A Wringing-Based Proof of a Second-Order Converse for the Multiple-Access Channel under Maximal Error Probability

Fei Wei
Arizona State University
fwei16@asu.edu

Oliver Kosut
Arizona State University
okosut@asu.edu

***Abstract*—The second-order converse bound of multiple access channels is an intriguing problem in information theory. In this work, in the setting of the two-user discrete memoryless multiple access channel (DM-MAC) under the maximal error probability criterion, we investigate the gap between the best achievable rates and the asymptotic capacity region. With "wringing techniques" and meta-converse arguments, we show that gap at blocklength n is upper bounded by $O(1/\sqrt{n})$.**

I. Introduction

The *multiple-access channel* (MAC) is a fundamental information theory problem, which admits multiple independent users (transmitters) to deliver signal to a noisy channel, and the signal sent by corresponding sources must be recovered at the receiver.

The MAC was first alluded to by Shannon [1] more than half century ago. Since then, it attracted significant amount of attention and became a fruitful research field [2]. For the *discrete memoryless multiple-access channel* (DM-MAC), the capacity region was derived by Liao [3] and Ahlswede [4]. For the *Gaussian MAC*, the capacity region was introduced by Wyner [5] and Cover [6]. Dueck [7] showed that for maximal errors the capacity region of the MAC (and also for Shannon's two-way channel) can be strictly smaller than the one for average errors. However, these results were in the context of first-order asymptotic, that is, the error probability approaches to zero and and blocklength grows to infinity. Naturally, one may ask that, if we have finite blocklength and constrained error probability, how does the capacity region change?

The *strong converse* shows that with error probability fixed to be greater than zero, the set of achievable coding rates approaches to the standard capacity region when blocklength grows to infinity. For the DM-MAC, the strong converse was first given by Dueck [8] using the *blowing-up lemma* and a *wringing step*. Ahlswede [9] presented an alternative proof using Augustin's converse argument [10] instead of the blowing-up lemma and also improved the *wringing technique*. For the Gaussian MAC, a strong converse was given by Fong and Tan [11] by applying Ahlswede's wringing technique. Based on the strong converse, one may further ask, given fixed error probability, how quickly do the achievable coding rates approach to the standard capacity region as the blocklength grows?

The answer to this question is the *second-order coding rate*, which is the gap between the best achievable rates and the asymptotic capacity region given finite blocklength. Strassen [12] showed that the gap at blocklength n for point-to-point channel is $O(1/\sqrt{n})$. This result was refined by Polyanskiy et al. [13] and also brought renewed interest regarding the second-order rates. Second-order achievable results for the MAC under average probability of error were found in [14]–[18]. For the problem variant with degraded message sets, second-order results, including matching converses, were derived in [19], [20]. However, for the standard MAC, existing second-order achievability and converse bounds do not coincide. Second-order converse bounds have been found by Kosut [21] under the average error probability criterion, and by Moulin [22] under the maximal error probability criterion. However, we have been unable to verify the proof of [22] (specifically, the transition from equation (27) to (28)). Thus, in this work, we present an alternative proof of a second-order converse for MAC under the maximal error probability criterion. Our work has a different approach which mainly uses *wringing-based* techniques (which will be introduced in detail later), and claims similar results as of [22].

II. Preliminaries

A. Basic Notations

For a random variable X, we use calligraphic letter $\mathcal{X}$ to represent the alphabet, and the lower case x to represent a realization. We denote the set of all possible distributions on $\mathcal{X}$ as $\mathcal{P}(\mathcal{X})$. Given a distribution P_X and a subset $A \subset \mathcal{X}$, we may represent $\sum_{x\in A} P_X(x)$ as $P_X(A)$. At times, for simplicity we may omit the subscript, e.g., we may use $P(x)$ instead of $P_X(x)$. And we denote a vector of random variables $(X_1, \ldots, X_n)$ as X^n, and set $\{1, \ldots, n\}$ as $[n]$.

Given distributions $P, Q \in \mathcal{P}(\mathcal{X})$, the Kullback-Leibler divergence is given by

$$D(P\|Q) = \mathbb{E}_P \left[\log \frac{P(X)}{Q(X)}\right].$$

The Rényi divergence of order ∞ is given by

$$D_\infty(P\|Q) = \sup_{A\subset\mathcal{X}} \log \frac{P(A)}{Q(A)}.$$

The divergence variance is defined as

$$V(P\|Q) = \mathrm{Var}_P\left[\log\frac{P(X)}{Q(X)}\right],$$

and the conditional divergence variance is defined as

$$V(P_{Y|X}\|Q_{Y|X}|U_X) = \sum_{x\in\mathcal{X}} U_X(x)V(P_{Y|X=x}\|Q_{Y|X=x}).$$

We take the standard $\mathcal{Q}^{-1}(p)$ as the inverse of the complementary Gaussian CDF. We also adopt the standard $O(\cdot)$ notation, that is, $g(n) = O(f(n))$ means $\limsup_{n\mapsto\infty}\left|\frac{g(n)}{f(n)}\right| < \infty$.

B. Multiple Access Channel

In this work, we consider a *discrete memoryless multiple-access channel* (DM-MAC) with two independent users (transmitters) S_1 and S_2. For $i = 1, 2$, S_i independently generates message $m_i \in \mathcal{M}_i$ ($|\mathcal{M}_i| = M_i$), then m_i is encoded into channel input x_i via deterministic encoding function $X_i : \mathcal{M}_i \mapsto \mathcal{X}_i$. The MAC is defined by channel law

$$W_{Y|X_1,X_2} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \mapsto \mathcal{Y}),$$

where $\mathcal{X}_1$, $\mathcal{X}_2$ are input alphabets and $\mathcal{Y}$ is the output alphabet. In what follows, given a message tuple (m_1, m_2) and encoding functions X_1, X_2, for simplicity we may use W_Y to represent $W_{Y|X_1,X_2}$, use $W(y|m_1, m_2)$ to represent $W_{Y|X_1,X_2}(y|X_1(m_1), X_2(m_2))$, and use $W_{Y|m_1,m_2}$ to represent $W_{Y|X_1(m_1),X_2(m_2)}$. Given a one-shot channel W, the product channel of blocklength n is given by

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

The receiver has a decoding function $\phi_{de} : \mathcal{Y}^n \mapsto \mathcal{M}_1 \times \mathcal{M}_2$. The decoding error probability for each message tuple (m_1, m_2) is

$$\begin{aligned} e(m_1, m_2) = \sum_{y^n\in\mathcal{Y}^n} & W^n(y^n|X_1^n(m_1), X_2^n(m_2)) \\ & \cdot \mathbf{1}(\phi_{de}(y^n) \neq (m_1, m_2)). \end{aligned}$$

Here $\mathbf{1}(\cdot)$ is the indicator function. The maximal error probability e_{max} is defined as

$$e_{\text{max}} = \max_{m_1,m_2} e(m_1, m_2).$$

A (n, M_1, M_2) DM-MAC code consists of codebooks $\{X_i^n(m_i) : m_i \in \mathcal{M}_i\}$ for $i = 1, 2$. We say a rate pair (R_1, R_2) is (n, ϵ) achievable if there exists a (n, M_1, M_2) MAC code with maximal error probability ϵ, where $2^{nR_1} \leq M_1$ and $2^{nR_2} \leq M_2$.

III. Main Results

Assuming X_1, X_2 are independent, we define

$$C_{\text{sum}} \triangleq \max_{P_{X_1}P_{X_2}} I(X_1, X_2; Y).$$

We assume the maximum is achieved at unique $(P^*_{X_1}, P^*_{X_2})$, such that $W_{Y|X_1X_2}P^*_{X_1}P^*_{X_2}$ is our optimal joint distribution for (X_1, X_2, Y) which achieves C_{sum}. Therefore, the corresponding unique distribution of Y is

$$P^*_Y(y) = \sum_{x_1,x_2} W(y|x_1, x_2)P^*(x_1)P^*(x_2).$$

And we define

$$V^* \triangleq V(W_{Y|X_1,X_2}\|P^*_Y|P^*_{X_1}P^*_{X_2}).$$

Note that in general the sum capacity for the maximal error setting is unknown, so we choose to write our converse bounds in terms of the average-error sum capacity, which is C_{sum} as we defined. Our goal of this work is to prove the following theorem.

Theorem 1. *For every (n, M_1, M_2) DM-MAC code with maximal error probability $\epsilon \in (0, 1)$, it holds that*

$$\log M_1M_2 \leq nC_{sum} - \sqrt{nV^*}\mathcal{Q}^{-1}(\epsilon) + o(\sqrt{n}).$$

IV. Proof Tools

The proof consists of four steps. Step 1), we use an averaging argument with Markov's inequality to find a subset Ω of message pairs with a desired property. Step 2), we use the wringing technique to find subsets $\bar{\mathcal{M}}_1$, $\bar{\mathcal{M}}_2$ and define subset $\bar{\Omega} = (\bar{\mathcal{M}}_1 \times \bar{\mathcal{M}}_2) \cap \Omega$ such that inputs are almost independent when restricted to $\bar{\Omega}$. Step 3), we prove a converse bound on the subset $\bar{\Omega}$ and relate it back to the original code.

A. The Wringing Technique

The history of the *wringing technique* can be tracked back to the work of Dueck [8] and Ahlswede [9] in the 1980s, and it was refined by Kosut [21] recently. Roughly speaking, given any subset Ω of a MAC code, the goal of the wringing technique is to "wring out" the dependence in the subset by taking $\bar{\Omega} \subset \Omega$, such that inputs are "almost independent" when restricted to be taken from $\bar{\Omega}$.

Definition 1 (Wringing Dependence)**.** *Given random variables X, Y with joint distribution $P_{X,Y}$, the Wringing dependence between X and Y is given by*

$$\begin{aligned} \Delta(X; Y) = \inf_{P_X,P_Y} \sup_{A\subset\mathcal{X},B\subset\mathcal{Y}} & \inf\{\delta \geq 0 : \\ & P_{X,Y}(A, B)^{1+\delta} \leq P_X(A)P_Y(B)\} \end{aligned}$$

Lemma 1 (The Wringing Lemma in [21])**.** *Let $P_{X,Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, $P_X \in \mathcal{P}(\mathcal{X})$, and $P_Y \in \mathcal{P}(\mathcal{Y})$ be distributions such that*

$$D_\infty(P_{X,Y}\|P_XP_Y) \leq \sigma$$

where σ is finite. For any $\delta > 0$, there exist sets $\bar{\mathcal{X}} \subset \mathcal{X}$, $\bar{\mathcal{Y}} \subset \mathcal{Y}$ such that

$$P_{X,Y}(\bar{\mathcal{X}} \times \bar{\mathcal{Y}}) \geq \exp\left\{-\frac{\sigma}{\delta}\right\}$$

and

$$\Delta(\bar{X}; \bar{Y}) \leq \delta$$

where $(\bar{X}, \bar{Y})$ are distributed according to $P_{X,Y|X\in\bar{\mathcal{X}},Y\in\bar{\mathcal{Y}}}$.

B. Hypothesis Testing

Definition 2. *Consider a hypothesis testing between two distributions P and Q. Let $\psi : \mathcal{X} \mapsto [0,1]$ be a randomized decision rule that $\psi(x) = \Pr[$Give $P|X = x]$. The* Type-II error probability of the Neyman-Pearson test at significance level $1-\epsilon$ *is given by*

$$\beta_{1-\epsilon}(P,Q) = \min_{\psi:\mathbb{E}_P[\psi(X)]\geq 1-\epsilon} \mathbb{E}_Q[\psi(X)].$$

Lemma 2 (Proposition 2.1 from [22])**.** *For any (n, M_1, M_2)-MAC code with maximal error probability ϵ over channel W^n, the following holds for any subset Ω^* of $\mathcal{M}_1 \times \mathcal{M}_2$ and any distribution $P_{Y^n} \in \mathcal{P}(\mathcal{Y}^n)$,*

$$\frac{1}{|\Omega^*|} \geq \frac{1}{|\Omega^*|} \sum_{(m_1,m_2)\in\Omega^*} \beta_{1-\epsilon}(W^n_{Y^n|X_1^n(m_1),X_2^n(m_2)}, P_{Y^n}).$$

V. Proof of the Theorem

We let (m_1, m_2) be drawn uniformly at random from $\mathcal{M}_1 \times \mathcal{M}_2$. This is an artificial distribution for the purposes of the proof, rather than an assumption on the true distribution of the messages. Consequently, from encoding functions $\{X_{1i} : i \in [n]\}$ and $\{X_{2i} : i \in [n]\}$, we obtain distributions $Q_{X_{1i}} \in \mathcal{P}(\mathcal{X}_1)$ and $Q_{X_{2i}} \in \mathcal{P}(\mathcal{X}_2)$ as follows. For any $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$,

$$Q_{X_{1i}}(x_1) \triangleq \sum_{m_1} P(m_1)\cdot \mathbf{1}(X_{1i}(m_1) = x_1),$$
$$Q_{X_{2i}}(x_2) \triangleq \sum_{m_2} P(m_2)\cdot \mathbf{1}(X_{2i}(m_2) = x_2).$$

Note that $Q_{X_{1i},X_{2i}}(x_1, x_2) = Q_{X_{1i}}(x_1)\cdot Q_{X_{2i}}(x_2)$, and

$$Q_{Y_i}(y) \triangleq \sum_{x_1,x_2} W(y|x_1, x_2)\cdot Q_{X_{1i},X_{2i}}(x_1, x_2).$$

For each message tuple (m_1, m_2), we define the *average divergence* as

$$D(m_1, m_2) \triangleq \frac{1}{n}\sum_{i=1}^{n} D(W_{Y_i|X_{1i}(m_1),X_{2i}(m_2)}\|Q_{Y_i}),$$

and define the *divergence variance* as

$$V(m_1, m_2) \triangleq \frac{1}{n}\sum_{i=1}^{n} V(W_{Y_i|X_{1i}(m_1),X_{2i}(m_2)}\|Q_{Y_i}),$$

and we assume $V(m_1, m_2) > 0$. Additionally, for $X_{1i} \sim Q_{X_{1i}}$ and $X_{2i} \sim Q_{X_{2i}}$, $Y_i \sim Q_{Y_i}$ for each $i \in [n]$ we define

$$I_3 \triangleq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i).$$

and we have the following lemma.

Lemma 3.

$$I_3 = \mathbb{E}_{m_1,m_2}[D(m_1, m_2)].$$

Proof.

$$\begin{aligned} I_3 =&\frac{1}{n}\sum_{i=1}^{n}\sum_{x_{1i},x_{2i},y_i} P(x_{1i}, x_{2i}, y_i)\log\frac{W(y_i|x_{1i}, x_{2i})}{Q(y_i)}\\ =&\sum_{m_1,m_2} P(m_1, m_2)\frac{1}{n}\sum_{i=1}^{n} D(W_{Y_i|X_{1i}(m_1),X_{2i}(m_2)}\|Q_{Y_i})\\ =&\mathbb{E}_{P_{m_1,m_2}}[D(m_1, m_2)]. \end{aligned}$$

□

Lemma 4. *Let $\Omega \triangleq \{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2 : D(m_1, m_2) \leq I_3 + \frac{1}{n}\}$, it holds that $|\Omega| \geq \frac{M_1M_2}{nI_3+1}$.*

Proof. As (m_1, m_2) is uniform over $\mathcal{M}_1 \times \mathcal{M}_2$, by Markov's inequality and Lemma 3, we have

$$\begin{aligned} \frac{|\Omega|}{M_1M_2} =& \Pr\left(D(m_1, m_2) \leq I_3 + \frac{1}{n}\right)\\ \geq&1 - \frac{\mathbb{E}_{P_{m_1,m_2}}[D(m_1, m_2)]}{I_3 + \frac{1}{n}}\\ =&1 - \frac{I_3}{I_3 + \frac{1}{n}}. \end{aligned}$$

□

Given Ω and P_{m_1,m_2}, we define P'_{m_1,m_2} as

$$P'(m_1, m_2) \triangleq P_{m_1,m_2|(m_1,m_2)\in\Omega}(m_1, m_2) = \frac{1}{|\Omega|}$$

for each $(m_1, m_2) \in \Omega$, and $P'(m_1, m_2) = 0$ otherwise. Accordingly, for any (m_1, m_2) it holds that

$$\log\frac{P'(m_1, m_2)}{P(m_1)P(m_2)} \leq \log(nI_3 + 1).$$

Therefore,

$$D_\infty(P'_{m_1,m_2}\|P_{m_1}P_{m_2}) \leq \log(nI_3 + 1) \triangleq \sigma.$$

Given P_{m_1}, P_{m_2}, and $P'_{m_1m_2}$, by Lemma 1, for any $\delta > 0$ there exists $\bar{\mathcal{M}}_1 \subseteq \mathcal{M}_1$ and $\bar{\mathcal{M}}_2 \subseteq \mathcal{M}_2$, such that

$$P'_{m_1,m_2}(\bar{\mathcal{M}}_1 \times \bar{\mathcal{M}}_2) \geq \exp\left(-\frac{\sigma}{\delta}\right), \tag{1}$$

and

$$\Delta(\bar{m}_1; \bar{m}_2) \leq \delta, \tag{2}$$

where $(\bar{m}_1, \bar{m}_2) \sim P'_{m_1,m_2|m_1\in\bar{\mathcal{M}}_1,m_2\in\bar{\mathcal{M}}_2}$.

We define $\bar{\Omega} \triangleq (\bar{\mathcal{M}}_1 \times \bar{\mathcal{M}}_2) \cap \Omega$. The definition of P'_{m_1,m_2} implies that $P'_{m_1,m_2}(\bar{\mathcal{M}}_1 \times \bar{\mathcal{M}}_2) = P'_{m_1,m_2}(\bar{\Omega})$, therefore (1) implies

$$P'_{m_1,m_2}(\bar{\Omega}) = \frac{|\bar{\Omega}|}{|\Omega|} \geq \exp\left(-\frac{\sigma}{\delta}\right). \tag{3}$$

We define $\bar{P}_{m_1,m_2}$ as

$$\bar{P}_{m_1,m_2}(m_1, m_2) \triangleq P'_{m_1,m_2|m_1\in\bar{\mathcal{M}}_1,m_2\in\bar{\mathcal{M}}_2}(m_1, m_2) = \frac{1}{|\bar{\Omega}|}$$

for each $(m_1, m_2) \in \bar{\Omega}$, and $\bar{P}_{m_1,m_2}(m_1, m_2) = 0$ otherwise.

By (3) and Lemma 4, it holds that

$$\begin{aligned} \frac{|\bar{\Omega}|}{M_1M_2} =&\frac{|\bar{\Omega}|}{|\Omega|}\cdot\frac{|\Omega|}{M_1M_2} = P'_{m_1,m_2}(\bar{\Omega})\cdot\frac{|\Omega|}{M_1M_2}\\ \geq& \exp\left(-\frac{\sigma}{\delta}\right)\cdot\frac{1}{nI_3 + 1}, \end{aligned}$$

which implies

$$\log|\bar{\Omega}| - \log(M_1M_2) \geq -\frac{\sigma}{\delta} - \log(nI_3+1) \stackrel{*}{=} -\left(1+\frac{1}{\delta}\right)\log(nI_3+1),$$

where (*) holds as we defined $\sigma = \log(nI_3+1)$. Thus,

$$\log M_1M_2 \leq \log|\bar{\Omega}| + \left(1+\frac{1}{\delta}\right)\log(nI_3+1). \tag{4}$$

We apply the meta-converse of Lemma 2 to the set $\bar{\Omega}$, with $P_{Y^n} = \prod_{i=1}^n Q_{Y_i}$. Thus

$$\frac{1}{|\bar{\Omega}|} \geq \frac{1}{|\bar{\Omega}|}\sum_{(m_1,m_2)\in\bar{\Omega}} \beta_{1-\epsilon}\left(W^n_{y^n|m_1,m_2}, \prod_{i=1}^n Q_{Y_i}\right). \tag{5}$$

For any (m_1,m_2), the following asymptotics hold for the Neyman–Pearson test [23]:

$$\begin{aligned} -\log\beta_{1-\epsilon}(W^n_{Y^n|m_1,m_2}, Q_{Y^n}) &= nD(m_1,m_2) - \sqrt{nV(m_1,m_2)}\mathcal{Q}^{-1}(\epsilon) \\ &\quad + \frac{1}{2}\log n + O(1). \end{aligned} \tag{6}$$

Combining (5) and (6) gives

$$\begin{aligned} \log|\bar{\Omega}| \leq &-\log\left[\frac{1}{|\bar{\Omega}|}\sum_{(m_1,m_2)\in\bar{\Omega}} \exp\left(-nD(m_1,m_2) + \sqrt{nV(m_1,m_2)}\mathcal{Q}^{-1}(\epsilon) - \frac{1}{2}\log n - O(1)\right)\right] \\ \stackrel{(a)}{\leq} &-\log\left[\frac{1}{|\bar{\Omega}|}\sum_{(m_1,m_2)\in\bar{\Omega}} \exp\left(\sqrt{nV(m_1,m_2)}\mathcal{Q}^{-1}(\epsilon)\right)\right] \\ &+ nI_3 + \frac{1}{2}\log n + O(1) \\ \stackrel{(b)}{\leq} &-\log\left[\exp\left(\sqrt{\frac{1}{|\bar{\Omega}|}\sum_{(m_1,m_2)\in\bar{\Omega}} nV(m_1,m_2)}\mathcal{Q}^{-1}(\epsilon)\right)\right] \\ &+ nI_3 + \frac{1}{2}\log n + O(1) \\ = &nI_3 - \sqrt{\sum_{(m_1,m_2)\in\bar{\Omega}} \frac{nV(m_1,m_2)}{|\bar{\Omega}|}}\mathcal{Q}^{-1}(\epsilon) + \frac{1}{2}\log n + O(1). \end{aligned} \tag{7}$$

Here, (a) holds by the definition of set Ω and $\bar{\Omega} \subset \Omega$. (b) holds by the fact that function $f(x) = e^{c\sqrt{x}}$ is convex for $x > 0$ when $c < 0$ and $x > c^{-2}$ when $c > 0$. Where as $\epsilon < 1/2$ implies $\mathcal{Q}^{-1}(\epsilon) > 0$ and we have $nV(m_1,m_2) > (\mathcal{Q}^{-1}(\epsilon))^{-2}$ when n is large. And $\mathcal{Q}^{-1}(\epsilon) \leq 0$ when $\epsilon \geq 1/2$.

Now combining (4) and (7) gives

$$\begin{aligned} \log M_1M_2 \leq &nI_3 - \sqrt{\sum_{(m_1,m_2)\in\bar{\Omega}} \frac{nV(m_1,m_2)}{|\bar{\Omega}|}}\mathcal{Q}^{-1}(\epsilon) \\ &+ \frac{1}{2}\log n + O(1) + \left(1+\frac{1}{\delta}\right)\log(nI_3+1). \end{aligned} \tag{8}$$

We want to prove that either I_3 is bounded away from C_{sum}, or

$$\sum_{(m_1,m_2)\in\bar{\Omega}} \frac{V(m_1,m_2)}{|\bar{\Omega}|} \approx V^*.$$

Given encoding functions $\{X_{1i} : i \in [n]\}$, $\{X_{2i} : i \in [n]\}$, let $\bar{P}_{X_1^n,X_2^n}$ be the distribution induced by $\bar{P}_{m_1,m_2}$, and $\bar{P}_{X_1^n}$, $\bar{P}_{X_2^n}$ be the marginals. That is,

$$\begin{aligned} &\bar{P}_{X_1^n,X_2^n}(x_1^n, x_2^n) \\ &\triangleq \sum_{(m_1,m_2)\in\bar{\Omega}} \bar{P}(m_1,m_2)\cdot\mathbf{1}(X_1^n(m_1) = x_1^n, X_2^n(m_2) = x_2^n). \end{aligned}$$

We first focus on the second-order term in (8), note that

$$\begin{aligned} &\sum_{i=1}^n V(W_{Y_i|X_{1i}X_{2i}}\|Q_{Y_i}|\bar{P}_{X_{1i}X_{2i}}) \\ &= \sum_{i=1}^n \sum_{x_{1i},x_{2i}} \bar{P}(x_{1i},x_{2i})V\left(W_{Y_i|x_{1i},x_{2i}}\|Q_{Y_i}\right) \\ &= \sum_{i=1}^n \sum_{x_{1i},x_{2i}} \sum_{(m_1,m_2)\in\bar{\Omega}} \mathbf{1}(X_{1i}(m_1) = x_{1i}, X_{2i}(m_2) = x_{2i}) \\ &\quad \cdot \bar{P}(m_1,m_2)V\left(W_{Y_i|x_{1i},x_{2i}}\|Q_{Y_i}\right) \\ &= \sum_{i=1}^n \sum_{(m_1,m_2)\in\bar{\Omega}} \frac{V(W_{Y_i|X_{1i}(m_1),X_{2i}(m_2)}\|Q_{Y_i})}{|\bar{\Omega}|} \\ &= \sum_{(m_1,m_2)\in\bar{\Omega}} \frac{nV(m_1,m_2)}{|\bar{\Omega}|}. \end{aligned} \tag{9}$$

Regarding each channel use i, for any $(x_1,x_2) \in \bar{\Omega}_x \triangleq \{(x_1,x_2) : x_1 = X_{1i}(m_1), x_2 = X_{2i}(m_2), (m_1,m_2) \in \bar{\Omega}\}$, let $A = \{m_1 : \mathbf{1}(X_{1i}(m_1) = x_1)\}$, $B = \{m_2 : \mathbf{1}(X_{2i}(m_2) = x_2)\}$. By (2) and Property 3 of Theorem 2 in [21], we have

$$|\bar{P}_{m_1,m_2}(A,B) - \bar{P}_{m_1}(A)\bar{P}_{m_2}(B)| \leq 2\delta.$$

By our definitions $\bar{P}_{X_1}(x_1) = \bar{P}_{m_1}(A)$, $\bar{P}_{X_2}(x_2) = \bar{P}_{m_2}(B)$, and $\bar{P}_{X_1,X_2}(x_1,x_2) = \bar{P}_{m_1,m_2}(A,B)$, such that for any $(x_1,x_2) \in \bar{\Omega}_x$ we have

$$|\bar{P}(x_1,x_2) - \bar{P}(x_1)\bar{P}(x_2)| \leq 2\delta. \tag{10}$$

For any distribution $P_Y \in \mathcal{P}(\mathcal{Y})$, by (10), for each $i \in [n]$

$$\begin{aligned} &\sum_{x_{1i},x_{2i}} \bar{P}(x_{1i},x_{2i})V(W_{Y_i|x_{1i},x_{2i}}\|P_Y) \\ &= \sum_{x_{1i},x_{2i}} \bar{P}(x_{1i},x_{2i})V(W_{Y_i|x_{1i},x_{2i}}\|P_Y), \\ &\geq \sum_{x_{1i},x_{2i}} \left(\bar{P}(x_{1i})\bar{P}(x_{2i}) - 2\delta\right)V(W_{Y_i|x_{1i},x_{2i}}\|P_Y) \\ &= V(W_{Y_i|X_{1i},X_{2i}}\|P_Y|\bar{P}_{X_{1i}}\bar{P}_{X_{2i}}) \\ &\quad - 2\delta\sum_{x_{1i},x_{2i}} V(W_{Y_i|x_{1i},x_{2i}}\|P_Y). \end{aligned} \tag{11}$$

By (9) and (11),

$$\sum_{(m_1,m_2)\in\bar{\Omega}} \frac{V(m_1,m_2)}{|\bar{\Omega}|} \geq \frac{1}{n}\sum_{i=1}^{n}\big[V(W_{Y_i|X_{1i},X_{2i}}\|Q_{Y_i}|\bar{P}_{X_{1i}}\bar{P}_{X_{2i}}) -2\delta V'\big] \tag{12}$$

where $V' = \sum_{x_{1i},x_{2i}} V(W_{Y_i|x_{1i},x_{2i}}\|Q_{Y_i})$.

It remains to prove that $Q_{Y_i} \approx P^*_Y$, $\bar{P}_{X_{1i}} \approx P^*_{X_1}$, and $\bar{P}_{X_{2i}} \approx P^*_{X_2}$. Regarding the first-order term in (7), if $nI_3 < nC_{\text{sum}} - \sqrt{nV^*}\mathcal{Q}^{-1}(\epsilon)$, then by (8) we are done with the proof. Hence we assume

$$nC_{\text{sum}} - \sqrt{nV^*}\mathcal{Q}^{-1}(\epsilon) \leq nI_3 = \sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i).$$

Recall that X_{1i} is independent of X_{2i}, so $I(X_{1i},X_{2i};Y_i) \leq C_{\text{sum}}$. Define the set

$$\mathcal{A} = \{i\in[n] : I(X_{1i},X_{2i};Y_i) < C_{\text{sum}} - n^{-1/4}\}.$$

We have

$$\begin{aligned} nC_{\text{sum}} - O(\sqrt{n}) &\leq \sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i) \\ &\leq |\mathcal{A}|(C_{\text{sum}} - n^{-1/4}) + (n-|\mathcal{A}|)C_{\text{sum}}. \end{aligned}$$

Thus $|\mathcal{A}|\leq O(n^{3/4})$. For any $i\notin\mathcal{A}$, we have

$$C_{\text{sum}} - n^{-1/4} \leq I(X_{1i},X_{2i};Y_i) \leq C_{\text{sum}}.$$

Recall that we have unique optimizer $(P^*_{X_1},P^*_{X_2}) = \arg\max_{P_{X_1},P_{X_2}} I(X_1,X_2;Y)$ and the mutual information can be considered as a continuous function of (P_{X_1},P_{X_2}). Thus, for any $i\notin\mathcal{A}$,

$$|Q_{Y_i}(y) - P^*_Y(y)|\leq o(1) \text{ for all } y.$$

Define the distribution

$$\bar{\bar{P}}_{Y_i}(y) = \sum_{x_1,x_2} W(y|x_1,x_2)\bar{P}_{X_{1i}}(x_1)\bar{P}_{X_{2i}}(x_2).$$

We again apply the meta-converse of Lemma 2 on $\bar{\Omega}$, but now with $P_{Y^n} = \prod_{i=1}^{n}\bar{\bar{P}}_{Y_i}$. By applying a similar technique as in (7), it holds that

$$\begin{aligned} \log|\bar{\Omega}| &\leq \sum_{i=1}^{n} D(W_{Y_i|X_{1i},X_{2i}}\|\bar{\bar{P}}_{Y_i}|\bar{P}_{X_{1i},X_{2i}}) - O(\sqrt{n}) \\ &\leq \sum_{i=1}^{n}\Big[D(W_{Y_i|X_{1i}X_{2i}}\|\bar{\bar{P}}_{Y_i}|\bar{P}_{X_{1i}}\bar{P}_{X_{2i}}) + O(\delta)\Big] - O(\sqrt{n}) \\ &\leq \sum_{i=1}^{n} I_{\bar{\bar{P}}}(X_{1i},X_{2i};Y_i) + O(n\delta) - O(\sqrt{n}). \end{aligned} \tag{13}$$

Where $I_{\bar{\bar{P}}}$ is the mutual information respect to $\bar{P}_{X_{1i}}$, $\bar{P}_{X_{2i}}$, and $\bar{\bar{P}}_{Y_i}$. If $\sum_{i=1}^{n} I_{\bar{\bar{P}}}(X_{1i},X_{2i};Y_i) + O(n\delta) < nC_{\text{sum}} - O(\sqrt{n})$ then we are done. Therefore we assume

$$\sum_{i=1}^{n} I_{\bar{\bar{P}}}(X_{1i},X_{2i};Y_i) + O(n\delta) \geq nC_{\text{sum}} - O(\sqrt{n}). \tag{14}$$

As described in detail below, we will choose $\delta = O(\sqrt{\log n}/n^{1/4})$. Thus

$$\sum_{i=1}^{n} I_{\bar{\bar{P}}}(X_{1i},X_{2i};Y_i) \geq nC_{\text{sum}} - O\left(n\frac{\sqrt{\log n}}{n^{1/4}}\right).$$

Applying a similar argument to above, there exists a set $\mathcal{B}\subset[n]$ where $|\mathcal{B}|= o(n)$ and for each $i\notin\mathcal{B}$,

$$I_{\bar{\bar{P}}}(X_{1i},X_{2i};Y_i) \geq C_{\text{sum}} - o(1).$$

Therefore, again by continuity of the mutual information, for all $i\notin\mathcal{B}$,

$$|\bar{P}_{X_{1i}}(x_1) - P^*_{X_1}(x_1)|\leq o(1) \text{ for all } x_1$$

and a similar result for $\bar{P}_{X_{2i}}$.

Thus, by (12), when n is large, it holds that

$$\begin{aligned} &\sum_{(m_1,m_2)\in\bar{\Omega}} \frac{V(m_1,m_2)}{|\bar{\Omega}|} \\ &\geq \frac{1}{n}\sum_{i\in(\mathcal{A}\cup\mathcal{B})^c} V(W_{Y_i|X_{1i},X_{2i}}\|Q_{Y_i}|\bar{P}_{X_{1i}},\bar{P}_{X_{2i}}) - 2\delta V' \\ &\geq \frac{n-|\mathcal{A}|-|\mathcal{B}|}{n}(V^* - o(1)) - 2\delta V' \\ &\geq V^* - o(1) - 2\delta V', \end{aligned} \tag{15}$$

where in the last inequality, we have used the fact that both $\mathcal{A}$ and $\mathcal{B}$ have cardinality that is $o(n)$. Applying the bound of (15) to (8), we have

$$\begin{aligned} \log M_1M_2 \leq & nC_{\text{sum}} - \sqrt{n(V^* - 2\delta V' - o(1))}\mathcal{Q}^{-1}(\epsilon) \\ &+ \frac{1}{2}\log n + O(1) + \left(1+\frac{1}{\delta}\right)\log(nI_3+1). \end{aligned}$$

By applying Taylor expansion with respect to δ at $\delta=0$, we have

$$\begin{aligned} &-\sqrt{n(V^* - 2\delta V' + o(1))}\mathcal{Q}^{-1}(\epsilon) \\ &\qquad = -\sqrt{n(V^*+o(1))}\mathcal{Q}^{-1}(\epsilon) + O(\delta\sqrt{n}). \end{aligned}$$

We achieve

$$\min_{\delta} O(\delta\sqrt{n}) + \frac{1}{\delta}\log(nI_3+1) = O(n^{\frac{1}{4}}\sqrt{\log n})$$

which is achieved by $\delta = O(\sqrt{\log n}/n^{1/4})$. Thus, the proof concludes as

$$\begin{aligned} &\log M_1M_2 \\ &\leq nC_{\text{sum}} - \sqrt{nV^* + o(n))}\mathcal{Q}^{-1}(\epsilon) + O(n^{\frac{1}{4}}\sqrt{\log n}) \\ &\leq nC_{\text{sum}} - \sqrt{nV^*}\mathcal{Q}^{-1}(\epsilon) + o(\sqrt{n}). \end{aligned}$$

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. CCF-1908725 and Grant No. CCF-1901243.

References

[1] C. E. Shannon *et al.*, "Two-way communication channels," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1961.

[2] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge university press, 2011.

[3] H. H.-J. Liao, "Multiple access channels," University of Hawaii, Honolulu, Tech. Rep., 1972.

[4] R. Ahlswede, "Multi-way communication channels," in *Second International Symposium on Information Theory: Tsahkadsor, Armenia, USSR, Sept. 2-8, 1971*, 1973.

[5] A. Wyner, "Recent results in the shannon theory," *IEEE Transactions on Information Theory*, vol. 20, no. 1, pp. 2–10, 1974.

[6] T. M. Cover, "Some advances in broadcast channels," in *Advances in Communication Systems*. Elsevier, 1975, vol. 4, pp. 229–260.

[7] G. Dueck, "Maximal error capacity regions are smaller than average error capacity regions for multiple-user channels," *Problems of Control and Information Theory*, vol. 7, pp. 11–19, 1978.

[8] ——, "The strong converse to the coding theorem for the multiple access channel," *Journal of Combinatorics, Information and System Sciences*, vol. 6, no. 3, pp. 187–196, 1981.

[9] R. Ahlswede, "An elementary proof of the strong converse theorem for the multiple access channel," *J. Combinatorics, Information and System Sciences*, vol. 7, no. 3, 1982.

[10] U. Augustin, "Gedächtnisfreie kanäle für diskrete zeit," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 6, no. 1, pp. 10–61, 1966.

[11] S. L. Fong and V. Y. Tan, "A proof of the strong converse theorem for gaussian multiple access channels," *IEEE Transactions on Information Theory*, vol. 62, no. 8, pp. 4376–4394, 2016.

[12] V. Strassen, "Asymptotic estimates in shannon's information theory," in *The 3rd Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, 1969, pp. 689–723.

[13] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[14] 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.

[15] E. Haim, Y. Kochman, and U. Erez, "A note on the dispersion of network problems," in *Electrical Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of*, Nov 2012, pp. 1–9.

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

[17] J. Scarlett, A. Martinez, and A. Guillén 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, Jan 2015.

[18] E. MolavianJazi and J. N. Laneman, "A second-order achievable rate region for gaussian multi-access channels via a central limit theorem for functions," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6719–6733, 2015.

[19] J. Scarlett and V. Y. Tan, "Second-order asymptotics for the discrete memoryless mac with degraded message sets," in *International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 2964–2968.

[20] ——, "Second-order asymptotics for the gaussian mac with degraded message sets," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6700–6718, 2015.

[21] O. Kosut, "A second-order converse bound for the multiple-access channel via wringing dependence," *arXiv preprint arXiv:2007.15664*, 2020.

[22] P. Moulin, "A new metaconverse and outer region for finite-blocklength macs," in *Information Theory and Applications Workshop (ITA)*. IEEE, 2013, pp. 1–8.

[23] V. Y. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *arXiv preprint arXiv:1504.02608*, 2015.