Second-Order Asymptotics for the Gaussian MAC with Degraded Message Sets

Jonathan Scarlett
Department of Engineering,
University of Cambridge
Email: jmscarlett@gmail.com

Vincent Y. F. Tan
Department of Electrical and Computer Engineering,
National University of Singapore
Email: vtan@nus.edu.sg

***Abstract*—This paper studies the second-order asymptotics of the Gaussian 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. We do so by elucidating the relationship between global and local notions of second-order asymptotics.**

I. Introduction

In this paper, we revisit the Gaussian multiple-access channel (MAC) with degraded message sets (DMS). This is a communication model in which two independent messages are to be sent from two sources to a common destination. One encoder, the cognitive or informed encoder (user 1, say), has access to both messages, while the uninformed encoder (user 2) only has access to its own message. The channel is assumed to be memoryless, and each use is described by

$$Y = X_1 + X_2 + Z, \tag{1}$$

where $Z \sim \mathcal{N}(0,1)$. Thus, the channel transition law is

$$W(y|x_1,x_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y-x_1-x_2)^2\right). \tag{2}$$

Let (m_1,m_2) be the message pair, and let $\mathbf{x}_1 = \mathbf{x}_1(m_1,m_2)$ and $\mathbf{x}_2 = \mathbf{x}_2(m_2)$ be the corresponding codewords. We assume the maximal power constraints

$$\|\mathbf{x}_1\|_2^2 \le nS_1, \quad \text{and} \quad \|\mathbf{x}_2\|_2^2 \le nS_2, \tag{3}$$

where S_1 and S_2 are arbitrary positive numbers.

The capacity region, i.e. the set of all pairs of achievable rates, is well-known (e.g. see [1, Ex. 5.18(b)]), and is given by the set of rate pairs (R_1,R_2) satisfying

$$R_1 \le \mathsf{C}\big((1-\rho^2)S_1\big) \tag{4}$$

$$R_1 + R_2 \le \mathsf{C}\big(S_1 + S_2 + 2\rho\sqrt{S_1S_2}\big) \tag{5}$$

for some $\rho \in [0,1]$, where $\mathsf{C}(x) := \frac{1}{2}\log(1+x)$ is the Gaussian capacity function. The capacity region for $S_1 = S_2 = 1$ is illustrated in Fig. 1. The boundary is parametrized by ρ.

While the capacity region is well-known, there is substantial motivation to understand the *second-order asymptotics* for this problem. For any given point (R_1^*, R_2^*) on the boundary of the capacity region, we study the rate of convergence to that point for an ε-reliable code. More precisely, we characterize the set of all (L_1, L_2) pairs that are achievable according to the following definition, similar to that of Nomura-Han [2].

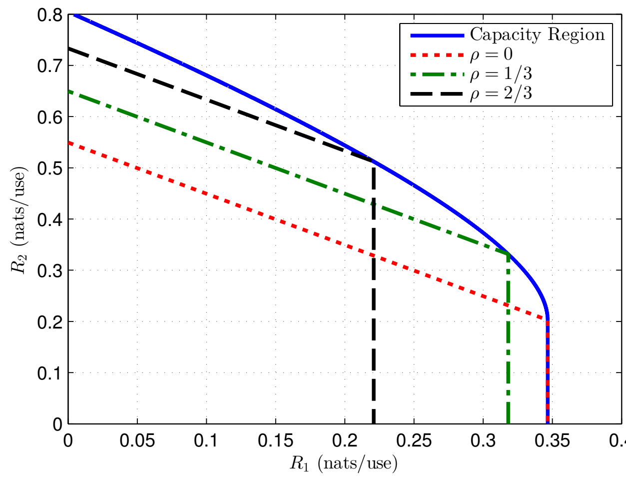

Fig. 1. Capacity region (CR) in nats/use of a Gaussian MAC with DMS where $S_1 = S_2 = 1$, i.e. 0 dB. Observe that $\rho \in [0,1]$ parametrizes points on the boundary. Every point on the curved part of the boundary is achieved by a unique input distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\rho))$.

Definition 1 (Second-Order Coding Rates). *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 user* $j = 1, 2$ *equal to* $M_{j,n}$*, and average error probability* ε_n*, such that*

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}}(\log M_{j,n} - nR_j^*) \ge L_j, \quad j = 1,2, \tag{6}$$

$$\limsup_{n\to\infty} \varepsilon_n \le \varepsilon. \tag{7}$$

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).

Definition 1 gives a *local* notation of second-order achievability. This is in contrast with the *global* asymptotics studied for various network information theory problems in [3]–[6], which we also study here as an initial step towards obtaining the local result. Similarly to Haim *et al.* [7], we believe that the study of local second-order asymptotics provides significantly greater insight into the system performance. To the best of our knowledge, our main result (Theorem 2) provides the first complete characterization of the second-order asymptotics of a multi-user information theory problem in which the boundary of the rate region is curved.

A. Notation

The i-th entry of a vector (e.g. $\mathbf{y}$) is denoted using a subscript (e.g. y_i). Given integers $l \le m$, we use the discrete interval notations $[l:m] := \{l,\dots,m\}$ and $[m] := [1:m]$. 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 $j \in [d]$. 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 $O(\cdot)$, $o(\cdot)$, $\Theta(\cdot)$, $\Omega(\cdot)$ and $\omega(\cdot)$.

II. Main Results

A. Preliminary Definitions

For a pair of rates (R_1,R_2), we define the *rate vector*

$$\mathbf{R} := \begin{bmatrix} R_1 \\ R_1+R_2 \end{bmatrix}. \tag{8}$$

The input distribution to achieve a point on the boundary characterized by some $\rho \in [0,1]$ is a 2-dimensional Gaussian distribution with zero mean and covariance matrix

$$\boldsymbol{\Sigma}(\rho) := \begin{bmatrix} S_1 & \rho\sqrt{S_1S_2} \\ \rho\sqrt{S_1S_2} & S_2 \end{bmatrix}. \tag{9}$$

The corresponding *mutual information vector* is given by

$$\mathbf{I}(\rho) = \begin{bmatrix} I_1(\rho) \\ I_{12}(\rho) \end{bmatrix} := \begin{bmatrix} \mathsf{C}\big(S_1(1-\rho^2)\big) \\ \mathsf{C}\big(S_1+S_2+2\rho\sqrt{S_1S_2}\big) \end{bmatrix}. \tag{10}$$

Let $\mathsf{V}(x,y) := \frac{x(y+2)}{2(x+1)(y+1)}$ be the *Gaussian cross-dispersion function* and let $\mathsf{V}(x) := \mathsf{V}(x,x)$ be the *Gaussian dispersion function* [8], [9]. For fixed $0 \le \rho \le 1$, define the *information-dispersion matrix*

$$\mathbf{V}(\rho) := \begin{bmatrix} V_1(\rho) & V_{1,12}(\rho) \\ V_{1,12}(\rho) & V_{12}(\rho) \end{bmatrix}, \tag{11}$$

where the elements of the matrix are

$$V_1(\rho) := \mathsf{V}\big(S_1(1-\rho^2)\big), \tag{12}$$

$$V_{1,12}(\rho) := \mathsf{V}\big(S_1(1-\rho^2), S_1+S_2+2\rho\sqrt{S_1S_2}\big), \tag{13}$$

$$V_{12}(\rho) := \mathsf{V}\big(S_1+S_2+2\rho\sqrt{S_1S_2}\big). \tag{14}$$

Let $(X_1,X_2) \sim P_{X_1,X_2} = \mathcal{N}(\mathbf{0};\boldsymbol{\Sigma}(\rho))$, and define $Q_{Y|X_2}$ and Q_Y to be Gaussian distributions induced by P_{X_1,X_2} and the channel W, namely

$$Q_{Y|X_2}(y|x_2) := \mathcal{N}\big(y; x_2(1+\rho\sqrt{S_1/S_2}), 1+S_1(1-\rho^2)\big), \tag{15}$$

$$Q_Y(y) := \mathcal{N}\big(y; 0, 1+S_1+S_2+2\rho\sqrt{S_1S_2}\big). \tag{16}$$

It should be noted that the random variables (X_1,X_2) and the densities $Q_{Y|X_2}$ and Q_Y all depend on ρ; this dependence is suppressed throughout the paper. The mutual information vector $\mathbf{I}(\rho)$ and information-dispersion matrix $\mathbf{V}(\rho)$ are the mean vector and conditional covariance matrix of the information density vector

$$\mathbf{j}(X_1,X_2,Y) := \begin{bmatrix} j_1(X_1,X_2,Y) \\ j_{12}(X_1,X_2,Y) \end{bmatrix} = \left[\log\frac{W(Y|X_1,X_2)}{Q_{Y|X_2}(Y|X_2)} \quad \log\frac{W(Y|X_1,X_2)}{Q_Y(Y)}\right]^T. \tag{17}$$

That is, we can write $\mathbf{I}(\rho)$ and $\mathbf{V}(\rho)$ as

$$\mathbf{I}(\rho) = \mathbb{E}\big[\mathbf{j}(X_1,X_2,Y)\big], \tag{18}$$

$$\mathbf{V}(\rho) = \mathbb{E}\big[\,\mathsf{Cov}\,\big(\mathbf{j}(X_1,X_2,Y)\,\big|\,X_1,X_2\big)\big] \tag{19}$$

with $(X_1,X_2,Y) \sim P_{X_1X_2} \times W$. For a given point $(z_1,z_2) \in \mathbb{R}^2$ and a (non-zero) positive semi-definite matrix $\mathbf{V}$, define

$$\Psi(z_1,z_2;\mathbf{V}) := \int_{-\infty}^{z_2}\int_{-\infty}^{z_1} \mathcal{N}(\mathbf{u};\mathbf{0},\mathbf{V})\,\mathrm{d}\mathbf{u}, \tag{20}$$

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

$$\Psi^{-1}(\mathbf{V},\varepsilon) := \big\{(z_1,z_2) \in \mathbb{R}^2 : \Psi(-z_1,-z_2;\mathbf{V}) \ge 1-\varepsilon\big\}. \tag{21}$$

B. Global Second-order Asymptotics

Here we provide inner and outer bounds on $\mathcal{C}(n,\varepsilon)$, defined to be the set of (R_1,R_2) pairs such that there exist codebooks of length n and rates at least R_1 and R_2 yielding an average error probability not exceeding ε. Let $\underline{g}(\rho,\varepsilon,n)$ and $\overline{g}(\rho,\varepsilon,n)$ be arbitrary functions of ρ, ε and n for now, and define

$$\mathcal{R}_{\mathrm{in}}(n,\varepsilon;\rho) := \bigg\{(R_1,R_2) \in \mathbb{R}^2 : \mathbf{R} \in \mathbf{I}(\rho) + \frac{\Psi^{-1}(\mathbf{V}(\rho),\varepsilon)}{\sqrt{n}} + \underline{g}(\rho,\varepsilon,n)\mathbf{1}\bigg\}, \tag{22}$$

$$\mathcal{R}_{\mathrm{out}}(n,\varepsilon;\rho) := \bigg\{(R_1,R_2) \in \mathbb{R}^2 : \mathbf{R} \in \mathbf{I}(\rho) + \frac{\Psi^{-1}(\mathbf{V}(\rho),\varepsilon)}{\sqrt{n}} + \overline{g}(\rho,\varepsilon,n)\mathbf{1}\bigg\}. \tag{23}$$

Theorem 1 (Global Bounds on the (n,ε)-Capacity Region)**.** *There exist functions $\underline{g}(\rho,\varepsilon,n)$ and $\overline{g}(\rho,\varepsilon,n)$ such that*

$$\bigcup_{0\le\rho\le1}\mathcal{R}_{\mathrm{in}}(n,\varepsilon;\rho) \subset \mathcal{C}(n,\varepsilon) \subset \bigcup_{-1\le\rho\le1}\mathcal{R}_{\mathrm{out}}(n,\varepsilon;\rho), \tag{24}$$

and such that $\underline{g}$ and $\overline{g}$ satisfy the following properties:

1) *For any $\varepsilon \in (0,1)$ and $\rho \in (-1,1)$, we have*

$$\underline{g}(\rho,\varepsilon,n) = O\left(\frac{\log n}{n}\right), \quad \overline{g}(\rho,\varepsilon,n) = O\left(\frac{\log n}{n}\right). \tag{25}$$

2) *For any $\varepsilon \in (0,1)$ and any sequence $\{\rho_n\}$ with $\rho_n \to \pm1$, we have*

$$\underline{g}(\rho_n,\varepsilon,n) = o\left(\frac{1}{\sqrt{n}}\right), \quad \overline{g}(\rho_n,\varepsilon,n) = o\left(\frac{1}{\sqrt{n}}\right). \tag{26}$$

Proof: See Section III. ∎

We remark that even though the union for the outer bound is taken over $\rho \in [-1,1]$, only the values $\rho \in [0,1]$ will play a role in establishing the local asymptotics, since negative values of ρ are not even first-order optimal, i.e. they fail to achieve a point on the boundary of the capacity region.

C. Local Second-order Asymptotics

Recall the definition of $\mathcal{L}(\varepsilon; R_1^*, R_2^*)$ given in Definition 1. We further define

$$\mathbf{D}(\rho) = \begin{bmatrix} D_1(\rho) \\ D_{12}(\rho) \end{bmatrix} := \frac{\partial}{\partial\rho} \begin{bmatrix} I_1(\rho) \\ I_{12}(\rho) \end{bmatrix}, \tag{27}$$

where the individual derivatives are given by

$$\frac{\partial I_1(\rho)}{\partial\rho} = \frac{-S_1\rho}{1+S_1(1-\rho^2)}, \tag{28}$$

$$\frac{\partial I_{12}(\rho)}{\partial\rho} = \frac{\sqrt{S_1S_2}}{1+S_1+S_2+2\rho\sqrt{S_1S_2}}. \tag{29}$$

Furthermore, for a vector $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$, we define

$$\mathbf{v}^- := \{(w_1,w_2) \in \mathbb{R}^2 : w_1 \le v_1, w_2 \le v_2\}. \tag{30}$$

Theorem 2 (Optimal Second-Order Coding Rate Region)**.** *Depending on (R_1^*, R_2^*), we have the following three cases:*

(i) *If $R_1^* = I_1(0)$ and $R_1^* + R_2^* \le I_{12}(0)$ (vertical segment of the boundary corresponding to $\rho = 0$), then*

$$\mathcal{L}(\varepsilon; R_1^*, R_2^*) = \Big\{(L_1, L_2) \in \mathbb{R}^2 : L_1 \le \sqrt{V_1(0)}\Phi^{-1}(\varepsilon)\Big\}. \tag{31}$$

(ii) *If $R_1^* = I_1(\rho)$ and $R_1^* + R_2^* = I_{12}(\rho)$ (curved segment of the boundary corresponding to $0 < \rho < 1$), then*

$$\mathcal{L}(\varepsilon; R_1^*, R_2^*) = \bigg\{(L_1, L_2) \in \mathbb{R}^2 : \begin{bmatrix} L_1 \\ L_1 + L_2 \end{bmatrix} \in \bigcup_{\beta\in\mathbb{R}} \Big\{\beta\,\mathbf{D}(\rho) + \Psi^{-1}(\mathbf{V}(\rho),\varepsilon)\Big\}\bigg\}. \tag{32}$$

(iii) *If $R_1^* = 0$ and $R_1^* + R_2^* = I_{12}(1)$ (point on the vertical axis corresponding to $\rho = 1$), then*

$$\mathcal{L}(\varepsilon; R_1^*, R_2^*) = \bigg\{(L_1, L_2) \in \mathbb{R}^2 : \begin{bmatrix} L_1 \\ L_1 + L_2 \end{bmatrix} \in \bigcup_{\beta\le 0} \bigg\{\beta\,\mathbf{D}(1) + \begin{bmatrix} 0 \\ \sqrt{V_{12}(1)}\Phi^{-1}(\varepsilon) \end{bmatrix}^-\bigg\}\bigg\}. \tag{33}$$

Proof: See Section IV. ∎

An example plot of $\mathcal{L}(\varepsilon; R_1^*, R_2^*)$ is shown in Fig. 2, using the parameters $S_1 = S_2 = 1$ and considering the point corresponding to $\rho = \frac{1}{2}$.

We proceed by discussing case (ii) in Theorem 2; a similar discussion applies to case (iii). As in Nomura-Han [2] and Tan-Kosut [3], the second-order asymptotics depend on $\mathbf{V}(\rho)$ and Ψ^{-1}. However, in our setting, the expression containing Ψ^{-1} alone (i.e. the expression obtained by setting $\beta = 0$ in (32)) corresponds to only considering the unique input distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\rho))$ achieving the point $(I_1(\rho), I_{12}(\rho) - I_1(\rho))$. From Fig. 2, this is *not* sufficient to achieve all second-order coding rates. Rather, we must allow the input distribution to vary with n. This is manifested in the $\beta\,\mathbf{D}(\rho)$ term in (32). Roughly speaking, this term arises by considering ρ_n converging to ρ at a rate $O\big(\frac{1}{\sqrt{n}}\big)$, and applying a first-order Taylor expansion.

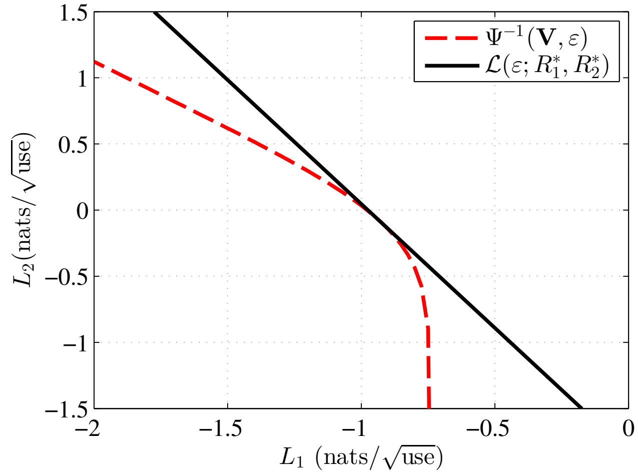

Fig. 2. Second-order coding rates in nats/$\sqrt{\text{use}}$ with $S_1 = S_2 = 1$, $\rho = \frac{1}{2}$ and $\varepsilon = 0.1$. The regions are to the bottom-left of the boundaries shown.

We observe that for $\rho \in (0,1)$, the set in (32) is a half-space which can alternatively be expressed as

$$\mathcal{L}(\varepsilon; R_1^*, R_2^*) = \big\{(L_1, L_2) \in \mathbb{R}^2 : L_2 \le a_\rho L_1 + b_{\rho,\varepsilon}\big\}, \tag{34}$$

where the slope and intercept are given by

$$a_\rho := \frac{D_{12}(\rho) - D_1(\rho)}{D_1(\rho)} \tag{35}$$

$$b_{\rho,\varepsilon} := \inf\big\{b \in \mathbb{R} : \exists L_1 \in \mathbb{R} \text{ s.t. } (L_1, (1+a_\rho)L_1 + b) \in \Psi^{-1}(\mathbf{V}(\rho),\varepsilon)\big\} \tag{36}$$

III. Proof of Theorem 1

Due to space constraints, we provide only an outline of the proof. Full details can be found in [10].

A. Converse

The steps of the converse proof are as follows.

1) A Reduction from Maximal to Equal Power Constraints: Using the argument in [8, Lem. 39], it suffices to consider codes such that the inequalities in (3) hold with equality.

2) A Reduction from Average to Maximal Error Probability: Using similar arguments to [11, Sec 3.4.4], it suffices to prove the converse for maximal (rather than average) error probability.[1] This is shown by starting with an average-error code, and then constructing a maximal-error code as follows: (i) Keep only the fraction $\frac{1}{\sqrt{n}}$ of user 2's messages with the smallest error probability (averaged over user 1's message); (ii) For each of user 2's messages, keep only the fraction $\frac{1}{\sqrt{n}}$ of user 1's messages with the smallest error probability.

3) Correlation Type Classes: Define $\mathcal{I}_0 := \{0\}$ and $\mathcal{I}_k := (\frac{k-1}{n}, \frac{k}{n}]$, $k \in [n]$, and let $\mathcal{I}_{-k} := -\mathcal{I}_k$ for $k \in [n]$. Consider the *correlation type classes* (or simply *type classes*)

$$\mathcal{T}_n(k) := \bigg\{(\mathbf{x}_1, \mathbf{x}_2) : \frac{\langle \mathbf{x}_1, \mathbf{x}_2\rangle}{\|\mathbf{x}_1\|_2\|\mathbf{x}_2\|_2} \in \mathcal{I}_k\bigg\} \tag{37}$$

[1] This argument is not valid for the standard MAC, but is possible here due to the partial cooperation (i.e. user 1 knowing both messages).

where $k \in [-n : n]$, and $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle := \sum_{i=1}^n x_{1i} x_{2i}$ is the standard inner product in $\mathbb{R}^n$. The total number of type classes is $2n+1$, which is polynomial in n analogously to the finite alphabet case [12, Ch. 2]. Using a similar argument to [12, Lem. 16.2], it suffices to consider codes for which all pairs $(\mathbf{x}_1, \mathbf{x}_2)$ are in a single type class, say indexed by k. We define $\hat{\rho} := \frac{k}{n}$ according to that type class.

4) A Verdú-Han-type Converse Bound: For any $\gamma > 0$, we can use standard arguments (e.g. [13, Lem. 2]) to prove the following non-asymptotic converse bound:

$$\varepsilon_n \geq \Pr(\mathcal{A}_n \cup \mathcal{B}_n) - 2\exp(-n\gamma), \tag{38}$$

where

$$\mathcal{A}_n := \left\{ \frac{1}{n} \sum_{i=1}^n j_1(X_{1i}, X_{2i}, Y_i) \leq \frac{1}{n} \log M_{1,n} - \gamma \right\} \tag{39}$$

$$\mathcal{B}_n := \left\{ \frac{1}{n} \sum_{i=1}^n j_{12}(X_{1i}, X_{2i}, Y_i) \leq \frac{1}{n} \log \big(M_{1,n} M_{2,n} \big) - \gamma \right\}, \tag{40}$$

with $Y_i \,|\, \{X_{1i} = x_1, X_{2i} = x_2\} \sim W(\cdot|x_1, x_2)$. The value of ρ used in the information densities j_1 and j_2 is arbitrary, and is chosen to be $\hat{\rho}$. The distribution of $(\mathbf{X}_1, \mathbf{X}_2)$ is that induced by the codebook, and by the preceding steps, its support is restricted to $\{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{T}_n(k) : \|\mathbf{x}_j\|_2^2 = nS_j, j = 1, 2\}$. We henceforth let $(\mathbf{x}_1, \mathbf{x}_2)$ be fixed on this support set.

5) Evaluation of the Verdú-Han Bound for $\hat{\rho} \in (-1, 1)$: Using the definitions of $\mathcal{T}_n(k)$ and the information densities in (17), we show the following in [10]:

$$\left\| \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \mathbf{j}(x_{1i}, x_{2i}, Y_i) \right] - \mathbf{I}(\hat{\rho}) \right\|_\infty \leq \frac{\xi_1}{n} \tag{41}$$

$$\left\| \mathsf{Cov}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{j}(x_{1i}, x_{2i}, Y_i) \right] - \mathbf{V}(\hat{\rho}) \right\|_\infty \leq \frac{\xi_2}{n} \tag{42}$$

for some $\xi_1 > 0$ and $\xi_2 > 0$ not depending on $\hat{\rho}$. Substituting these moments into (38) and using the fact that $\mathbf{V}(\rho)$ is non-singular for $\rho \in (-1, 1)$, we obtain the converse for the first part of the theorem using standard steps based on the multivariate Berry-Esseen theorem [14] (e.g. see [3])

6) Evaluation of the Verdú-Han Bound with $\hat{\rho}_n \to \pm 1$: The case $\hat{\rho}_n := \frac{k}{n} \to \pm 1$ is treated separately because $\mathbf{V}(1)$ is singular due to the fact that $V_1(1) = V_{1,12}(1) = 0$. The overall approach is similar to the previous step, but the univariate Berry-Esseen theorem [15, Sec. XVI.5] is used in place of the multivariate version. We focus our attention on $\hat{\rho}_n \to 1$.

Let $R_{j,n} := \frac{1}{n} \log M_{j,n}$ for $j = 1, 2$. Our aim is to show

$$\begin{bmatrix} R_{1,n} \\ R_{1,n} + R_{2,n} \end{bmatrix} \in \mathbf{I}(\hat{\rho}_n) + \frac{\Psi^{-1}\big(\mathbf{V}(\hat{\rho}_n), \varepsilon\big)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}} \right) \mathbf{1}. \tag{43}$$

From the assumption $\hat{\rho} \to 1$ and the structure of the matrix $\mathbf{V}(1)$, it is not difficult to prove that it suffices to show

$$\begin{bmatrix} R_{1,n} \\ R_{1,n} + R_{2,n} \end{bmatrix} \leq \mathbf{I}(\hat{\rho}_n) + \sqrt{\frac{V_{12}(1)}{n}} \begin{bmatrix} 0 \\ \Phi^{-1}(\varepsilon) \end{bmatrix} + o\left(\frac{1}{\sqrt{n}} \right) \mathbf{1}. \tag{44}$$

We further lower bound (38) by writing $\Pr(\mathcal{A}_n \cup \mathcal{B}_n) \geq \max\big\{ \Pr(\mathcal{A}_n), \Pr(\mathcal{B}_n) \big\}$. Since $V_{12}(1) > 0$, we can handle the $\Pr(\mathcal{B}_n)$ term in the same way as the single-user setting to obtain the second element-wise inequality in (44).

Using an argument based on Taylor expansions, we show in [10] that the mean and variance of $\sum_{i=1}^n j_1(x_{1i}, x_{2i}, Y_i)$ behave as $\Theta(1 - \hat{\rho}_n)$, and similarly for the sum of third absolute moments. Using this observation, we treat two cases separately: $1 - \hat{\rho}_n = \omega\big(\frac{1}{n}\big)$ and $1 - \hat{\rho}_n = O\big(\frac{1}{n}\big)$. In the former case, the first element-wise inequality in (44) is obtained by again using the Berry-Esseen theorem. In the latter case, the same result is obtained using Chebyshev's inequality.

B. Achievability

1) Random-Coding Ensemble: We use superposition coding [1, Ch. 5], with random codewords of the form

$$\left\{ \Big(\mathbf{X}_2(m_2), \{\mathbf{X}_1(m_1, m_2)\}_{m_1=1}^{M_{1,n}} \Big) \right\}_{m_2=1}^{M_{2,n}}$$
$$\sim \prod_{m_2=1}^{M_{2,n}} \left(P_{\mathbf{X}_2}(\mathbf{x}_2(m_2)) \prod_{m_1=1}^{M_{1,n}} P_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1(m_1, m_2)|\mathbf{x}_2(m_2)) \right). \tag{45}$$

The codeword distributions $P_{\mathbf{X}_2}$ and $P_{\mathbf{X}_1|\mathbf{X}_2}$ are chosen as

$$P_{\mathbf{X}_2}(\mathbf{x}_2) = \frac{1}{\mu_{2,n}} \prod_{i=1}^n P_{X_2}(x_{2i}) \mathbf{1}\big\{ \mathbf{x}_2 \in \mathcal{D}_{2,n} \big\}, \tag{46}$$

$$P_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2) = \frac{1}{\mu_{1,n}(\mathbf{x}_2)} \times \prod_{i=1}^n P_{X_1|X_2}(x_{1i}|x_{2i}) \mathbf{1}\big\{ (\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{D}_n \big\}, \tag{47}$$

where $P_{X_1, X_2} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\rho))$ (see (9)), $\mu_{2,n}$ and $\mu_{1,n}(\mathbf{x}_2)$ are normalizing constants, and

$$\mathcal{D}_n := \left\{ \big(\mathbf{x}_1, \mathbf{x}_2 \big) : \|\mathbf{x}_1\|_2^2 \leq nS_1, \|\mathbf{x}_2\|_2^2 \leq nS_2, \left| \frac{1}{n} \sum_{i=1}^n a_k(x_{1i}, x_{2i}) - \mathbb{E}[a_k(X_1, X_2)] \right| \leq \frac{\delta}{n}, \ k \in [K] \right\} \tag{48}$$

$$\mathcal{D}_{2,n} := \left\{ \mathbf{x}_2 \,:\, \Pr\big((\mathbf{X}_1', \mathbf{x}_2) \in \mathcal{D}_n \,\big|\, \mathbf{X}_2' = \mathbf{x}_2 \big) \geq \frac{\psi(n)}{2} \right\}. \tag{49}$$

Here $\{a_k(x_1, x_2)\}_{k=1}^K$ are *auxiliary cost functions* (assumed to be arbitrary for now), δ and $\psi(n)$ are constants, and we define $(\mathbf{X}_1', \mathbf{X}_2') \sim \prod_{i=1}^n P_{X_1, X_2}(x_{1i}', x_{2i}')$. This is an extension of the ensemble studied in [16] to superposition coding.

Building on the analysis of [16], we show in [10] that, as long as $\mathbb{E}[a_k(X_1, X_2)^2] < \infty$ for each k, $\psi(n)$ and δ can be chosen such that $\mu_{2,n} = \Omega(n^{-(K+2)/2})$, and $\mu_{1,n}(\mathbf{x}_2) = \Omega(n^{-(K+2)/2})$ for all $\mathbf{x}_2$ on the support of $P_{\mathbf{X}_2}$.

2) A Feinstein-type Achievability Bound: Using the above-mentioned properties of $\mu_{1,n}$ and $\mu_{2,n}$, we can use standard threshold-based arguments (e.g. [5, Thm. 4]) to obtain

$$\varepsilon_n \leq \Pr(\mathcal{F}_n \cup \mathcal{G}_n) + O\left(\frac{1}{\sqrt{n}} \right), \tag{50}$$

where

$$\mathcal{F}_n := \left\{ \frac{1}{n} \sum_{i=1}^{n} j_1(X_{1i}, X_{2i}, Y_i) \le \frac{1}{n} \log M_{1,n} + \gamma \right\} \tag{51}$$

$$\mathcal{G}_n := \left\{ \frac{1}{n} \sum_{i=1}^{n} j_{12}(X_{1i}, X_{2i}, Y_i) \le \frac{1}{n} \log \big(M_{1,n} M_{2,n}\big) + \gamma \right\} \tag{52}$$

and $\gamma = O\big(\frac{\log n}{n}\big)$.

3) Evaluation of the Bound: The remainder of the proof follows similar steps to the converse upon suitable choices of the auxiliary costs. In particular, let us define

$$\mathbf{j}(x_1, x_2) := \mathbb{E}[\mathbf{j}(x_1, x_2, Y)] \tag{53}$$

$$\mathbf{v}(x_1, x_2) := \mathsf{Cov}[\mathbf{j}(x_1, x_2, Y)], \tag{54}$$

where $(Y|x_1, x_2) \sim W(\cdot|x_1, x_2)$. For both $\rho \in [0, 1)$ and $\rho \to 1$, we let the first 5 auxiliary costs equal the entries of the vector and matrix in (53)–(54) (2 for $\mathbf{j}(x_1, x_2)$ and 3 for the symmetric matrix $\mathbf{v}(x_1, x_2)$). It follows from (48) that

$$\left\| \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \mathbf{j}(x_1, x_2, Y_i) \right] - \mathbf{I}(\rho) \right\|_\infty \le \frac{\delta}{n}, \tag{55}$$

$$\left\| \mathsf{Cov}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{j}(x_1, x_2, Y_i) \right] - \mathbf{V}(\rho) \right\|_\infty \le \frac{\delta}{n} \tag{56}$$

for all $(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{D}_n$, where the expectations and covariance are taken with respect to $W^n(\cdot|\mathbf{x}_1, \mathbf{x}_2)$. The third moment(s) can be bounded similarly by a suitable choice of the remaining auxiliary cost(s) ($K = 6$ for the case $\rho \in [0, 1)$, and $K = 7$ for the case $\rho \to 1$). This allows us to apply the Berry-Esseen theorem and Chebyshev's inequality as in the converse proof.

IV. Proof of Theorem 2

A. Converse

The proof for case (i) ($\rho = 0$) is similar to the single-user case [8], and is thus omitted. We consider cases (ii) and (iii).

1) Establishing The Convergence of ρ_n to ρ: Fix a correlation coefficient $\rho \in (0, 1]$, and consider any sequence of codes satisfying (6)–(7) for some (R_1^*, R_2^*) on the boundary parametrized by ρ, i.e. $R_1^* = I_1(\rho)$ and $R_1^* + R_2^* = I_{12}(\rho)$. Letting $\mathbf{R}_n := [R_{1,n}, R_{1,n}+R_{2,n}]^T$, it follows from the global converse result that

$$\mathbf{R}_n \in \mathbf{I}(\rho_n) + \frac{\Psi^{-1}(\mathbf{V}(\rho_n), \varepsilon)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\mathbf{1} \tag{57}$$

for some (possibly non-unique) sequence $\{\rho_n\}_{n\ge1}$. We claim that every such sequence $\{\rho_n\}_{n\ge1}$ converges to ρ. Indeed, since the boundary of the capacity region is curved and uniquely parametrized by ρ for $\rho \in (0, 1]$, $\rho_n \not\to \rho$ implies for some $\delta > 0$ and for all sufficiently large n that either $I_1(\rho_n) \le I_1(\rho) - \delta$ or $I_{12}(\rho_n) \le I_{12}(\rho) - \delta$. Combining this with (57), we deduce that $R_{1,n} \le I_1(\rho) - \frac{\delta}{2}$ or $R_{1,n} + R_{2,n} \le I_{12}(\rho) - \frac{\delta}{2}$ for sufficiently large n. This, in turn, contradicts the convergence of $(R_{1,n}, R_{2,n})$ to (R_1^*, R_2^*) implied by (6).

2) Establishing The Convergence Rate of ρ_n to ρ: Since $\mathbf{I}(\rho)$ is twice continuously differentiable, we have

$$\mathbf{I}(\rho_n) = \mathbf{I}(\rho) + \mathbf{D}(\rho)(\rho_n - \rho) + O\big((\rho_n - \rho)^2\big)\mathbf{1}. \tag{58}$$

In the case that $\rho_n - \rho = \omega\big(\frac{1}{\sqrt{n}}\big)$, (57) and (58) imply

$$\mathbf{R}_n \le \mathbf{I}(\rho) + \mathbf{D}(\rho)(\rho_n - \rho) + o(\rho_n - \rho)\mathbf{1}. \tag{59}$$

Since the first entry of $\mathbf{D}(\rho)$ is negative and the second entry is positive, (59) states that $L_1 = +\infty$ (i.e. a large addition to R_1^*) only if $L_1 + L_2 = -\infty$ (i.e. a large backoff from $R_1^* + R_2^*$), and $L_1 + L_2 = +\infty$ only if $L_1 = -\infty$. Thus, this case does not play a role in the characterization of $\mathcal{L}(\varepsilon; R_1^*, R_2^*)$, and we may proceed by focusing on the case that $\rho_n - \rho = O\big(\frac{1}{\sqrt{n}}\big)$.

3) Completion of the Proof: First consider case (ii), i.e. $\rho \in (0, 1)$. Assuming now that $\rho_n - \rho = O\big(\frac{1}{\sqrt{n}}\big)$, we can use the Bolzano-Weierstrass theorem [17, Thm. 3.6(b)] to conclude that there exists a subsequence $\{n_k\}_{k\ge1}$ such that $\sqrt{n_k}(\rho_{n_k} - \rho) \to \beta$ for some $\beta \in \mathbb{R}$. Combining this observation with (57) and (58) yields (32). Case (iii) is handled similarly, except that $\beta \le 0$ (since ρ_n may only approach one from below), and the set Ψ^{-1} takes an alternative form similarly to (44).

B. Achievability

The achievability part is similar to the converse part, yet simpler. Specifically, we can simply choose $\rho_n := \rho + \frac{\beta}{\sqrt{n}}$, and apply the above arguments based on Taylor expansions.

References

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

[2] R. Nomura and T. S. Han. Second-order Slepian-Wolf coding theorems for non-mixed and mixed sources. In *Int. Symp. Inf. Th.*, Istanbul, Turkey, 2013. arXiv:1207.2505 [cs.IT].

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

[4] Y.-W. Huang and P. Moulin. Finite blocklength coding for multiple access channels. In *Int. Symp. Inf. Th.*, Boston, MA, 2012.

[5] E. MolavianJazi and J. N. Laneman. A finite-blocklength perspective on Gaussian multi-access channels. arXiv:1309.2343 [cs.IT], Sep 2013.

[6] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas. Second-order rate region of constant-composition codes for the multiple-access channel. In *Allerton Conference*, 2013. arXiv:1303.6167 [cs.IT].

[7] E. Haim, Y. Kochman, and U. Erez. A note on the dispersion of network problems. In *Convention of Electrical and Electronics Engineers in Israel (IEEEI)*, 2012.

[8] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding in the finite blocklength regime. *IEEE Trans. on Inf. Th.*, 56:2307–59, May 2010.

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

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

[11] Y. Polyanskiy. *Channel coding: Non-asymptotic fundamental limits*. PhD thesis, Princeton University, 2010.

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

[13] S. Boucheron and M. R. Salamatian. About priority encoding transmission. *IEEE Trans. on Inf. Th.*, 46(2):699–705, 2000.

[14] F. Götze. On the rate of convergence in the multivariate CLT. *The Annals of Probability*, 19(2):721–739, 1991.

[15] W. Feller. *An Introduction to Probability Theory and Its Applications*. John Wiley and Sons, 2nd edition, 1971.

[16] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas. Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations. *IEEE Trans. on Inf. Th.* (to appear), 2014.

[17] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.