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

Jonathan Scarlett
University of Cambridge
jms265@cam.ac.uk

Alfonso Martinez
Universitat Pompeu Fabra
alfonso.martinez@ieee.org

Albert Guillén i Fàbregas
ICREA & Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

***Abstract*—This paper presents an achievable second-order rate region for the discrete memoryless multiple-access channel. The result is obtained using a random-coding ensemble in which each user's codebook contains codewords of a fixed composition. It is shown that this ensemble performs at least as well as i.i.d. random coding in terms of second-order asymptotics, and an example is given where a strict improvement is observed.**

I. Introduction

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

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

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

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

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

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Union's 7th Framework Programme (PEOPLE-2011-CIG) under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03.

A. Notation

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]) on $\mathcal{X}^n$ 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 $T^n(P_X)$. Given a sequence $\boldsymbol{x} \in T^n(P_X)$ and a conditional distribution $P_{Y|X}$, we define $T^n_{\boldsymbol{x}}(P_{Y|X})$ to be the set of sequences $\boldsymbol{y}$ such that $(\boldsymbol{x},\boldsymbol{y}) \in T^n(P_X \times P_{Y|X})$.

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

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

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

B. System Setup

We consider a 2-user DM-MAC $W(y|x_1,x_2)$ with input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ and output alphabet $\mathcal{Y}$. The encoders and decoder operate as follows. Encoder $\nu = 1,2$ takes as input a message m_ν equiprobable on the set $\{1,\dots,M_\nu\}$, and transmits the corresponding codeword $\boldsymbol{x}_\nu^{(m_\nu)}$ from the codebook $\mathcal{C}_\nu = \{\boldsymbol{x}_\nu^{(1)},\dots,\boldsymbol{x}_\nu^{(M_\nu)}\}$. Upon receiving $\boldsymbol{y}$ at the output of the channel, the decoder forms an estimate $(\hat{m}_1,\hat{m}_2)$ of the messages. An error is said to have occurred if the estimate $(\hat{m}_1, \hat{m}_2)$ differs from (m_1, m_2). A rate pair (R_1, R_2) is said to be (n,ϵ)-achievable if there exist codebooks with $M_1 \geq \exp(nR_1)$ and $M_2 \geq \exp(nR_2)$ codewords of length n for users 1 and 2 respectively such that the average error probability does not exceed ϵ.

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

$$P_{\boldsymbol{X}_\nu|\boldsymbol{U}}(\boldsymbol{x}_\nu|\boldsymbol{u}) = \frac{1}{|T^n_{\boldsymbol{u}}(Q_{\nu,n})|}\mathbb{1}\Big\{\boldsymbol{x}_\nu \in T^n_{\boldsymbol{u}}(Q_{\nu,n})\Big\}. \quad (2)$$

Throughout the paper, we define the joint distribution

$$P_{UX_1X_2Y}(u,x_1,x_2,y) \triangleq Q_U(u)Q_1(x_1|u)Q_2(x_2|u)W(y|x_1,x_2) \quad (3)$$

and denote the induced marginal distributions by $P_{Y|X_1U}$, $P_{Y|X_2U}$, etc. Furthermore, we define the rate vector

$$\boldsymbol{R} \triangleq \begin{bmatrix} R_1 \\ R_2 \\ R_1+R_2 \end{bmatrix} \quad (4)$$

and the information density vector

$$\boldsymbol{i}(u,x_1,x_2,y) \triangleq \begin{bmatrix} i_1(u,x_1,x_2,y) \\ i_2(u,x_1,x_2,y) \\ i_{12}(u,x_1,x_2,y) \end{bmatrix}, \quad (5)$$

where

$$i_1(u,x_1,x_2,y) \triangleq \log\frac{W(y|x_1,x_2)}{P_{Y|X_2U}(y|x_2,u)} \quad (6)$$

$$i_2(u,x_1,x_2,y) \triangleq \log\frac{W(y|x_1,x_2)}{P_{Y|X_1U}(y|x_1,u)} \quad (7)$$

$$i_{12}(u,x_1,x_2,y) \triangleq \log\frac{W(y|x_1,x_2)}{P_{Y|U}(y|u)}. \quad (8)$$

The averages of (6)–(8) with respect to $P_{UX_1X_2Y}$ are respectively given by $I(X_1;Y|X_2,U)$, $I(X_2;Y|X_1,U)$ and $I(X_1,X_2;Y|U)$.

C. Existing Results

We define the set

$$\mathsf{Q}_{\mathrm{inv}}(\boldsymbol{V},\epsilon) \triangleq \Big\{\boldsymbol{z} \in \mathbb{R}^3 : \mathbb{P}\big[\boldsymbol{Z} \preceq \boldsymbol{z}\big] \geq 1-\epsilon\Big\}, \quad (9)$$

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

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

$$n\boldsymbol{R} \in n\boldsymbol{I} - \sqrt{n}\mathsf{Q}_{\mathrm{inv}}(\boldsymbol{V},\epsilon) + g(n)\boldsymbol{1}, \quad (10)$$

are (n,ϵ)-achievable, where $\boldsymbol{R}$ is defined in (4).

Asymptotic expansions of the form (10) are somewhat more difficult to interpret than the scalar counterpart in (1) (e.g. see Haim *et al.* [12] for discussion). Roughly speaking, given a vector $\boldsymbol{I}$ and two covariance matrices $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$, $\boldsymbol{V}_1 \prec \boldsymbol{V}_2$ implies that $\boldsymbol{V}_1$ yields faster convergence to the pentagonal achievable rate region corresponding to $\boldsymbol{I}$ as n increases with $\epsilon < \frac{1}{2}$ fixed, at least in terms of second-order asymptotics.

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

$$\boldsymbol{V}^{\mathrm{iid}} = \mathbb{E}\Big[\mathrm{Cov}\big[\boldsymbol{i}(U,X_1,X_2,Y)\,\big|\,U\big]\Big] \quad (11)$$

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

For certain classes of channels, the present problem can be reduced to a single-user problem in order to obtain a matching converse to the above achievability results [12]. A more general converse containing variances of the form $\mathbb{E}[\mathrm{Var}[i_\nu(U,X_1,X_2,Y)\,|\,U,X_1,X_2]]$ $(\nu = 1,2,12)$ has recently been reported by Moulin [13], [14].

II. Main Result

The main result of this paper is the following theorem. Along with (5)–(8), we define the vectors

$$\boldsymbol{i}^{(1)}(u,x_1) \triangleq \mathbb{E}\big[\boldsymbol{i}(U,X_1,X_2,Y)\,\big|\,(U,X_1)=(u,x_1)\big] \quad (12)$$

$$\boldsymbol{i}^{(2)}(u,x_2) \triangleq \mathbb{E}\big[\boldsymbol{i}(U,X_1,X_2,Y)\,\big|\,(U,X_2)=(u,x_2)\big] \quad (13)$$

whose entries are given by

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

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

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

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

$$\boldsymbol{I} = \mathbb{E}\big[\boldsymbol{i}(U, X_1, X_2, Y)\big] \quad (16)$$

$$\boldsymbol{V} = \mathbb{E}\Big[\mathrm{Cov}\big[\boldsymbol{i}(U, X_1, X_2, Y) \,\big|\, U\big] - \mathrm{Cov}\big[\boldsymbol{i}^{(1)}(U, X_1) \,\big|\, U\big] - \mathrm{Cov}\big[\boldsymbol{i}^{(2)}(U, X_2) \,\big|\, U\big]\Big]. \quad (17)$$

Furthermore, the function $g(n)$ in (10) satisfies $g(n) = O(\log n)$ if the argument to the expectation in (17) has full rank for all $u \in \mathcal{U}$, and $g(n) = O(n^{\frac{1}{6}})$ more generally.

Proof: See Section IV. ■

The covariance matrix $\boldsymbol{V}$ in (17) can be interpreted as follows. The term $\mathrm{Cov}[\boldsymbol{i}]$ represents the variations in (X_1, X_2, Y) in the i.i.d. case, and the terms $\mathrm{Cov}[\boldsymbol{i}^{(1)}]$ and $\mathrm{Cov}[\boldsymbol{i}^{(2)}]$ represent the reduced variations in X_1 and X_2 respectively, resulting from the codewords having a fixed composition. From (11) and (17), we clearly have $\boldsymbol{V} \preceq \boldsymbol{V}^{\mathrm{iid}}$.

It is interesting to compare (17) with the conditional covariance matrix

$$\boldsymbol{V}^{\mathrm{joint}} = \mathbb{E}\bigg[\mathrm{Cov}\Big[\boldsymbol{i}(U, X_1, X_2, Y), \,\Big|\, U\Big] - \mathrm{Cov}\Big[\boldsymbol{i}^{(12)}(U, X_1, X_2) \,\Big|\, U\Big]\bigg] \quad (18)$$

$$= \mathbb{E}\bigg[\mathrm{Cov}\Big[\boldsymbol{i}(U, X_1, X_2, Y) \,\Big|\, U, X_1, X_2\Big]\bigg], \quad (19)$$

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

$$\mathrm{Cov}\Big[\boldsymbol{i}^{(12)}(u, X_1, X_2)\Big] = \mathrm{Cov}\Big[\mathbb{E}\big[\boldsymbol{i}^{(12)}(u, X_1, X_2) \,\big|\, X_1\big]\Big] + \mathbb{E}\Big[\mathrm{Cov}\big[\boldsymbol{i}^{(12)}(u, X_1, X_2) \,\big|\, X_1\big]\Big] \quad (20)$$

$$= \mathrm{Cov}\big[\boldsymbol{i}^{(1)}(u, X_1)\big] + \mathbb{E}\Big[\mathrm{Cov}\big[\boldsymbol{i}^{(12)}(u, X_1, X_2) \,\big|\, X_1\big]\Big], \quad (21)$$

where each expression is implicitly conditioned on $U = u$. The second term in (21) can be lower bounded (in the positive semidefinite sense) by $\mathrm{Cov}[\boldsymbol{i}^{(2)}(u, X_1)]$ by substituting the definitions of expectation and covariance, and using the identity $\mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}^T] \succeq \mathbb{E}[\boldsymbol{Z}]\mathbb{E}[\boldsymbol{Z}]^T$. Combined with (17) and (18), this yields the desired result.

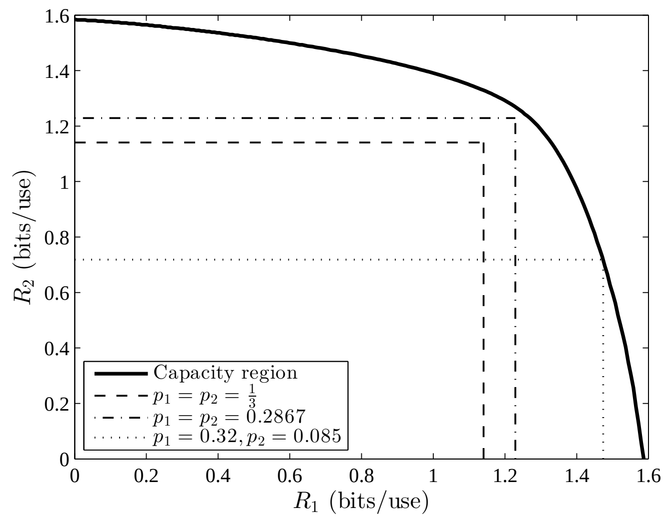

Figure 1. Capacity region of the collision channel.

III. Example: The Collision Channel

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

$$W(y|x_1, x_2) = \begin{cases} 1 & y = (x_1, x_2) \text{ and } \min\{x_1, x_2\} = 0 \\ 1 & y = \mathsf{c} \text{ and } \min\{x_1, x_2\} \neq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

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

We recall the following observations by Gallager [15]: (i) The capacity region can be obtained without time sharing;[1] (ii) By symmetry, the points on the boundary of the capacity region are achieved using $\mathcal{U} = \emptyset$ and input distributions of the form $Q_1 = (1 - 2p_1, p_1, p_1)$ and $Q_2 = (1 - 2p_2, p_2, p_2)$; (iii) The achievable rate region corresponding to any such (Q_1, Q_2) pair is rectangular. To illustrate these observations, we plot the capacity region in Figure 1, along with three achievable rate regions corresponding to particular choices of p_1 and p_2.

As n grows large, the second-order term in (10) becomes insignificant compared to the first-order term. Thus, the second-order asymptotics are most relevant for input distributions which are first-order optimal, in the sense that they achieve a point on the boundary of the capacity region. We proceed by showing that the diagonal entries of $\boldsymbol{V}$ are strictly smaller than those of $\boldsymbol{V}^{\mathrm{iid}}$ in (11) under all such input distributions. It suffices to consider the case $\mathcal{U} = \emptyset$, since otherwise the variances are simply weighted sums of the corresponding variances under $(Q_1(\cdot|u), Q_2(\cdot|u))$, weighted by Q_U. In fact, as stated above, it suffices to

[1] On the other hand, for the collision channel with K non-zero symbols, time-sharing is required for $K \geq 8$ [15].

consider distributions of the form $Q_1 = (1-2p_1, p_1, p_1)$ and $Q_2 = (1-2p_2, p_2, p_2)$.

Denote the diagonal entries of $\boldsymbol{V}$ by (V_1, V_2, V_{12}), and those of $\boldsymbol{V}^{\rm iid}$ by $(V_1^{\rm iid}, V_2^{\rm iid}, V_{12}^{\rm iid})$. We observe from (17) and (11) that for $\nu = 1,2,12$, $V_\nu \le V_\nu^{\rm iid}$ with equality if and only if $\mathrm{Var}\big[i_\nu^{(1)}(X_1)\big] = 0$ and $\mathrm{Var}\big[i_\nu^{(2)}(X_2)\big] = 0$, where the quantities $i_\nu^{(1)}$ and $i_\nu^{(2)}$ are defined as in (14)–(15) with $\mathcal{U} = \emptyset$. By a direct calculation, it can be shown that

$$i_{12}^{(1)}(x_1) = (1-2p_2)\log\frac{1}{1-2p_2} + 2p_2\log\frac{1}{p_2} + \log\frac{1}{Q_1(x_1)}, \tag{23}$$

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

IV. Proof of Theorem 1

For clarity of exposition, we present the proof in the absence of time-sharing, and we assume that the input distributions Q_1 and Q_2 are types (i.e. $Q_\nu \in \mathcal{P}_n(\mathcal{X}_\nu)$ for $\nu = 1,2$), and that $\boldsymbol{V}$ has full rank and hence $\boldsymbol{V} \succ \boldsymbol{0}$. In Section IV-C, we provide some comments regarding the general case. For $\nu = 1,2,12$, we write $i_\nu(x_1,x_2,y)$ to denote the quantities in (6)–(8) with the conditioning on u removed, and similarly for $\boldsymbol{i}(x_1,x_2,y)$.

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

$$\begin{aligned}(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}, \overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2) &\sim P_{\boldsymbol{X}_1}(\boldsymbol{x}_1)P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) \\ &\quad\times W^n(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2)P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1)P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2),\end{aligned} \tag{24}$$

where $W^n(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2) \triangleq \prod_{i=1}^n W(y_i|x_{1,i},x_{2,i})$. We make use of the threshold-based bound on the random-coding error probability $\overline{p}_e$ given in [16, Thm. 3], which is written in terms of three arbitrary output distributions $Q_{\boldsymbol{Y}|\boldsymbol{X}_2}$, $Q_{\boldsymbol{Y}|\boldsymbol{X}_1}$ and $Q_{\boldsymbol{Y}}$. Choosing these to be i.i.d. on the corresponding marginals of (3) (e.g. $P_{Y|X_2}$), we obtain

$$\begin{aligned}\overline{p}_e &\le 1 - \mathbb{P}\Big[\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}) \succ \boldsymbol{\gamma}\Big] \\ &\quad + \frac{M_1-1}{2}\mathbb{P}\Big[i_1^n(\overline{\boldsymbol{X}}_1,\boldsymbol{X}_2,\boldsymbol{Y}) > \gamma_1\Big] \\ &\quad + \frac{M_2-1}{2}\mathbb{P}\Big[i_2^n(\boldsymbol{X}_1,\overline{\boldsymbol{X}}_2,\boldsymbol{Y}) > \gamma_2\Big] \\ &\quad + \frac{(M_1-1)(M_2-1)}{2}\mathbb{P}\Big[i_{12}^n(\overline{\boldsymbol{X}}_1,\overline{\boldsymbol{X}}_2,\boldsymbol{Y}) > \gamma_{12}\Big],\end{aligned} \tag{25}$$

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

$$\boldsymbol{i}^n(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}) \triangleq \sum_{i=1}^n \boldsymbol{i}(x_{1,i},x_{2,i},y_i) \tag{26}$$

$$i_\nu^n(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}) \triangleq \sum_{i=1}^n i_\nu(x_{1,i},x_{2,i},y_i). \tag{27}$$

For $\nu = 1,2$, we have from [17, Eq. (2.4)] that the constant-composition codeword distribution $P_{\boldsymbol{X}_\nu}(\boldsymbol{x}_\nu)$ is upper bounded by a polynomial times $Q_\nu^n(\boldsymbol{x}_\nu) \triangleq \prod_{i=1}^n Q_\nu(x_{\nu,i})$. Applying this upper bound to the second, third and fourth terms in (25) and using an identical argument to [7, Eqs. (5)-(6)], we obtain

$$\overline{p}_e \le 1 - \mathbb{P}\Big[\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}) \succ \boldsymbol{\gamma}\Big] + p_0(n)\sum_{\nu=1,2,12} M_\nu e^{-\gamma_\nu}, \tag{28}$$

where $M_{12} \triangleq M_1 M_2$, and $p_0(n)$ is polynomial in n.

Using (28), the statement of the theorem will follow using identical steps to [6, Thm. 2] once we prove the following:

1) The mean and covariance of $\boldsymbol{i}^n$ respectively satisfy $\mathbb{E}[\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})] = n\boldsymbol{I} + O\big(\frac{\log n}{n}\big)$ and $\mathrm{Cov}[\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})] = n\boldsymbol{V} + O\big(\frac{\log n}{\sqrt{n}}\big)$, where $(\boldsymbol{I},\boldsymbol{V})$ are given by (16)–(17).
2) The probability on the right-hand side of (28) can be approximated using a multivariate Berry-Esseen theorem.

We prove these statements in Sections IV-A and IV-B respectively. The remaining details of the proof of Theorem 1 are omitted to avoid repetition with [6]. It should be noted that the growth rates $O\big(\frac{\log n}{n}\big)$ and $O\big(\frac{\log n}{\sqrt{n}}\big)$ in the first statement ensure that $g(n) = O(\log n)$ in (10), as stated in the theorem.

A. Calculation of Moments

The first moment of $\boldsymbol{i}^n$ (defined in (26)) is easily found by writing

$$\mathbb{E}\big[\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})\big] = \sum_{i=1}^n \mathbb{E}\big[\boldsymbol{i}(X_{1,i},X_{2,i},Y_i)\big] = n\boldsymbol{I}, \tag{29}$$

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

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

$$\begin{aligned}&\mathrm{Cov}\Big[\boldsymbol{i}^n(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y})\Big] \\ &= \mathrm{Cov}\Big[\sum_{i=1}^n \boldsymbol{i}(X_{1,i},X_{2,i},Y_i)\Big] &(30) \\ &= \sum_{i=1}^n\sum_{j=1}^n \mathrm{Cov}\Big[\boldsymbol{i}(X_{1,i},X_{2,i},Y_i), \boldsymbol{i}(X_{1,j},X_{2,j},Y_j)\Big] &(31) \\ &= n\mathrm{Cov}\Big[\boldsymbol{i}(X_1,X_2,Y)\Big] \\ &\qquad + (n^2-n)\mathrm{Cov}\Big[\boldsymbol{i}(X_1,X_2,Y), \boldsymbol{i}(X_1',X_2',Y')\Big], &(32)\end{aligned}$$

where (X_1, X_2, Y) and (X_1', X_2', Y') correspond to two arbitrary but different indices in $\{1, \cdots, n\}$. In (32), we have used the fact that, by the symmetry of the codebook construction, the n terms in (31) with $i = j$ are equal, and similarly for the $n^2 - n$ terms with $i \neq j$.

To compute the cross-covariance matrix in (32), we need the joint distribution of (X_1, X_2, Y) and (X_1', X_2', Y'). This distribution is easily understood by considering each codeword $\boldsymbol{X}_\nu$ as being generated by applying a random permutation to an arbitrary sequence $\boldsymbol{x}_\nu$ with $nQ_\nu(x_\nu)$ elements of each x_ν, i.e. $\boldsymbol{x}_\nu \in T^n(Q_\nu)$. Since such a random permutation can be done via sampling without replacement, we have

$$\mathbb{P}[X_\nu = x_\nu] = Q_\nu(x_\nu) \tag{33}$$

$$\mathbb{P}[X_\nu' = x_\nu' | X_\nu = x_\nu] = \frac{nQ_\nu(x_\nu') - \mathbb{1}\{x_\nu = x_\nu'\}}{n-1} \tag{34}$$

for $\nu = 1, 2$. Letting $Q_\nu'(x_\nu'|x_\nu)$ denote the right-hand side of (34), the cross-covariance matrix in (32) is given by

$$\begin{aligned}
&\mathrm{Cov}\Big[\boldsymbol{i}(X_1, X_2, Y), \boldsymbol{i}(X_1', X_2', Y')\Big] \\
&= \mathbb{E}\Big[\big(\boldsymbol{i}(X_1, X_2, Y) - \boldsymbol{I}\big)\big(\boldsymbol{i}(X_1', X_2', Y') - \boldsymbol{I}\big)^T\Big] && (35) \\
&= \sum_{x_1, x_2, y} Q_1(x_1)Q_2(x_2)W(y|x_1, x_2) \\
&\qquad \times \sum_{x_1', x_2', y'} Q_1'(x_1'|x_1)Q_2'(x_2'|x_2)W(y'|x_1', x_2') \\
&\qquad \times \big(\boldsymbol{i}(x_1, x_2, y) - \boldsymbol{I}\big)\big(\boldsymbol{i}(x_1', x_2', y') - \boldsymbol{I}\big)^T && (36) \\
&= \boldsymbol{M}_1 + \boldsymbol{M}_2 + \boldsymbol{M}_3 + \boldsymbol{M}_4, && (37)
\end{aligned}$$

where the four terms in (37) correspond to the four terms in the expansion of $\big(nQ_1(x_1') - \mathbb{1}\{x_1 = x_1'\}\big)\big(nQ_2(x_2') - \mathbb{1}\{x_2 = x_2'\}\big)$ resulting from (34). Specifically, we obtain

$$\boldsymbol{M}_1 = \frac{n^2}{(n-1)^2}\mathbb{E}\Big[\boldsymbol{i}(X_1, X_2, Y) - \boldsymbol{I}\Big]\mathbb{E}\Big[\boldsymbol{i}(X_1, X_2, Y) - \boldsymbol{I}\Big]^T \tag{38}$$

$$\boldsymbol{M}_2 = \frac{-n}{(n-1)^2}\mathbb{E}\Big[\big(\boldsymbol{i}(X_1, X_2, Y) - \boldsymbol{I}\big)\big(\boldsymbol{i}(\overline{X}_1, X_2, \overline{Y}) - \boldsymbol{I}\big)^T\Big] \tag{39}$$

$$\boldsymbol{M}_3 = \frac{-n}{(n-1)^2}\mathbb{E}\Big[\big(\boldsymbol{i}(X_1, X_2, Y) - \boldsymbol{I}\big)\big(\boldsymbol{i}(X_1, \overline{X}_2, \overline{\overline{Y}}) - \boldsymbol{I}\big)^T\Big] \tag{40}$$

$$\boldsymbol{M}_4 = \frac{1}{(n-1)^2}\mathbb{E}\Big[\big(\boldsymbol{i}(X_1, X_2, Y) - \boldsymbol{I}\big)\big(\boldsymbol{i}(X_1, X_2, \widetilde{Y}) - \boldsymbol{I}\big)^T\Big] \tag{41}$$

under the joint distribution

$$\begin{aligned}
&(X_1, X_2, Y, \overline{X}_1, \overline{X}_2, \overline{Y}, \overline{\overline{Y}}, \widetilde{Y}) \sim \\
&\quad Q_1(x_1)Q_2(x_2)W(y|x_1, x_2)Q_1(\overline{x}_1)Q_2(\overline{x}_2) \\
&\quad \times W(\overline{y}|\overline{x}_1, x_2)W(\overline{\overline{y}}|x_1, \overline{x}_2)W(\widetilde{y}|x_1, x_2). && (42)
\end{aligned}$$

We observe that $\boldsymbol{M}_1$ is the zero matrix, and $\boldsymbol{M}_4 = O\big(\frac{1}{n^2}\big)$. Furthermore, recalling the definitions of $\boldsymbol{i}^{(1)}$ and $\boldsymbol{i}^{(2)}$ in (14)–(15), we have

$$\begin{aligned}
\frac{-(n-1)^2}{n}\boldsymbol{M}_2 &= \mathbb{E}\Big[\mathbb{E}\Big[\big(\boldsymbol{i}(X_1, X_2, Y) - \boldsymbol{I}\big) \,\Big|\, X_2\Big] \\
&\qquad \times \mathbb{E}\Big[\big(\boldsymbol{i}(\overline{X}_1, X_2, \overline{Y}) - \boldsymbol{I}\big) \,\Big|\, X_2\Big]^T\Big] && (43) \\
&= \mathrm{Cov}\big[\boldsymbol{i}^{(2)}(X_2)\big]. && (44)
\end{aligned}$$

It follows that

$$\boldsymbol{M}_2 = \frac{-n}{(n-1)^2}\mathrm{Cov}\big[\boldsymbol{i}^{(2)}(X_2)\big], \tag{45}$$

and we similarly have

$$\boldsymbol{M}_3 = \frac{-n}{(n-1)^2}\mathrm{Cov}\big[\boldsymbol{i}^{(1)}(X_1)\big]. \tag{46}$$

Using the identity $\frac{n}{(n-1)^2} = \frac{1}{n} + O\big(\frac{1}{n^2}\big)$ and combining (32), (37), (45) and (46), we obtain

$$\mathrm{Cov}\big[\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})\big] = n\boldsymbol{V} + O(1), \tag{47}$$

where $\boldsymbol{V}$ is defined as in (17) with $\mathcal{U} = \emptyset$.

B. A Combinatorial Berry-Esseen Theorem

Before stating the required Berry-Esseen theorem, we outline some of the relevant literature. A combinatorial CLT was given by Hoeffding [10], who proved the asymptotic normality of random variables of the form $\sum_{j=1}^n f_n(j, \pi(j))$, where f_n is a real-valued function taking arguments on $1, \cdots, n$, and $\pi(\cdot)$ is uniformly distributed on the set of permutations of $\{1, \cdots, n\}$. A Berry-Esseen theorem was given by Bolthausen [18], and an extension to the multivariate setting was given by Bolthausen and Götze [19].

A more general setting is that in which each $f_n(j_1, j_2)$ is replaced by a random variable $Z_n(j_1, j_2)$, independent of $\pi(\cdot)$, such that $Z_n(j_1, j_2)$ is independent of $Z_n(j_1', j_2')$ whenever $(j_1, j_2) \neq (j_1', j_2')$. Berry-Esseen theorems for this setting were given by von Bahr [20] and Ho and Chen [21]. The analysis of each scalar quantity $i_\nu^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$ (see (24) and (27)) falls into this setting upon identifying

$$Z_n(j_1, j_2) = i_\nu\big(x_{1,j_1}, x_{2,j_2}, Y_n(j_1, j_2)\big), \tag{48}$$

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

In our setting, each $Z_n(j_1, j_2)$ must be replaced by a random vector $\boldsymbol{Z}_n(j_1, j_2)$ in $\mathbb{R}^3$. The desired Berry-Esseen theorem is a special case of a more general result by Loh [22, Thm. 2] for a problem known as Latin hypercube sampling. We define

$$\boldsymbol{\Sigma}_n \triangleq \frac{1}{n}\mathrm{Cov}\big[\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})\big] \tag{49}$$

$$\widehat{\boldsymbol{S}}_n \triangleq \boldsymbol{\Sigma}_n^{-\frac{1}{2}}\bigg(\frac{1}{\sqrt{n}}\big(\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) - n\boldsymbol{I}\big)\bigg), \tag{50}$$

where $(\cdot)^{-\frac{1}{2}}$ denotes the inverse of the positive semidefinite square root. From (47), we have $\mathbf{\Sigma}_n = \boldsymbol{V} + O(n^{-1})$.

Theorem 2. (Corollary of [22, Thm. 2]) *Let the input distributions Q_1 and Q_2 be given, and consider the quantities $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$, $(\boldsymbol{I}, \boldsymbol{V})$ and $(\mathbf{\Sigma}_n, \widehat{\boldsymbol{S}}_n)$ respectively defined in* (24), (16)–(17) *and* (49)–(50). *If $\boldsymbol{V} \succ \mathbf{0}$, then we have for sufficiently large n that*

$$\left|\mathbb{P}\big[\widehat{\boldsymbol{S}}_n \in \mathcal{A}\big] - \mathbb{P}\big[\boldsymbol{Z} \in \mathcal{A}\big]\right| \le \frac{K}{\sqrt{n}} \tag{51}$$

for any convex, Borel measurable set $\mathcal{A} \subseteq \mathbb{R}^d$, where $\boldsymbol{Z} \sim N(\mathbf{0}, \mathbb{I}_{3\times 3})$, and K is a constant depending only on $\boldsymbol{V}$ and the alphabet sizes $|\mathcal{X}_1|$, $|\mathcal{X}_2|$ and $|\mathcal{Y}|$.

Recovering Theorem 2 from [22, Thm. 2] is non-trivial, and the details are omitted here for the sake of brevity. In the more general setting of [22], the constant K is written in terms of the third moment of a random variable. However, in the present setting, this third moment can be uniformly bounded in terms of the alphabet sizes [6, Appendix D].

C. General Case

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

In general, the dispersion matrix $\boldsymbol{V}$ may not have full rank, in which case Theorem 2 does not directly apply. However, we can deal with this case by reducing the problem to a lower dimension, similarly to [6, Sec. VIII-A]. The argument is slightly more involved in our setting, since $n\boldsymbol{V}$ is not necessarily the exact covariance matrix of $\boldsymbol{i}^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$, due to the additional $O(1)$ term in (47). By carefully handling this remainder term, we obtain the same result with a third-order term of $g(n) = O(n^{\frac{1}{6}})$, as stated in Theorem 1.

Finally, consider the case that $\mathcal{U} \neq \emptyset$, and thus the codewords are drawn uniformly over the conditional type class $T_{\boldsymbol{u}}(Q_\nu)$ for some $\boldsymbol{u} \in T^n(Q_U)$. In this case, the procedure described in Section IV-A for generating a codeword uniformly over the type class should be modified as follows. Let $\boldsymbol{x}$ be an arbitrary element of the conditional type class $T_{\boldsymbol{u}}(\cdot)$. Instead of randomly permuting the entire sequence $\boldsymbol{x}$, a random permutation of the subsequence $\boldsymbol{x}^{(u)}$ corresponding to the indices where $\boldsymbol{u}$ equals u is applied independently for each value of $u \in \mathcal{U}$. Repeating the analysis of this section with each such subsequence handled separately, we obtain the more general result of Theorem 1.

Acknowledgments

We thank Vincent Tan and Pierre Moulin for helpful discussions.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

[14] P. Moulin, Personal Communication.

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

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

[17] G. Poltyrev, "Random coding bounds for discrete memoryless channels," *Prob. Inf. Transm.*, vol. 18, no. 1, pp. 9–21, 1982.

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

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

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

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

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