A New Metaconverse and Outer Region for Finite-Blocklength MACs

Pierre Moulin
Dept of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Abstract—An outer rate region for the Discrete Memoryless Multiple Access Channel is given, and is shown to coincide with a recently derived inner rate region within a $o(1/\sqrt{n})$ term—except in the vicinity of the corner points of the Ahlswede-Liao pentagonal region, where the gap is $O(1/\sqrt{n})$. The derivation is based on a new metaconverse under the maximum error probability criterion and on strong large deviations for Neyman-Pearson tests.

I. INTRODUCTION

The study of the fundamental communication limits in the finite-blocklength regime has received tremendous attention in the information theory community, following the works by Strassen [1],Polyanskiy et al. [2], and Hayashi [3]. Extension of the techniques therein to multiuser settings are quite challenging [4], [5], [6], [7], [8], [9], [10]. In this paper we derive a new outer region for multiple-access channels which essentially matches the inner region of [7]. More precisely, we obtain the same second-order coding rate as in [7], except in the vicinity of the corner points of the Ahlswede-Liao pentagonal region.

To obtain the outer region, we derive a strong converse under the maximum error probability criterion. A strong converse for the Ahlswede-Liao region was given by Ahlswede [11] under the average error probability criterion. However Ahlswede’s method requires the use of sophisticated wringing techniques which seem difficult to extend for a more refined asymptotic analysis. It should be noted that while the outer regions for the DM-MAC under the maximum and average error probability criteria do not coincide in general for deterministic codes [12], they do coincide if randomization of the code is allowed. 1 As in [11], the primary technical difficulty is to manage the pairs of codewords that result in atypically good decoding performance. To this end, we derive a new metaconverse, which is inspired by Strassen’s approach for deterministic codes on single-user channels [1].

The problem is stated as follows. Consider a discrete memoryless multiple access channel (DM-MAC) with two input alphabets $X_1$ and $X_2$, output alphabet $Y$, and channel law $W(y|x_1, x_2)$, $x_1 \in X_1$, $x_2 \in X_2$, $y \in Y$. We will often denote by $W_{x_1, x_2}$ the conditional distribution $W(\cdot|x_1, x_2)$ over $Y$. Let $M_1 = \{1, 2, \ldots, M_1\}$ and $M_2 = \{1, 2, \ldots, M_2\}$. A $(n, M_1, M_2)$ MAC code $C$ consists of two codebooks $\{x_1(m_1), m_1 \in M_1\}$ and $\{x_2(m_2), m_2 \in M_2\}$ and a (possibly randomized) decoder $\phi : Y^n \rightarrow M_1 \times M_2$. The decoding error probability for message pair $(m_1, m_2)$ is

$$e(m_1, m_2; C) = \sum_{y \in Y^n} W^n(y|x_1(m_1), x_2(m_2)) \times \mathbb{1}\{\phi(y) \neq (m_1, m_2)\}.$$ 

The average error probability for the code is $e_{\text{avg}}(C) \triangleq \frac{1}{M_1 M_2} \sum_{m_1, m_2} e(m_1, m_2; C)$. The maximum error probability for the code is $e_{\text{max}}(C) \triangleq \max_{m_1, m_2} e(m_1, m_2; C)$. A rate pair $(R_1, R_2)$ is $(n, \varepsilon)$ achievable if there exists a $(n, M_1, M_2)$ MAC code with maximum error probability $\varepsilon$.

Given a finite alphabet $U$ and a joint distribution $P_{U X_1 X_2 Y} = P_U P_{X_1|U} P_{X_2|U} P_Y$, define the random

1In particular, randomized permutation of codeword assignments equalizes the decoding error probability for all possible message pairs.
The following result was given in [7].

Let 

\[ I = \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} I(X_1; Y | X_2 U) \\ I(X_2; Y | X_1 U) \\ I(X_1 X_2; Y | U) \end{pmatrix} \]

and its conditional covariance matrix given \( U \) is

\[ V = \text{Cov}[L|U] = \sum_{u \in \Omega} P_U(u) \text{Cov}[L|U = u]. \tag{3} \]

The diagonal elements of \( V \) are the conditional variances

\[
\begin{align*}
V_1 &= V(X_1; Y | X_2 U) \triangleq \text{Var} \left[ \ln \frac{W(Y | X_1 X_2)}{P(Y | X_2 U)} \right] \\
V_2 &= V(X_2; Y | X_1 U) \triangleq \text{Var} \left[ \ln \frac{W(Y | X_1 X_2)}{P(Y | X_1 U)} \right] \\
V_3 &= V(X_1 X_2; Y | U) \triangleq \text{Var} \left[ \ln \frac{W(Y | X_1 X_2)}{P(Y | U)} \right].
\end{align*} \tag{4} \]

For any two distributions \( P \) and \( Q \) on \( \mathcal{Y} \), let

\[ D(P || Q) = \mathbb{E}_P[\ln \frac{P(Y)}{Q(Y)}] \]

and \( V(P || Q) = \text{Var}_P[\ln \frac{P(Y)}{Q(Y)}] \).

Let \( Z \) be a three-dimensional Gaussian vector with mean zero and covariance matrix \( V \). Define the region \( \mathcal{Q}_{\text{inv}}(V, \epsilon) \triangleq \{ z \in \mathbb{R}^3 : P(Z \leq z) \geq 1 - \epsilon \} \) where the inequality \( Z \leq z \) holds componentwise. The following result was given in [7].

**Theorem 1.1:** For any joint probability distribution \( P_U P_{X_1 | U} P_{X_2 | U} W \), the vectors

\[
\begin{pmatrix} \log M_1 \\ \log M_2 \\ \log(M_1 M_2) \end{pmatrix}
\]

multiplied by \( \frac{1}{n} \), are \((n, \epsilon)\) achievable rate vectors under the average error probability criterion.

Our main result in this paper is as follows.

**Theorem 1.2:** Assume the maximization problems

\[
\begin{align*}
\max_{P_{X_1}, P_{X_2}} I(X_1; Y | X_2) &\triangleq I_1^* \\
\max_{P_{X_1}, P_{X_2}} I(X_2; Y | X_1) &\triangleq I_2^* \\
\max_{P_{X_1}, P_{X_2}} I(X_1 X_2; Y) &\triangleq I_3^*
\end{align*}
\]

have unique solutions. The following holds for every \((n, M_1, M_2)\) MAC code with maximum error probability \( \epsilon \). There exists a joint probability distribution \( P_U P_{X_1 | U} P_{X_2 | U} W \) such that \(|U| \leq 3\) and

\[
\begin{align*}
\log M_1 &\leq nI(X_1; Y | X_2, U) - \sqrt{nV(X_1; Y | X_2, U)} Q^{-1}(\epsilon) + o(\sqrt{n}) \\
\log M_2 &\leq nI(X_2; Y | X_1, U) - \sqrt{nV(X_2; Y | X_1, U)} Q^{-1}(\epsilon) + o(\sqrt{n}) \\
\log(M_1 M_2) &\leq nI(X_1 X_2; Y | U) - \sqrt{nV(X_1 X_2; Y | U)} Q^{-1}(\epsilon) + o(\sqrt{n}).
\end{align*}
\]

By the properties of the function \( Q_{\text{inv}}(V, \epsilon) \), the inner region of Theorem 1.1 and the outer region of Theorem 1.2 agree up to \( o(1/\sqrt{n}) \) except in the vicinity of the corners of the Ahlswede-Liao pentagonal region, where the gap is \( O(1/\sqrt{n}) \).

**II. Meta Converse**

Given two probability distributions \( P \) and \( Q \) for a random variable \( Z \in \mathcal{Z} \), denote by \( \delta : \mathcal{Z} \to [0, 1] \) a randomized decision rule returning \( \delta(z) = \Pr[\text{Say } P | Z = z], z \in \mathcal{Z} \) \), and by \( \beta_{1-\epsilon}(P, Q) \) the type-II error probability of the Neyman-Pearson test at significance level \( 1 - \epsilon \), i.e.

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

**Proposition 2.1:** For any \((n, M_1, M_2)\) MAC code with maximum error probability \( \epsilon \) over channel \( W^n \), the following holds for any subset \( \Omega \) of \( \mathcal{M}_1 \times \mathcal{M}_2 \) and any distribution \( Q_Y \) over \( \mathcal{Y}^n \):

\[
\frac{1}{|\Omega|} \geq \frac{1}{|\Omega|_{(m_1, m_2) \in \Omega}} \sum_{(m_1, m_2) \in \Omega} \beta_{1-\epsilon}(W^n(\cdot|x_1(m_1), x_2(m_2)), Q_Y). \tag{5}
\]

In particular, for any three distributions \( Q_Y, Q_{Y|X_2} \),
and $Q_{Y|X}$, we have

$$\frac{1}{|\Omega|} \geq \frac{1}{|\Omega|} \sum_{(m_1, m_2) \in \Omega} \beta_{1-\epsilon}(W^n(|x_1(m_1), x_2(m_2)|, Q_Y) \tag{6}$$

$$\frac{M_2}{|\Omega|} \geq \frac{1}{|\Omega|} \sum_{(m_1, m_2) \in \Omega} \beta_{1-\epsilon}(W^n(|x_1(m_1), x_2(m_2)|, Q_Y|X=x_2(m_2)) \tag{7}$$

$$\frac{M_1}{|\Omega|} \geq \frac{1}{|\Omega|} \sum_{(m_1, m_2) \in \Omega} \beta_{1-\epsilon}(W^n(|x_1(m_1), x_2(m_2)|, Q_Y|X=x_1(m_1)) \tag{8}$$

This strengthens the metaconverse of [7, Theorem 5], which is based on the single-user metaconverse of [2] and states

$$\frac{1}{|\Omega|} \geq \min_{m_1, m_2 \in \Omega} \beta_{1-\epsilon}(W^n(|x_1(m_1), x_2(m_2)|, Q_Y) \tag{9}$$

in particular,

$$\frac{1}{M_1 M_2} \geq \min_{m_1, m_2} \beta_{1-\epsilon}(W^n(|x_1(m_1), x_2(m_2)|, Q_Y) \tag{10}$$

$$\frac{1}{M_1} \geq \min_{m_1, m_2} \beta_{1-\epsilon}(W^n(|x_1(m_1), x_2(m_2)|, Q_Y|X=x_2(m_2)) \tag{11}$$

$$\frac{1}{M_2} \geq \min_{m_1, m_2} \beta_{1-\epsilon}(W^n(|x_1(m_1), x_2(m_2)|, Q_Y|X=x_1(m_1)) \tag{12}$$

but is insufficient to derive the results claimed in this paper. The disadvantage of (10)—(12) is that there may exist a few codeword pairs for which $\beta_{1-\epsilon}$ is unusually low; to eliminate those, one typically seeks a large subset $\Omega$ of the joint message set that excludes them. As we shall see in Sec. III-A, this is sufficient to get a strong converse and a second-order coding rate, but not the best second-order coding rate. The advantage of the new bound (6)—(8) is that the pairs of concern contribute weakly to the sum and that no joint message set pruning technique is needed.

Proof of Prop. 2.1. The proof is based on Strassen’s method for the converse [1, Sec. 4] and is easiest to follow in the case of deterministic encoding and decoding rules. In that case, the space $Y^n$ is partitioned into decoding regions $D(m_1, m_2)$ such that $W^n(D(m_1, m_2)|x_1(m_1), x_2(m_2)) \geq 1-\epsilon$ for each $(m_1, m_2)$. For any $\Omega \subset M_1 \times M_2$ and any distribution $Q_Y$ over $Y^n$, we therefore have

$$1 \geq \sum_{(m_1, m_2) \in \Omega} Q_Y(D(m_1, m_2)) \tag{a}$$

$$\geq \sum_{(m_1, m_2) \in \Omega} \min_{D \subseteq Y^n : W^n(D|x_1(m_1), x_2(m_2)) \geq 1-\epsilon} Q_Y(D) \tag{b}$$

$$\geq \sum_{(m_1, m_2) \in \Omega} \beta_{1-\epsilon}(W^n(|x_1(m_1), x_2(m_2)|, Q_Y) \tag{c}$$

where (a) holds because the decoding regions are disjoint, and (b) holds by definition of the function $\beta_{1-\epsilon}$. This proves (5), which reduces to (6) when $\Omega = M_1 \times M_2$.

Similarly, let $M_1(m_2) \triangleq \{m_1 : (m_1, m_2) \in \Omega\}$. For any distribution $Q_{Y|X_2}$, we have, for each $m_2$,

$$1 \geq \sum_{m_1 \in M_1(m_2)} Q_{Y|X_2=x_2(m_2)}(D(m_1, m_2)) \tag{13}$$

$$\geq \sum_{m_1 \in M_1(m_2)} \min_{D \subseteq Y^n : W^n(D|x_2(m_2)) \geq 1-\epsilon} Q_{Y|X_2=x_2(m_2)}(D(m_1, m_2)) \tag{14}$$

$$\geq \sum_{m_1 \in M_1(m_2)} \beta_{1-\epsilon}(W^n(|x_1(m_1), x_2(m_2)|, Q_Y|X=x_2(m_2)) \tag{15}$$

Summing these inequalities over $m_2$ proves (7). The final inequality (8) is proved the same way.

The claim holds when the encoders and decoder are stochastic mappings $P_{X_1|m_1}$, $P_{X_2|m_2}$, and $P_{Y|M_1,M_2|Y}$, respectively. To derive (6) in that case, let

$$W(y|m_1, m_2) \triangleq \sum_{x_1, x_2} W^n(y|x_1 x_2) P_{X_1|m_1}(x_1) P_{X_2|m_2}(x_2).$$

For each $(m_1, m_2)$ we have

$$\sum_{y} P_{M_1,M_2|Y}(m_1, m_2|y) W(y|m_1, m_2) \geq 1-\epsilon.$$
For any \((m_1, m_2)\) and any distribution \(Q_Y\) we have
\[
\sum_y P_{\tilde{M}_1, \tilde{M}_2 | Y}(m_1, m_2 | y) Q_Y(y) \\
\geq P_{\tilde{M}_1, \tilde{M}_2 | Y} \cdot \sum_y P_{\tilde{M}_1, \tilde{M}_2 | Y}(m_1, m_2 | y) \mathbb{W}(y | m_1, m_2) \geq 1 - \epsilon \\
\sum_y P_{M_1, M_2 | Y}(m_1, m_2 | y) Q_Y(y) \\
= \beta_{1-\epsilon}(\mathbb{W}(\cdot | m_1, m_2), Q_Y).
\]
Summing over all \((m_1, m_2) \in \Omega\) we obtain
\[
1 \geq \sum_{(m_1, m_2) \in \Omega} \sum_y P_{M_1, M_2 | Y}(m_1, m_2 | y) Q_Y(y) \\
\geq \sum_{(m_1, m_2) \in \Omega} \beta_{1-\epsilon}(\mathbb{W}(\cdot | m_1, m_2), Q_Y)
\]
and similarly for any conditional distributions \(Q_Y_{|m_2}\) and \(Q_Y_{|m_1}\). 

\[1 \geq \sum_{m_1 \in \mathcal{M}_1(m_2)} \beta_{1-\epsilon}(\mathbb{W}(\cdot | m_1, m_2), Q_Y_{|m_2}), \]
\[1 \geq \sum_{m_2 \in \mathcal{M}_2(m_1)} \beta_{1-\epsilon}(\mathbb{W}(\cdot | m_1, m_2), Q_Y_{|m_1}).\]

### III. EARLY ATTEMPTS

Before sketching the proof of Theorem 1.2, we present two instructive attempts at deriving a strong converse and a second-order coding rate. The first is based on the metaconverse of [7, Theorem 5], the second on the new metaconverse of Prop 2.1.

#### A. First Metaconverse

The following technique was presented in [8] and yields an \(O(1/\sqrt{n})\) backoff from the Ahlswede-Liao capacity region, albeit not with the desired second-order coding rate. Define \(n\) pairs of random variables \((X_{1i}, X_{2i}), 1 \leq i \leq n\) that are equal to \((x_{1i}(m_1), x_{2i}(m_2))\) where \((m_1, m_2)\) is drawn uniformly from \(\mathcal{M}_1 \times \mathcal{M}_2\). The joint distribution of \((X_{1i}, X_{2i}, Y_i)\) is given by

\[
P_{X_{1i}, X_{2i}, Y_i}(x_1, x_2, y) \]

\[= P_{X_{1i}}(x_1) P_{X_{2i}}(x_2) W(y | x_1, x_2), \quad 1 \leq i \leq n.\]  

Let \(I_1 \triangleq \sum_{i=1}^n I(X_{1i}; Y_i | X_{2i}), I_2 \triangleq \sum_{i=1}^n I(X_{2i}; Y_i | X_{1i}), I_3 \triangleq \sum_{i=1}^n I(X_{1i}; X_{2i}; Y_i),\) and \(Q_Y = \prod_{i=1}^n P_Y\) and
\[
D(m_1, m_2) = \frac{1}{n} \sum_{i=1}^n D(W_{x_{1i}(m_1), x_{2i}(m_2)} \| P_Y), \quad V(m_1, m_2) = \frac{1}{n} \sum_{i=1}^n V(W_{x_{1i}(m_1), x_{2i}(m_2)} \| P_Y).
\]

The following asymptotics hold for the NP test:
\[
- \log \beta_{1-\epsilon}(W^n(\cdot | x_1, x_2), Q_Y) = nD(m_1, m_2) - \sqrt{\log n} V(m_1, m_2) Q^{-1}(\epsilon) + \frac{1}{2} \log n + O(1). \quad (14)
\]

We have
\[
\frac{1}{M_1 M_2} \sum_{m_1, m_2} D(m_1, m_2) = I_3
\]
hence by Markov’s inequality, there exists a set \(\Omega \subset \mathcal{M}_1 \times \mathcal{M}_2\) such that
\[
\max_{(m_1, m_2) \in \Omega} D(m_1, m_2) < I_3 + \frac{1}{n} \quad \text{and} \quad |\Omega| > \frac{M_1 M_2}{n I_3 + 1} \quad (15)
\]

Then
\[
\log |\Omega| \leq \max_{(m_1, m_2) \in \Omega} - \log \beta_{1-\epsilon}(W^n(\cdot | x_1(m_1), x_2(m_2)), Q_Y) \]
\[= \max_{(m_1, m_2) \in \Omega} \left\{ nD(m_1, m_2) \right\} - \sqrt{n} V(m_1, m_2) Q^{-1}(\epsilon) + \frac{1}{2} \log n + O(1) \]
\[\leq n I_3 - \sqrt{n} V_{min,3} Q^{-1}(\epsilon) + O(\log n)\]

hence
\[
\log(M_1 M_2) \leq n I_3 - \sqrt{n V_{min,3}} Q^{-1}(\epsilon) + O(\log n).
\]
The same technique can be used to derive the bounds
\[
\log M_1 \leq n I_3 - \sqrt{n V_{min,1}} Q^{-1}(\epsilon) + O(\log n) \quad \text{log} M_2 \leq n I_3 - \sqrt{n V_{min,2}} Q^{-1}(\epsilon) + O(\log n).
\]

By taking the supremum with respect to all possible distributions \(P_{X_{1i}}, P_{X_{2i}}, 1 \leq i \leq n\), we obtain a
strong converse for the DM-MAC under the maximum error probability criterion, with a $O(1/\sqrt{n})$ backoff from the Ahlswede-Liao bound on the sum-rate.

Hence the codeword pairs $(m_1, m_2)$ that have atypically large $D(m_1, m_2)$ can be managed by excluding them from the set $\Omega$. It would be nice to use the same technique again to find a subset $\Omega' \subset \Omega$ from which the codeword pairs $(m_1, m_2)$ that have atypically small $V(m_1, m_2)$ are excluded as well. Unfortunately such a set might not be sufficiently large, in fact might be empty.

B. New Metaconverse

**Lemma 3.1:** Fix $A > 0$ and $n > 0$. The function $f(t, u) \triangleq e^{-nt+Au\sqrt{n}}$, $t, u \geq 0$, is convex in $t$ for all $t > 0$ and convex in $u$ for all $u > 1/A$ but is not jointly convex in $(t, u)$.

**Proof:** The partial first derivatives of $f$ are $\frac{\partial f(t,u)}{\partial t} = -nf(t, u)$ and $\frac{\partial f(t,u)}{\partial u} = \frac{A}{2\sqrt{n}} f(t, u)$ and its Hessian is

$$\nabla^2 f(t, u) = \left( \begin{array}{cc} 1 & \frac{-A}{2\sqrt{n}} \\ \frac{-A}{4nu} & \frac{-A^{2}}{4nu} \left( 1 - \frac{1}{A\sqrt{n}} \right) \end{array} \right) n^2 f(t, u).$$

The partial second derivatives of $f$ with respect to $t$ and $u$ are positive over their range, hence $f$ is convex in $t$ and $u$ separately. The determinant of the Hessian equals $\frac{A^2}{4nu^2} f^2(t, u) < 0$, hence $f$ is not jointly convex in $(t, u)$. \hfill \Box

**Remark:** As $n \to \infty$, the (unnormalized) eigenvectors of the Hessian are asymptotic to $(1, -\frac{A}{2\sqrt{n}})$ and $(\frac{A^2}{4nu^2}, 1)$ and are associated with the positive and negative eigenvalues $[1 + \frac{A^2}{4nu} + o(n^{-1})]n^2 f(t, u)$ and $[-\frac{A^2}{4nu^2} + o(n^{-1/2})]n^2 f(t, u)$, respectively.

**Lemma 3.2:** Let $f$ be defined as in Lemma 3.1 and $\mathbb{E}_P f(T, U)$ be a function of a probability distribution $P$, to be minimized subject to the mean-value constraints $\mathbb{E}_P(T) = D$ and $\mathbb{E}_P(U) = V$ and the domain constraint $(t, u) \in \mathbb{R}^+ \times [0, u_{\max}]$. The minimum is equal to $f^* = e^{-nD+AV\sqrt{n/u_{\max}}}$ and achieved by $P$ that has support at the points

$$(t_1, u_1) = \left( D - \frac{AV}{\sqrt{n}u_{\max}}, 0 \right) \quad \text{and}$$

$$(t_2, u_2) = \left( D + \frac{A}{\sqrt{n}} \left( \sqrt{u_{\max}} - \frac{V}{\sqrt{u_{\max}}} \right), u_{\max} \right),$$

(16) the probability of the first point being $1 - \frac{V}{u_{\max}}$.

**Proof:** By the fundamental theorem of linear programming, the minimum is achieved by $P$ that has support at three points $(t_j, u_j), j = 1, 2, 3$ at most, with respective probabilities $\alpha_j, j = 1, 2, 3$. It may be verified that (16) satisfies the first- and second-order Kuhn-Tucker conditions for a minimum, with $f(t_1, u_1) = f(t_2, u_2) = f^*$. \hfill \Box

Lemma 3.2 suggests a second attempt at deriving a second-order coding rate. Using the metaconverse of Prop. 2.1 and the asymptotics of the NP test (14), we obtain

$$M_1 M_2 \leq \frac{1}{M_1 M_2} \sum_{m_1, m_2} \exp \left\{ n D(m_1, m_2) - \sqrt{n V(m_1, m_2)} Q^{-1}(\epsilon) + \frac{1}{2} \log n + O(1) \right\}.$$  

Applying Lemma 3.2 yields a strong converse again, but still not the desired second-order coding rate.

IV. SKETCH OF PROOF OF THEOREM 1.2

The desired coding rate can be obtained by working with random variables that are defined over blocks of samples instead of single samples as done above. Specifically, we partition the codewords into blocks of length $k = \log^2 n$ and consider the output distribution over such blocks. The auxiliary distribution over $Y^n$ for the NP test is blockwise memoryless with block size $k$.

To prove Theorem 1.2, we will apply Prop. 2.1 three times and use the asymptotics of the corresponding NP tests.

**Step 1.** Let $k = \log^2 n$ and partition the set $\{1, \cdots, n\}$ into $n/k$ blocks $B(j), 1 \leq j \leq \frac{n}{k}$ of size $k$. Given any sequence $x \in \mathcal{X}^n$, denote by $x(j) \in \mathcal{X}^k$ its restriction to block $B(j)$, i.e., $x(j) = \{x_i, i \in B(j)\}$. We denote by $x_1(j, m_1) \in \mathcal{X}_1^k$ the subsequence associated with codeword $x_1(m_1)$, and likewise $x_2(j, m_2) \in \mathcal{X}_2^k$ the subsequence associated with codeword $x_2(m_2)$. Define the pairs of random variables $(X_1(j), X_2(j)), 1 \leq j \leq \frac{n}{k}$ that take values $(x_1(j, m_1), x_2(j, m_2))$ where $(m_1, m_2)$ is drawn uniformly from $M_1 \times M_2 \times Mathbb{N}$. Hence $X_1(j)$ and $X_2(j)$ are independent for each $j$. Let $Y(j)$ be the restriction
to $B(j)$ of the random sequence $Y$ at the channel output, i.e., $Y(j) = \{Y_i, i \in B(j)\}$. Hence
\[
P_{Y(j)}(y(j)) = \frac{1}{M_1M_2} \sum_{m_1,m_2} W^k(y(j)|x_1(j,m_1),x_2(j,m_2)).
\]
The joint distribution of $(X_1(j), X_2(j), Y(j))$ is given by
\[
P_{X_1(j),x_2(j),Y(j)}(x_1(j),x_2(j),y(j)) = P_{X_1(j)}(x_1(j)) P_{X_2(j)}(x_2(j)) \times W^k(y(j)|x_1(j),x_2(j)).
\]
Fix the following three product distributions:
\[
Q_Y \triangleq \prod_{i=j}^{n/k} P_{Y(j)} \quad Q_Y|X_2 = \prod_{j=1}^{n/k} P_{Y(j)}|x_2(j),
\]
\[
Q_Y|X_1 = \prod_{j=1}^{n/k} P_{Y(j)}|x_1(j).
\]

**Step 2. Asymptotics of NP tests.** For $Q_Y$ defined in (17), we have
\[
- \log \beta_1 - \epsilon (W^n(\cdot|X_1,X_2), Q_Y)
\]
\[
= \sum_{j=1}^{n/k} D(W_{x_1(j),x_2(j)} || P_{Y(j)}
\]
\[
- \sqrt{\frac{1}{\log n} \sum_{j=1}^{n/k} V(W_{x_1(j),x_2(j)} || P_{Y(j)}|x_1=x_2)}
\]
\[
+ \frac{1}{2} \log n + O(1) \forall x_1 \in X_1^n, x_2 \in X_2^n.
\]
Simlarly
\[
- \log \beta_1 - \epsilon (W^n(\cdot|X_1,X_2), Q_Y|X_2=x_2)
\]
\[
= \sum_{j=1}^{n/k} D(W_{x_1(j),x_2(j)} || P_{Y(j)}|x_2=x_2(j))
\]
\[
- \sqrt{\frac{1}{\log n} \sum_{j=1}^{n/k} V(W_{x_1(j),x_2(j)} || P_{Y(j)}|x_2=x_2(j))} \cdot
\]
\[
+ \frac{1}{2} \log n + O(1) \forall x_1 \in X_1^n, x_2 \in X_2^n.
\]
and
\[
- \log \beta_1 - \epsilon (W^n(\cdot|X_1,X_2), Q_Y|X_1=x_1)
\]
\[
= \sum_{j=1}^{n/k} D(W_{x_1(j),x_2(j)} || P_{Y(j)}|x_1=x_1(j))
\]
\[
- \sqrt{\frac{1}{\log n} \sum_{j=1}^{n/k} V(W_{x_1(j),x_2(j)} || P_{Y(j)}|x_1=x_1(j))} \cdot
\]
\[
+ \frac{1}{2} \log n + O(1) \forall x_1 \in X_1^n, x_2 \in X_2^n.
\]

**Step 3. Let**
\[
d_j(m_1,m_2) \triangleq D(W_{x_1(j,m_1),x_2(j,m_2)} || P_{Y(j)})
\]
\[
v_j(m_1,m_2) \triangleq V(W_{x_1(j,m_1),x_2(j,m_2)} || P_{Y(j)})
\]
\[
T_j \triangleq I(X_1(j),X_2(j); Y(j))
\]
\[
= \frac{1}{M_1M_2} \sum_{m_1,m_2} d_j(m_1,m_2)
\]
\[
V_j \triangleq V(X_1(j),X_2(j); Y(j))
\]
\[
= \frac{1}{M_1M_2} \sum_{m_1,m_2} v_j(m_1,m_2)
\]
\[
I^*_3 = \sup_{P_{X_1}, P_{X_2}} I(X_1,X_2; Y)
\]
achieved at unique $(P^*_{X_1}, P^*_{X_2})$.

and define the set of block indices
\[
\mathcal{J} = \left\{ j \in \{1,2,\cdots,n/k\} : \frac{1}{k} T_j \geq I^*_3 - \frac{1}{\log^{3/2} n} \right\}
\]

By definition of $(P^*_{X_1}, P^*_{X_2})$, we have
\[
j \in \mathcal{J} \Rightarrow \frac{1}{k} T_j > V^*_3 - O\left(\frac{1}{\log^{3/2} n}\right).
\]

For $j \in \mathcal{J}$, we have $k I^*_3 \geq T_j \geq k I^*_3 - \frac{1}{\log n}$ and thus
\[
D(P_{X_1(j)} P_{X_2(j)} || (P^*_{X_1})^k (P^*_{X_2})^k) \to 0 \text{ as } k \to \infty.
\]
For any such block $j \in \mathcal{J}$, the random variable
\[
V_j \triangleq V(W_{x_1(j,m_1),x_2(j,m_2)} || P_{Y(j)}) = v_j(m_1,m_2),
\]
multiplied by $\frac{1}{k}$, converges to an average of $k$ iid random variables and therefore converges in probability
to its expectation \( \frac{1}{k} V_j \) as \( k \to \infty \). Moreover
\[
\begin{align*}
P \left[ V_j < V_j - \frac{k}{\log k} \right] & \leq \exp \left\{ -O \left( \frac{k}{\log^2 k} \right) \right\} \\
\forall j & \in \mathcal{J}.
\end{align*}
\]

The probability of the event
\[
\mathcal{E} = \left\{ \sum_{j \in \mathcal{J}} V_j - \sum_{j \in \mathcal{J}} \frac{k|\mathcal{J}|}{\log k} \right\}
\]

is upper-bounded by
\[
\begin{align*}
P[\mathcal{E}] & \leq P \left[ \exists j \in \mathcal{J} : V_j < V_j - \frac{k}{\log k} \right] \\
& \leq |\mathcal{J}| \max_{j \in \mathcal{J}} P \left[ V_j < V_j - \frac{k}{\log k} \right] \\
& \leq |\mathcal{J}| \exp \left\{ -O \left( \frac{k}{\log^2 k} \right) \right\} \\
& = \exp \left\{ -O \left( \frac{\log^2 n}{\log^2 \log n} \right) \right\}
\end{align*}
\]

which vanishes superpolynomially fast as \( n \to \infty \).

Recalling (29), for \( \mathcal{J} = \emptyset \) let \( \Omega = \mathcal{M}_1 \times \mathcal{M}_2 \), and for \( \mathcal{J} \neq \emptyset \) let
\[
\Omega = \left\{ (m_1, m_2) : \sum_{j \in \mathcal{J}} v_j(m_1, m_2) \geq \sum_{j \in \mathcal{J}} V_j - \frac{k|\mathcal{J}|}{\log k} \right\}
\]

where (30) implies that
\[
1 - \frac{\Omega}{\mathcal{M}_1 \mathcal{M}_2} = \exp \left\{ -O \left( \frac{\log^2 n}{\log^2 \log n} \right) \right\}.
\]

In all cases, applying (23) we have
\[
\begin{align*}
\frac{1}{\Omega} \sum_{(m_1, m_2) \in \Omega} \sum_{j=1}^{n/k} d_j(m_1, m_2) \\
& = \sum_{j=1}^{n/k} T_j + \exp \left\{ -O \left( \frac{\log^2 n}{\log^2 \log n} \right) \right\}.
\end{align*}
\]

**Step 4.** We have
\[
\begin{align*}
& \frac{1}{\Omega} \sum_{(m_1, m_2) \in \Omega} \beta_1 - \epsilon (W^n(x_1(m_1), x_2(m_2)), Q_Y) \\
& \geq \frac{1}{\Omega} \sum_{(m_1, m_2) \in \Omega} \exp \left\{ - \sum_{j=1}^{n/k} d_j(m_1, m_2) \\
& \quad + \sum_{j \in \mathcal{J}} v_j(m_1, m_2) Q^{-1}(\epsilon) - \frac{1}{2} \log n + O(1) \right\} \\
& \geq \exp \left\{ - \sum_{j=1}^{n/k} T_j + \sum_{j \in \mathcal{J}} V_j - k|\mathcal{J}| Q^{-1}(\epsilon) - \frac{1}{2} \log n + O(1) \right\} \\
& \geq \exp \left\{ - \sum_{j=1}^{n/k} T_j - |\mathcal{J}| (I_3^* - \frac{1}{\log^3 n}) \\
& \quad + \sum_{j \in \mathcal{J}} V_j - k|\mathcal{J}| Q^{-1}(\epsilon) - \frac{1}{2} \log n + O(1) \right\} \\
& \geq \exp \left\{ -nI_3^* + \frac{|\mathcal{J}|}{\log^3 n} \\
& \quad + \sum_{j \in \mathcal{J}} V_j - k|\mathcal{J}| Q^{-1}(\epsilon) - \frac{1}{2} \log n + O(1) \right\} \\
& \geq \exp \left\{ -nI_3^* + \sqrt{n} V_3^* Q^{-1}(\epsilon) + o(\sqrt{n}) \right\}
\end{align*}
\]
where (a) follows from (5), (b) from (18) (21) (22), (c) from definition of the set \( \Omega \) in (31), (d) from the convexity of the \( \exp(\cdot) \) function, (e) from (23) and (33), (f) from the definition of \( \mathcal{J} \) in (27), (g) from (28) and the fact that \( \bar{T}_j \leq kI^*_j \), and (h) holds because the minimum over \( \mathcal{J} \), namely, \( \mathcal{J} = \{1, \cdots, \frac{2}{k}\} \).

By (32), we obtain
\[
\log(M_1M_2) \leq nI^*_3 - \sqrt{nV^*_3} Q^{-1}(\epsilon) + o(\sqrt{n})
= \sup_{P_{X_1}, P_{X_2}} [nI(X_1X_2; Y)
- \sqrt{nV(X_1X_2; Y)} Q^{-1}(\epsilon)]
+ o(\sqrt{n}).
\tag{34}
\]

**Step 4.** Similarly to (34), we obtain
\[
\log M_1 \leq \sup_{P_{X_1}, P_{X_2}} [nI(X_1; Y|X_2)
- \sqrt{nV(X_1; Y|X_2)} Q^{-1}(\epsilon)]
+ o(\sqrt{n}).
\tag{35}
\]
and
\[
\log M_2 \leq \sup_{P_{X_1}, P_{X_2}} [nI(X_2; Y|X_1)
- \sqrt{nV(X_2; Y|X_1)} Q^{-1}(\epsilon)]
+ o(\sqrt{n}).
\tag{36}
\]

**Step 5.** Choosing \( \mathcal{U} = \{1, 2, 3\} \) and defining the three pairs \( (P_{X_1|U=u}, P_{X_2|U=u}) \), \( u = 1, 2, 3 \) as those that achieve the suprema in (35), (36), and (34) respectively, proves the claim. \( \square \)

**V. Example**

Consider the binary erasure MAC channel with input alphabets \( X_1 = X_2 = \{0, 1\} \), output alphabet \( Y = \{0, 1, 2, e\} \) and \( Y = \{ X_1 + X_2 \wp 1 - \lambda \, e \wp \lambda \} \) (erasure probability \( \lambda \)). No time-sharing is needed; the optimal input distributions \( P_{X_1}^* \) and \( P_{X_2}^* \) are uniform, and
\[
P_Y^* = (P_{X_1}^* P_{X_2}^* W) = \left\{ \frac{1 - \lambda}{4}, \frac{1 - \lambda}{2}, \frac{1 - \lambda}{4}, \lambda \right\}.
\]
The likelihood ratio vectors over \( Y \) are given by
\[
\frac{W_{0,0}}{P_Y^*} = \{4, 0, 0, 1\}, \quad \frac{W_{0,1}}{P_Y^*} = \{0, 2, 0, 1\},
\frac{W_{1,0}}{P_Y^*} = \{0, 0, 2, 1\}, \quad \frac{W_{1,1}}{P_Y^*} = \{0, 0, 4, 1\}.
\]
The mean and variance of the loglikelihood ratios are given by
\[
D(W_{00}||P_Y^*) = D(W_{11}||P_Y^*) = 2(1 - \lambda)
\]
\[
D(W_{01}||P_Y^*) = D(W_{10}||P_Y^*) = 1 - \lambda \tag{37}
\]
\[
V(W_{00}||P_Y^*) = V(W_{11}||P_Y^*) = 4\lambda(1 - \lambda)
\]
\[
V(W_{01}||P_Y^*) = V(W_{10}||P_Y^*) = \lambda(1 - \lambda) \tag{38}
\]
hence \( I^*_3 = I(P_{X_1}^* P_{X_2}^* W) = \frac{\lambda}{2} (1 - \lambda) \) and \( V^*_3 = V(P_{X_1}^* P_{X_2}^* W) = \frac{4}{2} \lambda (1 - \lambda) \). Note from (37) that the most informative input pairs are \((x_1, x_2) = (0, 0)\) and \((1, 1)\), however (38) shows that the information variance is also largest in this case.

**Acknowledgements.** This work was supported by NSF under grant CCF 12-19145. The author thanks Yen-Wei Huang for valuable comments about an earlier version of this paper.

**References**


