

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

Jonathan Scarlett, *Student Member, IEEE*, Alfonso Martinez, *Senior Member, IEEE*,
and Albert Guillén i Fàbregas, *Senior Member, IEEE*

Abstract—This paper studies the second-order asymptotics of coding rates for the discrete memoryless multiple-access channel (MAC) with a fixed target error probability. Using constant-composition random coding, coded time-sharing, and a variant of Hoeffding’s combinatorial central limit theorem, an inner bound on the set of locally achievable second-order coding rates is given for each point on the boundary of the capacity region. It is shown that the inner bound for constant-composition random coding includes that recovered by independent identically distributed random coding, and that the inclusion may be strict. The inner bound is extended to the Gaussian MAC via an increasingly fine quantization of the inputs.

Index Terms—Multiple-access channels, second-order coding rate, channel dispersion, constant-composition random coding, combinatorial central limit theorem.

I. INTRODUCTION

THE channel capacity describes the highest 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 settings. For discrete memoryless channels, the maximum number of codewords of length n yielding an error probability not exceeding $\epsilon \in (0, 1)$, denoted by $M^*(n, \epsilon)$, satisfies [3]

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

where C is the channel capacity, $Q^{-1}(\cdot)$ is the functional inverse of the standard Gaussian tail probability $Q(z) \triangleq \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$, and V is known as the channel

Manuscript received June 4, 2014; accepted October 23, 2014. Date of publication November 14, 2014; date of current version December 22, 2014. This work was supported in part by the European Research Council under Grant 259663, in part by the European Union’s 7th Framework Programme under Grant 303633, and in part by the Spanish Ministry of Economy and Competitiveness under Grant RYC-2011-08150 and Grant TEC2012-38800-C03-03. This paper was presented at the 51st Allerton Conference on Communication, Computing and Control in 2013.

J. Scarlett was with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. He is now with the Laboratory for Information and Inference Systems, École Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland (e-mail: jmscarlett@gmail.com).

A. Martinez is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain (e-mail: alfonso.martinez@ieee.org).

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

Communicated by I. Sason, Associate Editor at Large.

Digital Object Identifier 10.1109/TIT.2014.2371026

dispersion. Expansions of the form (1) provide additional insight into the system performance beyond the capacity alone by quantifying the rate of convergence.

In this paper, we study the second-order asymptotics of the multiple-access channel (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], whereas we demonstrate improved asymptotic bounds via the use of constant-composition random coding [1, Ch. 9]. A key tool in our analysis is a Berry-Esseen theorem associated with a variant of Hoeffding’s combinatorial central limit theorem (CLT) [9]. We consider a local notion of second-order achievability proposed by Nomura and Han [10], in which the second-order coding rates (e.g. $-\sqrt{V}Q^{-1}(\epsilon)$ in (1)) of the users are sought for a fixed point on the boundary of the capacity region.

A. Notation

The set of all probability distributions on an alphabet \mathcal{X} is denoted by $\mathcal{P}(\mathcal{X})$, and the set of conditional distributions on \mathcal{Y} given \mathcal{X} is denoted by $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. Given a distribution $Q(x)$ and a conditional distribution $W(y|x)$, the joint distribution $Q(x)W(y|x)$ is denoted by $Q \times W$. The set of all empirical distributions (i.e. types [11, Ch. 2]) for sequences in \mathcal{X}^n is denoted by $\mathcal{P}_n(\mathcal{X})$. 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 $\mathbf{x} \in T^n(P_X)$ and a conditional distribution $P_{Y|X}$, we define $T_{\mathbf{x}}^n(P_{Y|X})$ to be the set of sequences \mathbf{y} such that $(\mathbf{x}, \mathbf{y}) \in T^n(P_X \times P_{Y|X})$.

Bold symbols are used for vectors and matrices (e.g. \mathbf{x}), and the corresponding i -th entry of a vector is written using a subscript (e.g. x_i). The vectors (or matrices) of all zeros and all ones are denoted by $\mathbf{0}$ and $\mathbf{1}$ respectively, and the identity matrix is denoted by \mathbb{I} ; the sizes will be clear from the context. The symbols \prec , \preceq , etc. denote element-wise inequalities for vectors, and inequalities on the positive semidefinite cone for matrices (e.g. $\mathbf{V} \succ \mathbf{0}$ means \mathbf{V} is positive definite). We denote the ℓ_2 -norm of a vector by $\|\cdot\|$, and the maximum absolute value of the entries of a vector or matrix by $\|\cdot\|_\infty$. 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 matrix of two random vectors by $\text{Cov}[\mathbf{Z}_1, \mathbf{Z}_2] = \mathbb{E}[(\mathbf{Z}_1 - \mathbb{E}[\mathbf{Z}_1])(\mathbf{Z}_2 - \mathbb{E}[\mathbf{Z}_2])^T]$, and we write $\text{Cov}[\mathbf{Z}]$ in place of $\text{Cov}[\mathbf{Z}, \mathbf{Z}]$. The variance of a

scalar random variable is denoted by $\text{Var}[\cdot]$. 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 a set \mathcal{S} of real numbers (or vectors) and a constant (or vector) c , we write $\mathcal{S} + c$ (or $c + \mathcal{S}$) to denote the set $\{s + c : s \in \mathcal{S}\}$. We similarly write $c\mathcal{S} = \{cs : s \in \mathcal{S}\}$ for a given constant c .

For two sequences f_n and g_n , we write $f_n = O(g_n)$ if $|f_n| \leq c|g_n|$ for some c and sufficiently large n , and $f_n = o(g_n)$ if $\lim_{n \rightarrow \infty} \frac{f_n}{g_n} = 0$. We write $f_n = \Theta(g_n)$ if $f_n = O(g_n)$ and $g_n = O(f_n)$.

B. System Setup and Definitions

We consider a two-user discrete memoryless MAC (DM-MAC) $W(y|x_1, x_2)$ with input alphabets \mathcal{X}_1 and \mathcal{X}_2 and output alphabet \mathcal{Y} , yielding an n -letter transition law given by $W^n(y|\mathbf{x}_1, \mathbf{x}_2) \triangleq \prod_{i=1}^n W(y_i|x_{1,i}, x_{2,i})$. The encoders and decoder operate as follows. Encoder $v = 1, 2$ takes as input a message m_v equiprobable on the set $\{1, \dots, M_v\}$, and transmits the corresponding codeword $\mathbf{x}_v^{(m_v)}$ from the codebook $\mathcal{C}_v = \{\mathbf{x}_v^{(1)}, \dots, \mathbf{x}_v^{(M_v)}\}$. The decoder forms an estimate (\hat{m}_1, \hat{m}_2) of the message pair using the output sequence \mathbf{y} and the two codebooks. An error is said to have occurred if $(\hat{m}_1, \hat{m}_2) \neq (m_1, m_2)$. A rate pair (R_1, R_2) is said to be (n, ϵ) -achievable if there exist codebooks with $M_1 \geq e^{nR_1}$ and $M_2 \geq e^{nR_2}$ codewords of length n for users 1 and 2 respectively, such that the average error probability does not exceed ϵ . The capacity region \mathcal{R}^* is defined to be the closure of the set of rate pairs (R_1, R_2) that are (n, ϵ) -achievable for any $\epsilon \in (0, 1)$ and sufficiently large n .

Our results are proved using constant-composition random coding with coded time-sharing [12]. The precise description of the ensemble is postponed until Section IV; here we simply provide the definitions required to state the results. We fix a finite 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 define the joint distribution

$$P_{UX_1X_2Y}(u, x_1, x_2, y) \triangleq Q_U(u)Q_1(x_1|u)Q_2(x_2|u)W(y|x_1, x_2), \quad (2)$$

and denote the induced marginal distributions by $P_{Y|X_1U}$, $P_{Y|U}$, etc. Defining the rate vector

$$\mathbf{R} \triangleq \begin{bmatrix} R_1 \\ R_2 \\ R_1 + R_2 \end{bmatrix} \quad (3)$$

and the mutual information vector (implicitly dependent on Q_U , Q_1 , Q_2 and W)

$$\mathbf{I} \triangleq \begin{bmatrix} I(X_1; Y|X_2, U) \\ I(X_2; Y|X_1, U) \\ I(X_1, X_2; Y|U) \end{bmatrix}, \quad (4)$$

we have [13]–[15]

$$\mathcal{R}^* = \bigcup_U \bigcup_{Q_U, Q_1, Q_2} \left\{ (R_1, R_2) : \mathbf{R} \preceq \mathbf{I} \right\}. \quad (5)$$

Moreover, the union over \mathcal{U} may be restricted to satisfy $|\mathcal{U}| \leq 2$. The three conditions in the element-wise inequality

$\mathbf{R} \preceq \mathbf{I}$ correspond to a treatment of the error event as a union of three error types:

$$\begin{aligned} (\text{Type 1}) \quad & \hat{m}_1 = m_1 \text{ and } \hat{m}_2 \neq m_2, \\ (\text{Type 2}) \quad & \hat{m}_1 \neq m_1 \text{ and } \hat{m}_2 = m_2, \\ (\text{Type 12}) \quad & \hat{m}_1 \neq m_1 \text{ and } \hat{m}_2 \neq m_2. \end{aligned}$$

A key quantity in our analysis is the information density vector [6], [8]

$$\mathbf{i}(u, x_1, x_2, y) \triangleq \begin{bmatrix} i_1(u, x_1, x_2, y) \\ i_2(u, x_1, x_2, y) \\ i_{12}(u, x_1, x_2, y) \end{bmatrix}, \quad (6)$$

where

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

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

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

Averaging \mathbf{i} with respect to the distribution in (2) yields the mutual information vector in (4).

We consider the local notion of second-order asymptotics introduced by Nomura and Han for the Slepian-Wolf problem [10]; see also Hayashi [5] for the analogous definitions in the single-user setting. We proceed by presenting similar definitions for the present setting, albeit in a slightly different form. A pair (L_1, L_2) is said to be $(n, \epsilon, R_1^*, R_2^*)$ -achievable if there exist codebooks with $M_v \geq e^{nR_v^* + \sqrt{n}L_v}$ codewords of length n for $v = 1, 2$ such that the average error probability p_e does not exceed ϵ . The *second-order rate region* $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ is defined as the closure of the set of pairs (L_1, L_2) that are $(n, \epsilon, R_1^*, R_2^*)$ -achievable for sufficiently large n . In other words, $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ is the set of all (L_1, L_2) pairs for which there exists an ϵ -reliable code with $\log M_v = nR_v^* + \sqrt{n}L_v + o(\sqrt{n})$ for $v = 1, 2$. While this definition is valid for any pair (R_1^*, R_2^*) , our focus will be on pairs on the boundary of the capacity region \mathcal{R}^* ; in all other cases we have either $\mathcal{L} = \emptyset$ or $\mathcal{L} = \mathbb{R}^2$ due to the fact that the strong converse holds [31]. We will see that both negative and positive values of L_v arise; the former can be thought of as a backoff from the first-order term, and the latter an addition to the first-order term.

Finally, we define the set

$$\mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon) \triangleq \{\mathbf{z} \in \mathbb{R}^d : \mathbb{P}[\mathbf{Z} \preceq \mathbf{z}] \geq 1 - \epsilon\}, \quad (10)$$

where $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{V})$, and \mathbf{V} is a $d \times d$ positive semi-definite matrix. This definition applies for an arbitrary dimension d , which is dictated by the first argument.

C. Previous Work

The second-order rate region \mathcal{L} has been characterized for very few multi-user problems [10], [16], [17]. The one most relevant to this paper is the Gaussian MAC with degraded message sets [16], which has the notable feature of having a curved capacity region, giving rise to a non-standard derivative

term in the expression for \mathcal{L} . Our analysis will yield similar terms using different techniques.

Tan and Kosut [6] and Haim *et al.* [18] performed second-order asymptotic studies for various multi-user problems using different notions of achievability to those above. In particular, both of these works considered the problem of finding the backoff from the rates when a point (R_1^*, R_2^*) is approached from a given angle. As demonstrated in [16], this problem can be solved numerically in a straightforward fashion once $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ is characterized.

Other previous works on the DM-MAC have taken an alternative approach to characterizing the second-order asymptotics, namely, seeking *global* asymptotic expansions of the following form: For any triplet (Q_U, Q_1, Q_2) , rate vectors \mathbf{R} satisfying

$$n\mathbf{R} \in n\mathbf{I} - \sqrt{n}\mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon) + g(n)\mathbf{1}, \quad (11)$$

are (n, ϵ) -achievable for some *dispersion matrix* \mathbf{V} and function $g(n) = o(\sqrt{n})$, where \mathbf{I} is given in (4).

The first global result for the DM-MAC was given in [6], where i.i.d. random coding was used to obtain (11) with $\mathbf{I} = \mathbb{E}[\mathbf{i}(U, X_1, X_2, Y)]$ and $\mathbf{V} = \text{Cov}[\mathbf{i}(U, X_1, X_2, Y)]$ (see also [7]). Expansions of a similar form were given by MolavianJazi and Laneman [7]. By using a constant-composition time-sharing sequence, Huang and Moulin [8] showed that the dispersion matrix can be improved to

$$\mathbf{V}^{\text{iid}} \triangleq \mathbb{E}[\text{Cov}[\mathbf{i}(U, X_1, X_2, Y) | U]]. \quad (12)$$

As discussed by Haim *et al.* [18], expansions of the form (11) are more difficult to interpret than the scalar counterpart in (1), as the notion of the convergence of a region is inherently less concrete than that of the convergence of a scalar. While the scalar dispersion V in (1) corresponds to a concrete operational definition [4], it appears difficult to directly give any such meaning to the matrix \mathbf{V} based on global results. In fact, non-global asymptotic studies in [6] indicate that, in most cases of interest, entries (1, 2) and (2, 1) of the matrix do not play a fundamental role in characterizing the performance. Furthermore, it may be difficult to compare two dispersion matrices, since the partial positive definite ordering does not guarantee that at least one of $V_1 \leq V_2$ or $V_2 \leq V_1$ hold. These issues are even more troublesome when one considers the union over all input distributions; for example, standard proofs often yield a non-uniform remainder term $g(n)$ in (11). These limitations motivate the study of local asymptotics, such as $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ defined above. However, global results often prove useful as an intermediate step towards the local results.

D. Contributions

The main result of this paper is an inner bound on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ for the discrete memoryless MAC. The result is proved using constant-composition random coding and a variant of Hoeffding's combinatorial CLT [9], [19]. Since coding with fixed input distributions (not varying with n) is not sufficient to achieve all (L_1, L_2) pairs in network information theory problems [16], we apply coded

time-sharing [15, Sec. 4.5.3] between input distributions corresponding to two points on the boundary of the capacity region, with one of the points only corresponding to a fraction $O(\frac{1}{\sqrt{n}})$ of the block length. Several examples are provided, including (i) a case where constant-composition random coding yields a strictly larger inner bound than that of i.i.d. random coding, and (ii) an application to the Gaussian MAC via a quantization argument.

II. MAIN RESULT

A. Further Definitions

Our main result is written in terms of a dispersion matrix of the form

$$\mathbf{V} \triangleq \mathbb{E} \left[\text{Cov}[\mathbf{i}(U, X_1, X_2, Y) | U] - \text{Cov}[\mathbf{i}^{(1)}(U, X_1) | U] - \text{Cov}[\mathbf{i}^{(2)}(U, X_2) | U] \right], \quad (13)$$

where

$$\mathbf{i}^{(1)}(u, x_1) \triangleq \mathbb{E}[\mathbf{i}(U, X_1, X_2, Y) | (U, X_1) = (u, x_1)] \quad (14)$$

$$\mathbf{i}^{(2)}(u, x_2) \triangleq \mathbb{E}[\mathbf{i}(U, X_1, X_2, Y) | (U, X_2) = (u, x_2)]. \quad (15)$$

We can interpret (13) as follows: The term $\text{Cov}[\mathbf{i}]$ represents the variations in (X_1, X_2, Y) in the i.i.d. case (cf. (12)), and the terms $\text{Cov}[\mathbf{i}^{(1)}]$ and $\text{Cov}[\mathbf{i}^{(2)}]$ represent the reduced variations in X_1 and X_2 respectively, resulting from the codewords having a fixed composition. Since all covariance matrices are positive semidefinite, we clearly have $\mathbf{V} \leq \mathbf{V}^{\text{iid}}$. We henceforth write the entries of \mathbf{I} (see (4)) and \mathbf{V} using subscripts:

$$\mathbf{I} = \begin{bmatrix} I_1 \\ I_2 \\ I_{12} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} V_1 & V_{1,2} & V_{1,12} \\ V_{1,2} & V_2 & V_{2,12} \\ V_{1,12} & V_{2,12} & V_{12} \end{bmatrix}. \quad (16)$$

For a fixed point (R_1^*, R_2^*) on the boundary of \mathcal{R}^* , we let $\hat{\mathbf{T}}_- \triangleq \hat{\mathbf{T}}_-(R_1^*, R_2^*)$ and $\hat{\mathbf{T}}_+ \triangleq \hat{\mathbf{T}}_+(R_1^*, R_2^*)$ respectively denote the left and right unit tangent vectors along the boundary of \mathcal{R}^* in (R_1, R_2) -space; see Figure 1. We let $\hat{\mathbf{T}}_-$ (respectively, $\hat{\mathbf{T}}_+$) be undefined when $R_1^* = 0$ (respectively, $R_2^* = 0$); in all other cases, the vectors are well-defined due to the convexity of the capacity region. The case $\hat{\mathbf{T}}_- = -\hat{\mathbf{T}}_+$ corresponds to a curved or straight-line part of the boundary, whereas $\hat{\mathbf{T}}_- \neq -\hat{\mathbf{T}}_+$ corresponds to a sudden change in slope (e.g. at a corner point).

We construct the following vectors in the same way as (3):

$$\mathbf{T}_- \triangleq \begin{bmatrix} \hat{T}_{-,1} \\ \hat{T}_{-,2} \\ \hat{T}_{-,1} + \hat{T}_{-,2} \end{bmatrix}, \quad \mathbf{T}_+ \triangleq \begin{bmatrix} \hat{T}_{+,1} \\ \hat{T}_{+,2} \\ \hat{T}_{+,1} + \hat{T}_{+,2} \end{bmatrix}, \quad (17)$$

where $\hat{T}_{(\cdot),i}$ denotes the i -th entry of the corresponding unit tangent vector. To ease some of the subsequent discussions, we define the following scalars that correspond to $\hat{\mathbf{T}}_-$ and $\hat{\mathbf{T}}_+$ in a one-to-one fashion:

$$D_- \triangleq \frac{\hat{T}_{-,2}}{\hat{T}_{-,1}}, \quad D_+ \triangleq \frac{\hat{T}_{+,2}}{\hat{T}_{+,1}}. \quad (18)$$

These are the left and right derivatives of R_2^* as a function of R_1^* . They are non-positive, and are understood to equal $-\infty$

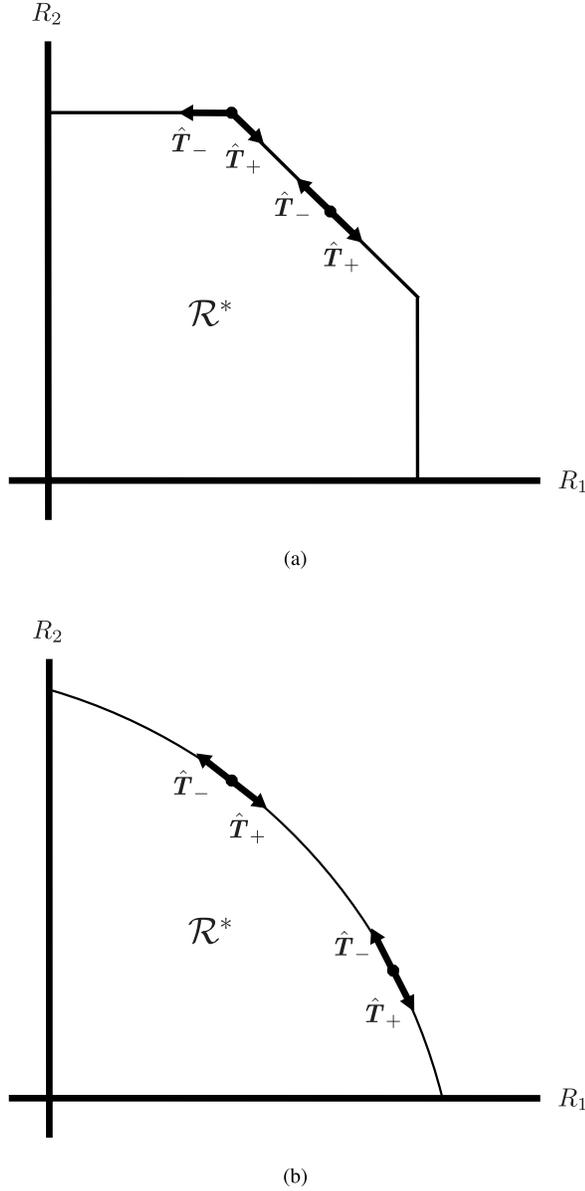


Fig. 1. Illustration of the vectors \hat{T}_- and \hat{T}_+ for various boundary points of two hypothetical capacity regions. (a) Pentagonal region. (b) Curved region.

when $\hat{T}_{(\cdot),1} = 0$, corresponding to a vertical part of the capacity region. Observe that \hat{T}_- is obtained by normalizing $[-1 - D_-]$ and \hat{T}_+ is obtained by normalizing $[1 D_+]$, and hence $\hat{T}_- = -\hat{T}_+$ if and only if $D_- = D_+$.

Given the pairs (R_1^*, R_2^*) and (L_1, L_2) , we define

$$\mathbf{R}^* \triangleq \begin{bmatrix} R_1^* \\ R_2^* \\ R_1^* + R_2^* \end{bmatrix}, \quad \mathbf{L} \triangleq \begin{bmatrix} L_1 \\ L_2 \\ L_1 + L_2 \end{bmatrix}. \quad (19)$$

For a non-empty index set $\mathcal{K} \subseteq \{1, 2, 12\}$, we let $\mathbf{L}^{(\mathcal{K})}$ denote the subvector of \mathbf{L} where only the indices corresponding to \mathcal{K} are kept, and similarly for $\mathbf{R}^{(\mathcal{K})}$, $\mathbf{R}^{*(\mathcal{K})}$, $\mathbf{T}_-^{(\mathcal{K})}$, $\mathbf{T}_+^{(\mathcal{K})}$ and $\mathbf{I}^{(\mathcal{K})}$. Similarly, $\mathbf{V}^{(\mathcal{K})}$ denotes the $|\mathcal{K}| \times |\mathcal{K}|$ submatrix of \mathbf{V} where only the rows and columns indexed by \mathcal{K} are kept.

B. Statement of Main Result

We say that the triplet (Q_U, Q_1, Q_2) achieves the rate pair (R_1, R_2) if $\mathbf{R} \preceq \mathbf{I}$; from (5), every point in \mathcal{R}^* (including

those on the boundary) is achieved by at least one such triplet. In the following theorem, \mathcal{K} can be thought of as the set of error types that are active for a given input distribution and boundary point (e.g. if the boundary point is achieved by the corner point of the pentagonal region corresponding to the type-2 and type-12 conditions, then $\mathcal{K} = \{2, 12\}$).

Theorem 1: Fix $\epsilon \in (0, 1)$, let (R_1^*, R_2^*) be a point on the boundary of the capacity region \mathcal{R}^* in (5), let (Q_U, Q_1, Q_2) be an arbitrary triplet achieving that point, and consider \mathbf{I} and \mathbf{V} in (4) and (13) respectively. Letting $\mathcal{K} \subseteq \{1, 2, 12\}$ be the set of indices of the largest cardinality such that $\mathbf{R}^{*(\mathcal{K})} = \mathbf{I}^{(\mathcal{K})}$, we have

$$\begin{aligned} \mathcal{L}(\epsilon, R_1^*, R_2^*) &\supseteq \left\{ (L_1, L_2) : \mathbf{L}^{(\mathcal{K})} \in \bigcup_{\beta \geq 0} \left\{ \beta \mathbf{T}_-^{(\mathcal{K})} - \mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon) \right\} \right\} \\ &\cup \left\{ (L_1, L_2) : \mathbf{L}^{(\mathcal{K})} \in \bigcup_{\beta \geq 0} \left\{ \beta \mathbf{T}_+^{(\mathcal{K})} - \mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon) \right\} \right\}, \end{aligned} \quad (20)$$

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

Proof: See Section IV-A. ■

We make the following remarks on Theorem 1:

- 1) In the case that $\hat{T}_- = -\hat{T}_+$, or equivalently $D_- = D_+$ (i.e. a curved or straight-line part of the boundary), the two sets in (20) can be combined into a single set containing a coefficient $\beta \in \mathbb{R}$ (with negative values allowed) and the vector $\mathbf{T}_+^{(\mathcal{K})}$. In this case, the inner bound on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ is a half-space. We will see in Section III-B that this does not always occur, and combinations other than $(D_-, D_+) = (0, -1)$ and $(D_-, D_+) = (-1, -\infty)$ are possible (these are the combinations that are observed for standard pentagonal regions).
- 2) In the case that \mathcal{K} contains only a single entry $\nu \in \{1, 2, 12\}$, the unions over β can be replaced by $\beta = 0$, yielding a simpler inner bound given by

$$\mathcal{L}(\epsilon, R_1^*, R_2^*) \supseteq \left\{ (L_1, L_2) : L_\nu \leq -\sqrt{V_\nu} \mathbf{Q}^{-1}(\epsilon) \right\}, \quad (21)$$

where $L_{12} \triangleq L_1 + L_2$. The fact that $\beta = 0$ suffices is shown in the same way for each ν , so we consider the case $\nu = 12$. Since (R_1^*, R_2^*) lies on the diagonal part of the pentagonal region corresponding to (Q_U, Q_1, Q_2) (and away from the corners), both $(R_1^* - \delta, R_2^* + \delta)$ and $(R_1^* + \delta, R_2^* - \delta)$ are achievable for sufficiently small δ , and hence $D_- \in [-\infty, -1]$ and $D_+ \in [-1, 0]$. The convexity of the capacity region implies that $D_+ \leq D_-$, and it follows that $D_- = D_+ = -1$. From (17), we see that $D_- = D_+ = -1$ implies $\mathbf{T}_-^{(\mathcal{K})} = \mathbf{T}_+^{(\mathcal{K})} = 0$, and hence each coefficient β in (20) is multiplying zero.

- 3) More generally, if the input distribution achieving (R_1^*, R_2^*) also achieves all of the boundary points in a neighborhood of (R_1^*, R_2^*) , then the unions over β can be replaced by $\beta = 0$. In particular, this is true when the entire capacity region is achieved by a single

input distribution. This will be observed for the Gaussian MAC in Section III-C.

- 4) All non-empty subsets \mathcal{K} of $\{1, 2, 12\}$ can occur with the exception of $\{1, 2\}$. Focusing on the case that time-sharing is absent, the case $\mathcal{K} = \{1, 2\}$ is impossible since

$$I(X_1, X_2; Y) = I(X_1; Y) + I(X_2; Y|X_1) \quad (22)$$

$$\leq I(X_1; X_2, Y) + I(X_2; Y|X_1) \quad (23)$$

$$= I(X_1; Y|X_2) + I(X_2; Y|X_1), \quad (24)$$

where (24) follows since X_1 and X_2 are independent. Whenever \mathcal{K} includes $\{1, 2\}$, we have $R_1^* = I(X_1; Y|X_2)$ and $R_2^* = I(X_2; Y|X_1)$, and it follows from (24) that $R_1^* + R_2^* = I(X_1, X_2; Y)$. Therefore, we have $\mathcal{K} = \{1, 2, 12\}$, corresponding to a rectangular achievable rate region.

- 5) The inner bound in (20) is of a similar form to the set appearing in [16, Thm. 3] for the Gaussian MAC with degraded message sets. The main differences are (i) The left and right tangent vectors are treated separately here, since unlike in [16], the two do not have the same slope in general (e.g. see Figure 1 and the example in Section III-B); (ii) There are six cases here corresponding to the different subsets \mathcal{K} of $\{1, 2, 12\}$ (see the previous item), whereas in [16] there are only two possibilities for the set of active rate conditions.
- 6) The proof of Theorem 1 can be followed using i.i.d. codeword distributions, yielding an analogous result with \mathbf{V}^{iid} (see (12)) in place of \mathbf{V} . Using the fact that $\mathbf{V} \leq \mathbf{V}^{\text{iid}}$, it is not difficult to show that the inner bounds on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ obtained using \mathbf{V} include those obtained using \mathbf{V}^{iid} whenever $\epsilon < \frac{1}{2}$. In Section III-A, we will see that the inclusion can be strict.
- 7) It is also of interest to compare \mathbf{V} to a hypothetical dispersion matrix of the form

$$\mathbf{V}^{\text{joint}} \triangleq \mathbb{E}[\text{Cov}[\mathbf{i}(U, X_1, X_2, Y) | U, X_1, X_2]]. \quad (25)$$

This is the matrix that would be obtained if the *joint* composition of (U, X_1, X_2) were fixed, which is impossible in the absence of cooperation between the users. As we show in [20], we have $\mathbf{V}^{\text{joint}} \leq \mathbf{V}$; this is proved using the matrix version of the law of total variance, along with the identity $\mathbb{E}[\mathbf{Z}\mathbf{Z}^T] \succeq \mathbb{E}[\mathbf{Z}]\mathbb{E}[\mathbf{Z}^T]$.

- 8) We make no attempt to present analytical expressions for $\hat{\mathbf{T}}_-$ and $\hat{\mathbf{T}}_+$, but two numerical approaches to their computation are presented in Section III. In general, if the capacity region is characterized numerically, then one can easily obtain numerical bounds or approximations for these tangent vectors.

III. EXAMPLES

A. The Collision Channel

We begin with a simple deterministic example that will permit us to compare i.i.d. and constant-composition random coding, and to discuss the role of the off-diagonal entries of the corresponding dispersion matrices.

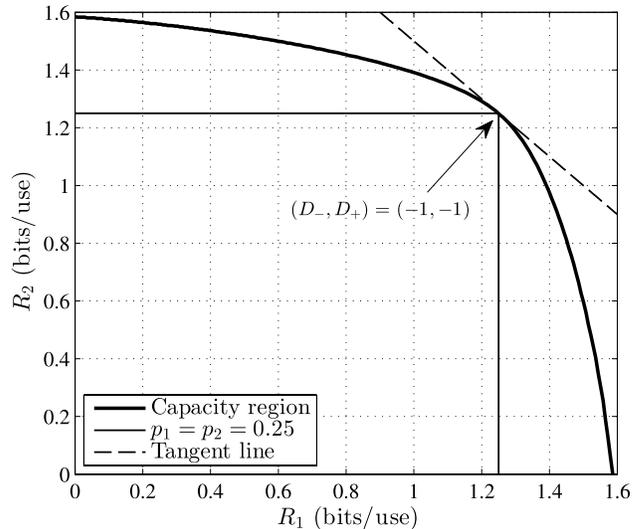


Fig. 2. Capacity region of the collision channel. The tangent line corresponds to the boundary point $(R_1^*, R_2^*) = (1.25, 1.25)$, and we have $D_- = D_+ = -1$.

Setting $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1, 2\}$ and $\mathcal{Y} = \{(0, 0), (0, 1), (0, 2), (1, 0), (2, 0), \mathbf{c}\}$, the channel is given by

$$W(y|x_1, x_2) = \begin{cases} 1 & y = (x_1, x_2) \text{ and } \min\{x_1, x_2\} = 0 \\ 1 & y = \mathbf{c} \text{ and } \min\{x_1, x_2\} \neq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (26)$$

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 \mathbf{c} , meaning ‘‘collision’’.

We recall the following observations by Gallager [21]: (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 by input distributions of the form $Q_1 = (1 - 2p_1, p_1, p_1)$ and $Q_2 = (1 - 2p_2, p_2, p_2)$, where $\mathcal{U} = \emptyset$; (iii) The achievable rate region corresponding to any such (Q_1, Q_2) pair is rectangular. The capacity region is shown in Figure 2. The left and right tangent vectors coincide (i.e. $\hat{\mathbf{T}}_- = -\hat{\mathbf{T}}_+$ and hence $D_- = D_+ \triangleq D$) at all boundary points (R_1^*, R_2^*) with $R_1^* > 0$ and $R_2^* > 0$, and the case of interest in Theorem 1 is $\mathcal{K} = \{1, 2, 12\}$.

One approach to computing the inner bound in (20) for a given boundary point (R_1^*, R_2^*) is to first find the pair (p_1, p_2) achieving that point, and then calculate $\hat{\mathbf{T}}_-$ and $\hat{\mathbf{T}}_+$ (see the example in Section III-B). In this example, the reverse approach turns out to be more convenient: We start with a given derivative $D < 0$, and perform an optimization over (p_1, p_2) to find the corresponding (unique) boundary point (R_1^*, R_2^*) .

As stated above, the achievable rate region for a given pair (p_1, p_2) is a rectangle with a corner point given by (I_1, I_2) . The straight line of slope D passing through this point is given by $R_2 = D(R_1 - I_1) + I_2$. Thus, finding the desired boundary point (R_1^*, R_2^*) simply amounts to maximizing $I_2 - DI_1$ with

¹On the other hand, for the collision channel with K non-zero symbols, time-sharing is required for $K \geq 8$ [21].

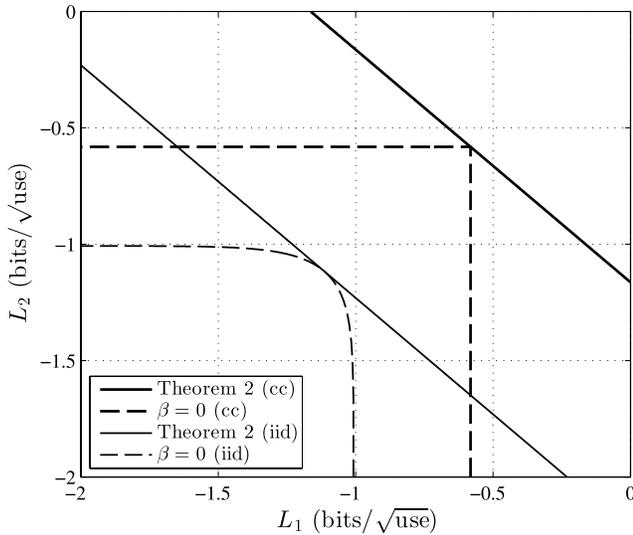


Fig. 3. Boundaries of the inner bounds on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ for the collision channel, with $R_1^* = R_2^* = 1.25$ and $\epsilon = 0.01$. The regions are to the bottom-left of the boundaries shown.

respect to (p_1, p_2) , which is a straightforward optimization problem.

For concreteness, we provide a numerical example for the case that $D = -1$, which corresponds to $\hat{T}_- = -\hat{T}_+ = \left[\frac{-1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}\right]^T$. Using a brute force search to three decimal places, we found the optimal parameters to be $p_1 = p_2 = 0.25$, yielding $R_1^* = R_2^* = 1.25$ bits/use. The inner bound on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ from Theorem 1, and its counterpart for i.i.d. random coding (cf. (12)), are shown in Figure 3, where we set $\epsilon = 0.01$. For comparison, we also plot the weaker inner bounds in which β is set to zero and only $-\mathbf{Q}_{\text{inv}}$ remains (cf. (20)). The boundaries of these regions are shown, and the regions lie to the bottom-left of these boundaries.

We see that the region resulting from constant-composition random coding is strictly larger than that resulting from i.i.d. random coding. In this example, the strict inclusion holds for all points on the boundary of the capacity region other than the endpoints corresponding to $R_1^* = 0$ or $R_2^* = 0$. This gain is analogous to a similar gain in the error exponent [12]. In contrast, in the single-user setting, the two ensembles yield the same second-order term and error exponent after the optimization of the input distribution [4], [22].

We conclude by discussing the roles of the various entries of the covariance matrices. In this example, the diagonal entries V_1 and V_2 determine the locations of the vertical and horizontal asymptotes in Figure 3. We see from Figure 3 that the off-diagonal terms also play a role. In particular, the rectangular shape of the region with $\beta = 0$ for constant-composition coding is an extreme case corresponding to a singular dispersion matrix \mathbf{V} , and this is in fact the most favorable shape possible (in terms of enlarging $\mathcal{L}(\epsilon; R_1^*, R_2^*)$) given fixed locations of the vertical and horizontal asymptotes. In contrast, the off-diagonal terms of \mathbf{V}^{iid} yield a more standard curved region for $\beta = 0$, which is less favorable. Thus, at least in this example, the enlarged second-order region for constant-composition codes is not only due to the smaller

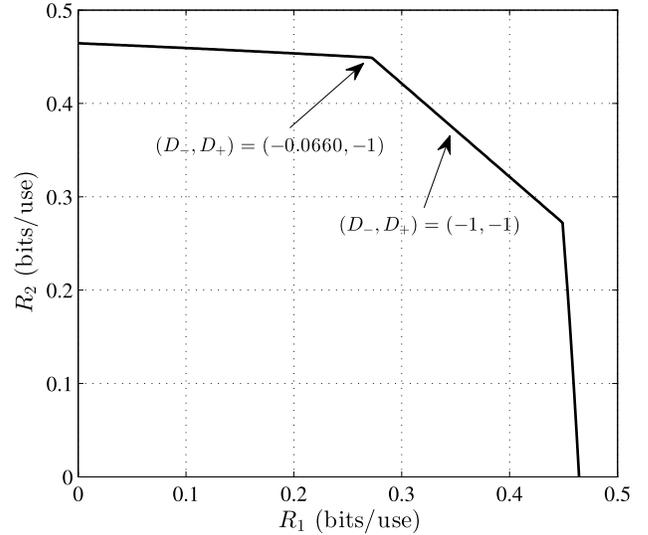


Fig. 4. Capacity region of the channel described by (27), and example values of (D_-, D_+) .

diagonal entries, but also due to a more favorable covariance matrix structure.

B. A Non-Deterministic Example

Here we provide an example showing that the two unions in (20) cannot, in general, be combined into one. In other words, it is necessary to consider the left and right tangent vectors separately. We set $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \{0, 1\}$, and

$$W(y|x_1, x_2) = \begin{cases} 1 & x_1 = x_2 = y \\ 0.5 & x_1 \neq x_2 \\ 0 & \text{otherwise.} \end{cases} \quad (27)$$

Thus, the channel is noiseless if $x_1 = x_2$, and completely noisy if $x_1 \neq x_2$. We write the input distributions as $Q_1 = (1 - p_1, p_1)$ and $Q_2 = (1 - p_2, p_2)$.

The capacity region is shown in Figure 4. Observe that there are two ‘‘corner points’’, but unlike those of standard pentagonal regions, neither of them corresponds to a change in angle of 45 degrees. More precisely, the middle segment shown in the plot has slope -1 , but the other two segments are neither vertical nor horizontal (in fact, they are not even straight line segments, even though they may appear to be). Both corner points are achieved by $p_1 = p_2 = 0.5$.

Here we focus on characterizing the set $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ for the upper corner point $(R_1^*, R_2^*) = (0.272, 0.449)$; identical arguments apply for the lower corner point. The case of interest in Theorem 1 is $\mathcal{K} = \{2, 12\}$. Since the middle segment in Figure 4 has slope -1 , we have $\hat{T}_+ = \left[\frac{1}{\sqrt{2}} \quad \frac{-1}{\sqrt{2}}\right]$. The idea used to compute \hat{T}_- is to shift the point $(p_1, p_2) = (0.5, 0.5)$ by a small amount $(\Delta \cos \theta, \Delta \sin \theta)$, and observe the behavior of the corner point $(I_{12} - I_2, I_2)$ for $\theta \in [0, 2\pi)$, where for each θ we are interested in the limiting behavior as $\Delta \rightarrow 0$. Making the dependence of I_2 and I_{12} on (p_1, p_2)

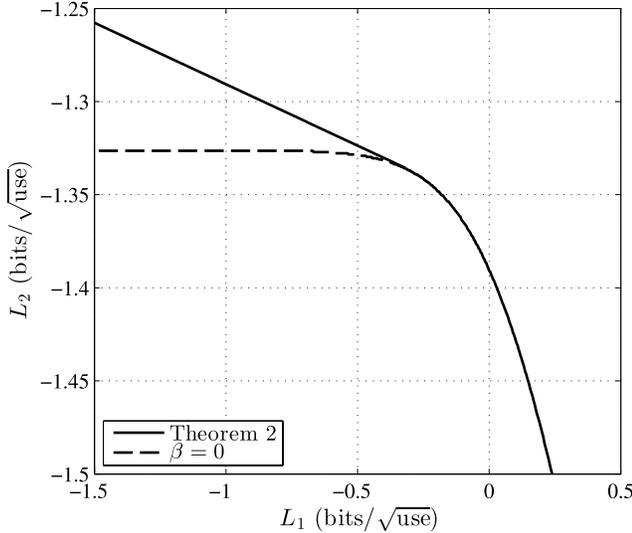


Fig. 5. Boundaries of the inner bounds on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ for the channel described by (27), with $(R_1^*, R_2^*) = (0.272, 0.449)$ and $\epsilon = 0.01$. The regions are to the bottom-left of the boundaries shown.

explicit, a second-order Taylor expansion yields

$$I_\nu(p_1 + \Delta \cos \theta, p_2 + \Delta \sin \theta) = I_\nu(p_1, p_2) + f_\nu(\theta) \Delta^2 + O(\Delta^3), \quad (28)$$

for $\nu = 2, 12$, where

$$f_\nu(\theta) \triangleq \begin{bmatrix} \cos \theta & \sin \theta \end{bmatrix} \begin{bmatrix} \frac{\partial^2 I_\nu}{\partial p_1^2} & \frac{\partial^2 I_\nu}{\partial p_1 \partial p_2} \\ \frac{\partial^2 I_\nu}{\partial p_1 \partial p_2} & \frac{\partial^2 I_\nu}{\partial p_2^2} \end{bmatrix} \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}. \quad (29)$$

Note that the first-order term in (28) is absent, since the derivatives $\frac{\partial I_\nu}{\partial p_1}$ and $\frac{\partial I_\nu}{\partial p_2}$ are zero at $(p_1, p_2) = (0.5, 0.5)$ for $\nu = 2, 12$. We conclude from (28) that for a fixed choice of θ , $(I_{12} - I_2, I_2)$ moves in the direction $(f_{12}(\theta) - f_2(\theta), f_2(\theta))$ in the limit as $\Delta \rightarrow 0$. Evaluating the direction for 10000 equally spaced angles over the range $\theta \in [0, 2\pi]$, we obtained $\hat{T}_- = [-0.99783 \ 0.06586]$, corresponding to $\theta = 0.221$ radians and $D_- = -0.0660$.

Figure 5 shows the resulting inner bound on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ given in Theorem 1, with $\epsilon = 0.01$. In this example, the region is identical for i.i.d. random coding and constant-composition random coding. It is interesting to observe the different shape of the region compared to the previous example, resulting from the differing left and right tangent vectors. It is only the former that plays a role in enlarging the achievable region, since the set $-\mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon)$ (with $\mathcal{K} = \{2, 12\}$) already satisfies the property that if a given point is in the set, so are all points on the right-hand side of the line of slope $D_+ = -1$ passing through that point.

C. Gaussian Multiple-Access Channel

We have focused our attention on the DM-MAC, which permits an analysis based on combinatorial arguments. We now discuss how Theorem 1 can be extended to the Gaussian MAC via an increasingly fine quantization of the inputs, similarly to

Hayashi [5, Th. 3] and Tan [23]. Each use of the channel is described by

$$Y = \sqrt{P_1} X_1 + \sqrt{P_2} X_2 + Z, \quad (30)$$

where $Z \sim N(0, 1)$, and where each codeword \mathbf{x}_ν for user $\nu = 1, 2$ is constrained to satisfy $\frac{1}{n} \|\mathbf{x}_\nu\|^2 \leq 1$. The quantities P_1 and P_2 represent the signal-to-noise ratios for users 1 and 2 respectively.

The capacity region is pentagonal [24, Sec. 15.1], and is achieved using Gaussian input distributions, namely $Q_1, Q_2 \sim N(0, 1)$. The quantities \mathbf{I} and \mathbf{V} in (4) and (13) can be written explicitly; for $\nu = 1, 2, 12$, we have

$$I_\nu = \frac{1}{2} \log(1 + P_\nu) \quad (31)$$

with $P_{12} \triangleq P_1 + P_2$. Moreover, for $\nu = 1, 2$, we have

$$V_\nu = \frac{P_\nu(2 + P_\nu)}{2(1 + P_\nu)^2} \quad (32)$$

$$V_{\nu,12} = \frac{P_\nu(2 + P_1 + P_2)}{2(1 + P_\nu)(1 + P_1 + P_2)}, \quad (33)$$

and the remaining entries of \mathbf{V} are given by

$$V_{12} = \frac{(P_1 + P_2)(2 + P_1 + P_2) + 2P_1P_2}{2(1 + P_1 + P_2)^2} \quad (34)$$

$$V_{1,2} = \frac{P_1P_2}{2(1 + P_1)(1 + P_2)}. \quad (35)$$

A brief outline of how these expressions are obtained is given in Appendix B.

We claim that the following inner bound holds using the notation of Theorem 1 (with the additional condition that the codewords must satisfy the above power constraints in the definition of \mathcal{L} in Section I-B):

$$\mathcal{L}(\epsilon, R_1^*, R_2^*) \supseteq \left\{ (L_1, L_2) : \mathbf{L}^{(\mathcal{K})} \in -\mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon) \right\}. \quad (36)$$

This result was first derived by MolavianJazi and Laneman [25], who used random coding according to the uniform distribution over the surface of a sphere. The techniques of this paper provide an alternative approach to deriving the result. Extending Theorem 1 accordingly is non-trivial, but it is done using well-established techniques; we provide an outline in Appendix B. In contrast with Theorem 1, no form of time-sharing is used in the proof, and no tangent vectors appear in (36). This is due to the fact that every point on the boundary of the capacity region is simultaneously achieved by Gaussian inputs.

IV. PROOF OF THEOREM 1

The random-coding ensemble used in the proof depends on two triplets (Q_U, Q_1, Q_2) and (Q'_U, Q'_1, Q'_2) of probability distributions on the same alphabets. We define $P'_{U'X_1X_2Y}$, \mathbf{i}' , $\mathbf{i}'^{(1)}$, $\mathbf{i}'^{(2)}$, \mathbf{I}' and \mathbf{V}' in the same way as Sections I-B and II-A, with (Q'_U, Q'_1, Q'_2) replacing (Q_U, Q_1, Q_2) . In particular, we have

$$\mathbf{I}' \triangleq \mathbb{E}[\mathbf{i}'(U', X'_1, X'_2, Y')] \quad (37)$$

$$\mathbf{V}' \triangleq \mathbb{E}[\text{Cov}[\mathbf{i}'(U', X'_1, X'_2, Y') | U'] - \text{Cov}[\mathbf{i}'^{(1)}(U', X'_1) | U'] - \text{Cov}[\mathbf{i}'^{(2)}(U', X'_2) | U']], \quad (38)$$

where $(U', X'_1, X'_2, Y') \sim P'_{UX_1X_2Y}$. As an intermediate step towards obtaining our local result, we present the following global result.

Theorem 2: Fix any finite time-sharing alphabet \mathcal{U} and the input distributions (Q_U, Q_1, Q_2) and (Q'_U, Q'_1, Q'_2) . For any $\beta \geq 0$ and $\epsilon \in (0, 1)$, there exists a function $g(n) = O(n^{\frac{1}{4}})$ (depending on (Q_U, Q_1, Q_2) , (Q'_U, Q'_1, Q'_2) , β and ϵ) such that all rate pairs (R_1, R_2) satisfying

$$n\mathbf{R} \in n\mathbf{I} + \sqrt{n}(\beta(\mathbf{I}' - \mathbf{I}) - \mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon)) + g(n)\mathbf{1} \quad (39)$$

are (n, ϵ) -achievable.

Proof: For clarity of exposition, we present the proof in two parts. In Section IV-B, we handle the case that $\beta = 0$ and $\mathcal{U} = \emptyset$. The changes required to handle the general case are presented in Section IV-C. ■

In the case that $\beta = 0$, Theorem 2 gives the second-order asymptotics for a fixed triplet (Q_U, Q_1, Q_2) . In this case, the proof reveals that the behavior $g(n) = O(n^{\frac{1}{4}})$ can be strengthened to $g(n) = O(\log n)$ if the argument to the expectation in (13) has full rank for all u , and $g(n) = O(n^{\frac{1}{6}})$ more generally.

The case $\beta > 0$ is proved by applying an extended version of coded time-sharing [15, Sec. 4.5.3] in which a fraction $1 - \frac{\beta}{\sqrt{n}}$ of the symbols are generated using (Q_U, Q_1, Q_2) , and the remaining symbols are generated using (Q'_U, Q'_1, Q'_2) .

A. Proof of Theorem 1 Based on Theorem 2

Throughout the proof, we write $R_{12} \triangleq R_1 + R_2$ and $R_{12}^* \triangleq R_1^* + R_2^*$ (recall that (R_1^*, R_2^*) is the boundary point of interest). We have from the definition of \mathcal{K} that $R_v^* < I_v$ for all $v \in \mathcal{K}^c$, and we can thus weaken (39) to

$$n\mathbf{R} \in n\mathbf{R}^* + n\delta\mathbf{1}_{\mathcal{K}^c} + \sqrt{n}(\beta(\mathbf{I}' - \mathbf{I}) - \mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon)) + O(n^{\frac{1}{4}})\mathbf{1}, \quad (40)$$

where $\delta \triangleq \min_{v \in \mathcal{K}^c} I_v - R_v^* > 0$, and $\mathbf{1}_{\mathcal{K}^c}$ contains ones at the indices corresponding to \mathcal{K}^c , and zeros elsewhere.

Let \mathbf{z} be a 3×1 vector, and let $\mathbf{z}^{(\mathcal{K})}$ and $\mathbf{z}^{(\mathcal{K}^c)}$ be the corresponding subvectors indexed by the superscript. From the definition of \mathbf{Q}_{inv} , the set $-\mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon)$ contains the vectors $\mathbf{z}^{(\mathcal{K})}$ such that $\mathbf{z} \in -\mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon)$ in the limit as the entries of $\mathbf{z}^{(\mathcal{K}^c)}$ tend towards $-\infty$ [10]. Moreover, since $n\delta$ grows faster than $O(\sqrt{n})$, the elements of $n\mathbf{R}$ corresponding to \mathcal{K}^c in (40) may incur an additional $L\sqrt{n}$ term for any value of L . Combining these observations with the definition of \mathcal{L} , we obtain

$$\mathcal{L}(\epsilon, R_1^*, R_2^*) \supseteq \left\{ (L_1, L_2) : \right. \\ \left. L^{(\mathcal{K})} \in \bigcup_{\beta \geq 0} \left\{ \beta(\mathbf{I}'^{(\mathcal{K})} - \mathbf{I}^{(\mathcal{K})}) - \mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon) \right\} \right\}. \quad (41)$$

Suppose for the time being that $R_2^* > 0$, and let (Q'_U, Q'_1, Q'_2) be chosen to achieve a boundary point (R'_1, R'_2) to the right of (R_1^*, R_2^*) (more precisely, one such that $R'_1 \geq R_1^*$ and $R'_2 \leq R_2^*$, with at least one of the inequalities being strict), and define R'_{12} , \mathbf{R}' and $\mathbf{R}'^{(\mathcal{K})}$ in the same way as R_{12} , \mathbf{R} and $\mathbf{R}^{(\mathcal{K})}$. Since (R'_1, R'_2) is achieved by (Q'_U, Q'_1, Q'_2) ,

we have $\mathbf{R}' \preceq \mathbf{I}'$, and in particular, $\mathbf{R}'^{(\mathcal{K})} \preceq \mathbf{I}'^{(\mathcal{K})}$. Moreover, from the definition of \mathcal{K} , we have $\mathbf{I}^{(\mathcal{K})} = \mathbf{R}^{*(\mathcal{K})}$. Combining these, we deduce from (41) that

$$\mathcal{L}(\epsilon, R_1^*, R_2^*) \supseteq \left\{ (L_1, L_2) : L^{(\mathcal{K})} \in \bigcup_{\beta \geq 0} \left\{ \beta(\mathbf{R}'^{(\mathcal{K})} - \mathbf{R}^{*(\mathcal{K})}) - \mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon) \right\} \right\}. \quad (42)$$

By the definition of T_+ (see (17)), the direction² of the vector $\mathbf{R}'^{(\mathcal{K})} - \mathbf{R}^{*(\mathcal{K})}$ approaches that of $T_+^{(\mathcal{K})}$ as (R'_1, R'_2) approaches (R_1^*, R_2^*) along the boundary from the right. By taking this limiting choice and using the fact that \mathcal{L} is defined using a closure operation, we obtain the second set in (20). Provided that $R_1^* > 0$, the first set is obtained in an identical fashion by letting (Q'_U, Q'_1, Q'_2) achieve a boundary point approaching (R_1^*, R_2^*) from the left.

B. Proof of Theorem 2 ($\beta = 0$, $\mathcal{U} = \emptyset$)

In this subsection, we consider the case that $\beta = 0$ and $\mathcal{U} = \emptyset$, and we omit the arguments u to the functions defined in Section II-A (e.g. $\mathbf{i}(u, x_1, x_2, y)$ is replaced by $\mathbf{i}(x_1, x_2, y)$).

For $v = 1, 2$, we are given the input distribution $Q_v \in \mathcal{P}(\mathcal{X}_v)$, and we let $Q_{v,n} \in \mathcal{P}_n(\mathcal{X}_v)$ be a type with the same support as Q_v such that $\max_{x_v} |Q_v(x_v) - Q_{v,n}(x_v)| \leq \frac{1}{n}$. We generate the $M_v \triangleq e^{nR_v}$ codewords of user $v = 1, 2$ independently according to the uniform distribution on $T^n(Q_{v,n})$, i.e.

$$P_{X_v}(x_v) = \frac{1}{|T^n(Q_{v,n})|} \mathbb{1}\{x_v \in T^n(Q_{v,n})\}. \quad (43)$$

For clarity of exposition, we assume that Q_1 and Q_2 are themselves types, and hence $Q_{1,n} = Q_1$ and $Q_{2,n} = Q_2$; the analysis for the more general case introduces an additive $O(1)$ term that can be incorporated into $g(n)$ in (39).

We define the random variables

$$\begin{aligned} (X_1, X_2, Y, \bar{X}_1, \bar{X}_2) &\sim P_{X_1}(x_1)P_{X_2}(x_2)W^n(y|x_1, x_2) \\ &\times P_{X_1}(\bar{x}_1)P_{X_2}(\bar{x}_2). \end{aligned} \quad (44)$$

Using a threshold-based decoder and standard bounding techniques (e.g. see [25]), we can upper bound the random-coding error probability \bar{p}_e as follows:

$$\begin{aligned} \bar{p}_e &\leq 1 - \mathbb{P}\left[\mathbf{i}^n(X_1, X_2, Y) \succ \boldsymbol{\gamma}\right] \\ &\quad + M_1 \mathbb{P}\left[i_1^n(\bar{X}_1, X_2, Y) > \gamma_1\right] \\ &\quad + M_2 \mathbb{P}\left[i_2^n(X_1, \bar{X}_2, Y) > \gamma_2\right] \\ &\quad + M_1 M_2 \mathbb{P}\left[i_{12}^n(\bar{X}_1, \bar{X}_2, Y) > \gamma_{12}\right], \end{aligned} \quad (45)$$

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

$$\mathbf{i}^n(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \triangleq \sum_{i=1}^n \mathbf{i}(x_{1,i}, x_{2,i}, y_i) \quad (46)$$

$$i_v^n(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \triangleq \sum_{i=1}^n i_v(x_{1,i}, x_{2,i}, y_i). \quad (47)$$

²If $|\mathcal{K}| = 1$, the ‘‘direction’’ should be interpreted as being the sign.

By applying a standard change of measure from constant-composition to i.i.d. [5], [11], we can upper bound the second, third and fourth terms of (45) by $M_\nu p_0(n)e^{-\gamma_\nu}$ for $\nu = 1, 2, 12$ respectively, where $p_0(n) \triangleq (n+1)^{|\mathcal{X}_1|+|\mathcal{X}_2|-2}$, and $M_{12} \triangleq M_1 M_2$. We thus obtain

$$\bar{p}_e \leq 1 - \mathbb{P}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \succ \boldsymbol{\gamma}] + p_0(n) \sum_{\nu=1,2,12} M_\nu e^{-\gamma_\nu}. \quad (48)$$

Using this bound with

$$\gamma_\nu = \log M_\nu + \left(d + \frac{1}{2}\right) \log n, \quad (49)$$

where $d \triangleq |\mathcal{X}_1| + |\mathcal{X}_2| - 2$ is the order of the polynomial $p_0(n)$, the desired result in (39) (with $\beta = 0$ and $\mathcal{U} = \emptyset$) will follow using nearly identical steps to [6, Th. 4] once we prove the following:

- 1) $\mathbb{E}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] = n\mathbf{I}$ and $\text{Cov}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] = n\mathbf{V} + \mathbf{\Delta}_n$, where \mathbf{I} and \mathbf{V} are given in (4) and (13), and $\mathbf{\Delta}_n$ has $O(1)$ entries.
- 2) In the case that $\mathbf{V} \succ \mathbf{0}$, the probability on the right-hand side of (48) can be approximated using a multivariate Berry-Esseen theorem with $O(\frac{1}{\sqrt{n}})$ convergence.
- 3) In the case that \mathbf{V} is singular, the problem can be reduced to a lower dimension using Chebyshev's inequality, and the Berry-Esseen theorem can again be applied.

We formalize and prove these statements in the remainder of this subsection; the remaining details of the proof of (39) are omitted to avoid repetition with [6].

1) *Calculation of Moments:* The first moment of \mathbf{i}^n is easily found by writing

$$\mathbb{E}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] = \sum_{i=1}^n \mathbb{E}[\mathbf{i}(X_{1,i}, X_{2,i}, Y_i)] = n\mathbf{I}, \quad (50)$$

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 \mathbf{i}^n , we write

$$\begin{aligned} & \text{Cov}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] \\ &= \text{Cov}\left[\sum_{i=1}^n \mathbf{i}(X_{1,i}, X_{2,i}, Y_i)\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}[\mathbf{i}(X_{1,i}, X_{2,i}, Y_i), \mathbf{i}(X_{1,j}, X_{2,j}, Y_j)] \\ &= n\text{Cov}[\mathbf{i}(X_1, X_2, Y)] \\ & \quad + (n^2 - n)\text{Cov}[\mathbf{i}(X_1, X_2, Y), \mathbf{i}(X_1^\dagger, X_2^\dagger, Y^\dagger)], \end{aligned} \quad (51)$$

where (X_1, X_2, Y) and $(X_1^\dagger, X_2^\dagger, Y^\dagger)$ correspond to two arbitrary but different indices in $\{1, \dots, n\}$ (e.g. one can set $(X_1, X_2, Y) = (X_{1,1}, X_{2,1}, Y_1)$ and $(X_1^\dagger, X_2^\dagger, Y^\dagger) = (X_{1,2}, X_{2,2}, Y_2)$). Equation (53) follows by noting that the symmetry of the codebook construction implies that the n terms in (52) with $i = j$ are equal, and similarly for the $n^2 - n$ terms with $i \neq j$.

To compute the cross-covariance matrix in (53), we need the joint distribution of (X_1, X_2, Y) and $(X_1^\dagger, X_2^\dagger, Y^\dagger)$.

This distribution can be understood by noting that the uniform distribution on $T^n(Q)$ is obtained by randomly permuting the symbols of an arbitrary sequence $\mathbf{x} \in T^n(Q)$. This, in turn, can be interpreted as successively performing uniform sampling from a collection of symbols without replacement (n times in total), where the initial collection contains $nQ(x)$ occurrences of each symbol $x \in \mathcal{X}$. By considering the first two steps of such a procedure, we readily obtain

$$\mathbb{P}[X_\nu = x_\nu] = Q_\nu(x_\nu) \quad (54)$$

$$\mathbb{P}[X_\nu^\dagger = x_\nu^\dagger | X_\nu = x_\nu] = \frac{nQ_\nu(x_\nu^\dagger) - \mathbb{1}\{x_\nu = x_\nu^\dagger\}}{n-1} \quad (55)$$

for $\nu = 1, 2$. Letting $P_\nu^\dagger(x_\nu^\dagger | x_\nu)$ denote the right-hand side of (55), the cross-covariance matrix in (53) is given by

$$\begin{aligned} & \text{Cov}[\mathbf{i}(X_1, X_2, Y), \mathbf{i}(X_1^\dagger, X_2^\dagger, Y^\dagger)] \\ &= \mathbb{E}\left[(\mathbf{i}(X_1, X_2, Y) - \mathbf{I})(\mathbf{i}(X_1^\dagger, X_2^\dagger, Y^\dagger) - \mathbf{I})^T\right] \\ &= \sum_{x_1, x_2, y} Q_1(x_1)Q_2(x_2)W(y|x_1, x_2) \\ & \quad \times \sum_{x_1^\dagger, x_2^\dagger, y^\dagger} P_1^\dagger(x_1^\dagger | x_1)P_2^\dagger(x_2^\dagger | x_2)W(y^\dagger | x_1^\dagger, x_2^\dagger) \\ & \quad \times (\mathbf{i}(x_1, x_2, y) - \mathbf{I})(\mathbf{i}(x_1^\dagger, x_2^\dagger, y^\dagger) - \mathbf{I})^T \\ & \triangleq \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4, \end{aligned} \quad (56)$$

where the four terms in (58) correspond to the four terms in the expansion of $(nQ_1(x_1^\dagger) - \mathbb{1}\{x_1 = x_1^\dagger\})(nQ_2(x_2^\dagger) - \mathbb{1}\{x_2 = x_2^\dagger\})$ resulting from (55). These can be written as

$$\mathbf{F}_1 = \frac{n^2}{(n-1)^2} \mathbb{E}[\mathbf{i}(X_1, X_2, Y) - \mathbf{I}] \mathbb{E}[\mathbf{i}(X_1, X_2, Y) - \mathbf{I}]^T \quad (59)$$

$$\mathbf{F}_2 = -\frac{n}{(n-1)^2} \mathbb{E}[(\mathbf{i}(X_1, X_2, Y) - \mathbf{I})(\mathbf{i}(\bar{X}_1, X_2, \bar{Y}) - \mathbf{I})^T] \quad (60)$$

$$\mathbf{F}_3 = -\frac{n}{(n-1)^2} \mathbb{E}[(\mathbf{i}(X_1, X_2, Y) - \mathbf{I})(\mathbf{i}(X_1, \bar{X}_2, \bar{Y}) - \mathbf{I})^T] \quad (61)$$

$$\mathbf{F}_4 = \frac{1}{(n-1)^2} \mathbb{E}[(\mathbf{i}(X_1, X_2, Y) - \mathbf{I})(\mathbf{i}(X_1, X_2, \tilde{Y}) - \mathbf{I})^T], \quad (62)$$

where

$$\begin{aligned} & (X_1, X_2, Y, \bar{X}_1, \bar{X}_2, \bar{Y}, \tilde{Y}) \\ & \sim Q_1(x_1)Q_2(x_2)W(y|x_1, x_2)Q_1(\bar{x}_1)Q_2(\bar{x}_2) \\ & \quad \times W(\bar{y}|\bar{x}_1, x_2)W(\tilde{y}|x_1, \bar{x}_2)W(\tilde{y}|x_1, x_2). \end{aligned} \quad (63)$$

Observe that \mathbf{F}_1 is the zero matrix since \mathbf{I} is the mean of \mathbf{i} , and \mathbf{F}_4 has $O(\frac{1}{n^2})$ entries since the expectation does not depend on n . Furthermore, recalling the definition of $\mathbf{i}^{(2)}$ in (15), the expectation in (60) can be written as

$$\begin{aligned} & \mathbb{E}\left[\mathbb{E}[(\mathbf{i}(X_1, X_2, Y) - \mathbf{I})|X_2] \mathbb{E}[(\mathbf{i}(\bar{X}_1, X_2, \bar{Y}) - \mathbf{I})|X_2]^T\right] \\ &= \text{Cov}[\mathbf{i}^{(2)}(X_2)]. \end{aligned} \quad (64)$$

It follows that

$$F_2 = \frac{-n}{(n-1)^2} \text{Cov}[\mathbf{i}^{(2)}(X_2)], \quad (66)$$

and we similarly have

$$F_3 = \frac{-n}{(n-1)^2} \text{Cov}[\mathbf{i}^{(1)}(X_1)]. \quad (67)$$

Using the identity $\frac{n}{(n-1)^2} = \frac{1}{n} + O(\frac{1}{n^2})$ and combining (53), (58), (66) and (67), we obtain

$$\text{Cov}[\mathbf{i}^n(X_1, X_2, Y)] = n\mathbf{V} + \mathbf{\Delta}_n, \quad (68)$$

where \mathbf{V} is defined as in (13) with $\mathcal{U} = \emptyset$, and $\mathbf{\Delta}_n$ has $O(1)$ entries.

2) *A Combinatorial Berry-Esseen Theorem:* The Berry-Esseen theorem required to bound the probability in (48) is a special case of a result by Loh [19, Th. 2] for a problem known as Latin hypercube sampling. This result builds on a combinatorial central limit theorem due to Hoeffding [9]; for other related works, see [26]–[29] and the references therein.

We define the quantities

$$\mathbf{\Sigma}_n \triangleq \frac{1}{n} \text{Cov}[\mathbf{i}^n(X_1, X_2, Y)], \quad (69)$$

$$\widehat{\mathbf{S}}_n \triangleq \frac{1}{\sqrt{n}} \mathbf{\Sigma}_n^{-\frac{1}{2}} (\mathbf{i}^n(X_1, X_2, Y) - n\mathbf{I}) \quad (70)$$

$$\mathbf{\Lambda}_n(x_1, x_2) \triangleq \mathbf{\Sigma}_n^{-\frac{1}{2}} \left(\mathbf{i}(x_1, x_2, Y(x_1, x_2)) - \mathbf{i}^{(1)}(x_1) - \mathbf{i}^{(2)}(x_2) + \mathbf{I} \right) \quad (71)$$

$$\zeta_n \triangleq \sum_{x_1, x_2} Q_1(x_1) Q_2(x_2) \mathbb{E}[\|\mathbf{\Lambda}_n(x_1, x_2)\|^3], \quad (72)$$

where $Y(x_1, x_2) \sim W(\cdot|x_1, x_2)$. From (68), the matrix $\mathbf{\Sigma}_n - \mathbf{V}$ has $O(\frac{1}{n})$ entries. It follows that whenever $\mathbf{V} \succ \mathbf{0}$, we have for sufficiently large n that $\mathbf{\Sigma}_n \succ \mathbf{0}$, and hence $\mathbf{\Sigma}_n^{-\frac{1}{2}}$ is well-defined.

Theorem 3 (Corollary of [19, Th. 2]): *Let the input distributions Q_1 and Q_2 be given, and consider the quantities (X_1, X_2, Y) , \mathbf{I} , \mathbf{V} , $\widehat{\mathbf{S}}_n$ and ζ_n respectively defined in (44), (4), (13), (70) and (72) (with $\mathcal{U} = \emptyset$). If $\mathbf{V} \succ \mathbf{0}$, then we have for sufficiently large n that*

$$\left| \mathbb{P}[\widehat{\mathbf{S}}_n \in \mathcal{A}] - \mathbb{P}[\mathbf{Z} \in \mathcal{A}] \right| \leq \frac{1}{\sqrt{n}} \frac{K}{\zeta_n} \quad (73)$$

for any convex, Borel-measurable set $\mathcal{A} \subseteq \mathbb{R}^3$, where $\mathbf{Z} \sim N(\mathbf{0}, \mathbb{I})$, and K is a universal constant.

Obtaining this result from [19, Th. 2] is non-trivial, and the details are provided in Appendix A.

In the discrete setting under consideration, one can show that $\zeta_n = \Theta(1)$ using the fact that the relevant third moments are finite (e.g. they can be uniformly bounded in terms of the alphabet sizes [6, Appendix D]). Thus, we obtain the desired $O(\frac{1}{\sqrt{n}})$ convergence in (73).

When $\mathbf{V} \succ \mathbf{0}$, we can use Theorem 3 to bound the probability in (48) by writing

$$\begin{aligned} & \mathbb{P}[\mathbf{i}^n(X_1, X_2, Y) \succ \boldsymbol{\gamma}] \\ &= \mathbb{P}\left[\frac{1}{\sqrt{n}}(\mathbf{i}^n(X_1, X_2, Y) - n\mathbf{I}) \succ \frac{1}{\sqrt{n}}(\boldsymbol{\gamma} - n\mathbf{I})\right] \end{aligned} \quad (74)$$

$$= \mathbb{P}\left[\frac{1}{\sqrt{n}} \mathbf{\Sigma}_n^{-\frac{1}{2}} (\mathbf{i}^n(X_1, X_2, Y) - n\mathbf{I}) \in \mathcal{A}_n\right] \quad (75)$$

$$= \mathbb{P}[\mathbf{Z} \in \mathcal{A}_n] + O\left(\frac{1}{\sqrt{n}}\right) \quad (76)$$

$$= \mathbb{P}\left[\mathbf{\Sigma}_n^{\frac{1}{2}} \mathbf{Z} \succ \frac{1}{\sqrt{n}}(\boldsymbol{\gamma} - n\mathbf{I})\right] + O\left(\frac{1}{\sqrt{n}}\right) \quad (77)$$

where (75) follows by defining \mathcal{A}_n to be the image of the rectangular region in (74) under $\mathbf{\Sigma}_n^{-\frac{1}{2}}$, (76) follows with $\mathbf{Z} \sim N(\mathbf{0}, \mathbb{I})$ from Theorem 3, and (77) follows by reversing the step in (75). These steps are similar to [6, Appendix B], where a Cholesky decomposition is used.

3) *Singular Dispersion Matrices:* In general, the dispersion matrix \mathbf{V} may not have full rank, in which case Theorem 3 does not directly apply. 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 $n\mathbf{V}$ is not necessarily the exact covariance matrix of $\mathbf{i}^n(X_1, X_2, Y)$, due to the additional $O(1)$ term in (68).

Suppose that \mathbf{V} has rank $r < 3$, and consider the matrix $\mathbf{\Sigma}_n$ in (69). Using an eigenvalue decomposition along with (68), we obtain

$$\mathbf{\Sigma}_n = \mathbf{\Phi}_n \mathbf{\Omega}_n \mathbf{\Phi}_n^T, \quad (78)$$

where $\mathbf{\Phi}_n$ is a unitary matrix, and $\mathbf{\Omega}_n$ is a diagonal matrix whose first r diagonals are $\Theta(1)$, and whose last $3 - r$ diagonals are $O(\frac{1}{n})$. Noting that $\mathbf{\Sigma}_n$ is the covariance matrix of $\mathbf{A}_n \triangleq \frac{1}{\sqrt{n}}(\mathbf{i}^n(X_1, X_2, Y) - n\mathbf{I})$, we see that $\mathbf{\Omega}_n$ is the covariance matrix of $\widetilde{\mathbf{A}}_n \triangleq \mathbf{\Phi}_n^T \mathbf{A}_n$.

Suppose for the time being that $r \geq 1$. From the structure of $\mathbf{\Omega}_n$, we conclude that

$$\widetilde{\mathbf{A}}_n = \begin{bmatrix} \widetilde{\mathbf{A}}_n^{(1)} \\ \widetilde{\mathbf{A}}_n^{(2)} \end{bmatrix}, \quad (79)$$

where $\widetilde{\mathbf{A}}_n^{(1)}$ and $\widetilde{\mathbf{A}}_n^{(2)}$ have dimension r and $3 - r$ respectively, and the covariance matrix of $\widetilde{\mathbf{A}}_n^{(2)}$ has $O(\frac{1}{n})$ entries. Since $\mathbf{\Phi}_n$ is unitary (i.e. $\mathbf{\Phi}_n \mathbf{\Phi}_n^T = \mathbb{I}$), we have $\mathbf{A}_n = \mathbf{\Phi}_n \widetilde{\mathbf{A}}_n$, and hence

$$\mathbf{A}_n = \mathbf{\Phi}_n \begin{bmatrix} \widetilde{\mathbf{A}}_n^{(1)} \\ \mathbf{0} \end{bmatrix} + \mathbf{\Phi}_n \begin{bmatrix} \mathbf{0} \\ \widetilde{\mathbf{A}}_n^{(2)} \end{bmatrix} \quad (80)$$

$$\triangleq \mathbf{\Phi}'_n \widetilde{\mathbf{A}}_n^{(1)} + \widetilde{\mathbf{\Delta}}_n, \quad (81)$$

where $\mathbf{\Phi}'_n$ is obtained from $\mathbf{\Phi}_n$ by keeping only the first r columns, and $\widetilde{\mathbf{\Delta}}_n$ denotes the second term in (80). Since \mathbf{A}_n has mean zero by construction, we conclude that the same is true of $\widetilde{\mathbf{A}}_n$, and hence of $\widetilde{\mathbf{\Delta}}_n$. Furthermore, since $\text{Cov}[\widetilde{\mathbf{A}}_n^{(2)}]$ has $O(\frac{1}{n})$ entries and $\widetilde{\mathbf{\Delta}}_n$ is obtained from $\widetilde{\mathbf{A}}_n^{(2)}$ via the unitary (and hence uniformly bounded) matrix $\mathbf{\Phi}_n$, $\text{Cov}[\widetilde{\mathbf{\Delta}}_n]$ also has

$O(\frac{1}{n})$ entries. Thus, Chebyshev's inequality implies for any $\delta_n > 0$ that

$$\mathbb{P}[\|\tilde{\Delta}_n\|_\infty \geq \delta_n] \leq \frac{\max_i \text{Var}[\tilde{\Delta}_{i,n}]}{\delta_n^2} = O\left(\frac{1}{n\delta_n^2}\right), \quad (82)$$

where $\tilde{\Delta}_{i,n}$ is the i -th entry of $\tilde{\Delta}$.

We are now in a position to bound the probability appearing in (48). The following holds for any $\delta_n > 0$:

$$\begin{aligned} & \mathbb{P}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \succ \boldsymbol{\gamma}] \\ &= \mathbb{P}\left[\mathbf{A}_n \succ \frac{1}{\sqrt{n}}(\boldsymbol{\gamma} - n\mathbf{I})\right] \end{aligned} \quad (83)$$

$$= \mathbb{P}\left[\Phi'_n \tilde{\mathbf{A}}_n^{(1)} + \tilde{\Delta}_n \succ \frac{1}{\sqrt{n}}(\boldsymbol{\gamma} - n\mathbf{I})\right] \quad (84)$$

$$\geq \mathbb{P}\left[\Phi'_n \tilde{\mathbf{A}}_n^{(1)} \succ \frac{1}{\sqrt{n}}(\boldsymbol{\gamma} - n\mathbf{I}) + \delta_n \mathbf{1}\right] - \mathbb{P}[\|\tilde{\Delta}_n\|_\infty \geq \delta_n] \quad (85)$$

$$= \mathbb{P}\left[\Phi'_n \tilde{\mathbf{A}}_n^{(1)} \succ \frac{1}{\sqrt{n}}(\boldsymbol{\gamma} - n\mathbf{I}) + \delta_n \mathbf{1}\right] + O\left(\frac{1}{n\delta_n^2}\right), \quad (86)$$

where the final three steps respectively make use of (81), [6, Lemma 9], and (82). Since the entries of $\tilde{\mathbf{A}}_n^{(1)}$ are shifted and weighted sums of the entries of $\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$, and since the corresponding covariance matrix is positive definite by construction, we can analyze (86) in the same way as the case $\mathbf{V} \succ \mathbf{0}$ (other than the terms $\delta_n \mathbf{1}$ and $O(\frac{1}{n\delta_n^2})$, which are handled in the following paragraph).

The fact that the resulting second-order term can be written as $-\sqrt{n}\mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon)$ follows in the same way as the i.i.d. case [6, p. 894].

The remainder term in (86) contributes an additive $O(\frac{1}{\sqrt{n}\delta_n^2})$ term to the expansion in (39), whereas the addition of $\delta_n \mathbf{1}$ in the first probability in (86) contributes an additive $O(\delta_n \sqrt{n})$ term to the expansion. The overall contribution $O(\frac{1}{\sqrt{n}\delta_n^2} + \delta_n \sqrt{n})$ is minimized by $\delta_n = \Theta(n^{-\frac{1}{3}})$, yielding $g(n) = O(n^{\frac{1}{6}})$, as stated following Theorem 2.

The case $r = 0$ (i.e. $\mathbf{V} = \mathbf{0}$) is handled similarly; by following the steps of (83)–(86), we readily obtain the following analog of (86):

$$\begin{aligned} & \mathbb{P}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \succ \boldsymbol{\gamma}] \\ & \geq \mathbb{1}\left\{\mathbf{0} \succ \frac{1}{\sqrt{n}}(\boldsymbol{\gamma} - n\mathbf{I}) + \delta_n \mathbf{1}\right\} + O\left(\frac{1}{n\delta_n^2}\right) \end{aligned} \quad (87)$$

Using this bound, we can obtain (39) (with $\beta = 0$ and $\mathcal{U} = \emptyset$) in the same way as the case $r > 0$ by noting that $\mathbf{Q}_{\text{inv}}(\mathbf{0}, \epsilon)$ equals the set of all vectors with non-negative components.

C. Proof of Theorem 2 (General Case)

Here we provide the changes required in the previous subsection to prove Theorem 2 in full generality. We recall the definitions of $P'_{UX_1X_2Y}$, \mathbf{i}' , etc. at the beginning of the section.

1) *Coded Time-Sharing With $\beta = 0$* : In the case that $\beta = 0$ but $\mathcal{U} \neq \emptyset$, we modify the constant-composition random-coding ensemble (cf. (43)) as follows. We let $Q_{U,n}$, $Q_{1,n}$ and $Q_{2,n}$ be (conditional) types that respectively approximate

Q_U , Q_1 and Q_2 . We fix an arbitrary time-sharing sequence \mathbf{u} with type $Q_{U,n}$, and generate the $M_v \triangleq e^{nR_v}$ codewords of user $v = 1, 2$ independently according to the uniform distribution on $T_u^n(Q_{v,n})$, i.e.

$$P_{X_v|U}(\mathbf{x}_v|\mathbf{u}) = \frac{1}{|T_u^n(Q_{v,n})|} \mathbb{1}\{\mathbf{x}_v \in T_u^n(Q_{v,n})\}. \quad (88)$$

Similarly to (44), we define the random variables

$$\begin{aligned} & (\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2) \\ & \sim P_{X_1|U}(\mathbf{x}_1|\mathbf{u})P_{X_2|U}(\mathbf{x}_2|\mathbf{u}) \\ & \quad \times W^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)P_{X_1|U}(\bar{\mathbf{x}}_1|\mathbf{u})P_{X_2|U}(\bar{\mathbf{x}}_2|\mathbf{u}). \end{aligned} \quad (89)$$

The procedure described in Section IV-B.1 for generating a codeword uniformly over the type class is modified as follows. Let \mathbf{x} be an arbitrary element of the conditional type class $T_u(\cdot)$. Instead of randomly permuting the entire sequence \mathbf{x} , a random permutation of the subsequence $\mathbf{x}^{(u)}$ corresponding to the indices where \mathbf{u} equals u is applied *independently* for each value of $u \in \mathcal{U}$. Due to this independence, we can handle the summation

$$\mathbf{i}^n(\mathbf{u}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \triangleq \sum_{i=1}^n \mathbf{i}(u_i, X_{1,i}, X_{2,i}, Y_i) \quad (90)$$

$$= \sum_u \sum_{i=1}^{nQ_U(u)} \mathbf{i}(u, X_{1,i}^{(u)}, X_{2,i}^{(u)}, Y_i^{(u)}) \quad (91)$$

by considering each value of $u \in \mathcal{U}$ separately. For the values of u corresponding to singular dispersion matrices, we can perform a reduction to a lower dimension as shown following (78). From Theorem 3, we conclude that each inner summation in (91) is asymptotically normal with $O(\frac{1}{\sqrt{nQ_U(u)}}) = O(\frac{1}{\sqrt{n}})$ convergence. It follows that the overall sum is also asymptotically normal with $O(\frac{1}{\sqrt{n}})$ convergence. To see this, we let $\hat{\mathbf{S}}_{1,n}$ and $\hat{\mathbf{S}}_{2,n}$ be asymptotically normal in the sense of (73), and let \mathbf{Z}_1 and \mathbf{Z}_2 be the associated Gaussian random variables. We then have

$$\begin{aligned} & \mathbb{P}[\hat{\mathbf{S}}_{1,n} + \hat{\mathbf{S}}_{2,n} \in \mathcal{A}] \\ &= \mathbb{E}[\mathbb{P}[\hat{\mathbf{S}}_{1,n} + \hat{\mathbf{S}}_{2,n} \in \mathcal{A} | \hat{\mathbf{S}}_{2,n}]] \end{aligned} \quad (92)$$

$$= \mathbb{E}[\mathbb{P}[\mathbf{Z}_1 + \hat{\mathbf{S}}_{2,n} \in \mathcal{A} | \hat{\mathbf{S}}_{2,n}]] + O\left(\frac{1}{\sqrt{n}}\right) \quad (93)$$

$$= \mathbb{P}[\mathbf{Z}_1 + \hat{\mathbf{S}}_{2,n} \in \mathcal{A}] + O\left(\frac{1}{\sqrt{n}}\right) \quad (94)$$

$$= \mathbb{P}[\mathbf{Z}_1 + \mathbf{Z}_2 \in \mathcal{A}] + O\left(\frac{1}{\sqrt{n}}\right), \quad (95)$$

where (95) follows using similar steps to (92)–(94). Using this observation and repeating the analysis of Section IV-B and [6], we obtain the more general result of Theorem 2 with $\beta = 0$.

2) *Coded Time-Sharing With $\beta > 0$* : In the case that $\beta > 0$, we apply a variant of coded time-sharing depending on an extended alphabet $\tilde{\mathcal{U}} \triangleq \mathcal{U} \times \{1, 2\}$ and the triplets (Q_U, Q_1, Q_2) and $(Q'_{U,2}, Q'_1, Q'_2)$. Specifically, we define the triplet $(\tilde{Q}_{U,n}, \tilde{Q}_{1,n}, \tilde{Q}_{2,n})$ with $\tilde{Q}_{U,n} \in \mathcal{P}(\tilde{\mathcal{U}})$,

$\tilde{Q}_{1,n} \in \mathcal{P}(\mathcal{X}_1 | \tilde{\mathcal{U}})$ and $\tilde{Q}_{2,n} \in \mathcal{P}(\mathcal{X}_2 | \tilde{\mathcal{U}})$ as follows:

$$\tilde{Q}_{U,n}(u, 1) = \left(1 - \frac{\beta}{\sqrt{n}}\right) Q_U(u) \quad (96)$$

$$\tilde{Q}_{U,n}(u, 2) = \frac{\beta}{\sqrt{n}} Q'_U(u) \quad (97)$$

$$\tilde{Q}_{v,n}(x_v | u, 1) = Q_v(x_v | u) \quad (98)$$

$$\tilde{Q}_{v,n}(x_v | u, 2) = Q'_v(x_v | u) \quad (99)$$

for $v = 1, 2$. We consider the constant-composition ensemble in (88) with $(\tilde{Q}_{U,n}, \tilde{Q}_{1,n}, \tilde{Q}_{2,n})$ playing the role of (Q_U, Q_1, Q_2) , and we set $\tilde{P}_{UX_1X_2Y,n} \triangleq \tilde{Q}_{U,n} \times \tilde{Q}_{1,n} \times \tilde{Q}_{2,n} \times W$. We follow the same arguments as the case that $\beta = 0$ and $\mathcal{U} \neq \emptyset$, but with care taken to handle the fact that some of the time-sharing values $\tilde{u} \in \tilde{\mathcal{U}}$ correspond to subsequences of \mathbf{u} having length $\Theta(\sqrt{n})$, rather than $\Theta(n)$. Since we are interested in the limit of large n , we may assume that $1 - \frac{\beta}{\sqrt{n}} > 0$.

The role of the information density vector \mathbf{i} is now played by the quantity

$$\tilde{\mathbf{i}}(\tilde{u}, x_1, x_2, y) \triangleq \begin{cases} \mathbf{i}(u, x_1, x_2, y) & \tilde{u} = (u, 1) \\ \mathbf{i}'(u, x_1, x_2, y) & \tilde{u} = (u, 2). \end{cases} \quad (100)$$

The corresponding mean vector and dispersion matrix with respect to $\tilde{P}_{UX_1X_2Y,n}$ are given by

$$\tilde{\mathbf{I}} \triangleq \left(1 - \frac{\beta}{\sqrt{n}}\right) \mathbf{I} + \frac{\beta}{\sqrt{n}} \mathbf{I}' \quad (101)$$

$$\tilde{\mathbf{V}} \triangleq \left(1 - \frac{\beta}{\sqrt{n}}\right) \mathbf{V} + \frac{\beta}{\sqrt{n}} \mathbf{V}', \quad (102)$$

where \mathbf{I}' and \mathbf{V}' are defined in (37)–(38). With these definitions, the analysis proceeds in the same way as the above analysis for $\mathcal{U} \neq \emptyset$. Analogously to (91) (and using analogous notation), the additive n -letter extension $\tilde{\mathbf{i}}^n$ of (100) admits the decomposition

$$\tilde{\mathbf{i}}^n(\mathbf{u}, X_1, X_2, Y) = \sum_{\tilde{u}} \sum_{i=1}^{n\tilde{Q}_{U,n}(\tilde{u})} \tilde{\mathbf{i}}(\tilde{u}, X_{1,i}^{(\tilde{u})}, X_{2,i}^{(\tilde{u})}, Y_i^{(\tilde{u})}). \quad (103)$$

The inner summations corresponding to $\tilde{u} = (u, 2)$ only contain $\Theta(\sqrt{n})$ terms (rather than $\Theta(n)$ terms). Since the remainder term in the Berry-Esseen theorem decays as the inverse of the square root of the number of terms, we get $O(n^{\frac{1}{4}})$ in place of $O(\frac{1}{\sqrt{n}})$ in (77).

On the other hand, the remainder term in (86) is unchanged despite the presence of \tilde{u} values with corresponding subsequences of length $\Theta(\sqrt{n})$. To see this, we first write

$$\mathbb{P}[\tilde{\mathbf{i}}^n > \boldsymbol{\gamma}] = \mathbb{P}\left[\sum_{\tilde{u}} \frac{1}{\sqrt{n}} (\tilde{\mathbf{i}}_{\tilde{u}}^{n_{\tilde{u}}} - n_{\tilde{u}} \tilde{\mathbf{I}}_{\tilde{u}}) > \frac{1}{\sqrt{n}} (\boldsymbol{\gamma} - n\tilde{\mathbf{I}})\right], \quad (104)$$

where $\tilde{\mathbf{i}}^n$ denotes the left-hand side of (103) (with implicit arguments), $\tilde{\mathbf{i}}_{\tilde{u}}^{n_{\tilde{u}}}$ denotes the inner summation corresponding to a given \tilde{u} on the right-hand side of (103), $n_{\tilde{u}} \triangleq n\tilde{Q}_{U,n}(\tilde{u})$ denotes the number of terms in the summation, and $\tilde{\mathbf{I}}_{\tilde{u}}$ denotes the mean of each summand therein. These definitions, along with those in (96)–(100), readily yield $\tilde{\mathbf{I}} = \sum_{\tilde{u}} \tilde{Q}_{U,n}(\tilde{u}) \tilde{\mathbf{I}}_{\tilde{u}}$.

Using (104), one can follow the steps in (83)–(86) and end up with the same remainder term as (86), regardless of which values of \tilde{u} have corresponding dispersion matrices that are singular.

Combining the preceding observations and following the steps of the previous subsections and [6], we obtain the following condition for (n, ϵ) -achievable:

$$n\mathbf{R} \in n\tilde{\mathbf{I}} - \sqrt{n} \mathbf{Q}_{\text{inv}}(\tilde{\mathbf{V}}, \epsilon) + O(n^{\frac{1}{4}}) \mathbf{1}. \quad (105)$$

The proof is concluded by substituting (101)–(102) into (105) and using a Taylor expansion of \mathbf{Q}_{inv} (e.g. see [16, Lemma 6]) to replace $\mathbf{Q}_{\text{inv}}(\tilde{\mathbf{V}}, \epsilon)$ by $\mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon)$.

V. CONCLUSION

We have characterized the second-order asymptotics of the DM-MAC using constant-composition random coding and a combinatorial Berry-Esseen theorem. Applying an extended version of coded time-sharing, we have presented a new method for obtaining the derivative (or tangent vector) terms in the second-order rate region, which first appeared in [16]. Analogously to the random-coding error exponents [12], we have observed improved bounds for constant-composition random coding compared to i.i.d. random coding. While we focused primarily on unconstrained channels, our results are directly applicable to discrete channels with input constraints, thus providing another advantage over i.i.d. codes. We have also presented an extension of our main result to the Gaussian setting via an increasingly fine quantization of the inputs.

A highly challenging open problem is the development of outer bounds on \mathcal{L} . The converse analysis for the Gaussian MAC with degraded message sets [16] relied on a reduction from average error to maximal error, but it is well-known that such a reduction is not possible for the standard MAC [30]. The “wringing techniques” used in Ahlswede’s derivation of the strong converse circumvent this issue [31], but still fail to exhibit $O(\frac{1}{\sqrt{n}})$ convergence rates to the boundary points, as is required to get a non-trivial outer bound on \mathcal{L} .

APPENDIX A

PROOF OF THEOREM 3

Here we outline the problem studied by Loh [19] and state the result that recovers Theorem 3, adapting the notation therein to be more consistent with ours. We write $A_1 \stackrel{d}{=} A_2$ if the random variables A_1 and A_2 have the same distribution.

The “dimensionality” in [19] corresponds to the number of users of the MAC, so we let it equal 2. Let $\pi_1(\cdot)$ and $\pi_2(\cdot)$ be independent random permutations of $\{1, \dots, n\}$, uniformly distributed over the $n!$ possible permutations. For $v = 1, 2$ and $j_1, j_2 = 1, \dots, n$, define the random variables $U_v(j_1, j_2)$ uniformly distributed on $(0, 1)$ independently of each other and of $\pi_1(\cdot)$ and $\pi_2(\cdot)$, and set

$$B_v(j_1, j_2) \triangleq \frac{j_v - U_v(j_1, j_2)}{n}, \quad v = 1, 2. \quad (106)$$

The summation of interest is written as follows:

$$S_n \triangleq \sum_{j=1}^n f\left(\mathbf{B}(\pi_1(j), \pi_2(j))\right), \quad (107)$$

where $\mathbf{B}(\cdot, \cdot) = [B_1(\cdot, \cdot) \ B_2(\cdot, \cdot)]^T$, and $\mathbf{f}(\cdot)$ is a function with a two-dimensional vector argument and a three-dimensional vector output.

Using (106) and the fact that $U_v(j_1, j_2) \in (0, 1)$ almost surely, any realization b_v of B_v uniquely determines both the index j_v and the variable U_v in the numerator of (106). Thus, overloading the symbol \mathbf{f} , we write the following for $\mathbf{b} = [b_1 \ b_2]^T$:

$$\mathbf{f}(\mathbf{b}) = \mathbf{f}(b_1, b_2) = \mathbf{f}(j_1, j_2, u_1, u_2), \quad (108)$$

where the final four arguments are deterministically deduced from \mathbf{b} .

We now provide the key definitions needed to state [19, Th. 2]. The notation here should be treated as being separate from the rest of this paper for now, but we will shortly see that the definitions of all re-used symbols are consistent. The quantities related to first moments are as follows:

$$\boldsymbol{\mu} \triangleq \frac{1}{n} \mathbb{E}[\mathbf{S}_n] \quad (109)$$

$$\boldsymbol{\mu}(j_1, j_2) \triangleq \mathbb{E}[\mathbf{f}(\mathbf{B}(j_1, j_2))] \quad (110)$$

$$\boldsymbol{\mu}^{(1)}(j_1) \triangleq \frac{1}{n} \sum_{j_2=1}^n \boldsymbol{\mu}(j_1, j_2) \quad (111)$$

$$\boldsymbol{\mu}^{(2)}(j_2) \triangleq \frac{1}{n} \sum_{j_1=1}^n \boldsymbol{\mu}(j_1, j_2). \quad (112)$$

The quantities related to second moments are as follows:

$$\boldsymbol{\Sigma}_n \triangleq \frac{1}{n} \text{Cov}[\mathbf{S}_n] \quad (113)$$

$$\widehat{\mathbf{S}}_n \triangleq \frac{1}{\sqrt{n}} \boldsymbol{\Sigma}_n^{-\frac{1}{2}} (\mathbf{S}_n - n\boldsymbol{\mu}) \quad (114)$$

$$\mathbf{f}^{(1)}(b_1) \triangleq \int_0^1 \mathbf{f}(b_1, b_2) db_2 \quad (115)$$

$$\mathbf{f}^{(2)}(b_2) \triangleq \int_0^1 \mathbf{f}(b_1, b_2) db_1 \quad (116)$$

$$\mathbf{f}_{\text{rem}}(b_1, b_2) \triangleq \mathbf{f}(b_1, b_2) - \mathbf{f}^{(1)}(b_1) - \mathbf{f}^{(2)}(b_2) + \boldsymbol{\mu} \quad (117)$$

$$\mathbf{V}_f \triangleq \int_0^1 \int_0^1 \mathbf{f}_{\text{rem}}(b_1, b_2) \mathbf{f}_{\text{rem}}(b_1, b_2)^T db_1 db_2 \quad (118)$$

Finally, the quantities related to third moments are as follows:

$$\boldsymbol{\Lambda}(j_1, j_2) \triangleq \boldsymbol{\Sigma}_n^{-\frac{1}{2}} (\mathbf{f}(\mathbf{B}(j_1, j_2)) - \boldsymbol{\mu}^{(1)}(j_1) - \boldsymbol{\mu}^{(2)}(j_2) + \boldsymbol{\mu}) \quad (119)$$

$$\xi_n \triangleq \frac{1}{n^2} \sum_{j_1=1}^n \sum_{j_2=1}^n \mathbb{E}[\|\boldsymbol{\Lambda}(j_1, j_2)\|^3]. \quad (120)$$

We now have the following.

Theorem 4 (Combinatorial Berry-Esseen Theorem [19, Th. 2]): *If $\mathbf{V}_f \succ \mathbf{0}$, then the following holds for sufficiently large n :*

$$\left| \mathbb{P}[\widehat{\mathbf{S}}_n \in \mathcal{A}] - \mathbb{P}[\mathbf{Z} \in \mathcal{A}] \right| \leq \frac{1}{\sqrt{n}} \frac{K}{\xi_n} \quad (121)$$

for any convex, Borel-measurable set $\mathcal{A} \subseteq \mathbb{R}^3$, where $\mathbf{Z} \sim N(\mathbf{0}, \mathbb{I})$, and K is a universal constant.

We now show that Theorem 3 is recovered by a suitable choice of \mathbf{f} . Let $\mathbf{x}_1 = (x_{1,1}, \dots, x_{1,n})$ and $\mathbf{x}_2 = (x_{2,1}, \dots, x_{2,n})$ be arbitrary sequences having type \mathcal{Q}_1 and \mathcal{Q}_2 respectively, and define $Y(j_1, j_2) \sim W(\cdot | x_{1,j_1}, x_{2,j_2})$ with independence between different (j_1, j_2) pairs. We set

$$\mathbf{f}(j_1, j_2, u_1, u_2) = \mathbf{i}(x_{1,j_1}, x_{2,j_2}, F_{Y(j_1, j_2)}^{-1}(u_1 \oplus u_2)), \quad (122)$$

where $F_{Y(j_1, j_2)}^{-1}(u) = \inf \{y : F_{Y(j_1, j_2)}(y) \geq u\}$ is the inverse cumulative distribution function (CDF) of $Y(j_1, j_2)$, and \oplus denotes real addition modulo one.

We first evaluate the quantities in (109)–(112). Clearly $U_1 \oplus U_2$ is uniform on $(0, 1)$, and since $F_Z^{-1}(U) \stackrel{d}{=} Z$ for any random variable Z with CDF F_Z , it follows that $F_{Y(j_1, j_2)}^{-1}(U_1 \oplus U_2) \stackrel{d}{=} Y(j_1, j_2)$, and hence

$$\mathbf{f}(\mathbf{B}(j_1, j_2)) = \mathbf{f}(j_1, j_2, U_1, U_2) \stackrel{d}{=} \mathbf{i}(x_{1,j_1}, x_{2,j_2}, Y(j_1, j_2)), \quad (123)$$

Since drawing a codeword uniformly over a type class is equivalent to randomly permuting any codeword of the given type, it follows that \mathbf{S}_n in (107) has the same distribution as $\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ in (48). Using (123), the fact that $\mathbf{x}_v \in T^n(\mathcal{Q}_v)$ ($v = 1, 2$), and the definitions of $\mathbf{i}^{(1)}$ and $\mathbf{i}^{(2)}$ in (14)–(15), we readily obtain $\boldsymbol{\mu} = \mathbf{I}$, $\boldsymbol{\mu}^{(1)}(j_1) = \mathbf{i}^{(1)}(x_{1,j_1})$, and $\boldsymbol{\mu}^{(2)}(j_2) = \mathbf{i}^{(2)}(x_{2,j_2})$.

Next, we consider the quantities in (115)–(118). Recalling that the pair (j_v, u_v) is uniquely determined by b_v for $v = 1, 2$, we have

$$\mathbf{f}^{(1)}(b_1) = \int_0^1 \mathbf{f}(b_1, b_2) db_2 \quad (124)$$

$$= \frac{1}{n} \sum_{j_2=1}^n \mathbb{E}[\mathbf{f}(j_1, j_2, u_1, U_2)] \quad (125)$$

$$= \frac{1}{n} \sum_{j_2=1}^n \mathbb{E}[\mathbf{i}(x_{1,j_1}, x_{2,j_2}, F_{Y(j_1, j_2)}^{-1}(u_1 \oplus U_2))] \quad (126)$$

$$= \frac{1}{n} \sum_{j_2=1}^n \mathbb{E}[\mathbf{i}(x_{1,j_1}, x_{2,j_2}, Y(j_1, j_2))] \quad (127)$$

$$= \mathbf{i}^{(1)}(x_{1,j_1}), \quad (128)$$

where (125) follows by interpreting the uniform averaging over $b_2 \in [0, 1]$ as an averaging over n segments of length $\frac{1}{n}$ along with an averaging over u_2 within each segment, (126) follows from (122), (127) follows since $u_1 \oplus U_2$ is uniform on $[0, 1]$ for any u_1 , and (128) follows from the definition of $\mathbf{i}^{(1)}$ in (14) and the fact that $\mathbf{x}_2 \in T^n(\mathcal{Q}_2)$. An identical argument reveals that $\mathbf{f}^{(2)}(b_2) = \mathbf{i}^{(2)}(x_{2,j_2})$, where j_2 is uniquely determined by b_2 .

The only remaining quantity whose evaluation is non-trivial is \mathbf{V}_f in (118). By writing (117) as

$$\mathbf{f}_{\text{rem}}(b_1, b_2) = (\mathbf{f}(b_1, b_2) - \boldsymbol{\mu}) - (\mathbf{f}^{(1)}(b_1) - \boldsymbol{\mu}) - (\mathbf{f}^{(2)}(b_2) - \boldsymbol{\mu}), \quad (129)$$

we can express (118) as the sum of $3^2 = 9$ integrals. The desired identity $\mathbf{V}_f = \mathbf{V}$ is obtained by showing that these

evaluate to

$$\mathbf{V}_f = \mathbf{V}^{(12)} + \mathbf{V}^{(1)} + \mathbf{V}^{(2)} - \mathbf{V}^{(1)} - \mathbf{V}^{(1)} - \mathbf{V}^{(2)} - \mathbf{V}^{(2)} + \mathbf{0} + \mathbf{0} \quad (130)$$

$$= \mathbf{V}^{(12)} - \mathbf{V}^{(1)} - \mathbf{V}^{(2)} \quad (131)$$

$$= \mathbf{V}, \quad (132)$$

where $\mathbf{V}^{(12)}$, $\mathbf{V}^{(1)}$ and $\mathbf{V}^{(2)}$ are the three covariance matrices appearing on the right-hand side of (13) (with $\mathcal{U} = \emptyset$). For brevity, we provide details for only one of the 9 terms in (130); the others are handled similarly. We have

$$\begin{aligned} & \int_0^1 \int_0^1 (\mathbf{f}(b_1, b_2) - \boldsymbol{\mu})(\mathbf{f}^{(1)}(b_1) - \boldsymbol{\mu})^T db_1 db_2 \\ &= \int_0^1 \left(\int_0^1 (\mathbf{f}(b_1, b_2) - \boldsymbol{\mu}) db_2 \right) (\mathbf{f}^{(1)}(b_1) - \boldsymbol{\mu})^T db_1 \end{aligned} \quad (133)$$

$$= \int_0^1 (\mathbf{f}^{(1)}(b_1) - \boldsymbol{\mu})(\mathbf{f}^{(1)}(b_1) - \boldsymbol{\mu})^T db_1 \quad (134)$$

$$= \text{Cov}[\mathbf{i}^{(1)}(X_1)] \quad (135)$$

$$= \mathbf{V}^{(1)}, \quad (136)$$

where (134) follows from (115), and (135) follows from (128) and by interpreting the integral as an average.

We refer the interested reader to [32, Sec. 4.5.6] for a list of identities between quantities appearing in the present paper and the paper of Loh [19].

APPENDIX B

EXTENSION TO THE GAUSSIAN SETTING

A. Evaluation of \mathbf{I} and \mathbf{V}

The expressions in (31)–(35) are derived from (4) and (13) by forming an explicit expression for $\mathbf{i}(x_1, x_2, y)$ in (6) (with $\mathcal{U} = \emptyset$ and $Q_1, Q_2 \sim N(0, 1)$), performing averaging in order to obtain $\mathbf{i}^{(1)}(x_1)$ and $\mathbf{i}^{(2)}(x_2)$ in (14)–(15), and then computing the corresponding means, variances, and covariances. For concreteness, we provide a brief outline of this process for the bottom-right entry of \mathbf{V} , namely V_{12} .

When Q_1 and Q_2 are $N(0, 1)$, the output distribution is $P_Y \sim N(0, 1 + P_1 + P_2)$, yielding

$$i_{12}(X_1, X_2, Y) = I_{12} - \frac{Z^2}{2} + \frac{(\sqrt{P_1}X_1 + \sqrt{P_2}X_2 + Z)^2}{2(1 + P_1 + P_2)}. \quad (137)$$

Averaging over (X_2, Y) and (X_1, Y) respectively, we obtain the following:

$$i_{12}^{(1)}(X_1) = I_{12} + \frac{P_1(X_1^2 - 1)}{2(1 + P_1 + P_2)} \quad (138)$$

$$i_{12}^{(2)}(X_2) = I_{12} + \frac{P_2(X_2^2 - 1)}{2(1 + P_1 + P_2)}. \quad (139)$$

Each of the three preceding quantities has mean I_{12} . Using the fact that the second and fourth moments of an $N(0, 1)$ random variable are 1 and 3 respectively, the corresponding variances are easily calculated to be $\frac{P_1 + P_2}{1 + P_1 + P_2}$, $\frac{P_1^2}{2(1 + P_1 + P_2)^2}$ and $\frac{P_2^2}{2(1 + P_1 + P_2)^2}$. Substituting these into (13) yields $V_{12} = \frac{(P_1 + P_2)(2 + P_1 + P_2) + 2P_1P_2}{2(1 + P_1 + P_2)^2}$, as desired.

B. Derivation of the Achievable Second-Order Rate Region

Recall that the entries of \mathbf{I} and \mathbf{V} can be written in the forms given in (31)–(35) respectively. The key result used in obtaining (36) is the following lemma, which states that there exists 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}) (see (31)–(35)), with the convergence $\mathbf{I}_m \rightarrow \mathbf{I}$ being exponentially fast in m . This generalizes a result by Wu and Verdú for the single-user setting [33], and is proved similarly.

Lemma 1: *There exist sequences 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 (4) and (13) such that (i) $\|\mathbf{I}_m - \mathbf{I}\|_\infty \leq e^{-\gamma m}$ for some $\gamma > 0$ and sufficiently large m , (ii) $\|\mathbf{V}_m - \mathbf{V}\|_\infty \rightarrow 0$, 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: The proof closely follows that of [33, Th. 8], so we only explain the differences. We choose Q_{m1} and Q_{m2} according to the Gauss quadrature rule Q_g [33, Sec. II], which satisfies the property of having the same moments as those of a standard Gaussian random variable up to order $2m - 1$ [33, Thm. 2]. Since Q_g converges weakly to $N(0, 1)$ [33], we immediately obtain parts (ii) and (iii) of Lemma 1, so it remains to prove part (i).

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 [24, Eq. (15.142)]

$$I(X_{m1}; Y_m | X_{m2}) = H(\sqrt{P_1}X_{m1} + Z) - H(Z), \quad (140)$$

we see that the convergence of the first entry of \mathbf{I}_m to that of \mathbf{I} is precisely that studied in [33], and similarly for the second entry. It remains to study the third entry, i.e. to show that $I(X_{m1}, X_{m2}; Y_m) \rightarrow I(X_1, X_2; Y)$ exponentially fast. Analogously to [33, Eq. (5)], we have

$$\begin{aligned} I(X_1, X_2; Y) - I(X_{m1}, X_{m2}; Y_m) \\ = D(\sqrt{P_1}X_1 + \sqrt{P_2}X_2 + Z \| \sqrt{P_1}X_{m1} + \sqrt{P_2}X_{m2} + Z) \end{aligned} \quad (141)$$

$$\triangleq D_m. \quad (142)$$

Using nearly identical arguments to [33, Sec. V] with an “optimal” output distribution of $N(0, 1 + P_1 + P_2)$, we obtain analogously to [33, Eq. (54)] that

$$\begin{aligned} D_m \leq \sum_{k \geq 1} \frac{1}{k!} \left(\frac{P_1 + P_2}{1 + P_1 + P_2} \right)^k \\ \times \left| \mathbb{E} \left[H_k \left(\frac{\sqrt{P_1}X_{m1} + \sqrt{P_2}X_{m2}}{\sqrt{P_1 + P_2}} \right) \right] \right|, \end{aligned} \quad (143)$$

where H_k is the Hermite polynomial of degree k (see [33, Eq. (15)]). As shown in [33], we obtain the desired exponential convergence rate of the mutual information provided that the expectation appearing in (143) is zero for odd values of k , and also for $k \leq 2m - 1$. For odd values of k , we use the same symmetry argument as that of [33]; 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

$$H_k(a + b) = \sum_{i=0}^k \sum_{j=0}^k c_{ij} a^i b^j \quad (144)$$

for some constants c_{ij} , which follows since H_k has degree k . By the independence of X_{m1} and X_{m2} , the expectation $\mathbb{E}\left[H_k\left(\frac{\sqrt{P_1}X_{m1} + \sqrt{P_2}X_{m2}}{\sqrt{P_1 + P_2}}\right)\right]$ depends only on the first k moments of X_{m1} and X_{m2} . Since the i -th moment of X_{mv} coincides with the corresponding moment of $X_v \sim N(0, P_v)$ for $i = 1, \dots, 2m - 1$ [33, Th. 2], we have for $k \leq 2m - 1$ that

$$\begin{aligned} & \mathbb{E}\left[H_k\left(\frac{\sqrt{P_1}X_{m1} + \sqrt{P_2}X_{m2}}{\sqrt{P_1 + P_2}}\right)\right] \\ &= \mathbb{E}\left[H_k\left(\frac{\sqrt{P_1}X_1 + \sqrt{P_2}X_2}{\sqrt{P_1 + P_2}}\right)\right] \end{aligned} \quad (145)$$

$$= 0, \quad (146)$$

where (146) follows since for any k , we have $H_k(X) = 0$ under $X \sim N(0, 1)$ [33]. ■

We proceed by proving that, analogously to Theorem 2, there exists $g(n) = o(\sqrt{n})$ such that all rate pairs (R_1, R_2) satisfying

$$\mathbf{R} \in n\mathbf{I} - \sqrt{n} \mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon) + g(n)\mathbf{1} \quad (147)$$

are (n, ϵ) -achievable. This is done by following the analysis of Section IV-B: We consider random coding with the constant-composition codeword distribution in (43), using (Q_{m1}, Q_{m2}) as the input distribution pair. As was done in [5, Th. 3], [23], we set $m = n^{\frac{1}{4}}$. By part (i) of Lemma 1, we have $\|n\mathbf{I}_m - n\mathbf{I}\|_{\infty} \leq ne^{-\gamma n^{\frac{1}{4}}}$, which behaves as $o(\sqrt{n})$. Similarly, parts (ii) and (iii) of Lemma 1 show that $\mathbf{V}_m \rightarrow \mathbf{V}$ and the relevant third moments associated with \mathbf{i} are bounded. The analysis of Section IV-B reveals that the remainder term $g(n)$ in (147) depends on the alphabet sizes through $(|\mathcal{X}_1| + |\mathcal{X}_2| + 2) \log n$ (see the choice of d following (49)), which is again $o(\sqrt{n})$ due to the fact that $|\mathcal{X}_1| = |\mathcal{X}_2| = n^{\frac{1}{4}}$.

Finally, using (147), we obtain (36) using identical steps to Section IV-B with $\beta = 0$ (see also [10]).

ACKNOWLEDGMENTS

The authors would like to thank Vincent Tan and Pierre Moulin for helpful discussions. In particular, the use of quantization arguments for the Gaussian case in Section III-C was recommended by Vincent.

REFERENCES

[1] R. M. Fano, *Transmission of Information: A Statistical Theory of Communication*. Cambridge, MA, USA: MIT Press, 1961.
 [2] Y. Altuğ and A. B. Wagner. (2012). "Moderate deviations in channel coding." [Online]. Available: <http://arxiv.org/abs/1208.1924>
 [3] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," in *Proc. 3rd Prague Conf. Inf. Theory*, 1962, pp. 689–723. [Online]. Available: <http://www.math.wustl.edu/~luthy/strassen.pdf>
 [4] Y. Polyanskiy, H. 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," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 881–903, Feb. 2014.
 [7] E. MolavianJazi and J. N. Laneman, "Simpler achievable rate regions for multiaccess with finite blocklength," in *Proc. IEEE Int. Symp. Inf. Theory*, Boston, MA, USA, Jul. 2012, pp. 36–40.
 [8] Y.-W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Boston, MA, USA, Jul. 2012, pp. 831–835.
 [9] W. Hoeffding, "A combinatorial central limit theorem," *Ann. Math. Statist.*, vol. 22, no. 4, pp. 558–566, 1951.
 [10] R. Nomura and T. S. Han, "Second-order Slepian–Wolf coding theorems for non-mixed and mixed sources," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 1974–1978.
 [11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.
 [12] Y.-S. Liu and B. L. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 376–386, Mar. 1996.
 [13] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inf. Theory*, Tsaghkadzor, Armenia, 1971, pp. 23–51.
 [14] H. Liao, "Multiple-access channels," Ph.D. dissertation, Dept. Elect. Eng., Univ. Hawaii, Honolulu, HI, USA, 1972.
 [15] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
 [16] J. Scarlett and V. Y. F. Tan. "Second-order asymptotics for the Gaussian MAC with degraded message sets." [Online]. Available: <http://arxiv.org/abs/1310.1197>
 [17] S.-Q. Le, V. Y. F. Tan, and M. Motani. (2014). "A case where interference does not affect the channel dispersion." [Online]. Available: <http://arxiv.org/abs/1404.0255>
 [18] E. Haim, Y. Kochman, and U. Erez, "A note on the dispersion of network problems," in *Proc. IEEE 27th Conv. Elect. Electron. Eng. Israel*, Nov. 2012, pp. 1–9.
 [19] W.-L. Loh, "On Latin hypercube sampling," *Ann. Statist.*, vol. 24, no. 5, pp. 2058–2080, 1996.
 [20] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Second-order rate region of constant-composition codes for the multiple-access channel," in *Proc. 51st Annu. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Oct. 2013, pp. 588–593.
 [21] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, Mar. 1985.
 [22] R. G. Gallager, "Fixed composition arguments and lower bounds to error probability," [Online]. Available: <http://web.mit.edu/gallager/www/notes/notes5.pdf>
 [23] V. Y. F. Tan, "Achievable second-order coding rates for the wiretap channel," in *Proc. IEEE Int. Conf. Commun. Syst.*, Singapore, Nov. 2012, pp. 65–69.
 [24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2001.
 [25] E. MolavianJazi and J. N. Laneman. (2014). "A finite-blocklength perspective on Gaussian multi-access channels." [Online]. Available: <http://arxiv.org/abs/1309.2343>
 [26] E. Bolthausen, "An estimate of the remainder in a combinatorial central limit theorem," *Probab. Theory Rel. Fields*, vol. 66, no. 3, pp. 379–386, 1984.
 [27] S.-T. Ho and L. H. Y. Chen, "An L_p bound for the remainder in a combinatorial central limit theorem," *Ann. Probab.*, vol. 6, no. 2, pp. 231–249, 1978.
 [28] E. Bolthausen and F. Götze, "The rate of convergence for multivariate sampling statistics," *Ann. Statist.*, vol. 21, no. 4, pp. 1692–1710, 1993.
 [29] B. von Bahr, "Remainder term estimate in a combinatorial limit theorem," *Probab. Theory Rel. Fields*, vol. 35, no. 2, pp. 131–139, 1976.
 [30] G. Dueck, "Maximal error capacity regions are smaller than average error capacity regions for multi-user channels," *Problems Control Inf. Theory*, vol. 7, no. 1, pp. 11–19, 1978.
 [31] R. Ahlswede, "An elementary proof of the strong converse theorem for the multiple-access channel," *J. Combinat. Inf. Syst. Sci.*, vol. 7, no. 3, pp. 216–230, 1982.
 [32] J. Scarlett, "Reliable communication under mismatched decoding," Ph.D. dissertation, Dept. Eng., Univ. Cambridge, Cambridge, U.K., 2014. [Online]. Available: <http://itc.upf.edu/biblio/1061>
 [33] Y. Wu and S. Verdú, "The impact of constellation cardinality on Gaussian channel capacity," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, 2010, pp. 620–628.

Jonathan Scarlett (S'14) was born in Melbourne, Australia, in 1988. In 2010, he received the B.Eng. degree in electrical engineering and the B.Sci. degree in computer science from the University of Melbourne, Australia. In 2011, he was a research assistant at the Department of Electrical & Electronic Engineering, University of Melbourne. From October 2011 to August 2014, he was a Ph.D. student in the Signal Processing and Communications Group at the University of Cambridge, United Kingdom. He is now a post-doctoral researcher with the Laboratory for Information and Inference Systems at the École Polytechnique Fédérale de Lausanne, Switzerland. His research interests are in the areas of information theory, signal processing, and high-dimensional statistics.

Alfonso Martinez (SM'11) was born in Zaragoza, Spain, in October 1973. He is currently a Ramón y Cajal Research Fellow at Universitat Pompeu Fabra, Barcelona, Spain. He obtained his Telecommunications Engineering degree from the University of Zaragoza in 1997. In 1998-2003 he was a Systems Engineer at the research centre of the European Space Agency (ESAESTEC) in Noordwijk, The Netherlands. His work on APSK modulation was instrumental in the definition of the physical layer of DVB-S2. From 2003 to 2007 he was a Research and Teaching Assistant at Technische Universiteit Eindhoven, The Netherlands, where he conducted research on digital signal processing for MIMO optical systems and on optical communication theory. Between 2008 and 2010 he was a post-doctoral fellow with the Informationtheoretic Learning Group at Centrum Wiskunde & Informatica (CWI), in Amsterdam, The Netherlands. In 2011 he was a Research Associate with the Signal Processing and Communications Lab at the Department of Engineering, University of Cambridge, Cambridge, U.K.

His research interests lie in the fields of information theory and coding, with emphasis on digital modulation and the analysis of mismatched decoding; in this area he has coauthored a monograph on *Bit-Interleaved Coded Modulation*. More generally, he is intrigued by the connections between information theory, optical communications, and physics, particularly by the links between classical and quantum information theory.

Albert Guillén i Fàbregas (S'01–M'05–SM'09) was born in Barcelona, Catalunya, Spain, in 1974. In 1999 he received the Telecommunication Engineering Degree and the Electronics Engineering Degree from Universitat Politècnica de Catalunya and Politecnico di Torino, respectively, and the Ph.D. in Communication Systems from École Polytechnique Fédérale de Lausanne (EPFL) in 2004.

Since 2011 he has been a Research Professor of the Institució Catalana de Recerca i Estudis Avançats (ICREA) hosted at the Department of Information and Communication Technologies, Universitat Pompeu Fabra. He is also an Adjunct Researcher at the Department of Engineering, University of Cambridge. He has held appointments at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, University of South Australia, University of Cambridge where he was a Reader and a Fellow of Trinity Hall, as well as visiting appointments at EPFL, École Nationale des Télécommunications (Paris), Universitat Pompeu Fabra, University of South Australia, Centrum Wiskunde & Informatica and Texas A&M University in Qatar. His specific research interests are in the areas of information theory, communication theory, coding theory, digital modulation and signal processing techniques.

Dr. Guillén i Fàbregas received the Starting Grant from the European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference, the 2004 Best Doctoral Thesis Award from the Spanish Institution of Telecommunications Engineers, and a Research Fellowship of the Spanish Government to join ESA. He is a Member of the Young Academy of Europe. He is a co-author of the monograph book *Bit-Interleaved Coded Modulation*. He is also an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY, an Editor of the *Foundations and Trends in Communications and Information Theory*, Now Publishers and was an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007-2011).