Quasi Structured Codes for Multi-Terminal Communications

Mohsen Heidari, Farhad Shirani, and S. Sandeep Pradhan, *Senior Member, IEEE*

***Abstract*— A new class of structured codes called quasi group codes (QGCs) is introduced. A QGC is a subset of a group code. In contrast with the group codes, QGCs are not closed under group addition. The parameters of the QGC can be chosen, such that the size of $\mathcal{C}+\mathcal{C}$ is equal to any number between $|\mathcal{C}|$ and $|\mathcal{C}|^2$. We analyze the performance of a specific class of QGCs. This class of QGCs is constructed by assigning single-letter distributions to the indices of the codewords in a group code. Then, the QGC is defined as the set of codewords whose index is in the typical set corresponding to these single-letter distributions. The asymptotic performance limits of this class of QGCs are characterized using single-letter information quantities. Corresponding covering and packing bounds are derived. It is shown that the point-to-point channel capacity and optimal rate-distortion function are achievable using QGCs. Coding strategies based on QGCs are introduced for three fundamental multi-terminal problems: the Körner-Marton problem for modulo prime-power sums, computation over the multiple access channel (MAC), and MAC with distributed states. For each problem, a single-letter achievable rate-region is derived. It is shown, through examples, that the coding strategies improve upon the previous strategies based on the unstructured codes, linear codes, and group codes.**

***Index Terms*— Quasi structured codes, distributed source coding, computation over multiple access channel (MAC), MAC with states, multi-terminal communication.**

I. Introduction

THE conventional technique of deriving the performance limits for any communication problem in information theory is via random coding [1] involving so-called Independent Identically Distributed (IID) random codebooks. Since such a code possesses only single-letter empirical properties, coding techniques are constrained to exploit only these for enabling efficient communication. We refer to them as unstructured codes. These techniques have been proven to achieve capacity for point-to-point (PtP) channels and particular multi-terminal channels such as multiple-access channel (MAC) and degraded broadcast channel. Based on these initial successes, it was widely believed that one can achieve the capacity of any network communication problem using IID codebooks.

Stepping beyond this conventional technique, Körner and Marton [2] proposed a technique based on statistically correlated codebooks (in particular, identical random linear codes) possessing algebraic closure properties, henceforth referred to as (random) structured codes, that outperformed all techniques based on (random) unstructured codes. This technique was proposed for the problem of distributed computation of the modulo two sum of two correlated symmetric binary sources [2]. Applications of structured codes were also studied for various multi-terminal communication systems, including, but not limited to, distributed source coding [3]–[6], computation over MAC [7]–[13], MAC with side information [4], [14]–[17], the joint source-channel coding over MAC [18], multiple-descriptions [19], interference channel [20]–[26], broadcast channel [27] and MAC with Feedback [28]. In these works, algebraic structures are exploited to design new coding schemes which outperform all coding schemes solely based on random unstructured codes. The emerging opinion in this regard is that even if computational complexity is a non-issue, algebraic structured codes may be necessary, in a deeply fundamental way, to achieve optimality in transmission and storage of information in networks.

There are several algebraic structures such as fields, ring and groups. Linear codes are defined over finite fields. The focus of this work is on structured codes defined over the ring of modulo-m integers, that is $\mathbb{Z}_m$. Group codes are a class of structured codes constructed over $\mathbb{Z}_m$, and were first studied by Slepian [29] for the Gaussian channel. A group code over $\mathbb{Z}_m$ is defined as a set of codewords that is closed under the element-wise modulo-m addition. Linear codes are a special case of group codes (the case when m is a prime). There are two main incentives to study group codes. First, linear codes are defined only over finite fields, and finite fields exists only when alphabet sizes equal to a prime power, i.e., $\mathbb{Z}_{p^r}$. Second, there are several communications problems in which group codes have superior performance limits compared to linear codes. As an example, group codes over $\mathbb{Z}_8$ have better error correcting properties than linear codes for communications over an additive white Gaussian noise channel with 8-PSK constellation [30]. As an another example, construction of polar codes over alphabets of size equal to a prime power p^r, is more efficient with a module structure rather than a vector space structure [31]–[34]. Bounds on the achievable rates of group codes in PtP communications were studied in [30], [35]–[39]. Como [38] derived the largest achievable rate using group codes for certain PtP channels. In [35], Ahlswede showed that group codes do not achieve the capacity of a general discrete memoryless channel. In [39], Sahebi, *et al.*, unified the previously known works, and characterized the ensemble of all group codes over finite commutative groups.

Manuscript received August 1, 2017; revised November 14, 2018; accepted June 19, 2019. Date of current version September 13, 2019. This work was supported by NSF under Grant CCF 1717299. This paper was presented in part at the 2016 and 2017 IEEE International Symposium on Information Theory (ISIT).

The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: mohsenhd@umich.edu; fshirani@umich.edu; pradhanv@umich.edu).

Communicated by V. M. Prabhakaran, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2019.2930591

In addition, the authors derived the optimum asymptotic performance limits of group codes for PtP channel/source coding problems.

Körner and Marton suggested the use of identical linear codes for compression of two correlated binary sources when the objective is to reconstruct the modulo-two sum of the sources. However, if the objective is to have the full reconstruction of both the sources at the decoder (Slepian-Wolf setting [40]), one may use independent unstructured binning of the sources using Shannon-style unstructured code ensembles [1]. Similar observations were made regarding the interference channel [20], [26], [41]. In such settings, despite the rate penalties that individual users may pay, the use of structured codes is preferred to achieve a common goal in a network. A selfish user intent on maximizing individual throughput is suggested to adopt Shannon-style unstructured code ensembles. This observation points to a trade-off between cooperation and communication/compression in networks.

A randomly generated codebook $\mathcal{C}$ in Shannon-style ensembles is completely unstructured (complete lack of structure) in the sense that, with high probability, the size of $\mathcal{C}+\mathcal{C}$ nearly *equals the square of the size* of $\mathcal{C}$. A linear code, group code or lattice code $\mathcal{C}$ is completely structured in the sense that the size of $\mathcal{C}+\mathcal{C}$ *equals the size* of $\mathcal{C}$. This gap between completely structured codes and completely unstructured codes leads to the following question: Is there a spectrum of strategies involving partially structured codes or partially unstructured codes that lie between these two extremes? Based on this line of thought, we consider a new class of codes which are not fully closed with respect to any algebraic structure but maintain a degree of "closedness" with respect to some. In our earlier works [9], [10], it was observed that adding a certain set of codewords to a group code improves the performance of the code. Based on these observations,[1] we introduce a new class of structured code ensembles called Quasi Group Codes (QGC) whose *closedness can be controlled*. A QGC is a subset of a group code. The degree of closedness of a QGC can be controlled in the sense that the size of $\mathcal{C}+\mathcal{C}$ can be any number between the size of $\mathcal{C}$ and the square of the size of $\mathcal{C}$. We provide a method for constructing specific subsets of these codes by putting single-letter distributions on the indices of the codewords. We are able to analyze the performance of the resulting code ensemble, and characterize the asymptotic performance using single-letter information quantities. By choosing the single-letter distribution on the indices one can operate anywhere in the spectrum between the two extremes: group codes and unstructured codes.

The contributions of this work are as follows. A new class of codes over groups called Quasi Group Codes (QGC) is introduced. These codes are constructed by taking subsets of group codes. This work considers QGCs over cyclic groups $\mathbb{Z}_{p^r}$. One can use the fundamental theorem of finitely generated Abelian groups to generalize the results of this paper to QGCs over non-cyclic finite Abelian groups. Information-theoretic characterizations for the asymptotic performance limits and properties of QGCs for source coding and channel coding problems are derived in terms of single-letter information quantities. Covering and packing bounds are derived for an ensemble of QGCs. Next, a binning technique for the QGCs is developed by constructing nested QGCs. As a result of these bounds, the PtP channel capacity and optimal rate-distortion function of sources are shown to be achievable using nested QGCs. The applications of QGCs in some multi-terminal communications problems are considered. More specifically our study includes the following problems:

Distributed Source Coding: A more general version of Körner-Marton problem is considered. In this problem, there are two distributed sources taking values from $\mathbb{Z}_{p^r}$. The sources are to be compressed in a distributed fashion. The decoder wishes to compute the modulo p^r-addition of the sources losslessly.

Computation over MAC: In this problem, two transmitters wish to communicate independent information to a receiver over a MAC. The objective is to decode the modulo-p^r sum of the codewords sent by the transmitters at the receiver. This problem is of interest in its own right. Moreover, this problem finds applications as an intermediate step in the study of other fundamental problems such as the interference channel and broadcast channel [27], [42].

MAC with Distributed States: In this problem, two transmitters wish to communicate independent information to a receiver over a MAC. The transition probability between the output and the inputs depends on states S_1, and S_2 corresponding to the two transmitters. The state sequences are generated IID according to some fixed joint probability distribution. Each encoder observes the corresponding state sequence non-causally. The objective of the receiver is to decode the messages of both transmitters.

These problems are formally defined in the sequel. For each problem, a coding scheme based on (nested) QGCs is introduced and a new single-letter achievable rate-region is characterized. It is shown, through examples, that QGCs improve upon coding strategies that are solely based on completely unstructured/structured codes.

The rest of this paper is organized as follows: Section II provides the preliminaries and notations. In Section III, we introduce QGC's and define an ensemble of QGCs. Section IV characterizes basic properties of QGCs. Section V describes a method for binning using QGCs. In Section VI and Section VII, we discuss the applications of QGC's in distributed source coding and computation over MAC, respectively. In Section VIII we investigate applications of nested QGCs in the problem of MAC with states. Finally, Section IX concludes the paper.

II. Preliminaries

A. Notations

We denote (i) vectors using lowercase bold letters such as $\mathbf{b}, \mathbf{u}$, (ii) matrices using uppercase bold letters such as $\mathbf{G}$, (iii) random variables using capital letters such as X, Y, (iv) numbers, realizations of random variables and elements of

[1]The motivation for this work comes from our earlier work on multi-level polar codes based on $\mathbb{Z}_{p^r}$ [32]. A multi-level polar code is not a group code. But it is a subset of a nontrivial group code.

sets using lowercase letters such as a, x. Calligraphic letters such as $\mathcal{C}$ and $\mathcal{U}$ are used to represent sets. For shorthand, we denote the set $\{1, 2, \ldots, m\}$ by $[1 : m]$.

B. Definitions

A group is a set equipped with a binary operation denoted by "+". All groups in this paper are Abelian. Given a prime power p^r, the group of integers modulo-p^r is denoted by $\mathbb{Z}_{p^r}$, where the underlying set is $\{0, 1, \cdots, p^r - 1\}$, and the addition is modulo-p^r addition. Given a group M, a subgroup is a subset H which is closed under the group addition. For $s \in [0 : r]$, define

$$H_s = p^s \mathbb{Z}_{p^r} = \{0, p^s, 2p^s, \cdots, (p^{r-s} - 1)p^s\},$$

and $T_s = \{0, 1, \cdots, p^s - 1\}$. For example, $H_0 = \mathbb{Z}_{p^r}, T_0 = \{0\}$, whereas $H_r = \{0\}, T_r = \mathbb{Z}_{p^r}$. Note, H_s is a subgroup of $\mathbb{Z}_{p^r}$, for $s \in [0 : r]$. Given H_s and T_s, each element a of $\mathbb{Z}_{p^r}$ can be represented uniquely as a sum $a = t + h$, where $h \in H_s$ and $t \in T_s$. We denote such t by $[a]_s$. Note that $[a]_s = a \mod p^s$, for $s \in [0, r]$. Therefore, with this notation, $[\cdot]_s$ is a function from $\mathbb{Z}_{p^r} \to T_s$. Note that this function satisfies the distributive property:

$$[a + b]_s = \Big[[a]_s + [b]_s\Big]_s$$

For any elements $a, b \in \mathbb{Z}_{p^r}$, we define the multiplication $a \cdot b$ by adding a with itself b times. Given a positive integer n, denote $\mathbb{Z}_{p^r}^n = \bigotimes_{i=1}^n \mathbb{Z}_{p^r}$. Note that $\mathbb{Z}_{p^r}^n$ is a group, whose addition is element-wise and its underlying set is $\{0, 1, \ldots, p^r - 1\}^n$. We follow the definition of shifted group codes on $\mathbb{Z}_{p^r}$ as in [39], [3].

Definition 1 (Shifted Group Codes)**.** An (n, k)-*shifted group code* over $\mathbb{Z}_{p^r}$ is defined as

$$\mathcal{C} = \{\mathbf{uG} + \mathbf{b} : \mathbf{u} \in \mathbb{Z}_{p^r}^k\}, \tag{1}$$

where $\mathbf{b} \in \mathbb{Z}_{p^r}^n$ is the translation (dither) vector and $\mathbf{G}$ is a $k \times n$ generator matrix with elements in $\mathbb{Z}_{p^r}$.

We follow the definition of typicality as in [43].

Definition 2. For any probability distribution P on $\mathcal{X}$ and $\epsilon > 0$, a sequence $\mathbf{x}^n \in \mathcal{X}^n$ is said to be ϵ-typical with respect to P if

$$\left|\frac{1}{n} N(a|\mathbf{x}^n) - P(a)\right| \leqslant \frac{\epsilon}{|\mathcal{X}|}, \ \forall a \in \mathcal{X},$$

and, in addition, no $a \in \mathcal{X}$ with $P(a) = 0$ occurs in $\mathbf{x}^n$. Note that $N(a|x^n)$ is the number of the occurrences of a in the sequence $\mathbf{x}^n$. The set of all ϵ-typical sequences with respect to a probability distribution P on $\mathcal{X}$ is denoted by $A_\epsilon^{(n)}(X)$.

The above definition can be extended to define joint typicality with respect to a joint probability distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$. A pair of sequences $(\mathbf{x}^n, \mathbf{y}^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ is said to be jointly ϵ-typical with respect to P_{XY} if

$$\left|\frac{1}{n} N(a, b|\mathbf{x}^n, \mathbf{y}^n) - P_{XY}(a, b)\right| \leqslant \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|}, \ \forall (a, b) \in \mathcal{X} \times \mathcal{Y}$$

such that none of (a, b) with $P_{XY}(a, b) = 0$ occurs in $(\mathbf{x}^n, \mathbf{y}^n)$. The set of all such pairs is denoted by $A_\epsilon^{(n)}(X, Y)$.

III. Quasi Group Codes

Linear codes and group codes are two classes of structured codes. These codes are closed under the addition of the underlying group or field. It is known in the literature that coding schemes based on linear codes and group codes improve upon unstructured random coding strategies [2]. In this section, we propose a new class of structured codes called *quasi-group codes*.

A QGC is defined as a subset of a group code. Therefore, QGCs are not necessarily closed under the addition of the underlying group. An (n, k) shifted group code over $\mathbb{Z}_{p^r}$ is defined as the image of a linear mapping from $\mathbb{Z}_{p^r}^k$ to $\mathbb{Z}_{p^r}^n$ as in Definition 1. Let $\mathcal{U}$ be an arbitrary subset of $\mathbb{Z}_{p^r}^k$. Then a QGC is defined as

$$\mathcal{C} = \{\mathbf{uG} + \mathbf{b} : \mathbf{u} \in \mathcal{U}\}, \tag{2}$$

where $\mathbf{G}$ is a $k \times n$ matrix and $\mathbf{b}$ is an element of $\mathbb{Z}_{p^r}^n$. If $\mathcal{U} = \mathbb{Z}_{p^r}^k$, then $\mathcal{C}$ is a shifted group code. As we will show, by changing the subset $\mathcal{U}$, the code $\mathcal{C}$ ranges from completely structured codes (such as group codes and linear codes) where $|\mathcal{C} + \mathcal{C}| = |\mathcal{C}|$ to completely unstructured codes where $|\mathcal{C} + \mathcal{C}| \approx |\mathcal{C}|^2$. For a general subset $\mathcal{U}$, it is difficult to derive a single-letter characterization of the asymptotic performance of such codes. To address this issue, we present a special type of subsets $\mathcal{U}$ for which single-letter characterization of their performance is possible.

Construction of $\mathcal{U}$: Given a positive integer m, consider m mutually independent random variables $U_1, U_2, \cdots, U_m$. Suppose each U_i takes values from $\mathbb{Z}_{p^r}$ with distribution $P_{U_i}, i \in [1 : m]$. For $\epsilon > 0$, and positive integers k_i, define $\mathcal{U}$ as a Cartesian product of the ϵ-typical sets of $U_i, i \in [1 : m]$. More precisely,

$$\mathcal{U} \triangleq \bigotimes_{i=1}^{m} A_\epsilon^{(k_i)}(U_i). \tag{3}$$

In this construction, set $\mathcal{U}$ is determined by m, k_i, ϵ, and the PMFs $P_{U_i}, i \in [1 : m]$. An example of such construction for $m = 1$ is given in the following.

Example 1. Let U be a random variable over $\mathbb{Z}_{p^r}$ with PMF P_U. For $\epsilon > 0$, let $\mathcal{U}$ to be the set of all ϵ-typical sequences $\mathbf{u}^k$. More precisely, define $\mathcal{U} = A_\epsilon^{(k)}(U)$. In this case, $\mathcal{U}$ is determined by the PMF P_U and ϵ. For instance, if U is uniform over $\mathbb{Z}_{p^r}$, then $\mathcal{U} = \mathbb{Z}_{p^r}^k$.

In what follows, we provide an alternative representation for the construction given in (3). Let $k \triangleq \sum_{i=1}^m k_i$ and denote $q_i \triangleq \frac{k_i}{k}$. With this notation, $q_i, i \in [1, m]$ form a probability distribution; because, $q_i \geqslant 0$ and $\sum_i q_i = 1$. Therefore, we can define a random variable Q with $P(Q = i) = q_i$. Define a random variable U with the conditional distribution

$$P(U = a|Q = i) = P(U_i = a)$$

for all $a \in \mathbb{Z}_{p^r}, i \in [1 : m]$. With this notation the set $\mathcal{U}$ in the above construction is characterized by a finite set $\mathcal{Q}$, a pair of random variables (U, Q) distributed over $\mathbb{Z}_{p^r} \times \mathcal{Q}$, an integer k, and $\epsilon > 0$. The joint distribution of U and Q is denoted by P_{UQ}. Note that we assume $P_Q(q) > 0$ for all $q \in \mathcal{Q}$. For a more concise notation, we identify the set $\mathcal{U}$ without explicitly specifying ϵ. Q can be interpreted as a *time sharing* random variable. It determines the contribution of U_i, measured by $\frac{k_i}{k}$, in the construction of $\mathcal{U}$. With the notation given for the construction of $\mathcal{U}$, we define its corresponding QGC.

Definition 3. An (n, k)- QGC $\mathcal{C}$ over $\mathbb{Z}_{p^r}$ is defined as in (2) and (3), and is characterized by a matrix $\mathbf{G} \in \mathbb{Z}_{p^r}^{k\times n}$, a translation $\mathbf{b} \in \mathbb{Z}_{p^r}^n$, and a pair of random variables (U, Q) distributed over the finite set $\mathbb{Z}_{p^r} \times \mathcal{Q}$. The set $\mathcal{U}$ in (3) is defined as the index set of $\mathcal{C}$.

Remark 1. Any shifted group code over $\mathbb{Z}_{p^r}$ is a QGC.

Remark 2. Let $\mathcal{C}$ be a random (n, k)-QGC constructed by selecting the elements of its generator matrix and translation vector randomly independently with uniform distribution from $\mathbb{Z}_{p^r}, r > 1$. In contrast to linear codes, codewords of $\mathcal{C}$ are not necessarily pairwise independent.

Information theoretic analysis of coding strategies are usually carried out by constructing ensembles of randomly generated codebooks [1], [44]. Following the same approach, we construct ensembles of QGCs with different blocklengths.

Fix positive integers (n, k) and random variables (U, Q). We create an ensemble of codes by taking the collection of all (n, k)-QGCs with random variables (U, Q), for all matrices $\mathbf{G}$ and translations $\mathbf{b}$. A random codebook $\mathcal{C}$ from this ensemble is chosen by selecting the elements of $\mathbf{G}$ and $\mathbf{b}$ randomly and uniformly from $\mathbb{Z}_{p^r}$. In order to characterize the asymptotic performance limits of QGCs, we need to define sequences of ensembles of QGCs. For any positive integer n, let $k_n = cn$, where $c > 0$ is a constant. Consider the sequence of the ensembles of (n, k_n)-QGCs with random variables (U, Q). In the next two lemmas, we characterize the size of randomly selected codebooks from these ensembles. The first lemma shows that the index set $\mathcal{U}$ for an ensemble of QGCs approximately equals to $2^{kH(U|Q)}$.

Lemma 1. *Let $\mathcal{U}_n$ be the index set associated with the ensemble of (n, k_n)-QGCs with random variables (U, Q) and $\epsilon > 0$, where $k_n = cn$ for a constant $c > 0$. Then there exists $N > 0$, such that for all $n > N$,*

$$\left|\frac{1}{k_n}\log_2|\mathcal{U}_n| - H(U|Q)\right| \leqslant \epsilon',$$

where ϵ' is a continuous function of ϵ, and $\epsilon' \to 0$ as $\epsilon \to 0$.

Proof: The proof is given in Appendix A-A ■

Remark 3. As an immediate consequence of Lemma 1, we provide an upper-bound on the size of a QGC. For that, let $\mathcal{C}_n$ be an (n, k_n)-QGC with random variables (U, Q). Then, for large enough n,

$$\frac{1}{n}\log_2|\mathcal{C}_n| \leqslant \frac{k_n}{n}H(U|Q) + \epsilon'. \tag{4}$$

To explain inequality (4), note that a codebook $\mathcal{C}_n$ is the image of the index set $\mathcal{U}_n$ under the mapping

$$\Phi_n(\mathbf{u}) = \mathbf{u}\mathbf{G}_n + \mathbf{b}^n.$$

Therefore, the bound in (4) is due to the fact that Φ_n is, in general, a many-to-one mapping. In the case of linear codes $(r = 1)$, it is assumed that $k < n$. In this case, for sufficiently large n, Φ_n is injective with high probability. This implies that the size of a random linear code approximately equals $\approx 2^k$. Consequently, $\frac{k}{n}$ is a relevant measure for the rate of a (k, n) linear code. However, for a QGC (general $r \geqslant 2$), even if $k \geqslant n$, under certain conditions, Φ_n is "almost" injective with high probability. In what follows, we characterize these conditions. We begin by defining α-injectivity.

Definition 4. A mapping $\phi : \mathcal{U} \to \mathcal{X}$, defined on finite sets $(\mathcal{U}, \mathcal{X})$, is said to be α-injective, if there exists a subset $\mathcal{A} \subseteq \mathcal{U}$ with cardinality at least $\alpha|\mathcal{U}|$ such that restriction of ϕ to $\mathcal{A}$ is injective.

By the above definition, any 1-injective map is one-to-one. The next lemma shows that under particular conditions on (U, Q) and for sufficiently large n, the mapping Φ_n is α-injective with high probability, where $\alpha \approx 1$.

Lemma 2. *Let $\mathcal{U}_n$ be the index set associated with the ensemble of (n, k_n)-QGCs with random variables (U, Q), where $k_n = cn$ for a constant $c > 0$. Define a map*

$$\Phi_n : \mathcal{U}_n \to \mathbb{Z}_{p^r}^n,$$

$\Phi_n(\mathbf{u}) = \mathbf{u}\mathbf{G}_n$ for all $\mathbf{u} \in \mathcal{U}_n$, where $\mathbf{G}_n$ is a $k_n \times n$ matrix whose elements are chosen randomly and uniformly from $\mathbb{Z}_{p^r}$. Suppose

$$H(U|[U]_s, Q) \leqslant \frac{1}{c}(r - s)\log_2 p - \epsilon,$$

for all $s \in [0 : r - 1]$. Then, for any $\gamma, \delta > 0$ and sufficiently large n, the mapping Φ_n is $(1 - \delta)$-injective with probability at least $(1 - \gamma)$.[2]

Proof: The proof is provided in Appendix A-B. ■

As a result, under the conditions given in Lemma 2, the rate of a random codebook selected from ensemble of (n, k)-QGCs with random variables (U, Q) approximately equals $R \approx \frac{k}{n}H(U|Q)$, with high probability. The condition in Lemma 2 can viewed as a restriction on the size of the index set, that is

$$\frac{k}{n}H(U|[U]_s, Q) \leqslant (r - s)\log_2 p - \epsilon, \quad 0 \leqslant s \leqslant r - 1. \tag{5}$$

We refer to this condition as the *injectivity* condition.

IV. Properties of Quasi Group Codes

It is known that if $\mathcal{C}$ is a random unstructured codebook, then $|\mathcal{C} + \mathcal{C}| \approx |\mathcal{C}|^2$ with high probability. Group codes on the other hand are closed under the addition, which means $|\mathcal{C} + \mathcal{C}| = |\mathcal{C}|$. Comparing to unstructured codes, when the structure of the group codes matches with that of a multi-terminal channel/source coding problem, it turns out that higher/lower transmission rates are obtained. However, in certain problems, the structure of the group codes is too restrictive. More precisely, when the underlying group is $\mathbb{Z}_{p^r}$

[2]Note that the map Φ_n in the lemma does not have any translation, i.e., $\mathbf{b} = 0$. It is sufficient to prove the lemma for $\mathbf{b} = 0$. This is due to the fact that if Φ_n is $(1 - \delta)$-injective, then so is $\Phi_n + \mathbf{b}$, for any translation $\mathbf{b}$.

for $r \geqslant 2$, there are several nontrivial subgroups. These subgroups cause a penalty on the rate of a group code. This results in lower transmission rates in channel coding and higher transmission rates in source coding.

Quasi group codes balance the trade-off between the structure of the group codes and that of the unstructured codes. More precisely, when $\mathcal{C}$ is a QGC, then $|\mathcal{C}+\mathcal{C}|$ is a number between $|\mathcal{C}|$ and $|\mathcal{C}|^2$. This results in a more flexible algebraic structure to match better with the structure of the channel or source. This trade-off is shown more precisely in the following lemma.

Lemma 3. *Let $\mathcal{C}_i, i=1,2$ be an (n,k_i)-QGC over $\mathbb{Z}_{p^r}$ with random variables (U_i, Q). Suppose, $P_{U_1,U_2,Q}$ is such that the Markov chain $U_1 \leftrightarrow Q \leftrightarrow U_2$ holds and that the injectivity condition in* (5) *is satisfied for (U_1,Q) and (U_2,Q).*

1) *Suppose $k_1=k_2=k$, and the generator matrices of $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{D}$ are identical. Let $\mathcal{D}$ be an (n,k)-QGC with random variables (U_1+U_2, Q) and the same generator matrix as for $\mathcal{C}_1$ and $\mathcal{C}_2$. Suppose $\mathbf{U}_i$ is selected randomly and uniformly from the index set (see Definition 3) of $\mathcal{C}_i, i=1,2$. Let $\mathbf{X}_i$ be the codeword of $\mathcal{C}_i$ corresponding to $\mathbf{U}_i, i=1,2$. Then, for all $\epsilon>0$ and sufficiently large n,*
$$P\{\mathbf{X}_1+\mathbf{X}_2 \in \mathcal{D}\} \geqslant 1-\delta(\epsilon),$$
where $\delta(\epsilon)\to 0$ as $\epsilon \to 0$.
2) *$\mathcal{C}_1+\mathcal{C}_2$ is an (n,k_1+k_2)-QGC with random variables $(U_I,(Q,I))$, where $I\in\{1,2\}$. If $I=i$, then $U_I=U_i$, $i=1,2$. In addition, the joint PMF of these random variables is given by*
$$P(I=i, Q=q,\, U_I=a) = \frac{k_i}{k_1+k_2} P(Q=q)P(U_i=a|Q=q), \tag{6}$$
for all $a\in\mathbb{Z}_{p^r}, q\in\mathcal{Q}$ and $i=1,2$.

Proof: Suppose $\mathcal{U}_i$ is the index set, $\mathbf{G}_i$ is the matrix, and $\mathbf{b}_i$ is the translation of $\mathcal{C}_i, i=1,2$.

We prove the first statement for the case when time sharing random variable Q is trivial. The proof for general Q follows from similar steps. If Q is trivial, the index sets satisfy $\mathcal{U}_i = A_\epsilon^{(k)}(U_i), i=1,2$. Since $k_1=k_2$ and $\mathbf{G}_1=\mathbf{G}_2$, then
$$\mathbf{X}_i = \mathbf{U}_i\mathbf{G}+\mathbf{b}_i, \quad i=1,2.$$
With this notation, $\mathbf{X}_1+\mathbf{X}_2 = (\mathbf{U}_1+\mathbf{U}_2)\mathbf{G}+\mathbf{b}_1+\mathbf{b}_2$. From Lemma 10, with probability at least $1-2^{-n\epsilon/p^r}$, we have $(\mathbf{U}_1,\mathbf{U}_2)\in A_{\delta(\epsilon)}^{(k)}(U_1,U_2)$, where δ is a function as in Lemma 10. Therefore, $\mathbf{U}_1+\mathbf{U}_2 \in A_{\delta(\epsilon)}^{(k)}(U_1+U_2)$ with probability at least $1-2^{-n\epsilon/p^r}$. The proof is complete by noting that the index set of $\mathcal{D}$ is defined as $\mathcal{U}_d \triangleq A_{\delta(\epsilon)}^{(k)}(U_1+U_2)$.

For the second statement, we have
$$\mathcal{C}_1+\mathcal{C}_2 = \{[\mathbf{u}_1,\mathbf{u}_2]\begin{bmatrix}\mathbf{G}_1\\ \mathbf{G}_2\end{bmatrix}+\mathbf{b}_1+\mathbf{b}_2 : \mathbf{u}_i\in\mathcal{U}_i, i=1,2\}.$$
Therefore, $\mathcal{C}_1+\mathcal{C}_2$ is an (n,k_1+k_2)-QGC. Note that $\mathcal{U}_1\times\mathcal{U}_2$ is the index set associated with this codebook. The statement follows, since each subset $\mathcal{U}_i, i=1,2$ is a Cartesian product of ϵ-typical sets of $U_{i,q}, q\in\mathcal{Q}$. The random variables $(U_I,(Q,I))$ describes such a Cartesian product. ∎

We explain the intuition behind the lemma. Suppose $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{D}$ are QGCs with identical generator matrices and with random variables U_1, U_2 and U_1+U_2, respectively. Then $\mathcal{D} = \mathcal{C}_1+\mathcal{C}_2$ with probability approaching one.

Remark 4. If $\mathcal{C}_1$ and $\mathcal{C}_2$ are the QGCs as in Lemma 3, then from standard counting arguments we have
$$\max\{|\mathcal{C}_1|,|\mathcal{C}_2|\} \leqslant |\mathcal{C}_1+\mathcal{C}_2| \leqslant \min\{p^{rn}, |\mathcal{C}_1|\cdot|\mathcal{C}_2|\}$$

In what follows, we derive a packing bound and a covering bound for a QGC with matrices and translation chosen randomly and uniformly. Fix a PMF P_{XY}, and suppose an ϵ-typical sequence $\mathbf{y}$ is given with respect to the marginal distribution P_Y. Consider the set of all codewords in a QGC that are jointly typical with $\mathbf{y}$ with respect to P_{XY}. In the packing lemma, we characterize the conditions under which the probability of this set is small. This implies the existence of a "good-channel" code which is also a QGC. In the covering lemma, we derive the conditions for which, with high probability, there exists at least one such codeword in a QGC. In this case a "good-source" code exists which is also a QGC. These conditions are provided in the next two lemmas.

For any positive integer n, let $k_n=cn$, where $c>0$ is a constant. Let $\mathcal{C}_n$ be a sequence of (n,k_n)-QGCs with random variables (U,Q), $\epsilon>0$. By R_n denote the rate of $\mathcal{C}_n$. Suppose the elements of the generator matrix and the translation of $\mathcal{C}_n$ are chosen randomly and uniformly from $\mathbb{Z}_{p^r}$.

Lemma 4 (Packing)**.** *Let $(X,Y)\sim P_{XY}$. By $\mathbf{c}_n(\theta)$ denote the θth codeword of $\mathcal{C}_n$. Let $\tilde{\mathbf{Y}}^n$ be a random sequence distributed according to $\prod_{i=1}^n P_{Y|X}(\tilde{y}_i|c_{n,i}(\theta))$. Suppose, conditioned on $\mathbf{c}_n(\theta)$, $\tilde{\mathbf{Y}}^n$ is independent of all other codewords in $\mathcal{C}_n$. Then, for any $\theta\in[1:|\mathcal{C}_n|]$, and $\delta>0$, $\exists N>0$ such that for all $n>N$,*
$$P\{\exists \mathbf{x}\in\mathcal{C}_n : (\mathbf{x},\tilde{\mathbf{Y}}^n)\in A_\epsilon^{(n)}(X,Y), \mathbf{x}\neq\mathbf{c}_n(\theta)\} < \delta,$$
if the following bounds hold
$$R_n < \min_{0\leqslant s\leqslant r-1} \frac{H(U|Q)}{H(U|Q,[U]_s)}\Big(\log_2 p^{r-s} - H(X|Y,[X]_s)+\eta(\epsilon)\Big), \tag{7}$$
where $\eta(\epsilon)\to 0$ as $\epsilon\to 0$.

Proof: See Appendix B. ∎

Lemma 5 (Covering)**.** *Let $(X,\hat{X})\sim P_{X\hat{X}}$, where $\hat{X}$ takes values from $\mathbb{Z}_{p^r}$. Let $\mathbf{X}^n$ be a random sequence distributed according to $\prod_{i=1}^n P_X(x_i)$. Then, for any $\delta>0$, $\exists N>0$ such that for all $n>N$,*
$$P\{\exists \hat{\mathbf{x}}\in\mathcal{C}_n : (\mathbf{X}^n,\hat{\mathbf{x}})\in A_\epsilon^{(n)}(X,\hat{X})\} > 1-\delta$$
if the following inequalities hold
$$R_n > \max_{1\leqslant s\leqslant r}\frac{H(U|Q)}{H([U]_s|Q)}\big(\log_2 p^s - H([\hat{X}]_s|X)+\eta(\epsilon)\big). \tag{8}$$

Proof: See Appendix C. ∎

Remark 5. The covering and packing bounds for the special case $r = 1$ are simplified to

$$\text{Packing: } R_n < \log_2 p - H(X|Y),$$
$$\text{Covering: } R_n > \log_2 \ p - H(\hat{X}|X).$$

Lemma 3, 4 and Lemma 5 provide a tool to derive inner bounds for achievable rates using quasi group codes in multi-terminal channel coding and source coding problems.

V. Binning Using QGC

Note that in a randomly generated QGC, all codewords have uniform distribution over $\mathbb{Z}_{p^r}^n$. However, in many communication setups we require application of codes with non-uniform distributions. In addition, we require binning techniques for various multi-terminal communications. In this section, we present a method for random binning of QGCs. In the next sections, we will use random binning of QGCs to propose coding schemes for various multi-terminal problems.

We introduce nested quasi group codes using which we propose a random binning technique. A QGC $\mathcal{C}_I$ is said to be nested in a QGC $\mathcal{C}_O$, if $\mathcal{C}_I \subset \mathcal{C}_O + \mathbf{b}$, for some translation $\mathbf{b}$. Suppose $\mathcal{C}_O$ is an $(n, k+l)$-QGC with the following structure,

$$\mathcal{C}_O \triangleq \{\mathbf{uG} + \mathbf{v}\tilde{\mathbf{G}} + \mathbf{b} : \mathbf{u} \in \mathcal{U}, \mathbf{v} \in \mathcal{V}\}, \tag{9}$$

where $\mathcal{U}$ and $\mathcal{V}$ are subsets of $\mathbb{Z}_{p^r}^k$, and $\mathbb{Z}_{p^r}^l$, respectively. Define the inner-code as

$$\mathcal{C}_I \triangleq \{\mathbf{uG} + \mathbf{b} : \mathbf{u} \in \mathcal{U}\}.$$

By Definition 3, $\mathcal{C}_I$ is an (n, k)-QGC. In addition, there exists $\mathbf{a} \in \mathbb{Z}_{p^r}^n$ such that $\mathcal{C}_I \subset \mathcal{C}_O + \mathbf{a}$. The pair $(\mathcal{C}_I, \mathcal{C}_O)$ is called a nested QGC. For any fixed element $\mathbf{v} \in \mathcal{V}$, we define its corresponding bin as the set

$$\mathcal{B}(\mathbf{v}) \triangleq \{\mathbf{uG} + \mathbf{v}\tilde{\mathbf{G}} + \mathbf{b} : \mathbf{u} \in \mathcal{U}\}. \tag{10}$$

Definition 5. An (n, k, l)-nested QGC is defined as a pair $(\mathcal{C}_I, \mathcal{C}_O)$, where $\mathcal{C}_I$ is an (n, k)-QGC, and

$$\mathcal{C}_O = \{\mathbf{x}_I + \bar{\mathbf{x}} : \mathbf{x}_I \in \mathcal{C}_I, \bar{\mathbf{x}} \in \bar{\mathcal{C}}\},$$

where $\bar{\mathcal{C}}$ is an (n, l)-QGC. Let the random variables corresponding to $\mathcal{C}_I$ and $\bar{\mathcal{C}}$ are (U, Q) and (V, Q), respectively. $\mathcal{C}_I, \mathcal{C}_O$ and $\bar{\mathcal{C}}$ are called the inner, the outer and the shift codes, respectively. Then, $\mathcal{C}_O$ is characterized by (U, V, Q).

In a nested QGC both the outer-code and the inner-code are themselves QGCs. More precisely we have the following remark.

Remark 6. Let $(\mathcal{C}_I, \mathcal{C}_O)$ be an (n, k_1, k_2)-nested QGC with random variables (U_1, U_2, Q). Suppose the joint distribution among (U_1, U_2, Q) is the one that satisfies the Markov chain $U_1 \leftrightarrow Q \leftrightarrow U_2$. Then by Lemma 3 $\mathcal{C}_O$ is an $(n, k_1 + k_2)$-QGC with random variables $(U_I, (Q, I))$, where I is a random index variable taking values in $\{1, 2\}$, and the joint PMF of the random variables (U_I, Q, I) is given by (6).

Note that with equation (10), $\mathcal{B}(\mathbf{v}) = \mathcal{C}_I + \mathbf{v}\tilde{\mathbf{G}}$. As a result, each bin is a shifted version of the inner-code. Thus, each bin in an (n, k, l)-nested QGC is also an (n, k)-QGC.

Remark 7. Suppose $(\mathcal{C}_I, \mathcal{C}_O)$ is an (n, k_1, k_2)-nested QGC with random matrices and translations. Assume the injectivity condition (5) holds for C_I and C_O. By R_O and R_I denote the rates of $\mathcal{C}_O$ and $\mathcal{C}_I$, respectively. Let ρ denote the binning rate (the rate of $\bar{\mathcal{C}}$ as in Definition 5). Using Remark 6 and 3, for large enough n, with probability close to one, $|R_O - R_I - \rho| \leqslant o(\epsilon)$.

Intuitively, as a result of this remark, $R_O \approx R_I + \rho$. Furthermore, since the injectivity condition holds, then with probability close to one, we obtain

$$R_O \approx \frac{k}{n} H(U|Q) + \frac{l}{n} H(V|Q),$$
$$R_I \approx \frac{k}{n} H(U|Q),$$
$$\rho \approx \frac{l}{n} H(V|Q).$$

This implies that the bins $\mathcal{B}(\mathbf{v})$ corresponding to different $\mathbf{v} \in \bar{\mathcal{C}}$ are "almost disjoint". In this method for binning, since both the inner-code and the outer-code are QGCs, the structure of the inner-code, bins and the outer-code can be determined using the PMFs of the related random variables (that is U, V and Q as in Definition 5).

We established a set of lemmas (Lemma 1- 5) that are used to derive achievable rates for coding strategies based on QGCs. In the following, we introduce a coding strategy using QGCs and show the achievability of the Shannon performance limits for PtP channel and source coding problem. For that, we first provide a set of definitions to model PtP channel and source coding problem.

Channel Model: A discrete memoryless channel is characterized by the triple $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, where the two finite sets $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively, and $P_{Y|X}$ is the channel transition probability matrix.

Definition 6. An (n, Θ)-code for a channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ is a pair of mappings (e, f) where $e : [1 : \Theta] \rightarrow \mathcal{X}^n$ and $f : \mathcal{Y}^n \rightarrow [1 : \Theta]$.

Definition 7. For a given channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, a rate R is said to be achievable if for any $\epsilon > 0$ and for all sufficiently large n, there exists an (n, Θ)-code such that :

$$\frac{1}{\Theta} \sum_{i=1}^{\Theta} P_{Y|X}^n (f(Y^n) \neq i | X^n = e(i)) < \epsilon, \quad \frac{1}{n} \log \Theta > R - \epsilon.$$

The channel capacity is defined as the supremum of all achievable rates.

Source Model: A discrete memoryless source is a tuple $(\mathcal{X}, \hat{\mathcal{X}}, P_X, d)$, where the two finite sets $\mathcal{X}$ and $\hat{\mathcal{X}}$ are the source and reconstruction alphabets, respectively, P_X is the source probability distribution, and $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}^+$ is the (bounded) distortion function.

Definition 8. An (n, Θ)-code for a source $(\mathcal{X}, \hat{\mathcal{X}}, P_X, d)$ is a pair of mappings (e, f) where

$$f : \mathcal{X}^n \rightarrow [1 : \Theta]$$

and

$$e : [1 : \Theta] \rightarrow \hat{\mathcal{X}}^n.$$

Definition 9. For a given source $(\mathcal{X}, \hat{\mathcal{X}}, P_X, d)$, a rate-distortion pair (R, D) is said to be achievable if for any $\epsilon > 0$ and for all sufficiently large n, there exists an (n, Θ)-code such that :

$$\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i) < D + \epsilon, \quad \frac{1}{n}\log\Theta < R + \epsilon,$$

where $\hat{X}^n = e(f(X^n))$. The optimal rate-distortion region is defined as the set of all achievable rate-distortion pairs.

Definition 10. An (n, Θ)-code is said to be based on nested QGCs, if there exists an (n, k, l)-nested QGC with random variables (U, V, Q) such that a) $\Theta = |\mathcal{V}|$, where $\mathcal{V}$ is the index set associated with the codebook $\bar{\mathcal{C}}$ (see Definition 5), b) for any $\mathbf{v} \in \mathcal{V}$, the output of the mapping $e(\mathbf{v})$ is in $\mathcal{B}(\mathbf{v})$, where $\mathcal{B}(\mathbf{v})$ is the bin associated with $\mathbf{v}$, and is defined as in (10).

Definition 11. For a channel, a rate R is said to be achievable using nested QGCs if for any $\epsilon > 0$ and all sufficiently large n, there exists an (n, Θ)-code based on nested QGCs such that:

$$\frac{1}{\Theta}\sum_{i=1}^{\Theta} P(f(Y^n) \neq i | X^n = e(i)) < \epsilon, \quad \frac{1}{n}\log\Theta > R - \epsilon.$$

For a source, a rate-distortion pair (R, D) is said to be achievable using nested QGCs, if for any $\epsilon > 0$ and for all sufficiently large n, there exists an (n, Θ)-code based on nested QGCs such that:

$$\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i) < D + \epsilon, \quad \frac{1}{n}\log\Theta < R + \epsilon,$$

where $\hat{X}^n = e(f(X^n))$.

Theorem 1. *The PtP channel capacity and the optimal rate-distortion region of sources are achievable using nested QGCs.*

In what follows, we introduce an achievable scheme using nested QGCs and provide an outline of the proof for the theorem.

Channel coding using QGCs : Consider a memoryless channel with input alphabet $\mathcal{X}$ and conditional distribution $P_{Y|X}$. Let the prime power p^r be such that $|\mathcal{X}| \leqslant p^r$. Fix a PMF P_X on $\mathcal{X}$, and set $l = nR$, where R will be determined later. Let $(\mathcal{C}_I, \mathcal{C}_O)$ be an (n, k, l)-nested QGC with random variables (U, V, Q). Let Q be a trivial random variable, and U and V be independent with uniform distribution over $\{0, 1\}$. The elements of the generator matrix and the translation used for the nested QGC are drawn randomly and uniformly from $\mathbb{Z}_{p^r}$. Let R_I and R_O denote the rate of the inner-code $\mathcal{C}_I$ and the outer-code $\mathcal{C}_O$, respectively. According to Remark 7, with probability close to one, $R_O \approx R_I + R$ and the binning rate approximately equals to $\frac{l}{n}H(V) = R$.

Suppose the messages are drawn randomly and uniformly from $\{0, 1\}^l$. Upon receiving a message $\mathbf{v}$, the encoder first calculates its bin, that is $\mathcal{B}(\mathbf{v})$. Then it finds $\mathbf{x} \in \mathcal{B}(\mathbf{v})$ such that $\mathbf{x} \in A_\epsilon^{(n)}(X)$. If $\mathbf{x}$ was found, it is transmitted to the channel. Otherwise, an encoding error is declared. Upon receiving $\mathbf{y}$ from the channel, the decoder finds all $\tilde{\mathbf{c}} \in \mathcal{C}_O$ such that

$$(\tilde{\mathbf{c}}, \mathbf{y}) \in A_\epsilon^{(n)}(X, Y).$$

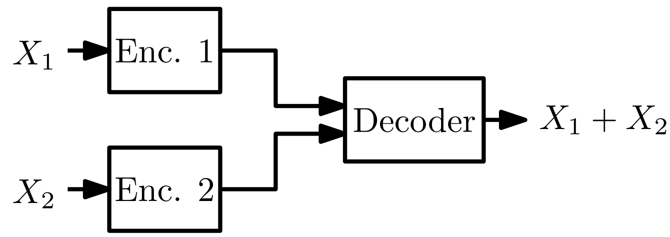

Fig. 1. An example for the problem of distributed source coding. In this setup, the sources X_1 and X_2 take values from $\mathbb{Z}_{p^r}$. The decoder reconstructs $X_1 + X_2$ losslessly.

Then, the decoder lists the bin number for any of such $\tilde{\mathbf{c}}$. If the bin number is unique, it is declared as the decoded message. Otherwise, an encoding error will be declared.

The effective transmission of the above coding strategy equals the binning rate, i.e., R. Using the covering lemma (Lemma 5), the probability of the error at the encoder approaches zero, if

$$R_I \geqslant \log p^r - H(X).$$

Using the packing lemma (Lemma 4), the probability of error at the decoder approaches zero, if

$$R_O \leqslant \log p^r - H(X|Y).$$

As a result, the effective transmission rate $R \leqslant I(X; Y)$ is achievable.

Source coding using QGCs : We use the same nested QGC constructed for the channel coding problem. Given a distortion level D, consider a random variable $\hat{X}$ such that $\mathbb{E}\{d(X, \hat{X})\} \leqslant D$. Let $\mathbf{x}$ be a typical sequence from the source. The encoder finds a codeword $\mathbf{c} \in \mathcal{C}_O$ such that $(\mathbf{x}, \mathbf{c})$ is jointly ϵ-typical with respect to $P_X P_{\hat{X}|X}$. If no such $\mathbf{c}$ was found, an encoding error will be declared. Otherwise, the encoder sends the bin index $\mathbf{v}$ for which $\mathbf{c} \in \mathcal{B}(\mathbf{v})$. Given $\mathbf{v}$, the decoder finds $\tilde{\mathbf{c}} \in \mathcal{B}(\mathbf{v})$ such that $\tilde{\mathbf{c}}$ is ϵ-typical with respect to $P_{\hat{X}}$. An error occurs, if no unique codeword $\tilde{\mathbf{c}}$ was found.

Note that with high probability the effective transmission rate approximately equals to R. Using Lemma 5, the encoding error approaches zero, if

$$R_O \geqslant \log p^r - H(\hat{X}|X).$$

Using Lemma 4, the decoding error approaches zero, if

$$R_I \leqslant \log p^r - H(\hat{X}).$$

As a result the rate $R \geqslant I(X; \hat{X})$ and distortion D is achievable.

VI. DISTRIBUTED SOURCE CODING

In this section, we consider a distributed source coding problem described as follows. Suppose X_1 and X_2 are sources with alphabet $\mathbb{Z}_{p^r}$ and with joint PMF $P_{X_1 X_2}$. The jth encoder compresses X_j and sends it to a central decoder. The decoder wishes to reconstruct $X_1 + X_2$ losslessly, where the addition is modulo-p^r. Figure 1 depicts the diagram of this setup.

It is assumed that n IID copies of the sources are made available at the encoders, where n is called the blocklength. In what follows, we define the encoding and decoding processes and formulate the problem setup.

Definition 12. An (n, Θ_1, Θ_2)-code consists of two encoding functions

$$f_i : \mathbb{Z}_{p^r}^n \to \{1, 2, \cdots, \Theta_i\}, \quad i = 1, 2,$$

and a decoding function

$$g : \{1, 2, \cdots, \Theta_1\} \times \{1, 2, \cdots, \Theta_2\} \to \mathbb{Z}_{p^r}^n$$

Definition 13. Given a pair of sources $(X_1, X_2) \sim P_{X_1X_2}$ with values over $\mathbb{Z}_{p^r} \times \mathbb{Z}_{p^r}$, a pair (R_1, R_2) is said to be achievable if for any $\epsilon > 0$ and sufficiently large n, there exists an (n, Θ_1, Θ_2)-code such that,

$$\frac{1}{n} \log_2 M_i < R_i + \epsilon, \quad \text{for} \quad i = 1, 2,$$

and

$$P\{\mathbf{X_1}^n + \mathbf{X_2}^n \neq g(f_1(\mathbf{X_1}^n), f_2(\mathbf{X_2}^n))\} \leqslant \epsilon.$$

For this problem, we adopt nested QGCs and propose a new coding scheme. The following theorem presents an achievable rate region for the defined setup.

Theorem 2. *For a pair of sources $(X_1, X_2) \sim P_{X_1X_2}$ with values from $\mathbb{Z}_{p^r}$, lossless reconstruction of the modulo-p^r sum $X_1 + X_2$ is possible with transmission rate-pair (R_1, R_2), if there exist random variables (W_1, W_2, Q) such that the following bound holds*

$$R_i \geqslant \log_2 p^r - \min_{0 \leqslant s \leqslant r-1} \frac{H(W_i|Q)}{H(W_1 + W_2|[W_1 + W_2]_s, Q)} \Big(\log_2 p^{(r-s)} - H(X_1 + X_2|[X_1 + X_2]_s) \Big), \quad (11)$$

where $i = 1, 2$, (W_1, W_2) take values from $\mathbb{Z}_{p^r}$, the Markov chain $W_1 - Q - W_2$ holds, and the injectivity condition (5) *is satisfied for each pair (W_1, Q) and (W_2, Q). In addition, $|\mathcal{Q}| \leqslant r$ is sufficient to achieve the above bounds.*

Proof: See Appendix D. ■

Remark 8. The intuition for the rate-region can be briefly explained as follows. Each source is encoded using a nested QGC. The source covering task constrains the rate of the outer code. The packing task induced by the need to recover the sum $(X_1 + X_2)$ at the decoder constrains the rate of the inner code. The overall rates of transmission is given by the difference between these two rates.

Every linear code and group code is a QGC. Therefore, the achievable rate region given in Theorem 2 subsumes the one achieved using linear codes or group codes with jointly typical encoding/decoding techniques. We show, through the following example, that the inclusion is strict.

Example 2. Consider a distributed source coding problem in which X_1 and X_2 are sources over $\mathbb{Z}_4$ and lossless reconstruction of $X_1 \oplus_4 X_2$ is required at the decoder. Assume X_1 is uniform over $\mathbb{Z}_4$. X_2 is related to X_1 via the equation $X_2 = N - X_1$, where N is a random variable which is independent of X_1. The distribution of N is presented in Table I.

TABLE I
DISTRIBUTION OF N

N	0	1	2	3
P_N	0.06	0.54	0.04	0.36

Using random unstructured codes, the rates (R_1, R_2) such that

$$R_1 + R_2 \geqslant H(X_1, X_2)$$

are achievable [40]. It is also possible to use linear codes for the reconstruction of $X_1 \oplus_4 X_2$. For that, the decoder first reconstructs the modulo-7 sum of X_1 and X_2, then from $X_1 \oplus_7 X_2$ the modulo-4 sum is retrieved. This is because linear codes are built only over finite fields, and $\mathbb{Z}_7$ is the smallest field in which the modulo-4 addition can be embedded. Therefore, the rates

$$R_1 = R_2 \geqslant H(X_1 \oplus_7 X_2)$$

is achievable using linear codes over the field $\mathbb{Z}_7$ [2]. As is shown in [39], group codes in this example outperform linear codes. The largest achievable region using group codes is described by all rate pair (R_1, R_2) such that

$$R_i \geqslant \max\{H(Z), 2\ H(Z|[Z]_1)\}, \ i = 1, 2,$$

where $Z = X_1 \oplus_4 X_2$. It is shown in [9] that using transversal group codes the rates (R_1, R_2) such that

$$R_i \geqslant \max\{H(Z), 1/2\ H(Z) + H(Z|[Z]_1)\}$$

are achievable. An achievable rate region using nested QGC's can be obtained from Theorem 2. Let Q be a trivial random variable and set

$$P(W_1 = 0) = P(W_2 = 0) = 0.95$$

and

$$P(W_1 = 1) = P(W_2 = 1) = 0.05.$$

As a result one can verify that the following is achievable:

$$R_j \geqslant 2 - \min\{0.6(2 - H(Z)), 5.7(2 - 2H(Z|[Z]_1)\}.$$

Note that the factors 0.6 and 5.7 are determined by the specific choice of the probability distribution on (W_1, Q) and (W_2, Q). Different factor are obtained by changing the probability distributions. We compare the achievable rates of these schemes. The result are presented in Table II.

TABLE II
ACHIEVABLE SUM-RATE USING DIFFERENT CODING SCHEMES FOR EXAMPLE 2. NOTE THAT $Z \triangleq X_1 \oplus_4 X_2$

Scheme	Achievable Rate	
Unstructured Codes	$H(X_1, X_2)$	3.44
Linear Codes	$H(X_1 \oplus_7 X_2)$	4.12
Group Codes	$\max\{H(Z), 2H(Z\|[Z]_1)\}$	3.88
QGCs	$2 - \min\{0.6(2 - H(Z)), 5.7(2 - 2H(Z\|[Z]_1)\}$	3.34

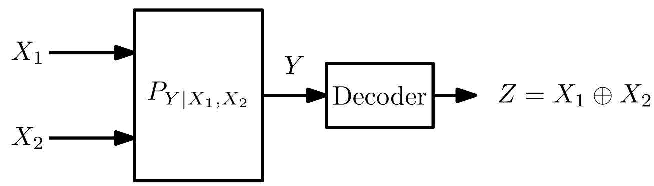

Fig. 2. An example for the problem of computation over MAC. The channel input alphabets belong to $\mathbb{Z}_{p^r}$. The receiver decodes $X_1 + X_2$ which is the modulo-p^r sum of the inputs of the MAC.

VII. Computation Over MAC

In this section, we consider the problem of computation over MAC. Figure 2 depicts an example of this problem. In this setup X_1 and X_2 are the channel's inputs, and take values from $\mathbb{Z}_{p^r}$. Two distributed encoders map their messages to X_1^n and X_2^n. Upon receiving the channel output the decoder wishes to decode $X_1^n + X_2^n$ losslessly. The definition of a code for computation over MAC, and an achievable rate are given in Definition 15 and 16, respectively. Applications of this problem are found in various multi-user communication setups such as interference and broadcast channels.

Definition 14. A two-user MAC is a tuple $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, P_{Y|X_1X_2})$, where the finite sets $\mathcal{X}_1, \mathcal{X}_2$ are the inputs alphabets, $\mathcal{Y}$ is the output alphabet, and $P_{Y|X_1\ X_2}$ is the channel transition probability matrix. Without loss of generality, it is assumed that $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{Z}_{p^r}$, for a prime-power p^r.

Definition 15 (Codes for computation over MAC)**.** An (n, Θ_1, Θ_2)-code for computation over a MAC $(\mathbb{Z}_{p^r}, \mathbb{Z}_{p^r}, \mathcal{Y}, P_{Y|X_1X_2})$ consists of two encoding functions and one decoding function $f_i : [1 : \Theta_i] \to \mathbb{Z}_{p^r}^n$, for $i = 1, 2$, and $g : \mathcal{Y}^n \to \mathbb{Z}_{p^r}^n$, respectively.

Definition 16 (Achievable Rate)**.** (R_1, R_2) is said to be achievable, if for any $\epsilon > 0$, there exists for all sufficiently large n an (n, Θ_1, Θ_2)-code such that

$$P\{g(Y^n) \neq f_1(M_1) + f_2(M_2)\} \leqslant \epsilon,$$
$$R_i - \epsilon \leqslant \frac{1}{n} \log \Theta_i,$$
$$H(M_i|f_i(M_i)) \leqslant \epsilon, \quad i = 1, 2,$$

where M_1 and M_2 are independent random variables and $P(M_i = m_i) = \frac{1}{\Theta_i}$ for all $m_i \in [1 : \Theta_i], i = 1, 2$.

For the above setup, we use QGCs to derive an achievable rate region.

Theorem 3. *Given a MAC $(\mathbb{Z}_{p^r}, \mathbb{Z}_{p^r}, \mathcal{Y}, P_{Y|X_1X_2})$, rate-pair (R_1, R_2) is achievable according to Definition 16, if there exist random variables $(Q, X_1, X_2, V_1, V_2, W_1, W_2)$ such that the following bounds hold*

$$R_i \leqslant \min_{0 \leqslant s \leqslant r} \frac{H(V_i|Q)}{H(V|[V]_s, Q)} \Big(\log_2\ p^{r-s} - H(X|Y, [X]_s) - \max_{\substack{1 \leqslant t \leqslant r \\ j=0,1}} \frac{H(W|[W]_s, Q)}{H([W_j]_t|Q)} \left(\log_2 p^t - H([X_j]_t)\right) \Big)$$

where $i = 1, 2$, (V_1, V_2, W_1, W_2) take values from $\mathbb{Z}_{p^r}$, and $W = W_1 + W_2, V = V_1 + V_2, X = X_1 + X_2$. Moreover, the injectivity condition (5) *is satisfied for each pair $(W_1, Q), (W_2, Q), (V_1, Q)$, and (V_2, Q) and the joint PMF of all the random variables factors as*

$$P_{QX_1X_2V_1V_2W_1W_2Y} = P_{X_1}P_{X_2}P_Q P_{Y|X_1X_2} \prod_{i=1}^{2} P_{V_i|Q}P_{W_i|Q}.$$

Remark 9. The cardinality bound $|\mathcal{Q}| \leqslant r^2$ is sufficient to achieve the rate region in the theorem.

Proof: See Appendix E. ∎

Corollary 1. *A special case of the theorem is when X_1 and X_2 are distributed uniformly over $\mathbb{Z}_{p^r}$. In this case, the following is achievable*

$$R_i \leqslant \min_{0 \leqslant s \leqslant r} \frac{H(V_i|Q)}{H(V_1+V_2|[V_1+V_2]_s, Q)} I(X_1 + X_2; Y|[X_1 + X_2]_s), \tag{12}$$

where $i = 1, 2$.

We show, through the following example, that QGC outperforms the previously known schemes.

Example 3. Consider the MAC described by $Y = X_1 + X_2 + N$, where X_1 and X_2 are the channel inputs with alphabet $\mathbb{Z}_4$. N is independent of X_1 and X_2 with the distribution given in Table I.

Using standard unstructured codes the rate pair (R_1, R_2) satisfying

$$R_1 + R_2 \leqslant I(X_1X_2; Y)$$

are achievable. Note that the modulo-4 addition can be embedded in a larger field such as $\mathbb{Z}_7$. For that linear codes over $\mathbb{Z}_7$ can be used. In this case, the following rates are achievable:

$$R_1 = R_2 = \max_{P_{X_1}P_{X_2}: X_1, X_2 \in \mathbb{Z}_4} \min\left\{H(X_1), H(X_2)\right\} - H(X_1 \oplus_7\ X_2|Y),$$

where the maximization is taken over all probability distribution $P_{X_1}P_{X_2}$ on $\mathbb{Z}_7 \times \mathbb{Z}_7$ such that $P(X_i \in \mathbb{Z}_4) = 1, , i = 1, 2$. This is because, $\mathbb{Z}_4$ is the input alphabet of the channel.

It is shown in [39] that the largest achievable region using group codes is

$$R_i \leqslant \min\{I(Z; Y), 2I(Z; Y|[Z]_1)\},$$

where $Z = X_1 + X_2$ and X_1 and X_2 are uniform over $\mathbb{Z}_4$. Using Corollary 1, QGC's achieve

$$R_i \leqslant \min\{0.6\ I(Z; Y), 5.7\ I(Z; Y|[Z]_1)\}.$$

This can be verified by checking (12) when Q is a trivial random variable,

$$P(V_1 = 0) = P(V_2 = 0) = 0.95$$

and

$$P(V_1 = 1) = P(V_2 = 1) = 0.05.$$

TABLE III

ACHIEVABLE RATES USING DIFFERENT CODING SCHEMES FOR EXAMPLE 3. NOTE THAT $Z \triangleq X_1 + X_2$

Scheme	Achievable Rate ($R_1 = R_2$)	
Unstructured Codes	$I(X_1X_2;Y)/2$	0.28
Linear codes	$\min\{H(X_1), H(X_2)\} - H(X_1 \oplus_7 X_2 \vert Y)$	0.079
Group Codes	$\min\{I(Z;Y), 2I(Z;Y\vert[Z]_1)\}$	0.06
QGCs	$\min\{0.6I(Z;Y), 5.7I(Z;Y\vert[Z]_1)\}$	0.33

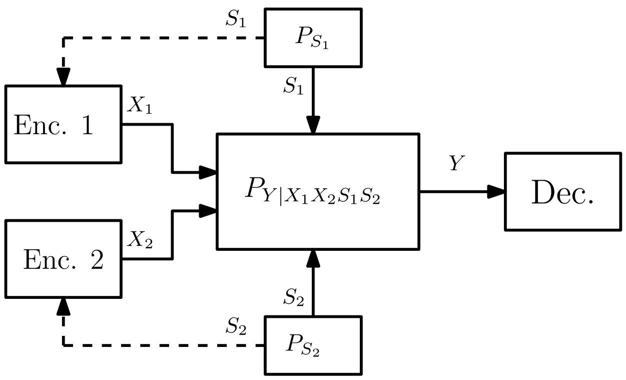

Fig. 3. A two-user MAC with distributed states. The states (S_1, S_2) are generated randomly according to $P_{S_1S_2}$. The entire sequence of each state S_i is available non-casually at the ith transmitter, where $i = 1, 2$.

Note that the factors 0.6 and 5.7 are determined by the specific choice of the probability distribution on (W_1, Q) and (W_2, Q). Different factors can be obtained by changing the probability distributions. We compare the achievable rates of these schemes for the explained setup. The result are presented in Table III.

VIII. MAC WITH STATES

A. Model

Consider a two-user discrete memoryless MAC with input alphabets $\mathcal{X}_1, \mathcal{X}_2$, and output alphabet $\mathcal{Y}$. The transition probabilities between the input and the output of the channel depends on a random vector (S_1, S_2) which is called state. Figure 3 demonstrates such setup. Each state S_i takes values from a set $\mathcal{S}_i$, where $i = 1, 2$. The sequence of the states is generated randomly according to the probability distribution $\prod_{i=1}^n P_{S_1S_2}$. The entire sequence of the state S_i is known at the ith transmitter, $i = 1, 2$, non-causally. The conditional distribution of Y given the inputs and the state is $P_{Y|X_1X_2S_1S_2}$. Each input X_i is associated with a state dependent cost function $c_i : \mathcal{X}_i \times \mathcal{S}_i \to [0, +\infty)$.[3] The cost associated with the sequences x_i^n and s_i^n is given by

$$\bar{c}_i(x_i^n, s_i^n) = \frac{1}{n}\sum_{j=1}^n c_i(x_{ij}, s_{ij}).$$

Definition 17. An (n, Θ_1, Θ_2)-code for reliable communication over a given two-user MAC with states is defined by two encoding functions

$$f_i : \{1, 2, \ldots, \Theta_i\} \times \mathcal{S}_i^n \to \mathcal{Y}^n, \quad i = 1, 2,$$

and a decoding function

$$g : \mathcal{Y}^n \to \{1, 2, \ldots, \Theta_1\} \times \{1, 2, \ldots, \Theta_2\}.$$

Definition 18. For a given MAC with state, the rate-cost tuple $(R_1, R_2, \tau_1, \tau_2)$ is said to be achievable, if for any $\epsilon > 0$, and for all large enough n there exists an (n, Θ_1, Θ_2)-code such that

$$P\{g(Y^n) \neq (M_1, M_2)\} \leqslant \epsilon, \quad \frac{1}{n}\log\Theta_i \geqslant R_i - \epsilon,$$

and

$$\mathbb{E}\{\bar{c}_i(f_i(M_i), S_i^n)\} \leqslant \tau_i + \epsilon,$$

for $i = 1, 2$, where a) M_1, M_2 are independent random variables with distribution $P(M_i = m_i) = \frac{1}{\Theta_i}$ for all $m_i \in [1 : \Theta_i]$, b) (M_1, M_2) is independent of the states (S_1, S_2). Given τ_1, τ_2, the capacity region $\mathcal{C}_{\tau_1,\tau_2}$ is defined as the set of all rates (R_1, R_2) such that the rate-cost $(R_1, R_2, \tau_1, \tau_2)$ is achievable.

B. Achievable Rates

We propose a structured coding scheme that builds upon QGC. Then we present the single-letter characterization of the achievable region of this coding scheme. Using this binning method, a rate region is given in the following theorem.

Theorem 4. *For a given MAC $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, P_{Y|X_1X_2})$ with independent states (S_1, S_2) and cost functions c_1, c_2 the following rates are achievable using nested-QGC*

$$R_1 + R_2 \leqslant r\log_2 p - H(Z_1 + Z_2|Y, Q) - \max_{\substack{i=1,2\\ 1\leqslant t\leqslant r}} \left\{\frac{H(V_1+V_2|Q)}{H([V_i]_t|Q)}\Big(\log_2\ p^t - H([Z_i]_t|Q, S_i)\Big)\right\},$$

where the joint distribution of the above random variables factors as

$$P_{S_1S_2}P_Q P_{Y|X_1X_2}\prod_{i=1,2}P_{V_i|Q}P_{Z_i|QS_i}P_{X_i|QZ_iS_i}.$$

Proof: Let $\mathcal{C}_{I,j}$ be an (n, k)-QGC with matrix $\mathbf{G}_j$, translation $\mathbf{b}_j$, and random variables (W_j, Q), where W_j is uniform over $\{0, 1\}$, and $j = 1, 2$. Denote $\mathcal{W}_1$ and $\mathcal{W}_2$ as the index sets associated with $\mathcal{C}_{I,1}$ and $\mathcal{C}_{I,1}$, as in (2). Let $\bar{\mathcal{C}}_1, \bar{\mathcal{C}}_2$ and $\bar{\mathcal{D}}$ be three (n, l) QGC with identical matrices $\bar{\mathbf{G}}$ and identical translations $\bar{\mathbf{b}}$. Suppose (V_j, Q) are the random variables associated with $\bar{\mathcal{C}}_j$, where $j = 1, 2$. Furthermore, let $(V_1 + V_2, Q)$ is the random variable associated with $\bar{\mathcal{D}}$. Suppose that the elements of all the matrices and the translations are selected randomly and uniformly from $\mathbb{Z}_{p^r}$. Rate of $\bar{\mathcal{C}}_i$ is denoted by ρ_i, rate of $\bar{\mathcal{D}}$ is denoted by ρ, and that of $\mathcal{C}_{I,i}$ is $R_i, i = 1, 2$. For each, sequence $\mathbf{z}_i$ and $\mathbf{s}_i$, generate a sequence $\mathbf{x}_i$ randomly with IID distribution according to $P^n_{X_i|Z_iS_i}, i = 1, 2$. Denote such sequence by $x_i(\mathbf{s}_i, \mathbf{z}_i)$.

Codebook Construction: For each encoder we use a nested QGC. For the first encoder, we use the (n, k, l)-nested QGC

[3] We use a cost function for this problem because, in many cases without a cost function the problem has a trivial solution.

generated by $\mathcal{C}_{I,1}$ and $\bar{\mathcal{C}}_1$. For the second encoder, we use the (n,k,l)-nested QGC characterized by $\mathcal{C}_{I,2}$ and $\bar{\mathcal{C}}_2$. The codebook used in the decoder is $\mathcal{C}_{I,1}+\mathcal{C}_{I,2}+\bar{\mathcal{D}}$. By Lemma 3, this codebook is an $(n,2k+l)$-QGC. In addition, the rate of such code is $R_1+R_2+\rho$

Encoding: For $i=1,2$, the ith encoder is given a message θ_i, and an state sequence $\mathbf{s}_i$. The encoder first calculates the bin associated with θ_i. Then it finds a codeword $\mathbf{z}_i$ in that bin such $(\mathbf{z}_i,\mathbf{s}_i)$ are jointly ϵ-typical with respect to $P_{Z_iS_i}$. If no such sequence was found, the error event E_i will be declared. The encoder calculates $\mathbf{x}_i(\mathbf{s}_i,\mathbf{z}_i)$, and sends it through the channel. Define the event E_c as the event in which $(\mathbf{Z}_1,\mathbf{Z}_2,\mathbf{s}_1,\mathbf{s}_2)$ are not jointly ϵ'- typical with respect to the joint distribution $P_{Z_1Z_2S_1S_2}$.

Decoding: The decoder receives y^n from the channel. Then it finds $\tilde{\mathbf{w}}_1\in\mathcal{W}_1$, $\tilde{\mathbf{w}}_2\in\mathcal{W}_2$, and $\tilde{\mathbf{v}}\in A_\epsilon^{(n)}(V_1+V_2)$ such that the corresponding codeword defined as

$$\tilde{\mathbf{z}}=\tilde{\mathbf{w}}_1\mathbf{G}_1+\tilde{\mathbf{w}}_2\mathbf{G}_2+\tilde{\mathbf{v}}\bar{\mathbf{G}}+\mathbf{b}_1+\mathbf{b}_2+\bar{\mathbf{b}}$$

is jointly $\tilde{\epsilon}$-typical with $\mathbf{Y}$ with respect to $P_{Z_1+Z_2,Y}$. If $\tilde{\mathbf{w}}_1,\tilde{\mathbf{w}}_2$ are unique, then they are considered as the decoded messages. Otherwise an error event E_d will be declared.

Error Analysis: We use Lemma 5 for E_1 and E_2. For that in the covering bound given in (8) set $R=\rho_i, U=V_i, Q=\bar{Q}, \hat{X}=X_i$, and $X=S_i$, where $i=1,2$. As a result, $P(E_1)$ and $P(E_2)$ approaches zero as $n\to\infty$, if the covering bound holds:

$$\rho_i>\max_{1\leqslant t\leqslant r}\frac{H(V_i|\bar{Q})}{H([V_i]_t|\bar{Q})}(\log_2\ p^t-H([Z]_t|S_i)).$$

Note that by Remark 3, $\rho_i\leqslant\frac{l}{n}H(V_i|\bar{Q})+\delta(\epsilon)$. Thus, the above bound gives the following bound

$$\frac{l}{n}H([V_i]_t|\bar{Q})>\log_2 p^t-H([Z]_t|S_i),\qquad(13)$$

where $1\leqslant t\leqslant r,\ i=1,2$.

Analysis of $E_c\bigcap E_1^c\bigcap E_2^c$: Define the set

$$\mathcal{E}_{\mathbf{s}_1,\mathbf{s}_2}\triangleq\Big\{(\mathbf{z}_1,\mathbf{z}_2)\in\mathbb{Z}_{p^r}^n\times\mathbb{Z}_{p^r}^n:(\mathbf{z}_i,\mathbf{s}_i)\in A_\epsilon^{(n)}(Z_i,S_i),$$
$$(\mathbf{z}_1,\mathbf{z}_2,\mathbf{s}_1,\mathbf{s}_2)\notin A_\epsilon^{(n)}(Z_1,Z_2,S_1,S_2), i=1,2\Big\}.$$

Therefore, probability of $E_c\bigcap E_1^c\bigcap E_2^c$ can be written as

$$P(E_c\bigcap E_1^c\bigcap E_2^c)=\sum_{(\mathbf{s}_1,\mathbf{s}_2)\in A_\epsilon^{(n)}(S_1,S_2)}P^n_{S_1,S_2}(\mathbf{s}_1,\mathbf{s}_2)\sum_{(\mathbf{z}_1,\mathbf{z}_2)\in\mathcal{E}_{\mathbf{s}_1,\mathbf{s}_2}}$$
$$P(e_1(\Theta_1,\mathbf{s}_1)=\mathbf{x}_1,e_2(\Theta_2,\mathbf{s}_2)=\mathbf{x}_2),$$

where e_i is the output of the ith encoder, and Θ_i is the random message to be transmitted by encoder i, where $i=1,2$. To bound $P(E_c\bigcap E_1^c\bigcap E_2^c)$, we use a similar argument as in the proof of Theorem 3. We can show that, $\mathbb{E}\{P(E_c\bigcap E_1^c\bigcap E_2^c)\}\to 0$ as $n\to\infty$.

Analysis of $E_d\bigcap(E_c\bigcup E_1\bigcup E_2)^c$:

Next, we use Lemma 4 to provide an upper-bound on $P(E_d\bigcap(E_c\bigcup E_1\bigcup E_2)^c)$. Conditioned on $E_1^c\bigcap E_2^c$, the event E_d is the same as the event of interest in Lemma 4. Set $\mathcal{C}_n=\mathcal{C}_{I,1}+\mathcal{C}_{I,2}+\bar{\mathcal{D}}$, and $R=R_1+R_2+\rho$. It can be shown that $P(E_d\bigcap(E_c\bigcup E_1\bigcup E_2)^c)$ approaches zero, if the packing bound in (7) holds. Since W_i is uniform over $\{0,1\}$, then $H(W_i|Q,[W_i]_t)=0$ for all $t>0$. Therefore, the packing bound is simplified to

$$R_1+R_2+\rho\leqslant\log_2 p^r-H(Z_1+Z_2|Y).\qquad(14)$$

Note that $\rho\leqslant\frac{l}{n}H(V_1+V_2|Q)$. Therefore, if the bound

$$R_1+R_2\leqslant\log_2 p^r-H(Z_1+Z_2|Y)-\frac{l}{n}H(V_1+V_2|Q),\qquad(15)$$

holds on R_1+R_2, then (14) holds too. Using (13), we establish a lower bound on $\frac{l}{n}H(V_1+V_2|Q)$. We have

$$\frac{l}{n}H(V_1+V_2|Q)>\frac{H(V_1+V_2|Q)}{H([V_i]_t|\bar{Q})}\left(\log_2 p^t-H([Z]_t|S_i)\right),\qquad(16)$$

where $1\leqslant t\leqslant r,\ i=1,2$. Then combining (15) and (16) gives the following:

$$R_1+R_2\leqslant\log_2 p^r-H(Z_1+Z_2|Y)$$
$$-\frac{H(V_1+V_2|Q)}{H([V_i]_t|\bar{Q})}\left(\log_2\ p^t-H([Z]_t|S_i)\right).$$

Since these bounds hold for $i=1,2$, and $1\leqslant t\leqslant r$, we get the bound in the theorem. ■

Lemma 6. *The rate region given in Theorem 4 contains the achievable rate region using group codes and linear codes. For that let $V_i, i=1,2$ be distributed uniformly over $\mathbb{Z}_{p^r}$. Therefore, we get the bound*

$$R_1+R_2\leqslant\min_{\substack{i=1,2\\1\leqslant t\leqslant r}}\{\frac{r}{t}H([Z_i]_t|QS_i)\}-H(Z_1+Z_2|YQ).$$

Jafar [45] used the Gel'fand-Pinsker approach for the point-to-point channel coding with states, and proposed a coding scheme using unstructured random codes. Using this scheme a single-letter and computable rate region is characterized.

Definition 19. For a MAC $(\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y},P_{Y|X_1X_2})$ with states (S_1,S_2) and cost functions c_1,c_2, define $\mathscr{R}_{GP}$ as

$$\max\Big\{I(U_1,U_2;Y|Q)-I(U_1;S_1|Q)-I(U_2;S_2|Q)\Big\},\qquad(17)$$

where the maximization is taken over all joint probability distributions $P_{S_1S_2QU_1U_2X_1X_2Y}$ satisfying $\mathbb{E}\{c_i(X_i,S_i)\}\leqslant\tau_i$ for $i=1,2$, and factoring as

$$P_QP_{S_1S_2}P_{Y|X_1X_2}\prod_{i=1,2}P_{U_iX_i|S_iQ}.$$

The collection of all such PMFs $P_{S_1S_2QU_1U_2X_1X_2Y}$ is denoted by $\mathscr{P}_{GP}$.

To the best of our knowledge, $\mathscr{R}_{GP}$ is the current largest achievable rate region using unstructured codes for the problem of MAC with states [45].

C. An Example

We present a MAC with state setup for which $\mathscr{R}_{GP}$ is strictly contained in the region characterized in Theorem 4.

Example 4. Consider a noiseless MAC given in the following

$$Y = X_1 \oplus_4 \ S_1 \oplus_4 \ X_2 \oplus_4 \ S_2,$$

where X_1, X_2 are the inputs, Y is the output, and S_1, S_2 are the states. All the random variables take values from $\mathbb{Z}_4$. The states S_1 and S_2 are mutually independent, and are distributed uniformly over $\mathbb{Z}_4$. The cost function at the first encoder is defined as

$$c_1(x) \triangleq \begin{cases} 1 & \text{if } \ x \in \{1,3\} \\ 0 & \text{otherwise,} \end{cases}$$

whereas, for the second encoder the cost function is

$$c_2(x) \triangleq \begin{cases} 1 & \text{if } \ x \in \{2,3\} \\ 0 & \text{otherwise.} \end{cases}$$

We are interested in satisfying the cost constraints $\mathbb{E}\{c_1(X_1)\} = \mathbb{E}\{c_2(X_2)\} = 0$. This implies that, with probability one, $X_1 \in \{0,2\}$, and $X_2 \in \{0,1\}$.

Lemma 7. *For the setup in Example 4, an outer-bound for $\mathscr{R}_{GP}$ is the set of all rate pairs (R_1, R_2) such that $R_1+R_2 < 1$.*

Proof: See Appendix F. ■

Using numerical analysis, we can provide a tighter bound on the sum-rate which is $R_1 + R_2 \leqslant 0.32$. However, the bound in Lemma 7 is sufficient for the purpose of this paper.

Corollary 2. *For the MAC with states problem in Example 4, the rate pairs (R_1, R_2) satisfying $R_1 + R_2 = 1$ is achievable.*

Proof: The proof follows using Theorem 4 with appropriately selected distributions $P_{V_i|Q}$, $P_{Z_i|QS_i}$, and $P_{X_i|QZ_iS_i}$ for $i = 1, 2$. For that, let Q be a trivial random variable and (V_1, V_2) be IID random variables uniform distribution over $\{0, 1\}$. Conditioned on S_1, the distributions of Z_1 is given by

$$P_{Z_1|S_1}(z_1|s_1) \triangleq \begin{cases} 1/2 & \text{if } z_1 = -s_1, \text{or } z_1 = -s_1 + 2 \\ 0 & \text{otherwise,} \end{cases}$$

The distribution of Z_2 conditioned on S_2 is

$$P_{Z_2|S_2}(z_2|s_2) \triangleq \begin{cases} 1/2 & \text{if } z_2 = s_2, \text{or } z_2 = s_2 + 1 \\ 0 & \text{otherwise,} \end{cases}$$

The conditional distributions of X_i given (S_i, Z_i), $i = 1, 2$, are governed by the relation $X_i = Z_i \ominus S_i$, $i = 1, 2$. As a result, $X_1 \in \{0, 2\}$, and $X_2 \in \{0, 1\}$, with probability one. Hence, the cost constraints for (c_1, c_2) are satisfied. Therefore, for the defined distributions, the sum-rate given in the Theorem is simplified to $R_1+R_2 \leqslant 1$. As a result the sum-rate $R_1+R_2 = 1$ is achievable. ■

IX. Conclusion

A new class of structured codes called Quasi Group Codes was introduced, and basic properties and performance limits of such codes were investigate. The asymptotic performance limits of QGCs was characterized using single-letter information quantities. The PtP channel capacity and optimal rate-distortion function are achievable using QGCs. Coding strategies based on QGCs were studied for three multi-terminal problems: the Körner-Marton problem for modulo prime-power sums, computation over MAC, and MAC with States. For each problem, a coding scheme based on (nested) QGCs was introduced, and a single-letter achievable rate-region was derived. The results show that the coding scheme improves upon coding strategies based on unstructured codes, linear codes and group codes.

Appendix A

A. Proof of Lemma 1

Proof: Using (3) we get

$$\mathcal{U}_n = \bigotimes_{q\in\mathcal{Q}} A_\epsilon^{(k_{q,n})}(U_q),$$

where $k_{q,n} = P_Q(q)k_n$, and the distribution of U_q is the same as the conditional distribution of U given $Q = q$. Using well-known results on the size of ϵ-typical sets we can provide a bound on $|A_\epsilon^{(k_{q,n})}(U_q)|$. More precisely, there exists N_q such that for all $k_{q,n} > cN_q$, we have

$$\left|\frac{1}{k_{q,n}} \log_2 |A_\epsilon^{(k_{q,n})}(U_q)| - H(U_q)\right| \leqslant 2\epsilon'_q,$$

where using the same argument as in [43]

$$\epsilon'_q = -\frac{\epsilon}{p^r} \sum_{a\in\mathbb{Z}_{p^r}, P(U_q=a)>0} \log_2 \ P(U_q = a).$$

Therefore,

$$\begin{aligned}\frac{1}{k_n}\log_2|\mathcal{U}_n| &= \frac{1}{k_n}\sum_{q\in\mathcal{Q}}\log_2|A_\epsilon^{(k_{q,n})}(U_q)| \\ &\leqslant \sum_{q\in\mathcal{Q}}\frac{k_{q,n}}{k_n}(H(U_q)+2\epsilon'_q) \\ &\overset{(a)}{=} H(U|Q) + \sum_{q\in\mathcal{Q}} P_Q(q)2\epsilon'_q \leqslant H(U|Q)+\epsilon',\end{aligned}$$

where $\epsilon' \triangleq 2\max_{q\in Q}\epsilon'_q$. Note that (a) holds as $P_Q(q) = k_{q,n}/k_n$. Using a similar argument we can show that

$$\frac{1}{k_n}\log_2|\mathcal{U}_n| \geqslant H(U|Q) - \epsilon'.$$

Finally, by setting $N = \max_q N_q$, and combining the bounds on $\frac{1}{k_n}\log_2|\mathcal{U}_n|$ the proof is completed. ■

B. Proof of Lemma 2

Proof: For any $\mathbf{u} \in \mathcal{U}_n$, define

$$\theta(\mathbf{u}) \triangleq \sum_{\substack{\mathbf{u}'\in\mathcal{U}_n \\ \mathbf{u}'\neq\mathbf{u}}} \mathbb{1}\{\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})\}.$$

Note that $\theta(\mathbf{u})$ is the number of vectors $\mathbf{u}' \in \mathcal{U}_n$ that have the same output as for $\mathbf{u}$, i.e., $\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})$. Let

$$\mathcal{A} \triangleq \{\mathbf{u}\in\mathcal{U}_n : \ \theta(\mathbf{u}) = 0\}.$$

Note that $\mathcal{A}$ is a subset over which Φ_n is injective. We show that $|\mathcal{A}^c| \leqslant \delta\ |\mathcal{U}_n|$ with high probability. Using Markov inequality:

$$\mathbb{P}\{|\mathcal{A}^c| \geqslant \delta|\mathcal{U}_n|\} \leqslant \frac{\mathbb{E}[|\mathcal{A}^c|]}{\delta|\mathcal{U}_n|},$$

where the expectation is taken with respect to the distribution on random mapping Φ_n. Note that

$$|\mathcal{A}^c| = \sum_{u\in\mathcal{U}_n} \mathbb{1}\{\theta(u) > 0\} \leqslant \sum_{u\in\mathcal{U}_n} \theta(u)$$

Hence,

$$\mathbb{P}\{|\mathcal{A}^c| \geqslant \delta|\mathcal{U}_n|\} \leqslant \frac{1}{\delta|\mathcal{U}_n|} \sum_{u\in\mathcal{U}_n} \mathbb{E}[\theta(u)]. \tag{18}$$

By definition, $\mathbb{E}[\theta(u)] = \sum_{\mathbf{u}'\neq\mathbf{u}} \mathbb{P}\{\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})\}$. We provide an upper bound on $\mathbb{E}[\theta(u)]$.

Let $H_s = p^s\mathbb{Z}_{p^r}$ be a subgroup of $\mathbb{Z}_{p^r}$, where $s \in [0 : r-1]$. If $a \in \mathbb{Z}_{p^r} - \{0\}$, then there exists a maximum $s \in [0 : r-1]$ such that $a \in H_s$. That is $a \in H_s$ and $a \notin H_t$ for all $t > s$. As a result, for any $\mathbf{u}' \in \mathcal{U}_n$ there are r cases for the maximum s such that $u - u' \in H_s^{k_n}$. Considering these cases, we obtain

$$\sum_{\substack{\mathbf{u}'\in\mathcal{U}_n\\ \mathbf{u}'\neq\mathbf{u}}} P\{\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})\} = \sum_{s=0}^{r-1} \sum_{\substack{\mathbf{u}'\in\mathcal{U}_n\\ \mathbf{u}'-\mathbf{u}\in H_s^{k_n}\setminus H_{s+1}^{k_n}}} P\{\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})\} \tag{19}$$

Since Φ_n is a linear map, we have

$$P\{\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})\} = P\{\Phi_n(\mathbf{u}' - \mathbf{u}) = 0\}.$$

Next, we use Lemma 11 (see Appendix H). Since

$$\mathbf{u}' - \mathbf{u} \in H_s^{k_n}\setminus H_{s+1}^{k_n},$$

then

$$P\{\Phi_n(\mathbf{u}' - \mathbf{u}) = 0\} = p^{-n(r-s)}.$$

Therefore, using (19) and the expression for $\mathbb{E}[\theta(u)]$, we get

$$\mathbb{E}[\theta(u)] \leqslant \sum_{s=0}^{r-1} \sum_{\substack{\mathbf{u}'\in\mathcal{U}_n\\ \mathbf{u}'-\mathbf{u}\in H_s^{k_n}}} p^{-n(r-s)} \tag{20}$$

Next, we replace the summation over $\mathbf{u}'$ with the size of the set $\mathcal{U}_n \bigcap(\mathbf{u}+H_s^{k_n})$. Since $\mathcal{U}_n$ is a Cartesian product of typical sets, we use Lemma 12 (see Appendix H) to obtain the following bound

$$|\mathcal{U}_n \bigcap(\mathbf{u} + H_s^{k_n})| \leqslant \prod_q 2^{k_{q,n}\left(H(U_q|[U_q]_s)+\epsilon'_q\right)},$$

where $k_{q,n} = P_Q(q)k_n$. Therefore, the following bound holds:

$$\mathbb{E}[\theta(u)] \leqslant \sum_{s=0}^{r-1} 2^{k_n(H(U|Q[U]_s)+\epsilon')} p^{-n(r-s)} \tag{21}$$

By assumption,

$$H(U|[U]_s, Q) \leqslant \frac{1}{c}(r-s)\log_2 p - \epsilon, \forall s \in [0 : r-1].$$

Therefore, for appropriate choice of ϵ and for sufficiently large n, the right-hand side of (21) can be made arbitrary small (say smaller than $\delta\gamma$). Therefore, from Markov inequality given in (18), we obtain

$$\mathbb{P}\{|\mathcal{A}^c| \geqslant \delta|\mathcal{U}_n|\} \leqslant \frac{1}{\delta|\mathcal{U}_n|} \sum_{u\in\mathcal{U}_n} \gamma\,\delta = \gamma.$$

■

Appendix B
Proof of Lemma 4

Proof: Let $\mathcal{C}_n$ be the random (n, k_n)-QGC as in Lemma 4. For shorthand, for any $\mathbf{u} \in \mathcal{U}_n$, denote $\Phi_n(\mathbf{u}) = \mathbf{u}\mathbf{G}_n$, where $\mathbf{G}_n$ is the random matrix corresponding to $\mathcal{C}_n$. Fix $\mathbf{u}_0 \in \mathcal{U}_n$. Without loss of generality assume $\mathbf{c}(\theta) = \Phi_n(\mathbf{u}_0) + B$, where B is the translation associated with $\mathcal{C}_n$. Define the event

$$\mathcal{E}_n(\mathbf{u}) := \{(\Phi_n(\mathbf{u}) + B, \tilde{\mathbf{Y}}) \in A_\epsilon^{(n)}(X, Y)\},$$

and let $\mathcal{E}_n$ be the event of interest as given in the lemma. Then $\mathcal{E}_n$ is the union of $\mathcal{E}_n(\mathbf{u})$ for all $\mathbf{u} \in \mathcal{U}_n\setminus\{\mathbf{u}_0\}$. By the union bound, the probability of $\mathcal{E}_n$ is bounded as

$$P(\mathcal{E}_n) \leqslant \sum_{\substack{\mathbf{u}\in\mathcal{U}_n\\ \mathbf{u}\neq\mathbf{u}_0}} P(\mathcal{E}_n(\mathbf{u})) \tag{22}$$

For any $\mathbf{u} \in \mathcal{U}_n$, the probability of $\mathcal{E}_n(\mathbf{u})$, can be calculated as,

$$\begin{aligned} P(\mathcal{E}_n(\mathbf{u})) &= \sum_{\mathbf{x}_0\in\mathbb{Z}_{p^r}^n} \sum_{\mathbf{y}\in\mathcal{Y}^n} P(\Phi_n(\mathbf{u}_0)+B = \mathbf{x}_0, \tilde{\mathbf{Y}}=\mathbf{y}, \mathcal{E}_n(\mathbf{u}))\\ &= \sum_{\mathbf{x}_0\in\mathbb{Z}_{p^r}^n} \sum_{\mathbf{y}\in A_\epsilon^{(n)}(Y)} \sum_{\substack{\mathbf{x}:\\ (\mathbf{x},\mathbf{y})\in A_\epsilon^{(n)}(X,Y)}} P(\Phi_n(\mathbf{u}_0) + B = \mathbf{x}_0, \tilde{\mathbf{Y}} = \mathbf{y}, \Phi_n(\mathbf{u}) + B = \mathbf{x}). \end{aligned} \tag{23}$$

By assumption, conditioned on $\Phi_n(\mathbf{u}_0) + B$, the random variable $\tilde{\mathbf{Y}}$ is independent of $\Phi_n(\mathbf{u}) + B$. Therefore, the summand in (23) is simplified to

$$P(\Phi_n(\mathbf{u}_0) + B = \mathbf{x}_0, \Phi_n(\mathbf{u}) + B = \mathbf{x})P^n_{Y|X}(\mathbf{y}|\mathbf{x}_0). \tag{24}$$

Since B is uniform over $\mathbb{Z}^n_{p^r}$, and is independent of other random variables,

$$\begin{aligned} &P(\Phi_n(\mathbf{u}_0) + B = \mathbf{x}_0,\Phi_n(\mathbf{u}) + B = \mathbf{x})\\ &\qquad = p^{-nr}P(\Phi_n(\mathbf{u} - \mathbf{u}_0) = \mathbf{x} - \mathbf{x}_0). \end{aligned}$$

Using Lemma 11 (in Appendix H), if $\mathbf{u} - \mathbf{u}_0 \in H_s^{k_n}\setminus H_{s+1}^{k_n}$, then

$$P(\Phi_n(\mathbf{u} - \mathbf{u}_0) = \mathbf{x} - \mathbf{x}_0) = p^{-n(r-s)}\mathbb{1}\{\mathbf{x} - \mathbf{x}_0 \in H_s^{k_n}\}.$$

Therefore, using (23), and for $\mathbf{u} - \mathbf{u}_0 \in H_s^{k_n}\setminus H_{s+1}^{k_n}$ we obtain

$$P(\mathcal{E}_n(\mathbf{u})) = \sum_{\mathbf{x}_0\in\mathbb{Z}_{p^r}^n} \sum_{\mathbf{y}\in A_\epsilon^{(n)}(Y)} \sum_{\substack{\mathbf{x}:\\ (\mathbf{x},\mathbf{y})\in A_\epsilon^{(n)}(X,Y)\\ \mathbf{x}-\mathbf{x}_0\in H_s^n}} p^{-nr}P^n_{Y|X}(\mathbf{y}|\mathbf{x}_0)p^{-n(r-s)}$$

Denote

$$\mathcal{A} \triangleq \{\mathbf{x} : (\mathbf{x}, \mathbf{y}) \in A_\epsilon^{(n)}(X, Y),\ \mathbf{x} - \mathbf{x}_0 \in H_s^n\}.$$

Note that if $([\mathbf{x}_0]_s, \mathbf{y}) \notin A_\epsilon^{(n)}([X]_s Y)$, then $\mathcal{A} = \varnothing$. Therefore,

$$P(\mathcal{E}_n(\mathbf{u})) = \sum_{\substack{(\mathbf{x}_0,\mathbf{y}):\\ ([\mathbf{x}_0]_s,\mathbf{y})\in A_\epsilon^{(n)}([X]_s Y)}} \sum_{\mathbf{x}\in\mathcal{A}} p^{-nr} P^n_{Y|X}(\mathbf{y}|\mathbf{x}_0) p^{-n(r-s)} \tag{25}$$

Next, we replace the summation over $\mathbf{x}$ with the size of the set $\mathcal{A}$. We bound the size of $\mathcal{A}$ using Lemma 12. Therefore, an upper-bound on (25) is

$$\begin{aligned}
P(\mathcal{E}_n(\mathbf{u})) &\leqslant \sum_{\substack{(\mathbf{x}_0,\mathbf{y}):\\ ([\mathbf{x}_0]_s,\mathbf{y})\in A_\epsilon^{(n)}([X]_s Y)}} p^{-nr} P^n_{Y|X}(\mathbf{y}|\mathbf{x}_0) p^{-n(r-s)} 2^{n(H(X|Y,[X]_s)+\delta(4\epsilon))}\\
&\leqslant \sum_{\mathbf{x}_0\in\mathbb{Z}^n_{p^r}} \sum_{\mathbf{y}\in\mathcal{Y}^n} p^{-nr} P^n_{Y|X}(\mathbf{y}|\mathbf{x}_0) p^{-n(r-s)} 2^{n(H(X|Y,[X]_s)+\delta(4\epsilon))}\\
&\leqslant p^{-n(r-s)} 2^{n(H(X|Y,[X]_s)+\delta(4\epsilon))}.
\end{aligned} \tag{26}$$

Note that if $\mathbf{a} \in \mathbb{Z}^k_{p^r}, \mathbf{a} \neq \mathbf{0}$ then there exists $s \in [0 : r-1]$ such that $\mathbf{a} \in H^k_s \backslash H^k_{s+1}$. Therefore, there are r different cases for each value of s. Using (26), and considering these cases, we obtain

$$\begin{aligned}
P(\mathcal{E}_n) &\leqslant \sum_{s=0}^{r-1} \sum_{\substack{\mathbf{u}\in\mathcal{U}_n\\ \mathbf{u}-\mathbf{u}_0\in H_s^{kn}\backslash H_{s+1}^{kn}}} P(\mathcal{E}_n(\mathbf{u}))\\
&\leqslant \sum_{s=0}^{r-1} \sum_{\substack{\mathbf{u}\in\mathcal{U}_n\\ \mathbf{u}-\mathbf{u}_0\in H_s^{kn}\backslash H_{s+1}^{kn}}} 2^{n(H(X|Y[X]_s)+\delta(4\epsilon))} p^{-n(r-s)}\\
&\leqslant \sum_{s=0}^{r-1} |\mathcal{U}_n \textstyle\bigcap (\mathbf{u}_0 + H^k_s)| 2^{n(H(X|Y[X]_s)+\delta(4\epsilon))} p^{-n(r-s)}.
\end{aligned}$$

Note that $\mathcal{U}_n$ is the Cartesian product of ϵ-typical sets $A_\epsilon^{(p(q)k_n)}(U_q)$, where $q \in \mathcal{Q}$ and $k_n = cn$. For each component q of $\mathcal{U}_n$, we can apply Lemma 12. Therefore,

$$\begin{aligned}
|\mathcal{U}_n \cap (\mathbf{u}_0 + H^k_s)| &\leqslant 2^{\sum_q p(q)k_n(H(U_q|[U_q]_s)+\delta(2\epsilon))}\\
&= 2^{k_n(H(U|[U]_s,Q)+\delta(2\epsilon))}.
\end{aligned}$$

Finally,

$$P(\mathcal{E}_n) \leqslant \sum_{s=0}^{r-1} 2^{n\left(\frac{k_n}{n}(H(U|[U]_s,Q)+H(X|Y,[X]_s)+\frac{k_n}{n}\delta(2\epsilon)+\delta(4\epsilon)\right)} p^{-n(r-s)}.$$

As a result $\lim_{n\to\infty} P(\mathcal{E}_n) = 0$, if the inequality

$$cH(U|[U]_s, Q) \leqslant \log_2 p^{r-s} - H(X|Y,[X]_s) - 2(2+c)\delta(\epsilon),$$

holds for all $0 \leqslant s \leqslant r-1$. Multiply each side of this inequality by $\frac{H(U|Q)}{H(U|Q,[U]_s)}$. This gives the following bound

$$cH(U|Q) \leqslant \frac{H(U|Q)}{H(U|Q,[U]_s)}\Big(\log_2 p^{r-s} - H(X|Y,[X]_s) - 2(2+c)\delta(\epsilon)\Big).$$

By definition $R_n = \frac{1}{n}\log_2|\mathcal{C}_n| \leqslant cH(U|Q) + \epsilon'$. Therefore,

$$R_n \leqslant \frac{H(U|Q)}{H(U|Q,[U]_s)}(\log_2 p^{r-s} - H(X|Y,[X]_s) - 2(2+c)\delta(\epsilon)),$$

and the proof is completed. ■

Appendix C
Proof of Lemma 5

Proof: We use the same notation as in the proof of Lemma 4. For any typical sequence $\mathbf{x}$ define

$$\lambda_n(\mathbf{x}) = \sum_{\hat{\mathbf{x}}\in A_\epsilon^{(n)}(\hat{X}|\mathbf{x})} \sum_{\mathbf{u}\in\mathcal{U}_n} \mathbb{1}\{\Phi_n(\mathbf{u}) + B = \hat{x}\}.$$

Note $\lambda_n(\mathbf{x})$ counts the number of codewords that are conditionally typical with $\mathbf{x}$ with respect to $p(\hat{\mathbf{x}}|\mathbf{x})$. We show that

$$\lim_{n\to\infty} P(\lambda_n(\mathbf{x}) = 0) = 0$$

for any ϵ-typical sequence $\mathbf{x}$. This implies that

$$\lim_{n\to\infty} P(\lambda_n(\mathbf{X}^n) = 0) = 0,$$

where $\mathbf{X}^n \sim \prod_{i=1}^n p(x)$. This proves the statements of the Lemma. Hence, it suffices to show that $\lim_{n\to\infty} P(\lambda_n(\mathbf{x}) = 0) = 0$. We have,

$$\begin{aligned}
P\{\lambda_n(\mathbf{x}) = 0\} &\leqslant P\Big\{\lambda_n(\mathbf{x}) \leqslant \frac{1}{2}E(\lambda_n(x))\Big\}\\
&\leqslant P\Big\{|\lambda_n(x) - E(\lambda_n(x))| \geqslant \frac{1}{2}E(\lambda_n(x))\Big\}
\end{aligned} \tag{27}$$

Hence, by Chebyshev's inequality,

$$P\{\lambda_n(\mathbf{x}) = 0\} \leqslant \frac{4\ Var(\lambda_n(x))}{E(\lambda_n(x))^2}.$$

Note that

$$E(\lambda_n(x)) = \sum_{\hat{\mathbf{x}}\in A_\epsilon^{(n)}(\hat{X}|\mathbf{x})} \sum_{\mathbf{u}\in\mathcal{U}_n} P\{\Phi(\mathbf{u}) + B = \hat{\mathbf{x}}\} \tag{28}$$

Since B is uniform over $\mathbb{Z}^n_{p^r}$, we get

$$E(\lambda_n(x)) = |A_\epsilon^{(n)}(X|\hat{\mathbf{x}})||\mathcal{U}_n|p^{-rn}. \tag{29}$$

Note that

$$2^{k_n(H(U|Q)-2\epsilon')} \leqslant |\mathcal{U}_n| \leqslant 2^{k_n(H(U|Q)+2\epsilon')},$$

where

$$\epsilon' = -\frac{\epsilon}{p^r}\sum_{q\in\mathcal{Q}} P_Q(q) \sum_{a\in\mathbb{Z}_{p^r}:P_{U|Q}(a|q)>0} \log P_{U|Q}(a|q).$$

Therefore,

$$E(\lambda_n(x)) \geqslant 2^{n(H(\hat{X}|X)-2\tilde{\epsilon})} 2^{k_n(H(U|Q)-2\epsilon')} p^{-rn} \tag{30a}$$

$$E(\lambda_n(x)) \leqslant 2^{n(H(\hat{X}|X)+2\tilde{\epsilon})} 2^{k_n(H(U|Q)+2\epsilon')} p^{-rn}, \tag{30b}$$

To calculate the variance, we start with

$$E(\lambda_n(x)^2) = \sum_{\hat{\mathbf{x}},\hat{\mathbf{x}}'\in A_\epsilon^{(n)}(\hat{X}|\mathbf{x})} \sum_{\mathbf{u},\mathbf{u}'\in\mathcal{U}_n} P\{\Phi(\mathbf{u}) + B = \hat{\mathbf{x}}, \Phi(\mathbf{u}') + B = \hat{\mathbf{x}'}\}.$$

Since B is independent of other random variables, the most inner term in the above summations is simplified to

$$p^{-nr} P\{\Phi(\mathbf{u}-\mathbf{u}') = \hat{\mathbf{x}} - \hat{\mathbf{x}}'\}.$$

Using Lemma 11 (in Appendix H), if $\mathbf{u}-\mathbf{u}' \in H_s^{k_n} \backslash H_{s+1}^{k_n}$, then

$$P\{\Phi(\mathbf{u}-\mathbf{u}') = \hat{\mathbf{x}} - \hat{\mathbf{x}}'\} = p^{-n(r-s)} \mathbb{1}\{\hat{\mathbf{x}} - \hat{\mathbf{x}}' \in H_s^n\}$$

Considering all the cases for the values of s, we get

$$E(\lambda_n(x)^2) = \sum_{s=0}^{r} \sum_{\substack{\mathbf{u},\mathbf{u}' \in \mathcal{U}_n \\ \mathbf{u}-\mathbf{u}' \in H_s^{k_n} \backslash H_{s+1}^{k_n}}} \sum_{\substack{\mathbf{x},\hat{\mathbf{x}}' \in A_\epsilon^{(n)}(\hat{X}|\mathbf{x}) \\ \hat{\mathbf{x}} - \hat{\mathbf{x}}' \in H_s^n}} p^{-nr} p^{-n(r-s)}$$

Since the innermost terms in the above summations do not depend on the individual values of $\mathbf{x}, \hat{\mathbf{x}}, \mathbf{u}, \mathbf{u}'$, the corresponding summations can be replaced by the size of the associated sets. Moreover, we provide an upper bound on the summation over $\mathbf{u}, \mathbf{u}'$ by replacing $H_s^{k_n} \backslash H_{s+1}^{k_n}$ with $H_s^{k_n}$. Using Lemma 12 for $\mathbf{x}, \hat{\mathbf{x}}$, we get

$$E(\lambda_n(x)^2) \leqslant \sum_{s=0}^{r} \sum_{\mathbf{u} \in \mathcal{U}_n} \sum_{\substack{\mathbf{u}' \in \mathcal{U}_n \\ \mathbf{u}-\mathbf{u}' \in H_s^{k_n}}} 2^{n(H(\hat{X}|X)+\tilde{\epsilon}+H(\hat{X}|X,[\hat{X}]_s)+\delta(4\epsilon))} p^{-nr} p^{-n(r-s)}.$$

For any $\mathbf{u} \in \mathcal{U}_n$, by applying Lemma 12 we get

$$|\mathcal{U}_n \bigcap (\mathbf{u} + H_s^{k_n})| \leqslant 2^{k_n(H(U|Q,[U]_s)+\delta(4\epsilon))}.$$

As a result,

$$\begin{aligned} E(\lambda_n(x)^2) \leqslant \sum_{s=0}^{r} & 2^{k_n(H(U|Q,[U]_s)+\delta(4\epsilon))} 2^{k_n(H(U|Q)+\epsilon')} \\ & \times 2^{n(H(\hat{X}|X)+\tilde{\epsilon}+H(\hat{X}|X,[\hat{X}]_s)+\delta(4\epsilon))} p^{-nr} p^{-n(r-s)}. \end{aligned}$$

Note that the case $s=0$ gives $E^2(\lambda_n(x))$. Therefore,

$$\begin{aligned} Var(\lambda_n(x)^2) &\leqslant p^{-nr} \sum_{s=1}^{r} 2^{k_n(H(U|Q)+H(U|Q,[U]_s))} \\ &\times 2^{n(H(\hat{X}|X)+H(\hat{X}|X,[\hat{X}]_s))} 2^{n(1+c)(\epsilon+\delta(4\epsilon))} p^{-n(r-s)} \end{aligned} \tag{31}$$

Finally, using (30), (31) and the Chebyshev's inequality as argued before, we get

$$\begin{aligned} P\{\lambda_n(\mathbf{x}) = 0\} &\leqslant 4 \sum_{s=1}^{r} 2^{k_n(-H(U|Q)+H(U|Q,[U]_s))} \\ &\times 2^{n(-H(\hat{X}|X)+H(\hat{X}|X,[\hat{X}]_s))} 2^{n(1+c)(\epsilon+\delta(4\epsilon))} p^{nr} p^{-n(r-s)} \\ &= 4\, 2^{n(1+c)(\epsilon+\delta(4\epsilon))} \sum_{s=1}^{r} 2^{-k_n H([U]_s|Q)} 2^{-nH([\hat{X}]_s|X)} p^{ns}. \end{aligned}$$

The second equality follows, because the equality

$$H(V|W) - H(V|[V]_s, W) = H([V]_s|W)$$

holds for any random variables V and W. Therefore, $P\{\lambda_n(\mathbf{x})\}$ approaches zero, as $n \to \infty$, if the inequality

$$cH([U]_s|Q) \geqslant \log_2 \; p^s - H([\hat{X}]_s|X) + (1+c)(\epsilon + \delta(4\epsilon)),$$

holds for $1 \leqslant s \leqslant r$. By the definition of rate and the above inequalities the proof is completed. ∎

Appendix D
Proof of Theorem 2

Fix a positive integer n, and define $l_1 \triangleq c_1 n$, $l_2 \triangleq c_2 n$, and $k \triangleq \tilde{c}n$, where $\tilde{c}, c_1$ and c_2 are positive real numbers such that l_1, l_2 and k are integers.

Codebook Generation: We use two nested QGC's, one for each encoder. The codebook for Encoder 1 is an (n, k, l_1) nested QGC (as in Definition 5) with random variables (W_1, V_1, Q). Let $\mathcal{C}_{I,1}, \bar{\mathcal{C}}_1$, and $\mathcal{C}_{O,1}$ denote the corresponding inner code, shift code and the outer code (as in Definition 5), respectively. The codebook for Encoder 2 is an (n, k, l_2) nested QGC with random variables (W_2, V_2, Q), inner code $\mathcal{C}_{I,2}$, shift code $\bar{\mathcal{C}}_2$, and outer code $\mathcal{C}_{O,2}$. The codebook at the decoder is denoted by $\mathcal{C}_d$ which is an (n, k) QGC with random variables $(W_1 + W_2, Q)$.

Conditioned on Q, the random variables (W_1, W_2, V_1, V_2) are mutually independent. The random variable V_i is uniform over $\{0, 1\}$, and is independent of Q.

The nested QGCs and $\mathcal{C}_d$ have identical generator matrices but different translations and index random variables. Note that each nested QGC has two generator matrices/translations, one for the inner code and one for the shift code as in Definition 5. The generator matrix and the translation for the inner codes $\mathcal{C}_{I,i}, i = 1, 2$, are denoted by $\mathbf{G}$ and $\mathbf{b}$, respectively. The generator matrix and the translation used for shift code $\mathcal{C}_{I,i}$, are denoted by $\bar{\mathbf{G}}$ and $\bar{\mathbf{b}}_i$, respectively, where $i = 1, 2$. The elements of $\mathbf{G}, \bar{\mathbf{G}}, \mathbf{b}$, and $\bar{\mathbf{b}}_i, i = 1, 2$ are generated randomly and independently from $\mathbb{Z}_{p^r}$.

By $R_{O,i}$ and $R_{I,i}$ denote the rate of the inner code and outer code defined for the ith nested QGC. Define $R_i \triangleq R_{O,i} - R_{I,i}, i = 1, 2$.

Encoding: Suppose $(\mathbf{x}_1, \mathbf{x}_2)$ is a realization of (X_1^n, X_2^n). The first encoder checks if $\mathbf{x}_1$ is ϵ-typical and $\mathbf{x}_1 \in \mathcal{C}_{O,1}$. If not, an encoding error E_1 is declared. In the case of no encoding error, by Definition 5, $\mathbf{x}_1 = \mathbf{c}_{I,1} + \bar{\mathbf{c}}_1$, where $\mathbf{c}_{I,1} \in \mathcal{C}_{I,1}$ and $\bar{\mathbf{c}}_1 \in \bar{\mathcal{C}}_1$. The first encoder sends the index of $\bar{\mathbf{c}}_1$. Note $\bar{\mathbf{c}}_1$ determines the index of the bin which contains $\mathbf{x}_1$. Similarly, if $\mathbf{x}_2 \in A_\epsilon^{(n)}(X_2)$ and $\mathbf{x}_2 \in \mathcal{C}_{O,2}$, the second encoder sends finds $\mathbf{c}_{I,2} \in \mathcal{C}_{I,2}$ and $\bar{\mathbf{c}}_2 \in \bar{\mathcal{C}}_2$ such that $\mathbf{x}_2 = \mathbf{c}_{I,2} + \bar{\mathbf{c}}_2$. Then it sends the index of $\bar{\mathbf{c}}_2$. If no such $\mathbf{c}_{I,2}$ and $\bar{\mathbf{c}}_2$ are found, an error event E_2 is declared.

Decoding: The decoder wishes to reconstruct $\mathbf{x}_1 + \mathbf{x}_2$. Assume there is no encoding error. Upon receiving the bin numbers from the encoders, the decoder calculates $\bar{\mathbf{c}}_1$ and $\bar{\mathbf{c}}_2$. Then, it finds $\tilde{\mathbf{c}} \in \mathcal{C}_d$ such that

$$\tilde{\mathbf{c}} + \bar{\mathbf{c}}_1 + \bar{\mathbf{c}}_2 \in A_\epsilon^{(n)}(X_1 + X_2).$$

If $\tilde{\mathbf{c}}$ is unique, then

$$\tilde{\mathbf{c}} + \bar{\mathbf{c}}_1 + \bar{\mathbf{c}}_2$$

is declared as a reconstruction of $\mathbf{x}_1 + \mathbf{x}_2$. An error event E_d occurs, if no unique $\tilde{\mathbf{c}}$ was found.

We need to find conditions for which the probability of the error events E_1, E_2 and E_d approach zero. By $\mathcal{W}_i$ denote the index set of $\mathcal{C}_{I,i}$, and let $\mathcal{V}_i$ be the index set of $\bar{\mathcal{C}}_i, i = 1, 2$.

Error: Let $(f_1(\cdot), f_2(\cdot))$ and $g(\cdot,\cdot)$ denote the encoding and decoding functions corresponding to the above coding scheme. The overall error event is defined as

$$E \triangleq \{\mathbf{X}_1^n + \mathbf{X}_2^n \neq g(f_1(\mathbf{X}_1^n), f_2(\mathbf{X}_2^n))\}$$

For the achievability, we need to show that $P(E)$ can be made arbitrary small for sufficiently large n. For that, using the aforementioned encoding and decoding error events we have

$$P(E) \leqslant P(E_1 \cup E_2 \bigcup E_d) + P(E|E_1^c \bigcap E_2^c \bigcap E_d^c)$$

Using standard arguments for typical sequences, we can show that when there is no encoding and decoding error (i.e., $E_1^c \bigcap E_2^c \bigcap E_d^c$) the error probability $P(E|E_1^c \bigcap E_2^c \bigcap E_d^c)$ approaches 0 as $n \to \infty$. As a result, the second term above is sufficiently small for large enough n. Therefore, for sufficiently large n and from the union bound on the first term we obtain,

$$P(E) \leqslant P(E_1) + P(E_2) + P(E_d) + \epsilon.$$

A. Analysis of E_1, E_2

In what follows, we apply the covering lemma (Lemma 5) to bound the probability of the encoding errors. For that the outer code $\mathcal{C}_{O,i}$ is used to "cover" the source X_i. Note that $\mathcal{C}_{O,i}$ is the outer code for the (n, k, l) nested QGC used at Encoder i, $i = 1, 2$. Therefore, $\mathcal{C}_{O,i}$ is a $(n, k+l)$ QGC with appropriately defined index random variables (as is defined in Lemma 3). The random variables defined for $\mathcal{C}_{O,i}$ are $(U_i, (Q, J_i))$, where given $J_i = 1$ we have $U_i = W_i$, and given $J_i = 2$ we get $U_i = V_i$. In addition, $P(J_i = 0) = \frac{k}{l_i+k}$, and $P(J_i = 1) = \frac{l_i}{l_i+k}$. We apply Lemma 5 to bound the probability of E_i. In this lemma set $\hat{X} = X = X_i$ with probability one, $\mathcal{C}_n = \mathcal{C}_{O,i}$, and $R_n = R_{O,i}, i = 1, 2$. Using Lemma 5, $P(E_i)$ is sufficiently small for large blocklength n if

$$R_{O,i} \geqslant \max_{1\leqslant s\leqslant r} \frac{H(U_i|Q, J_i)}{H([U_i]_s|Q, J_i)}(\log_2 \ p^s + o(\epsilon)).$$

Using Remark 3, and the above bound we get

$$\frac{k+l_i}{n} H([U_i]_s|Q, J_i) \geqslant \log_2 \ p^s + o(\epsilon)$$

for $s \in [1 : r]$. Therefore, by the definition of U_i and J_i, we get

$$\frac{k}{n} H([W_i]_s|Q) + \frac{l_i}{n} H(V_i|Q) \geqslant \log_2 \ p^s + o(\epsilon), \ 1 \leqslant s \leqslant r.$$

Note that in this bound we use the equality $H([V_i]_s) = H(V_i)$. This equality holds because V_i takes values from $\{0, 1\}$. Again using Remark 3, we get $|R_i - \frac{l_i}{n} H(V_i|Q)| \leqslant o(\epsilon)$. Hence, if the following holds

$$\frac{k}{n} H([W_i]_s|Q) + R_i \geqslant \log_2 p^s + o(\epsilon), \tag{32}$$

for $1 \leqslant s \leqslant r$ and $i = 1, 2$, then $P(E_i) \to 0$ as $n \to \infty$.

B. Analysis of E_d

Upon receiving the bin numbers, the decoder calculates $\bar{\mathbf{c}}_1$ and $\bar{\mathbf{c}}_2$. The decoding error consists of two events: 1) no typical sequence $\tilde{\mathbf{z}}$ was found, and 2) multiple typical sequences $\tilde{\mathbf{z}}$ were found. Using standard arguments, one can show that the probability of the first event is sufficiently small for large enough n. In what follows, we bound the probability of the second event, i.e., $E_{d,2}$. This event occurs, if there exist more than one $\tilde{\mathbf{c}} \in \mathcal{C}_{I,1} + \mathcal{C}_{I,2}$ such that $\tilde{\mathbf{c}} + \bar{\mathbf{c}}_1 + \bar{\mathbf{c}}_2$ is ϵ-typical with respect to $P_{X_1+X_2}$.

To bound $P(E_{d,2})$ we need to take into account whether there is an encoding error or not. For that, first we provide an alternative representation for the encoding errors. For any sequence $\mathbf{x}_i \in \mathbb{Z}_{p^r}^n$ define

$$\lambda_i(\mathbf{x}_i) = \sum_{\mathbf{w}_i \in \mathcal{W}_i} \sum_{\mathbf{v}_i \in \mathcal{V}_i} \mathbb{1}\{\mathbf{x}_i = \mathbf{w}_i \mathbf{G} + \mathbf{v}_i \bar{\mathbf{G}} + \mathbf{b} + \bar{\mathbf{b}}_i\},$$

where $i = 1, 2$ and $(\mathbf{G}, \bar{\mathbf{G}}, \mathbf{b}, \bar{\mathbf{b}}_i)$ are the generator matrices and translations defined for the ith nested QGC. With this notation, E_i occurs if $\lambda_i(\mathbf{x}_i) = 0$, where $(\mathbf{x}_1, \mathbf{x}_2)$ is a realization of the sources. Next, we define a super-set of the encoding error events as

$$E_i' \triangleq \{\lambda_i(\mathbf{x}_i) < \frac{1}{2} E(\lambda_i(\mathbf{x}_i))\}, \qquad i = 1, 2, \tag{33}$$

where $E(\lambda_i(\mathbf{x}_i))$ is the expected value of $\lambda_i(\mathbf{x}_i)$. Note that $E_i \subseteq E_i', i = 1, 2$.

For the modified encoding error events (E_1', E_2') given in (33) we have

$$\begin{aligned} P(E_{d,2}) &\leqslant P(E_1' \bigcup E_2') + P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c) \\ &\leqslant P(E_1') + P(E_2') + P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c) \end{aligned}$$

For the first two terms above, based on the proof of Lemma 5, we can showed that $P(E_i') \to 0$ as $n \to \infty$. Note that $P(E_i')$ is the same as the second term in (27) in the proof of the covering. In fact, for the proof of the covering bound, we showed that such probability approaches 0 as $n \to \infty$.

In what follow, we show that the second probability in the above approaches 0 as $n \to \infty$.

Analysis of $\mathbf{P(E_{d,2}|E_1'^c \bigcap E_2'^c)}$: Note that $E_1'^c \bigcap E_2'^c$ implies that there is no encoding error; because

$$\lambda_i(\mathbf{x}_i) > 1/2 \ E(\lambda_i(\mathbf{x}_i)).$$

Since there is no error at the encoding stage, $\mathbf{x}_i \in \mathcal{C}_{O,i}, i = 1, 2$. By Definition 5, every codeword in $\mathcal{C}_{O,i}$ is characterized by a pair $(\mathbf{v}_i, \mathbf{w}_i)$, where $\mathbf{v}_i \in \mathcal{V}_i, \mathbf{w}_i \in \mathcal{W}_i, i = 1, 2$. Given $\mathbf{x}_i$, if more than one pair was found at the ith encoder, select one randomly and uniformly. By $P(\mathbf{v}_i, \mathbf{w}_i|\mathbf{x}_i)$ denote the probability that $(\mathbf{v}_i, \mathbf{w}_i)$ is selected at the ith encoder. Then,

$$P(\mathbf{v}_i, \mathbf{w}_i|\mathbf{x}_i) = \frac{1}{\lambda_i(\mathbf{x}_i)} \mathbb{1}\{\mathbf{w}_i \mathbf{G} + \mathbf{v}_i \bar{\mathbf{G}} + \mathbf{b} + \bar{\mathbf{b}}_i = \mathbf{x}_i\}.$$

Fix $\mathbf{G}$, $\tilde{\mathbf{G}}_i, \mathbf{b}$ and $\bar{\mathbf{b}}_i, i = 1, 2$. Suppose $\mathbf{x}_1$ and $\mathbf{x}_2$ are the realizations of the sources X_1 and X_2, respectively. Moreover, suppose $(\mathbf{x}_1, \mathbf{x}_2) \in A_\epsilon^{(n)}(X_1, X_2)$. Therefore,

$$\begin{aligned}&P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c|\mathbf{x}_1, \mathbf{x}_2)\\ &= \mathbb{1}\Big\{\lambda_i(\mathbf{x_i}) \geqslant \frac{1}{2}E(\lambda_i(\mathbf{x}_i)), i = 1, 2\Big\}\times\\ &\left[\prod_{j=1}^{2}\sum_{\mathbf{v}_j\in\mathcal{V}_j}\sum_{\mathbf{w}_j\in\mathcal{W}_j} P(\mathbf{v}_j, \mathbf{w}_j|\mathbf{x}_j)\right] P(E_{d,2}|\mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2).\end{aligned}$$

In what follows, we bound $P(E_{d,2}|\mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2)$, $P(\mathbf{v}_1, \mathbf{w}_1|\mathbf{x}_1)$, and $P(\mathbf{v}_2, \mathbf{w}_2|\mathbf{x}_2)$. For the first conditional probability we have

$$\begin{aligned}P(E_{d,2}|\mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i =1, 2) = \mathbb{1}\big\{\exists\tilde{\mathbf{z}} \in A_\epsilon^{(n)}(X_1 + X_2) :\\ \tilde{\mathbf{z}} \neq \mathbf{x}_1 + \mathbf{x}_2, \tilde{\mathbf{z}} \in \mathcal{C}_{I,1} + \mathcal{C}_{I,2} + \bar{\mathbf{c}}_1 + \bar{\mathbf{c}}_2\big\},\end{aligned}$$

where, $\bar{\mathbf{c}}_i = \mathbf{v}_i\bar{\mathbf{G}} + \bar{\mathbf{b}}_i$, $i = 1, 2$. Let $\mathcal{W} = \mathcal{W}_1 + \mathcal{W}_2$, and define $Z \triangleq X_1 + X_2$. Using the union bound, we have

$$\begin{aligned}&P(E_{d,2}|\mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2)\\ &\leqslant \sum_{\tilde{\mathbf{w}}\in\mathcal{W}} \sum_{\substack{\tilde{\mathbf{z}}\in A_\epsilon^{(n)}(Z)\\ \tilde{\mathbf{z}}\neq\mathbf{x}_1+\mathbf{x}_2}} \mathbb{1}\{\tilde{\mathbf{w}}\mathbf{G} + (\mathbf{v}_1 + \mathbf{v}_2)\bar{\mathbf{G}} + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 = \tilde{\mathbf{z}}\}\\ &\leqslant \sum_{\substack{\tilde{\mathbf{w}}\in\mathcal{W}\\ \tilde{\mathbf{w}}\neq\mathbf{w}_1+\mathbf{w}_2}} \sum_{\tilde{\mathbf{z}}\in A_\epsilon^{(n)}(Z)} \mathbb{1}\{\tilde{\mathbf{w}}\mathbf{G} + (\mathbf{v}_1 + \mathbf{v}_2)\bar{\mathbf{G}} + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 = \tilde{\mathbf{z}}\}.\end{aligned} \tag{34}$$

The second inequality follows, because the condition $\tilde{\mathbf{w}} \neq \mathbf{w}_1 + \mathbf{w}_2$ is less restrictive than $\tilde{\mathbf{z}} \neq \mathbf{x}_1 + \mathbf{x}_2$. This is due to the fact that $\mathbf{G}$ is not injective necessarily.

Next, we provide an upper-bound on $P(\mathbf{v}_i, \mathbf{w}_i|\mathbf{x}_i)$, $i = 1, 2$. Since $E_1'^c \bigcap E_2'^c$ is in the conditioning, $\lambda_i(\mathbf{x}_i) \geqslant \frac{1}{2}E(\lambda_i(\mathbf{x}_i))$. As a result,

$$P(\mathbf{v}_i, \mathbf{w}_i|\mathbf{x}_i) \leqslant \frac{2}{E(\lambda_i(\mathbf{x}_i))}\mathbb{1}\{\mathbf{w}_i\mathbf{G} + \mathbf{v}_i\bar{\mathbf{G}} + \mathbf{b} + \bar{\mathbf{b}}_i = \mathbf{x}_i\} \tag{35}$$

Using the bounds given in (34) and (35), we get

$$\begin{aligned}&P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c|\mathbf{x}_1, \mathbf{x}_2) \leqslant\\ &\left[\prod_{j=1}^{2}\sum_{\substack{\mathbf{v}_j\in\mathcal{V}_j\\ \mathbf{w}_j\in\mathcal{W}_j}} \frac{2}{E(\lambda_j(\mathbf{x}_j))}\mathbb{1}\{\mathbf{w}_j\mathbf{G} + \mathbf{v}_j\bar{\mathbf{G}} + \mathbf{b} + \bar{\mathbf{b}}_j = \mathbf{x}_j\}\right]\\ &\times \sum_{\substack{\tilde{\mathbf{w}}\in\mathcal{W}\\ \tilde{\mathbf{w}}\neq\mathbf{w}_1+\mathbf{w}_2}} \sum_{\tilde{\mathbf{z}}\in A_\epsilon^{(n)}(Z)} \mathbb{1}\{\tilde{\mathbf{w}}\mathbf{G} + (\mathbf{v}_1 + \mathbf{v}_2)\bar{\mathbf{G}} + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 = \tilde{\mathbf{z}}\}.\end{aligned}$$

Next, we average $P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c|\mathbf{x}_1, \mathbf{x}_2)$ over all possible choices of $\mathbf{G}$, $\bar{\mathbf{G}}$, $\mathbf{b}$, $\bar{\mathbf{b}}_1$, and $\bar{\mathbf{b}}_2$. We obtain

$$\begin{aligned}&\mathbb{E}\{P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c|\mathbf{x}_1, \mathbf{x}_2)\} \leqslant\\ &\sum_{\substack{\mathbf{v}_1\in\mathcal{V}_1\\ \mathbf{w}_1\in\mathcal{W}_1}} \frac{2}{E(\lambda_1(\mathbf{x}_1))} \sum_{\substack{\mathbf{v}_2\in\mathcal{V}_2\\ \mathbf{w}_2\in\mathcal{W}_2}} \frac{2}{E(\lambda_2(\mathbf{x}_2))} \sum_{\substack{\tilde{\mathbf{w}}\in\mathcal{W}\\ \tilde{\mathbf{w}}\neq\mathbf{w}_1+\mathbf{w}_2}} \sum_{\tilde{\mathbf{z}}\in A_\epsilon^{(n)}(Z)}\\ &P\Big\{\tilde{\mathbf{w}}\mathbf{G} + (\mathbf{v}_1 + \mathbf{v}_2)\bar{\mathbf{G}} + 2\mathbf{B} + \bar{\mathbf{B}}_1 + \bar{\mathbf{B}}_2 = \tilde{\mathbf{z}},\\ &\qquad \mathbf{w}_i\mathbf{G} + \mathbf{v}_i\bar{\mathbf{G}} + \mathbf{B} + \bar{\mathbf{B}}_i = \mathbf{x}_i, i = 1, 2\Big\}.\end{aligned}$$

Note $\bar{\mathbf{B}}_1$ and $\bar{\mathbf{B}}_2$ are independent random variables with uniformly distributed over $\mathbb{Z}_{p^r}^n$. Therefore, the innermost term in the above summations equals

$$p^{-2nr}P\{(\tilde{\mathbf{w}} - \mathbf{w}_1 - \mathbf{w}_2)\mathbf{G} = \tilde{\mathbf{z}} - \mathbf{x}_1 - \mathbf{x}_2\}. \tag{36}$$

We apply Lemma 11 (in Appendix H), to calculate the above probability. If $\tilde{\mathbf{w}} - \mathbf{w}_1 - \mathbf{w}_2 \in H_s^k\backslash H_{s+1}^k$, then (36) equals to

$$p^{-2nr}p^{-n(r-s)}\mathbb{1}\{\tilde{\mathbf{z}} - \mathbf{x}_1 - \mathbf{x}_2 \in H_s^k\}. \tag{37}$$

As a result, we have

$$\begin{aligned}&\mathbb{E}\{P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c|\mathbf{x}_1, \mathbf{x}_2)\} \leqslant\\ &\sum_{\substack{\mathbf{v}_1\in\mathcal{V}_1\\ \mathbf{w}_1\in\mathcal{W}_1}} \frac{2}{E(\lambda_1(\mathbf{x}_1))} \sum_{\substack{\mathbf{v}_2\in\mathcal{V}_2\\ \mathbf{w}_2\in\mathcal{W}_2}} \frac{2}{E(\lambda_2(\mathbf{x}_2))} \sum_{s=0}^{r-1} \sum_{\substack{\tilde{\mathbf{w}}\in\mathcal{W}\\ \tilde{\mathbf{w}}-\mathbf{w}_1-\mathbf{w}_2\in H_s^k\backslash H_{s+1}^k}}\\ &\sum_{\substack{\tilde{\mathbf{z}}\in A_\epsilon^{(n)}(Z)\\ \tilde{\mathbf{z}}-\mathbf{x}_1-\mathbf{x}_2\in H_s^n}} p^{-2nr}p^{-n(r-s)}.\end{aligned}$$

Since the innermost terms in the above summations depend only on s, we can replace the summations over $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{z}}$ with the size of the associated sets. We apply Lemma 12 to bound the size of these sets. Also, we can replace the summations over $\mathbf{v}_i$ and $\mathbf{w}_i$, $i = 1, 2$ with the size of the related sets. Define $W \triangleq W_1 + W_2$, we get,

$$\begin{aligned}&\mathbb{E}\{P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c|\mathbf{x}_1, \mathbf{x}_2)\} \leqslant\\ &|\mathcal{W}_1||\mathcal{V}_1|\frac{2}{E(\lambda_1(\mathbf{x}_1))}|\mathcal{W}_2||\mathcal{V}_2|\frac{2}{E(\lambda_2(\mathbf{x}_2))}\times\\ &\sum_{s=0}^{r-1} 2^{n(H(Z|[Z]_s)+o(\epsilon))}2^{k(H(W|Q,[W]_s)+o(\epsilon))}p^{-2nr}p^{-n(r-s)}.\end{aligned}$$

Note that from (29) in the proof of Lemma 5,

$$E(\lambda_i(\mathbf{x}_i)) = |\mathcal{W}_i||\mathcal{V}_i|p^{-nr}, i = 1, 2.$$

Therefore, we have

$$\begin{aligned}&\mathbb{E}\{P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c|\mathbf{x}_1, \mathbf{x}_2)\} \leqslant\\ &4\sum_{s=0}^{r-1} 2^{n(H(Z|[Z]_s)+o(\epsilon))}2^{k(H(W|Q,[W]_s)+o(\epsilon))}p^{-n(r-s)}.\end{aligned}$$

Note that the above bound does not depend on ϵ-typical sequences $\mathbf{x}_1$ and $\mathbf{x}_2$. Using standard arguments for ϵ-typical sets, the probability that $(\mathbf{X}_1^n, \mathbf{X}_2^n) \notin A_\epsilon^{(n)}(X_1, X_2)$ is upper-bounded by $\frac{c}{n\epsilon^2}$, where $c = \frac{p^{6r}}{4}$. Hence, we have

$$\begin{aligned}&\mathbb{E}\{P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c)\} \leqslant \frac{c}{n\epsilon^2} + 4(1 - \frac{c}{n\epsilon^2})\sum_{s=0}^{r-1}\\ &\qquad 2^{n(H(Z|[Z]_s)+o(\epsilon))}2^{k(H(W|Q,[W]_s)+o(\epsilon))}p^{-n(r-s)}.\end{aligned}$$

Therefore, $\mathbb{E}\{P(E_{d,2} \bigcap E_1'^c \bigcap E_2'^c)\}$ tends to zero as $n \to \infty$, if for all $s \in [0 : r-1]$,

$$\frac{k}{n}H(W|Q, [W]_s) < \log_2 p^{(r-s)} - H(Z|[Z]_s) - o(\epsilon). \tag{38}$$

Next, we use (38) to show that the bounds in (32) are redundant except the following:

$$R_i + \frac{k}{n} H(W_i|Q) = \log_2 p^r. \tag{39}$$

For that, we compare (39) with the bounds in (32) for different values of s. Noting that

$$H(W_i|Q) = H([W_i]_s|Q) + H(W_i|Q[W_i]_s),$$

it is sufficient to show that

$$\frac{k}{n} H(W_i|Q, [W_i]_s) \leqslant \log_2 \ p^{r-s}.$$

For that, we first prove the following inequality

$$H(W_i|Q, [W_i]_s) \leqslant H(W_1 + W_2|Q, [W_1 + W_2]_s), \tag{40}$$

where $i = 1, 2$, and $0 \leqslant s \leqslant r$. Then, using (38), we get

$$\frac{k}{n} H(W_i|Q, [W_i]_s) \leqslant \log_2 \ p^{r-s}.$$

In what follows, we prove (40). For that

$$\begin{aligned}
&H(W_1 + W_2|Q, [W_1 + W_2]_s) \\
&\quad = H(W_1 + W_2|Q, [[W_1]_s + [W_2]_s]_s) \\
&\quad \geqslant H(W_1 + W_2|Q, [W_1]_s, [W_2]_s) \\
&\quad = H(W_1, W_2|Q, [W_1]_s, [W_2]_s) \\
&\quad\quad - H(W_1|Q, [W_1]_s, [W_2]_s, W_1 + W_2) \\
&\quad \overset{(a)}{=} H(W_2|Q, [W_2]_s) + H(W_1|Q, [W_1]_s) \\
&\quad\quad - H(W_1|Q, [W_1]_s, [W_2]_s, W_1 + W_2) \\
&\quad \overset{(b)}{=} H(W_2|Q, [W_2]_s) + I(W_1; W_1 + W_2|Q, [W_1]_s, [W_2]) \\
&\quad \geqslant H(W_2|Q, [W_2]_s),
\end{aligned}$$

where (a) and (b) hold because of the Markov chain $W_1 \leftrightarrow Q \leftrightarrow W_2$. Similarly, we can show that

$$H(W_1 + W_2|Q, [W_1 + W_2]_s) \geqslant H(W_1|Q, [W_1]_s).$$

Finally, using (39) and (38) the following holds

$$\begin{aligned}
R_i &\geqslant \log_2 p^r - \\
&\min_{0 \leqslant s \leqslant r-1} \frac{H(W_i|Q)}{H(W_1 + W_2|Q, [W_1 + W_2]_s)} \Big(\log_2 p^{(r-s)} \\
&\qquad\qquad - H(Z|[Z]_s) \Big),
\end{aligned} \tag{41}$$

where we minimize the above bound over all PMFs of the form

$$P_{QW_1V_1W_2V_2} = P_Q \prod_i \left(P_{V_i|Q} P_{W_i|Q} \right),$$

such that $p(q)$ is a rational number for all $q \in \mathcal{Q}$. Since rational numbers are dense in $\mathbb{R}$, one can consider arbitrary PMF $p(q)$. Lastly, in the next lemma, we show that the cardinality bound $|\mathcal{Q}| \leqslant r$ is sufficient to optimize (41).

Lemma 8. *The cardinality of $\mathcal{Q}$ is bounded by $|\mathcal{Q}| \leqslant r$.*

Proof: Note that (38) and (39) give an alternative characterization of the achievable region. Using these equations, observe that this region is convex in $\mathbb{R}^2$. As a result, we can characterize the achievable region by its supporting hyper-planes. Let

$$\bar{R}_i \triangleq \log_2 \ p^r - R_i, \quad i = 1, 2.$$

Using (41) for any $0 \leqslant \alpha \leqslant 1$ the corresponding supporting hyper-plane is characterized by

$$\begin{aligned}
&\big(\alpha \bar{R}_1 + (1-\alpha)\bar{R}_2\big) H(W|Q, [W]_s) - \\
&\Big(\alpha H(W_1|Q) + (1-\alpha) H(W_2|Q)\Big) \Big(\log_2 p^{(r-s)} \\
&\qquad\qquad - H(Z|[Z]_s) \Big) \leqslant 0,
\end{aligned} \tag{42}$$

where $s \in [0, r-1]$. We use the support lemma for the above inequalities to bound $|\mathcal{Q}|$. To this end, we first show that the left-hand side of these inequalities are continuous functions of conditional PMF's of W_1 and W_2 given Q. Let $\mathscr{P}_r$ denote the set of all product PMF's on $\mathbb{Z}_{p^r} \times \mathbb{Z}_{p^r}$. Note $\mathscr{P}_r$ is a compact set. Fix $q \in \mathcal{Q}$. Denote

$$f(p(w_1|q)p(w_2|q)) = \alpha H(W_1|Q = q) + (1-\alpha) H(W_2|Q = q)$$

and

$$g_s(p(w_1|q)p(w_2|q)) = H(W_1 + W_2|Q = q, [W_1 + W_2]_s),$$

where $s \in [0 : r-1]$. We show that $f(\cdot), g_s(\cdot)$ are real valued continuous functions of $\mathscr{P}_r$. Since the entropy function is continuous then so is f. We can write

$$\begin{aligned}
g_s(p(w_1|q)p(w_2|q)) &= H(W_1 + W_2|Q = q) \\
&\quad - H([W_1 + W_2]_s|Q = q).
\end{aligned}$$

Note that $[\cdot]_s$ is a continuous function from $\mathscr{P}_r$ to $\mathscr{P}_r$. This implies that $H([\cdot]_s)$ is also continuous. So g_s is continuous. As a result, the left-hand side of the bounds in (42) are real valued continuous functions of $\mathscr{P}_r$. Therefore, we can apply the support lemma [44]. Since there are r bounds for different values of s, then $|\mathcal{Q}| \leqslant r$. ■

Appendix E
Proof of Theorem 3

Fix positive integer n, and define $l \triangleq cn$, and $k \triangleq \tilde{c}n$, where $\tilde{c}$ and c are positive real numbers such that l and k are integers.

Codebook Generation: We use two nested QGC's, one for each encoder. The codebook for Encoder 1 is an (n, k, l) nested QGC (as in Definition 5) with random variables (W_1, V_1, Q). Let $\mathcal{C}_{I,1}, \bar{\mathcal{C}}_1$, and $\mathcal{C}_{O,1}$ denote the corresponding inner code, shift code and the outer code (as in Definition 5), respectively. The codebook for Encoder 2 is an (n, k, l) nested QGC with random variables (W_2, V_2, Q), inner code $\mathcal{C}_{I,2}$, shift code $\bar{\mathcal{C}}_2$, and outer code $\mathcal{C}_{O,2}$. For the decoder, we use $\mathcal{C}_{O,1} + \mathcal{C}_{O,2}$ as a codebook. Conditioned on Q, the random variables (W_1, W_2, V_1, V_2) are mutually independent.

The nested QGCs and $\mathcal{C}_d$ have identical generator matrices but different translations and index random variables. Note that each nested QGC has two generator matrices/translations, one for the inner code and one for the shift code as in Definition 5. The generator matrix and the translation for the inner codes $\mathcal{C}_{I,i}, i = 1, 2$, are denoted by $\mathbf{G}$ and $\mathbf{b}$, respectively. The generator matrix and the translation used for shift code $\mathcal{C}_{I,i}$, are denoted by $\bar{\mathbf{G}}$ and $\bar{\mathbf{b}}_i$, respectively, where $i = 1, 2$. The elements of $\mathbf{G}, \bar{\mathbf{G}}, \mathbf{b}$, and $\bar{\mathbf{b}}_i, i = 1, 2$ are generated randomly and independently from $\mathbb{Z}_{p^r}$. By R_i denote the rate of $\bar{\mathcal{C}}_i$, and let $R_{I,i}$ be the rate of $\mathcal{C}_{I,i}$, where $i = 1, 2$.

Encoding: Index the codewords of $\bar{\mathcal{C}}_i, i = 1, 2$. Upon receiving a message index θ_i, the ith encoder finds the codeword $\mathbf{c}_i \in \bar{\mathcal{C}}_i$ with that index. Then it finds $\mathbf{c}_{I,i} \in \mathcal{C}_{I,i}$ such that $\mathbf{c}_i + \mathbf{c}_{I,i}$ is ϵ-typical with respect to P_{X_i}. If such codeword was found, the encoder i sends $\mathbf{x}_i = \mathbf{c}_i + \mathbf{c}_{I,i}, i = 1, 2$. Otherwise, an error event $E_i, i = 1, 2$ is declared.

Decoding: The channel takes $\mathbf{x}_1$ and $\mathbf{x}_2$ and produces $\mathbf{y}$. Upon receiving $\mathbf{y}$ from the channel, the decoder wishes to decode $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$. It finds $\tilde{\mathbf{x}} \in \mathcal{C}_{O,1} + \mathcal{C}_{O,2}$ such that $\tilde{\mathbf{x}}$ and $\mathbf{y}$ are jointly $\tilde{\epsilon}$-typical with respect to the distribution $P_{X_1+X_2,Y}$. An error event E_d is declared, if no unique $\tilde{\mathbf{x}}$ was found.

Probability of Error: Let $(f_1(\cdot), f_2(\cdot))$ and $g(\cdot,\cdot)$ denote the encoding and decoding functions corresponding to the above coding scheme. The overall error event is defined as

$$E \triangleq \{g(Y^n) \neq f_1(M_1) + f_2(M_2)\}.$$

For the achievability, we need to show that $P(E)$ can be made arbitrary small for sufficiently large n. If (X_1^n, X_2^n) denote the outputs of the encoders, define an error event E_c as the event in which $(X_1^n, X_2^n) \notin A_\epsilon^{(n)}(X_1, X_2)$. Next, using the aforementioned encoding and decoding error events we have

$$\begin{aligned} P(E) \leqslant & P(E_1 \bigcup E_2 \bigcup E_d \bigcup E_c) \\ & + P(E|E_1^c \bigcap E_2^c \bigcap E_d^c \bigcap E_c^c). \end{aligned}$$

Using standard arguments for typical sequences, we can show that when there is no encoding and decoding error (i.e., $E_1^c \bigcap E_2^c \bigcap E_d^c \bigcap E_c^c$) the error probability $P(E|E_1^c \bigcap E_2^c \bigcap E_d^c \bigcap E_c^c)$ approaches 0 as $n \to \infty$. As a result, the second term above is sufficiently small for large enough n. Therefore, for sufficiently large n and from the union bound on the first term we obtain,

$$P(E) \leqslant P(E_1) + P(E_2) + P(E_d) + P(E_c) + \epsilon.$$

We need to find conditions for which the probability of the error events E_1, E_2, E_d and E_c approach zero. For any $\mathbf{a} \in \mathbb{Z}_{p^r}^k$ and $\bar{\mathbf{a}} \in \mathbb{Z}_{p^r}^l$ define the map $\phi(\mathbf{a}, \bar{\mathbf{a}}) = \mathbf{a}\mathbf{G} + \bar{\mathbf{a}}\bar{\mathbf{G}}$. By $\Phi(\cdot,\cdot)$ denote the map ϕ whose matrices are selected randomly and uniformly.

A. Analysis of E_1, E_2

For any sequence $\mathbf{v}_i \in \mathcal{V}_i$ define

$$\lambda_i(\mathbf{v}_i) = \sum_{\mathbf{w}_i \in \mathcal{W}_i} \sum_{\mathbf{x}_i \in A_\epsilon^{(n)}(X_i)} \mathbb{1}\{\mathbf{x}_i = \phi(\mathbf{w}_i, \mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i\},$$

where $i = 1, 2$. Therefore, E_i occurs if $\lambda_i(\mathbf{v}_i) = 0$. For more convenience, we weaken the definition of event E_i. We say E_i occurs, if $\lambda_i(\mathbf{v}_i) < \frac{1}{2}E(\lambda_i(v_i))$. Using Lemma 5 we can show that $P(E_i) \to 0$ as $n \to \infty$, if

$$\frac{k}{n}H([W_i]_t|Q) \geqslant \log_2 p^t - H([X_i]_t) + \gamma(\epsilon), \tag{43}$$

holds for $i = 1, 2$, and $1 \leqslant t \leqslant r$, where γ is a function satisfying $\lim_{\epsilon\to 0} \gamma(\epsilon) = 0$.

B. Analysis of E_c

Define the set

$$\mathcal{E} \triangleq \Big\{(\mathbf{x}_1, \mathbf{x}_2) \in A_\epsilon^{(n)}(X_1) \times A_\epsilon^{(n)}(X_2) : (\mathbf{x}_1, \mathbf{x}_2) \notin A_\epsilon^{(n)}(X_1, X_2)\Big\}.$$

Therefore, probability of E_c can be written as

$$P(E_c|E_1^c \bigcap E_2^c) = \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} P(e_1(\Theta_1) = \mathbf{x}_1, e_2(\Theta_2) = \mathbf{x}_2),$$

where e_i is the output of the ith encoder, and Θ_i is the random message to be transmitted by encoder i, where $i = 1, 2$. By $P(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i)$ denote the probability that $(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i)$ is selected at the ith encoder. Then,

$$P(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i) = \frac{1}{|\mathcal{V}_i|}\frac{1}{\lambda_i(\mathbf{v}_i)}\mathbb{1}\{\phi(\mathbf{w}_i, \mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i = \mathbf{x}_i\}.$$

By the definition of $\phi_1(\cdot)$ and $\phi_2(\cdot)$, we have

$$\begin{aligned} P(E_c|E_1^c \bigcap E_2^c) = & \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} \prod_{i=1}^{2}\Bigg[\sum_{\mathbf{v}_i\in\mathcal{V}_i}\sum_{\mathbf{w}_i\in\mathcal{W}_i}\frac{1}{|\mathcal{V}_i|}\frac{1}{\lambda_i(\mathbf{v}_i)} \times \\ & \mathbb{1}\Big\{\mathbf{x}_i = \phi_i(\mathbf{w}_i, \mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i\Big\}\Bigg]. \end{aligned}$$

Since there is no encoding error (for the modified version), then $\lambda_i(\mathbf{v}_i) \geqslant \frac{1}{2}E[\lambda_i(\mathbf{v}_i)], i = 1, 2$. Therefore, replacing $\lambda_i(\mathbf{v}_i)$ in the above expression with $\frac{1}{2}E[\lambda_i(\mathbf{v}_i)]$ gives an upper bound on $P(E_c|E_1^c \bigcap E_2^c)$. Next, we take expectation over all ϕ_1 and ϕ_2. We have

$$\begin{aligned} & \mathbb{E}\{P(E_c|E_1^c \bigcap E_2^c)\} \leqslant \\ & \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} \sum_{\substack{\mathbf{v}_i\in\mathcal{V}_i \\ i=1,2}} \sum_{\substack{\mathbf{w}_i\in\mathcal{W}_i \\ i=1,2}} \left[\prod_{j=1}^{2}\frac{4}{|\mathcal{V}_j|E[\lambda_j(\mathbf{v}_j)]}\right] \times \\ & \qquad P\Big\{\mathbf{x}_i = \Phi_i(\mathbf{w}_i, \mathbf{v}_i) + \mathbf{B} + \bar{\mathbf{B}}_i, i = 1, 2\Big\} \\ & \overset{(a)}{=} \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} \sum_{\substack{\mathbf{v}_i\in\mathcal{V}_i \\ i=1,2}} \sum_{\substack{\mathbf{w}_i\in\mathcal{W}_i \\ i=1,2}} \left[\prod_{j=1}^{2}\frac{4}{|\mathcal{V}_j|E[\lambda_j(\mathbf{v}_j)]}\right] p^{-2nr} \\ & = \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} |\mathcal{W}_1||\mathcal{W}_2|\frac{4}{E[\lambda_1(\mathbf{v}_1)]E[\lambda_2(\mathbf{v}_2)]}p^{-2nr}. \end{aligned} \tag{44}$$

Note that (a) is because $\mathbf{B}_1$ and $\mathbf{B}_2$ are independent random vectors with uniform distribution over $\mathbb{Z}_{p^r}^n$. From the definition of $\lambda_j(\mathbf{v}_j), j = 1, 2$, we have

$$E[\lambda_j(\mathbf{v}_j)] = |\mathcal{W}_j||A_\epsilon^{(n)}(X_i)|p^{-nr}.$$

As a result of the above equation and (44),

$$\mathbb{E}\{P(E_c|E_1^c \bigcap E_2^c)\} \leqslant \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} 4|A_\epsilon^{(n)}(X_1)|^{-1}|A_\epsilon^{(n)}(X_2)|^{-1}.$$

There exists a continuous function $\delta(\epsilon) > 0$ with $\delta(0) = 0$ such that for any $\mathbf{x}_i \in A_\epsilon^{(n)}(X_i)$, we have

$$P_{X_i}^n(\mathbf{x}_i) \geqslant |A_\epsilon^{(n)}(X_i)|^{-1}2^{-\delta(\epsilon)}.$$

Thus,

$$\mathbb{E}\{P(E_c \bigcap E_1^c \bigcap E_2^c)\} \leqslant \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} P^n_{X_1}(\mathbf{x}_1)P^n_{X_2}(\mathbf{x}_2)2^{n2\delta(\epsilon)} = 2^{n2\delta(\epsilon)}P^n_{X_1X_2}(\mathcal{E}).$$

Thus, $\mathbb{E}\{P(E_c|E_1^c \bigcap E_2^c)\} \to 0$ as $n \to \infty$.

C. Analysis of E_d

In what follows, to make the analysis tractable, we define an alternative decoding error. Upon receiving $\mathbf{y}$, the decoder finds $\tilde{\mathbf{w}} \in A_\epsilon^{(n)}(W_1+W_2)$ and $\tilde{\mathbf{v}} \in A_\epsilon^{(n)}(V_1+V_2)$ such that $\phi(\tilde{\mathbf{w}},\tilde{\mathbf{v}}) + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2$ is jointly typical with $\mathbf{y}$ with respect to $P_{X_1+X_2,Y}$. For the alternative decoder, we define a new decoding error. A decoding error E'_d occurs, if $(\tilde{\mathbf{w}},\tilde{\mathbf{v}})$ is not unique. With this definition $E_d \subseteq E'_d$. Because, the mapping $\mathbf{x}_i = \phi(\mathbf{w}_i,\mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i$ is not necessarily injective. Note that the new decoder is required to decode $\mathbf{w}_1+\mathbf{w}_2$ and $\mathbf{v}_1+\mathbf{v}_2$. This is a more restrictive condition than decoding $\mathbf{x}_1+\mathbf{x}_2$. Therefore, it is sufficient to show that $P(E'_d) \to 0$ as $n \to \infty$. In what follows, we provide an upper bound on $P(E'_d)$.

Since the probability of the encoding errors E_1, E_2 and E_c are sufficiently small, then

$$P(E'_d) \approx P(E'_d \cap E_1^c \bigcap E_2^c \bigcap E_c^c).$$

We show that this probability approaches zero as $n \to \infty$. Fix $\phi, \mathbf{b}$ and $\bar{\mathbf{b}}_i, i = 1,2$. Note that By $P(\mathbf{v}_i,\mathbf{w}_i,\mathbf{x}_i)$ denote the probability that $(\mathbf{v}_i,\mathbf{w}_i,\mathbf{x}_i)$ is selected at the ith encoder. Then,

$$P(\mathbf{v}_i,\mathbf{w}_i,\mathbf{x}_i) = \frac{1}{|\mathcal{V}_i|}\frac{1}{\lambda_i(\mathbf{v}_i)}\mathbb{1}\{\phi(\mathbf{w}_i,\mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i = \mathbf{x}_i\}.$$

Then the probability of $E'_d \cap E_1^c \bigcap E_2^c \bigcap E_c^c$ equals

$$\begin{aligned}&P(E'_d \cap E_1^c \bigcap E_2^c \bigcap E_c^c) = \\ &\left[\prod_{j=1}^{2}\sum_{\mathbf{v}_j\in\mathcal{V}_j}\sum_{\mathbf{w}_j\in\mathcal{W}_j}\mathbb{1}\Big\{\lambda_i(\mathbf{v}_i) \geqslant 1/2\ E(\lambda_i(\mathbf{v}_i)), i = 1,2\Big\}\right] \\ &\times \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in A_\epsilon^{(n)}(X_1,X_2)}\sum_{\mathbf{y}\in\mathcal{Y}^n} P(\mathbf{v}_i,\mathbf{w}_i,\mathbf{x}_i, i = 1,2) \\ &P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2)P\Big(E_d \mid E_1^c \bigcap E_2^c \bigcap E_c^c, \\ &\qquad \mathbf{y},\mathbf{x}_i,\mathbf{v}_i,\mathbf{w}_i, i = 1,2\Big).\end{aligned}$$

Next, we bound $P(E'_d \mid E_1^c \bigcap E_2^c \bigcap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1,2)$, and $P(\mathbf{v}_i\mathbf{w}_i,\mathbf{x}_i, i = 1,2)$.

$$\begin{aligned}&P(E'_d \mid E_1^c \bigcap E_2^c \bigcap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1,2) = \\ &\mathbb{1}\Big\{\exists\ (\tilde{\mathbf{w}},\tilde{\mathbf{v}}) \in \mathcal{W}\times\mathcal{V} : (\tilde{\mathbf{w}},\tilde{\mathbf{v}}) \neq (\mathbf{w}_1+\mathbf{w}_2, \mathbf{v}_1+\mathbf{v}_2), \\ &\qquad \phi(\tilde{\mathbf{w}},\tilde{\mathbf{v}}) + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 \in A^n_{\epsilon'}(Z|\mathbf{y})\Big\},\end{aligned}$$

where $\mathcal{W} \triangleq A_\epsilon^{(n)}(W_1+W_2)$, $\mathcal{V} \triangleq A_\epsilon^{(n)}(V_1+V_2)$, and $Z \triangleq X_1 + X_2$. Using the union bound, we have

$$\begin{aligned}&P(E'_d \mid E_1^c \bigcap E_2^c \bigcap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1,2) \leqslant \\ &\sum_{\substack{\tilde{\mathbf{w}}\in\mathcal{W}\\ \tilde{\mathbf{w}}\neq\mathbf{w}_1+\mathbf{w}_2}}\sum_{\substack{\tilde{\mathbf{v}}\in\mathcal{V}\\ \tilde{\mathbf{v}}\neq\mathbf{v}_1+\mathbf{v}_2}}\sum_{\tilde{\mathbf{z}}\in A^{(n)}_{\epsilon'}(Z|\mathbf{y})}\mathbb{1}\{\phi(\tilde{\mathbf{w}},\tilde{\mathbf{v}}) + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 = \tilde{\mathbf{z}}\}\end{aligned} \tag{45}$$

Note that $P(\mathbf{v}_i,\mathbf{w}_i,\mathbf{x}_i, i = 1,2) = \prod_{i=1,2}P(\mathbf{v}_i,\mathbf{w}_i,\mathbf{x}_i)$. Since there is no encoding error, $\lambda_i(\mathbf{v}_i) \geqslant \frac{1}{2}E(\lambda_i(\mathbf{v}_i))$. As a result,

$$P(\mathbf{v}_i,\mathbf{w}_i,\mathbf{x}_i) \leqslant \frac{1}{|\mathcal{V}_i|}\frac{2}{E(\lambda_i(\mathbf{v}_i))}\mathbb{1}\{\phi(\mathbf{w}_i,\mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i = \mathbf{x}_i\} \tag{46}$$

Therefore, using (46), we have

$$\begin{aligned}&P(E'_d \bigcap E_1^c \bigcap E_2^c \bigcap E_c^c) \leqslant \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in A_\epsilon^{(n)}(X_1,X_2)} \\ &\left[\prod_{j=1}^{2}\sum_{\mathbf{v}_j\in\mathcal{V}_j}\sum_{\mathbf{w}_j\in\mathcal{W}_j}\mathbb{1}\Big\{\lambda_j(\mathbf{v}_j) \geqslant 1/2\ E(\lambda_j(\mathbf{v}_j))\Big\}\frac{1}{|\mathcal{V}_j|}\frac{2}{E(\lambda_i(\mathbf{v}_j))}\right. \\ &\left.\times\,\mathbb{1}\{\phi(\mathbf{w}_j,\mathbf{v}_j) + \mathbf{b} + \bar{\mathbf{b}}_j = \mathbf{x}_j\}\right]\sum_{\mathbf{y}\in\mathcal{Y}^n}P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) \\ &\times P\Big(E'_d \mid E_1^c \bigcap E_2^c \bigcap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1,2\Big) \\ &\leqslant \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in A_\epsilon^{(n)}(X_1,X_2)}\left[\prod_{j=1}^{2}\sum_{\mathbf{v}_j\in\mathcal{V}_j}\sum_{\mathbf{w}_j\in\mathcal{W}_j}\frac{1}{|\mathcal{V}_j|}\frac{2}{E(\lambda_i(\mathbf{v}_j))}\right. \\ &\left.\times\,\mathbb{1}\{\phi(\mathbf{w}_j,\mathbf{v}_j) + \mathbf{b} + \bar{\mathbf{b}}_j = \mathbf{x}_j\}\right]\sum_{\mathbf{y}\in\mathcal{Y}^n}P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) \\ &\times P(E'_d \mid E_1^c \bigcap E_2^c \bigcap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1,2).\end{aligned} \tag{47}$$

The last inequality follows by eliminating the indicator function on $\big\{\lambda_i(\mathbf{v}_i) \geqslant 1/2\ E(\lambda_i(\mathbf{v}_i)), i = 1,2\big\}$. Note that for jointly ϵ-typical sequences $\mathbf{x}_1, \mathbf{x}_2$ and large enough n, we have

$$P(\mathbf{Y}^n \notin A_\epsilon^{(n)}(Y|\mathbf{x}_1,\mathbf{x}_2)) \leqslant \frac{c}{n\tilde{\epsilon}^2},$$

where c is a constant. This follows from the standard arguments on typical sets. Thus, using (47) and (45) we get

$$\begin{aligned}&P(E'_d \bigcap E_1^c \bigcap E_2^c \bigcap E_c^c) \leqslant \frac{c}{n\tilde{\epsilon}^2} + \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in A_\epsilon^{(n)}(X_1,X_2)} \\ &\left[\prod_{j=1}^{2}\sum_{\mathbf{v}_j\in\mathcal{V}_j}\sum_{\mathbf{w}_j\in\mathcal{W}_j}\frac{1}{|\mathcal{V}_j|}\frac{2\mathbb{1}\{\phi(\mathbf{w}_j,\mathbf{v}_j) + \mathbf{b} + \bar{\mathbf{b}}_j = \mathbf{x}_j\}}{E(\lambda_i(\mathbf{v}_j))}\right] \\ &\times \sum_{\mathbf{y}\in A_\epsilon^n(Y|\mathbf{x}_1,\mathbf{x}_2)}P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2)\sum_{\substack{\tilde{\mathbf{w}}\in\mathcal{W}\\ \tilde{\mathbf{w}}\neq\mathbf{w}_1+\mathbf{w}_2}}\sum_{\substack{\tilde{\mathbf{v}}\in\mathcal{V}\\ \tilde{\mathbf{v}}\neq\mathbf{v}_1+\mathbf{v}_2}} \\ &\sum_{\tilde{\mathbf{z}}\in A^{(n)}_{\epsilon'}(Z|\mathbf{y})}\mathbb{1}\Big\{\phi(\tilde{\mathbf{w}},\tilde{\mathbf{v}}) + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 = \tilde{\mathbf{z}}\Big\}.\end{aligned}$$

Next, we take the average of the above expression over all maps ϕ, and all vectors $\mathbf{b}, \bar{\mathbf{b}}_i, i = 1, 2$.

$$\mathbb{E}\{P(E'_d \bigcap E_1^c \bigcap E_2^c \bigcap E_c^c)\} \leqslant \frac{c}{n\tilde{\epsilon}^2} + \left[\prod_{j=1}^{2} \sum_{\mathbf{v}_j \in \mathcal{V}_j} \sum_{\mathbf{w}_j \in \mathcal{W}_j} \frac{1}{|\mathcal{V}_j|} \frac{2}{E(\lambda_j(\mathbf{v}_j))}\right] \times \sum_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}) \in A_\epsilon^{(n)}(X_1,X_2,Y)} P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) \sum_{\substack{\tilde{\mathbf{w}} \in \mathcal{W} \\ \tilde{\mathbf{w}} \neq \mathbf{w}_1+\mathbf{w}_2}} \sum_{\substack{\tilde{\mathbf{v}} \in \mathcal{V} \\ \tilde{\mathbf{v}} \neq \mathbf{v}_1+\mathbf{v}_2}} \sum_{\tilde{\mathbf{z}} \in A_{\epsilon'}^{(n)}(Z|\mathbf{y})} P\Big\{\tilde{z} = \Phi(\tilde{\mathbf{w}}, \tilde{\mathbf{v}}) + 2\mathbf{B} + \bar{\mathbf{B}}_1 + \bar{\mathbf{B}}_1, x_1 = \Phi(\mathbf{w}_1, \mathbf{v}_1) + \mathbf{B} + \bar{\mathbf{B}}_1, x_2 = \Phi(\mathbf{w}_2, \mathbf{v}_2) + \mathbf{B} + \bar{\mathbf{B}}_1\Big\}.$$

Notice that $\mathbf{B}, \bar{\mathbf{B}}_1$, and are $\bar{\mathbf{B}}_1$ are uniform over $\mathbb{Z}^n_{p^r}$ and independent of other random variables. Hence, the innermost term in the above summations is simplified to

$$p^{-2nr} P\{\mathbf{z} - \tilde{x_1} - x_2 = \Phi(\tilde{\mathbf{w}} - (\mathbf{w_1} + \mathbf{w_2}), \tilde{\mathbf{v}} - (\mathbf{v_1} + \mathbf{v_2}))\} \tag{48}$$

Using Lemma 11, if

$$\tilde{\mathbf{w}} - (\mathbf{w_1} + \mathbf{w_2}), \tilde{\mathbf{v}} - (\mathbf{v_1} + \mathbf{v_2}) \in H_s^k \backslash H_{s+1}^k,$$

the expression in (48) equals

$$p^{-2nr} p^{-n(r-s)} \mathbb{1}\{\tilde{z} - \mathbf{x_1} - \mathbf{x_2} \in H_s^n\},$$

where $0 \leqslant s \leqslant r-1$. Therefore, $\mathbb{E}\{P(E'_d \bigcap E_1^c \bigcap E_2^c \bigcap E_c^c)\}$ is upper-bounded as

$$\mathbb{E}\{P(E'_d \bigcap E_1^c \bigcap E_2^c \bigcap E_c^c)\} \leqslant \frac{c}{n\tilde{\epsilon}^2} + \left[\prod_{j=1}^{2} \sum_{\mathbf{v}_j \in \mathcal{V}_j} \sum_{\mathbf{w}_j \in \mathcal{W}_j} \frac{1}{|\mathcal{V}_j|} \frac{2}{E(\lambda_j(\mathbf{v}_j))}\right] \times \sum_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}) \in A_{\tilde{\epsilon}}^{(n)}(X_1,X_2,Y)} P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) \sum_{s=0}^{r-1} \sum_{\substack{\tilde{\mathbf{w}} \in \mathcal{W} \\ \tilde{\mathbf{w}} - (\mathbf{w_1}+\mathbf{w_2}) \in H_s^k}} \sum_{\substack{\tilde{\mathbf{v}} \in \mathcal{V} \\ \tilde{\mathbf{v}} - (\mathbf{v_1}+\mathbf{v_2}) \in H_s^k}} \sum_{\substack{\tilde{z} \in A_\epsilon^n(Z|y) \\ \mathbf{z} - \tilde{x_1} - x_2 \in H_s^n}} p^{-2nr} p^{-n(r-s)}. \tag{49}$$

Note that the most inner term in the above summations does not depend on the value of $\tilde{\mathbf{z}}, \tilde{\mathbf{v}}$ and $\tilde{\mathbf{w}}$. Hence, we replace those summations by the size of the corresponding subsets. Using Lemma 12 we can bound the size of these subsets and get the following bound on the probability of error

$$\mathbb{E}\{P(E'_d \bigcap E_1^c \bigcap E_2^c \bigcap E_c^c)\} \leqslant \frac{c}{n\tilde{\epsilon}^2} + \left[\prod_{j=1}^{2} \sum_{\mathbf{v}_j \in \mathcal{V}_j} \sum_{\mathbf{w}_j \in \mathcal{W}_j} \frac{1}{|\mathcal{V}_j|} \frac{2}{E(\lambda_j(\mathbf{v}_j))}\right] \times \sum_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}) \in A_{\tilde{\epsilon}}^{(n)}(X_1,X_2,Y)} P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) \sum_{s=0}^{r-1} 2^{k(H(W|Q,[W]_s)+\eta_1(\epsilon))} \times 2^{l(H(V|Q,[V]_s)+\eta_2(\epsilon))}\, 2^{n(H(Z|Y[Z]_s)+\eta_3(\epsilon))} p^{-2nr} p^{-n(r-s)},$$

where $W = W_1 + W_2, V = V_1 + V_2$, and $\lim_{\epsilon \to 0} \eta_i(\epsilon) = 0, i = 1, 2, 3$. Note that

$$E(\lambda_i(\mathbf{v}_i)) = |\mathcal{W}_i||A_\epsilon^{(n)}(X_i)|p^{-nr}, i = 1, 2.$$

As the terms in the above expression do not depend on the values of $\mathbf{w}_i, \mathbf{v}_i, \mathbf{x}_i, i = 1, 2$ and $\mathbf{y}$, we can replace the summations over them with the corresponding sets. As a result, we have

$$\mathbb{E}\{P(E'_d \bigcap E_1^c \bigcap E_2^c E_c^c)\} \leqslant \frac{c}{n\epsilon^2} + 4 \sum_{s=0}^{r-1} p^{-n(r-s)} 2^{kH(W|Q,[W]_s)} 2^{lH(V|Q,[V]_s)}\, 2^{n(H(Z|Y,[Z]_s)+\delta'(\epsilon))},$$

where $\lim_{\epsilon \to 0} \delta'(\epsilon) = 0$. Therefore, the right-hand side of the above inequality approaches zero as $n \to \infty$, if the following bounds hold:

$$\begin{aligned} &\frac{k}{n} H(W|Q,[W]_s) + \frac{l}{n} H(V|Q,[V]_s) \\ &\leqslant \log_2 p^{r-s} - H(Z|Y[Z]_s) - \delta(\epsilon), \end{aligned} \tag{50}$$

for $0 \leqslant s \leqslant r-1$. Next, we apply Fourier-Motzkin technique [44] to eliminate $\frac{k}{n}$ from (43) and (50). We get

$$\begin{aligned} \frac{l}{n} H(V|Q,[V]_s) &\leqslant \log_2\ p^{r-s} - H(Z|Y[Z]_s) \\ &- \frac{H(W|Q,[W]_s)}{H([W_i]_t|Q)} (\log_2\ p^t - H([X_i]_t)) - o(\epsilon), \end{aligned}$$

where $i = 1, 2,\quad 0 \leqslant s \leqslant r-1$, and $1 \leqslant t \leqslant r$. Note by definition

$$R_i = \frac{1}{n} \log_2 |\bar{\mathcal{C}}_i| \leqslant \frac{1}{n} \log_2 |\mathcal{V}_i| \leqslant \frac{l}{n} H(V_i|Q).$$

Therefore, we obtain the bounds in the theorem. Using the same argument as in Lemma 8, we can bound the cardinality of Q by $|\mathcal{Q}| \leqslant r^2$. This completes the proof.

Appendix F
Proof of Lemma 7

Proof: Consider the bound on the sum-rate given in (17). The set of all (R_1, R_2) satisfying only this bound is an outer-bound for $\mathscr{R}_{GP}$. The time-sharing random variable Q is trivial for this outer-bound, because there is only one inequality on the rates, and because of the cost constraints $\mathbb{E}\{c_i(X_i)\} = 0, i = 1, 2$. For any distribution $P \in \mathscr{P}_{GP}$, we obtain

$$\begin{aligned} R_1 + R_2 &\leqslant I(U_1, U_2; Y) - I(U_1; S_1) - I(U_2; S_2) \\ &= H(Y) - H(Y|U_1, U_2) - H(S_1) \\ &\quad + H(S_1|U_1) - H(S_2) + H(S_2|U_2) \\ &\leqslant H(S_1|U_1) + H(S_2|U_2) - H(Y|U_1, U_2) - 2 \\ &= \max_{P \in \mathscr{P}_{GP}} \sum_{u_1 \in \mathcal{U}_1} \sum_{u_2 \in \mathcal{U}_2} p(u_1, u_2)\Big(H(S_1|u_1) \\ &\qquad + H(S_2|u_2) - H(Y|u_1, u_2) - 2\Big), \end{aligned} \tag{51}$$

where the second inequality holds, as $H(Y) \leqslant 2$, and $H(S_i) = 2$ for $i = 1, 2$. In the next step, we relax the conditions in $\mathscr{P}_{GP}$, and provide an upper-bound on (51). For

$i = 1, 2$, and any $u_i \in \mathcal{U}_i$, define $\mathscr{P}_{u_i}$ as the collection of all conditional PMFs $p(s_i, x_i|u_i)$ on $\mathbb{Z}_4^2$ such that

1) $X_i = f_i(S_i, u_i)$ for some function f_i,
2) $E(c_i(X_i)|u_i) = 0$.

In the first condition, given u_i, $f_i(s_i, u_i)$ can be thought as a function g_{u_i} of s_i. For different u_i's we have different functions $g_{u_i}(s_i)$. The second condition is implied from the cost constraint $E(c_i(X_i)) = 0$, because without loss of generality we assume $p(u_i) > 0$ for all $u_i \in \mathcal{U}_i$. Also, note that we removed the condition that S_i is uniform over $\mathbb{Z}_4$. Hence, $\mathscr{P}_{GP}$ is a subset of the set of all PMFs of the form $P = \prod_{i=1}^2 p(u_i)p(s_i, x_i|u_i)$, where $p(s_i, x_i|u_i) \in \mathscr{P}_{u_i}$, $i = 1, 2$.

As a result, (51) is upper-bounded by

$$\begin{aligned} R_1 + R_2 \leqslant \max_{p(u_1),p(u_2)} \max_{\substack{p(s_i,x_i|u_i)\in\mathscr{P}_{u_i}\\ i=1,2}} \\ \sum_{\substack{u_1\in\mathcal{U}_1\\ u_2\in\mathcal{U}_2}} p(u_1,u_2)\Big(H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1,u_2) - 2\Big) \\ \leqslant \max_{u_1\in\mathcal{U}_1, u_2\in\mathcal{U}_2} \max_{\substack{p(s_i,x_i|u_i)\in\mathscr{P}_{u_i}\\ i=1,2}} \\ \Big(H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1,u_2) - 2\Big). \end{aligned}$$

Fix $u_2 \in \mathcal{U}_2$ and $p(s_2, x_2|u_2) \in \mathscr{P}_{u_2}$. We maximize over all $u_1 \in \mathcal{U}_1$ and $p(s_1, x_1|u_1) \in \mathscr{P}_{u_1}$. Let $N = X_2 + S_2$, where X_2 and S_2 are distributed according to $p(s_2, x_2|u_2)$. For fixed $u_2 \in \mathcal{U}_2$, by $Q_{u_2} \in \mathscr{P}_{u_2}$ denote the PMF $p(s_2, x_2|u_2)$. This maximization problem is equivalent to finding

$$\begin{aligned} R(u_2, Q_{u_2}) \triangleq H(S_2|u_2) + \max_{u_1\in\mathcal{U}_1} \\ \max_{p(s_1,x_1|u_1)\in\mathscr{P}_{u_1}} H(S_1|u_1) - H(X_1 + S_1 + N|u_1) - 2. \quad (52) \end{aligned}$$

Consider the problem of PtP channel with state, where the channel is $Y = X_1 + S_1 + N$. It can be shown that

$$R(u_2, Q_{u_2}) - H(S_2|u_2)$$

is an upper-bound on the capacity of this problem. We proceed by the following lemma.

Lemma 9. *The following bound holds $R(u_2, Q_{u_2}) < 1$ for all $u_2 \in \mathcal{U}_2$ and $Q_{u_2} \in \mathscr{P}_{u_2}$.*

Proof: The proof is given in Appendix G. ∎

Finally, as a result of the above lemma the proof is completed. ∎

Appendix G
Proof of Lemma 9

Proof: Note that for any fixed $u_2 \in \mathcal{U}_2$, the distribution of N depends on the conditional PMF $p(s_1|u_1)$, and the function $x_1 = f_1(s_1, u_1)$. For any $u \in \mathcal{U}_2$ define

$$\mathcal{L}_u := \{f_2(u, s) + s : s \in \mathbb{Z}_4\}.$$

For any given $i \in \{1, 2, 3, 4\}$, define

$$\mathcal{B}_i \triangleq \{u \in \mathcal{U}_2 : |\mathcal{L}_u| = i\}.$$

Note that $\mathcal{B}_i$'s are disjoint and $\mathcal{U}_2 = \bigcup_i \mathcal{B}_i$. Depending on u_2, we consider four cases. In what follows, for each case, we derive an upper bound on (52). Consider the PMF $p(\omega)$ on $\mathbb{Z}_4$. For brevity, we represent this PMF by the vector $\mathbf{p} := (p(0), p(1), p(2), p(3))$.

Case 1: $u_2 \in \mathcal{B}_1$

Since $|\mathcal{L}_{u_2}| = 1$, then for all $s_2 \in \mathbb{Z}_4$ the following holds

$$s_2 + f_2(s_2, u_2) = a,$$

where $a \in \mathbb{Z}_4$ is a constant that only depends on u_2. This implies that conditioned on u_2, $X_2 + S_2$ equals to a constant a, with probability one. Therefore,

$$\begin{aligned} H(X_1 + S_1 + X_2 + S_2|u_2, u_1) &= H(X_1 + S_1 + a|u_1, u_2) \\ &= H(X_1 + S_1|u_1). \end{aligned}$$

Moreover,

$$H(S_2|u_2) = H(a \ominus X_2|u_2) = H(X_2|u_2).$$

By assumption $p(u_2) > 0$. Therefore, the cost constraint $\mathbb{E}(c_2(X_2)) = 0$ implies that $\mathbb{E}(c_2(X_2)|U_2 = u_2) = 0$. Hence, given $U_2 = u_2$, the random variable X_2 takes at most two values with positive probabilities. As a result, $H(X_2|u_2) \leqslant 1$. Given this inequality, we obtain

$$R(u_2, Q_{u_2}) \leqslant H(S_1|u_1) - H(X_1 + S_1|u_1) - 1 \leqslant 0$$

where the last inequality follows by Lemma 14 in Appendix H.

Case 2: $u_2 \in \mathcal{B}_2$

For any fixed $u_2 \in \mathcal{B}_2$, $f_2(s_2, u_2) + s_2$ takes two values for all $s_2 \in \mathbb{Z}_4$. Assume these values are $a, b \in \mathbb{Z}_4$, where $a \neq b$. Given u_2 the random variable $X_2 + S_2$ is distributed over $\{a, b\}$. Therefore, $X_2 + S_2 \ominus a$ is distributed over $\{0, b \ominus a\}$, and

$$\begin{aligned} H(X_1 + S_1 + X_2 + S_2|u_2, u_1) = \\ H(X_1 + S_1 + X_2 + S_2 \ominus a|u_2, u_1). \end{aligned}$$

As a result, the case $\{a, b\}$ gives the same bound as $\{0, b \ominus a\}$, and we need to consider only the case in which $a = 0$. For the case in which $a = 0$, and $b = 3$, consider $X_2 + S_2 + 1$. Using a similar argument as above, we can show that when $b = 3$, we get the same bound when $b = 1$. Therefore, we only need to consider the cases in which $a = 0$, and $b \in \{1, 2\}$. We address these cases in the next Claim.

Claim 1. *Let $P(X_2 + S_2 = 0|u_1) = p_0$. The following holds:*
1) If $b = 2$, then

$$\begin{aligned} R(u_2, Q_{u_2}) \leqslant \beta\Big(H(S_1|u_1) - H(X_1 + S_1 + N_{(2/3,0,1/3,0)}|u_1)\Big) \\ + (1-\beta)\Big(H(S_1|u_1) - H(X_1 + S_1 + N_{(1/3,0,2/3,0)}|u_1)\Big) \\ + H(S_2|u_2) - 2. \end{aligned}$$

2) If $b = 1$, then

$$R(u_2, Q_{u_2}) \leqslant \beta\Big(H(S_1|u_1) - H(X_1+S_1+N_{(2/3,1/3,0,0)}|u_1)\Big) + (1-\beta)\Big(H(S_1|u_1) - H(X_1+S_1+N_{(1/3,2/3,0,0)}|u_1)\Big) + H(S_2|u_2) - 2.$$

Proof: The proof is given in Appendix I. ∎

Using the claim and applying Lemma 14, we have

$$R(u_2, Q_{u_2}) < 1 + H(S_2|u_2) - 2 \leqslant 1.$$

Case 3: $u_2 \in \mathcal{B}_3$

We need only to consider the case when $\mathbf{p} = (p_0, p_1, p_2, 0)$. We proceed by the following claim.

Claim 2. *If $u_2 \in \mathcal{B}_3$, the following bound holds*

$$\begin{aligned} R(u_2, Q_{u_2}) &\leqslant \beta_0\Big(H(S_1|u_1) - H(X_1+S_1+N_{(2/4,1/4,1/4,0)}|u_1)\Big) \\ &+ \beta_1\Big(H(S_1|u_1) - H(X_1+S_1+N_{(1/4,2/4,1/4,0)}|u_1)\Big) \\ &+ \beta_2\Big(H(S_1|u_1) - H(X_1+S_1+N_{(1/4,1/4,2/4,0)}|u_1)\Big) \\ &+ H(S_2|u_2) - 2, \end{aligned}$$

where $\beta_i = 4p_i - 1,\ i = 0, 1, 2$.

Proof: Similar to Claim 1, we can write $\mathbf{p}$ as a linear combination of three distributions of the form

$$\mathbf{p} = \beta_0 \times [2/4, 1/4, 1/4, 0] + \beta_1 \times [1/4, 2/4, 1/4, 0] + \beta_2 \times [1/4, 1/4, 2/4, 0],$$

where $\beta_i = 4p_i - 1,\ i = 0, 1, 2$. The proof then follows from the concavity of the entropy. ∎

Therefore, by Lemma 14, we obtain

$$R(u_2, Q_{u_2}) < 1 + H(S_2|u_2) - 2 \leqslant 1.$$

Case 4: $u_2 \in \mathcal{B}_4$

In this case, there is a 1-1 correspondence between $x_2(s_2, u_2) + s_2$ and s_2. Therefore

$$H(S_2|u_1, u_2) = H(S_2 + X_2|u_1, u_2),$$

and we obtain

$$\begin{aligned} &H(S_2|u_1, u_2) - H(X_1 + S_1 + X_2 + S_2|u_1, u_2) \\ &= H(S_2 + X_2|u_1, u_2) - H(X_1 + S_1 + X_2 + S_2|u_1, u_2) \\ &\leqslant 0. \end{aligned}$$

Therefore

$$H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1u_2) - 2 \leqslant H(S_1|u_1) - 2 \leqslant 0.$$

Finally, considering all four cases $R(u_2, Q_{u_2}) < 1$ for all $u_2 \in \mathcal{U}_2$. This completes the proof. ∎

Appendix H
Useful Lemmas

Lemma 10. *Let X and Y be independent random variables with marginal distributions P_X and P_Y, respectively. Suppose X and Y take values from a group $\mathbb{Z}_m$. Then*

1) $A_{\epsilon/2}^{(n)}(X+Y) \subseteq A_\epsilon^{(n)}(X) + A_\epsilon^{(n)}(Y)$,
2) *there exists a function $\delta(\cdot)$ with $\lim_{\epsilon\to 0}\delta(\epsilon) = 0$ such that*

$$\frac{\left|A_{\delta(\epsilon)}^{(n)}(X,Y)\right|}{\left|A_\epsilon^{(n)}(X)\right|\left|A_\epsilon^{(n)}(Y)\right|} \geqslant 1 - 2^{-n\frac{\epsilon}{m}}.$$

Proof: For the first statement take an arbitrary element $\mathbf{z} \in A_{\epsilon/2}^{(n)}(X+Y)$. We show that such an element can be written as $\mathbf{z} = \mathbf{x} + \mathbf{y}$ for some element $\mathbf{x} \in A_\epsilon^{(n)}(X)$ and $\mathbf{y} \in A_\epsilon^{(n)}(Y)$. For that, select an arbitrary $\mathbf{y} \in A_{\epsilon/2}^{(n)}(Y|\mathbf{z})$. From standard arguments on typical sequences, $\mathbf{y}$ is $\epsilon/2$- typical with respect to P_Y. In addition, $(\mathbf{z}, \mathbf{y}) \in A_\epsilon^{(n)}(X+Y, Y)$. As a result,

$$(\mathbf{z} - \mathbf{y}, \mathbf{y}) \in A_\epsilon^{(n)}(X, Y).$$

Set $\mathbf{x} = \mathbf{z} - \mathbf{y}$. We showed that, $(\mathbf{x}, \mathbf{y}) \in A_\epsilon^{(n)}(X, Y)$, and $\mathbf{x} + \mathbf{y} = \mathbf{z}$. Since $\mathbf{x}$ and $\mathbf{y}$ are jointly ϵ-typical, then $\mathbf{x} \in A_\epsilon^{(n)}(X)$ and $\mathbf{y} \in A_\epsilon^{(n)}(Y)$. This completes the proof for the first statement.

For the second statement, given $\tilde{\epsilon} > 0$ we have

$$1 - \frac{\left|A_{\tilde{\epsilon}}^{(n)}(X,Y)\right|}{\left|A_\epsilon^{(n)}(X)\right|\left|A_\epsilon^{(n)}(Y)\right|} \leqslant \frac{\left|A_{\tilde{\epsilon}}^{(n)}(X,Y)^c\right|}{\left|A_\epsilon^{(n)}(X)\right|\left|A_\epsilon^{(n)}(Y)\right|} = \sum_{(\mathbf{x},\mathbf{y})\notin A_{\tilde{\epsilon}}^{(n)}(X,Y)} \frac{1}{\left|A_\epsilon^{(n)}(X)\right|\left|A_\epsilon^{(n)}(Y)\right|}$$

Let $P_{X,Y}^n = \prod_{i=1}^n P_X P_Y$. From standard arguments for ϵ-typical sequences the above expression does not exceed

$$\sum_{(\mathbf{x},\mathbf{y})\notin A_{\tilde{\epsilon}}^{(n)}(X,Y)} 2^{n\epsilon\frac{\alpha}{m}} P_{X,Y}^n(\mathbf{x},\mathbf{y}) = P_{X,Y}^n\{A_{\tilde{\epsilon}}^{(n)}(X,Y)^c\} 2^{n\epsilon\frac{\alpha}{m}} \leqslant 2^{n\epsilon\frac{\alpha}{m}} 2^{-\frac{\tilde{\epsilon}^2 n}{m^2 \ln 4}},$$

where

$$\alpha = -\frac{3}{m} \sum_{\substack{a,b\in\mathbb{Z}_m \\ P_{X,Y}(a,b)>0}} \log P_{X,Y}(a,b).$$

The last inequality holds as (X, Y) are independent. Define the function $\delta(\epsilon) =\triangleq [m\epsilon(1+\alpha)\ln 4]^{1/2}$ and set $\tilde{\epsilon} = \delta(\epsilon)$. As a result, the right-hand side of the above inequality is simplified to $2^{-n\frac{\epsilon}{m}}$. Thus, the second statement of the lemma is established. ∎

Lemma 11 ([39]). *Suppose that $\mathbf{G}$ is a $k \times n$ matrix with elements generated randomly and uniformly from $\mathbb{Z}_{p^r}$. If $\mathbf{u} \in H_s^k \backslash H_{s+1}^k$, then*

$$P\{\mathbf{u}\mathbf{G}_i = \mathbf{x}\} = p^{-n(r-s)}\mathbb{1}\{x \in H_s^n\}.$$

Lemma 12. *Given $(X, Y) \sim P_{XY}$, and sequences $\mathbf{x}, \mathbf{y}$ such that $([\mathbf{x}]_s, \mathbf{y}) \in A_\epsilon^{(n)}([X]_s, Y)$, let*

$$\mathcal{A} \triangleq \{\mathbf{x}' \mid (\mathbf{x}', \mathbf{y}) \in A_\epsilon^n(XY), \mathbf{x}' - \mathbf{x} \in H_s^n\}.$$

Then

$$A_{c_1\epsilon}^{(n)}(X|[\mathbf{x}]_s, \mathbf{y}) \subseteq \mathcal{A} \subseteq A_{c_2\epsilon}^{(n)}(X|[\mathbf{x}]_s, \mathbf{y}),$$

and we have,

$$|\mathcal{A}| \geqslant (1 - c_1\epsilon)2^{n(H(X|Y[X]_s) - c_1\delta(\epsilon))}$$
$$|\mathcal{A}| \leqslant 2^{n(H(X|Y[X]_s) + c_2\delta(\epsilon))},$$

where

$$\delta(\epsilon) = \frac{\epsilon}{|\mathcal{Y}|} \sum_{a\in\mathcal{X}} \sum_{b\in\mathcal{Y}:p(b|a)>0} \log_2 \ p(b|a),$$

and $c_1 = \frac{1}{|\mathcal{X}|+|\mathcal{Y}|}$, and $c_2 = p^{r-s}\frac{|\mathcal{X}|+1}{|\mathcal{Y}|}$.

Proof: Suppose $\mathbf{x}' \in \mathcal{A}$. Then $\mathbf{x}' - \mathbf{x} \in H_s^n$, which implies $[\mathbf{x}']_s = [\mathbf{x}]_s$. In addition, $(\mathbf{x}', \mathbf{y}) \in A_\epsilon^{(n)}(X, Y)$. Therefore,

$$(\mathbf{x}', [\mathbf{x}]_s, \mathbf{y}) \in A_{\epsilon'}^{(n)}(X, [X], Y),$$

where $\epsilon' = \epsilon p^{r-s}$. Thus,

$$\mathbf{x}' \in A_{\epsilon''}^{(n)}(X|[\mathbf{x}]_s, \mathbf{y}),$$

where $\epsilon'' = \frac{|\mathcal{X}|+1}{|\mathcal{Y}|}\epsilon'$. On the other hand, if $\mathbf{x}' \in A_{\tilde{\epsilon}}^{(n)}(X|[\mathbf{x}]_s\mathbf{y})$, then $[\mathbf{x}']_s = [\mathbf{x}]_s$, and $\mathbf{x}' \in A_\epsilon^{(n)}(X|\mathbf{y})$, where $\epsilon = \tilde{\epsilon}(|\mathcal{X}| + |\mathcal{Y}|)$. ■

Lemma 13. *Let X and Y be two independent random variables over $\mathbb{Z}_m$ with distributions $\mathbf{p} = (p_0, p_1, \ldots, p_{m-1})$ and $\mathbf{q} = (q_0, q_1, \ldots, q_{m-1})$, respectively. Then $H(X \oplus_m Y) = H(Y)$ if and only if there exists $i \in [1 : m]$ such that $\mathbf{p} \circledast_m \mathbf{q} = \pi^i(\mathbf{q})$, where $\circledast_m$ is the circular convolution and is defined as*

$$(\mathbf{p} \circledast_m \mathbf{q})(a) \triangleq \sum_{b\in\mathbb{Z}_m} p_b q_{a\ominus b}, \quad \forall a \in \mathbb{Z}_m,$$

$\pi((q_0, q_1, \ldots, q_{m-1})) = (q_{m-1}, q_0, q_1, \ldots, q_{m-2})$, and π^i is the composition of the function π with itself for i times.

Proof: First note that as X is independent of Y, we have

$$H(X \oplus_m Y) - H(Y) = I(X; X \oplus_m Y) \geqslant 0.$$

We want to find all distributions $\mathbf{p}$ and $\mathbf{q}$ for which the right-hand side equals zero. We first fix a distribution $\mathbf{q}$ and find all $\mathbf{p}$ such that the equality holds. This is equivalent to the solution of the following minimization problem:

$$\min_{\mathbf{p}\in\Delta_m} H(\mathbf{p} \circledast_m \mathbf{q}) - H(\mathbf{q}), \tag{53}$$

where

$$\Delta_m \triangleq \Big\{(q_0, q_1, \ldots, q_{m-1}) \in \mathbb{R}^m : \sum_{i=0}^{m-1} q_i = 1,\ q_i \geqslant 0,\ i \in [0 : m-1]\Big\}.$$

Note that Δ_m is a $m-1$-dimensional simplex in $\mathbb{R}^m$. Define the map

$$\varphi_\mathbf{q} : \Delta_m \mapsto \Delta_m,\ \varphi_\mathbf{q}(\mathbf{p}) = \mathbf{p} \circledast_m \mathbf{q}$$

for all $\mathbf{p}, \mathbf{q} \in \Delta_m$. Note that $\varphi_\mathbf{q}$ is a linear map. Let $\varphi_\mathbf{q}(\Delta_m)$ denote the image of Δ_m under $\varphi_\mathbf{q}$. Since $\varphi_\mathbf{q}$ is a linear map, $\varphi_\mathbf{q}(\Delta_m)$ is a simplex. Therefore, (53) is equivalent to

$$\min_{\mathbf{p}'\in\varphi_\mathbf{q}(\Delta_m)} H(\mathbf{p}') - H(\mathbf{q}).$$

It is well-known that the entropy function is strictly concave. Hence, the minimum points are the extreme points of the simplex $\varphi_\mathbf{q}(\Delta_m)$. Extreme points of $\varphi_\mathbf{q}(\Delta_m)$ are the image of the extreme points of Δ_m. Define the map $\pi : \Delta_m \mapsto \Delta_m$ as in the statement of the lemma. Extreme points of $\varphi_\mathbf{q}(\Delta_m)$ are characterized by $\pi^i(\mathbf{q}), i \in [1 : m]$, where π^i is the composition of π with itself for i times. Therefore, the minimum points of (53) are described as $\bigcup_{i=1}^m \varphi_\mathbf{q}^{-1}(\pi^i(\mathbf{q}))$, where $\varphi^{-1}(\mathbf{a})$ is the pre-image of $\mathbf{a}, \forall \mathbf{a} \in \Delta_m$.

Next, we range over all $\mathbf{q} \in \Delta_m$. Define the set

$$\mathcal{A}_i \triangleq \{(\mathbf{p}, \mathbf{q}) \in \Delta_m \times \Delta_m : \mathbf{p} \circledast_m \mathbf{q} = \pi^i(\mathbf{q})\}.$$

Then, the set of all $(\mathbf{p}, \mathbf{q})$ such that $H(\mathbf{p} \circledast_m \mathbf{q}) = H(\mathbf{q})$ is characterized by the set $\bigcup_{i=1}^m \mathcal{A}_i$. This is equivalent to the statement of the lemma. ■

Lemma 14. *Suppose S and $N_\mathbf{p}$ are independent random variables over $\mathbb{Z}_4$, where $\mathbf{p}$ is the distribution of $N_\mathbf{p}$. Let $f : \mathbb{Z}_4 \mapsto \mathbb{Z}_4$ be a function of S, and denote $X \triangleq f(S)$. Suppose for the cost functions (c_1, c_2) given in Example 4, the equality $\mathbb{E}\{c_1(X)\} = 0$ holds. Then the following bounds hold:*

$$H(S) - H(X + S) \leqslant 1$$
$$H(S) - H(X + S + N_\mathbf{p}) < 1,$$

where

$$\mathbf{p} \in \Big\{[1/3, 0, 2/3, 0], [1/3, 2/3, 0, 0], [1/4, 1/4, 1/2, 0]\Big\}.$$

Proof: For the first equality, we start with the following relations

$$\begin{aligned} H(X + S) &= H(X, S) - H(X|X + S) \\ &= H(S) - H(X|X + S). \end{aligned}$$

Therefore, we obtain

$$H(S) - H(X + S) = H(X|X + S) \leqslant H(X) \overset{(a)}{\leqslant} 1.$$

Note (a) is true, because X takes at most two values with positive probabilities.

For the second inequality we have

$$\begin{aligned} H(S) - H(X + S + N_\mathbf{p}) &= H(S) - H(X + S) + H(X + S) \\ &\quad - H(X + S + N_\mathbf{p}) \\ &\leqslant 1 - (H(X + S + N_\mathbf{p}) - H(X + S)) \\ &\leqslant 1. \end{aligned} \tag{54}$$

TABLE IV
THE CONDITIONS ON $x(\cdot)$ AND S

$X+S$	0	1	2	3
$(s, x(s))$	$(0,0),(2,2)$	$(1,0),(3,2)$	$(0,2),(2,0)$	$(1,2),(3,0)$

Let $\mathbf{q}$ be the distribution of $X+S$. We find the conditions on $\mathbf{p}$ and $\mathbf{q}$ for which

$$H(X+S+N_{\mathbf{p}}) - H(X+S) = 0.$$

Since $N_{\mathbf{p}}$ is independent of $X+S$, we can use Lemma 13 in which $Y = N_{\mathbf{p}}$ and $X = X+S$. Therefore,

$$H(X+S+N_{\mathbf{p}}) = H(X+S),$$

if and only if $\mathbf{p} \circledast_4 \mathbf{q} = \pi^i(\mathbf{q})$ for some $i \in [1:4]$. For fixed i and $\mathbf{p}$, the map defined by

$$\mathbf{q} \mapsto \mathbf{p} \circledast_4 \mathbf{q} - \pi^i(\mathbf{q})$$

is a linear map. In addition, the null space of this map characterizes the set of all $\mathbf{q}$ that satisfies the equality in Lemma 13. For $\mathbf{p} = [1/3, 0, 2/3, 0]$ this map can be represented by the matrix

$$A_{i,[1/3,0,2/3,0]} = \begin{bmatrix} -\frac{2}{3} & 0 & \frac{2}{3} & 0 \\ 0 & -\frac{2}{3} & 0 & \frac{2}{3} \\ \frac{2}{3} & 0 & -\frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & -\frac{2}{3} \end{bmatrix}$$

The null space of $\mathbf{A}_{i,[1/3,0,2/3,0]}$ is the subspace spanned by $[1/2, 0, 1/2, 0]$ and $[1/4, 1/4, 1/4, 1/4]$. Using the same approach, we can show that for any $i \in [1:4]$ and

$$\mathbf{p} \in \Big\{[1/3, 0, 2/3, 0], [1/3, 2/3, 0, 0], [1/4, 1/4, 1/2, 0]\Big\},$$

the null space of $\mathbf{A}_{i,\mathbf{p}}$ is contained in the subspace spanned by $[1/2, 0, 1/2, 0]$ and $[1/4, 1/4, 1/4, 1/4]$. This implies that $q_0 = q_2$ and $q_1 = q_3$.

Note $\mathbf{q}$ is the distribution of $x(S) + S$. Next, we find all functions $x(\cdot)$ and random variables S such that $q_0 = q_2$ and $q_1 = q_3$. For each $a \in \mathbb{Z}_4$, we characterize $(s, x(s))$ such that $x(s) + s = a$, where $x(s) \in \{0, 2\}$. We present such a characterization in Table IV. Using Table IV, if $q_0 > 0$, then

$$\mathbb{P}(S=0) = \mathbb{P}(S=2) = q_0$$

and $x(0) = x(2)$. Similarly, if $q_1 > 0$, then

$$\mathbb{P}(S=1) = \mathbb{P}(S=3) = q_1$$

and $x(1) = x(3)$. Therefore, if $q_0, q_1 > 0$, the distribution of S equals to $\mathbf{q} = [q_0, q_1, q_0, q_1]$. If $q_0 = 0$, then $q_1 = 1/2$. This implies

$$\mathbb{P}(S=1) = \mathbb{P}(S=3) = \frac{1}{2}.$$

Similarly, If $q_1 = 0$, then

$$\mathbb{P}(S=0) = \mathbb{P}(S=2) = q_1 = \frac{1}{2}.$$

As a result of this argument, $H(S) = H(X+S)$. Also by Lemma 13, the equality

$$H(X+S) = H(X+S+N_{\mathbf{p}})$$

holds. Therefore, in this case,

$$H(S) - H(X+S+N_{\mathbf{p}}) = 0.$$

To sum-up, we proved that if

$$\mathbf{p} \in \Big\{[1/3, 0, 2/3, 0], [1/3, 2/3, 0, 0], [1/4, 1/4, 1/2, 0]\Big\},$$

and

$$H(X+S) = H(X+S+N_{\mathbf{p}}),$$

then

$$H(S) - H(X+S+N_{\mathbf{p}}) = 0.$$

Therefore, using this argument and (54), we proved that if

$$\mathbf{p} \in \Big\{[1/3, 0, 2/3, 0], [1/3, 2/3, 0, 0], [1/4, 1/4, 1/2, 0]\Big\},$$

then

$$H(X+S) - H(X+S+N_{\mathbf{p}}) < 1.$$

■

APPENDIX I
PROOF OF CLAIM 1

Proof:

1): Let $a = 0, b = 2$, and $P(X_2 + S_2 = 0|u_1) = p_0$, and $P(X_2 + S_2 = 2|u_1) = 1 - p_0$. We represent this PMF by the vector $\mathbf{p} = [p_0, 0, 1-p_0, 0]$. This probability distribution is a linear combination of the form

$$\mathbf{p} = \beta\big[2/3, 0, 1/3, 0\big] + (1-\beta)\big[1/3, 0, 2/3, 0\big], \tag{55}$$

where $\beta = 3p_0 - 1$.

Remark 10. Let $Z = X + Y$, where the PMF of X is $\mathbf{p} = [p_0, p_1, p_2, p_3]$, and the PMF of Y is $\mathbf{q} = [q_0, q_1, q_2, q_3]$. If $\mathbf{t}$ is the PMF of Z, then $\mathbf{t} = \mathbf{p} \circledast_4 \mathbf{q}$, where $\circledast_4$ is the circular convolution in $\mathbb{Z}_4$. In addition, the map

$$(\mathbf{p}, \mathbf{q}) \longmapsto \mathbf{p} \circledast_4 \mathbf{q}$$

is bi-linear.

Let

$$t_i = \mathbb{P}\big(X_1 + S_1 + X_2 + S_2 = i|u_1 u_2\big),$$

and

$$q_i = \mathbb{P}\big(X_1 + S_1 = i|u_1\big)$$

for all $i \in \mathbb{Z}_4$. Also denote $\mathbf{q} = [q_0, q_1, q_2, q_3]$, and $\mathbf{t} = [t_0, t_1, t_2, t_3]$. Using Remark 10 and equation (55) we obtain

$$\mathbf{t} = \beta\big([2/3, 0, 1/3, 0] \circledast_4 \mathbf{q}\big) + (1-\beta)\big([1/3, 0, 2/3, 0] \circledast_4 \mathbf{q}\big).$$

This implies that, $\mathbf{t}$ is also a linear combination of two PMFs. From the concavity of entropy, we get the following lower-bound:

$$\begin{aligned}
&H(X_1 + S_1 + X_2 + S_2|u_1 u_2) = H(\mathbf{t}) \\
&= H\Big(\beta\big([2/3, 0, 1/3, 0] \circledast_4 \mathbf{q}\big) + (1-\beta)\big([1/3, 0, 2/3, 0] \circledast_4 \mathbf{q}\big)\Big) \\
&\geqslant \beta H\Big([2/3, 0, 1/3, 0] \circledast_4 \mathbf{q}\Big) + (1-\beta) H\Big([1/3, 0, 2/3, 0] \circledast_4 \mathbf{q}\Big) \\
&= \beta H\big(X_1 + S_1 + N_{[2/3,0,1/3,0]}|u_1\big) \\
&+ (1-\beta) H\big(X_1 + S_1 + N_{[1/3,0,2/3,0]}|u_1\big),
\end{aligned}$$

where in the last equality, $N_{[\lambda_0,\lambda_1,\lambda_2,\lambda_3]}$ denotes a random variable with PMF $[\lambda_0, \lambda_1, \lambda_2, \lambda_3]$ that is also independent of u_1 and $X_1 + S_1$. As a result of the above argument, equation (51) is bounded by

$$\begin{aligned}
&H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1u_2) - 2\\
&\leqslant H(S_1|u_1) + H(S_2|u_2) - \beta H(X_1 + S_1 + N_{[2/3,0,1/3,0]}|u_1)\\
&- (1-\beta)H(X_1 + S_1 + N_{[1/3,0,2/3,0]}|u_1) - 2\\
&= \beta\Big(H(S_1|u_1) - H(X_1 + S_1 + N_{[2/3,0,1/3,0]}|u_1)\Big)\\
&+ (1-\beta)\Big(H(S_1|u_1) - H(X_1 + S_1 + N_{[1/3,0,2/3,0]}|u_1)\Big)\\
&+ H(S_2|u_2) - 2.
\end{aligned}$$

2): Let $a = 0, b = 2$, and $P(X_2 + S_2 = 0|u_1) = p_0$, and $P(X_2+S_2 = 1|u_1) = 1-p_0$. In this case $\mathbf{p} = [p_0, 1-p_0, 0, 0]$. Also,

$$\mathbf{p} = \beta[2/3, 1/3, 0, 0] + (1-\beta)[1/3, 2/3, 0, 0],$$

where $\beta = 3p_0 - 1$. Similar to case 1), we use Remark 10 and the concavity of the entropy to get,

$$\begin{aligned}
&H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1u_2) - 2\\
&\leqslant \beta\Big(H(S_1|u_1) - H(X_1 + S_1 + N_{[2/3,1/3,0,0]}|u_1)\Big)\\
&+ (1-\beta)\Big(H(S_1|u_1) - H(X_1 + S_1 + N_{[1/3,2/3,0,0]}|u_1)\Big)\\
&+ H(S_2|u_2) - 2
\end{aligned}$$

■

ACKNOWLEDGMENT

We wish to thank the reviewers and the associate editor for their comments and suggestions which contributed significantly to this manuscript.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

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

[3] 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, Mar. 2011.

[4] 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. IT-29, no. 3, pp. 396–412, May 1983.

[5] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 3, pp. 277–288, May 1980.

[6] T. S. Han and K. Kobayashi, "A dichotomy of functions $f(X, Y)$ of correlated sources (X, Y)," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 1, pp. 69–76, Jan. 1987.

[7] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[8] S. H. Lim, C. Feng, A. Pastore, B. Nazer, and M. Gastpar, "A joint typicality approach to algebraic network information theory," 2016, *arXiv:1606.09548*. [Online]. Available: https://arxiv.org/abs/1606.09548

[9] M. Heidari, F. Shirani, and S. S. Pradhan, "Beyond group capacity in multi-terminal communications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 2081–2085.

[10] M. Heidari and S. S. Pradhan, "How to compute modulo prime-power sums," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1824–1828.

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

[12] J. Zhan, S. Y. Park, M. Gastpar, and A. Sahai, "Linear function computation in networks: Duality and constant gap results," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 620–638, Apr. 2013.

[13] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Linear codes, target function classes, and network computing capacity," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5741–5753, Sep. 2013.

[14] T. Philosof, R. Zamir, U. Erez, and A. J. Khisti, "Lattice strategies for the dirty multiple access channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5006–5035, Aug. 2011.

[15] T. Philosof and R. Zamir, "On the loss of single-letter characterization: The dirty multiple access channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2442–2454, Jun. 2009.

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

[17] M. Heidari, F. Shirani, and S. S. Pradhan, "A new achievable rate region for multiple-access channel with states," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 36–40.

[18] M. Heidari, F. Shirani, and S. S. Pradhan, "New sufficient conditions for multiple-access channel with correlated sources," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2019–2023.

[19] F. Shirani, M. Heidari, and S. S. Pradhan, "New lattice codes for multiple-descriptions," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 1580–1584.

[20] S. Sridharan, A. Jafarian, S. Vishwanath, S. Jafar, and S. Shamai (Shitz), "A layered lattice coding scheme for a class of three user Gaussian interference channels," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2008, pp. 531–538.

[21] S.-N. Hong and G. Caire, "On interference networks over finite fields," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4902–4921, Aug. 2014.

[22] G. Bresler, A. Parekh, and D. N. C. Tse, "The approximate capacity of the many-to-one and one-to-many Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4566–4592, Sep. 2010.

[23] U. Niesen and M. A. Maddah-Ali, "Interference alignment: From degrees of freedom to constant-gap capacity approximations," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4855–4888, Aug. 2013.

[24] A. Jafarian and S. Vishwanath, "Achievable rates for K-user Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4367–4380, Jul. 2012.

[25] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian K-user interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 2072–2076.

[26] F. Shirani and S. S. Pradhan, "Trade-off between communication and cooperation in the interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2214–2218.

[27] A. Padakandla and S. S. Pradhan, "Achievable rate region for three user discrete broadcast channel based on coset codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 1277–1281.

[28] M. Heidari, F. Shirani, and S. S. Pradhan, "On the necessity of structured codes for communications over MAC with feedback," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2298–2302.

[29] D. Slepian, "Group codes for the Gaussian channel," *Bell Syst. Tech. J.*, vol. 47, no. 4, pp. 575–602, Apr. 1968.

[30] H.-A. Loeliger, "Signal sets matched to groups," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1675–1682, Nov. 1991.

[31] E. Şaşoğlu, E. Telatar, and E. Arikan, "Polarization for arbitrary discrete memoryless channels," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2009, pp. 144–148.

[32] A. G. Sahebi and S. S. Pradhan, "Multilevel channel polarization for arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7839–7857, Dec. 2013.

[33] E. Abbe and E. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5437–5448, Aug. 2012.

[34] W. C. Park and A. Barg, "Polar codes for q-ary channels, $q = 2^r$," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 955–969, Feb. 2013.

[35] R. Ahlswede, "Group codes do not achieve Shannon's channel capacity for general discrete channels," *Ann. Math. Statist.*, vol. 42, no. 1, pp. 224–240, Feb. 1971.

[36] R. Ahlswede and J. Gemma, “Bounds on algebraic code capacities for noisy channels. I,” *Inf. Control*, vol. 19, no. 2, pp. 124–145, 1971.
[37] R. Ahlswede and J. Gemma, “Bounds on algebraic code capacities for noisy channels. II,” *Inf. Control*, vol. 19, no. 2, pp. 146–158, 1971.
[38] G. Como and F. Fagnani, “The capacity of finite Abelian group codes over symmetric memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2037–2054, May 2009.
[39] A. G. Sahebi and S. S. Pradhan, “Abelian group codes for channel coding and source coding,” *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2399–2414, May 2015.
[40] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
[41] S. A. Jafar and S. Vishwanath, “Generalized degrees of freedom of the symmetric Gaussian K user interference channel,” *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3297–3303, Jul. 2010.
[42] A. Padakandla, A. G. Sahebi, and S. S. Pradhan, “A new achievable rate region for the 3-user discrete memoryless interference channel,” in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 2256–2260.
[43] I. Csiszar and J. Korner, *Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[44] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[45] S. Jafar, “Capacity with causal and noncausal side information: A unified view,” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5474, Dec. 2006.

Mohsen Heidari received his BSc degree and MSc in Electrical Engineering from the Sharif University of Technology, Iran, in 2011 and 2013, respectively. He received a MSc degree in Applied Mathematics in 2017 and a PhD degree in Electrical Engineering in 2019, both from the University of Michigan. He is currently a postdoctoral research fellow in the department of Electrical Engineering, the University of Michigan. Mohsen's research interests lie in information theory, communication theory and quantum information theory.

Farhad Shirani obtained his B.Sc. degree from Sharif University of Technology in 2011 and Ph.D. from the University of Michigan at Ann Arbor in 2016. Currently, he is a Research Assistant Professor at the New York University at New York. His research interests include multiterminal communication systems, coding theory, privacy and security.

S. Sandeep Pradhan (M'99–SM'14) obtained his M.E. degree from the Indian Institute of Science in 1996 and Ph.D. from the University of California at Berkeley in 2001. From 2002 to 2008 he was an assistant professor, and from 2008 to 2015 he was an associate professor in the Department of Electrical Engineering and Computer Science at the University of Michigan at Ann Arbor, where he is currently a professor. He is the recipient of 2001 Eliahu Jury award given by the University of California at Berkeley for outstanding research in the areas of systems, signal processing, communications and control, the CAREER award given by the National Science Foundation (NSF), and the Outstanding Achievement award for the year 2009 from the University of Michigan. He was an associate editor for IEEE TRANSACTIONS ON INFORMATION THEORY in the area of Shannon theory from 2014-2016. His research interests include network information theory, coding theory, and quantum information theory.