Y) \neq (m_1, m_2) \big| (X_1, X_2) = (X_1(m_1, m_2), X_2(m_1, m_2))\big).$$

We also consider codes for the n-length product channel, where $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}$ are replaced by $\mathcal{X}_1^n, \mathcal{X}_2^n, \mathcal{Y}^n$ respectively, and

$$p_{Y^n|X_1^n,X_2^n}(y^n|x_1^n, x_2^n) = \prod_{i=1}^{n} p_{Y|X_1,X_2}(y_i|x_{1i}, x_{2i}).$$

An (M_1, M_2, K) code for the n-length channel achieving average probability of error at most ϵ is called an $(n, M_1, M_2, K, \epsilon)$ code. The fundamental limit for the sum-rate, which will be our primary interest, is given by

$$R_{\text{sum}}(n, \epsilon, K) = \sup\left\{ \tfrac{\log(M_1 M_2)}{n} : \exists (n, M_1, M_2, K, \epsilon) \text{ code} \right\}.$$

The following notation will be useful. Given a MAC, the sum-capacity without cooperation is given by

$$C_{\text{sum}} = \max_{p_{X_1} p_{X_2}} I(X_1, X_2; Y). \tag{1}$$

Let $\mathcal{P}^\star$ be the set of product distributions $p_{X_1} p_{X_2}$ achieving the maximum in (1). For any $p_{X_1} p_{X_2} \in \mathcal{P}^\star$, let p_Y be the resulting marginal on the channel output, giving

$$p_Y(y) = \sum_{(x_1,x_2) \in \mathcal{X}_1 \times \mathcal{X}_2} p_{X_1}(x_1) p_{X_2}(x_2) p_{Y|X_1,X_2}(y|x_1, x_2)$$

for all $y \in \mathcal{Y}$. We use $i(x_1, x_2; y)$, $i(x_1; y|x_2)$ and $i(x_2; y|x_1)$ to represent the joint and conditional information densities

$$\begin{aligned} i(x_1, x_2; y) &= \log\left(\frac{p_{Y|X_1,X_2}(y|x_1, x_2)}{p_Y(y)}\right) \\ i(x_1; y|x_2) &= \log\left(\frac{p_{Y|X_1,X_2}(y|x_1, x_2)}{p_{Y|X_2}(y|x_2)}\right) \\ i(x_2; y|x_1) &= \log\left(\frac{p_{Y|X_1,X_2}(y|x_1, x_2)}{p_{Y|X_1}(y|x_1)}\right), \end{aligned}$$

where $p_{Y|X_1}$ and $p_{Y|X_2}$ are conditional marginals on Y under joint distribution $p_{X_1,X_2,Y} = p_{X_1} p_{X_2} p_{Y|X_1,X_2}$. We denote the 3-vector of all three quantities as

$$\mathbf{i}(x_1, x_2; y) = \begin{bmatrix} i(x_1, x_2; y) \\ i(x_1; y|x_2) \\ i(x_2; y|x_1) \end{bmatrix}.$$

It will be convenient to define

$$\begin{aligned} i(x_1, x_2) &= E[i(x_1, x_2; Y)|(X_1, X_2) = (x_1, x_2)] \\ &= D(p_{Y|X_1=x_1,X_2=x_2} \| p_Y). \end{aligned}$$

Let

$$V_1 = \text{Var}(i(X_1, X_2)), \tag{2}$$
$$V_2 = E[\text{Var}(i(X_1, X_2; Y)|X_1, X_2)]. \tag{3}$$

Roughly speaking, V_1 represents the information-variance of the codewords, whereas V_2 represents the information-variance of the channel noise. Given two distributions p_X, q_X, let the divergence-variance be

$$V(p_X \| q_X) = \text{Var}_{p_X}\left(\log \frac{p_X(X)}{q_X(X)}\right).$$

III. Main Results

In the literature on second-order rates, there are typically two types of results: (i) finite blocklength results, with no asymptotic terms, that are typically written in terms of abstract alphabets, and (ii) asymptotic results that derive from these finite blocklength results, which are typically easier to understand. The following is an achievable result which has some flavor of both: the channel noise is dealt with via an asymptotic analysis, but the dependence on the randomness in the codewords is written as in a finite blocklength result. We provide this "intermediate" result because, depending on the CF parameter K, the relevant aspect of the codeword distribution may be in the central limit, moderate deviations, or large deviations regime. Thus, in this form one may plug in any concentration bound to derive an achievable bound. Subsequently, Theorem 2 gives specific achievable results based on two different concentration bounds. We also prove another achievable bound, Theorem 3, which does not rely on Theorem 1, but instead uses an approach based on the method of types that applies at larger values of K.

Theorem 1. *Assume $\log K = o(n)$. For any distribution p_{X_1}, p_{X_2}, let $X_j^n(k)$ be an i.i.d. sequence from p_{X_j} for each $k \in [K]$, with all sequences mutually independent. There exists an $(n, M_1, M_2, K, \epsilon)$ code if*

$$\begin{aligned} \epsilon \geq \Pr\Bigg(& \max_{k \in [K]} \sum_{i=1}^{n} i(X_{1i}(k), X_{2i}(k)) + \sqrt{n V_2}\, Z_0 \\ & < \log(M_1 M_2 K) + \frac{1}{2} \log n \Bigg) \\ & + O\left(\sqrt{\frac{\log n}{n}}\right) + O\left(\sqrt{\frac{\log K}{n}}\right) \end{aligned} \tag{4}$$

$$\log M_1 \leq n I(X_1; Y|X_2) - c\sqrt{n \log K + n \log n} \tag{5}$$

$$\log M_2 \leq n I(X_2; Y|X_1) - c\sqrt{n \log K + n \log n} \tag{6}$$

where Z_0 is a standard Gaussian, and where c is a constant.

Since we are primarily interested in the sum-rate, constraint (4) is more relevant than (5)–(6). By selecting a rate pair that is close to the sum-capacity but away from the corner points, the constraints on the individual rates are easily satisfied.

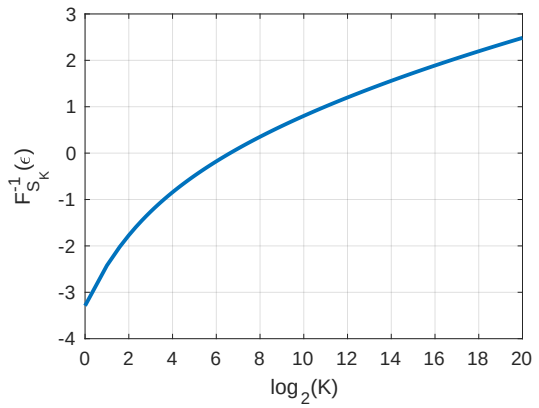

Fig. 1. The inverse CDF $F^{-1}_{S_K}(\epsilon)$ for $\epsilon = 0.01$, for $V_1 = V_2 = 1$ across a range of K. Note that the horizontal axis is $\log_2 K$, i.e., the number of bits transmitted from the CF.

For fixed K, let $Z_0, \ldots, Z_K$ be drawn i.i.d. from $\mathcal{N}(0,1)$. Let

$$S_K = \sqrt{V_2}\, Z_0 + \sqrt{V_1} \max_{k\in[1:K]} Z_k,$$

and define the CDF of S_K as

$$F_{S_K}(s) = \Pr(S_K \le s).$$

Also let $F^{-1}_{S_K}$ be the inverse of the CDF; that is,

$$F^{-1}_{S_K}(p) = \sup\{s : F_{S_K}(s) \le p\} \text{ for } p \in [0,1].$$

In what follows we use Theorem 1 and the function $F^{-1}_{S_K}$ to explicitly bound from below the benefit in sum-rate when cooperating with varying measures of K. A numerical computation of $F^{-1}_{S_K}(\epsilon)$ as a function of K is shown in Fig. 1. Lemma 1 gives lower and upper bounds on $F^{-1}_{S_K}$; we defer the proof to the extended version [1].

Lemma 1. *For K and ϵ that satisfy $K > e^3\sqrt{2\pi}\ln(4/\epsilon)$, $F^{-1}_{S_K}(\epsilon)$ is at least*

$$\sqrt{V_1(2\ln K - 2\ln\ln(4/\epsilon) - \ln\ln K - \ln(4\pi))} - \sqrt{2V_2\ln(2/\epsilon)}.$$

Moreover, for all K and ϵ,

$$F^{-1}_{S_K}(1-\epsilon) \le \sqrt{2V_1\ln K} + \sqrt{2V_1\ln(4/\epsilon)} + \sqrt{2V_2\ln(2/\epsilon)}.$$

Theorem 2. *For any $p_{X_1}p_{X_2} \in \mathcal{P}^\star$ and the associated constants V_1 and V_2, if $\log K = o(n^{1/3})$, then*

$$R_{\text{sum}}(n,\epsilon,K) \ge C_{\text{sum}} + \frac{1}{\sqrt{n}} F^{-1}_{S_K}(\epsilon) - \theta_n$$

where

$$\theta_n = O\left(\frac{\log n}{n}\right), \qquad \text{if } K \le \log n \tag{7}$$

$$\theta_n = O\left(\frac{K}{n}\right), \qquad \text{if } \log n \le K \le \log^{3/2} n \tag{8}$$

$$\theta_n = O\left(\frac{\log^{3/2} n}{n}\right), \qquad \text{if } \log^{3/2} n \le K \le n \tag{9}$$

$$\theta_n = O\left(\frac{\log^{3/2} K}{n}\right), \qquad \text{if } K \ge n. \tag{10}$$

For larger K, our achievability bound employs the function

$$\Delta(a) = \max_{p_{X_1,X_2}: I(X_1;X_2)\le a} I(X_1,X_2;Y) - C_{\text{sum}}.$$

Note that $\Delta(0) = 0$. Lemma 2 captures the behavior of $\Delta(a)$ for small a. (See [1] for the proof.)

Lemma 2. *In the limit as $a \to 0$,*

$$\Delta(a) = \sqrt{a\, 2V_1^\star \ln 2} + o(\sqrt{a})$$

where

$$V_1^\star = \max_{p_{X_1}p_{X_2}\in\mathcal{P}^\star} \text{Var}(i(X_1,X_2)). \tag{11}$$

Theorem 3. *For any K such that $\log K = \omega(\log n)$,*

$$R_{sum}(n,\epsilon,K) \ge C_{sum} + \Delta\left(\tfrac{\log K}{n} - O\left(\tfrac{\log n}{n}\right)\right) - O\left(\tfrac{1}{\sqrt{n}}\right).$$

Remark 1. *While Theorems 2 and 3 appear quite different, Lemmas 1 and 2 imply that for mid-range K values, they give similar results. In particular, if $\log n \ll \log K \ll n^{1/3}$, then applying Theorem 2, and choosing the distribution $p_{X_1}p_{X_2} \in \mathcal{P}^\star$ that achieves the maximum in (11) gives*

$$R_{\text{sum}}(n,\epsilon,K) - C_{\text{sum}} \ge \frac{1}{\sqrt{n}} F^{-1}_{S_K}(\epsilon) - \theta_n \approx \sqrt{\frac{2V_1^\star \ln K}{n}}.$$

For the same range of K, Theorem 3 gives

$$R_{\text{sum}}(n,\epsilon,K) - C_{\text{sum}} \ge \Delta\left(\tfrac{\log K}{n} - O\left(\tfrac{\log n}{n}\right)\right) - O\left(\tfrac{1}{\sqrt{n}}\right) \approx \sqrt{\frac{V_1^\star \log K}{n} 2\ln 2} = \sqrt{\frac{2V_1^\star \ln K}{n}}.$$

A. Comparison to prior work

In [5], an analog to Theorem 3 is proven for the asymptotic blocklength regime. Namely, in our notation, [5] proves that for any $\epsilon > 0$ and $\delta > 0$, if we set $K = 2^{\Omega(n)}$ then there exist n such that,

$$R_{\text{sum}}(n,\epsilon,K) - C_{\text{sum}} > \Delta\left(\frac{\log K}{n}\right) - \delta.$$

Similarly, in [5], [8], an analog to Lemma 2 is shown for asymptotic blocklength. Specifically, it is shown that the existence of distributions $p_{X_1}p_{X_2} \in \mathcal{P}^\star$ and $p_{\tilde{X}_1\tilde{X}_2}$ over $\mathcal{X}_1 \times \mathcal{X}_2$ such that (a) the support of $p_{\tilde{X}_1\tilde{X}_2}$ is included in that of $p_{X_1}p_{X_2}$, and (b)

$$I(\tilde{X}_1,\tilde{X}_2,\tilde{Y}) + D(p_{\tilde{X}_1\tilde{X}_2}\|p_{X_1}p_{X_2}) > I(X_1,X_2;Y)$$

for

$$\begin{aligned} &p_{X_1,X_2,\tilde{X}_1,\tilde{X}_2,Y,\tilde{Y}}(x_1,x_2,\tilde{x}_1,\tilde{x}_2,y,\tilde{y}) \\ &= p_{X_1}(x_1)p_{X_2}(x_2)p_{\tilde{X}_1,\tilde{X}_2}(\tilde{x}_1,\tilde{x}_2) \\ &\quad \cdot p_{Y|X_1,X_2}(y|x_1,x_2)p_{Y|X_1,X_2}(\tilde{y}|\tilde{x}_1,\tilde{x}_2), \end{aligned}$$

imply that there exists a constant σ_0 such that

$$\lim_{a\to 0} \Delta(a) \ge \sigma_0\sqrt{a}.$$

Although Theorem 3 and Lemma 2 (and their proof techniques) are similar in nature to those of [5], [8], the analysis

presented here is refined in that it captures higher order behavior in blocklength n and further optimized to address the challenges in studying values of K that are sub-exponential in the blocklength n.

We may also compare our results against prior achievable bounds without cooperation. Note that the standard MAC, with no cooperation, corresponds to $K = 1$. In fact, in this case Theorem 2 gives the same second-order term as the best-known achievable bound for the MAC sum-rate [9]–[13]. This can be seen by noting that $S_1 \sim \mathcal{N}(0, V_1 + V_2)$, and so $F_{S_1}^{-1}(\epsilon) = \sqrt{V_1 + V_2}\Phi^{-1}(\epsilon)$ where Φ is the CDF of $\mathcal{N}(0,1)$. Thus Theorem 2 gives

$$R_{\text{sum}}(n, \epsilon, 1) \geq C_{\text{sum}} + \sqrt{\frac{V_1 + V_2}{n}}\Phi^{-1}(\epsilon) - O\left(\frac{\log n}{n}\right).$$

Moreover, $V_1+V_2 = \text{Var}(i(X_1, X_2; Y))$ which, for the optimal input distribution, is precisely the best-known achievable dispersion. The proof of Theorem 2 uses i.i.d. codebooks, which, as shown in [12], can be outperformed in terms of second-order rate by constant combination codebooks. However, as pointed out in [13, Sec. III-B], the two approaches give the same bounds on the sum-rate itself.

Another interesting conclusion comes from comparing the no cooperation case ($K = 1$) with a *single bit* of cooperation ($K = 2$). As long as $V_1^\star > 0$, it is easy to see that $F_{S_2}^{-1}(\epsilon) > F_{S_1}^{-1}(\epsilon)$ for any $\epsilon \in (0, 1)$ (Fig. 1 shows an example). Thus, the second-order coefficient in Theorem 2 for $K = 2$ is strictly improved compared to $K = 1$. Therefore, even a single bit of cooperation allows for $O(\sqrt{n})$ additional message bits.

IV. Proof Sketches of Main Results

Due to space limitations, we provide only sketches of the proofs of Theorems 1–3; full proofs are in [1].

Proof sketch of Theorem 1: We use random code design, beginning with independent design of the codewords for both transmitters. For encoder $j \in \{1, 2\}$, draw each codeword $f_j(m_j, k)$ for $m_j \in [M_j], k \in [K]$ from the n-length i.i.d. distribution from p_{X_j}. The codewords are mutually independent. The facilitator code $e(m_1, m_2)$ for $(m_1, m_2) \in [M_1] \times [M_2]$ is given by

$$e(m_1, m_2) = \arg\max_{k \in [K]} s(f_1(m_1, k), f_2(m_2, k)),$$

where the score function s is $s(x_1^n, x_2^n) = \sum_{i=1}^n i(x_{1i}, x_{2i})$. While maximum likelihood decoding is expected to give the best performance, instead we here employ a threshold decoder for simplicity. For notational efficiency, let

$$\begin{aligned}&(X_1^n, X_2^n)(m_1, m_2)\\ &= (f_1(m_1, e(m_1, m_2)), f_2(m_2, e(m_1, m_2))).\end{aligned}$$

Given a constant vector $\mathbf{c}^\star = [c_{12}^\star, c_1^\star, c_2^\star]^T$, the decoder $g(y)$ chooses the unique message pair (m_1, m_2) such that

$$\mathbf{i}((X_1^n, X_2^n)(m_1, m_2); y) \geq \mathbf{c}^\star,$$

where the vector inequality means that all three inequalities must hold simultaneously. If the number of message pairs that meet this constraint is not one, we declare an error.

To write a finite blocklength bound derived from this code design, we define the following random variables. Let $(\tilde{X}_1^n, \tilde{X}_2^n)$ be the joint distribution of $(X_1^n, X_2^n)(1, 1)$ that results from our choice of CF. Also let $\tilde{Y}^n$ be the channel output when $(\tilde{X}_1^n, \tilde{X}_2^n)$ are the inputs. By analyzing various types of errors, and applying commonly-used finite blocklength bounding techniques, we find that the expected error probability satisfies

$$\begin{aligned}E[P_e] \leq{}& \Pr\left(i(\tilde{X}_1^n, \tilde{X}_2^n; \tilde{Y}^n) < \log(M_1 M_2 K) + \frac{1}{2}\log n\right)\\ &+ \Pr\left(i(\tilde{X}_1^n; \tilde{Y}^n | \tilde{X}_2^n) < \log M_1 + \frac{1}{2}\log n\right)\\ &+ \Pr\left(i(\tilde{X}_2^n; \tilde{Y}^n | \tilde{X}_1^n) < \log M_2 + \frac{1}{2}\log n\right) + \frac{3}{\sqrt{n}}. \quad (12)\end{aligned}$$

By Hoeffding's inequality and the assumptions in (5)–(6), the second and third terms in (12) are each no more than $1/\sqrt{n}$ for a suitable constant c.

Now we consider the first term in (12). For fixed x_1^n, x_2^n,

$$E[i(x_1^n, x_2^n; Y^n)] = \sum_{i=1}^n i(x_{1i}, x_{2i}) = s(x_1^n, x_2^n).$$

Thus we can apply the Berry-Esseen theorem (see, e.g. [14, Ch. XVI.5]) to write

$$\begin{aligned}&\Pr\left(i(x_1^n, x_2^n; Y^n) < c_{12}^\star \middle| (X_1^n, X_2^n) = (x_1^n, x_2^n)\right)\\ &\leq \Pr\left(s(x_1^n, x_2^n) + \sqrt{\textstyle\sum_i V(p(y|x_{1i}, x_{2i}) \| p_Y)} Z_0 < c_{12}^\star\right) + \frac{B}{\sqrt{n}}\end{aligned}$$

where $Z_0 \sim \mathcal{N}(0, 1)$ and B is a constant. To complete the proof of the theorem requires bounding the divergence-variance quantity, which can be done via Hoeffding's inequality, and some technical manipulations on the Gaussian distribution. ∎

Proof sketch of Theorem 2: Given ϵ, our goal is to choose M_1, M_2 to satisfy the conditions of Theorem 1, while

$$\log(M_1 M_2) = nC_{\text{sum}} + \sqrt{n}\, F_{S_K}^{-1}(\epsilon) - n\theta_n \quad (13)$$

where θ_n satisfies one of (7)–(10) depending on K. By choosing a point on the border of the capacity region that achieves the sum-capacity but is away from the corner points, we can easily satisfy (5)–(6). It remains to prove (4). We can write the probability in (4) as

$$\int_{-\infty}^{\infty} \phi(z) \Pr\left(\sum_{i=1}^n i(X_{1i}, X_{2i}) < c_{12}^\star - \sqrt{nV_2}\, z\right)^K dz \quad (14)$$

where ϕ is the PDF of $\mathcal{N}(0, 1)$, and $c_{12}^\star = \log(M_1 M_2 K) + \frac{1}{2}\log n$. Note that the random quantity in (14) is an i.i.d. sum where each term has expectation

$$E[i(X_1, X_2)] = I(X_1, X_2; Y) = C_{\text{sum}}$$

and variance V_1. We divide the remainder of the proof into two cases.

Case 1: $K \leq \log^{3/2} n$. By the Berry-Esseen theorem, (14) is no more than

$$\int_{-\infty}^{\infty} \phi(z) \Big[\Pr\left(nC_{\text{sum}} + \sqrt{n}\,\sigma Z_1 < c_{12}^\star - \sqrt{nV_2}\, z\right)$$

$$+ O\Big(\tfrac{1}{\sqrt{n}}\Big)\Big]^K dz$$
$$\leq \Pr(nC_{\text{sum}} + \sqrt{n}\, S_K < c_{12}^\star) + O\left(\frac{K}{\sqrt{n}}\right).$$

This bound leads to (7)–(8).

Case 2: $K \geq \log^{3/2} n$ and $\log K = o(n^{1/3})$. In this regime, we bound the probability in (14) using the following lemma.

Lemma 3 (Moderate deviations [15])**.** *Let $X_1, X_2, \ldots$ be i.i.d. random variables with zero mean and unit variance, and let $W = \sum_{i=1}^n X_i/\sqrt{n}$ where $c = E[e^{t|X_1|}] < \infty$ for some $t > 0$. There exist constants a_0 and b_0 depending only on t and c such that, for any $0 \leq w \leq a_0 n^{1/6}$,*

$$\left|\frac{\Pr(W \geq w)}{Q(w)} - 1\right| \leq \frac{b_0(1+w^3)}{\sqrt{n}},$$

where $Q(w) = 1 - \Phi(w)$ is the complementary CDF of the standard Gaussian distribution.

Since $\Pr(|Z_0| > \sqrt{\ln n}) \leq 2/\sqrt{n}$, we only need to consider the probability in (14) for $|z| \leq \sqrt{\ln n}$. From the target for $M_1 M_2$ in (13) and the bound on $F_{S_K}^{-1}(\epsilon)$ from Lemma 1, it can be shown that for this range of z, in fact the probability in (14) falls into the regime of the moderate deviations bound. Thus, (14) is no more than

$$\int_{-\sqrt{\ln n}}^{\sqrt{\ln n}} \phi(z) \left(1 - Q\left(\tfrac{c_{12}^\star - \sqrt{nV_2}\, z - nC_{\text{sum}}}{\sqrt{nV_1}}\right)(1-\lambda_n)\right)^K dz$$
$$+ O\left(\tfrac{1}{\sqrt{n}}\right), \text{ where } \lambda_n = O\left(\tfrac{\log^{3/2} K}{\sqrt{n}}\right) + O\left(\tfrac{\log^{3/2} n}{\sqrt{n}}\right)$$

To further bound this quantity, we use the fact that for any $w \geq 0$ and any $0 \leq \lambda \leq 3/4$, $Q(w)(1-\lambda) \geq Q(w + 2\lambda)$. Thus, (14) is no more than

$$F_{S_K}\left(\frac{c_{12}^\star - nC_{\text{sum}}}{\sqrt{n}} + 2\sqrt{V_1}\lambda_n\right)$$
$$+ Q\left(\frac{c_{12}^\star - nC_{\text{sum}}}{\sqrt{nV_2}}\right) + O\left(\frac{1}{\sqrt{n}}\right).$$

This is the essential bound that leads to (9)–(10). ∎

Proof sketch of Theorem 3: This proof uses the method of types. A probability mass function p_X is an n-length type on alphabet $\mathcal{X}$ if $p_X(x)$ is a multiple of $1/n$ for each $x \in \mathcal{X}$. For an n-length type p_X, the type class is denoted $T(p_X)$.

Let p_{X_1,X_2} be an n-length joint type on alphabet $\mathcal{X}_1 \times \mathcal{X}_2$. Note that the marginal distributions p_{X_1} and p_{X_2} are also n-length types. We employ the following random code construction. Draw codewords uniformly from the type classes $T(p_{X_1})$ and $T(p_{X_2})$. Given message pair (m_1, m_2), the cooperation facilitator chooses uniformly from the set of $k \in [K]$ where

$$(f(m_1,k), f(m_2,k)) \in T(p_{X_1,X_2}).$$

If there is no such k, the CF chooses k uniformly at random. These random choices at the CF are taken to be part of the random code design. For the purposes of this proof, the three information densities employ the joint distribution p_{X_1,X_2}. The quantity V_2 is also defined as in (3) using information density for this joint distribution. The decoder is as follows. Given y^n, choose the unique message pair (m_1, m_2) such that

1) $\mathbf{i}((X_1^n, X_2^n)(m_1,m_2); y^n) \geq \mathbf{c}^\star$,
2) $((X_1^n, X_2^n)(m_1,m_2)) \in T(p_{X_1,X_2})$

for a constant vector $\mathbf{c}^\star = [c_{12}^\star, c_1^\star, c_2^\star]^T$ to be determined. If there is no message pair or more than one satisfying these conditions, declare an error. Note that, given

$$(X_1^n, X_2^n)(m_1,m_2) \in T(p_{X_1,X_2}),$$

$(X_1^n, X_2^n)(m_1,m_2)$ is uniformly distributed on $T(p_{X_1,X_2})$. Let $q(x_1^n, x_2^n)$ be the uniform distribution on the type class $T(p_{X_1,X_2})$, with corresponding conditional distributions $q(x_1^n|x_2^n)$ and $q(x_2^n|x_1^n)$. Define random variables X_1^n, X_2^n, Y^n to have distribution

$$p_{X_1^n,X_2^n,Y^n}(x_1^n, x_2^n, y^n) = q(x_1^n, x_2^n) p_{Y^n|X_1^n,X_2^n}(y^n|x_1^n, x_2^n).$$

Applying some tools from the method of types, we may derive the finite blocklength bound

$$\begin{aligned} E[P_e] \leq\ & \Pr((X_1^n, X_2^n)(1,1) \notin T(p_{X_1,X_2})) \\ & + \Pr(i(X_1^n, X_2^n; Y^n) < \log(M_1 M_2) + d) \\ & + \Pr(i(X_1^n; Y^n|X_2^n) < \log M_1 + d) \\ & + \Pr(i(X_2^n; Y^n|X_2^n) < \log M_2 + d) + \frac{3}{\sqrt{n}}. \end{aligned} \tag{15}$$

where $d = \frac{1}{2}\log n + |\mathcal{X}_1| \cdot |\mathcal{X}_2| \log(n+1)$. Our goal is to show that the bound in (15) is at most ϵ for

$$\log(M_1 M_2) = nI(X_1, X_2; Y) - \sqrt{nV_2}\, Q^{-1}(\epsilon) - n\theta_n. \tag{16}$$

where $\theta_n = O(\frac{\log n}{n})$ is an error term to be chosen below.

Consider the first term in (15). Note that $(X_1^n, X_2^n)(1,1) \notin T(p_{X_1,X_2})$ only if

$$(f_1(1,k), f_2(1,k)) \notin T(p_{X_1,X_2}) \text{ for all } k \in [K].$$

This occurs with probability bounded as

$$\begin{aligned} &\Pr((X_1^n, X_2^n)(1,1) \notin T(p_{X_1,X_2})) \\ &= \left(1 - \frac{|T(p_{X_1,X_2})|}{|T(p_{X_1})| \cdot |T(p_{X_2})|}\right)^K \\ &\leq \left(1 - (n+1)^{-|\mathcal{X}_1|\cdot|\mathcal{X}_2|} 2^{-nI(X_1;X_2)}\right)^K \\ &\leq \exp\left\{-K(n+1)^{-|\mathcal{X}_1|\cdot|\mathcal{X}_2|} 2^{-nI(X_1;X_2)}\right\} \leq \frac{1}{\sqrt{n}}, \end{aligned}$$

where the last inequality holds if

$$nI(X_1;X_2) \leq \log K - |\mathcal{X}_1| \cdot |\mathcal{X}_2| \log(n+1) - \log\left(\tfrac{1}{2}\ln n\right). \tag{17}$$

The second term in (15) can be bounded using the Berry-Esseen inequality by

$$\begin{aligned} &\Pr(i(X_1^n, X_2^n; Y^n) < \log(M_1 M_2) + d) \\ &\leq Q\left(\frac{nI(X_1,X_2;Y) - \log(M_1 M_2) + d}{\sqrt{nV_2}}\right) + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

As in the proof of Thm. 2, we can use Hoeffding's inequality to bound the third and fourth terms in (15) from above by $1/\sqrt{n}$. Plugging in for $M_1 M_2$ from (16), we find

$$E[P_e] \leq Q\left(Q^{-1}(\epsilon) + \sqrt{\frac{n}{V_2}}\,\theta_n - O\left(\frac{\log n}{\sqrt{n}}\right)\right) + O\left(\frac{1}{\sqrt{n}}\right).$$

Therefore, there exists a choice for $\theta_n = O(\frac{\log n}{n})$ where this bound is no greater than ϵ. ∎

REFERENCES

[1] O. Kosut, M. Effros, and M. Langberg, "Every bit counts: Second-order analysis of cooperation in the multiple-access channel," *arXiv preprint arXiv:2102.01247*, 2021.

[2] F. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 441–445, 1983.

[3] F. Willems and E. Van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Transactions on Information Theory*, vol. 31, no. 3, pp. 313–327, 1985.

[4] P. Noorzad, M. Effros, M. Langberg, and T. Ho, "On the power of cooperation: Can a little help a lot?" in *IEEE International Symposium on Information Theory*, 2014, pp. 3132–3136.

[5] P. Noorzad, M. Effros, and M. Langberg, "The unbounded benefit of encoder cooperation for the k-user MAC," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3655–3678, 2018.

[6] M. Langberg and M. Effros, "On the capacity advantage of a single bit," in *2016 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2016, pp. 1–6.

[7] P. Noorzad, M. Effros, and M. Langberg, "Can negligible cooperation increase capacity? the average-error case," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 1256–1260.

[8] P. Noorzad, M. Langberg, and M. Effros, "Negligible Cooperation: Contrasting the Maximal- and Average-Error Cases," *arXiv preprint arXiv:1911.10449*, 2019.

[9] Y.-W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *2012 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2012, pp. 831–835.

[10] E. M. Jazi and J. N. Laneman, "Simpler achievable rate regions for multiaccess with finite blocklength," in *2012 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2012, pp. 36–40.

[11] V. Y. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 881–903, 2014.

[12] J. Scarlett, A. Martinez, and A. G. i Fàbregas, "Second-order rate region of constant-composition codes for the multiple-access channel," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 157–172, 2015.

[13] R. C. Yavas, V. Kostina, and M. Effros, "Random access channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 67, no. 4, pp. 2115–2140, 2021.

[14] W. Feller, *An introduction to probability theory and its applications. Vol. II.*, ser. Second edition. New York: John Wiley & Sons Inc., 1971.

[15] L. H. Y. Chen, X. Fang, and Q.-M. Shao, "From Stein identities to moderate deviations," *Ann. Probab.*, vol. 41, no. 1, pp. 262–293, 01 2013.