Second-Order Asymptotics for the Discrete Memoryless MAC with Degraded Message Sets

Jonathan Scarlett
Laboratory for Information and Inference Systems
École Polytechnique Fédérale de Lausanne
Email: jmscarlett@gmail.com

Vincent Y. F. Tan
Dept. of Elec. and Comp. Eng. and Dept. of Mathematics,
National University of Singapore
Email: vtan@nus.edu.sg

***Abstract*—This paper studies the second-order asymptotics of the discrete memoryless multiple-access channel with degraded message sets. For a fixed average error probability $\varepsilon \in (0,1)$ and an arbitrary point on the boundary of the capacity region, we characterize the speed of convergence of rate pairs that converge to that point for codes that have asymptotic error probability no larger than ε, thus complementing an analogous result given previously for the Gaussian setting.**

I. Introduction

In recent years, there has been great interest in characterizing the fixed-error asymptotics (e.g. dispersion, the Gaussian approximation) of source coding and channel coding problems, and the behavior is well-understood for a variety of single-user settings [1]–[3]. On the other hand, analogous studies of multi-user problems have generally had significantly less success, with the main exceptions being Slepian-Wolf coding [4], [5], the Gaussian interference channel with strictly very strong interference [6], and the Gaussian multiple-access channel (MAC) with degraded message sets [7].

In this paper, we complement our work on the latter problem by considering its discrete counterpart. By obtaining matching achievability and converse results, we provide the first complete characterization of the second-order asymptotics for a discrete channel-type network information theory problem.

A. System Setup

We consider the two-user discrete memoryless MAC (DM-MAC) with degraded message sets [8, Ex. 5.18], with input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ and output alphabet $\mathcal{Y}$. As usual, there are two messages m_1 and m_2, equiprobable on the sets $\{1,\dots,M_1\}$ and $\{1,\dots,M_2\}$ respectively. The first user knows both messages, whereas the second user only knows m_2. Given these messages, the users transmit the codewords $\mathbf{x}_1(m_1,m_2)$ and $\mathbf{x}_2(m_2)$ from their respective codebooks, and the decoder receives a noisy output sequence which is generated according to the memoryless transition law $W^n(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) = \prod_{i=1}^n W(y_i|x_{1,i},x_{2,i})$. An estimate $(\hat{m}_1,\hat{m}_2)$ is formed, and an error is said to have occurred if $(\hat{m}_1,\hat{m}_2) \ne (m_1,m_2)$.

The capacity region $\mathcal{C}$ is given by the set of rate pairs (R_1,R_2) satisfying [8, Ex. 5.18]

$$R_1 \le I(X_1;Y|X_2) \tag{1}$$

$$R_1 + R_2 \le I(X_1,X_2;Y) \tag{2}$$

for some input joint distribution $P_{X_1X_2}$, where the mutual information quantities are with respect to $P_{X_1X_2}(x_1,x_2)W(y|x_1,x_2)$. The achievability part is proved using superposition coding.

We formulate the second-order asymptotics according to the following definition [5].

Definition 1 (Second-Order Coding Rates)**.** *Fix $\varepsilon \in (0,1)$, and let (R_1^*,R_2^*) be a pair of rates on the boundary of $\mathcal{C}$. A pair (L_1,L_2) is $(\varepsilon,R_1^*,R_2^*)$*-achievable *if there exists a sequence of codes with length n, number of codewords for message $j=1,2$ equal to $M_{j,n}$, and average error probability ε_n, such that*

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}}(\log M_{j,n} - nR_j^*) \ge L_j, \quad j=1,2, \tag{3}$$

$$\limsup_{n\to\infty} \varepsilon_n \le \varepsilon. \tag{4}$$

The $(\varepsilon,R_1^*,R_2^*)$-optimal second-order coding rate region *$\mathcal{L}(\varepsilon;R_1^*,R_2^*) \subset \mathbb{R}^2$ is defined to be the closure of the set of all $(\varepsilon,R_1^*,R_2^*)$-achievable rate pairs (L_1,L_2).*

Throughout the paper, we write non-asymptotic rates as $R_{1,n} := \frac{1}{n}\log M_{1,n}$ and $R_{2,n} := \frac{1}{n}\log M_{2,n}$. Roughly speaking, the preceding definition is concerned with ε-reliable codes such that $R_{j,n} \ge R_j^* + \frac{1}{\sqrt{n}}L_j + o\big(\frac{1}{\sqrt{n}}\big)$ for $j=1,2$.

We will also use the following standard definition: A rate pair (R_1,R_2) is (n,ε)-*achievable* if there exists a length-n code having an average error probability no higher than ε, and whose rate is at least R_j for message $j=1,2$.

B. Notation

Except where stated otherwise,[1] the i-th entry of a vector (e.g. $\mathbf{y}$) is denoted using a subscript (e.g. y_i). For two vectors of the same length $\mathbf{a},\mathbf{b} \in \mathbb{R}^d$, the notation $\mathbf{a} \le \mathbf{b}$ means that $a_j \le b_j$ for all j. The notation $\mathcal{N}(\mathbf{u};\boldsymbol{\mu},\boldsymbol{\Lambda})$ denotes the multivariate Gaussian probability density function (pdf) with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Lambda}$. We use the standard asymptotic notations $O(\cdot)$, $o(\cdot)$, $\Theta(\cdot)$, and $\omega(\cdot)$. All logarithms have base e, and all rates have units of nats. The closure operation is denoted by $\mathrm{cl}(\cdot)$.

The set of all probability distributions on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$, and the set of all types [9, Ch. 2] is

[1]For example, the vectors in (5)–(6), do not adhere to this convention.

denoted by $\mathcal{P}_n(\mathcal{X})$. For a given type $Q_X \in \mathcal{P}_n(\mathcal{X})$, we define the type class $T^n(Q_X)$ to be the set of sequences having type Q_X. Similarly, given a conditional type $Q_{Y|X}$ and a sequence $\mathbf{x} \in T^n(Q_X)$, we define $T^n_{\mathbf{x}}(Q_{Y|X})$ to be the set of sequences $\mathbf{y}$ such that $(\mathbf{x},\mathbf{y}) \in T^n(Q_X \times Q_{Y|X})$.

II. Main Result

A. Preliminary Definitions

Given the rate pairs $(R_{1,n}, R_{2,n})$ and (R_1^*, R_2^*), we define

$$\mathbf{R}_n := \begin{bmatrix} R_{1,n} \\ R_{1,n}+R_{2,n} \end{bmatrix}, \quad \mathbf{R}^* := \begin{bmatrix} R_1^* \\ R_1^*+R_2^* \end{bmatrix} \tag{5}$$

Similarly, given the second-order rate pair (L_1, L_2), we write

$$\mathbf{L} := \begin{bmatrix} L_1 \\ L_1+L_2 \end{bmatrix} \tag{6}$$

Given a joint input distribution $P_{X_1X_2} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$, we define $P_{X_1X_2Y} := P_{X_1X_2} \times W$, and denote the induced marginals by $P_{Y|X_1}$, P_Y, etc. We define the following *information density vector*, which implicitly depends on $P_{X_1X_2}$:

$$\mathbf{j}(x_1,x_2,y) := \begin{bmatrix} j_1(x_1,x_2,y) & j_{12}(x_1,x_2,y) \end{bmatrix}^T \tag{7}$$

$$= \begin{bmatrix} \log \dfrac{W(y|x_1,x_2)}{P_{Y|X_2}(y|x_2)} & \log \dfrac{W(y|x_1,x_2)}{P_Y(y)} \end{bmatrix}^T. \tag{8}$$

The mean and conditional covariance matrix are given by

$$\mathbf{I}(P_{X_1X_2}) = \mathbb{E}\big[\mathbf{j}(X_1,X_2,Y)\big], \tag{9}$$

$$\mathbf{V}(P_{X_1X_2}) = \mathbb{E}\big[\operatorname{Cov}\big(\mathbf{j}(X_1,X_2,Y)\,\big|\,X_1,X_2\big)\big]. \tag{10}$$

Observe that the entries of $\mathbf{I}(P_{X_1X_2})$ are the mutual informations appearing in (1)–(2). We write the entries of $\mathbf{I}$ and $\mathbf{V}$ using subscripts as follows:

$$\mathbf{I}(P_{X_1X_2}) = \begin{bmatrix} I_1(P_{X_1X_2}) \\ I_{12}(P_{X_1X_2}) \end{bmatrix}, \tag{11}$$

$$\mathbf{V}(P_{X_1X_2}) = \begin{bmatrix} V_1(P_{X_1X_2}) & V_{1,12}(P_{X_1X_2}) \\ V_{1,12}(P_{X_1X_2}) & V_{12}(P_{X_1X_2}) \end{bmatrix}, \tag{12}$$

For a given point $(z_1,z_2) \in \mathbb{R}^2$ and a positive semi-definite matrix $\mathbf{V}$, we define the multivariate Gaussian cumulative distribution function (CDF)

$$\Psi(z_1,z_2;\mathbf{V}) := \int_{-\infty}^{z_2}\int_{-\infty}^{z_1} \mathcal{N}(\mathbf{u};\mathbf{0},\mathbf{V})\,\mathrm{d}\mathbf{u}, \tag{13}$$

and for a given $\varepsilon \in (0,1)$, we define the corresponding "inverse" set

$$\Psi^{-1}(\mathbf{V},\varepsilon) := \big\{(z_1,z_2) \in \mathbb{R}^2 : \Psi(-z_1,-z_2;\mathbf{V}) \ge 1-\varepsilon\big\}. \tag{14}$$

Similarly, we let $\Phi(\cdot)$ denote the standard Gaussian CDF, and we denote its functional inverse by $\Phi^{-1}(\cdot)$. Moreover, we let

$$\Pi(R_1^*,R_2^*) := \big\{P_{X_1X_2} : \mathbf{I}(P_{X_1X_2}) \ge \mathbf{R}^*\big\} \tag{15}$$

be the set of input distributions achieving a given point (R_1^*, R_2^*) of the boundary of $\mathcal{C}$. Note that in contrast with the single-user setting [1]–[3], this definition uses an inequality rather than an equality, as one of the mutual information quantities may be strictly larger than the corresponding entry of $\mathbf{R}^*$ and yet be first-order optimal. For example, assuming that the capacity region on the left of Figure 1 is achieved by a single input distribution, all points (R_1^*, R_2^*) on the vertical boundary satisfy $I_{12}(P_{X_1X_2}) > R_1^* + R_2^*$.

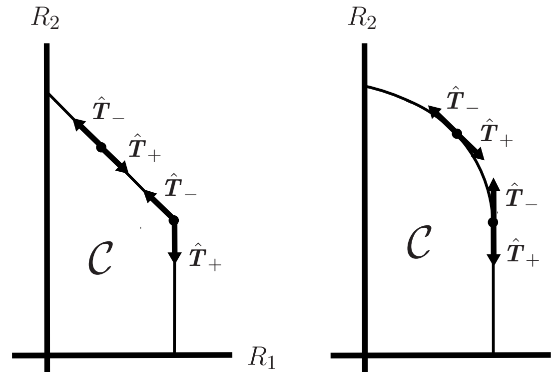

Fig. 1. Unit tangent vectors $\hat{\mathbf{T}}_-$ and $\hat{\mathbf{T}}_+$ for two boundary points (R_1^*, R_2^*) of two hypothetical capacity regions.

The preceding definitions are analogous to those appearing in previous works such as [4], while the remaining definitions are somewhat less standard. Given the boundary point (R_1^*, R_2^*), we let $\hat{\mathbf{T}}_- := \hat{\mathbf{T}}_-(R_1^*, R_2^*)$ and $\hat{\mathbf{T}}_+ := \hat{\mathbf{T}}_+(R_1^*, R_2^*)$ denote the left and right unit tangent vectors along the boundary of $\mathcal{C}$ in (R_1, R_2) space; see Figure 1 for an illustration. Furthermore, we define

$$\mathbf{T}_- := \begin{bmatrix} \hat{T}_{-,1} \\ \hat{T}_{-,1}+\hat{T}_{-,2} \end{bmatrix}, \quad \mathbf{T}_+ := \begin{bmatrix} \hat{T}_{+,1} \\ \hat{T}_{+,1}+\hat{T}_{+,2} \end{bmatrix}. \tag{16}$$

It is understood that $\hat{\mathbf{T}}_-$ and $\mathbf{T}_-$ (respectively, $\hat{\mathbf{T}}_+$ and $\mathbf{T}_+$) are undefined when $R_1^* = 0$ (respectively, $R_2^* = 0$). As is observed in Figure 1, we have $\hat{\mathbf{T}}_- = -\hat{\mathbf{T}}_+$ on the curved and straight-line parts of $\mathcal{C}$, and $\hat{\mathbf{T}}_- \neq -\hat{\mathbf{T}}_+$ when there is a sudden change in slope (e.g. at a corner point).

The following set of vectors can be thought of as those that point strictly inside $\mathcal{C}$ when placed at (R_1^*, R_2^*):

$$\hat{\mathcal{V}}(R_1^*,R_2^*) := \big\{\mathbf{v} \in \mathbb{R}^2 : (R_1^*,R_2^*) + \alpha\mathbf{v} \in \mathcal{C} \text{ for some } \alpha > 0\big\}. \tag{17}$$

Using this definition, we set

$$\mathcal{V}(R_1^*,R_2^*) := \mathrm{cl}\Bigg(\bigcup_{(v_1,v_2)\in\hat{\mathcal{V}}(R_1^*,R_2^*)} \big\{(v_1, v_1+v_2)\big\}\Bigg). \tag{18}$$

Due to the closure operation, it is readily verified that $\mathbf{T}_- \in \mathcal{V}$ and $\mathbf{T}_+ \in \mathcal{V}$.

B. Statement of Main Result

For a given boundary point (R_1^*, R_2^*) and input distribution $P_{X_1X_2} \in \Pi(R_1^*, R_2^*)$, we define the set $\mathcal{L}_0(\varepsilon; R_1^*, R_2^*, P_{X_1X_2})$ separately for the following three cases:

(i) If $R_1^* = I_1(P_{X_1X_2})$ and $R_1^* + R_2^* < I_{12}(P_{X_1X_2})$, then

$$\mathcal{L}_0 = \Big\{(L_1,L_2) : L_1 \le \sqrt{V_1(P_{X_1X_2})}\,\Phi^{-1}(\varepsilon)\Big\} \tag{19}$$

(ii) If $R_1^* < I_1(P_{X_1X_2})$ and $R_1^* + R_2^* = I_{12}(P_{X_1X_2})$, then

$$\mathcal{L}_0 = \Big\{(L_1, L_2) \,:\, L_1 + L_2 \le \sqrt{V_{12}(P_{X_1X_2})}\Phi^{-1}(\varepsilon)\Big\} \quad (20)$$

(iii) If $R_1^* = I_1(P_{X_1X_2})$ and $R_1^* + R_2^* = I_{12}(P_{X_1X_2})$, then

$$\mathcal{L}_0 = \Big\{(L_1, L_2) \,:\, \mathbf{L} \in \bigcup_{\beta \ge 0} \big\{\beta \mathbf{T}_- + \Psi^{-1}(\mathbf{V}(P_{X_1X_2}), \varepsilon)\big\}\Big\}$$
$$\cup \Big\{(L_1, L_2) \,:\, \mathbf{L} \in \bigcup_{\beta \ge 0} \big\{\beta \mathbf{T}_+ + \Psi^{-1}(\mathbf{V}(P_{X_1X_2}), \varepsilon)\big\}\Big\}, \quad (21)$$

where the first (respectively, second) set in the union is understood to be empty when $R_1^* = 0$ (respectively, $R_2^* = 0$).

We are now in a position to state our main result.

Theorem 1. *For any point (R_1^*, R_2^*) on the boundary of the capacity region, and any $\varepsilon \in (0, 1)$, we have*

$$\mathcal{L}(\varepsilon; R_1^*, R_2^*) = \bigcup_{P_{X_1X_2} \in \Pi(R_1^*, R_2^*)} \mathcal{L}_0(\varepsilon; R_1^*, R_2^*, P_{X_1X_2}). \quad (22)$$

Proof: See Section III. ■

Suppose that $\mathcal{X}_1 = \emptyset$ and $(R_1^*, R_2^*) = (0, C)$, where $C := \max_{P_{X_2}} I(P_{X_2}, W)$ and $W : \mathcal{X}_2 \to \mathcal{Y}$. Clearly L_1 plays no role, and Theorem 1 states that the achievable values of L_2 are precisely those in the set

$$\mathcal{L}_2(\varepsilon) := \bigcup_{P_{X_2} \in \Pi} \Big\{L_2 : L_2 \le \sqrt{V(P_{X_2})}\Phi^{-1}(\varepsilon)\Big\}, \quad (23)$$

where $\Pi := \{P_{X_2} : I(P_{X_2}, W) = C\}$, and $V(\cdot) := V_{12}(\cdot)$ is the conditional information variance [3]. Letting $L^* := \sup \mathcal{L}_2(\varepsilon)$ be the *second-order coding rate* [2] of the discrete memoryless channel (DMC) $W : \mathcal{X}_2 \to \mathcal{Y}$, we readily obtain

$$L^* = \begin{cases} \sqrt{\min_{P_{X_2} \in \Pi} V(P_{X_2})}\, \Phi^{-1}(\varepsilon) & \varepsilon < \frac{1}{2} \\ \sqrt{\max_{P_{X_2} \in \Pi} V(P_{X_2})}\, \Phi^{-1}(\varepsilon) & \varepsilon \ge \frac{1}{2}. \end{cases} \quad (24)$$

Thus, our main result reduces to the classical result of Strassen [1, Thm. 3.1] for the single-user setting (see also [2], [3]). This illustrates the necessity of the set $\Pi(R_1^*, R_2^*)$ in the characterization of $\mathcal{L}(\varepsilon; R_1^*, R_2^*)$ in Theorem 1. Such a set is not needed in the Gaussian setting [7], as every boundary point is achieved uniquely by a single multivariate Gaussian distribution. Another notable difference in Theorem 1 compared to [7] is the use of left and right tangent vectors instead of a single derivative vector.

Both of the preceding differences were also recently observed in an achievability result for the standard MAC [10]. However, no converse results were given in [10], and the main novelty of the present paper is in the converse proof.

It is not difficult to show that $\mathcal{L}$ equals a half-space whenever $\hat{\mathbf{T}}_- = -\hat{\mathbf{T}}_+$, as was observed in [7], [10]. A less obvious fact is that the unions over β in (21) can be replaced by $\beta = 0$ whenever the corresponding input distribution $P_{X_1X_2}$ achieves all of the boundary points in a neighborhood of (R_1^*, R_2^*). We refer the reader to [7], [10] for further discussions and illustrative numerical examples.

III. Proof of Theorem 1

Due to space constraints, we do not attempt to make the proof self-contained. We avoid repeating the parts in common with [7], [10], and we focus on the most novel aspects.

A. Achievability

The achievability part of Theorem 1 is proved using a similar (yet simpler) argument to that of the standard MAC given in [10], so we only provide a brief outline.

We use constant-composition superposition coding with coded time sharing [8, Sec. 4.5.3]. We set $\mathcal{U} := \{1, 2\}$, fix a joint distribution $Q_{UX_1X_2}$ (to be specified shortly), and let $Q_{UX_1X_2,n}$ be the closest corresponding joint type. We write the marginal distributions in the usual way (e.g. $Q_{X_1|U,n}$). We let $\mathbf{u}$ be a deterministic time-sharing sequence with $nQ_{U,n}(1)$ ones and $nQ_{U,n}(2)$ twos. We first generate the $M_{2,n}$ codewords of user 2 independently according to the uniform distribution on $T_{\mathbf{u}}^n(Q_{X_1|U,n})$. For each m_2, we generate $M_{1,n}$ codewords for user 1 conditionally independently according to the uniform distribution on $T_{\mathbf{u}\mathbf{x}_2}^n(Q_{X_1|X_2U,n})$, where $\mathbf{x}_2$ is the codeword for user 2 corresponding to m_2.

We fix $\beta \ge 0$ and choose $Q_{UX_1X_2}$ such that $Q_U(1) = 1 - \frac{\beta}{\sqrt{n}}$ and $Q_U(2) = \frac{\beta}{\sqrt{n}}$, let $Q_{X_1X_2|U=1}$ be an input distribution $P_{X_1X_2}$ achieving the boundary point of interest, and let $Q_{X_1X_2|U=2}$ be an input distribution $P'_{X_1X_2}$ achieving a different boundary point. We define the shorthands $\mathbf{I} := \mathbf{I}(P_{X_1X_2})$, $\mathbf{V} := \mathbf{V}(P_{X_1X_2})$ and $\mathbf{I}' := \mathbf{I}(P'_{X_1X_2})$. Using the generalized Feinstein bound given in [7] along with the multivariate Berry-Esseen theorem, we can use the arguments of [10] to conclude that all rate pairs $(R_{1,n}, R_{2,n})$ satisfying

$$\mathbf{R}_n \in \mathbf{I} + \frac{1}{\sqrt{n}}\Big(\beta(\mathbf{I}' - \mathbf{I}) + \Psi^{-1}(\mathbf{V}, \varepsilon)\Big) + g(n)\mathbf{1} \quad (25)$$

are (n, ε)-achievable for some $g(n) = O\big(n^{1/4}\big)$ depending on ε, β, $P_{X_1X_2}$ and $P'_{X_1X_2}$. Note that this argument may require a reduction to a lower dimension for singular dispersion matrices; an analogous reduction will be given in the converse proof below.

The achievability part of Theorem 1 now follows as in [10]. In the cases corresponding to (19)–(20), we eliminate one of the two element-wise inequalities from (25) to obtain the desired result. For the remaining case corresponding to (21), we obtain the first (respectively, second) term in the union by letting $P'_{X_1X_2}$ achieve a boundary point approaching (R_1^*, R_2^*) from the left (respectively, right).

B. Converse

The converse proof builds on that for the Gaussian case [7], but contains more new ideas compared to the achievability part. We thus provide a more detailed treatment.

1) A Reduction from Average Error to Maximal Error: Using an identical argument to the Gaussian case [7] (which itself builds on [9, Cor. 16.2]), we can show that $\mathcal{L}(\varepsilon; R_1^*, R_2^*)$ is identical when the average error probability is replaced by the maximal error probability in Def. 1. We may thus proceed by considering the maximal error probability. Note that neither this step nor the following step are possible for the standard MAC; the assumption of degraded message sets is crucial.

2) A Reduction to Constant-Composition Codes: Using the previous step and the fact that the number of joint types on $\mathcal{X}_1 \times \mathcal{X}_2$ is polynomial in n, we can again follow an identical argument to the Gaussian case [7] to show that $\mathcal{L}(\varepsilon; R_1^*, R_2^*)$ is unchanged when the codebook is restricted to contain codeword pairs $(\mathbf{x}_1, \mathbf{x}_2)$ sharing a common joint type. We thus limit our attention to such codebooks; we denote the corresponding sequence of joint types by $\{P_{X_1 X_2, n}\}_{n \geq 1}$.

3) Passage to a Convergent Subsequence: Since $\mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ is compact, the sequence $\{P_{X_1 X_2, n}\}_{n \geq 1}$ must have a convergent subsequence, say indexed by a sequence $\{n_k\}_{k \geq 1}$ of block lengths. We henceforth limit our attention to values of n on this subsequence. To avoid cumbersome notation, we continue writing n instead of n_k. However, it should be understood that asymptotic notations such as $O(\cdot)$ and $(\cdot)_n \to (\cdot)$ are taken with respect to this subsequence. The idea is that it suffices to prove the converse result only for values of n on an arbitrary subsequence of $(1, 2, 3, \dots)$, since we used the $\liminf$ in (3) and the $\limsup$ in (4).

4) A Verdú-Han-Type Converse Bound: We make use of the following non-asymptotic converse bound from [7]:

$$\varepsilon_n \geq 1 - \Pr\left(\frac{1}{n}\sum_{i=1}^n \mathbf{j}(X_{1,i}, X_{2,i}, Y_i) \geq \mathbf{R}_n - \gamma \mathbf{1}\right) - 2e^{-n\gamma}, \tag{26}$$

where γ is an arbitrary constant, $(\mathbf{X}_1, \mathbf{X}_2)$ is the random pair induced by the codebook, and $\mathbf{Y}$ is the resulting output. The output distributions defining $\mathbf{j}$ are those induced by the fixed input joint type $P_{X_1 X_2, n}$. By the above constant-composition reduction and a simple symmetry argument, we may replace $(\mathbf{X}_1, \mathbf{X}_2)$ by a fixed pair $(\mathbf{x}_1, \mathbf{x}_2) \in T^n(P_{X_1 X_2, n})$.

5) Handling Singular Dispersion Matrices: Directly applying the multivariate Berry-Esseen theorem (e.g. see [4, Sec. VI]) to (26) is problematic, since the dispersion matrix $\mathbf{V}(P_{X_1 X_2, n})$ may be singular or asymptotically singular. We therefore proceed by handling such matrices, and reducing the problem to a lower dimension as necessary.

We henceforth use the shorthands $\mathbf{I}_n := \mathbf{I}(P_{X_1 X_2, n})$ and $\mathbf{V}_n := \mathbf{V}(P_{X_1 X_2, n})$. An eigenvalue decomposition yields

$$\mathbf{V}_n = \mathbf{U}_n \mathbf{D}_n \mathbf{U}_n^T, \tag{27}$$

where $\mathbf{U}_n$ is unitary (i.e. $\mathbf{U}_n \mathbf{U}_n^T$ is the identity matrix) and $\mathbf{D}_n$ is diagonal. Since we passed to a convergent subsequence in Step 3 and the eigenvalue decomposition map $\mathbf{V}_n \to (\mathbf{U}_n, \mathbf{D}_n)$ is continuous, we conclude that both $\mathbf{U}_n$ and $\mathbf{D}_n$ converge, say to $\mathbf{U}_\infty$ and $\mathbf{D}_\infty$. When $\mathrm{rank}(\mathbf{D}_\infty) = 2$ (i.e. $\mathbf{D}_\infty$ has full rank), we directly use the multivariate Berry-Esseen theorem. We proceed by discussing lower rank matrices.

Since $\mathbf{V}_n$ is the covariance matrix of $\mathbf{A}_n := \frac{1}{\sqrt{n}}\big(\sum_{i=1}^n \mathbf{j}(x_{1,i}, x_{2,i}, Y_i) - n\mathbf{I}_n\big)$ (with $Y_i \sim W(\cdot|x_{1,i}, x_{2,i})$), we see that $\mathbf{D}_n$ is the covariance matrix of $\tilde{\mathbf{A}}_n := \mathbf{U}_n^T \mathbf{A}_n$. In the case that $\mathrm{rank}(\mathbf{D}_\infty) = 1$, we may write

$$\tilde{\mathbf{A}}_n := [\tilde{A}_{n,1} \ \tilde{A}_{n,2}]^T, \tag{28}$$

where $\mathsf{Var}[\tilde{A}_{n,1}]$ is bounded away from zero, and $\mathsf{Var}[\tilde{A}_{n,2}] \to 0$. Since $\mathbf{U}_n$ is unitary, we have

$$\mathbf{A}_n = \mathbf{U}_n \tilde{\mathbf{A}}_n = \mathbf{U}_{n,1} \tilde{A}_{n,1} + \mathbf{\Delta}_n, \tag{29}$$

where $\mathbf{U}_{n,i}$ denotes the i-th column of $\mathbf{U}_n$, and $\mathbf{\Delta}_n := \mathbf{U}_{n,2} \tilde{A}_{n,2}$. Since $\mathbf{A}_n$ has mean zero by construction, the same is true of $\tilde{\mathbf{A}}_n$ and hence $\mathbf{\Delta}_n$. Moreover, since $\tilde{A}_{n,1}$ has vanishing variance, the same is true of each entry of $\mathbf{\Delta}_n$. Thus, Chebyshev's inequality implies that, for any $\delta_n > 0$,

$$\Pr\left(\|\mathbf{\Delta}_n\|_\infty \geq \delta_n\right) \leq \frac{\psi_n}{\delta_n^2}, \tag{30}$$

where $\psi_n := \max_{i=1,2} \mathsf{Var}[\Delta_{n,i}] \to 0$.

We can now bound the probability in (26) as follows:

$$\begin{aligned}
&\Pr\left(\frac{1}{n}\sum_{i=1}^n \mathbf{j}(X_{1,i}, X_{2,i}, Y_i) \geq \mathbf{R}_n - \gamma \mathbf{1}\right) \\
&= \Pr\Big(\mathbf{A}_n \geq \sqrt{n}\big(\mathbf{R}_n - \mathbf{I}_n - \gamma\mathbf{1}\big)\Big) &(31)\\
&= \Pr\Big(\mathbf{U}_{n,1}\tilde{A}_{n,1} + \mathbf{\Delta}_n \geq \sqrt{n}\big(\mathbf{R}_n - \mathbf{I}_n - \gamma\mathbf{1}\big)\Big) &(32)\\
&\leq \Pr\Big(\mathbf{U}_{n,1}\tilde{A}_{n,1} \geq \sqrt{n}\big(\mathbf{R}_n - \mathbf{I}_n - \gamma\mathbf{1}\big) - \delta_n\mathbf{1}\Big) \\
&\qquad + \Pr\left(\|\mathbf{\Delta}_n\|_\infty \geq \delta_n\right) &(33)\\
&\leq \Pr\Big(\mathbf{U}_{n,1}\tilde{A}_{n,1} \geq \sqrt{n}\big(\mathbf{R}_n - \mathbf{I}_n - \gamma\mathbf{1}\big) - \delta_n\mathbf{1}\Big) + \frac{\psi_n}{\delta_n^2}, &(34)
\end{aligned}$$

where the last three steps respectively follow from (29), [4, Lemma 9], and (30). We now choose $\delta_n = \psi_n^{1/3}$, so that both δ_n and $\frac{\psi_n}{\delta_n^2}$ are vanishing. Equation (34) permits an application of the *univariate* Berry-Esseen theorem, since the variance of $\tilde{A}_{n,1}$ is bounded away from zero.

The case $\mathrm{rank}(\mathbf{D}_\infty) = 0$ is handled similarly using Chebyshev's inequality, and we thus omit the details and merely state that (34) is replaced by

$$\mathbb{1}\Big(\sqrt{n}\big(\mathbf{R}_n - \mathbf{I}_n - \gamma\mathbf{1}\big) \leq \delta_n \mathbf{1}\Big) + \delta_n' \tag{35}$$

where $\delta_n \to 0$ and $\delta_n' \to 0$.

6) Application of the Berry-Esseen Theorem: Let $\mathbf{I}_\infty$ and $\mathbf{V}_\infty$ denote the limiting values (on the convergent subsequence of block lengths) of $\mathbf{I}_n$ and $\mathbf{V}_n$. In this step, we will use the fact that $\Psi^{-1}(\cdot, \varepsilon)$ is continuous in the following sense:

$$\Psi^{-1}(\mathbf{V}_n, \varepsilon) - \delta\mathbf{1} \subset \Psi^{-1}(\mathbf{V}_\infty, \varepsilon) \subset \Psi^{-1}(\mathbf{V}_n, \varepsilon) + \delta\mathbf{1} \tag{36}$$

for any $\delta > 0$ and sufficiently large n. This is proved using a Taylor expansion when $\mathbf{V}_\infty$ has full rank, and is proved similarly to [7, Lemma 6] when $\mathbf{V}_\infty$ is singular.

We claim that the preceding two steps, along with the choice $\gamma := \frac{\log n}{n}$, imply that the rate pair $(R_{1,n}, R_{2,n})$ satisfies

$$\mathbf{R}_n \in \mathbf{I}_n + \frac{1}{\sqrt{n}}\Psi^{-1}(\mathbf{V}_\infty, \varepsilon) + g(n)\mathbf{1} \tag{37}$$

for some $g(n) = o\big(\frac{1}{\sqrt{n}}\big)$ depending on $P_{X_1, X_2, n}$ and ε. In the case $\mathrm{rank}(\mathbf{D}_\infty) = 2$ (see the preceding step), this follows by applying the multivariate Berry-Esseen theorem with a positive definite covariance matrix, re-arranging to obtain (37) with $\mathbf{V}_n$ in place of $\mathbf{V}_\infty$, and then using (36).

In the case $\mathrm{rank}(\mathbf{V}_\infty) = 1$, we obtain (37) by applying the univariate Berry-Esseen theorem to (34) and similarly applying rearrangements and (36). The resulting expression can be written in the multivariate form in (37) by a similar argument to [4, p. 894].

When $\mathrm{rank}(\mathbf{V}_\infty) = 0$, we have $\mathbf{V}_\infty = \mathbf{0}$, and $\Psi^{-1}(\mathbf{V}_\infty, \varepsilon)$ is simply the quadrant $\{\mathbf{z} : \mathbf{z} \le \mathbf{0}\}$. We thus obtain (37) by noting that the indicator function in (35) is zero for sufficiently large n whenever either entry of $\mathbf{R}_n$ exceeds the corresponding entry of $\mathbf{I}_n$ by $\Theta\big(\frac{1}{\sqrt{n}}\big)$.

7) Establishing the Convergence to $\Pi(R_1^*, R_2^*)$*:* We use a proof by contradiction to show that the limiting value $P_{X_1X_2,\infty}$ of $P_{X_1X_2,n}$ (on the convergent subsequence of block lengths) must lie within $\Pi(R_1^*, R_2^*)$. Assuming the contrary, we observe from (15) that at least one of the strict inequalities $I_1(P_{X_1X_2,\infty}) < R_1^*$ and $I_{12}(P_{X_1X_2,\infty}) < R_1^* + R_2^*$ must hold. It thus follows from (37) and the continuity of $\mathbf{I}(P_{X_1X_2})$ that there exists $\delta > 0$ such that either $R_{1,n} \le R_1^* - \delta$ or $R_{1,n} + R_{2,n} \le R_1^* + R_2^* - \delta$ for sufficiently large n, in contradiction with the convergence of $(R_{1,n}, R_{2,n})$ to (R_1^*, R_2^*) implied by (3).

8) Completion of the Proof for Cases (i) and (ii): Here we handle distributions $P_{X_1X_2,\infty}$ corresponding to the cases in (19)–(20). We focus on case (ii), since case (i) is handled similarly.

It is easily verified from (14) that each point $\mathbf{z}$ in $\Psi^{-1}(\mathbf{V}, \varepsilon)$ satisfies $z_1 + z_2 \le \sqrt{V_{12}}\Phi^{-1}(\varepsilon)$. We can thus weaken (37) to

$$R_{1,n} + R_{2,n} \le I_{12}(P_{X_1X_2,n}) + \sqrt{\frac{V_{\infty,12}}{n}}\Phi^{-1}(\varepsilon) + g(n). \tag{38}$$

We will complete the proof by showing that $I_{12}(P_{X_1X_2,n}) \le R_1^* + R_2^*$ for all n. Since $\bigcup_{P_{X_1X_2}} \big\{ I_{12}(P_{X_1X_2}) \big\}$ is the set of all achievable (first-order) sum rates, it suffices to show that any boundary point corresponding to (20) is one maximizing the sum rate. We proceed by establishing that this is true.

The conditions stated before (20) state that (R_1^*, R_2^*) lies on the diagonal part of the achievable trapezium corresponding to $P_{X_1X_2}$, and away from the corner point. It follows that $\mathbf{p}_1 := (R_1^* - \delta, R_2^* + \delta)$ and $\mathbf{p}_2 := (R_1^* + \delta, R_2^* - \delta)$ are achievable for sufficiently small δ. If another point $\mathbf{p}_0$ with a strictly higher sum rate were achievable, then all points within the triangle with corners defined by $\mathbf{p}_0$, $\mathbf{p}_1$ and $\mathbf{p}_2$ would also be achievable. This would imply the achievability of $(R_1^* + \delta', R_2^* + \delta')$ for sufficiently small $\delta' > 0$, which contradicts the assumption that (R_1^*, R_2^*) is a boundary point.

9) Completion of the Proof for Case (iii)): We now turn to the remaining case in (21), corresponding to $\mathbf{I}_\infty = \mathbf{R}^*$. Again using the fact that $\mathbf{I}(P_{X_1X_2})$ is continuous in $P_{X_1X_2}$, we have

$$\mathbf{I}_n = \mathbf{R}^* + \mathbf{\Delta}(P_{X_1X_2,n}), \tag{39}$$

where $\|\mathbf{\Delta}(P_{X_1X_2,n})\|_\infty \to 0$. We claim that $\mathbf{\Delta}(P_{X_1X_2,n}) \in \mathcal{V}(R_1^*, R_2^*)$ (see (18)). Indeed, if this were not the case, then (39) would imply that the pair $(I_{n,1}, I_{n,12} - I_{n,1})$ lies outside the capacity region, in contradiction with the fact that rates satisfying (1)–(2) are (first-order) achievable for *all* $P_{X_1X_2}$.

Assuming for the time being that $\|\mathbf{\Delta}(P_{X_1X_2,n})\|_\infty = O\big(\frac{1}{\sqrt{n}}\big)$, we immediately obtain the outer bound

$$\mathcal{L}(\varepsilon, R_1^*, R_2^*) \subseteq \bigg\{ (L_1, L_2) : \mathbf{L} \in \bigcup_{P_{X_1X_2} \in \Pi, \mathbf{T} \in \mathcal{V}} \Big\{ \Psi^{-1}(\mathbf{V}(P_{X_1X_2}), \varepsilon) + \mathbf{T} \Big\} \bigg\}. \tag{40}$$

The set in (40) clearly includes $\mathcal{L}_0$ in (21). We proceed by showing that the reverse inclusion holds, and hence the two sets are identical. Since $\hat{\mathbf{T}}_-$ and $\hat{\mathbf{T}}_+$ are tangent vectors, any vector $\mathbf{T} \in \mathcal{V}$ can have one or more of its components increased to yield a vector $\mathbf{T}'$ whose direction coincides with either $\mathbf{T}_-$ or $\mathbf{T}_+$. The fact that the magnitude of $\mathbf{T}'$ may be arbitrary is captured by the unions over $\beta \ge 0$ in (21).

It remains to handle the case that $\|\mathbf{\Delta}(P_{X_1X_2,n})\|_\infty$ is not $O\big(\frac{1}{\sqrt{n}}\big)$. By performing another pass to a subsequence of block lengths if necessary, we can assume that $\|\mathbf{\Delta}(P_{X_1X_2,n})\|_\infty = \omega\big(\frac{1}{\sqrt{n}}\big)$. Such scalings can be shown to play no role in characterizing $\mathcal{L}$, similarly to [7]; we provide only an outline here. Let $\Delta_{n,1}$ and $\Delta_{n,12}$ denote the entries of $\mathbf{\Delta}(P_{X_1X_2,n})$, and let $\Delta_{n,2} := \Delta_{n,12} - \Delta_{n,1}$. If either $\Delta_{n,1}$ or $\Delta_{n,2}$ is negative and decays with a rate $\omega\big(\frac{1}{\sqrt{n}}\big)$, then no value of the corresponding L_j ($j \in \{1, 2\}$) can satisfy the condition in (3), so the converse is trivial. On the other hand, if either $\Delta_{n,1}$ or $\Delta_{n,2}$ is positive and $\omega\big(\frac{1}{\sqrt{n}}\big)$, we simply recover the right-hand side of (40) in the limiting case that either T_1 or $T_{12} - T_1$ (where $\mathbf{T} := [T_1 \ \ T_{12}]^T$) grows unbounded. Thus, the required converse statement for this case is already captured by (40).

Acknowledgments: The second author is supported by NUS startup grants WBS R-263-000-A98-750/133.

References

[1] V. Strassen. Asymptotische Abschätzungen in Shannons Informationstheorie. In *Trans. Third Prague Conf. Inf. Theory*, pages 689–723, Prague, 1962.

[2] M. Hayashi. Information spectrum approach to second-order coding rate in channel coding. *IEEE Trans. Inf. Th.*, 55(10):4947–66, 2009.

[3] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Th.*, 56(5):2307–59, 2010.

[4] V. Y. F. Tan and O. Kosut. On the dispersions of three network information theory problems. *IEEE Trans. Inf. Th.*, 60(2):881–903, 2014.

[5] R. Nomura and T. S. Han. Second-order Slepian-Wolf coding theorems for non-mixed and mixed sources. *IEEE Trans. Inf. Th.*, 60(9):5553–5572, 2014.

[6] S.-Q. Le, V. Y. F. Tan, and M. Motani. Second-order asymptotics for the Gaussian interference channel with strictly very strong interference. In *Int. Symp. Inf. Th.*, Honolulu, HI, July 2014.

[7] J. Scarlett and V. Y. F. Tan. Second-order asymptotics for the Gaussian MAC with degraded message sets. arXiv:1310.1197 [cs.IT].

[8] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, Cambridge, U.K., 2012.

[9] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[10] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas. Second-order rate region of constant-composition codes for the multiple-access channel. *IEEE Trans. Inf. Th.*, 61(1):157–172, Jan. 2015.