Optimal Achievable Rates for Computation With Random Homologous Codes

Pinar Sen
Department of Electrical and Computer Engineering
University of California, San Diego, USA
Email: psen@ucsd.edu

Sung Hoon Lim
Korea Institute of Ocean Science and Technology
Busan, Korea
Email: shlim@kiost.ac.kr

Young-Han Kim
Department of Electrical and Computer Engineering
University of California, San Diego, USA
Email: yhk@ucsd.edu

***Abstract*—Recent studies by Padakandla and Pradhan, and by Lim, Feng, Pastore, Nazer, and Gastpar built the framework of nested coset codes for the computation problem, namely, computing a desired linear combination of sources over a multiple access channel. This paper presents an outer bound on the optimal rate region for the computation problem when the encoding strategy is restricted to random ensembles of homologous codes, namely, structured nested coset codes from the same generator matrix and individual shaping functions based on joint typicality encoding. The optimal rate region is characterized when the desired linear combination and the channel structure are matched. Under this condition, a suboptimal joint typicality decoding rule is shown to achieve the optimal rate region. This result implies that the performance of random homologous code ensembles cannot be improved by using the optimal maximum likelihood decoder for the aforementioned class of computation problems.**

I. Introduction

Consider a communication problem over a two-sender multiple access channel (MAC) where the receiver is interested in computing a function of the transmitted sources. One trivial approach to this *computation* problem would be to decode for the arguments of the function, i.e., individual sources, and then compute the function from the recovered sources. For the first communication step of this approach, the conventional random independently and identically distributed (i.i.d.) code ensembles achieve the optimal performance [1]–[3]. Studying the dual problem of encoding modulo-two sum of distributed binary sources, however, Körner and Marton [4] showed that using the same linear code at multiple users can achieve strictly better performance than random i.i.d. code ensembles for a class of source distributions. Building on this observation, Nazer and Gastpar [5] developed a channel coding technique for the computation problem over *linear* MACs that is based on linear codes and outperforms previously known schemes.

Nested coset codes [6], [7] bring a new dimension to such structured coding schemes. Using these codes, Padakandla and Pradhan [7] developed a linear computation scheme for an *arbitrary* MAC. In this coding scheme, a coset code of a rate higher than the target is first generated randomly. A codeword of a desired property (such as type or joint type) is then selected from a subset (a coset of a subcode). Similar selection techniques were also developed in the context of lattice codes in [8]. Although reminiscent of the multicoding scheme in Gelfand–Pinsker coding for channels with state and Marton coding for broadcast channels, this construction is more fundamental in the sense that the scheme is useful even for single-user communication. When dealing with multiple users, the structure between users' codebooks can be controlled via common and individual parts of the underlying linear code. Recent efforts exploited the benefit of nested coset codes for a broader class of applications: such as interference channels [9], [10], multiple access channels [11], [12], and multiple access channels with state [13].

With the main motivation of developing a unified compute–forward framework for relay networks, Lim, Feng, Pastore, Nazer, and Gastpar [14], [15] generalized the nested coset codes of the same generator matrix (which we referred to as *homologous* codes [11]) for asymmetric rate pairs. They also developed stronger analysis tools for *simultaneous decoding* of random homologous code ensembles that led to a unified scheme with performance improvement from its predecessors within the context of linear computation problems. The resulting performance of homologous codes, when adapted to the Gaussian case, improves upon the lattice codes, another structured coding scheme proposed for compute–forward [16].

We are now equipped with random homologous codes—superior alternatives to conventional random codes—the use of which brings us one step closer to understanding the fundamental limits of computation problems. Nonetheless, several questions remain open: What is the simultaneously achievable rate pairs for reliable computation (even linear computation)? Which scheme achieves this computation capacity region? The answers require a joint optimization of encoder and decoder designs, which is in terra incognita.

In this paper, we instead concentrate on the performance of the optimal maximum likelihood decoder for linear computation when the encoder is restricted to realizations of a given random homologous code ensemble. We characterize the optimal rate region when the desired linear combination and the channel structure are matched, which is the case in which the benefit of computation can be realized to the full

The work in this paper was supported in part by the Electronics and Telecommunications Research Institute through Grant 17ZF1100 from the Korean Ministry of Science, ICT, and Future Planning, and in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology under Grant NRF-2017R1C1B1004192.

978-1-5386-4780-6/18/$31.00©2018 IEEE

extent as indicated by [17]. This result, inter alia, implies that the suboptimal joint typicality decoding rule proposed in [14], [15] achieves this optimal rate region. Thus, the performance of random homologous code ensembles cannot be improved by the maximum likelihood decoder for the aforementioned class of linear computation problems.

We adapt the notation in [1], [2]. The set of integers $\{1, 2, \ldots, n\}$ is denoted by $[1 : n]$. For a length-n sequence (vector) $x^n = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$, we define its type as $\pi(x|x^n) = |\{i\colon x_i = x\}|/n$ for $x \in \mathcal{X}$. Upper case letters, e.g., X, Y, denote random variables. For $\epsilon \in (0, 1)$, we define $\mathcal{T}_\epsilon^{(n)}(X) = \{x^n\colon |p(x) - \pi(x|x^n)| \le \epsilon p(x),\ x \in \mathcal{X}\}$. The function $\mathbb{1}_{\mathcal{S}} : \mathcal{X} \to \{0, 1\}$ is defined as $\mathbb{1}_{\mathcal{S}}(x) = 1$ if $x \in \mathcal{S}$ and 0 otherwise. A length-n vector of all zeros is denoted by $\mathbf{0}_n$, where the subscript is omitted when it is clear in the context. $\mathbb{F}_q$ denotes a finite field of size q, where $\hat{\mathbb{F}}_q \doteq \mathbb{F}_q \backslash \{0\}$. For set $\mathcal{A}$, $\mathrm{cl}(\mathcal{A})$ denotes the closure of $\mathcal{A}$. We use $\epsilon_n \ge 0$ to denote a generic function of n that tends to zero as $n \to \infty$. Throughout information measures are in log base q.

II. Formal Statement of the Problem

Consider the two-sender finite-field input memoryless multiple access channel (MAC)

$$(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y}),$$

which consists of two sender alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{F}_q$, where $\mathbb{F}_q$ denotes a finite field of size q, a receiver alphabet $\mathcal{Y}$, and a collection of conditional probability distributions $p_{Y|X_1,X_2}(y|x_1, x_2)$.

Each sender $j = 1, 2$ encodes a message $M_j \in \mathbb{F}_q^{nR_j}$ into a codeword $X_j^n = x_j^n(M_j) \in \mathbb{F}_q^n$ and transmits X_j^n over the channel. The goal of communication is to convey a linear combination of the codewords rather than the messages. Hence, the receiver finds an estimate $\hat{W}_{\mathbf{a}}^n = \hat{w}_{\mathbf{a}}^n(Y^n) \in \mathbb{F}_q^n$ of

$$W_{\mathbf{a}}^n \triangleq a_1 X_1^n \oplus a_2 X_2^n,$$

for a desired (nonzero) vector $\mathbf{a} = [a_1\ a_2]$ over $\mathbb{F}_q$. The encoding maps $x_j^n(m_j)$, $j = 1, 2$, and the decoding map $\hat{w}_{\mathbf{a}}^n(y^n)$ define an (n, nR_1, nR_2) *computation code* of the multiple access channel. The set $\mathsf{C}_n = \{(x_1^n(m_1), x_2^n(m_2)) : m_1 \in \mathbb{F}_q^{(nR_1)\times(nR_2)}\}$ is referred to as the *codebook* associated with the (n, nR_1, nR_2) code.

Remark 1: For the simplicity of presentation, we consider the case $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{F}_q$, but our arguments can be extended to arbitrary $\mathcal{X}_1$ and $\mathcal{X}_2$ through the channel transformation technique by Gallager [18, Sec. 6.2]. Given a pair of symbol-by-symbol mappings $\varphi_j : \mathbb{F}_q \to \mathcal{X}_j$, $j = 1, 2$, consider the *virtual channel* with finite field inputs, $p(y|v_1, v_2) = p_{Y|X_1,X_2}(y|\varphi_1(v_1), \varphi_2(v_2))$, for which a computation code is to be defined. The goal of the communication is to convey $W_{\mathbf{a}} = a_1 V_1^n \oplus a_2 V_2^n$, where $V_j^n = v_j^n(M_j) \in \mathbb{F}_q^n$ is the virtual codeword mapped to message M_j at sender $j = 1, 2$. The results can be applied to this computation problem defined on the virtual channel.

The performance of a computation code is measured by the average probability of error

$$P_e^{(n)}(\mathsf{C}_n) = \mathsf{P}(\hat{W}_{\mathbf{a}}^n \ne W_{\mathbf{a}}^n | \mathsf{C}_n),$$

when M_1 and M_2 are independent and uniformly distributed. A rate pair (R_1, R_2) is said to be *achievable* if there exists a sequence of (n, nR_1, nR_2) computation codes such that

$$\lim_{n\to\infty} P_e^{(n)}(\mathsf{C}_n) = 0$$

and

$$\lim_{n\to\infty} H(M_j | x_j^n(M_j), \mathsf{C}_n) = 0, \quad j \in \{1, 2\} \text{ with } a_j \ne 0. \tag{1}$$

Note that without the condition in (1), the problem is trivial and an arbitrarily large rate pair is achievable.

We now consider a random ensemble of computation codes with the following structure. Let $p = p(x_1)p(x_2)$ be a given input pmf on $\mathbb{F}_q \times \mathbb{F}_q$, and let $\epsilon > 0$. Suppose that the codewords $X_1^n(m_1)$, $m_1 \in \mathbb{F}_q^{nR_1}$, and $X_2^n(m_2)$, $m_2 \in \mathbb{F}_q^{nR_2}$ that constitute the codebook are generated randomly as follows:

- Let $\hat{R}_j = D(p_{X_j} || \mathrm{Unif}(\mathbb{F}_q)) + \epsilon$, $j = 1, 2$.
- Let $\kappa = \max\{nR_1 + n\hat{R}_1, nR_2 + n\hat{R}_2\}$. Randomly generate a $\kappa \times n$ matrix G, and two vectors D_1^n and D_2^n such that elements of G, D_1^n, and D_2^n are i.i.d. $\mathrm{Unif}(\mathbb{F}_q)$.
- Given realizations G, d_1^n, and d_2^n for random matrix G, and random dithers D_1^n and D_2^n, let

  $$u_j^n(m_j, l_j) = [m_j\ l_j\ \mathbf{0}]\mathsf{G} + d_j^n, \quad m_j \in \mathbb{F}_q^{nR_j}, l_j \in \mathbb{F}_q^{n\hat{R}_j},$$

  for $j = 1, 2$. At sender $j = 1, 2$, assign the codeword $X_j^n(m_j) = U_j^n(m_j, L_j(m_j))$ to each message $m_j \in \mathbb{F}_q^{nR_j}$, where $L_j(m_j)$ is drawn uniformly at random among l_j indices satisfying $u_j^n(m_j, l_j) \in \mathcal{T}_\epsilon^{(n)}(X_j)$ if there exists one, or among $\mathbb{F}_q^{n\hat{R}_j}$ otherwise.

With a slight abuse of terminology, we refer to the random tuple $\mathcal{C}_n = (G, D_1^n, D_2^n, (L_1(m_1) : m_1 \in \mathbb{F}_q^{nR_1}), (L_2(m_2) : m_2 \in \mathbb{F}_q^{nR_2}))$ as the *random homologous codebook*. Each realization of the random homologous codebook $\mathcal{C}_n$ results in one instance $\{x_1^n(m_1), x_2^n(m_2) : (m_1, m_2) \in \mathbb{F}_q^{nR_1} \times \mathbb{F}_q^{nR_2}\}$ of such generated codebooks, which constitutes an (n, nR_1, nR_2) computation code along with the optimal decoder. The random code ensemble generated in this manner is referred to as an $(n, nR_1, nR_2; p, \epsilon)$ *random homologous code ensemble*. A rate pair (R_1, R_2) is said to be *achievable by the (p, ϵ)-distributed random homologous code ensemble* if there exits a sequence of $(n, nR_1, nR_2; p, \epsilon)$ random homologous code ensembles such that

$$\lim_{n\to\infty} \mathsf{E}_{\mathcal{C}_n}[P_e^{(n)}(\mathcal{C}_n)] = 0$$

and

$$\lim_{n\to\infty} H(M_j | X_j^n(M_j), \mathcal{C}_n) = 0, \ \ j \in \{1, 2\} \text{ with } a_j \ne 0,$$

where the expectation is with respect to the random homologous codebook $\mathcal{C}_n$. Given (p, ϵ), let $\mathscr{R}^*(p, \epsilon)$ be the set

of all rate pairs achievable by the (p,ϵ)-distributed random homologous code ensemble. We define the limit

$$\lim_{\epsilon\to 0}\mathscr{R}^*(p,\epsilon)=\bigcup_{\epsilon>0}\bigcap_{0<\gamma<\epsilon}\mathscr{R}^*(p,\gamma)\overset{(a)}{=}\bigcap_{\epsilon>0}\bigcup_{0<\gamma<\epsilon}\mathscr{R}^*(p,\gamma), \tag{2}$$

which exists if (a) holds. Given p, the optimal rate region $\mathscr{R}^*(p)$, when it exists, is defined as

$$\mathscr{R}^*(p)=\mathrm{cl}\left[\lim_{\epsilon\to 0}\mathscr{R}^*(p,\epsilon)\right].$$

III. The Main Result

A linear combination $W_{\mathbf{a}}=a_1X_1\oplus a_2X_2$ is said to be *natural* if

$$H(W_{\mathbf{a}}|Y)=\min_{\mathbf{b}\neq\mathbf{0}}H(W_{\mathbf{b}}|Y),$$

where $\mathbf{b}=[b_1\ b_2]$, and $W_{\mathbf{b}}=b_1X_1\oplus b_2X_2$. In a sense, a natural combination $W_{\mathbf{a}}$ is the best matched to the channel structure and thus the easiest to recover at the receiver.

Theorem 1: Given an input pmf $p=p(x_1)p(x_2)$, the optimal rate region $\mathscr{R}^*(p)$ for computation of a natural combination $W_{\mathbf{a}}$ is the set of rate pairs (R_1,R_2) such that

$$\begin{aligned}R_j&\le I(X_j;Y|X_{j^c}),\\ R_j&\le I(X_1,X_2;Y)-\min\{R_{j^c},I(X_{j^c};W_{\mathbf{a}},Y)\},\end{aligned}$$

for every $j\in\{1,2\}$ with $a_j\neq 0$, where $j^c=\{1,2\}\setminus\{j\}$.

We present an equivalent characterization of the optimal rate region in Theorem 1. Let $\mathscr{R}_{\mathrm{CF}}(p)$ be the set of rate pairs (R_1,R_2) such that

$$R_j\le H(X_j)-H(W_{\mathbf{a}}|Y),\quad \forall j\in\{1,2\}:a_j\neq 0.$$

Let $\mathscr{R}_{\mathrm{MAC}}(p)$ be the set of rate pairs (R_1,R_2) such that

$$\begin{aligned}R_1&\le I(X_1;Y|X_2),\\ R_2&\le I(X_2;Y|X_1),\\ R_1+R_2&\le I(X_1,X_2;Y).\end{aligned}$$

Proposition 1 ([19]): Given an input pmf $p=p(x_1)p(x_2)$, the optimal rate region $\mathscr{R}^*(p)$ for computation of a natural combination $W_{\mathbf{a}}$ can be equivalently expressed as

$$\mathscr{R}^*(p)=[\mathscr{R}_{\mathrm{CF}}(p)\cup\mathscr{R}_{\mathrm{MAC}}(p)].$$

IV. Proof of Theorem 1

We prove Theorem 1 in three steps: 1) we first prove the achievability in Section IV-A, where we follow the results in [14], [15] that studied the rate region achievable by random homologous code ensemble using a suboptimal joint typicality decoding rule, 2) we then show by Lemma 1 that the achievable rate region is equivalent to $\mathscr{R}^*(p)$ in Proposition 1 if $W_{\mathbf{a}}$ is a natural combination, and 3) we present the proof of the converse in Section IV-B by showing that if a rate pair (R_1,R_2) is achievable by the (p,ϵ)-distributed random homologous code ensemble for arbitrarily small ϵ, then (R_1,R_2) must lie in $\mathscr{R}^*(p)$.

A. Proof of Achievability

The performance of random homologous code ensembles was studied using a suboptimal *joint typicality* decoder by Lim et al. in [14], [15]. We prove the achievability of $\mathscr{R}^*(p)$ using the results in [14], [15]. For completeness, we first describe the joint typicality decoding rule and then characterize the rate region achievable by the (p,ϵ)-distributed random homologous code ensemble when decoder is specialized to the joint typicality decoder. We then concentrate on arbitrarily small ϵ to provide a lower bound on the rate region $\mathscr{R}^*(p)$.

Upon receiving y^n, the ϵ'-joint typicality decoder, $\epsilon'>0$, looks for a unique vector $s\in\mathbb{F}_q^{\kappa}$ such that

$$s=a_1[m_1\ l_1\ \mathbf{0}]\oplus a_2[m_2\ l_2\ \mathbf{0}],$$

for some $(m_1,l_1,m_2,l_2)\in\mathbb{F}_q^{nR_1}\times\mathbb{F}_q^{n\hat{R}_1}\times\mathbb{F}_q^{nR_2}\times\mathbb{F}_q^{n\hat{R}_2}$ that satisfies

$$(u_1^n(m_1,l_1),u_2^n(m_2,l_2),y^n)\in\mathcal{T}_{\epsilon'}^{(n)}(X_1,X_2,Y).$$

If the decoder finds such s, then it declares $\hat{w}_{\mathbf{a}}^n=s\mathsf{G}\oplus a_1d_1^n\oplus a_2d_2^n$ as an estimate; otherwise, it declares an error.

To describe the performance of the joint typicality decoder, we define $\mathscr{R}_{\mathrm{CF}}(p,\delta)$ for a given input pmf p and $\delta\ge 0$ as the set of rate pairs (R_1,R_2) such that

$$R_j\le H(X_j)-H(W_{\mathbf{a}}|Y)-\delta,\quad \forall j\in\{1,2\}:a_j\neq 0.$$

Similarly, we define $\mathscr{R}_1(p,\delta)$ as the set of rate pairs (R_1,R_2) such that

$$\begin{aligned}R_1&\le I(X_1;Y|X_2)-\delta,\\ R_2&\le I(X_2;Y|X_1)-\delta,\\ R_1+R_2&\le I(X_1,X_2;Y)-\delta,\\ R_1&\le I(X_1,X_2;Y)-H(X_2)+\min_{\mathbf{b}\in\hat{\mathbb{F}}_q^{1\times 2}}H(W_{\mathbf{b}}|Y)-\delta,\end{aligned}$$

and $\mathscr{R}_2(p,\delta)$ as the set of rate pairs (R_1,R_2) such that

$$\begin{aligned}R_1&\le I(X_1;Y|X_2)-\delta,\\ R_2&\le I(X_2;Y|X_1)-\delta,\\ R_1+R_2&\le I(X_1,X_2;Y)-\delta,\\ R_2&\le I(X_1,X_2;Y)-H(X_1)+\min_{\mathbf{b}\in\hat{\mathbb{F}}_q^{1\times 2}}H(W_{\mathbf{b}}|Y)-\delta,\end{aligned}$$

where $\mathbf{b}=[b_1\ b_2]$, and $W_{\mathbf{b}}=b_1X_1\oplus b_2X_2$. We are now ready to state the rate region achievable by the random homologous code ensembles, which follows by [14] and [15].

Theorem 2 ([14], [15]): Let $p=p(x_1)p(x_2)$ be an input pmf and $\delta>0$. Then, there exists $\eta(\delta)<\delta$ such that for every $\epsilon<\eta(\delta)$, a rate pair

$$(R_1,R_2)\in\mathscr{R}_{\mathrm{CF}}(p,\delta)\cup\mathscr{R}_1(p,\delta)\cup\mathscr{R}_2(p,\delta)$$

is achievable by the (p,ϵ)-distributed random homologous code ensemble along with the ϵ'-joint typicality decoder for some ϵ' such that $\epsilon<\epsilon'<\delta$.

Taking the union over all $\delta>0$, Theorem 2 implies

$$\bigcup_{\delta>0}[\mathscr{R}_{\mathrm{CF}}(p,\delta)\cup\mathscr{R}_1(p,\delta)\cup\mathscr{R}_2(p,\delta)]\subseteq\bigcup_{\epsilon>0}\bigcap_{0<\gamma<\epsilon}\mathscr{R}^*(p,\gamma).$$

Therefore, the closure of the left hand side, which is equal to the rate region $\mathscr{R}_{\rm CF}(p) \cup \mathscr{R}_1(p) \cup \mathscr{R}_2(p)$, constitutes an inner bound on the optimal rate region $\mathscr{R}^*(p)$, where the region $\mathscr{R}_{\rm CF}(p) = \mathscr{R}_{\rm CF}(p, \delta = 0)$ is as defined in Section III, and $\mathscr{R}_j(p)$ denotes the region $\mathscr{R}_j(p, \delta = 0)$ for $j = 1, 2$.

The achievability proof of Theorem 1 (for the equivalent rate region in Proposition 1) finally can be established by the next lemma.

Lemma 1 ([19]): If the desired linear combination $W_{\mathbf{a}} = a_1 X_1 \oplus a_2 X_2$ for $(a_1, a_2) \neq (0, 0)$ is natural, then

$$[\mathscr{R}_{\rm CF}(p) \cup \mathscr{R}_1(p) \cup \mathscr{R}_2(p)] = [\mathscr{R}_{\rm CF}(p) \cup \mathscr{R}_{\rm MAC}(p)].$$

B. Proof of the Converse

We first present an outer bound on the rate region $\mathscr{R}^*(p, \epsilon)$ for a fixed input pmf p and $\epsilon > 0$. We then discuss the limit of this outer bound as $\epsilon \to 0$ to establish an outer bound on the rate region $\mathscr{R}^*(p)$. Throughout this section, $\delta_i(\epsilon) \geq 0$, $i \in \mathbb{Z}^+$, denotes a continuous function of ϵ that tends to zero as $\epsilon \to 0$. Given an input pmf p and $\delta > 0$, we define the rate region $\mathscr{R}_{\rm OUT}(p, \delta)$ as the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_j &\leq I(X_j; Y | X_{j^c}) + \delta, \\ R_j &\leq I(X_1, X_2; Y) - \min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} + \delta, \end{aligned} \quad (3)$$

for every $j \in \{1, 2\}$ with $a_j \neq 0$, where $j^c = \{1, 2\} \setminus \{j\}$.

We are now ready to state the main theorem that leads to the proof of the converse for the optimal rate region $\mathscr{R}^*(p)$.

Theorem 3: Let $p = p(x_1)p(x_2)$ be an input pmf and $\epsilon > 0$. If a rate pair (R_1, R_2) is achievable by the (p, ϵ)-distributed random homologous code ensemble, then there exists a continuous $\delta'(\epsilon)$ that monotonically tends to zero as $\epsilon \to 0$ such that

$$(R_1, R_2) \in \mathscr{R}_{\rm OUT}(p, \delta'(\epsilon)).$$

Proof: Let $\epsilon' > \epsilon$. Define the indicator random variable

$$E_n = \mathbb{1}_{\{(X_1^n(M_1), X_2^n(M_2)) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2)\}}. \quad (4)$$

Since $\hat{R}_i = D(p_{X_i} || \mathrm{Unif}(\mathbb{F}_q)) + \epsilon$, $i = 1, 2$, by Lemma 12 in [14], $\mathsf{P}(E_n = 0)$ tends to zero as $n \to \infty$ if $\epsilon' \in (\epsilon \ \ \delta_1(\epsilon))$ is sufficiently large. Let $j \in \{1, 2\}$ such that $a_j \neq 0$. Then, for n sufficiently large, we have

$$\begin{aligned} nR_j &= H(M_j | M_{j^c}, \mathcal{C}_n) \\ &= I(M_j; Y^n | M_{j^c}, \mathcal{C}_n) + H(M_j | Y^n, M_{j^c}, \mathcal{C}_n) \\ &\overset{(a)}{\leq} I(M_j; Y^n | M_{j^c}, \mathcal{C}_n, E_n = 1) + n\epsilon_n \\ &= \sum_{i=1}^{n} I(M_j; Y_i | Y^{i-1}, M_{j^c}, \mathcal{C}_n, X_{j^c i}, E_n = 1) + n\epsilon_n \\ &\leq \sum_{i=1}^{n} I(M_1, M_2, X_{ji}, Y^{i-1}, \mathcal{C}_n; Y_i | X_{j^c i}, E_n = 1) + n\epsilon_n \\ &\overset{(b)}{=} \sum_{i=1}^{n} I(X_{ji}; Y_i | X_{j^c i}, E_n = 1) + n\epsilon_n \\ &\overset{(c)}{\leq} n(I(X_j; Y | X_{j^c}) + \delta_2(\epsilon)) + n\epsilon_n, \end{aligned} \quad (5)$$

where (a) follows for large n by the fact that E_n is a binary random variable with $\mathsf{P}(E_n = 0) \to 0$ as $n \to \infty$, and by Lemma 3 in Appendix A, (b) follows since $(M_1, M_2, Y^{i-1}, \mathcal{C}_n) \to (X_{1i}, X_{2i}) \to Y_i$ form a Markov chain for $i \in [1 : n]$, and (c) follows by Lemma 4 in Appendix A.

For the proof of (3), we start with

$$\begin{aligned} nR_j &= H(M_j | M_{j^c}, \mathcal{C}_n) \\ &\overset{(a)}{\leq} I(M_j; Y^n | M_{j^c}, \mathcal{C}_n) + n\epsilon_n \\ &= I(M_1, M_2; Y^n | \mathcal{C}_n) - I(M_{j^c}; Y^n | \mathcal{C}_n) + n\epsilon_n, \end{aligned} \quad (6)$$

where (a) follows by Lemma 3 in Appendix A. Following arguments similar to (5), the first term can be bounded as

$$I(M_1, M_2; Y^n | \mathcal{C}_n) \leq n(I(X_1, X_2; Y) + \delta_3(\epsilon)) + n\epsilon_n. \quad (7)$$

For the second term in (6), we need the following lemma, which is proved in Appendix B.

Lemma 2: For every $\epsilon'' > \epsilon'$, for n sufficiently large,

$$I(M_{j^c}; Y^n | \mathcal{C}_n) \geq n[\min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} - \delta_4(\epsilon'')] - n\epsilon_n.$$

Combining (6), (7), and Lemma 2 with $\epsilon'' = \delta_1(\epsilon)$, we have

$$\begin{aligned} nR_j \leq & n(I(X_1, X_2; Y) + \delta_3(\epsilon)) \\ & - n[\min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} - \delta_5(\epsilon)] + 2n\epsilon_n, \end{aligned} \quad (8)$$

for n sufficiently large. Letting $n \to \infty$ in (5) and (8) establishes

$$\begin{aligned} R_j &\leq I(X_j; Y | X_{j^c}) + \delta_2(\epsilon), \\ R_j &\leq I(X_1, X_2; Y) - \min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} + \delta_6(\epsilon). \end{aligned}$$

The claim follows by taking a continuous monotonic function $\delta'(\epsilon) \geq \max\{\delta_2(\epsilon), \delta_6(\epsilon)\}$ that tends to zero as $\epsilon \to 0$. ■

The proof of the converse for the optimal rate region $\mathscr{R}^*(p)$ in Theorem 1 follows by letting $\epsilon \to 0$ since

$$\lim_{\epsilon \to 0} \mathscr{R}_{\rm OUT}(p, \delta'(\epsilon)) = \mathscr{R}^*(p),$$

where the limit on the left hand side is defined in a similar manner to (2).

Remark 2: Given an input pmf p, the rate region defined in Theorem 1 is in general an outer bound on the optimal rate region $\mathscr{R}^*(p)$ for the computation problem of an *arbitrary* linear combination (not necessarily natural).

APPENDIX A

Next lemma is a variation of Fano's inequality for random homologous code ensembles.

Lemma 3 ([19]): Suppose that the average probability of error $\mathsf{E}_{\mathcal{C}_n}[P_e^{(n)}(\mathcal{C}_n)]$ and the entropy $H(M_i | X_i^n(M_i), \mathcal{C}_n)$ tends to zero as $n \to \infty$ for $i \in \{1, 2\}$ such that $a_i \neq 0$. Then,

$$H(W_{\mathbf{a}}^n | Y^n, \mathcal{C}_n) \leq n\epsilon_n,$$

and for every $j \in \{1, 2\}$ with $a_j \neq 0$

$$H(M_j | Y^n, M_{j^c}, \mathcal{C}_n) \leq n\epsilon_n,$$

for some $\epsilon_n \to 0$ as $n \to \infty$.

The permutation invariance of the distribution of random homologous codewords implies the following.

Lemma 4 ([19]): Let $(X,Y) \sim p(x,y)$ on $\mathbb{F}_q \times \mathcal{Y}$ and fix $\epsilon > 0$. Let $X^n(m)$ be the random codeword assigned to message $m \in \mathbb{F}_q^{nR}$ by $(n, nR; p(x), \epsilon)$ random homologous code ensemble. Further let Y^n be a random sequence distributed according to $\prod_{i=1}^n p_{Y|X}(y_i|x_i)$. Then, for every $(x,y) \in \mathbb{F}_q \times \mathcal{Y}$, and for every $i \in [1:n]$,

$$\begin{aligned}(1-\epsilon)p(x,y) &\le \mathsf{P}(X_i = x, Y_i = y | X^n \in \mathcal{T}_\epsilon^{(n)}(X)) \\ &\le (1+\epsilon)p(x,y).\end{aligned}$$

Appendix B: Proof of Lemma 2

Let $\epsilon'' > \epsilon'$. Let $j \in \{1,2\}$ be such that $a_j \neq 0$, and $j^c = \{1,2\} \setminus \{j\}$. First, by Lemma 3, we have

$$I(M_{j^c}; Y^n | \mathcal{C}_n) \ge I(M_{j^c}; W_{\mathbf{a}}^n, Y^n | \mathcal{C}_n) - n\epsilon_n.$$

Therefore, it suffices to prove that for n sufficiently large,

$$\begin{aligned}&I(M_{j^c}; W_{\mathbf{a}}^n, Y^n | \mathcal{C}_n) \\ &\qquad \ge n[\min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} - \delta_4(\epsilon'') - 2\epsilon_n].\end{aligned}$$

Similar to [20], we will show that given $W_{\mathbf{a}}^n, Y^n$ and $\mathcal{C}_n$, a relatively short list $\mathcal{L} \subseteq \mathbb{F}_q^{nR_{j^c}}$ can be constructed that contains M_{j^c} with high probability. Define a random set

$$\mathcal{L} = \{m \in \mathbb{F}_q^{nR_{j^c}} : (X_{j^c}^n(m), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y)\}.$$

It can be shown [19] that for each $m \neq M_{j^c}$,

$$\mathsf{P}(m \in \mathcal{L}, E_n = 1) \le q^{-n(I(X_{j^c}; W_{\mathbf{a}}, Y) - \delta_4(\epsilon''))},$$

where random variable E_n is as defined in (4). Since $\mathsf{P}(E_n = 1)$ tends to one as $n \to \infty$, for n sufficiently large, we have

$$\begin{aligned}\mathsf{E}(|\mathcal{L}||E_n = 1) &\le 1 + \sum_{m \neq M_{j^c}} \mathsf{P}(m \in \mathcal{L} | E_n = 1) \\ &\le 1 + q^{n(R_{j^c} - I(X_{j^c}; W_{\mathbf{a}}, Y) + \delta_4(\epsilon'') + \epsilon_n)}. \quad (9)\end{aligned}$$

Define another indicator random variable $F_n = \mathbb{1}_{\{M_{j^c} \in \mathcal{L}\}}$. Since $\epsilon'' > \epsilon'$ and $\mathsf{P}(E_n = 1)$ tends to one as $n \to \infty$, by the conditional typicality lemma in [1, p.27], $\mathsf{P}(F_n = 1)$ tends to one as $n \to \infty$. Then, for n sufficiently large, we have

$$\begin{aligned}&H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n) \\ &\qquad \le H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n, F_n = 1, E_n) + n\epsilon_n.\end{aligned}$$

We now use the fact that if $M_{j^c} \in \mathcal{L}$, then the conditional entropy cannot exceed $\log(|\mathcal{L}|)$:

$$\begin{aligned}&H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n, F_n = 1, E_n) \\ &\overset{(a)}{=} H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n, F_n = 1, E_n, \mathcal{L}, |\mathcal{L}|) \\ &\le H(M_{j^c} | F_n = 1, E_n, \mathcal{L}, |\mathcal{L}|) \\ &\overset{(b)}{\le} \sum_{l=0}^{q^{nR_{j^c}}} \mathsf{P}(|\mathcal{L}| = l | E_n = 1) \log(l) + n\epsilon_n \\ &= \mathsf{E}[\log(|\mathcal{L}|) | E_n = 1] + n\epsilon_n \\ &\overset{(c)}{\le} \log(\mathsf{E}[|\mathcal{L}||E_n = 1]) + n\epsilon_n \\ &\overset{(d)}{\le} \max\{0, n(R_{j^c} - I(X_{j^c}; W_{\mathbf{a}}, Y))\} + n\delta_4(\epsilon'') + n\epsilon_n,\end{aligned}$$

where (a) follows since the set $\mathcal{L}$ and its cardinality $|\mathcal{L}|$ are functions of $(\mathcal{C}_n, W_{\mathbf{a}}^n, Y^n)$, (b) follows for large n since $\mathsf{P}(E_n = 0)$ tends to zero as $n \to \infty$, (c) follows by Jensen's inequality, and (d) follows by (9) and the softmax interpretation of the log-sum-exp function [21, p.72]. Substituting back gives

$$\begin{aligned}&I(M_{j^c}; W_{\mathbf{a}}^n, Y^n | \mathcal{C}_n) \\ &= H(M_{j^c} | \mathcal{C}_n) - H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n) \\ &= nR_{j^c} - H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n) \\ &\ge n[R_{j^c} - \max\{0, (R_{j^c} - I(X_{j^c}; W_{\mathbf{a}}, Y))\} - \delta_4(\epsilon'') - 2\epsilon_n] \\ &= n[\min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} - \delta_4(\epsilon'') - 2\epsilon_n].\end{aligned}$$

References

[1] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[3] G. Kramer, "Topics in multi-user information theory," *Found. Trends Comm. Inf. Theory*, vol. 4, no. 4/5, pp. 265–444, 2007.

[4] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 219–221, 1979.

[5] B. Nazer and M. Gastpar, "Computation over gaussian multiple-access channels," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2007, pp. 2391–2395.

[6] S. Miyake, "Coding theorems for point-to-point communication systems using sparse matrix codes," Ph.D. dissertation, 2010.

[7] A. Padakandla and S. S. Pradhan, "Computing sum of sources over an arbitrary multiple access channel," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2013, pp. 2144–2148.

[8] T. Gariby and U. Erez, "On general lattice quantization noise," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2008, pp. 2717–2721.

[9] A. Padakandla, A. G. Sahebi, and S. S. Pradhan, "A new achievable rate region for the 3-user discrete memoryless interference channel," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2012, pp. 2256–2260.

[10] ——, "An achievable rate region for the three-user interference channel based on coset codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1250–1279, March 2016.

[11] P. Sen and Y. H. Kim, "Homologous codes for multiple access channels," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2017, pp. 874–878.

[12] ——, "Homologous codes for multiple access channels," 2018, submitted to *IEEE Trans. Inf. Theory*.

[13] A. Padakandla and S. S. Pradhan, "Achievable rate region based on coset codes for multiple access channel with states," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2013, pp. 2641–2645.

[14] S. H. Lim, C. Feng, A. Pastore, B. Nazer, and M. Gastpar, "A joint typicality approach to algebraic network information theory," *CoRR*, vol. abs/1606.09548, 2016.

[15] ——, "Towards an algebraic network information theory: Simultaneous joint typicality decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2017, pp. 1818–1822.

[16] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, 2011.

[17] N. Karamchandani, U. Niesen, and S. Diggavi, "Computation over mismatched channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 666–677, April 2013.

[18] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[19] P. Sen, S. H. Lim, and Y. H. Kim, "Optimal achievable rates for computation with random homologous codes," 2018, submitted to *IEEE Trans. Inf. Theory*. [Online]. Available: http://arxiv.org/abs/1805.03338

[20] B. Bandemer, A. E. Gamal, and Y. H. Kim, "Optimal achievable rates for interference networks with random codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6536–6549, Dec 2015.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge: Cambridge University Press, 2004.