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

Jonathan Scarlett and Vincent Y. F. Tan

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 boundary point for codes that have asymptotic error probability no larger than ε. We do so by elucidating clear relationships between the rate pairs achievable at large blocklengths and the local second-order asymptotics, i.e. the second-order behavior of these rate pairs in the neighborhood of a boundary point. We provide a numerical example to illustrate how the angle of approach to a boundary point affects the second-order coding rate. This is the first conclusive characterization of the second-order asymptotics of a network information theory problem in which the capacity region is not a polygon.

Index Terms

Gaussian multiple-access channel, Degraded message sets, Superposition coding, Strong converse, Finite blocklengths, Second-order coding rates, Dispersion.

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, has access to both messages, while the uninformed encoder only has access to its own message. Both transmitted signals are assumed to be power limited, and their sum is corrupted by additive white Gaussian noise (AWGN). See Fig. 1.

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

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

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

for some $\rho \in [0,1]$, where S_1 and S_2 are the admissible transmit powers, and $\mathsf{C}(x) := \frac{1}{2}\log(1+x)$ is the Gaussian capacity function. The capacity region for $S_1 = S_2 = 1$ is illustrated in Fig. 2. The boundary is parametrized by ρ, and the direct part is proved using superposition coding [2].

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, known as second-order coding rates [3]–[6], for which there exist sequences of codes whose asymptotic error probability does not exceed ε, and whose code sizes $M_{1,n}$ and $M_{2,n}$ behave as

$$\log M_{j,n} \ge nR_j^* + \sqrt{n}L_j + o\big(\sqrt{n}\big), \quad j = 1,2. \tag{3}$$

This study allows us to understand the fundamental tradeoffs between the rates of transmission and average error probability from a perspective different from the study of error exponents. Here, instead of fixing a pair of rates and studying the exponential decay of the error probability ε, we fix ε and study the speed at which a sequence of rate pairs approaches an information-theoretic limit as the blocklength grows.

J. Scarlett is with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (email: jmscarlett@gmail.com). V. Y. F. Tan is with the Institute for Infocomm Research (I^{2}R), A*STAR, Singapore and with the Department of Electrical and Computer Engineering, National University of Singapore. (email: vtan@nus.edu.sg).

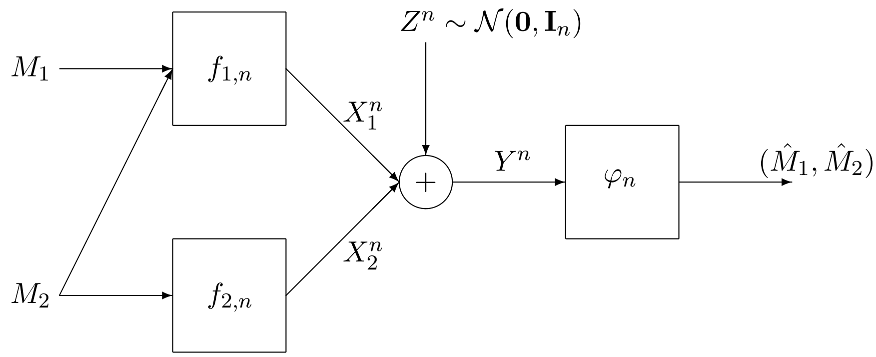

Fig. 1. The model for the Gaussian MAC with degraded message sets (DMS).

A. Related Work

The study of the second-order asymptotics of channel coding for discrete memoryless channels was initiated by Strassen [7]. For the single-user AWGN channel with (equal or maximal) power constraint S, a specialization of our model with $M_{2,n} = 1$, Hayashi [4] and Polyanskiy *et al.* [8] showed that the optimum (highest) second-order coding rate is $\sqrt{\mathsf{V}(S)}\Phi^{-1}(\varepsilon)$, where $\mathsf{V}(x) := \frac{x(x+2)}{2(x+1)^2}$ is the *Gaussian dispersion function*. More precisely, Polyanskiy *et al.* [8, Thm. 54] showed the asymptotic expansion

$$\log M^*(n, \varepsilon) = n\mathsf{C}(S) + \sqrt{n\mathsf{V}(S)}\Phi^{-1}(\varepsilon) + g_n, \tag{4}$$

where $M^*(n,\varepsilon)$ is the maximum size of a length-n block code with (either average or maximum) error probability ε, and $O(1) \le g_n \le \frac{1}{2}\log n + O(1)$. In fact, the expression for $\mathsf{V}(S)$ was already known to Shannon [9, Sec. X], who analyzed the reliability function of the AWGN channel for rates close to capacity.

There have been numerous attempts to study the finite blocklength behavior and second-order asymptotics for MACs [10]–[18], but most of these works focus on inner bounds (the direct part). The development of tight and easily-evaluated converse bounds remains more modest, and those available do not match the direct part in general or are very restrictive (e.g. product channels were considered in [18]). We will see that the assumption of Gaussianity of the channel model together with the degradedness of the message sets allows us to circumvent some of the difficulties in proving second-order converses for the MAC, thus allowing us to obtain a *conclusive* second-order result for the Gaussian MAC with DMS.

We focus primarily on *local* second-order asymptotics propounded by Haim *et al.* [18] for general network information theory problems, where a boundary point is fixed and the rate of approach is characterized. This is different from the global asymptotics studied in [10]–[17], which we also study here as an initial step towards obtaining the local result. Similarly to Haim *et al.* [18], we believe that the study of local second-order asymptotics provides significantly greater insight into the system performance.

B. Main Contributions

Our main contribution is the characterization of the set of admissible second-order coding rates (L_1, L_2) in the curved part of the boundary of the capacity region. For a point on the boundary characterized by $\rho \in (0,1)$, we show that the set of achievable second-order rate pairs (L_1, L_2) is given by those satisfying

$$\begin{bmatrix} L_1 \\ L_1 + L_2 \end{bmatrix} \in \bigcup_{\beta \in \mathbb{R}} \big\{\beta\, \mathbf{D}(\rho) + \Psi^{-1}(\mathbf{V}(\rho), \varepsilon)\big\}, \tag{5}$$

where the entries of $\mathbf{D}(\rho)$ are the derivatives of the capacities in (1)–(2), $\mathbf{V}(\rho)$ is the *dispersion matrix* [10], [11], and Ψ^{-1} is the 2-dimensional generalization of the inverse of the cumulative distribution of a Gaussian. (All quantities are defined precisely in the sequel.) Thus, the contribution from the Gaussian approximation $\Psi^{-1}(\mathbf{V}(\rho), \varepsilon)$ is insufficient for characterizing the second-order asymptotics of multi-terminal channel coding problems in general, and we need to take into account contributions from the first-order term in terms of its slope $\mathbf{D}(\rho)$. This is in stark

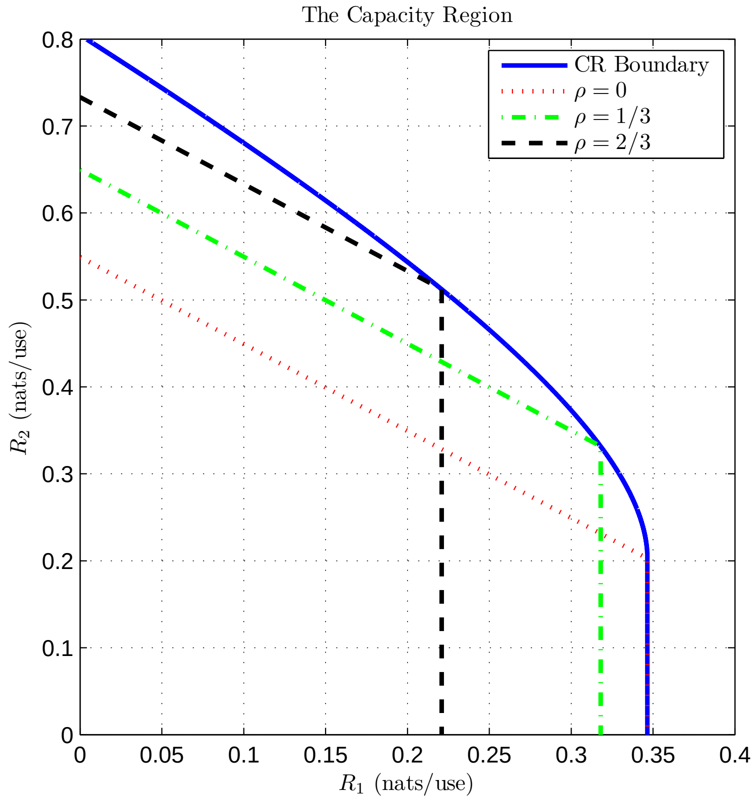

Fig. 2. Capacity region (CR) in nats/use of a Gaussian MAC with DMS where $S_1 = S_2 = 1$, i.e. $0\,\mathrm{dB}$. Observe that $\rho \in [0, 1]$ parametrizes points on the boundary. Every point on the curved part of the boundary is achieved by a unique input distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\rho))$.

contrast to single-user problems (e.g. [3], [4], [6]–[8]) and the (two-encoder) Slepian-Wolf problem [5], [10] where the Gaussian approximation in terms of a dispersion quantity is sufficient for second-order asymptotics.

We first derive global second-order results [10], [18] and then use them to obtain local second-order results. As in [18], we make a strong connection between these two perspectives. To the best of our knowledge, our main result (Theorem 3) provides the first complete characterization of the second-order asymptotics of a multi-user information theory problem in which the boundary of the capacity region (or optimal rate region for source coding problems) is curved.

II. Problem Setting and Definitions

In this section, we state the channel model, various definitions and some known results.

Notation: Given integers $l \leq m$, we use the discrete interval [1] notations $[l : m] := \{l, \ldots, m\}$ and $[m] := [1 : m]$. All log's and exp's are with respect to the natural base e. The ℓ_p-norm of the vectorized version of matrix $\mathbf{A}$ is denoted by $\|\mathbf{A}\|_p := \big(\sum_{i,j} |a_{i,j}|^p\big)^{1/p}$. For two vectors of the same length $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$, the notation $\mathbf{a} \leq \mathbf{b}$ means that $a_j \leq b_j$ for all $j \in [d]$. The notation $\mathcal{N}(\mathbf{u}; \boldsymbol{\mu}, \mathbf{\Lambda})$ denotes the multivariate Gaussian probability density function (pdf) with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Lambda}$. The argument $\mathbf{u}$ will often remain unspecified. We use standard asymptotic notations: $f_n \in O(g_n)$ if and only if (iff) $\limsup_{n\to\infty} \big|f_n/g_n\big| < \infty$; $f_n \in \Omega(g_n)$ iff $g_n \in O(f_n)$; $f_n \in \Theta(g_n)$ iff $f_n \in O(g_n) \cap \Omega(g_n)$; $f_n \in o(g_n)$ iff $\limsup_{n\to\infty} \big|f_n/g_n\big| = 0$; and $f_n \in \omega(g_n)$ iff $\liminf_{n\to\infty} \big|f_n/g_n\big| = \infty$.

A. Channel Model

The signal model for the Gaussian MAC is given by

$$Y = X_1 + X_2 + Z, \tag{6}$$

where X_1 and X_2 represent the inputs to the channel, $Z \sim \mathcal{N}(0,1)$ is additive Gaussian noise with zero mean and unit variance, and Y is the output of the channel. Thus, the channel from (X_1, X_2) to Y can be written as

$$W(y|x_1,x_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y-x_1-x_2)^2\right). \tag{7}$$

The channel is used n times in a memoryless manner without feedback. The channel inputs (i.e., the transmitted codewords) $\mathbf{x}_1 = (x_{11}, \ldots, x_{1n})$ and $\mathbf{x}_2 = (x_{21}, \ldots, x_{2n})$ are required to satisfy the maximal power constraints

$$\|\mathbf{x}_1\|_2^2 \le nS_1, \quad \text{and} \quad \|\mathbf{x}_2\|_2^2 \le nS_2, \tag{8}$$

where S_1 and S_2 are arbitrary positive numbers. We do not incorporate multiplicative gains g_1 and g_2 to X_1 and X_2 in the channel model in (6); this is without loss of generality, since in the presence of these gains we may equivalently redefine (8) with $S_j' := S_j/g_j^2$ for $j = 1,2$.

B. Definitions

Definition 1 (Code)**.** *An* $(n, M_{1,n}, M_{2,n}, S_1, S_2, \varepsilon_n)$-code *for the Gaussian MAC with DMS consists of two encoders* $f_{1,n}, f_{2,n}$ *and a decoder* φ_n *of the form* $f_{1,n} : [M_{1,n}] \times [M_{2,n}] \to \mathbb{R}^n$, $f_{2,n} : [M_{2,n}] \to \mathbb{R}^n$ *and* $\varphi_n : \mathbb{R}^n \to [M_{1,n}] \times [M_{2,n}]$ *which satisfy*

$$\|f_{1,n}(m_1,m_2)\|_2^2 \le nS_1 \quad \forall\,(m_1,m_2) \in [M_{1,n}] \times [M_{2,n}], \tag{9}$$

$$\|f_{2,n}(m_2)\|_2^2 \le nS_2 \quad \forall\,m_2 \in [M_{2,n}], \tag{10}$$

$$\Pr\left((\mathsf{M}_1,\mathsf{M}_2) \ne (\hat{\mathsf{M}}_1,\hat{\mathsf{M}}_2)\right) \le \varepsilon_n, \tag{11}$$

where the messages M_1 *and* M_2 *are uniformly distributed on* $[M_{1,n}]$ *and* $[M_{2,n}]$ *respectively, and* $(\hat{\mathsf{M}}_1,\hat{\mathsf{M}}_2) := \varphi_n(Y^n)$ *are the decoded messages.*

Since S_1 and S_2 are fixed positive numbers, we suppress the dependence of the subsequent definitions, results and parameters on these constants. We will often make reference to (n,ε)-*codes*; this is the family of $(n, M_{1,n}, M_{2,n}, S_1, S_2, \varepsilon)$-codes where the sizes $M_{1,n}, M_{2,n}$ are left unspecified.

Definition 2 ((n,ε)-Achievability)**.** *A pair of non-negative numbers* (R_1,R_2) *is* (n,ε)-achievable *if there exists an* $(n, M_{1,n}, M_{2,n}, S_1, S_2, \varepsilon_n)$*-code such that*

$$\frac{1}{n}\log M_{j,n} \ge R_j, \quad j=1,2, \quad \textit{and} \quad \varepsilon_n \le \varepsilon. \tag{12}$$

The (n,ε)-capacity region $\mathcal{C}(n,\varepsilon) \subset \mathbb{R}_+^2$ *is defined to be the set of all* (n,ε)*-achievable rate pairs* (R_1,R_2)*.*

Definition 2 is a non-asymptotic one that is used primarily for the global second-order results. We now introduce definitions that involve the existence of *sequences* of codes, implying that these are asymptotic-type definitions.

Definition 3 (First-Order Coding Rates)**.** *A pair of non-negative numbers* (R_1,R_2) *is* ε-achievable *if there exists a sequence of* $(n, M_{1,n}, M_{2,n}, S_1, S_2, \varepsilon_n)$*-codes such that*

$$\liminf_{n\to\infty} \frac{1}{n}\log M_{j,n} \ge R_j, \quad j=1,2, \quad \textit{and} \quad \limsup_{n\to\infty} \varepsilon_n \le \varepsilon. \tag{13}$$

The ε-capacity region $\mathcal{C}(\varepsilon) \subset \mathbb{R}_+^2$ *is defined to be the closure of the set of all* ε*-achievable rate pairs* (R_1,R_2)*. The* capacity-region $\mathcal{C}$ *is defined as*

$$\mathcal{C} := \bigcap_{\varepsilon>0} \mathcal{C}(\varepsilon) = \lim_{\varepsilon\to 0} \mathcal{C}(\varepsilon), \tag{14}$$

where the limit exists because of the monotonicity of $\mathcal{C}(\varepsilon)$*.*

Next, we state the most important definitions concerning local second-order coding rates in the spirit of Nomura-Han [5] and Tan-Kosut [10]. We will spend the majority of the paper developing tools to characterize these rates. Here (R_1^*, R_2^*) is a pair of rates on the boundary of $\mathcal{C}(\varepsilon)$.

Definition 4 (Second-Order Coding Rates)**.** *A pair of numbers* (L_1, L_2) *is* $(\varepsilon, R_1^*, R_2^*)$-achievable *if there exists a sequence of* $(n, M_{1,n}, M_{2,n}, S_1, S_2, \varepsilon_n)$*-codes such that*

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}}(\log M_{j,n} - nR_j^*) \geq L_j, \quad j = 1, 2, \quad \textit{and} \quad \limsup_{n\to\infty} \varepsilon_n \leq \varepsilon. \tag{15}$$

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)*.*

Stated differently, if (L_1, L_2) is $(\varepsilon, R_1^*, R_2^*)$-achievable, then there are codes whose error probabilities are asymptotically no larger than ε, and whose sizes $(M_{1,n}, M_{2,n})$ satisfy the asymptotic relation in (3). Even though we refer to L_1 and L_2 as "rates", they may be negative [3]–[6]. A negative value corresponds to a backoff from the first-order term, whereas a positive value corresponds to an addition to the first-order term.

C. Existing First-Order Results

To put things in context, let us review some existing results concerning the ε-capacity region. To state the result compactly, we define the *mutual information vector* as

$$\mathbf{I}(\rho) = \begin{bmatrix} I_1(\rho) \\ I_{12}(\rho) \end{bmatrix} := \begin{bmatrix} \mathsf{C}\big(S_1(1-\rho^2)\big) \\ \mathsf{C}\big(S_1 + S_2 + 2\rho\sqrt{S_1 S_2}\big) \end{bmatrix}. \tag{16}$$

For a pair of rates (R_1, R_2), let the *rate vector* be

$$\mathbf{R} := \begin{bmatrix} R_1 \\ R_1 + R_2 \end{bmatrix}. \tag{17}$$

A statement of the following result is provided in [1, Ex. 5.18(b)]. A weak converse was proved for the more general Gaussian MAC with common message in [19].

Proposition 1 (Capacity Region)**.** *The capacity region of the Gaussian MAC with DMS is given by*

$$\mathcal{C} = \bigcup_{0\leq\rho\leq 1} \left\{(R_1, R_2) \in \mathbb{R}_+^2 : \mathbf{R} \leq \mathbf{I}(\rho)\right\}. \tag{18}$$

The union on the right is a subset of $\mathcal{C}(\varepsilon)$ for every $\varepsilon \in (0, 1)$. However, only the weak converse is implied by (18). To the best of the authors' knowledge, the strong converse for the Gaussian MAC with DMS has not been demonstrated. A by-product of the derivation of the second-order asymptotics in this paper is the strong converse, allowing us to assert that for all $\varepsilon \in (0, 1)$,

$$\mathcal{C} = \mathcal{C}(\varepsilon). \tag{19}$$

The direct part of Proposition 1 can be proved using superposition coding [2]. Treat X_2 as the cloud center and X_1 as the satellite codeword. The input distribution to achieve a point on the boundary characterized by some $\rho \in [0, 1]$ is a 2-dimensional Gaussian with zero mean and covariance matrix

$$\mathbf{\Sigma}(\rho) := \begin{bmatrix} S_1 & \rho\sqrt{S_1 S_2} \\ \rho\sqrt{S_1 S_2} & S_2 \end{bmatrix}. \tag{20}$$

Thus, the parameter ρ represents the correlation between the two users' codewords.

III. Global Second-Order Results

In this section, we present inner and outer bounds to $\mathcal{C}(n, \varepsilon)$. We begin with some definitions. 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* [4], [8], [9] for a single-user additive white Gaussian noise channel with signal-to-noise ratio x. For fixed $0 \leq \rho \leq 1$, define the *information-dispersion matrix*

$$\mathbf{V}(\rho) := \begin{bmatrix} V_1(\rho) & V_{1,12}(\rho) \\ V_{1,12}(\rho) & V_{12}(\rho) \end{bmatrix}, \tag{21}$$

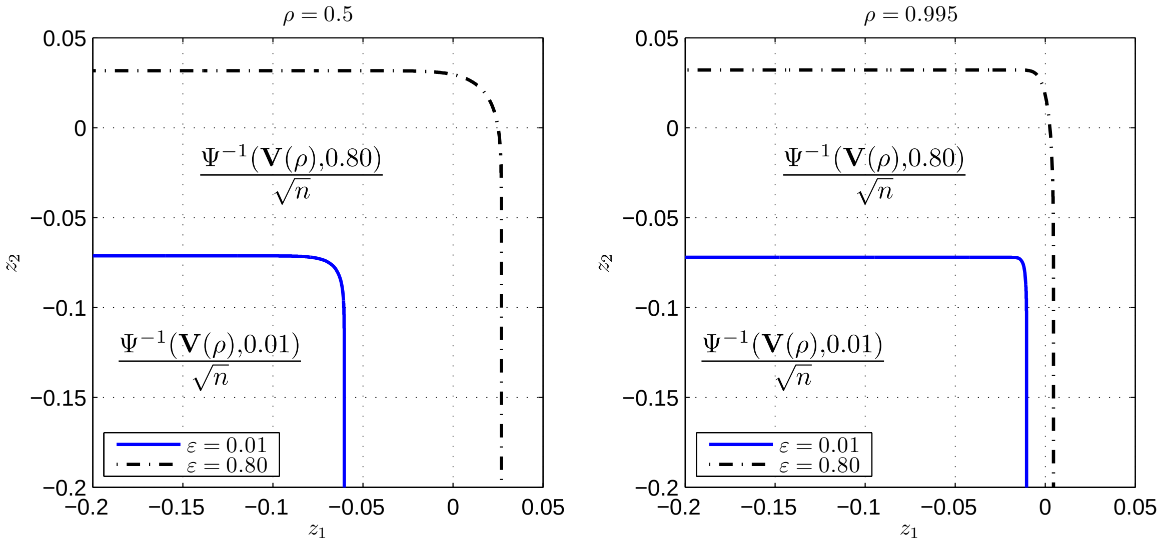

Fig. 3. Illustration of the set $\Psi^{-1}(\mathbf{V}(\rho),\varepsilon)/\sqrt{n}$ with $n = 500$, $S_1 = S_2 = 1$, $\rho = 0.5$ (small condition number) and $\rho = 0.995$ (large condition number). The information dispersion matrix $\mathbf{V}(\rho)$ is defined in (21). In the plots, ε takes two values, 0.01 and 0.80. For the former, the set is completely contained in the third quadrant and this corresponds to approaching the boundary from *inside* the capacity region.

where the elements of the matrix are

$$V_1(\rho) := \mathsf{V}\big(S_1(1-\rho^2)\big), \tag{22}$$

$$V_{1,12}(\rho) := \mathsf{V}\big(S_1(1-\rho^2), S_1 + S_2 + 2\rho\sqrt{S_1 S_2}\big), \tag{23}$$

$$V_{12}(\rho) := \mathsf{V}\big(S_1 + S_2 + 2\rho\sqrt{S_1 S_2}\big). \tag{24}$$

Let $(X_1, X_2) \sim P_{X_1,X_2} = \mathcal{N}(\mathbf{0}; \mathbf{\Sigma}(\rho))$, 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

$$Q_{Y|X_2}(y|x_2) := \mathcal{N}\big(y; x_2(1 + \rho\sqrt{S_1/S_2}), 1 + S_1(1-\rho^2)\big), \tag{25}$$

$$Q_Y(y) := \mathcal{N}\big(y; 0, 1 + S_1 + S_2 + 2\rho\sqrt{S_1 S_2}\big). \tag{26}$$

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

$$\mathbf{j}(X_1, X_2, Y) := \begin{bmatrix} j_1(X_1, X_2, Y) \\ j_{12}(X_1, X_2, Y) \end{bmatrix} = \begin{bmatrix} \log \frac{W(Y|X_1, X_2)}{Q_{Y|X_2}(Y|X_2)} & \log \frac{W(Y|X_1, X_2)}{Q_Y(Y)} \end{bmatrix}^T. \tag{27}$$

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

$$\mathbf{I}(\rho) = \mathbb{E}\big[\mathbf{j}(X_1, X_2, Y)\big], \tag{28}$$

$$\mathbf{V}(\rho) = \mathbb{E}\big[\,\mathsf{Cov}\left(\mathbf{j}(X_1, X_2, Y) \,\big|\, X_1, X_2\right)\big]. \tag{29}$$

For a given point $(z_1, z_2) \in \mathbb{R}^2$ and a (non-zero) positive semi-definite matrix $\mathbf{V}$, define

$$\Psi(z_1, z_2; \mathbf{V}) := \int_{-\infty}^{z_2} \int_{-\infty}^{z_1} \mathcal{N}(\mathbf{u}; \mathbf{0}, \mathbf{V}) \, \mathrm{d}\mathbf{u}, \tag{30}$$

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

$$\Psi^{-1}(\mathbf{V}, \varepsilon) := \big\{(z_1, z_2) \in \mathbb{R}^2 : \Psi(-z_1, -z_2; \mathbf{V}) \geq 1 - \varepsilon\big\}. \tag{31}$$

These quantities can be thought of as the generalization of the cumulative distribution function (cdf) of the standard Gaussian $\Phi(z) := \int_{-\infty}^{z} \mathcal{N}(u; 0, 1) \, \mathrm{d}u$ and its inverse $\Phi^{-1}(\varepsilon) := \sup\big\{z \in \mathbb{R} : \Phi(-z) \geq 1 - \varepsilon\big\}$ to the bivariate case.

For $\varepsilon < \frac{1}{2}$, the points contained in $\Psi^{-1}(\mathbf{V}, \varepsilon)$ have negative coordinates. See Fig. 3 for an illustration of (scaled versions of) $\Psi^{-1}(\mathbf{V}(\rho), \varepsilon)$.

Let $\underline{g}(\rho, \varepsilon, n)$ and $\overline{g}(\rho, \varepsilon, n)$ be arbitrary functions of ρ, ε and n for now, and define the inner and outer regions

$$\mathcal{R}_{\text{in}}(n, \varepsilon; \rho) := \left\{ (R_1, R_2) \in \mathbb{R}^2 \,:\, \mathbf{R} \in \mathbf{I}(\rho) + \frac{\Psi^{-1}(\mathbf{V}(\rho), \varepsilon)}{\sqrt{n}} + \underline{g}(\rho, \varepsilon, n)\mathbf{1} \right\}, \tag{32}$$

$$\mathcal{R}_{\text{out}}(n, \varepsilon; \rho) := \left\{ (R_1, R_2) \in \mathbb{R}^2 \,:\, \mathbf{R} \in \mathbf{I}(\rho) + \frac{\Psi^{-1}(\mathbf{V}(\rho), \varepsilon)}{\sqrt{n}} + \overline{g}(\rho, \varepsilon, n)\mathbf{1} \right\}. \tag{33}$$

Theorem 2 (Global Bounds on the (n, ε)-Capacity Region)**.** *There exist functions $\underline{g}(\rho, \varepsilon, n)$ and $\overline{g}(\rho, \varepsilon, n)$ such that the (n, ε)-capacity region satisfies*

$$\bigcup_{0 \le \rho \le 1} \mathcal{R}_{\text{in}}(n, \varepsilon; \rho) \subset \mathcal{C}(n, \varepsilon) \subset \bigcup_{-1 \le \rho \le 1} \mathcal{R}_{\text{out}}(n, \varepsilon; \rho), \tag{34}$$

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*

$$\underline{g}(\rho, \varepsilon, n) = O\left(\frac{\log n}{n}\right), \quad \textit{and} \quad \overline{g}(\rho, \varepsilon, n) = O\left(\frac{\log n}{n}\right). \tag{35}$$

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

$$\underline{g}(\rho_n, \varepsilon, n) = o\left(\frac{1}{\sqrt{n}}\right), \quad \textit{and} \quad \overline{g}(\rho_n, \varepsilon, n) = o\left(\frac{1}{\sqrt{n}}\right). \tag{36}$$

The proof of Theorem 2 is provided in Section VI. 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 in Section IV, 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.

IV. Local Second-Order Coding Rates

In this section, we present our main result, namely, the characterization of the $(\varepsilon, R_1^*, R_2^*)$-optimal second-order coding rate region $\mathcal{L}(\varepsilon; R_1^*, R_2^*)$ (see Definition 4), where (R_1^*, R_2^*) is an arbitrary point on the boundary of $\mathcal{C}$. Our result is stated in terms of the derivative of the mutual information vector with respect to ρ, namely

$$\mathbf{D}(\rho) = \begin{bmatrix} D_1(\rho) \\ D_{12}(\rho) \end{bmatrix} := \frac{\partial}{\partial \rho} \begin{bmatrix} I_1(\rho) \\ I_{12}(\rho) \end{bmatrix}, \tag{37}$$

where the individual derivatives are given by

$$\frac{\partial I_1(\rho)}{\partial \rho} = \frac{-S_1 \rho}{1 + S_1(1 - \rho^2)}, \tag{38}$$

$$\frac{\partial I_{12}(\rho)}{\partial \rho} = \frac{\sqrt{S_1 S_2}}{1 + S_1 + S_2 + 2\rho\sqrt{S_1 S_2}}. \tag{39}$$

Note that $\rho \in (0, 1]$ represents the strictly concave part of the boundary (see Fig. 2), and in this interval we have $D_1(\rho) < 0$ and $D_{12}(\rho) > 0$.

We introduce the following notation: For a vector $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$, define the *down-set* of $\mathbf{v}$ as

$$\mathbf{v}^- := \{(w_1, w_2) \in \mathbb{R}^2 : w_1 \le v_1, w_2 \le v_2\}. \tag{40}$$

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

Theorem 3 (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*

$$\mathcal{L}(\varepsilon; R_1^*, R_2^*) = \left\{ (L_1, L_2) \in \mathbb{R}^2 : L_1 \le \sqrt{V_1(0)}\Phi^{-1}(\varepsilon) \right\}. \tag{41}$$

(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*

$$\mathcal{L}(\varepsilon; R_1^*, R_2^*) = \left\{ (L_1, L_2) \in \mathbb{R}^2 : \begin{bmatrix} L_1 \\ L_1 + L_2 \end{bmatrix} \in \bigcup_{\beta \in \mathbb{R}} \left\{ \beta\, \mathbf{D}(\rho) + \Psi^{-1}(\mathbf{V}(\rho), \varepsilon) \right\} \right\}. \tag{42}$$

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

$$\mathcal{L}(\varepsilon; R_1^*, R_2^*) = \left\{ (L_1, L_2) \in \mathbb{R}^2 \,:\, \begin{bmatrix} L_1 \\ L_1 + L_2 \end{bmatrix} \in \bigcup_{\beta \le 0} \left\{ \beta\, \mathbf{D}(1) + \begin{bmatrix} 0 \\ \sqrt{V_{12}(1)}\Phi^{-1}(\varepsilon) \end{bmatrix}^{-} \right\} \right\}. \tag{43}$$

The proof of Theorem 3 is provided in Section VII. It leverages on the global second-order result in Theorem 2.

A. Discussion

Observe that in case (i), the second-order region is simply characterized by a scalar dispersion term $V_1(0)$ and the inverse of the Gaussian cdf Φ^{-1}. Roughly speaking, in this part of the boundary, there is effectively only a single rate constraint in terms of R_1, since we are operating "far away" from the sum rate constraint. This results in a large deviations-type event for the sum rate constraint which has no bearing on second-order asymptotics; see further discussions in [5], [10] and [18].

Cases (ii)–(iii) are more interesting, and their proofs are non-trivial. As in Nomura-Han [5] and Tan-Kosut [10], the second-order asymptotics for case (ii) depend on the dispersion matrix $\mathbf{V}(\rho)$ and the 2-dimensional analogue of the inverse of the Gaussian cdf Ψ^{-1}, since *both* rate constraints are active at a point on the boundary parametrized by $\rho \in (0, 1)$. However, in our setting, the expression containing Ψ^{-1} alone (i.e. the expression obtained by setting $\beta = 0$ in (42)) 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, since there are non-empty regions within the capacity region that are not contained in the trapezium of rate pairs achievable using $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\rho))$.

Thus, to achieve all (L_1, L_2) pairs, we must allow the input distribution to vary with the blocklength n. This is manifested in the $\beta\, \mathbf{D}(\rho)$ term. Our proof of the direct part involves random coding with an input distribution of the form $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\rho_n))$, where $\rho_n - \rho = O\big(\frac{1}{\sqrt{n}}\big)$. By a Taylor series, the resulting mutual information vector $\mathbf{I}(\rho_n)$ is approximately $\mathbf{I}(\rho) + (\rho_n - \rho)\mathbf{D}(\rho)$. Since $\rho_n - \rho = O\big(\frac{1}{\sqrt{n}}\big)$, the gradient term $(\rho_n - \rho)\mathbf{D}(\rho)$ also contributes to the second-order behavior, together with the Gaussian approximation term $\Psi^{-1}(\mathbf{V}(\rho), \varepsilon)$.

For the converse, we consider an arbitrary sequence of codes with rate pairs $\{(R_{1,n}, R_{2,n})\}_{n \ge 1}$ converging to $(I_1(\rho), I_{12}(\rho) - I_1(\rho))$ with second-order behavior given by (15). From the global result, we know $[R_{1,n}, R_{1,n} + R_{2,n}]^T \in \mathcal{R}_{\mathrm{out}}(n, \varepsilon; \rho_n)$ for some sequence of ρ_n. We then establish, using the definition of the second-order coding rates in (15), that $\rho_n = \rho + O\big(\frac{1}{\sqrt{n}}\big)$. Finally, by the Bolzano-Weierstrass theorem [20, Thm. 3.6(b)], we may pass to a subsequence of ρ_n (if necessary), and this establishes the converse.

A similar discussion holds true for case (iii); the main differences are that the covariance matrix is singular, and that the union in (43) is taken over $\beta \le 0$ only, since ρ_n can only approach one from below.

B. Second-Order Asymptotics for a Given Angle of Approach

Here we study the second-order behavior when a point on the boundary is approached from a given angle, as was done in Tan-Kosut [10]. If $\varepsilon < \frac{1}{2}$ (resp. $\varepsilon > \frac{1}{2}$), we approach a boundary point from inside (resp. outside) the capacity region. We focus on the most interesting case in Theorem 3, namely, case (ii) corresponding to $\rho \in (0, 1)$. Case (iii) can be handled similarly, and in case (i) the angle of approach is of little interest since L_2 can be arbitrarily large or small.

First, we present an alternative expression for the set $\mathcal{L} = \mathcal{L}(\varepsilon; R_1^*, R_2^*)$ given in (42) with $R_1^* = I_1(\rho)$ and $R_1^* + R_2^* = I_{12}(\rho)$ for some $\rho \in (0, 1)$. It is easily seen that $(L_1, L_2) \in \mathcal{L}$ implies $(L_1 + \beta D_1(\rho), L_2 + \beta D_2(\rho)) \in \mathcal{L}$, where $D_2(\rho) := D_{12}(\rho) - D_1(\rho)$. It follows that $\mathcal{L}$ equals the set of all points lying below a straight line with slope $\frac{D_2(\rho)}{D_1(\rho)}$ which intersects the boundary of $\Psi^{-1}(\mathbf{V}(\rho), \varepsilon)$. That is,

$$\mathcal{L}(\varepsilon; R_1^*, R_2^*) = \Big\{ (L_1, L_2) \,:\, L_2 \le a_\rho L_1 + b_{\rho, \varepsilon} \Big\}, \tag{44}$$

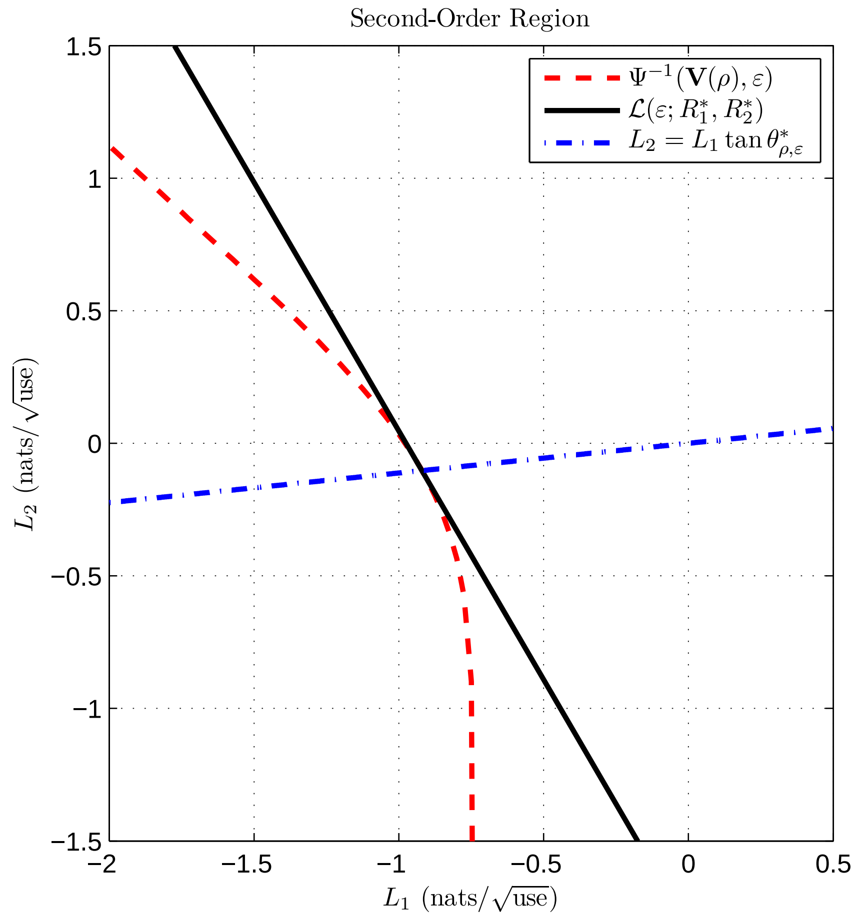

Fig. 4. Second-order coding rates in nats/$\sqrt{\text{use}}$ with $S_1 = S_2 = 1$, $\rho = \frac{1}{2}$ and $\varepsilon = 0.1$. The regions $\Psi^{-1}(\mathbf{V}(\rho),\varepsilon)$ and $\mathcal{L}(\varepsilon;R_1^*,R_2^*)$ are to the bottom left of the boundaries. We also plot the line $L_2 = L_1 \tan\theta^*_{\rho,\varepsilon}$, where $\theta^*_{\rho,\varepsilon}$ is the unique angle θ for which the intersection of the boundary of $\mathcal{L}(\varepsilon;R_1^*,R_2^*)$ and the line $L_2 = L_1\tan\theta$ coincides with the boundary of $\Psi^{-1}(\mathbf{V}(\rho),\varepsilon)$.

where

$$a_\rho := \frac{D_2(\rho)}{D_1(\rho)}, \quad \text{and} \quad b_{\rho,\varepsilon} := \inf\Big\{b \,:\, \big(L_1, a_\rho L_1 + b\big) \in \Psi^{-1}(\mathbf{V}(\rho),\varepsilon) \text{ for some } L_1 \in \mathbb{R}\Big\}. \tag{45}$$

We provide an example in Fig. 4 with the parameters $S_1 = S_2 = 1$, $\rho = \frac{1}{2}$ and $\varepsilon = 0.1$. Since $\varepsilon < \frac{1}{2}$, the boundary point (R_1^*, R_2^*) is approached from the inside. See Fig. 3, where for $\varepsilon < \frac{1}{2}$, the set $\Psi^{-1}(\mathbf{V},\varepsilon)$ only contains points with negative coordinates.

Given the gradient a_ρ, the offset $b_{\rho,\varepsilon}$, and an angle θ (measured with respect to the horizontal axis), we seek the pair (L_1, L_2) on the boundary of $\mathcal{L}(\varepsilon;R_1^*,R_2^*)$ such that $L_2 = L_1\tan\theta$. It is easily seen that this point is obtained by solving for the intersection of the line $L_2 = a_\rho L_1 + b_{\rho,\varepsilon}$ with $L_2 = L_1 \tan\theta$. The two lines coincide when

$$L_1 = \frac{b_{\rho,\varepsilon}}{\tan\theta - a_\rho}, \quad \text{and} \quad L_2 = \frac{b_{\rho,\varepsilon}\tan\theta}{\tan\theta - a_\rho}. \tag{46}$$

In Fig. 4, we see that there is only a single angle $\theta^*_{\rho,\varepsilon} \approx 3.253$ rads for which the point of intersection in (46) is also on the boundary of $\Psi^{-1}(\mathbf{V}(\rho),\varepsilon)$, yielding $(L_1, L_2) \approx (-0.920, -0.103)$. In other words, there is only one angle for which coding with a fixed (not varying with n) input distribution $\mathcal{N}(\mathbf{0},\mathbf{V}(\rho))$ is optimal in the second-order sense (i.e. for which the added term $\beta\mathbf{D}(\rho)$ in (42) is of no additional help and $\beta = 0$ is optimal). For all the other angles, we should choose a non-zero coefficient β, which corresponds to choosing an input distribution which varies with n.

Finally, in Fig. 5, we plot the norm of the vector of second-order rates $[L_1, L_2]^T$ in (46) against θ, the angle of approach. For $\varepsilon < \frac{1}{2}$, the point $[L_1, L_2]^T$ may be interpreted as that corresponding to the "smallest backoff" from

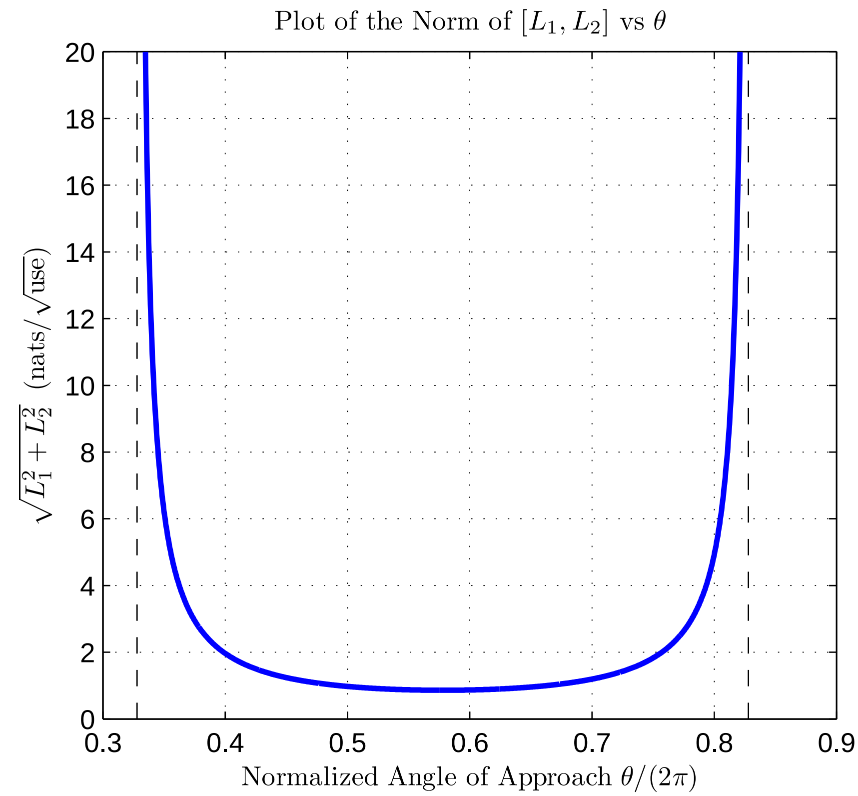

Fig. 5. Plot of $\sqrt{L_1^2+L_2^2}$ against the angle of approach θ with the same parameters as in Fig. 4. The second-order rates L_1, L_2, as functions of θ, are defined in (46). Here, $\pi + \arctan(a_\rho) \approx 0.328 \times (2\pi)$ and $2\pi + \arctan(a_\rho) \approx 0.828 \times (2\pi)$ are the critical angles where $\sqrt{L_1^2+L_2^2}$ diverges.

the first-order optimal rates.[1] Thus, $\sqrt{L_1^2+L_2^2}$ is a measure of the total backoff. For $\varepsilon > \frac{1}{2}$, $[L_1, L_2]^T$ corresponds to the "largest addition" to the first-order rates. It is noted that the norm tends to infinity when the angle tends to $\pi + \arctan(a_\rho)$ (from above) or $2\pi + \arctan(a_\rho)$ (from below). This corresponds to an approach almost parallel to the gradient at the point on the boundary parametrized by ρ. A similar phenomenon was observed for the Slepian-Wolf problem [10].

V. Concluding Remarks

We have studied the second-order asymptotics (i.e. identified the optimal second-order coding rate region) of the Gaussian MAC with DMS. There are two reasons as to why the analysis here is more tractable vis-à-vis finite blocklength or second-order analysis for the the discrete memoryless MAC (DM-MAC) studied extensively in [10]–[12], [16]–[18]. Gaussianity allows us to identify the boundary of the capacity region and associate each point on the boundary with an input distribution parametrized by ρ. Because for the DM-MAC, one needs to take the convex closure of the union over input distributions P_{X_1,X_2} to define the capacity region [1, Sec. 4.5], the boundary points are more difficult to characterize. In addition, in the absence of the DMS assumption, one needs to ensure in a converse proof (possibly related to the *wringing technique* of Ahlswede [21]) that the codewords pairs are almost orthogonal. By leveraging on the DMS assumption, we circumvent this requirement.

For future investigations, we note that the Gaussian broadcast channel [1, Sec. 5.5] is a problem which is very similar to the Gaussian MAC with DMS (both require superposition coding and each point on the boundary is achieved by a unique input distribution). As such, we expect that some of the second-order analysis techniques contained herein may be applicable to the Gaussian broadcast channel.

[1]There may be some imprecision in the use of the word "backoff" here as for angles in the second (resp. fourth) quadrant, L_2 (resp. L_1) is positive. On the other hand, one could generally refer to "backoff" as moving in some inward direction (relative to the capacity region boundary) even if it is in a direction where one of the second-order rates increases. The same goes for the term "addition".

VI. Proof of Theorem 2: Global Second-Order Result

A. Converse Part

We first prove the outer bound in (34). The analysis is split into seven steps.

1) A Reduction from Maximal to Equal Power Constraints: Let $\mathcal{C}_{\mathrm{eq}}(n,\varepsilon)$ be the (n,ε)-capacity region in the case that (9) and (10) are equality constraints, i.e., $\|f_{1,n}(m_1,m_2)\|_2^2 = nS_1$ and $\|f_{2,n}(m_2)\|_2^2 = nS_2$ for all (m_1,m_2). We claim that

$$\mathcal{C}_{\mathrm{eq}}(n,\varepsilon) \subset \mathcal{C}(n,\varepsilon) \subset \mathcal{C}_{\mathrm{eq}}(n+1,\varepsilon). \tag{47}$$

The lower bound is obvious. The upper bound follows by noting that the decoder for the length-$(n+1)$ code can ignore the last symbol, which can be chosen to equalize the powers.

It follows from (47) that for the purpose of second-order asymptotics, $\mathcal{C}_{\mathrm{eq}}(n,\varepsilon)$ and $\mathcal{C}(n,\varepsilon)$ are equivalent. This argument was also used in [8, Lem. 39] and [9, Sec. XIII]. Henceforth, we assume that all codewords $(\mathbf{x}_1,\mathbf{x}_2)$ have empirical powers *exactly* equal to (S_1,S_2).

2) A Reduction from Average to Maximal Error Probability: Let $\mathcal{C}_{\max}(n,\varepsilon)$ be the (n,ε)-capacity region in the case that, along with the replacements in the previous step, (11) is replaced by

$$\max_{m_1\in[M_{1,n}],m_2\in[M_{2,n}]} \Pr\left((\mathsf{M}_1,\mathsf{M}_2)\neq(\hat{\mathsf{M}}_1,\hat{\mathsf{M}}_2)\,\middle|\,(\mathsf{M}_1,\mathsf{M}_2)=(m_1,m_2)\right)\le\varepsilon_n. \tag{48}$$

That is, the average error probability is replaced by the maximal error probability. Here we show that $\mathcal{C}(n,\varepsilon)$ and $\mathcal{C}_{\max}(n,\varepsilon)$ are equivalent for the purposes of second-order asymptotics, thus allowing us to focus on the maximal error probability for the converse proof. It should be noted that this argument fails for the regular MAC [22].

Using similar arguments to [23, Sec 3.4.4], we will start with the average-error code, and use a standard expurgation argument to obtain a maximal-error code having the same asymptotic rates and error probability. Let $\varepsilon_n(m_1,m_2)$ be the error probability given that the message pair (m_1,m_2) is encoded, and let

$$\varepsilon_n(m_2) := \frac{1}{M_{1,n}}\sum_{m_1=1}^{M_{1,n}}\varepsilon_n(m_1,m_2) \tag{49}$$

be the error probability for message m_2, averaged over M_1.

Consider a sequence of codes with message sets $\mathcal{M}_{1,n}$ and $\mathcal{M}_{2,n}$, having an error probability not exceeding ε_n. Let $\tilde{\mathcal{M}}_{2,n}$ contain the fraction $\frac{1}{\sqrt{n}}$ of the messages $m_2\in\mathcal{M}_{2,n}$ with the highest values of $\varepsilon_n(m_2)$ (here and subsequently, we ignore rounding issues, since these do not affect the argument). It follows that $\varepsilon_n(m_2)\le\frac{\varepsilon_n}{1-\frac{1}{\sqrt{n}}}$, since otherwise the codewords not appearing in $\tilde{\mathcal{M}}_{2,n}$ would contribute more than ε_n to the average error probability of the original code, causing a contradiction.

Next, for each $m_2\in\tilde{\mathcal{M}}_{2,n}$, let $\tilde{\mathcal{M}}_{1,n}(m_2)$ contain the fraction $\frac{1}{\sqrt{n}}$ of the messages m_1 with the highest values of $\varepsilon_n(m_1,m_2)$. Since user 1 knows both messages, we may relabel the messages so that $\tilde{\mathcal{M}}_{1,n}:=\tilde{\mathcal{M}}_{1,n}(m_2)$ is the same for each m_2. Repeating the above argument, we conclude that

$$\varepsilon_n(m_1,m_2)\le\frac{\varepsilon_n(m_2)}{1-\frac{1}{\sqrt{n}}}\le\frac{\varepsilon_n}{\left(1-\frac{1}{\sqrt{n}}\right)^2}=\varepsilon_n+O\left(\frac{1}{\sqrt{n}}\right) \tag{50}$$

for all $m_1\in\tilde{\mathcal{M}}_{1,n}$ and $m_2\in\tilde{\mathcal{M}}_{2,n}$. Moreover, we have by construction that

$$\frac{1}{n}\log\left|\tilde{\mathcal{M}}_{j,n}\right| = \frac{1}{n}\log\left|\mathcal{M}_{j,n}\right| - \frac{1}{2}\frac{\log n}{n} \tag{51}$$

for $j=1,2$. By absorbing the remainder terms in (50) and (51) into the third-order term $\overline{g}(\rho,\varepsilon,n)$ in (33), we see that it suffices to prove the converse result for the maximal 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 $k \in [n]$. We see that the family $\{\mathcal{I}_k : k \in [-n : n]\}$ forms a partition of $[-1, 1]$. Consider the *correlation type classes* (or simply *type classes*)

$$\mathcal{T}_n(k) := \left\{(\mathbf{x}_1, \mathbf{x}_2) : \frac{\langle \mathbf{x}_1, \mathbf{x}_2\rangle}{\|\mathbf{x}_1\|_2 \|\mathbf{x}_2\|_2} \in \mathcal{I}_k\right\} \tag{52}$$

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 case of discrete alphabets [24, Ch. 2].

We use an argument similar to that of Csiszár-Körner [24, Lem. 16.2] to perform a further reduction (along with those in the first two steps) to codes for which all codeword pairs have the same type. Let the codebook $\mathcal{C} := \{(\mathbf{x}_1(m_1, m_2), \mathbf{x}_2(m_2)) : m_1 \in \mathcal{M}_{1,n}, m_2 \in \mathcal{M}_{2,n}\}$ be given; in accordance with the previous two steps, we assume that it has codewords meeting the power constraints with equality, and *maximal* error probability not exceeding ε_n. For each $m_2 \in \mathcal{M}_{2,n}$, we can find a set $\tilde{\mathcal{M}}_{1,n}(m_2) \subset \mathcal{M}_{1,n}$ (re-using the notation of the previous step) such that all pairs of codewords $(\mathbf{x}_1(m_1, m_2), \mathbf{x}_2(m_2))$, $m_1 \in \tilde{\mathcal{M}}_{1,n}(m_2)$ have the same type, say indexed by $k(m_2) \in [-n : n]$, and

$$\frac{1}{n} \log \big|\tilde{\mathcal{M}}_{1,n}(m_2)\big| \ge \frac{1}{n} \log \big|\mathcal{M}_{1,n}(m_2)\big| - \frac{\log(2n+1)}{n}, \quad \forall\, m_2 \in \mathcal{M}_{2,n}. \tag{53}$$

We may assume that all the sets $\tilde{\mathcal{M}}_{1,n}(m_2), m_2 \in \mathcal{M}_{2,n}$ have the same cardinality; otherwise, we can remove extra codeword pairs from some sets $\tilde{\mathcal{M}}_{1,n}(m_2)$ and (53) will still be satisfied. We may also assume (by relabeling if necessary) that $\tilde{\mathcal{M}}_{1,n} := \tilde{\mathcal{M}}_{1,n}(m_2)$ is the same for each m_2. Now, we have a subcodebook $\tilde{\mathcal{C}}_1 := \{(\mathbf{x}_1(m_1, m_2), \mathbf{x}_2(m_2)) : m_1 \in \tilde{\mathcal{M}}_{1,n}, m_2 \in \mathcal{M}_{2,n}\}$, where for each m_2, all the codeword pairs have the same type and (53) is satisfied. Across the m_2's, there may be different types indexed by $k(m_2) \in [-n : n]$, but there exists a dominant type indexed by $k^* \in \{k(m_2) : m_2 \in \mathcal{M}_{2,n}\}$ and a set $\tilde{\mathcal{M}}_{2,n} \subset \mathcal{M}_{2,n}$ such that

$$\frac{1}{n} \log \big|\tilde{\mathcal{M}}_{2,n}\big| \ge \frac{1}{n} \log \big|\mathcal{M}_{2,n}\big| - \frac{\log(2n+1)}{n}. \tag{54}$$

As such, we have shown that there exists a subcodebook $\tilde{\mathcal{C}}_{12} := \{(\mathbf{x}_1(m_1, m_2), \mathbf{x}_2(m_2)) : m_1 \in \tilde{\mathcal{M}}_{1,n}, m_2 \in \tilde{\mathcal{M}}_{2,n}\}$ of constant type indexed by k^* where the sum rate satisfies

$$\frac{1}{n} \log \big|\tilde{\mathcal{M}}_{1,n} \times \tilde{\mathcal{M}}_{2,n}\big| \ge \frac{1}{n} \log \big|\mathcal{M}_{1,n} \times \mathcal{M}_{2,n}\big| - \frac{2\log(2n+1)}{n}. \tag{55}$$

The reduced code clearly has maximal error probability no higher than that of $\mathcal{C}$. Combining this observation with (54) and (55), we see that the converse part of Theorem 2 for fixed-type codes implies the same for general codes, since the additional $O\big(\frac{\log n}{n}\big)$ factors in (54) and (55) can be absorbed into the third-order term $\overline{g}(\rho, \varepsilon, n)$. Thus, in the remainder of the proof, we limit our attention to fixed-type codes. For each n, the type is indexed by $k \in [-n : n]$, and we define $\hat{\rho} := \frac{k}{n} \in [-1, 1]$. In some cases, we will be interested in *sequences* of such values, in which case we will make the dependence on n explicit by writing $\hat{\rho}_n$.

4) A Verdú-Han-type Converse Bound for the Gaussian MAC with DMS: We now state a non-asymptotic converse bound for the Gaussian MAC with DMS based on analogous bounds in Han's work on the information spectrum approach for the general MAC [25, Lem. 4] and in Boucheron-Salamatian's work on the information spectrum approach for the general broadcast channel with DMS [26, Lem. 2]. This result can be proved similarly to [25], [26], so we omit its proof.

The bound only requires that the *average* error probability is no higher than ε_n, which is guaranteed by the fact that the maximal error probability is no higher than ε_n. That is, the reduction to the maximal error probability in Section VI-A2 was performed for the sole purpose of making the reduction to fixed types in Section VI-A3 possible.

Proposition 4. *Fix a blocklength $n \ge 1$, auxiliary output distributions $Q_{\mathbf{Y}|\mathbf{X}_2}$ and $Q_{\mathbf{Y}}$, and a constant $\gamma > 0$. For any $(n, M_{1,n}, M_{2,n}, S_1, S_2, \varepsilon_n)$-code with codewords of fixed empirical powers S_1 and S_2 falling into a single correlation type class $\mathcal{T}_n(k)$, there exist random vectors $(\mathbf{X}_1, \mathbf{X}_2)$ with joint distribution $P_{\mathbf{X}_1,\mathbf{X}_2}$ supported on $\{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{T}_n(k) : \|\mathbf{x}_j\|_2^2 = nS_j, j = 1, 2\}$ such that*

$$\varepsilon_n \ge \Pr(\mathcal{A}_n \cup \mathcal{B}_n) - 2\exp(-n\gamma), \tag{56}$$

where

$$\mathcal{A}_n := \left\{ \frac{1}{n} \log \frac{W^n(\mathbf{Y}|\mathbf{X}_1, \mathbf{X}_2)}{Q_{\mathbf{Y}|\mathbf{X}_2}(\mathbf{Y}|\mathbf{X}_2)} \le \frac{1}{n} \log M_{1,n} - \gamma \right\} \tag{57}$$

$$\mathcal{B}_n := \left\{ \frac{1}{n} \log \frac{W^n(\mathbf{Y}|\mathbf{X}_1, \mathbf{X}_2)}{Q_{\mathbf{Y}}(\mathbf{Y})} \le \frac{1}{n} \log \big(M_{1,n} M_{2,n}\big) - \gamma \right\}, \tag{58}$$

with $\mathbf{Y} \,|\, \{\mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2\} \sim W^n(\cdot|\mathbf{x}_1, \mathbf{x}_2)$.

There are several differences in Proposition 4 compared to [25, Lem. 4]. First, in our work, there are cost constraints on the codewords, and thus the support of the input distribution $P_{\mathbf{X}_1,\mathbf{X}_2}$ is specified to reflect this constraint. Second, there are two (instead of three) events in the probability in (56) because the informed encoder $f_{1,n}$ has access to both messages. Third, we can choose arbitrary output distributions $Q_{\mathbf{Y}|\mathbf{X}_2}$ and $Q_{\mathbf{Y}}$. This generalization is analogous to the non-asymptotic converse bound by Hayashi and Nagaoka for classical-quantum channels [27, Lem. 4]. The freedom to choose the output distribution is crucial in both our problem and [27].

5) Evaluation of the Verdú-Han Bound for $\hat{\rho} \in (-1, 1)$: Recall from Sections VI-A1 and VI-A3 that the codewords satisfy exact power constraints and belong to a single type class $\mathcal{T}_n(k)$. In this subsection, we consider the case that $\hat{\rho} := \frac{k}{n} \in (-1, 1)$, and we derive bounds which will be useful for sequences $\hat{\rho}_n$ uniformly bounded away from -1 and 1. In Section VI-A6, we present alternative bounds to handle the remaining cases.

We evaluate (56) for a single correlation type class parametrized by k. Let $\gamma := \frac{\log n}{2n}$, yielding $2\exp(-n\gamma) = \frac{2}{\sqrt{n}}$. We choose the output distributions $Q_{\mathbf{Y}|\mathbf{X}_2}$ and $Q_{\mathbf{Y}}$ to be the n-fold products of $Q_{Y|X_2}$ and Q_Y, defined in (25)–(26) respectively, with $\hat{\rho}$ in place of ρ.

We now characterize the statistics of the first and second moments of $\mathbf{j}(x_{1i}, x_{2i}, Y_i)$ in (27) for fixed sequences $(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{T}_n(k)$. From Appendix A, these moments can be expressed as continuously differentiable functions of the empirical powers $\frac{1}{n}\|\mathbf{x}_1\|_2^2$, $\frac{1}{n}\|\mathbf{x}_2\|_2^2$ and the empirical correlation coefficient $\frac{\langle \mathbf{x}_1, \mathbf{x}_2 \rangle}{\|\mathbf{x}_1\|_2 \|\mathbf{x}_2\|_2}$. The former two quantities are fixed due to the reduction in Section VI-A1, and the latter is within $\frac{1}{n}$ of $\hat{\rho}$ by the assumption that $(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{T}_n(k)$. Thus, Taylor expanding (A.6) and (A.12) in Appendix A, we obtain

$$\left\| \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \mathbf{j}(x_{1i}, x_{2i}, Y_i) \right] - \mathbf{I}(\hat{\rho}) \right\|_\infty \le \frac{\xi_1}{n} \tag{59}$$

$$\left\| \mathsf{Cov}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{j}(x_{1i}, x_{2i}, Y_i) \right] - \mathbf{V}(\hat{\rho}) \right\|_\infty \le \frac{\xi_2}{n} \tag{60}$$

for some $\xi_1 > 0$ and $\xi_2 > 0$ which can taken to be independent of $\hat{\rho}$ (since the corresponding derivatives are uniformly bounded).

Let $R_{j,n} := \frac{1}{n} \log M_{j,n}$ for $j = 1, 2$, and let $\mathbf{R}_n := [R_{1,n}, R_{1,n} + R_{2,n}]^T$. We have

$$\Pr(\mathcal{A}_n \cup \mathcal{B}_n) = 1 - \Pr(\mathcal{A}_n^c \cap \mathcal{B}_n^c) = 1 - \mathbb{E}_{\mathbf{X}_1,\mathbf{X}_2}\big[\Pr(\mathcal{A}_n^c \cap \mathcal{B}_n^c | \mathbf{X}_1, \mathbf{X}_2) \big] \tag{61}$$

and in particular, using the definition of $\mathbf{j}(x_1, x_2, y)$ in (27) and the fact that $Q_{\mathbf{Y}|\mathbf{X}_2}$ and $Q_{\mathbf{Y}}$ are product distributions,

$$\Pr(\mathcal{A}_n^c \cap \mathcal{B}_n^c | \mathbf{x}_1, \mathbf{x}_2) = \Pr\left(\frac{1}{n} \sum_{i=1}^n \mathbf{j}(x_{1i}, x_{2i}, Y_i) > \mathbf{R}_n - \gamma \mathbf{1} \right) \tag{62}$$

$$\le \Pr\left(\frac{1}{n} \sum_{i=1}^n \Big(\mathbf{j}(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[\mathbf{j}(x_{1i}, x_{2i}, Y_i)] \Big) > \mathbf{R}_n - \mathbf{I}(\hat{\rho}) - \gamma \mathbf{1} - \frac{\xi_1}{n} \mathbf{1} \right), \tag{63}$$

where (63) follows from (59).

We are now in a position to apply the multivariate Berry-Esseen theorem [28], [29] (see Appendix B). The first two moments are bounded according to (59)–(60), and in Appendix A we show that, upon replacing the given $(\mathbf{x}_1, \mathbf{x}_2)$ pair with a different pair yielding the same statistics of $\sum_{i=1}^n \mathbf{j}(x_{1i}, x_{2i}, Y_i)$ if necessary (cf. Lemma 9), the required third moment is uniformly bounded (cf. Lemma 10). It follows that

$$\begin{aligned}&\Pr(\mathcal{A}_n^c \cap \mathcal{B}_n^c|\mathbf{x}_1,\mathbf{x}_2)\\&\le \Psi\left(\sqrt{n}\Big(I_1(\hat\rho)+\gamma+\frac{\xi_1}{n}-R_{1,n}\Big), \sqrt{n}\Big(I_{12}(\hat\rho)+\gamma+\frac{\xi_1}{n}-(R_{1,n}+R_{2,n})\Big); \mathsf{Cov}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{j}(x_{1i},x_{2i},Y_i)\right]\right)+\frac{\psi(\hat\rho)}{\sqrt{n}}.\end{aligned} \tag{64}$$

By Taylor expanding the continuously differentiable function $(z_1,z_2,\mathbf{V})\mapsto\Psi(z_1,z_2;\mathbf{V})$, using the approximation in (60) and the fact that $\det(\mathbf{V}(\hat\rho))>0$ for $\hat\rho\in(-1,1)$, we obtain

$$\Pr(\mathcal{A}_n^c \cap \mathcal{B}_n^c|\mathbf{x}_1,\mathbf{x}_2)\le\Psi\left(\sqrt{n}\big(I_1(\hat\rho)-R_{1,n}\big),\sqrt{n}\big(I_{12}(\hat\rho)-(R_{1,n}+R_{2,n})\big);\mathbf{V}(\hat\rho)\right)+\frac{\eta(\hat\rho)\log n}{\sqrt{n}}. \tag{65}$$

The remainder terms in (64) and (65) are dependent on $\hat\rho$, so we denote them by $\psi(\hat\rho)$ and $\eta(\hat\rho)$. These remainder terms satisfy $\psi(\hat\rho),\eta(\hat\rho)\to\infty$ as $\hat\rho\to\pm1$, since $\mathbf{V}(\hat\rho)$ becomes singular as $\hat\rho\to\pm1$. Despite this non-uniformity, we conclude from (56) and (65) that any (n,ε)-code with codewords in $\mathcal{T}_n(k)$ must, for large enough n, have rates that satisfy

$$\begin{bmatrix}R_{1,n}\\R_{1,n}+R_{2,n}\end{bmatrix}\in\mathbf{I}(\hat\rho)+\frac{\Psi^{-1}\Big(\mathbf{V}(\hat\rho),\varepsilon+\frac{2}{\sqrt{n}}+\frac{\eta(\hat\rho)\log n}{\sqrt{n}}\Big)}{\sqrt{n}}. \tag{66}$$

The following "continuity" lemma for $\varepsilon\mapsto\Psi^{-1}(\mathbf{V},\varepsilon)$ is proved in Appendix C.

Lemma 5. *Fix $0<\varepsilon<1$ and a positive sequence $\lambda_n=o(1)$. Let $\mathbf{V}$ be a non-zero positive semi-definite matrix. There exists a function $h(\mathbf{V},\varepsilon)$ such that*

$$\Psi^{-1}\big(\mathbf{V},\varepsilon+\lambda_n\big)\subset\Psi^{-1}\big(\mathbf{V},\varepsilon\big)+h(\mathbf{V},\varepsilon)\,\lambda_n\,\mathbf{1}. \tag{67}$$

We conclude from Lemma 5 that

$$\Psi^{-1}\Big(\mathbf{V}(\hat\rho),\varepsilon+\frac{2}{\sqrt{n}}+\frac{\eta(\hat\rho)\log n}{\sqrt{n}}\Big)\subset\Psi^{-1}\big(\mathbf{V}(\hat\rho),\varepsilon\big)+\frac{h(\hat\rho,\varepsilon)\log n}{\sqrt{n}}\mathbf{1} \tag{68}$$

for some function $h(\hat\rho,\varepsilon):=h(\mathbf{V}(\hat\rho),\varepsilon)$ which diverges only as $\hat\rho\to\pm1$. Uniting (66) and (68), we deduce that

$$\begin{bmatrix}R_{1,n}\\R_{1,n}+R_{2,n}\end{bmatrix}\in\mathbf{I}(\hat\rho)+\frac{\Psi^{-1}\big(\mathbf{V}(\hat\rho),\varepsilon\big)}{\sqrt{n}}+\frac{h(\hat\rho,\varepsilon)\log n}{n}\mathbf{1}. \tag{69}$$

6) Evaluation of the Verdú-Han Bound with $\hat\rho_n\to\pm1$: Here we consider a sequence of codes of a single type indexed by k_n such that $\hat\rho_n:=\frac{k_n}{n}\to1$. The case $\hat\rho_n\to-1$ is handled similarly, and the details are thus omitted.

Our aim is to show that

$$\begin{bmatrix}R_{1,n}\\R_{1,n}+R_{2,n}\end{bmatrix}\in\mathbf{I}(\hat\rho_n)+\frac{\Psi^{-1}\big(\mathbf{V}(\hat\rho_n),\varepsilon\big)}{\sqrt{n}}+o\left(\frac{1}{\sqrt{n}}\right)\mathbf{1}. \tag{70}$$

The following lemma states that as $\hat\rho_n\to1$, the set $\Psi^{-1}\big(\mathbf{V}(\hat\rho_n),\varepsilon\big)$ in (70) can be approximated by $\Psi^{-1}\big(\mathbf{V}(1),\varepsilon\big)$, which is a simpler rectangular set. The proof of the lemma is provided in Appendix D.

Lemma 6. *Fix $0<\varepsilon<1$ and let $\hat\rho_n\to1$. There exist positive sequences $a_n,b_n=\Theta((1-\hat\rho_n)^{1/4})$ and $c_n=\Theta((1-\hat\rho_n)^{1/2})$ satisfying*

$$\begin{bmatrix}0\\\sqrt{V_{12}(1)}\Phi^{-1}(\varepsilon+a_n)\end{bmatrix}^{-}-b_n\mathbf{1}\subset\Psi^{-1}(\mathbf{V}(\hat\rho_n),\varepsilon)\subset\begin{bmatrix}0\\\sqrt{V_{12}(1)}\Phi^{-1}(\varepsilon)\end{bmatrix}^{-}+c_n\mathbf{1}. \tag{71}$$

The down-set notation $[\mathbf{v}]^-$ is defined in (40).

From the inner bound in Lemma 6, in order to show (70) it suffices to show

$$\begin{bmatrix}R_{1,n}\\R_{1,n}+R_{2,n}\end{bmatrix}\le\mathbf{I}(\hat\rho_n)+\sqrt{\frac{V_{12}(1)}{n}}\begin{bmatrix}0\\\Phi^{-1}(\varepsilon)\end{bmatrix}+o\left(\frac{1}{\sqrt{n}}\right)\mathbf{1}, \tag{72}$$

where we absorbed the sequences a_n, b_n into the $o\big(\frac{1}{\sqrt{n}}\big)$ term.

We return to the step in (63), which when combined with the Verdú-Han bound (with $\gamma := \frac{\log n}{2n}$) yields for some $(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{T}_n(k)$ that

$$\varepsilon_n \geq 1 - \Pr\left(\frac{1}{n}\sum_{i=1}^{n}\Big(\mathbf{j}(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[\mathbf{j}(x_{1i}, x_{2i}, Y_i)]\Big) > \mathbf{R}_n - \mathbf{I}(\hat{\rho}_n) - \gamma\mathbf{1} - \frac{\xi_1}{n}\mathbf{1}\right) - \frac{2}{\sqrt{n}} \tag{73}$$

$$\geq \max\Bigg\{\Pr\left(\frac{1}{n}\sum_{i=1}^{n}\Big(j_1(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[j_1(x_{1i}, x_{2i}, Y_i)]\Big) \leq R_{1,n} - I_1(\hat{\rho}_n) - \gamma - \frac{\xi_1}{n}\right),$$
$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n}\Big(j_{12}(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[j_{12}(x_{1i}, x_{2i}, Y_i)]\Big) \leq R_{1,n} + R_{2,n} - I_{12}(\hat{\rho}_n) - \gamma - \frac{\xi_1}{n}\right)\Bigg\} - \frac{2}{\sqrt{n}}. \tag{74}$$

From (60) and the assumption that $\hat{\rho}_n \to 1$, the variance of $\sum_{i=1}^{n} j_{12}(x_{1i}, x_{2i}, Y_i)$ equals $n(V_{12}(1) + o(1))$. Since $V_{12}(1) > 0$, we can treat the second term in the maximum in (74) in an identical fashion to the single-user setting [7], [8] to obtain the second of the element-wise inequalities in (72). It remains to prove the first, i.e. to show that no $\Theta\big(\frac{1}{\sqrt{n}}\big)$ addition to $R_{1,n}$ is possible for $\varepsilon \in (0,1)$. We will make use of the identities

$$I_1(\hat{\rho}_n) = \Theta(1 - \hat{\rho}_n) \tag{75}$$
$$V_1(\hat{\rho}_n) = \Theta(1 - \hat{\rho}_n) \tag{76}$$

which follow since $I_1(1) = V_1(1) = 0$, and by applying Taylor expansions in the same way as (59)–(60) (see also Appendix A). It is easily verified that the corresponding derivatives (e.g. $\frac{\partial I_1}{\partial \rho}$) at $\rho = 1$ are strictly negative, hence justifying the use of $\Theta(\cdot)$ instead of $O(\cdot)$ in (75)–(76).

We treat the cases $1 - \hat{\rho}_n = \omega\big(\frac{1}{n}\big)$ and $1 - \hat{\rho}_n = O\big(\frac{1}{n}\big)$ separately. In the former case, we combine (59)–(60) and (75)–(76) to conclude that

$$\tilde{I}_{1,n} := \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} j_1(x_{1i}, x_{2i}, Y_i)\right] = \Theta\left(1 - \hat{\rho}_n\right), \tag{77}$$

$$\tilde{V}_{1,n} := \mathsf{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} j_1(x_{1i}, x_{2i}, Y_i)\right] = \Theta\left(1 - \hat{\rho}_n\right). \tag{78}$$

Furthermore, we show in Appendix A that, upon replacing the given $(\mathbf{x}_1, \mathbf{x}_2)$ pair with a different pair yielding the same statistics of $\sum_{i=1}^{n} j_1(x_{1i}, x_{2i}, Y_i)$ if necessary (cf. Lemma 9), we have (cf. Lemma 11)

$$\tilde{T}_{1,n} := \sum_{i=1}^{n}\mathbb{E}\left[\Big|\frac{1}{\sqrt{n}}\Big(j_1(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[j_1(x_{1i}, x_{2i}, Y_i)]\Big)\Big|^3\right] = O\left(\frac{1 - \hat{\rho}_n}{\sqrt{n}}\right). \tag{79}$$

Thus, an application of the (univariate) Berry-Esseen theorem [30, Sec. XVI.5] yields

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n}\Big(j_1(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[j_1(x_{1i}, x_{2i}, Y_i)]\Big) \leq R_{1,n} - I_1(\hat{\rho}_n) - \gamma - \frac{\xi_1}{n}\right)$$
$$\geq \Phi\left(\frac{\sqrt{n}\big(R_{1,n} - I_1(\hat{\rho}_n) - \gamma - \frac{\xi_1}{n}\big)}{\sqrt{\tilde{V}_{1,n}}}\right) - \frac{6\,\tilde{T}_{1,n}}{\tilde{V}_{1,n}^{3/2}}. \tag{80}$$

We see from (78)–(79) that the remainder term above scales as $O\big([n(1 - \hat{\rho}_n)]^{-1/2}\big)$. Since it is assumed that $1 - \hat{\rho}_n = \omega\big(\frac{1}{n}\big)$, this remainder term vanishes. Thus, combining (80) with (74) and the fact that $\gamma = O\big(\frac{\log n}{n}\big)$, and inverting the relationship between rate and error probability, we obtain

$$R_{1,n} \leq I_1(\hat{\rho}_n) + \sqrt{\frac{\tilde{V}_{1,n}}{n}}\Phi^{-1}(\varepsilon) + o\left(\frac{1}{\sqrt{n}}\right). \tag{81}$$

Furthermore, $\hat{\rho}_n \to 1$ implies that $\tilde{V}_{1,n} \to 0$, and hence we have for any $\varepsilon \in (0,1)$ that

$$R_{1,n} \le I_1(\hat{\rho}_n) + o\left(\frac{1}{\sqrt{n}}\right), \tag{82}$$

thus yielding (72) and hence (70).

It remains to show that (82) also holds when $1 - \hat{\rho}_n = O\big(\frac{1}{n}\big)$. In this case, we can combine (60) and (76) to conclude that the variance of $\sum_{i=1}^n j_1(x_{1i}, x_{2i}, Y_i)$ is $O(1)$, and we thus have from Chebyshev's inequality that

$$\Pr\left(\frac{1}{n}\sum_{i=1}^n \Big(j_1(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[j_1(x_{1i}, x_{2i}, Y_i)]\Big) \le \frac{c}{\sqrt{n}}\right) \to 1 \tag{83}$$

for all $c > 0$. Substituting (83) into (74) and taking $c \to 0$ yields (82), as desired.

7) Completion of the Proof: Combining (69) and (70), we conclude that for any sequence of codes with error probability not exceeding $\varepsilon \in (0,1)$, we have for some sequence $\hat{\rho}_n \in [-1,1]$ that

$$\begin{bmatrix} R_{1,n} \\ R_{1,n} + R_{2,n} \end{bmatrix} \in \mathbf{I}(\hat{\rho}_n) + \frac{\Psi^{-1}\big(\mathbf{V}(\hat{\rho}_n), \varepsilon\big)}{\sqrt{n}} + \overline{g}(\hat{\rho}_n, \varepsilon, n)\mathbf{1}, \tag{84}$$

where $\overline{g}(\rho, \varepsilon, n)$ satisfies the conditions in the theorem statement. Specifically, the first condition follows from (69) (with $\bar{g}(\rho,\varepsilon,n) := h(\rho,\varepsilon)\frac{\log n}{n}$), and the second from (70) (with $\bar{g}(\rho,\varepsilon,n) := o\big(\frac{1}{\sqrt{n}}\big)$). This concludes the proof of the global converse.

B. Direct Part

We now prove the inner bound in (34). At a high level, we modify the key ideas in the analysis of the cost-constrained ensemble by Scarlett *et al.* [31] so that they are applicable to superposition codes. This approach can be applied to input-constrained MACs with DMS in significantly greater generality than the Gaussian case.

1) Random-Coding Ensemble: For user $j = 1, 2$, we introduce K *auxiliary cost functions* $\{a_k(x_1, x_2)\}_{k=1}^K$, which are assumed to be arbitrary for now. The ensemble will be defined in such a way that, with probability one, each codeword pair falls into the set

$$\mathcal{D}_n := \left\{ \big(\mathbf{x}_1, \mathbf{x}_2\big) : \|\mathbf{x}_1\|_2^2 \le nS_1, \|\mathbf{x}_2\|_2^2 \le nS_2, \left| \frac{1}{n}\sum_{i=1}^n a_k(x_{1i}, x_{2i}) - \mathbb{E}[a_k(X_1, X_2)] \right| \le \frac{\delta}{n},\ \forall\, k \in [K] \right\}, \tag{85}$$

where δ is a positive constant, and (X_1, X_2) are jointly Gaussian with a covariance matrix of the form given in (20). Roughly speaking, the set $\mathcal{D}_n$ contains codewords satisfying the power constraints such that the empirical expectations of the K auxiliary cost functions are δ/n-close to the true expectations.

Before defining the ensemble, we present the following straightforward variation of [31, Prop. 1]. We make use of the fact that $\|\mathbf{x}\|_2^2 = \sum_{i=1}^n x_i^2$, i.e. the power constraints are additive in the same way as the auxiliary costs.

Proposition 7. *Fix $\rho \in [0,1]$, and let $(\mathbf{X}_1', \mathbf{X}_2')$ be jointly distributed according to the n-fold product distribution of $P_{X_1,X_2} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\rho))$ (see (20)), i.e. $(\mathbf{X}_1', \mathbf{X}_2') \sim \prod_{i=1}^n P_{X_1,X_2}(x_{1i}', x_{2i}')$. If $\mathbb{E}[a_k(X_1, X_2)^2] < \infty$ for all $k \in [K]$, then there exists a choice of $\delta > 0$ such that*

$$\Pr\Big((\mathbf{X}_1', \mathbf{X}_2') \in \mathcal{D}_n\Big) \ge \psi(n), \tag{86}$$

where $\psi(n) = \Omega(n^{-(K+2)/2})$.

Proof: In the case that the power constraints are absent from $\mathcal{D}_n$, this is a special case of the statement of [31, Prop. 1], which was proved using the following steps: (i) Find a subset of $K' \le K$ linearly independent auxiliary cost functions, linear combinations of which can be used to construct the remaining $K - K'$ functions; (ii) Apply a local limit theorem (e.g. [32]) to bound the probability that all K constraints in $\mathcal{D}_n$ are satisfied. The second step relies on the constraints being two-sided, i.e. allowing for deviations on both sides of the mean in $\mathcal{D}_n$. This is false in the presence of the power constraints. However, for $\rho \in [0,1)$ we have $\det(\mathbf{\Sigma}(\rho)) > 0$, thus ensuring that the functions x_1^2 and x_2^2 can be included in the set of $K' \le K + 2$ linearly independent functions, and ensuring the validity of the second step above. For the remaining case $\rho = 1$, the codewords $\mathbf{X}_1'$ and $\mathbf{X}_2'$ are scalar multiples of each other, and thus the power constraints for the two users are equivalent (i.e. one implies the other). We can thus remove one of the power constraints from $\mathcal{D}_n$ without changing the statement of the proposition, and the remaining constraint can be included in the set of $K' \leq K+1$ linearly independent functions. This concludes the proof. ■

We now define the random-coding ensemble. We use superposition coding, in which the codewords are generated according to

$$\left\{\Big(\mathbf{X}_2(m_2), \{\mathbf{X}_1(m_1,m_2)\}_{m_1=1}^{M_{1,n}}\Big)\right\}_{m_2=1}^{M_{2,n}} \sim \prod_{m_2=1}^{M_{2,n}} \bigg(P_{\mathbf{X}_2}(\mathbf{x}_2(m_2)) \prod_{m_1=1}^{M_{1,n}} P_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1(m_1,m_2)|\mathbf{x}_2(m_2))\bigg) \tag{87}$$

for some codeword distributions $P_{\mathbf{X}_2}$ and $P_{\mathbf{X}_1|\mathbf{X}_2}$. These distributions will be chosen such that all codewords fall into the set $\mathcal{D}_n$ with probability one. The auxiliary costs will be chosen to satisfy the assumptions of Proposition 7, and we assume that δ is chosen such that (86) holds, in accordance with the proposition statement.

Let the i.i.d. codewords $(\mathbf{X}_1', \mathbf{X}_2')$ be defined as in Proposition 7. Defining the set

$$\mathcal{D}_{2,n} := \left\{\mathbf{x}_2 \,:\, \Pr\big((\mathbf{X}_1', \mathbf{x}_2) \in \mathcal{D}_n \,\big|\, \mathbf{X}_2' = \mathbf{x}_2\big) \geq \frac{1}{2}\psi(n)\right\}, \tag{88}$$

the codeword distributions are given by

$$P_{\mathbf{X}_2}(\mathbf{x}_2) = \frac{1}{\mu_{2,n}} \prod_{i=1}^{n} P_{X_2}(x_{2i}) \mathbf{1}\big\{\mathbf{x}_2 \in \mathcal{D}_{2,n}\big\}, \tag{89}$$

$$P_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2) = \frac{1}{\mu_{1,n}(\mathbf{x}_2)} \prod_{i=1}^{n} P_{X_1|X_2}(x_{1i}|x_{2i}) \mathbf{1}\big\{(\mathbf{x}_1,\mathbf{x}_2) \in \mathcal{D}_n\big\}, \tag{90}$$

where $\mu_{2,n}$ and $\mu_{1,n}(\mathbf{x}_2)$ are normalizing constants. We see that $\mathbf{X}_2 \in \mathcal{D}_{2,n}$ with probability one, and the definition of $\mathcal{D}_{2,n}$ in (88) implies that $\mu_{1,n}(\mathbf{x}_2) \geq \frac{1}{2}\psi(n) = \Omega(n^{-(K+2)/2})$ for all $\mathbf{x}_2 \in \mathcal{D}_{2,n}$. It will prove useful to show that we similarly have $\mu_{2,n} = \Omega(n^{-(K+2)/2})$. To see this, we write (86) as

$$\mu_{2,n} \Pr\big((\mathbf{X}_1', \mathbf{X}_2') \in \mathcal{D}_n \,\big|\, \mathbf{X}_2' \in \mathcal{D}_{2,n}\big) + (1-\mu_{2,n}) \Pr\big((\mathbf{X}_1', \mathbf{X}_2') \in \mathcal{D}_n \,\big|\, \mathbf{X}_2' \notin \mathcal{D}_{2,n}\big) \geq \psi(n). \tag{91}$$

Upper bounding the first probability by one and the second according to (88), we obtain

$$\mu_{2,n} + \frac{1}{2}\psi(n) \geq \psi(n), \tag{92}$$

where we have also used $1-\mu_{2,n} \leq 1$. It follows that $\mu_{2,n} \geq \frac{1}{2}\psi(n) = \Omega(n^{-(K+2)/2})$, as desired. Finally, upper bounding the indicator functions in (89)–(90) by one, we obtain

$$P_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{x}_1,\mathbf{x}_2) \leq \frac{4}{\psi(n)^2} \prod_{i=1}^{n} P_{X_1,X_2}(x_{1i}, x_{2i}). \tag{93}$$

2) A Feinstein-type Achievability Bound for the MAC with DMS: We now state a non-asymptotic achievability based on an analogous bound for the MAC [25, Lem. 3]. This bound can be considered as a dual of Proposition 4. Define

$$P_{\mathbf{X}_1|\mathbf{X}_2}W^n(\mathbf{y}|\mathbf{x}_2) := \int_{\mathbb{R}^n} P_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2) W^n(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2)\,\mathrm{d}\mathbf{x}_1, \tag{94}$$

$$P_{\mathbf{X}_1,\mathbf{X}_2}W^n(\mathbf{y}) := \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} P_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{x}_1,\mathbf{x}_2) W^n(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2)\,\mathrm{d}\mathbf{x}_1\,\mathrm{d}\mathbf{x}_2 \tag{95}$$

to be output distributions induced by an input distribution $P_{\mathbf{X}_1,\mathbf{X}_2}$ and the channel W^n. We have the following proposition, whose proof is omitted since it is uses standard arguments (e.g. see [14, Thm. 4]).

Proposition 8. *Fix a blocklength $n \geq 1$, random vectors $(\mathbf{X}_1, \mathbf{X}_2)$ with joint distribution $P_{\mathbf{X}_1,\mathbf{X}_2}$ such that $\|\mathbf{X}_1\|_2^2 \leq nS_1$ and $\|\mathbf{X}_2\|_2^2 \leq nS_2$ almost surely, auxiliary output distributions $Q_{\mathbf{Y}|\mathbf{X}_2}$ and $Q_{\mathbf{Y}}$, and a constant $\gamma > 0$. Then there exists an $(n, M_{1,n}, M_{2,n}, S_1, S_2, \varepsilon_n)$-code for which*

$$\varepsilon_n \leq \Pr(\mathcal{F}_n \cup \mathcal{G}_n) + \Lambda_1 \exp(-n\gamma) + \Lambda_{12}\exp(-n\gamma), \tag{96}$$

where

$$\Lambda_1 := \sup_{\mathbf{x}_2,\mathbf{y}} \frac{\mathrm{d}(P_{\mathbf{X}_1|\mathbf{X}_2}W^n)(\mathbf{y}|\mathbf{x}_2)}{\mathrm{d}Q_{\mathbf{Y}|\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_2)}, \quad \Lambda_{12} := \sup_{\mathbf{y}} \frac{\mathrm{d}(P_{\mathbf{X}_1,\mathbf{X}_2}W^n)(\mathbf{y})}{\mathrm{d}Q_{\mathbf{Y}}(\mathbf{y})}, \tag{97}$$

and

$$\mathcal{F}_n := \left\{\frac{1}{n}\log\frac{W^n(\mathbf{Y}|\mathbf{X}_1,\mathbf{X}_2)}{Q_{\mathbf{Y}|\mathbf{X}_2}(\mathbf{Y}|\mathbf{X}_2)} \le \frac{1}{n}\log M_{1,n} + \gamma\right\} \tag{98}$$

$$\mathcal{G}_n := \left\{\frac{1}{n}\log\frac{W^n(\mathbf{Y}|\mathbf{X}_1,\mathbf{X}_2)}{Q_{\mathbf{Y}}(\mathbf{Y})} \le \frac{1}{n}\log\big(M_{1,n}M_{2,n}\big) + \gamma\right\} \tag{99}$$

with $\mathbf{Y} \,|\, \{\mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2\} \sim W^n(\cdot|\mathbf{x}_1,\mathbf{x}_2)$.

The main difference between (96) and traditional Feinstein-type threshold decoding bounds (e.g. [25, Lem. 3], [33, Lem. 1]) is that we have the freedom to choose arbitrary output distributions $Q_{\mathbf{Y}|\mathbf{X}_2}$ and $Q_{\mathbf{Y}}$; this comes at the cost of introducing the multiplicative factors Λ_1 and Λ_{12} which depend on the maximum value of the Radon-Nikodym derivatives in (97). These multiplicative factors result from a standard change of measure argument.

3) Analysis of the Random-Coding Error Probability for fixed $\rho \in [0,1)$: We now use Proposition 8 with the codeword input distribution $P_{\mathbf{X}_1,\mathbf{X}_2}$ in (89)-(90). By construction, the probability of either codeword violating the power constraint is zero. We choose the output distributions $Q_{\mathbf{Y}|\mathbf{X}_2} := (P_{X_1|X_2}W)^n$ and $Q_{\mathbf{Y}} := (P_{X_1,X_2}W)^n$ to be of the convenient product form. As such, by (88), (90) and (93), we can choose $\Lambda_1 = 2\psi(n)^{-1}$ and $\Lambda_{12} = 4\psi(n)^{-2}$ in (97). Hence,

$$\varepsilon_n \le 1 - \Pr\left(\frac{1}{n}\sum_{i=1}^n \mathbf{j}(X_{1i},X_{2i},Y_i) > \mathbf{R}_n + \gamma\mathbf{1}\right) + \frac{2\exp(-n\gamma)}{\psi(n)} + \frac{4\exp(-n\gamma)}{\psi(n)^2}, \tag{100}$$

where the information density vector $\mathbf{j}(x_1,x_2,y)$ is defined with respect to $P_{X_1|X_2}W(y|x_2)$ and $P_{X_1,X_2}W(y)$, which coincide with $Q_{Y|X_2}$ and Q_Y in (25)–(26). Choosing

$$\gamma := \max\left\{-\frac{1}{n}\log\left(\frac{\psi(n)}{\sqrt{n}}\right), -\frac{1}{n}\log\left(\frac{\psi(n)^2}{\sqrt{n}}\right)\right\}, \tag{101}$$

we notice that the sum of the third and fourth terms in (100) is upper bounded by $\frac{6}{\sqrt{n}}$. The fact that $\psi(n) = \Omega(n^{-(K+2)/2})$ (see Proposition 7) implies that $\gamma = O\big(\frac{\log n}{n}\big)$, and we obtain

$$\varepsilon_n \le \max_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{D}_n} 1 - \Pr\left(\frac{1}{n}\sum_{i=1}^n \mathbf{j}(x_{1i},x_{2i},Y_i) > \mathbf{R}_n + \gamma\mathbf{1}\right) + \frac{6}{\sqrt{n}}. \tag{102}$$

We now make use of the key idea proposed in [31], namely, choosing the auxiliary costs in terms of the moments of $\mathbf{j}(x_1,x_2,Y)$ in order to ensure that the Berry-Esseen theorem can be applied to (102). We define

$$\mathbf{j}(x_1,x_2) := \mathbb{E}[\mathbf{j}(x_1,x_2,Y)], \tag{103}$$

$$\mathbf{v}(x_1,x_2) := \mathsf{Cov}[\mathbf{j}(x_1,x_2,Y)], \tag{104}$$

$$t(x_1,x_2) := \mathbb{E}\Big[\big\|\mathbf{j}(x_1,x_2,Y) - \mathbb{E}[\mathbf{j}(x_1,x_2,Y)]\big\|_2^3\Big], \tag{105}$$

where the expectations and covariance are taken with respect to $W(\cdot|x_1,x_2)$. We set $K=6$ and let the auxiliary costs equal the entries of the vectors and matrix in (103)–(105) (2 entries for $\mathbf{j}(x_1,x_2)$, 3 unique entries for the symmetric matrix $\mathbf{v}(x_1,x_2)$ and 1 entry for the scalar $t(x_1,x_2)$).[2] The assumptions of Proposition 7 are satisfied,

[2]An alternative approach would be to set $K=3$ and choose the system costs $a_1(x_1,x_2) = x_1^2$, $a_2(x_1,x_2) = x_2^2$ and $a_3(x_1,x_2) = x_1x_2$. Under these choices, all codeword pairs have roughly the same powers and empirical correlations, thus allowing us to bound the moments associated with $\sum_{i=1}^n \mathbf{j}(x_{1i},x_{2i},Y_i)$ similarly to Section VI-A. Furthermore, the uniform bound on the third absolute moment given in Lemma 10 in Appendix A can be used in place of the auxiliary cost corresponding to $t(x_1,x_2)$. On the other hand, the approach of choosing the auxiliary costs according to (103)–(105) remains applicable in significantly greater generality beyond the Gaussian setting. See [31] for more details.

since all moments of a Gaussian random variable are finite. We have from (28)–(29), (85) and (103)–(104) that

$$\left\| \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{j}(x_1,x_2,Y_i)\right] - \mathbf{I}(\rho)\right\|_{\infty} \le \frac{\delta}{n}, \tag{106}$$

$$\left\| \mathsf{Cov}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{j}(x_1,x_2,Y_i)\right] - \mathbf{V}(\rho)\right\|_{\infty} \le \frac{\delta}{n} \tag{107}$$

for all $(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{D}_n$, where the expectations and covariance are taken with respect to $W^n(\cdot|\mathbf{x}_1,\mathbf{x}_2)$. Similarly, defining $T(\rho):=\mathbb{E}[t(X_1,X_2)]$, we have from (85) and (105) that

$$\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\Big\|\mathbf{j}(x_{1i},x_{2i},Y_i)-\mathbb{E}[\mathbf{j}(x_{1i},x_{2i},Y_i)]\Big\|_2^3\right] - T(\rho)\right| \le \frac{\delta}{n}. \tag{108}$$

Thus, by applying the multivariate Berry-Esseen theorem [28], [29] (see Appendix B) to (102) and performing Taylor expansions similarly to Section VI-A5, we obtain

$$\varepsilon_n \le 1 - \Psi\Big(\sqrt{n}\big(I_1(\rho)-R_{1,n}\big), \sqrt{n}\big(I_{12}(\rho)-(R_{1,n}+R_{2,n})\big); \mathbf{V}(\rho)\Big) + \frac{\zeta(\rho,\delta)\log n}{\sqrt{n}}, \tag{109}$$

where $\zeta(\rho,\delta)$ is a function depending only on ρ and δ. By inverting the relationship between the rates and the error probability similarly to Section VI-A5, we obtain the desired result for any given $\rho\in[0,1)$, i.e. the first part of the theorem.

4) Analysis of the Random-Coding Error Probability for $\rho_n\to 1$: In order to prove the second part of the theorem, we use the cost-constrained ensemble with ρ varying with n, namely $\rho_n\to 1$. Similarly to (72), it suffices to show the achievability of $(R_{1,n},R_{2,n})$ satisfying

$$\begin{bmatrix} R_{1,n} \\ R_{1,n}+R_{2,n}\end{bmatrix} \ge \mathbf{I}(\rho_n) + \sqrt{\frac{V_{12}(1)}{n}}\begin{bmatrix} 0 \\ \Phi^{-1}(\varepsilon)\end{bmatrix} + o\left(\frac{1}{\sqrt{n}}\right)\mathbf{1}, \tag{110}$$

rather than the equivalent form given by (84). See the outer bound in Lemma 6.

We set $K=7$ and choose the first five auxiliary costs to be the same as those above, while letting a_6 and a_7 equal the two entries of

$$\mathbf{t}(x_1,x_2) := \mathbb{E}\Big[\big|\mathbf{j}(x_1,x_2,Y)-\mathbb{E}[\mathbf{j}(x_1,x_2,Y)]\big|^3\Big], \tag{111}$$

where the absolute value is applied element-wise. The reasons for the different choices of auxiliary costs here (compared to the case of fixed $\rho\in[0,1)$) are to obtain a sharper bound on the third absolute moment corresponding to $\sum_{i=1}^{n} j_1(x_{1i},x_{2i},Y_i)$, and since we will use two applications of the scalar Berry-Esseen theorem instead of one application of the vector version.

Defining

$$\mathbf{T}(\rho) = \begin{bmatrix} T_1(\rho) \\ T_{12}(\rho)\end{bmatrix} := \mathbb{E}[\mathbf{t}(X_1,X_2)], \tag{112}$$

we have similarly to (106)–(108) that

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\Big|\mathbf{j}(x_{1i},x_{2i},Y_i)-\mathbb{E}[\mathbf{j}(x_{1i},x_{2i},Y_i)]\Big|^3\right] - \mathbf{T}(\rho)\right\|_{\infty} \le \frac{\delta}{n}. \tag{113}$$

We have $I_1(1)=V_1(1)=T_1(1)=0$, and the behaviors of $I_1(\rho_n)$ and $V_1(\rho_n)$ as $\rho_n\to 1$ are characterized by (75)–(76). Furthermore, Lemma 12 in Appendix A states that

$$T_1(\rho_n) = O(1-\rho_n) \tag{114}$$

analogously to (75)–(76) (except with $O(\cdot)$ in place of $\Theta(\cdot)$).

Since Proposition 7 holds for all $\rho\in[0,1]$, it also holds for ρ varying within this range, and (102) remains true with $\gamma = O\big(\frac{\log n}{n}\big)$. Applying the union bound to (102), we obtain

$$\varepsilon_n \le \Pr\left(\frac{1}{n}\sum_{i=1}^{n} j_1(x_{1i},x_{2i},Y_i) \le R_{1,n}+\gamma\right) + \Pr\left(\frac{1}{n}\sum_{i=1}^{n} j_1(x_{1i},x_{2i},Y_i) \le R_{1,n}+R_{2,n}+\gamma\right) + \frac{6}{\sqrt{n}}. \tag{115}$$

for some $(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{D}_n$ The subsequent arguments are similar to Section VI-A6, so we only provide an outline. We treat the cases $1 - \rho_n = \omega\big(\frac{1}{n}\big)$ and $1 - \rho_n = O\big(\frac{1}{n}\big)$ separately. For the former, we fix a small $c > 0$ and choose

$$R_{1,n} = I_1(\rho_n) - \frac{c}{\sqrt{n}} - \gamma. \tag{116}$$

Using the bounds in (106)–(107) and (113), and applying the (scalar) Berry-Esseen theorem, we have similarly to (80) that

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n} j_1(x_{1i}, x_{2i}, Y_i) \le R_{1,n} + \gamma\right) \le \Phi\left(\frac{-\frac{c}{\sqrt{n}} + O\big(\frac{\log n}{n}\big)}{\sqrt{\frac{1}{n}\big(V_1(\rho_n) + O\big(\frac{1}{n}\big)\big)}}\right) + o(1), \tag{117}$$

where $V_1(\rho_n) = \omega\big(\frac{1}{n}\big)$, and the remainder term is $o(1)$ due to (76) and (114). Since $\rho_n \to 1$, we see that $V_1(\rho_n) \to 0$, and hence the right-hand side of (117) vanishes for any $c > 0$. Thus, applying the Berry-Esseen theorem to (115) and inverting the relationship between the rates and error probability, we obtain the following bound on the sum rate:

$$R_{1,n} + R_{2,n} \ge I_{12}(\rho_n) + \sqrt{\frac{V_{12}(1)}{n}}\Phi^{-1}(\varepsilon) + o\left(\frac{1}{\sqrt{n}}\right). \tag{118}$$

Combining (116) and (118), we obtain (110) upon taking $c \to 0$.

Using (106)–(107) and applying Chebyshev's inequality similarly to (83), we see that the left-hand side of (117) also vanishes for any $c > 0$ in the case that $1 - \rho_n = O\big(\frac{1}{n}\big)$. It follows that (110) remains true for this case, thus completing the proof of the second part of Theorem 2.

VII. Proof of Theorem 3: Local Second-Order Result

A. Converse Part

We now present the proof of the converse part of Theorem 3.

1) Proof for case (i) ($\rho = 0$): To prove the converse part for case (i), it suffices to consider the most optimistic case, namely $M_{2,n} = 1$ (i.e. no information is sent by the uninformed user). From the single-user dispersion result given in [4], [8] (cf. (4)), the number of messages for user 1 must satisfy

$$\log M_{1,n} \le nI_1(0) + \sqrt{nV_1(0)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}), \tag{119}$$

thus proving the converse part of (41).

2) Establishing The Convergence of ρ_n to ρ: Fix a correlation coefficient $\rho \in (0, 1]$, and consider any sequence of $(n, M_{1,n}, M_{2,n}, S_1, S_2, \varepsilon_n)$-codes for the Gaussian MAC with DMS satisfying (15). Let us consider the associated rates $\{(R_{1,n}, R_{2,n})\}_{n\ge1}$, where $R_{j,n} = \frac{1}{n}\log M_{j,n}$ for $j = 1, 2$. As required by Definition 4, we suppose that these codes satisfy

$$\liminf_{n\to\infty} R_{j,n} \ge R_j^*, \tag{120}$$

$$\liminf_{n\to\infty} \sqrt{n}\big(R_{j,n} - R_j^*\big) \ge L_j, \quad j = 1, 2, \tag{121}$$

$$\limsup_{n\to\infty} \varepsilon_n \le \varepsilon \tag{122}$$

for some (R_1^*, R_2^*) on the boundary parametrized by ρ, i.e. $R_1^* = I_1(\rho)$ and $R_1^* + R_2^* = I_{12}(\rho)$. The first-order optimality condition in (120) is not explicitly required by Definition 4, but it can easily be seen that (121), which is required by Definition 4, implies (120). Letting $\mathbf{R}_n := [R_{1,n}, R_{1,n} + R_{2,n}]^T$, we have from the global converse bound in (34) that there exists a sequence $\{\rho_n\}_{n\ge1} \subset [-1, 1]$ such that

$$\mathbf{R}_n \in \mathbf{I}(\rho_n) + \frac{\Psi^{-1}(\mathbf{V}(\rho_n), \varepsilon)}{\sqrt{n}} + \overline{g}(\rho_n, \varepsilon, n)\mathbf{1}. \tag{123}$$

Although $\overline{g}(\rho_n, \varepsilon, n)$ depends on ρ_n, we know from Theorem 2 that it is $o\big(\frac{1}{\sqrt{n}}\big)$ for both $\rho_n \to \pm1$ and ρ_n bounded away from ±1. It follows that

$$\mathbf{R}_n \in \mathbf{I}(\rho_n) + \frac{\Psi^{-1}(\mathbf{V}(\rho_n), \varepsilon)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\mathbf{1} \tag{124}$$

for all sequences $\{\rho_n\}_{n\geq 1}$. We claim that this result implies that ρ_n 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 $I_1(\rho_n) \leq I_1(\rho) - \delta$ or $I_{12}(\rho_n) \leq I_{12}(\rho) - \delta$. It follows from (124) that $R_{1,n} \leq I_1(\rho_n) + \frac{\delta}{2}$ and $R_{1,n} + R_{2,n} \leq I_{12}(\rho_n) + \frac{\delta}{2}$ for n large enough. As such, we deduce that $R_{1,n} \leq I_1(\rho) - \frac{\delta}{2}$ or $R_{1,n} + R_{2,n} \leq I_{12}(\rho) - \frac{\delta}{2}$ for all sufficiently large n. This, in turn, contradicts the first-order optimality conditions in (120).

3) Establishing The Convergence Rate of ρ_n to ρ: Because each entry of $\mathbf{I}(\rho)$ is twice continuously differentiable, a Taylor expansion yields

$$\mathbf{I}(\rho_n) = \mathbf{I}(\rho) + \mathbf{D}(\rho)(\rho_n - \rho) + O\big((\rho_n - \rho)^2\big)\mathbf{1}, \tag{125}$$

where $\mathbf{D}(\rho)$ is the derivative of $\mathbf{I}$ defined in (37). In the same way, since each entry of $\mathbf{V}(\rho)$ is continuously differentiable in ρ, we have

$$\|\mathbf{V}(\rho_n) - \mathbf{V}(\rho)\|_\infty = O(\rho_n - \rho). \tag{126}$$

We claim that these expansions, along with (124), imply that

$$\mathbf{R}_n \in \mathbf{I}(\rho) + \mathbf{D}(\rho)(\rho_n - \rho) + \frac{\Psi^{-1}(\mathbf{V}(\rho), \varepsilon)}{\sqrt{n}} + \left[o\left(\frac{1}{\sqrt{n}}\right) + O\big((\rho_n - \rho)^2\big) + O\left(\frac{(\rho_n - \rho)^{1/2}}{\sqrt{n}}\right) \right] \mathbf{1}. \tag{127}$$

The final term in the square parentheses results from the outer bound in Lemma 6 for the case $\rho = 1$. For $\rho \in (0,1)$ a standard Taylor expansion yields (127) with the last term replaced by $O\big(\frac{\rho_n - \rho}{\sqrt{n}}\big)$, and it follows that (127) holds for *all* $\rho \in (0,1]$.

We treat two cases separately: (I) $\rho_n - \rho = O(\frac{1}{\sqrt{n}})$, and (II) $\rho_n - \rho = \omega(\frac{1}{\sqrt{n}})$. We first show that Case (II) is not of interest in our study of local second-order coding rates (Definition 4). If Case (II) holds, intuitively in (127), $\frac{1}{\sqrt{n}}\Psi^{-1}(\mathbf{V}(\rho), \varepsilon)$ is dominated by $\mathbf{D}(\rho)(\rho_n - \rho)$ and hence, the second-order term scales as $\omega(\frac{1}{\sqrt{n}})$ instead of the desired $\Theta(\frac{1}{\sqrt{n}})$. To be more precise, because

$$\Psi^{-1}(\mathbf{V}(\rho), \varepsilon) \subset \begin{bmatrix} \sqrt{V_1(\rho)}\Phi^{-1}(\varepsilon) \\ \sqrt{V_{12}(\rho)}\Phi^{-1}(\varepsilon) \end{bmatrix}^-, \tag{128}$$

the bound in (127) implies that $\mathbf{R}_n$ must satisfy

$$\mathbf{R}_n \in \mathbf{I}(\rho) + \mathbf{D}(\rho)(\rho_n - \rho) + \frac{1}{\sqrt{n}} \begin{bmatrix} \sqrt{V_1(\rho)}\Phi^{-1}(\varepsilon) \\ \sqrt{V_{12}(\rho)}\Phi^{-1}(\varepsilon) \end{bmatrix}^- + o(\rho_n - \rho)\mathbf{1}. \tag{129}$$

In other words, (127) and $\rho_n - \rho = \omega(\frac{1}{\sqrt{n}})$ imply that $\mathbf{R}_n$ must satisfy

$$\mathbf{R}_n \leq \mathbf{I}(\rho) + \mathbf{D}(\rho)(\rho_n - \rho) + o(\rho_n - \rho)\mathbf{1}. \tag{130}$$

Since the first entry of $\mathbf{D}(\rho)$ is negative and the second entry is positive, (130) 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 $L_1 = -\infty$. This is due the fact that we consider second-order terms scaling as $\Theta\big(\frac{1}{\sqrt{n}}\big)$. Thus, only Case (I) is of interest, i.e. $\rho_n - \rho = O\big(\frac{1}{\sqrt{n}}\big)$.

4) Completing the Proof for Case (ii) ($\rho \in (0,1)$): Assuming now that $\rho_n - \rho = O(\frac{1}{\sqrt{n}})$, it follows that $\tau_n := \sqrt{n}(\rho_n - \rho)$ is a bounded sequence. By the Bolzano-Weierstrass theorem [20, Thm. 3.6(b)], $\{\tau_n\}_{n\geq 1}$ contains a convergent subsequence $\{\tau_{n_k}\}_{k\geq 1}$. Fix any convergent subsequence $\{\tau_{n_k}\}_{k\geq 1}$ and let the limit of this subsequence be $\beta \in \mathbb{R}$, i.e. $\beta := \lim_{k\to\infty} \tau_{n_k}$. Then, for the blocklengths indexed by n_k, we know from (127) that

$$\sqrt{n_k}\big(\mathbf{R}_{n_k} - \mathbf{I}(\rho)\big) \in \beta\, \mathbf{D}(\rho) + \Psi^{-1}(\mathbf{V}(\rho), \varepsilon) + o(1)\, \mathbf{1}, \tag{131}$$

where the $o(1)$ term combines the $o\big(\frac{1}{\sqrt{n}}\big)$ term in (127) and the deviation $(\tau_{n_k} - \beta)\max\{-D_1(\rho), D_{12}(\rho)\}$. By referring to the second-order optimality condition in (121), and applying the definition of the limit inferior in [20, Def. 3.16], we know that *every* convergent subsequence of $\{R_{j,n}\}_{n\geq 1}$ has a subsequential limit that satisfies $\lim_{k\to\infty}\sqrt{n_k}\big(R_{j,n_k}-R_j^*\big)\geq L_j$ for $j=1,2$. In other words, for all $\gamma>0$, there exist an integer K_1 such that

$$\sqrt{n_k}\big(R_{1,n_k}-I_1(\rho)\big)\geq L_1-\gamma \tag{132}$$

$$\sqrt{n_k}\big(R_{1,n_k}+R_{1,n_k}-I_{12}(\rho)\big)\geq L_1+L_2-2\gamma \tag{133}$$

for all $k\geq K_1$. Thus, we may lower bound each component in the vector on the left of (131) with $L_1-\gamma$ and $L_1+L_2-2\gamma$. There also exists an integer K_2 such that the $o(1)$ terms are upper bounded by γ for all $k\geq K_2$. We conclude that any pair of $(\varepsilon,R_1^*,R_2^*)$-achievable second-order coding rates (L_1,L_2) must satisfy

$$\begin{bmatrix} L_1-2\gamma \\ L_1+L_2-3\gamma\end{bmatrix}\in\bigcup_{\beta\in\mathbb{R}}\big\{\beta\,\mathbf{D}(\rho)+\Psi^{-1}(\mathbf{V}(\rho),\varepsilon)\big\}\,. \tag{134}$$

Finally, since $\gamma>0$ is arbitrary, we can take $\gamma\downarrow 0$, thus completing the converse proof for case (ii).

5) Completing the Proof for Case (iii) ($\rho=1$): The case $\rho=1$ is handled in essentially the same way as $\rho\in(0,1)$, so we only state the differences. Since β represents the difference between ρ_n and ρ, and since $\rho_n\leq 1$, we should only consider the case that $\beta\leq 0$. Furthermore, for $\rho=1$ the set $\Psi^{-1}(\mathbf{V}(\rho),\varepsilon)$ can be written in a simpler form; see Lemma 6. Using this form, we readily obtain (43).

B. Direct Part

We obtain the local result from the global result using a similar (yet simpler) argument to the converse part in Section VII-A. For fixed $\rho\in[0,1]$ and $\beta\in\mathbb{R}$, let

$$\rho_n:=\rho+\frac{\beta}{\sqrt{n}}, \tag{135}$$

where we require $\beta\geq 0$ (resp. $\beta\leq 0$) when $\rho=0$ (resp. $\rho=1$). By Theorem 2, we can achieve $(R_{1,n},R_{2,n})$ satisfying

$$\mathbf{R}_n\in\mathbf{I}(\rho_n)+\frac{\Psi^{-1}(\mathbf{V}(\rho_n),\varepsilon)}{\sqrt{n}}+o\left(\frac{1}{\sqrt{n}}\right)\mathbf{1}. \tag{136}$$

Substituting (135) into (136) and performing Taylor expansions in an identical fashion to the converse part (cf. the argument from (125) to (127)), we obtain

$$\mathbf{R}_n\in\mathbf{I}(\rho)+\frac{\beta\,\mathbf{D}(\rho)}{\sqrt{n}}+\frac{\Psi^{-1}(\mathbf{V}(\rho),\varepsilon)}{\sqrt{n}}+o\left(\frac{1}{\sqrt{n}}\right)\mathbf{1}. \tag{137}$$

We immediately obtain the desired result for case (ii) where $\rho\in[0,1)$. We also obtain the desired result for case (iii) where $\rho=1$ using the alternative form of $\Psi^{-1}(\mathbf{V}(1),\varepsilon)$ (see Lemma 6), similarly to the converse proof.

For case (i), we substitute $\rho=0$ into (38) and (39) to obtain $\mathbf{D}(\rho)=[0\ D_{12}(\rho)]^T$ with $D_{12}(\rho)>0$. Since β can be arbitrarily large, it follows from (137) that L_2 can take any real value. Furthermore, the set $\Psi^{-1}(\mathbf{V}(0),\varepsilon)$ contains vectors with a first entry arbitrarily close to $\sqrt{V_1(0)}\Phi^{-1}(\varepsilon)$ (provided that the other entry is sufficiently negative), and we thus obtain (41).

Appendix A
Moments of the Information Density Vector

Let $\rho\in[-1,1]$ be given, and recall the definition of the information density vector in (27), and the choices of $Q_{Y|X_2}$ and Q_Y in (25)–(26). For a given pair of sequences $(\mathbf{x}_1,\mathbf{x}_2)$, form the random vector

$$\mathbf{A}_n:=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{j}(x_{1i},x_{2i},Y_i), \tag{A.1}$$

where $Y_i|\{X_1 = x_{1i}, X_2 = x_{2i}\} \sim W(\cdot|x_{1i}, x_{2i})$. Define the constants $\alpha := S_1(1-\rho^2)$, $\vartheta := S_1 + S_2 + 2\rho\sqrt{S_1 S_2}$ and $\kappa := \rho\sqrt{S_1/S_2}$. Then, it can be verified that

$$j_1(x_1, x_2, Y) = \frac{1}{2}\log(1+\alpha) - \frac{Z^2}{2} + \frac{(x_1 - \kappa x_2 + Z)^2}{2(1+\alpha)} = \frac{-\alpha Z^2 + 2(x_1 - \kappa x_2)Z}{2(1+\alpha)} + f_1(x_1, x_2), \tag{A.2}$$

$$j_{12}(x_1, x_2, Y) = \frac{1}{2}\log(1+\vartheta) - \frac{Z^2}{2} + \frac{(x_1 + x_2 + Z)^2}{2(1+\vartheta)} = \frac{-\vartheta Z^2 + 2(x_1 + x_2)Z}{2(1+\vartheta)} + f_{12}(x_1, x_2), \tag{A.3}$$

where $Z := Y - x_1 - x_2 \sim \mathcal{N}(0,1)$ and $f_1(x_1, x_2)$ and $f_{12}(x_1, x_2)$ are some deterministic functions that will not affect the covariance matrix. Taking the expectation, we obtain

$$\mathbb{E}\big[j_1(x_1, x_2, Y)\big] = \frac{1}{2}\log(1+\alpha) - \frac{1}{2} + \frac{1 + (x_1 - \kappa x_2)^2}{2(1+\alpha)} = \frac{1}{2}\log(1+\alpha) + \frac{(x_1 - \kappa x_2)^2 - \alpha}{2(1+\alpha)}, \tag{A.4}$$

$$\mathbb{E}\big[j_{12}(x_1, x_2, Y)\big] = \frac{1}{2}\log(1+\vartheta) - \frac{1}{2} + \frac{1 + (x_1 + x_2)^2}{2(1+\vartheta)} = \frac{1}{2}\log(1+\vartheta) + \frac{(x_1 + x_2)^2 - \vartheta}{2(1+\vartheta)}. \tag{A.5}$$

Setting $x_1 \leftarrow x_{1i}$, $x_2 \leftarrow x_{2i}$ and $Y \leftarrow Y_i$ in (A.4) and (A.5) and summing over all i, we conclude that the mean vector of $\mathbf{A}_n$ is

$$\mathbb{E}\big[\mathbf{A}_n\big] = \sqrt{n}\left[\mathsf{C}(\alpha) + \frac{\|\mathbf{x}_1 - \kappa\mathbf{x}_2\|_2^2 - n\alpha}{2n(1+\alpha)} \qquad \mathsf{C}(\vartheta) + \frac{\|\mathbf{x}_1 + \mathbf{x}_2\|_2^2 - n\vartheta}{2n(1+\vartheta)}\right]^T. \tag{A.6}$$

From (A.2) and (A.3), we deduce that the variances are

$$\mathsf{Var}\big[j_1(x_1, x_2, Y)\big] = \mathsf{Var}\left[\frac{-\alpha Z^2 + 2(x_1 - \kappa x_2)Z}{2(1+\alpha)}\right] = \frac{\alpha^2 + 2(x_1 - \kappa x_2)^2}{2(1+\alpha)^2}, \tag{A.7}$$

$$\mathsf{Var}\big[j_{12}(x_1, x_2, Y)\big] = \mathsf{Var}\left[\frac{-\vartheta Z^2 + 2(x_1 + x_2)Z}{2(1+\vartheta)}\right] = \frac{\vartheta^2 + 2(x_1 + x_2)^2}{2(1+\vartheta)^2}, \tag{A.8}$$

where we have used $\mathsf{Var}[Z^2] = 2$ and $\mathsf{Cov}[Z^2, Z] = \mathbb{E}Z^3 - (\mathbb{E}Z)(\mathbb{E}Z^2) = 0$. The covariance is

$$\mathsf{Cov}\big[j_1(x_1, x_2, Y), j_{12}(x_1, x_2, Y)\big] = \mathsf{Cov}\left[\frac{-\alpha Z^2 + 2(x_1 - \kappa x_2)Z}{2(1+\alpha)}, \frac{-\vartheta Z^2 + 2(x_1 + x_2)Z}{2(1+\vartheta)}\right] \tag{A.9}$$

$$= \frac{1}{4(1+\alpha)(1+\vartheta)}\Big\{\mathbb{E}\big[(-\alpha Z^2 + 2(x_1 - \kappa x_2)Z)(-\vartheta Z^2 + 2(x_1 + x_2)Z)\big] - \mathbb{E}\big[-\alpha Z^2 + 2(x_1 - \kappa x_2)Z\big]\mathbb{E}\big[-\vartheta Z^2 + 2(x_1 + x_2)Z\big]\Big\} \tag{A.10}$$

$$= \frac{3\alpha\vartheta + 4(x_1 - \kappa x_2)(x_1 + x_2) - \alpha\vartheta}{4(1+\alpha)(1+\vartheta)} = \frac{\alpha\vartheta + 2(x_1^2 + (1-\kappa)x_1 x_2 - \kappa x_2^2)}{2(1+\alpha)(1+\vartheta)} \tag{A.11}$$

Setting $x_1 \leftarrow x_{1i}$, $x_2 \leftarrow x_{2i}$ and $Y \leftarrow Y_i$ in (A.7), (A.8) and (A.11) and summing over all i, we conclude that covariance matrix of $\mathbf{A}_n$ is

$$\mathsf{Cov}\big[\mathbf{A}_n\big] = \begin{bmatrix} \frac{n\alpha^2 + 2\|\mathbf{x}_1 - \kappa\mathbf{x}_2\|_2^2}{2n(1+\alpha)^2} & \frac{n\alpha\vartheta + 2(\|\mathbf{x}_1\|_2^2 + (1-\kappa)\langle\mathbf{x}_1, \mathbf{x}_2\rangle - \kappa\|\mathbf{x}_2\|_2^2)}{2n(1+\alpha)(1+\vartheta)} \\ \frac{n\alpha\vartheta + 2(\|\mathbf{x}_1\|_2^2 + (1-\kappa)\langle\mathbf{x}_1, \mathbf{x}_2\rangle - \kappa\|\mathbf{x}_2\|_2^2)}{2n(1+\alpha)(1+\vartheta)} & \frac{n\vartheta^2 + 2\|\mathbf{x}_1 + \mathbf{x}_2\|_2^2}{2n(1+\vartheta)^2} \end{bmatrix}. \tag{A.12}$$

In the remainder of the section, we analyze several third absolute moments associated with $\mathbf{A}_n$ appearing in the univariate [30, Sec. XVI.5] and multivariate Berry-Esseen theorems [28], [29] (see Appendix B). The following lemma will be used to replace any given $(\mathbf{x}_1, \mathbf{x}_2)$ pair by an "equivalent" pair (in the sense that the statistics of $\mathbf{A}_n$ are unchanged) for which the corresponding third moments have the desired behavior. This is analogous to Polyanskiy *et al.* [8], where for the AWGN channel, one can use a spherical symmetry argument to replace any given sequence $\mathbf{x}$ such that $\|\mathbf{x}\|_2^2 = nS$ with a fixed sequence $(\sqrt{S}, \cdots, \sqrt{S})$.

Lemma 9. *The joint distribution of $\mathbf{A}_n$ depends on $(\mathbf{x}_1, \mathbf{x}_2)$ only through the powers $\|\mathbf{x}_1\|_2^2$, $\|\mathbf{x}_2\|_2^2$ and the inner product $\langle\mathbf{x}_1, \mathbf{x}_2\rangle$.*

Proof: This follows by substituting (A.2)–(A.3) into (A.1) and using the symmetry of the additive noise sequence $\mathbf{Z} = (Z_1, \dots, Z_n)$. For example, from (A.2), the first entry of $\mathbf{A}_n$ can be written as

$$\frac{1}{\sqrt{n}}\left(\frac{n}{2}\log(1+\alpha) - \frac{1}{2}\|\mathbf{Z}\|_2^2 + \frac{1}{2(1+\alpha)}\|\mathbf{x}_1 - \kappa\mathbf{x}_2 + \mathbf{Z}\|_2^2\right), \tag{A.13}$$

and the desired result follows by writing

$$\|\mathbf{x}_1 - \kappa\mathbf{x}_2 + \mathbf{Z}\|^2 = \|\mathbf{x}_1\|^2 + \kappa^2\|\mathbf{x}_2\|^2 + \|\mathbf{Z}\|^2 - 2\kappa\langle\mathbf{x}_1, \mathbf{x}_2\rangle + 2\langle\mathbf{x}_1, \mathbf{Z}\rangle - 2\kappa\langle\mathbf{x}_2, \mathbf{Z}\rangle. \tag{A.14}$$

Since $\mathbf{Z}$ is i.i.d. Gaussian, the last two terms depend on $(\mathbf{x}_1, \mathbf{x}_2)$ only through $\|\mathbf{x}_1\|_2^2$ and $\|\mathbf{x}_2\|_2^2$. ■

We now provide lemmas showing that, upon replacing a given pair $(\mathbf{x}_1, \mathbf{x}_2)$ with an equivalent pair using Lemma 9 if necessary, the corresponding third moments have the desired behavior. It will prove useful to work with the empirical correlation coefficient

$$\rho_{\mathrm{emp}}(\mathbf{x}_1, \mathbf{x}_2) := \frac{\langle\mathbf{x}_1, \mathbf{x}_2\rangle}{\|\mathbf{x}_1\|_2\|\mathbf{x}_2\|_2}. \tag{A.15}$$

It is easily seen that Lemma 9 remains true when the inner product $\langle\mathbf{x}_1, \mathbf{x}_2\rangle$ is replaced by this normalized quantity.

Lemma 10. *For any fixed $\tilde{\rho} \in [-1, 1]$, there exists a sequence of pairs $(\mathbf{x}_1, \mathbf{x}_2)$ (indexed by increasing lengths n) such that $\|\mathbf{x}_1\|_2^2 = nS_1$, $\|\mathbf{x}_2\|_2^2 = nS_2$, $\rho_{\mathrm{emp}}(\mathbf{x}_1, \mathbf{x}_2) = \tilde{\rho}$, and*

$$\tilde{T}_n := \sum_{i=1}^n \mathbb{E}\left[\left\|\frac{1}{\sqrt{n}}\big(\mathbf{j}(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[\mathbf{j}(x_{1i}, x_{2i}, Y_i)]\big)\right\|_2^3\right] = O\left(\frac{1}{\sqrt{n}}\right), \tag{A.16}$$

where the $O\big(\frac{1}{\sqrt{n}}\big)$ term is uniform in $\tilde{\rho} \in [-1, 1]$.

Proof: By using the fact that $\|\mathbf{v}\|_2 \le \|\mathbf{v}\|_1$ and $(|a| + |b|)^3 \le 4|a|^3 + 4|b|^3$, the following bounds hold:

$$\begin{aligned}\tilde{T}_n &\le \sum_{i=1}^n \mathbb{E}\left[\left\|\frac{1}{\sqrt{n}}\big(\mathbf{j}(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[\mathbf{j}(x_{1i}, x_{2i}, Y_i)]\big)\right\|_1^3\right] &\text{(A.17)}\\ &\le 4\sum_{i=1}^n \mathbb{E}\left[\left|\frac{1}{\sqrt{n}}\big(j_1(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[j_1(x_{1i}, x_{2i}, Y_i)]\big)\right|^3\right] \\ &\quad + 4\sum_{i=1}^n \mathbb{E}\left[\left|\frac{1}{\sqrt{n}}\big(j_{12}(x_{1i}, x_{2i}, Y_i) - \mathbb{E}[j_{12}(x_{1i}, x_{2i}, Y_i)]\big)\right|^3\right]. &\text{(A.18)}\end{aligned}$$

We now specify $(\mathbf{x}_1, \mathbf{x}_2)$ whose powers and correlation match those given in the lemma statement. Assuming for the time being that $|\tilde{\rho}| \le \frac{n-1}{n}$, we choose

$$\mathbf{x}_1 = \big(\sqrt{S_1}, \cdots, \sqrt{S_1}\big) \tag{A.19}$$

$$\mathbf{x}_2 = \big(\sqrt{S_2(1+\eta)}, \sqrt{S_2}, \cdots, \sqrt{S_2}, -\sqrt{S_2(1-\eta)}, -\sqrt{S_2}, \cdots, -\sqrt{S_2}\big) \tag{A.20}$$

where $\eta \in (-1, 1)$, and $\mathbf{x}_2$ contains $k \ge 1$ negative entries and $n - k \ge 1$ positive entries. It is easily seen that $\|\mathbf{x}_1\|_2^2 = nS_1$ and $\|\mathbf{x}_2\|_2^2 = nS_2$, as desired. Furthermore, we can choose k and η to obtain the desired correlation since

$$\langle\mathbf{x}_1, \mathbf{x}_2\rangle = \big(n - 2(k-1) + \sqrt{1+\eta} - \sqrt{1-\eta}\big)\sqrt{S_1S_2}, \tag{A.21}$$

and since the range of the function $f(\eta) := \sqrt{1+\eta} - \sqrt{1-\eta}$ for $\eta \in (-1, 1)$ is given by $(-\sqrt{2}, \sqrt{2})$.

Using (A.2)–(A.3), it can easily be verified that the third absolute moment of each entry of $\mathbf{j}(x_1, x_2, Y)$ (i.e. $\mathbb{E}\big|j_1(x_1, x_2, Y) - \mathbb{E}[j_1(x_1, x_2, Y)]\big|^3$ and $\mathbb{E}\big|j_{12}(x_1, x_2, Y) - \mathbb{E}[j_{12}(x_1, x_2, Y)]\big|^3$) is bounded above by some constant for any $(x_1, x_2) = (\sqrt{S_1}, \pm\sqrt{cS_2})$ ($c \in (0, 2)$). We thus obtain (A.16) using (A.18). The proof is concluded by noting that a similar argument applies for the case $\tilde{\rho} \in (\frac{n-1}{n}, 1]$ by replacing (A.20) by

$$\mathbf{x}_2 = \big(\sqrt{S_2(1+\eta)}, \sqrt{S_2(1-\eta)}, \sqrt{S_2}, \cdots, \sqrt{S_2}\big), \tag{A.22}$$

and similarly (with negative entries) when $\tilde{\rho} \in [-1, \frac{n-1}{n})$. ■

Lemma 11. *For any pair of sequences $\{\rho_n\}_{n\geq 1}\subset[-1,1]$ and $\{\tilde{\rho}_n\}_{n\geq 1}\subset[-1,1]$ such that $\rho_n,\tilde{\rho}_n\to 1$ with $1-\tilde{\rho}_n=\Theta(1-\rho_n)$, there exists a sequence of vectors $(\mathbf{x}_1,\mathbf{x}_2)$ (indexed by increasing lengths n) with the properties $\|\mathbf{x}_1\|_2^2=nS_1$, $\|\mathbf{x}_2\|_2^2=nS_2$, $\rho_{\mathrm{emp}}(\mathbf{x}_1,\mathbf{x}_2)=\tilde{\rho}_n$, and they satisfy*

$$\tilde{T}_{1,n}:=\sum_{i=1}^{n}\mathbb{E}\left[\Big|\frac{1}{\sqrt{n}}\big(j_1(x_{1i},x_{2i},Y_i)-\mathbb{E}[j_1(x_{1i},x_{2i},Y_i)]\big)\Big|^3\right]=O\left(\frac{1-\tilde{\rho}_n}{\sqrt{n}}\right), \tag{A.23}$$

where the information density is computed with respect to ρ_n, i.e. j_1 is defined in (27) *with input distribution $P_{X_1,X_2}=\mathcal{N}(\mathbf{0},\mathbf{\Sigma}(\rho_n))$.*

Proof: We make use of the notation and definitions in the proof of Lemma 10, using a subscript to denote dependence on n, e.g. k_n. We again focus on the case that $|\tilde{\rho}_n|\leq\frac{n-1}{n}$ and we choose $(\mathbf{x}_1,\mathbf{x}_2)$ according to (A.19)–(A.20). The remaining cases are handled similarly (e.g. using (A.22)) and are thus omitted.

Inverting (A.21), we have $k_n=\frac{n}{2}(1-\tilde{\rho}_n)+O(1)$, and thus the entries corresponding to $x_2\neq\sqrt{S_2}$ in (A.20) contribute an additive $O\big(\frac{1-\tilde{\rho}_n}{\sqrt{n}}\big)$ term to the summation in (A.23). For the remaining entries, where $(x_1,x_2)=(\sqrt{S_1},\sqrt{S_2})$, we note from (A.2) that the third absolute moment of $j(x_1,x_2,Y)$ is given by

$$\mathbb{E}\Big[\big|j(x_1,x_2,Y)-\mathbb{E}\big[j(x_1,x_2,Y)\big]\big|^3\Big]=\mathbb{E}\left[\left|\frac{-\alpha(Z^2-1)+2(x_1-\kappa x_2)Z}{2(1+\alpha)}\right|^3\right] \tag{A.24}$$

$$\leq 2\left(|\alpha|^3\mathbb{E}\big[|Z^2-1|^3\big]+8\big(\sqrt{S_1}-\kappa\sqrt{S_2}\big)^3\mathbb{E}\big[|Z|^3\big]\right) \tag{A.25}$$

$$=\Theta\left(|\alpha|^3+\big(\sqrt{S_1}-\kappa\sqrt{S_2}\big)^3\right) \tag{A.26}$$

$$=\Theta(1-\tilde{\rho}_n), \tag{A.27}$$

where (A.25) follows because $\alpha\geq 0$ and $|a+b|^3\leq 4(|a|^3+|b|^3)$, and (A.27) follows by substituting $\alpha=S_1(1-\rho^2)$ and $\kappa=\rho\sqrt{S_1/S_2}$, and applying first-order Taylor expansions in $1-\tilde{\rho}_n$ (recall that $1-\rho_n=\Theta(1-\tilde{\rho}_n)$ by assumption). From (A.27), we conclude that the entries where $(x_1,x_2)=(\sqrt{S_1},\sqrt{S_2})$ in (A.19)–(A.20) contribute an additive $O\big(\frac{1-\tilde{\rho}_n}{\sqrt{n}}\big)$ term to the summation in (A.23), thus yielding the desired result. ■

Recall the definition of $T_1(\rho)$ in (112) which we restate here for the reader's convenience:

$$T_1(\rho):=\mathbb{E}\left[\mathbb{E}\Big[\big|j_1(X_1,X_2,Y)-\mathbb{E}[j_1(X_1,X_2,Y)]\big|^3\,\big|\,X_1,X_2\Big]\right]. \tag{A.28}$$

The distribution of (X_1,X_2) above is $\mathcal{N}(\mathbf{0},\mathbf{\Sigma}(\rho))$ and the information density j_1 is also defined with respect to ρ.

Lemma 12. *For any sequence $\{\rho_n\}_{n\geq 1}$ satisfying $\rho_n\to 1$, we have*

$$T_1(\rho_n)=O(1-\rho_n). \tag{A.29}$$

Proof: We upper bound $t_1(x_1,x_2)$ using the Cauchy-Schwarz and arithmetic-geometric mean inequalities as follows:

$$t_1(x_1,x_2)=\mathbb{E}\big[\big|j_1(x_1,x_2,Y)-\mathbb{E}[j_1(x_1,x_2,Y)]\big|^3\big] \tag{A.30}$$

$$\leq\sqrt{\mathbb{E}\Big[\big(j_1(x_1,x_2,Y)-\mathbb{E}[j_1(x_1,x_2,Y)]\big)^2\Big]\mathbb{E}\Big[\big(j_1(x_1,x_2,Y)-\mathbb{E}[j_1(x_1,x_2,Y)]\big)^4\Big]} \tag{A.31}$$

$$\leq\frac{1}{2}\left(\mathbb{E}\Big[\big(j_1(x_1,x_2,Y)-\mathbb{E}[j_1(x_1,x_2,Y)]\big)^2\Big]+\mathbb{E}\Big[\big(j_1(x_1,x_2,Y)-\mathbb{E}[j_1(x_1,x_2,Y)]\big)^4\Big]\right). \tag{A.32}$$

Now, by (A.7), the variance of $j_1(x_1,x_2,Y)$ is

$$m_2(x_1,x_2):=\mathbb{E}\Big[\big(j_1(x_1,x_2,Y)-\mathbb{E}[j_1(x_1,x_2,Y)]\big)^2\Big]=\frac{\alpha^2+2(x_1-\kappa x_2)^2}{2(1+\alpha)^2}. \tag{A.33}$$

By a similar calculation, we deduce that the centralized fourth moment of $j_1(x_1,x_2,Y)$ is

$$m_4(x_1,x_2):=\mathbb{E}\Big[\big(j_1(x_1,x_2,Y)-\mathbb{E}[j_1(x_1,x_2,Y)]\big)^4\Big]=\frac{15\alpha^4+60\alpha^2(x_1-\kappa x_2)^2+12(x_1-\kappa x_2)^4}{4(1+\alpha)^4}. \tag{A.34}$$

Taking the expectation of (A.33) with respect to $(X_1, X_2) \sim \mathcal{N}(0, \mathbf{\Sigma}(\rho))$ yields

$$\mathbb{E}[m_2(X_1, X_2)] = \frac{\alpha(2+\alpha)}{2(1+\alpha)^2} \tag{A.35}$$

where we used the fact that $X_1 - \kappa X_2 \sim \mathcal{N}(0, \alpha)$. Similarly, taking the expectation of (A.34) yields

$$\mathbb{E}[m_4(X_1, X_2)] = \frac{3\alpha^2(5\alpha^2 + 20\alpha + 12)}{4(1+\alpha)^4} \tag{A.36}$$

where we used the fact that $\mathbb{E}\big[(X_1 - \kappa X_2)^4\big] = 3\alpha^2$. Now, observe that since $\alpha = \alpha(\rho) = S_1(1-\rho^2)$ is continuously differentiable in ρ, so are $\mathbb{E}[m_2(X_1, X_2)]$ and $\mathbb{E}[m_4(X_1, X_2)]$. Moreover $\mathbb{E}[m_2(X_1, X_2)]$ and $\mathbb{E}[m_4(X_1, X_2)]$ are equal to zero when $\rho = 1$ (and hence $\alpha = 0$). Consequently, by (A.32) we know that $T_1(\rho)$ is upper bounded by a continuously differentiable function in ρ and it evaluates to zero when $\rho = 1$. We conclude, by first-order Taylor expansions in ρ in the vicinity of $\rho = 1$, that (A.29) holds. ■

Appendix B
The Multivariate Berry-Esseen Theorem

In this section, we state a version of the multivariate Berry-Esseen theorem [28], [29] that is suited for our needs in this paper. The following is a restatement of Corollary 38 in [34].

Theorem 13. *Let $\mathbf{U}_1, \ldots, \mathbf{U}_n$ be independent, zero-mean random vectors in $\mathbb{R}^d$. Let $\mathbf{L}_n := \frac{1}{\sqrt{n}}(\mathbf{U}_1 + \cdots + \mathbf{U}_n)$, Assume $\mathbf{V} := \mathsf{Cov}(\mathbf{L}_n)$ is positive definite with minimum eigenvalue $\lambda_{\min}(\mathbf{V}) > 0$. Let $t := \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\|\mathbf{U}_i\|_2^3\big]$ and let $\mathbf{Z}$ be a zero-mean Gaussian random vector with covariance $\mathbf{V}$. Then, for all $n \in \mathbb{N}$,*

$$\sup_{\mathscr{C} \in \mathfrak{C}_d} \big| \Pr(\mathbf{L}_n \in \mathscr{C}) - \Pr(\mathbf{Z} \in \mathscr{C}) \big| \le \frac{k_d\, t}{\lambda_{\min}(\mathbf{V})^{3/2}\sqrt{n}}, \tag{B.1}$$

where $\mathfrak{C}_d$ is the family of all convex, Borel measurable subsets of $\mathbb{R}^d$, and k_d is a function only of the dimension d (e.g., $k_2 = 265$).

Appendix C
Proof of Lemma 5

Fix $(z_1, z_2) \in \Psi^{-1}\big(\mathbf{V}, \varepsilon + \lambda_n\big)$ and define $\mathbf{Z} = (Z_1, Z_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$. Since $\Psi^{-1}\big(\mathbf{V}, \varepsilon\big)$ is monotonic in the sense that $\Psi^{-1}\big(\mathbf{V}, \varepsilon\big) \subset \Psi^{-1}\big(\mathbf{V}, \varepsilon'\big)$ for $\varepsilon \le \varepsilon'$, it suffices to verify that (z_1, z_2) belongs to the set on the right-hand side of (67) for those (z_1, z_2) on the boundary of $\Psi^{-1}\big(\mathbf{V}, \varepsilon + \lambda_n\big)$. That is (cf. (31)),

$$\Pr\big(Z_1 \le -z_1, Z_2 \le -z_2\big) = 1 - (\varepsilon + \lambda_n). \tag{C.1}$$

Define $\nu_n := \inf\big\{\nu > 0 : (-z_1 - \nu, -z_2 - \nu) \in \Psi^{-1}(\mathbf{V}, \varepsilon)\big\}$. We need to show that $\nu_n = o(1)$ is bounded above by some linear function of λ_n. By using (C.1) and the definition of ν_n, we see that

$$\lambda_n = \Pr\big(Z_1 \in [-z_1 - \nu_n, -z_1] \cup Z_2 \in [-z_2 - \nu_n, -z_2]\big) \tag{C.2}$$

$$\ge \max_{j=1,2}\left\{\Phi\left(\frac{-z_j}{\sqrt{V_{jj}}}\right) - \Phi\left(\frac{-z_j - \nu_n}{\sqrt{V_{jj}}}\right)\right\}. \tag{C.3}$$

The assumption that $\mathbf{V}$ is a non-zero positive-semidefinite matrix ensures that at least one of $V_{jj}, j = 1, 2$ is non-zero. We have the lower bound

$$\Phi\left(\frac{-z}{\sqrt{V}}\right) - \Phi\left(\frac{-z - \nu_n}{\sqrt{V}}\right) \ge \frac{\nu_n}{\sqrt{V}} \min\left\{\mathcal{N}(z; 0, V), \mathcal{N}(z + \nu_n; 0, V)\right\}. \tag{C.4}$$

Hence, for all n large enough, each of the terms in $\{\cdot\}$ in (C.3) is bounded below by $\nu_n f(z_j, V_{jj})$ for $j = 1, 2$ where $f(z, V) := \frac{1}{2\sqrt{V}}\mathcal{N}(z; 0, V)$ satisfies $\lim_{z \to \pm\infty} f(z, V) = 0$. Hence, $\nu_n \le \lambda_n \min_{j=1,2}\{f(z_j, V_{jj})^{-1}\}$. For every fixed $\varepsilon \in (0, 1)$, every $(z_1, z_2) \in \Psi^{-1}\big(\mathbf{V}, \varepsilon + \lambda_n\big)$ satisfies $\min\{|z_1|, |z_2|\} < \infty$, and hence $\min_{j=1,2}\{f(z_j, V_{jj})^{-1}\}$ is finite. This concludes the proof.

APPENDIX D
PROOF OF LEMMA 6

Recall that $\hat{\rho}_n \to 1$. We start by proving the inner bound on $\Psi^{-1}(\mathbf{V}(\hat{\rho}_n), \varepsilon)$. Let (w_1, w_2) be an arbitrary element of the left-hand-side of (71), i.e. $w_1 \le -b_n$ and $w_2 \le \sqrt{V_{12}(1)}\Phi^{-1}(\varepsilon + a_n) - b_n$. Define the random variables $(Z_{1,n}, Z_{2,n}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}(\hat{\rho}_n))$ and the sequence $b_n := (1-\hat{\rho}_n)^{1/4}$. Consider

$$\Pr\left(Z_{1,n} \le -w_1, Z_{2,n} \le -w_2\right) \ge \Pr\left(Z_{1,n} \le b_n, Z_{2,n} \le -\left(\sqrt{V_{12}(1)}\Phi^{-1}(\varepsilon + a_n) - b_n\right)\right) \tag{D.1}$$

$$\ge \Pr\left(Z_{2,n} \le -\left(\sqrt{V_{12}(1)}\Phi^{-1}(\varepsilon + a_n) - b_n\right)\right) - \Pr\left(Z_{1,n} > b_n\right) \tag{D.2}$$

$$= \Phi\left(\frac{-\left(\sqrt{V_{12}(1)}\Phi^{-1}(\varepsilon + a_n) - b_n\right)}{\sqrt{V_{12}(\hat{\rho}_n)}}\right) - \Phi\left(\frac{-b_n}{\sqrt{V_1(\hat{\rho}_n)}}\right). \tag{D.3}$$

From the choice of b_n and the fact that $\sqrt{V_1(\hat{\rho}_n)} = \Theta(\sqrt{1-\hat{\rho}_n})$ (cf. (76)), the argument of the second term scales as $-(1-\hat{\rho}_n)^{-1/4}$ which tends to $-\infty$. Hence, the second term vanishes. We may thus choose a vanishing sequence a_n so that the expression in (D.3) equals $1-\varepsilon$. Such a choice satisfies $a_n = \Theta(b_n) = \Theta((1-\hat{\rho}_n)^{1/4})$, in accordance with the lemma statement. From the definition in (31), we have proved that $(w_1, w_2) \in \Psi^{-1}(\mathbf{V}(\hat{\rho}_n), \varepsilon)$ for this choice of (a_n, b_n).

For the outer bound on $\Psi^{-1}(\mathbf{V}(\hat{\rho}_n), \varepsilon)$, let (u_1, u_2) be an arbitrary element of $\Psi^{-1}(\mathbf{V}(\hat{\rho}_n), \varepsilon)$. By definition,

$$\Pr(Z_{1,n} \le -u_1, Z_{2,n} \le -u_2) \ge 1 - \varepsilon, \tag{D.4}$$

where $(Z_{1,n}, Z_{2,n}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}(\hat{\rho}_n))$ as above. Thus,

$$1 - \varepsilon \le \Pr(Z_{2,n} \le -u_2) = \Phi\left(\frac{-u_2}{\sqrt{V_{12}(\hat{\rho}_n)}}\right). \tag{D.5}$$

This leads to

$$u_2 \le \sqrt{V_{12}(\hat{\rho}_n)}\Phi^{-1}(\varepsilon) = \sqrt{V_{12}(1)}\Phi^{-1}(\varepsilon) + c_n' \tag{D.6}$$

for some $c_n' = \Theta(1-\hat{\rho}_n)$, since $\rho \mapsto \sqrt{V_{12}(\rho)}$ is continuously differentiable and its derivative does not vanish at $\rho = 1$. Similarly, we have

$$u_1 \le \sqrt{V_1(\hat{\rho}_n)}\Phi^{-1}(\varepsilon) = c_n'' \tag{D.7}$$

for some $c_n'' = \Theta(\sqrt{1-\hat{\rho}_n})$, since $V_1(1) = 0$ and $\sqrt{V_1(\hat{\rho}_n)} = \Theta(\sqrt{1-\hat{\rho}_n})$ by (76). Letting $c_n := \max\{|c_n'|, |c_n''|\} = \Theta(\sqrt{1-\hat{\rho}_n})$, we deduce that (u_1, u_2) belongs to the rightmost set in (71). This completes the proof.

ACKNOWLEDGMENT

We are grateful to Ebrahim MolavianJazi for pointing us to a minor error in an earlier version of the paper.

REFERENCES

[1] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2012.
[2] T. Cover, “Broadcast channels,” *IEEE Trans. on Inf. Th.*, vol. 18, no. 1, pp. 2–14, 1972.
[3] M. Hayashi, “Second-order asymptotics in fixed-length source coding and intrinsic randomness,” *IEEE Trans. on Inf. Th.*, vol. 54, pp. 4619–37, Oct 2008.
[4] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” *IEEE Trans. on Inf. Th.*, vol. 55, pp. 4947–66, Nov 2009.
[5] 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]`.
[6] R. Nomura and T. S. Han, “Second-order resolvability, intrinsic randomness, and fixed-length source coding for mixed sources: Information spectrum approach,” *IEEE Trans. on Inf. Th.*, vol. 59, pp. 1–16, Jan 2013.
[7] V. Strassen, “Asymptotische Abschätzungen in Shannons Informationstheorie,” in *Trans. Third Prague Conf. Inf. Theory*, (Prague), pp. 689–723, 1962.
[8] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding in the finite blocklength regime,” *IEEE Trans. on Inf. Th.*, vol. 56, pp. 2307–59, May 2010.
[9] C. E. Shannon, “Probability of error for optimal codes in a Gaussian channel,” *Bell Systems Technical Journal*, vol. 38, pp. 611–656, 1959.

[10] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," in *Conference on Information Sciences and Systems (CISS)*, (Princeton, NJ), 2012. arXiv:1201.3901 [cs.IT].

[11] Y.-W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *Int. Symp. Inf. Th.*, (Boston, MA), 2012.

[12] E. MolavianJazi and J. N. Laneman, "Simpler achievable rate regions for multiaccess with finite blocklength," in *Int. Symp. Inf. Th.*, (Boston, MA), 2012.

[13] E. MolavianJazi and J. N. Laneman, "A random coding approach to Gaussian multiple access channels with finite blocklength," in *Allerton Conference*, (Monticello, IL), 2012.

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

[15] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *Allerton Conference*, (Monticello, IL), 2012.

[16] P. Moulin, "A new metaconverse and outer region for finite-blocklength MACs," in *Info. Th. and Applications (ITA) Workshop*, (San Diego, CA), 2013.

[17] 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].

[18] 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.

[19] S. I. Bross, A. Lapidoth, and M. Wigger, "Dirty-paper coding for the Gaussian multiaccess channel with conferencing," *IEEE Trans. on Inf. Th.*, vol. 58, no. 9, pp. 5640–5668, 2012.

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

[21] R. Ahlswede, "An elementary proof of the strong converse theorem for the multiple access channel," *J. of Combinatorics, Information & System Sciences*, pp. 216–230, 1982.

[22] G. Dueck, "Maximal error capacity regions are smaller than average error capacity regions for multi-user channels," *Probl. Control Inf. Theory*, vol. 7, pp. 11–19, 1978.

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

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

[25] T. S. Han, "An information-spectrum approach to capacity theorems for the general multiple-access channel," *IEEE Trans. on Inf. Th.*, vol. 44, pp. 2773–95, Jul 1998.

[26] S. Boucheron and M. R. Salamatian, "About priority encoding transmission," *IEEE Trans. on Inf. Th.*, vol. 46, no. 2, pp. 699–705, 2000.

[27] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Trans. on Inf. Th.*, vol. 49, pp. 1753–68, Jul 2003.

[28] F. Götze, "On the rate of convergence in the multivariate CLT," *The Annals of Probability*, vol. 19, no. 2, pp. 721–739, 1991.

[29] R. Bhattacharya and S. Holmes, "An exposition of Götze's estimation of the rate of convergence in the multivariate central limit theorem," tech. rep., Stanford University, 2010. arxiv:1003.4254 [math.ST].

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

[31] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Finite-length bounds, error exponents and approximations," arXiv:1303.6166 [cs.IT], Mar 2013.

[32] C. Stone, "On local and ratio limit theorems," in *Fifth Berkeley Symp. Math. Stat. Prob.*, pp. 217–224, 1965.

[33] J. N. Laneman, "On the distribution of mutual information," in *Information Theory and Applications Workshop*, 2006.

[34] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Non-asymptotic and second-order achievability bounds for coding with side-information," arXiv:1301.6467 [cs.IT], Apr 2013.