Compute-Forward for DMCs: Simultaneous Decoding of Multiple Combinations

Sung Hoon Lim®, Member, IEEE, Chen Feng®, Member, IEEE, Adriano Pastore®, Senior Member, IEEE, Bobak Nazer®, Senior Member, IEEE, and Michael Gastpar®, Fellow, IEEE

Abstract—Algebraic network information theory is an emerging facet of network information theory, studying the achievable rates of random code ensembles that have algebraic structure, such as random linear codes. A distinguishing feature is that linear combinations of codewords can sometimes be decoded more efficiently than codewords themselves. The present work further develops this framework by studying the simultaneous decoding of multiple messages. Specifically, consider a receiver in a multi-user network that wishes to decode several messages. Simultaneous joint typicality decoding is one of the most powerful techniques for determining the fundamental limits at which reliable decoding is possible. This technique has historically been used in conjunction with random i.i.d. codebooks to establish achievable rate regions for networks. Recently, it has been shown that, in certain scenarios, nested linear codebooks in conjunction with “single-user” or sequential decoding can yield better achievable rates. For instance, the compute–forward problem examines the scenario of recovering \( L \leq K \) linear combinations of transmitted codewords over a \( K \)-user multiple-access channel (MAC), and it is well established that linear codebooks can yield higher rates. This paper develops bounds for simultaneous joint typicality decoding used in conjunction with nested linear codebooks, and applies them to obtain a larger achievable region for compute–forward over a \( K \)-user discrete memoryless MAC. The key technical challenge is that competing codeword tuples that are linearly dependent on the true codeword tuple introduce statistical dependencies, which requires careful partitioning of the associated error events.

Index Terms—Compute–forward, joint decoding, linear codes, multiple-access channel.

I. INTRODUCTION

FOR several decades, decode–forward [1], compress–forward [1], and amplify–forward [2] have served as the fundamental building blocks of transmission strategies for relay networks. These three relaying strategies were initially developed on canonical network models such as the relay channel and the diamond relay network using random independent and identically distributed (i.i.d.) codebooks and joint typicality decoding arguments. Subsequently, these strategies were generalized to \( N \)-user relay networks [3]–[10] that also relied upon random i.i.d. codebooks and joint typicality decoding.

Beginning with the many-help-one source coding work of Körner and Marton [11] followed by a series of recent papers [12]–[37], it has been observed that random i.i.d. codebooks may not suffice to attain the capacity region of certain networks. Instead, codes with some form of algebraic structure, such as nested linear or lattice codes, can sometimes attain larger rate regions. In the context of relaying, this has led to a fourth relaying paradigm known as compute–forward [12]–[18]. The key idea is that, if all users employ the same linear or lattice codebook, then linear combinations of codewords are themselves codewords, and can often be recovered at higher rates as compared to recovering one (or more) codewords. After the relays recover linear combinations, they forward them to the destinations, which then obtain their desired codewords by solving a system of linear equations. This strategy was originally proposed for Gaussian channels with equal rates and power constraints using (random) nested lattice codes combined with “single-user” lattice decoding [14]. It was subsequently generalized to include unequal power constraints and rates as well as sequential decoding [22], [38], [39].

Much of the prior work that demonstrates the rate gains of random linear or lattice codes over random i.i.d. codes has focused on either binary or Gaussian channels. Inspired by these examples, there is now a concerted effort to generalize these results into proof techniques with the objective to develop an algebraic network information theory based
on codes with algebraic structure. (See the textbook of El Gamal and Kim for the state-of-the-art rate regions for random i.i.d. codes [7].) As demonstrated by Padakandla et al. [24] and Padakandla and Pradhan [33], [34], random nested linear codes, when combined with joint typicality encoding and decoding, can be used to generalize the aforementioned examples to discrete memoryless networks. The key insight is that, although a straightforward application of a random linear codebook will lead to a uniform input distribution, joint typicality encoding (i.e., multicoding) can be used to shape a random nested linear codebook to induce any input distribution while preserving the linear structure of the nested linear codes. Generic shaping of linear (lattice) codes was independently discovered in the context of trellis shaping by Forney [40], lattice codes by Gariby and Erez [41], and sparse linear codes by Miyake [42].

In this paper, we develop techniques for bounding the error probability for *simultaneous joint typicality decoding* when used in conjunction with nested linear codebooks. The main technical difficulty is that (exponentially many) competing codewords are linearly dependent on the true codewords, and thus create statistical dependencies that are not handled by classical bounding techniques. We partition error events based on a particular rank criterion, which in turn enables us to characterize the rate penalties that stem from these linear dependencies. We apply our bounds towards deriving an achievable rate region for the general compute–forward problem of recovering \( L \leq K \) linear combinations over a \( K \)-user discrete memoryless MAC. In prior work, we derived an achievable region for the special case of \( K = 2 \) users and, in the process, generalized technical lemmas from network information theory (e.g., packing, covering, Markov) so that they apply to random nested linear codes [43]. We employ these lemmas as part of our derivations for the \( K \geq 2 \) setting, and find that our achievable rate region improves upon our previous results for the \( K = 2 \) case. Overall, simultaneous decoding has played an important role in the development of many results in classical network information theory, and the simultaneous decoding bounds developed herein may also prove useful beyond the compute–forward setting.

The rest of the paper is organized as follows. In the next section, we formally give the problem statement. In Section III, we state our main results on the joint decoding rate region for computing multiple linear combinations (Theorem 1). In Section IV, we give the proof of Theorem 1 and finally, in Section V, we conclude with some discussions.

We closely follow the notation in [7]. Let \( \mathcal{X} \) denote a discrete set and \( x^n \) a length-\( n \) sequence whose elements belong to \( \mathcal{X} \). We use uppercase letters to denote random variables. For instance, \( X \) is a random variable that takes values in \( \mathcal{X} \). We follow standard notation for probability measures. Specifically, we denote the probability of an event \( A \) by \( P\{A\} \) and use \( p_X(x) \) to denote probability mass functions (pmf).

For a discrete set \( \mathcal{X} \), the type of \( x^n \) is defined to be \( \pi(x^n) \) := \( \{ \xi : x_\xi = x \} / n \) for \( x \in \mathcal{X} \). Let \( X \) be a discrete random variable over \( \mathcal{X} \) with probability mass function \( p_X(x) \). For any parameter \( \epsilon \in (0, 1) \), we define the set of \( \epsilon \)-typical \( n \)-sequences \( x^n \) (or the typical set in short) [44] as \( T_{\epsilon}(n)(\mathcal{X}) = \{ x^n : |\pi(x^n) - p_X(x) | \leq \epsilon p_X(x) \text{ for all } x \in \mathcal{X} \} \). We use \( \delta(\epsilon) > 0 \) to denote a generic function of \( \epsilon > 0 \) that tends to zero as \( \epsilon \to 0 \). One notable departure is that we define sets of message indices starting at zero rather than one with shorthand \( \{\pi\} := \{0, \ldots, n - 1\} \). We also define \( 1:n = \{1, \ldots, n\} \) and reserve \( K = [1:K] \) to denote the full set of users.

We use the notation \( F_q \) to denote a finite field of order \( q \). We denote deterministic row vectors with lowercase, boldface font (e.g., \( a \in F_q^K \)). Note that row vectors can also be written as a sequence (e.g., \( u^n \in F_q^n \)). We will denote random sequences using uppercase font (e.g., \( U^n \in F_q^n \)). For a subset of sequences, we define \( U(S) = \{U_k : k \in S\} \). Random matrices will be denoted with uppercase, boldface font (e.g., \( G \in F_q^{n \times k} \)) and we will use uppercase, sans-serif font to denote realizations of random matrices (e.g., \( G \in F_q^{n \times k} \)) or deterministic matrices. We denote by \( \Theta_k \in F_q^k \) the standard basis (row) vector where the \( k \)-th element is 1 and the rest of the elements are all zero.

Define the matrix \( I(S) = \Pi_{q \times K} \) as a subset of the identity matrix \( I \in F_q^{K \times K} \) composed of the standard basis vectors \( \Theta_k \), \( k \in S \), i.e., the rows of \( I(S) \) are \( \Theta_k \), \( k \in S \). Likewise, for any matrix \( A \), we define \( A(S) \) as the submatrix containing only those rows of \( A \) whose index is in \( S \), i.e., \( A(S) = I(S)A \). Specifically for vectors, we will frequently use the shorthand \( A_k \) for \( A(\{k\}) \). We denote the row span of \( A \) by \( \text{span}(A) \) as well as its nullspace by \( \text{null}(A) \). Throughout the paper, we assume that all rates \( R_k \), \( k \in K \) are non-negative and are subject to constraints \( R_k \geq 0 \).

We define an empty matrix as a matrix with zero rows or zero columns (or both). We will assume that an empty matrix is full rank with rank 0. The product of an empty matrix and another matrix is an empty matrix, e.g., if \( A \) is a \( 0 \times 3 \) empty matrix and \( B \) is a \( 3 \times 5 \) matrix, then \( AB \) is an empty matrix of size \( 0 \times 5 \).

II. PROBLEM STATEMENT

We now give a formal problem statement for compute–forward. Consider the \( K \)-user discrete memoryless multiple-access channel (DM-MAC)

\[
(\mathcal{X}_1 \times \cdots \times \mathcal{X}_K, P_{Y|X_1,\ldots,X_K}, Y)
\]

(1)

which consists of \( K \) input alphabets \( \mathcal{X}_k \), \( k \in [1:K] \), one receiver alphabet \( Y \), and a collection of conditional pmfs \( p_{Y|X_1,\ldots,X_K} \). See Figure 1 for an illustration.

Consider a finite field \( F_q \) and let \( A_1, \ldots, A_L \in F_q^{L \times K} \) denote coefficient vectors. Define

\[
A = \begin{bmatrix} A_1 \\ \vdots \\ A_L \end{bmatrix} \in F_q^{L \times K}
\]

(2)

as a coefficient matrix, with \( L \leq K \).

A \((2^{nR_1},\ldots,2^{nR_K}, n; A)\) code for compute–forward consists of

- \( K \) message sets \( [2^{nR_k}], k \in [1:K] \)
- \( K \) encoders, where encoder \( k \) maps each message \( m_k \in [2^{nR_k}] \) to a pair of sequences \( (u^n_k, x^n_k)(m_k) \in F_q^n \times \mathcal{X}^n_k \) such that \( u^n_k(m_k) \) is injective,
For a coefficient matrix $F \in \mathbb{F}_q^{L_F \times K}$, let us define the notation
\[
W_F = F \begin{bmatrix} U_1 & \cdots & U_K \end{bmatrix}^T
\]
for a vector of linear combinations of $(U_1, \ldots, U_K) \in \mathbb{F}_q^K$.

The following theorem establishes our main result on computing $L$ linear combinations.

**Theorem 1 (Compute–Forward for the DM-MAC):** A rate tuple $(R_1, \ldots, R_K)$ is achievable for recovering the $L$ linear combinations with coefficient matrix $A \in \mathbb{F}_q^{L_F \times K}$ if, for some pmf $\prod_{k=1}^{K} P(u_k)$ and symbol mappings $x_k(u_k), k \in K$, it is contained in
\[
\mathcal{R}_{\text{joint}} = \bigcup_{B,C,S,T} \bigcap_{k=1}^{K} \left\{ (R_1, \ldots, R_K) \in \mathbb{R}_+^K : \sum_{k \in T} R_k < H(U(T)) - H(W_B|Y, W_{CB}) \right\}
\]
where the set operations are over all tuples $(B, C, S, T)$ satisfying the following constraints:

1) $B \in \mathbb{F}_q^{L_B \times K}$ runs over all full-rank matrices such that $1 \leq L_B \leq K$ and span$(B) \supseteq \text{span}(A)$,
2) $C \in \mathbb{F}_q^{L_C \times L_B}$ runs over all full-rank matrices (including empty matrices) such that $0 \leq L_C < L_B$,
3) $S \subseteq [1:L_B]$ runs over all index sets of size $|S| = L_B - L_C$ satisfying
\[
\text{rank} \left( \begin{bmatrix} C \mid I(S) \end{bmatrix} \right) = L_B,
\]
4) $T \subseteq K$ runs over all index sets of size $|T| = L_B - L_C$ satisfying
\[
\text{rank} \left( \begin{bmatrix} B(S) \mid I(K \setminus T) \end{bmatrix} \right) = K.
\]

The coding strategy and error analysis are provided in Section IV. In the following, we give some remarks on Theorem 1.

**Remark 2:** Note that the rate region in curly braces in (5) is a function of $B$, $C$ and $T$ but not of $S$. Rather, in the context of the intersection $\cap_T$, $T$ runs over a set that depends on $S$ (cf. (7)).

**Remark 3:** By the Steinitz Lemma [45], there always exists at least one $S \subseteq K$ and $T \subseteq K$ such that (6) and (7) are satisfied, respectively.
Remark 4: Without loss of generality, in the evaluation of $\mathcal{R}_{\text{joint}}$ we can restrict C to being in reduced row echelon form [46] since the right-hand side of (5) only depends on C via span(C). Equivalently, we have that $I(U_W;Y,W_{BC}) = I(U_W;Y,W_{BC})$ for any $C,C'$ such that span(C) = span(C') since C and $C'$ are deterministic functions of one another. This simplification can be applied for any of the corollaries of Theorem 1 that follow.

Theorem 1 admits a direct generalization to multiple receivers. For instance, assume there are $K$ transmitters that communicate with $N$ receivers across the discrete memoryless channel $p_{Y_1,\ldots,Y_N|X_1,\ldots,X_K}$ and that the $i$'th receiver observes channel output $Y_i$ and wants the linear combinations (multicoding) and decoding steps with coefficient matrix $A^{(i)}$. Let $\mathcal{R}_{\text{joint}}^{(i)}$ denote (5) evaluated with $A^{(i)}$ in place of A and $Y_i$ in place of Y. Then, a rate tuple $(R_1,\ldots,R_K)$ is achievable if, for some pmf $\prod_{k=1}^K p(u_k)$ and symbol mappings $x_k(u_k), k \in K$, it is contained in $\bigcap_{i=1}^N \mathcal{R}_{\text{joint}}^{(i)}$.

The rate region in Theorem 1 can be extended to include coded time-sharing by introducing a time-sharing random variable (say, Q). However, since we are using linearly generated codes, the coded time sharing sequence $q^n$ cannot be directly applied to the codewords using superposition coding. For nested linear codes, a time sharing random variable can be included by generating a random sequence from $p(q)$ independently, and by applying the joint typicality encoding (multicoding) and decoding steps with $\prod_{k=1}^K p(q)p(u_k|q)$ and $x_k(u_k,q)$. The resulting rate region in (5) is then extended to,

$$\sum_{k \in T} R_k < H(U(T)|Q) - H(W_B|Y,W_{CB},Q).$$

The rate region in Theorem 1 can also be extended to the case of finite chain rings by using the theoretical tools developed in several existing papers, such as [47]–[50]. This extension (which is left for future work) generally leads to a rate penalty not encountered in the finite field case and further complicates the rate region.

The following corollary simplifies Theorem 1 for computing one linear combination over a two-user DM-MAC, i.e., $A \in \mathbb{F}_{q}^{1 \times 2}$ and $K = 2$. In particular, we consider the cases with $A^{(i)} = [a_1, a_2] = a$ where $a_1 \neq 0$ and $a_2 \neq 0$ to avoid degenerate cases. The case when rank(A) = 2 and $K = 2$ will be considered afterwards.

Corollary 1 (Two Users, One Linear Combination): Consider the case with $K = 2$ and $L = 1$. A rate pair $(R_1, R_2)$ is achievable for computing one linear combination with respect to the coefficients $A^{(i)} = [a_1, a_2] = a$ over a two-user DM-MAC if

$$(R_1, R_2) \in (\mathcal{R}_{\text{CF}} \cup \mathcal{R}_{\text{LMAC}})$$

for some pmf $p(u_1)p(u_2)$ and symbol mappings $x_1(u_1), x_2(u_2)$, where

$$\mathcal{R}_{\text{CF}} = \{(R_1, R_2); R_1 < H(U_1) - H(W_1|Y) \cup R_2 < H(U_2) - H(W_1|Y)\},$$

$$\mathcal{R}_{\text{LMAC}} = (\mathcal{R}_1 \cup \mathcal{R}_2),$$

and $\mathcal{R}_{\text{CF}}$ denote the two-user multiple-access achievable rate region (for a fixed input distribution and without time sharing). As shown in [43, App. E], $\mathcal{R}_{\text{LMAC}} \subseteq \mathcal{R}_{\text{CF}} \cup \mathcal{R}_{\text{LMAC}}$ since $\mathcal{R}_{\text{CF}}$ fills in the defect in $\mathcal{R}_{\text{LMAC}}$. Since the relation holds for both regions without time sharing, the inclusion relation obviously extends to the time-sharing case. An illustration of the rate regions is given in Fig. 2.

For the special case of $A = I$, the computation problem reduces to the conventional multiple-access problem, that is, we recover all $K$ messages individually. In the following corollary, we specialize Theorem 1 by fixing $B = 1$ for the multiple-access case. Note that, since our proposed coding scheme is constrained by the use of nested linear codes, our achievable rate region does not always match the multiple-access capacity region.

Corollary 2 (Multiple Access via Nested Linear Codes): A rate tuple $(R_1,\ldots,R_K)$ is achievable for multiple access with nested linear codes if there exists some pmf $\prod_{k=1}^K p(u_k)$ and symbol mappings $x_k(u_k), k \in K$ such that, for each natural number $0 \leq L_C < K$ and each full-rank matrix $C \in \mathbb{F}_{q}^{L_C \times K}$, we can select a subset $S \subseteq K$ (that can depend on C) of size $|S| = K - L_C$ satisfying

$$R(S) < I(U(S);Y,W_{\text{CF}}),$$

and

$$\text{rank} \left( \begin{bmatrix} C \\ I(|S|) \end{bmatrix} \right) = K.$$
Corollary 3 strictly improves upon our previous results [43, Theorem 5] for the two-user case, i.e., $\mathcal{R}_{\text{LMAC,old}}$ is contained in $\mathcal{R}_{\text{LMAC}}$.

In general, simultaneous decoding offers better performance than sequential decoding. However, for some applications a sequential decoder may offer a better compromise by lowering the implementation complexity, perhaps at the expense of rate (cf. successive cancellation decoding vs. joint decoding for multiple access). For notational convenience, let $A_k$ denote $A(\{k\})$ and $A^K = A(\{1, \ldots, K\})$. Extending the basic idea of successive cancellation to the computation problem, a decoder could first recover the linear combination corresponding to the first row of $A$, i.e., $W_{A_1}^n$, then use the channel output and the linear combination pair $(W_{A_1}^n, Y^n)$ to recover a second linear combination corresponding to the second row $A_2$ and so on (see Figure 3). Based on this sequential decoding strategy, the following theorem establishes a sequential decoding rate region for computing multiple linear combinations.

**Theorem 2 (Sequential Decoding):** A rate tuple $(R_1, \ldots, R_K)$ is achievable for computing the linear combinations with coefficient matrix $A \in \mathbb{F}_q^{L \times K}$ if, for some pmf $\prod_{k=1}^K p(u_k)$, symbol mappings $x_k(u_k)$, $k \in \mathcal{K}$ and full-rank matrix $B \in \mathbb{F}_q^{L_B \times K}$, $L \leq L_B \leq K$ satisfying $\text{span}(A) \subseteq \text{span}(B)$, we have that

$$R_k < H(U_k) - H(W_{B_j}|Y, W_{B_{j-1}}),$$

for all $1 \leq j \leq L_B$ and $k \in \mathcal{K}(B_j)$ where $B_j$ is the $j$-th row of $B$ and $B' = B(\{1:j\})$, and $\mathcal{K}(B_j) = \{ k \in \mathcal{K}: B_{jk} \neq 0 \}$.

**Proof:** Consider the case of recovering a single linear combination ($L = 1$) corresponding to a vector $\tilde{A} \in \mathbb{F}_q^{1 \times K}$.

For this case, we evaluate Theorem 1 by fixing $A = B = \tilde{A}$. The resulting region is the set of rates $(R_1, \ldots, R_K)$ such that for all $1 \leq j \leq L_B$ and $k \in \mathcal{K}(B_j)$,

$$R_k < H(U_k) - H(W_{B_j}|Y, W_{CB}),$$

with $\text{span}(A) \subseteq \text{span}(B)$.

Let

$$\mathcal{R}_{\text{joint}}(B) = \bigcup_{C} \bigcup_{S} \bigcup_{T} \left\{ (R_1, \ldots, R_K) \in \mathbb{R}_+^K : \sum_{k \in T} R_k < H(U(T)) - H(W_{B_j}|Y, W_{CB}) \right\}$$

and

$$\mathcal{R}_{\text{seq}}(B) = \bigcup_{(j,k): B_{jk} \neq 0 \atop B_{j-1}} \left\{ (R_1, \ldots, R_K) \in \mathbb{R}_+^K : R_k < H(U_k) - H(W_{B_j}|Y, W_{B_{j-1}}) \right\}$$

denote the partial rate regions involved in Theorems 1 and 2, respectively, prior to computing the union over all matrices $B$ satisfying $\text{span}(B) \supseteq \text{span}(A)$. All set operations (unions and intersections) in (16)–(17) are to be taken over the sets specified in the statements of Theorems 1 and 2, respectively.

Thus, we have that

$$\mathcal{R}_{\text{joint}} = \bigcup_B \mathcal{R}_{\text{joint}}(B)$$

and we similarly define

$$\mathcal{R}_{\text{seq}} = \bigcup_B \mathcal{R}_{\text{seq}}(B).$$

**Theorem 3:** For any $B$, it holds that

$$\mathcal{R}_{\text{seq}}(B) \subseteq \mathcal{R}_{\text{joint}}(B).$$

In particular, it follows that $\mathcal{R}_{\text{seq}} \subseteq \mathcal{R}_{\text{joint}}$.  

---

Fig. 2. An illustration of $\mathcal{R}_{\text{CF}}$, which is the $\mathcal{R}_{\text{CF}}$ rate region (9) evaluated with respect to a coefficient vector $a$ that minimizes $H(W_n|Y)$. For the two-user rate region $\mathcal{R}_{\text{MAC}} = \mathcal{R}_1 \cup \mathcal{R}_2$ in Corollary 1, if $\mathcal{R}_{\text{CF}}$ is contained in $\mathcal{R}_{\text{MAC}}$ as in (a), then $\mathcal{R}_{\text{MAC}}$ and $\mathcal{R}_{\text{MAC}}$ coincide. Otherwise, if $\mathcal{R}_{\text{CF}}$ protrudes out from $\mathcal{R}_{\text{MAC}}$ as in (b), then $\mathcal{R}_{\text{MAC}}$ is obtained by mirroring the protruding part along the dominant face, and removing it from $\mathcal{R}_{\text{MAC}}$.

Fig. 3. Sequential decoder for recovering multiple linear combinations.
The proof of Theorem 3 is given in Appendix B.

**Example 1:** Consider a $K = 3$ user DM-MAC with

$$Y = \left[ \sum_{k=1}^{3} X_k + Z \right] \mod 4,$$

where $X_k = \{0, 1\}$, $Y = Z = \{0, 1, 2, 3\}$, and $Z$ is an additive random noise generated with pmf $p_Z(0) = 1 - p$ and $p_Z(1) = p_Z(2) = p_Z(3) = p/3$.

Figure 5 depicts an inner bound on the joint decoding rate region $\mathcal{R}_{\text{joint}}$ for the channel from Example 1. The bound is obtained by uniform distribution inputs, taking the union over one rank-1 matrix $B = [1, 1, 1]$, one rank-3 matrix $B = I$, and three rank-2 matrices,

$$B \in \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \right\}.$$  

We note that the rate region is exhausted by the specified $B$ matrices. For simplicity, we have only shown the rate region with a fixed distribution $p(u_k) = \text{Unif}(F_2)$ and $X_k = U_k$. Taking the union over different distributions will give a larger region (not shown for simplicity, cf. Example 2), however, the maximum sum rate is attained with the uniform distribution.

The sequential decoding points in Theorem 2 with $B = [1, 1, 1]$ and

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix},$$

are marked as $a$ and $b$, respectively. The region connecting the corner points $c$ is the multiple-access capacity (recovering the messages separately) of the channel.

**Example 2:** Consider the same channel as in Example 1 but with

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

i.e., jointly computing two linear combinations. The inner bound evaluated with Theorem 1 is shown in Figure 6. The bound is obtained by fixing the input distributions to the uniform distribution and taking the union over rank-2 matrices such that $\text{span}(B) = \text{span}(A)$ and one rank-3 matrix $B = I$. The rate region of Figure 6 can be enlarged by taking the union over both uniform and non-uniform input distributions as illustrated in Figure 7.

Before concluding the section, we explain how our DMC results are related to the lattice compute–forward strategy by Nazer and Gastpar [14] by specializing Corollary 1 to the two-user Gaussian MAC,

$$Y^n = h \begin{bmatrix} x_1^n \\ x_2^n \end{bmatrix} + Z^n$$

where $h = [h_1, h_2]$ is the vector of channel gains, the noise $Z^n$ is i.i.d. $\mathcal{N}(0, 1)$, and the channel inputs are subject to average power constraints $\sum_{i=1}^{n} x_{ki}^2 \leq nP_k$. The goal is to
recover the linear combination with integer\(^1\) coefficient vector
\(A = [a_1, a_2] = a \in \mathbb{Z}^{1 \times 2}\) again assuming that \(a_1 \neq 0\) and \(a_2 \neq 0\) to avoid degenerate cases. In [43], we have shown via a discretization method that the rate region \(R_C\) in Corollary 1 can be specialized to the Gaussian case in the form of

\[
R_C = \left\{ (R_1, R_2) : \begin{array}{c}
R_1 < h(U_1) - h(W_a|Y) + \log \gcd(|a_1|, |a_2|), \\
R_2 < h(U_2) - h(W_a|Y) + \log \gcd(|a_1|, |a_2|)
\end{array} \right\},
\]

where \(U_k \sim \mathcal{N}(0, P_k)\), the symbol mappings are \(X_k = U_k\), and \(\gcd(|a_1|, |a_2|)\) is the greatest common divisor of \(|a_1|\) and \(|a_2|\). Specifically, the inequalities in (26) evaluate to

\[
\begin{align}
R_1 &< \frac{1}{2} \log \left( \frac{P_1}{a(\Sigma^{-1} + h^{-1}h)^{-1}a^T} \right) + \log \gcd(|a_1|, |a_2|), \\
R_2 &< \frac{1}{2} \log \left( \frac{P_2}{a(\Sigma^{-1} + h^{-1}h)^{-1}a^T} \right) + \log \gcd(|a_1|, |a_2|),
\end{align}
\]

(27a)

(27b)

where \(\Sigma = \text{diag}(P_1, P_2)\). The rate region given by the inequalities in (27) is the compute–forward rate region for asymmetric powers from [38]. Thus, the joint typicality approach can recover the best-known achievable rate region based on nested lattice codes.

\(^1\)It can be shown that, if the channel coefficients and power constraints are bounded, then we can select a large enough finite field such that any integer-linear combination of codewords (with a positive sum rate) has a corresponding finite field combination [14, Theorem 11]. Thus, we can evaluate the rate region by solving a special case of the shortest vector problem, which can be efficiently solved for \(K = 2\) by Gauss’ algorithm as well as for \(K > 2\) by the algorithm proposed in [51, Theorems 1, 2].

IV. PROOF OF THEOREM 1

We begin by specifying the nested linear codes that will be used as our encoding functions in this paper, starting with some definitions. For compatibility with linear codes, we define the q-ary expansion of the messages \(m_k \in [2^nR_k]\) by \(m_k \in \mathbb{F}_{q}^\kappa_k\), where \(\kappa_k = nR_k/\log(q)\). In addition to the messages, we use auxiliary indices \(l_k \in [2^n\hat{R}_k]\), \(k = 1, \ldots, K\), and similarly define their q-ary expansion by \(l_k \in \mathbb{F}_{q}^\hat{\kappa}_k\), where \(\hat{\kappa}_k = n\hat{R}_k/\log(q)\). We define \(R_k := R_k + \hat{R}_k\), \(\hat{R}_k := \max\{R_1, R_2, \ldots, R_K\}\) and \(R_k := \max\{R_1, R_2, \ldots, R_K\}\).

For notational convenience, we assume that \(n\hat{R}_k/\log(q)\) and \(n\hat{R}_k/\log(q)\) are integers for all rates in the sequel. Further define

\[
m_k(m_k, l_k) = [m_k, l_k, 0], \quad k \in \mathcal{K},
\]

where \(m_k(m_k, l_k) \in \mathbb{F}_{q}^\kappa\), \(\kappa = n\hat{R}_k/\log(q)\), and 0 is a vector of zeros with length \(n(\hat{R}_k - R_k)/\log(q)\). Note that all \(m_k(m_k, l_k)\) have the same length due to zero padding. When it is clear from the context, we will simply write \(m_k\) in place of \(m_k(m_k, l_k)\). Moreover, since the set \([2^nR_k]\) has a one-to-one correspondence to \(\mathbb{F}_{q}^\kappa\), with some abuse of notation and for simplicity, we will often denote \(m_k\) as a member of the set \([2^nR_k]\), i.e., \(m_k \in [2^nR_k]\).

We define a \((2^nR_1, \ldots, 2^nR_K, 2^n\hat{R}_1, \ldots, 2^n\hat{R}_K, \mathbb{F}_q, n)\) nested linear code as the collection of \(K\) codebooks generated by the following procedure.

**Codebook generation.** Fix a finite field \(\mathbb{F}_q\) and a parameter \(c' \in (0, 1)\). Randomly generate a \(\kappa \times n\) matrix, \(G \in \mathbb{F}_{q}^{\kappa \times n}\), and sequences \(d_k^n \in \mathbb{F}_q^n\), \(k = 1, \ldots, K\) where each element of \(G\) and \(d_k^n\) are randomly and independently generated according to \(\text{Unif}(\mathbb{F}_q)\).

For each \(k \in \mathcal{K}\), generate a linear code \(C_k\) with parameters \((R_k, \hat{R}_k, n, q)\) by

\[
u_k^n(m_k, l_k) = u_k^n(m_k(m_k, l_k)) = m_k(m_k, l_k)G \oplus d_k^n, \quad (28)
\]

for \(m_k \in [2^nR_k]\), \(l_k \in [2^n\hat{R}_k]\). Since we have a one-to-one correspondence between \((m_k, l_k)\) and \(m_k\), we will frequently use \(u_k^n(m_k)\) to denote \(u_k^n(m_k, l_k)\).

As an alternative representation, we write the codebook construction in (28) by

\[
\begin{bmatrix}
u_1^n(m_1) \\
\vdots \\
u_K^n(m_K)
\end{bmatrix} = MG \oplus D, \quad (29)
\]

where

\[
M = \begin{bmatrix} m_1 \\
\vdots \\
m_K \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} d_1^n \\
\vdots \\
d_K^n \end{bmatrix}. \quad (30)
\]

Throughout the proof, we will be interested in the linear dependency between \(m_1, \ldots, m_K\), and representation of messages in matrix form \(M\) will be useful. Note that from this construction, each codeword is i.i.d. uniformly distributed (i.e., \(\prod_{i=1}^K \text{Unif}(\mathbb{F}_q)\)), and the codewords are pairwise independent.
Encoding. Fix an arbitrary pmf $\prod_{k=1}^{K} p(u_k)$, and functions $x_k(u_k)$, $k \in K$. For $k \in K$, given $m_k \in [2^{N_k}]$, find an index $l_k \in [2^{N_k}]$ such that $u_k^m(m_k, l_k) \in T_{e}^{(n)}(U_k)$. If there is more than one, select one randomly and uniformly. If there is none, randomly choose an index from $[2^{N_k}]$. Node $k$ transmits $x_k(u_k)$, $i = 1, \ldots, n$.

Define the collection of $M$ matrices by the set $\mathcal{I}$ and the subset of $M \in \mathcal{I}$ with respect to the coefficient matrix $A$ by

$$\mathcal{I}_{\text{subset}}(A) = \{ M_A : M_A = A M, M \in \mathcal{I} \}. \quad (31)$$

Decoding. Let $\epsilon^e < \epsilon$. Upon receiving $y^n$, the decoder searches for a unique index tuple $M_A \in \mathcal{I}_{\text{subset}}(A)$ such that $M_A = AM$ and

$$\{ u_1^n(m_1), \ldots, u_K^n(m_K), y^n \} \in T_{\epsilon}^{(n)}, \quad (32)$$

for some $M \in \mathcal{I}$. If it finds a unique index tuple, it declares

$$\hat{M}_A = M_A \oplus AD \quad (33)$$

as its estimate. Otherwise, if there is no such index tuple, or more than one, the decoder declares an error.

Analysis of the probability of error. Let $M_1, \ldots, M_K$ be the messages, and $L_1, \ldots, L_K$ be the indices chosen by the encoders. With some abuse of notation, denote by the random variable $M_k$ the true sum of the indices $m_k(M_k, L_k)$, $k \in K$ with respect to the coefficients $A$. Then, the decoder makes an error only if one or more of the following events occur,

$$\mathcal{E}_1 = \{ U_1^n(M_k, l_k) \notin T_{\epsilon}^{(n)} \text{ for all } l_k \}$$

for some $m_K \in K$, $k \in K$,

$$\mathcal{E}_2 = \{ (U_1^n(M_1, L_1), \ldots, U_K^n(M_K, L_K), Y^n) \notin T_{\epsilon}^{(n)} \}.$$

By Lemma 1 stated below, the probability $P(\mathcal{E}_1)$ tends to zero as $n \to \infty$ if

$$R_k > D(p_{u_k} || p_{\hat{u}_k}) + \delta(\epsilon'), \quad k = 1, \ldots, K. \quad (35)$$

Lemma 1 [43, Lemma 9]: Let $(X, \hat{X}, X, \hat{X}) \sim p_{X, \hat{X}}(x, \hat{x})$ and $\hat{p}_{\hat{X}}(\hat{x})$ be a distribution on $\hat{X}$ such that $D(\hat{p}_{\hat{X}} || p_{\hat{X}}) < \infty$. Let $X^n$ be a random sequence with $\lim_{n \to \infty} \prod_{i=1}^{n} X^n(m) \in T_{\epsilon}^{(n)}(X) = 1$ and let $X^n(m), m \in C$, where $|C| \geq 2^{nR}$, be pairwise independent and independent of $X^n$, each distributed according to $\prod_{i=1}^{n} \hat{p}_{\hat{X}}(\hat{x}_i)$. Then, there exists a $\delta(\epsilon)$ that tends to zero as $\epsilon \to 0$ such that

$$\lim_{n \to \infty} P(\{ X^n, \hat{X} \in \mathcal{C} \}) = 0,$$
that we cannot directly apply the standard packing lemma (e.g. [7, Lemma 3.1]). In the following, we resolve this difficulty by partitioning the sum index tuples $\mathbf{M}_B$ and present a joint typicality lemma that can be applied to each subset separately.

To this end, we proceed with some definitions. Define the set
\[
\mathcal{L}_B = \{ \mathbf{M}_B : \mathbf{M}_B \in \mathcal{I} \text{subset}(B), \mathbf{M}_B \neq 0 \}. \tag{37}
\]
We further divide the set $\mathcal{L}_B$ into the cover
\[
\mathcal{L}_B(r, C) = \{ \mathbf{M}_B : \mathbf{M}_B \in \mathcal{L}_B, \text{rank(}\mathbf{M}_B) = r, \text{CM}_B = 0 \} = \{ \mathbf{M}_B : \mathbf{M}_B \in \mathcal{L}_B, \text{null}(\mathbf{M}_B^T) = \text{span}(C) \},
\]
for $1 \leq r \leq L_B$ and $C \subset \mathbb{F}_q^{(L_n - r) \times L_B}$. For the case $r = L_B$, $C = \emptyset$.

\[
\mathcal{L}_B(L_B, \emptyset) = \{ \mathbf{M}_B : \mathbf{M}_B \in \mathcal{L}_B, \text{rank}(\mathbf{M}_B) = L_B \}.
\]

Note that we have $\mathcal{L}_B = \bigcup_{r \geq 1} \mathcal{L}_B(r, C)$. That is, $\mathcal{L}_B$ can be divided into non-overlapping subsets where each subset $\mathcal{L}_B(r, C)$ contains matrices of the same rank $r$ and the same nullspace specified by $C$.

We are now ready to proceed with the last probability term $P(\mathcal{E}_3 \cap \mathcal{E}_1^c | \mathcal{M})$ using the union of events bound. Continuing from (36),
\[
P(\mathcal{E}_3 \cap \mathcal{E}_1^c | \mathcal{M}) \leq \sum_{r=1}^{L_B} \sum_{C \subset \mathcal{L}_B(r, C)} \sum_{\mathbf{M}_B \in \mathcal{L}_B(r, C)} P\{ (W_B^n(\mathbf{M}_B), Y^n) \in \mathcal{T}_{\epsilon,n}^{(n)}, \mathcal{E}_1^c | \mathcal{M} \}, \tag{a)
\]
\[
\leq \sum_{r=1}^{L_B} \sum_{C \subset \mathcal{L}_B(r, C)} 2^n \max_r (\sum_{k \in T} R_k + \sum_{k \in T} \hat{R}_k)
\times 2^{-n(I(W_{B(S)}; Y, W_{CB}) + D(pw_{B(S)} || q^{|S|}) + \hat{D} - \delta(\epsilon))},
\]
where $\hat{D} = \sum_{k \in \mathcal{K}} (D(p_{U_k} || q_k) - \hat{R}_k)$ and (a) holds for any $\mathcal{S} \subset \mathcal{K}$ such that $|\mathcal{S}| = r$ and
\[
\text{rank} \left( \begin{bmatrix} \mathcal{C} \\ |l(\mathcal{S})| \end{bmatrix} \right) = L_B,
\]
the maximum in the exponent of the last inequality is over all subsets $T \subset \mathcal{K}$ such that $|T| = r$ and
\[
\text{rank} \left( \begin{bmatrix} \mathcal{B}(S) \\ l(\mathcal{K} \setminus T) \end{bmatrix} \right) = K. \tag{39}
\]
In step (a), we have applied the following two key lemmas that provide a cardinality bound on the set $\mathcal{L}_B(r, C)$ and a joint typicality lemma for nested linear codes. The proofs of these lemmas are deferred to Appendix C.

**Lemma 2 (Cardinality Bound):** Let $L_B$ and $r$ be integers such that $L_B \leq K$ and $1 \leq r \leq L_B$. Let $\mathcal{B} \subset \mathbb{F}_q^{L_n \times K}$, $\mathcal{L}_B \leq K$ and $C \subset \mathbb{F}_q^{(L_n - r) \times L_B}$ be full rank matrices. Then, for any $\mathcal{S} \subset \mathcal{K}$ such that $|\mathcal{S}| = r$ and
\[
\text{rank} \left( \begin{bmatrix} \mathcal{C} \\ |l(\mathcal{S})| \end{bmatrix} \right) = L_B,
\]
we have that
\[
|\mathcal{L}_B(r, C)| \leq \max_T 2^n (\sum_{k \in T} R_k + \sum_{k \in T} \hat{R}_k), \tag{41}
\]
where the maximization is over all subsets $T \subset \mathcal{K}$ such that $|T| = r$ and
\[
\text{rank} \left( \begin{bmatrix} \mathcal{B}(S) \\ l(\mathcal{K} \setminus T) \end{bmatrix} \right) = K.
\]

**Lemma 3 (Joint Typicality Lemma for Nested Linear Codes):** Consider $1 \leq r \leq L_B$ and $C \subset \mathbb{F}_q^{(L_n - r) \times L_B}$ such that rank($\mathbf{C}$) = $L_B$ - $r$ and assume that $\mathbf{M}_B \in \mathcal{L}_B(r, C)$. Then,
\[
P \left\{ (W_B^n(\mathbf{M}_B), Y^n) \in \mathcal{T}_{\epsilon,n}^{(n)}, \mathcal{E}_1^c | \mathcal{M} \right\}
\leq 2^{-n(I(W_{B(S)}; Y, W_{CB}) + D(pw_{B(S)} || q^{|S|}) + \hat{D} - \delta(\epsilon))},
\]
where $\hat{D} = \sum_{k \in \mathcal{K}} (D(p_{U_k} || q_k) - \hat{R}_k)$.

Thus, for any $\mathbf{B} \in \mathbb{F}_q^{L_n \times K}$, span($\mathcal{A}$) $\subset$ span($\mathcal{B}$), we have a bound on $P(\mathcal{E}_3 \cap \mathcal{E}_1^c | \mathcal{M})$ that tends to zero as $n \to \infty$ if for all full rank $\mathbf{C} \subset \mathbb{F}_q^{L_n \times L_B}$, $0 \leq L_C < L_B$, there exists an $\mathcal{S}$ that satisfies (38) and
\[
\hat{R}_k > D(p_{U_k} || q_k) + \delta(\epsilon), \quad k \in \mathcal{K},
\]
\[
\sum_{k \in \mathcal{T} \setminus \mathcal{T}} R_k - \sum_{k \in \mathcal{T} \setminus \mathcal{T}} \hat{R}_k < I(W(B_S); Y, W_{CB})
+ D(pw_{B(S)} || q^{|S|}) + \hat{D} - \delta(\epsilon),
\]
for all $\mathcal{T}$ which satisfies (39). To complete the proof, eliminate the auxiliary rates $\hat{R}_k$, $k \in \mathcal{K}$ and find that
\[
\sum_{k \in \mathcal{T}} R_k < I(W(B_S); Y, W_{CB}) + D(pw_{B(S)} || q^{|S|})
- \sum_{k \in \mathcal{T}} D(p_{U_k} || q_k) - \delta(\epsilon')
\]
\[
\stackrel{(a)}{=} I(W(B_S); Y, W_{CB}) - H(W(B_S))
+ \sum_{k \in \mathcal{T}} H(U_k) - \delta(\epsilon')
\]
\[
= H(U(\mathcal{T})) - H(W(B_S); Y, W_{CB}) - \delta(\epsilon'),
\]
where step (a) is from the relation $D(pw_{B(S)} || q^{|S|}) = |\mathcal{S}| \log(q) - H(W(B_S))$ as well as $D(p_{U_k} || q_k) = \log(q) - H(U_k)$ and the fact that $|\mathcal{S}| = |\mathcal{T}|$.

**V. DISCUSSIONS**

In this paper, we presented a framework for integrating structured code ensembles into the joint typicality framework. As a case, we generalized the compute--forward framework to discrete memoryless networks and established a joint decoding rate region for computing any number of linear combinations of the codewords. Our work provides the foreground for a general theorem on arbitrary networks and flows using nested linear codes.

In this sense, we view the compute--forward framework, not as a fourth paradigm for relaying, but as a new dimension in code construction for relaying strategies. This is a more general perspective that views compute--forward as an “algebraic” decode--forward strategy as originally suggested by Abbas El Gamal in his 2010 ISIT Plenary Talk [53]. He also...
posed several interesting questions on how joint typicality coding strategies can be combined with structured codes. Our framework provides the initial tools to redevelop and explore the coding strategies in network information theory using random nested linear code ensembles in place of random i.i.d. code ensembles.

One by-product of our analysis is an achievable rate region for multiple access via nested linear codes. Recent work [54] has further explored this rate region and shown, through a careful selection of both the finite field and symbol mappings, that it in fact corresponds to the full multiple-access capacity region.

Another important aspect of our framework is the resulting simultaneous joint decoding rate region for compute–forward. The joint typicality decoder presented in this paper was shown to be optimal with respect to nested linear codes in [55] for the $K = 2$, $L = 1$ case. On the other hand, the sharpest-known analysis for a lattice-based compute–forward strategy relies on suboptimal sequential decoding [38] due to the technical limitations in analyzing joint decoders for lattice codes [56].

In an effort to build a unifying compute–forward framework that includes all previously-known achievable rate regions (traditionally obtained via lattice codes), one important question is whether the DMC framework presented herein can be translated to the continuous case and to integer-linear combinations over the real field.

In [43, Theorems 4,7], we have already given a proof for the special case $K = 2$, $L = 1$. Our discretization method, which borrows a key result from [57], is more involved than most of the classic discretization approaches for information-theoretic quantities (e.g., [7], [58], [59]). The proof of the general case for arbitrary $K$, $L$, requires yet more steps and will appear in an upcoming publication [60]. Other interesting directions for future research include developing guidelines for optimizing the input distributions, reducing the search space of matrices and sets, finding a more compact representation of our achievable rate region, and possibly bounding the field size in terms of the cardinalities of the input alphabets.

APPENDIX A

PROOF OF COROLLARY 1

We particularize Theorem 1 by setting $K = 2$, $L = 1$ and $A = a = [a_1, a_2] \in \mathbb{F}_q^{1 \times 2}$ with $a_1 \neq 0$ and $a_2 \neq 0$. Since in the outermost union operation in (5), $B$ runs over all matrices satisfying span($B$) \supset span($A$), we infer that $B$ must run over all full-rank matrices $B \in \mathbb{F}_q^{2 \times 2}$ as well as over all those $B \in \mathbb{F}_q^{2 \times 2}$ that are scalar multiples of $a$ (for which case it suffices to consider $B = a$). For the sake of simplifying derivations (at the cost of possibly missing out on a part of the achievable rate region), out of all possible full-rank 2-by-2 matrices $B$, we shall only retain the identity matrix $B = I$. In summary, the union operation reduces to taking the union over only two matrices, namely $B = a$ and $B = I$.

For $B = a$, by virtue of the constraints on $C$, $S$, $T$ laid out in Theorem 1, in the set operations of (3) the matrix $C$ can only be the 0-by-1 empty matrix, $S$ can only be the singleton set $\{1\}$, and $T$ can be either $\{1\}$ or $\{2\}$. The resulting rate region is $\mathcal{R}_{\text{CF}}$ as defined in (9).

For $B = I$, $C$ runs over all full-rank (empty) 0-by-2 and 1-by-2 matrices:
- For $C \in \mathbb{F}_q^{0 \times 2}$, $S$ and $T$ can only be equal to $S = T = \{1, 2\}$, hence we obtain the sum-rate bound

$$R_1 + R_2 < H(U_1, U_2) - H(W_B|Y)$$

$$= I(U_1, U_2; Y)$$

$$= I(X_1, X_2; Y).$$

- For $C = [c_1, c_2] \in \mathbb{F}_q^{1 \times 2}$ a non-zero vector, we need to further distinguish three cases:
  - Case $c_1 \neq 0$ and $c_2 = 0$: The index sets can only be equal to $S = T = \{2\}$, hence we obtain the rate bound

$$R_2 < H(U_2) - H(U_1, U_2|Y, U_1)$$

$$= I(U_2; Y, U_1)$$

$$= I(X_2; Y|X_1).$$

(42)

- Case $c_1 = 0$ and $c_2 \neq 0$: similarly to the previous case, the index sets can only be equal to $S = T = \{1\}$, hence we obtain the rate bound

$$R_1 < I(X_1; Y|X_2).$$

(43)

- Case $c_1 \neq 0$ and $c_2 \neq 0$: the index sets can be either $S = T = \{1\}$ or $S = T = \{2\}$. For the former, we obtain

$$R_1 < H(U_1) - H(U_1, U_2|Y, W_C)$$

$$= H(U_1) - H(U_1|Y, W_C)$$

$$= I(U_1; Y, W_C)$$

(44)

For the latter, we obtain similarly

$$R_2 < I(U_2; Y, W_C).$$

(45)

The last two rate inequalities (44)–(45) are combined via a logical ‘or’ (due to the union over $S$). Recombining the above three case distinctions on the coefficient pair $(c_1, c_2)$ via a logical ‘and’ (due to union over $C$) yields the rate region $\mathcal{R}_{\text{LMAC}}$ as defined in (10). Finally, the union over $B$ yields the final rate region $\mathcal{R}_{\text{CF}} \cup \mathcal{R}_{\text{LMAC}}$ and proves Corollary 1.

APPENDIX B

PROOF OF THEOREM 3

Let us define $\mathcal{R}_{\text{joint}}(B)$ as the joint decoding region

$\mathcal{R}_{\text{joint}}(B)$

where, for each $C$, we fix an index set $S^*$ chosen according to Algorithm 1. (To streamline our notation, we do not show the dependence of $S^*$ on $C$ explicitly.) In other words, instead of taking the union over all $S$ in Theorem 1, for each $C$ we fix a set $S = S^*$, which leads to the relation

$$\mathcal{R}_{\text{joint}}(B) \subseteq \mathcal{R}_{\text{joint}}(B),$$

since by following the steps in Algorithm 1, $S^*$ satisfies

$$\text{rank}([C^T, I(S^*)^T]) = L_B.$$  

(46)

In the following, we prove the relation

$$\mathcal{R}_{\text{seq}}(B) \subseteq \mathcal{R}_{\text{joint}}(B)$$

(47)
by showing that the rate region $\mathcal{R}_{\text{seq}}(B)$ satisfies every inequality in $\mathcal{R}_{\text{join}}(B)$, namely, the set of inequalities

$$\sum_{k \in T} R_k < H(U(T)) - H(W_{B,(S')\setminus Y,W_{CB}}),$$

for all full rank $C \in \mathbb{F}_{Lc \times Lc}$, $S'$ chosen by Algorithm 1, and all $T \subseteq K$ such that $|T| = Lc - Lc$ with

$$\text{rank}(B(S')^T, I(K \setminus T)^T) = K.$$  \hfill (48)

Lemma 4: Let $S'$ be chosen according to Algorithm 1. Then, for $T$ such that (7) is satisfied, there exists a one-to-one mapping $\sigma : S' \to T$ such that for all $j \in S'$, $B_{j,\sigma(j)} \neq 0$.

**Proof:** Let

$$\hat{B} = \begin{bmatrix} B(S') \\ I(K \setminus T) \end{bmatrix}.$$  

Since $\hat{B}$ and $B(S')$ are full rank, we have that $|S'| = |T|$. Next, we define a submatrix $B(S',T)$ which is formed by taking the elements $B_{ij}$, $i \in S'$ and $j \in T$. Since $B(S',T)$ is a submatrix of $\hat{B}$, the existence of a permutation $\sigma : [1:|S'|] \to [1:|T|]$ such that $B_{j,\sigma(j)}(S',T) \neq 0$, $j \in [1:|S'|]$, implies the existence of a one-to-one mapping $\sigma : S' \to T$ such that $B_{j,\sigma(j)} \neq 0$, and thus equivalently, $B_{j,\sigma(j)} \neq 0$ for $j \in S'$. To this end, we will show that there exists such a permutation for $B(S',T)$ by contradiction. From the fact that $\hat{B}$ is full rank and $I(K \setminus T)$ is a collection of standard basis vectors, it is easy to see that the submatrix $B(S',T)$ is a full rank matrix. Since $B(S',T)$ is also a square matrix, it is invertible. Suppose that there does not exist such a permutation for $B(S',T)$. Then, for all possible permutations, $\prod_{j=1}^{|S'|} B_{j,\sigma(j)}(S',T) = 0$. Since this implies that the determinant of $B(S',T)$ is zero, it contradicts the fact that it is invertible.

By taking the sum over both sides of the inequalities

$$R_{\sigma(j)} < H(U_{\sigma(j)}) - H(W_{B,j,Y,W_{B,j-1}}), \ j \in S'$$

which are included in the region $\mathcal{R}_{\text{seq}}(B)$, we have

$$\sum_{k \in T} R_k \leq H(U(T)) - \sum_{j \in S'} H(W_{B,j,Y,W_{B,j-1}}) \leq H(U(T)) - \sum_{j \in S'} H(W_{B,j,Y,W_{CB},W_{B,j-1}}) \leq H(U(T)) \leq H(U(T)) - \sum_{j \in S'} H(W_{B,j,Y,W_{CB},W_{B,(S'\cap [1:j-1])}}) \leq H(U(T)) \leq H(U(T)) - H(W_{B,B,s}Y,W_{CB},W_{B,s}) \leq H(U(T)) - H(W_{B,B,s}Y,W_{CB},W_{B,s}),$$

where step (a) follows since there exists a one-to-one mapping $\sigma : S' \to T$ such that for all $j \in S$, $B_{j,\sigma(j)} \neq 0$ and step (b) follows from the fact that

$$\text{span}(CB,B^j) = \text{span}(CB[\{j(S'\cap [1:j])\}],)$$

since for $k \notin S'$ where $1 \leq k \leq j$,

$$\mathbf{e}_k \in \text{span}\left([I(S'\cap [1:k])]\right)$$

according to Algorithm 1.

Finally, since the relation holds for an arbitrary $C$, we have shown the relation (47).

**APPENDIX C**

**PROOF OF LEMMAS 2 AND 3**

**A. Proof of Lemma 2**

Recall the definition $B(S) = I(S)B$. First, we show that $B(S)M$ is full rank. By assumption, $\text{rank}(M) = r$, where $r \leq Lc$. Thus,

$$\text{rank}\left(\begin{bmatrix} CB \\ I(S)BM \end{bmatrix}\right) = \text{rank}\left(\begin{bmatrix} 0 \\ I(S)BM \end{bmatrix}\right) = r.$$  

Next define

$$L_B(r,C,S) = \{ M_{B(S)} : M_{B(S)} = B(S)M, \ M \in I, \text{rank}(M) = r, CM_B = 0 \},$$

$$\bar{L}_B(r,S) = \{ M_{B(S)} : M_{B(S)} = B(S)M, \ M \in I, \text{rank}(M_B) = r \}.$$  

Then,

$$|L_B(r,C,S)| \leq |L_B(r,C,S)| \leq |L_B(r,S)|,$$

where (a) follows from the fact that $\hat{C} := [C^T, I(S)T]^T$ is an invertible $Lc \times Lc$ matrix and $CBM = 0$, and thus, there is a one-to-one correspondence between $B(S)M \leftrightarrow \hat{CBM} \leftrightarrow BM$. The proof is a direct consequence of the fact that $B(S)M$ is full rank and applying the following lemma on $L_B(r,C,S)$.

**Lemma 5:** Consider a matrix $M \in I$, where the $k$-th row is $m_k \in [2^{nR_k}]$. Let $B \in \mathbb{F}_{nq \times K}$, $1 \leq L \leq K$ be a full-rank matrix, and define

$$A_B(L_B) = \{ M_B : M_B = BM, I \in I, \text{rank}(M) = L \}.$$  

Then,

$$|A_B(L_B)| \leq \max_L 2^{nR_k(T)}$$

where the maximum is over all $T$ such that $|T| = L_B$ and $\text{rank}(M) = L_B$.  

**Proof:** We will prove this upper bound by construction. First, we begin with a special case where the rates are ordered by $R_1 \geq \cdots \geq R_k$ and $B$ is in reduced echelon form. Let $T = \{t_1, t_2, \ldots, t_{L_B}\}$ be the set of pivot positions of $B$. Then the maximum number of non-zero entries in the $j$-th row of $M_B$ is the same as that in the $j$-th row of $M$, which is given by $[nR_j/\log(q)]$ for $j = 1, \ldots, L_B$. In other words, the $j$-th row of $M_B$ has at most $2^{nR_j}$ possibilities, because
the $t_j$-th row of $M$ has at most $2^nR(K)$ possibilities. Hence, $M_B$ has at most $2^n\hat{R}(T)$ possibilities. This gives an upper bound for $|A_B(L_B)|$ under the special case. Note that $\hat{T}$ constructed above satisfies the condition of $|T| = L_B$, rank([$B^T$, $I(K \setminus T)^T$]) = $K$. Therefore,

$$|A_B(L_B)| \leq 2^n\hat{R}(T) \leq \max_T 2^n\hat{R}(T).$$

That is, the upper bound indeed holds for this special case.

Next, we consider a more general case where $\hat{R}_1 \geq \cdots \geq \hat{R}_K$ and $B$ is not necessarily in reduced row echelon form. Let RRE(B) be the reduced row echelon form of $B$. Then RRE(B) = QB for some $L_B \times L_B$ invertible matrix $Q$. Since $Q$ is invertible, the number of distinct $M_B$ is equal to the number of distinct RRE(B).M. This reduces to our special case.

Finally, we consider the most general case where $\hat{R}_1, \ldots, \hat{R}_K$ can be in an arbitrary order. Then there exists a permutation $\pi: K \to K$ such that $R_{\pi(1)} \geq \cdots \geq R_{\pi(K)}$. In this case, we treat user $\pi(j)$ as our “virtual” user $j$ and apply our previous argument to these virtual users. In particular, we let $\hat{T}_\pi$ be the set of pivot positions of $B$ with respect to the virtual users. Then, we have $|A_B(L_B)| \leq 2^n\hat{R}(\hat{T}_\pi)$. Moreover, for any permutation $\pi$, $\hat{T}_\pi$ satisfies the condition of $|T| = L_B$, rank([$B^T$, $I(K \setminus T)^T$]) = $K$. Therefore,

$$|A_B(L_B)| \leq 2^n\hat{R}(\hat{T}_\pi) \leq \max_T 2^n\hat{R}(T).$$

This completes the proof.

B. Proof of Lemma 3

Consider $1 \leq r \leq L_B$ and $C \in \mathbb{F}_q^{(L_B-r) \times L_B}$ such that rank($C$) = $L_B - r$ and assume that $M_B \in L_B(r, C)$. Fix a set $S \subseteq K$ such that |$S$| = $L_B$ and

$$\text{rank}(\{C^T, I(S)^T\}) = L_B.$$ 

Let $\hat{C}_1 = \{U^n_k(0, 0) \in T^{(n)}_k, k \in K\}$. Then, we have

$$P\{(W^n_B(M_B), Y^n) \in T^{(n)}_r, \epsilon^n_1 | M\} = 2^{-n(I(W^n_B(M_B), Y^n; W_{CB}) + D(p_{W_B}(\cdot) || p_{W_{CB}}(\cdot)) + D - \delta(e))}$$

from equations (50) to (59), shown at the top of the page, where $U^n_k(0) = \{U^n_k(0) : k \in S\}$, step (a) follows from the fact that $W^n_B$ and $(W^n_B(M_B), W^n_{CB})$ are deterministic functions of each other, step (b) follows from the fact that $T^{(n)}_r \subseteq T^{(n)}$, step (c) follows from the fact that conditioned on $M$, $(Y^n, W^n_{CB}) \to U^n_k(0) \to W^n_B(M_B)$ forms a Markov chain, step (d) follows from [43, Lemma 11], and step (e) follows from [43, Lemma 7].

REFERENCES

[1] T. Cover and A. E. Gamal, “Capacity theorems for the relay channel,” IEEE Trans. Inf. Theory, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[2] B. Schein and R. G. Gallager, “The Gaussian parallel relay channel,” in Proc. IEEE Int. Symp. Inf. Theory, Sorrento, Italy, Jun. 2000, p. 22.

[3] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” IEEE Trans. Inf. Theory, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
LIM et al.: COMPUTE-FORWARD FOR DMCs: SIMULTANEOUS DECODING OF MULTIPLE COMBINATIONS

[55] 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), Paris, France, Jun. 2018, pp. 2351–2355.

[56] O. Ordentlich and U. Erez, “On the robustness of lattice interference alignment,” IEEE Trans. Inf. Theory, vol. 59, no. 5, pp. 2735–2759, May 2013.

[57] A. V. Makki and Y. Wu, “Equivalence of additive-combinatorial linear inequalities for Shannon entropy and differential entropy,” IEEE Trans. Inf. Theory, vol. 64, no. 5, pp. 3579–3589, May 2018.

[58] R. M. Gray, Entropy and Information Theory. New York, NY, USA: Springer, 1990.

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

[60] A. Pastore, S. H. Lim, C. Feng, B. Nazer, and M. Gastpar, “A unified discretization approach to compute-forward: From discrete to continuous inputs,” in preparation.

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. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST) in 2010 and 2011, respectively.

From August 2009 to July 2010, he was a Visiting Scholar at UCSD. From March 2010 to May 2014, he was with Samsung Electronics. From June 2012 to July 2016, he was a Post-Doctoral Associate with the Department of Electrical and Computer Engineering, University of Toronto, Canada, in 2009 and 2014, respectively.

From 2014 to 2015, he was a Post-Doctoral Fellow with Boston University, USA, and École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He joined the School of Engineering, The University of British Columbia, Kelowna, Canada, in July 2015, where he is currently an Assistant Professor and a Co-Cluster Lead of Blockchain@UBC. His research interests include coding theory, wireless communications, data compression, coding theory, and machine learning. He was a Gold Prize Recipient of the Samsung Humantech Paper Awards in 2011, a recipient of the 2016 NRF Postdoctoral Fellowship, and has served in the Technical Program Committee for the 2015 Information Theory Workshop, Jeju, South Korea.

Chen Feng (Member, IEEE) received the B.Eng. degree from the Department of Electronic and Communications Engineering, Shanghai Jiao Tong University, China, in 2006, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the Department of Electrical and Computer Engineering, University of Toronto, Canada, in 2009 and 2014, respectively.

From 2014 to 2015, he was a Post-Doctoral Fellow with Boston University, USA, and École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He joined the School of Engineering, The University of British Columbia, Kelowna, Canada, in July 2015, where he is currently an Assistant Professor and a Co-Cluster Lead of Blockchain@UBC. His research interests include coding theory and its applications in various fields, ranging from wireless communications to quantum communications, and from communication networks to blockchain systems.

Adriano Pastore (Senior Member, IEEE) received the Diplôme d’Ingénieur from the École Centrale Paris, the Dipl.Ing. degree in electrical engineering from the Technical University of Munich in 2009, and the Ph.D. degree from the Signal Theory and Communications Department, Universität Politécnica de Catalunya (UPC), in 2014. From 2014 to 2016, he was a Post-Doctoral Fellow at the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne. He is currently a Senior Researcher at the Department for Statistical Inference for Communications and Positioning, Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Spain. His research interests include network information theory, wireless communications, privacy, and machine learning for communications.

Bobak Nazer (Senior Member, IEEE) received the B.S.E.E. degree from Rice University, Houston, TX, USA, in 2003, and the M.S. and Ph.D. degrees from the University of California at Berkeley, Berkeley, CA, USA, in 2005 and 2009, respectively, all in electrical engineering.

From 2009 to 2010, he was a Post-Doctoral Associate with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI, USA. He is currently an Associate Professor with the Department of Electrical and Computer Engineering and a Distinguished Faculty Fellow with the College of Engineering, Boston University, Boston, MA, USA. His research interests include information theory, communications, signal processing, and neuroscience.

Dr. Nazer received the Eli Jury Award from the EECS Department at UC Berkeley in 2009, the Dean’s Catalyst Award from the College of Engineering at BU in 2011 and 2017, the NSF CAREER Award in 2013, the IEEE Communications Society and Information Theory Society Joint Paper Award in 2013, the BU ECE Faculty Service Award in 2017, and the BU ECE Outstanding Faculty Teaching Award in 2018. He was one of the co-organizers for the Spring 2016 Thematic Program at the Institut Henri Poincaré on the Nexus of Information and Computation Theories. He was the General Chair of the 2019 IEEE North American School of Information Theory.

Michael Gastpar (Fellow, IEEE) received the Dipl. El.-Ing. degree from the Eidgenössische Technische Hochschule (ETH), Zürich, Switzerland, in 1997, the M.S. degree in electrical engineering from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 1999, and the Doctorat ès Science degree from the École Polytechnique Fédérale (EPFL), Lausanne, Switzerland, in 2002. He was also a Student in engineering and philosophy at The University of Edinburgh and the University of Lausanne.

From 2003 to 2011, he was an Assistant and a tenured Associate Professor with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, Since 2011, he has been a Professor with the School of Computer and Communication Sciences, École Polytechnique Fédérale (EPFL), Lausanne, Switzerland. He was also a Professor at the Delft University of Technology, The Netherlands, and a Researcher with the Mathematics of Communications Department, Bell Labs, Lucent Technologies, Murray Hill, NJ, USA. His research interests include network information theory and related coding and signal processing techniques, with applications to sensor networks and neuroscience.

Dr. Gastpar received the IEEE Communications Society and Information Theory Society Joint Paper Award in 2013 and the EPFL Best Thesis Award in 2002. He was an Information Theory Society Distinguished Lecturer from 2009 to 2011, an Associate Editor for Shannon Theory for the IEEE Transactions on Information Theory from 2008 to 2011, and the Technical Program Committee Co-Chair for the 2010 International Symposium on Information Theory, Austin, TX, USA.