Second-Order Rate Region of Constant-Composition Codes for the Multiple-Access Channel

Jonathan Scarlett, Alfonso Martinez and Albert Guillén i Fàbregas

Abstract

This paper presents an achievable second-order rate region for the discrete memoryless multiple-access channel. The result is obtained using a random-coding ensemble in which each user's codebook contains codewords of a fixed composition. The improvement of the second-order rate region over existing ones is demonstrated both analytically and numerically. Finally, an achievable second-order rate region for the Gaussian multiple-access channel is derived via an increasingly fine quantization of the input.

I. Introduction

Shannon's channel capacity describes the largest possible rate of transmission with vanishing error probability in coded communication systems. Further characterizations of the system performance are given by error exponents [1, Ch. 9], moderate deviations results [2], and second-order coding rates [3]. The latter has regained significant attention in recent years [4], [5], and is well-understood for a variety of single-user channels. In particular, for discrete memoryless channels, the maximum number of codewords $M^*(n,\epsilon)$ of length n yielding an error probability not exceeding ϵ satisfies

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}\mathsf{Q}^{-1}(\epsilon) + o(\sqrt{n}), \tag{1}$$

where C is the channel capacity, $\mathsf{Q}^{-1}(\cdot)$ is the inverse of the Q-function, and V is known as the channel dispersion. From (1), we see that a higher dispersion V means that a larger backoff from capacity is needed to achieve a fixed $\epsilon < \frac{1}{2}$, at least in terms of second-order asymptotics.

J. Scarlett is with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: jmscarlett@gmail.com). A. Martinez is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain (e-mail: alfonso.martinez@ieee.org). A. Guillén i Fàbregas is with the Institució Catalana de Recerca i Estudis Avançats (ICREA), the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain, and also with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

In this paper, we study the second-order asymptotics for the discrete memoryless multiple-access channel (DM-MAC). Achievability results for this problem have previously been obtained using i.i.d. random coding with a random time-sharing sequence [6], [7] and a deterministic time-sharing sequence [8].

The main result of this paper is an achievable second-order rate region (see Definition 1) obtained using constant-composition random coding [1, Ch. 9]. We demonstrate an improvement over the achievability results of [6]–[8] even after the optimization of the input distributions. We can think of the improvement of constant-composition codes as being analogous to a similar gain in the random-coding error exponent for the MAC [9]. A key tool in our analysis is a Berry-Esseen theorem associated with a variant of Hoeffding's combinatorial central limit theorem (CLT) [10]; see Section IV-B for details.

A. Notation

The set of all probability distributions on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. Given a distribution $Q(x)$ and a conditional distribution $W(y|x)$, the joint distribution $Q(x)W(y|x)$ is denoted by $Q \times W$. We make use of the method of types [11, Ch. 2]. The set of all types of length n on an alphabet $\mathcal{A}$ is denoted by $\mathcal{P}_n(\mathcal{A})$. The set of all sequences of length n with a given type P_X is denoted by $T^n(P_X)$, and similarly for joint types. Given a sequence $\boldsymbol{x} \in T^n(P_X)$ and a conditional distribution $P_{Y|X}$, we define $T^n_{\boldsymbol{x}}(P_{Y|X})$ to be the set of sequences $\boldsymbol{y}$ such that $(\boldsymbol{x}, \boldsymbol{y}) \in T^n(P_X \times P_{Y|X})$.

Bold symbols are used for vectors and matrices (e.g. $\boldsymbol{x}$), and the corresponding i-th entry of a vector is denoted with a subscript (e.g. x_i). The vectors of all zeros and all ones are denoted by $\mathbf{0}$ and $\mathbf{1}$ respectively, and the $k \times k$ identity matrix is denoted by $\mathbb{I}_{k\times k}$. The symbols $\prec$, $\preceq$, etc. denote element-wise inequalities for vectors, and inequalities on the positive semidefinite cone for matrices (e.g. $\boldsymbol{V} \succ \mathbf{0}$ means $\boldsymbol{V}$ is positive definite). We denote the transpose of a vector or matrix by $(\cdot)^T$, the inverse of a matrix by $(\cdot)^{-1}$, the positive definite matrix square root by $(\cdot)^{\frac{1}{2}}$, and its inverse by $(\cdot)^{-\frac{1}{2}}$. The multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

We denote the cross-covariance of two random vectors by $\mathrm{Cov}[\boldsymbol{Z}_1, \boldsymbol{Z}_2] = \mathbb{E}\big[(\boldsymbol{Z}_1 - \mathbb{E}[\boldsymbol{Z}_1])(\boldsymbol{Z}_2 - \mathbb{E}[\boldsymbol{Z}_2])^T\big]$, and we write $\mathrm{Cov}[\boldsymbol{Z}]$ in place of $\mathrm{Cov}[\boldsymbol{Z}, \boldsymbol{Z}]$. The variance of a scalar random variable is denoted by $\mathrm{Var}[\cdot]$. Logarithms have base e, and all rates are in nats except in the examples, where bits are used. We denote the indicator function by $\mathbb{1}\{\cdot\}$.

For two sequences f_n and g_n, we write $f_n = O(g_n)$ if $|f_n| \le c|g_n|$ for some c and sufficiently large n, and $f_n = o(g_n)$ if $\lim_{n\to\infty} \frac{f_n}{g_n} = 0$. We write $f_n = \Theta(g_n)$ if $f_n = O(g_n)$ and $g_n = O(f_n)$. A vector or matrix is said to be $O(f_n)$ if all of its entries are $O(f_n)$ in the scalar sense.

B. System Setup

We consider a 2-user DM-MAC $W(y|x_1, x_2)$ with input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ and output alphabet $\mathcal{Y}$. The encoders and decoder operate as follows. Encoder $\nu = 1, 2$ takes as input a message m_ν equiprobable on the set $\{1, \dots, M_\nu\}$, and transmits the corresponding codeword $\boldsymbol{x}_\nu^{(m_\nu)}$ from the codebook $\mathcal{C}_\nu = \{\boldsymbol{x}_\nu^{(1)}, \dots, \boldsymbol{x}_\nu^{(M_\nu)}\}$. Upon receiving $\boldsymbol{y}$ at the output of the channel, the decoder forms an estimate $(\hat{m}_1, \hat{m}_2)$ of the messages. An error is said to have occurred if the estimate $(\hat{m}_1, \hat{m}_2)$ differs from (m_1, m_2).

A rate pair (R_1, R_2) is said to be (n, ϵ)-achievable if there exist codebooks with $M_1 \geq \exp(nR_1)$ and $M_2 \geq \exp(nR_2)$ codewords of length n for users 1 and 2 respectively such that the average error probability does not exceed ϵ. Given (R_1, R_2), we define the rate vector

$$\boldsymbol{R} \stackrel{\triangle}{=} \begin{bmatrix} R_1 \\ R_2 \\ R_1 + R_2 \end{bmatrix}. \tag{2}$$

The achievability part of the capacity result of Ahlswede [12] and Liao [13] states that for any $\epsilon \in (0,1)$, rates satisfying

$$\boldsymbol{R} \preceq \begin{bmatrix} I(X_1; Y|X_2, U) \\ I(X_2; Y|X_1, U) \\ I(X_1, X_2; Y|U) \end{bmatrix} + g_\epsilon(n)\mathbf{1} \tag{3}$$

are (n, ϵ)-achievable for some $g_\epsilon(n)$ which vanishes as $n \to \infty$, under any joint distribution of the form $(U, X_1, X_2, Y) \sim P_U \times P_{X_1|U} \times P_{X_2|U} \times W$. Equation (3) is said to describe a first-order achievable rate region. In this paper, we are concerned with second-order expansions, in which the $g_\epsilon(n)\mathbf{1}$ term is replaced by the sum of a second-order term and an asymptotic third-order term.

We consider constant-composition random coding, as considered by Liu and Hughes [9], among others. We fix a time-sharing alphabet $\mathcal{U}$, as well as the input distributions $Q_U(u)$, $Q_1(x_1|u)$ and $Q_2(x_2|u)$. We let $Q_{U,n}$, $Q_{1,n}$ and $Q_{2,n}$ denote (conditional) types which are closest to Q_U, Q_1 and Q_2 respectively in terms of L_∞ norm. We fix an arbitrary time-sharing sequence $\boldsymbol{u}$ with type $Q_{U,n}$ and generate the $M_\nu \stackrel{\triangle}{=} e^{nR_\nu}$ codewords of user $\nu = 1, 2$ conditionally independently according to the uniform distribution on $T^n_{\boldsymbol{u}}(Q_{\nu,n})$, i.e.

$$P_{\boldsymbol{X}_\nu|\boldsymbol{U}}(\boldsymbol{x}_\nu|\boldsymbol{u}) = \frac{1}{|T^n_{\boldsymbol{u}}(Q_{\nu,n})|} \mathbb{1}\Big\{\boldsymbol{x}_\nu \in T^n_{\boldsymbol{u}}(Q_{\nu,n})\Big\}. \tag{4}$$

Throughout the paper, we define the joint distribution

$$P_{UX_1X_2Y}(u, x_1, x_2, y) \stackrel{\triangle}{=} Q_U(u) Q_1(x_1|u) Q_2(x_2|u) W(y|x_1, x_2) \tag{5}$$

and denote the induced marginal distributions by $P_{Y|X_1U}$, $P_{Y|X_2U}$, etc. A key quantity in our analysis is the information density vector

$$\boldsymbol{i}(u, x_1, x_2, y) \stackrel{\triangle}{=} \begin{bmatrix} i_1(u, x_1, x_2, y) \\ i_2(u, x_1, x_2, y) \\ i_{12}(u, x_1, x_2, y) \end{bmatrix}, \tag{6}$$

where

$$i_1(u,x_1,x_2,y) \triangleq \log \frac{W(y|x_1,x_2)}{P_{Y|X_2U}(y|x_2,u)} \tag{7}$$

$$i_2(u,x_1,x_2,y) \triangleq \log \frac{W(y|x_1,x_2)}{P_{Y|X_1U}(y|x_1,u)} \tag{8}$$

$$i_{12}(u,x_1,x_2,y) \triangleq \log \frac{W(y|x_1,x_2)}{P_{Y|U}(y|u)}. \tag{9}$$

Averaging these quantities with respect to the distribution in (5) yields the mutual information quantities appearing in (3).

C. Existing Results

We define the set

$$\mathsf{Q}_{\mathrm{inv}}(\boldsymbol{V},\epsilon) \triangleq \Big\{ \boldsymbol{z} \in \mathbb{R}^3 \,:\, \mathbb{P}\big[\boldsymbol{Z} \preceq \boldsymbol{z}\big] \geq 1-\epsilon \Big\}, \tag{10}$$

where $\boldsymbol{Z} \sim N(\boldsymbol{0},\boldsymbol{V})$. Since the existing second-order rate regions (and the one given in this paper) are written in a similar form in terms of a matrix, a vector, and the set $\mathsf{Q}_{\mathrm{inv}}$, we define the following notion of achievability.

Definition 1. Let $\boldsymbol{I}$ be a 3×1 non-negative vector, and let $\boldsymbol{V}$ be a 3×3 positive semidefinite matrix. The pair $(\boldsymbol{I},\boldsymbol{V})$ is said to be second-order achievable if, for all $\epsilon \in (0,1)$, there exists a sequence $g(n) = o(\sqrt{n})$ such that all pairs (R_1,R_2) satisfying

$$n\boldsymbol{R} \in n\boldsymbol{I} - \sqrt{n}\mathsf{Q}_{\mathrm{inv}}(\boldsymbol{V},\epsilon) + g(n)\boldsymbol{1}, \tag{11}$$

are (n,ϵ)-achievable, where $\boldsymbol{R}$ is defined in (2).

Expansions of the form (11) are somewhat more difficult to interpret than the scalar counterpart in (1). Roughly speaking, given a vector $\boldsymbol{I}$ and two covariance matrices $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$, $\boldsymbol{V}_1 \prec \boldsymbol{V}_2$ implies that $\boldsymbol{V}_1$ yields faster convergence to the achievable rate region corresponding to $\boldsymbol{I}$ as n increases with $\epsilon < \frac{1}{2}$ fixed, at least in terms of second-order asymptotics.

The first study of the problem under consideration was by Tan and Kosut [6], who used i.i.d. random coding to prove that $(\boldsymbol{I},\boldsymbol{V})$ with $\boldsymbol{I} = \mathbb{E}[\boldsymbol{i}(U,X_1,X_2,Y)]$ and $\boldsymbol{V} = \mathrm{Cov}[\boldsymbol{i}(U,X_1,X_2,Y)]$ is second-order achievable for any choice of $\mathcal{U}$ and (Q_U,Q_1,Q_2). MolavianJazi and Laneman [7] obtained second-order asymptotic results by treating the three error events separately rather than jointly, and using just three variance terms instead of a full 3×3 covariance matrix. Huang and Moulin [8] showed that the covariance matrix can be improved to

$$\boldsymbol{V}^{\mathrm{iid}} = \mathbb{E}\Big[\mathrm{Cov}\Big[\boldsymbol{i}(U,X_1,X_2,Y) \,\big|\, U\Big]\Big] \tag{12}$$

by fixing a constant-composition time-sharing sequence $\boldsymbol{u}$, rather than generating one at random. This result improves on that of [6] due to the fact that conditioning reduces variance.

A simple improvement on the achievability result of [8] can be obtained by letting one user's codebook be constant-composition and the other i.i.d., yielding a covariance matrix of the form $\boldsymbol{V} = \mathbb{E}\big[\mathrm{Cov}[\boldsymbol{i}(U,X_1,X_2,Y) \,|\, U,X_1]\big]$

or $\boldsymbol{V} = \mathbb{E}\big[\mathrm{Cov}[\boldsymbol{i}(U, X_1, X_2, Y) \,|\, U, X_2]\big]$. The covariance matrix obtained in this paper improves further on each of these.

Our main result is closely related to a recent result by MolavianJazi and Laneman [14], who derived an achievable second-order rate region for the Gaussian MAC using random coding with a uniform distribution over the surface of a sphere. In Section V, we give an alternative proof of the main result of [14] via an increasingly fine quantization of the input.

For certain classes of channels, the present problem can be reduced to a single-user problem in order to obtain a matching converse to the above achievability results [15]. A more general converse containing variances of the form $\mathbb{E}[\mathrm{Var}[i_\nu(U, X_1, X_2, Y) \,|\, U, X_1, X_2]]$ ($\nu = 1, 2, 12$) has recently been reported by Moulin [16], [17].

An alternative to considering expansions of the form (11) is to consider the second-order asymptotics as a particular point on the boundary of the capacity region is approached from a certain angle. We do not pursue this approach in this paper; see [6], [15] for further discussion.

II. Main Result

The main result of this paper is the following theorem. Along with (6)–(9), we define the quantities

$$\boldsymbol{i}^{(1)}(u, x_1) \triangleq \mathbb{E}\big[\boldsymbol{i}(U, X_1, X_2, Y) \,\big|\, (U, X_1) = (u, x_1)\big] \tag{13}$$

$$\boldsymbol{i}^{(2)}(u, x_2) \triangleq \mathbb{E}\big[\boldsymbol{i}(U, X_1, X_2, Y) \,\big|\, (U, X_1) = (u, x_2)\big] \tag{14}$$

whose entries are given by

$$i_\nu^{(1)}(u, x_1) \triangleq \mathbb{E}\big[i_\nu(U, X_1, X_2, Y) \,\big|\, (U, X_1) = (u, x_1)\big] \tag{15}$$

$$i_\nu^{(2)}(u, x_2) \triangleq \mathbb{E}\big[i_\nu(U, X_1, X_2, Y) \,\big|\, (U, X_1) = (u, x_2)\big] \tag{16}$$

for $\nu = 1, 2, 12$.

Theorem 1. *Fix any finite time-sharing alphabet $\mathcal{U}$ and the input distributions (Q_U, Q_1, Q_2). The pair $(\boldsymbol{I}, \boldsymbol{V})$ is second-order achievable, where*

$$\boldsymbol{I} = \mathbb{E}\big[\boldsymbol{i}(U, X_1, X_2, Y)\big] \tag{17}$$

$$\boldsymbol{V} = \mathbb{E}\Big[\mathrm{Cov}\big[\boldsymbol{i}(U, X_1, X_2, Y) \,\big|\, U\big] - \mathrm{Cov}\big[\boldsymbol{i}^{(1)}(U, X_1) \,\big|\, U\big] - \mathrm{Cov}\big[\boldsymbol{i}^{(2)}(U, X_2) \,\big|\, U\big]\Big]. \tag{18}$$

Furthermore, the function $g(n)$ in (11) satisfies $g(n) = O(\log n)$ if the argument to the expectation in (18) has full rank for all $u \in \mathcal{U}$, and $g(n) = O(n^{\frac{1}{6}})$ more generally.

Proof: See Section IV. ∎

The covariance matrix $\boldsymbol{V}$ in (18) can be interpreted as follows. The term $\mathrm{Cov}[\boldsymbol{i}]$ represents the variations in (X_1, X_2, Y) in the i.i.d. case, and the terms $\mathrm{Cov}[\boldsymbol{i}^{(1)}]$ and $\mathrm{Cov}[\boldsymbol{i}^{(2)}]$ represent the reduced variations in X_1 and X_2 respectively, resulting from the codewords having a fixed composition. In particular, using constant-composition coding for user 1 and i.i.d. coding for user 2, we instead obtain the covariance matrix

$$\boldsymbol{V}^{\text{cc-iid}} = \mathbb{E}\Big[\mathrm{Cov}\Big[\boldsymbol{i}(U,X_1,X_2,Y)\,\Big|\,U\Big] - \mathrm{Cov}\Big[\boldsymbol{i}^{(1)}(U,X_1)\,\Big|\,U\Big]\Big] \tag{19}$$

$$= \mathbb{E}\Big[\mathrm{Cov}\Big[\boldsymbol{i}(U,X_1,X_2,Y)\,\Big|\,U,X_1\Big]\Big], \tag{20}$$

thus recovering the result stated in Section I-C. Since all covariance matrices are positive semidefinite, we clearly have $\boldsymbol{V} \preceq \boldsymbol{V}^{\text{cc-iid}} \preceq \boldsymbol{V}^{\text{iid}}$.

It is interesting to compare (18) with the conditional covariance matrix

$$\boldsymbol{V}^{\text{joint}} = \mathbb{E}\Big[\mathrm{Cov}\Big[\boldsymbol{i}(U,X_1,X_2,Y),\,\Big|\,U\Big] - \mathrm{Cov}\Big[\boldsymbol{i}^{(12)}(U,X_1,X_2)\,\Big|\,U\Big]\Big] \tag{21}$$

$$= \mathbb{E}\Big[\mathrm{Cov}\Big[\boldsymbol{i}(U,X_1,X_2,Y)\,\Big|\,U,X_1,X_2\Big]\Big], \tag{22}$$

where $\boldsymbol{i}^{(12)}(U,X_1,X_2) \triangleq \mathbb{E}\big[\boldsymbol{i}(U,X_1,X_2,Y)\,|\,U,X_1,X_2\big]$. Roughly speaking, this is the covariance matrix which we would obtain if the *joint* composition of (U,X_1,X_2) were fixed, which is impossible in general in the absence of cooperation. Based on this observation, we expect that $\boldsymbol{V}^{\text{joint}} \preceq \boldsymbol{V}$. To show that this is true, we use the matrix version of the law of total variance to write

$$\mathrm{Cov}\Big[\boldsymbol{i}^{(12)}(u,X_1,X_2)\Big] = \mathrm{Cov}\Big[\mathbb{E}\big[\boldsymbol{i}^{(12)}(u,X_1,X_2)\,\big|\,X_1\big]\Big] + \mathbb{E}\Big[\mathrm{Cov}\big[\boldsymbol{i}^{(12)}(u,X_1,X_2)\,\big|\,X_1\big]\Big] \tag{23}$$

$$= \mathrm{Cov}\big[\boldsymbol{i}^{(1)}(u,X_1)\big] + \mathbb{E}\Big[\mathrm{Cov}\big[\boldsymbol{i}^{(12)}(u,X_1,X_2)\,\big|\,X_1\big]\Big], \tag{24}$$

where each expression is implicitly conditioned on $U=u$. The second quantity in (24) can be weakened as follows:

$$\mathbb{E}\Big[\mathrm{Cov}\big[\boldsymbol{i}^{(12)}(u,X_1,X_2)\,\big|\,X_1\big]\Big]$$

$$= \sum_{x_1,x_2} Q_1(x_1|u)Q_2(x_2|u)\Big(\boldsymbol{i}^{(12)}(u,X_1,X_2) - \mathbb{E}\big[\boldsymbol{i}^{(12)}(u,x_1,X_2)\big]\Big)\Big(\boldsymbol{i}^{(12)}(u,X_1,X_2) - \mathbb{E}\big[\boldsymbol{i}^{(12)}(u,x_1,X_2)\big]\Big)^T \tag{25}$$

$$\succeq \sum_{x_1,x_2} Q_2(x_2|u)\Big(\boldsymbol{i}^{(2)}(u,X_2) - \mathbb{E}\big[\boldsymbol{i}^{(2)}(u,X_2)\big]\Big)\Big(\boldsymbol{i}^{(2)}(u,X_2) - \mathbb{E}\big[\boldsymbol{i}^{(2)}(u,X_2)\big]\Big)^T \tag{26}$$

$$= \mathrm{Cov}\big[\boldsymbol{i}^{(2)}(u,X_2)\big], \tag{27}$$

where (26) follows using the identity $\mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}^T] \succeq \mathbb{E}[\boldsymbol{Z}]\mathbb{E}[\boldsymbol{Z}]^T$. Combining (18), (21), (24) and (27), we obtain the desired result.

III. Example: The Collision Channel

In this section, we consider the channel with $\mathcal{X}_1 = \mathcal{X}_2 = \{0,1,2\}$, $\mathcal{Y} = \{(0,0),(0,1),(0,2),(1,0),(2,0),\mathsf{c}\}$ and

$$W(y|x_1,x_2) = \begin{cases} 1 & y = (x_1,x_2) \text{ and } \min\{x_1,x_2\} = 0 \\ 1 & y = \mathsf{c} \text{ and } \min\{x_1,x_2\} \neq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{28}$$

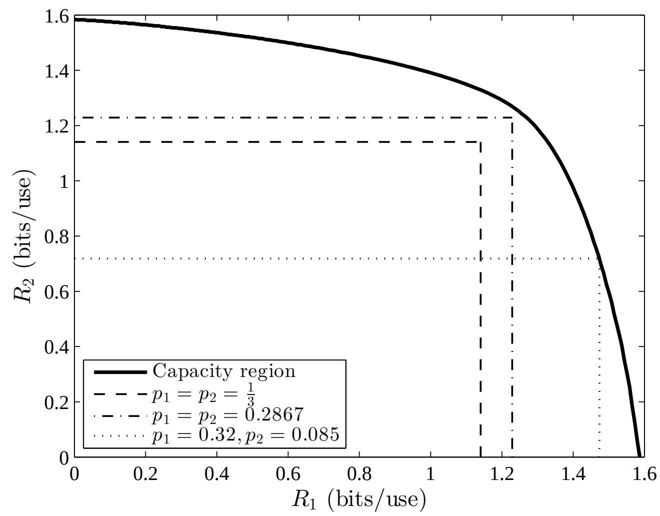

Figure 1. Capacity region of the collision channel.

In words, if either user transmits the zero symbol then the pair (x_1, x_2) is received noiselessly, whereas if both users transmit a non-zero symbol then the output is c, meaning "collision".

We recall the following observations by Gallager [18]: (i) The capacity region can be obtained without time sharing;[1] (ii) By symmetry, the points on the boundary of the capacity region are achieved using $\mathcal{U} = \emptyset$ and input distributions of the form $Q_1 = (1 - 2p_1, p_1, p_1)$ and $Q_2 = (1 - 2p_2, p_2, p_2)$; (iii) The achievable rate region corresponding to any such (Q_1, Q_2) pair is rectangular. To illustrate these observations, we plot the capacity region in Figure 1, along with three achievable rate regions corresponding to particular choices of p_1 and p_2.

We first compare the various random-coding schemes with fixed input distributions. Figure 2 plots the second-order regions with $p_1 = p_2 = 0.2$, $n = 50$ and $\epsilon = 0.01$, and with the third-order $o(\sqrt{n})$ terms ignored. It should be noted that these ignored terms can be significant at finite block lengths, and thus the resulting curves should only be viewed as approximations.

The improvement over [8] obtained by letting user 1's codewords be constant-composition is insignificant at small values of R_1 but significant at high values of R_1. A similar observation applies for user 2. The region resulting from Theorem 1, obtained using constant-composition codes for both users, is strictly larger than all of the others, and yields particularly large gains near the corner point. It is interesting to note that it is the only one which yields a rectangular second-order region for this particular choice of p_1 and p_2. This results from a rank-one dispersion matrix; see [6] for further discussion.

The preceding example shows that constant-composition codes can perform better than i.i.d. codes for a given choice of Q_1 and Q_2. In the remainder of this section, we argue that the gains remain present even after the full optimization, as is the case for the random-coding error exponents of certain MACs [9]. In contrast, in the

[1] On the other hand, for the collision channel with K non-zero symbols, time-sharing is required for $K \geq 8$ [18].

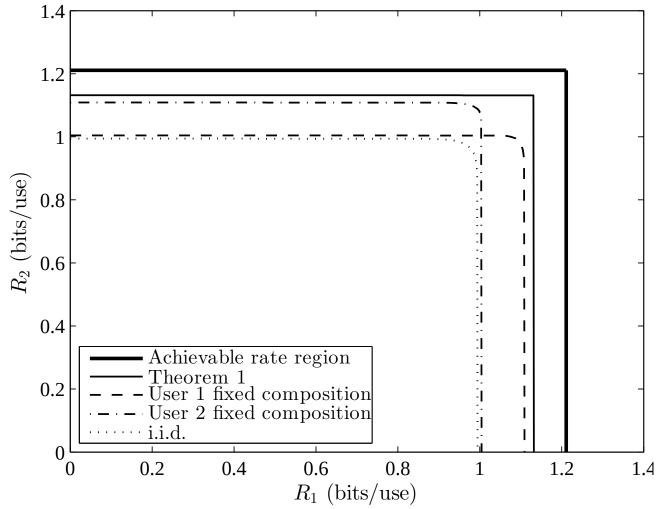

Figure 2. Second-order rate regions for the collision channel with $p_1 = p_2 = 0.2$, $n = 50$ and $\epsilon = 0.01$.

single-user setting, constant-composition codes yield higher error exponents and second-order rates for a given input distribution, but no gain after the optimization of the input distribution [11, Ex. 10.33] [4] [5].

For any given n, one can take the union of the achievable second-order regions in Theorem 1 (with the third-order term ignored) over all (Q_U, Q_1, Q_2). We denote the resulting region by $\mathcal{R}_n^*$, and we say that (Q_U, Q_1, Q_2) is first-order optimal (respectively, second-order optimal) if it achieves a point on the boundary of the capacity region (respectively, the boundary of $\mathcal{R}_n^*$). As n grows large, the second-order term in (11) becomes insignificant compared to the first-order term, and we conclude that any sequence of second-order optimal input distributions must be asymptotically first-order optimal. Thus, we will obtain the desired result by showing that the diagonal entries of $\boldsymbol{V}$ are strictly smaller than those of $\boldsymbol{V}^{\mathrm{iid}}$ in (12) under all first-order optimal input distributions. It suffices to consider the case $\mathcal{U} = \emptyset$, since otherwise these variances are simply weighted sums of the corresponding variances under $(Q_1(\cdot|u), Q_2(\cdot|u))$, weighted by Q_U. In fact, as stated above, it suffices to consider distributions of the form $Q_1 = (1 - 2p_1, p_1, p_1)$ and $Q_2 = (1 - 2p_2, p_2, p_2)$.

Denote the diagonal entries of $\boldsymbol{V}$ by (V_1, V_2, V_{12}), and those of $\boldsymbol{V}^{\mathrm{iid}}$ by $(V_1^{\mathrm{iid}}, V_2^{\mathrm{iid}}, V_{12}^{\mathrm{iid}})$. We observe from (12) and (18) that for $\nu = 1, 2, 12$, $V_\nu \le V_\nu^{\mathrm{iid}}$ with equality if and only if $\mathrm{Var}\big[i_\nu^{(1)}(X_1)\big] = 0$ and $\mathrm{Var}\big[i_\nu^{(2)}(X_2)\big] = 0$; the quantities $i_\nu^{(1)}$ and $i_\nu^{(2)}$ are defined as in (15)–(16) with $\mathcal{U} = \emptyset$. By a direct calculation, it can be shown that

$$i_{12}^{(1)}(x_1) = (1 - 2p_2) \log \frac{1}{1 - 2p_2} + 2p_2 \log \frac{1}{p_2} + \log \frac{1}{Q_1(x_1)},$$

which yields zero variance if and only if $p_1 = \frac{1}{3}$, i.e. $Q_1 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Similarly, $i_{12}^{(2)}(X_2)$ has zero variance if and only if $p_2 = \frac{1}{3}$. However, from Figure 1, we know that $p_1 = p_2 = \frac{1}{3}$ is not first-order optimal. A similar argument holds for $i_1^{(1)}$, $i_1^{(2)}$, $i_2^{(1)}$ and $i_2^{(2)}$, except that the condition $p_1 = p_2 = \frac{1}{3}$ is replaced by $p_1 = p_2 = 0.2867$. Once again, we see from Figure 1 that this choice is not first-order optimal. Thus, for $\nu = 1, 2, 12$, we have $V_\nu < V_\nu^{\mathrm{iid}}$ for all first-order optimal input distributions.

IV. Proof of Theorem 1

For clarity of exposition, we present the proof in the absence of time-sharing, and we assume that the input distributions Q_1 and Q_2 are types (i.e. $Q_\nu \in \mathcal{P}_n(\mathcal{X}_\nu)$ for $\nu = 1, 2$), and that $\boldsymbol{V}$ has full rank and hence $\boldsymbol{V} \succ \boldsymbol{0}$. In Section IV-C, we state the changes in the proof required to handle the general case. For $\nu = 1, 2, 12$, we write $i_\nu(x_1, x_2, y)$ to denote the quantities in (7)–(9) with the conditioning on u removed, and similarly for $\boldsymbol{i}(x_1, x_2, y)$.

Using the notation of Section I-B with the time-sharing sequence removed, we define the random variables

$$(\boldsymbol{X}_1, \boldsymbol{X}_2, \overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{Y}) \sim P_{\boldsymbol{X}_1}(\boldsymbol{x}_1)P_{\boldsymbol{X}_2}(\boldsymbol{x}_2)P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1)P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2)W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2), \tag{29}$$

where $W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \triangleq \prod_{i=1}^n W(y_i|x_{1,i}, x_{2,i})$. We make use of the threshold-based bound on the random-coding error probability $\overline{p}_e$ given in [19, Thm. 3], which is written in terms of three arbitrary output distributions $Q_{\boldsymbol{Y}|\boldsymbol{X}_2}$, $Q_{\boldsymbol{Y}|\boldsymbol{X}_1}$ and $Q_{\boldsymbol{Y}}$. Choosing these to be i.i.d. on the corresponding marginals of (5) (e.g. $P_{Y|X_2}$), we obtain

$$\begin{aligned}\overline{p}_e \le 1 - \mathbb{P}\Big[\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) \succ \boldsymbol{\gamma}\Big] + \frac{M_1 - 1}{2}\mathbb{P}\Big[i_1^n(\overline{\boldsymbol{X}}_1, \boldsymbol{X}_2, \boldsymbol{Y}) > \gamma_1\Big] \\ + \frac{M_2 - 1}{2}\mathbb{P}\Big[i_2^n(\boldsymbol{X}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{Y}) > \gamma_2\Big] + \frac{(M_1 - 1)(M_2 - 1)}{2}\mathbb{P}\Big[i_{12}^n(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{Y}) > \gamma_{12}\Big],\end{aligned} \tag{30}$$

where $\boldsymbol{\gamma} = [\gamma_1\ \gamma_2\ \gamma_{12}]^T$ is arbitrary, and

$$\boldsymbol{i}^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \triangleq \sum_{i=1}^n \boldsymbol{i}(x_{1,i}, x_{2,i}, y_i) \tag{31}$$

$$i_\nu^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \triangleq \sum_{i=1}^n i_\nu(x_{1,i}, x_{2,i}, y_i) \tag{32}$$

with $\boldsymbol{i}$ and i_ν defined in (6)–(9).

We claim that the second, third and fourth terms of (30) can be upper bounded by $M_\nu p_0(n) e^{-\gamma_\nu}$ for $\nu = 1, 2, 12$ respectively, where $p_0(n)$ is a polynomial depending only on the alphabet sizes, and

$$M_{12} \triangleq M_1 M_2. \tag{33}$$

We prove this for $\nu = 12$ only, since the other two are handled similarly. We write $Q_\nu^n(\boldsymbol{x}_\nu) \triangleq \prod_{i=1}^n Q_\nu(x_{\nu,i})$ for $\nu = 1, 2$, and make use of the fact that

$$P_{\boldsymbol{X}_\nu}(\boldsymbol{x}_\nu) = \frac{1}{\mu_{\nu,n}} Q_\nu^n(\boldsymbol{x}_\nu) \mathbb{1}\big\{\boldsymbol{x}_\nu \in T^n(Q_\nu)\big\}, \tag{34}$$

where $\mu_{\nu,n} \triangleq \mathbb{P}\big[\boldsymbol{X}'_\nu \in T^n(Q_\nu)\big]$ with $\boldsymbol{X}'_\nu \sim Q_\nu^n(\boldsymbol{x}_\nu)$. Using standard properties of types, we have $\mu_{\nu,n} \ge$

$(n+1)^{-(|\mathcal{X}_\nu|-1)}$ [11, pp. 17]. We therefore obtain

$$M_1 M_2 \mathbb{P}\Big[i_1^n(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{Y}) > \gamma_{12}\Big] = M_1 M_2 \sum_{\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y}} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \mathbb{1}\big\{i_{12}^n(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y}) > \gamma_{12}\big\} \tag{35}$$

$$\leq \frac{M_1 M_2}{\mu_{1,n}\mu_{2,n}} \sum_{\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y}} Q_1^n(\overline{\boldsymbol{x}}_1) Q_2^n(\overline{\boldsymbol{x}}_2) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \mathbb{1}\big\{i_{12}^n(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y}) > \gamma_{12}\big\} \tag{36}$$

$$\leq \frac{M_1 M_2}{\mu_{1,n}\mu_{2,n}} \sum_{\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y}} Q_1^n(\overline{\boldsymbol{x}}_1) Q_2^n(\overline{\boldsymbol{x}}_2) \bigg(\prod_{i=1}^n P_Y(y_i)\bigg) e^{-\gamma_{12}} \tag{37}$$

$$\leq M_1 M_2 p_0(n) e^{-\gamma_{12}}, \tag{38}$$

where (36) follows from (34) and by summing over all sequences instead of just those in $T^n(Q_\nu)$, (37) follows by using the definition of i_{12} and upper bounding the indicator function, and we have defined $p_0(n) \triangleq (n+1)^{(|\mathcal{X}_1|+|\mathcal{X}_2|-2)}$.

Returning to (30), we have thus far shown that

$$\overline{p}_e \leq 1 - \mathbb{P}\big[\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) \succ \boldsymbol{\gamma}\big] + p_0(n) \sum_{\nu=1,2,12} M_\nu e^{-\gamma_\nu}. \tag{39}$$

Using this bound with

$$\gamma_\nu = \log M_\nu + \Big(d + \frac{1}{2}\Big) \log n, \tag{40}$$

where $d = |\mathcal{X}_1| + |\mathcal{X}_2| - 2$ is the order of the polynomial $p_0(n)$, the statement of the theorem will follow using identical steps to [6, Thm. 2] once we prove the following:

1) $\mathbb{E}[\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})] = n\boldsymbol{I} + O\big(\frac{\log n}{n}\big)$ and $\mathrm{Cov}[\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})] = n\boldsymbol{V} + O\big(\frac{\log n}{\sqrt{n}}\big)$, where $(\boldsymbol{I}, \boldsymbol{V})$ are given by (17)–(18).
2) The probability on the right-hand side of (39) can be approximated using a multivariate Berry-Esseen theorem.

We prove these statements in Sections IV-A and IV-B respectively. The remaining details of the proof of Theorem 1 are omitted to avoid repetition with [6]. It should be noted that the growth rates $O\big(\frac{\log n}{n}\big)$ and $O\big(\frac{\log n}{\sqrt{n}}\big)$ in the former statement ensure that $g(n) = O(\log n)$, as stated in the theorem.

A. Calculation of Moments

Let $X_{1,i}$ denote the i-th entry of $\boldsymbol{X}_1$, and similarly for $X_{2,i}$ and Y_i. The first moment of $\boldsymbol{i}^n$ is easily found by writing

$$\mathbb{E}\big[\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})\big] = \sum_{i=1}^n \mathbb{E}\big[\boldsymbol{i}(X_{1,i}, X_{2,i}, Y_i)\big] = n\boldsymbol{I}, \tag{41}$$

where the last equality follows since, by symmetry, $X_{1,i} \sim Q_1$ and $X_{2,i} \sim Q_2$ for all i.

To compute the covariance matrix of $\boldsymbol{i}^n$, we write

$$\mathrm{Cov}\Big[\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})\Big] = \mathrm{Cov}\Bigg[\sum_{i=1}^{n}\boldsymbol{i}(X_{1,i},X_{2,i},Y_i)\Bigg] \tag{42}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}\Big[\boldsymbol{i}(X_{1,i},X_{2,i},Y_i),\boldsymbol{i}(X_{1,j},X_{2,j},Y_j)\Big] \tag{43}$$

$$= n\mathrm{Cov}\Big[\boldsymbol{i}(X_1,X_2,Y)\Big] + (n^2-n)\mathrm{Cov}\Big[\boldsymbol{i}(X_1,X_2,Y),\boldsymbol{i}(X_1',X_2',Y')\Big], \tag{44}$$

where (X_1,X_2,Y) and (X_1',X_2',Y') correspond to two arbitrary but different indices in $\{1,\cdots,n\}$ (e.g. one can set $(X_1,X_2,Y)=(X_{1,1},X_{2,1},Y_1)$ and $(X_1',X_2',Y')=(X_{1,2},X_{2,2},Y_2)$). In (44), we have used the fact that, by the symmetry of the codebook construction, the n terms in (43) with $i=j$ are equal, and similarly for the n^2-n terms with $i\neq j$.

To compute the cross-covariance matrix in (44), we need the joint distribution of (X_1,X_2,Y) and (X_1',X_2',Y'). This distribution is easily understood by considering the following procedure for generating a codeword uniformly over $T^n(Q)$: (i) Fix an arbitrary sequence $\boldsymbol{x}=(x_1,\cdots,x_n)$ with composition Q; (ii) Randomly choose a symbol from the n symbols of $\boldsymbol{x}$ (each with probability $\frac{1}{n}$) and place it in position 1 of the codeword, (iii) From the $n-1$ remaining symbols of $\boldsymbol{x}$, randomly choose one (each with probability $\frac{1}{n-1}$) and place it in position 2 of the codeword; (iv) Continue until all n symbols have been placed. Stated more compactly, this procedure generates a codeword uniformly over $T^n(Q)$ by randomly permuting the symbols of an arbitrary sequence $\boldsymbol{x}\in T^n(Q)$.

From the above procedure, we conclude that

$$\mathbb{P}[X_\nu = x_\nu] = Q_\nu(x_\nu) \tag{45}$$

$$\mathbb{P}[X_\nu' = x_\nu' | X_\nu = x_\nu] = \frac{nQ_\nu(x_\nu') - \mathbb{1}\{x_\nu = x_\nu'\}}{n-1} \tag{46}$$

for $\nu=1,2$. Let $Q_\nu'(x_\nu'|x_\nu)$ denote the right-hand side of (46). The cross-covariance matrix in (44) is given by

$$\mathrm{Cov}\Big[\boldsymbol{i}(X_1,X_2,Y),\boldsymbol{i}(X_1',X_2',Y')\Big]$$

$$= \mathbb{E}\Big[\big(\boldsymbol{i}(X_1,X_2,Y)-\boldsymbol{I}\big)\big(\boldsymbol{i}(X_1',X_2',Y')-\boldsymbol{I}\big)^T\Big] \tag{47}$$

$$= \sum_{x_1,x_2,y} Q_1(x_1)Q_2(x_2)W(y|x_1,x_2)$$

$$\times \sum_{x_1',x_2',y'} Q_1'(x_1'|x_1)Q_2'(x_2'|x_2)W(y'|x_1',x_2')\big(\boldsymbol{i}(x_1,x_2,y)-\boldsymbol{I}\big)\big(\boldsymbol{i}(x_1',x_2',y')-\boldsymbol{I}\big)^T \tag{48}$$

$$= \boldsymbol{M}_1+\boldsymbol{M}_2+\boldsymbol{M}_3+\boldsymbol{M}_4, \tag{49}$$

where the four terms in (49) correspond to the four terms in the expansion of $\big(nQ_1(x_1')-\mathbb{1}\{x_1=x_1'\}\big)\big(nQ_2(x_2')-$

$\mathbb{1}\{x_2 = x_2'\}\big)$ resulting from (46). Specifically, we obtain

$$\boldsymbol{M}_1 = \frac{n^2}{(n-1)^2}\mathbb{E}\Big[\boldsymbol{i}(X_1,X_2,Y) - \boldsymbol{I}\Big]\mathbb{E}\Big[\boldsymbol{i}(X_1,X_2,Y) - \boldsymbol{I}\Big]^T \tag{50}$$

$$\boldsymbol{M}_2 = -\frac{n}{(n-1)^2}\mathbb{E}\Big[\big(\boldsymbol{i}(X_1,X_2,Y) - \boldsymbol{I}\big)\big(\boldsymbol{i}(\overline{X}_1,X_2,\overline{Y}) - \boldsymbol{I}\big)^T\Big] \tag{51}$$

$$\boldsymbol{M}_3 = -\frac{n}{(n-1)^2}\mathbb{E}\Big[\big(\boldsymbol{i}(X_1,X_2,Y) - \boldsymbol{I}\big)\big(\boldsymbol{i}(X_1,\overline{X}_2,\overline{\overline{Y}}) - \boldsymbol{I}\big)^T\Big] \tag{52}$$

$$\boldsymbol{M}_4 = \frac{1}{(n-1)^2}\mathbb{E}\Big[\big(\boldsymbol{i}(X_1,X_2,Y) - \boldsymbol{I}\big)\big(\boldsymbol{i}(X_1,X_2,\widetilde{Y}) - \boldsymbol{I}\big)^T\Big] \tag{53}$$

under the joint distribution

$$(U,X_1,X_2,Y,\overline{X}_1,\overline{X}_2,\overline{Y},\overline{\overline{Y}},\widetilde{Y}) \sim Q_U(u)Q_1(x_1|u)Q_2(x_2|u)W(y|x_1,x_2) \\ \times Q_1(\overline{x}_1|u)Q_2(\overline{x}_2|u)W(\overline{y}|\overline{x}_1,x_2)W(\overline{\overline{y}}|x_1,\overline{x}_2)W(\widetilde{y}|x_1,x_2). \tag{54}$$

We observe that $\boldsymbol{M}_1$ is the zero matrix, and $\boldsymbol{M}_4 = O\big(\frac{1}{n^2}\big)$. Furthermore, recalling the definitions of $\boldsymbol{i}^{(1)}$ and $\boldsymbol{i}^{(2)}$ in (15)–(16), we have

$$\frac{-(n-1)^2}{n}\boldsymbol{M}_2 = \mathbb{E}\bigg[\mathbb{E}\Big[\big(\boldsymbol{i}(X_1,X_2,Y) - \boldsymbol{I}\big)\,\Big|\,X_2\Big]\mathbb{E}\Big[\big(\boldsymbol{i}(\overline{X}_1,X_2,\overline{Y}) - \boldsymbol{I}\big)\,\Big|\,X_2\Big]^T\bigg] \tag{55}$$

$$= \mathrm{Cov}\big[\boldsymbol{i}^{(2)}(X_2)\big]. \tag{56}$$

It follows that

$$\boldsymbol{M}_2 = \frac{-n}{(n-1)^2}\mathrm{Cov}\big[\boldsymbol{i}^{(2)}(X_2)\big], \tag{57}$$

and we similarly have

$$\boldsymbol{M}_3 = \frac{-n}{(n-1)^2}\mathrm{Cov}\big[\boldsymbol{i}^{(1)}(X_1)\big]. \tag{58}$$

Using the identity $\frac{n}{(n-1)^2} = \frac{1}{n} + O\big(\frac{1}{n^2}\big)$ and combining (44), (49), (57) and (58), we obtain

$$\mathrm{Cov}\big[\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})\big] = n\boldsymbol{V} + O(1), \tag{59}$$

where $\boldsymbol{V}$ is defined as in (18) with $\mathcal{U} = \emptyset$.

B. A Combinatorial Berry-Esseen Theorem

Before stating the required Berry-Esseen theorem, we outline some of the relevant literature. A combinatorial CLT was given by Hoeffding [10], who proved the asymptotic normality of random variables of the form $\sum_{j=1}^n f_n(j,\pi(j))$, where f_n is a real-valued function taking arguments on $1,\cdots,n$, and $\pi(\cdot)$ is uniformly distributed on the set of permutations of $\{1,\cdots,n\}$. The rate of convergence (i.e. Berry-Esseen theorem) was studied by Bolthausen [20], who proved $O\big(\frac{1}{\sqrt{n}}\big)$ convergence under fairly general conditions. An extension to the multivariate setting was given by Bolthausen and Götze [21].

A more general setting is that in which each $f_n(j_1,j_2)$ is replaced by a random variable $Z_n(j_1,j_2)$, independent of $\pi(\cdot)$, such that $Z_n(j_1,j_2)$ is independent of $Z_n(j_1',j_2')$ whenever $(j_1,j_2) \ne (j_1',j_2')$. Berry-Esseen theorems for this setting were given by von Bahr [22] and Ho and Chen [23]. The analysis of each scalar quantity $i_\nu^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$ (see (29) and (32)) falls into this setting upon identifying

$$Z_n(j_1, j_2) = i_\nu\big(x_{1,j_1}, x_{2,j_2}, Y_n(j_1, j_2)\big), \tag{60}$$

where $\boldsymbol{x}_1 = (x_{1,1}, \cdots, x_{1,n})$ and $\boldsymbol{x}_2 = (x_{2,1}, \cdots, x_{2,n})$ are arbitrary sequences of type Q_1 and Q_2 respectively, and $Y_n(j_1, j_2) \sim W(\cdot|x_{1,j_1}, x_{2,j_2})$. Under this choice, the permutation $\pi(\cdot)$ applied to $\boldsymbol{x}_2$ induces the uniform distribution on $T^n(Q_2)$, as desired. By symmetry, we can let $\boldsymbol{x}_1$ be an arbitrary element of $T^n(Q_1)$ (e.g. see [10, Thm. 5]).

In our case, a multivariate generalization to random vectors $\boldsymbol{Z}_n(j_1, j_2)$ in $\mathbb{R}^3$ is required. The desired Berry-Esseen theorem is a special case of a more general result by Loh [24, Thm. 2] for a problem known as Latin hypercube sampling. When specialized to our setting, we obtain the following theorem. Details on the how to specialize [24, Thm. 2] are provided in Appendix A. We define the quantities

$$\boldsymbol{\Sigma}_n \triangleq \frac{1}{n}\mathrm{Cov}\big[\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})\big], \tag{61}$$

$$\widehat{\boldsymbol{S}}_n \triangleq \boldsymbol{\Sigma}_n^{-\frac{1}{2}}\left(\frac{1}{\sqrt{n}}\big(\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) - n\boldsymbol{I}\big)\right) \tag{62}$$

$$\boldsymbol{T}_n(x_1, x_2) \triangleq \boldsymbol{\Sigma}_n^{-\frac{1}{2}}\Big(\boldsymbol{i}\big(x_1, x_2, Y(x_1, x_2)\big) - \boldsymbol{i}^{(1)}(x_1) - \boldsymbol{i}^{(2)}(x_2) + \boldsymbol{I}\Big) \tag{63}$$

$$\beta_n \triangleq \sum_{x_1, x_2} Q_1(x_1)Q_2(x_2)\mathbb{E}\Big[\|\boldsymbol{T}_n(x_1, x_2)\|^3\Big], \tag{64}$$

where $Y(x_1, x_2) \sim W(\cdot|x_1, x_2)$. From (59), we have $\boldsymbol{\Sigma}_n = \boldsymbol{V} + O(n^{-1})$.

Theorem 2. (Corollary of [24, Thm. 2]) *Let the input distributions Q_1 and Q_2 be given, and consider the quantities $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$, $(\boldsymbol{I}, \boldsymbol{V})$ and $(\boldsymbol{\Sigma}_n, \widehat{\boldsymbol{S}}_n)$ respectively defined in* (29), (17)–(18) *and* (61)–(62). *Suppose that $\boldsymbol{V} \succ \boldsymbol{0}$, and that n is sufficiently large so that $\boldsymbol{\Sigma}_n \succ \boldsymbol{0}$. Then*

$$\Big|\mathbb{P}\big[\widehat{\boldsymbol{S}}_n \in \mathcal{A}\big] - \mathbb{P}\big[\boldsymbol{Z} \in \mathcal{A}\big]\Big| \le \frac{1}{\sqrt{n}}\frac{K}{\beta_n} \tag{65}$$

for any convex, Borel measurable set $\mathcal{A} \subseteq \mathbb{R}^d$, where $\boldsymbol{Z} \sim N(\boldsymbol{0}, \mathbb{I}_{3\times 3})$, K is a universal constant,[2] *and β_n is defined in* (64).

In the discrete memoryless setting, one can show that $\beta_n = \Theta(1)$ using (61) and the uniform bounding techniques in [6, Appendix D]. Thus, we obtain the required $O\big(\frac{1}{\sqrt{n}}\big)$ convergence in (65). Assuming that $\boldsymbol{V} \succ \boldsymbol{0}$ and n is

[2]In the extension of this theorem to the MAC with several users, this constant may depend on the number of users.

sufficiently large so that $\mathbf{\Sigma}_n \succ \mathbf{0}$, we can use Theorem 2 to bound the probability in (39) by writing

$$\mathbb{P}\big[\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}) \succ \boldsymbol{\gamma}\big] = \mathbb{P}\Big[\frac{1}{\sqrt{n}}\big(\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}) - n\boldsymbol{I}\big) \succ \frac{1}{\sqrt{n}}\big(\boldsymbol{\gamma} - n\boldsymbol{I}\big)\Big] \tag{66}$$

$$= \mathbb{P}\Big[\mathbf{\Sigma}_n^{-\frac{1}{2}}\Big(\frac{1}{\sqrt{n}}\big(\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}) - n\boldsymbol{I}\big)\Big) \in \mathcal{A}_n\Big] \tag{67}$$

$$= \mathbb{P}\big[\boldsymbol{Z} \in \mathcal{A}_n\big] + O\bigg(\frac{1}{\sqrt{n}}\bigg) \tag{68}$$

$$= \mathbb{P}\Big[\mathbf{\Sigma}_n^{\frac{1}{2}}\boldsymbol{Z} \succ \frac{1}{\sqrt{n}}\big(\boldsymbol{\gamma} - \boldsymbol{I}\big)\Big] + O\bigg(\frac{1}{\sqrt{n}}\bigg), \tag{69}$$

where (67) follows by defining $\mathcal{A}_n$ to be the image of the rectangular region in (66) under $\mathbf{\Sigma}_n^{-\frac{1}{2}}$, (68) follows by defining $\boldsymbol{Z} \sim N(\mathbf{0}, \mathbb{I}_{3\times 3})$ and applying Theorem 2, and (69) follows by reversing the step from (66) to (67). These steps are similar in nature to [6, Appendix B], where a Cholesky decomposition is used.

C. General Case

In the case that Q_1 and Q_2 do not correspond to types of length n, we can simply repeat the above derivation using $Q_{1,n}$ and $Q_{2,n}$, defined in Section I-B. In this case, each type differs from its corresponding distribution by at most $O\big(\frac{1}{n}\big)$ in each entry, which does not affect the second-order asymptotics.

In general, the dispersion matrix $\boldsymbol{V}$ may not have full rank, in which case Theorem 2 does not directly apply. However, we can deal with this case by reducing the problem to a lower dimension, similarly to [6, Sec. VIII-A]. The argument here is slightly more involved, since $n\boldsymbol{V}$ is not necessarily the exact covariance matrix of $\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})$, due to the additional $O(1)$ term in (59). If $\boldsymbol{V}$ has rank $r < 3$, then there exist matrices $\boldsymbol{T}$ and $\tilde{\boldsymbol{V}}$ of dimension $3 \times r$ and $r \times r$ respectively such that $\boldsymbol{V} = \boldsymbol{T}\tilde{\boldsymbol{V}}\boldsymbol{T}^T$, and such that

$$\mathbb{E}\Big[\boldsymbol{T}\tilde{\boldsymbol{i}}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})\Big] = n\boldsymbol{I} \tag{70}$$

$$\mathrm{Cov}\Big[\tilde{\boldsymbol{i}}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})\Big] = n\tilde{\boldsymbol{V}} + O(1) \tag{71}$$

for some $r \times 1$ subvector $\tilde{\boldsymbol{i}}^n$ of $\boldsymbol{i}^n$. It follows that

$$\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}) = \boldsymbol{T}\tilde{\boldsymbol{i}}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}) + \mathbf{\Delta}, \tag{72}$$

where $\mathbf{\Delta}$ is a 3×1 random vector (not necessarily independent of $\boldsymbol{i}'^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})$) with $\mathbb{E}[\mathbf{\Delta}] = 0$ and $\mathrm{Cov}[\mathbf{\Delta}] = O(1)$. We can thus write

$$\mathbb{P}\big[\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}) \succ \boldsymbol{\gamma}\big] = \mathbb{P}\big[\boldsymbol{T}\tilde{\boldsymbol{i}}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}) + \mathbf{\Delta} \succ \boldsymbol{\gamma}\big] \tag{73}$$

$$\geq \mathbb{P}\big[\boldsymbol{T}\tilde{\boldsymbol{i}}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}) \succ \boldsymbol{\gamma} + \delta_n\mathbf{1}\big] - \mathbb{P}\big[\|\mathbf{\Delta}\|_\infty \geq \delta_n\big], \tag{74}$$

where (74) holds for an arbitrary constant $\delta_n \geq 0$ by [6, Lemma 8]. Using Chebsyhev's inequality and (71), we have

$$\mathbb{P}\big[\|\mathbf{\Delta}\|_\infty \geq \delta_n\big] \leq O\big(\delta_n^{-2}\big). \tag{75}$$

Following the analysis of this section and [6], we obtain that the $O(\delta_n^{-2})$ term in (75) contributes an additive $O(\sqrt{n}\delta_n^{-2})$ term to the expansion in (11), whereas the addition of $\delta_n\mathbf{1}$ in the first probability in (74) contributes an additive $O(\delta_n)$ term to the expansion. The overall third-order term $O(\sqrt{n}\delta_n^{-2}+\delta_n)$ is minimized by $\delta_n = O(n^{\frac{1}{6}})$, yielding $g(n) = O(n^{\frac{1}{6}})$ as stated in Theorem 1.

Finally, we consider the case that $\mathcal{U} \neq \emptyset$, and thus the codewords are drawn uniformly over the conditional type class $T_{\boldsymbol{u}}(Q_\nu)$ for some $\boldsymbol{u} \in T^n(Q_U)$. In this case, the procedure described in Section IV-A for generating a codeword uniformly over the type class should be replaced by the following procedure. Let $\boldsymbol{x}$ be an arbitrary element of the conditional type class $T_{\boldsymbol{u}}(\cdot)$. Instead of randomly permuting the entire sequence $\boldsymbol{x}$, a random permutation of the subsequence $\boldsymbol{x}^{(u)}$ corresponding to the indices where $\boldsymbol{u}$ equals u is applied independently for each value of $u \in \mathcal{U}$. Due to this independence, the covariance matrices in (43) are zero between symbols with different corresponding u values. Within each subsequence, the joint distribution between two symbols is similar to that of (45)–(46), with $Q_1(\cdot|u)$ replacing Q_1 and $nQ_U(u)$ replacing n. The quantity $\boldsymbol{i}^n$ in (31) is replaced by

$$\boldsymbol{i}^n(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}) \stackrel{\triangle}{=} \sum_{i=1}^{n} \boldsymbol{i}(u_i, x_{1,i}, x_{2,i}, y_i) \tag{76}$$

$$= \sum_{u} \sum_{i=1}^{nQ_U(u)} \boldsymbol{i}(u, x_{1,i}^{(u)}, x_{2,i}^{(u)}, y_i^{(u)}), \tag{77}$$

where $x_{1,i}^{(u)}$ is the i-th entry of $\boldsymbol{x}_1$ for which the corresponding $\boldsymbol{u}$ entry equals u, and similarly for $x_{2,i}^{(u)}$ and $y_i^{(u)}$. From Theorem 2, we conclude that under $(\boldsymbol{u},\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})$, each inner summation in (77) is asymptotically normal with $O\Big(\frac{1}{\sqrt{nQ_U(u)}}\Big) = O\Big(\frac{1}{\sqrt{n}}\Big)$ convergence. It follows that the overall sum is also asymptotically normal with $O\Big(\frac{1}{\sqrt{n}}\Big)$ convergence.

Using the above observations and repeating the analysis of this section, we obtain the more general result of Theorem 1.

V. Application to the Gaussian MAC

Thus far, we have limited our attention to the DM-MAC, thus permitting an analysis based on types and combinatorial methods. In this section, we show that the same techniques can be used to derive second-order asymptotics for the Gaussian MAC using an increasingly fine quantization of the input, similarly to Hayashi [5, Thm. 3] and Tan [25]. The channel is described by

$$Y = \sqrt{P_1}X_1 + \sqrt{P_2}X_2 + Z, \tag{78}$$

where $Z \sim N(0,1)$, and where the codewords for user $\nu = 1,2$ are constrained to satisfy $\|\mathbf{x}_\nu\|^2 \le n$. We can think of P_1 and P_2 as representing the powers for users 1 and 2 respectively, though for convenience we have factored them into the channel itself. The capacity region is pentagonal [26, Sec. 15.1], and is achieved by the Gaussian input distributions, namely $Q_1, Q_2 \sim N(0,1)$. Under this choice, the pair $(\mathbf{I},\mathbf{V})$ in (17)–(18) can be written explicitly as follows:

$$\mathbf{I} = \begin{bmatrix} \frac{1}{2}\log(1+P_1) \\ \frac{1}{2}\log(1+P_2) \\ \frac{1}{2}\log(1+P_1+P_2) \end{bmatrix} \tag{79}$$

$$\mathbf{V} = \begin{bmatrix} \frac{P_1(2+P_1)}{2(1+P_1)^2} & \frac{P_1P_2}{2(1+P_1)(1+P_2)} & \frac{P_1(2+P_1+P_2)}{2(1+P_1)(1+P_1+P_2)} \\ \frac{P_1P_2}{2(1+P_1)(1+P_2)} & \frac{P_2(2+P_2)}{2(1+P_2)^2} & \frac{P_2(2+P_1+P_2)}{2(1+P_2)(1+P_1+P_2)} \\ \frac{P_1(2+P_1+P_2)}{2(1+P_1)(1+P_1+P_2)} & \frac{P_2(2+P_1+P_2)}{2(1+P_2)(1+P_1+P_2)} & \frac{(P_1+P_2)(2+P_1+P_2)+2P_1P_2}{2(1+P_1+P_2)^2} \end{bmatrix}. \tag{80}$$

The goal of this section is to present an alternative proof of the following result, which was recently derived by MolavianJazi and Laneman using random coding with a uniform distribution over the surface of a sphere [14].

Theorem 3. *The pair* $(\mathbf{I}, \mathbf{V})$ *in* (79)–(80) *is second-order achievable for the Gaussian MAC described by* (78).

Theorem 3 follows in a straightforward fashion from Theorem 1 and the following lemma, which states the existence of a sequence of discrete input distributions Q_{m1} and Q_{m2} of cardinality m such that the corresponding vector-matrix pair $(\mathbf{I}_m, \mathbf{V}_m)$ converges to $(\mathbf{I}, \mathbf{V})$, with the convergence $\mathbf{I}_m \to \mathbf{I}$ being exponentially fast in m. This generalizes a result by Wu and Verdú for the scalar case [27], and is proved similarly.

Lemma 1. *There exists a sequence of discrete input distributions* Q_{m1} *and* Q_{m2} *of cardinality* m *with a corresponding matrix-vector pair* $(\mathbf{I}_m, \mathbf{V}_m)$ *defined according to* (17)–(18) *under the channel law in* (78) *such that (i)* $\|\mathbf{I}_m - \mathbf{I}\|_\infty \le e^{-\psi m}$ *for some* $\psi > 0$ *and sufficiently large* m*, (ii)* $\|\mathbf{V}_m - \mathbf{V}\|_\infty \to 0$*, and (iii) the third absolute moment of each entry of* $\mathbf{i}(X_{m1}, X_{m2}, Y)$ *under* $Q_{m1} \times Q_{m2} \times Y$ *is uniformly bounded in* m.

Proof: See Appendix B. ∎

We are now in a position to prove Theorem 3. Setting $m = n^{\frac{1}{4}}$, we have from parts (i)–(ii) of Lemma 1 that $\|\mathbf{I}_m - \mathbf{I}\|_\infty \le e^{-\psi n^{\frac{1}{4}}}$ and $\|\mathbf{V}_m - \mathbf{V}\|_\infty \to 0$, thus yielding

$$n\boldsymbol{I}_m - \sqrt{n}\mathsf{Q}_{\mathrm{inv}}(\boldsymbol{V}_m, \epsilon) = n\mathbf{I} - \sqrt{n}\mathsf{Q}_{\mathrm{inv}}(\boldsymbol{V}, \epsilon) + o(\sqrt{n}), \tag{81}$$

where we have used the continuity of $\mathsf{Q}_{\mathrm{inv}}$ for $\mathbf{V} \succ 0$. It remains to show that under $(\mathbf{I}_m, \mathbf{V}_m)$, the third-order term $g(n)$ (see (11)) from the proof of Theorem 3 is also $o(\sqrt{n})$. This term is affected by two remainder terms appearing in Section IV: the second term in (40), and the remainder term in the Berry-Esseen theorem in (65). For the former term, we note that $d = |\mathcal{X}_1| + |\mathcal{X}_2| - 2$ as stated following (40), and we thus have $(d + \frac{1}{2}) \log n = O(n^{\frac{1}{4}} \log n) = o(\sqrt{n})$, as desired. For the latter term, we note from (64) and the third part of Lemma 1 that $\frac{1}{\sqrt{n}} \frac{K}{\beta_n} = O\big(\frac{1}{\sqrt{n}}\big)$, thus yielding the same growth rate as that for the discrete memoryless case.

APPENDIX

A. Proof of Theorem 2

In this section we outline the problem studied by Loh [24] and describe how Theorem 2 can be attained. To avoid a clash of notation, we use the symbol Z in place of X defined in [24], and (j_1, j_2) in place of (i_1, i_2). We use the symbol U defined in [24] even though it clashes with the time-sharing variable in the present paper, since time-sharing is not considered in this section. Throughout this section, we write $A \overset{d}{=} B$ if the random variables A and B have the same distribution.

The dimension d of the hypercube in [24] corresponds to the number of users of the MAC, so we focus on the case $d = 2$. Let $\pi_1(\cdot)$ and $\pi_2(\cdot)$ be independent random permutations of $\{1, \cdots, n\}$, uniformly distributed over the $n!$ possible permutations. For $\nu = 1, 2$ and $j_1, j_2 = 1, \cdots, n$, define the random variables $U_\nu(j_1, j_2) \sim \mathrm{Uniform}(0, 1)$, independent of each other and of $\pi_1(\cdot)$ and $\pi_2(\cdot)$. Define

$$Z_\nu(j_1, j_2) \triangleq \frac{j_\nu - U_\nu(j_1, j_2)}{n},\ \nu = 1, 2 \tag{82}$$

$$\boldsymbol{Z}(j_1, j_2) \triangleq \begin{bmatrix} Z_1(j_1, j_2) \\ Z_2(j_1, j_2) \end{bmatrix}. \tag{83}$$

Loh considers sample means of the form

$$\hat{\mu}_n \triangleq \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{f}\Big(\boldsymbol{Z}\big(\pi_1(j), \pi_2(j)\big)\Big), \tag{84}$$

where $\boldsymbol{f}(\cdot)$ is a vector-valued function with two arguments. We henceforth write $\boldsymbol{f}(z_1, z_2)$ and $\boldsymbol{f}([z_1\, z_2]^T)$ interchangeably. We wish to choose $\boldsymbol{f}$ in such a way that, for each (j_1, j_2), we obtain

$$\boldsymbol{f}\big(\boldsymbol{Z}(j_1, j_2)\big) \stackrel{d}{=} \boldsymbol{i}\big(x_{1,j_1}, x_{2,j_2}, Y_n(j_1, j_2)\big),$$

where $\boldsymbol{x}_1 = (x_{1,1}, \cdots, x_{1,n})$ and $\boldsymbol{x}_2 = (x_{2,1}, \cdots, x_{2,n})$ are arbitrary sequences of type Q_1 and Q_2 respectively, and $Y_n(j_1, j_2) \sim W(\cdot|x_{1,j_1}, x_{2,j_2})$ with independence between different (j_1, j_2) pairs. Since drawing a codeword uniformly from a type class is equivalent to randomly permuting any codeword of the given type, such an identification will yield $\hat{\mu}_n$ in (84) with the same distribution as $\frac{1}{n}\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$ in (31).

Using (82) and the fact that $U_\nu(j_1, j_2) \in (0, 1)$ almost surely, any realization z_ν of Z_ν uniquely determines both the index j_ν and the variable U_ν in the numerator of (82). Thus, overloading the symbol $\boldsymbol{f}$, we can write $\boldsymbol{f}(z_1, z_2) = \boldsymbol{f}(j_1, j_2, u_1, u_2)$ without ambiguity. We choose

$$\boldsymbol{f}(j_1, j_2, u_1, u_2) = \boldsymbol{i}\big(x_{1,j_1}, x_{2,j_2}, F^{-1}_{Y_n(i_1,i_2)}(u_1 \oplus u_2)\big), \tag{85}$$

where $F^{-1}_{Y_n(j_1,j_2)}(u) = \inf\big\{y \,:\, F_{Y_n(j_1,j_2)}(y) \geq u\big\}$ is the inverse CDF of $Y_n(j_1, j_2)$, and $\oplus$ denotes real addition modulo one (e.g. $1.1 \oplus 2.3 = 0.4$). Clearly $U_1 \sim \mathrm{Uniform}(0, 1)$ and $U_2 \sim \mathrm{Uniform}(0, 1)$ implies $U_1 \oplus U_2 \sim \mathrm{Uniform}(0, 1)$, and since $F_X^{-1}(U) \stackrel{d}{=} X$ for any random variable X (discrete or continuous) with CDF F_X, it follows that $F^{-1}_{Y_n(j_1,j_2)}(U_1 \oplus U_2) \stackrel{d}{=} Y_n(j_1, j_2)$ as desired. While the choice of the function in (85) to recover the desired distribution is not unique, it yields a simpler analysis when either u_1 or u_2 is fixed. In particular, we have

$$F^{-1}_{Y_n(j_1,j_2)}(u_1 \oplus U_2(j_1, j_2)) \stackrel{d}{=} Y_n(j_1, j_2) \tag{86}$$

$$F^{-1}_{Y_n(j_1,j_2)}(U_1(j_1, j_2) \oplus u_2) \stackrel{d}{=} Y_n(j_1, j_2) \tag{87}$$

for all (u_1, u_2). This observation simplifies the evaluation of several quantities in [24].

Using the choice of $\boldsymbol{f}$ in (85), one can obtain the following equivalences between quantities defined in [24] (left-hand sides) and the present paper (right-hand sides):

$$\hat{\mu} = \frac{1}{n}\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) \tag{88}$$

$$\mu = \boldsymbol{I} \tag{89}$$

$$\int_{[0,1]^2} f_{\text{rem}}(\boldsymbol{z})f_{\text{rem}}(\boldsymbol{z})^T d\boldsymbol{z} = \boldsymbol{V} \tag{90}$$

$$\Sigma_{\text{lhs}} = \frac{1}{n}\boldsymbol{\Sigma}_n \tag{91}$$

$$\mu_{-\nu}(j_\nu) = \boldsymbol{i}^{(\nu)}(x_{\nu,j_\nu}),\ \nu = 1,2 \tag{92}$$

$$W = \widehat{\boldsymbol{S}}_n \tag{93}$$

$$Y(j_1, j_2) = \frac{1}{\sqrt{n}}\boldsymbol{T}_n(x_{1,j_1}, x_{2,j_2}) \tag{94}$$

$$C_{d,p} = K \tag{95}$$

$$\beta_3 = \frac{\beta_n}{\sqrt{n}}. \tag{96}$$

Using these equivalences, we see that Theorem 2 coincides with [24, Thm. 2].

B. *Proof of Lemma 1*

The arguments in the proof are very similar to [27], so we only explain the differences. We choose Q_{m1} and Q_{m2} according the Gauss quadrature rule Q_g [27, Sec. II], which satisfies the property of having moments which coincide with that of a standard Gaussian random variable for all moments up to order $2m-1$ [27, Thm. 2]. Using this property along with the fact that Q_g converges weakly to $N(0,1)$, we immediately obtain parts (ii) and (iii) of Lemma 1.

Define $(X_{m1}, X_{m2}, Y_m) \sim Q_{m1} \times Q_{m2} \times W$ and $(X_1, X_2, Y) \sim Q_1 \times Q_2 \times W$, where $Q_1, Q_2 \sim N(0,1)$. Using the identity [26, Eq. (15.142)]

$$I(X_{m1}; Y_m | X_{m2}) = H\big(\sqrt{P_1}X_{m1} + Z\big) - H(Z), \tag{97}$$

we see that the convergence of the first entry of $\mathbf{I}_m$ to $\mathbf{I}$ is precisely that studied in [27], and similarly for the second entry. It remains to study the third entry, i.e. to show that $I(X_{m1}, X_{m2}; Y_m) \to I(X_1, X_2; Y)$ exponentially fast. Analogously to [27, Eq. (5)], we have

$$I(X_1, X_2; Y) - I(X_{m1}, X_{m2}; Y_m) = D\big(\sqrt{P_1}X_1 + \sqrt{P_2}X_2 + Z \| \sqrt{P_1}X_{m1} + \sqrt{P_2}X_{m2} + Z\big) \tag{98}$$

$$\triangleq D_m. \tag{99}$$

Using nearly identical arguments to [27, Sec. V] with an "optimal" output distribution of $N(0, 1+P_1+P_2)$ (instead of $N(0, 1+\text{snr})$), we obtain analogously to [27, Eq. (54)] that

$$D_m \le \sum_{k\ge1} \frac{1}{k!}\Big(\frac{P_1+P_2}{1+P_1+P_2}\Big)^k \Big|\mathbb{E}\Big[H_k\Big(\frac{\sqrt{P_1}X_{m1} + \sqrt{P_2}X_{m2}}{\sqrt{P_1+P_2}}\Big)\Big]\Big|, \tag{100}$$

where H_k is the Hermite polynomial of degree k; see [27, Eq. (15)]. As shown in [27], we obtain the desired convergence rate provided that the expectation appearing in (100) is zero for odd values of k, and also for $k \leq 2m-1$. For odd values of k, we simply use the same symmetry argument as [27]; since the distributions of X_{m1} and X_{m2} are both symmetric, so is that of their weighted sum. To handle the remaining values $k \leq 2m-1$, we write

$$H_k(a+b) = \sum_{i=0}^{k}\sum_{j=0}^{k} c_{ij} a^i b^j$$

for some constants c_{ij}. By the independence of X_{m1} and X_{m2}, the expectation $\mathbb{E}\Big[H_k\Big(\frac{\sqrt{P_1}X_{m1}+\sqrt{P_2}X_{m2}}{\sqrt{P_1+P_2}}\Big)\Big]$ depends only on the first k moments of X_{m1} and X_{m2}. Since the i-th moment of $X_{m\nu}$ coincides with the corresponding moment of $X_\nu \sim N(0,P_\nu)$ for $i=1,\cdots,2m-1$ [27, Thm. 2], we have for $k \leq 2m-1$ that

$$\mathbb{E}\Big[H_k\Big(\frac{\sqrt{P_1}X_{m1}+\sqrt{P_2}X_{m2}}{\sqrt{P_1+P_2}}\Big)\Big] = \mathbb{E}\Big[H_k\Big(\frac{\sqrt{P_1}X_1+\sqrt{P_2}X_2}{\sqrt{P_1+P_2}}\Big)\Big] \tag{101}$$

$$= 0, \tag{102}$$

where (102) follows since for any k we have $H_k(X)=0$ under $X \sim N(0,1)$ [27].

Acknowledgments

We thank Vincent Tan and Pierre Moulin for helpful discussions. In particular, the use of quantization arguments for the Gaussian case in Section V was recommended by Vincent.

References

[1] R. Fano, *Transmission of information: A statistical theory of communications*. MIT Press, 1961.

[2] Y. Altuğ and A. B. Wagner, "Moderate deviations in channel coding," http://arxiv.org/abs/1208.1924.

[3] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," in *Trans. 3rd Prague Conf. on Inf. Theory*, 1962, pp. 689–723, English Translation: http://www.math.wustl.edu/~luthy/strassen.pdf.

[4] Y. Polyanskiy, V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[5] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[6] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," 2012, arXiv:1201.3901v2 [cs.IT].

[7] E. MolavianJazi and J. Laneman, "Simpler achievable rate regions for multiaccess with finite blocklength," in *IEEE Int. Symp. Inf. Theory*, Boston, MA, July 2012.

[8] Y. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *IEEE Int. Symp. Inf. Theory*, Boston, MA, July 2012.

[9] Y. Liu and B. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 376–386, March 1996.

[10] W. Hoeffding, "A combinatorial central limit theorem," *Annals Math. Stats.*, vol. 22, no. 4, pp. 558–566, 1951.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[12] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inf. Theory*, 1971, pp. 23–52.

[13] H. Liao, "Multiple-access channels," Ph.D. dissertation, Dept. Elec. Eng. Univ. Hawaii, Honolulu, 1972.

[14] E. MolavianJazi and J. N. Laneman, "A finite-blocklength perspective on gaussian multi-access channels," arXiv:1309.2343v1 [cs.IT].

[15] E. Haim, Y. Kochman, and U. Erez, "A note on the dispersion of network problems," in *IEEE Conv. Elec. Eng. in Israel*, 2012.

[16] P. Moulin, "A new metaconverse and outer region for finite-blocklength MACs," in *Inf. Theory and Apps. Workshop*, San Diego, CA, 2013.

[17] P. Moulin, Personal Communication.

[18] R. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, March 1985.

[19] E. MolavianJazi and J. N. Laneman, "A random coding approach to Gaussian multiple access channels with finite blocklength," in *50th Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, Oct. 2012.

[20] E. Bolthausen, "An estimate of the remainder in a combinatorial central limit theorem," *Prob. Theory and Rel. Fields*, vol. 66, pp. 379–386, 1984.

[21] E. Bolthausen and F. Götze, "The rate of convergence for multivariate sampling statistics," *Annals of Statistics*, vol. 21, no. 4, pp. 1692–1710, 1993.

[22] B. von Bahr, "Remainder term estimate in a combinatorial limit theorem," *Prob. Theory and Rel. Fields*, vol. 35, no. 2, pp. 131–139, 1976.

[23] S. Ho and L. H. Y. Chen, "An L_p bound for the remainder in a combinatorial central limit theorem," *Annals of Probability*, vol. 6, no. 2, pp. 231–249, 1978.

[24] W. Loh, "On Latin hypercube sampling," *Annals of Statistics*, vol. 24, no. 5, pp. 2058–2080, 1996.

[25] V. Tan, "Achievable second-order coding rates for the wiretap channel," in *IEEE Int. Conf. Comm. Sys.*, Singapore, 2012, pp. 65–69.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 2001.

[27] Y. Wu and S. Verdú, "The impact of constellation cardinality on Gaussian channel capacity," in *48th Annual Allerton Conf. on Comm., Control and Comp.*, 2010, pp. 620–628.