

On the Optimal Achievable Rates for Linear Computation With Random Homologous Codes

Pinar Sen^{ID}, *Student Member, IEEE*, Sung Hoon Lim^{ID}, *Member, IEEE*, and Young-Han Kim^{ID}, *Fellow, IEEE*

Abstract—The problem of computing a linear combination of sources over a multiple access channel is studied. Inner and outer bounds on the optimal tradeoff between the communication rates are established when encoding is restricted to random ensembles of *homologous codes*, namely, structured nested coset codes from the same generator matrix and individual shaping functions, but when decoding is optimized with respect to the realization of the encoders. For the special case in which the desired linear combination is “matched” to the structure of the multiple access channel in a natural sense, these inner and outer bounds coincide. This result indicates that most, if not all, coding schemes for computation in the literature that rely on random construction of nested coset codes cannot be improved by using more powerful decoders such as the maximum likelihood decoder. The proof techniques are adapted to characterize the rate region for broadcast channels achieved by Marton’s (random) coding scheme under maximum likelihood decoding. By generalizing some of the techniques, a single-letter outer bound for the capacity region of the computation problem is presented and compared with the inner bound achieved by homologous codes.

Index Terms—Multiple access channel, broadcast channel, structured coding, linear code, maximum likelihood decoding, simultaneous decoding, Marton’s coding.

I. INTRODUCTION

CONSIDER a multiple access channel (MAC) with two senders and one receiver, in which the receiver wishes to reliably estimate a linear function of transmitted codewords from the senders (see Figure 1). One trivial approach to this *computation* problem involves two steps: first recover the individual codewords and then compute the function from the recovered codewords. When the problem is isolated to the first communication step of this plug-in approach, using the conventional random independently and identically distributed (i.i.d.) code ensembles achieves the optimal rates of communicating independent codewords [1], [2]. For the problem as

a whole, however, the use of random i.i.d. code ensembles is strictly suboptimal even for a trivial MAC. The earlier studies on computation concentrated on recovering/utilizing a linear combination of *messages* at the receiver. As shown by Körner and Marton [3] for the problem of encoding a modulo-two sum of distributed dependent binary sequences, using the *same* random ensemble of linear codes at multiple encoders can achieve strictly better rates than using independently generated ensembles of codes. Building on this observation, Nazer and Gastpar [4] developed a channel coding scheme that uses the same random ensemble of linear codes at multiple encoders and showed that this *structured* coding scheme outperforms conventional random coding schemes for computing a linear combination of transmitted messages over a linear MAC. This influential work led to the development of the *compute–forward* strategy for relay networks [5]–[7]. Over the past decade, the compute–forward strategy based on lattice codes for Gaussian channels and its extensions have shown to provide higher achievable rates for several communication problems over relay networks [5]–[11].

More recently, *nested coset codes* [12], [13] were proposed as more flexible alternatives for achieving the desired linear structure at multiple encoders. In particular, Padakandla and Pradhan [13] developed a fascinating coding scheme for the computation problem of transmitted messages over an *arbitrary* MAC. In this coding scheme, a coset code with a rate higher than the target (message) rate is first generated randomly. Next, in the *shaping* step, a codeword of a desired property (such as type or joint type) is selected from a subset of codewords (a coset of a subcode). Although reminiscent of the multicoding scheme of Gelfand and Pinsker [14] for channels with state, and Marton’s coding scheme [15] for broadcast channels, this construction is more fundamental in some sense, since the scheme is directly applicable even for classical point-to-point communication channels. A similar shaping technique was also developed for lattice codes in [16]. For multiple encoders, the desired common structure is obtained by using coset codes with the same generator matrix. Recent efforts exploited the benefit of such constructions for a broader class of channel models, such as interference channels [17], [18], multiple access channels [19], [20], and multiple access channels with state [21].

In those earlier studies, *message* computation at the receiver is closely related (if not equivalent) to *codeword* computation due to the underlying linearity shared among users’ codebooks. Motivated by the physical meaning of compute–forward and

Manuscript received October 29, 2018; revised November 27, 2019; accepted May 19, 2020. Date of publication July 20, 2020; date of current version September 22, 2020. This work was supported in part by the Electronics and Telecommunications Research Institute through the Korean Ministry of Science, ICT, and Future Planning under Grant 17ZF1100; and in part by the National Research Foundation (NRF) of South Korea Funded by the Ministry of Education, Science and Technology under Grant NRF-2017R1C1B1004192. This article was presented in part at the 2018 IEEE International Symposium of Information Theory. (*Corresponding author: Pinar Sen.*)

Pinar Sen and Young-Han Kim are with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: psen@ucsd.edu; yhk@ucsd.edu).

Sung Hoon Lim is with the School of Software, Hallym University, Chuncheon 24252, South Korea (e-mail: shlim@hallym.ac.kr).

Communicated by S. Watanabe, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2020.3010253

0018-9448 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

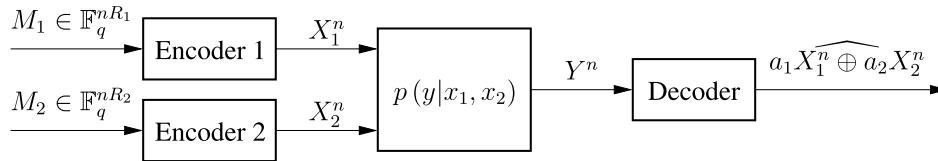


Fig. 1. Linear computation over two-sender multiple access channel.

interference alignment, where a linear combination of codewords is to be utilized at the receiver to cancel out the interfering codewords, Lim, Feng, Pastore, Nazer, and Gastpar [22], [23] tackled codeword computation and generalized the nested coset codes constructed with the same generator matrix to asymmetric rate pairs. We referred to this generalized version, together with the shaping step, as *homologous codes* [19], [20], [24]. This terminology is motivated from its biological definition, i.e., the structures modified from the same ancestry (underlying linear code) to adapt to different purposes (desired shape). Lim *et al.* [22], [23] further analyzed *simultaneous decoding* of random ensembles of homologous codes and showed that it can achieve rates higher than existing approaches to computation problems. For instance, when adapted to the Gaussian MAC, the resulting achievable rates improve upon those of lattice codes [7]. With mathematical rate expressions in single-letter mutual information terms that work for a general memoryless channel and result in better performances than those of lattice codes for Gaussian channels, homologous codes have a potential to bringing a deeper understanding of the fundamental limits of the computation problem.

Several open questions remain, however. What is the optimal tradeoff between achievable rates for reliable computation? Which scheme achieves this computation capacity region? The answers require joint optimization of encoder and decoder designs, which seems to be intractable as in many other network information theory problems.

In this paper, we instead concentrate on the performance of the optimal maximum likelihood decoder when the encoder is restricted to a given random ensemble of homologous codes. We characterize the optimal rate region when the desired linear combination and the channel structure are “matched” (see Definition 1 in Section III), which is the case in which the benefit of computation can be realized to the fullest extent as indicated by [25]. This result, *inter alia*, implies that the suboptimal joint typicality decoding rule proposed in [22], [23] achieves this optimal rate region. Thus, the performance of random ensembles of homologous codes cannot be improved by the maximum likelihood decoder.

The main contribution lies in the outer bound on the optimal rate region (Theorem 3), which characterizes the necessary condition that a rate pair must satisfy if the average probability of decoding error vanishes asymptotically. The proof of this bound relies on two key observations. First, the distribution of a given random ensemble of homologous codes converges asymptotically to the product of the desired input distribution. Second, given the channel output, a relatively short list of messages can be constructed that includes the actually

transmitted message with high probability. The second observation, which is adapted from the analysis in [26] for the optimal rate region of interference networks with random i.i.d. code ensembles, seems to be a recurring path to establishing the optimal performance of random code ensembles.

As hinted earlier, the construction of random ensemble of homologous codes has many similarities to Marton’s coding scheme [15], one of the fundamental coding schemes in network information theory. As a result, adapting the proof techniques that we developed for homologous codes, we can establish an outer bound on the optimal rate region for broadcast channels with Marton’s coding scheme (Proposition 2). The resulting outer bound coincides with the inner bound that is achieved by *simultaneous nonunique decoding*, thus characterizing the optimal rate region of a two-receiver general broadcast channel achieved by a given random code ensemble.

The rest of the paper is organized as follows. Section II formally defines the computation problem. Section III presents the main result of the paper—the optimal rate region achievable by a random ensemble of homologous codes. The inner and the outer bounds on this region are presented in Sections IV and V, respectively. Section VI establishes the optimal rate region for a broadcast channel achievable by Marton’s coding scheme. Section VII concludes the paper with some discussion on the *capacity region* of computation problem by providing a single-letter outer bound for a given (fixed) computation code and comparing it to the inner bound that is achieved by homologous codes.

We adapt the notation in [27], [28]. The set of integers $\{1, 2, \dots, n\}$ is denoted by $[n]$. For a length- n sequence (row vector) $x^n = (x_1, x_2, \dots, x_n) \in \mathcal{X}^n$, we define its type as $\pi(x|x^n) = |\{i: x_i = x\}|/n$ for $x \in \mathcal{X}$. Upper case letters X, Y, \dots denote random variables. For $\epsilon \in (0, 1)$, we define the ϵ -typical set of n -sequences (or the typical set in short) as $\mathcal{T}_\epsilon^{(n)}(X) = \{x^n: |p(x) - \pi(x|x^n)| \leq \epsilon p(x), x \in \mathcal{X}\}$. The indicator function $\mathbb{1}_S: \mathcal{X} \rightarrow \{0, 1\}$ for $S \subseteq \mathcal{X}$ is defined as $\mathbb{1}_S(x) = 1$ if $x \in S$ and 0 otherwise. A length- n row vector of all zeros is denoted by $\mathbf{0}_n$, where the subscript is omitted when it is clear from the context. We denote by \mathbb{F}_q a finite field of size q , \mathbb{F}_q^* is the set of nonzero elements in \mathbb{F}_q , and \mathbb{F}_q^d is the d -dimensional vector space over \mathbb{F}_q . The limit of a collection of sets $\{\mathcal{A}(\epsilon)\}$ indexed by $\epsilon > 0$ is defined as

$$\lim_{\epsilon \rightarrow 0} \mathcal{A}(\epsilon) := \bigcup_{\epsilon > 0} \bigcap_{0 < \gamma < \epsilon} \mathcal{A}(\gamma) \stackrel{(a)}{=} \bigcap_{\epsilon > 0} \bigcup_{0 < \gamma < \epsilon} \mathcal{A}(\gamma), \quad (1)$$

which exists if (a) holds. The closure $\text{cl}(\mathcal{A})$ of a set $\mathcal{A} \subseteq \mathbb{R}^d$ denotes the smallest closed superset of \mathcal{A} . The convex hull $\text{conv}(\mathcal{A})$ of a set \mathcal{A} denotes the smallest convex and closed

superset of \mathcal{A} . We use $\epsilon_n \geq 0$ to denote a generic sequence of n that tends to zero as $n \rightarrow \infty$, and use $\delta_i(\epsilon) \geq 0$, $i \in \mathbb{Z}^+$, to denote a continuous function of ϵ that tends to zero as $\epsilon \rightarrow 0$. For the q -ary computation problem studied in Sections II through V, information measures are in logarithm base q .

II. FORMAL STATEMENT OF THE PROBLEM

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

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

in Figure 1, which consists of two sender alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{F}_q$, 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. Here and henceforth, we assume without loss of generality that nR_1 and nR_2 are integers. The goal of communication is to convey a linear combination of the codewords. Hence, upon receiving the sequence Y^n , the decoder finds an estimate $\hat{W}_a^n = \hat{w}_a^n(Y^n) \in \mathbb{F}_q^n$ of

$$W_a^n := a_1 X_1^n \oplus a_2 X_2^n$$

for a desired (nonzero) vector $\mathbf{a} = [a_1 \ a_2]$ over \mathbb{F}_q , where the operator \oplus denotes the q -ary addition. Formally, an (n, nR_1, nR_2) computation code for the multiple access channel consists of two encoders that map $x_j^n(m_j)$, $j = 1, 2$, and a decoder that maps $\hat{w}_a^n(y^n)$. The collection of codewords $\mathcal{C}_n := \{(x_1^n(m_1), x_2^n(m_2)) : (m_1, m_2) \in \mathbb{F}_q^{(nR_1) \times (nR_2)}\}$ is referred to as the *codebook* associated with the (n, nR_1, nR_2) code.

Remark 1: For 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 [29, Sec. 6.2]. More specifically, given a pair of symbol-by-symbol mappings $\varphi_j : \mathbb{F}_q \rightarrow \mathcal{X}_j$, $j = 1, 2$, 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 $W_a := a_1 V_1^n \oplus a_2 V_2^n$, 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 $j = 1, 2$. Our results can be readily applied to this computation problem defined on the virtual channel.

The performance of a given computation code with a fixed desired vector \mathbf{a} and with codebook \mathcal{C}_n is measured by the average probability of error

$$P_e^{(n)}(\mathcal{C}_n) = \mathbf{P}(\hat{W}_a^n \neq W_a^n | \mathcal{C}_n),$$

when M_1 and M_2 are independent and uniformly distributed. Message M_j , $j = 1, 2$, is said to be *confusable* if $x_j^n(M_j) = x_j^n(m_j)$ for some $m_j \neq M_j \in \mathbb{F}_q^{nR_j}$. A rate pair (R_1, R_2) is said to be *achievable for \mathbf{a} -computation* if there exists a sequence of (n, nR_1, nR_2) computation codes such that

$$\lim_{n \rightarrow \infty} P_e^{(n)}(\mathcal{C}_n) = 0$$

and

$$\lim_{n \rightarrow \infty} \mathbf{P}(M_j \text{ is confusable} | \mathcal{C}_n) = 0, \quad (2)$$

for every $j \in \{1, 2\}$ with $a_j \neq 0$. Note that without the condition in (2), the problem is trivial and an arbitrarily large rate pair is achievable.

We now define the random ensemble of computation codes referred to as homologous codes. 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 according to the following steps:

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

$$u_j^n(m_j, l_j) = [m_j \ l_j \ \mathbf{0}_{\kappa - n(R_j + \hat{R}_j)}] G + d_j^n,$$

for every $m_j \in \mathbb{F}_q^{nR_j}$ and every $l_j \in \mathbb{F}_q^{n\hat{R}_j}$, $j = 1, 2$. At sender $j = 1, 2$, assign a 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 a random variable that is drawn uniformly at random among all l_j vectors 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.

The intuition behind this construction can be identified as follows. In step 3), sender $j = 1, 2$ constructs a coset code of rate $R_j + \hat{R}_j$, which is larger than its target rate R_j . The redundancy, the amount of which is determined by \hat{R}_j , provides the existence of a codeword within the typical set $\mathcal{T}_\epsilon^{(n)}(X_j)$ with high probability. The fact that the codewords of different senders are built from the same underlying linear code benefits computation in the sense that a linear combination of codewords is a codeword from a coset of the same underlying linear code.

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*¹. 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*, where p is the given input pmf and $\epsilon > 0$ is the parameter used in steps 1 and 3 in codebook generation. A rate pair (R_1, R_2) is said to be *achievable for \mathbf{a} -computation by the (p, ϵ) -distributed random homologous code ensemble* if there exists a sequence of $(n, nR_1, nR_2; p, \epsilon)$ random homologous code ensembles such that

$$\lim_{n \rightarrow \infty} \mathbf{E}_{\mathcal{C}_n} [P_e^{(n)}(\mathcal{C}_n)] = 0 \quad (3)$$

¹The codebook that each sender is equipped with through steps 1)-3) is referred to as the nested coset codebook in the literature [13], [22], [23]. In that sense, a homologous codebook can be seen as a family of nested coset codebooks constructed with the same generator matrix G with individual dithers D_j^n and codebook distributions $p(x_j)$, $j = 1, 2$.

and

$$\lim_{n \rightarrow \infty} \mathbf{E}_{\mathcal{C}_n} [\mathbf{P}(M_j \text{ is confusable} | \mathcal{C}_n)] = 0, \quad (4)$$

for every $j \in \{1, 2\}$ with $a_j \neq 0$. Here the expectations are with respect to the random homologous codebook \mathcal{C}_n , i.e., $(G, D_1^n, D_2^n, (L_1(m_1) : m_1 \in \mathbb{F}_q^{nR_1}), (L_2(m_2) : m_2 \in \mathbb{F}_q^{nR_2}))$. Given $(p, \epsilon, \mathbf{a})$, let $\mathcal{R}^*(p, \epsilon, \mathbf{a})$ be the set of all rate pairs achievable for \mathbf{a} -computation by the (p, ϵ) -distributed random homologous code ensemble. Given the input pmf p and the desired vector $\mathbf{a} \neq \mathbf{0} \in \mathbb{F}_q^2$, the optimal rate region $\mathcal{R}^*(p, \mathbf{a})$, when it exists, is defined as

$$\mathcal{R}^*(p, \mathbf{a}) := \text{cl} \left[\lim_{\epsilon \rightarrow 0} \mathcal{R}^*(p, \epsilon, \mathbf{a}) \right].$$

Remark 2: Instead of (4), one may consider alternative notions for the *confusability* of the transmitted message, such as

$$\lim_{n \rightarrow \infty} \frac{H(M_j | X_j^n(M_j), \mathcal{C}_n)}{n} = 0, \quad (5)$$

or

$$\lim_{n \rightarrow \infty} \mathbf{E}_{\mathcal{C}_n} [\mathbf{P}(G \text{ is rank deficient} | \mathcal{C}_n)] = 0. \quad (6)$$

It is easy to show that our results for the optimal rate region $\mathcal{R}^*(p, \mathbf{a})$ under (4) still apply if we change the confusability notion with (5) or (6).

III. THE MAIN RESULT

In this section, we present a single-letter characterization of the optimal rate region when the target linear combination is in the following class.

Definition 1: A linear combination $W_{\mathbf{a}} = a_1 X_1 \oplus a_2 X_2$ for some $\mathbf{a} = [a_1 \ a_2] \in \mathbb{F}_q^2 \setminus \{\mathbf{0}\}$ is said to be *natural* if

$$H(W_{\mathbf{a}} | Y) = \min_{\mathbf{b} \neq \mathbf{0}} H(W_{\mathbf{b}} | Y), \quad (7)$$

where $\mathbf{b} = [b_1 \ b_2]$ and $W_{\mathbf{b}} = b_1 X_1 \oplus b_2 X_2$ are over \mathbb{F}_q .

Intuitively, a natural combination $W_{\mathbf{a}}$ is the easiest to recover at the receiver and thus, in some sense, is the best linear combination that is matched to the channel structure.

We are now ready to present the optimal rate region for computing natural linear combinations.

Theorem 1: Given an input pmf $p = p(x_1)p(x_2)$ and a vector $\mathbf{a} \neq \mathbf{0} \in \mathbb{F}_q^2$ such that $W_{\mathbf{a}}$ is a natural combination, the optimal rate region $\mathcal{R}^*(p, \mathbf{a})$ is the set of rate pairs (R_1, R_2) such that

$$R_j \leq I(X_j; Y | X_{j^c}), \quad (8a)$$

$$R_j \leq I(X_1, X_2; Y) - \min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} \quad (8b)$$

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

The rate region in (8) in Theorem 1, which we will denote as $\mathcal{R}^{**}(p, \mathbf{a})$, can be equivalently characterized in terms of well-known rate regions for compute-forward and message communication. Let $\mathcal{R}_{\text{CF}}(p, \mathbf{a})$ be the set of rate pairs (R_1, R_2) such that

$$R_j \leq H(X_j) - H(W_{\mathbf{a}} | Y), \quad \forall j \in \{1, 2\} \text{ with } a_j \neq 0. \quad (9)$$

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

$$R_1 \leq I(X_1; Y | X_2),$$

$$R_2 \leq I(X_2; Y | X_1),$$

$$R_1 + R_2 \leq I(X_1, X_2; Y).$$

Proposition 1: For any input pmf $p = p(x_1)p(x_2)$ and any linear combination $W_{\mathbf{a}}$,

$$\mathcal{R}^{**}(p, \mathbf{a}) = \mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_{\text{MAC}}(p).$$

The proof of Proposition 1 is relegated to Appendix A.

We prove Theorem 1 in three steps: 1) we first present a general (not necessarily for natural combinations) inner bound on the optimal rate region in Section IV, where we follow the results in [22], [23] that studied the rate region achievable by random homologous code ensembles using a suboptimal joint typicality decoding rule, 2) we then show by Lemma 1 in Section IV that this inner bound is equivalent to $\mathcal{R}^{**}(p, \mathbf{a})$ in Proposition 1 if $W_{\mathbf{a}}$ is a natural combination, and 3) we present a general (not necessarily for natural combinations) outer bound on the optimal rate region in Section V by showing that if a rate pair (R_1, R_2) is achievable for \mathbf{a} -computation by the (p, ϵ) -distributed random homologous code ensemble for arbitrarily small ϵ , then (R_1, R_2) must lie in $\mathcal{R}^{**}(p, \mathbf{a})$ in Theorem 1.

Remark 3: Due to the underlying linearity shared between different users' codebooks, the computation problem defined in Section II is closely related to the *message computation*. Indeed, one may redefine the computation problem over messages where the goal of transmission is to convey a linear combination $a_1 M_1 \oplus a_2 M_2$ of messages for $R_1 = R_2$ and redefine the achievability for \mathbf{a} -computation by the (p, ϵ) -distributed random homologous code ensemble and optimal symmetric rate $R^*(p, \mathbf{a})$ in a similar manner but based on condition (3) only, then $R^*(p, \mathbf{a})$ is equal to the largest symmetric rate satisfying (8) in Theorem 1. The achievability simply follows from the inner bound in Section IV. To see this, note that a linear combination of codewords is of the form

$$\begin{aligned} & a_1 X_1^n(M_1) \oplus a_2 X_2^n(M_2) \\ &= (a_1 [M_1 \ L_1 \ \mathbf{0}_{\kappa - n(R_1 + \hat{R}_1)}] \oplus a_2 [M_2 \ L_1 \ \mathbf{0}_{\kappa - n(R_2 + \hat{R}_2)}]) G \\ & \quad \oplus a_1 D_1^n \oplus a_2 D_2^n. \end{aligned}$$

Since the generator matrix G is full rank almost surely as $n \rightarrow \infty$ by Lemma 6 under the rate constraints in Theorem 1, $(a_1 [M_1 \ L_1 \ \mathbf{0}] \oplus a_2 [M_2 \ L_1 \ \mathbf{0}])$ can be recovered from $a_1 X_1^n(M_1) \oplus a_2 X_2^n(M_2)$ almost surely. When $R_1 = R_2 = R$, the first nR bits of $(a_1 [M_1 \ L_1 \ \mathbf{0}] \oplus a_2 [M_2 \ L_1 \ \mathbf{0}])$ would give $a_1 M_1 \oplus a_2 M_2$ as desired. To prove the optimality, an outer bound can be obtained by following similar steps with Section V.

IV. AN INNER BOUND

The linear computation performance of random homologous code ensembles was studied using a suboptimal *joint typicality* decoder in [22], [23]. For completeness, we first describe the joint typicality decoding rule and then characterize the rate

region achievable for \mathbf{a} -computation by the (p, ϵ) -distributed random homologous code ensemble *under this joint typicality decoding rule*. We then concentrate on an arbitrarily small ϵ to provide an inner bound on the optimal rate region $\mathcal{R}^*(p, \mathbf{a})$. We will omit the steps that were already established in [22], [23] and instead provide detailed references.

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

$$s = a_1[m_1 l_1 \mathbf{0}_{\kappa-n(R_1+\hat{R}_1)}] \oplus a_2[m_2 l_2 \mathbf{0}_{\kappa-n(R_2+\hat{R}_2)}],$$

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

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

where $u_i^n(m_i, l_i) = [m_i l_i \mathbf{0}_{\kappa-n(R_i+\hat{R}_i)}]G \oplus d_i^n$ is the auxiliary codeword defined in step 3) of the code construction in Section II. If the decoder finds such s , then it declares $\hat{w}_{\mathbf{a}}^n = sG \oplus a_1 d_1^n \oplus a_2 d_2^n$ as an estimate; otherwise, it declares an error.

To describe the performance of the joint typicality decoder, we define $\mathcal{R}_{CF}(p, \delta, \mathbf{a})$ for a given input pmf p , $\delta \geq 0$, and nonzero vector $\mathbf{a} \in \mathbb{F}_q^2$ as the set of rate pairs (R_1, R_2) such that

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

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

$$R_1 \leq I(X_1; Y|X_2) - \delta, \quad (10a)$$

$$R_2 \leq I(X_2; Y|X_1) - \delta, \quad (10b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y) - \delta, \quad (10c)$$

$$R_1 \leq I(X_1, X_2; Y) - H(X_2) + \min_{b_1, b_2 \in \mathbb{F}_q^*} H(W_{\mathbf{b}}|Y) - \delta, \quad (10d)$$

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

$$R_1 \leq I(X_1; Y|X_2) - \delta, \quad (11a)$$

$$R_2 \leq I(X_2; Y|X_1) - \delta, \quad (11b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y) - \delta, \quad (11c)$$

$$R_2 \leq I(X_1, X_2; Y) - H(X_1) + \min_{b_1, b_2 \in \mathbb{F}_q^*} H(W_{\mathbf{b}}|Y) - \delta, \quad (11d)$$

where $\mathbf{b} = [b_1 b_2]$ and $W_{\mathbf{b}} = b_1 X_1 \oplus b_2 X_2$ are over \mathbb{F}_q . Note that the region $\mathcal{R}_{CF}(p, \mathbf{a}) = \mathcal{R}_{CF}(p, \delta = 0, \mathbf{a})$, as defined in (9) in Section III. Similarly, let $\mathcal{R}_j(p)$ denote the region $\mathcal{R}_j(p, \delta = 0)$ for $j = 1, 2$ in (10) and (11).

We are now ready to state the rate region achievable by the random homologous code ensembles that combines the inner bounds in [22, Theorem 1] and [23, Corollary 1].

Theorem 2: Let $p = p(x_1)p(x_2)$ be an input pmf, $\delta > 0$, and $\mathbf{a} \in \mathbb{F}_q^2$ be a nonzero vector. Then, there exists $\epsilon' < \delta$ such that for every $\epsilon < \epsilon'$ sufficiently small, a rate pair

$$(R_1, R_2) \in \mathcal{R}_{CF}(p, \delta, \mathbf{a}) \cup \mathcal{R}_1(p, \delta) \cup \mathcal{R}_2(p, \delta) \quad (12)$$

is achievable for \mathbf{a} -computation by the (p, ϵ) -distributed random homologous code ensemble along with the ϵ' -joint typicality decoder. In particular,

$$[\mathcal{R}_{CF}(p, \mathbf{a}) \cup \mathcal{R}_1(p) \cup \mathcal{R}_2(p)] \subseteq \mathcal{R}^*(p, \mathbf{a}). \quad (13)$$

Proof: The proof of [22, Theorem 1] analyzes the average probability of error for \mathbf{a} -computation by the (p, ϵ) -distributed random homologous code ensemble paired with the ϵ' -joint typicality decoder for $\epsilon' > \epsilon > 0$. Two upper bounds on the average probability of error were given. The first one, direct decoding bound, captures the error event that incorrect linear combinations are confused with the correct one and shows that for sufficiently small $\epsilon < \epsilon' < \delta$, the average probability of error tends to zero as $n \rightarrow \infty$ if

$$(R_1, R_2) \in \mathcal{R}_{CF}(p, \delta, \mathbf{a}). \quad (14)$$

The second one, multiple access bound, captures the error event that incorrect message pairs (codeword pairs) are confused with the correct one. This bound was later improved in the proof of [23, Corollary 1]. The improved version shows that for every $\mathbf{a} \in \mathbb{F}_q^2$, the average probability of error for \mathbf{a} -computation tends to zero as $n \rightarrow \infty$ if

$$(R_1, R_2) \in \mathcal{R}_1(p, \delta) \cup \mathcal{R}_2(p, \delta). \quad (15)$$

Combining (14) and (15) establishes (12).

We still need to show that the condition in (4) holds. Suppose that $a_j \neq 0$. For a given codebook \mathcal{C}_n , let G_j denote the submatrix that consists of the first $(nR_j + n\hat{R}_j)$ rows of G within \mathcal{C}_n and $s_j(G)$ be the indicator variable such that $s_j = 1$ if G_j is full rank. Then,

$$\begin{aligned} & \mathbb{E}_{\mathcal{C}_n}[\mathbb{P}(M_j \text{ is confusable} | \mathcal{C}_n)] \\ &= \sum_{\mathcal{C}_n} \mathbb{P}(\mathcal{C}_n = \mathcal{C}_n) \mathbb{P}(M_j \text{ is confusable} | \mathcal{C}_n = \mathcal{C}_n) \\ &= \sum_{\substack{\mathcal{C}_n: \\ s_j(G)=0}} \mathbb{P}(\mathcal{C}_n = \mathcal{C}_n) \mathbb{P}(M_j \text{ is confusable} | \mathcal{C}_n = \mathcal{C}_n) \\ &\leq \sum_{\substack{\mathcal{C}_n: \\ s_j(G)=0}} \mathbb{P}(\mathcal{C}_n = \mathcal{C}_n) \\ &= \mathbb{P}(S_j(G) = 0). \end{aligned}$$

Now, by Lemma 6 in Appendix B (with $R \leftarrow R_j + \hat{R}_j$), the term $\mathbb{P}(S_j(G) = 0)$ tends to zero as $n \rightarrow \infty$ if $R_j + \hat{R}_j < 1$. By definition, $\hat{R}_j = D(p_{X_j} \| \text{Unif}(\mathbb{F}_q)) + \epsilon$, which reduces the constraint to the form of $R_j < H(X_j) - \epsilon$. Since this condition is satisfied if (12) holds, the proof of (12) follows.

The proof of (13) follows by taking the closure of the union of (12) over all $\delta > 0$, which completes the proof of Theorem 1. ■

The inner bound (13) in Theorem 1 is valid for computing an arbitrary linear combination, which may not be equal to the rate region $\mathcal{R}^{**}(p, \mathbf{a})$ in Theorem 1 for every $\mathbf{a} \in \mathbb{F}_q^2$, in general. For computing a *natural* linear combination, however, the following lemma shows that the equivalent rate region in Proposition 1 is achievable.

Lemma 1: 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

$$[\mathcal{R}_{CF}(p, \mathbf{a}) \cup \mathcal{R}_1(p) \cup \mathcal{R}_2(p)] = [\mathcal{R}_{CF}(p, \mathbf{a}) \cup \mathcal{R}_{MAC}(p)].$$

The proof of Lemma 1 is relegated to Appendix C.

V. AN OUTER BOUND

We start with presenting an outer bound on the rate region $\mathcal{R}^*(p, \epsilon, \mathbf{a})$ for a fixed input pmf p , $\epsilon > 0$, and nonzero vector $\mathbf{a} \in \mathbb{F}_q^2$. We then discuss the limit of this outer bound as $\epsilon \rightarrow 0$ to establish an outer bound on the rate region $\mathcal{R}^*(p, \mathbf{a})$. Given an input pmf p , $\delta > 0$, and nonzero vector $\mathbf{a} \in \mathbb{F}_q^2$, we define the rate region $\mathcal{R}^{**}(p, \delta, \mathbf{a})$ as the set of rate pairs (R_1, R_2) such that

$$R_j \leq I(X_j; Y | X_{j^c}) + \delta, \quad (16a)$$

$$R_j \leq I(X_1, X_2; Y) - \min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} + \delta, \quad (16b)$$

for every $j \in \{1, 2\}$ with $a_j \neq 0$, where $j^c = \{1, 2\} \setminus \{j\}$. Note that $\mathcal{R}^{**}(p, \delta = 0, \mathbf{a})$ is equal to $\mathcal{R}^{**}(p, \mathbf{a})$ as defined in (8).

We are now ready to state an outer bound on the optimal rate region for computing an *arbitrary* linear combination, which is also an outer bound on $\mathcal{R}^*(p, \mathbf{a})$ in Theorem 1 for computing a *natural* combination.

Theorem 3: Let $p = p(x_1)p(x_2)$ be an input pmf, $\epsilon > 0$, and $\mathbf{a} \in \mathbb{F}_q^2$ be a nonzero vector. If a rate pair (R_1, R_2) is achievable for \mathbf{a} -computation by the (p, ϵ) -distributed random homologous code ensemble, then there exists a continuous $\delta'(\epsilon)$ that tends to zero monotonically as $\epsilon \rightarrow 0$ such that

$$(R_1, R_2) \in \mathcal{R}^{**}(p, \delta'(\epsilon), \mathbf{a}). \quad (17)$$

In particular,

$$\mathcal{R}^*(p, \mathbf{a}) \subseteq \mathcal{R}^{**}(p, \mathbf{a}). \quad (18)$$

Proof: We first start with an averaged version of Fano's inequality for a random homologous code ensemble \mathcal{C}_n (recall the notation in Section II).

Lemma 2: If

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

and

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} [\mathbf{P}(M_j \text{ is confusable} | \mathcal{C}_n)] = 0 \quad (19)$$

for every $j \in \{1, 2\}$ with $a_j \neq 0$, then 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 $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

The proof of Lemma 2 is relegated to Appendix D.

We next define the indicator random variable

$$E_n = \mathbb{1}_{\{(X_1^n(M_1), X_2^n(M_2)) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2)\}} \quad (20)$$

for $\epsilon' > 0$. Since $\hat{R}_i = D(p_{X_i} \| \text{Unif}(\mathbb{F}_q)) + \epsilon$, $i = 1, 2$, by the Markov lemma [22, Lemma 12] for homologous codes, $\mathbf{P}(E_n = 0)$ tends to zero as $n \rightarrow \infty$ if ϵ' is sufficiently large compared to ϵ . Let $\epsilon' = \delta_1(\epsilon)$, which still tends to zero as

$\epsilon \rightarrow 0$. Suppose that $a_j \neq 0$. Then, for n sufficiently large,

$$\begin{aligned} nR_j &= H(M_j | M_{j^c}, \mathcal{C}_n) \\ &\stackrel{(a)}{\leq} I(M_j; Y^n | M_{j^c}, \mathcal{C}_n) + n\epsilon_n \\ &\leq I(M_j, E_n; Y^n | M_{j^c}, \mathcal{C}_n) + n\epsilon_n \\ &\stackrel{(b)}{\leq} \log_q 2 + I(M_j; Y^n | M_{j^c}, \mathcal{C}_n, E_n) + n\epsilon_n \\ &\leq \log_q 2 + I(M_j; Y^n | M_{j^c}, \mathcal{C}_n, E_n = 0) \mathbf{P}(E_n = 0) \\ &\quad + I(M_j; Y^n | M_{j^c}, \mathcal{C}_n, E_n = 1) \mathbf{P}(E_n = 1) + n\epsilon_n \\ &\leq \log_q 2 + nR_j \mathbf{P}(E_n = 0) \\ &\quad + I(M_j; Y^n | M_{j^c}, \mathcal{C}_n, E_n = 1) + n\epsilon_n \\ &= \log_q 2 + nR_j \mathbf{P}(E_n = 0) + n\epsilon_n \\ &\quad + \sum_{i=1}^n I(M_j; Y_i | Y^{i-1}, M_{j^c}, \mathcal{C}_n, X_{j^c i}, E_n = 1) \\ &\leq \log_q 2 + nR_j \mathbf{P}(E_n = 0) + n\epsilon_n \\ &\quad + \sum_{i=1}^n I(M_j, X_{ji}, Y^{i-1}, M_{j^c}, \mathcal{C}_n; Y_i | X_{j^c i}, E_n = 1) \\ &\stackrel{(c)}{=} \log_q 2 + nR_j \mathbf{P}(E_n = 0) + n\epsilon_n \\ &\quad + \sum_{i=1}^n I(X_{ji}; Y_i | X_{j^c i}, E_n = 1), \end{aligned} \quad (21)$$

where (a) follows by Lemma 2, (b) follows since E_n is a binary random variable, and (c) follows since

$$(M_1, M_2, Y^{i-1}, \mathcal{C}_n, E_n) \rightarrow (X_{1i}, X_{2i}) \rightarrow Y_i$$

form a Markov chain for every $i \in [n]$. To further upper bound (21), we make a connection between the distribution of the random homologous codebook and the input pmf p as follows.

Lemma 3: Let $(X_1, X_2, Y) \sim p(x_1)p(x_2)p(y|x_1, x_2)$ on $\mathbb{F}_q \times \mathbb{F}_q \times \mathcal{Y}$ and $\epsilon, \epsilon' > 0$. Let $(X_1^n(m_1), X_2^n(m_2))$ be the random codeword pair assigned to message pair $(m_1, m_2) \in \mathbb{F}_q^{nR_1} \times \mathbb{F}_q^{nR_2}$ by an $(n, nR_1, nR_2; p, \epsilon)$ random homologous code ensemble, where $p = p(x_1)p(x_2)$ is the input pmf. Further let Y^n be a random sequence distributed according to $\prod_{i=1}^n p_{Y|X_1, X_2}(y_i | x_{1i}, x_{2i})$. Then, for every $(x_1, x_2, y) \in \mathbb{F}_q \times \mathbb{F}_q \times \mathcal{Y}$ and for every $i = 1, 2, \dots, n$,

$$\begin{aligned} &(1 - \epsilon')p(x_1, x_2, y) \\ &\leq \mathbf{P}(X_{1i} = x_1, X_{2i} = x_2, Y_i = y | (X_1^n, X_2^n) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2)) \\ &\leq (1 + \epsilon')p(x_1, x_2, y). \end{aligned}$$

The proof of Lemma 3 is relegated to Appendix E.

Back to the proof of Theorem 3, we are now ready to establish (16a). By Lemma 3, each term $I(X_{ji}; Y_i | X_{j^c i}, E_n = 1)$ is close to $I(X_j; Y | X_{j^c})$ upto a function of ϵ' that vanishes as $\epsilon' \rightarrow 0$. Therefore, combining (21) with Lemma 3, we have

$$\begin{aligned} nR_j &\leq \log_q 2 + nR_j \mathbf{P}(E_n = 0) + n\epsilon_n \\ &\quad + n(I(X_j; Y | X_{j^c}) + \delta_2(\epsilon')) \\ &\stackrel{(d)}{\leq} n(I(X_j; Y | X_{j^c}) + \delta_2(\epsilon')) + 2n\epsilon_n \\ &\stackrel{(e)}{\leq} n(I(X_j; Y | X_{j^c}) + \delta_3(\epsilon)) + 2n\epsilon_n, \end{aligned} \quad (22)$$

where (d) follows since $\mathbf{P}(E_n = 0)$ tends to zero as $n \rightarrow \infty$ and (e) follows since $\epsilon' = \delta_1(\epsilon)$.

For the proof of (16b), we start with

$$\begin{aligned} nR_j &= H(M_j | M_{j^c}, \mathcal{C}_n) \\ &\stackrel{(a)}{\leq} I(M_j; Y^n | M_{j^c}, \mathcal{C}_n) + n\epsilon_n \\ &= I(M_1, M_2; Y^n | \mathcal{C}_n) - I(M_{j^c}; Y^n | \mathcal{C}_n) + n\epsilon_n, \end{aligned} \quad (23)$$

where (a) follows by Lemma 2. Following arguments similar to (22), the first term in (23) can be bounded as

$$\begin{aligned} &I(M_1, M_2; Y^n | \mathcal{C}_n) \\ &\leq \log_q 2 + n(R_1 + R_2) \mathbf{P}(E_n = 0) \\ &\quad + \sum_{i=1}^n I(M_1, M_2; Y_i | \mathcal{C}_n, Y^{i-1}, E_n = 1) \\ &\leq n\epsilon_n + \sum_{i=1}^n I(M_1, M_2, \mathcal{C}_n, Y^{i-1}; Y_i | E_n = 1) \\ &= n\epsilon_n + \sum_{i=1}^n I(M_1, M_2, \mathcal{C}_n, Y^{i-1}, X_{1i}, X_{2i}; Y_i | E_n = 1) \\ &= n\epsilon_n + \sum_{i=1}^n I(X_{1i}, X_{2i}; Y_i | E_n = 1) \\ &\leq n\epsilon_n + n(I(X_1, X_2; Y) + \delta_4(\epsilon)). \end{aligned} \quad (24)$$

To bound the second term in (23), we need the following lemma, which is proved in Appendix F.

Lemma 4: For every $\epsilon'' > \epsilon'$ and for n sufficiently large,

$$\begin{aligned} &I(M_{j^c}; Y^n | \mathcal{C}_n) \\ &\geq n[\min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} - \delta_5(\epsilon'')] - n\epsilon_n. \end{aligned}$$

Combining (23), (24), and Lemma 4 with $\epsilon'' = 2\delta_1(\epsilon)$, we have

$$\begin{aligned} nR_j &\leq n(I(X_1, X_2; Y) + \delta_4(\epsilon)) \\ &\quad - n[\min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} - \delta_6(\epsilon)] + 2n\epsilon_n \end{aligned} \quad (25)$$

for n sufficiently large. Letting $n \rightarrow \infty$ in (22) and (25) establishes

$$\begin{aligned} R_j &\leq I(X_j; Y | X_{j^c}) + \delta_3(\epsilon), \\ R_j &\leq I(X_1, X_2; Y) - \min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} + \delta_7(\epsilon). \end{aligned}$$

The proof of (17) follows by taking a continuous monotonic function $\delta'(\epsilon) \geq \max\{\delta_3(\epsilon), \delta_7(\epsilon)\}$ that tends to zero as $\epsilon \rightarrow 0$. Letting $\epsilon \rightarrow 0$ in (17) establishes (18), which completes the proof of Theorem 3. ■

The arguments used in the proof of (16a) starting from Fano's inequality can be generalized for a fixed (n, nR_1, nR_2) computation code to provide a general outer bound on the achievable rate pairs for \mathbf{a} -computation. It seems, however, difficult to generalize the arguments used in the proof of (16b). In particular, it is unclear whether Lemma 4 can be generalized to a fixed computation code. In Section VII, we present a single-letter outer bound on the achievable rate pairs for \mathbf{a} -computation and compare that with the inner bound implied by Theorem 2.

VI. OPTIMAL ACHIEVABLE RATES FOR BROADCAST CHANNELS WITH MARTON CODING

In this section, we apply the techniques developed in the previous sections to establish the optimal rate region for broadcast channels by Marton coding. Consider the two-receiver discrete memoryless broadcast channel (DM-BC) $(\mathcal{X}, p(y_1, y_2 | x), \mathcal{Y}_1 \times \mathcal{Y}_2)$ in Fig. 2, where the sender communicates independent messages M_1 and M_2 to respective receivers (see [15], [30], [31] for the formal definition of the communication problem over the broadcast channel). Throughout this section, information measures are in log base 2 to follow a similar notation with the existing literature.

Let $p = p(u_1, u_2)$ be a given pmf on some finite set $\mathcal{U}_1 \times \mathcal{U}_2$, and $x = x(u_1, u_2)$ be a function from $\mathcal{U}_1 \times \mathcal{U}_2$ to \mathcal{X} , and let $\epsilon > 0$ and $\alpha \in [0, 1]$. The random ensemble of Marton codes [15] is generated according to the following steps:

- 1) Let $\hat{R}_1 = \alpha(I(U_1; U_2) + 10\epsilon H(U_1, U_2))$ and $\hat{R}_2 = \bar{\alpha}(I(U_1; U_2) + 10\epsilon H(U_1, U_2))$, where $\bar{\alpha} := (1 - \alpha)^2$.
- 2) For every $m_1 \in [2^{n\hat{R}_1}]$, generate *auxiliary* codewords $u_1^n(m_1, l_1), l_1 \in [2^{n\hat{R}_1}]$, each drawn i.i.d. from $p(u_1)$. Similarly, for every $m_2 \in [2^{n\hat{R}_2}]$, generate *auxiliary* codewords $u_2^n(m_2, l_2), l_2 \in [2^{n\hat{R}_2}]$, each drawn i.i.d. from $p(u_2)$.
- 3) At the sender, for every message pair, $(m_1, m_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]$, find an index pair $(l_1, l_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]$ such that

$$(u_1^n(m_1, l_1), u_2^n(m_2, l_2)) \in \mathcal{T}_\epsilon^{(n)}(U_1, U_2),$$

and assign the codeword $x^n(m_1, m_2)$ as $x_i(u_{1i}(m_1, l_1), u_{2i}(m_2, l_2)), i \in [n]$. If there are more than one such pair of (l_1, l_2) , choose one of them uniformly at random; otherwise, choose one uniformly at random from $[2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]$.

We refer to the random tuple

$$\begin{aligned} \mathcal{C}_n &:= ((U_1^n(m_1, l_1) : m_1 \in [2^{n\hat{R}_1}], l_1 \in [2^{n\hat{R}_1}]), \\ &\quad (U_2^n(m_2, l_2) : m_2 \in [2^{n\hat{R}_2}], l_2 \in [2^{n\hat{R}_2}]), \\ &\quad ((L_1, L_2, x)(m_1, m_2) : m_1 \in [2^{n\hat{R}_1}], m_2 \in [2^{n\hat{R}_2}])) \end{aligned}$$

as the *Marton random codebook*. Each realization of the Marton random codebook \mathcal{C}_n results in one instance

$$\{x^n(m_1, m_2) : (m_1, m_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]\}$$

of such generated codebooks, which constitutes an (n, nR_1, nR_2) code for the DM-BC along with the optimal decoder. The random code ensemble generated in this manner is referred to as an $(n, nR_1, nR_2; p, x, \alpha, \epsilon)$ *Marton random code ensemble*, where $p = p(u_1, u_2)$ is the given pmf, $x = x(u_1, u_2)$ is the given function from $\mathcal{U}_1 \times \mathcal{U}_2$ to \mathcal{X} , $\alpha \in [0, 1]$ is the parameter used in step (1), and $\epsilon > 0$ is the parameter used in steps (1) and (3). A rate pair (R_1, R_2) is said to be *achievable by the (p, x, α, ϵ) -distributed Marton random code*

²One can obtain the same results by following a similar analysis when the rates \hat{R}_1 and \hat{R}_2 are chosen as $\alpha(I(U_1; U_2) + \delta(\epsilon)H(U_1, U_2))$ and $\bar{\alpha}(I(U_1; U_2) + \delta(\epsilon)H(U_1, U_2))$ respectively, for a sufficiently large $\delta(\epsilon)$ that tends to zero as $\epsilon \rightarrow 0$.

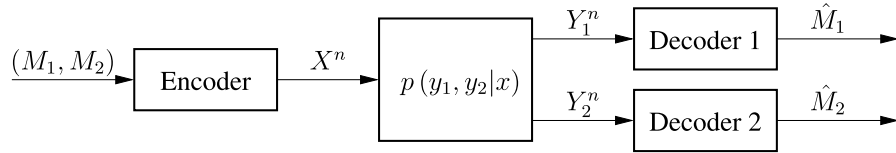


Fig. 2. Two-receiver broadcast channel.

ensemble if there exists a sequence of $(n, nR_1, nR_2; p, x, \alpha, \epsilon)$ Marton random code ensembles such that

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

where the expectation is with respect to the Marton random codebook \mathcal{C}_n . Given (p, x, α, ϵ) , let $\mathcal{R}_{\text{BC}}^*(p, x, \alpha, \epsilon)$ be the set of all rate pairs achievable by the (p, x, α, ϵ) -distributed Marton random code ensemble. Given pmf $p = p(u_1, u_2)$ and function $x = x(u_1, u_2)$, the optimal rate region $\mathcal{R}_{\text{BC}}^*(p, x)$, when it exists, is defined as

$$\mathcal{R}_{\text{BC}}^*(p, x) := \text{cl} \left[\bigcup_{\alpha \in [0, 1]} \lim_{\epsilon \rightarrow 0} \mathcal{R}_{\text{BC}}^*(p, x, \alpha, \epsilon) \right].$$

We are now ready to state main result of this section.

Theorem 4: Given a pmf $p(u_1, u_2)$ and a function $x = x(u_1, u_2)$, the optimal rate region $\mathcal{R}_{\text{BC}}^*(p, x)$ for the broadcast channel $p(y_1, y_2|x)$ is the closure of the set of rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U_1; Y_1, U_2) - \alpha I(U_1; U_2), \quad (26a)$$

$$R_1 \leq I(U_1, U_2; Y_1) - \min\{I(U_1, U_2; Y_1), I(U_2; Y_1, U_1) - \bar{\alpha}I(U_1; U_2), R_2\}, \quad (26b)$$

$$R_2 \leq I(U_2; Y_2, U_1) - \bar{\alpha}I(U_1; U_2), \quad (26c)$$

$$R_2 \leq I(U_1, U_2; Y_2) - \min\{I(U_1, U_2; Y_2), I(U_1; Y_2, U_2) - \alpha I(U_1; U_2), R_1\} \quad (26d)$$

for some $\alpha \in [0, 1]$.

We prove Theorem 4 by showing that given a pmf $p(u_1, u_2)$, a function $x(u_1, u_2)$, and $\alpha \in [0, 1]$, the rate region $\mathcal{R}_{\text{BC}}^*(p, x, \alpha) := \text{cl}[\lim_{\epsilon \rightarrow 0} \mathcal{R}_{\text{BC}}^*(p, x, \alpha, \epsilon)]$ is equal to the rate region characterized by (26), which we will denote as $\mathcal{R}_{\text{BC}}^{**}(p, x, \alpha)$. We take a two-step approach similar to Sections IV and V, and establish the inner and the outer bounds on the rate region $\mathcal{R}_{\text{BC}}^*(p, x, \alpha)$, respectively.

The inner bound is relegated to Appendix G. For the outer bound, given a fixed pmf $p = p(u_1, u_2)$, a function $x = x(u_1, u_2)$ from $\mathcal{U}_1 \times \mathcal{U}_2$ to \mathcal{X} , $\alpha \in [0, 1]$, and $\delta > 0$, we define the rate region $\mathcal{R}_{\text{BC}}^{**}(p, x, \alpha, \delta)$ as the set of rate pairs (R_1, R_2) such that

$$R_1 \leq I(U_1; Y_1, U_2) - \alpha I(U_1; U_2) + \delta, \quad (27a)$$

$$R_1 \leq I(U_1, U_2; Y_1) - \min \left\{ \begin{array}{l} I(U_1, U_2; Y_1), \\ I(U_2; Y_1, U_1) - \bar{\alpha}I(U_1; U_2), R_2 \end{array} \right\} + \delta, \quad (27b)$$

$$R_2 \leq I(U_2; Y_2, U_1) - \bar{\alpha}I(U_1; U_2) + \delta, \quad (27c)$$

$$R_2 \leq I(U_1, U_2; Y_2) - \min \left\{ \begin{array}{l} I(U_1, U_2; Y_2), \\ I(U_1; Y_2, U_2) - \alpha I(U_1; U_2), R_1 \end{array} \right\} + \delta. \quad (27d)$$

Note that the region $\mathcal{R}_{\text{BC}}^{**}(p, x, \alpha, \delta = 0)$ is equal to $\mathcal{R}_{\text{BC}}^*(p, x, \alpha)$ as defined in (26).

Proposition 2: Let $p = p(u_1, u_2)$ be a pmf, $x = x(u_1, u_2)$ be a function, $\alpha \in [0, 1]$, and $\epsilon > 0$. If a rate pair (R_1, R_2) is achievable by the (p, x, α, ϵ) -distributed Marton random code ensemble, then there exists a continuous $\delta'(\epsilon)$ that tends to zero monotonically as $\epsilon \rightarrow 0$ such that

$$(R_1, R_2) \in \mathcal{R}_{\text{BC}}^{**}(p, x, \alpha, \delta'(\epsilon)). \quad (28)$$

In particular,

$$\mathcal{R}_{\text{BC}}^*(p, x, \alpha) \subseteq \mathcal{R}_{\text{BC}}^{**}(p, x, \alpha). \quad (29)$$

Proof: We first start with an averaged version of Fano's inequality for a Marton random code ensemble \mathcal{C}_n . Consider a fixed codebook $\mathcal{C}_n = \mathcal{c}_n$. By Fano's inequality,

$$H(M_j | Y_j^n, \mathcal{C}_n = \mathcal{c}_n) \leq 1 + nR_j P_e^{(n)}(\mathcal{C}_n), \quad j = 1, 2.$$

Taking the expectation over Marton random codebook \mathcal{C}_n , it follows that

$$H(M_j | Y_j^n, \mathcal{C}_n) \leq 1 + nR_j \mathbf{E}_{\mathcal{C}_n} [P_e^{(n)}(\mathcal{C}_n)] \leq n\epsilon_n, \quad j = 1, 2, \quad (30)$$

for some $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$ since $\mathbf{E}_{\mathcal{C}_n} [P_e^{(n)}(\mathcal{C}_n)] \rightarrow 0$.

We next define the indicator random variable

$$\tilde{E}_n = \mathbb{1}_{\{(U_1^n(M_1, L_1), U_2^n(M_2, L_2)) \in \mathcal{T}_\epsilon^{(n)}(U_1, U_2)\}}. \quad (31)$$

Since $\hat{R}_1 + \hat{R}_2 = I(U_1; U_2) + 10\epsilon H(U_1, U_2)$, $\mathbf{P}(\tilde{E}_n = 0)$ tends to zero as $n \rightarrow \infty$ by the mutual covering lemma in [28, p. 208].

We are now ready to establish (27a). For n sufficiently large, we have

$$\begin{aligned} nR_1 &= H(M_1 | M_2, \mathcal{C}_n) \\ &\stackrel{(a)}{\leq} I(M_1; Y_1^n | M_2, \mathcal{C}_n) + n\epsilon_n \\ &\leq I(M_1, \tilde{E}_n; Y_1^n | M_2, \mathcal{C}_n) + n\epsilon_n \\ &\stackrel{(b)}{\leq} 1 + I(M_1; Y_1^n | M_2, \mathcal{C}_n, \tilde{E}_n) + n\epsilon_n \\ &\leq 1 + I(M_1; Y_1^n | M_2, \mathcal{C}_n, \tilde{E}_n = 0) \mathbf{P}(\tilde{E}_n = 0) \\ &\quad + I(M_1; Y_1^n | M_2, \mathcal{C}_n, \tilde{E}_n = 1) \mathbf{P}(\tilde{E}_n = 1) + n\epsilon_n \\ &\leq 1 + nR_1 \mathbf{P}(\tilde{E}_n = 0) + I(M_1; Y_1^n | M_2, \mathcal{C}_n, \tilde{E}_n = 1) + n\epsilon_n \\ &\leq 1 + nR_1 \mathbf{P}(\tilde{E}_n = 0) \\ &\quad + I(M_1, L_2; Y_1^n | M_2, \mathcal{C}_n, \tilde{E}_n = 1) + n\epsilon_n \\ &\leq 1 + nR_1 \mathbf{P}(\tilde{E}_n = 0) + n\hat{R}_2 \\ &\quad + I(M_1; Y_1^n | M_2, L_2, \mathcal{C}_n, \tilde{E}_n = 1) + n\epsilon_n \end{aligned}$$

$$\begin{aligned}
&= 1 + nR_1 \mathbf{P}(\tilde{E}_n = 0) + n\hat{R}_2 + n\epsilon_n \\
&\quad + \sum_{i=1}^n I(M_1; Y_{1i} | Y_1^{i-1}, M_2, L_2, \mathcal{C}_n, U_{2i}, \tilde{E}_n = 1) \\
&\leq 1 + nR_1 \mathbf{P}(\tilde{E}_n = 0) + n\hat{R}_2 + n\epsilon_n \\
&\quad + \sum_{i=1}^n I(M_1, U_{1i}, Y_1^{i-1}, M_2, L_2, \mathcal{C}_n; Y_{1i} | U_{2i}, \tilde{E}_n = 1) \\
&\stackrel{(c)}{=} 1 + nR_1 \mathbf{P}(\tilde{E}_n = 0) + n\hat{R}_2 + n\epsilon_n \\
&\quad + \sum_{i=1}^n I(U_{1i}; Y_{1i} | U_{2i}, \tilde{E}_n = 1) \\
&\stackrel{(d)}{\leq} 1 + nR_1 \mathbf{P}(\tilde{E}_n = 0) + n\hat{R}_2 + n\epsilon_n \\
&\quad + n(I(U_1; Y_1 | U_2) + \delta_1(\epsilon)), \\
&\leq 1 + nR_1 \mathbf{P}(\tilde{E}_n = 0) + n\bar{\alpha}(I(U_1; U_2) + \delta_2(\epsilon)) + n\epsilon_n \\
&\quad + n(I(U_1; Y_1 | U_2) + \delta_1(\epsilon)), \\
&\stackrel{(e)}{\leq} n(I(U_1; Y_1, U_2) - \alpha I(U_1; U_2) + \delta_3(\epsilon)) + 2n\epsilon_n, \quad (32)
\end{aligned}$$

where (a) follows by (the averaged version of) Fano's inequality in (30), (b) follows since \tilde{E}_n is a binary random variable, (c) follows since $(M_1, M_2, Y_1^{i-1}, \mathcal{C}_n, \tilde{E}_n) \rightarrow (U_{1i}, U_{2i}) \rightarrow Y_{1i}$ form a Markov chain for every $i \in [n]$, (d) follows by the memoryless property of the channel and by Lemma 9 in Appendix E since the distribution of $(U_1^n(M_1, L_1), U_2^n(M_2, L_2))$ is permutation invariant by construction, and (e) follows since $\mathbf{P}(\tilde{E}_n = 0)$ tends to zero as $n \rightarrow \infty$.

For the proof of (27b), we start with

$$\begin{aligned}
nR_1 &= H(M_1 | M_2, \mathcal{C}_n) \\
&\stackrel{(a)}{\leq} I(M_1; Y_1^n | M_2, \mathcal{C}_n) + n\epsilon_n \\
&= I(M_1, M_2; Y_1^n | \mathcal{C}_n) - I(M_2; Y_1^n | \mathcal{C}_n) + n\epsilon_n, \quad (33)
\end{aligned}$$

where (a) follows by (the averaged version of) Fano's inequality in (30). Following arguments similar to (32), the first term in (33) can be bounded as

$$\begin{aligned}
&I(M_1, M_2; Y_1^n | \mathcal{C}_n) \\
&\leq 1 + n(R_1 + R_2) \mathbf{P}(\tilde{E}_n = 0) \\
&\quad + \sum_{i=1}^n I(M_1, M_2; Y_{1i} | \mathcal{C}_n, Y_1^{i-1}, \tilde{E}_n = 1) \\
&\leq n\epsilon_n + \sum_{i=1}^n I(M_1, M_2, \mathcal{C}_n, Y_1^{i-1}; Y_{1i} | \tilde{E}_n = 1) \\
&= n\epsilon_n + \sum_{i=1}^n I(M_1, M_2, \mathcal{C}_n, Y_1^{i-1}, U_{1i}, U_{2i}; Y_{1i} | \tilde{E}_n = 1) \\
&= n\epsilon_n + \sum_{i=1}^n I(U_{1i}, U_{2i}; Y_{1i} | \tilde{E}_n = 1), \\
&\leq n\epsilon_n + n(I(U_1, U_2; Y_1) + \delta_4(\epsilon)). \quad (34)
\end{aligned}$$

For the second term in (33), we need the following lemma, which is proved in Appendix H. This lemma is a version of Lemma 4 for Marton random code ensembles.

Lemma 5: For every $\epsilon' > \epsilon$ and for n sufficiently large,

$$\begin{aligned}
&I(M_2; Y_1^n | \mathcal{C}_n) \\
&\geq n \left[\min \left\{ \begin{array}{l} R_2, I(U_1, U_2; Y_1), \\ I(U_2; Y_1, U_1) - \bar{\alpha} I(U_1; U_2) \end{array} \right\} - \delta_5(\epsilon') \right] - n\epsilon_n.
\end{aligned}$$

Combining (33), (34), and Lemma 5 with $\epsilon' = 2\epsilon$, we have

$$\begin{aligned}
nR_1 &\leq nI(U_1, U_2; Y_1) + n\delta_6(\epsilon) + 2n\epsilon_n \\
&\quad - n \min \left\{ \begin{array}{l} R_2, I(U_1, U_2; Y_1), \\ I(U_2; Y_1, U_1) - \bar{\alpha} I(U_1; U_2) \end{array} \right\} \quad (35)
\end{aligned}$$

for n sufficiently large.

For (27c) and (27d), we establish similarly for decoder 2

$$nR_2 \leq n(I(U_2; Y_2, U_1) - \bar{\alpha} I(U_1; U_2) + \delta_7(\epsilon)) + 2n\epsilon_n \quad (36)$$

and

$$\begin{aligned}
nR_2 &\leq nI(U_1, U_2; Y_2) + n\delta_8(\epsilon) + 2n\epsilon_n \\
&\quad - n \min \left\{ \begin{array}{l} R_1, I(U_1, U_2; Y_2), \\ I(U_1; Y_2, U_2) - \alpha I(U_1; U_2) \end{array} \right\} \quad (37)
\end{aligned}$$

for n sufficiently large. The proof of (28) follows by letting $n \rightarrow \infty$ in (32), (35), (36), and (37), and taking a continuous monotonic function $\delta'(\epsilon) \geq \max\{\delta_3(\epsilon), \delta_6(\epsilon), \delta_7(\epsilon), \delta_8(\epsilon)\}$ that tends to zero as $\epsilon \rightarrow 0$. Letting $\epsilon \rightarrow 0$ in (28) establishes (29), which completes the proof of Proposition 2. ■

Remark 4: Marton coding we have analyzed involves two codewords. Marton's original coding scheme [15] uses rate splitting and superposition coding, and involves an additional codeword that carries messages for both receivers (see also [28, Proposition 8.1]). Our technique can be similarly adapted to this general version of Marton coding.

VII. DISCUSSION

For the linear computation problem, the outer bound on the optimal rate region presented in Section V is valid for *any* computation, not only for natural computation. The inner bound presented in Theorem 2, however, matches this outer bound only for *natural* computation. It is an interesting but difficult problem to characterize the optimal rate region for *arbitrary* linear computation by a random homologous code ensemble. At this point, it is unclear whether it is the inner or outer bound that is loose (or both). The extension of the results in this paper to more than two senders is also a challenging question.

A more fundamental question is to characterize the *capacity region* of the linear computation problem. The following presents an outer bound on the rate pairs (R_1, R_2) that is achievable for *a*-computation by *any code*. The proof is deferred to Appendix I.

Proposition 3 (A general outer bound): Given a vector $\mathbf{a} = [a_1 \ a_2] \in \mathbb{F}_q^2$ with $a_1, a_2 \neq 0$, if a rate pair (R_1, R_2) is

achievable for \mathbf{a} -computation, then it must satisfy

$$R_1 \leq \min \left\{ I(X_1; Y | X_2, Q), \right. \\ \left. I(X_1, X_2; Y | Q) - I(X_2; W_{\mathbf{a}}, Y | T, Q) \right\}, \quad (38a)$$

$$R_2 \leq \min \left\{ I(X_2; Y | X_1, Q), \right. \\ \left. I(X_1, X_2; Y | Q) - I(X_1; W_{\mathbf{a}}, Y | T, Q) \right\}, \quad (38b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y | Q) + I(X_1, X_2; W_{\mathbf{a}}, Y | T, Q) \\ - I(X_1; W_{\mathbf{a}}, Y | T, Q) - I(X_2; W_{\mathbf{a}}, Y | T, Q), \quad (38c)$$

for some $p(q)p(x_1|q)p(x_2|q)p(t|x_1, x_2, q)$ such that $(T, Q) \rightarrow (X_1, X_2) \rightarrow W_{\mathbf{a}}$ and $(T, Q, W_{\mathbf{a}}) \rightarrow (X_1, X_2) \rightarrow Y$ each form a Markov chain, and

$$I(X_1; W_{\mathbf{a}}, Y | T, Q) + I(X_2; W_{\mathbf{a}}, Y | T, Q) \\ \leq I(X_1, X_2; W_{\mathbf{a}}, Y | T, Q). \quad (39)$$

Note that (39) is equivalent to

$$I(X_1; X_2 | T, Q) \leq I(X_1; X_2 | W_{\mathbf{a}}, Y, T, Q),$$

which is a variation of dependence-balance condition for two-way channels [32].

We next take a closer look at achievability. First, note that by Theorem 2, there exists a sequence of (fixed) (n, nR_1, nR_2) computation codes that have vanishing error probability and satisfy (2) if

$$(R_1, R_2) \in \mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_1(p) \cup \mathcal{R}_2(p)$$

for some input pmf $p = p(x_1)p(x_2)$. We now convexify this achievable rate region to get the following general inner bound on the capacity region for \mathbf{a} -computation.

Proposition 4 (A general inner bound): Given a vector $\mathbf{a} = [a_1 \ a_2] \in \mathbb{F}_q^2$ with $a_1, a_2 \neq 0$, a rate pair (R_1, R_2) is achievable for \mathbf{a} -computation if

$$R_1 \leq \min \left\{ I(X_1; Y | X_2, Q), \right. \\ \left. I(X_1, X_2; Y | Q) - I(X_2; W_{\mathbf{a}}, Y | T, Q) \right\}, \quad (40a)$$

$$R_2 \leq \min \left\{ I(X_2; Y | X_1, Q), \right. \\ \left. I(X_1, X_2; Y | Q) - I(X_1; W_{\mathbf{a}}, Y | T, Q) \right\}, \quad (40b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y | Q) + I(X_1, X_2; W_{\mathbf{a}}, Y | T, Q) \\ - I(X_1; W_{\mathbf{a}}, Y | T, Q) - I(X_2; W_{\mathbf{a}}, Y | T, Q), \quad (40c)$$

for some $p(q)p(x_1|q)p(x_2|q)p(t|x_1, x_2, q)$ such that

$$T|x_1, x_2, q \sim \begin{cases} (x_1, x_2) & \text{with probability } \beta \\ \emptyset & \text{with probability } 1 - \beta \end{cases} \quad (41)$$

for some $\beta \in [0, 1]$.

Remark 5: One can notice the structural similarity of the rate regions in Propositions 3 and 4. Indeed, the rate region in Proposition 4 is a special case of the one in Proposition 3 since every pmf $p(q)p(x_1|q)p(x_2|q)p(t|x_1, x_2, q)$ such that $T|x_1, x_2, q$ is conditionally randomly drawn according to (41)

satisfies the Markov chains in Proposition 3 as well as the condition in (39).

Proof: [Proof of Proposition 4] Taking the convex hull of the rate region in Theorem 2, we know that the rate region

$$\text{conv} \left(\bigcup_{p=p(x_1)p(x_2)} [\mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_1(p) \cup \mathcal{R}_2(p)] \right) \\ = \text{conv} \left(\bigcup_{p=p(x_1)p(x_2)} \text{conv} [\mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_1(p) \cup \mathcal{R}_2(p)] \right) \\ \stackrel{(a)}{=} \text{conv} \left(\bigcup_{p=p(x_1)p(x_2)} \text{conv} [\mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_{\text{MAC}}(p)] \right)$$

is achievable, where (a) follows since for every pmf $p = p(x_1)p(x_2)$, $\text{conv}(\mathcal{R}_1(p) \cup \mathcal{R}_2(p)) = \mathcal{R}_{\text{MAC}}(p)$. We now prove that this achievable rate region is equivalent to the rate region in Proposition 4. Consider a fixed $Q = q$ and let $p_q := p(x_1|q)p(x_2|q)$ and $(X_{1q}, X_{2q}) \sim p_q$. It suffices to show that the rate region defined by (40) evaluated for $Q = q$ and p_q is equivalent to

$$\text{conv}[\mathcal{R}_{\text{CF}}(p_q, \mathbf{a}) \cup \mathcal{R}_{\text{MAC}}(p_q)].$$

To see this, note that when $T = (X_{1q}, X_{2q})$, the rate region defined by (40) reduces to $\mathcal{R}_{\text{MAC}}(p_q)$. Similarly, when $T = \emptyset$, the rate region defined by (40) reduces to $\mathcal{R}_{\text{CF}}(p_q, \mathbf{a})$. In words, the random variable T for different $\beta \in [0, 1]$ values plays the role of time-sharing between the rate regions $\mathcal{R}_{\text{MAC}}(p_q)$ and $\mathcal{R}_{\text{CF}}(p_q, \mathbf{a})$. Therefore, taking the union over $\beta \in [0, 1]$ results in $\text{conv}[\mathcal{R}_{\text{CF}}(p_q, \mathbf{a}) \cup \mathcal{R}_{\text{MAC}}(p_q)]$, which completes the proof. ■

APPENDIX A

PROOF OF PROPOSITION 1

Fix a pmf $p = p(x_1)p(x_2)$ and a nonzero vector $\mathbf{a} \in \mathbb{F}_q^2$. We first show that $[\mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_{\text{MAC}}(p)] \subseteq \mathcal{R}^*(p, \mathbf{a})$. Suppose that the rate pair $(R_1, R_2) \in \mathcal{R}_{\text{CF}}(p, \mathbf{a})$. Then, for every $j \in \{1, 2\}$ with $a_j \neq 0$, the rate pair (R_1, R_2) satisfies

$$R_j \leq H(X_j) - H(W_{\mathbf{a}}|Y) \\ \leq H(X_j) - H(W_{\mathbf{a}}|Y, X_{j^c}) \\ = I(X_j; Y | X_{j^c}),$$

and

$$R_j \leq H(X_j) - H(W_{\mathbf{a}}|Y) \\ = I(X_1, X_2; Y) - I(X_{j^c}; W_{\mathbf{a}}, Y) \\ \leq I(X_1, X_2; Y) - \min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\},$$

which implies that the rate pair $(R_1, R_2) \in \mathcal{R}^*(p, \mathbf{a})$. It follows that $\mathcal{R}_{\text{CF}}(p, \mathbf{a}) \subseteq \mathcal{R}^*(p, \mathbf{a})$. Similarly, suppose that the rate pair $(R_1, R_2) \in \mathcal{R}_{\text{MAC}}(p)$. Then, for every $j \in \{1, 2\}$ with $a_j \neq 0$, the rate pair (R_1, R_2) satisfies

$$R_j \leq I(X_j; Y | X_{j^c}),$$

and

$$R_j \leq I(X_1, X_2; Y) - R_{j^c} \\ \leq I(X_1, X_2; Y) - \min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\},$$

which implies that the rate pair $(R_1, R_2) \in \mathcal{R}^*(p, \mathbf{a})$. Therefore, $\mathcal{R}_{\text{MAC}}(p) \subseteq \mathcal{R}^*(p, \mathbf{a})$.

Next, we show that $\mathcal{R}^*(p, \mathbf{a}) \subseteq [\mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_{\text{MAC}}(p)]$. Suppose that the rate pair $(R_1, R_2) \in \mathcal{R}^*(p, \mathbf{a})$ such that $R_{j^c} > I(X_{j^c}; W_{\mathbf{a}}, Y)$ for every $j \in \{1, 2\}$ with $a_j \neq 0$. Then, (R_1, R_2) satisfies

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

for every $j \in \{1, 2\}$ with $a_j \neq 0$. Then, $(R_1, R_2) \in \mathcal{R}_{\text{CF}}(p, \mathbf{a})$. It is easy to see that the rate pair $(R_1, R_2) \in \mathcal{R}^*(p, \mathbf{a})$ that satisfies $R_{j^c} \leq I(X_{j^c}; W_{\mathbf{a}}, Y)$ for some $j \in \{1, 2\}$ with $a_j \neq 0$, is included in $\mathcal{R}_{\text{MAC}}(p)$. Thus, $\mathcal{R}^*(p, \mathbf{a}) \subseteq [\mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_{\text{MAC}}(p)]$, which completes the proof.

APPENDIX B

Lemma 6: Let G be an $nR \times n$ random matrix over \mathbb{F}_q with $R < 1$ where each element is drawn i.i.d. $\text{Unif}(\mathbb{F}_q)$. Then,

$$\mathbf{P}(G \text{ is not full rank}) \leq q^{-n(1-R-\epsilon_n)},$$

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

Proof: Probability of choosing nR linearly independent rows can be written as

$$\begin{aligned} \mathbf{P}(G \text{ is full rank}) &= \frac{\prod_{j=1}^{nR} (q^n - q^{j-1})}{(q^n)^{nR}} \\ &= \prod_{j=1}^{nR} (1 - q^{j-1-n}) \\ &\geq (1 - q^{-n(1-R)})^{nR} \\ &\stackrel{(a)}{\geq} 1 - nRq^{-n(1-R)}, \end{aligned}$$

where (a) follows by Bernoulli's inequality for n large enough since $R < 1$. Using this relation, we have

$$\begin{aligned} \mathbf{P}(G \text{ is not full rank}) &= 1 - \mathbf{P}(G \text{ is full rank}) \\ &\leq nRq^{-n(1-R)}. \end{aligned}$$

Defining $\epsilon_n = \frac{\log_q(nR)}{n}$ completes the proof. ■

APPENDIX C PROOF OF LEMMA 1

Fix a pmf $p = p(x_1)p(x_2)$ and a nonzero vector $\mathbf{a} \in \mathbb{F}_q^2$. We will show that if the condition in (7) holds, then $\mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_1(p) \cup \mathcal{R}_2(p) = \mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_{\text{MAC}}(p)$. To start with, note that the rate regions $\mathcal{R}_1(p)$ and $\mathcal{R}_2(p)$ have one additional rate constraint compared to $\mathcal{R}_{\text{MAC}}(p)$. Therefore, $\mathcal{R}_j(p) \subseteq \mathcal{R}_{\text{MAC}}(p)$ for $j = 1, 2$ and it follows that $\mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_1(p) \cup \mathcal{R}_2(p) \subseteq \mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_{\text{MAC}}(p)$ holds in general. Then, it suffices to show that if the condition in (7) holds, then $\mathcal{R}_{\text{MAC}}(p) \subseteq [\mathcal{R}_{\text{CF}}(p, \mathbf{a}) \cup \mathcal{R}_1(p) \cup \mathcal{R}_2(p)]$. Suppose that the condition in (7) is satisfied. Let the rate pair $(R_1, R_2) \in \mathcal{R}_{\text{MAC}}(p)$ be such that $R_{j^c} > I(X_{j^c}; W_{\mathbf{a}}, Y)$ for every $j \in \{1, 2\}$ with $a_j \neq 0$. Then, (R_1, R_2) satisfies

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

for every $j \in \{1, 2\}$ with $a_j \neq 0$, implying that $(R_1, R_2) \in \mathcal{R}_{\text{CF}}(p, \mathbf{a})$. Now, let the rate pair $(R_1, R_2) \in \mathcal{R}_{\text{MAC}}(p)$ be such that $R_{j^c} \leq I(X_{j^c}; W_{\mathbf{a}}, Y)$ for some $j \in \{1, 2\}$ with $a_j \neq 0$. By condition (7), we have

$$\begin{aligned} I(X_{j^c}; W_{\mathbf{a}}, Y) &= I(X_1, X_2; Y) - H(X_j) + H(W_{\mathbf{a}}|Y) \\ &= I(X_1, X_2; Y) - H(X_j) + \min_{\mathbf{b} \neq 0} H(W_{\mathbf{b}}|Y) \\ &\leq I(X_1, X_2; Y) - H(X_j) + \min_{\substack{b_1, b_2 \\ \in \mathbb{F}_q^*}} H(W_{\mathbf{b}}|Y). \end{aligned}$$

Then, the rate pair $(R_1, R_2) \in \mathcal{R}_1(p) \cup \mathcal{R}_2(p)$, which completes the proof.

APPENDIX D PROOF OF LEMMA 2

Note that for $j = 1, 2$,

$$\begin{aligned} H(M_j|Y^n, M_{j^c}, \mathcal{C}_n) &= I(M_j; W_{\mathbf{a}}^n|Y^n, M_{j^c}, \mathcal{C}_n) + H(M_j|W_{\mathbf{a}}^n, Y^n, M_{j^c}, \mathcal{C}_n) \\ &\leq H(W_{\mathbf{a}}^n|Y^n, \mathcal{C}_n) + H(M_j|W_{\mathbf{a}}^n, Y^n, M_{j^c}, \mathcal{C}_n). \end{aligned} \quad (42)$$

To bound the first term in (42), we need a version of Fano's inequality for computation.

Lemma 7: If the average probability of error $\mathbf{E}_{\mathcal{C}_n}[P_e^{(n)}(\mathcal{C}_n)]$ tends to zero as $n \rightarrow \infty$, then

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

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

Proof: For fixed codebook $\mathcal{C}_n = \mathcal{C}_n$, by Fano's inequality

$$H(W_{\mathbf{a}}^n|Y^n, \mathcal{C}_n = \mathcal{C}_n) \leq 1 + nP_e^{(n)}(\mathcal{C}_n).$$

Taking the expectation over the random homologous codebook \mathcal{C}_n , we have

$$H(W_{\mathbf{a}}^n|Y^n, \mathcal{C}_n) \leq 1 + n\mathbf{E}_{\mathcal{C}_n}[P_e^{(n)}(\mathcal{C}_n)] \stackrel{(a)}{\leq} n\epsilon_n,$$

where (a) follows since $\mathbf{E}_{\mathcal{C}_n}[P_e^{(n)}(\mathcal{C}_n)]$ tends to zero as $n \rightarrow \infty$. ■

Suppose that $a_j \neq 0$. Define indicator variable θ_j , $j = 1, 2$, such that $\theta_j = 1$ if M_j is confusable. Combining (42) with Lemma 7, we have

$$\begin{aligned} H(M_j|Y^n, M_{j^c}, \mathcal{C}_n) &\leq n\epsilon_n + H(M_j|W_{\mathbf{a}}^n, Y^n, M_{j^c}, \mathcal{C}_n) \\ &\stackrel{(a)}{=} n\epsilon_n + H(M_j|W_{\mathbf{a}}^n, X_{j^c}^n(M_{j^c}), Y^n, M_{j^c}, \mathcal{C}_n) \\ &\stackrel{(b)}{=} n\epsilon_n + H(M_j|W_{\mathbf{a}}^n, X_j^n(M_j), X_{j^c}^n(M_{j^c}), Y^n, M_{j^c}, \mathcal{C}_n) \\ &\leq n\epsilon_n + H(M_j|X_j^n(M_j), \mathcal{C}_n) \\ &\leq n\epsilon_n + H(M_j, \theta_j|X_j^n(M_j), \mathcal{C}_n) \\ &\stackrel{(c)}{\leq} n\epsilon_n + \log_q 2 + H(M_j|X_j^n(M_j), \mathcal{C}_n, \theta_j) \\ &= n\epsilon_n + \log_q 2 + H(M_j|X_j^n(M_j), \mathcal{C}_n, \theta_j = 1) \mathbf{P}(\theta_j = 1) \\ &\leq n\epsilon_n + \log_q 2 + nR_j \mathbf{P}(\theta_j = 1) \\ &\stackrel{(d)}{\leq} n\epsilon_n + \log_q 2 + nR_j \epsilon_n \\ &= n(\epsilon_n + \frac{\log_q 2}{n} + R_j \epsilon_n), \end{aligned}$$

where (a) follows since $X_{j^c}^{n_j}(M_{j^c})$ is a function of (M_{j^c}, \mathcal{C}_n) , (b) follows since $X_j^n(M_j)$ is a function of $(X_{j^c}^{n_j}(M_{j^c}), W_{\mathbf{a}}^n)$ when $a_j \neq 0$, (c) follows since θ_j is a binary random variable, and (d) follows by the assumption in (19) in Lemma 2.

APPENDIX E PROOF OF LEMMA 3

For the simplicity of the exposition without loss of generality, we provide a proof from a single-sender perspective and the memoryless point-to-point channel $p_{Y|X}(y|x)$ with $\mathcal{X} = \mathbb{F}_q$. Let $\epsilon > 0$ and $p = p(x)$ be a pmf on \mathbb{F}_q . Define an $(n, nR; p_X, \epsilon)$ random nested coset code ensemble following steps 1)-3) for a single sender. Let X^n be the codeword sent through the channel, $\epsilon' > 0$, $i \in [n]$, and $(x, y) \in \mathbb{F}_q \times \mathcal{Y}$. Then,

$$\begin{aligned} & \mathbf{P}(X_i = x, Y_i = y | X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)) \\ &= \mathbf{P}(X_i = x | X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)) \\ & \quad \times \mathbf{P}(Y_i = y | X_i = x, X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)) \\ &= \mathbf{P}(X_i = x | X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)) p_{Y|X}(y|x). \end{aligned} \quad (43)$$

We make a connection between the conditional distribution of X_i given $\{X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)\}$ and the input pmf $p(x)$. Therefore, we start with exploring the conditional distribution of X_i given $\{X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)\}$.

Lemma 8: Let $p(x)$ be a pmf on \mathbb{F}_q , and $\epsilon, \epsilon' > 0$. Define $\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)$ as the set of elements in $\mathcal{T}_{\epsilon'}^{(n)}(X)$ with type Θ . Suppose $X^n(m) = U^n(m, L(m))$ denote the random codeword assigned to message m by $(n, nR; p(x), \epsilon)$ random nested coset code ensemble. Then,

$$U^n(m, L) | \{U^n(m, L) \in \mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)\} \sim \text{Unif}(\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)),$$

for every $m \in \mathbb{F}_q^{nR}$.

Proof: Without loss of generality, we drop index m . It suffices to show that the distribution of $U^n(L)$ is permutation invariant. Let u^n, v^n have the same type (typical or not) and let $v^n = \sigma(u^n)$ for some permutation σ . Then, we have

$$\begin{aligned} & \mathbf{P}(U^n(L) = u^n) \\ &= \sum_l \sum_G \mathbf{P}(L = l, G = G, D^n = u^n \ominus lG) \\ &\stackrel{(a)}{=} \sum_l \sum_G \mathbf{P}(L = l, G = \sigma(G), D^n = v^n \ominus l\sigma(G)) \\ &= \mathbf{P}(U^n(L) = v^n), \end{aligned}$$

where $\sigma(G)$ is the matrix constructed by applying permutation σ to the columns of G and (a) follows since a permutation applied to a coset code preserves the type of each codeword. ■

Building on top of Lemma 8, we next establish that the conditional distribution of X_i given $\{X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)\}$ is close to the input pmf $p(x)$.

Lemma 9: Let $\epsilon' > 0$. Define $\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)$ in a similar way to Lemma 8. Suppose that the distribution of X^n is uniform within $\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)$, namely,

$$X^n | \{X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)\} \sim \text{Unif}(\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)) \quad (44)$$

for every type Θ such that $\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta) \neq \emptyset$. Then, conditioned on the typical set, X_i 's have identical distribution that satisfies

$$(1 - \epsilon')p(x) \leq \mathbf{P}(X_i = x | X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)) \leq (1 + \epsilon')p(x),$$

for every $x \in \mathcal{X}$.

Proof: Let $x \in \mathcal{X}$. For a type Θ , let Θ_x denote the empirical mode of x within type Θ . Then, for every type Θ such that $\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta) \neq \emptyset$, we have

$$\begin{aligned} & \mathbf{P}(X_i = x | X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)) \\ &= \sum_{\substack{x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X, \Theta) \\ \text{s.t. } x_i = x}} \mathbf{P}(X^n = x^n | X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)) \\ &\stackrel{(a)}{=} \sum_{\substack{x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X, \Theta) \\ x_i = x}} \frac{1}{|\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)|} \\ &\stackrel{(b)}{=} \Theta_x | \mathcal{T}_{\epsilon'}^{(n)}(X, \Theta) | \frac{1}{|\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)|} \\ &= \Theta_x, \end{aligned}$$

where (a) follows since X^n is conditionally uniform over $\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)$, and (b) follows since $\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)$ is closed under permutation. Combining this observation with the fact that Θ is the type of a typical sequence, we get

$$(1 - \epsilon')p(x) \leq \mathbf{P}(X_i = x | X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)) \leq (1 + \epsilon')p(x),$$

for every $x \in \mathcal{X}$. Since $\mathcal{T}_{\epsilon'}^{(n)}(X)$ is the disjoint union of $\mathcal{T}_{\epsilon'}^{(n)}(X, \Theta)$ over all types, multiplying each side with $\mathbf{P}(X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X, \Theta))$ and then summing over Θ gives

$$\begin{aligned} (1 - \epsilon')p(x) \mathbf{P}(X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)) \\ \leq \mathbf{P}(X_i = x, X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)) \\ \leq (1 + \epsilon')p(x) \mathbf{P}(X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)), \end{aligned}$$

for every $x \in \mathcal{X}$. The claim follows from dividing each side by $\mathbf{P}(X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X))$. ■

Back to the proof of Lemma 3, we have by Lemma 8 that the distribution of X^n (codeword from an $(n, nR; p(x), \epsilon)$ random nested coset code ensemble) satisfies the condition in (44) in Lemma 9. Therefore, combining (43) with Lemma 9 completes the proof.

APPENDIX F PROOF OF LEMMA 4

Let $\epsilon'' > \epsilon'$. Suppose that $a_j \neq 0$, and $j^c = \{1, 2\} \setminus \{j\}$. First, by Lemma 7, we have

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

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

$$\begin{aligned} & I(M_{j^c}; W_{\mathbf{a}}^n, Y^n | \mathcal{C}_n) \\ & \geq n[\min\{R_{j^c}, I(X_{j^c}; W_{\mathbf{a}}, Y)\} - \delta_5(\epsilon'') - \epsilon_n]. \end{aligned}$$

Similar to [26], 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

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

Note that the set \mathcal{L} is random with the underlying distribution on $(W_{\mathbf{a}}^n, Y^n, \mathcal{C}_n)$, which is induced by drawing a random homologous codebook \mathcal{C}_n and using this codebook to encode $X_1^n(M_1)$ and $X_2^n(M_2)$ that lead to $W_{\mathbf{a}}^n = a_1 X_1^n(M_1) \oplus a_2 X_2^n(M_2)$ and Y^n through the finite-field input memoryless MAC $p(y|x_1, x_2)$. We first bound the probability that an incorrect message is in the random set \mathcal{L} . Define two events $\mathcal{M}_1 = \{M_1 = M_2 = \mathbf{0}\}$ and $\mathcal{M}_2 = \{L_1(M_1) = L_2(M_2) = \mathbf{0}\}$. The indicator random variable E_n is as defined in (20). By the symmetry of the codebook generation, for every $m \neq \mathbf{0} \in \mathbb{F}_q^{nR_{j^c}}$, we have

$$\mathbf{P}(m \oplus M_{j^c} \in \mathcal{L}, E_n = 1) = \mathbf{P}(m \in \mathcal{L}, E_n = 1 | \mathcal{M}_1, \mathcal{M}_2). \quad (45)$$

The proof of this claim is given in (46)-(53), shown at the bottom of the page, where (48) follows since (M_1, M_2) is independent from (G, D_1^n, D_2^n) , (49) follows since (M_1, M_2) is uniformly distributed and

$$(G, D_1^n, D_2^n) \stackrel{d}{=} (G, [m_1 \ l_1 \ \mathbf{0}]G \oplus D_1^n, [m_2 \ l_2 \ \mathbf{0}]G \oplus D_2^n)$$

result in two equivalent codebooks (i.e., the same set of codebooks with permuted mappings from messages to codewords), and (53) follows by the fact proved in [22, Lemma 11] that (M_1, L_1, M_2, L_2) is uniformly distributed over its support.

Continuing from (53), the probability in (45) is bounded in (54)-(66), shown at the top of the next page, where step (a) follows since $\epsilon'' > \epsilon'$, (b) follows since conditioned on \mathcal{M}_1 and \mathcal{M}_2 , $U_{j^c}^n \rightarrow (D_1^n, D_2^n) \rightarrow Y^n$ form a Markov chain, (c) follows by [22, Lemma 11] since (G, D_1^n, D_2^n) is independent from (M_1, M_2) , and (d) follows by the construction of the random homologous codebook \mathcal{C}_n with $\hat{R}_i = D(p_{X_i} \| \text{Unif}(\mathbb{F}_q)) + \epsilon$. Since $\mathbf{P}(E_n = 1)$ tends to one as $n \rightarrow \infty$, for n sufficiently large we have $\mathbf{P}(E_n = 1) \geq q^{-\epsilon}$. Therefore, for n sufficiently large, the conditional probability is bounded as

$$\begin{aligned} \mathbf{P}(m \oplus M_{j^c} \in \mathcal{L} | E_n = 1) &= \frac{\mathbf{P}(m \oplus M_{j^c} \in \mathcal{L}, E_n = 1)}{\mathbf{P}(E_n = 1)} \\ &\leq \mathbf{P}(m \oplus M_{j^c} \in \mathcal{L}, E_n = 1) q^\epsilon \\ &\leq q^{-n(I(X_{j^c}; W_{\mathbf{a}}, Y) - \delta_5(\epsilon''))} q^\epsilon. \end{aligned}$$

The expected size of \mathcal{L} given $\{E_n = 1\}$ is then bounded as

$$\begin{aligned} \mathbf{E}(|\mathcal{L}| | E_n = 1) &\leq 1 + \sum_{m \neq \mathbf{0}} \mathbf{P}(m \oplus M_{j^c} \in \mathcal{L} | E_n = 1) \\ &\leq 1 + q^{n(R_{j^c} - I(X_{j^c}; W_{\mathbf{a}}, Y) + \delta_5(\epsilon'') + \frac{\epsilon}{n})} \\ &= 1 + q^{n(R_{j^c} - I(X_{j^c}; W_{\mathbf{a}}, Y) + \delta_5(\epsilon'') + \epsilon_n)}, \quad (67) \end{aligned}$$

for n sufficiently large. Define another indicator random variable $F_n = \mathbb{1}_{\{M_{j^c} \in \mathcal{L}\}}$. Since $\epsilon'' > \epsilon'$ and $\mathbf{P}(E_n = 1)$

$$\begin{aligned} &\mathbf{P}(m \oplus M_{j^c} \in \mathcal{L}, E_n = 1) \\ &= \mathbf{P}((X_{j^c}^n(m \oplus M_{j^c}), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y), (X_1^n(M_1), X_2^n(M_2)) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2)) \quad (46) \end{aligned}$$

$$= \sum_{\substack{m_1, l_1 \\ m_2, l_2}} \sum_{\substack{G \\ d_1^n, d_2^n}} \mathbf{P} \left(\begin{array}{l} (M_1, M_2) = (m_1, m_2), (L_1(M_1), L_2(M_2)) = (l_1, l_2), G = G, D_1^n = d_1^n, D_2^n = d_2^n, \\ (X_{j^c}^n(m \oplus m_{j^c}), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y), (X_1^n(m_1), X_2^n(m_2)) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) \end{array} \right) \quad (47)$$

$$\begin{aligned} &= \sum_{\substack{m_1, l_1 \\ m_2, l_2}} \sum_{\substack{G \\ d_1^n, d_2^n}} \mathbf{P}(M_1 = m_1, M_2 = m_2) \mathbf{P}(G = G, D_1^n = d_1^n, D_2^n = d_2^n) \\ &\quad \mathbf{P} \left(\begin{array}{l} L_1(M_1) = l_1, L_2(M_2) = l_2, \\ (X_{j^c}^n(m \oplus m_{j^c}), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y), \\ (X_1^n(m_1), X_2^n(m_2)) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) \end{array} \middle| \begin{array}{l} (M_1, M_2) = (m_1, m_2), \\ G = G, D_1^n = d_1^n, D_2^n = d_2^n \end{array} \right) \quad (48) \end{aligned}$$

$$\begin{aligned} &= \sum_{\substack{m_1, l_1 \\ m_2, l_2}} \sum_{\substack{G \\ d_1^n, d_2^n}} \mathbf{P}(M_1 = \mathbf{0}, M_2 = \mathbf{0}) \mathbf{P}(G = G, D_1^n = [m_1 \ l_1 \ \mathbf{0}]G \oplus d_1^n, D_2^n = [m_2 \ l_2 \ \mathbf{0}]G \oplus d_2^n) \\ &\quad \mathbf{P} \left(\begin{array}{l} L_1(M_1) = \mathbf{0}, L_2(M_2) = \mathbf{0}, \\ (X_{j^c}^n(m), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y), \\ (X_1^n(\mathbf{0}), X_2^n(\mathbf{0})) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) \end{array} \middle| \begin{array}{l} (M_1, M_2) = (\mathbf{0}, \mathbf{0}), \\ G = G, D_1^n = [m_1 \ l_1 \ \mathbf{0}]G \oplus d_1^n, \\ D_2^n = [m_2 \ l_2 \ \mathbf{0}]G \oplus d_2^n \end{array} \right) \quad (49) \end{aligned}$$

$$\begin{aligned} &= \sum_{\substack{m_1, l_1 \\ m_2, l_2}} \sum_{\substack{G \\ d_1^n, d_2^n}} \mathbf{P} \left(\begin{array}{l} (M_1, M_2) = (\mathbf{0}, \mathbf{0}), (L_1(M_1), L_2(M_2)) = (\mathbf{0}, \mathbf{0}), \\ G = G, D_1^n = [m_1 \ l_1 \ \mathbf{0}]G \oplus d_1^n, D_2^n = [m_2 \ l_2 \ \mathbf{0}]G \oplus d_2^n, \\ (X_{j^c}^n(m), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y), (X_1^n(\mathbf{0}), X_2^n(\mathbf{0})) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) \end{array} \right) \quad (50) \end{aligned}$$

$$= \sum_{\substack{m_1, l_1 \\ m_2, l_2}} \mathbf{P} \left(\begin{array}{l} (M_1, M_2) = (\mathbf{0}, \mathbf{0}), (L_1(M_1), L_2(M_2)) = (\mathbf{0}, \mathbf{0}), \\ (X_{j^c}^n(m), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y), (X_1^n(\mathbf{0}), X_2^n(\mathbf{0})) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) \end{array} \right) \quad (51)$$

$$= \sum_{\substack{m_1, l_1 \\ m_2, l_2}} \mathbf{P}(\mathcal{M}_1, \mathcal{M}_2) \mathbf{P} \left(\begin{array}{l} (X_{j^c}^n(m), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y), \\ (X_1^n(\mathbf{0}), X_2^n(\mathbf{0})) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) \end{array} \middle| \begin{array}{l} \mathcal{M}_1 \\ \mathcal{M}_2 \end{array} \right) \quad (52)$$

$$= \mathbf{P}((X_{j^c}^n(m), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y), (X_1^n(\mathbf{0}), X_2^n(\mathbf{0})) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) | \mathcal{M}_1, \mathcal{M}_2), \quad (53)$$

$$\begin{aligned} & \mathbf{P}(m \in \mathcal{L}, E_n = 1 | \mathcal{M}_1, \mathcal{M}_2) \\ &= \mathbf{P}((X_{j^c}^n(m), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y) | (X_1^n(\mathbf{0}), X_2^n(\mathbf{0})) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) | \mathcal{M}_1, \mathcal{M}_2) \end{aligned} \quad (54)$$

$$\leq \mathbf{P} \left(\begin{array}{c} (U_{j^c}^n(m, l), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y) \text{ for some } l \in \mathbb{F}_q^{n\hat{R}_{j^c}}, \\ (U_1^n(\mathbf{0}, \mathbf{0}), U_2^n(\mathbf{0}, \mathbf{0})) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) \end{array} \middle| \begin{array}{c} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) \quad (55)$$

$$\stackrel{(a)}{\leq} \mathbf{P} \left(\begin{array}{c} (U_{j^c}^n(m, l), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y) \text{ for some } l \in \mathbb{F}_q^{n\hat{R}_{j^c}}, \\ (U_1^n(\mathbf{0}, \mathbf{0}), U_2^n(\mathbf{0}, \mathbf{0})) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) \end{array} \middle| \begin{array}{c} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) \quad (56)$$

$$\leq \sum_l \mathbf{P}((U_{j^c}^n(m, l), W_{\mathbf{a}}^n, Y^n) \in \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y), (U_1^n(\mathbf{0}, \mathbf{0}), U_2^n(\mathbf{0}, \mathbf{0})) \in \mathcal{T}_{\epsilon'}^{(n)}(X_1, X_2) | \mathcal{M}_1, \mathcal{M}_2) \quad (57)$$

$$\leq \sum_l \sum_{\substack{(x_1^n, x_2^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_1, X_2)}} \sum_{\substack{(u^n, w^n, y^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y)}} \mathbf{P} \left(\begin{array}{c} U_{j^c}^n(m, l) = u^n, W_{\mathbf{a}}^n = w^n, Y^n = y^n, \\ U_1^n(\mathbf{0}, \mathbf{0}) = x_1^n, U_2^n(\mathbf{0}, \mathbf{0}) = x_2^n \end{array} \middle| \begin{array}{c} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) \quad (58)$$

$$= \sum_l \sum_{\substack{(x_1^n, x_2^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_1, X_2)}} \sum_{\substack{(u^n, w^n, y^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y)}} \mathbf{P} \left(\begin{array}{c} U_{j^c}^n(m, l) = u^n, a_1 D_1^n \oplus a_2 D_2^n = w^n, \\ Y^n = y^n, D_1^n = x_1^n, D_2^n = x_2^n \end{array} \middle| \begin{array}{c} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) \quad (59)$$

$$\stackrel{(b)}{=} \sum_l \sum_{\substack{(x_1^n, x_2^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_1, X_2)}} \sum_{\substack{(u^n, w^n, y^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y)}} \mathbf{P} \left(\begin{array}{c} U_{j^c}^n(m, l) = u^n, a_1 D_1^n \oplus a_2 D_2^n = w^n, \\ D_1^n = x_1^n, D_2^n = x_2^n \end{array} \middle| \begin{array}{c} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) p(y^n | x_1^n, x_2^n) \quad (60)$$

$$\stackrel{(c)}{\leq} q^{n(\hat{R}_1 + \hat{R}_2)} \sum_l \sum_{\substack{(x_1^n, x_2^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_1, X_2)}} \sum_{\substack{(u^n, w^n, y^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y)}} \mathbf{P} \left(\begin{array}{c} U_{j^c}^n(m, l) = u^n, a_1 D_1^n \oplus a_2 D_2^n = w^n, \\ D_1^n = x_1^n, D_2^n = x_2^n \end{array} \right) p(y^n | x_1^n, x_2^n) \quad (61)$$

$$= q^{n(\hat{R}_1 + \hat{R}_2)} \sum_l \sum_{\substack{(x_1^n, x_2^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_1, X_2)}} \sum_{\substack{(u^n, w^n, y^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_{j^c}, W_{\mathbf{a}}, Y)}} \mathbf{P} \left(\begin{array}{c} [m \ l] G \oplus D_{j^c}^n = u^n, \\ D_1^n = x_1^n, D_2^n = x_2^n \end{array} \right) p(y^n | x_1^n, x_2^n) \mathbb{1}_{\{w^n = a_1 x_1^n \oplus a_2 x_2^n\}} \quad (62)$$

$$= q^{n(\hat{R}_1 + \hat{R}_2)} \sum_l \sum_{\substack{(x_1^n, x_2^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(X_1, X_2)}} \sum_{\substack{(w^n, y^n) \in \\ \mathcal{T}_{\epsilon''}^{(n)}(W_{\mathbf{a}}, Y)}} \sum_{u^n \in} q^{-3n} p(y^n | x_1^n, x_2^n) \mathbb{1}_{\{w^n = a_1 x_1^n \oplus a_2 x_2^n\}} \quad (63)$$

$$\leq q^{n(\hat{R}_1 + \hat{R}_2 + \hat{R}_{j^c})} q^{-3n} q^{n(H(X_{j^c} | W_{\mathbf{a}}, Y) + H(X_1, X_2) + \delta(\epsilon''))} \quad (64)$$

$$\stackrel{(d)}{\leq} q^{-n(I(X_{j^c}; W_{\mathbf{a}}, Y) - \delta(\epsilon'') - 3\epsilon)} \quad (65)$$

$$\leq q^{-n(I(X_{j^c}; W_{\mathbf{a}}, Y) - \delta_5(\epsilon''))}, \quad (66)$$

tends to one as $n \rightarrow \infty$, by the conditional typicality lemma in [28, p. 27], $\mathbf{P}(F_n = 1)$ tends to one as $n \rightarrow \infty$. Then, for n sufficiently large, we have

$$\begin{aligned} & H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n) \\ &= H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n, E_n, F_n) \\ &\quad + I(M_{j^c}; E_n, F_n | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n) \\ &\leq H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n, E_n, F_n) + 2 \log_q 2 \\ &\leq 2 \log_q 2 + \mathbf{P}(F_n = 0) H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n, F_n = 0, E_n) \\ &\quad + H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n, F_n = 1, E_n) \\ &\leq 2 \log_q 2 + nR_{j^c} \mathbf{P}(F_n = 0) \\ &\quad + H(M_{j^c} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n, F_n = 1, E_n). \end{aligned} \quad (68)$$

For the last term in (68), we use the fact that if $M_{j^c} \in \mathcal{L}$, then the conditional entropy cannot exceed $\log(|\mathcal{L}|)$:

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

$$\begin{aligned} & \leq H(M_{j^c} | F_n = 1, E_n, \mathcal{L}, |\mathcal{L}|) \\ &= \sum_{l=0}^{q^{nR_{j^c}}} \mathbf{P}(|\mathcal{L}| = l, E_n = 1) \\ &\quad \times H(M_{j^c} | E_n = 1, F_n = 1, \mathcal{L}, |\mathcal{L}| = l) \\ &\quad + \sum_{l=0}^{q^{nR_{j^c}}} \mathbf{P}(|\mathcal{L}| = l, E_n = 0) \\ &\quad \times H(M_{j^c} | E_n = 0, F_n = 1, \mathcal{L}, |\mathcal{L}| = l) \\ &\leq \sum_{l=0}^{q^{nR_{j^c}}} \mathbf{P}(|\mathcal{L}| = l, E_n = 1) \\ &\quad \times H(M_{j^c} | E_n = 1, F_n = 1, \mathcal{L}, |\mathcal{L}| = l) \\ &\quad + \mathbf{P}(E_n = 0) nR_{j^c} \\ &\leq \sum_{l=0}^{q^{nR_{j^c}}} \mathbf{P}(|\mathcal{L}| = l, E_n = 1) \log_q(l) + nR_{j^c} \mathbf{P}(E_n = 0) \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{l=0}^{q^{nR_{jc}}} \mathbf{P}(|\mathcal{L}| = l | E_n = 1) \log_q(l) + nR_{jc} \mathbf{P}(E_n = 0) \\
&= \mathbf{E}[\log_q(|\mathcal{L}|) | E_n = 1] + nR_{jc} \mathbf{P}(E_n = 0) \\
&\stackrel{(b)}{\leq} \log_q(\mathbf{E}[|\mathcal{L}| | E_n = 1]) + nR_{jc} \mathbf{P}(E_n = 0) \\
&\stackrel{(c)}{\leq} \log_q 2 + \max\{0, n(R_{jc} - I(X_{jc}; W_{\mathbf{a}}, Y) + \delta_5(\epsilon'') + \epsilon_n)\} \\
&\quad + nR_{jc} \mathbf{P}(E_n = 0) \\
&\leq \log_q 2 + \max\{0, n(R_{jc} - I(X_{jc}; W_{\mathbf{a}}, Y))\} \\
&\quad + n\delta_5(\epsilon'') + n\epsilon_n + nR_{jc} \mathbf{P}(E_n = 0),
\end{aligned}$$

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 by Jensen's inequality, and (c) follows by (67) and the soft-max interpretation of the log-sum-exp function [33, p. 72]. Substituting back gives

$$\begin{aligned}
&I(M_{jc}; W_{\mathbf{a}}^n, Y^n | \mathcal{C}_n) \\
&= H(M_{jc} | \mathcal{C}_n) - H(M_{jc} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n) \\
&= nR_{jc} - H(M_{jc} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n) \\
&\geq nR_{jc} - 2\log_q 2 - nR_{jc} \mathbf{P}(F_n = 0) \\
&\quad - H(M_{jc} | \mathcal{C}_n, W_{\mathbf{a}}^n, Y^n, F_n = 1, E_n) \\
&\geq nR_{jc} - 3\log_q 2 - nR_{jc}(\mathbf{P}(E_n = 0) + \mathbf{P}(F_n = 0)) \\
&\quad - \max\{0, n(R_{jc} - I(X_{jc}; W_{\mathbf{a}}, Y))\} - n\delta_5(\epsilon'') - n\epsilon_n \\
&= n[\min\{R_{jc}, I(X_{jc}; W_{\mathbf{a}}, Y)\} - \delta_5(\epsilon'') - \epsilon_n] \\
&\quad - 3 - nR_{jc}(\mathbf{P}(E = 0) + \mathbf{P}(F = 0)) \\
&\stackrel{(d)}{=} n[\min\{R_{jc}, I(X_{jc}; W_{\mathbf{a}}, Y)\} - \delta_5(\epsilon'') - 2\epsilon_n],
\end{aligned}$$

where (d) follows for large n since both probabilities $\mathbf{P}(E_n = 0)$ and $\mathbf{P}(F_n = 0)$ tend to zero as $n \rightarrow \infty$.

APPENDIX G

PROOF OF ACHIEVABILITY FOR THEOREM 4

Let parameters $\alpha \in [0, 1]$ and $\epsilon > 0$. Given a pmf $p(u_1, u_2)$ and a function $x(u_1, u_2)$, consider an $(n, nR_1, nR_2; p, x, \alpha, \epsilon)$ Marton random code ensemble. We use the nonunique simultaneous joint typicality decoding rule in [34] to establish the achievability. Let $\epsilon' > \epsilon$. Upon receiving the sequence y_j^n , the ϵ' -joint typicality decoder $j = 1, 2$ looks for a unique $m_j \in [2^{nR_j}]$ such that

$$(u_1^n(m_1, l_1), u_2^n(m_2, l_2), y_j^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_j),$$

for some $l_1 \in [2^{n\hat{R}_1}]$, $l_2 \in [2^{n\hat{R}_2}]$, and $m_{j^c} \in [2^{nR_{j^c}}]$, where j^c denotes $\{1, 2\} \setminus j$. If decoder $j = 1, 2$ finds such m_j , then it declares m_j as an estimate; otherwise, it declares an error.

We analyze the probability of error. It suffices to consider decoder 1, which declares an error only if one or more of the following events occur

$$\begin{aligned}
\mathcal{E}_0 &= \{(U_1^n(M_1, l_1), U_2^n(M_2, l_2)) \notin \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2) \\
&\quad \text{for every } (l_1, l_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]\}, \\
\mathcal{E}_1 &= \{(U_1^n(M_1, L_1), U_2^n(M_2, L_2), Y_1^n) \notin \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1)\}, \\
\mathcal{E}_2 &= \{(U_1^n(m_1, l_1), U_2^n(m_2, l_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) \\
&\quad \text{for some } m_1 \neq M_1 \in [2^{nR_1}] \text{ and} \\
&\quad \text{for some } (m_2, l_1, l_2) \in [2^{nR_2}] \times [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]\}.
\end{aligned}$$

By the union of events bound, $\mathbf{P}_e^{(n)}(\mathcal{C}_n) \leq \mathbf{P}(\mathcal{E}_0) + \mathbf{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c) + \mathbf{P}(\mathcal{E}_2 \cap \mathcal{E}_0^c)$. Since $\hat{R}_1 + \hat{R}_2 = I(U_1; U_2) + 10\epsilon H(U_1, U_2)$, by the mutual covering lemma in [28, p. 208], the probability $\mathbf{P}(\mathcal{E}_0)$ tends to zero as $n \rightarrow \infty$. By the conditional typicality lemma in [28, p. 27], the probability $\mathbf{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c)$ tends to zero as $n \rightarrow \infty$. The last term can be bounded by two ways. First, by the symmetric codebook generation,

$$\begin{aligned}
\mathbf{P}(\mathcal{E}_2 \cap \mathcal{E}_0^c) &\leq \mathbf{P}(\mathcal{E}_2) \\
&= \mathbf{P}(\mathcal{E}_2 | M_1 = M_2 = 1) \\
&\leq \mathbf{P}((U_1^n(m_1, l_1), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, Y_1) \\
&\quad \text{for some } m_1 \neq 1, \text{ for some } l_1 \in [2^{n\hat{R}_1}] | M_1 = 1),
\end{aligned}$$

which tends to zero as $n \rightarrow \infty$ if $R_1 + \hat{R}_1 \leq I(U_1; Y_1) - \delta(\epsilon')$ by the packing lemma in [28]. Letting $\hat{R}_1 = \alpha(I(U_1; U_2) + 10\epsilon H(U_1, U_2))$, we have

$$R_1 \leq \max\{0, I(U_1; Y_1) - \alpha I(U_1; U_2) - 2\delta(\epsilon')\}. \quad (69)$$

Secondly, we can decompose the event $\mathcal{E}_2 = \mathcal{E}_{21} \cup \mathcal{E}_{22}$ such that

$$\begin{aligned}
\mathcal{E}_{21} &= \{(U_1^n(m_1, l_1), U_2^n(M_2, l_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) \\
&\quad \text{for some } m_1 \neq M_1, \text{ for some } (l_1, l_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]\}, \\
\mathcal{E}_{22} &= \{(U_1^n(m_1, l_1), U_2^n(m_2, l_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) \\
&\quad \text{for some } m_1 \neq M_1, \text{ for some } m_2 \neq M_2, \\
&\quad \text{for some } (l_1, l_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]\}.
\end{aligned}$$

The probability $\mathbf{P}(\mathcal{E}_{22})$ is bounded in (70)-(77), shown at the top of the next page, where (74) follows since given $\{M_1 = M_2 = 1\}$, the pair $(U_1^n(m_1, l_1), U_2^n(m_2, l_2))$ for $m_1 \neq 1, m_2 \neq 1$ is i.i.d. with respect to the product pmf $p(u_1)p(u_2)$ and is independent from Y_1^n . Substituting $\hat{R}_1 + \hat{R}_2 = I(U_1; U_2) + 10\epsilon H(U_1, U_2)$ into (77), it follows that $\mathbf{P}(\mathcal{E}_{22})$ tends to zero as $n \rightarrow \infty$ if $R_1 + R_2 \leq I(U_1, U_2; Y_1) - 3\delta(\epsilon')$.

We next bound the probability $\mathbf{P}(\mathcal{E}_{21} \cap \mathcal{E}_0^c)$. Define the events $\mathcal{M}_1 := \{M_1 = M_2 = 1\}$ and $\mathcal{M}_2 := \{L_1 = L_2 = 1\}$. By the symmetry of the codebook generation,

$$\mathbf{P}(\mathcal{E}_{21} \cap \mathcal{E}_0^c) = \mathbf{P}(\mathcal{E}_{21} \cap \mathcal{E}_0^c | \mathcal{M}_1, \mathcal{M}_2). \quad (78)$$

To see this, define the tuple of auxiliary codewords for sender $j = 1, 2$ as $\tilde{\mathcal{C}}_n(j) := (U_j^n(m_j, l_j) : m_j \in [2^{nR_j}], l_j \in [2^{n\hat{R}_j}])$. We first show that (M_1, M_2, L_1, L_2) is uniformly distributed over its support. It suffices to show that for every $(m_1, m_2, l_1, l_2) \in [2^{nR_1}] \times [2^{nR_2}] \times [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]$,

$$\begin{aligned}
&\mathbf{P}(M_1 = m_1, M_2 = m_2, L_1 = l_1, L_2 = l_2) \\
&= \mathbf{P}(M_1 = 1, M_2 = 1, L_1 = 1, L_2 = 1).
\end{aligned}$$

Fix a tuple $(m_1, m_2, l_1, l_2) \in [2^{nR_1}] \times [2^{nR_2}] \times [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]$. Given $\tilde{\mathcal{C}}_n(j) = \mathcal{C}_j$, let $\sigma_j(c_j)$ denote the permuted version of \mathcal{C}_j such that

$$\begin{aligned}
&\{u_j^n(m_j, l'_j) \in \mathcal{C}_j : l'_j \in [2^{n\hat{R}_j}]\} \\
&= \{\tilde{u}_j^n(1, l'_j) \in \sigma_j(\mathcal{C}_j) : l'_j \in [2^{n\hat{R}_j}]\}
\end{aligned}$$

and $u_j^n(m_j, l_j) \in \mathcal{C}_j$ and $\tilde{u}_j^n(1, 1) \in \sigma_j(\mathcal{C}_j)$ satisfy

$$u_j^n(m_j, l_j) = \tilde{u}_j^n(1, 1).$$

$$\mathbf{P}(\mathcal{E}_{22}) = \mathbf{P}(\mathcal{E}_{22} | M_1 = M_2 = 1) \quad (70)$$

$$= \mathbf{P}((U_1^n(m_1, l_1), U_2^n(m_2, l_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) \text{ for some } m_1 \neq 1, \text{ for some } m_2 \neq 1, \\ \text{for some } (l_1, l_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}] | M_1 = M_2 = 1) \quad (71)$$

$$\leq \sum_{m_1 \neq 1} \sum_{l_1} \sum_{m_2 \neq 1} \sum_{l_2} \mathbf{P}((U_1^n(m_1, l_1), U_2^n(m_2, l_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) | M_1 = M_2 = 1) \quad (72)$$

$$\leq \sum_{m_1 \neq 1} \sum_{l_1} \sum_{m_2 \neq 1} \sum_{l_2} \sum_{\substack{(u_1^n, u_2^n, y_1^n) \in \\ \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1)}} \mathbf{P}(U_1^n(m_1, l_1) = u_1^n, U_2^n(m_2, l_2) = u_2^n, Y_1^n = y_1^n | M_1 = M_2 = 1) \quad (73)$$

$$= \sum_{m_1 \neq 1} \sum_{l_1} \sum_{m_2 \neq 1} \sum_{l_2} \sum_{\substack{(u_1^n, u_2^n, y_1^n) \in \\ \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1)}} p(y_1^n | M_1 = M_2 = 1) \prod_{i=1}^n p_{U_1}(u_{1i}) p_{U_2}(u_{2i}) \quad (74)$$

$$\leq \sum_{m_1 \neq 1} \sum_{l_1} \sum_{m_2 \neq 1} \sum_{l_2} \sum_{\substack{(u_1^n, u_2^n, y_1^n) \in \\ \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1)}} p(y_1^n | M_1 = M_2 = 1) 2^{-n(H(U_1) + H(U_2) - \delta(\epsilon'))} \quad (75)$$

$$\leq \sum_{m_1 \neq 1} \sum_{l_1} \sum_{m_2 \neq 1} \sum_{l_2} 2^{-n(H(U_1) + H(U_2) - H(U_1, U_2 | Y_1) - 2\delta(\epsilon'))} \quad (76)$$

$$\leq 2^{n(R_1 + R_2 + \hat{R}_1 + \hat{R}_2)} 2^{-n(H(U_1) + H(U_2) - H(U_1, U_2 | Y_1) - 2\delta(\epsilon'))}, \quad (77)$$

Then, we have

$$\begin{aligned} & \mathbf{P}(M_1 = m_1, M_2 = m_2, L_1 = l_1, L_2 = l_2) \\ &= \sum_{\mathcal{C}_1, \mathcal{C}_2} \mathbf{P} \left(\begin{array}{c} M_1 = m_1, M_2 = m_2, L_1 = l_1, L_2 = l_2, \\ \tilde{\mathcal{C}}_n(1) = c_1, \tilde{\mathcal{C}}_n(2) = c_2 \end{array} \right) \\ &\stackrel{(a)}{=} \sum_{\mathcal{C}_1, \mathcal{C}_2} \mathbf{P}(M_1 = m_1, M_2 = m_2) \mathbf{P}(\tilde{\mathcal{C}}_n(1) = c_1) \\ & \quad \mathbf{P}(\tilde{\mathcal{C}}_n(2) = c_2) \mathbf{P} \left(\begin{array}{c} L_1 = l_1, \\ L_2 = l_2 \end{array} \middle| \begin{array}{c} M_1 = m_1, M_2 = m_2, \\ \tilde{\mathcal{C}}_n(1) = c_1, \tilde{\mathcal{C}}_n(2) = c_2 \end{array} \right) \\ &\stackrel{(b)}{=} \sum_{\mathcal{C}_1, \mathcal{C}_2} \mathbf{P}(M_1 = 1, M_2 = 1) \mathbf{P}(\tilde{\mathcal{C}}_n(1) = \sigma_1(c_1)) \\ & \quad \mathbf{P}(\tilde{\mathcal{C}}_n(2) = \sigma_2(c_2)) \mathbf{P} \left(\begin{array}{c} L_1 = 1, \\ L_2 = 1 \end{array} \middle| \begin{array}{c} M_1 = 1, M_2 = 1, \\ \tilde{\mathcal{C}}_n(1) = \sigma_1(c_1), \\ \tilde{\mathcal{C}}_n(2) = \sigma_2(c_2) \end{array} \right) \\ &= \sum_{\mathcal{C}_1, \mathcal{C}_2} \mathbf{P} \left(\begin{array}{c} M_1 = 1, M_2 = 1, L_1 = 1, L_2 = 1, \\ \tilde{\mathcal{C}}_n(1) = \sigma_1(c_1), \tilde{\mathcal{C}}_n(2) = \sigma_2(c_2) \end{array} \right) \\ &= \mathbf{P}(M_1 = 1, M_2 = 1, L_1 = 1, L_2 = 1), \end{aligned} \quad (87)$$

where (a) follows since $(M_1, M_2, \tilde{\mathcal{C}}_n(1), \tilde{\mathcal{C}}_n(2))$ are independent, (b) follows since (M_1, M_2) is uniformly distributed and $\tilde{\mathcal{C}}_n(j) \stackrel{d}{=} \sigma_j(\tilde{\mathcal{C}}_n(j))$, $j = 1, 2$.

Following similar arguments, we can now prove the claim in (78), the proof of which is given in (79)-(86), shown at the top of the next page, where (86) follows since (M_1, M_2, L_1, L_2) is uniformly distributed.

To bound $\mathbf{P}(\mathcal{E}_{21} \cap \mathcal{E}_0^c)$, we continue from (78) as follows.

$$\begin{aligned} & \mathbf{P}(\mathcal{E}_{21} \cap \mathcal{E}_0^c | \mathcal{M}_1, \mathcal{M}_2) \\ &\leq \sum_{m_1 \neq 1} \sum_{l_1, l_2} \mathbf{P} \left(\begin{array}{c} (U_1^n(m_1, l_1), U_2^n(1, l_2), Y_1^n) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y), \\ (U_1^n(1, 1), U_2^n(1, 1)) \\ \in \mathcal{T}_{\epsilon}^{(n)}(U_1, U_2) \end{array} \middle| \begin{array}{c} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) \\ & \quad \sum_{\substack{(\tilde{u}_1^n, y_1^n) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, Y_1 | u_2^n)}} \mathbf{P} \left(\begin{array}{c} U_1^n(m_1, l_1) = \tilde{u}_1^n, \\ U_1^n(1, 1) = u_1^n, \\ U_2^n(1, 1) = u_2^n, \\ Y_1^n = y_1^n \end{array} \middle| \begin{array}{c} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) \end{aligned}$$

where (a) follows since $\epsilon' > \epsilon$. The first summation term in (87) can be bounded as

$$\begin{aligned} & \sum_{m_1 \neq 1} \sum_{l_1} \mathbf{P} \left(\begin{array}{c} (U_1^n(m_1, l_1), U_2^n(1, 1), Y_1^n) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y), \\ (U_1^n(1, 1), U_2^n(1, 1)) \\ \in \mathcal{T}_{\epsilon}^{(n)}(U_1, U_2) \end{array} \middle| \begin{array}{c} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) \\ &\leq \sum_{m_1 \neq 1} \sum_{l_1} \sum_{\substack{(u_1^n, u_2^n) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2)}} \end{aligned}$$

$$\mathbf{P}(\mathcal{E}_{21} \cap \mathcal{E}_0^c) = \mathbf{P}((U_1^n(m'_1, l'_1), U_2^n(M_2, l'_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) \text{ for some } m'_1 \neq M_1, \text{ for some } (l'_1, l'_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}], (U_1^n(M_1, L_1), U_2^n(M_2, L_2)) \in \mathcal{T}_{\epsilon}^{(n)}(U_1, U_2)) \quad (79)$$

$$= \sum_{\substack{m_1, m_2 \\ l_1, l_2}} \sum_{\mathcal{C}_1, \mathcal{C}_2} \mathbf{P} \left(\begin{array}{l} (U_1^n(m'_1, l'_1), U_2^n(M_2, l'_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) \text{ for some } m'_1 \neq M_1, \\ \text{for some } (l'_1, l'_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}], (U_1^n(M_1, L_1), U_2^n(M_2, L_2)) \in \mathcal{T}_{\epsilon}^{(n)}(U_1, U_2), \\ (M_1, M_2, L_1, L_2) = (m_1, m_2, l_1, l_2), (\tilde{\mathcal{C}}_n(1), \tilde{\mathcal{C}}_n(2)) = (\mathcal{C}_1, \mathcal{C}_2) \end{array} \right) \quad (80)$$

$$= \sum_{\substack{m_1, m_2 \\ l_1, l_2}} \sum_{\mathcal{C}_1, \mathcal{C}_2} \mathbf{P}(M_1 = m_1, M_2 = m_2) \mathbf{P}((\tilde{\mathcal{C}}_n(1), \tilde{\mathcal{C}}_n(2)) = (\mathcal{C}_1, \mathcal{C}_2)) \\ \mathbf{P} \left(\begin{array}{l} (U_1^n(m'_1, l'_1), U_2^n(M_2, l'_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) \\ \text{for some } m'_1 \neq M_1, \text{ for some } (l'_1, l'_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}], \\ (U_1^n(M_1, L_1), U_2^n(M_2, L_2)) \in \mathcal{T}_{\epsilon}^{(n)}(U_1, U_2), \\ (L_1, L_2) = (l_1, l_2) \end{array} \middle| \begin{array}{l} M_1 = m_1 \\ M_2 = m_2 \\ \tilde{\mathcal{C}}_n(1) = \mathcal{C}_1 \\ \tilde{\mathcal{C}}_n(2) = \mathcal{C}_2 \end{array} \right) \quad (81)$$

$$= \sum_{\substack{m_1, m_2 \\ l_1, l_2}} \sum_{\mathcal{C}_1, \mathcal{C}_2} \mathbf{P}(M_1 = 1, M_2 = 1) \mathbf{P}((\tilde{\mathcal{C}}_n(1), \tilde{\mathcal{C}}_n(2)) = (\sigma_1(\mathcal{C}_1), \sigma_2(\mathcal{C}_2))) \\ \mathbf{P} \left(\begin{array}{l} (U_1^n(m'_1, l'_1), U_2^n(M_2, l'_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) \\ \text{for some } m'_1 \neq M_1, \text{ for some } (l'_1, l'_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}], \\ (U_1^n(M_1, L_1), U_2^n(M_2, L_2)) \in \mathcal{T}_{\epsilon}^{(n)}(U_1, U_2), \\ (L_1, L_2) = (1, 1) \end{array} \middle| \begin{array}{l} M_1 = 1 \\ M_2 = 1 \\ \tilde{\mathcal{C}}_n(1) = \sigma_1(\mathcal{C}_1) \\ \tilde{\mathcal{C}}_n(2) = \sigma_2(\mathcal{C}_2) \end{array} \right) \quad (82)$$

$$= \sum_{\substack{m_1, m_2 \\ l_1, l_2}} \sum_{\mathcal{C}_1, \mathcal{C}_2} \mathbf{P} \left(\begin{array}{l} (U_1^n(m'_1, l'_1), U_2^n(M_2, l'_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) \text{ for some } m'_1 \neq M_1, \\ \text{for some } (l'_1, l'_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}], (U_1^n(M_1, L_1), U_2^n(M_2, L_2)) \in \mathcal{T}_{\epsilon}^{(n)}(U_1, U_2), \\ (M_1, M_2, L_1, L_2) = (1, 1, 1, 1), (\tilde{\mathcal{C}}_n(1), \tilde{\mathcal{C}}_n(2)) = (\sigma_1(\mathcal{C}_1), \sigma_2(\mathcal{C}_2)) \end{array} \right) \quad (83)$$

$$= \sum_{\substack{m_1, m_2 \\ l_1, l_2}} \mathbf{P}(\mathcal{E}_{21} \cap \mathcal{E}_0^c, (M_1, M_2, L_1, L_2) = (1, 1, 1, 1)) \quad (84)$$

$$= \sum_{\substack{m_1, m_2 \\ l_1, l_2}} \mathbf{P}((M_1, M_2, L_1, L_2) = (1, 1, 1, 1)) \mathbf{P}(\mathcal{E}_{21} \cap \mathcal{E}_0^c | \mathcal{M}_1, \mathcal{M}_2) \quad (85)$$

$$= \mathbf{P}(\mathcal{E}_{21} \cap \mathcal{E}_0^c | \mathcal{M}_1, \mathcal{M}_2) \quad (86)$$

$$\stackrel{(a)}{=} \sum_{m_1 \neq 1} \sum_{l_1} \sum_{\substack{(u_1^n, u_2^n) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2)}} \mathbf{P} \left(\begin{array}{l} U_1^n(m_1, l_1) = \tilde{u}_1^n, \\ U_1^n(1, 1) = u_1^n, \\ U_2^n(1, 1) = u_2^n \end{array} \middle| \begin{array}{l} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) p(y_1^n | u_1^n, u_2^n) \\ \leq 2^{n(\hat{R}_1 + \hat{R}_2)} \sum_{m_1 \neq 1} \sum_{l_1} \sum_{\substack{(u_1^n, u_2^n) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2)}} \mathbf{P} \left(\begin{array}{l} U_1^n(m_1, l_1) = \tilde{u}_1^n, \\ U_1^n(1, 1) = u_1^n, \\ U_2^n(1, 1) = u_2^n \end{array} \right) p(y_1^n | u_1^n, u_2^n) \\ \leq 2^{n(R_1 + 2\hat{R}_1 + \hat{R}_2 + H(U_1, U_2) + H(U_1 | Y_1, U_2) - 2H(U_1) - H(U_2) + 3\delta(\epsilon'))} \\ = 2^{n(R_1 + 2\hat{R}_1 + \hat{R}_2 - I(U_1; U_2) - I(U_1; Y_1, U_2) + 3\delta(\epsilon'))}, \quad (87)$$

$$\stackrel{(b)}{\leq} 2^{n(\hat{R}_1 + \hat{R}_2)} \sum_{m_1 \neq 1} \sum_{l_1} \sum_{\substack{(u_1^n, u_2^n) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2)}} \mathbf{P} \left(\begin{array}{l} U_1^n(m_1, l_1) = \tilde{u}_1^n, \\ U_1^n(1, 1) = u_1^n, \\ U_2^n(1, 1) = u_2^n \end{array} \right) p(y_1^n | u_1^n, u_2^n) \\ \stackrel{(c)}{\leq} 2^{n(\hat{R}_1 + \hat{R}_2)} \sum_{m_1 \neq 1} \sum_{l_1} \sum_{\substack{(u_1^n, u_2^n) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2)}} p(y_1^n | u_1^n, u_2^n) 2^{-n(2H(U_1) + H(U_2) - \delta(\epsilon'))} \\ \leq 2^{n(\hat{R}_1 + \hat{R}_2)} \sum_{m_1 \neq 1} \sum_{l_1} \sum_{\substack{(u_1^n, u_2^n) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2)}} p(y_1^n | u_1^n, u_2^n) 2^{-n(2H(U_1) + H(U_2) - \delta(\epsilon'))} \\ \leq 2^{n(H(U_1 | Y_1, U_2) + \delta(\epsilon'))} 2^{-n(2H(U_1) + H(U_2) - \delta(\epsilon'))} \\ \leq 2^{n(R_1 + 2\hat{R}_1 + \hat{R}_2 - 2I(U_1; U_2) - I(U_1, U_2; Y_1) + 3\delta(\epsilon'))}.$$

where (a) follows since given $(\mathcal{M}_1, \mathcal{M}_2)$ the tuple $U_1^n(m_1, l_1) \rightarrow (U_1^n(1, 1), U_2^n(1, 1)) \rightarrow Y_1^n$ form a Markov chain, (b) follows by [22, Lemma 11] since the tuple $(U_1^n(m_1, l_1), U_1^n(1, 1), U_2^n(1, 1))$ is independent of the event \mathcal{M}_1 and (M_1, M_2, L_1, L_2) is uniformly distributed, and (c) follows since the tuple $(U_1^n(m_1, l_1), U_1^n(1, 1), U_2^n(1, 1))$ is i.i.d. with respect to the product pmf $p(u_1)p(u_2)$.

Similarly, the second summation term in (87) can be bounded as

$$\sum_{m_1 \neq 1} \sum_{l_1} \sum_{l_2 \neq 1} \mathbf{P} \left(\begin{array}{l} (U_1^n(m_1, l_1), U_2^n(1, l_2), Y_1^n) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y), \\ (U_1^n(1, 1), U_2^n(1, 1)) \\ \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2) \end{array} \middle| \begin{array}{l} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) \\ \leq 2^{n(R_1 + 2\hat{R}_1 + 2\hat{R}_2 - 2I(U_1; U_2) - I(U_1, U_2; Y_1) + 3\delta(\epsilon'))}.$$

Therefore, $\mathbf{P}(\mathcal{E}_{21} \cap \mathcal{E}_0^c)$ tends to zero as $n \rightarrow \infty$ if $R_1 + 2\hat{R}_1 + \hat{R}_2 \leq I(U_1; U_2) + I(U_1; Y_1, U_2) - 3\delta(\epsilon')$ and

$R_1 + 2\hat{R}_1 + 2\hat{R}_2 \leq 2I(U_1; U_2) + I(U_1, U_2; Y_1) - 3\delta(\epsilon')$. Letting $\hat{R}_1 = \alpha(I(U_1; U_2) + 10\epsilon H(U_1, U_2))$ and $\hat{R}_2 = \bar{\alpha}(I(U_1; U_2) + 10\epsilon H(U_1, U_2))$ results in $R_1 \leq I(U_1; Y_1, U_2) - \alpha I(U_1; U_2) - 4\delta(\epsilon')$ and $R_1 \leq I(U_1, U_2; Y_1) - 4\delta(\epsilon')$.

Combining with (69), the probability of error at decoder 1 tends to zero as $n \rightarrow \infty$ if

$$R_1 \leq \max\{0, I(U_1; Y_1) - \alpha I(U_1; U_2) - 4\delta(\epsilon')\}, \quad (88)$$

or

$$R_1 \leq I(U_1; Y_1, U_2) - \alpha I(U_1; U_2) - 4\delta(\epsilon'), \quad (89a)$$

$$R_1 + R_2 \leq I(U_1, U_2; Y_1) - 4\delta(\epsilon'). \quad (89b)$$

Repeating similar steps, the probability of error at decoder 2 tends to zero as $n \rightarrow \infty$ if

$$R_2 \leq \max\{0, I(U_2; Y_2) - \bar{\alpha} I(U_1; U_2) - 4\delta(\epsilon')\}, \quad (90)$$

or

$$R_2 \leq I(U_2; Y_2, U_1) - \bar{\alpha} I(U_1; U_2) - 4\delta(\epsilon'), \quad (91a)$$

$$R_1 + R_2 \leq I(U_1, U_2; Y_2) - 4\delta(\epsilon'). \quad (91b)$$

If we denote the set of rate pairs satisfying (88) or (89) as $\mathcal{R}_{\text{BC},1}(p, x, \alpha, \delta(\epsilon'))$, and denote the set of rate pairs satisfying (90) or (91) as $\mathcal{R}_{\text{BC},2}(p, x, \alpha, \delta(\epsilon'))$, then the rate region $\mathcal{R}_{\text{BC},1}(p, x, \alpha, \delta(\epsilon')) \cap \mathcal{R}_{\text{BC},2}(p, x, \alpha, \delta(\epsilon'))$ is achievable by the ϵ' -typicality decoders. Define the rate regions $\mathcal{R}_{\text{BC},j}(p, x, \alpha) := \mathcal{R}_{\text{BC},j}(p, x, \alpha, \delta(\epsilon') = 0)$, $j = 1, 2$. Let $\epsilon' = 2\epsilon$. Taking $\epsilon \rightarrow 0$ and then taking the closure implies

$$\mathcal{R}_{\text{BC},1}(p, x, \alpha) \cap \mathcal{R}_{\text{BC},2}(p, x, \alpha) \subseteq \mathcal{R}_{\text{BC}}^*(p, x, \alpha).$$

The achievability proof follows from the next lemma that provides an equivalent characterization for the rate region in Theorem 4.

Lemma 10: For any input pmf $p = p(u_1, u_2)$, function $x = x(u_1, u_2)$, and $\alpha \in [0, 1]$,

$$\mathcal{R}_{\text{BC}}^{**}(p, x, \alpha) = \mathcal{R}_{\text{BC},1}(p, x, \alpha) \cap \mathcal{R}_{\text{BC},2}(p, x, \alpha).$$

Proof: Fix pmf $p = p(u_1, u_2)$, function $x = x(u_1, u_2)$ and $\alpha \in [0, 1]$. It suffices to show that the rate region $\mathcal{R}_{\text{BC},1}(p, x, \alpha)$ is equivalent to the set of rate pairs (R_1, R_2) that satisfy (26a)-(26b). We first show that any rate pair in $\mathcal{R}_{\text{BC},1}(p, x, \alpha)$ satisfies (26a)-(26b). Suppose that the rate pair $(R_1, R_2) \in \mathcal{R}_{\text{BC},1}(p, x, \alpha)$, which implies that

$$R_1 \leq I(U_1; Y_1, U_2) - \alpha I(U_1; U_2),$$

and

$$\begin{aligned} R_1 &\leq \max\{0, I(U_1; Y_1) - \alpha I(U_1; U_2), I(U_1, U_2; Y_1) - R_2\} \\ &= I(U_1, U_2; Y_1) \\ &\quad - \min\{I(U_1, U_2; Y_1), I(U_2; Y_1, U_1) - \bar{\alpha} I(U_1; U_2), R_2\}. \end{aligned}$$

Therefore, (R_1, R_2) satisfies (26a)-(26b).

We now prove the other direction. Suppose that the rate pair (R_1, R_2) satisfies (26a)-(26b). Assume also that

$R_2 < \min\{I(U_2; Y_1, U_1) - \bar{\alpha} I(U_1; U_2), I(U_1, U_2; Y_1)\}$. It then follows that

$$R_1 \leq I(U_1; Y_1, U_2) - \alpha I(U_1; U_2),$$

$$R_1 \leq I(U_1, U_2; Y_1) - R_2.$$

So, the rate pair $(R_1, R_2) \in \mathcal{R}_{\text{BC},1}(p, x, \alpha)$. If instead $R_2 \geq \min\{I(U_2; Y_1, U_1) - \bar{\alpha} I(U_1; U_2), I(U_1, U_2; Y_1)\}$, then

$$R_1 \leq I(U_1; Y_1, U_2) - \alpha I(U_1; U_2),$$

$$R_1 \leq I(U_1, U_2; Y_1)$$

$$- \min\{I(U_2; Y_1, U_1) - \bar{\alpha} I(U_1; U_2), I(U_1, U_2; Y_1)\}$$

$$= \max\{0, I(U_1; Y_1) - \alpha I(U_1; U_2)\}.$$

Therefore, $(R_1, R_2) \in \mathcal{R}_{\text{BC},1}(p, x, \alpha)$, which completes the proof of the lemma. \blacksquare

APPENDIX H PROOF OF LEMMA 5

Let $\epsilon' > \epsilon$. First, by (the averaged version of) Fano's lemma in (30), we have

$$I(M_2; Y_1^n | \mathcal{C}_n) \geq I(M_2; M_1, Y_1^n | \mathcal{C}_n) - n\epsilon_n.$$

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

$$\begin{aligned} &I(M_2; M_1, Y_1^n | \mathcal{C}_n) \\ &\geq n \left[\min \left\{ \begin{array}{l} R_2, I(U_1, U_2; Y_1), \\ I(U_2; Y_1, U_1) - \bar{\alpha} I(U_1; U_2) \end{array} \right\} - \delta(\epsilon') - 2\epsilon_n \right], \end{aligned}$$

for some $\delta(\epsilon')$ that tends to zero as $\epsilon' \rightarrow 0$.

Similar to [26], we will show that given M_1, Y_1^n and \mathcal{C}_n , a relatively short list $\mathcal{L} \subseteq [2^{nR_2}]$ can be constructed that contains M_2 with high probability. Define a random set

$$\begin{aligned} \mathcal{L} &= \{m_2 \in [2^{nR_2}] : \\ &\quad (U_1^n(M_1, l_1), U_2^n(m_2, l_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1) \\ &\quad \text{for some } (l_1, l_2) \in [2^{n\hat{R}_1}] \times [2^{n\hat{R}_2}]\}. \end{aligned}$$

Note that the set \mathcal{L} is random with the underlying distribution on $(M_1, Y_1^n, \mathcal{C}_n)$, which is induced by drawing a Marton random codebook \mathcal{C}_n and using this codebook to encode $U_1^n(M_1, L_1)$ and $U_2^n(M_2, L_2)$ into $X^n(M_1, M_2)$ that lead to Y_1^n through the DM-BC $p(y_1, y_2 | x)$. We first bound the probability that an incorrect message is in the random set \mathcal{L} . Define the events $\mathcal{M}_1 = \{M_1 = M_2 = 1\}$ and $\mathcal{M}_2 = \{L_1 = L_2 = 1\}$. The indicator random variable \tilde{E}_n is as defined in (31). By the symmetry of the codebook generation discussed in Appendix G, for every $m_2 \neq M_2 \in [2^{nR_2}]$

$$\mathbb{P}(m_2 \in \mathcal{L}, \tilde{E}_n = 1) = \mathbb{P}(m_2 \in \mathcal{L}, \tilde{E}_n = 1 | \mathcal{M}_1, \mathcal{M}_2), \quad (92)$$

which is easy to see following similar steps to the proof of (78). We will use the conditioned version to bound the

probability term in (92). For every $m_2 \neq 1 \in [2^{nR_2}]$,

$$\begin{aligned}
& \mathbf{P}(m_2 \in \mathcal{L}, \tilde{E}_n = 1 | \mathcal{M}_1, \mathcal{M}_2) \\
& \stackrel{(a)}{=} \mathbf{P} \left(\begin{array}{l} (U_1^n(1, l_1), U_2^n(m_2, l_2), Y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)} \\ \text{for some } (l_1, l_2) \in [2^{nR_1}] \times [2^{nR_2}], \\ (U_1^n(1, 1), U_2^n(1, 1)) \in \mathcal{T}_{\epsilon}^{(n)} \end{array} \middle| \begin{array}{l} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) \\
& \stackrel{(b)}{\leq} \sum_{l_2} \sum_{(u_1^n, u_2^n) \in \mathcal{T}_{\epsilon}^{(n)}(U_1, U_2)} \\
& \quad \sum_{(\tilde{u}_2^n, y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_2, Y_1 | u_1^n)} \mathbf{P} \left(\begin{array}{l} U_1^n(1, 1) = u_1^n, \\ U_2^n(1, 1) = u_2^n, \\ U_2^n(m_2, l_2) = \tilde{u}_2^n, \\ Y_1^n = y_1^n \end{array} \middle| \begin{array}{l} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right) \\
& + \sum_{l_1 \neq 1} \sum_{l_2} \sum_{(u_1^n, u_2^n) \in \mathcal{T}_{\epsilon}^{(n)}(U_1, U_2)} \\
& \quad \sum_{(\tilde{u}_1^n, \tilde{u}_2^n, y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, U_2, Y_1)} \mathbf{P} \left(\begin{array}{l} U_1^n(1, 1) = u_1^n, \\ U_2^n(1, 1) = u_2^n, \\ U_1^n(m_1, l_1) = \tilde{u}_1^n, \\ U_2^n(m_2, l_2) = \tilde{u}_2^n, \\ Y_1^n = y_1^n \end{array} \middle| \begin{array}{l} \mathcal{M}_1, \\ \mathcal{M}_2 \end{array} \right), \quad (93)
\end{aligned}$$

where (b) follows by the union of events bound and by decomposing the event in (a) onto two sets: $\{l_1 = 1\}$ and $\{l_1 \neq 1\}$. Two summation terms on the right hand side of (93) can be bounded by using similar arguments to the proof of the inner bound for Theorem 4 (refer to the bounds on the two summation terms in (87) in Appendix G) to get

$$\begin{aligned}
\mathbf{P}(m_2 \in \mathcal{L}, \tilde{E}_n = 1) & \leq 2^{-n(I(U_2; Y_1, U_1) - \bar{\alpha}I(U_1; U_2) - 4\delta(\epsilon'))} \\
& + 2^{-n(I(U_1, U_2; Y_1) - 4\delta(\epsilon'))}.
\end{aligned}$$

Since $\mathbf{P}(\tilde{E}_n = 1)$ tends to one as $n \rightarrow \infty$, for n sufficiently large, $\mathbf{P}(m_2 \in \mathcal{L} | \tilde{E}_n = 1) \leq \mathbf{P}(m_2 \in \mathcal{L}, \tilde{E}_n = 1)2^\epsilon$. The expected cardinality of \mathcal{L} given $\{\tilde{E}_n = 1\}$ is then bounded as

$$\begin{aligned}
\mathbf{E}(|\mathcal{L}| | \tilde{E}_n = 1) & \leq 1 + \sum_{m_2 \neq M_2} \mathbf{P}(m_2 \in \mathcal{L} | \tilde{E}_n = 1) \\
& \leq 1 + 2^{n(R_2 - I(U_2; Y_1, U_1) + \bar{\alpha}I(U_1; U_2) + 4\delta(\epsilon') + \frac{\epsilon}{n})} \\
& \quad + 2^{n(R_2 - I(U_1, U_2; Y_1) + 4\delta(\epsilon') + \frac{\epsilon}{n})} \\
& = 1 + 2^{n(R_2 - I(U_2; Y_1, U_1) + \bar{\alpha}I(U_1; U_2) + 4\delta(\epsilon') + \epsilon_n)} \\
& \quad + 2^{n(R_2 - I(U_1, U_2; Y_1) + 4\delta(\epsilon') + \epsilon_n)}, \quad (94)
\end{aligned}$$

for n sufficiently large.

Define another indicator random variable $\tilde{F}_n = \mathbb{1}_{\{M_2 \in \mathcal{L}\}}$. Since $\epsilon' > \epsilon$ and $\mathbf{P}(\tilde{E}_n = 1)$ tends to one as $n \rightarrow \infty$, by the conditional typicality lemma in [28, p. 27], $\mathbf{P}(\tilde{F}_n = 1)$ tends to one as $n \rightarrow \infty$. Then, for n sufficiently large, we have

$$\begin{aligned}
& H(M_2 | \mathcal{C}_n, M_1, Y_1^n) \\
& = H(M_2 | \mathcal{C}_n, M_1, Y_1^n, \tilde{E}_n, \tilde{F}_n) + I(M_2; \tilde{E}_n, \tilde{F}_n | \mathcal{C}_n, M_1, Y_1^n) \\
& \leq H(M_2 | \mathcal{C}_n, M_1, Y_1^n, \tilde{E}_n, \tilde{F}_n) + 2 \\
& \leq 2 + \mathbf{P}(\tilde{F}_n = 0)H(M_2 | \mathcal{C}_n, M_1, Y_1^n, \tilde{E}_n, \tilde{F}_n = 0) \\
& \quad + H(M_2 | \mathcal{C}_n, M_1, Y_1^n, \tilde{E}_n, \tilde{F}_n = 1) \\
& \leq 2 + nR_2 \mathbf{P}(\tilde{F}_n = 0) + H(M_2 | \mathcal{C}_n, M_1^n, Y_1^n, \tilde{E}_n, \tilde{F}_n = 1).
\end{aligned}$$

For the last term, we use the fact that if $M_2 \in \mathcal{L}$, then the conditional entropy cannot exceed $\log(|\mathcal{L}|)$:

$$\begin{aligned}
& H(M_2 | \mathcal{C}_n, M_1, Y_1^n, \tilde{E}_n, \tilde{F}_n = 1) \\
& \stackrel{(a)}{=} H(M_2 | \mathcal{C}_n, M_1, Y_1^n, \tilde{E}_n, \tilde{F}_n = 1, \mathcal{L}, |\mathcal{L}|) \\
& \leq H(M_2 | \tilde{E}_n, \tilde{F}_n = 1, \mathcal{L}, |\mathcal{L}|) \\
& = \sum_{l=0}^{2^{nR_2}} \mathbf{P}(|\mathcal{L}| = l, \tilde{E}_n = 1) \\
& \quad \times H(M_2 | \tilde{E}_n = 1, \tilde{F}_n = 1, \mathcal{L}, |\mathcal{L}| = l) \\
& + \sum_{l=0}^{2^{nR_2}} \mathbf{P}(|\mathcal{L}| = l, \tilde{E}_n = 0) \\
& \quad \times H(M_2 | \tilde{E}_n = 0, \tilde{F}_n = 1, \mathcal{L}, |\mathcal{L}| = l) \\
& \leq \sum_{l=0}^{2^{nR_2}} \mathbf{P}(|\mathcal{L}| = l, \tilde{E}_n = 1) \\
& \quad \times H(M_2 | \tilde{E}_n = 1, \tilde{F}_n = 1, \mathcal{L}, |\mathcal{L}| = l) \\
& + nR_2 \mathbf{P}(\tilde{E}_n = 0) \\
& \leq \sum_{l=0}^{2^{nR_2}} \mathbf{P}(|\mathcal{L}| = l, \tilde{E}_n = 1) \log(l) + nR_2 \mathbf{P}(\tilde{E}_n = 0) \\
& \leq \sum_{l=0}^{2^{nR_2}} \mathbf{P}(|\mathcal{L}| = l | \tilde{E}_n = 1) \log(l) + nR_2 \mathbf{P}(\tilde{E}_n = 0) \\
& = \mathbf{E}[\log(|\mathcal{L}|) | \tilde{E}_n = 1] + nR_2 \mathbf{P}(\tilde{E}_n = 0) \\
& \stackrel{(b)}{\leq} \log(\mathbf{E}[|\mathcal{L}| | \tilde{E}_n = 1]) + nR_2 \mathbf{P}(\tilde{E}_n = 0) \\
& \stackrel{(c)}{\leq} \max \{0, n(R_2 - I(U_1, U_2; Y_1) + 4\delta(\epsilon') + \epsilon_n), \\
& \quad n(R_2 - I(U_2; Y_1, U_1) + \bar{\alpha}I(U_1; U_2) + 4\delta(\epsilon') + \epsilon_n)\} \\
& \quad + nR_2 \mathbf{P}(\tilde{E}_n = 0) \\
& \leq n \cdot \max \{0, R_2 - I(U_1, U_2; Y_1), \\
& \quad R_2 - I(U_2; Y_1, U_1) + \bar{\alpha}I(U_1; U_2)\} \\
& \quad + n4\delta(\epsilon') + n\epsilon_n + nR_2 \mathbf{P}(\tilde{E}_n = 0),
\end{aligned}$$

where (a) follows since the set \mathcal{L} and its cardinality $|\mathcal{L}|$ are functions of $(\mathcal{C}_n, M_1, Y_1^n)$, (b) follows by Jensen's inequality, and (c) follows by (94) and the soft-max interpretation of the log-sum-exp function [33, p. 72]. Substituting back gives

$$\begin{aligned}
& I(M_2; M_1, Y_1^n | \mathcal{C}_n) \\
& = H(M_2 | \mathcal{C}_n) - H(M_2 | \mathcal{C}_n, M_1, Y_1^n) \\
& = nR_2 - H(M_2 | \mathcal{C}_n, M_1, Y_1^n) \\
& \geq nR_2 - 2 - nR_2 \mathbf{P}(\tilde{F}_n = 0) \\
& \quad - H(M_2 | \mathcal{C}_n, M_1^n, Y_1^n, \tilde{E}_n, \tilde{F}_n = 1) \\
& \geq nR_2 - 2 - nR_2 \mathbf{P}(\tilde{F}_n = 0) - n4\delta(\epsilon') - n\epsilon_n \\
& \quad - nR_2 \mathbf{P}(\tilde{E}_n = 0) - n \cdot \max \{0, R_2 - I(U_1, U_2; Y_1), \\
& \quad R_2 - I(U_2; Y_1, U_1) + \bar{\alpha}I(U_1; U_2)\} \\
& \stackrel{(a)}{=} n[\min\{R_2, I(U_2; Y_1, U_1) - \bar{\alpha}I(U_1; U_2), I(U_1, U_2; Y_1)\} \\
& \quad - 4\delta(\epsilon') - 2\epsilon_n],
\end{aligned}$$

where (a) follows since both of the probabilities $\mathbf{P}(\tilde{E}_n = 0)$ and $\mathbf{P}(\tilde{F}_n = 0)$ tend to zero as $n \rightarrow \infty$.

APPENDIX I

PROOF OF PROPOSITION 3

We start with a version of Fano's inequality for computation, similar to Lemma 2 but for a fixed codebook this time.

Lemma 11: If

$$\lim_{n \rightarrow \infty} P_e^{(n)} = 0$$

and

$$\lim_{n \rightarrow \infty} \mathbf{P}(M_j \text{ is confusable}) = 0,$$

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

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

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

Proof: First note that for every $j \in \{1, 2\}$, we have

$$\begin{aligned} H(M_j | Y^n, M_{j^c}) &\leq H(M_j, W_{\mathbf{a}}^n | Y^n, M_{j^c}) \\ &= H(W_{\mathbf{a}}^n | Y^n, M_{j^c}) + H(M_j | W_{\mathbf{a}}^n, Y^n, M_{j^c}) \\ &\stackrel{(a)}{\leq} n\epsilon_n + H(M_j | W_{\mathbf{a}}^n, Y^n, M_{j^c}), \end{aligned}$$

where (a) follows by Fano's inequality. To bound the second term in (a), let j be such that $a_j \neq 0$ and let θ_j be an indicator random variable which is 1 if M_j is confusable. Then, we get

$$\begin{aligned} H(M_j | W_{\mathbf{a}}^n, Y^n, M_{j^c}) &\stackrel{(b)}{=} H(M_j | W_{\mathbf{a}}^n, Y^n, M_{j^c}, X_1^n, X_2^n) \\ &\leq H(M_j | X_j^n) \\ &\leq H(M_j, \theta_j | X_j^n) \\ &\stackrel{(c)}{\leq} \log_q 2 + H(M_j | \theta_j, X_j^n) \\ &= \log_q 2 + H(M_j | \theta_j = 1, X_j^n) \mathbf{P}(\theta_j = 1) \\ &\leq \log_q 2 + nR_j \mathbf{P}(\theta_j = 1) \\ &\stackrel{(d)}{\leq} n\epsilon_n, \end{aligned}$$

where (b) follows since (X_1^n, X_2^n) is a function of $(M_{j^c}, W_{\mathbf{a}}^n)$ when $a_j \neq 0$, (c) follows since θ_j is a binary random variable, and (d) follows since $\mathbf{P}(\theta_j = 1)$ tends to zero as $n \rightarrow \infty$. ■

Suppose now that a rate pair (R_1, R_2) is achievable. Let j be such that $a_j \neq 0$. Then,

$$\begin{aligned} nR_j &= H(M_j | M_{j^c}) \\ &\stackrel{(a)}{\leq} I(M_j; Y^n | M_{j^c}) + n\epsilon_n \\ &= \sum_{i=1}^n I(M_j; Y_i | Y^{i-1}, M_{j^c}) + n\epsilon_n \\ &= \sum_{i=1}^n I(M_j, X_{ji}; Y_i | Y^{i-1}, M_{j^c}, X_{j^c i}) + n\epsilon_n \\ &\leq \sum_{i=1}^n I(M_j, Y^{i-1}, M_{j^c}, X_{ji}; Y_i | X_{j^c i}) + n\epsilon_n \end{aligned}$$

$$\begin{aligned} &\stackrel{(b)}{=} \sum_{i=1}^n I(X_{ji}; Y_i | X_{j^c i}) + n\epsilon_n \\ &\stackrel{(c)}{=} nI(X_{jQ}; Y_Q | X_{j^c Q}, Q) + n\epsilon_n, \end{aligned}$$

where step (a) follows by Lemma 11, (b) follows since $(M_1, M_2, Y^{i-1}) \rightarrow (X_{1i}, X_{2i}) \rightarrow Y_i$ form a Markov chain, and (c) follows by defining a time sharing random variable Q that is uniform on $[n]$ and independent from (X_1^n, X_2^n, Y^n) .

We can continue from (a) above to provide another bound on nR_j as follows.

$$\begin{aligned} nR_j &\leq I(M_j; Y^n | M_{j^c}) + n\epsilon_n \\ &= I(M_1, M_2; Y^n) - I(M_{j^c}; Y^n) + n\epsilon_n \\ &\stackrel{(d)}{\leq} I(M_1, M_2; Y^n) - I(M_{j^c}; W_{\mathbf{a}}^n, Y^n) + 2n\epsilon_n \\ &= \sum_{i=1}^n I(M_1, M_2; Y_i | Y^{i-1}) \\ &\quad - \sum_{i=1}^n I(M_{j^c}; W_{\mathbf{a}, i}, Y_i | W_{\mathbf{a}}^{i-1}, Y^{i-1}) + 2n\epsilon_n \\ &= \sum_{i=1}^n I(M_1, M_2, X_{1i}, X_{2i}; Y_i | Y^{i-1}) \\ &\quad - \sum_{i=1}^n I(M_{j^c}, X_{j^c i}; W_{\mathbf{a}, i}, Y_i | W_{\mathbf{a}}^{i-1}, Y^{i-1}) + 2n\epsilon_n \\ &\stackrel{(e)}{=} \sum_{i=1}^n I(M_1, M_2, X_{1i}, X_{2i}; Y_i | Y^{i-1}) \\ &\quad - \sum_{i=1}^n I(M_{j^c}, X_{j^c i}; W_{\mathbf{a}, i}, Y_i | T_i) + 2n\epsilon_n \\ &\leq \sum_{i=1}^n I(M_1, M_2, Y^{i-1}, X_{1i}, X_{2i}; Y_i) \\ &\quad - \sum_{i=1}^n I(X_{j^c i}; W_{\mathbf{a}, i}, Y_i | T_i) + 2n\epsilon_n \\ &\stackrel{(f)}{=} \sum_{i=1}^n I(X_{1i}, X_{2i}; Y_i) \\ &\quad - \sum_{i=1}^n I(X_{j^c i}; W_{\mathbf{a}, i}, Y_i | T_i) + 2n\epsilon_n \\ &= nI(X_{1Q}, X_{2Q}; Y_Q | Q) \\ &\quad - nI(X_{j^c Q}; W_{\mathbf{a}, Q}, Y_Q | T_Q, Q) + 2n\epsilon_n, \end{aligned}$$

where step (d) follows by Fano's inequality, (e) follows by defining $T_i := (W_{\mathbf{a}}^{i-1}, Y^{i-1})$, and (f) follows since $(M_1, M_2, Y^{i-1}) \rightarrow (X_{1i}, X_{2i}) \rightarrow Y_i$ form a Markov chain. Note that both $T_i \rightarrow (X_{1i}, X_{2i}) \rightarrow W_{\mathbf{a}, i}$ and $(T_i, W_{\mathbf{a}, i}) \rightarrow (X_{1i}, X_{2i}) \rightarrow Y_i$ each form a Markov chain.

We next bound the sum rate using the fact that $a_1, a_2 \neq 0$ as follows.

$$\begin{aligned} n(R_1 + R_2) &= H(M_1, M_2) \\ &= I(M_1, M_2; Y^n) + H(M_1, M_2, W_{\mathbf{a}}^n | Y^n) \\ &\stackrel{(g)}{\leq} I(M_1, M_2; Y^n) + H(M_1, M_2 | W_{\mathbf{a}}^n, Y^n) + n\epsilon_n \end{aligned}$$

$$\begin{aligned}
&= I(M_1, M_2; Y^n) + H(M_1|W_{\mathbf{a}}^n, Y^n) \\
&\quad + H(M_2|M_1, W_{\mathbf{a}}^n, Y^n) + n\epsilon_n \\
&\stackrel{(h)}{\leq} I(M_1, M_2; Y^n) + H(M_1|W_{\mathbf{a}}^n, Y^n) + 2n\epsilon_n \\
&= I(M_1, M_2; Y^n) + H(M_1|W_{\mathbf{a}}^n, Y^n) + H(M_2|W_{\mathbf{a}}^n, Y^n) \\
&\quad - H(M_1, M_2|W_{\mathbf{a}}^n, Y^n) + H(M_1|W_{\mathbf{a}}^n, Y^n, M_2) + 2n\epsilon_n, \tag{95}
\end{aligned}$$

where (g) follows by Fano's inequality and (h) follows by Lemma 11 since $a_2 \neq 0$. Note that since $a_1 \neq 0$, by Lemma 11, we also have

$$H(M_1|W_{\mathbf{a}}^n, Y^n, M_2) \leq n\epsilon_n.$$

Utilizing this observation in (95), we continue with

$$\begin{aligned}
&n(R_1 + R_2) \\
&\leq I(M_1, M_2; Y^n) + H(M_1|W_{\mathbf{a}}^n, Y^n) \\
&\quad + H(M_2|W_{\mathbf{a}}^n, Y^n) - H(M_1, M_2|W_{\mathbf{a}}^n, Y^n) + 3n\epsilon_n \\
&= I(M_1, M_2; Y^n) + I(M_1, M_2; W_{\mathbf{a}}^n, Y^n) \\
&\quad - I(M_1; W_{\mathbf{a}}^n, Y^n) - I(M_2; W_{\mathbf{a}}^n, Y^n) + 3n\epsilon_n \\
&= \sum_{i=1}^n I(M_1, M_2; Y_i|Y^{i-1}) \\
&\quad + \sum_{i=1}^n I(M_1, M_2; W_{\mathbf{a},i}, Y_i|W_{\mathbf{a}}^{i-1}, Y^{i-1}) \\
&\quad - \sum_{i=1}^n I(M_1; W_{\mathbf{a},i}, Y_i|W_{\mathbf{a}}^{i-1}, Y^{i-1}) \\
&\quad - \sum_{i=1}^n I(M_2; W_{\mathbf{a},i}, Y_i|W_{\mathbf{a}}^{i-1}, Y^{i-1}) + 3n\epsilon_n \\
&= \sum_{i=1}^n I(M_1, M_2, X_{1i}, X_{2i}; Y_i|Y^{i-1}) \\
&\quad + \sum_{i=1}^n I(M_1, M_2, X_{1i}, X_{2i}; W_{\mathbf{a},i}, Y_i|T_i) \\
&\quad - \sum_{i=1}^n I(M_1, X_{1i}; W_{\mathbf{a},i}, Y_i|T_i) \\
&\quad - \sum_{i=1}^n I(M_2, X_{2i}; W_{\mathbf{a},i}, Y_i|T_i) + 3n\epsilon_n \\
&\leq \sum_{i=1}^n I(M_1, M_2, Y^{i-1}, X_{1i}, X_{2i}; Y_i) \\
&\quad + \sum_{i=1}^n I(M_1, M_2, X_{1i}, X_{2i}; W_{\mathbf{a},i}, Y_i|T_i) \\
&\quad - \sum_{i=1}^n I(X_{1i}; W_{\mathbf{a},i}, Y_i|T_i) \\
&\quad - \sum_{i=1}^n I(X_{2i}; W_{\mathbf{a},i}, Y_i|T_i) + 3n\epsilon_n \\
&\stackrel{(k)}{=} \sum_{i=1}^n I(X_{1i}, X_{2i}; Y_i) + \sum_{i=1}^n I(X_{1i}, X_{2i}; W_{\mathbf{a},i}, Y_i|T_i) \\
&\quad - \sum_{i=1}^n I(X_{1i}; W_{\mathbf{a},i}, Y_i|T_i) - \sum_{i=1}^n I(X_{2i}; W_{\mathbf{a},i}, Y_i|T_i) + 3n\epsilon_n
\end{aligned}$$

$$\begin{aligned}
&= nI(X_{1Q}, X_{2Q}; Y_Q|Q) + nI(X_{1Q}, X_{2Q}; W_{\mathbf{a},Q}, Y_Q|T_Q, Q) \\
&\quad - nI(X_{1Q}; W_{\mathbf{a},Q}, Y_Q|T_Q, Q) \\
&\quad - nI(X_{2Q}; W_{\mathbf{a},Q}, Y_Q|T_Q, Q) + 3n\epsilon_n,
\end{aligned}$$

where step (k) follows since $(M_1, M_2, W_{\mathbf{a}}^{i-1}, Y^{i-1}) \rightarrow (X_{1i}, X_{2i}) \rightarrow (W_{\mathbf{a},i}, Y_i)$ form a Markov chain.

It remains to show the dependence balance condition in (39).

$$\begin{aligned}
0 &\leq I(M_1; M_2|W_{\mathbf{a}}^n, Y^n) \\
&\stackrel{(a)}{=} I(M_1; M_2|W_{\mathbf{a}}^n, Y^n) - I(M_1; M_2) \\
&= H(M_1|W_{\mathbf{a}}^n, Y^n) - H(M_1|M_2, W_{\mathbf{a}}^n, Y^n) \\
&\quad - H(M_1) + H(M_1|M_2) \\
&= I(M_1; W_{\mathbf{a}}^n, Y^n|M_2) - I(M_1; W_{\mathbf{a}}^n, Y^n) \\
&= \sum_{i=1}^n I(M_1; W_{\mathbf{a},i}, Y_i|M_2, W_{\mathbf{a}}^{i-1}, Y^{i-1}) \\
&\quad - \sum_{i=1}^n I(M_1; W_{\mathbf{a},i}, Y_i|W_{\mathbf{a}}^{i-1}, Y^{i-1}) \\
&= \sum_{i=1}^n I(M_1, X_{1i}; W_{\mathbf{a},i}, Y_i|M_2, X_{2i}, W_{\mathbf{a}}^{i-1}, Y^{i-1}) \\
&\quad - \sum_{i=1}^n I(M_1, X_{1i}; W_{\mathbf{a},i}, Y_i|W_{\mathbf{a}}^{i-1}, Y^{i-1}) \\
&\leq \sum_{i=1}^n I(M_1, M_2, X_{1i}; W_{\mathbf{a},i}, Y_i|X_{2i}, W_{\mathbf{a}}^{i-1}, Y^{i-1}) \\
&\quad - \sum_{i=1}^n I(X_{1i}; W_{\mathbf{a},i}, Y_i|W_{\mathbf{a}}^{i-1}, Y^{i-1}) \\
&\stackrel{(b)}{=} \sum_{i=1}^n I(X_{1i}; W_{\mathbf{a},i}, Y_i|X_{2i}, W_{\mathbf{a}}^{i-1}, Y^{i-1}) \\
&\quad - \sum_{i=1}^n I(X_{1i}; W_{\mathbf{a},i}, Y_i|W_{\mathbf{a}}^{i-1}, Y^{i-1}) \\
&= \sum_{i=1}^n I(X_{1i}; W_{\mathbf{a},i}, Y_i|X_{2i}, T_i) - \sum_{i=1}^n I(X_{1i}; W_{\mathbf{a},i}, Y_i|T_i) \\
&= nI(X_{1Q}; W_{\mathbf{a},Q}, Y_Q|X_{2Q}, T_Q, Q) \\
&\quad - nI(X_{1Q}; W_{\mathbf{a},Q}, Y_Q|T_Q, Q),
\end{aligned}$$

where (a) follows since M_1 and M_2 are independent and (b) follows since $(M_1, M_2, W_{\mathbf{a}}^{i-1}, Y^{i-1}) \rightarrow (X_{1i}, X_{2i}) \rightarrow (W_{\mathbf{a},i}, Y_i)$ form a Markov chain.

Letting $X_1 = X_{1Q}, X_2 = X_{2Q}, W_{\mathbf{a}} = W_{\mathbf{a},Q}, Y = Y_Q$, and $T = T_Q$ and $n \rightarrow \infty$ completes the proof.

REFERENCES

- [1] R. Ahlswede, "Multiway communication channels," in *Proc. 2nd Int. Symp. Inf. Theory*, Tsahkadsor, Armenian, 1971, pp. 23–52.
- [2] H. H. J. Liao, "Multiple access channels," Ph.D. dissertation, Dept. Elect. Eng., Univ. Hawaii, Honolulu, HI, USA, Sep. 1972.
- [3] J. Korner and K. Marton, "How to encode the modulo-two sum of binary sources (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 2, pp. 219–221, Mar. 1979.
- [4] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.
- [5] M. P. Wilson, K. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5641–5654, Nov. 2010.

- [6] W. Nam, S.-Y. Chung, and Y. H. Lee, "Capacity of the Gaussian two-way relay channel to within $\frac{1}{2}$ bit," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5488–5494, Nov. 2010.
- [7] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.
- [8] U. Niesen and P. Whiting, "The degrees of freedom of compute-and-forward," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5214–5232, Aug. 2012.
- [9] Y. Song and N. Devroye, "Lattice codes for the Gaussian relay channel: Decode-and-forward and compress-and-forward," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4927–4948, Aug. 2013.
- [10] S.-N. Hong and G. Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227–5243, Sep. 2013.
- [11] Z. Ren, J. Goseling, J. H. Weber, and M. Gastpar, "Maximum throughput gain of compute-and-forward for multiple unicast," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1111–1113, Jul. 2014.
- [12] S. Miyake, "Coding theorems for point-to-point communication systems using sparse matrix codes," Ph.D. dissertation, Dept. Elect. Eng. Inf. Syst., Univ. Tokyo, Tokyo, Japan, 2010.
- [13] A. Padakandla and S. S. Pradhan, "Computing sum of sources over an arbitrary multiple access channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 2144–2148.
- [14] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [15] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, May 1979.
- [16] T. Gariby and U. Erez, "On general lattice quantization noise," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 2717–2721.
- [17] 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*, Jul. 2012, pp. 2256–2260.
- [18] A. Padakandla, A. G. Sahebi, and S. S. Pradhan, "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, Mar. 2016.
- [19] P. Sen and Y.-H. Kim, "Homologous codes for multiple access channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 874–878.
- [20] P. Sen and Y.-H. Kim, "Homologous codes for multiple access channels," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1549–1571, Mar. 2020.
- [21] 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*, Jul. 2013, pp. 2641–2645.
- [22] S. H. Lim, C. Feng, A. Pastore, B. Nazer, and M. Gastpar, "A joint typicality approach to compute-forward," *IEEE Trans. Inf. Theory*, vol. 64, no. 12, pp. 7657–7685, Dec. 2018.
- [23] S. H. Lim, C. Feng, A. Pastore, B. Nazer, and M. Gastpar, "Towards an algebraic network information theory: Simultaneous joint typicality decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 1818–1822.
- [24] P. Sen, S. H. Lim, and Y.-H. Kim, "Optimal achievable rates for computation with random homologous codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2351–2355.
- [25] N. Karamchandani, U. Niesen, and S. Diggavi, "Computation over mismatched channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 666–677, Apr. 2013.
- [26] B. Bandemer, A. El 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.
- [27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
- [28] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [29] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
- [30] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.
- [31] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2524–2530, Oct. 1998.
- [32] A. P. Hekstra and F. M. J. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 44–53, Jan. 1989.
- [33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [34] L. Wang, E. Şaşoğlu, B. Bandemer, and Y.-H. Kim, "A comparison of superposition coding schemes," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2970–2974.

Pinar Sen (Student Member, IEEE) received the B.S. and M.S. degrees (Hons.) in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 2011 and 2014, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California at San Diego, La Jolla, CA, USA, in 2020. Her current research interests include coding and information theory in multi-user networks, with applications in wireless communications.

Sung Hoon Lim (Member, IEEE) received the B.S. degree (Hons.) in electrical and computer engineering from Korea University, South Korea, in 2005, and the M.S. degree in electrical engineering and Ph.D. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST) in 2007 and 2011, respectively. From August 2009 to July 2010, he was a Visiting Scholar with UCSD. From March 2012 to May 2014, he was with Samsung Electronics. From June 2014 to July 2016, he was a Post-Doctoral Associate with the School of Computer and Communication Sciences, École Polytechnique Fédérale (EPFL), Lausanne, Switzerland. From August 2016 to August 2019, he was with the Korea Institute of Ocean Science and Technology (KIOST), Busan, South Korea. He is currently an Assistant Professor with the School of Software, Hallym University, Chuncheon, South Korea. His research interests include information theory, wireless communications, data compression, coding theory, and machine learning. He was a Gold prize recipient of the Samsung Humantech paper awards in 2011 and the 2016 NRF Postdoctoral Fellowship. He has served as a Technical Program Committee for the 2015 Information Theory Workshop, Jeju, South Korea.

Young-Han Kim (Fellow, IEEE) received the B.S. degree (Hons.) in electrical engineering from Seoul National University, Seoul, South Korea, in 1996, and the M.S. degrees in electrical engineering and in statistics and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2001, 2006, and 2006, respectively. In 2006, he joined the University of California at San Diego, La Jolla, CA, USA, where he is currently a Professor with the Department of Electrical and Computer Engineering. He has coauthored the book *Network Information Theory* (Cambridge University Press, 2011). His current research interests include information theory, communication engineering, and data science. He was a recipient of the 2008 NSF Faculty Early Career Development Award, the 2009 U.S.–Israel Binational Science Foundation Bergmann Memorial Award, the 2012 IEEE Information Theory Paper Award, and the 2015 IEEE Information Theory Society James L. Massey Research and Teaching Award for Young Scholars. He served as an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and a Distinguished Lecturer for the IEEE Information Theory Society.