

Distributed Structured Matrix Multiplication

Derya Malak

Communication Systems Department, EURECOM, Biot Sophia Antipolis, FRANCE
derya.malak@eurecom.fr

Abstract—We devise achievable encoding schemes for distributed source compression for computing inner products, symmetric matrix products, and more generally, square matrix products, which are a class of nonlinear transformations. To that end, our approach relies on devising nonlinear mappings of distributed sources, which are then followed by the structured linear encoding scheme, introduced by Körner and Marton. For different computation scenarios, we contrast our findings on the achievable sum rate with the state of the art to demonstrate the possible savings in compression rate. When the sources have special correlation structures, it is possible to achieve unbounded gains, as demonstrated by the analysis and numerical simulations.

Index Terms—Distributed computation, inner product, structured codes, matrix-vector multiplication, matrix multiplication.

I. INTRODUCTION

The inner product operation between two vectors captures the similarity between vectors and allows us to describe the lengths, angles, projections, vector norms, matrix norms induced by vector norms, orthogonality of vectors, polynomials, and a variety of other functions as well [1]. Inner products are widely used in geometry and trigonometry using linear algebra and in applications spanning physics, engineering, and mathematics, e.g., to determine the convolution of functions [1], and the Fourier transform approximations, machine learning [2] and pattern recognition [3], e.g., the linear regression and the least squares models [1], and quantum computing, e.g., to describe the overlap between the two quantum states [4].

With the advent of edge computing systems, massive parallelization techniques, coded computation frameworks, and modern distributed computing systems, e.g., MapReduce, Hadoop, and Spark, have been devised to implement the computationally intensive task of distributed matrix multiplication with low communication and computation cost [5]. To that end, novel coded matrix-multiplication constructions, e.g., polynomial codes [6], [7], gradient coding [5], and Lagrange coded computing [8], [9], for mitigating the costs, faulty nodes, and stragglers. Hence, the inner product is relevant as it is a building block of distributed matrix multiplication.

Co-funded by the European Union (ERC, SENSIBILITÉ, 101077361). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

This research was partially supported by a Huawei France-funded Chair towards Future Wireless Networks, and supported by the program “PEPR Networks of the Future” of France 2030.

In this paper, we devise structured encoding schemes for distributed computing of inner products and symmetric and, more generally, square matrices via distributed matrix products. Our main contributions are summarized as follows:

- We devise a distributed encoding scheme that performs structured coding on nonlinear mappings of two distributed sources \mathbf{A} and \mathbf{B} to compute their inner product.
- We showcase the conditions for which the sum rate achieved by this structured coding is strictly less than the sum rate of distributed unstructured encoding of the sources \mathbf{A} and \mathbf{B} [10]. Here, the performance criterion is that both \mathbf{A} and \mathbf{B} cannot be decoded by the receiver.
- We derive achievable rate regions for distributed computation of symmetric and square matrices, a class of nonlinear transformations, with a vanishing probability of error, and determine example scenarios — detailed in Corollaries 1-2 — with special correlation structures to guarantee, via the structured coding scheme of Körner and Marton [11], significant savings over [10].
- We contrast the achievable rates with the existing approaches (e.g., [10], [12], [13]) via numerical examples.

Connections to the state of the art. Slepian and Wolf have provided an unstructured coding technique for the asymptotic lossless compression of distributed source variables X_1 and X_2 at the minimum rate needed, i.e., $H(X_1, X_2)$ [10]. Han-Kobayashi [14] have provided a characterization to determine whether computing a general bivariate function $f(X_1, X_2)$ of two random sequences $\{X_{1i}\}$ and $\{X_{2i}\}$ from two correlated memoryless sources requires a smaller rate than $H(X_1, X_2)$. For distributed coding of a finite alphabet source X with side information Y , Orlitsky and Roche have devised an unstructured coding scheme to achieve the minimum rate at which source X has to compress for distributed computing of $f(X, Y)$ with vanishing error [15], exploiting Körner’s characteristic graph G_X and its entropy [16]. This scheme is equivalent to performing Slepian-Wolf encoding on the colors of the sufficiently large OR powers of G_X given Y [17]–[19].

Körner and Marton have devised a structured encoding strategy that minimizes the sum rate for distributed computing the modulo-two sum of doubly symmetric binary source (DSBS) sequences with a low probability of error [11]. Ahlswede and Han have tightened the rate region for general binary sources [20] that embed the regions of [11] and [10]. The rate region for this problem has been extended to a larger class of source distributions [21], and for reconstructing the modulo- q sum of

the two sources in a q -ary prime finite field \mathbb{F}_q at a sum rate of $2H(X_1 \oplus_q X_2)$ [14]. For computing a nonlinear function, the embedding of the function in a sufficiently large prime \mathbb{F}_q [12], [13], and finding an injective mapping between the function and $X_1 \oplus_q X_2$, known as *structured binning*, may provide savings over [10].

In this paper, leveraging these fundamental principles, we demonstrate further savings in compression for computing inner products as well as matrix products through devising nonlinear mappings of the sources followed by the linear encoding scheme in [11], while requiring a smaller prime field size versus [12], [13].

Notation. We denote by $H(X)$ the Shannon entropy of a discrete random variable X , which is drawn from probability mass function (PMF) P_X . Similarly, $H(\cdot, \cdot)$ and $H(\cdot | \cdot)$ denote the joint and conditional entropies, given a joint PMF P_{X_1, X_2} . Let $h(\epsilon)$ denote the binary entropy function for Bernoulli distributed X with parameter $\epsilon \in [0, 1]$, i.e., $X \sim \text{Bern}(\epsilon)$.

We denote by X^n the length n realization of X . The boldface notation \mathbf{X} denotes a random matrix with elements in \mathbb{F}_q and \mathbf{X}^\top is its transpose. $\mathbf{1}_m$ and $\mathbf{0}_m$ are length m all-ones and all-zeros vectors, respectively. $\mathbb{P}(A)$ is the probability of an event A , and $\mathbb{1}_{x \in A}$ is the indicator function which takes the value 1 if $x \in A$, and 0 otherwise. The inner product of two vectors in the vector space V over a field F is a scalar, which is a map $\langle \cdot, \cdot \rangle : V \times V \rightarrow F$.

II. SYSTEM MODEL AND MAIN RESULTS

We consider a distributed coding scenario with two sources and a receiver. The two sources are assigned matrix variables $\mathbf{A} \in \mathbb{F}_q^{m \times l}$ and $\mathbf{B} \in \mathbb{F}_q^{m \times l}$, respectively, that model two statistically dependent independent and identically distributed (i.i.d.) finite alphabet source matrix sequences with entries from a field whose characteristic is $q \geq 2$. The objective of the receiver is to compute a function $f(\mathbf{A}, \mathbf{B})$, which is the matrix product of \mathbf{A} and \mathbf{B} , i.e., $\mathbf{A}^\top \mathbf{B} : \mathbb{F}_q^{m \times l} \times \mathbb{F}_q^{m \times l} \rightarrow \mathbb{F}_q^{l \times l}$. To that end, we devise a distributed encoding scheme that performs structured coding on nonlinear source mappings.

We demonstrate different regimes where the sum rate achieved by *structured coding* is strictly less than $H(\mathbf{A}, \mathbf{B})$ [10] so that the receiver can recover $\mathbf{A}^\top \mathbf{B}$, while it is desired that sources \mathbf{A} and \mathbf{B} cannot be decoded by the receiver. Exploiting [11], we first study the special scenario of distributed computing of inner products, where the key idea is to implement linear structured coding of [11] on vector-wise embeddings of the sources (Propositions 1 and 2), and propose a hybrid scheme that combines [11] with the *unstructured coding* technique of [15]. We contrast the achievable rate savings over unstructured coding [10], and over linear embedding of $\mathbf{A}^\top \mathbf{B}$ in a prime finite field [12] and [13]. We generalize our scheme for computing symmetric matrices via distributed multiplication (Proposition 3) and provide a condition for not being able to recover \mathbf{A} and \mathbf{B} (Proposition 4). Finally, we consider the distributed computation of square matrices (Proposition 5 and Corollary 2). We showcase the

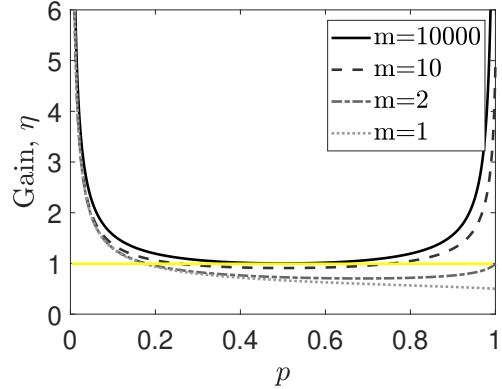


Fig. 1. Gain, η (cf. (3) in Corollary 1. The flat (yellow) line marks $\eta = 1$.

compression performance of our schemes versus the existing results, demonstrating the achievable savings.

We next detail our main achievability results.

A. Distributed Computation of Inner Products of Sources

The distributed sources hold the even-length vector variables $\mathbf{A} = [a_1 \ a_2 \ \dots \ a_m]^\top$ and $\mathbf{B} = [b_1 \ b_2 \ \dots \ b_m]^\top$, respectively, with entries chosen from a prime field \mathbb{F}_q . The receiver aims to compute the inner product $f(\mathbf{A}, \mathbf{B}) = \langle \mathbf{A}, \mathbf{B} \rangle$.

We next present an achievable coding scheme for distributed computation of $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^m a_i b_i$. The coding scheme relies on devising nonlinear mappings from each source and using the linear encoding scheme of Körner-Martón in [11].

Proposition 1. (Distributed inner product computation.)

Given two sequences of random vectors $\mathbf{A} = [\mathbf{A}_1^\top \ \mathbf{A}_2^\top]^\top \in \mathbb{F}_q^{m \times 1}$ and $\mathbf{B} = [\mathbf{B}_1^\top \ \mathbf{B}_2^\top]^\top \in \mathbb{F}_q^{m \times 1}$ of even length m , generated by two correlated memoryless q -ary sources, where $\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_1, \mathbf{B}_2 \in \mathbb{F}_q^{m/2 \times 1}$, the following sum rate is achievable by the separate encoding of the sources for the receiver to recover $\langle \mathbf{A}, \mathbf{B} \rangle$ with a small probability of error:

$$R_{\text{KM}}^\Sigma = 2H(\mathbf{U}, \mathbf{V}, W), \quad (1)$$

where $\mathbf{U}, \mathbf{V} \in \mathbb{F}_q^{m/2 \times 1}$ are vector variables, and $W \in \mathbb{F}_q$ is a random variable, and they satisfy the following relations:

$$\begin{aligned} \mathbf{U} &= \mathbf{A}_2 \oplus_q \mathbf{B}_1, \\ \mathbf{V} &= \mathbf{A}_1 \oplus_q \mathbf{B}_2, \\ W &= \mathbf{A}_2^\top \mathbf{A}_1 \oplus_q \mathbf{B}_1^\top \mathbf{B}_2. \end{aligned} \quad (2)$$

Proof. See Appendix A. \square

We next contrast the sum rate for computing $\langle \mathbf{A}, \mathbf{B} \rangle$ given by Proposition 1 with the sum rate of Slepian-Wolf in [10] to demonstrate the gap between R_{KM}^Σ and R_{SW}^Σ . The following result indicates that the rate needed to compute $\langle \mathbf{A}, \mathbf{B} \rangle$ may be substantially less than $H(\mathbf{A}, \mathbf{B})$ for binary-valued sources.

Corollary 1. Consider two sequences of correlated source vectors $\mathbf{A} \in \mathbb{F}_2^{m \times 1}$ and $\mathbf{B} \in \mathbb{F}_2^{m \times 1}$, with entries a_i and b_i that are i.i.d. across $i = 1, \dots, m$ each, where $a_i \sim \text{Bern}(1/2)$

and $b_i \sim \text{Bern}(1/2)$ are correlated for a given $i = 1, \dots, m$. Assume that \mathbf{U} and \mathbf{V} , as defined in (2), have entries $u_i, v_i \sim \text{Bern}(p)$, that are i.i.d. across $i = 1, \dots, m/2$.

For this setting, the gain of the encoding technique in Proposition 1 with a sum rate R_{KM}^Σ given in (1) over the sum rate R_{SW}^Σ for lossless compression of the sources is

$$\eta = \frac{R_{\text{SW}}^\Sigma}{R_{\text{KM}}^\Sigma} = \frac{m(1+h(p))}{2mh(p) + 2(1-(1-p)^m)}. \quad (3)$$

Proof. The proof follows from evaluating (1) and contrasting it with $R_{\text{SW}}^\Sigma = H(\mathbf{A}, \mathbf{B}) = m(1+h(p))$. For details, we refer the reader to Appendix B. \square

Note from Corollary 1 that when $\eta > 1$, the receiver can compute $\langle \mathbf{A}, \mathbf{B} \rangle$ without recovering (\mathbf{A}, \mathbf{B}) . It holds that

$$\lim_{p \rightarrow 0} \eta = \infty, \quad \lim_{p \rightarrow 1} \eta = \frac{m}{2}, \quad \lim_{m \rightarrow \infty} \eta = \frac{1+h(p)}{2h(p)}, \quad (4)$$

where the limit $\lim_{m \rightarrow \infty} \eta$ is the same as the gain for the DSBS model studied in [11], which tends to infinity as $p \rightarrow \{0, 1\}$.

We illustrate the gain η as a function of (m, p) in Figure 1, indicating that R_{KM}^Σ may be substantially less than the joint entropy of the sources for this special class of source PMFs.

It has been shown in [12], [13] that via embedding the nonlinear function $D_k = a_k b_k$, where $a_k, b_k \in \mathbb{F}_q$, in a sufficiently large prime \mathbb{F}_q , the decoder can reconstruct $\tilde{D}^n = \{a_k \oplus_q b_k\}_{k=1}^n$, and hence, compute $D^n = \{a_k b_k\}_{k=1}^n$ with high probability. For instance, if $a_k, b_k \in \mathbb{F}_2$, we can reconstruct D^n from $\tilde{D}^n = \{a_k \oplus_3 b_k\}_{k=1}^n$ using a sum rate of $R_S^\Sigma = 2H(\mathbf{A} \oplus_3 \mathbf{B})$.

Motivated by the notion of embedding in [12] and [13], we next devise an achievability scheme for computing $\langle \mathbf{A}, \mathbf{B} \rangle$, where the key idea is to compress the vector-wise embeddings of the sources vectors \mathbf{A} and \mathbf{B} via employing the linear structured encoding scheme of [11], in contrast to entry-wise embeddings that require a sum rate of R_S^Σ (cf. [12] and [13]).

Proposition 2. (Vector-wise embeddings followed by linear encoding for distributed computation of $\langle \mathbf{A}, \mathbf{B} \rangle$.) Given two sequences of vectors $\mathbf{A} \in \mathbb{F}_q^{m \times 1}$ and $\mathbf{B} \in \mathbb{F}_q^{m \times 1}$, generated by two correlated memoryless sources, with entries from a field \mathbb{F}_q with $q > 2$, restricting the source mappings to be linear; the following sum rate is achievable via the Körner-Martón's scheme to recover $\langle \mathbf{A}, \mathbf{B} \rangle$ at the receiver with a small probability of error:

$$R_{\text{SV}}^\Sigma = 2H\left(\left\{a_i \oplus_r b_i\right\}_{i=1}^m, \bigoplus_{i=1}^m a_i^2 \oplus_q b_i^2\right), \quad (5)$$

where $r = 2(q-1)m$ for m even, and $r = 2(q-1)m + 1$ for m odd, respectively, and \bigoplus_q denotes a modulo- q addition.

Proof. The proof follows from noting that the receiver, upon receiving $\{a_i \oplus_r b_i\}_{i=1}^m$ and $\bigoplus_{i=1}^m a_i^2 \oplus_q b_i^2$, can reconstruct

$$2c = qk + \left(\sum_{i=1}^m (a_i \oplus_r b_i)^2 - \bigoplus_{i=1}^m (a_i^2 \oplus_q b_i^2) \right) \bmod q,$$

where there is a unique $k \in \mathbb{F}_q$ for which $c = \langle \mathbf{A}, \mathbf{B} \rangle \in \mathbb{F}_q$. For details, we refer the reader to Appendix C. \square

We next describe a hybrid encoding scheme. Note that $\mathbf{A}^\top \mathbf{B}$ can be rewritten as $g(\mathbf{A}, \mathbf{Y}) = \mathbf{A}^\top (\mathbf{Y} - \mathbf{A}) \bmod q$ if $\mathbf{Y} = \mathbf{A} \oplus_q \mathbf{B}$ is known. Exploiting Körner's characteristic graphs [16] to enable nonlinear encoding of \mathbf{A}^1 , the minimum compression rate of \mathbf{A} for computing $g(\mathbf{A}, \mathbf{Y})$ given side information \mathbf{Y} equals the conditional graph entropy of \mathbf{A} given \mathbf{Y} , denoted by $H_{G_A}(\mathbf{A} | \mathbf{Y})$, as introduced by Orlitsky and Roche [15]. Via concatenating the structured coding scheme of Körner and Marton [11] to first compute \mathbf{Y} and then the unstructured coding model of Orlitsky and Roche [15] to next determine $g(\mathbf{A}, \mathbf{Y})$, it is possible to achieve a sum rate $R_{\text{KM-OR}}^\Sigma = 2H(\mathbf{Y}) + H_{G_A}(\mathbf{A} | \mathbf{Y})$. When $\mathbf{Y} = \mathbf{B}$, the required rate is $H(\mathbf{B}) + H_{G_A}(\mathbf{A} | \mathbf{B})$, which is smaller than R_{SW}^Σ because $H_{G_A}(\mathbf{A} | \mathbf{B}) \leq H(\mathbf{A} | \mathbf{B})$ [15].

In Figure 2, we contrast the sum rate performance of Propositions 1-2 for distributed computing of $\langle \mathbf{A}, \mathbf{B} \rangle$ of vectors $\mathbf{A}, \mathbf{B} \in \mathbb{F}_2^{m \times 1}$ with a small probability of error, with existing unstructured and structured coding schemes. In Figure 2-(Left), we use the PMF in Corollary 1 where $m = 2$, i.e., the pairs (a_1, b_2) and (a_2, b_1) represent DSBSs, each with a crossover probability p . Note that the sum rates R_S^Σ and $R_{\text{KM-OR}}^\Sigma$ perform poorly versus R_{SV}^Σ , and are not indicated. The sum rate R_{KM}^Σ converges to $H(\langle \mathbf{A}, \mathbf{B} \rangle)$ at low p .

In Figure 2-(Middle), we use $m = 1$, and the pair (a, b) is a DSBS with a crossover probability p . At low p , $R_{\text{KM-OR}}^\Sigma$ and R_{SW}^Σ converge to $H(\langle a, b \rangle)$ whereas R_S^Σ performs poorly. For large p , structured coding yields low rates (R_S^Σ and $R_{\text{KM-OR}}^\Sigma$).

In Figure 2-(Right), we use $m = 2$, and the pairs (a_1, b_1) and (a_2, b_2) are DSBSs, each with a crossover probability p . The performance of R_{KM}^Σ is worse than R_{SW}^Σ and not shown. Similarly, R_{SV}^Σ is higher than $R_{\text{KM-OR}}^\Sigma$, and is not indicated. For any given p value, R_S^Σ is always smaller than R_{SW}^Σ , and $R_{\text{KM-OR}}^\Sigma$ approaches $H(\langle \mathbf{A}, \mathbf{B} \rangle)$ for small and large p .

B. Distributed Computation of Symmetric Matrices

We next consider a generalization of Proposition 1 for distributed computing of inner products to distributed computing of a square symmetric matrix $\mathbf{D} = (D_{ij}) \in \mathbb{F}_q^{l \times l}$, given by the product $\mathbf{D} = \mathbf{A}^\top \mathbf{B}$, where $\mathbf{A}, \mathbf{B} \in \mathbb{F}_q^{m \times l}$, for $q > 2$ and $l \geq 1$, and symmetry implies that $D_{ji} = D_{ij}$ for every i, j .

Proposition 3. (Computing symmetric matrices via distributed multiplication.) Given two sequences of random matrices $\mathbf{A} = [\mathbf{A}_1^\top \ \mathbf{A}_2^\top]^\top \in \mathbb{F}_q^{m \times l}$ and $\mathbf{B} = [\mathbf{B}_1^\top \ \mathbf{B}_2^\top]^\top \in \mathbb{F}_q^{m \times l}$ generated by two correlated memoryless q -ary sources, respectively, where $\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_1, \mathbf{B}_2 \in \mathbb{F}_q^{m/2 \times l}$ and $q > 2$, the achievable sum rate by the separate encoding of the sources for the receiver to recover the symmetric matrix $\mathbf{D} = \mathbf{A}^\top \mathbf{B} \in \mathbb{F}_q^{l \times l}$ with vanishing error is given as

$$R_{\text{KM}}^\Sigma = 2H(\mathbf{U}, \mathbf{V}, \mathbf{W}), \quad (6)$$

¹For a detailed description of characteristic graphs and their entropies, we refer the reader to [17]–[19].

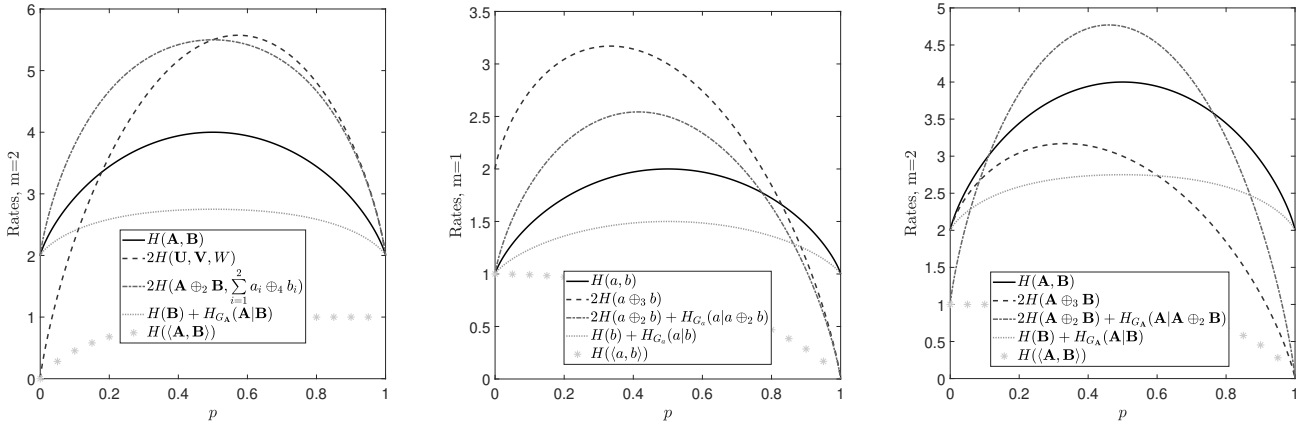


Fig. 2. Rate comparisons for various source PMFs. (Left) Corollary 1 for $m = 2$. (Middle) $m = 1$, and a, b are DSBSs. (Right) $m = 2$, and $\{a_i, b_i\}_{i=1}^2$ are DSBSs.

where $\mathbf{U} = \mathbf{A}_2 \oplus_q \mathbf{B}_1 \in \mathbb{F}_q^{m/2 \times l}$, $\mathbf{V} = \mathbf{A}_1 \oplus_q \mathbf{B}_2 \in \mathbb{F}_q^{m/2 \times l}$, and $\mathbf{W} = \mathbf{A}_2^T \mathbf{A}_1 \oplus_q \mathbf{B}_1^T \mathbf{B}_2 \in \mathbb{F}_q^{l \times l}$ are matrix variables.

Proof. The proof is on similar lines as Appendix A. For details, we refer the reader to Appendix D. \square

In Figure 3-(Left), we showcase the sum rates R_{KM}^Σ and R_{SW}^Σ versus p (in log scale) for distributed computing of symmetric matrices $\mathbf{D} = \mathbf{A}^T \mathbf{B}$ for $q = 2$ under assumptions²: i) $\mathbf{A}_1^T \mathbf{B}_1 = \mathbf{B}_1^T \mathbf{A}_1$, i.e., $\mathbf{W} = \mathbf{0}_{l \times l}$, and ii) $\mathbf{A}_2^T \mathbf{A}_1 = \mathbf{B}_1^T \mathbf{B}_2$. Because \mathbf{D} is symmetric, i) and ii) lead to

$$\begin{aligned} \mathbf{U}^T \mathbf{V} &= (\mathbf{A}_2 \oplus_2 \mathbf{B}_1)^T (\mathbf{A}_1 \oplus_2 \mathbf{B}_2) \\ &\stackrel{(a)}{=} \mathbf{A}_2^T \mathbf{A}_1 \oplus_2 \mathbf{A}_2^T \mathbf{B}_2 \oplus_2 \mathbf{A}_1^T \mathbf{B}_1 \oplus_2 \mathbf{B}_1^T \mathbf{B}_2 \\ &\stackrel{(b)}{=} \mathbf{A}_1^T \mathbf{B}_1 \oplus_2 \mathbf{A}_2^T \mathbf{B}_2 = \mathbf{D}, \end{aligned}$$

where (a) and (b) follow from assumptions i) and ii), respectively. These assumptions ensure a rate gain of $\eta = R_{\text{SW}}^\Sigma / R_{\text{KM}}^\Sigma$ that grows exponentially fast, as p tends to $\{0, 1\}$.

We next state a necessary condition for successful recovery of $\mathbf{D} = \mathbf{A}^T \mathbf{B}$, where $\mathbf{A}, \mathbf{B} \in \mathbb{F}_q^{m \times l}$, without recovering \mathbf{A} and \mathbf{B} . This result holds true for any symmetric $\mathbf{D} \in \mathbb{F}_q^{l \times l}$.

Proposition 4. (A necessary condition for the nonrecovery of the sources in distributed computation of $\mathbf{A}^T \mathbf{B}$.) For distributed encoding of $\mathbf{A}, \mathbf{B} \in \mathbb{F}_q^{m \times l}$ to compute $\mathbf{D} = \mathbf{A}^T \mathbf{B}$, which is symmetric, the below condition ensures that the sum rate in (1) is less than the achievable sum rate of [10]:

$$H(\mathbf{A}^T \mathbf{B}) + H(\mathbf{Q} | \mathbf{A}^T \mathbf{B}) < H(\mathbf{A} | \mathbf{Q}, \mathbf{A}^T \mathbf{B}), \quad (7)$$

where $\mathbf{Q} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} \in \mathbb{F}_q^{m \times 1}$, and \mathbf{U} and \mathbf{V} are defined in (2).

Proof. The inequality in (7) can be obtained using the expansions of R_{KM}^Σ and R_{SW}^Σ . For the technical steps, we refer the reader to Appendix E. \square

²For $q > 2$, it is clear (from Appendix D of Proposition 3) that the choices of \mathbf{U} , \mathbf{V} , and \mathbf{W} guarantee the recovery of $\mathbf{A}^T \mathbf{B}$ without assumptions i)-ii).

It is clear, using the expansion (24) of R_{KM}^Σ in Appendix E, that the sum rate $R_{\text{KM}}^\Sigma = 2H(\mathbf{A}^T \mathbf{B})$ is achievable when $H(\mathbf{Q} | \mathbf{A}^T \mathbf{B}) = 0$ in (7). In this case, (7) implies that $H(\mathbf{A}^T \mathbf{B}) < H(\mathbf{A} | \mathbf{Q}, \mathbf{A}^T \mathbf{B})$, meaning that it is possible to recover the inner product $\mathbf{A}^T \mathbf{B}$ while keeping \mathbf{A} and \mathbf{B} unknown to the receiver.

C. Distributed Computation of Square Matrices

Recall that Proposition 3 does not capture non-symmetric matrix products. We here consider distributed computing of a square matrix $\mathbf{D} = (D_{ij}) \in \mathbb{F}_q^{l \times l}$, given by the product $\mathbf{D} = \mathbf{A}^T \mathbf{B}$, where \mathbf{D} is not symmetric, and $\mathbf{A}, \mathbf{B} \in \mathbb{F}_q^{m \times l}$, for $l, m > 1$, and $q > 2$. The following proposition gives an achievable distributed encoding scheme of the sources \mathbf{A} and \mathbf{B} towards computing \mathbf{D} .

Proposition 5. (Computing square matrices via distributed matrix multiplication.) Given two sequences of random matrices $\mathbf{A} \in \mathbb{F}_q^{m \times l}$ and $\mathbf{B} \in \mathbb{F}_q^{m \times l}$ generated by two correlated memoryless q -ary sources, where $q > 2$, the following sum rate is achievable by the separate encoding of the sources for the receiver to recover a general square matrix $\mathbf{D} = \mathbf{A}^T \mathbf{B} \in \mathbb{F}_q^{l \times l}$ with vanishing error:

$$R_{\text{KM}}^\Sigma = 2H(\{\mathbf{A} \oplus_q \tilde{\mathbf{B}}_j\}_{j=1}^l, \{\tilde{\mathbf{A}}_j^T \tilde{\mathbf{A}}_j\}_{j=1}^l), \quad (8)$$

where we use the shorthand notation $\tilde{\mathbf{A}}_j^T \tilde{\mathbf{A}}_j = \mathbf{A}^T \mathbf{A} \oplus_q \tilde{\mathbf{B}}_j^T \tilde{\mathbf{B}}_j$ for $j \in \{1, 2, \dots, l\}$, where $\tilde{\mathbf{B}}_j = \mathbf{B}_j \mathbf{1}_{1 \times l} \in \mathbb{F}_q^{m \times l}$ are matrix variables, $\mathbf{1}_{1 \times l}$ is a length l row vector of all ones, where $\mathbf{B} = [\mathbf{B}_1 \ \mathbf{B}_2 \ \dots \ \mathbf{B}_l]$ with $\mathbf{B}_j \in \mathbb{F}_q^{m \times 1}$ for $j \in \{1, 2, \dots, l\}$.

Proof. We refer the reader to Appendix F. \square

To demonstrate the performance of Proposition 5, we next consider a corollary, where $l = 2$, $m > 1$, and with $q = 3$.

Corollary 2. Given $\mathbf{A} = [\mathbf{A}_1 \ \mathbf{A}_2] \in \mathbb{F}_3^{m \times 2}$ and $\mathbf{B} = [\mathbf{B}_1 \ \mathbf{B}_2] \in \mathbb{F}_3^{m \times 2}$ with entries $a_{ij} \sim (\frac{1}{2} - \epsilon, 2\epsilon, \frac{1}{2} - \epsilon)$ for

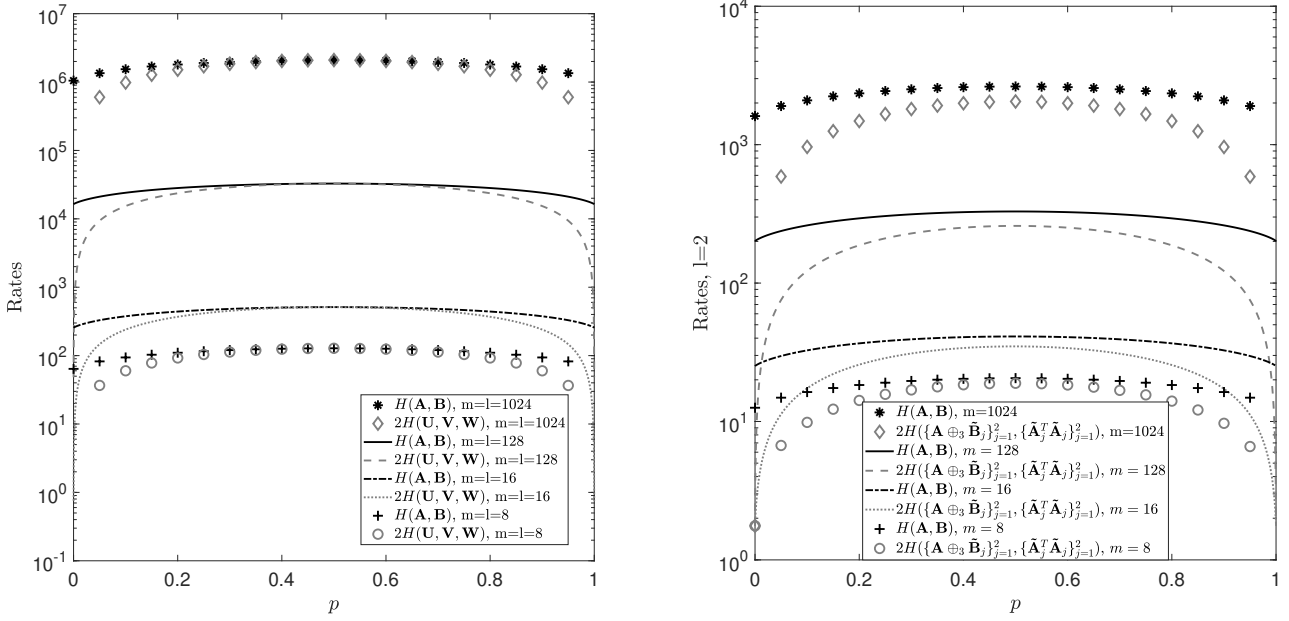


Fig. 3. Rate (in log scale) versus p for distributed computing of (Left) symmetric matrices $\mathbf{A}^T \mathbf{B} = \mathbf{B}^T \mathbf{A}$ via distributed multiplication of matrices $\mathbf{A}, \mathbf{B} \in \mathbb{F}_2^{m \times m}$ for different m , and (Right) square matrices via distributed matrix multiplication for different m and $l = 2$, where the joint source PMF is given in Corollary 2.

some $\epsilon \in [0, \frac{1}{2}]$, and $b_{i1} = b_{i2} = -a_{i2}$ that are i.i.d. across $i = 1, \dots, m$ and the joint PMF of a_{i1}, b_{i1} satisfies

$$P_{a_{i1}, b_{i1}} = \begin{bmatrix} (\frac{1}{2} - \epsilon)(1 - p) & (\frac{1}{2} - \epsilon)p & 0 \\ 2\epsilon p & 0 & 2\epsilon(1 - p) \\ 0 & (\frac{1}{2} - \epsilon)(1 - p) & (\frac{1}{2} - \epsilon)p \end{bmatrix}.$$

The sum rate for distributed encoding of (\mathbf{A}, \mathbf{B}) is given as

$$R_{\text{SW}}^{\Sigma} = m(h(2\epsilon) + (1 - 2\epsilon) + h(p)). \quad (9)$$

Exploiting Proposition 5 to compute $\mathbf{A}^T \mathbf{B}$, we can achieve

$$R_{\text{KM}}^{\Sigma} \leq 2mh \left(2 \left(\frac{1}{2} - \epsilon \right) (1 - p) + 2\epsilon(1 - p), \right. \\ \left. 2 \left(\frac{1}{2} - \epsilon \right) p + 2\epsilon p \right) + 2 \log_2(3). \quad (10)$$

Proof. We refer the reader to Appendix G. \square

In Figure 3-(Right), we demonstrate the sum rate performance of Proposition 5 (in log scale) versus p via contrasting the sum rates in (9) and (10) for the joint PMF model in Corollary 2, where we assume that $\epsilon = 0.2$. We observe that the rate gain $\eta = R_{\text{SW}}^{\Sigma} / R_{\text{KM}}^{\Sigma}$ grows exponentially fast, as p tends to $\{0, 1\}$.

Discussion. We proposed structured coding techniques for nonlinear mappings of distributed sources in a q -ary prime finite field to perform inner product-based matrix computation toward realizing distributed multiplication for special matrix classes, e.g., symmetric and square matrices, through imposing structural constraints on sources. Our future work includes the study of general matrix products, providing insights into the

problems of distributed rank computation, trace computation, and low-rank matrix factorization, as well as the derivation of tighter achievability bounds for the distributed multiplication of general matrices and higher dimensional matrices or tensors.

ACKNOWLEDGMENT

The author gratefully acknowledges the constructive discussions with Prof. Arun Padakandla at EURECOM.

APPENDIX

A. Proof of Proposition 1

The receiver aims to compute the inner product $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^m a_i b_i$, with entries from a field of characteristic $q \geq 2$. Here, we focus on $q = 2$, and the generalization to $q > 2$ is straightforward [14].

Encoding: Sources devise mappings $g_1 : \mathbf{A} \rightarrow \mathbf{X}_1$ and $g_2 : \mathbf{B} \rightarrow \mathbf{X}_2$, respectively, defined below, to determine the binary-valued column vectors

$$\mathbf{X}_1 = g_1(\mathbf{A}) = \begin{bmatrix} \mathbf{A}_2 \\ \mathbf{A}_1 \\ \mathbf{A}_2^T \mathbf{A}_1 \end{bmatrix} \in \mathbb{F}_2^{(m+1) \times 1}, \quad (11)$$

$$\mathbf{X}_2 = g_2(\mathbf{B}) = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \mathbf{B}_1^T \mathbf{B}_2 \end{bmatrix} \in \mathbb{F}_2^{(m+1) \times 1}. \quad (12)$$

We denote by $\mathbf{X}_1^n, \mathbf{X}_2^n \in \mathbb{F}_2^{(m+1) \times n}$ the length n source vectors. The encoders of the sources $\{\mathbf{X}_{1i}\}$ and $\{\mathbf{X}_{2i}\}$ are defined by functions $f_1 : \mathbf{X}_1^n \rightarrow \mathcal{R}_{f_1}$ and $f_2 : \mathbf{X}_2^n \rightarrow \mathcal{R}_{f_2}$, where \mathcal{R}_{f_1} and \mathcal{R}_{f_2} denote the ranges of f_1 and f_2 , respectively. The pair

of functions (f_1, f_2) is called an (n, ϵ) -coding scheme if there exists a function $\phi : \mathcal{R}_{f_1} \times \mathcal{R}_{f_2} \rightarrow \mathcal{Z}^n$ such that by letting

$$\hat{\mathbf{Z}}^n \triangleq \phi(f_1(\mathbf{X}_1^n), f_2(\mathbf{X}_2^n)), \quad (13)$$

we have $\mathbb{P}(\hat{\mathbf{Z}}^n \neq \mathbf{Z}^n) < \epsilon$. Here, \mathbf{Z} is the modulo-two sum of \mathbf{X}_1 and \mathbf{X}_2 , i.e., $\mathbf{Z} = \mathbf{X}_1 \oplus_2 \mathbf{X}_2 \in \mathbb{F}_2^{(m+1) \times 1}$.

Our encoding scheme requires a well-known lemma of Elias [22], which showed that linear codes achieve the capacity of binary symmetric channels, and its adaption to the problem of computing the modulo-two sum of DSBSs in [11]. Using a simple generalization of this result to vector variables, for fixed $\epsilon > 0$ and for sufficiently large n , there exists a binary matrix $\mathbf{C} \in \mathbb{F}_2^{k \times n}$, where $f_1(\mathbf{X}_1^n) \triangleq \mathbf{C}(\mathbf{X}_1^n) = \mathbf{C} \cdot (\mathbf{X}_1^n)^\top \in \mathbb{F}_2^{k \times (m+1)}$ and $f_2(\mathbf{X}_2^n) \triangleq \mathbf{C} \cdot (\mathbf{X}_2^n)^\top \in \mathbb{F}_2^{k \times (m+1)}$ denote the modulo-two product of the matrix \mathbf{C} with the transpose of the binary vector sequences \mathbf{X}_1^n and \mathbf{X}_2^n , respectively, and a decoding function $\psi : \{0, 1\}^{k \times (m+1)} \rightarrow \{0, 1\}^{n \times (m+1)}$ that satisfy

$$\phi(f_1(\mathbf{X}_1^n), f_2(\mathbf{X}_2^n)) \triangleq \psi(f_1(\mathbf{X}_1^n) + f_2(\mathbf{X}_2^n))$$

such that i) $k < n(H(\mathbf{Z}) + \epsilon)$, and ii) $\mathbb{P}(\psi(\mathbf{C}(\mathbf{Z}^n)) \neq \mathbf{Z}^n) < \epsilon$. Hence, application of Elias's lemma [22] and [11] yields that (\mathbf{C}, \mathbf{C}) is an (n, ϵ) -coding scheme.

Decoding: Exploiting the achievability result of Körner-Martón [11], the sum rate needed for the receiver to recover the vector sequence $\mathbf{Z}^n = \mathbf{X}_1^n \oplus_2 \mathbf{X}_2^n \in \mathbb{F}_2^{(m+1) \times n}$ with a vanishing error probability can be determined as [11]:

$$R_{\text{KM}}^\Sigma = 2H(\mathbf{U}, \mathbf{V}, W). \quad (14)$$

Using ψ given in ii), the receiver can recover \mathbf{Z}^n . However, lossless decoding of \mathbf{X}_1^n and \mathbf{X}_2^n is not guaranteed.

We next show that the sum rate in (14) is sufficient to recover $\langle \mathbf{A}, \mathbf{B} \rangle$. Using $\hat{\mathbf{Z}}^n$, the receiver computes

$$\begin{aligned} \mathbf{U}^\top \mathbf{V} - W &= (\mathbf{A}_2^\top \oplus_q \mathbf{B}_1^\top)(\mathbf{A}_1 \oplus_q \mathbf{B}_2) \\ &\quad - (\mathbf{A}_2^\top \mathbf{A}_1 \oplus_q \mathbf{B}_1^\top \mathbf{B}_2) \\ &= \mathbf{B}_1^\top \mathbf{A}_1 \oplus_q \mathbf{A}_2^\top \mathbf{B}_2 \\ &\stackrel{(a)}{=} \mathbf{A}_1^\top \mathbf{B}_1 \oplus_q \mathbf{A}_2^\top \mathbf{B}_2 = \langle \mathbf{A}, \mathbf{B} \rangle \end{aligned}, \quad (15)$$

where (a) follows from $\mathbf{B}_1^\top \mathbf{A}_1 = \mathbf{A}_1^\top \mathbf{B}_1 \in \mathbb{F}_q$.

Hence, (14) with (15) gives the achievability result we seek.

B. Proof of Corollary 1

Employing the definitions of \mathbf{X}_1 and \mathbf{X}_2 in (11), (1) and [11], we can determine the sum rate needed for the receiver to recover $(\mathbf{U}, \mathbf{V}, W)$ in an asymptotic manner:

$$\begin{aligned} R_{\text{KM}}^\Sigma &= 2H(\mathbf{U}, \mathbf{V}, W) \\ &= 2H(\mathbf{U}) + 2H(\mathbf{V}) + 2H(\mathbf{A}_2^\top \mathbf{A}_1 \oplus_2 \mathbf{B}_1^\top \mathbf{B}_2 | \mathbf{U}, \mathbf{V}) \\ &= 2\frac{m}{2}h(p) + 2\frac{m}{2}h(p) + 2H(\mathbf{A}_2^\top \mathbf{A}_1 \oplus_2 \mathbf{B}_1^\top \mathbf{B}_2 | \mathbf{U}, \mathbf{V}) \\ &\stackrel{(a)}{=} 2mh(p) + 2H(\mathbf{U}^\top \mathbf{A}_1 \oplus_2 \mathbf{A}_2^\top \mathbf{V} | \mathbf{U}, \mathbf{V}) \\ &\stackrel{(b)}{=} 2mh(p) + 2H(\mathbf{Q}^\top \mathbf{A} | \mathbf{U}, \mathbf{V}) \end{aligned}$$

$$\begin{aligned} &\stackrel{(c)}{=} 2mh(p) + 2H\left(\sum_{i=1}^m a_i z_i | \mathbf{Q}\right) \\ &\stackrel{(d)}{=} 2mh(p) + 2\sum_{j=1}^m \binom{m}{j} p^j (1-p)^{m-j} H\left(\sum_{i=1}^j a_i\right) \\ &\stackrel{(e)}{=} 2mh(p) + 2(1 - (1-p)^m), \end{aligned} \quad (16)$$

where (a) follows from employing the relations $\mathbf{U} = \mathbf{A}_2 \oplus_2 \mathbf{B}_1 = [u_1 \ u_2 \ \dots \ u_{m/2}]^\top$ and $\mathbf{V} = \mathbf{A}_1 \oplus_2 \mathbf{B}_2 = [v_1 \ v_2 \ \dots \ v_{m/2}]^\top$, and simplification using conditioning, (b) follows from employing $\mathbf{Q} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} \in \mathbb{F}_2^{m \times 1}$ and $\mathbf{A}_2^\top \mathbf{V} = \mathbf{V}^\top \mathbf{A}_2$, (c) from the definition of \mathbf{Q} , and (d) from $H(\mathbf{Q}^\top \mathbf{A} | \mathbf{Q}) \leq H(\mathbf{Q}^\top \mathbf{A} | \mathbf{Q}^\top \mathbf{1}_m)$, where $\mathbf{Q}^\top \mathbf{1}_m = \sum_{i=1}^m y_i \sim B(m, p)$, such that $H(\mathbf{Q}^\top \mathbf{A} | \mathbf{Q}^\top \mathbf{1}_m = j) = H\left(\sum_{i=1}^j a_i\right)$, $j \geq 1$ exploiting that $a_i \stackrel{i.i.d.}{\sim} \text{Bern}(1/2)$, and $H(\mathbf{Q}^\top \mathbf{A} | \mathbf{Q}^\top \mathbf{1}_m = 0) = 0$. Finally, incorporating $H\left(\sum_{i=1}^j a_i\right) = 1$, $j \geq 1$ we obtain (e).

The encoding rate for asymptotic lossless compression of \mathbf{A} and \mathbf{B} is given by the Slepian-Wolf theorem [10]:

$$\begin{aligned} R_{\text{SW}}^\Sigma &= H(\mathbf{A}, \mathbf{B}) \\ &\stackrel{(a)}{=} 2H(\mathbf{A}_1, \mathbf{B}_2) \\ &= 2\frac{m}{2}(1 + h(p)) \\ &= m(1 + h(p)), \end{aligned} \quad (17)$$

where (a) follows from using $\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix}$, with $\mathbf{A}_1 \perp \mathbf{A}_2$, $\mathbf{B}_1 \perp \mathbf{B}_2$, and \mathbf{A} and \mathbf{B} having i.i.d. entries.

From (16) and (17), $\eta = R_{\text{SW}}^\Sigma / R_{\text{KM}}^\Sigma$ is given by (3).

C. Proof of Proposition 2

Here, we provide a sketch of the proof. For binary sources, i.e., $q = 2$, it is easy to verify that

$$\langle \mathbf{A}, \mathbf{B} \rangle = \left[\frac{1}{2} \left(\sum_{i=1}^m a_i \oplus_r b_i - \sum_{i=1}^m (a_i \oplus_2 b_i) \right) \right] \pmod{2},$$

where $r = 2m$ for m even, and $r = 2m + 1$ for m odd, respectively. Hence, the following sum rate is achievable:

$$R_S^\Sigma = 2H\left(\sum_{i=1}^m a_i \oplus_r b_i, \{a_i \oplus_2 b_i\}_{i=1}^m\right). \quad (18)$$

When the data is generated by two correlated memoryless q -ary sources for $q \geq 2$, it is possible to achieve a sum rate

$$R_S^\Sigma = 2H\left(\left\{a_i \oplus_r b_i\right\}_{i=1}^m, \bigoplus_{i=1}^m a_i^2 \oplus_q b_i^2\right),$$

where $r = 2(q-1)m$ and $r = 2(q-1)m + 1$ for even and odd m , respectively. Upon receiving $\{a_i \oplus_r b_i\}_{i=1}^m$ and

$\bigoplus_{i=1}^m a_i^2 \oplus_q b_i^2$, the receiver can reconstruct

$$2c = qk + \left(\sum_{i=1}^m (a_i \oplus_r b_i)^2 - \bigoplus_{i=1}^m (a_i^2 \oplus_q b_i^2) \right) \bmod q ,$$

where there is a unique $k \in \mathbb{F}_q$ for which $c = \langle \mathbf{A}, \mathbf{B} \rangle \in \mathbb{F}_q$.

D. Proof of Proposition 3

Given two sequences of random matrices $\mathbf{A}, \mathbf{B} \in \mathbb{F}_q^{m \times l}$, the receiver aims to compute $\mathbf{A}^\top \mathbf{B} = \mathbf{B}^\top \mathbf{A} \in \mathbb{F}_q^{l \times l}$.

Encoding: Each source uses mappings $g_1 : \mathbf{A} \rightarrow \mathbf{X}_1$ and $g_2 : \mathbf{B} \rightarrow \mathbf{X}_2$, to determine the respective matrices:

$$\mathbf{X}_1 = g_1(\mathbf{A}) = \begin{bmatrix} \mathbf{A}_2 \\ \mathbf{A}_1 \\ \mathbf{A}_2^\top \mathbf{A}_1 \end{bmatrix} \in \mathbb{F}_q^{(m+l) \times l} , \quad (19)$$

$$\mathbf{X}_2 = g_2(\mathbf{B}) = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \mathbf{B}_1^\top \mathbf{B}_2 \end{bmatrix} \in \mathbb{F}_q^{(m+l) \times l} . \quad (20)$$

Following the steps of Appendix A, there exists an (n, ϵ) -coding scheme (\mathbf{C}, \mathbf{C}) for a matrix $\mathbf{C} \in \mathbb{F}_q^{k \times n}$, where $k = (m+l) \times l$, for decoding $\mathbf{Z} = \mathbf{X}_1 \oplus_q \mathbf{X}_2 \in \mathbb{F}_q^{(m+l) \times l}$ with a small probability of error.

Decoding: Exploiting the achievability result of Körner-Martón [11], the sum rate needed for the receiver to recover the matrix sequence $\mathbf{Z}^n = \mathbf{X}_1^n \oplus_q \mathbf{X}_2^n$ with a vanishing error probability can be determined as [11]:

$$R_{\text{KM}}^\Sigma = 2H(\mathbf{U}, \mathbf{V}, \mathbf{W}) . \quad (21)$$

For $q > 2$, given a symmetric matrix $\mathbf{D} = \mathbf{A}^\top \mathbf{B} \in \mathbb{F}_q^{l \times l}$, the following relation holds:

$$\mathbf{D} = \frac{1}{2}(\mathbf{D} \oplus_q \mathbf{D}^\top) = \mathbf{A}^\top \mathbf{B} = \mathbf{B}^\top \mathbf{A} . \quad (22)$$

Using $\hat{\mathbf{Z}}^n$, the receiver computes

$$\begin{aligned} & \frac{1}{2}((\mathbf{U}^\top \mathbf{V} - \mathbf{W}) \oplus_q (\mathbf{U}^\top \mathbf{V} - \mathbf{W})^\top) \\ &= \frac{1}{2}((\mathbf{A}_2^\top \oplus_q \mathbf{B}_1^\top)(\mathbf{A}_1 \oplus_q \mathbf{B}_2) - (\mathbf{A}_2^\top \mathbf{A}_1 \oplus_q \mathbf{B}_1^\top \mathbf{B}_2) \\ & \quad \oplus_q (\mathbf{A}_1^\top \oplus_q \mathbf{B}_2^\top)(\mathbf{A}_2 \oplus_q \mathbf{B}_1) - (\mathbf{A}_1^\top \mathbf{A}_2 \oplus_q \mathbf{B}_2^\top \mathbf{B}_1)) \\ &= \frac{1}{2}((\mathbf{B}_1^\top \mathbf{A}_1 \oplus_q \mathbf{A}_2^\top \mathbf{B}_2) \oplus_q (\mathbf{A}_1^\top \mathbf{B}_1 \oplus_q \mathbf{B}_2^\top \mathbf{A}_2)) \\ &= \frac{1}{2}((\mathbf{A}_1^\top \mathbf{B}_1 \oplus_q \mathbf{A}_2^\top \mathbf{B}_2) \oplus_q (\mathbf{A}_1^\top \mathbf{B}_1 \oplus_q \mathbf{A}_2^\top \mathbf{B}_2)^\top) \\ &= \frac{1}{2}(\mathbf{A}^\top \mathbf{B} \oplus_q (\mathbf{A}^\top \mathbf{B})^\top) = \mathbf{A}^\top \mathbf{B} , \end{aligned} \quad (23)$$

where the last equality follows from that $\mathbf{D} = \mathbf{A}^\top \mathbf{B}$ is a symmetric matrix and it satisfies (22).

Combining (21) with (23) gives the achievability result.

E. Proof of Proposition 4

We first show that the encoding scheme of Proposition 1 does not allow the recovery of $\mathbf{A}, \mathbf{B} \in \mathbb{F}_q^{m \times l}$ by the receiver, i.e., $H(\mathbf{A}, \mathbf{B} | \mathbf{A}^\top \mathbf{B}, \mathbf{Q}) > 0$, for $m > 1$.

The receiver can recover $\mathbf{Q} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} \in \mathbb{F}_q^{m \times 1}$ and $W \in \mathbb{F}_q$ with a small probability of error. The extra rate needed from the encoders for the receiver to determine $\mathbf{A}, \mathbf{B} \in \mathbb{F}_q^{m \times l}$ is

$$\begin{aligned} H(\mathbf{A}, \mathbf{B} | \mathbf{A}^\top \mathbf{B}, \mathbf{Q}) &\stackrel{(a)}{=} H(\mathbf{A}, \mathbf{B}, \mathbf{A}^\top \mathbf{B}, \mathbf{Q}) - H(\mathbf{A}^\top \mathbf{B}, \mathbf{Q}) \\ &\stackrel{(b)}{=} H(\mathbf{A}, \mathbf{B}, \mathbf{Q}) - H(\mathbf{A}^\top \mathbf{B}, \mathbf{Q}) \\ &\stackrel{(c)}{=} H(\mathbf{A}, \mathbf{Q}) - H(\mathbf{A}^\top \mathbf{Q}, \mathbf{Q}) \\ &\stackrel{(d)}{=} H(\mathbf{A}, \mathbf{A}^\top \mathbf{Q}, \mathbf{Q}) - H(\mathbf{A}^\top \mathbf{Q}, \mathbf{Q}) \\ &\stackrel{(e)}{=} H(\mathbf{A} | \mathbf{A}^\top \mathbf{Q}, \mathbf{Q}) \stackrel{(f)}{\geq} 0 , \end{aligned}$$

where (a) follows from using the definition of conditional entropy and that the receiver can compute $\mathbf{A}^\top \mathbf{B} = \mathbf{U}^\top \mathbf{V} - W$ from \mathbf{Q} and W , (b) from $H(\mathbf{A}^\top \mathbf{B} | \mathbf{Q}, \mathbf{A}, \mathbf{B}) = 0$, (c) from $H(\mathbf{B} | \mathbf{A}, \mathbf{Q}) = 0$ and $\mathbf{A}^\top \mathbf{B} = \mathbf{A}^\top \mathbf{Q}$ given \mathbf{Q} , (d) from $H(\mathbf{A}^\top \mathbf{Q} | \mathbf{A}, \mathbf{Q}) = 0$, (e) from employing the definition of conditional entropy, and (f) holds with equality if the function $g(\mathbf{A}, \mathbf{Q}) = \mathbf{A}^\top \mathbf{Q}$ is partially invertible, i.e., $H(\mathbf{A} | g(\mathbf{A}, \mathbf{Q}), \mathbf{Q}) = 0$, e.g., when g is the arithmetic sum or the modulo sum of the two vectors. Hence, the inequality in (f) is strict for inferring $\mathbf{A}, \mathbf{B} \in \mathbb{F}_q^{m \times l}$ from \mathbf{Q} and W .

We next prove the main result of the proposition. Given matrix variables $\mathbf{A}, \mathbf{B} \in \mathbb{F}_q^{m \times l}$ such that $\mathbf{D} = \mathbf{A}^\top \mathbf{B}$ is symmetric, letting $\mathbf{Q} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}$, we first expand R_{KM}^Σ as

$$\begin{aligned} R_{\text{KM}}^\Sigma &= 2H(\mathbf{U}, \mathbf{V}, W) \\ &= 2H(\mathbf{U}, \mathbf{V}, \mathbf{A}^\top \mathbf{B}) \\ &= 2H(\mathbf{A}^\top \mathbf{B}) + 2H(\mathbf{Q} | \mathbf{A}^\top \mathbf{B}) . \end{aligned} \quad (24)$$

We next expand R_{SW}^Σ as

$$\begin{aligned} R_{\text{SW}}^\Sigma &= H(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}, \mathbf{B}, \mathbf{U}, \mathbf{V}, W, \mathbf{A}^\top \mathbf{B}) \\ &= H(\mathbf{U}, \mathbf{V}, \mathbf{A}^\top \mathbf{B}) + H(\mathbf{A}_1, \mathbf{B}_1, \mathbf{A}_2, \mathbf{B}_2 | \mathbf{A}_2 \oplus_q \mathbf{B}_1, \\ & \quad \mathbf{A}_1 \oplus_q \mathbf{B}_2, \mathbf{A}_1^\top \mathbf{B}_1 \oplus_q \mathbf{A}_2^\top \mathbf{B}_2) \\ &= H(\mathbf{U}, \mathbf{V}, \mathbf{A}^\top \mathbf{B}) + H(\mathbf{A}_1, \mathbf{A}_2 | \mathbf{U}, \mathbf{V}, \mathbf{A}_1^\top (\mathbf{U} \oplus_q \mathbf{A}_2) \\ & \quad \oplus_q \mathbf{A}_2^\top (\mathbf{V} \oplus_q \mathbf{A}_1)) \\ &= H(\mathbf{Q}, \mathbf{A}^\top \mathbf{B}) + H(\mathbf{A} | \mathbf{Q}, \mathbf{A}^\top \mathbf{Q} \oplus_q \mathbf{A}_1^\top \mathbf{A}_2 \oplus_q \mathbf{A}_2^\top \mathbf{A}_1) \\ &= H(\mathbf{A}^\top \mathbf{B}) + H(\mathbf{Q} | \mathbf{A}^\top \mathbf{B}) \\ & \quad + H(\mathbf{A} | \mathbf{Q}, \mathbf{A}^\top \mathbf{Q} \oplus_q \mathbf{A}_1^\top \mathbf{A}_2 \oplus_q \mathbf{A}_2^\top \mathbf{A}_1) , \end{aligned} \quad (25)$$

where it is easy to observe that $\mathbf{A}^\top \mathbf{B} = \mathbf{A}_1^\top (\mathbf{U} + \mathbf{A}_2) + \mathbf{A}_2^\top (\mathbf{V} + \mathbf{A}_1) = \mathbf{A}^\top \mathbf{Q} \oplus_q \mathbf{A}_1^\top \mathbf{A}_2 \oplus_q \mathbf{A}_2^\top \mathbf{A}_1$.

Similarly, via exploiting $\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{V} \\ \mathbf{U} \end{bmatrix}$, we can show that

$$\begin{aligned} R_{\text{SW}}^\Sigma &= H(\mathbf{A}^\top \mathbf{B}) + H(\mathbf{Q} | \mathbf{A}^\top \mathbf{B}) \\ & \quad + H(\mathbf{B} | \tilde{\mathbf{Q}}, \tilde{\mathbf{Q}}^\top \mathbf{B} \oplus_q \mathbf{B}_1^\top \mathbf{B}_2 + \mathbf{B}_2^\top \mathbf{B}_1) , \end{aligned} \quad (26)$$

where $\mathbf{A}^\top \mathbf{B} = \tilde{\mathbf{Q}}^\top \mathbf{B} \oplus_q \mathbf{B}_1^\top \mathbf{B}_2 \oplus_q \mathbf{B}_2^\top \mathbf{B}_1$.

From (25) and (26), we note that $H(\mathbf{A} | \mathbf{A}^\top \mathbf{Q}) = H(\mathbf{B} | \tilde{\mathbf{Q}}^\top \mathbf{B})$. Contrasting (24) with (25), the condition

$$H(\mathbf{A}^\top \mathbf{B}) + H(\mathbf{Q} | \mathbf{A}^\top \mathbf{B}) < H(\mathbf{A} | \mathbf{Q}, \mathbf{A}^\top \mathbf{B})$$

$$= H(\mathbf{B} | \tilde{\mathbf{Q}}, \mathbf{A}^\top \mathbf{B})$$

ensures that $R_{\text{KM}}^\Sigma < R_{\text{SW}}^\Sigma$.

When $l = 1$, we have $\mathbf{A}_1^\top \mathbf{A}_2 = \mathbf{A}_2^\top \mathbf{A}_1$ and $\mathbf{B}_1^\top \mathbf{B}_2 = \mathbf{B}_2^\top \mathbf{B}_1$, hence, the above condition is equivalent to

$$\begin{aligned} H(\mathbf{A}^\top \mathbf{B}) + H(\mathbf{Q} | \mathbf{A}^\top \mathbf{B}) &< H(\mathbf{A} | \mathbf{Q}, \mathbf{A}^\top \mathbf{Q}) \\ &= H(\mathbf{B} | \tilde{\mathbf{Q}}, \tilde{\mathbf{Q}}^\top \mathbf{B}). \end{aligned}$$

F. Proof of Proposition 5

Note that $\mathbf{D}_j = \mathbf{A}^\top \mathbf{B}_j$ for $j \in \{1, \dots, l\}$, where $\mathbf{D}_j = [d_{1j} \ d_{2j} \ \dots \ d_{lj}]^\top \in \mathbb{F}_q^{l \times 1}$. Following the steps in Appendix H, the receiver can recover $\{\mathbf{A} \oplus_q \tilde{\mathbf{B}}_j\}_{j=1}^l$, $\{\mathbf{A}^\top \mathbf{A} \oplus_q \tilde{\mathbf{B}}_j^\top \tilde{\mathbf{B}}_j\}_{j=1}^l$, and then compute the following $l \times l$ matrix:

$$\begin{aligned} &(\mathbf{A} \oplus_q \tilde{\mathbf{B}}_j)^\top (\mathbf{A} \oplus_q \tilde{\mathbf{B}}_j) - (\mathbf{A}^\top \mathbf{A} \oplus_q \tilde{\mathbf{B}}_j^\top \tilde{\mathbf{B}}_j) \\ &= \mathbf{A}^\top \tilde{\mathbf{B}}_j \oplus_q \tilde{\mathbf{B}}_j^\top \mathbf{A} \\ &= \begin{bmatrix} d_{1j} \oplus_q d_{1j} & d_{1j} \oplus_q d_{2j} & \dots & d_{1j} \oplus_q d_{lj} \\ d_{2j} \oplus_q d_{1j} & d_{2j} \oplus_q d_{2j} & \dots & d_{2j} \oplus_q d_{lj} \\ \vdots & \vdots & \ddots & \vdots \\ d_{lj} \oplus_q d_{1j} & d_{lj} \oplus_q d_{2j} & \dots & d_{lj} \oplus_q d_{lj} \end{bmatrix}, \end{aligned}$$

which is a symmetric matrix with l unknowns and $\frac{l(l-1)}{2} \geq l$ linearly independent equations for $l \geq 2$ and $q > 2$. Hence, \mathbf{D}_j , for each $j = 1, \dots, l$, as well as \mathbf{D} can be recovered.

G. Proof of Corollary 2

The sum rate for distributed encoding of (\mathbf{A}, \mathbf{B}) is given as

$$\begin{aligned} R_{\text{SW}}^\Sigma &= H(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}_1, \mathbf{B}_1, \mathbf{A}_2, \mathbf{B}_2) \\ &= H(\mathbf{A}_1, \mathbf{B}_1) \\ &= m(h(2\epsilon) + (1 - 2\epsilon) + h(p)). \end{aligned}$$

Exploiting Proposition 5 to compute $\mathbf{A}^\top \mathbf{B}$, we can achieve

$$\begin{aligned} R_{\text{KM}}^\Sigma &= 2H(\mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1, \mathbf{A}^\top \mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1^\top \tilde{\mathbf{B}}_1) \\ &= 2mh \left(2 \left(\frac{1}{2} - \epsilon \right) (1 - p) + 2\epsilon(1 - p), \right. \\ &\quad \left. 2 \left(\frac{1}{2} - \epsilon \right) p + 2\epsilon p \right) \\ &\quad + 2h(\mathbf{A}^\top \mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1^\top \tilde{\mathbf{B}}_1 | \mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1) \\ &\leq 2mh \left(2 \left(\frac{1}{2} - \epsilon \right) (1 - p) + 2\epsilon(1 - p), \right. \\ &\quad \left. 2 \left(\frac{1}{2} - \epsilon \right) p + 2\epsilon p \right) + 2 \log_2(3), \end{aligned}$$

where the last step follows from using that

$$\mathbf{A}^\top \mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1^\top \tilde{\mathbf{B}}_1 = \begin{bmatrix} \mathbf{A}_1^\top \mathbf{A}_1 \oplus_3 \mathbf{B}_1^\top \mathbf{B}_1 & \mathbf{A}_1^\top \mathbf{A}_2 \oplus_3 \mathbf{B}_1^\top \mathbf{B}_1 \\ \mathbf{A}_2^\top \mathbf{A}_1 \oplus_3 \mathbf{B}_1^\top \mathbf{B}_1 & \mathbf{A}_2^\top \mathbf{A}_1 \oplus_3 \mathbf{B}_1^\top \mathbf{B}_1 \end{bmatrix},$$

and evaluating the conditional entropy as

$$\begin{aligned} &h(\mathbf{A}^\top \mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1^\top \tilde{\mathbf{B}}_1 | \mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1) \\ &= h(\mathbf{A}_1^\top \mathbf{A}_1 \oplus_3 \mathbf{B}_1^\top \mathbf{B}_1, \mathbf{A}_1^\top \mathbf{A}_2 \oplus_3 \mathbf{B}_1^\top \mathbf{B}_1, \\ &\quad \mathbf{A}_2^\top \mathbf{A}_1 \oplus_3 \mathbf{B}_1^\top \mathbf{B}_1 | \mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1) \\ &= h(\{ \sum_{i=1}^m a_{ij}^2 \oplus_3 b_{i1}^2 \}_{j=1}^2, \sum_{i=1}^m a_{i1} a_{i2} \oplus_3 b_{i1}^2 | \mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1) \end{aligned}$$

$$\stackrel{(a)}{=} h(\sum_{i=1}^m a_{i1}^2 \oplus_3 b_{i1}^2, \sum_{i=1}^m a_{i2}^2 \oplus_3 b_{i1}^2 | \mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1)$$

$$\stackrel{(b)}{\leq} 2 \log_2(3),$$

where (a) follows from that $a_{i1} a_{i2} \oplus_3 b_{i1}^2$ can be recovered given the side information $\mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1 = \{a_{ij} \oplus_3 b_{i1}\}_{i,j}$ for $i = 1, 2, \dots, m$ and $j = 1, 2$, and given $\{a_{ij}^2 \oplus_3 b_{i1}^2\}_{j=1}^2$, for $j = 1, 2$. More specifically, the receiver can recover

$$2a_{ij} b_{i1} = (a_{ij} \oplus_3 b_{i1})^2 - (a_{ij}^2 \oplus_3 b_{i1}^2), \quad j = 1, 2. \quad (27)$$

Hence, using the side information $\mathbf{A} \oplus_3 \tilde{\mathbf{B}}_1$ and (27), the receiver can recover $a_{i1} a_{i2} \oplus_3 b_{i1}^2$ as follows:

$$\begin{aligned} a_{i1} a_{i2} \oplus_3 b_{i1}^2 &= (a_{i1} \oplus_3 b_{i1})(a_{i2} \oplus_3 b_{i1}) \\ &\quad - (a_{i1} b_{i1} \oplus_3 a_{i2} b_{i1}). \end{aligned}$$

Finally, step (b) follows from exploiting that both variable $\sum_{i=1}^m a_{i1}^2 \oplus_3 b_{i1}^2$ and variable $\sum_{i=1}^m a_{i2}^2 \oplus_3 b_{i1}^2$ reside in \mathbb{F}_3 .

REFERENCES

- [1] G. Strang, *Introduction to linear algebra*. SIAM, 2022.
- [2] M. H. Mousavi, M. A. Maddah-Ali, and M. Mirmohseni, "Private inner product retrieval for distributed machine learning," in *Proc., IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 355–359.
- [3] É. Bouscatié, G. Castagnos, and O. Sanders, "Pattern matching in encrypted stream from inner product encryption," in *Proc., IACR Int. Conf. Public-Key Cryptogr.* Springer, May 2023, pp. 774–801.
- [4] H. Buhrman, R. Cleve, J. Watrous, and R. De Wolf, "Quantum fingerprinting," *Physical Review Letters*, vol. 87, no. 16, p. 167902, Sep. 2001.
- [5] R. Tandon, Q. Lei, A. G. Dimakis, and N. Karampatziakis, "Gradient coding: Avoiding stragglers in distributed learning," in *International Conference on Machine Learning*. PMLR, Jul. 2017, pp. 3368–3376.
- [6] Q. Yu, M. Maddah-Ali, and S. Avestimehr, "Polynomial codes: An optimal design for high-dimensional coded matrix multiplication," in *Proc., Advances in Neural Information Processing Systems*, vol. 30, Long Beach, CA, Dec. 2017, pp. 4403–4413.
- [7] S. Dutta, M. Fahim, F. Haddadpour, H. Jeong, V. Cadambe, and P. Grover, "On the optimal recovery threshold of coded matrix multiplication," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 278–301, Jul. 2019.
- [8] M. Soleymani, H. MahdaviFar, and A. S. Avestimehr, "Analog lagrange coded computing," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 283–295, Feb. 2021.
- [9] Q. Yu, S. Li, N. Raviv, S. M. M. Kalan, M. Soltanolkotabi, and S. A. Avestimehr, "Lagrange coded computing: Optimal design for resiliency, security, and privacy," in *Proc., Int. Conf. Artif. Intell. Stat.* Naha, Okinawa, Japan: PMLR, Apr. 2019, pp. 1215–1225.
- [10] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [11] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources (corresp.)," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 219–221, Mar. 1979.
- [12] D. Krithivasan and S. S. Pradhan, "Distributed source coding using abelian group codes: A new achievable rate-distortion region," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1495–1519, Feb. 2011.
- [13] S. S. Pradhan, A. Padakandla, F. Shirani *et al.*, "An algebraic and probabilistic framework for network information theory," *Found. Trends Commun. Inf. Theory*, vol. 18, no. 2, pp. 173–379, Dec. 2020.
- [14] T. Han and K. Kobayashi, "A dichotomy of functions F(X, Y) of correlated sources (X, Y)," *IEEE Trans. Inf. Theory*, vol. 33, no. 1, pp. 69–76, Jan. 1987.
- [15] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, p. 903–917, Mar. 2001.

- [16] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Proc., 6th Prague Conf. Inf. Theory*, Prague, Czech Republic, Sep. 1973, pp. 411–425.
- [17] S. Feizi and M. Médard, "On network functional compression," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5387–5401, Sep. 2014.
- [18] D. Malak, "Fractional graph coloring for functional compression with side information," in *Proc., IEEE Inf. Theory Wksh.*, Mumbai, India, Nov. 2022, pp. 750–755.
- [19] M. R. Deylam Salehi and D. Malak, "An achievable low complexity encoding scheme for coloring cyclic graphs," in *Proc., Annu. Allerton Conf. Commun. Control Comput.*, Monticello, IL, USA, Sep. 2023, pp. 1–8.
- [20] R. Ahlswede and T. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 396–412, May 1983.
- [21] C. Nair and Y. N. Wang, "On optimal weighted-sum rates for the modulo sum problem," in *Proc., IEEE Int. Symp. Inf. Theory*, Jun. 2020, pp. 2416–2420.
- [22] R. G. Gallager, *Information theory and reliable communication*. Wiley: New York, Jan. 1968, vol. 588.