On Linear Coding over Finite Rings and Applications to Computing

Sheng Huang, Mikael Skoglund, *Senior Member, IEEE*

Abstract

This paper presents a coding theorem for linear coding over finite rings, in the setting of the Slepian–Wolf source coding problem. This theorem covers corresponding achievability theorems of Elias [1] and Csiszár [2] for linear coding over finite fields as special cases. In addition, it is shown that, for any set of finite correlated discrete memoryless sources, there always exists a sequence of linear encoders over some finite non-field rings which achieves the data compression limit, the Slepian–Wolf region. Hence, the optimality problem regarding linear coding over finite non-field rings for data compression is closed with positive confirmation with respect to existence.

For application, we addressed the problem of source coding for computing, where the decoder is interested in recovering a discrete function of the data generated and independently encoded by several correlated i.i.d. random sources. We propose linear coding over finite rings as an alternative solution to this problem. Results in Körner–Marton [3] and Ahlswede–Han [4, Theorem 10] are generalized to cases for encoding (pseudo) nomographic functions (over rings). Since a discrete function with a finite domain always admits a nomographic presentation, we conclude that both generalizations universally apply for encoding all discrete functions of finite domains. Based on these, we demonstrate that linear coding over finite rings strictly outperforms its field counterpart in terms of achieving better coding rates and reducing the required alphabet sizes of the encoders for encoding infinitely many discrete functions.

Index Terms

Linear Coding, Source Coding, Ring, Field, Source Coding for Computing.

I. Introduction

The problem of *source coding for computing* can be defined as follows.

Problem 1 (Source Coding for Computing)**.** Given $\mathcal{S} = \{1, 2, \cdots, s\}$ and $(X_1, X_2, \cdots, X_s) \sim p$. Let t_i $(i \in \mathcal{S})$ be a *discrete memoryless source* that randomly generates i.i.d. discrete data $X_i^{(1)}, X_i^{(2)}, \cdots, X_i^{(n)}, \cdots$, where $X_i^{(n)}$ has a finite sample space $\mathscr{X}_i$ and $\left[X_1^{(n)}, X_2^{(n)}, \cdots, X_s^{(n)}\right] \sim p, \forall\, n \in \mathbb{N}^+$. For a *discrete function* $g : \prod_{i\in\mathcal{S}} \mathscr{X}_i \to \Omega$, what is the largest region $\mathcal{R}[g] \subset \mathbb{R}^s$, such that, $\forall\, (R_1, R_2, \cdots, R_s) \in \mathcal{R}[g]$ and $\forall\, \epsilon > 0$, there exists an $N_0 \in \mathbb{N}^+$, such

S. Huang and M. Skoglund are with the Communication Theory Lab, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, 10044, Sweden e-mail: (sheng.huang@ee.kth.se; skoglund@ee.kth.se).

This work was funded in part by the Swedish Research Council.

that for all $n > N_0$, there exist s *encoders* $\phi_i : \mathscr{X}_i^n \to \left[1, 2^{nR_i}\right], i \in \mathcal{S}$, and one *decoder* $\psi : \prod_{i\in\mathcal{S}} \left[1, 2^{nR_i}\right] \to \Omega^n$, with

$$\Pr\left\{\vec{g}\left(X_1^n, \cdots, X_s^n\right) \neq \psi\left[\phi_1\left(X_1^n\right), \cdots, \phi_s\left(X_s^n\right)\right]\right\} < \epsilon,$$

where $X_i^n = \left[X_i^{(1)}, X_i^{(2)}, \cdots, X_i^{(n)}\right]$ and

$$\vec{g}\left(X_1^n, \cdots, X_s^n\right) = \begin{bmatrix} g\left(X_1^{(1)}, \cdots, X_s^{(1)}\right) \\ \vdots \\ g\left(X_1^{(n)}, \cdots, X_s^{(n)}\right) \end{bmatrix} \in \Omega^n ?$$

The region $\mathcal{R}[g]$ is called the *achievable coding rate region* for computing g. A rate tuple $\mathbf{R} \in \mathbb{R}^s$ is said to be *achievable* for computing g (or simply achievable) if and only if $\mathbf{R} \in \mathcal{R}[g]$. A region $\mathcal{R} \subset \mathbb{R}^s$ is said to be *achievable* for computing g (or simply achievable) if and only if $\mathcal{R} \subseteq \mathcal{R}[g]$.

If g is an *identity function*, the computing problem, Problem 1, is known as the *Slepian–Wolf* (SW) *source coding* problem. $\mathcal{R}[g]$ is then the *SW region* [5], namely

$$\mathcal{R}[X_1, X_2, \cdots, X_s] = \left\{(R_1, R_2, \cdots, R_s) \in \mathbb{R}^s \;\middle|\; \sum_{j\in T} R_j > H(X_T|X_{T^c}), \forall\, \emptyset \neq T \subseteq \mathcal{S}\right\},$$

where T^c is the *complement* of T in $\mathcal{S}$ and $X_T\left(X_{T^c}\right)$ is the random variable array $\prod_{j\in T} X_j \left(\prod_{j\in T^c} X_j\right)$. However, from [5] it is hard to draw conclusions regarding the structure of the optimal encoders, as the corresponding mappings are chosen randomly among all feasible mappings. This limits the scope of their potential applications. As a completion, *linear coding over finite fields* (LCoF), namely $\mathscr{X}_i$'s are injectively mapped into some subsets of some finite *fields* and the ϕ_i's are chosen as *linear mappings* over these fields, is considered. It is shown that LCoF achieves the same encoding limit, the SW region [1], [2]. Although it seems straightforward to study linear mappings over *rings* (non-field rings in particular), it has not been proved (nor denied) that linear encoding over non-field rings can be equally optimal.

For an arbitrary discrete function g, Problem 1 remains open in general, and $\mathcal{R}[X_1, X_2, \cdots, X_s] \subseteq \mathcal{R}[g]$ obviously. Making use of Elias' theorem on *binary linear codes* [1], Körner–Marton [3] shows that $\mathcal{R}[\oplus_2]$ ("$\oplus_2$" is the *modulo-two sum*) contains the region

$$\mathcal{R}_{\oplus_2} = \left\{(R_1, R_2) \in \mathbb{R}^2 \mid R_1, R_2 > H(X_1 \oplus_2 X_2)\right\}.$$

This region is not contained in the SW region for certain distributions. In other words, $\mathcal{R}[\oplus_2] \supsetneq \mathcal{R}[X_1, X_2]$. Combining the standard random coding technique and Elias' result, [4] shows that $\mathcal{R}[\oplus_2]$ can be strictly larger than the convex hull of the union $\mathcal{R}[X_1, X_2] \cup \mathcal{R}_{\oplus_2}$. However, the functions considered in these works are relatively simple. With a *polynomial approach*, [6], [7] generalize the result of Ahlswede–Han [4, Theorem 10] to the scenario of g being arbitrary. Making use of the fact that a discrete function is essentially a *polynomial function* (see Definition II.2) over some finite field, an achievable region is given for computing an arbitrary discrete function. Such a region contains and can be strictly larger (depending on the precise function and distribution under consideration) than the

SW region. Conditions under which $\mathcal{R}[g]$ is strictly larger than the SW region are presented in [8] and [6] from different perspectives, respectively.

The present work proposes replacing the linear encoders over finite fields from Elias [1] and Csiszár [2] with linear encoders over finite rings in the case of the problems accounted for above. Achievability theorems related to *linear coding over finite rings* (LCoR) for SW data compression are presented, covering the results in [1], [2] as special cases in the sense of characterizing the achievable region. In addition, it is proved that there always exists a sequence of linear encoders over some finite non-field rings that achieves the SW region for any scenario of SW. Therefore, the issue of optimality of linear coding over finite non-field rings for data compression is closed with respect to existence. Furthermore, we also consider LCoR as an alternative technique for the general computing problem, Problem 1. Results from Körner–Marton [3], Ahlswede–Han [4, Theorem 10] and [7] are generalized to corresponding ring versions for encoding (*pseudo*) *nomographic functions* (over rings). Since any discrete function with a finite domain admits a *nomographic presentation*, we conclude that our results universally apply for encoding all discrete functions of finite domains. Finally, it is shown that our ring approach dominates its field counterpart in terms of achieving better coding rates and reducing alphabet sizes of the encoders for encoding some discrete function. The proof is done by taking advantage of the fact that the *characteristic* of a ring can be any positive integer while the characteristic of a field must be a prime. From this observation used in the proof, it is seen that there are actually infinite many such functions.

II. Rings, Ideals and Linear Mappings

We start by introducing some fundamental algebraic concepts and related properties. Readers who are already familiar with this material may still choose to go through quickly to identify our notation.

Definition II.1. The touple $[\mathfrak{R}, +, \cdot]$ is called a *ring* if the following criteria are met:

1) $[\mathfrak{R}, +]$ is an *Abelian group*;
2) There exists a *multiplicative identity*[1] $1 \in \mathfrak{R}$, namely, $1 \cdot a = a \cdot 1 = a$, $\forall\, a \in \mathfrak{R}$;
3) $\forall\, a, b, c \in \mathfrak{R}$, $a \cdot b \in \mathfrak{R}$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$;
4) $\forall\, a, b, c \in \mathfrak{R}$, $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and $(b + c) \cdot a = (b \cdot a) + (c \cdot a)$.

We often write $\mathfrak{R}$ for $[\mathfrak{R}, +, \cdot]$ when the *operations* considered are known from the context. The operation "$\cdot$" is usually written by juxtaposition, ab for $a \cdot b$, for all $a, b \in \mathfrak{R}$.

A ring $[\mathfrak{R}, +, \cdot]$ is said to be *commutative* if $\forall\, a, b \in \mathfrak{R}$, $a \cdot b = b \cdot a$. In Definition II.1, the *identity* of the group $[\mathfrak{R}, +]$, denoted by 0, is called the *zero*. A ring $[\mathfrak{R}, +, \cdot]$ is said to be *finite* if the cardinality $|\mathfrak{R}|$ is finite, and $|\mathfrak{R}|$ is called the *order* of $\mathfrak{R}$. The set $\mathbb{Z}_q$ of integers modulo q is a commutative finite ring with respect to the *modular arithmetic*. For any ring $\mathfrak{R}$, the set of all *polynomials* of s *indeterminants* over $\mathfrak{R}$ is an infinite ring.

[1] Sometimes a ring without a multiplicative identity is considered. Such a structure has been called a *rng*. We consider rings with multiplicative identities in this paper. However, similar results remain valid when considering rngs instead. Although we will occasionally comment on such results, they are not fully considered in the present work.

Definition II.2. A *polynomial function*[2] of k *variables* over a finite ring $\mathfrak{R}$ is a function $g : \mathfrak{R}^k \to \mathfrak{R}$ of the form

$$g(x_1, x_2, \cdots, x_k) = \sum_{j=0}^{m} a_j x_1^{m_{1j}} x_2^{m_{2j}} \cdots x_k^{m_{kj}}, \tag{1}$$

where $a_j \in \mathfrak{R}$ and m and m_{ij}'s are non-negative integers. The set of all the polynomial functions of k variables over ring $\mathfrak{R}$ is designated by $\mathfrak{R}[k]$.

Remark 1. Polynomial and polynomial function are sometimes only defined over a commutative ring [9], [10]. It is a very delicate matter to define them over a non-commutative ring [11], [12], due to the fact that $x_1 x_2$ and $x_2 x_1$ can become different objects. We choose to define "polynomial functions" with formula (1) because those functions are within the scope of this paper's interest.

Proposition II.3. *Given s rings $\mathfrak{R}_1, \mathfrak{R}_2, \cdots, \mathfrak{R}_s$, for any non-empty set $T \subseteq \{1, 2, \cdots, s\}$, the* Cartesian product *(see [9]) $\mathfrak{R}_T = \prod_{i \in T} \mathfrak{R}_i$ forms a new ring $[\mathfrak{R}_T, +, \cdot]$ with respect to the component-wise operations defined as follows:*

$$\mathbf{a}' + \mathbf{a}'' = \left[a'_1 + a''_1, a'_2 + a''_2, \cdots, a'_{|T|} + a''_{|T|}\right],$$
$$\mathbf{a}' \cdot \mathbf{a}'' = \left[a'_1 a''_1, a'_2 a''_2, \cdots, a'_{|T|} a''_{|T|}\right],$$

$\forall\, \mathbf{a}' = \left[a'_1, a'_2, \cdots, a'_{|T|}\right], \mathbf{a}'' = \left[a''_1, a''_2, \cdots, a''_{|T|}\right] \in \mathfrak{R}_T$.

Remark 2. In Proposition II.3, $[\mathfrak{R}_T, +, \cdot]$ is called the *direct product* of $\{\mathfrak{R}_i | i \in T\}$. It can be easily seen that $[0, 0, \cdots, 0]$ and $[1, 1, \cdots, 1]$ are the zero and the multiplicative identity of $[\mathfrak{R}_T, +, \cdot]$, respectively.

Definition II.4. A non-zero element a of a ring $\mathfrak{R}$ is said to be *invertible*, if and only if there exists $b \in \mathfrak{R}$, such that $ab = ba = 1$. b is called the *inverse* of a, denoted by a^{-1}. An invertible element of a ring is called a *unit*.

Remark 3. It can be proved that the inverse of a unit is unique. By definition, the multiplicative identity is the inverse of itself.

Let $\mathfrak{R}^* = \mathfrak{R} \setminus \{0\}$. The ring $[\mathfrak{R}, +, \cdot]$ is a *field* if and only if $[\mathfrak{R}^*, \cdot]$ is an Abelian group. In other words, all non-zero elements of $\mathfrak{R}$ are invertible. All fields are commutative rings. $\mathbb{Z}_q$ is a field if and only if q is a *prime*. All finite fields of the same order are *isomorphic* to each other [13, pp. 549]. This "unique" field of order q is denoted by $\mathbb{F}_q$. It is necessary that q is a power of a prime. More details regarding finite fields can be found in [13, Ch. 14.3].

Theorem II.5 (Wedderburn's little theorem c.f. Theorem 7.13 of [9])**.** *Let $\mathfrak{R}$ be a finite ring. $\mathfrak{R}$ is a field if and only if all non-zero elements of $\mathfrak{R}$ are invertible.*

[2]Polynomial and polynomial function are distinct concepts.

Remark 4. Wedderburn's little theorem guarantees commutativity for a finite ring if all of its non-zero elements are invertible. Hence, a finite ring is either a field or at least one of its elements has no inverse. However, a finite commutative ring is not necessary a field, e.g. $\mathbb{Z}_q$ is not a field if q is not a prime.

Definition II.6 (c.f. [13])**.** The *characteristic* of a finite ring $\mathfrak{R}$ is defined to be the smallest positive integer m, such that $\sum_{j=1}^{m} 1 = 0$, where 0 and 1 are the zero and the multiplicative identity of $\mathfrak{R}$, respectively. The characteristic of $\mathfrak{R}$ is often denoted by Char($\mathfrak{R}$).

Remark 5. Clearly, Char($\mathbb{Z}_q$) $= q$. For a finite field $\mathbb{F}_q$, Char($\mathbb{F}_q$) is always the prime q_0 such that $q = q_0^n$ for some integer n [9, Proposition 2.137].

Proposition II.7. *Let $\mathbb{F}_q$ be a finite field. For any $0 \neq a \in \mathbb{F}_q$, $m =$ Char($\mathbb{F}_q$) if and only if m is the smallest positive integer such that $\sum_{j=1}^{m} a = 0$.*

Proof: Since $a \neq 0$,

$$\sum_{j=1}^{m} a = 0 \Rightarrow a^{-1} \sum_{j=1}^{m} a = a^{-1} \cdot 0 \Rightarrow \sum_{j=1}^{m} 1 = 0 \Rightarrow \sum_{j=1}^{m} a = 0$$

The statement is proved. ■

Definition II.8. A subset $\mathfrak{I}$ of a ring $[\mathfrak{R}, +, \cdot]$ is said to be a *left ideal* of $\mathfrak{R}$, denoted by $\mathfrak{I} \leq_l \mathfrak{R}$, if and only if

1) $[\mathfrak{I}, +]$ is a subgroup of $[\mathfrak{R}, +]$;
2) $\forall\, x \in \mathfrak{I}$ and $\forall\, a \in \mathfrak{R}$, $a \cdot x \in \mathfrak{I}$.

If condition 2) is replaced by

3) $\forall\, x \in \mathfrak{I}$ and $\forall\, a \in \mathfrak{R}$, $x \cdot a \in \mathfrak{I}$,

then $\mathfrak{I}$ is called a *right ideal* of $\mathfrak{R}$, denoted by $\mathfrak{I} \leq_r \mathfrak{R}$. $\{0\}$ is a *trivial* left (right) ideal, usually denoted by 0.

The cardinality $|\mathfrak{I}|$ is called the *order* of a finite left (right) ideal $\mathfrak{I}$.

Remark 6. Let $\{a_1, a_2, \cdots, a_n\}$ be a non-empty set of elements of some ring $\mathfrak{R}$. It is easy to verify that $\langle a_1, a_2, \cdots, a_n \rangle_r = \left\{ \sum_{i=1}^{n} a_i b_i \middle| b_i \in \mathfrak{R}, \forall\ 1 \leq i \leq n \right\}$ is a right ideal and $\langle a_1, a_2, \cdots, a_n \rangle_l = \left\{ \sum_{i=1}^{n} b_i a_i \middle| b_i \in \mathfrak{R}, \forall\ 1 \leq i \leq n \right\}$ is a left ideal. Furthermore, $\langle a_1, a_2, \cdots, a_n \rangle_r = \mathfrak{R}$ and $\langle a_1, a_2, \cdots, a_n \rangle_l = \mathfrak{R}$ if a_i is a unit for some $1 \leq i \leq n$.

It is well-known that if $\mathfrak{I} \leq_l \mathfrak{R}$, then $\mathfrak{R}$ is divided into disjoint *cosets* which are of equal size (cardinality). For any coset $\mathfrak{J}$, $\mathfrak{J} = x + \mathfrak{I} = \{x + y | y \in \mathfrak{I}\}$, $\forall\, x \in \mathfrak{J}$. The set of all cosets forms a *left module* over $\mathfrak{R}$, denoted by $\mathfrak{R}/\mathfrak{I}$. Similarly, $\mathfrak{R}/\mathfrak{I}$ becomes a *right module* over $\mathfrak{R}$ if $\mathfrak{I} \leq_r \mathfrak{R}$ [14]. Of course, $\mathfrak{R}/\mathfrak{I}$ can also be considered as a *quotient group* [9, Ch. 1.6 and Ch. 2.9]. However, its structure is well richer than simply being a quotient group.

Proposition II.9. *Let $\mathfrak{R}_i$ $(1 \leq i \leq s)$ be a ring and $\mathfrak{R} = \prod_{i=1}^{s} \mathfrak{R}_i$. For any $\mathfrak{A} \subseteq \mathfrak{R}$, $\mathfrak{A} \leq_l \mathfrak{R}$ (or $\mathfrak{A} \leq_r \mathfrak{R}$) if and only if $\mathfrak{A} = \prod_{i=1}^{s} \mathfrak{A}_i$ and $\mathfrak{A}_i \leq_l \mathfrak{R}_i$ (or $\mathfrak{A}_i \leq_r \mathfrak{R}_i$), $\forall\, 1 \leq i \leq s$.*

Proof: It suffices to complete the proof for $\leq_l$ only. Let π_i $(1 \leq i \leq s)$ be the coordinate function assigning every element in $\mathfrak{R}$ its ith component. Then $\mathfrak{A} \subseteq \prod_{i=1}^{s} \mathfrak{A}_i$, where $\mathfrak{A}_i = \pi_i(\mathfrak{A})$. Moreover, for any

$$\mathbf{x} = (\pi_1(\mathbf{x}_1), \pi_2(\mathbf{x}_2), \cdots, \pi_s(\mathbf{x}_s)) \in \prod_{i=1}^{s} \mathfrak{A}_i,$$

where $\mathbf{x}_i \in \mathfrak{A}$ for all feasible i, we have that

$$\mathbf{x} = \sum_{i=1}^{s} \mathbf{e}_i \mathbf{x}_i,$$

where $\mathbf{e}_i \in \mathfrak{R}$ has the ith coordinate being 1 and others being 0. If $\mathfrak{A} \leq_l \mathfrak{R}$, then $\mathbf{x} \in \mathfrak{A}$ by definition. Therefore, $\prod_{i=1}^{s} \mathfrak{A}_i \subseteq \mathfrak{A}$. Consequently, $\mathfrak{A} = \prod_{i=1}^{s} \mathfrak{A}_i$. Since π_i is a homomorphism, we also have that $\mathfrak{A}_i \leq_l \mathfrak{R}_i$ for all feasible i. The other direction is easily verified by definition. ■

Remark 7. It is worthwhile to point out that Proposition II.9 does not hold for infinite index set, namely, $\mathfrak{R} = \prod_{i \in I} \mathfrak{R}_i$, where I is not finite.

For any $\emptyset \neq T \subseteq \mathcal{S}$, Proposition II.9 states that any left (right) ideal of $\mathfrak{R}_T$ is a Cartesian product of some left (right) ideals of $\mathfrak{R}_i$, $i \in T$. Let $\mathfrak{I}_i$ be a left (right) ideal of ring $\mathfrak{R}_i$ $(1 \leq i \leq s)$. We define $\mathfrak{I}_T$ to be the left (right) ideal $\prod_{i \in T} \mathfrak{I}_i$ of $\mathfrak{R}_T$.

Definition II.10. A mapping $f : \mathfrak{R}^n \to \mathfrak{R}^m$ given as:

$$f(x_1, x_2, \cdots, x_n) = \left(\sum_{j=1}^{n} a_{1,j} x_j, \cdots, \sum_{j=1}^{n} a_{m,j} x_j \right)^t, \forall\, (x_1, x_2, \cdots, x_n) \in \mathfrak{R}^n, \tag{2}$$

where t stands for transposition and $a_{i,j} \in \mathfrak{R}$ for all feasible i and j, is called a *left linear mapping* over ring $\mathfrak{R}$. Similarly,

$$f(x_1, x_2, \cdots, x_n) = \left(\sum_{j=1}^{n} x_j a_{1,j}, \cdots, \sum_{j=1}^{n} x_j a_{m,j} \right)^t, \forall\, (x_1, x_2, \cdots, x_n) \in \mathfrak{R}^n,$$

defines a *right linear mapping* over ring $\mathfrak{R}$. If $m = 1$, then f is called a *left* (*right*) *linear function* over $\mathfrak{R}$.

From now on, left linear mapping (function) or right linear mapping (function) are simply called *linear mapping* (*function*). This will not lead to any confusion since the intended use can usually be clearly distinguished from the context.

Remark 8. The mapping f in Definition II.10 is called *linear* in accordance with the definition of *linear mapping* (*function*) over field. In fact, the two structures have several similar properties. Moreover, (2) is equivalent to

$$f(x_1, x_2, \cdots, x_n) = \mathbf{A} (x_1, x_2, \cdots, x_n)^t, \forall\, (x_1, x_2, \cdots, x_n) \in \mathfrak{R}^n, \tag{3}$$

where $\mathbf{A}$ is an $m \times n$ *matrix* over $\mathfrak{R}$ and $[\mathbf{A}]_{i,j} = a_{i,j}$ for all feasible i and j. $\mathbf{A}$ is named the *coefficient matrix*. It is easy to prove that a linear mapping is uniquely determined by its coefficient matrix, and vice versa. The linear mapping f is said to be *trivial*, denoted by 0, if $\mathbf{A}$ is the *zero matrix*, i.e. $[\mathbf{A}]_{i,j} = 0$ for all feasible i and j.

Let $\mathbf{A}$ be an $m \times n$ matrix over ring $\mathfrak{R}$ and $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$, $\forall\, \mathbf{x} \in \mathfrak{R}^n$. For the *system of linear equations*

$$f(\mathbf{x}) = \mathbf{A}\mathbf{x} = \mathbf{0}, \text{ where } \mathbf{0} = (0, 0, \cdots, 0)^t \in \mathfrak{R}^m,$$

let $\mathfrak{S}(f)$ be the set of all *solutions*, namely $\mathfrak{S}(f) = \{\mathbf{x} \in \mathfrak{R}^n | f(\mathbf{x}) = \mathbf{0}\}$. It is obvious that $\mathfrak{S}(f) = \mathfrak{R}^n$ if f is trivial, i.e. $\mathbf{A}$ is the zero matrix. If $\mathfrak{R}$ is a field, then $\mathfrak{S}(f)$ is a *subspace* of $\mathfrak{R}^n$. We conclude this section with a lemma regarding the cardinalities of $\mathfrak{R}^n$ and $\mathfrak{S}(f)$ in the following.

Lemma II.11. *For a finite ring $\mathfrak{R}$ and a linear function*

$$f : \mathbf{x} \mapsto (a_1, a_2, \cdots, a_n)\mathbf{x}$$

$$\left(f : \mathbf{x} \mapsto \mathbf{x}^t (a_1, a_2, \cdots, a_n)^t\right), \forall\, \mathbf{x} \in \mathfrak{R}^n,$$

we have

$$\frac{|\mathfrak{S}(f)|}{|\mathfrak{R}|^n} = \frac{1}{|\mathfrak{I}|},$$

where $\mathfrak{I} = \langle a_1, a_2, \cdots, a_n \rangle_r$ ($\mathfrak{I} = \langle a_1, a_2, \cdots, a_n \rangle_l$). In particular, if a_i is invertible for some $1 \leq i \leq n$, then $|\mathfrak{S}(f)| = |\mathfrak{R}|^{n-1}$.

Proof: It is obvious that the image $f(\mathfrak{R}^n) = \mathfrak{I}$ by definition. Moreover, $\forall\, x \neq y \in \mathfrak{I}$, the pre-images $f^{-1}(x)$ $f^{-1}(y)$ satisfy $f^{-1}(x) \cap f^{-1}(y) = \emptyset$ and $\left|f^{-1}(x)\right| = \left|f^{-1}(y)\right| = |\mathfrak{S}(f)|$. Therefore, $|\mathfrak{I}|\,|\mathfrak{S}(f)| = |\mathfrak{R}|^n$, i.e. $\frac{|\mathfrak{S}(f)|}{|\mathfrak{R}|^n} = \frac{1}{|\mathfrak{I}|}$. Moreover, if a_i is a unit, then $\mathfrak{I} = \mathfrak{R}$, thus, $|\mathfrak{S}(f)| = |\mathfrak{R}|^n / |\mathfrak{R}| = |\mathfrak{R}|^{n-1}$. ■

III. Linear Coding over Finite Rings

In this section, we will present a coding rate region achieved with LCoR for the SW source coding problem, i.e. g is an identity function in Problem 1. This region is exactly the SW region if all the rings considered are fields. However, being field is not necessary as seen in Section V, where the issue of optimality is addressed.

Before proceeding, a subtlety needs to be cleared out. It is assumed that a source, say t_i, generates data taking values from a finite sample space $\mathscr{X}_i$, while $\mathscr{X}_i$ does not necessarily admit any algebraic structure. We have to either assume that $\mathscr{X}_i$ is with a certain algebraic structure, for instance $\mathscr{X}_i$ is a ring, or injectively map elements of $\mathscr{X}_i$ into some algebraic structure. In our subsequent discussions, we assume that $\mathscr{X}_i$ is mapped into a finite ring $\mathfrak{R}_i$ of order at least $|\mathscr{X}_i|$ by some injection Φ_i. Hence, $\mathscr{X}_i$ can simply be treated as a subset $\Phi_i(\mathscr{X}_i) \subseteq \mathfrak{R}_i$ for a fixed Φ_i. When required, Φ_i can also be selected to obtain desired outcomes.

To facilitate our discussion, the following notation is used. For $\emptyset \neq T \subseteq \mathcal{S}$, X_T (x_T and $\mathscr{X}_T$ resp.) is defined to be the Cartesian product

$$\prod_{i \in T} X_i \left(\prod_{i \in T} x_i \text{ and } \prod_{i \in T} \mathscr{X}_i \text{ resp.} \right),$$

where $x_i \in \mathscr{X}_i$ is a realization of X_i. If $(X_1, X_2, \cdots, X_s) \sim p$, we denote the *marginal* of p with respect to X_T by p_{X_T}, i.e. $X_T \sim p_{X_T}$, define

$$H(p_{X_T}) = H(X_T) \text{ and}$$
$$\text{supp}(p_{X_T}) = \{ x_T \in \mathscr{X}_T | \, p_{X_T}(x_T) > 0 \}.$$

For simplicity, $\mathscr{M}(\mathscr{X}_{\mathcal{S}}, \mathfrak{R}_{\mathcal{S}})$ is defined to be

$$\{ [\Phi_1, \Phi_2, \cdots, \Phi_s] | \, \Phi_i : \mathscr{X}_i \to \mathfrak{R}_i \text{ is injective}, \forall \, i \in \mathcal{S} \}$$

($|\mathfrak{R}_i| \geq |\mathscr{X}_i|$ is implicitly assumed), and $\Phi(x_T) = \prod_{i \in T} \Phi_i(x_i)$ for any $\Phi \in \mathscr{M}(\mathscr{X}_{\mathcal{S}}, \mathfrak{R}_{\mathcal{S}})$ and $x_T \in \mathscr{X}_T$. For any $\Phi \in \mathscr{M}(\mathscr{X}_{\mathcal{S}}, \mathfrak{R}_{\mathcal{S}})$, let

$$\mathcal{R}_\Phi = \left\{ [R_1, R_2, \cdots, R_s] \in \mathbb{R}^s \left| \sum_{i \in T} \frac{R_i \log |\mathfrak{I}_i|}{\log |\mathfrak{R}_i|} > r(T, \mathfrak{I}_T), \right. \right.$$
$$\left. \left. \forall \, \emptyset \neq T \subseteq \mathcal{S}, \forall \, 0 \neq \mathfrak{I}_i \leq_l \mathfrak{R}_i \right\}, \right. \tag{4}$$

where $r(T, \mathfrak{I}_T) = H(X_T | X_{T^c}) - H(Y_{\mathfrak{R}_T/\mathfrak{I}_T} | X_{T^c}) = H(X_T | Y_{\mathfrak{R}_T/\mathfrak{I}_T}, X_{T^c})$ and $Y_{\mathfrak{R}_T/\mathfrak{I}_T} = \Phi(X_T) + \mathfrak{I}_T$ is a random variable with sample space $\mathfrak{R}_T/\mathfrak{I}_T$.

Theorem III.1. *$\mathcal{R}_\Phi$ is achievable with linear coding over the finite rings $\mathfrak{R}_1, \mathfrak{R}_2, \cdots, \mathfrak{R}_s$. In exact terms, $\forall \, \epsilon > 0$, there exists $N_0 \in \mathbb{N}^+$, for all $n > N_0$, there exist linear encoders (left linear mappings to be more precise) $\phi_i : \Phi(\mathscr{X}_i)^n \to \mathfrak{R}_i^{k_i}$ ($i \in \mathcal{S}$) and a decoder ψ, such that*

$$\Pr \left\{ \psi \left(\prod_{i \in S} \phi_i (\mathbf{X}_i) \right) \neq \prod_{i \in S} \mathbf{X}_i \right\} < \epsilon,$$

where $\mathbf{X}_i = \left[\Phi\left(X_i^{(1)}\right), \Phi\left(X_i^{(2)}\right), \cdots, \Phi\left(X_i^{(n)}\right) \right]^t$, as long as

$$\left[\frac{k_1 \log |\mathfrak{R}_1|}{n}, \frac{k_2 \log |\mathfrak{R}_2|}{n}, \cdots, \frac{k_s \log |\mathfrak{R}_s|}{n} \right] \in \mathcal{R}_\Phi.$$

Proof: The proof is given in Section IV. ■

The following is a concrete example helping to interpret this theorem.

Example III.2. Consider the single source scenario, where $X_1 \sim p$ and $\mathscr{X}_1 = \mathbb{Z}_6$, specified as follows.

X_1	0	1	2	3	4	5
$p(X_1)$	0.05	0.1	0.15	0.2	0.2	0.3

By Theorem III.1,

$$\mathcal{R} = \{ R_1 \in \mathbb{R} | R_1 > \max\{2.40869, 2.34486, 2.24686\} \}$$
$$= \{ R_1 \in \mathbb{R} | R_1 > 2.40869 = H(X_1) \}$$

is achievable with linear coding over ring $\mathbb{Z}_6$. Obviously, $\mathcal{R}$ is just the SW region $\mathcal{R}[X_1]$. Optimality is claimed.

Besides, we would like to point out that some of the inequalities defining (4) are not active for specific scenarios. Two classes of these scenarios are discussed in the following theorems.

Theorem III.3. *Suppose $\mathfrak{R}_i$ $(1 \leq i \leq s)$ is a (finite) product ring $\prod_{l=1}^{k_i} \mathfrak{R}_{l,i}$ of finite rings $\mathfrak{R}_{l,i}$'s, and the sample space $\mathscr{X}_i$ satisfies $|\mathscr{X}_i| \leq |\mathfrak{R}_{l,i}|$ for all feasible i and l. Given injections $\Phi_{l,i} : \mathscr{X}_i \to \mathfrak{R}_{l,i}$ and let*

$$\Phi = [\Phi_1, \Phi_2, \cdots, \Phi_s],$$

where $\Phi_i = \prod_{l=1}^{k_i} \Phi_{l,i}$ is defined as

$$\Phi_i : x_i \mapsto (\Phi_{1,i}(x_i), \Phi_{2,i}(x_i), \cdots, \Phi_{k_i,i}(x_i)) \in \mathfrak{R}_i, \forall\, x_i \in \mathscr{X}_i.$$

We have that

$$\mathcal{R}_{\Phi,\text{prod}} = \left\{ [R_1, R_2, \cdots, R_s] \in \mathbb{R}^s \middle| \sum_{i \in T} \frac{R_i \log |\mathfrak{I}_i|}{\log |\mathfrak{R}_i|} > H(X_T | Y_{\mathfrak{R}_T/\mathfrak{I}_T}, X_{T^c}), \right.$$
$$\left. \forall\, \emptyset \neq T \subseteq \mathcal{S}, \forall\, \mathfrak{I}_i = \prod_{l=1}^{k_i} \mathfrak{I}_{l,i} \text{ with } 0 \neq \mathfrak{I}_{l,i} \leq_l \mathfrak{R}_{l,i} \right\}, \tag{5}$$

where $Y_{\mathfrak{R}_T/\mathfrak{I}_T} = \Phi(X_T) + \mathfrak{I}_T$, is achievable with linear coding over $\mathfrak{R}_1, \mathfrak{R}_2, \cdots, \mathfrak{R}_s$. Moreover, $\mathcal{R}_\Phi \subseteq \mathcal{R}_{\Phi,\text{prod}}$.

Proof: The proof is found in Section IV. ∎

Let $\mathfrak{R}$ be a finite ring and

$$\mathbb{M}_{L,\mathfrak{R},m} = \left\{ \begin{bmatrix} a_1 & 0 & & 0 \\ a_2 & a_1 & & 0 \\ & & \ddots & \\ a_m & a_{m-1} & & a_1 \end{bmatrix} \middle| a_1, a_2, \cdots, a_m \in \mathfrak{R} \right\},$$

where m is a positive integer. It is easy to verify that $\mathbb{M}_{L,\mathfrak{R},m}$ is a ring with respect to matrix operations. Moreover, $\mathfrak{I}$ is a left ideal of $\mathbb{M}_{L,\mathfrak{R},m}$ if and only if

$$\mathfrak{I} = \left\{ \begin{bmatrix} a_1 & 0 & & 0 \\ a_2 & a_1 & & 0 \\ & & \ddots & \\ a_m & a_{m-1} & & a_1 \end{bmatrix} \middle| \begin{array}{r} a_j \in \mathfrak{I}_j \leq_l \mathfrak{R}, \forall\, 1 \leq j \leq m; \\ \mathfrak{I}_j \subseteq \mathfrak{I}_{j+1}, \forall\, 1 \leq j < m \end{array} \right\}.$$

Let $\mathfrak{O}(\mathbb{M}_{L,\mathfrak{R},m})$ be the set of all left ideals of the form

$$\left\{ \begin{bmatrix} a_1 & 0 & & 0 \\ a_2 & a_1 & & 0 \\ & & \ddots & \\ a_m & a_{m-1} & & a_1 \end{bmatrix} \middle| \begin{array}{r} a_j \in \mathfrak{I}_j \leq_l \mathfrak{R}, \forall\, 1 \leq j \leq m; \\ \mathfrak{I}_j \subseteq \mathfrak{I}_{j+1}, \forall\, 1 \leq j < m; \\ \mathfrak{I}_i = 0 \text{ for some } 1 \leq i \leq m \end{array} \right\}.$$

Theorem III.4. *Let $\mathfrak{R}_i$ $(1 \leq i \leq s)$ be a finite ring such that $|\mathscr{X}_i| \leq |\mathfrak{R}_i|$. For any injections $\Phi_i' : \mathscr{X}_i \to \mathfrak{R}_i$, let*

$$\Phi = [\Phi_1, \Phi_2, \cdots, \Phi_s],$$

where $\Phi_i : \mathscr{X}_i \to \mathbb{M}_{L,\mathfrak{R}_i,m_i}$ is defined as

$$\Phi_i : x_i \mapsto \begin{bmatrix} \Phi_i'(x_i) & 0 & & 0 \\ \Phi_i'(x_i) & \Phi_i'(x_i) & & 0 \\ & & \ddots & \\ \Phi_i'(x_i) & \Phi_i'(x_i) & & \Phi_i'(x_i) \end{bmatrix}, \forall\, x_i \in \mathscr{X}_i.$$

We have that

$$\mathcal{R}_{\Phi,\mathrm{m}} = \Bigg\{ [R_1, R_2, \cdots, R_s] \in \mathbb{R}^s \Bigg| \sum_{i \in T} \frac{R_i \log |\mathfrak{I}_i|}{\log |\mathfrak{R}_i|} > H(X_T | Y_{\mathfrak{R}_T/\mathfrak{I}_T}, X_{T^c}), \\ \forall\, \emptyset \neq T \subseteq \mathcal{S}, \forall\, \mathfrak{I}_i \leq_l \mathbb{M}_{L,\mathfrak{R}_i,m_i} \text{ and } \mathfrak{I}_i \notin \mathfrak{O}(\mathbb{M}_{L,\mathfrak{R}_i,m_i}) \Bigg\}, \tag{6}$$

where $Y_{\mathfrak{R}_T/\mathfrak{I}_T} = \Phi(X_T) + \mathfrak{I}_T$, is achievable with linear coding over $\mathbb{M}_{L,\mathfrak{R}_1,m_1}, \mathbb{M}_{L,\mathfrak{R}_2,m_2}, \cdots, \mathbb{M}_{L,\mathfrak{R}_s,m_s}$. Moreover, $\mathcal{R}_\Phi \subseteq \mathcal{R}_{\Phi,\mathrm{m}}$.

Proof: The proof is found in Section IV. ■

Remark 9. The difference between (4), (5) and (6) lies in their restrictions defining $\mathfrak{I}_i$'s, respectively, as highlighted in the proofs given in Section IV.

Remark 10. Without much effort, one can see that $\mathcal{R}_\Phi$ ($\mathcal{R}_{\Phi,\mathrm{prod}}$ and $\mathcal{R}_{\Phi,\mathrm{m}}$, resp.) in Theorem III.1 (Theorem III.3 and Theorem III.4, resp.) depends on Φ via random variables $Y_{\mathfrak{R}_T/\mathfrak{I}_T}$'s whose distributions are determined by Φ. For each $i \in \mathcal{S}$, there exist $\dfrac{|\mathfrak{R}_i|!}{(|\mathfrak{R}_i| - |\mathscr{X}_i|)!}$ distinct injections from $\mathscr{X}_i$ to a ring $\mathfrak{R}_i$ of order at least $|\mathscr{X}_i|$. Let $\mathrm{cov}(A)$ be the convex hull of a set $A \subseteq \mathbb{R}^s$. By a straightforward time sharing argument, we have that

$$\mathcal{R}_l = \mathrm{cov}\left(\bigcup_{\Phi \in \mathscr{M}(\mathscr{X}_\mathcal{S}, \mathfrak{R}_\mathcal{S})} \mathcal{R}_\Phi \right) \tag{7}$$

is achievable with linear coding over $\mathfrak{R}_1, \mathfrak{R}_2, \cdots, \mathfrak{R}_s$.

Remark 11. From Theorem V.1, one will see that (4) and (7) are the same when all the rings are fields. Actually, both are identical to the SW region. However, (7) can be strictly larger than (4) (see Section V), when not all the rings are fields. This implies that, in order to achieve the desired rate, a suitable injection is required. However, be reminded that taking the convex hull in (7) is not always needed for optimality as shown in Example III.2. A more sophisticated elaboration on this issue is found in Section V.

The rest of this section provides key supporting lemmata and concepts used to prove Theorem III.1, Theorem III.3 and Theorem III.4. The final proofs are presented in Section IV.

Lemma III.5. *Let $\mathbf{x}, \mathbf{y} \in \mathfrak{R}^n$ be two distinct sequences, where $\mathfrak{R}$ is a finite ring, and assume that $\mathbf{y} - \mathbf{x} = (a_1, a_2, \cdots, a_n)^t$. If $f : \mathfrak{R}^n \to \mathfrak{R}^k$ is a random linear mapping chosen uniformly at random, i.e. generate the $k \times n$ coefficient matrix $\mathbf{A}$ of f by independently choosing each entry of $\mathbf{A}$ from $\mathfrak{R}$ uniformly at random, then*

$$\Pr\{f(\mathbf{x}) = f(\mathbf{y})\} = |\mathfrak{I}|^{-k}, \tag{8}$$

where $\mathfrak{I} = \langle a_1, a_2, \cdots, a_n \rangle_l$.

Proof: Let $f = (f_1, f_2, \cdots, f_k)^t$, where $f_i : \mathfrak{R}^n \to \mathfrak{R}$ is a random linear function. Then

$$\begin{aligned}\Pr\{f(\mathbf{x}) = f(\mathbf{y})\} &= \Pr\left\{\bigcap_{i=1}^{k} \{f_i(\mathbf{x}) = f_i(\mathbf{y})\}\right\} \\ &= \prod_{i=1}^{k} \Pr\{f_i(\mathbf{x} - \mathbf{y}) = 0\},\end{aligned}$$

since the f_i's are independent from each other. The statement follows from Lemma II.11 which assure that $\Pr\{f_i(\mathbf{x} - \mathbf{y}) = 0\} = |\mathfrak{I}|^{-1}$. ■

Remark 12. In Lemma III.5, if $\mathfrak{R}$ is a field and $\mathbf{x} \neq \mathbf{y}$, then $\mathfrak{I} = \mathfrak{R}$ because every non-zero a_i is a unit. Thus, $\Pr\{f(\mathbf{x}) = f(\mathbf{y})\} = |\mathfrak{R}|^{-k}$.

Definition III.6 (c.f. [15])**.** Let $X \sim p_X$ be a discrete random variable with sample space $\mathscr{X}$. The set $\mathcal{T}_\epsilon(n, X)$ of *strongly ϵ-typical sequences* of length n with respect to X is defined to be

$$\left\{\mathbf{x} \in \mathscr{X}^n \left| \left| \frac{N(x; \mathbf{x})}{n} - p_X(x) \right| \leq \epsilon, \forall\, x \in \mathscr{X} \right.\right\},$$

where $N(x; \mathbf{x})$ is the number of occurrences of x in the sequence $\mathbf{x}$.

The notation $\mathcal{T}_\epsilon(n, X)$ is sometimes replaced by $\mathcal{T}_\epsilon$ when the length n and the random variable X referred to are clear from the context.

Now we conclude this section with the following lemma. It is a crucial part for our proofs of the achievability theorems. It generalizes the classic conditional typicality lemma [16, Theorem 15.2.2], yet at the same time distinguishes our argument from the one for the field version.

Lemma III.7. *Let $(X_1, X_2) \sim p$ be a jointly random variable whose sample space is a finite ring $\mathfrak{R} = \mathfrak{R}_1 \times \mathfrak{R}_2$. For any $\eta > 0$, there exists $\epsilon > 0$, such that, $\forall\, (\mathbf{x}_1, \mathbf{x}_2)^t \in \mathcal{T}_\epsilon(n, (X_1, X_2))$ and $\forall\, \mathfrak{I} \leq_l \mathfrak{R}_1$,*

$$|D_\epsilon(\mathbf{x}_1, \mathfrak{I} | \mathbf{x}_2)| < 2^{n\left[H(X_1 | Y_{\mathfrak{R}_1/\mathfrak{I}}, X_2) + \eta\right]}, \tag{9}$$

where

$$D_\epsilon(\mathbf{x}_1, \mathfrak{I} | \mathbf{x}_2) = \left\{ (\mathbf{y}, \mathbf{x}_2)^t \in \mathcal{T}_\epsilon \middle| \, \mathbf{y} - \mathbf{x}_1 \in \mathfrak{I}^n \right\}$$

and $Y_{\mathfrak{R}_1/\mathfrak{I}} = X_1 + \mathfrak{I}$ is a random variable with sample space $\mathfrak{R}_1/\mathfrak{I}$.

First Proof [3]*:* Let $\mathfrak{R}_1/\mathfrak{I} = \{a_1+\mathfrak{I}, a_2+\mathfrak{I}, \cdots, a_m+\mathfrak{I}\}$, where $m = |\mathfrak{R}_1|/|\mathfrak{I}|$. For arbitrary $\epsilon > 0$ and integer n, without loss of generality, assume that

$$\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} = \begin{bmatrix}\mathbf{x}_{1,1}, \mathbf{x}_{1,2}, \cdots, \mathbf{x}_{1,m}\\ \mathbf{x}_{2,1}, \mathbf{x}_{2,2}, \cdots, \mathbf{x}_{2,m}\end{bmatrix} = \begin{bmatrix}x_1^{(1)}, & x_1^{(2)}, & \cdots, & x_1^{(n)}\\ x_2^{(1)}, & x_2^{(2)}, & \cdots, & x_2^{(n)}\end{bmatrix}$$

and

$$\mathbf{z}_j = \begin{bmatrix}\mathbf{x}_{1,j}\\ \mathbf{x}_{2,j}\end{bmatrix} = \begin{bmatrix}x_1^{(\sum_{k=0}^{j-1} c_k+1)}, & x_1^{(\sum_{k=0}^{j-1} c_k+2)}, & \cdots, & x_1^{(\sum_{k=0}^{j} c_k)}\\ x_2^{(\sum_{k=0}^{j-1} c_k+1)}, & x_2^{(\sum_{k=0}^{j-1} c_k+2)}, & \cdots, & x_2^{(\sum_{k=0}^{j} c_k)}\end{bmatrix} \in (a_j + \mathfrak{I}\times\mathfrak{R}_2)^{c_j},$$

where $c_0 = 0$ and $c_j = \sum_{r\in a_j+\mathfrak{I}\times\mathfrak{R}_2} N\left(r, (\mathbf{x}_1,\mathbf{x}_2)^t\right)$, $1 \le j \le m$. For any $\mathbf{y} = \left[y^{(1)}, y^{(2)}, \cdots, y^{(n)}\right]$ with $(\mathbf{y},\mathbf{x}_2)^t \in D_\epsilon(\mathbf{x}_1, \mathfrak{I}|\mathbf{x}_2)$, we have $y^{(i)} - x_1^{(i)} \in \mathfrak{I}$, $\forall\ 1 \le i \le n$, by definition. Thus, $y^{(i)}$ and $x_1^{(i)}$ belong to the same coset, i.e. $y^{(\sum_{k=0}^{j-1} c_k+1)}, y^{(\sum_{k=0}^{j-1} c_k+2)}, \cdots, y^{(\sum_{k=0}^{j} c_k)} \in a_j + \mathfrak{I}, \forall\ 1 \le j \le m$. Furthermore, $\forall\ r \in \mathfrak{R}$,

$$\begin{aligned}&\left|N\left(r, (\mathbf{x}_1,\mathbf{x}_2)^t\right)/n - p(r)\right| \le \epsilon \text{ and}\\ &\left|N\left(r, (\mathbf{y},\mathbf{x}_2)^t\right)/n - p(r)\right| \le \epsilon\\ \Longrightarrow &\left|\frac{N\left(r, (\mathbf{y},\mathbf{x}_2)^t\right)}{n} - \frac{N\left(r, (\mathbf{x}_1,\mathbf{x}_2)^t\right)}{n}\right| \le 2\epsilon,\end{aligned}$$

since $(\mathbf{x}_1,\mathbf{x}_2)^t, (\mathbf{y},\mathbf{x}_2)^t \in \mathcal{T}_\epsilon$. As a consequence,

$$\mathbf{z}'_j = \begin{bmatrix}y^{(\sum_{k=0}^{j-1} c_k+1)}, & y^{(\sum_{k=0}^{j-1} c_k+2)}, & \cdots, & y^{(\sum_{k=0}^{j} c_k)}\\ x_2^{(\sum_{k=0}^{j-1} c_k+1)}, & x_2^{(\sum_{k=0}^{j-1} c_k+2)}, & \cdots, & x_2^{(\sum_{k=0}^{j} c_k)}\end{bmatrix} \in (a_j + \mathfrak{I}\times\mathfrak{R}_2)^{c_j}$$

is a strongly 2ϵ-typical sequence of length c_j with respect to the random variable $Z_j \sim p_j = \text{emp}(\mathbf{z}_j)$ (the *empirical distribution* of $\mathbf{z}_j$). The sample space of Z_j is $a_j + \mathfrak{I}\times\mathfrak{R}_2$. Therefore, the number of all possible $\mathbf{z}'_j$'s (namely, all elements $\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix} \in \mathcal{T}_{2\epsilon}(c_j, Z_j)$ such that $\mathbf{w}_2 = \mathbf{x}_{2,j}$) is upper bounded by $2^{c_j[H(p_j)-H(p_{j,2})+2\epsilon]}$, where $p_{j,2}$ is the marginal of p_j with respect to the second coordinate, by [15, Theorem 6.10]. Consequently,

$$|D_\epsilon(\mathbf{x}_1, \mathfrak{I}|\mathbf{x}_2)| \le 2^{\sum_{j=1}^m c_j[H(p_j)-H(p_{j,2})+2\epsilon]}. \tag{10}$$

Direct computation yields

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^m c_j H(p_j) &= \sum_{j=1}^m \frac{c_j}{n} \sum_{r\in a_j+\mathfrak{I}\times\mathfrak{R}_2} \frac{N\left(r, (\mathbf{x}_1,\mathbf{x}_2)^t\right)}{c_j} \log \frac{c_j}{N\left(r, (\mathbf{x}_1,\mathbf{x}_2)^t\right)}\\ &= \sum_{r\in\mathfrak{R}} \frac{N\left(r, (\mathbf{x}_1,\mathbf{x}_2)^t\right)}{n} \log \frac{n}{N\left(r, (\mathbf{x}_1,\mathbf{x}_2)^t\right)} - \sum_{j=1}^m \frac{c_j}{n}\log\frac{n}{c_j}\end{aligned}$$

[3] An alternative, second, proof was suggested by an anonymous reviewer for our paper [17], and is presented in Appendix C for completeness.

and

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{m} c_j H(p_{j,2}) \\
=&\sum_{j=1}^{m}\frac{c_j}{n}\left[\sum_{r_2\in\mathfrak{R}_2}\frac{\sum_{r_1\in a_j+\mathfrak{I}} N\left((r_1,r_2),(\mathbf{x}_1,\mathbf{x}_2)^t\right)}{c_j}\times\log\frac{c_j}{\sum_{r_1\in a_j+\mathfrak{I}} N\left((r_1,r_2),(\mathbf{x}_1,\mathbf{x}_2)^t\right)}\right] \\
=&\sum_{j=1}^{m}\sum_{r_2\in\mathfrak{R}_2}\frac{\sum_{r_1\in a_j+\mathfrak{I}} N((r_1,r_2),(\mathbf{x}_1,\mathbf{x}_2)^t)}{n}\times\log\frac{n}{\sum_{r_1\in a_j+\mathfrak{I}} N((r_1,r_2),(\mathbf{x}_1,\mathbf{x}_2)^t)}-\sum_{j=1}^{m}\frac{c_j}{n}\log\frac{n}{c_j}.
\end{aligned}$$

Since the entropy H is a continuous function, there exists some small $0<\epsilon<\eta/4$, such that

$$\left|\sum_{r\in\mathfrak{R}}\frac{N\left(r,(\mathbf{x}_1,\mathbf{x}_2)^t\right)}{n}\log\frac{n}{N\left(r,(\mathbf{x}_1,\mathbf{x}_2)^t\right)}-H(X_1,X_2)\right|<\eta/8,$$

$$\left|\sum_{j=1}^{m}\frac{c_j}{n}\log\frac{n}{c_j}-H(Y_{\mathfrak{R}_1/\mathfrak{I}})\right|<\eta/8 \text{ and}$$

$$\left|\sum_{j=1}^{m}\sum_{r_2\in\mathfrak{R}_2}\frac{\sum_{r_1\in a_j+\mathfrak{I}} N((r_1,r_2),(\mathbf{x}_1,\mathbf{x}_2)^t)}{n}\times\log\frac{n}{\sum_{r_1\in a_j+\mathfrak{I}} N((r_1,r_2),(\mathbf{x}_1,\mathbf{x}_2)^t)}-H(X_2,Y_{\mathfrak{R}_1/\mathfrak{I}})\right|<\eta/8.$$

Therefore,

$$\frac{1}{n}\sum_{j=1}^{m} c_j H(p_j) < H(X_1,X_2)-H(Y_{\mathfrak{R}_1/\mathfrak{I}})+\eta/4 \tag{11}$$

$$\frac{1}{n}\sum_{j=1}^{m} c_j H(p_{j,2}) > H(X_2,Y_{\mathfrak{R}_1/\mathfrak{I}})-H(Y_{\mathfrak{R}_1/\mathfrak{I}})-\eta/4 \tag{12}$$

where (11) and (12) are guaranteed for small $0<\epsilon<\eta/4$. Substituting (11) and (12) into (10), (9) follows. ■

Remark 13. Assume that $\mathbf{y}-\mathbf{x}=(a_1,a_2,\cdots,a_n)^t$, then $\mathbf{y}-\mathbf{x}\in\mathfrak{I}^n$ is equivalent to $\langle a_1,a_2,\cdots,a_n\rangle_l\subseteq\mathfrak{I}$.

IV. Proof of the Achievability Theorems

A. Proof of Theorem III.1

As mentioned, $\mathscr{X}_i$ can be seen as a subset of $\mathfrak{R}_i$ for a fixed $\Phi=[\Phi_1,\cdots,\Phi_s]$. In this section, we assume that X_i has sample space $\mathfrak{R}_i$, which makes sense since Φ_i is injective.

Let $\mathbf{R}=[R_1,R_2,\cdots,R_s]$ and $k_i=\left\lfloor\frac{nR_i}{\log|\mathfrak{R}_i|}\right\rfloor$, $\forall\ i\in\mathcal{S}$, where n is the length of the data sequences. If $\mathbf{R}\in\mathcal{R}_\Phi$, then $\sum_{i\in T}\frac{R_i\log|\mathfrak{I}_i|}{\log|\mathfrak{R}_i|}>r\left(T,\mathfrak{I}_T\right)$, (this implies that $\frac{1}{n}\sum_{i\in T}k_i\log|\mathfrak{I}_i|-r\left(T,\mathfrak{I}_T\right)>2\eta$ for some small constant $\eta>0$ and large enough n), $\forall\ \emptyset\neq T\subseteq\mathcal{S},\forall\ 0\neq\mathfrak{I}_i\leq_l\mathfrak{R}_i$. We claim that $\mathbf{R}$ is achievable by linear coding over $\mathfrak{R}_1,\mathfrak{R}_2,\cdots,\mathfrak{R}_s$.

Encoding:

For every $i\in\mathcal{S}$, randomly generate a $k_i\times n$ matrix $\mathbf{A}_i$ based on a uniform distribution, i.e. independently choose each entry of $\mathbf{A}_i$ uniformly at random from $\mathfrak{R}_i$. Define a linear encoder $\phi_i:\mathfrak{R}_i^n\to\mathfrak{R}_i^{k_i}$ such that

$$\phi_i:\mathbf{x}\mapsto\mathbf{A}_i\mathbf{x},\forall\ \mathbf{x}\in\mathfrak{R}_i^n.$$

Obviously the coding rate of this encoder is $\frac{1}{n}\log|\phi_i(\mathfrak{R}_i^n)| \le \frac{1}{n}\log|\mathfrak{R}_i|^{k_i} = \frac{\log|\mathfrak{R}_i|}{n}\left\lfloor\frac{nR_i}{\log|\mathfrak{R}_i|}\right\rfloor \le R_i$.

Decoding:

Subject to observing $\mathbf{y}_i \in \mathfrak{R}_i^{k_i}$ $(i \in \mathcal{S})$ from the ith encoder, the decoder claims that $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_s]^t \in \prod_{i=1}^{s} \mathfrak{R}_i^n$ is the array of the encoded data sequences, if and only if:

1) $\mathbf{x} \in \mathcal{T}_\epsilon$; and
2) $\forall\, \mathbf{x}' = [\mathbf{x}'_1, \mathbf{x}'_2, \cdots, \mathbf{x}'_s]^t \in \mathcal{T}_\epsilon$, if $\mathbf{x}' \neq \mathbf{x}$, then $\phi_j(\mathbf{x}'_j) \neq \mathbf{y}_j$, for some j.

Error:

Assume that $\mathbf{X}_i \in \mathfrak{R}_i^n$ $(i \in \mathcal{S})$ is the original data sequence generated by the ith source. It is readily seen that an error occurs if and only if one of the following events occurs:

E_1: $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_s]^t \notin \mathcal{T}_\epsilon$;

E_2: There exists $\mathbf{X} \neq [\mathbf{x}'_1, \mathbf{x}'_2, \cdots, \mathbf{x}'_s]^t \in \mathcal{T}_\epsilon$, such that $\phi_i(\mathbf{x}'_i) = \phi_i(\mathbf{X}_i)$, $\forall\, i \in \mathcal{S}$.

Error Probability:

By the joint asymptotic equipartition principle (AEP) [15, Theorem 6.9], $\Pr\{E_1\} \to 0$, $n \to \infty$.

Additionally, for $\emptyset \neq T \subseteq \mathcal{S}$, let

$$D_\epsilon(\mathbf{X}; T) = \left\{ [\mathbf{x}'_1, \mathbf{x}'_2, \cdots, \mathbf{x}'_s]^t \in \mathcal{T}_\epsilon \middle| \mathbf{x}'_i \neq \mathbf{X}_i \text{ if and only if } i \in T \right\}.$$

We have

$$D_\epsilon(\mathbf{X}; T) = \bigcup_{0 \neq \mathfrak{I} \le_l \mathfrak{R}_T} \left[D_\epsilon(\mathbf{X}_T, \mathfrak{I} | \mathbf{X}_{T^c}) \setminus \{\mathbf{X}\} \right], \tag{13}$$

where $\mathbf{X}_T = \prod_{i \in T} \mathbf{X}_i$ and $\mathbf{X}_{T^c} = \prod_{i \in T^c} \mathbf{X}_i$, since $\mathfrak{I}$ goes over all possible non-trivial left ideals. Consequently,

$$\begin{aligned}
\Pr\{E_2 | E_1^c\} &= \sum_{[\mathbf{x}'_1, \cdots, \mathbf{x}'_s]^t \in \mathcal{T}_\epsilon \setminus \{\mathbf{X}\}} \prod_{i \in \mathcal{S}} \Pr\{\phi_i(\mathbf{x}'_i) = \phi_i(\mathbf{X}_i) | E_1^c\} \\
&= \sum_{\emptyset \neq T \subseteq \mathcal{S}} \sum_{\substack{[\mathbf{x}'_1, \cdots, \mathbf{x}'_s]^t \\ \in D_\epsilon(\mathbf{X}; T)}} \prod_{i \in T} \Pr\{\phi_i(\mathbf{x}'_i) = \phi_i(\mathbf{X}_i) | E_1^c\} && (14) \\
&\le \sum_{\emptyset \neq T \subseteq \mathcal{S}} \sum_{0 \neq \mathfrak{I} \le_l \mathfrak{R}_T} \sum_{\substack{[\mathbf{x}'_1, \cdots, \mathbf{x}'_s]^t \\ \in D_\epsilon(\mathbf{X}_T, \mathfrak{I} | \mathbf{X}_{T^c}) \setminus \{\mathbf{X}\}}} \prod_{i \in T} \Pr\{\phi_i(\mathbf{x}'_i) = \phi_i(\mathbf{X}_i) | E_1^c\} && (15) \\
&< \sum_{\emptyset \neq T \subseteq \mathcal{S}} \sum_{\substack{0 \neq \prod_{i \in T} \mathfrak{I}_i \\ \le_l \mathfrak{R}_T}} \left(2^{n[r(T, \mathfrak{I}) + \eta]} - 1 \right) \prod_{i \in T} |\mathfrak{I}_i|^{-k_i} && (16) \\
&< (2^s - 1) \left(2^{|R_{\mathcal{S}}|} - 2 \right) \times \max_{\substack{\emptyset \neq T \in \mathcal{S}, \\ 0 \neq \prod_{i \in T} \mathfrak{I}_i \le_l \mathfrak{R}_T}} 2^{-n\left[\frac{1}{n}\sum_{i \in T} k_i \log|\mathfrak{I}_i| - [r(T, \mathfrak{I}) + \eta]\right]}, && (17)
\end{aligned}$$

where

(14) is from the fact that $\mathcal{T}_\epsilon \setminus \{\mathbf{X}\} = \coprod_{\emptyset \neq T \subseteq \mathcal{S}} D_\epsilon(\mathbf{X}; T)$ (disjoint union);

(15) follows from (13) by the union bound (Boole's inequality);

(16) is from Lemma III.5 and Lemma III.7, as well as the fact that every left ideal of $\mathfrak{R}_T$ is a Cartesian product of some left ideals $\mathfrak{I}_i$ of $\mathfrak{R}_i$, $i \in T$ (see Proposition II.9). At the same time, ϵ is required to be sufficiently small;

(17) is due to the facts that the number of non-empty subsets of $\mathcal{S}$ is $2^s - 1$ and the number of non-trivial left ideals of the finite ring $\mathfrak{R}_T$ is less than $2^{|R_S|} - 1$, which is the number of non-empty subsets of $\mathfrak{R}_S$ ($\supseteq \mathfrak{R}_T$).

Thus, $\Pr\{E_2|E_1^c\} \to 0$, when $n \to \infty$, from (17), since for sufficiently large n and small ϵ, $\frac{1}{n}\sum_{i\in T} k_i \log|\mathfrak{I}_i| - [r(T,\mathfrak{I}) + \eta] > \eta > 0$.

Therefore, $\Pr\{E_1 \cup E_2\} = \Pr\{E_1\} + \Pr\{E_2|E_1^c\} \to 0$ as $\epsilon \to 0$ and $n \to \infty$.

B. Proof of Theorem III.3

The proof follows almost the same steps as in proving Theorem III.1, except that the performance analysis only focuses on sequences $(a_{i,1}, a_{i,2}, \cdots, a_{i,n}) \in \mathfrak{R}_i^n$ $(1 \le i \le s)$ such that

$$a_{i,j} = \left(\Phi_{1,i}\left(x_i^{(j)}\right), \Phi_{2,i}\left(x_i^{(j)}\right), \cdots, \Phi_{k_i,i}\left(x_i^{(j)}\right)\right) \in \prod_{l=1}^{k_i} \mathfrak{R}_{l,i}$$

for some $x_i^{(j)} \in \mathscr{X}_i$. Let $\mathbf{X}_i, \mathbf{Y}_i$ be any two such sequences satisfying $\mathbf{X}_i - \mathbf{Y}_i \in \mathfrak{I}_i^n$ for some $\mathfrak{I}_i \le_l \mathfrak{R}_i$. Based on the special structure of $\mathbf{X}_i$ and $\mathbf{Y}_i$, it is easy to verify that $\mathfrak{I}_i \ne 0 \Leftrightarrow \mathfrak{I}_i = \prod_{l=1}^{k_i} \mathfrak{I}_{l,i}$ and $0 \ne \mathfrak{I}_{l,i} \le_l \mathfrak{R}_{l,i}$, for all $1 \le l \le k_i$. (This causes the difference between (4) and (5).) In addition, it is obvious that $\mathcal{R}_\Phi \subseteq \mathcal{R}_{\Phi,\text{prod}}$ by their definitions.

C. Proof of Theorem III.4

The proof is similar to that for Theorem III.1, except that it only focuses on sequences $(a_{i,1}, a_{i,2}, \cdots, a_{i,n}) \in \mathbb{M}_{L,\mathfrak{R}_i,m_i}^n$ $(1 \le i \le s)$ such that $a_{i,j} \in \mathbb{M}_{L,\mathfrak{R}_i,m_i}$ satisfies $[a_{i,j}]_{u,v} = \begin{cases} a, & u \le v; \\ 0, & \text{otherwise,} \end{cases}$ for some $a \in \mathfrak{R}_i$. Let $\mathbf{X}_i, \mathbf{Y}_i$ be any two such sequences such that $\mathbf{X}_i - \mathbf{Y}_i \in \mathfrak{I}_i^n$ for some $\mathfrak{I}_i \le_l \mathbb{M}_{L,\mathfrak{R}_i,m_i}$. It is easily seen that $\mathfrak{I}_i \ne 0$ if and only if $\mathfrak{I}_i \notin \mathfrak{O}(\mathbb{M}_{L,\mathfrak{R}_i,m_i})$. (This causes the difference between (4) and (6).) In addition, it is obvious that $\mathcal{R}_\Phi \subseteq \mathcal{R}_{\Phi,\text{m}}$ by their definitions.

V. Optimality

Obviously, Theorem III.1 specializes to its field counterpart if all rings considered are fields, as summarized in the following theorem.

Theorem V.1. *Region* (4) *is the SW region if $\mathfrak{R}_i$ contains no proper non-trivial left ideal, equivalently*[4]*, $\mathfrak{R}_i$ is a field, for all $i \in \mathcal{S}$. As a consequence, region* (7) *is the SW region.*

[4]Equivalency does not necessarily hold for rngs.

Proof: In Theorem III.1, random variable $Y_{\mathfrak{R}_T/\mathfrak{I}_T}$ admits a sample space of cardinality 1 for all $\emptyset \neq T \subseteq \mathcal{S}$, since the only non-trivial left ideal of $\mathfrak{R}_i$ is itself for all feasible i. Thus, $0 = H(Y_{\mathfrak{R}_T/\mathfrak{I}_T}) \geq H(Y_{\mathfrak{R}_T/\mathfrak{I}_T}|X_{T^c}) \geq 0$. Consequently,

$$\mathcal{R}_\Phi = \Big\{ [R_1, R_2, \cdots, R_s] \in \mathbb{R}^s \Big| \sum_{i\in T} R_i > H(X_T|X_{T^c}), \forall\, \emptyset \neq T \subseteq \mathcal{S} \Big\},$$

which is the SW region $\mathcal{R}[X_1, X_2, \cdots, X_s]$. Therefore, region (7) is also the SW region.

If $\mathfrak{R}_i$ is a field, then obviously it has no proper non-trivial left (right) ideal. Conversely, $\forall\ 0 \neq a \in \mathfrak{R}_i$, $\langle a \rangle_l = \mathfrak{R}_i$ implies that $\exists\ 0 \neq b \in \mathfrak{R}_i$, such that $ba = 1$. Similarly, $\exists\ 0 \neq c \in \mathfrak{R}_i$, such that $cb = 1$. Moreover, $c = c \cdot 1 = cba = 1 \cdot a = a$. Hence, $ab = cb = 1$. b is the inverse of a. By Wedderburn's little theorem, $\mathfrak{R}_i$ is a field. ■

One important question to address is whether linear coding over finite non-field rings can be equally optimal for data compression. Hereby, we claim that, for any SW scenario, there always exist linear encoders over some finite non-field rings which achieve the data compression limit. Therefore, optimality of linear coding over finite non-field rings for data compression is established in the sense of existence.

A. Existence Theorem I: Single Source

For any single source scenario, the assertion that there always exists a finite ring $\mathfrak{R}_1$, such that $\mathcal{R}_l$ is in fact the SW region

$$\mathcal{R}[X_1] = \{R_1 \in \mathbb{R} | R_1 > H(X_1)\},$$

is equivalent to the existence of a finite ring $\mathfrak{R}_1$ and an injection $\Phi_1 : \mathscr{X}_1 \to \mathfrak{R}_1$, such that

$$\max_{0 \neq \mathfrak{I}_1 \leq_l \mathfrak{R}_1} \frac{\log|\mathfrak{R}_1|}{\log|\mathfrak{I}_1|} \left[H(X_1) - H(Y_{\mathfrak{R}_1/\mathfrak{I}_1})\right] = H(X_1), \tag{18}$$

where $Y_{\mathfrak{R}_1/\mathfrak{I}_1} = \Phi_1(X_1) + \mathfrak{I}_1$.

Theorem V.2. *Let $\mathfrak{R}_1$ be a finite ring of order $|\mathfrak{R}_1| \geq |\mathscr{X}_1|$. If $\mathfrak{R}_1$ contains one and only one proper non-trivial left ideal $\mathfrak{I}_0$ and $|\mathfrak{I}_0| = \sqrt{|\mathfrak{R}_1|}$, then region* (7) *coincides with the SW region, i.e. there exists an injection $\Phi_1 : \mathscr{X}_1 \to \mathfrak{R}_1$, such that* (18) *holds.*

Remark 14. Examples of such a non-field ring $\mathfrak{R}_1$ in the above theorem include

$$\mathbb{M}_{L,p} = \left\{ \begin{bmatrix} x & 0 \\ y & x \end{bmatrix} \middle| x, y \in \mathbb{Z}_p \right\}$$

($\mathbb{M}_{L,p}$ is a ring with respect to matrix addition and multiplication) and $\mathbb{Z}_{p^2}$, where p is any prime. For any single source scenario, one can always choose $\mathfrak{R}_1$ to be either $\mathbb{M}_{L,p}$ or $\mathbb{Z}_{p^2}$. Consequently, optimality is attained.

Proof of Theorem V.2: Notice that the random variable $Y_{\mathfrak{R}_1/\mathfrak{I}_0}$ depends on the injection Φ_1, so does its entropy $H(Y_{\mathfrak{R}_1/\mathfrak{I}_0})$. Obviously $H(Y_{\mathfrak{R}_1/\mathfrak{R}_1}) = 0$, since the sample space of the random variable $Y_{\mathfrak{R}_1/\mathfrak{R}_1}$ contains only one

element. Therefore,

$$\frac{\log|\mathfrak{R}_1|}{\log|\mathfrak{R}_1|}\left[H(X_1)-H(Y_{\mathfrak{R}_1/\mathfrak{R}_1})\right]=H(X_1).$$

Consequently, (18) is equivalent to

$$\begin{aligned}&\frac{\log|\mathfrak{R}_1|}{\log|\mathfrak{I}_0|}\left[H(X_1)-H(Y_{\mathfrak{R}_1/\mathfrak{I}_0})\right]\le H(X_1)\\ \Leftrightarrow &H(X_1)\le 2H(Y_{\mathfrak{R}_1/\mathfrak{I}_0}),\end{aligned}\tag{19}$$

since $|\mathfrak{I}_0|=\sqrt{|\mathfrak{R}_1|}$. By Lemma A.1, there exists injection $\tilde{\Phi}_1:\mathscr{X}_1\to\mathfrak{R}_1$ such that (19) holds if $\Phi_1=\tilde{\Phi}_1$. The statement follows. ■

Up to isomorphism, there are exactly 4 distinct rings of order p^2 for a given prime p. They include 3 non-field rings, $\mathbb{Z}_p\times\mathbb{Z}_p$, $\mathbb{M}_{L,p}$ and $\mathbb{Z}_{p^2}$, in addition to the field $\mathbb{F}_{p^2}$. It has been proved that, using linear encoders over the last three, optimality can always be achieved in the single source scenario. Actually, the same holds true for all multiple sources scenarios.

B. Existence Theorem II: Multiple Sources

Theorem V.3. *Let $\mathfrak{R}_1,\mathfrak{R}_2,\cdots,\mathfrak{R}_s$ be s finite rings with $|\mathfrak{R}_i|\ge|\mathscr{X}_i|$. If $\mathfrak{R}_i$ is isomorphic to either*

1) *a field, i.e. $\mathfrak{R}_i$ contains no proper non-trivial left (right) ideal; or*
2) *a ring containing one and only one proper non-trivial left ideal $\mathfrak{I}_{0i}$ and $|\mathfrak{I}_{0i}|=\sqrt{|\mathfrak{R}_i|}$,*

for all feasible i, then (7) *coincides with the SW region $\mathcal{R}[X_1,X_2,\cdots,X_s]$.*

Remark 15. It is obvious that Theorem V.3 includes Theorem V.2 as a special case. In fact, its proof resembles the one of Theorem V.2. Examples of $\mathfrak{R}_i$'s include all finite fields, $\mathbb{M}_{L,p}$ and $\mathbb{Z}_{p^2}$, where p is a prime. However, Theorem V.3 does not guarantee that all rates, except the *vertexes*, in the *polytope* of the SW region are "directly" achievable for the multiple sources case. A time sharing scheme is required in our current proof. Nevertheless, all rates are "directly" achievable if $\mathfrak{R}_i$'s are fields or if $s=1$. This is partially the reason that the two theorems are stated separately.

Remark 16. Theorem V.3 also includes Theorem V.1 as a special case. However, Theorem V.1 admits a simpler proof compared to the one for Theorem V.3.

Proof of Theorem V.3: It suffices to prove that, for any $\mathbf{R}=[R_1,R_2,\cdots,R_s]\in\mathbb{R}^s$ satisfies

$$R_i>H(X_i|X_{i-1},X_{i-2},\cdots,X_1),\forall\ 1\le i\le s,$$

$\mathbf{R}\in\mathcal{R}_\Phi$ for some set of injections $\Phi=[\Phi_1,\Phi_2,\cdots,\Phi_s]$, where $\Phi_i:\mathscr{X}_i\to\mathfrak{R}_i$. Let $\tilde{\Phi}=[\tilde{\Phi}_1,\tilde{\Phi}_2,\cdots,\tilde{\Phi}_s]$ be the set of injections, where, if

(i) $\mathfrak{R}_i$ is a field, $\tilde{\Phi}_i$ is any injection;

(ii) $\mathfrak{R}_i$ satisfies 2), $\tilde{\Phi}_i$ is the injection such that

$$H(X_i|X_{i-1},X_{i-2},\cdots,X_1)\le 2H(Y_{\mathfrak{R}_i/\mathfrak{I}_{0i}}|X_{i-1},X_{i-2},\cdots,X_1),$$

when $\Phi_i = \tilde{\Phi}_i$. The existence of $\tilde{\Phi}_i$ is guaranteed by Lemma A.1.

If $\Phi = \tilde{\Phi}$, then

$$\begin{aligned}\frac{\log|\mathfrak{I}_i|}{\log|\mathfrak{R}_i|}H(X_i|X_{i-1},X_{i-2},\cdots,X_1) \geq & H(X_i|X_{i-1},X_{i-2},\cdots,X_1) - H(Y_{\mathfrak{R}_i/\mathfrak{I}_i}|X_{i-1},X_{i-2},\cdots,X_1)\\ = & H(X_i|Y_{\mathfrak{R}_i/\mathfrak{I}_i},X_{i-1},X_{i-2},\cdots,X_1),\end{aligned}$$

for all $1 \leq i \leq s$ and $0 \neq \mathfrak{I}_i \leq_l \mathfrak{R}_i$. As a consequence,

$$\begin{aligned}\sum_{i\in T}\frac{R_i\log|\mathfrak{I}_i|}{\log|\mathfrak{R}_i|} > & \sum_{i\in T}\frac{\log|\mathfrak{I}_i|}{\log|\mathfrak{R}_i|}H(X_i|X_{i-1},X_{i-2},\cdots,X_1)\\ \geq & \sum_{i\in T}\left[H(X_i|Y_{\mathfrak{R}_i/\mathfrak{I}_i},X_{i-1},X_{i-2},\cdots,X_1)\right]\\ \geq & \sum_{i\in T}\left[H(X_i|Y_{\mathfrak{R}_T/\mathfrak{I}_T},X_{T^c},X_{i-1},X_{i-2},\cdots,X_1)\right]\\ \geq & H\left(X_T\left|Y_{\mathfrak{R}_T/\mathfrak{I}_T},X_{T^c}\right.\right)\\ = & H\left(X_T|X_{T^c}\right) - H\left(Y_{\mathfrak{R}_T/\mathfrak{I}_T}|X_{T^c}\right),\end{aligned}$$

for all $\emptyset \neq T \subseteq \{1,2,\cdots,s\}$. Thus, $\mathbf{R} \in \mathcal{R}_{\tilde{\Phi}}$. ■

By Theorem V.1, Theorem V.2 and Theorem V.3, we draw the conclusion that

Corollary V.4. *For any SW scenario, there always exists a sequence of linear encoders over some finite rings (fields or non-field rings) which achieves the data compression limit, the SW region.*

In fact, LCoR can be optimal even for rings beyond those stated in the above theorems (see Example III.2). We classify some of these scenarios in the remaining parts of this section.

C. Product Rings

Theorem V.5. *Let $\mathfrak{R}_{l,1},\mathfrak{R}_{l,2},\cdots,\mathfrak{R}_{l,s}$ $(l=1,2)$ be a set of finite rings of equal size, and $\mathfrak{R}_i = \mathfrak{R}_{1,i}\times\mathfrak{R}_{2,i}$ for all feasible i. If the coding rate $\mathbf{R}\in\mathbb{R}^s$ is achievable with linear encoders over $\mathfrak{R}_{l,1},\mathfrak{R}_{l,2},\cdots,\mathfrak{R}_{l,s}$ $(l=1,2)$, then $\mathbf{R}$ is achievable with linear encoders over $\mathfrak{R}_1,\mathfrak{R}_2,\cdots,\mathfrak{R}_s$.*

Proof: By definition, $\mathbf{R}$ is a convex combination of coding rates which are achieved by different linear encoding schemes over $\mathfrak{R}_{l,1},\mathfrak{R}_{l,2},\cdots,\mathfrak{R}_{l,s}$ $(l=1,2)$, respectively. To be more precise, there exist $\mathbf{R}_1,\mathbf{R}_2,\cdots,\mathbf{R}_m\in\mathbb{R}^s$ and positive numbers $w_1,w_2,\cdots,w_m$ with $\sum_{j=1}^m w_j = 1$, such that $\mathbf{R} = \sum_{j=1}^m w_j\mathbf{R}_j$. Moreover, there exist injections $\Phi_l = [\Phi_{l,1},\Phi_{l,2},\cdots,\Phi_{l,s}]$ $(l=1,2)$, where $\Phi_{l,i}:\mathscr{X}_i\to\mathfrak{R}_{l,i}$, such that

$$\begin{aligned}\mathbf{R}_j\in\mathcal{R}_{\Phi_l} = \Bigg\{[R_1,R_2,\cdots,R_s]\in\mathbb{R}^s\Bigg| & \sum_{i\in T}\frac{R_i\log|\mathfrak{I}_{l,i}|}{\log|\mathfrak{R}_{l,i}|} > H(X_T|X_{T^c}) - H(Y_{\mathfrak{R}_{l,T}/\mathfrak{I}_{l,T}}|X_{T^c}),\\ & \forall\,\emptyset\neq T\subseteq\mathcal{S},\forall\,0\neq\mathfrak{I}_{l,i}\leq_l\mathfrak{R}_{l,i}\Bigg\},\end{aligned}\tag{20}$$

where $\mathfrak{R}_{l,T}=\prod_{i\in T}\mathfrak{R}_{l,i}$, $\mathfrak{I}_{l,T}=\prod_{i\in T}\mathfrak{I}_{l,i}$ and $Y_{\mathfrak{R}_{l,T}/\mathfrak{I}_{l,T}}=\Phi_l(X_T)+\mathfrak{I}_{l,T}$ is a random variable with sample space $\mathfrak{R}_{l,T}/\mathfrak{I}_{l,T}$. To show that $\mathbf{R}$ is achievable with linear encoders over $\mathfrak{R}_1,\mathfrak{R}_2,\cdots,\mathfrak{R}_s$, it suffices to prove that $\mathbf{R}_j$ is achievable with linear encoders over $\mathfrak{R}_1,\mathfrak{R}_2,\cdots,\mathfrak{R}_s$ for all feasible j. Let $\mathbf{R}_j=[R_{j,1},R_{j,2},\cdots,R_{j,s}]$. For all $\emptyset\neq T\subseteq\mathcal{S}$ and $0\neq\mathfrak{I}_i=\mathfrak{I}_{1,i}\times\mathfrak{I}_{2,i}\leq_l\mathfrak{R}_i$ with $0\neq\mathfrak{I}_{l,i}\leq_l\mathfrak{R}_{l,i}$ $(l=1,2)$, we have

$$\sum_{i\in T}\frac{R_{j,i}\log|\mathfrak{I}_i|}{\log|\mathfrak{R}_i|}=\sum_{i\in T}\frac{R_{j,i}\log|\mathfrak{I}_{1,i}|}{\log|\mathfrak{R}_{1,i}|}\frac{c_1}{c_1+c_2}+\sum_{i\in T}\frac{R_{j,i}\log|\mathfrak{I}_{2,i}|}{\log|\mathfrak{R}_{2,i}|}\frac{c_2}{c_1+c_2},$$

where $c_l=\log|\mathfrak{R}_{l,1}|$. By (20), it can be easily seen that

$$\sum_{i\in T}\frac{R_{j,i}\log|\mathfrak{I}_i|}{\log|\mathfrak{R}_i|}>H(X_T|X_{T^c})-\frac{1}{c_1+c_2}\sum_{l=1}^{2}c_lH(Y_{\mathfrak{R}_{l,T}/\mathfrak{I}_{l,T}}|X_{T^c}).$$

Meanwhile, let $\mathfrak{R}_T=\prod_{i\in T}\mathfrak{R}_i$, $\mathfrak{I}_T=\prod_{i\in T}\mathfrak{I}_i$, $\Phi=[\Phi_{1,1}\times\Phi_{2,1},\Phi_{1,2}\times\Phi_{2,2},\cdots,\Phi_{1,s}\times\Phi_{2,s}]$ (Note:

$$\Phi_{1,i}\times\Phi_{2,i}:x_i\mapsto(\Phi_{1,i}(x_i),\Phi_{2,i}(x_i))\in\mathfrak{R}_i$$

for all $x_i\in\mathscr{X}_i$.) and $Y_{\mathfrak{R}_T/\mathfrak{I}_T}=\Phi(X_T)+\mathfrak{I}_T$. It can be verified that $Y_{\mathfrak{R}_{l,T}/\mathfrak{I}_{l,T}}$ $(l=1,2)$ is a function of $Y_{\mathfrak{R}_T/\mathfrak{I}_T}$, hence, $H(Y_{\mathfrak{R}_T/\mathfrak{I}_T}|X_{T^c})\geq H(Y_{\mathfrak{R}_{l,T}/\mathfrak{I}_{l,T}}|X_{T^c})$. Consequently,

$$\sum_{i\in T}\frac{R_{j,i}\log|\mathfrak{I}_i|}{\log|\mathfrak{R}_i|}>H(X_T|X_{T^c})-H(Y_{\mathfrak{R}_T/\mathfrak{I}_T}|X_{T^c}),$$

which implies that $\mathbf{R}_j\in\mathcal{R}_{\Phi,\text{prod}}$ by Theorem III.3. We therefore conclude that $\mathbf{R}_j$ is achievable with linear encoders over $\mathfrak{R}_1,\mathfrak{R}_2,\cdots,\mathfrak{R}_s$ for all feasible j, so is $\mathbf{R}$. ■

Obviously, $\mathfrak{R}_1,\mathfrak{R}_2,\cdots,\mathfrak{R}_s$ in Theorem V.5 are of the same size. Inductively, one can verify the following without any difficulty.

Theorem V.6. *Let $\mathscr{L}$ be any finite index set, $\mathfrak{R}_{l,1},\mathfrak{R}_{l,2},\cdots,\mathfrak{R}_{l,s}$ $(l\in\mathscr{L})$ be a set of finite rings of equal size, and $\mathfrak{R}_i=\prod_{l\in\mathscr{L}}\mathfrak{R}_{l,i}$ for all feasible i. If the coding rate $\mathbf{R}\in\mathbb{R}^s$ is achievable with linear encoders over $\mathfrak{R}_{l,1},\mathfrak{R}_{l,2},\cdots,\mathfrak{R}_{l,s}$ $(l\in\mathscr{L})$, then $\mathbf{R}$ is achievable with linear encoders over $\mathfrak{R}_1,\mathfrak{R}_2,\cdots,\mathfrak{R}_s$.*

Remark 17. There are delicate issues to the situation Theorem V.6 (Theorem V.5) illustrates. Let $\mathscr{X}_i$ $(1\leq i\leq s)$ be the set of all symbols generated by the ith source. The hypothesis of Theorem V.6 (Theorem V.5) implicitly implies the alphabet constraint $|\mathscr{X}_i|\leq|\mathfrak{R}_{l,i}|$ for all feasible i and l.

Let $\mathfrak{R}_1,\mathfrak{R}_2,\cdots,\mathfrak{R}_s$ be s finite rings each of which is isomorphic to either

1) a ring $\mathfrak{R}$ containing one and only one proper non-trivial left ideal whose order is $\sqrt{|\mathfrak{R}|}$, e.g. $\mathbb{M}_{L,p}$ and $\mathbb{Z}_{p^2}$ (p is a prime); or
2) a ring of a finite product of finite field(s) and/or ring(s) satisfying 1), e.g. $\mathbb{M}_{L,p}\times\prod_{j=1}^{m}\mathbb{Z}_{p_j}$ (p and p_j's are prime) and $\prod_{i=1}^{m'}\mathbb{M}_{L,p_i}\times\prod_{j=1}^{m''}\mathbb{F}_{q_j}$ (m' and m'' are non-negative, p_i's are prime and q_j's are power of primes).

Theorem V.3 and Theorem V.6 ensure that linear encoders over ring $\mathfrak{R}_1,\mathfrak{R}_2,\cdots,\mathfrak{R}_s$ are always optimal in any applicable (subject to the condition specified in the corresponding theorem) SW coding scenario. As a very special

case, $\mathbb{Z}_p \times \mathbb{Z}_p$, where p is a prime, is always optimal in any (single source or multiple sources) scenario with alphabet size less than or equal to p. However, using a field or product rings is not necessary. As shown in Theorem V.2, neither $\mathbb{M}_{L,p}$ nor $\mathbb{Z}_{p^2}$ is (isomorphic to) a product of rings nor a field. It is also not required to have a restriction on the alphabet size (see Theorem V.3), even for product rings (see Example III.2 for a case of $\mathbb{Z}_2 \times \mathbb{Z}_3$).

D. Trivial Case: Uniform Distributions

The following theorem is trivial, however we include it for completeness.

Theorem V.7. *Regardless which set of rings $\mathfrak{R}_1, \mathfrak{R}_2, \cdots, \mathfrak{R}_s$ is chosen, as long as $|\mathfrak{R}_i| = |\mathscr{X}_i|$ for all feasible i, region* (4) *is the SW region if $(X_1, X_2, \cdots, X_s) \sim p$ is a uniform distribution.*

Proof: If p is uniform, then, for any $\emptyset \neq T \subseteq \mathcal{S}$ and $0 \neq \mathfrak{I}_T \leq_l \mathfrak{R}_T$, $Y_{\mathfrak{R}_T/\mathfrak{I}_T}$ is uniformly distributed on $\mathfrak{R}_T/\mathfrak{I}_T$. Moreover, X_T and X_{T^c} are independent, so are $Y_{\mathfrak{R}_T/\mathfrak{I}_T}$ and X_{T^c}. Therefore, $H(X_T|X_{T^c}) = H(X_T) = \log|\mathfrak{R}_T|$ and $H(Y_{\mathfrak{R}_T/\mathfrak{I}_T}|X_{T^c}) = H(Y_{\mathfrak{R}_T/\mathfrak{I}_T}) = \log\dfrac{|\mathfrak{R}_T|}{|\mathfrak{I}_T|}$. Consequently,

$$r(T, \mathfrak{I}_T) = H(X_T|X_{T^c}) - H(Y_{\mathfrak{R}_T/\mathfrak{I}_T}|X_{T^c}) = \log|\mathfrak{I}_T|.$$

Region (4) is the SW region. ■

Remark 18. When p is uniform, it is obvious that the uncoded strategy (all encoders are one-to-one mappings) is optimal in the SW source coding problem. However, optimality stated in Theorem V.7 does not come from deliberately fixing the linear encoding mappings, but generating them randomly.

So far, we have only shown that there exist linear encoders over finite non-field rings that are equally good as their field counterparts. In next section, Problem 1 is considered with an arbitrary g. It will be demonstrated that linear coding over finite non-field rings can *strictly outperform* its field counterpart for encoding some discrete functions, and there are infinitely many such functions.

VI. Application: Source Coding for Computing

The problem of Source Coding for Computing, Problem 1, with an arbitrary g is addressed in this section. Some advantages of LCoR (compared to LCoF) will be demonstrated. We begin with establishing the following theorem which can be recognized as a generalization of Körner–Marton [3].

Theorem VI.1. *Let $\mathfrak{R}$ be a finite ring, and*

$$\hat{g} = h \circ k, \text{ where } k(x_1, x_2, \cdots, x_s) = \sum_{i=1}^{s} k_i(x_i) \tag{21}$$

and h, k_i's are functions mapping $\mathfrak{R}$ to $\mathfrak{R}$. Then

$$\mathcal{R}_{\hat{g}} = \left\{(r, r, \cdots, r) \in \mathbb{R}^s \middle| r > \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log|\mathfrak{R}|}{\log|\mathfrak{I}|}\left[H(X) - H(Y_{\mathfrak{R}/\mathfrak{I}})\right]\right\} \subseteq \mathcal{R}[\hat{g}], \tag{22}$$

where $X = k(X_1, X_2, \cdots, X_s)$ and $Y_{\mathfrak{R}/\mathfrak{I}} = X + \mathfrak{I}$.

Proof: By Theorem III.1, $\forall\, \epsilon > 0$, there exists a large enough n, an $m \times n$ matrix $\mathbf{A} \in \mathfrak{R}^{m\times n}$ and a decoder ψ, such that $\Pr\{X^n \neq \psi(\mathbf{A}X^n)\} < \epsilon$, if $m > \max\limits_{0\neq\mathfrak{I}\leq_l\mathfrak{R}} \dfrac{n(H(X)-H(Y_{\mathfrak{R}/\mathfrak{I}}))}{\log|\mathfrak{I}|}$. Let $\phi_i = \mathbf{A}\circ\vec{k}_i$ $(1\leq i\leq s)$ be the encoder of the ith source. Upon receiving $\phi_i(X_i^n)$ from the ith source, the decoder claims that $\vec{h}\left(\hat{X}^n\right)$, where $\hat{X}^n = \psi\left[\sum\limits_{i=1}^{s}\phi_i\left(X_i^n\right)\right]$, is the function, namely $\hat{g}$, subject to computation. The probability of decoding error is

$$\begin{aligned}
\Pr\left\{\vec{h}\left[\vec{k}\left(X_1^n, X_2^n, \cdots, X_s^n\right)\right] \neq \vec{h}\left(\hat{X}^n\right)\right\} &\leq \Pr\left\{X^n \neq \hat{X}^n\right\} \\
&= \Pr\left\{X^n \neq \psi\left[\sum_{i=1}^{s}\phi_i\left(X_i^n\right)\right]\right\} \\
&= \Pr\left\{X^n \neq \psi\left[\sum_{i=1}^{s}\mathbf{A}\vec{k}_i\left(X_i^n\right)\right]\right\} \\
&= \Pr\left\{X^n \neq \psi\left[\mathbf{A}\sum_{i=1}^{s}\vec{k}_i\left(X_i^n\right)\right]\right\} \\
&= \Pr\left\{X^n \neq \psi\left[\mathbf{A}\vec{k}\left(X_1^n, X_2^n, \cdots, X_s^n\right)\right]\right\} \\
&= \Pr\left\{X^n \neq \psi\left(\mathbf{A}X^n\right)\right\} < \epsilon.
\end{aligned}$$

Therefore, all $(r, r, \cdots, r) \in \mathbb{R}^s$, where $r = \dfrac{m\log|\mathfrak{R}|}{n} > \max\limits_{0\neq\mathfrak{I}\leq_l\mathfrak{R}} \dfrac{\log|\mathfrak{R}|}{\log|\mathfrak{I}|}\left[H(X) - H(Y_{\mathfrak{R}/\mathfrak{I}})\right]$, is achievable, i.e. $\mathcal{R}_{\hat{g}} \subseteq \mathcal{R}[\hat{g}]$. ∎

Corollary VI.2. *In Theorem VI.1, let $X = k(X_1, X_2, \cdots, X_s) \sim p_X$. We have*

$$\mathcal{R}_{\hat{g}} = \left\{(r, r, \cdots, r) \in \mathbb{R}^s \middle|\, r > H(X)\right\} \subseteq \mathcal{R}[\hat{g}],$$

if either of the following conditions holds:

1) *$\mathfrak{R}$ is isomorphic to a finite field;*
2) *$\mathfrak{R}$ is isomorphic to a ring containing one and only one proper non-trivial left ideal $\mathfrak{I}_0$ with $|\mathfrak{I}_0| = \sqrt{|\mathfrak{R}|}$, and*

$$H(X) \leq 2H(X + \mathfrak{I}_0).$$

Proof: If either 1) or 2) holds, then it is guaranteed that

$$\max_{0\neq\mathfrak{I}\leq_l\mathfrak{R}} \frac{\log|\mathfrak{R}|}{\log|\mathfrak{I}|}\left[H(X) - H(Y_{\mathfrak{R}/\mathfrak{I}})\right] = H(X)$$

in Theorem VI.1. The statement follows. ∎

Remark 19. By Lemma A.2, examples of non-field rings satisfying 2) in Corollary VI.2 includes

(1) $\mathbb{Z}_4$ with $p_X(0) = p_1, p_X(1) = p_2, p_X(3) = p_3$ and $p_X(2) = p_4$ satisfying

$$0 \leq \max\{p_2, p_3\} \not< \min\{p_1, p_4\} \leq 1 \text{ and } 0 \leq \max\{p_1, p_4\} \not< \min\{p_2, p_3\} \leq 1, \tag{23}$$

(2) $\mathbb{M}_{L,2}$ with

$$p_X\left(\begin{bmatrix}0&0\\0&0\end{bmatrix}\right)=p_1, p_X\left(\begin{bmatrix}1&0\\0&1\end{bmatrix}\right)=p_2, p_X\left(\begin{bmatrix}1&0\\1&1\end{bmatrix}\right)=p_3 \text{ and } p_X\left(\begin{bmatrix}0&0\\1&0\end{bmatrix}\right)=p_4$$

satisfying (23) and etc.

Interested readers can figure out even more explicit examples deduced from Lemma A.1.

Remark 20. If $\mathfrak{R}$ is isomorphic to $\mathbb{Z}_2$ and $\hat{g}$ is the modulo-two sum, then Corollary VI.2 recovers the theorem of Körner–Marton [3]. While if $\mathfrak{R}$ is (isomorphic to) a field, it becomes a special case of [7, Theorem III.1]. Actually, almost all the results in [6] and [7] can be reproved in the setting of rings in a parallel fashion.

We claim that there are functions g for which LCoR outperforms LCoF; in fact, there are infinite many such g's. To prove this, some definitions are required for the mechanics of our argument.

Definition VI.3. Let $g_1 : \prod_{i=1}^{s} \mathscr{X}_i \to \Omega_1$ and $g_2 : \prod_{i=1}^{s} \mathscr{Y}_i \to \Omega_2$ be two functions. If there exist bijections $\mu_i : \mathscr{X}_i \to \mathscr{Y}_i$, $\forall\ 1 \le i \le s$, and $\nu : \Omega_1 \to \Omega_2$, such that

$$g_1(x_1, x_2, \cdots, x_s) = \nu^{-1}(g_2(\mu_1(x_1), \mu_2(x_2), \cdots, \mu_s(x_s))),$$

then g_1 and g_2 are said to be *equivalent* (via $\mu_1, \mu_2, \cdots, \mu_s$ and ν).

Definition VI.4. Given function $g : \mathscr{D} \to \Omega$, and let $\emptyset \neq \mathscr{S} \subseteq \mathscr{D}$. The *restriction* of g on $\mathscr{S}$ is defined to be the function $g|_{\mathscr{S}} : \mathscr{S} \to \Omega$ such that $g|_{\mathscr{S}} : x \mapsto g(x), \forall\ x \in \mathscr{S}$.

Lemma VI.5. *For any discrete function $g : \prod_{i=1}^{k} \mathscr{X}_i \to \Omega$ with $\mathscr{X}_i$'s and Ω being finite, there always exist a finite ring (field) and a polynomial function $\hat{g} \in \mathfrak{R}[k]$ such that*

$$\nu\left(g\left(x_1, x_2, \cdots, x_k\right)\right) = \hat{g}\left(\mu_1(x_1), \mu_2(x_2), \cdots, \mu_k(x_k)\right)$$

for some injections $\mu_i : \mathscr{X}_i \to \mathfrak{R}$ $(1 \le i \le k)$ and $\nu : \Omega \to \mathfrak{R}$.

Proof: There are several possible proofs of this lemma. One is provided in appendix B. ■

Remark 21. Up to equivalence, a function can be presented in many different formats. For example, the function $\min\{x, y\}$ defined on $\{0,1\} \times \{0,1\}$ (with ordering $0 \le 1$) can either be seen as $F_1(x, y) = xy$ on $\mathbb{Z}_2^2$ or be treated as the restriction of $F_2(x, y) = x + y - (x + y)^2$ defined on $\mathbb{Z}_3^2$ to the domain $\{0,1\} \times \{0,1\} \subsetneq \mathbb{Z}_3^2$.

Lemma VI.5 implies that any discrete function defined on a finite domain is equivalent to a restriction of some polynomial function over some finite ring (field). As a consequence, we can restrict Problem 1 to all polynomial functions. This polynomial approach offers valuable insight into the general problem, because the algebraic structure of a polynomial function is clearer than that of an arbitrary function. We often call $\hat{g}$ in Lemma VI.5 a *polynomial presentation* of g. On the other hand, the $\hat{g}$ given by (21) is named a *nomographic function* over $\mathfrak{R}$ (by terminology borrowed from [18]), it is said to be a *nomographic presentation* of g if g is equivalent to a restriction of it.

Lemma VI.6. *Let $\mathscr{X}_1, \mathscr{X}_2, \cdots, \mathscr{X}_s$ and Ω be some finite sets. For any discrete function $g : \prod_{i=1}^{s} \mathscr{X}_i \to \Omega$, there exists a nomographic function $\hat{g}$ over some finite ring (field) $\mathfrak{R}$ such that*

$$\nu\left(g\left(x_1, x_2, \cdots, x_k\right)\right) = \hat{g}\left(\mu_1(x_1), \mu_2(x_2), \cdots, \mu_k(x_k)\right)$$

for some injections $\mu_i : \mathscr{X}_i \to \mathfrak{R}$ $(1 \leq i \leq k)$ and $\nu : \Omega \to \mathfrak{R}$.

Proof: There are several proofs of this lemma. One is provided in appendix B. ■

Lemma VI.6 advances Lemma VI.5 by claiming that a discrete function with a finite domain is always equivalent to a restriction of some nomographic function. From this, it is seen that Theorem VI.1 and Corollary VI.2 have presented a universal solution to Problem 1.

Given some finite ring $\mathfrak{R}$, let $\hat{g}$ of format (21) be a nomographic presentation of g. We say that the region $\mathcal{R}_{\hat{g}}$ given by (22) is achievable for computing g in the sense of Körner–Marton. From Theorem VI.11 given later, we know that $\mathcal{R}_{\hat{g}}$ might not be the largest achievable region one can obtain for computing g. However, $\mathcal{R}_{\hat{g}}$ still captures the ability of linear coding over $\mathfrak{R}$ when used for computing g. In other words, $\mathcal{R}_{\hat{g}}$ is the region purely achieved with linear coding over $\mathfrak{R}$ for computing g. On the other hand, regions from Theorem VI.11 are achieved by combining the linear coding and the standard random coding techniques. Therefore, it is reasonable to compare LCoR with LCoF in the sense of Körner–Marton.

We show that linear coding over finite rings, non-field rings in particular, strictly outperforms its field counterpart, LCoF, in the following example.

Example VI.7 ([19])**.** Let $g : \{\alpha_0, \alpha_1\}^3 \to \{\beta_0, \beta_1, \beta_2, \beta_3\}$ (Fig 1) be a function such that

$$\begin{aligned}
&g : (\alpha_0, \alpha_0, \alpha_0) \mapsto \beta_0; \quad g : (\alpha_0, \alpha_0, \alpha_1) \mapsto \beta_3;\\
&g : (\alpha_0, \alpha_1, \alpha_0) \mapsto \beta_2; \quad g : (\alpha_0, \alpha_1, \alpha_1) \mapsto \beta_1;\\
&g : (\alpha_1, \alpha_0, \alpha_0) \mapsto \beta_1; \quad g : (\alpha_1, \alpha_0, \alpha_1) \mapsto \beta_0;\\
&g : (\alpha_1, \alpha_1, \alpha_0) \mapsto \beta_3; \quad g : (\alpha_1, \alpha_1, \alpha_1) \mapsto \beta_2.
\end{aligned} \tag{24}$$

Define $\mu : \{\alpha_0, \alpha_1\} \to \mathbb{Z}_4$ and $\nu : \{\beta_0, \beta_1, \beta_2, \beta_3\} \to \mathbb{Z}_4$ by

$$\begin{aligned}
&\mu : \alpha_j \mapsto j, \quad \forall\, j \in \{0, 1\}, \text{ and}\\
&\nu : \beta_j \mapsto j, \quad \forall\, j \in \{0, 1, 2, 3\},
\end{aligned} \tag{25}$$

respectively. Obviously, g is equivalent to $x + 2y + 3z \in \mathbb{Z}_4[3]$ (Fig 2) via $\mu_1 = \mu_2 = \mu_3 = \mu$ and ν. However, by Proposition VI.8, there exists no $\hat{g} \in \mathbb{F}_4[3]$ of format (21) so that g is equivalent to any restriction of $\hat{g}$. Although, Lemma VI.6 ensures that there always exists a bigger field $\mathbb{F}_q$ such that g admits a presentation $\hat{g} \in \mathbb{F}_q[3]$ of format (21), the size q must be strictly bigger than 4. For instance, let

$$\hat{h}(x) = \sum_{a \in \mathbb{Z}_5} a\left[1 - (x - a)^4\right] - \left[1 - (x - 4)^4\right] \in \mathbb{Z}_5[1].$$

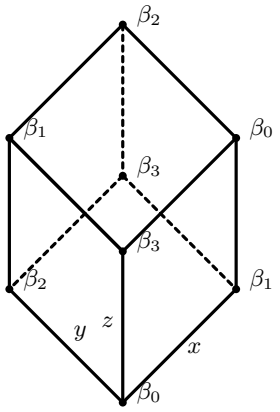

Fig 1: $g : \{\alpha_0, \alpha_1\}^3 \to \{\beta_0, \beta_1, \beta_2, \beta_3\}$

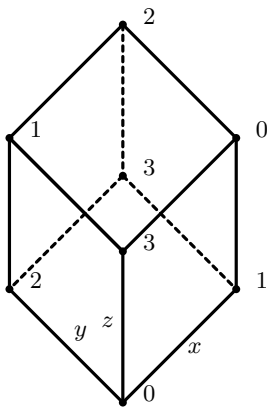

Fig 2: $x + 2y + 3z \in \mathbb{Z}_4[3]$

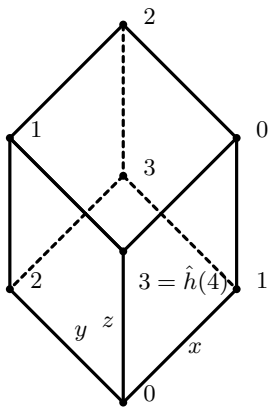

Fig 3: $\hat{h}(x + 2y + 4z) \in \mathbb{Z}_5[3]$

Then, g has presentation $\hat{h}(x + 2y + 4z) \in \mathbb{Z}_5[3]$ (Fig 3) via $\mu_1 = \mu_2 = \mu_3 = \mu : \{\alpha_0, \alpha_1\} \to \mathbb{Z}_5$ and $\nu : \{\beta_0, \beta_1, \beta_2, \beta_3\} \to \mathbb{Z}_5$ defined (symbolic-wise) by (25).

Proposition VI.8. *There exists no polynomial function $\hat{g} \in \mathbb{F}_4[3]$ of format* (21)*, such that a restriction of $\hat{g}$ is equivalent to the function g defined by* (24)*.*

Proof: Suppose $\nu \circ g = \hat{g} \circ (\mu_1, \mu_2, \mu_3)$, where $\mu_1, \mu_1, \mu_3 : \{\alpha_0, \alpha_1\} \to \mathbb{F}_4$, $\nu : \{\beta_0, \cdots, \beta_3\} \to \mathbb{F}_4$ are injections, and $\hat{g} = h \circ (k_1 + k_2 + k_3)$ with $h, k_i \in \mathbb{F}_4[1]$ for all feasible i. We claim that $\hat{g}$ and h are both surjective, since $\left|g\left(\{\alpha_0, \alpha_1\}^3\right)\right| = |\{\beta_0, \beta_1, \beta_2, \beta_3\}| = 4 = |\mathbb{F}_4|$. In particular, h is bijective. Therefore, $h^{-1} \circ \nu \circ g = k_1 \circ \mu_1 + k_2 \circ \mu_2 + k_3 \circ \mu_3$, i.e. g admits a presentation $k_1(x) + k_2(y) + k_3(z) \in \mathbb{F}_4[3]$. A contradiction to Lemma A.3. ■

As a consequence of Proposition VI.8, in the sense of Körner–Marton, in order to use LCoF to encode function g, the alphabet sizes of the three encoders need to be at least 5. However, LCoR offers a solution in which the alphabet sizes are 4, strictly smaller than using LCoF. Most importantly, the region achieved with linear coding over any finite field $\mathbb{F}_q$, is always a subset of the one achieved with linear coding over $\mathbb{Z}_4$. This is proved in the following proposition.

Proposition VI.9. *Let g be the function defined by* (24)*, $\{\alpha_0, \alpha_1\}^3$ be the sample space of $(X_1, X_2, X_3) \sim p$ and p_X be the distribution of $X = g(X_1, X_2, X_3)$. If*

$$p_X(\beta_0) = p_1, p_X(\beta_1) = p_2, p_X(\beta_3) = p_3 \text{ and } p_X(\beta_2) = p_4$$

satisfying (23)*, then, in the sense of Körner–Marton, the region $\mathcal{R}_1$ achieved with linear coding over $\mathbb{Z}_4$ contains the one, that is $\mathcal{R}_2$, obtained with linear coding over any finite field $\mathbb{F}_q$ for computing g. Moreover, if* $\mathrm{supp}(p)$ *is the whole domain of g, then $\mathcal{R}_1 \supsetneq \mathcal{R}_2$.*

Proof: Let $\hat{g} = h \circ k \in \mathbb{F}_q[3]$ be a polynomial presentation of g with format (21). By Corollary VI.2 and

(X_1, X_2, X_3)	p	(X_1, X_2, X_3)	p
$(\alpha_0, \alpha_0, \alpha_0)$	$1/90$	$(\alpha_0, \alpha_1, \alpha_0)$	$1/90$
$(\alpha_1, \alpha_0, \alpha_1)$	$1/90$	$(\alpha_1, \alpha_1, \alpha_1)$	$1/90$
$(\alpha_1, \alpha_0, \alpha_0)$	$42/90$	$(\alpha_0, \alpha_0, \alpha_1)$	$1/90$
$(\alpha_0, \alpha_1, \alpha_1)$	$42/90$	$(\alpha_1, \alpha_1, \alpha_0)$	$1/90$

Table 1

Remark 19, we have

$$\mathcal{R}_1 = \left\{ (R_1, R_2, R_3) \in \mathbb{R}^3 \middle| R_i > H(X_1 + 2X_2 + 3X_3) \right\},$$
$$\mathcal{R}_2 = \left\{ (R_1, R_2, R_3) \in \mathbb{R}^3 \middle| R_i > H(k(X_1, X_2, X_3)) \right\}.$$

Assume that $\nu \circ g = h \circ k \circ (\mu_1, \mu_2, \mu_3)$, where $\mu_1, \mu_1, \mu_3 : \{\alpha_0, \alpha_1\} \to \mathbb{F}_q$ and $\nu : \{\beta_0, \cdots, \beta_3\} \to \mathbb{F}_q$ are injections. Obviously, $g(X_1, X_2, X_3)$ is a function of $k(X_1, X_2, X_3)$. Hence,

$$H(k(X_1, X_2, X_3)) \geq H(g(X_1, X_2, X_3)). \tag{26}$$

On the other hand, $H(X_1 + 2X_2 + 3X_3) = H(g(X_1, X_2, X_3))$. Therefore,

$$H(k(X_1, X_2, X_3)) \geq H(X_1 + 2X_2 + 3X_3), \tag{27}$$

and $\mathcal{R}_1 \supseteq \mathcal{R}_2$. In addition, we claim that $h|_{\mathscr{S}}$, where $\mathscr{S} = k\left(\prod_{j=1}^{3} \mu_j\{\alpha_0, \alpha_1\}\right)$, is not injective. Otherwise, $h : \mathscr{S} \to \mathscr{S}'$, where $\mathscr{S}' = h(\mathscr{S})$, is bijective, hence, $(h|_{\mathscr{S}'})^{-1} \circ \nu \circ g = k \circ (\mu_1, \mu_2, \mu_3) = k_1 \circ \mu_1 + k_2 \circ \mu_2 + k_3 \circ \mu_3$. A contradiction to Lemma A.3. Consequently, $|\mathscr{S}| > |\mathscr{S}'| = |\nu(\{\beta_0, \cdots, \beta_3\})| = 4$. If $\text{supp}(p) = \{\alpha_0, \alpha_1\}^3$, then (26) as well as (27) hold strictly, thus, $\mathcal{R}_1 \supsetneq \mathcal{R}_2$. ■

A more intuitive comparison (which is not as conclusive as Proposition VI.9) can be identified from the presentations of g given in Fig 2 and Fig 3. According to Corollary VI.2, linear encoders over field $\mathbb{Z}_5$ achieve

$$\mathcal{R}_{\mathbb{Z}_5} = \left\{ (R_1, R_2, R_3) \in \mathbb{R}^3 \middle| R_i > H(X_1 + 2X_2 + 4X_3) \right\}.$$

The one achieved by linear encoders over ring $\mathbb{Z}_4$ is

$$\mathcal{R}_{\mathbb{Z}_4} = \left\{ (R_1, R_2, R_3) \in \mathbb{R}^3 \middle| R_i > H(X_1 + 2X_2 + 3X_3) \right\}.$$

Clearly, $H(X_1 + 2X_2 + 3X_3) \leq H(X_1 + 2X_2 + 4X_3)$, thus, $\mathcal{R}_{\mathbb{Z}_4}$ contains $\mathcal{R}_{\mathbb{Z}_5}$. Furthermore, as long as

$$0 < \Pr(\alpha_0, \alpha_0, \alpha_1), \Pr(\alpha_1, \alpha_1, \alpha_0) < 1,$$

$\mathcal{R}_{\mathbb{Z}_4}$ is strictly larger than $\mathcal{R}_{\mathbb{Z}_5}$, since $H(X_1 + 2X_2 + 3X_3) < H(X_1 + 2X_2 + 4X_3)$. To be specific, assume that $(X_1, X_2, X_3) \sim p$ satisfies Table 1, we have

$$\mathcal{R}[X_1, X_2, X_3] \subsetneq \mathcal{R}_{\mathbb{Z}_5} = \left\{ (R_1, R_2, R_3) \in \mathbb{R}^3 \middle| R_i > 0.4812 \right\}$$
$$\subsetneq \mathcal{R}_{\mathbb{Z}_4} = \left\{ (R_1, R_2, R_3) \in \mathbb{R}^3 \middle| R_i > 0.4590 \right\}.$$

Based on Proposition VI.8 and Proposition VI.9, we conclude that LCoR dominates LCoF, in terms of achieving better coding rates with smaller alphabet sizes of the encoders for computing g. As a direct conclusion, we have:

Theorem VI.10. *In the sense of Körner–Marton, LCoF is not optimal.*

Remark 22. The key property underlying the proof of Proposition VI.9 is that the characteristic of a finite field must be a prime while the characteristic of a finite ring can be any positive integer larger than or equal to 2. This implies that it is possible to construct infinitely many discrete functions for which using LCoF always leads to a suboptimal achievable region compared to linear coding over finite non-field rings. Examples include $\sum_{i=1}^{s} x_i \in \mathbb{Z}_{2p}[s]$ for $s \geq 2$ and prime $p > 2$ (note: the characteristic of $\mathbb{Z}_{2p}$ is $2p$ which is not a prime). One can always find an explicit distribution of sources for which linear coding over $\mathbb{Z}_{2p}$ strictly dominates linear coding over each and every finite field.

As mentioned, $\mathcal{R}_{\hat{g}}$ given by (22) is sometimes strictly smaller than $\mathcal{R}[g]$. This was first shown by Ahlswede–Han [4] for the case of g being the modulo-two sum. Their approach combines the linear coding technique over binary field with the standard random coding technique. In the following, we generalize the result of Ahlswede–Han [4, Theorem 10] to the settings, where g is arbitrary, and, at the same time, LCoF is replaced by its generalized version, LCoR.

Consider function $\hat{g}$ admitting

$$\hat{g}(x_1, x_2, \cdots, x_s) = h\Big[k_0(x_1, x_2, \cdots, x_{s_0}), \sum_{j=s_0+1}^{s} k_j(x_j)\Big], 0 \leq s_0 < s, \tag{28}$$

where $k_0 : \mathfrak{R}^{s_0} \to \mathfrak{R}$ and h, k_j's are functions mapping $\mathfrak{R}$ to $\mathfrak{R}$. By Lemma VI.6, a discrete function with a finite domain is always equivalent to a restriction of some function of format (28). We call $\hat{g}$ from (28) a *pseudo nomographic function* over ring $\mathfrak{R}$.

Theorem VI.11. *Let $\mathcal{S}_0 = \{1, 2, \cdots, s_0\} \subseteq \mathcal{S} = \{1, 2, \cdots, s\}$. If $\hat{g}$ is of format* (28)*, and $\mathbf{R} = (R_1, R_2, \cdots, R_s) \in \mathbb{R}^s$ satisfying*

$$\sum_{j \in T} R_j > |T \setminus \mathcal{S}_0| \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log |\mathfrak{R}|}{\log |\mathfrak{I}|} \big[H(X|V_{\mathcal{S}}) - H(Y_{\mathfrak{R}/\mathfrak{I}}|V_{\mathcal{S}})\big] + I(Y_T; V_T|V_{T^c}), \forall\, \emptyset \neq T \subseteq \mathcal{S}, \tag{29}$$

where $\forall\, j \in \mathcal{S}_0$, $V_j = Y_j = X_j$; $\forall\, j \in \mathcal{S} \setminus \mathcal{S}_0$, $Y_j = k_j(X_j)$, V_j's are discrete random variables such that

$$p(y_1, y_2, \cdots, y_s, v_1, v_2, \cdots, v_s) = p(y_1, y_2, \cdots, y_s) \prod_{j=s_0+1}^{s} p(v_j|y_j), \tag{30}$$

and $X = \sum_{j=s_0+1}^{s} Y_j$, $Y_{\mathfrak{R}/\mathfrak{I}} = X + \mathfrak{I}$, then $\mathbf{R} \in \mathcal{R}[\hat{g}]$.

Proof: The proof can be completed by applying the tricks from Lemma III.5 and Lemma III.7 to the approach generalized from Ahlswede–Han [4, Theorem 10]. Details are found in Appendix D. ■

Remark 23. The achievable region given by (29) always contains the SW region. Moreover, it is in general larger than the $\mathcal{R}_{\hat{g}}$ from (22). If $\hat{g}$ is the modulo-two sum, namely $s_0 = 0$ and h, k_j's are identity functions for all $s_0 < j \leq s$, then (29) resumes the region of Ahlswede–Han [4, Theorem 10].

VII. Conclusion

A. Right Linearity

Careful readers might have noticed that the encoders we used so far are actually left linear mappings. By symmetry, almost all related statements can be easily reproved for right linear mappings (encoders). As an example, the following corresponds to Theorem III.1.

Theorem VII.1. *For any $\Phi \in \mathscr{M}\left(\mathscr{X}_{\mathcal{S}}, \mathfrak{R}_{\mathcal{S}}\right)$,*

$$\mathcal{R}'_{\Phi} = \left\{ [R_1, R_2, \cdots, R_s] \in \mathbb{R}^s \middle| \sum_{i \in T} \frac{R_i \log |\mathfrak{I}_i|}{\log |\mathfrak{R}_i|} > r\left(T, \mathfrak{I}_T\right), \forall\, \emptyset \neq T \subseteq \mathcal{S}, \forall\, 0 \neq \mathfrak{I}_i \leq_r \mathfrak{R}_i \right\}, \tag{31}$$

where $r\left(T, \mathfrak{I}_T\right) = H(X_T|X_{T^c}) - H(Y_{\mathfrak{R}_T/\mathfrak{I}_T}|X_{T^c})$ and $Y_{\mathfrak{R}_T/\mathfrak{I}_T} = \Phi(X_T) + \mathfrak{I}_T$, is achievable with (right) linear coding over the finite rings $\mathfrak{R}_1, \mathfrak{R}_2, \cdots, \mathfrak{R}_s$.

By time sharing,

$$\mathcal{R}_r = \mathrm{cov}\left(\bigcup_{\Phi \in \mathscr{M}\left(\mathscr{X}_{\mathcal{S}}, \mathfrak{R}_{\mathcal{S}}\right)} \mathcal{R}'_{\Phi} \right), \tag{32}$$

where $\mathcal{R}'_{\Phi}$ is given by (31), is achievable with (right) LCoR.

B. Field, Ring, Rng and Group

Conceptually speaking, LCoR is in fact a generalization of the linear coding technique proposed by Elias [1] and Csiszár [2] (LCoF), since a field must be a ring. However, as seen in Section IV, analyzing the decoding error for the ring version is in general substantially more challenging than in the case of the field version. Our approach crucially relies on the concept of ideals. A field contains no non-trivial ideal but itself. Because of this special property of fields, our general argument for finite rings will render to a simple one when only finite fields are considered.

Even though our analysis for the ring scenario is more complicated than the one for field, linear encoders working over some finite rings are in general considerably easier to implement in practice. This is because the implementation of *finite field arithmetic* can be quite demanding. Normally, a finite field is given by its *polynomial representation*, operations are carried out based on the polynomial operations (addition and multiplication) followed by the *polynomial long division algorithm*. In contrast, implementing arithmetic of many finite rings is a straightforward task. For instance, the arithmetic of *modulo integers ring* $\mathbb{Z}_q$, for any positive integer q, is simply the integer modulo q arithmetic.

In addition, it is also very interesting to consider instead linear coding over rngs. It will be even more intriguing should it turn out that the rng version outperforms the ring version in the computing problem (Problem 1), in the

same manner that the ring version outperforms its field counterpart. It will also be interesting to see whether the idea of using rng provides more understanding of the problems from [8] and [6].

Some works, including [20], [21], [22], have proposed to implement coding over a simpler algebraic structure, that of a group. Seemingly, this corresponds to a more universal approach since both fields and rings are also groups. However, one subtle issue is often overlooked in this context. Namely, the set of rings (or rngs) is not a subset of the set of groups, since several non-isomorphic rings (or rngs) can be defined on one and the same group. For instance, given two distinct primes p and q, up to isomorphism,

1) there are 2 finite rngs of order p, while there is only one group of order p;
2) there are 4 finite rngs of order pq;
3) there are 11 finite rngs of order p^2 (if $p = 2$, then 4 of them are rings, namely $\mathbb{F}_4$, $\mathbb{Z}_4$, $\mathbb{Z}_2 \times \mathbb{Z}_2$ and $\mathbb{M}_{L,2}$ [23]), while there are only 2 groups of order p^2, both of which are Abelian;
4) there are 22 finite rngs of order p^2q;
5) there are 52 finite rngs of order 8;
6) there are $3p+50$ finite rngs of order p^3 ($p > 2$), while there are 5 groups of order p^3, 3 of which are Abelian (More can be found from [24]).

Therefore, there is no one-to-one correspondence between rings (field or rngs) and groups, in either direction. Furthermore, from the point of view of formulating a *multivariate function*, one is highly restricted by using groups, compared to rings (rng or field). Specifically, it is well-known that every discrete function defined on a finite domain is essentially a restriction of some polynomial function over a finite ring (rng or field). Although non-Abelian structures (non-Abelian groups) have the potential to lead to important non-trivial results [25], they are very difficult to handle theoretically and in practice. The performance of non-Abelian group block codes can be quite bad [26].

C. Final Remarks

This paper establishes achievability theorems regarding linear coding over finite rings for Slepian–Wolf data compression. Our results include related work from Elias [1] and Csiszár [2] regarding linear coding over finite fields as special cases in the sense of characterizing the achievable region. We have also proved that, for any Slepian–Wolf scenario, there always exists a sequence of linear encoders over some finite rings (non-field rings in particular) that achieves the data compression limit, the Slepian–Wolf region. Thus, with regard to existence, the optimality issue of linear coding over finite non-field rings for data compression is confirmed positively.

In addition, we also address the problem of source coding for computing, Problem 1. Results of Körner–Marton [3], Ahlswede–Han [4, Theorem 10] and [7] are generalized to corresponding ring versions. Based on these, it is demonstrated that LCoR dominates its field counterpart for encoding (infinitely) many discrete functions.

APPENDIX A
SUPPORTING LEMMATA

Lemma A.1. *Let $\mathfrak{R}$ be a finite ring, X and Y be two correlated discrete random variables, and $\mathscr{X}$ be the sample space of X with $|\mathscr{X}| \leq |\mathfrak{R}|$. If $\mathfrak{R}$ contains one and only one proper non-trivial left ideal $\mathfrak{I}$ and $|\mathfrak{I}| = \sqrt{|\mathfrak{R}|}$, then there exists injection $\tilde{\Phi} : \mathscr{X} \to \mathfrak{R}$ such that*

$$H(X|Y) \leq 2H(\tilde{\Phi}(X) + \mathfrak{I}|Y). \tag{A.1}$$

Proof: Let

$$\tilde{\Phi} \in \arg\max_{\Phi \in \mathscr{M}} H(\Phi(X) + \mathfrak{I}|Y),$$

where $\mathscr{M}$ is the set of all possible Φ's (maximum can always be reached because $|\mathscr{M}| = \dfrac{|\mathfrak{R}|!}{(|\mathfrak{R}| - |\mathscr{X}|)!}$ is finite, but it is not uniquely attained by $\tilde{\Phi}$ in general). Assume that $\mathscr{Y}$ is the sample space (not necessarily finite) of Y. Let $q = |\mathfrak{I}|$, $\mathfrak{I} = \{r_1, r_2, \cdots, r_q\}$ and $\mathfrak{R}/\mathfrak{I} = \{a_1 + \mathfrak{I}, a_2 + \mathfrak{I}, \cdots, a_q + \mathfrak{I}\}$. We have that

$$H(X|Y) = -\sum_{y \in \mathscr{Y}} \sum_{i,j=1}^{q} p_{i,j,y} \log \frac{p_{i,j,y}}{p_y} \text{ and}$$

$$H(\tilde{\Phi}(X) + \mathfrak{I}|Y) = -\sum_{y \in \mathscr{Y}} \sum_{i=1}^{q} p_{i,y} \log \frac{p_{i,y}}{p_y},$$

where

$$p_{i,j,y} = \Pr\left\{\tilde{\Phi}(X) = a_i + r_j, Y = y\right\},$$

$$p_y = \sum_{i,j=1}^{q} p_{i,j,y},$$

$$p_{i,y} = \sum_{j=1}^{q} p_{i,j,y}.$$

(Note: $\Pr\left\{\tilde{\Phi}(X) = r\right\} = 0$ if $r \in \mathfrak{R} \setminus \tilde{\Phi}(\mathscr{X})$. In addition, every element in $\mathfrak{R}$ can be uniquely expressed as $a_i + r_j$.) Therefore, (A.1) is equivalent to

$$-\sum_{y \in \mathscr{Y}} \sum_{i,j=1}^{q} p_{i,j,y} \log \frac{p_{i,j,y}}{p_y} \leq -2 \sum_{y \in \mathscr{Y}} \sum_{i=1}^{q} p_{i,y} \log \frac{p_{i,y}}{p_y}$$

$$\Leftrightarrow \sum_{y \in \mathscr{Y}} p_y \sum_{i=1}^{q} \frac{p_{i,y}}{p_y} H\left(\frac{p_{i,1,y}}{p_{i,y}}, \frac{p_{i,2,y}}{p_{i,y}}, \cdots, \frac{p_{i,q,y}}{p_{i,y}}\right) \leq \sum_{y \in \mathscr{Y}} p_y H\left(\frac{p_{1,y}}{p_y}, \frac{p_{2,y}}{p_y}, \cdots, \frac{p_{q,y}}{p_y}\right), \tag{A.2}$$

where $H(v_1, v_2, \cdots, v_q) = -\sum_{j=1}^{q} v_j \log v_j$, by the *grouping rule for entropy* [16, pp. 49]. Let

$$A = \sum_{y \in \mathscr{Y}} p_y H\left(\sum_{i=1}^{q} \frac{p_{i,1,y}}{p_y}, \sum_{i=1}^{q} \frac{p_{i,2,y}}{p_y}, \cdots, \sum_{i=1}^{q} \frac{p_{i,q,y}}{p_y}\right).$$

The concavity of the function H implies that

$$\sum_{y \in \mathscr{Y}} p_y \sum_{i=1}^{q} \frac{p_{i,y}}{p_y} H\left(\frac{p_{i,1,y}}{p_{i,y}}, \frac{p_{i,2,y}}{p_{i,y}}, \cdots, \frac{p_{i,q,y}}{p_{i,y}}\right) \leq A. \tag{A.3}$$

At the same time,

$$\sum_{y\in\mathscr{Y}} p_y H\left(\frac{p_{1,y}}{p_y}, \frac{p_{2,y}}{p_y}, \cdots, \frac{p_{q,y}}{p_y}\right) = \max_{\Phi\in\mathscr{M}} H(\Phi(X)+\mathfrak{I}|Y)$$

by the definition of $\tilde{\Phi}$. We now claim that

$$A \leq \max_{\Phi\in\mathscr{M}} H(\Phi(X)+\mathfrak{I}|Y). \tag{A.4}$$

Suppose otherwise, i.e. $A > \sum_{y\in\mathscr{Y}} p_y H\left(\frac{p_{1,y}}{p_y}, \frac{p_{2,y}}{p_y}, \cdots, \frac{p_{q,y}}{p_y}\right)$. Let $\Phi' : \mathscr{X} \to \mathfrak{R}$ be defined as

$$\Phi' : x \mapsto a_j + r_i \Leftrightarrow \tilde{\Phi}(x) = a_i + r_j.$$

We have that

$$\begin{aligned} H(\Phi'(X)+\mathfrak{I}|Y) &= \sum_{y\in\mathscr{Y}} p_y H\left(\sum_{i=1}^{q}\frac{p_{i,1,y}}{p_y}, \sum_{i=1}^{q}\frac{p_{i,2,y}}{p_y}, \cdots, \sum_{i=1}^{q}\frac{p_{i,q,y}}{p_y}\right) = A \\ &> \sum_{y\in\mathscr{Y}} p_y H\left(\frac{p_{1,y}}{p_y}, \frac{p_{2,y}}{p_y}, \cdots, \frac{p_{q,y}}{p_y}\right) = \max_{\Phi\in\mathscr{M}} H(\Phi(X)+\mathfrak{I}|Y). \end{aligned}$$

It is absurd that $H(\Phi'(X)+\mathfrak{I}|Y) > \max_{\Phi\in\mathscr{M}} H(\Phi(X)+\mathfrak{I}|Y)$! Therefore, (A.2) is valid by (A.3) and (A.4), so is (A.1). ■

Lemma A.2. *If both*

$$0 \leq \max\{p_2, p_3\} \not< \min\{p_1, p_4\} \leq 1 \text{ and } 0 \leq \max\{p_1, p_4\} \not< \min\{p_2, p_3\} \leq 1$$

are valid, and $\sum_{j=1}^{4} p_j = 1$, *then*

$$-\sum_{j=1}^{4} p_j \log p_j \leq -2\big[(p_2+p_3)\log(p_2+p_3) + (p_1+p_4)\log(p_1+p_4)\big]. \tag{A.5}$$

Proof [27]: Without loss of generality, we assume that $0 \leq \max\{p_4, p_3\} \leq \min\{p_2, p_1\} \leq 1$ which implies that $p_1 + p_2 - 1/2 \geq |p_1 + p_4 - 1/2|$. Let $H_2(c) = -c\log c - (1-c)\log(1-c)$, $0 \leq c \leq 1$, be the binary entropy function. By the grouping rule for entropy [16, pp. 49], (A.5) equals to

$$\begin{aligned} &(p_1+p_4)\left(\frac{p_1}{p_1+p_4}\log\frac{p_1+p_4}{p_1} + \frac{p_4}{p_1+p_4}\log\frac{p_1+p_4}{p_4}\right) \\ &+(p_2+p_3)\left(\frac{p_2}{p_2+p_3}\log\frac{p_2+p_3}{p_2} + \frac{p_3}{p_2+p_3}\log\frac{p_2+p_3}{p_3}\right) \\ &\leq -(p_2+p_3)\log(p_2+p_3) - (p_1+p_4)\log(p_1+p_4) \\ &\quad\Leftrightarrow \\ A =&(p_1+p_4)H_2\left(\frac{p_1}{p_1+p_4}\right) + (p_2+p_3)H_2\left(\frac{p_2}{p_2+p_3}\right) \\ \leq&H_2(p_1+p_4). \end{aligned}$$

Since H_2 is a concave function and $\sum_{j=1}^{4} p_j = 1$, then

$$A \leq H_2\left(p_1 + p_2\right).$$

Moreover, $p_1 + p_2 - 1/2 \geq |p_1 + p_4 - 1/2|$ guarantees that

$$H_2\left(p_1 + p_2\right) \leq H_2\left(p_1 + p_4\right),$$

because $H_2(c) = H_2(1-c), \forall\ 0 \leq c \leq 1$, and $H_2(c') \leq H_2(c'')$ if $0 \leq c' \leq c'' \leq 1/2$. Therefore, $A \leq H_2\left(p_1 + p_4\right)$ and (A.5) holds. ■

Lemma A.3. *No matter which finite field $\mathbb{F}_q$ is chosen, g given by* (24) *admits no presentation $k_1(x)+k_2(y)+k_3(z)$, where $k_i \in \mathbb{F}_q[1]$ for all feasible i.*

Proof: Suppose otherwise, i.e. $k_1 \circ \mu_1 + k_2 \circ \mu_2 + k_3 \circ \mu_3 = \nu \circ g$ for some injections $\mu_1, \mu_1, \mu_3 : \{\alpha_0, \alpha_1\} \to \mathbb{F}_q$ and $\nu : \{\beta_0, \cdots, \beta_3\} \to \mathbb{F}_q$. By (24), we have

$$\begin{aligned}
\nu(\beta_1) =& (k_1 \circ \mu_1)(\alpha_1) + (k_2 \circ \mu_2)(\alpha_0) + (k_3 \circ \mu_3)(\alpha_0) \\
=& (k_1 \circ \mu_1)(\alpha_0) + (k_2 \circ \mu_2)(\alpha_1) + (k_3 \circ \mu_3)(\alpha_1) \\
\nu(\beta_3) =& (k_1 \circ \mu_1)(\alpha_1) + (k_2 \circ \mu_2)(\alpha_1) + (k_3 \circ \mu_3)(\alpha_0) \\
=& (k_1 \circ \mu_1)(\alpha_0) + (k_2 \circ \mu_2)(\alpha_0) + (k_3 \circ \mu_3)(\alpha_1) \\
\Longrightarrow& \nu(\beta_1) - \nu(\beta_3) = \tau = -\tau \\
\Longrightarrow& \tau + \tau = 0,
\end{aligned} \tag{A.6}$$

where $\tau = k_2(\mu_2(\alpha_0)) - k_2(\mu_2(\alpha_1))$. Since μ_2 is injective, (A.6) implies that either $\tau = 0$ or Char$(\mathbb{F}_q) = 2$ by Proposition II.7. Noticeable that $k_2(\mu_2(\alpha_0)) \neq k_2(\mu_2(\alpha_1))$, i.e. $\tau \neq 0$, otherwise, $\nu(\beta_1) = \nu(\beta_3)$ which contradicts the assumption that ν is injective. Thus, Char$(\mathbb{F}_q) = 2$. Let $\rho = (k_3 \circ \mu_3)(\alpha_0) - (k_3 \circ \mu_3)(\alpha_1)$. Obviously, $\rho \neq 0$ because of the same reason that $\tau \neq 0$, and $\rho + \rho = 0$ since Char$(\mathbb{F}_q) = 2$. Therefore,

$$\begin{aligned}
\nu(\beta_0) =& (k_1 \circ \mu_1)(\alpha_0) + (k_2 \circ \mu_2)(\alpha_0) + (k_3 \circ \mu_3)(\alpha_0) \\
=& (k_1 \circ \mu_1)(\alpha_0) + (k_2 \circ \mu_2)(\alpha_0) + (k_3 \circ \mu_3)(\alpha_1) + \rho \\
=& \nu(\beta_3) + \rho \\
=& (k_1 \circ \mu_1)(\alpha_1) + (k_2 \circ \mu_2)(\alpha_1) + (k_3 \circ \mu_3)(\alpha_0) + \rho \\
=& (k_1 \circ \mu_1)(\alpha_1) + (k_2 \circ \mu_2)(\alpha_1) + (k_3 \circ \mu_3)(\alpha_1) + \rho + \rho \\
=& \nu(\beta_2) + 0 = \nu(\beta_2).
\end{aligned}$$

This contradicts the assumption that ν is injective. ■

Remark 24. As a special case, this lemma implies that no matter which finite field $\mathbb{F}_q$ is chosen, g defined by (24) has no polynomial presentation that is linear over $\mathbb{F}_q$. In contrast, g admits presentation $x+2y+3z \in \mathbb{Z}_4[3]$ which is a linear function over $\mathbb{Z}_4$.

APPENDIX B
PROOFS OF LEMMA VI.5 AND LEMMA VI.6

A. Proof of Lemma VI.5

Let p be a prime such that $p^m \geq \max\{|\Omega|, |\mathscr{X}_i| \,|1 \leq i \leq k\}$ for some integer m, and choose $\mathfrak{R}$ to be a finite field of order p^m. By [28, Lemma 7.40], the number of polynomial functions in $\mathfrak{R}[k]$ is $p^{mp^{mk}}$. Moreover, the number of distinct functions with domain $\mathfrak{R}^k$ and codomain $\mathfrak{R}$ is also $|\mathfrak{R}|^{|\mathfrak{R}^k|} = p^{mp^{mk}}$. Hence, any function $g' : \mathfrak{R}^k \to \mathfrak{R}$ is a polynomial function.

In the meanwhile, any injections $\mu_i : \mathscr{X}_i \to \mathfrak{R}$ $(1 \leq i \leq k)$ and $\nu : \Omega \to \mathfrak{R}$ give rise to a function

$$\hat{g} = \nu \circ g\left(\mu_1', \mu_2', \cdots, \mu_k'\right) : \mathfrak{R}^k \to \mathfrak{R},$$

where μ_i' is the inverse mapping of $\mu_i : \mathscr{X}_i \to \mu_i(\mathscr{X}_i)$. Since $\hat{g}$ must be a polynomial function as shown, the statement is established.

Remark 25. Another proof involving Fermat's little theorem can be found in [6].

B. Proof of Lemma VI.6

Let $\mathbb{F}$ be a finite field such that $|\mathbb{F}| \geq |\mathscr{X}_i|$ for all $1 \leq i \leq s$ and $|\mathbb{F}|^s \geq |\Omega|$, and let $\mathfrak{R}$ be the *splitting field* of $\mathbb{F}$ of order $|\mathbb{F}|^s$ (one example of the pair $\mathbb{F}$ and $\mathfrak{R}$ is the $\mathbb{Z}_p$, where p is some prime, and its *Galois extension* of *degree* s). It is easily seen that $\mathfrak{R}$ is an s dimensional vector space over $\mathbb{F}$. Hence, there exist s vectors $v_1, v_2, \cdots, v_s \in \mathfrak{R}$ that are linearly independent. Let μ_i be an injection from $\mathscr{X}_i$ to the subspace generated by vector v_i. It is easy to verify that $k = \sum_{i=1}^{s} \mu_i$ is injective since $v_1, v_2, \cdots, v_s$ are linearly independent. Let k' be the inverse mapping of $k : \prod_{i=1}^{s} \mathscr{X}_i \to k\left(\prod_{i=1}^{s} \mathscr{X}_i\right)$ and $\nu : \Omega \to \mathfrak{R}$ be any injection. By [28, Lemma 7.40], there exists a polynomial function $h \in \mathfrak{R}[s]$ such that $h = \nu \circ g \circ k'$. Let $\hat{g}(x_1, x_2, \cdots, x_s) = h\left(\sum_{i=1}^{s} x_i\right)$. The statement is proved.

Remark 26. In the proof, k is chosen to be injective because the proof includes the case that g is an identity function. In general, k is not necessarily injective.

APPENDIX C
THE SECOND PROOF OF LEMMA III.7

Define the mapping $\Gamma : \mathfrak{R}_1 \to \mathfrak{R}_1/\mathfrak{I}$ by

$$\Gamma : x_1 \mapsto x_1 + \mathfrak{I}, \forall\, x_1 \in \mathfrak{R}_1.$$

Assume that $\mathbf{x}_1 = \left[x_1^{(1)}, x_1^{(2)}, \cdots, x_1^{(n)}\right]$, and let

$$\overline{\mathbf{y}} = \left[\Gamma\left(x_1^{(1)}\right), \Gamma\left(x_1^{(2)}\right), \cdots, \Gamma\left(x_1^{(n)}\right)\right].$$

By definition, $\forall\ (\mathbf{y}, \mathbf{x}_2)^t \in D_\epsilon(\mathbf{x}_1, \mathfrak{I}|\mathbf{x}_2)$, where $\mathbf{y} = \left[y^{(1)}, y^{(2)}, \cdots, y^{(n)}\right]$,

$$\left[\Gamma\left(y^{(1)}\right), \Gamma\left(y^{(2)}\right), \cdots, \Gamma\left(y^{(n)}\right)\right] = \overline{\mathbf{y}}.$$

Moreover,

$$(\mathbf{y}, \overline{\mathbf{y}}, \mathbf{x}_2)^t \in \mathcal{T}_\epsilon(n, (X_1, Y_{\mathfrak{R}_1/\mathfrak{I}}, X_2)),\ \text{and}$$

$$|D_\epsilon(\mathbf{x}_1, \mathfrak{I}|\mathbf{x}_2)| = \left|\left\{(\mathbf{y}, \overline{\mathbf{y}}, \mathbf{x}_2)^t \in \mathcal{T}_\epsilon \middle| \mathbf{y} - \mathbf{x}_1 \in \mathfrak{I}^n\right\}\right|.$$

For fixed $(\overline{\mathbf{y}}, \mathbf{x}_2)^t \in \mathcal{T}_\epsilon$, the number of strongly ϵ-typical sequences $\mathbf{y}$ such that $(\mathbf{y}, \overline{\mathbf{y}}, \mathbf{x}_2)^t$ is strongly ϵ-typical is strictly upper bounded by $2^{n\left[H(X_1|Y_{\mathfrak{R}_1/\mathfrak{I}}, X_2)+\eta\right]}$ if n is larger enough and ϵ is small. Therefore,

$$|D_\epsilon(\mathbf{x}_1, \mathfrak{I}|\mathbf{x}_2)| < 2^{n\left[H(X_1|Y_{\mathfrak{R}_1/\mathfrak{I}}, X_2)+\eta\right]}.$$

Remark 27. The mechanisms behind the first proof and the second one are in fact very different. However, this is not very clear for i.i.d. scenarios. For non-i.i.d. scenarios, the results proved by these two approaches diverse. Although the technique from the first proof is more complicated, it provides results with its own advantages. Henceforth, we deliberately put the first proof in the first place. Interested readers are kindly referred to [29] for more details of the differences.

APPENDIX D
PROOF OF THEOREM VI.11

Choose $\delta > 6\epsilon > 0$, such that $R_j = R'_j + R''_j$, $\forall\ j \in \mathcal{S}$, $\sum_{j\in T} R'_j > I(Y_T; V_T|V_{T^c}) + 2\,|T|\,\delta$, $\forall\ \emptyset \neq T \subseteq \mathcal{S}$, and $R''_j > r + 2\delta$, where $r = \max_{0\neq\mathfrak{I}\leq_l\mathfrak{R}} \frac{\log|\mathfrak{R}|}{\log|\mathfrak{I}|}\left[H(X|V_\mathcal{S}) - H(Y_{\mathfrak{R}/\mathfrak{I}}|V_\mathcal{S})\right]$, $\forall\ j \in \mathcal{S}\setminus\mathcal{S}_0$.

A. Encoding:

Fix the joint distribution p which satisfies (30). For all $j \in \mathcal{S}_0$, let $\mathscr{V}_{j,\epsilon} = \mathcal{T}_\epsilon(n, X_j)$. For all $j \in \mathcal{S}\setminus\mathcal{S}_0$, generate randomly $2^{n[I(Y_j;V_j)+\delta]}$ strongly ϵ-typical sequences according to distribution $p_{V_j^n}$ and let $\mathscr{V}_{j,\epsilon}$ be the set of these generated sequences. Define mapping $\phi'_j : \mathfrak{R}^n \to \mathscr{V}_{j,\epsilon}$ as follows:

1) If $j \in \mathcal{S}_0$, then, $\forall\ \mathbf{x} \in \mathfrak{R}^n$, $\phi'_j(\mathbf{x}) = \begin{cases}\mathbf{x}, & \text{if } \mathbf{x} \in \mathcal{T}_\epsilon;\\ \mathbf{x}_0, & \text{otherwise,}\end{cases}$ where $\mathbf{x}_0 \in \mathscr{V}_{j,\epsilon}$ is fixed.
2) If $j \in \mathcal{S}\setminus\mathcal{S}_0$, then for every $\mathbf{x} \in \mathfrak{R}^n$, let $Ł_\mathbf{x} = \{\mathbf{v} \in \mathscr{V}_{j,\epsilon} | (\vec{k}_j(\mathbf{x}), \mathbf{v}) \in \mathcal{T}_\epsilon\}$. If $\mathbf{x} \in \mathcal{T}_\epsilon$ and $Ł_\mathbf{x} \neq \emptyset$, then $\phi'_j(\mathbf{x})$ is set to be some element in $Ł_\mathbf{x}$; otherwise $\phi'_j(\mathbf{x})$ is some fixed $\mathbf{v}_0 \in \mathscr{V}_{j,\epsilon}$.

Define mapping $\eta_j : \mathscr{V}_{j,\epsilon} \to [1, 2^{nR'_j}]$ by randomly choosing the value for each $\mathbf{v} \in \mathscr{V}_{j,\epsilon}$ according to a uniform distribution.

Let $k = \min\limits_{j\in\mathcal{S}\setminus\mathcal{S}_0}\left\{\left\lfloor\frac{nR''_j}{\log|\mathfrak{R}|}\right\rfloor\right\}$. When n is big enough, we have $k > \frac{n[r+\delta]}{\log|\mathfrak{R}|}$. Randomly generate a $k\times n$ matrix $\mathbf{M}\in\mathfrak{R}^{k\times n}$, and let $\theta_j:\mathfrak{R}^n\to\mathfrak{R}^k$ ($j\in\mathcal{S}\setminus\mathcal{S}_0$) be the function $\theta_j:\mathbf{x}\mapsto\mathbf{M}\vec{k}_j(\mathbf{x}),\forall\,\mathbf{x}\in\mathfrak{R}^n$.

Define the encoder ϕ_j as the follows

$$\phi_j=\begin{cases}\eta_j\circ\phi'_j, & j\in\mathcal{S}_0;\\ (\eta_j\circ\phi'_j,\theta_j), & \text{otherwise.}\end{cases}$$

B. Decoding:

Upon observing $(a_1,a_2,\cdots,a_{s_0},(a_{s_0+1},b_{s_0+1}),\cdots,(a_s,b_s))$ at the decoder, the decoder claims that

$$\vec{h}\left[\vec{k}_0\left(\hat{V}_1^n,\hat{V}_2^n,\cdots,\hat{V}_{s_0}^n\right),\hat{X}^n\right]$$

is the function of the generated data, if and only if there exists one and only one

$$\hat{\mathbf{V}}=\left(\hat{V}_1^n,\hat{V}_2^n,\cdots,\hat{V}_s^n\right)\in\prod_{j=1}^{s}\mathscr{V}_{j,\epsilon},$$

such that $a_j=\eta_j(\hat{V}_j^n),\forall\,j\in\mathcal{S}$, and $\hat{X}^n$ is the only element in the set

$$\mathscr{L}_{\hat{\mathbf{V}}}=\left\{\mathbf{x}\in\mathfrak{R}^n\middle|(\mathbf{x},\hat{\mathbf{V}})\in\mathcal{T}_\epsilon,\mathbf{M}\mathbf{x}=\sum_{j=t+1}^{s}b_j\right\}.$$

C. Error:

Assume that X_j^n is the data generated by the jth source and let $X^n=\sum\limits_{j=s_0+1}^{s}\vec{k}_j\left(X_j^n\right)$. An error happens if and only if one of the following events happens.

E_1: $(X_1^n,X_2^n,\cdots,X_s^n,Y_1^n,Y_2^n,\cdots,Y_s^n,X^n)\notin\mathcal{T}_\epsilon$;

E_2: There exists some $j_0\in\mathcal{S}\setminus\mathcal{S}_0$, such that $Ł_{X_{j_0}^n}=\emptyset$;

E_3: $(Y_1^n,Y_2^n,\cdots,Y_s^n,X^n,\mathbf{V})\notin\mathcal{T}_\epsilon$, where $\mathbf{V}=(V_1^n,V_2^n,\cdots,V_s^n)$ and $V_j^n=\phi'_j(X_j^n),\forall\,j\in\mathcal{S}$;

E_4: There exists $\mathbf{V}'=(\mathbf{v}'_1,\mathbf{v}'_2,\cdots,\mathbf{v}'_s)\in\mathcal{T}_\epsilon\cap\prod\limits_{j=1}^{s}\mathscr{V}_{j,\epsilon}$, $\mathbf{V}'\neq\mathbf{V}$, such that $\eta_j(\mathbf{v}'_j)=\eta_j\left(V_j^n\right),\forall\,j\in\mathcal{S}$;

E_5: $X^n\notin\mathscr{L}_{\mathbf{V}}$ or $|\mathscr{L}_{\mathbf{V}}|>1$, i.e. there exists $X_0^n\in\mathfrak{R}^n$, $X_0^n\neq X^n$, such that $\mathbf{M}X_0^n=\mathbf{M}X^n$ and $(X_0^n,\mathbf{V})\in\mathcal{T}_\epsilon$.

Let $\gamma=\Pr\left\{\bigcup\limits_{l=1}^{5}E_l\right\}=\sum\limits_{l=1}^{5}\Pr\left\{E_l|E_{l,c}\right\}$, where $E_{1,c}=\emptyset$ and $E_{l,c}=\bigcap\limits_{\tau=1}^{l-1}E_\tau^c$ for $1<l\leq 5$. In the following, we show that $\gamma\to 0$, $n\to\infty$.

(a). By the joint AEP [15, Theorem 6.9], $\Pr\{E_1\}\to 0$, $n\to\infty$.

(b). Let $E_{2,j}=\left\{Ł_{X_j^n}=\emptyset\right\}$, $\forall j\in\mathcal{S}\setminus\mathcal{S}_0$. Then

$$\Pr\{E_2|E_{2,c}\}\leq\sum_{j\in\mathcal{S}\setminus\mathcal{S}_0}\Pr\left\{E_{2,j}|E_{2,c}\right\}. \tag{D.1}$$

For any $j\in\mathcal{S}\setminus\mathcal{S}_0$, because the sequence $\mathbf{v}\in\mathscr{V}_{j,\epsilon}$ and $Y_j^n=\vec{k}_j(X_j^n)$ are drawn independently, we have

$$\begin{aligned}\Pr\{(Y_j^n,\mathbf{v})\in\mathcal{T}_\epsilon\}\geq&(1-\epsilon)2^{-n[I(Y_j;V_j)+3\epsilon]}\\ =&(1-\epsilon)2^{-n[I(Y_j;V_j)+\delta/2]+n(\delta/2-3\epsilon)}\\ >&2^{-n[I(Y_j;V_j)+\delta/2]}\end{aligned}$$

when n is big enough. Thus,

$$\begin{aligned}\Pr\{E_{2,j}|E_{2,c}\} &= \Pr\left\{\text{Ł}_{X_j^n} = \emptyset \mid E_{2,c}\right\} \\ &= \prod_{\mathbf{v}\in\mathscr{V}_{j,\epsilon}} \Pr\left\{\left(\vec{k}_j(X_j^n), \mathbf{v}\right) \notin \mathcal{T}_\epsilon\right\} \\ &< \left\{1 - 2^{-n[I(Y_j;V_j)+\delta/2]}\right\}^{2^{n[I(Y_j;V_j)+\delta]}} \qquad \text{(D.2)} \\ &\to 0, n \to \infty.\end{aligned}$$

where (D.2) holds true for all big enough n and the limit follow from the fact that $(1-1/a)^a \to e^{-1}, a \to \infty$. Therefore, $\Pr\{E_2|E_{2,c}\} \to 0$, $n \to \infty$ by (D.1).

(c). By (30), it is obvious that $V_{J_1} - Y_{J_1} - Y_{J_2} - V_{J_2}$ forms a Markov chain for any two disjoint nonempty sets $J_1, J_2 \subsetneq \mathcal{S}$. Thus, if $(Y_j^n, V_j^n) \in \mathcal{T}_\epsilon$ for all $j \in \mathcal{S}$ and $(Y_1^n, Y_2^n, \cdots, Y_s^n) \in \mathcal{T}_\epsilon$, then $(Y_1^n, Y_2^n, \cdots, Y_s^n, \mathbf{V}) \in \mathcal{T}_\epsilon$. In the meantime, $X-(Y_1, Y_2, \cdots, Y_s)-(V_1, V_2, \cdots, V_s)$ is also a Markov chain. Hence, $(Y_1^n, Y_2^n, \cdots, Y_s^n, X^n, \mathbf{V}) \in \mathcal{T}_\epsilon$ if $(Y_1^n, Y_2^n, \cdots, Y_s^n, X^n) \in \mathcal{T}_\epsilon$. Therefore, $\Pr\{E_3|E_{3,c}\} = 0$.

(d). For all $\emptyset \neq J \subseteq \mathcal{S}$, let $J = \{j_1, j_2, \cdots, j_{|j|}\}$ and

$$\Gamma_J = \left\{\mathbf{V}' = (\mathbf{v}_1', \mathbf{v}_2', \cdots, \mathbf{v}_s') \in \prod_{j=1}^s \mathscr{V}_{j,\epsilon} \middle| \mathbf{v}_j' = V_j^n \text{ if and only if } j \in \mathcal{S}\setminus J\right\}.$$

By definition, $|\Gamma_J| = \prod_{j\in J} |\mathscr{V}_{j,\epsilon}| - 1 = 2^{n\left[\sum_{j\in J} I(Y_j;V_j)+|J|\delta\right]} - 1$ and

$$\begin{aligned}\Pr\{E_4|E_{4,c}\} &= \sum_{\emptyset\neq J\subseteq\mathcal{S}} \sum_{\mathbf{V}'\in\Gamma_J} \Pr\left\{\eta_j(\mathbf{v}_j') = \eta_j(V_j^n), \forall\, j \in J, \mathbf{V}' \in \mathcal{T}_\epsilon|E_{4,c}\right\} \\ &= \sum_{\emptyset\neq J\subseteq\mathcal{S}} \sum_{\mathbf{V}'\in\Gamma_J} \Pr\left\{\eta_j(\mathbf{v}_j') = \eta_j(V_j^n), \forall\, j \in J\right\} \times \Pr\left\{\mathbf{V}' \in \mathcal{T}_\epsilon|E_{4,c}\right\} \qquad \text{(D.3)} \\ &< \sum_{\emptyset\neq J\subseteq\mathcal{S}} \sum_{\mathbf{V}'\in\Gamma_J} 2^{-n\sum_{j\in J} R_j'} \times 2^{-n\left[\sum_{i=1}^{|J|} I(V_{j_i};V_{J^c},V_{j_1},\cdots,V_{j_{i-1}})-|J|\delta\right]} \qquad \text{(D.4)} \\ &< \sum_{\emptyset\neq J\subseteq\mathcal{S}} 2^{n\left[\sum_{j\in J} I(Y_j;V_j)+|j|\delta\right]} \times 2^{-n\sum_{j\in J} R_j'} \times 2^{-n\left[\sum_{i=1}^{|j|} I(V_{j_i};V_{J^c},V_{j_1},\cdots,V_{j_{i-1}})-|j|\delta\right]} \\ &\leq C \max_{\emptyset\neq J\subseteq\mathcal{N}} 2^{-n\left[\sum_{j\in J} R_j' - I(Y_J;V_J|V_{J^c})-2|j|\delta\right]} \qquad \text{(D.5)} \\ &\to 0, n \to \infty,\end{aligned}$$

where $C = 2^s - 1$. Equality (D.3) holds because the processes of choosing η_j's and generating $\mathbf{V}'$ are done independently. (D.4) follows from Lemma D.1 and the definitions of η_j's. (D.5) is from Lemma D.2.

Lemma D.1. *Let $[X_1, X_2, \cdots, X_l, Y] \sim q$. For any $\epsilon > 0$ and positive integer n, choose a sequence $\tilde{X}_j^n$ $(1 \leq j \leq l)$ randomly from $\mathcal{T}_\epsilon(n, X_j)$ based on a uniform distribution. If $\mathbf{y} \in \mathscr{Y}^n$ is an ϵ-typical sequence with respect to Y, then*

$$\Pr\left\{(\tilde{X}_1^n, \tilde{X}_2^n, \cdots, \tilde{X}_l^n, Y^n) \in \mathcal{T}_\epsilon|Y^n = \mathbf{y}\right\} \leq 2^{-n\left[\sum_{j=1}^l I(X_j;Y,X_1,X_2,\cdots,X_{j-1})-3l\epsilon\right]}.$$

Proof: Let F_j be the event $\{(\tilde{X}_1^n, \tilde{X}_2^n, \cdots, \tilde{X}_j^n, Y^n) \in \mathcal{T}_\epsilon\}$, $1 \le j \le l$, and $F_0 = \emptyset$. We have

$$\begin{aligned}\Pr\left\{(\tilde{X}_1^n, \tilde{X}_2^n, \cdots, \tilde{X}_l^n, Y^n) \in \mathcal{T}_\epsilon | Y^n = \mathbf{y}\right\} &= \prod_{j=1}^{l} \Pr\{F_j | Y^n = \mathbf{y}, F_{j-1}\} \\ &\le \prod_{j=1}^{l} 2^{-n[I(X_j;Y,X_1,X_2,\cdots,X_{j-1})-3\epsilon]} \\ &= 2^{-n\left[\sum_{j=1}^{l} I(X_j;Y,X_1,X_2,\cdots,X_{j-1})-3l\epsilon\right]},\end{aligned}$$

since $\tilde{X}_1^n, \tilde{X}_2^n, \cdots, \tilde{X}_l^n, \mathbf{y}$ are generated independent. ■

Lemma D.2. *If* $(Y_1, V_1, Y_2, V_2, \cdots, Y_s, V_s) \sim q$, *and*

$$q(y_1, v_1, y_2, v_2, \cdots, y_s, v_s) = q(y_1, y_2, \cdots, y_s) \prod_{i=1}^{s} q(v_i|y_i),$$

then, $\forall\ J = \{j_1, j_2, \cdots, j_{|j|}\} \subseteq \{1, 2, \cdots, s\}$,

$$I(Y_J; V_J | V_{J^c}) = \sum_{i=1}^{|j|} I(Y_{j_i}; V_{j_i}) - I(V_{j_i}; V_{J^c}, V_{j_1}, \cdots, V_{j_{i-1}}).$$

(e). Let $E_{5,1} = \{\mathscr{L}_{\mathbf{V}} = \emptyset\}$ and $E_{5,2} = \{|\mathscr{L}_{\mathbf{V}}| > 1\}$. We have $\Pr\{E_{5,1}|E_{5,c}\} = 0$, because $E_{5,c}$ contains the event that $(X^n, \mathbf{V}) \in \mathscr{L}_{\mathbf{V}}$ and $\mathbf{V}$ is unique. Therefore,

$$\begin{aligned}\Pr\{E_5|E_{5,c}\} &= \Pr\{E_{5,2}|E_{5,c}\} \\ &= \sum_{(X_0^n, \mathbf{V}) \in \mathcal{T}_\epsilon \setminus (X^n, \mathbf{V})} \Pr\{\mathbf{M}X_0^n = \mathbf{M}X^n\} \\ &< \sum_{0 \ne \mathfrak{I} \le_l \mathfrak{R}} \ \sum_{D_\epsilon(X^n, \mathfrak{I}|\mathbf{V}) \setminus (X^n, \mathbf{V})} \Pr\{\mathbf{M}X_0^n = \mathbf{M}X^n\}\end{aligned}$$

Choose a small $\eta > 0$ such that $\eta < \dfrac{\delta}{2\log|\mathfrak{R}|}$. Then

$$\begin{aligned}\Pr\{E_5|E_{5,c}\} &< \left(2^{|\mathfrak{R}|} - 2\right) \max_{0 \ne \mathfrak{I} \le_l \mathfrak{R}} 2^{n\left[H(X|V_{\mathcal{S}}) - H(Y_{\mathfrak{R}/\mathfrak{I}}|V_{\mathcal{S}}) + \eta\right]} \times 2^{-k\log|\mathfrak{I}|} \qquad \text{(D.6)} \\ &= \left(2^{|\mathfrak{R}|} - 2\right) \max_{0 \ne \mathfrak{I} \le_l \mathfrak{R}} 2^{-n\left[k\log|\mathfrak{I}|/n - H(X|V_{\mathcal{S}}) + H(Y_{\mathfrak{R}/\mathfrak{I}}|V_{\mathcal{S}}) - \eta\right]} \\ &< \left(2^{|\mathfrak{R}|} - 2\right) \max_{0 \ne \mathfrak{I} \le_l \mathfrak{R}} 2^{-n[\delta\log|\mathfrak{I}|/\log|\mathfrak{R}| - \eta]} \\ &< \left(2^{|\mathfrak{R}|} - 2\right) 2^{-n\delta/2\log|\mathfrak{R}|} \qquad \text{(D.7)} \\ &\to 0, n \to \infty,\end{aligned}$$

where (D.6) is from Lemma III.5 and Lemma III.7 (for all large enough n and small enough ϵ) and (D.7) is because $|\mathfrak{I}| \ge 2$ for all $\mathfrak{I} \ne 0$.

To summarize, by (a)–(e), we have $\gamma \to 0, n \to \infty$. The theorem is established.

APPENDIX E
CODING OVER ABELIAN GROUPS

Most of the coding literature has focused on coding over fields. Some both traditional and recent work, including [30], has also considered (Abelian) groups, while significantly fewer results are available for coding over rings. In this appendix we elaborate on the relation between coding over rings and groups in order to clearly show that our results in this paper are not subsumed by previous work on coding over groups. In fact previous work, e.g. [30], on "linear coding over finite Abelian groups" does not even include linear coding over finite fields as a special case.

(R1) Consider the example given in [30, Section VIII.B.1)] for reconstruction of the modulo-two sum of *binary symmetric sources* [3]. On [30, pp. 1509], it reads "Rate points achieved by embedding the function in the Abelian groups $\mathbb{Z}_3$, $\mathbb{Z}_4$ are *strictly worse* than that achieved by embedding the function in $\mathbb{Z}_2$ while embedding in $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ gives the Slepian–Wolf rate region for the *lossless* reconstruction of (X, Y)" [5].

[30] clearly states that group coding over $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ for encoding the modulo-two sum of symmetric sources gives only the Slepian–Wolf region. On the contrary, consider either the finite field $\mathbb{F}_4$ or the non-field ring

$$\mathbb{M}_{L,2} = \left\{ \left[\begin{matrix} a & 0 \\ b & a \end{matrix} \right] \middle| \, a, b \in \mathbb{Z}_2 \right\}$$

(note: the underlying Abelian group defined $\mathbb{F}_4$ and $\mathbb{M}_{L,2}$ is $\mathbb{Z}_2 \oplus \mathbb{Z}_2$). We claim that linear coding over either $\mathbb{F}_4$ or $\mathbb{M}_{L,2}$ for encoding the modulo-two sum of symmetric sources gives the Körner–Marton region [3]. This is because linear coding over finite field, e.g. $\mathbb{F}_4$, is always optimal for the Slepian–Wolf problem, so is linear coding over non-field ring $\mathbb{M}_{L,2}$ by Theorem V.3. Unfortunately, group coding over $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ is not.

It is well-known that the Körner–Marton region is often strictly larger than the Slepian–Wolf region. Linear coding over non-field ring $\mathbb{M}_{L,2}$ (field $\mathbb{F}_4$) as a special case "linear coding over Abelian group $\mathbb{Z}_2 \oplus \mathbb{Z}_2$" must not achieve a region larger than the Slepian–Wolf region, leading to a contradiction.

(R2) [30, row 2 of TABLE III] states that group coding over $\mathbb{Z}_4 \oplus \mathbb{Z}_4$ (achieving sum rate 3.5) is strictly worse than over group $\mathbb{Z}_4$ (achieving sum rate 3) for lossless encoding a quaternary function [30, Section VIII.A]. On the contrary, linear coding over ring $\mathbb{Z}_4 \times \mathbb{Z}_4$ (with underlying Abelian group $\mathbb{Z}_4 \oplus \mathbb{Z}_4$) always achieves region containing the one achieved by linear coding over ring $\mathbb{Z}_4$. This is implied by Theorem III.3. By direct calculation, we have that linear coding over ring $\mathbb{Z}_4 \times \mathbb{Z}_4$ (achieving sum rate 3) is strictly better than "linear coding over Abelian group $\mathbb{Z}_4 \oplus \mathbb{Z}_4$" (achieving sum rate 3.5). Again, a contradiction.

(R3) Linearity is often defined with regard to *linear combination*, which requires corresponding definitions of addition and multiplication. It is basic and well-known that over Abelian group $G = \mathbb{Z}_p \oplus \mathbb{Z}_p \oplus \mathbb{Z}_p \oplus \mathbb{Z}_p$ (p is a prime), there are at least three distinct definitions of multiplication to define rings over G. These rings are isomorphic to either

a) the field $\mathbb{F}_{p^4}$ which is commutative; or

[5]There is a typo at the end of the last sentence. (X, Y) should be $F(X, Y) = X \oplus_2 Y$ from the context, because coding over $\mathbb{Z}_3$ is not strictly worse than coding over $\mathbb{Z}_2$ for lossless reconstruct the original data (X, Y) [2].

b) the non-field ring

$$\mathbb{M}_p = \left\{ \left. \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right| a, b, c, d \in \mathbb{Z}_p \right\}$$

which is not commutative; or

c) the product ring $\mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p$ which is commutative.

Suppose "linear coding over Abelian group G" is defined with respect to some multiplicative operation "$*$", at the same time, this linear scheme over G includes the three distinct linear coding schemes defined over $\mathbb{F}_{p^4}$, $\mathbb{M}_p$ and $\mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p$ simultaneously. We then conclude that this newly defined multiplicative operation "$*$" is commutative and non-commutative at the same time. As a conclusion, "linear coding over Abelian group" does not make sense as a concept.

(R4) Finally, we emphasize that according to the Fundamental Theorem of (Finite) Abelian Group [9, Theorem 5.25], up to isomorphism, every finite Abelian group is a *direct sum* of cyclic groups of prime-power order [9, Proposition 5.27]. This implies that every finite Abelian group is can be represented via direct sum of modulo integers. However, many finite rings are not (isomorphic to) direct product of modulo integers, e.g. finite fields $\mathbb{F}_q$ (when q is a power of a prime but is not a prime), matrix rings $\mathbb{M}_{L,q'}$ (when $q' \geq 2$ is any positive integer) and all non-commutative rings. For a fixed order (e.g. p^2 with p being a prime), the number of finite rings is often significantly bigger than the number of finite Abelian groups. For instance, there are 4 rings of order 4 while there are 2 groups of order 4.

ACKNOWLEDGMENT

The authors would like to thank their colleagues Jinfeng Du and Mattias Andersson for assistance in proving Lemma A.2. They are also very grateful to an anonymous reviewer of the paper [17] for suggesting an alternative proof of Lemma III.7.

REFERENCES

[1] P. Elias, "Coding for noisy channels," *IRE Conv. Rec.*, pp. 37–46, Mar. 1955.

[2] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Transactions on Information Theory*, vol. 28, no. 4, pp. 585–592, Jul. 1982.

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

[4] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 396–411, May 1983.

[5] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[6] S. Huang and M. Skoglund, "Polynomials and computing functions of correlated sources," in *IEEE International Symposium on Information Theory*, Jul. 2012, pp. 771–775.

[7] ——, "Computing polynomial functions of correlated sources: Inner bounds," in *International Symposium on Information Theory and its Applications*, Oct. 2012, pp. 160–164.

[8] T. S. Han and K. Kobayashi, "A dichotomy of functions f(x, y) of correlated sources (x, y) from the viewpoint of the achievable rate region," *IEEE Transactions on Information Theory*, vol. 33, no. 1, pp. 69–76, Jan. 1987.

[9] J. J. Rotman, *Advanced Modern Algebra*, 2nd ed. American Mathematical Society, Aug. 2010.

[10] G. Mullen and H. Stevens, "Polynomial functions (mod m)," *Acta Mathematica Hungarica*, vol. 44, no. 3-4, pp. 237–241, Sep. 1984. [Online]. Available: http://link.springer.com/article/10.1007/BF01950276

[11] T. W. Hungerford, *Algebra (Graduate Texts in Mathematics)*. Springer, Dec. 1980.

[12] T.-Y. Lam, *A First Course in Noncommutative Rings*, 2nd ed. Springer, Jun. 2001.

[13] D. S. Dummit and R. M. Foote, *Abstract Algebra*, 3rd ed. Wiley, 2003.

[14] F. W. Anderson and K. R. Fuller, *Rings and Categories of Modules*, 2nd ed. Springer-Verlag, 1992.

[15] R. W. Yeung, *Information Theory and Network Coding*, 1st ed. Springer Publishing Company, Incorporated, Sep. 2008.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, Jul. 2006.

[17] S. Huang and M. Skoglund, "On achievability of linear source coding over finite rings," in *IEEE International Symposium on Information Theory*, July 2013.

[18] R. C. Buck, "Nomographic functions are nowhere dense," *Proceedings of the American Mathematical Society*, vol. 85, no. 2, pp. 195–199, Jun. 1982. [Online]. Available: http://www.jstor.org/stable/2044280

[19] S. Huang and M. Skoglund, "Linear source coding over rings and applications," in *IEEE Swedish Communication Technologies Workshop*, Oct. 2012, pp. 1–6.

[20] D. Slepian, "Group codes for the gaussian channel," *The Bell System Technical Journal*, 1968.

[21] R. Ahlswede, "Group codes do not achieve shannon's channel capacity for general discrete channels," *The Annals of Mathematical Statistics*, vol. 42, no. 1, pp. 224–240, Feb. 1971.

[22] G. D. Forney, Jr., "On the hamming distance properties of group codes," *IEEE Transactions on Information Theory*, vol. 38, no. 6, pp. 1797–1801, Nov. 1992.

[23] D. Singmaster and D. M. Bloom, "Rings of order four," *Mathematical Association of America*, vol. 71, no. 8, pp. 918–920, Oct. 1964.

[24] Wikimedia Foundation, Inc. (2012, Jul.) Finite ring. [Online]. Available: http://en.wikipedia.org/wiki/Finite_ring

[25] T. H. Chan and A. Grant, "Entropy vector and network codes," in *IEEE International Symposium on Information Theory*, Jun. 2007.

[26] J. C. Interlando, R. Palazzo, Jr., and M. Elia, "Group block codes over nonabelian groups are asymptotically bad," *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1277–1280, Jul. 1996.

[27] J. Du and M. Andersson, Private Communication, May 2012.

[28] R. Lidl and H. Niederreiter, *Finite Fields*, 2nd ed. New York: Gambridge University Press, 1997.

[29] S. Huang and M. Skoglund, *Coding for Computing Irreducible Markovian Functions of Sources with Memory*, KTH Royal Institute of Technology, May 2013. [Online]. Available: http://www.ee.kth.se/∼sheng11

[30] D. Krithivasan and S. Pradhan, "Distributed source coding using Abelian group codes: A new achievable rate-distortion region," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1495–1519, 2011.