On Multiplicative Matrix Channels over Finite Chain Rings

Roberto W. Nóbrega*, Chen Feng†, Danilo Silva*, Bartolomeu F. Uchôa-Filho*

*Department of Electrical Engineering, Federal University of Santa Catarina, Brazil
†Department of Electrical and Computer Engineering, University of Toronto, Canada
rwnobrega@eel.ufsc.br, cfeng@eecg.utoronto.ca, danilo@eel.ufsc.br, uchoa@eel.ufsc.br

***Abstract*—Motivated by nested-lattice-based physical-layer network coding, this paper considers communication in multiplicative matrix channels over finite chain rings. Such channels are defined by the law $Y = AX$, where X and Y are the input and output matrices, respectively, and A is called the transfer matrix. We assume that the instances of the transfer matrix are unknown to the transmitter, but available at the receiver. As contributions, we obtain a closed-form expression for the channel capacity, and we propose a coding scheme that can achieve this capacity with polynomial time complexity. Our results extend the corresponding ones for finite fields.**

***Index Terms*—Finite chain rings, multiplicative matrix channels, random linear network coding.**

I. Introduction

A *multiplicative matrix channel* (MMC) over a finite field $\mathbb{F}_q$ is a communication channel in which the input $\boldsymbol{X} \in \mathbb{F}_q^{n\times \ell}$ and the output $\boldsymbol{Y} \in \mathbb{F}_q^{m\times \ell}$ are related by

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{F}_q^{m\times n}$ is called the *transfer matrix*. Such channels turn out to be suitable models for the end-to-end communication channel between a source node and a sink node in an error-free, erasure-prone network performing random linear network coding [1]. In this context, $\boldsymbol{X}$ is the matrix whose rows are the n packets (of length ℓ) transmitted by the source node, $\boldsymbol{Y}$ is the matrix whose rows are the m packets received by the sink node, and $\boldsymbol{A}$ is a random matrix whose entries are determined by factors such as the network topology and the random choices of the network coding coefficients. Note that each packet can be viewed as an element of the message space $\Omega = \mathbb{F}_q^\ell$, a finite vector space.

The present work considers MMCs over *finite chain rings* (of which finite fields are a special case). The motivation comes from *physical-layer network coding* [2] employing *compute-and-forward* [3]. Indeed, [4] shows that in a wireless network employing compute-and-forward over a generic nested lattice, the end-to-end communication channel between a source node and a sink node can still be modeled[1] by the same channel law (1). In this case, though, the underlying ring is a not a finite field, but a *principal ideal domain* T (typically the integers $\mathbb{Z}$, the Gaussian integers $\mathbb{Z}[i]$, or the Eisenstein integers $\mathbb{Z}[\omega]$), with the corresponding message space Ω being a *finite T-module*. As such, $\Omega \cong T/\langle d_1\rangle \times T/\langle d_2\rangle \times \cdots \times T/\langle d_\ell\rangle$, where $d_1, d_2, \ldots, d_\ell \in T$ are non-zero non-unit elements satisfying $d_1 \mid d_2 \mid \cdots \mid d_\ell$. A special situation commonly found in practice is when the d_is are all powers of a given prime of T. In this case, the message space Ω can also be viewed as a *finite R-module*, where $R = T/\langle d_\ell\rangle$ is a *finite chain ring*.

Finite-field MMCs have been studied under a probabilistic approach according to different assumptions on the transfer matrix [5]–[10]. In this work, we consider finite-chain-ring MMCs with *channel side information at the receiver* (CSIR), that is, we assume that the instances of the transfer matrix $\boldsymbol{A}$ are unknown to the transmitter, but available at the receiver. Besides that, we impose no restrictions on the statistics of $\boldsymbol{A}$, except that it must be independent of $\boldsymbol{X}$. As contributions, we obtain a closed-form expression for the channel capacity (§V-A), and we propose a coding scheme for the channel, which combines several codes over a finite field to obtain a code over a finite chain ring (§V-B and §V-C). We then show that the scheme can achieve the channel capacity with polynomial time complexity, and that it does not necessarily require the complete knowledge of the probability distribution of $\boldsymbol{A}$; only the expected value of its rank (or, rather, its "shape"—see Section II) is needed (§V-D). The results are then naturally adapted to the non-coherent scenario, i.e., the case in which the instances of the transfer matrix are unknown to both the transmitter and receiver (§V-E). Our results extend (and make use of) some of those obtained by Yang et al. in [7], which addresses the finite field case. It is also worth mentioning that a generalization of the results in [6] from finite fields to finite chain rings is presented in [11].

The remainder of this paper is organized as follows. Section II reviews basic concepts on finite chain rings and linear algebra over them. Section III presents the channel model. Section IV reviews some of the existing results on MMCs over finite fields, and Section V contains our main contributions about MMCs over finite chain rings. Finally, Section VI concludes the paper.

The work of R. W. Nóbrega, D. Silva, and B. F. Uchôa-Filho was partly supported by CNPq-Brazil.

[1]Note that any additive error introduced at the physical layer may be avoided, at each relay node, by employing a linear error-detecting code over the underlying ring.

II. Background on Finite Chain Rings

We now present some basic results on finite chain rings and linear algebra over them. For more details, we refer the reader to [12]–[15]. By the term *ring* we always mean a commutative ring with identity $1 \neq 0$.

A. Finite Chain Rings

A ring R is called a *chain ring* if, for any two ideals I, J of R, either $I \subseteq J$ or $J \subseteq I$. It is known that a finite ring R is a chain ring if and only if R is both *principal* (i.e., all of its ideals are generated by a single element) and *local* (i.e., the ring has a unique maximal ideal). Let $\pi \in R$ be any generator for the maximal ideal of R, and let s be the nilpotency index of π (i.e., the smallest integer s such that $\pi^s = 0$). Then, one can show that R has precisely $s+1$ ideals, namely,

$$R = \langle \pi^0 \rangle \supset \langle \pi^1 \rangle \supset \cdots \supset \langle \pi^{s-1} \rangle \supset \langle \pi^s \rangle = \{0\},$$

where $\langle x \rangle$ denotes the ideal generated by $x \in R$. We call the invariant s the *depth* of the chain ring R. Furthermore, it is also known that the quotient $R/\langle \pi \rangle$ is a field, called the *residue field* of R. If $q = |R/\langle \pi \rangle|$, then the size of each ideal of R is $|\langle \pi^i \rangle| = q^{s-i}$, $0 \leq i \leq s$; in particular, $|R| = q^s$. Note that, if $s = 1$, then $\pi = 0$ and R is a finite field.

An example of a chain ring is $\mathbb{Z}_8 = \{0, 1, \ldots, 7\}$, the ring of integers modulo 8. Its ideals are $\langle 1 \rangle = \mathbb{Z}_8$, $\langle 2 \rangle = \{0, 2, 4, 6\}$, $\langle 4 \rangle = \{0, 4\}$, and $\langle 0 \rangle = \{0\}$ (so that $s = 3$), and its residue field is $\mathbb{Z}_8/\langle 2 \rangle \cong \mathbb{F}_2$ (so that $q = 2$).

For the remainder of this paper, R denotes a chain ring with depth s and residue field of size q. Also, $\pi \in R$ denotes a fixed generator for its maximal ideal, and $\Gamma \subseteq R$ denotes a fixed set of coset representatives for the residue field of R. Without loss of generality, assume $0 \in \Gamma$.[2]

Lemma 1. *Every element $x \in R$ can be written uniquely as*

$$x = \sum_{i=0}^{s-1} x^{(i)} \pi^i,$$

where $x^{(i)} \in \Gamma$, for $0 \leq i < s$.

The above expression is known as the *π-adic expansion of x* (with respect to Γ). For example, the 2-adic expansion of $6 \in \mathbb{Z}_8$ with respect to $\Gamma = \{0, 1\}$ is $6 = 0 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2$, i.e., the standard binary expansion of 6.

Note that the uniqueness of the π-adic expansion (given Γ) allows us to define the maps $(\cdot)^{(i)} : R \to \Gamma$, for $0 \leq i < s$. We also define

$$x^{\underline{i}} = \sum_{j=0}^{i-1} x^{(j)} \pi^j,$$

for $0 \leq i \leq s$. One can show that $x^{\underline{i}} \equiv_{\pi^i} x$ for all $x \in R$, where $\equiv_a$ denotes congruence modulo a (i.e., $x \equiv_a y$ if and only if $x - y \in \langle a \rangle$). In particular, $x^{(0)} = x^{\underline{1}} \equiv_\pi x$.

[2] A particularly nice, canonical choice for Γ is $\Gamma(R) = \{x \in R : x^q = x\}$, known as the *Teichmüller coordinate set* of R.

We next mention a few basic results that help us compute with π-adic expansions. The proof is omitted due to lack of space.

Lemma 2. *Let $x, y, z \in R$. Then,*

1) $\left(x\pi^i\right)^{(i+j)} = x^{(j)}$, *for* $0 \leq j < s - i$; *and*
2) $(x + y\pi^i + z\pi^{i+1})^{(i)} \equiv_\pi x^{(i)} + y^{(0)}$.

B. Modules over Finite Chain Rings

By an *s-shape* we mean a non-decreasing sequence of s non-negative integers. Let $\mu = (\mu_0, \mu_1, \ldots, \mu_{s-1})$ be an s-shape. We define

$$R^\mu \triangleq \underbrace{\langle 1 \rangle \times \cdots \times \langle 1 \rangle}_{\mu_0} \times \underbrace{\langle \pi \rangle \times \cdots \times \langle \pi \rangle}_{\mu_1 - \mu_0} \times \cdots \times \underbrace{\langle \pi^{s-1} \rangle \times \cdots \times \langle \pi^{s-1} \rangle}_{\mu_{s-1} - \mu_{s-2}}.$$

Clearly, being a Cartesian product of ideals, R^μ is a finite R-module. Conversely, every finite R-module M is isomorphic to R^μ for some unique s-shape μ. We call μ the *shape* of M, and write $\mu = \operatorname{shape} M$. We have

$$|R^\mu| = q^{|\mu|}, \tag{2}$$

where $|\mu|$ is defined as $|\mu| = \mu_0 + \mu_1 + \cdots + \mu_{s-1}$. For convenience, we set $\mu_{-1} = 0$.

Let μ be an s-shape. We define $\mu - n = (\mu_0 - n, \mu_1 - n, \ldots, \mu_{s-1} - n)$, which is also an s-shape. For convenience, we may write the s-shape $(m, m, \ldots, m)$ simply as m. According to this convention, R^m stands for the same object, whether m is interpreted as an integer or as an s-shape.

C. Matrices over Finite Chain Rings

For any subset $S \subseteq R$, we denote by $S^{m \times n}$ the collection of all $m \times n$ matrices with entries in S. The set of all invertible $n \times n$ matrices over R is called the *general linear group of degree n over R*, and is denoted by $\mathrm{GL}_n(R)$.

Let $A \in R^{m \times n}$, and set $r = \min\{n, m\}$. A diagonal matrix (not necessarily square)

$$D = \operatorname{diag}(d_1, d_2, \ldots, d_r) \in R^{m \times n}$$

is called a *Smith normal form* of A if there exist matrices $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ (not necessarily unique) such that $A = PDQ$ and $d_1 \mid d_2 \mid \cdots \mid d_r$. It is known that matrices over principal rings (in particular, finite chain rings) always have a Smith normal form, which is unique up to multiplication of the diagonal entries by units. In this work, we shall require such entries be powers of $\pi \in R$; by doing so, the Smith normal form becomes (absolutely) unique.

Let $\operatorname{row} A$ and $\operatorname{col} A$ denote the row and column span of $A \in R^{m \times n}$, respectively. By using the Smith normal form, we can easily prove that $\operatorname{row} A$ is isomorphic to $\operatorname{col} A$. We define the *shape* of A as $\operatorname{shape} A = \operatorname{shape}(\operatorname{row} A) = \operatorname{shape}(\operatorname{col} A)$. Moreover, $\mu = \operatorname{shape} A$ if and only if the Smith normal form of A is given by

$$\operatorname{diag}(\underbrace{1, \ldots, 1}_{\mu_0}, \underbrace{\pi, \ldots, \pi}_{\mu_1 - \mu_0}, \ldots, \underbrace{\pi^{s-1}, \ldots, \pi^{s-1}}_{\mu_{s-1} - \mu_{s-2}}, \underbrace{0, \ldots, 0}_{r - \mu_{s-1}}),$$

where $r = \min\{n, m\}$. For example, consider the matrix

$$A = \begin{bmatrix} 4 & 3 & 6 \\ 6 & 7 & 2 \end{bmatrix}$$

over $\mathbb{Z}_8$. Then, $A = PDQ$, where

$$P = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 4 & 3 & 6 \\ 1 & 2 & 6 \\ 5 & 6 & 3 \end{bmatrix},$$

so that $\operatorname{shape} A = (1, 2, 2)$. We also define the kernel of A as usual, that is, $\ker A = \{x \in R^n : Ax = 0\}$. From the first isomorphism theorem [16, §10.2], $\operatorname{col} A \cong R^n / \ker A$.

Let λ be an s-shape. We denote by $R^{n\times\lambda}$ the subset of matrices in $R^{n\times\ell}$ whose rows are elements of R^λ, where $\ell = \lambda_{s-1}$. Clearly, $|R^{n\times\lambda}| = q^{n|\lambda|}$. Finally, we extend the π-adic expansion map $(\cdot)^{(i)}$ to matrices over R in an element-wise fashion. Thus, $A \in R^{n\times\lambda}$ if and only if $A^{(i)} = \begin{bmatrix} B_i & 0 \end{bmatrix} \in \Gamma^{n\times\ell}$, for some $B_i \in \Gamma^{n\times\lambda_i}$, $0 \le i < s$.

III. Channel Model

Throughout this paper, bold symbols are used to represent random entities, while regular symbols are used for their samples.

Recall that a *discrete memoryless channel* (*DMC*) [17] with input $\boldsymbol{x}$ and output $\boldsymbol{y}$ is defined by a triplet $(\mathcal{X}, p_{\boldsymbol{y}|\boldsymbol{x}}, \mathcal{Y})$, where $\mathcal{X}$ and $\mathcal{Y}$ are the *channel input and output alphabets*, respectively, and $p_{\boldsymbol{y}|\boldsymbol{x}}$, called the *channel transition probability*, gives the probability that $\boldsymbol{y} = y \in \mathcal{Y}$ is received given that $\boldsymbol{x} = x \in \mathcal{X}$ is sent. The channel is memoryless in the sense that the output symbol at a given time depends only on the input at that time and is conditionally independent of previous inputs or outputs. The *capacity* of the DMC is given by

$$C = \max_{p_{\boldsymbol{x}}} I(\boldsymbol{x}; \boldsymbol{y}),$$

where $I(\boldsymbol{x}; \boldsymbol{y})$ is the mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$, and the maximization is over all possible input distributions $p_{\boldsymbol{x}}$.

Let R be a finite chain ring, let n and m be positive integers, and let λ be an s-shape. Also, let $p_{\boldsymbol{A}}$ be a probability distribution over $R^{m\times n}$. From these, we can define the MMC with CSIR over R as a DMC with input $\boldsymbol{X} \in R^{n\times\lambda}$, output $(\boldsymbol{Y}, \boldsymbol{A}) \in R^{m\times\lambda} \times R^{m\times n}$, and channel transition probability

$$p_{\boldsymbol{Y},\boldsymbol{A}|\boldsymbol{X}}(Y, A|X) = \begin{cases} p_{\boldsymbol{A}}(A), & \text{if } Y = AX, \\ 0, & \text{otherwise.} \end{cases}$$

In this work, we shall denote the channel just defined simply by $\mathrm{MMC}_{\mathrm{CSIR}}(\boldsymbol{A}, \lambda)$. We also make use of the random variable $\boldsymbol{\rho} = \operatorname{shape} \boldsymbol{A}$, distributed according to

$$p_{\boldsymbol{\rho}}(\rho) = \sum_{A:\, \operatorname{shape} A = \rho} p_{\boldsymbol{A}}(A),$$

Finally, set $\ell = \lambda_{s-1}$ (interpreted as the packet length).

An *matrix (block) code* of *length* N is defined by a pair $(\mathcal{C}, \Phi)$, where $\mathcal{C} \subseteq (R^{n\times\lambda})^N$ is called the *codebook*, and $\Phi : (R^{m\times\lambda} \times R^{m\times n})^N \to \mathcal{C}$ is called the *decoding function*. We sometimes abuse the notation and write $\mathcal{C}$ instead of $(\mathcal{C}, \Phi)$. The *rate* of the code $\mathcal{C}$ is defined by $\mathrm{R}(\mathcal{C}) = (\log|\mathcal{C}|)/N$, and its *probability of error*, denoted by $\mathrm{P_e}(\mathcal{C})$, is defined as usual [17]. When $N = 1$, we say that $\mathcal{C}$ is a *one-shot* code; otherwise, we say that $\mathcal{C}$ is a *multi-shot* code.

IV. Review of the MMC over a Finite Field

In this section, we briefly review some of the existing results about the MMC with CSIR over a finite field (i.e., $R = \mathbb{F}_q$). Note that, in this case, $s = 1$, $\lambda = \ell$, and $\boldsymbol{\rho} = \operatorname{rank} \boldsymbol{A} \triangleq \boldsymbol{r}$. The following result is due to Yang et al. [7], [8].

Theorem 3. *[7, Prop. 1] The capacity of $\mathrm{MMC}_{\mathrm{CSIR}}(\boldsymbol{A}, \ell)$ is given by*

$$C = \mathrm{E}[\boldsymbol{r}]\ell,$$

in q-ary digits per channel use, and is achieved if the input is uniformly distributed over $\mathbb{F}_q^{n\times\ell}$. In particular, the capacity depends on $p_{\boldsymbol{A}}$ only through $\mathrm{E}[\boldsymbol{r}]$.

Also in [7], [8], two multi-shot coding schemes for MMCs over finite fields are proposed, which are able to achieve the channel capacity given in Theorem 3. The first scheme makes use of rank-metric codes [18] and requires $\ell \ge n$ in order to be capacity-achieving; the second scheme is based on random coding and imposes no restriction on ℓ. Both schemes have polynomial time complexity. Also important, both coding schemes are "universal" in the sense that only the value of $\mathrm{E}[\boldsymbol{r}]$ is taken into account in the code construction (the full knowledge of $p_{\boldsymbol{A}}$, or even $p_{\boldsymbol{r}}$, is not required).

V. The MMC over a Finite Chain Ring

Consider again the case of a general finite chain ring R.

A. Channel Capacity

The following result generalizes Theorem 3.

Theorem 4. *The capacity of $\mathrm{MMC}_{\mathrm{CSIR}}(\boldsymbol{A}, \lambda)$ is given by*

$$C = \sum_{i=0}^{s-1} \mathrm{E}[\boldsymbol{\rho}_{s-i-1}]\lambda_i,$$

in q-ary digits per channel use, and is achieved if the input is uniformly distributed over $R^{n\times\lambda}$. In particular, the capacity depends on $p_{\boldsymbol{A}}$ only through $\mathrm{E}[\boldsymbol{\rho}]$.

In order to prove this theorem, we need the following lemma.

Lemma 5. *Let $\boldsymbol{X} \in R^{n\times\lambda}$ be a random matrix, let $A \in R^{m\times n}$ be any fixed matrix, and let $\rho = \operatorname{shape} A$. Define $\boldsymbol{Y} = A\boldsymbol{X} \in R^{m\times\lambda}$. Then,*

$$H(\boldsymbol{Y}) \le \sum_{i=0}^{s-1} \rho_{s-i-1}\lambda_i,$$

where equality holds if $\boldsymbol{X}$ is uniformly distributed over $R^{n\times\lambda}$.

Proof. Note that $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be expressed as

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_0 & \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_{s-1} \end{bmatrix},$$
$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_0 & \boldsymbol{Y}_1 & \cdots & \boldsymbol{Y}_{s-1} \end{bmatrix},$$

where $\boldsymbol{X}_i \in \langle \pi^i \rangle^{n\times(\lambda_i-\lambda_{i-1})}$ and $\boldsymbol{Y}_i \in \langle \pi^i \rangle^{m\times(\lambda_i-\lambda_{i-1})}$, for $0 \le i < s$. We have

$$\boldsymbol{Y}_i = A\boldsymbol{X}_i,$$

so that the support of each of the columns of $\boldsymbol{Y}_i$ is a subset of $\operatorname{col} \pi^i A$. We have $\operatorname{shape} \pi^i A = (0, \dots, 0, \rho_0, \dots, \rho_{s-i-1})$, so that, from (2), we have $|\operatorname{col} \pi^i A| = q^{\rho_0+\cdots+\rho_{s-i-1}}$. Therefore, the support of $\boldsymbol{Y}$ has size at most

$$\begin{aligned}\prod_{i=0}^{s-1} |\operatorname{col} \pi^i A|^{\lambda_i-\lambda_{i-1}} &= \prod_{i=0}^{s-1} q^{(\rho_0+\cdots+\rho_{s-i-1})(\lambda_i-\lambda_{i-1})} \\ &= q^{\sum_{i=0}^{s-1} \rho_{s-i-1}\lambda_i},\end{aligned}$$

from which the inequality follows.

Now suppose $\boldsymbol{X}$ is uniformly distributed over $R^{n\times\lambda}$. This means that $\boldsymbol{X}_i$ is uniformly distributed over $\langle \pi^i \rangle^{n\times(\lambda_i-\lambda_{i-1})}$. One may show that there exists $\boldsymbol{X}'_i$ uniformly distributed over $R^{n\times(\lambda_i-\lambda_{i-1})}$ such that $\boldsymbol{X}_i = \pi^i \boldsymbol{X}'_i$. Let $\boldsymbol{y}$ denote a column of $\boldsymbol{Y}_i$, whose support is $\operatorname{col} \pi^i A$. Since $\boldsymbol{Y}_i = A\boldsymbol{X}_i = \pi^i A\boldsymbol{X}'_i$, we have, for every $y \in \operatorname{col} \pi^i A$,

$$\begin{aligned}\Pr[\boldsymbol{y} = y] &= \frac{|\{x' \in R^n : \pi^i A x' = y\}|}{|R^n|} \\ &= \frac{|\ker \pi^i A|}{|R^n|} \\ &= \frac{1}{|\operatorname{col} \pi^i A|},\end{aligned}$$

that is, $\boldsymbol{y}$ is uniformly distributed over its support. Therefore, $\boldsymbol{Y}$ itself is also uniformly distributed over its support. This concludes the proof. □

We can now prove Theorem 4.

Proof of Theorem 4. The channel mutual information is given by

$$\begin{aligned}I(\boldsymbol{X}; \boldsymbol{Y}, \boldsymbol{A}) &= I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{A}) + I(\boldsymbol{X}; \boldsymbol{A}) \\ &= H(\boldsymbol{Y}|\boldsymbol{A}) - H(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{A}) + I(\boldsymbol{X}; \boldsymbol{A}) \\ &= H(\boldsymbol{Y}|\boldsymbol{A}),\end{aligned}$$

where $H(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{A}) = 0$ since $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$, and $I(\boldsymbol{X}; \boldsymbol{A}) = 0$ since $\boldsymbol{X}$ and $\boldsymbol{A}$ are independent. Thus,

$$I(\boldsymbol{X}; \boldsymbol{Y}, \boldsymbol{A}) = H(\boldsymbol{Y}|\boldsymbol{A}) = \sum_A p_{\boldsymbol{A}}(A) H(\boldsymbol{Y}|\boldsymbol{A} = A),$$

and the result follows from Lemma 5. □

B. Auxiliary Results

Before we describe the proposed coding scheme, we first present two simple lemmas regarding the solution of systems of linear equations over a finite chain ring. These results will serve as a basis for the coding scheme. The first problem turns a system of linear equations over the chain ring into multiple systems over the residue field.

Lemma 6. *Let $y \in R^n$ and $A \in \mathrm{GL}_n(R)$. Let $x \in R^n$ be the (unique) solution of $Ax = y$. Then, the π-adic expansion of x can be obtained recursively from*

$$A^{(0)} x^{(i)} \equiv_\pi y^{(i)} - \left(Ax^{\underline{i}}\right)^{(i)},$$

for $0 \le i < s$.

Proof. For $0 \le i < s$, we have

$$y = Ax = A\sum_{j=0}^{i-1} x^{(j)}\pi^j + Ax^{(i)}\pi^i + A\sum_{j=i+1}^{s-1} x^{(j)}\pi^j,$$

so that, from Lemma 2,

$$y^{(i)} \equiv_\pi \left(Ax^{\underline{i}}\right)^{(i)} + \left(Ax^{(i)}\right)^{(0)}.$$

After simplifying and rearranging we get the equation displayed on the lemma. Since $A^{(0)} \in \mathrm{GL}_n(\mathbb{F}_q/\langle\pi\rangle)$, we can compute, recursively, $x^{(0)}, x^{(1)}, \dots, x^{(s-1)}$. □

The second problem deals with the solution of diagonal systems of linear equations. Let $M_{j:j'}$ denote the sub-matrix of M consisting of rows j up to, *but not including*, j', where we index the matrix entries starting from 0.

Lemma 7. *Let $Y \in R^{m\times\lambda}$ and $D \in R^{m\times n}$, where D is the Smith normal form of itself and has shape ρ. If $Y = DX$, then*

$$X^{(i)}_{0:\rho_{s-i-1}} = \begin{bmatrix} Y^{(i)}_{0:\rho_0} \\ Y^{(i+1)}_{\rho_0:\rho_1} \\ \vdots \\ Y^{(i+s-1)}_{\rho_{s-i-2}:\rho_{s-i-1}} \end{bmatrix},$$

for $0 \le i < s$.

Proof. Note that $Y = DX$ is equivalent to

$$\begin{aligned} Y_{0:\rho_0} &= X_{0:\rho_0}, \\ Y_{\rho_0:\rho_1} &= \pi X_{\rho_0:\rho_1}, \\ &\vdots \\ Y_{\rho_{s-2}:\rho_{s-1}} &= \pi^{s-1} X_{\rho_{s-2}:\rho_{s-1}}. \end{aligned}$$

From Lemma 2, this implies

$$\begin{aligned} X^{(i)}_{0:\rho_0} &= Y^{(i)}_{0:\rho_0}, & 0 \le i < s, \\ X^{(i)}_{\rho_0:\rho_1} &= Y^{(i+1)}_{\rho_0:\rho_1}, & 0 \le i < s-1, \\ &\vdots & \vdots \\ X^{(i)}_{\rho_{s-2}:\rho_{s-1}} &= Y^{(i+s-1)}_{\rho_{s-2}:\rho_{s-1}}, & 0 \le i < 1, \end{aligned}$$

from which the result follows. □

Example: Let $R = \mathbb{Z}_8$, with $\pi = 2$ and $\Gamma = \{0, 1\}$. Let $n = 5$, $m = 4$, and $\lambda = (3, 4, 6)$. Suppose $\rho = (1, 3, 4)$, so that $D = \operatorname{diag}(1, 2, 2, 4) \in \mathbb{Z}_8^{4\times5}$. Also, suppose

$$Y = \begin{bmatrix} 6 & 7 & 1 & 2 & 0 & 4 \\ 6 & 4 & 2 & 0 & 0 & 0 \\ 0 & 2 & 6 & 4 & 0 & 0 \\ 4 & 0 & 4 & 0 & 0 & 0 \end{bmatrix}.$$

From this we can conclude that

$$X^{(0)} = \begin{bmatrix} 0 & 1 & 1 & & & \\ 1 & 0 & 1 & & & \\ 0 & 1 & 1 & & & \\ 1 & 0 & 1 & & & \\ * & * & * & & & \end{bmatrix},$$

$$X^{(1)} = \begin{bmatrix} 1 & 1 & 0 & 1 & & \\ 1 & 1 & 0 & 0 & & \\ 0 & 0 & 1 & 0 & & \\ * & * & * & * & & \\ * & * & * & * & & \end{bmatrix},$$

$$X^{(2)} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{bmatrix},$$

where $*$ denotes unknown entries and blank spaces denote zero entries. Note that the unknown entries are due to $\rho = \operatorname{shape} D$, while blank entries are due to λ (see §II-C).

C. Coding Scheme

We now propose a coding scheme for the channel, which is based on the π-adic expansion discussed in §II-A and in the ideas of the previous subsection. For simplicity of exposition, we will start by describing the scheme for the particular case of one-shot codes. The general case will be discussed afterwards. From now on, let $\mathbb{F}_q = R/\langle\pi\rangle$.

Let $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{s-1}$ (referred to as *component codes*) be a sequence of one-shot matrix codes over the residue field $\mathbb{F}_q$, where each $\mathcal{C}_i$, for $0 \le i < s$, is a code for $\mathrm{MMC}_{\mathrm{CSIR}}(\boldsymbol{A}_i, \lambda_i)$, for some transfer matrix $\boldsymbol{A}_i \in \mathbb{F}_q^{m\times n}$. We will combine these component codes to obtain a matrix code $\mathcal{C}$ over the chain ring R for $\mathrm{MMC}_{\mathrm{CSIR}}(\boldsymbol{A}, \lambda)$. We refer to $\mathcal{C}$ as the *composite code*.

1) Codebook: Denote by $\varphi : R \to \mathbb{F}_q$ the natural projection map from R onto $\mathbb{F}_q$. Also, denote by $\bar{\varphi} : \mathbb{F}_q \to \Gamma$ the coset representative selector map, with the property that $\varphi(\bar{\varphi}(x)) = x$ for all $x \in \mathbb{F}_q$. The codebook $\mathcal{C} \subseteq R^{n\times\lambda}$ is defined by

$$\mathcal{C} = \left\{ \sum_{i=0}^{s-1} X^{(i)} \pi^i : X_i \in \mathcal{C}_i, 0 \le i < s \right\},$$

where

$$X^{(i)} = \begin{bmatrix} \bar{\varphi}(X_i) & 0 \end{bmatrix} \in \Gamma^{n\times\ell}. \tag{3}$$

It should be clear that the codewords in $\mathcal{C}$ indeed satisfy the constraints of $R^{n\times\lambda}$.

2) Decoding: We now describe the decoding procedure. Intuitively, the decoder decomposes a single MMC over the chain ring into multiple MMCs over the residue field. In the following, $M_{j\times k}$ denotes the upper-left $j\times k$ sub-matrix of M.

Step 1. The decoder, which knows the transfer matrix A, starts by computing its Smith normal form $D \in R^{m\times n}$. It also computes $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $A = PDQ$.

Step 2. Let $\rho = \operatorname{shape} A = \operatorname{shape} D$. Define $\tilde{X} \triangleq QX \in R^{n\times\lambda}$ (which is unknown to the receiver) and $\tilde{Y} \triangleq P^{-1}Y \in R^{m\times\lambda}$ (which is calculated at the receiver), so that $Y = AX$ is equivalent to

$$\tilde{Y} = D\tilde{X}.$$

From this equation, the decoder can obtain partial information about $\tilde{X}$. More precisely, it can compute $\tilde{X}^{(i)}_{\rho_{s-i-1}\times\lambda_i}$, for $0 \le i < s$, according to Lemma 7.

Step 3. In possession of $\tilde{X}^{(i)}_{\rho_{s-i-1}\times\lambda_i}$, for $0 \le i < s$, the decoder will then try to decode X based on the equation

$$\tilde{X} = QX,$$

in a *multistage fashion*. Indeed, similarly to Lemma 6, we have, for $0 \le i < s$,

$$\tilde{X}^{(i)} - \left(QX^{\underline{i}}\right)^{(i)} \equiv_\pi Q^{(0)} X^{(i)}.$$

Considering only the ρ_{s-i-1} topmost rows (since the remaining rows are unknown), and keeping only the λ_i leftmost columns (since the remaining columns are already known to be zero), we get

$$\tilde{X}^{(i)}_{\rho_{s-i-1}\times\lambda_i} - \left(Q_{\rho_{s-i-1}\times n} X^{\underline{i}}_{n\times\lambda_i}\right)^{(i)} \equiv_\pi Q^{(0)}_{\rho_{s-i-1}\times n} X^{(i)}_{n\times\lambda_i}.$$

Finally, projecting into $\mathbb{F}_q$ (that is, applying φ to both sides), and appending enough zero rows (in order to obtain an $m\times n$ system) gives

$$Y_i = A_i X_i, \tag{4}$$

where $Y_i \in \mathbb{F}_q^{m\times\lambda_i}$ and $A_i \in \mathbb{F}_q^{m\times n}$ are defined by

$$Y_i = \begin{bmatrix} \varphi\left(\tilde{X}^{(i)}_{\rho_{s-i-1}\times\lambda_i}\right) - \varphi\left(\left(Q_{\rho_{s-i-1}\times n} X^{\underline{i}}_{n\times\lambda_i}\right)^{(i)}\right) \\ 0 \end{bmatrix}, \tag{5}$$

and

$$A_i = \begin{bmatrix} \varphi\left(Q_{\rho_{s-i-1}\times n}\right) \\ 0 \end{bmatrix}. \tag{6}$$

Note that Y_i can only be calculated after $X_0, X_1, \ldots, X_{i-1}$ are known. Therefore, in this step the decoder obtains, successively, estimates of $X_0, X_1, \ldots, X_{s-1}$ from (4). Finally, it computes an estimate of X according to (3) and the π-adic expansion.

3) Extension to the Multi-Shot Case: We finally consider the multi-shot case. Let $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{s-1}$ be a sequence of N-shot matrix codes (the component codes), where $\mathcal{C}_i \subseteq (\mathbb{F}_q^{n\times\lambda_i})^N$, for $0 \le i < s$. The codewords of the composite code $\mathcal{C}$ are then given by $(X(1), X(2), \ldots, X(N)) \in (R^{n\times\lambda})^N$, where $X(j)$ is obtained from the j-th coordinates of the codewords of the component codes, similarly to the one-shot case.

Proceeding similarly to Steps 1 and 2 above, the decoder obtains $\tilde{X}^{(i)}_{\rho_{s-i-1}\times\lambda_i}(j)$, for $0 \le i < s$ and $j = 1, \ldots, N$, and $Q(j)$, for $j = 1, \ldots, N$. Step 3 is also similar, with the important detail that the whole sequence $(X_i(1), X_i(2), \ldots, X_i(N)) \in \mathcal{C}_i$ is decoded from $(Y_i(1), Y_i(2), \ldots, Y_i(N))$ and $(A_i(1), A_i(2), \ldots, A_i(N))$ by using the decoder of $\mathcal{C}_i$, before proceeding to stage $i+1$.

D. Universality, Rate, Probability of Error, and Complexity

From the proposed coding scheme, it is now clear that the i-th component code should be aimed at $\mathrm{MMC}_{\mathrm{CSIR}}(\boldsymbol{A}_i, \lambda_i)$, where $\boldsymbol{A}_i \in \mathbb{F}_q^{m\times n}$ is defined in (6). In principle, we could calculate the probability distribution of $\boldsymbol{A}_i$ provided we have access to the probability distribution of $\boldsymbol{A}$. Nevertheless, if we employ one of the coding schemes proposed in [7] (see Section IV), then the particular probability distribution of $\boldsymbol{A}_i$ becomes unimportant once we know the expected value of its rank. From (6), we have $\operatorname{rank} \boldsymbol{A}_i = \boldsymbol{\rho}_{s-i-1}$, so that, in this case, only the knowledge of $\mathrm{E}[\boldsymbol{\rho}]$ is needed. Thus, the proposed coding scheme is "universal," provided the component codes are also universal.

Let $\mathcal{C}$ denote the composite code obtained from the component codes $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{s-1}$. Then, the $\mathcal{C}$ has a rate of

$$\mathrm{R}(\mathcal{C}) = \mathrm{R}(\mathcal{C}_0) + \mathrm{R}(\mathcal{C}_1) + \cdots + \mathrm{R}(\mathcal{C}_{s-1}),$$

in q-ary digits. Also, from the union bound, the probability of error is upper-bounded by

$$\mathrm{P_e}(\mathcal{C}) \leq \mathrm{P_e}(\mathcal{C}_0) + \mathrm{P_e}(\mathcal{C}_1) + \cdots + \mathrm{P_e}(\mathcal{C}_{s-1}).$$

Thus, if each $\mathcal{C}_i$ is capacity-achieving in $\mathrm{MMC}_{\mathrm{CSIR}}(\boldsymbol{A}_i, \lambda_i)$, we have $\mathrm{R}(\mathcal{C}_i)$ arbitrarily close to $\mathrm{E}[\boldsymbol{\rho}_{s-i-1}]\lambda_i$ and $\mathrm{P_e}(\mathcal{C}_i)$ arbitrarily close to zero, for $0 \leq i < s$. Therefore, $\mathrm{R}(\mathcal{C})$ is arbitrarily close to $\sum_i \mathrm{E}[\boldsymbol{\rho}_{s-i-1}]\lambda_i$ (the channel capacity), and $\mathrm{P_e}(\mathcal{C})$ is arbitrarily close to zero.

The computational complexity of the scheme is simply the sum of the individual computational complexities of each component code, plus the cost of calculating the Smith normal form of A (which can be done with $O(nm\min\{n,m\})$ operations in R), the cost of calculating $\tilde{Y}$ (taking $O(m^2(m+\ell))$ operations), and the cost of $s-1$ matrix multiplications and additions in (5) (taking $O(n^2\ell)$ operations each).

E. Extension to the Non-Coherent Scenario

So far we have only considered the case where channel side information is available at the receiver. Nevertheless, as mentioned in the introduction, we can still reuse the coding scheme proposed in §V even in a non-coherent scenario, by means of *channel sounding* (also known as *channel training*). In this technique, the instances of $\boldsymbol{A}$ are provided to the receiver by introducing headers in the transmitted matrix $\boldsymbol{X} \in R^{n\times\lambda}$, that is, by setting $\boldsymbol{X} = \begin{bmatrix} I & \boldsymbol{X'} \end{bmatrix}$, where $I \in R^{n\times n}$ is the identity matrix, and $\boldsymbol{X'} \in R^{n\times(\lambda-n)}$ is a payload matrix coming from a matrix code. For this to work, we clearly need $\lambda_0 \geq n$. Note that channel sounding introduces an overhead of n^2 symbols. Nevertheless, the overhead can be made negligible if we are allowed to arbitrarily increase the packet length, that is, the proposed scheme can be capacity-achieving in this asymptotic scenario.

VI. Conclusions

In this work we investigated multiplicative matrix channels over finite chain rings, which have practical applications in nested-lattice-based physical-layer network coding. As contributions, the channel capacity has been determined, which generalizes the corresponding result for finite fields. Furthermore, a capacity-achieving coding scheme was proposed, combining several codes over the residue field to obtain a new code over the chain ring.

Several points are still open. The capacity of the non-coherent MMC, a problem addressed in [7], [9] for the case of finite fields, still needs to be generalized for the case of finite chain rings. In this line, the "uniform given shape" distribution seems to be of fundamental importance. Also, designing capacity-achieving coding schemes for the non-coherent MMC with small λ is still an open problem, even in the finite-field case.

References

[1] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Transactions on Information Theory*, vol. 54, pp. 3579–3591, Aug. 2008.

[2] S. C. Liew, S. Zhang, and L. Lu, "Physical-layer network coding: Tutorial, survey, and beyond," *Physical Communication*, vol. 6, pp. 4–42, May 2013.

[3] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Transactions on Information Theory*, vol. 57, pp. 6463–6486, Oct. 2011.

[4] C. Feng, D. Silva, and F. R. Kschischang, "An algebraic approach to physical-layer network coding," *To appear in the IEEE Transactions on Information Theory*.

[5] M. Jafari Siavoshani, S. Mohajer, C. Fragouli, and S. Diggavi, "On the capacity of non-coherent network coding," *IEEE Transactions on Information Theory*, vol. 57, pp. 1046–1066, Feb. 2011.

[6] D. Silva, F. R. Kschischang, and R. Koetter, "Communication over finite-field matrix channels," *IEEE Transactions on Information Theory*, vol. 56, pp. 1296–1305, Mar. 2010.

[7] S. Yang, S.-W. Ho, J. Meng, E.-h. Yang, and R. W. Yeung, "Linear operator channels over finite fields," *Computing Research Repository (CoRR)*, vol. abs/1002.2293, Apr. 2010.

[8] S. Yang, J. Meng, and E.-h. Yang, "Coding for linear operator channels over finite fields," in *Proceedings of the 2010 IEEE International Symposium on Information Theory (ISIT'10)*, (Austin, Texas), pp. 2413–2417, June 2010.

[9] R. W. Nóbrega, D. Silva, and B. F. Uchôa-Filho, "On the capacity of multiplicative finite-field matrix channels," *Computing Research Repository (CoRR)*, vol. abs/1105.6115, Apr. 2013. To appear in the IEEE Transactions on Information Theory.

[10] S. Yang, S.-W. Ho, J. Meng, and E.-h. Yang, "Capacity analysis of linear operator channels over finite fields," *Computing Research Repository (CoRR)*, vol. abs/1108.4257, Dec. 2012.

[11] C. Feng, R. W. Nóbrega, F. R. Kschischang, and D. Silva, "Communication over finite-chain-ring matrix channels," *Computing Research Repository (CoRR)*, vol. abs/1304.2523, Apr. 2013. Submitted to the IEEE Transactions on Information Theory.

[12] B. R. McDonald, *Finite Rings with Identity*, vol. 28 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, Inc., 1974.

[13] A. A. Nechaev, "Finite rings with applications," in *Handbook of Algebra* (M. Hazewinkel, ed.), vol. 5, pp. 213–320, North-Holland, 2008.

[14] T. Honold and I. Landjev, "Linear codes over finite chain rings," *The Electronic Journal of Combinatorics*, vol. 7, 2000.

[15] W. C. Brown, *Matrices over Commutative Rings*, vol. 169 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, Inc., 1992.

[16] D. S. Dummit and R. M. Foote, *Abstract Algebra*. John Wiley and Sons, 3rd ed., 2004.

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

[18] D. Silva, F. R. Kschischang, and R. Koetter, "A rank-metric approach to error control in random network coding," *IEEE Transactions on Information Theory*, vol. 54, pp. 3951–3967, Sept. 2008.