On Achievability of Linear Source Coding over Finite Rings

Sheng Huang, Mikael Skoglund
Communication Theory Lab, School of Electrical Engineering
KTH Royal Institute of Technology
Stockholm, Sweden
Email: sheng.huang@ee.kth.se, skoglund@ee.kth.se

***Abstract*—We propose using linear mappings over finite rings as encoders in the Slepian–Wolf and the source coding for computing problems. It is known that the arithmetic of many finite rings is substantially easier to implement than the one of finite fields. Hence, one of the advantages of using linear mappings over rings, instead of its field counterparts, is reducing implementation complexity. More importantly, the ring version dominates the field version in terms of achieving strictly better coding rates with strictly smaller alphabet size in the source coding for computing problem [1].**

This paper is dedicated to proving an achievability theorem of linear source coding over finite rings in the Slepian–Wolf problem. This result includes those given by Elias [2] and Csiszár [3] saying that linear coding over finite fields is optimal, i.e. achieves the Slepian–Wolf region. Although the optimality issue remains open, it has been verified in various scenarios including particularly many cases use non-field rings [1], [4].

I. Introduction

Linear coding over finite fields (LCoF) is shown to be optimal in the Slepian–Wolf (SW) source coding problem [5], [2], [3]. In exact terms, every rate tuple in the SW region

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

where $\mathcal{S} = \{1, 2, \cdots, s\}$, $T^c = \mathcal{S} \setminus T$ and X_T is the random variable array $\prod_{j\in T} X_j$, is achievable by using linear mappings over finite fields as encoders. In addition, its linear structure plays a very important role in the source coding for computing problem (Problem 1), where the decoder is interested in recovering a (discrete) function of the original data. By exploring the linear property of the linear encoders, an achievable region, containing and potentially larger than the SW region, is obtained for computing an arbitrary function. Related work include [6], [7], [8], [9].

However, there are some drawbacks of using linear mappings over finite fields as encoders:

1) The implementation of finite field arithmetic is complicated (compared to the implementation of arithmetic of modulo integer rings for instance), since it involves implementing the polynomial long division algorithm;
2) Alphabet sizes of the encoders are usually larger than required, because the size of a finite field must be a power of a prime;
3) Most importantly, it is not optimal for the computing problem in general [1].

In order to overcome these weaknesses, we propose linear coding over finite rings (LCoR), namely using linear mappings over finite rings (Definition II.5) as encoders. Making use of LCoR, issues listed above are compromised to a certain extent:

1) Arithmetic of modulo integer rings is substantially simpler to implement;
2) There exists a finite ring of any size, e.g. modulo integer ring $\mathbb{Z}_q$ is of size q for any integer $q \geq 2$;
3) LCoR dominates LCoF in terms of achieving strictly larger achievable region with strictly smaller alphabet size in the computing problem [1].

This paper focuses on establishing an achievability theorem (Theorem IV.1) of LCoR for the SW source coding scenario. From this theorem, it is easily seen that LCoR is optimal if all the rings used are fields (Theorem IV.2). On this regard, Theorem IV.1 includes those results of Elias [2] and Csiszár [3]. Unfortunately, there is no general conclusive result (neither confirm nor deny) regarding optimality of LCoR in the SW problem yet. Nevertheless, it is demonstrated in [1] that LCoR is equally optimal in many scenarios of SW. Especially, [4] proves that, for any arbitrary scenario of SW, there always exist non-field rings over which linear coding is optimal. Therefore, it seems rather plausible that LCoR is as optimal as LCoF. Example IV.3 is given as a demonstration.

As mentioned, one of the major purposes of studying LCoR is to introduce improved encoding techniques for the computing problem (Problem 1). Based on our achievability result, it is demonstrated that LCoR outperforms LCoF in terms of achieving strictly larger achievable region with strictly smaller alphabet size in the computing problem. Because of the space limitation, this part of the discussion is omitted. Interested readers are kindly referred to [1, Section VI].

II. Preliminaries

This section introduces some fundamental concepts and notation. Readers who are already familiar with this material may still choose to go through quickly to identify our notation.

A. Source Coding for Computing

As a generalization of the SW problem, the computing problem considers the following scenario.

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

978-1-4799-0446-4/13/$31.00 ©2013 IEEE

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 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 $\mathcal{R} \subseteq \mathcal{R}[g]$.

Obviously, if g is an *identity function*, then Problem 1 equals to the SW problem. We mainly consider this special case in this paper. In particular, we focus on using linear mappings over finite rings as encoders.

B. Rings, Ideals and Linear Mappings

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

1) $[\mathfrak{R}, +]$ is an *Abelian group*;
2) There exists a *multiplicative identity* $1 \in \mathfrak{R}$, 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 operator "$\cdot$" is usually written by juxtaposition, ab for $a \cdot b$, for all $a, b \in \mathfrak{R}$. The *identity* of the group $[\mathfrak{R}, +]$, denoted by 0, is called the *zero*. An element a of a ring $\mathfrak{R}$ is said to be *invertible*, if and only if there exists $b \in \mathfrak{R}$, such that $ab = ba = 1$. An invertible element of a ring is called a *unit*.

Proposition II.2 ([1])**.** *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 $\mathfrak{R}_T = \prod_{i \in T} \mathfrak{R}_i$ forms a new ring $[\mathfrak{R}_T, +, \cdot]$ with respect to the component-wise operations.*

Definition II.3. 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 $\forall\, r \in \mathfrak{R}$, $r \cdot x \in \mathfrak{I}$.

If condition 2) is replaced by

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

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.

Remark 1. For a left (right) ideal $\mathfrak{I}$ of ring $\mathfrak{R}$, $\mathfrak{R}/\mathfrak{I}$ designates the *quotient group* $\mathfrak{R} \mod \mathfrak{I}$ whose elements are *cosets* of $\mathfrak{I}$.

Remark 2. 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 $\langle a_1, a_2, \cdots, a_n \rangle_r = \left\{\sum_{i=1}^n a_i r_i \middle|\, r_i \in \mathfrak{R}, \forall\, 1 \leq i \leq n\right\}$ is a right ideal. Furthermore, $\langle a_1, a_2, \cdots, a_n \rangle_r = \mathfrak{R}$ if a_i is a unit for some $1 \leq i \leq n$.

Proposition II.4 ([1])**.** *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 $\mathfrak{A}_i \leq_r \mathfrak{R}_i$), $\forall\, 1 \leq i \leq s$.*

Remark 3. For any $\emptyset \neq T \subseteq \mathcal{S}$, Proposition II.4 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.5. A mapping $f : \mathfrak{R}^n \to \mathfrak{R}^m$ given as:

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

where $a_{i,j} \in \mathfrak{R}$ for all feasible i and j, is called a *left* (*right*) *linear mapping* over ring $\mathfrak{R}$. If $m = 1$, then f is called a *left* (*right*) *linear function* over $\mathfrak{R}$. The matrix $\mathbf{A} = [a_{i,j}]$ is called the *coefficient matrix* of f.

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.

III. Some Properties of Random Linear Mappings

Several important lemmata regarding linear encoders over finite rings are proved in this section.

Lemma III.1. *For a finite ring $\mathfrak{R}$ and a linear function*

$$f : \mathbf{x} \mapsto [a_1, a_2, \cdots, a_n]\mathbf{x}, \quad \forall\, \mathbf{x} \in \mathfrak{R}^n,$$

we have $\dfrac{|\mathfrak{S}(f)|}{|\mathfrak{R}|^n} = \dfrac{1}{|\mathfrak{I}|}$, where $\mathfrak{S}(f) = \{\mathbf{x} \in \mathfrak{R}^n | f(\mathbf{x}) = 0\}$ and $\mathfrak{I} = \langle a_1, a_2, \cdots, a_n \rangle_r$. 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 $f^{-1}(x) \cap f^{-1}(y) = \emptyset$ 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. $\dfrac{|\mathfrak{S}(f)|}{|\mathfrak{R}|^n} = \dfrac{1}{|\mathfrak{I}|}$. Moreover, if a_i is a unit, then $\mathfrak{I} = \mathfrak{R}$, thus, $|\mathfrak{S}(f)| = |\mathfrak{R}|^n / |\mathfrak{R}| = |\mathfrak{R}|^{n-1}$. ■

Remark 4. For linear function

$$f : \mathbf{x} \mapsto \mathbf{x}^t[a_1, a_2, \cdots, a_n]^t, \quad \forall\, \mathbf{x} \in \mathfrak{R}^n,$$

Lemma III.1 holds true, if

$$\mathfrak{I} = \langle a_1, a_2, \cdots, a_n\rangle_l = \left\{ \textstyle\sum_{i=1}^n r_i a_i \middle| \, r_i \in \mathfrak{R}, \forall\, 1 \le i \le n \right\}$$

which is a left ideal of $\mathfrak{R}$.

Lemma III.2. *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 $\mathbf{A}$ uniformly at random, then*

$$\Pr\{f(\mathbf{x}) = f(\mathbf{y})\} = |\mathfrak{I}|^{-k}, \tag{1}$$

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

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

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

since f_i's are mutually independent. The statement follows from Lemma III.1 and Remark 4 which assure that $\Pr\{f_i(\mathbf{x} - \mathbf{y}) = 0\} = |\mathfrak{I}|^{-1}$. ■

Definition III.3. 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 $\left\{ \mathbf{x} \in \mathscr{X}^n \middle| \left| \frac{N(x;\mathbf{x})}{n} - p_X(x) \right| \le \epsilon, \forall\, x \in \mathscr{X} \right\}$, where $N(x; \mathbf{x})$ is the number of occurrences of x in the sequence $\mathbf{x}$.

$\mathcal{T}_\epsilon(n, X)$ is sometimes replaced by $\mathcal{T}_\epsilon$ when the length n and the random variable X referred to are clear from the context. Additionally, $H(p)$ is defined to be $H(X)$ for $X \sim p$.

The following lemma is a crucial part of our proof. It generalizes the classic conditional typicality lemma [10, Theorem 15.2.2], yet at the same time distinguishes our argument from the one for the field version.

Lemma III.4. *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 $\forall\, \mathfrak{I} \le_l \mathfrak{R}_1$,*

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

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 $\mathfrak{R}_1/\mathfrak{I}$.

Proof: 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

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

and

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

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,

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

is a strongly 2ϵ-typical sequences of length c_j with respect to the random variable $Z_j \sim p_j$, where p_j is the empirical distribution induced by $\mathbf{z}_j$. The sample space of Z_j is $a_j + \mathfrak{I} \times \mathfrak{R}_2$. Therefore, the number of such a $\mathbf{z}'_j$ of length c_j (all elements $\begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2\end{bmatrix}^t \in \mathcal{T}_{2\epsilon}(c_j, Z_j)$ 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. Consequently,

$$|D_\epsilon(\mathbf{x}_1, \mathfrak{I}|\mathbf{x}_2)| \le 2^{\sum_{j=1}^m c_j[H(p_j) - H(p_{j,2}) + 2\epsilon]}. \tag{3}$$

Direct computation yields

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

and

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

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

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

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

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

Therefore,

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

$$= H(X_1|X_2) - H(X_2) - H(Y_{\mathfrak{R}_1/\mathfrak{I}}) + \eta/4, \quad (5)$$

$$\frac{1}{n} \sum_{j=1}^{m} c_j H(p_{j,2}) > H(X_2, Y_{\mathfrak{R}_1/\mathfrak{I}}) - H(Y_{\mathfrak{R}_1/\mathfrak{I}}) - \eta/4 \quad (6)$$

$$= H(Y_{\mathfrak{R}_1/\mathfrak{I}}|X_2) - H(X_2) - H(Y_{\mathfrak{R}_1/\mathfrak{I}}) - \eta/4, \quad (7)$$

where (4) and (6) are guaranteed for some small $0 < \epsilon < \eta/4$. Substituting (5) and (7) into (3), (2) follows. ■

IV. The Achievability Theorem of LCoR

Generally speaking, the data generated by a source is not necessarily associated with any specific algebraic structure. In order to apply the linear encoders (over ring), we assume that there exists an array $\Phi = [\Phi_1, \Phi_2, \cdots, \Phi_s]$ of injections $\Phi_i : \mathscr{X}_i \to \mathfrak{R}_i$ mapping $\mathscr{X}_i$ to a finite ring $\mathfrak{R}_i$ of *order*[1] $|\mathfrak{R}_i| \geq |\mathscr{X}_i|$ for all $i \in \mathcal{S} = \{1, 2, \cdots, s\}$. Thus, $\mathscr{X}_i$ can be seen as a subset of $\mathfrak{R}_i$ for a fixed Φ. When required, Φ_i can also be selected to obtain desired outcomes (Remark 6). To facilitate our discussion, we define $\Phi(x_T) = \{\Phi_i(x_i)\}_{i \in T}$, where $x_T = \prod_{i \in T} x_i \in \prod_{i \in T} \mathscr{X}_i$, for any $\emptyset \neq T \subseteq \mathcal{S}$, and let $\mathscr{M}(\mathscr{X}_\mathcal{S}, \mathfrak{R}_\mathcal{S})$ be the set

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

($|\mathfrak{R}_i| \geq |\mathscr{X}_i|$ is implicitly assumed).

Theorem IV.1. *Given $\Phi \in \mathscr{M}(\mathscr{X}_\mathcal{S}, \mathfrak{R}_\mathcal{S})$ and let*

$$\mathcal{R}_\Phi = \left\{ [R_1, R_2, \cdots, R_s] \in \mathbb{R}^s \middle| \sum_{i \in T} \frac{R_i \log |\mathfrak{I}_i|}{\log |\mathfrak{R}_i|} > r(T, \mathfrak{I}_T), \forall \, \emptyset \neq T \subseteq \mathcal{S}, \forall \, 0 \neq \mathfrak{I}_i \leq_l \mathfrak{R}_i \right\}, \quad (8)$$

where $r(T, \mathfrak{I}_T) = H(X_T|X_{T^c}) - H(Y_{\mathfrak{R}_T/\mathfrak{I}_T}|X_{T^c})$ and $Y_{\mathfrak{R}_T/\mathfrak{I}_T} = \Phi(X_T) + \mathfrak{I}_T$ is a random variable with sample space $\mathfrak{R}_T/\mathfrak{I}_T$. Any rate in $\mathcal{R}_\Phi$ is achievable with linear coding over finite rings $\mathfrak{R}_1, \mathfrak{R}_2, \cdots, \mathfrak{R}_s$.

Theorem IV.2. *Region* (8) *is the SW region if $\mathfrak{R}_i$ contains no proper non-trivial left ideal, i.e. $\mathfrak{R}_i$ is a field, for all $i \in \mathcal{S}$.*

Remark 5. This theorem states that LCoF is optimal in achieving the SW region, i.e. results of [2] and [3] are reestablished in this respect. The proof can be found in [1].

For the non-field ring scenario, we demonstrate optimality with Example IV.3 in the following. Much more sophisticated analysis on this issue is found in [1] and [4].

Example IV.3. Consider the single source scenario, where $X_1 \sim p$ and $\mathscr{X}_1 = \mathbb{Z}_6$, satisfying the 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 IV.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.

Proof of Theorem IV.1: For a fixed $\Phi = [\Phi_1, \cdots, \Phi_s]$, we can 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 $\sum_{i \in T} \frac{R_i \log |\mathfrak{I}_i|}{\log |\mathfrak{R}_i|} > r(T, \mathfrak{I}_T)$, (this implies that $\frac{1}{n} \sum_{i \in T} k_i \log |\mathfrak{I}_i| - r(T, \mathfrak{I}_T) > 2\eta$ 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.

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. 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

$$\frac{1}{n} \log |\phi_i(\mathfrak{R}_i^n)| \leq \frac{1}{n} \log |\mathfrak{R}_i|^{k_i} = \frac{\log |\mathfrak{R}_i|}{n} \left\lfloor \frac{nR_i}{\log |\mathfrak{R}_i|} \right\rfloor \leq R_i.$$

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:

E_1: $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_s]^t \notin \mathcal{T}_\epsilon$; or
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 AEP, $\Pr\{E_1\} \to 0, n \to \infty$. Additionally, for $\emptyset \neq T \subseteq \mathcal{S}$, let

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

[1] The number of elements of a finite field / ring / left (right) ideal.

We have

$$D_\epsilon(\mathbf{X}; T) = \bigcup\nolimits_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}_T} \left[D_\epsilon(\mathbf{X}_T, \mathfrak{I} | \mathbf{X}_{T^c}) \setminus \{\mathbf{X}\} \right], \quad (9)$$

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,

$$\begin{aligned}
&\Pr\left\{E_2 | E_1^c\right\} \\
&= \sum_{\left[\mathbf{x}_1', \cdots, \mathbf{x}_s'\right]^t \in \mathcal{T}_\epsilon \setminus \{\mathbf{X}\}} \prod_{i \in \mathcal{S}} \Pr\left\{\phi_i(\mathbf{x}_i') = \phi_i(\mathbf{X}_i) | E_1^c\right\} \\
&= \sum_{\emptyset \neq T \subseteq \mathcal{S}} \sum_{\substack{\left[\mathbf{x}_1', \cdots, \mathbf{x}_s'\right]^t \\ \in D_\epsilon(\mathbf{X}; T)}} \prod_{i \in T} \Pr\left\{\phi_i(\mathbf{x}_i') = \phi_i(\mathbf{X}_i) | E_1^c\right\} \qquad (10) \\
&\leq \sum_{\emptyset \neq T \subseteq \mathcal{S}} \sum_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}_T} \sum_{\substack{\left[\mathbf{x}_1', \cdots, \mathbf{x}_s'\right]^t \\ \in D_\epsilon(\mathbf{X}_T, \mathfrak{I} | \mathbf{X}_{T^c}) \setminus \{\mathbf{X}\}}} \\
&\quad \prod_{i \in T} \Pr\left\{\phi_i(\mathbf{x}_i') = \phi_i(\mathbf{X}_i) | E_1^c\right\} \qquad (11) \\
&< \sum_{\emptyset \neq T \subseteq \mathcal{S}} \sum_{\substack{0 \neq \prod_{i \in T} \mathfrak{I}_i \\ \leq_l \mathfrak{R}_T}} \left(2^{n[r(T, \mathfrak{I}_T) + \eta]} - 1\right) \prod_{i \in T} |\mathfrak{I}_i|^{-k_i} \qquad (12) \\
&< (2^s - 1)\left(2^{|R_\mathcal{S}|} - 2\right) \times \\
&\quad \max_{\substack{\emptyset \neq T \in \mathcal{S}, \\ 0 \neq \prod_{i \in T} \mathfrak{I}_i \leq_l \mathfrak{R}_T}} 2^{-n\left[\frac{1}{n} \sum_{i \in T} k_i \log |\mathfrak{I}_i| - [r(T, \mathfrak{I}_T) + \eta]\right]}, \qquad (13)
\end{aligned}$$

where

(10) 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);

(11) follows from (9) by the union bound;

(12) is from Lemma III.2 and Lemma III.4, 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.4). At the same time, ϵ is required to be sufficiently small;

(13) 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 $2^{|R_\mathcal{S}|} - 1$, which is the number of non-empty subsets of $\mathfrak{R}_\mathcal{S}$ ($\supseteq \mathfrak{R}_T$).

Thus, $\Pr\left\{E_2 | E_1^c\right\} \to 0$, when $n \to \infty$, from (13), since for sufficiently large n and small ϵ, $\dfrac{1}{n} \sum_{i \in T} k_i \log |\mathfrak{I}_i| - [r(T, \mathfrak{I}_T) + \eta] > \eta > 0$.

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

Remark 6. Without much effort, one can see that $\mathcal{R}_\Phi$ in Theorem IV.1 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 $\operatorname{cov}(A)$ be the convex hull of a set $A \subseteq \mathbb{R}^s$. By a straightforward time sharing argument, we have that

$$\mathcal{R}_l = \operatorname{cov}\left(\bigcup\nolimits_{\Phi \in \mathscr{M}(\mathscr{X}_\mathcal{S}, \mathfrak{R}_\mathcal{S})} \mathcal{R}_\Phi\right) \qquad (14)$$

is achievable with LCoR. As mentioned in Theorem IV.2, $\mathcal{R}_\Phi$ is the SW region for all feasible Φ, so is $\mathcal{R}_l$, if $\mathfrak{R}_i' s$ are fields. Furthermore, it has also been proved that there always exist non-field rings (e.g. $\mathbb{Z}_{p^2}$, where p is any prime, and etc) such that $\mathcal{R}_l$ is exactly the SW region. In other words, there always exist non-field rings over which linear coding is optimal for any SW scenario [4].

Remark 7. Although our proof (for LCoR) is more complicated than the one proving LCoF, the encoding and decoding procedures are almost identical to the corresponding field versions. Considering the implementation of ring arithmetic is simple in many cases (e.g. $\mathbb{Z}_q$), using LCoR will reduce complexity in implementation. Furthermore, in many specific circumstances of Problem 1, LCoR (regarding non-field rings) dominates its field counterpart regardless which finite field is considered. More details on this are found in [1].

V. Conclusion

This paper proves the achievability part of a source coding theorem of LCoR. This provides a theoretical background for a simple practical encoder design (by making use of finite ring structures). In addition, it also offers a new encoding technique for the computing problem (Problem 1). [1] demonstrates that LCoR outperforms LCoF in achieving strictly larger achievable region with strictly smaller alphabet size in computing certain functions. Therefore, it is worthwhile to further investigate this coding technique.

On the other hand, although the optimality issue is closed in the sense of existence [4], this issue remains open in general.

Acknowledgment

The authors would like to thank the reviewers for their valuable inputs. In particular, we thank the reviewer who suggested a simplified version of the proof of Lemma III.4 that will be considered in future versions of [1].

References

[1] S. Huang and M. Skoglund, "On linear coding over finite rings and applications to computing," *IEEE Transactions on Information Theory*, Submitted. Available: http://www.ee.kth.se/∼sheng11

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

[3] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Transactions on Information Theory*, vol. 28, pp. 585–592, July 1982.

[4] S. Huang and M. Skoglund, *On Existence of Optimal Linear Encoders over Non-field Rings for Data Compression.* KTH Royal Institute of Technology. Available: http://www.ee.kth.se/∼sheng11

[5] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, July 1973.

[6] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Transactions on Information Theory*, vol. 25, pp. 219–221, Mar. 1979.

[7] M. Sefidgaran and A. Tchamkerten, "Computing a function of correlated sources: A rate region," in *IEEE International Symposium on Information Theory*, pp. 1856–1860, Aug. 2011.

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

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

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