UNIVERSITY OF CALIFORNIA

Los Angeles

Theoretical Bounds and Constructions of Codes in the Generalized Cayley Metric

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Electrical and Computer Engineering

by

Siyi Yang

2018

ABSTRACT OF THE THESIS

Theoretical Bounds and Constructions of Codes
in the Generalized Cayley Metric

by

Siyi Yang

Master of Science in Electrical and Computer Engineering
University of California, Los Angeles, 2018
Professor Lara Dolecek, Chair

Permutation codes have recently garnered substantial research interest due to their potential in various applications including cloud storage systems, genome resequencing and flash memories. In this paper, we study the theoretical bounds and constructions of permutation codes in the generalized Cayley metric. The generalized Cayley metric captures the number of generalized transposition errors in a permutation, and subsumes previously studied error types, including transpositions and translocations, without imposing restrictions on the lengths and positions of the translocated segments. Relying on the *breakpoint analysis* proposed by Chee and Vu, we first propose a coding scheme that is order-optimal albeit not constructive based on this method. We then develop another construction of permutation codes in the generalized Cayley distance. This scheme is both explicit and systematic. We also prove the existence of order- optimal systematic codes and offer a concrete construction based on this method. For the generalized Cayley metric, we prove that our coding schemes have less redundancy than the existing codes based on interleaving when the codelength is sufficiently large and the number of errors is relatively small.

The thesis of Siyi Yang is approved.

Suhas N. Diggavi

Raghu Meka

Lara Dolecek, Committee Chair

University of California, Los Angeles

2018

To mom and dad.

Table of Contents

Acknowledgment

Foremost, I would like to express my sincere gratitude to my advisor Prof. Lara Dolecek for the continuous support of my study and research. Her guidance helped me throughout my research and writing of this thesis.

I am grateful to my lab mate, Clayton Schoeny, for his profound suggestions on my research, and his assistance and patience throughout the process of writing the thesis.

I would also like to thank the rest of my thesis committee: Prof. Suhas Diggavi and Prof. Raghu Meka for their patience on reading and approving this work.

Finally, I must express my gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and research. This accomplishment would not have been possible without them. Thank you.

I also recognized that this work would not have been possible without the financial support NSF CAREER grant No.1150212 and NSF CIF grant No.1527130.

CHAPTER 1

Introduction

This thesis is concerned with theoretical analysis and constructions of permutation codes in the generalized Cayley metric. We open with a brief introduction on the background and applications of permutation codes.

1.1 Background of Permutation Codes

Permutation codes have recently garnered substantial interest due to their potential in flash memories. In recent years, a novel scheme called rank modulation has been popular in flash memories. In flash memories, there are two typical types of problems: the overshoot problem, in which a cell receives higher than normal charge, and the memory endurance problem, in which a defective cell loses its charge more quickly than normal. Aiming at eliminating the overshooting problem in programming cells and the memory endurance problem in aging devices, rank modulation was first proposed by Jiang et al. in [JMSB09]. In the rank modulation scheme, a set of n consecutive cells stores the information in the permutation induced by the relative ranks of the charge levels of the cells. Then the above charge leakage problems between cells all correspond to rearrangements of the stored permutations.

Research on permutation codes has been focused on constructions that proved to be robust to different kinds of errors, thus resulting in different metrics of distance for permutation codes. In recent years, permutation codes in the Kendall-τ metric [BYEB16, BE15, ZG16] and Ulam metric [FM14, FSM13, GLRS15], along with codes in the Levenshtein metric

[GYF+16, CVZ15] have been intensely studied. The Kendall-τ metric captures the number of transpositions required to transform a permutation into the other one, where a transposition refers to exchanging the positions of two adjacent symbols. The Ulam metric captures the number of translocations that is necessary for obtaining a permutation from the other one, where a translocation is equivalent to exchanging a substring and one of its neighboring symbols. In this thesis, however, we pave our way to a more general type of metric, the so-called generalized Cayley metric that captures the generalized transposition errors. Generalized transposition errors subsume transpositions and translocations that the Kendall-τ metric and Ulam metric describe, when no restrictions are imposed on the positions and lengths of the translocated segments [CV14].

Generalized transposition errors are also encountered in various applications including cloud storage systems and genome resequencing. Cloud storage applications such as Dropbox, OneDrive, iTunes, Google play, etc., are becoming increasingly popular, since they help to manage and synchronize the folders stored in different devices (PC) [DW15]. The items in the folders undergo rearrangements between two rounds of synchronizations. These operations, when the items are uniquely labeled and ordered, correspond to transpositions in permutations. In DNA resequencing, released genomes consist of collections of unassembled contigs and their organizations evolve over time by undergoing rearrangement operations. Gene order in a chromosome is subject to rearrangements including reversals, transpositions, translocations, block-interchanges, etc. [ZS17, CLCL16]. Errors encountered in the applications described above can be appropriately modeled by the generalized Cayley metric for permutation codes, introduced by Chee and Vu [CV14], capturing the number of generalized transpositions between two permutations.

Codes in the generalized Cayley metric were first studied in [CV14] using *breakpoint analysis*, wherein a coding scheme is constructed based on permutation codes, previously introduced in [FSM13], in the Ulam metric. Let N be the length of the codewords, and t be the maximum number of errors in the generalized Cayley metric. While the coding scheme

proposed in [CV14] is explicitly constructive and implementable, the interleaving technique used inevitably incurs a noticeable rate penalty of $\mathcal{O}\left(\frac{1}{\log N}\right)$, without even considering the number of errors the codes are able to correct. As we show later, the best possible rate of a length-N code that corrects t generalized transposition errors is $1-\mathcal{O}(\frac{t}{N})$. When t is small compared to $\mathcal{O}(\frac{N}{\log N})$, the gap between the rate of existing codes based on interleaving and the optimal rate increases with N, thus motivating the need to introduce other techniques that are not based on interleaving.

1.2 Outline of Contributions

The content of this thesis is organized as follows. In Section 2.1, we present the basic notation and properties for the generalized Cayley metric and the so-called block permutation metric, which is introduced for metric embedding. In Section 2.2, we define the notion of error-correcting codes in these two metrics and derive useful upper and lower bounds on their optimal rates. We prove the optimal rate to be $1-\mathcal{O}\left(\frac{t}{N}\right)$ and use these results to guide the construction of order-optimal codes in Chapter 3.

In Chapter 3, we propose constructions of order-optimal permutation codes in the generalized Cayley metric. The main idea of our coding schemes is to map each permutation on $[1:N]$ to an unique characteristic set on the Galois field $\mathbb{F}_q$, where q is a prime number such that $N^2-N<q<2N^2-2N$ and N is the codelength. We prove that knowledge of the boundaries of the unaltered segments is sufficient for recovering the permutation from its modified version, obtained through generalized transpositions. We exploit the fact that the symmetric difference of the characteristic sets of two distinct permutations corresponds to these boundaries. Given that the number of such boundaries is linearly upper bounded by the number of generalized transpositions, it is sufficient to find permutations with corresponding sets on $\mathbb{F}_q$ that have large enough set differences to ensure the error correction code property. Our proposed method provides a sufficient condition for ensuring the lower bound on the cardinality of these set differences, which in turn ensures a large enough minimum

distance of the resulting order-optimal code.

We present a method for constructing permutation codes in the generalized Cayley metric. We assign to each permutation of length N a parity check sum with elements chosen from a Galois field $\mathbb{F}_q$, where q is a prime number such that $N^2 - N < q < 2(N^2 - N)$. We prove that the permutations with the same parity check sums constitute a codebook, and we prove that the largest one is order-optimal.

Based on this method, in Chapter 4, we extend our research to the existence of systematic codes. In systematic codes, each codeword contains the information bits as a substring. Systematic codes have the advantages of easy implementation and low decoding complexity. We present an explicitly constructive coding scheme for order-optimal systematic permutation codes in the generalized Cayley metric. We then compare our work to existing schemes and prove that the rates of our proposed codes are higher than those of existing codes based on interleaving. We prove that our coding schemes are more rate efficient when N is large enough and t is relatively small.

Lastly, we conclude and summarize our main contributions in Chapter 5, and discuss future extensions.

CHAPTER 2

Analysis

2.1 Measure of Distance

In this thesis, we denote by $[N]$ the set $\{1, 2, \cdots, N\}$. We let $\mathbb{S}_N$ represent the set of all permutations on $[N]$, where each permutation $\sigma : [N] \to [N]$ is a bijection between $[N]$ and itself. The symbol $\circ$ denotes the composition of functions. Specifically, $\sigma \circ \pi$ denotes the composition of two permutations $\sigma, \pi \in \mathbb{S}_N$, i.e., $(\sigma \circ \pi)(i) = \sigma(\pi(i)), \forall\, i \in [N]$. We assign a vector $(\sigma(1), \sigma(2), \cdots, \sigma(N))$ to each permutation $\sigma \in \mathbb{S}_N$. Under this notation, we call $e = (1, 2, \cdots, N)$ the *identity permutation*. Additionally, σ^{-1} is the inverse permutation of σ. The subsequence of σ from position i to j is written as $\sigma[i; j] \triangleq (\sigma(i), \sigma(i+1), \cdots, \sigma(j))$. The symbol Δ refers to the symmetric difference. Let $\mathrm{GCD}(\cdot)$ and $\mathrm{LCM}(\cdot)$ be the greatest common divisor and the least common multiple, respectively. The symbol $\equiv$ denotes 'congruent modulo'.

2.1.1 Generalized Cayley Distance

A ***generalized transposition*** $\phi(i_1, j_1, i_2, j_2) \in \mathbb{S}_N$, where $i_1 \leq j_1 < i_2 \leq j_2 \in [N]$, refers to a permutation that is obtained from swapping two segments of the identity permutation

[CV14], i.e., $e\left[i_1, j_1\right]$ and $e\left[i_2, j_2\right]$, namely,

$$\begin{aligned}\phi\left(i_1, j_1, i_2, j_2\right) \triangleq \{&1, \cdots, i_1-1, i_2, \cdots, j_2, \\ &j_1+1, \cdots, i_2-1, i_1, \cdots, j_1, j_2+1, \cdots, N\}.\end{aligned} \tag{2.1}$$

Denote the set of all transformations of a single generalized transposition on a permutation of length N as $\mathbb{T}_N$. For each $\pi \in \mathbb{S}_N$ and $\phi\left(i_1, j_1, i_2, j_2\right) \in \mathbb{T}_N$, the permutation obtained from swapping the segments $\pi\left[i_1; j_1\right]$ and $\pi\left[i_2; j_2\right]$ is exactly $\pi \circ \phi$, i.e., the permutation

$$\begin{aligned}&(\pi(1), \cdots, \pi(i_1-1), \pi(i_2), \cdots, \pi(j_2), \pi(j_1+1), \\ &\cdots, \pi(i_2-1), \pi(i_1), \cdots, \pi(j_1), \pi(j_2+1), \cdots, \pi(N)).\end{aligned} \tag{2.2}$$

Example 1. *Let $\pi = (3, 5, 6, 7, 9, 8, 1, 2, 10, 4) \in \mathbb{S}_{10}$. Let the underlines mark the subsequences that are swapped by ϕ. Then we have,*

$$\begin{aligned}\pi &= \left(3, \underline{5, 6, 7, 9}, 8, \underline{1, 2}, 10, 4\right), \\ \phi(2, 5, 7, 8) &= \left(1, \underline{7, 8}, 6, \underline{2, 3, 4, 5}, 9, 10\right), \\ (\pi \circ \phi)(2, 5, 7, 8) &= \left(3, \underline{1, 2}, 8, \underline{5, 6, 7, 9}, 10, 4\right).\end{aligned}$$

Definition 1. (Generalized Cayley Distance, cf. [CV14]) *The **generalized Cayley distance** $d_G(\pi_1, \pi_2)$ is defined as the minimum number of generalized transpositions that is needed to obtain the permutation π_2 from π_1, i.e.,*

$$\begin{aligned}d_G(\pi_1, \pi_2) \triangleq \min_d \{&\exists\ \phi_1, \phi_2, \cdots, \phi_d \in \mathbb{T}_N, \\ &s.t.,\ \pi_2 = \pi_1 \circ \phi_1 \circ \phi_2 \cdots \circ \phi_d\}.\end{aligned} \tag{2.3}$$

Remark 1. (cf. [CV14]). *For all $\pi_1, \pi_2, \pi_3 \in \mathbb{S}_N$, the generalized Cayley distance d_G satisfies the following properties:*

1. *(Symmetry) $d_G(\pi_2, \pi_1) = d_G(\pi_1, \pi_2)$.*

2. *(Left-invariance)* $d_G(\pi_3 \circ \pi_1, \pi_3 \circ \pi_2) = d_G(\pi_1, \pi_2)$.

3. *(Triangle Inequality)* $d_G(\pi_1, \pi_3) \leq d_G(\pi_1, \pi_2) + d_G(\pi_2, \pi_3)$.

Notice that the generalized Cayley distance d_G between two permutations is hard to compute, which makes it difficult to construct codes in the generalized Cayley metric. The common method to address the difficulty of specifying the distances between permutations is metric embedding, where we find another metric that is computable and is of the same order as d_G. Then we are able to transform the construction of codes in d_G into that in the new metric. This new metric is the *block permutation distance* we introduce next.

2.1.2 Block Permutation Distance

We say a permutation $\pi \in \mathbb{S}_N$ is ***minimal***[1] if and only if no consecutive elements in π are also consecutive elements in the identity permutation e, i.e.,

$$\forall\ 1 \leq i < N,\ \pi(i+1) \neq \pi(i) + 1. \tag{2.4}$$

Denote the set of all minimal permutations of length N as $\mathbb{D}_N$. We then define the *block permutation distance* as follows.

Definition 2. *The* ***block permutation distance*** $d_B(\pi_1, \pi_2)$ *between two permutations* $\pi_1, \pi_2 \in \mathbb{S}_N$ *is equal to* d *if*

$$\begin{aligned} \pi_1 &= (\psi_1, \psi_2, \cdots, \psi_{d+1}), \\ \pi_2 &= \left(\psi_{\sigma(1)}, \psi_{\sigma(2)}, \cdots, \psi_{\sigma(d+1)}\right), \end{aligned} \tag{2.5}$$

where $\sigma \in \mathbb{D}_{d+1}$, $\psi_k = \pi_1[i_{k-1}+1 : i_k]$ *for some* $0 = i_0 < i_1 \cdots < i_d < i_{d+1} = N$, *and* $1 \leq k \leq d+1$.

Note that the block permutation distance is d if and only if $(d+1)$ is the minimum number of blocks the permutation π_1 needs to be divided into in order to obtain π_2 through block

[1] We note that this is different from the notion of minimal permutation specified in group theory.

level permutation. Here by block level permutation we refer to dividing the permutation into multiple segments and making a permutation of those segments.

Remark 2. *The block permutation distance d_B also satisfies the properties of symmetry and left-invariance, which are defined in* Remark 1.

Proof. We suppose $\pi_1, \pi_2 \in \mathbb{S}_N$ such that $d_B(\pi_1, \pi_2) = d$. According to the definition of the block permutation distance, π_1, π_2 satisfies (2.5) for some $\sigma \in \mathbb{S}_{d+1}$ and some $\psi_1, \psi_2, \cdots, \psi_{d+1}$.

To prove the symmetry, we define $\psi'_i = \psi_{\sigma(i)}$ for $1 \leq i \leq d+1$, and $\sigma' = \sigma^{-1}$. Then we will have

$$\begin{aligned}
\pi_2 &= \left(\psi'_1, \psi'_2, \cdots, \psi'_{d+1}\right), \\
\pi_1 &= \left(\psi'_{\sigma'(1)}, \psi'_{\sigma'(2)}, \cdots, \psi'_{\sigma'(d+1)}\right),
\end{aligned}$$

thus, $d_B(\pi_2, \pi_1) = d = d_B(\pi_1, \pi_2)$.

To prove the left-invariance, suppose the length of ψ_i is l_i and $\psi_i = (\psi_i(1), \psi_i(2), \cdots, \psi_i(l_i))$ for all $1 \leq i \leq d+1$. For any $\pi_3 \in \mathbb{S}_N$, we define $\tilde{\psi}_i = (\pi_3(\psi_i(1)), \pi_3(\psi_i(2)), \cdots, \pi_3(\psi_i(l_i)))$, for $1 \leq i \leq d+1$. Then we have

$$\begin{aligned}
\pi_3 \circ \pi_1 &= \left(\tilde{\psi}_1, \tilde{\psi}_2, \cdots, \tilde{\psi}_{d+1}\right), \\
\pi_3 \circ \pi_2 &= \left(\tilde{\psi}_{\sigma(1)}, \tilde{\psi}_{\sigma(2)}, \cdots, \tilde{\psi}_{\sigma(d+1)}\right).
\end{aligned}$$

Therefore $d_B(\pi_3 \circ \pi_1, \pi_3 \circ \pi_2) = d = d_B(\pi_1, \pi_2)$. □

Note that Definition 2 is an implicit representation of d_B. We next find other ways to characterize d_B explicitly.

Definition 3. *The* ***characteristic set*** $A(\pi)$ *for any $\pi \in \mathbb{S}_N$ is defined as set of all consecutive pairs in π, i.e.,*

$$A(\pi) \triangleq \{(\pi(i), \pi(i+1)) \,|\, 1 \leq i < N\}. \tag{2.6}$$

Definition 4. *The **block permutation weight** $w_B(\pi)$ is defined as the number of consecutive pairs in π that does not belong to $A(e)$ (w_B is exactly the number of so-called breakpoints in [CV14]), i.e.,*

$$w_B(\pi) \triangleq |A(\pi) \setminus A(e)|. \tag{2.7}$$

Here e refers to the identity permutation.

The next two lemmas state explicit representations of the block permutation distance d_B by the characteristic set and the block permutation weight, respectively. We apply Lemma 1 to construct the coding scheme later in Chapter 3, and Lemma 2 to derive the forthcoming relation between the generalized Cayley distance d_G and the block permutation distance d_B.

Lemma 1. *For all $\pi_1, \pi_2 \in \mathbb{S}_N$,*

$$d_B(\pi_1, \pi_2) = |A(\pi_2) \setminus A(\pi_1)| = |A(\pi_1) \setminus A(\pi_2)|. \tag{2.8}$$

Proof. The proof is in Appendix A. □

Remark 3. *From Lemma 1 and Definition 4, it is obvious that*

$$w_B(\pi) = d_B(e, \pi) = d_B(\pi, e). \tag{2.9}$$

For all $\pi_1, \pi_2 \in \mathbb{S}_N$, it follows immediately from the left-invariance property of d_B and (2.9) that

$$d_B(\pi_1, \pi_2) = w_B\left(\pi_1^{-1} \circ \pi_2\right). \tag{2.10}$$

In *Example 2*, we show how to compute the block permutation distance of two permutations from their characteristic sets, as is indicated in Lemma 1.

Example 2. *Let $\pi_1 = (3, 5, 6, 7, 9, 8, 1, 2, 10, 4)$, $\pi_2 = (3, 1, 2, 8, 5, 6, 7, 9, 10, 4)$. Denote*

$\psi_i, 1 \leq i \leq 4$ and σ as below,

$$\begin{aligned}
\psi_1 &= (3),\\
\psi_2 &= (5,6,7,9),\\
\psi_3 &= (8),\\
\psi_4 &= (1,2),\\
\psi_5 &= (10,4),\\
\sigma &= (1,4,3,2,5).
\end{aligned}$$

Then we have

$$\begin{aligned}
\pi_1 &= (\psi_1, \psi_2, \psi_3, \psi_4, \psi_5)\,,\\
\pi_2 &= \left(\psi_{\sigma(1)}, \psi_{\sigma(2)}, \psi_{\sigma(3)}, \psi_{\sigma(4)}, \psi_{\sigma(5)}\right),
\end{aligned} \tag{2.11}$$

thus, $d_B(\pi_1, \pi_2) = 4$.

Using the characteristic sets,

$$\begin{aligned}
A(\pi_1) = \{&(3,5),(5,6),(6,7),(7,9),\\
&(9,8),(8,1),(1,2),(2,10),(10,4)\},\\
A(\pi_2) = \{&(3,1),(1,2),(2,8),(8,5),\\
&(5,6),(6,7),(7,9),(9,10),(10,4)\},
\end{aligned}$$

we have that

$$|A(\pi_2) \setminus A(\pi_1)| = |\{(3,5),(9,8),(8,1),(2,10)\}| = 4.$$

This example is in accordance with Lemma 1.

2.1.3 Metric Embedding

In general, the generalized Cayley distance is difficult to compute, whereas the block permutation distance is easier to derive. In the next section, we apply metric embedding to transform the problem of code design in d_G into that in d_B, which is easier to deal with, using the following results.

Lemma 2. *For all $\pi_1, \pi_2 \in \mathbb{S}_N$, we have:*

$$w_B(\pi_1 \circ \pi_2) \leq w_B(\pi_1) + w_B(\pi_2). \tag{2.12}$$

Proof. The proof is in Appendix B. □

Remark 4. *It follows immediately from equation (2.10) and Lemma 2 that the block permutation distance satisfies the triangle inequality, i.e., $\forall\ \pi_1, \pi_2, \pi_3 \in \mathbb{S}_N$,*

$$d_B(\pi_1, \pi_3) \leq d_B(\pi_1, \pi_2) + d_B(\pi_2, \pi_3). \tag{2.13}$$

From Lemma 2 and the definitions of the generalized Cayley metric and the block permutation metric, we derive the following relation between d_B and d_G. This result is used later in Chapter 3.

Lemma 3. *For all $\pi_1, \pi_2 \in \mathbb{S}_N$, we have:*

$$d_G(\pi_1, \pi_2) \leq d_B(\pi_1, \pi_2) \leq 4 d_G(\pi_1, \pi_2). \tag{2.14}$$

Proof. To prove the upper bound, we consider two arbitrary permutations π_1, π_2, and let $k = d_G(\pi_1, \pi_2)$. We know from the definition of block permutation weight and generalized transpositions that for any generalized transposition $\phi \in \mathbb{T}_N$ ($\mathbb{T}_N$ is defined before as the set

of all generalized transposition with length N), we have:

$$w_B(\phi) \leq 4. \tag{2.15}$$

From the definition of generalized Cayley metric, we know that $\exists\ \{\phi_i\}_{i=1}^k$, such that

$$\pi_2 = \pi_1 \circ \phi_1 \circ \phi_2 \cdots \circ \phi_k.$$

Then from Lemma 2 and (2.15), we know that:

$$\begin{aligned} d_B(\pi_1, \pi_2) =& w_B\left(\pi_1^{-1} \circ \pi_2\right) \\ =& w_B(\phi_1 \circ \phi_2 \circ \cdots \circ \phi_k) \\ \leq& \sum_{i=1}^{k} w_B(\phi_i) \\ \leq& 4k = 4d_G(\pi_1, \pi_2). \end{aligned}$$

The upper bound is proved.

To prove the lower bound, we consider distinct permutations π_1 and π_2 such that $d_B(\pi_1, \pi_2) = d > 0$. Then, from the definition of the block permutation weight we know that there exists a minimal permutation σ (minimal permutation is defined in Section 2.1.2 and a partition $\{\psi_i\}_{i=1}^{d+1}$ of π_1 such that $\pi_1 = (\psi_1, \psi_2, \cdots, \psi_{d+1})$ and

$$\pi_2 = \left(\psi_{\sigma(1)}, \psi_{\sigma(2)}, \cdots, \psi_{\sigma(d+1)}\right).$$

Next, suppose l_0 is the smallest index l such that $\sigma(l) \neq l$, $1 \leq l \leq N$. Let $k_0 = \sigma^{-1}(l_0)$, we have $k_0 > l_0$. Let ϕ_1 be the generalized transposition that swaps the subsequences $\left(\psi_{\sigma(l_0)}, \psi_{\sigma(l_0+1)}, \cdots, \psi_{\sigma(k_0-1)}\right)$ and $\psi_{\sigma(k_0)} = \psi_{l_0}$ in π_2. Let $\pi_1^{(1)} = \pi_2 \circ \phi_1$ and

$\sigma^{(1)} = (1, 2, \cdots, l_0, \sigma(l_0), \sigma(l_0+1), \cdots, \sigma(k_0-1), \sigma(k_0+1), \cdots, \sigma(d+1)$. Then,

$$\pi_2^{(1)} = \left(\psi_{\sigma^{(1)}(1)}, \psi_{\sigma^{(1)}(2)}, \cdots, \psi_{\sigma^{(1)}(d+1)}\right).$$

Here $\pi_2^{(1)} = \pi_2 \circ \phi_1$. If $\pi_2^{(1)} = \pi_1$, then $\pi_1 = \pi_2 \circ \phi_1$. Otherwise we let l_1 be the smallest index l such that $\sigma^{(1)}(l) \neq l$, $1 \leq l \leq N$, and we know that $l_1 > l_0$.

We can find a series of generalized transpositions $\phi_1, \phi_2, \cdots, \phi_m$, $1 \leq m \leq d$ sequentially, by the following procedure, such that $\pi_2 \circ \phi_1 \circ \phi_2 \circ \cdots \circ \phi_m = \pi_1$. Suppose $\phi_1, \phi_2, \cdots, \phi_i$ are found. Let $\pi_2^{(i)} = \phi_1 \circ \phi_2 \circ \cdots \circ \phi_i = \left(\psi_{\sigma^{(i)}(1)}, \psi_{\sigma^{(i)}(2)}, \cdots, \psi_{\sigma^{(i)}(d+1)}\right)$. If $\pi_2^{(i)} = \pi_1$, then $\pi_1 = \pi_2 \circ \phi_1 \circ \phi_2 \circ \cdots \circ \phi_i$. Otherwise we let l_i be the smallest index such that $\sigma^{(i)}(l_i) \neq l_i$. Suppose $k_i = \left(\sigma^{(i)}\right)^{-1}(l_i)$, and we have $k_i > l_i$.

Denote the generalized transposition that swaps the subsequences $\left(\psi_{\sigma^{(i)}(1)}, \psi_{\sigma^{(i)}(2)}, \cdots, \psi_{\sigma^{(i)}(k_i-1)}\right)$ and $\psi_{\sigma^{(i)}(k_i)} = \psi_{l_i}$ in $\pi_2^{(i)}$ by ϕ_{i+1}. Let $\pi_2^{i+1} = \pi_2^{(i)} \circ \phi_{i+1}$, $\sigma^{(i+1)} = (1, 2, \cdots, l_i, \sigma^{(i)}(l_i), \sigma^{(i)}(l_i + 1), \cdots, \sigma^{(i)}(k_i - 1), \sigma^{(i)}(k_i + 1), \cdots, \sigma^{(i)}(d+1)$. Then,

$$\pi_2^{(i+1)} = \left(\psi_{\sigma^{(i+1)}(1)}, \psi_{\sigma^{(i+1)}(2)}, \cdots, \psi_{\sigma^{(i+1)}(d+1)}\right).$$

Follow this procedure, and suppose m is the smallest integer such that $\pi_2^{(m)} = \pi_1$. In this procedure, we find $l_0, \cdots, l_{m-1}$ sequentially, where $1 < l_0 < l_1 < \cdots < l_{m-1}$. We also know that $l_{m-1} \leq d$, otherwise we must have $\sigma^{(m-1)}(i) = i$ for all $1 \leq i \leq d$, and $\sigma^{(m-1)}(d+1) \neq d+1$, which leads to a contradiction. Therefore $d \geq l_{m-1} > \cdots > l_0 \geq 1$, which indicates that $m \leq d$. Note that $\pi_1 = \pi_2 \circ \phi_1 \circ \cdots \circ \phi_m$, from which we know that $d_G(\pi_1, \pi_2) \leq m \leq d = d_B(\pi_1, \pi_2)$.

The lemma is proved. □

2.2 Theoretical Bounds

A subset $\mathcal{C}_G(N,t)$ of $\mathbb{S}_N$ is called a ***t-generalized Cayley code*** if it can correct t generalized transposition errors. Any t-generalized Cayley code has minimum generalized Cayley distance $d_{G,min} \geq 2t+1$. Similarly, a subset $\mathcal{C}_B(N,t)$ of $\mathbb{S}_N$ is a ***t-block permutation code*** if its minimum block permutation distance $d_{B,min} \geq 2t+1$. Denote the code rate of $\mathcal{C}_G(N,t)$, $\mathcal{C}_B(N,t)$ as $R_G(N,t)$ and $R_B(N,t)$, respectively. Let $\mathcal{C}_{G,opt}(N,t)$ and $\mathcal{C}_{B,opt}(N,t)$ be t-generalized Cayley codes and t-block permutation codes with optimal rate, denoted as $R_{G,opt}(N,t)$ and $R_{B,opt}(N,t)$, respectively. We derive lower bounds and the upper bounds of $R_{G,opt}(N,t)$ and $R_{B,opt}(N,t)$.

For each $\pi \in \mathbb{S}_N$, we define the ***generalized Cayley ball*** $B_G(N,t,\pi)$ of radius t centered at π to be the set of all permutations in $\mathbb{S}_N$ that have a generalized Cayley distance from π not exceeding t. We know from the left-invariance property of d_G that the cardinality of $B_G(N,t,\pi)$ is independent of π; we denote $|B_G(N,t,\pi)|$ as $b_G(N,t)$. The ***block permutation ball*** $B_B(N,t,\pi)$ and the corresponding ball-size $b_B(N,t)$ are similarly defined.

We derive the lower and upper bounds of $b_B(N,t)$ and $b_G(N,t)$ in the following two lemmas, respectively. We build on these results and Lemma 6 to compute the bounds of the rate of optimal codes in d_G and d_B, proving that their redundancy is $\mathcal{O}(\frac{t}{N})$.

Lemma 4. *For all $N \in \mathbb{N}^*$, $t \leq N-\sqrt{N}-1$, $b_B(N,t)$ is bounded by the following inequality:*

$$\prod_{k=1}^{t}(N-k) \leq b_B(N,t) \leq \prod_{k=0}^{t}(N-k). \tag{2.16}$$

Proof. The proof is in Appendix C. □

Lemma 5. *For all $N \in \mathbb{N}^*$, $t \leq \min\{N-\sqrt{N}-1, \frac{N-1}{4}\}$, $b_G(N,t)$ is bounded as follows:*

$$\binom{N-1}{4t}\frac{(2t)!}{2^t t!} \leq b_G(N,t) \leq \prod_{k=0}^{4t}(N-k). \tag{2.17}$$

Proof. The proof is in Appendix D. □

As the metric d_B and d_G both satisfy the triangle inequality, it follows that a subset

$\mathcal{C} \subseteq \mathbb{S}_N$ is a t-generalized Cayley code if $d_G(x,y) > 2t$ for all $x, y \in \mathcal{C}$, $x \neq y$. Similarly, C is a t-block permutation code if $d_B(x,y) > 2t$ for all $x, y \in \mathcal{C}$, $x \neq y$. The cardinalities of the optimal codes $\mathcal{C}_{B,opt}(N,t)$ and $\mathcal{C}_{G,opt}(N,t)$ are bounded by the inequalities below.

$$\begin{aligned} \frac{N!}{b_B(N,2t)} \leq |\mathcal{C}_{B,opt}(N,t)| \leq \frac{N!}{b_B(N,t)}, \\ \frac{N!}{b_G(N,2t)} \leq |\mathcal{C}_{G,opt}(N,t)| \leq \frac{N!}{b_G(N,t)}. \end{aligned} \tag{2.18}$$

We formulate the optimal code rate as follows

$$\begin{aligned} R_{B,opt}(N,t) = \frac{\log|\mathcal{C}_{B,opt}(N,t)|}{\log N!}, \\ R_{G,opt}(N,t) = \frac{\log|\mathcal{C}_{G,opt}(N,t)|}{\log N!}. \end{aligned} \tag{2.19}$$

From [Rob55, (1)-(2)], we know that for all $N \in \mathbb{N}^*$,

$$N! = \sqrt{2\pi} N^{N+1/2} e^{-N} \cdot e^{r_N}, \tag{2.20}$$

where

$$\frac{1}{12N+1} < r_n < \frac{1}{12N}. \tag{2.21}$$

From (2.20) and (2.21), Lemma 6 follows.

Lemma 6. *For all $N \in \mathbb{N}^*$, it follows that*

$$(N+\frac{1}{2})\log N - (\log e)N < \sum_{n=1}^{N} \log n < (N+1)\log N - N + 2.$$

Theorem 1. *For fixed t and sufficiently large N, the optimal rates satisfy the following inequalities,*

$$\begin{aligned} 1 - c_1 \cdot \frac{2t+1}{N} \leq & R_{B,opt}(N,t) \leq 1 - \frac{t}{N}, \\ 1 - c_1 \cdot \frac{8t+1}{N} \leq & R_{G,opt}(N,t) \leq 1 - c_2 \cdot \frac{4t}{N}, \end{aligned} \tag{2.22}$$

where $c_1 = 1 + \frac{2\log e}{\log N}$, $c_2 = 1 - \frac{3(\log t+1)}{4(\log N-1)}$.

Proof. From (2.18) and (2.19), it follows that

$$\begin{aligned} 1 - \frac{\log b_B(N,2t)}{\log N!} \le R_{B,opt}(N,t) \le 1 - \frac{\log b_B(N,t)}{\log N!}, \\ 1 - \frac{\log b_G(N,2t)}{\log N!} \le R_{G,opt}(N,t) \le 1 - \frac{\log b_G(N,t)}{\log N!}. \end{aligned} \tag{2.23}$$

By applying Lemma 5 and Lemma 6 to (2.23), for fixed t and sufficiently large N, we have

$$\begin{aligned} R_{G,opt}(N,t) &\ge 1 - \frac{\log\left[\prod_{k=0}^{8t}(N-k)\right]}{\log N!} \\ &> 1 - \frac{(8t+1)\log N}{(N+\frac{1}{2})\log N - (\log e)N} \\ &> 1 - \frac{8t+1}{N}\left(1 + \frac{2\log e}{\log N}\right). \end{aligned}$$

$$\begin{aligned} R_{G,opt}(N,t) &\le 1 - \frac{\log\left[\binom{N-1}{4t}\frac{(2t)!}{2^t t!}\right]}{\log N!} \\ &= 1 - \frac{\log\left[\prod_{k=1}^{4t}(N-k)\right] + \log\left[\frac{(2t)!}{(4t)!2^t t!}\right]}{\log N!} \\ &< 1 - \frac{4t(\log N - \frac{1}{8}) - 3t\log(4t)}{(N+1)\log N - N + 2} \\ &= 1 - \frac{4t}{N}\left(\frac{\log N - \frac{1}{8} - \frac{3}{2} - \frac{3}{4}\log t}{\log N - 1 + \frac{2+\log N}{N}}\right) \\ &= 1 - \frac{4t}{N}\left(1 - \frac{\frac{5}{8} + \frac{2+\log N}{N} + \frac{3}{4}\log t}{\log N - 1 + \frac{2+\log N}{N}}\right) \\ &< 1 - \frac{4t}{N}\left(1 - \frac{\frac{3}{4}(\log t + 1)}{\log N - 1}\right) \\ &= 1 - \frac{4t}{N}\left(1 - \frac{3(\log t+1)}{4(\log N - 1)}\right). \end{aligned}$$

Similarly, by applying Lemma 4 and Lemma 6 to (2.23), we have

$$\begin{aligned} R_{B,opt}(N,t) &\geq 1 - \frac{\log\left[\prod_{k=0}^{2t}(N-k)\right]}{\log N!} \\ &> 1 - \frac{(2t+1)\log N}{(N+\frac{1}{2})\log N - (\log e)N} \\ &> 1 - \frac{(2t+1)\log N}{N(\log N - \log e)} \\ &> 1 - \frac{2t+1}{N}\left(1 + \frac{2\log e}{\log N}\right). \end{aligned}$$

$$\begin{aligned} R_{B,opt}(N,t) &\leq 1 - \frac{\log\left[\prod_{k=1}^{t}(N-k)\right]}{\log N!} \\ &< 1 - \frac{t(\log N - \frac{1}{2})}{(N+1)\log N - N + 2} \\ &< 1 - \frac{t(\log N - \frac{1}{2})}{N\log N - \frac{1}{2}N} \\ &= 1 - \frac{t}{N}. \end{aligned}$$

The theorem is proved. □

The above two inequalities in Theorem 1 indicate that the rate $R = 1 - \mathcal{O}\left(\frac{t}{N}\right)$ is order-optimal for both the t-generalized Cayley codes and the t-block permutation codes.

CHAPTER 3

Order-optimal Codes in the Generalized Cayley Metric

We discussed the optimal rate of t-generalized Cayley Codes and t-block permutation codes in the previous chapter. We now focus on the corresponding constructions of codes with the order-optimal rates. We know from Lemma 3 that any $4t$-block permutation code is also a t-generalized Cayley code. In the sequel, we thus focus on the construction of order-optimal t-block permutation codes. In Section 3.1, we develop a construction of order-optimal t-block permutation codes (Theorem 2). We then provide the corresponding decoding scheme of the proposed codes in Section 3.2.

3.1 Encoding Scheme

Denote the set of all ordered pairs of non-identical elements from $[N]$ as P, then $|P| = N^2 - N$. Suppose q is a prime number such that $q \geq |P|$. From *Bertrand's postulate*, we can always find a q such that $|P| \leq q \leq 2|P|$.

Let $\upsilon : \ P \to \mathbb{F}_q$ be an arbitrary injection from P to $\mathbb{F}_q$, where $\mathbb{F}_q$ is the Galois field of order q. Let $\mathcal{P}(\mathbb{F}_q)$ represent the power set of $\mathbb{F}_q$. We define an injection $\nu : \mathbb{S}_N \to \mathcal{P}(\mathbb{F}_q)$ as follows:

$$\nu(\pi) \triangleq \{\upsilon(p) | p \in A(\pi)\}. \tag{3.1}$$

Then ν is invertible, namely, we are able to compute a unique π from $\nu(\pi)$.

We then define a class of surjections $\alpha^{(q,d)} : \mathbb{S}_N \to \mathbb{F}_q^{2d-1}$ as follows:

$$\alpha^{(q,d)}(\pi) \triangleq (\alpha_1, \alpha_2, \cdots, \alpha_{2d-1}), \tag{3.2}$$

where

$$\begin{cases} \alpha_1 \equiv \sum\limits_{b\in\nu(\pi)} b & \mod q, \\ \alpha_2 \equiv \sum\limits_{b\in\nu(\pi)} b^2 & \mod q, \\ \quad\vdots & \\ \alpha_{2d-1} \equiv \sum\limits_{b\in\nu(\pi)} b^{2d-1} & \mod q. \end{cases} \tag{3.3}$$

The following Lemma 7 states that the cardinality of the symmetric difference of $\nu(\pi_1), \nu(\pi_2)$ for any two permutation $\pi_1, \pi_2 \in \mathbb{S}_N$, $\pi_1 \neq \pi_2$ is greater than $2d$ if their parity check sums $\alpha^{(q,d)}(\pi_1)$ and $\alpha^{(q,d)}(\pi_2)$ are identical. Therefore their block permutation distance is greater than d based on Lemma 1. This lemma is applied throughout this paper in the construction of order-optimal permutation codes in both the generalized Cayley metric and the block permutation distance.

Lemma 7. *For all π_1, $\pi_2 \in \mathbb{S}_N$, $\pi_1 \neq \pi_2$, if $\alpha^{(q,d)}(\pi_1) = \alpha^{(q,d)}(\pi_2)$, then*

$$|\nu(\pi_1)\Delta\nu(\pi_2)| > 2d. \tag{3.4}$$

Proof. The proof is in Appendix E. □

Note that the function $\alpha^{(q,2t)}$ induces a surjection from $\mathbb{S}_N$ to $\mathbb{F}_q^{4t-1}$ and divides $\mathbb{S}_N$ into q^{4t-1} subsets based on their parity check sums $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_{4t-1})$. We next prove that each such subset is a t-block permutation code, which is stated as the following theorem.

Theorem 2. *For all* $\boldsymbol{\alpha} \in \mathbb{F}_q^{4t-1}$*, suppose:*

$$\mathcal{C}_{\boldsymbol{\alpha}}(N,t) = \{\pi | \pi \in \mathbb{S}_N,\ \alpha^{(q,2t)}(\pi) = \boldsymbol{\alpha}\}, \tag{3.5}$$

where $\alpha^{(q,2t)}$ *is defined in (4.2). Then* $\forall\ \pi_1, \pi_2 \in \mathcal{C}_{\boldsymbol{\alpha}}(N,t)$*,* $\pi_1 \neq \pi_2$*, we have*

$$d_B(\pi_1, \pi_2) \geq 2t + 1. \tag{3.6}$$

Proof. Let $d = 2t$ in Lemma 1, then (3.1) implies that

$$\begin{aligned} d_B(\pi_1, \pi_2) &= \frac{1}{2}|A(\pi_1)\Delta A(\pi_2)| \\ &= \frac{1}{2}|\nu(\pi_1)\Delta\nu(\pi_2)| \\ &> \frac{1}{2}(2 \cdot 2t) = 2t, \end{aligned} \tag{3.7}$$

where Δ refers to the symmetric difference of sets. □

Theorem 2 implies that $\{\mathcal{C}_{\boldsymbol{\alpha}}(N,t) : \boldsymbol{\alpha} \in \mathbb{F}_q^{4t-1}\}$ is a partition of $\mathbb{S}_N$, where each component $\mathcal{C}_{\boldsymbol{\alpha}}(N,t)$ is a t-block permutation code indexed by $\boldsymbol{\alpha}$. Suppose $\mathcal{C}_{\boldsymbol{\alpha}_{\max}}(N,t)$ is the one with maximal cardinality and has parity check sum $\boldsymbol{\alpha}_{\max}$. It follows from *Pigeonhole Principle* that:

$$|\mathcal{C}_{\boldsymbol{\alpha}_{\max}}(N,t)| \geq \frac{N!}{|\mathbb{F}_q^{4t-1}|} = \frac{N!}{q^{4t-1}}. \tag{3.8}$$

Denote the rate of $\mathcal{C}_{\boldsymbol{\alpha}_{\max}}(N,t)$ by $R_B(N,t)$. Given that $N^2 - N = |P| \leq q < 2|P| =$

$2N^2 - 2N < 2N^2$, it follows from Lemma 6 that for sufficiently large N,

$$\begin{aligned} R_B(N,t) &\geq 1 - \frac{4t\log q}{\log N!} > 1 - \frac{8t\log N + 4t}{\log N!} \\ &> 1 - \frac{8t(\log N + \frac{1}{2})}{(N+\frac{1}{2})\log N - (\log e)N} \\ &> 1 - \frac{8t}{N}\left(\frac{\log N + \frac{1}{2}}{\log N - \log e}\right) \\ &> 1 - \frac{8t}{N}\left(1 + \frac{1}{2\log N}\right)\left(1 + \frac{2\log e}{\log N}\right) \\ &> 1 - \frac{8t}{N}\left(1 + \frac{2\log e + 1}{\log N}\right). \end{aligned} \tag{3.9}$$

Then $\mathcal{C}_{\boldsymbol{\alpha}_{\max}}(N,t)$ is an order-optimal t-block permutation code.

3.2 Decoding Scheme

In previous discussion, we map each permutation $\pi \in \mathbb{S}_N$ to a unique set $\nu(\pi) \in \mathcal{P}(\mathbb{F}_q)$ as defined in equation (3.1), where $N^2 - N \leq q \leq 2N^2 - 2N$ and $\mathcal{P}(\mathbb{F}_q)$ represents the power set of $\mathbb{F}_q$. In the decoding scheme, our objective is to compute $\nu(\pi)$ from a previously specified parity check sum $\boldsymbol{\alpha}$ and the received permutation π'. The strategy is, for every set $B \in \mathcal{P}(\mathbb{F}_q)$, map B to a polynomial $f(X;B)$ defined as follows:

$$f(X;B) \triangleq \prod_{b\in B}(X+b) = X^{N-1} + \sum_{i=1}^{N-1} e_i^B X^{N-1-i}. \tag{3.10}$$

We call $f(X;B)$ the ***characteristic function*** of the set B. All the polynomials as well as the polynomial operations are defined in $\mathbb{F}_q$.

Given the a priori agreement on the codebook, i.e., the choice of $\boldsymbol{\alpha}$, the value of the first $4t$ coefficients of $f(X;B)$ and $f(X;B')$ can be computed, where $B = \nu(\pi)$ and $B' = \nu(\pi')$. We then use these coefficients to derive $\nu(\pi) = B$. Note that this coding strategy bares resemblance to that proposed in [DA10], the key difference being that the coefficients of the polynomials we discussed are partially known, which making our decoding scheme more

complicated, whereas those in [DA10] are fully known.

Note that $e_i^B, 1 \leq i \leq N-1$ in (3.10) is the i-th elementary symmetric polynomial of the elements in B. Also note that the i-th component $\alpha_i, 1 \leq i \leq 4t-1$, of the value $\boldsymbol{\alpha} = \alpha^{(q,2t)}(\pi)$ is exactly the i-th power sum of the elements in $B = \nu(\pi)$. We know from *Newton's identities* that there exists a bijection between the $4t-1$ power sums and the first $4t-1$ elementary symmetric polynomials of elements in B, described below:

$$\begin{cases} e_0^B = 1, \\ e_1^B = \alpha_1, \\ e_2^B = 2^{-1}(e_1^B\alpha_1 - \alpha_2), \\ e_3^B = 3^{-1}(e_2^B\alpha_1 - e_1^B\alpha_2 + \alpha_3), \\ \vdots \\ e_{4t-1}^B = (4t-1)^{-1}(e_{4t-2}^B\alpha_1 - e_{4t-3}^B\alpha_2 + \cdots + \alpha_{4t-1}). \end{cases} \tag{3.11}$$

Denote the coefficient of X^{N-i-1} in $f(X;B)$ and $f(X;B')$ by a_i, a_i', respectively. Let $r(B) = (a_1, a_2, \cdots, a_{4t-1})$, $r(B') = (a_1', a_2', \cdots, a_{4t-1}')$. Suppose the transmitter sends $\pi \in \mathbb{S}_N$ and the receiver receives π', where $d_G(\pi, \pi') \leq t$. The receiver uses the knowledge of $\boldsymbol{\alpha}$ to compute $r(B)$ and to derive $r(B')$ from B', where $B = \nu(\pi)$ and $B' = \nu(\pi')$. Note that π can be computed from $B = \nu(\pi)$ since ν is an injection from $\mathbb{S}_N$ to $\mathcal{P}(\mathbb{F}_q)$. Thus the objective is to compute B from $r(B)$, $r(B')$ and B'.

Denote $f_1 = f(X;B), f_2 = f(X;B')$, where $f(X;B)$ is specified in (3.10). Our objective

is to compute B from f_1, f_2. Suppose $D_2 = B \setminus B'$, $D_1 = B' \setminus B$, $D_3 = B \cap B'$. Let

$$\begin{aligned} g_1(X) &= \frac{f_1}{GCD(f_1, f_2)} = \prod_{b \in D_2} (X + b), \\ g_2(X) &= \frac{f_2}{GCD(f_1, f_2)} = \prod_{b \in D_1} (X + b), \\ g_3(X) &= GCD(f_1, f_2) = \prod_{b \in D_3} (X + b). \end{aligned} \tag{3.12}$$

Notice that g_1, g_2, g_3 uniquely determine f_1, f_2, which indicates that they are sufficient for computing π. We next seek to compute g_1, g_2, g_3 from $r(B)$ and $f_2 = g_2 \cdot g_3$, from which $f_1 = g_1 \cdot g_3$ can be determined. Let $(h_1, h_2) = (X^{t-k} g_2, X^{t-k} g_1)$, where $k = \deg g_1 = \deg g_2 = |D_1| = |D_2| \leq t$. Then (h_1, h_2) satisfy $h_1 \cdot f_1 = h_2 \cdot f_2$. We will also prove later in Theorem 3 that g_1, g_2, g_3 can be computed from an arbitrary solution (h_1, h_2) of $h_1 \cdot f_1 = h_2 \cdot f_2$. Therefore any solution to $h_1 \cdot f_1 = h_2 \cdot f_2$ is sufficient for computing π. Also notice that the first $4t - 1$ coefficients of $h_1 \cdot f_1$ and $h_2 \cdot f_2$ uniquely determine $r(B)$ and $r(B')$, respectively, if h_1, h_2 are known. In order to compute g_1, g_2, g_3, it is sufficient to find (h_1, h_2) of degree t such that the first $4t - 1$ coefficients of $h_1 \cdot f_1$ and that of $h_2 \cdot f_2$ are equal, i.e., the following inequality holds,

$$\deg(h_1 \cdot f_1 - h_2 \cdot f_2) < N - 3t. \tag{3.13}$$

For each $\mathbf{c} \in \mathbb{F}_q^{2t}$, where

$$\mathbf{c} = \left(c_1, \cdots, c_t, -c'_1, \cdots, -c'_t \right)^T, \tag{3.14}$$

define the polynomials $h_1(\mathbf{c}), h_2(\mathbf{c})$ of degree t as follows,

$$\begin{aligned} h_1(\mathbf{c}) &\triangleq X^t + c_1 X^{t-1} + c_2 X^{t-2} + \cdots + c_t, \\ h_2(\mathbf{c}) &\triangleq X^t + c'_1 X^{t-1} + c'_2 X^{t-2} + \cdots + c'_t. \end{aligned} \tag{3.15}$$

Define

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ a_1 & 1 & \ddots & \vdots & a'_1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \ddots & 0 \\ a_{t-1} & a_{t-2} & \cdots & 1 & a'_{t-1} & a'_{t-2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{4t-2} & a_{4t-3} & \cdots & a_{3t-1} & a'_{4t-2} & a'_{4t-3} & \cdots & a'_{3t-1} \end{pmatrix}, \tag{3.16}$$

and

$$\mathbf{b} = \begin{pmatrix} a'_1, \cdots, a'_{4t-1} \end{pmatrix}^T - \begin{pmatrix} a_1, \cdots, a_{4t-1} \end{pmatrix}^T. \tag{3.17}$$

The following Lemma 8 provides an equivalent linear equation to find a solution that satisfies (3.13), and Theorem 3 shows how to compute π from this intermediate value.

Lemma 8. *Consider the following equation:*

$$\mathbf{Ac} = \mathbf{b}. \tag{3.18}$$

For any vector $\mathbf{c} \in \mathbb{F}_q^{2t}$, $\mathbf{c}$ is a solution to (3.18) iff $(h_1(\mathbf{c}), h_2(\mathbf{c}))$ satisfies (3.13).

Proof. The proof is in Appendix F. □

Theorem 3. *Let $\mathbf{c}$ be an arbitrary solution to (3.18), and $h_1 = h_1(\mathbf{c})$, $h_2 = h_2(\mathbf{c})$. Denote h, v_1, v_2 as follows.*

$$h = GCD(h_1, h_2), v_1 = \frac{h_1}{h},\ v_2 = \frac{h_2}{h}. \tag{3.19}$$

Suppose V_1, V_2 are the sets of the additive inverses of roots of v_1, v_2, respectively. Then π can be computed from the following equation:

$$\pi = \nu^{-1}\left(V_2 \cup (B' \setminus V_1)\right).$$

Proof. Note that $B = \nu(\pi)$ and g is an injection, we only need to prove that $B = V_2 \cup (B' \setminus V_1)$. From (3.12) it follows that

$$h_1 \cdot f_1 - h_2 \cdot f_2 = (h_1 \cdot g_1 - h_2 \cdot g_2) \cdot g_3,$$

where $\deg g_3 = |B \cap B'| \geq N - 1 - t$. From Lemma 8 we know that (3.13) is true, therefore

$$h_1 \cdot f_1 = h_2 \cdot f_2.$$

We know from (3.19) that

$$v_1 \cdot f_1 = v_2 \cdot f_2,$$

where $\mathrm{GCD}(v_1, v_2) = 1$. Then we have

$$v_1 | f_2,\ v_2 | f_1,\ \frac{f_1}{v_2} = \frac{f_2}{v_1} = f.$$

Suppose V_3 is the set of the additive inverses of roots of f. Then $V_2 \cup V_3 = B$, $V_1 \cup V_3 = B'$, thus $B = V_2 \cup V_3 = V_2 \cup (B' \setminus V_1)$. □

Note that V_1, V_2 computed in Theorem 3 are exactly identical to D_2, D_1 described before (3.12), respectively.

Example 3. *Suppose the sender transmits the permutation* $\pi_1 = (2, 4, 7, 3, 5, 1, 8, 6, 9, 10) \in \mathcal{C}_{\boldsymbol{\alpha}}(10, 2)$*, where* $\boldsymbol{\alpha} = (16, 0, 86, 44, 61, 9, 49)$*, and the receiver recives* $\pi' = (8, 6, 9, 10, 5, 1, 2, 4, 7, 3) \in \mathbb{S}_{10}$*. In the encoding scheme,* $q = 97 > 10^2 - 10$*, and for all* $i \neq j \in [10]$*,*

$$\upsilon(i, j) = 10(i - 1) + j - 1.$$

The receiver can apply Newton's Identities *to compute* $r(B) = (16, 31, 0, 42, 54, 94, 59)$ *from* $\boldsymbol{\alpha}$*, and derive* $r(B') = (80, 64, 83, 10, 72, 22, 26)$ *from* $B' = \nu(\pi') = \{75, 58, 89, 94, 40, 1, 13, 36, 62\}$.

Then

$$\mathbf{A} = \begin{pmatrix} 1 & 16 & 31 & 0 & 42 & 54 & 94 \\ 0 & 1 & 16 & 31 & 0 & 42 & 54 \\ 1 & 80 & 64 & 83 & 10 & 72 & 22 \\ 0 & 1 & 80 & 64 & 83 & 10 & 72 \end{pmatrix}^T, \tag{3.20}$$
$$\mathbf{b} = \begin{pmatrix} 64 & 33 & 83 & 65 & 18 & 25 & 64 \end{pmatrix}^T.$$

Notice that $\mathbf{c} = \begin{pmatrix} 95, 94, 66, 26 \end{pmatrix}$ *is a solution to* $\mathbf{Ac} = \mathbf{b}$. *Therefore* $h_1 = X^2+95X+94 = (X+1)(X+94)$, $h_2 = X^2+31X+71 = (X+24)(X+7)$. *The receiver then knows that* $D_1 = \{1, 94\}$, $D_2 = \{24, 7\}$. *Therefore* $\nu(\pi) = B = D_2 \cup (B' \backslash D_1) = \{13, 36, 62, 24, 40, 7, 75, 58, 89\}$. *Then it follows that* $A(\pi) = \{(2, 4), (4, 7), (7, 3), (3, 5), (5, 1), (1, 8), (8, 6), (6, 9), (9, 10)\}$. *From the definition of* f, *the receiver is able to decode* π *from* $A(\pi)$ *as* $\pi = (2, 4, 7, 3, 5, 1, 8, 6, 9, 10)$.

CHAPTER 4

Systematic Permutation Codes in the Generalized Cayley Metric

We presented a coding scheme for an order-optimal t-block permutation code in Chapter 3. However, we observe that it is difficult to identify a bijection between the transmitted messages and the codewords in the non-systematic codes in this configuration. We now develop order-optimal t-block permutation codes in the systematic form. We first provide in Section 4.1 a general construction of a systematic t-block permutation code that is not necessarily order-optimal. Then in Section 4.2, we provide the decoding scheme of this construction. We next prove the existence of an order-optimal version of this code and provide a specific construction of the systematic order-optimal code in Section 4.3. Finally, we compare the rate of our schemes with the existing constructions and prove that our codes have higher rates.

4.1 Encoding Scheme

Let messages be permutations in $\mathbb{S}_N$. In systematic codes, the codewords are permutations of length $N+M$. In our configuration, we derive each codeword $\sigma \in \mathbb{S}_{N+M}$ from a message $\pi \in \mathbb{S}_N$ by sequentially inserting values $N+1, N+2, \cdots, N+M$ into π in the positions specified by a sequence $S = (s_1, s_2, \cdots, s_M)$. We then prove in Lemma 9 that the block permutation distance between the resulting codewords cannot be smaller than that of their

original permutations. It follows from Theorem 2 that permutations with the same parity check sum $\alpha^{(q,2t)}$ defined in equation (4.2) have a block permutation distance of at least $2t+1$. Therefore, it suffices to show that the permutations with different parity check sums map to codewords that are sufficiently far apart under the block permutation distance.

In this section, we present an explicitly constructed encoding method in Theorem 2 based on a so-called *t-auxiliary set* we introduce in Definition 9. We start by presenting a collection of lemmas and definitions to support our results.

Definition 5. *For any permutation $\pi \in \mathbb{S}_N$ and the integer $i \in \mathbb{N}$, where $1 \leq s \leq N$, let $E(\pi, s)$ be a permutation in $\mathbb{S}_{N+1}$ derived by inserting the element $N+1$ after the element s in π, i.e.,*

$$E(\pi, s) \triangleq (\pi_1, \pi_2, \cdots, \pi_k, N+1, \pi_{k+1}, \cdots, \pi_N),$$

*where $k = \pi^{-1}(s)$. We call $E(\pi, s)$ the **extension** of π on the **extension point** s.*

*Consider a sequence $S = (s_1, s_2, \cdots, s_M)$, where $s_m \in [N]$ for all $1 \leq m \leq M$. The **extension** $E(\pi, S)$ of π on the **extension sequence** S is a permutation in $\mathbb{S}_{N+M}$ derived from inserting the elements $N+1, \cdots, N+M$ sequentially after the elements $s_1, \cdots, s_M$ in π, i.e.,*

$$E(\pi, S) \triangleq E(E(\cdots, E(E(\pi, s_1), s_2), \cdots, s_{M-1}), s_M).$$

Note that in Definition 5, the elements $s_1, \cdots, s_M$ in the extension sequence S are not necessarily distinct. If different symbols are sequentially inserted after the same element, then they are all placed right after this element in descending order, as is shown in the following example.

Example 4. *Suppose $\pi = (1, 4, 5, 7, 6, 2, 3)$, $I = \{4, 1, 2, 2\}$, then*

$$E(\pi, I) = (1, 9, 4, 8, 5, 7, 6, 2, 11, 10, 3).$$

The next Definition 6 presents the notion of the *jump points* of the extensions of two

permutations. Then Lemma 9states that the block permutation distance between two extensions is strictly larger than that of the original permutations when the extension point of one of them is a jump point. Based on this result, we further introduce the notion of *jump index* and *jump set* in Definition 7. We know that the block permutation distance of two permutations from $\mathbb{S}_N$ is lower bounded by the sum of that of their extensions and the cardinality of the jump set.

Definition 6. *Suppose $E(\pi_1, s_1)$, $E(\pi_2, s_2)$ are two arbitrary extensions of π_1 and π_2, respectively, where $\pi_1, \pi_2 \in \mathbb{S}_N$, $\pi_{1,k_1} = s_1$ and $\pi_{2,k_2} = s_2$. Then s_1 is called a **jump point** of $E(\pi_1, s_1)$ with respect to $E(\pi_2, s_2)$, if $s_1 \neq s_2$ and at least one of the following conditions is satisfied:*

1. *$k_1 = N$ or $k_2 = N$;*
2. *$k_1, k_2 < N$, and $\pi_{1,k_1+1} \neq \pi_{2,k_2+1}$.*

Lemma 9. *For any two extensions $E(\pi_1, s_1)$ and $E(\pi_2, s_2)$, if s_1 is a jump point, then*

$$d_B(E(\pi_1, s_1), E(\pi_2, s_2)) > d_B(\pi_1, \pi_2), \tag{4.1}$$

else

$$d_B(E(\pi_1, s_1), E(\pi_2, s_2)) = d_B(\pi_1, \pi_2). \tag{4.2}$$

Proof. The proof is in Appendix G. □

In the following Example 5, we provide examples of jump points that satisfy the two conditions indicated in the previous Definition 6. We also provide an example of an extension point that is not a jump point.

Example 5. *Suppose $\pi = (1, 5, 7, 2, 3, 6, 4)$, $\pi' = (2, 3, 1, 5, 7, 6, 4)$, $s_1 = 4$, $s_1' = 5$, $s_2 = 5$,*

$s_2' = 6$, $s_3 = 3$, $s_3' = 7$. *Then*

$$\begin{aligned}
\sigma_1 &= E(\pi, s_1) = (1, 5, 7, 2, 3, 6, 4, 8)\,, \\
\sigma_1' &= E(\pi', s_1') = (2, 3, 1, 5, 8, 7, 6, 4)\,, \\
\sigma_2 &= E(\pi, s_2) = (1, 5, 8, 7, 2, 3, 6, 4)\,, \\
\sigma_2' &= E(\pi', s_2') = (2, 3, 1, 5, 7, 6, 8, 4)\,, \\
\sigma_3 &= E(\pi, s_3) = (1, 5, 7, 2, 3, 8, 6, 4)\,, \\
\sigma_3' &= E(\pi', s_3') = (2, 3, 1, 5, 7, 8, 6, 4)\,.
\end{aligned}$$

Given that $d_B(\pi, \pi') = 2$, *we observe by inspection that*

$$\begin{aligned}
&d_B(\sigma_1, \sigma_1') = 4 > d_B(\pi, \pi'),\ s_1 \textit{ is a jump point;} \\
&d_B(\sigma_2, \sigma_2') = 5 > d_B(\pi, \pi'),\ s_2 \textit{ is a jump point;} \\
&d_B(\sigma_3, \sigma_3') = 2 = d_B(\pi, \pi'),\ s_3 \textit{ is not a jump point.}
\end{aligned}$$

Notice that s_1 *is a jump point that satisfies the first condition in Definition 6 and* s_2 *satisfies the second condition. This example is consistent with Lemma 9.*

Definition 7. *Suppose* $E(\pi_1, S_1)$ *and* $E(\pi_2, S_2)$ *are extensions of* π_1 *and* π_2 *on extension sequences* S_1 *and* S_2, *respectively, where* $\pi_1, \pi_2 \in \mathbb{S}_N$, $S_1 = (s_{1,1}, s_{1,2}, \cdots, s_{1,M})$ *and* $S_2 = (s_{2,1}, s_{2,2}, \cdots, s_{2,M})$. *Then, for any* $m \in [M]$, m *is called a* ***jump index*** *of* $E(\pi_1, S_1)$ *and* $E(\pi_2, S_2)$ *if* $s_{1,m}$ *is a jump point of* $E(E(\pi_1, J_{1,m}), s_{1,m})$ *with respect to* $E(E(\pi_2, J_{2,m}), s_{2,m})$, *where* $J_{1,m} = (s_{1,1}, s_{1,2}, \cdots, s_{1,m-1})$, $J_{2,m} = (s_{2,1}, s_{2,2}, \cdots, s_{2,m-1})$. *Define the* ***jump set*** $F(\pi_1, \pi_2, S_1, S_2)$ *as the set of all jump indices of* $E(\pi_1, S_1)$ *and* $E(\pi_2, S_2)$.

Remark 5. *For any extensions* $E(\pi_1, S_1)$, $E(\pi_2, S_2)$ *of* π_1, π_2 *on extension sequences* S_1, S_2, *respectively, it is obvious from the above Definition 7 and Lemma 9 that*

$$d_B(E(\pi_1, S_1), E(\pi_2, S_2)) \geq d_B(\pi_1, \pi_2) + |F(\pi_1, \pi_2, S_1, S_2)|. \tag{4.3}$$

In the following Example 6, we provide an example of how to identify the jump indices and compute the jump set. This example satisfies inequality (4.3).

Example 6. *Continuing with the values of π, π' specified in Example 5, let $S = (4, 6, 7)$, $S' = (5, 6, 5)$. Then*

$$\begin{aligned}
\sigma_0 = \pi &= (1, 5, 7, 2, 3, 6, 4)\,, \\
\sigma_0' = \pi' &= (2, 3, 1, 5, 7, 6, 4)\,, \\
\sigma_1 = E(\sigma_0, s_1) &= (1, 5, 7, 2, 3, 6, 4, 8)\,, \\
\sigma_1' = E(\sigma_0', s_1') &= (2, 3, 1, 5, 8, 7, 6, 4)\,, \\
\sigma_2 = E(\sigma_1, s_2) &= (1, 5, 7, 2, 3, 6, 9, 4, 8)\,, \\
\sigma_2' = E(\sigma_1', s_2') &= (2, 3, 1, 5, 8, 7, 6, 9, 4)\,, \\
\sigma_3 = E(\sigma_2, s_3) &= (1, 5, 7, 10, 2, 3, 6, 9, 4, 8)\,, \\
\sigma_3' = E(\sigma_2', s_3') &= (2, 3, 1, 5, 10, 8, 7, 6, 9, 4)\,.
\end{aligned}$$

It follows immediately that

$$\begin{aligned}
&d_B(\sigma_0, \sigma_0') = 2, \\
&d_B(\sigma_1, \sigma_1') = 4 > d_B(\sigma_0, \sigma_0'),\ 1 \textit{ is a jump index}; \\
&d_B(\sigma_2, \sigma_2') = 4 = d_B(\sigma_1, \sigma_1'),\ 2 \textit{ is not a jump index}; \\
&d_B(\sigma_3, \sigma_3') = 5 > d_B(\sigma_2, \sigma_2'),\ 3 \textit{ is a jump index}.
\end{aligned}$$

According to Definition 7, $F(\pi, \pi', S, S') = \{1, 3\}$. Moreover, $d_B(\sigma, \sigma') = d_B(\sigma_3, \sigma_3') = 5 > 4 = d_B(\pi, \pi') + |F(\pi, \pi', S, S')|$, which is in accordance with equation (4.3).

Next we prove in Lemma 10 that the right hand expression of equation (4.3) can be lower bounded by the cardinality of so-called *Hamming set* defined in the following Definition 8. Based on this result, we present a construction of systematic t-block permutation codes in Theorem 4 with the help of a so-called *t-auxiliary set*.

Definition 8. *For any sequences* $\mathbf{v}_1$, $\mathbf{v}_2$ *of integers, with length* M, *where* $\mathbf{v}_1 = (v_{1,1}, v_{1,2}, \cdots, v_{1,M})$ *and* $\mathbf{v}_2 = (v_{2,1}, v_{2,2}, \cdots, v_{2,M})$, *define the* ***Hamming set*** *of* $\mathbf{v}_1$ *with respect to* $\mathbf{v}_2$ *as follows,*

$$H(\mathbf{v}_1, \mathbf{v}_2) \triangleq \{v_{1,m} | v_{1,m} \neq v_{2,m}, m \in [M]\}. \tag{4.4}$$

Remark 6. *It is obvious that* $d_H(\mathbf{v}_1, \mathbf{v}_2) \geq |H(\mathbf{v}_1, \mathbf{v}_2)|$. *Additionally, for any three sequences* $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ *of integers, the following triangle inequality holds true:*

$$|H(\mathbf{v}_1, \mathbf{v}_3)| \leq |H(\mathbf{v}_1, \mathbf{v}_2)| + |H(\mathbf{v}_2, \mathbf{v}_3)|. \tag{4.5}$$

Lemma 10. *For any extensions* $E(\pi_1, S_1)$, $E(\pi_2, S_2)$ *of* π_1, π_2 *on extension sequences* S_1, S_2, *respectively, it follows that*

$$d_B(E(\pi_1, S_1), E(\pi_2, S_2)) \geq |H(S_1, S_2)|. \tag{4.6}$$

Proof. The proof is in Appendix H. □

Example 7. *Continuing on with the numerical values of* π, π', S, S' *as in Example 6, we conclude that,* $H(S, S') = \{4, 7\}$, $m(4) = 1$, $m(7) = 3$. *Then it follows that* $d_B(\sigma, \sigma') = 5 > 2 = |H(S, S')|$, *which is in accordance with the above Lemma 10.*

Definition 9. *Consider a set* $\mathcal{A}(N, K, t) \subset [N]^K$. *We call* $\mathcal{A}(N, K, t)$ *a* ***t*-auxiliary set** *of length* K *in range* $[N]$ *if for any* $\mathbf{c}_1 \neq \mathbf{c}_2 \in \mathcal{A}(N, K, t)$, $|H(\mathbf{c}_1, \mathbf{c}_2)| \geq 2t + 1$ *holds.*

Theorem 4. *For any* t*-auxiliary set* $\mathcal{A}(N, K, t)$ *with cardinality that is no less than* q^{4t-1}, *suppose* $\varphi : \ \alpha^{(q,2t)}(\mathbb{S}_N) \to \mathcal{A}(N, K, t)$ *is an arbitrary injection, where* q *is a prime number such that* $N^2 - N < q < 2(N^2 - N)$ *and the parity check sum* $\alpha^{(q,2t)}$ *is defined in equation (4.2). Then, the set* $\mathcal{C}_B^{\text{sys}}(N, K, t) = \{E(\pi, \varphi \circ \alpha^{(q,2t)}(\pi)) | \pi \in \mathbb{S}_N\}$ *is a systematic* t*-block permutation code.*

Proof. It is clear by choice of $E(\pi, S)$ that $\mathcal{C}_B^{\text{sys}}(N, K, t)$ is systematic. For any arbitrary two

messages $\pi_1, \pi_2 \in \mathbb{S}_N$, denote their corresponding codeword by $\sigma_1 = E(\pi_1, \varphi \circ \alpha^{(q,2t)}(\pi_1))$ and $\sigma_2 = E(\pi_2, \varphi \circ \alpha^{(q,2t)}(\pi_2))$, respectively. Suppose $\boldsymbol{\alpha}_1 = \alpha^{(q,2t)}(\pi_1)$, $\boldsymbol{\alpha}_2 = \alpha^{(q,2t)}(\pi_2)$, $S_1 = \varphi(\boldsymbol{\alpha}_1)$ and $S_2 = \varphi(\boldsymbol{\alpha}_2)$. Then $\sigma_1 = E(\pi_1, S_1)$, $\sigma_2 = E(\pi_2, S_2)$. Consider the following two cases:

1. $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2$. In this case, from Theorem 2 we know that $d_B(\pi_1, \pi_2) > 2t$. Then Lemma 9 implies that $d_B(\sigma_1, \sigma_2) \geq d_B(\pi_1, \pi_2) \geq 2t + 1$.

2. $\boldsymbol{\alpha}_1 \neq \boldsymbol{\alpha}_2$. In this case, $S_1 \neq S_2 \in \mathcal{A}(N, K, t)$. Then from Definition 9, $|H(S_1, S_2)| \geq 2t + 1$. Therefore from Lemma 10, $d_B(\sigma_1, \sigma_2) \geq |H(S_1, S_2)| \geq 2t + 1$.

From the above discussion, $d_B(\sigma_1, \sigma_2) \geq 2t+1$ is aways true, which means that $\mathcal{C}_B^{\text{sys}}(N, K, t)$ is indeed a systematic t-block permutation code. □

4.2 Decoding Scheme

Based on the construction and the notation in Theorem 4, suppose the sender sends a codeword $\sigma = E(\pi, \varphi \circ \alpha^{(q,2t)}(\pi))$ through the channel and the receiver receives a noisy version σ', where $d_B(\sigma, \sigma') \leq t$.

In this section, we prove in the forthcoming Lemma 11 that the extension sequence S of the codeword $E(\pi, S)$ is decodable given that $d_B(\sigma, \sigma') \leq t$, from which the parity check sum defined in (4.2) of the transmitted information π can be derived.

For convenience, we introduce the following definition of *truncation* and will use it throughout this section.

Definition 10. *For any permutation $\sigma \in \mathbb{S}_{N+1}$ and an integer $u \in [N+1]$, denote $T(\sigma, u)$ to be the sequence derived by removing the element u from σ, i.e.,*

$$T(\sigma, u) \triangleq (\sigma_1, \sigma_2, \cdots, \sigma_{k-1}, \sigma_{k+1}, \cdots, \sigma_N), \tag{4.7}$$

where $k = \pi^{-1}(u)$.

Then, for any permutation $\sigma \in \mathbb{S}_{N+M}$ and a set $U \subset [N+M]$, denote the ***truncation*** *$T(\sigma, U)$ of σ on set U to be the sequence derived by removing the elements contained in $U = \{u_1, u_2, \cdots, u_M\}$ from σ, i.e.,*

$$T(\sigma, U) \triangleq T(T(\cdots, T(T(\sigma, u_1), u_2), \cdots, u_{|U|-1}), u_{|U|}). \tag{4.8}$$

Note that in this Definition 10, the ordering of $u_1, \cdots, u_{|U|}$ has no impact on the value of $T(\sigma, U)$. The following is the example of the truncation of a permutation.

Example 8. *Suppose $\sigma = (1, 4, 5, 2, 3, 9, 8, 6, 7)$, $U = \{4, 5, 9\}$, then*

$$T(\sigma, U) = (1, 2, 3, 8, 6, 7)\,.$$

Our decoding scheme has two major steps. Recall that $\alpha^{(q,2t)}$ is defined in as the parity check sum of π. The first is to compute the parity check sum $\boldsymbol{\alpha} = \alpha^{(q,2t)}(\pi)$ of $\pi = T(\sigma, \{N+1, \cdots, N+M\})$ from σ'. The second step is to apply the decoding algorithm proposed in Theorem 3 to the subsequence $\pi' = T(\sigma', \{N+1, \cdots, N+M\})$ and compute π.

The following Lemma 11 proves the decodability of the sequence S from S', where S is the extension sequence of π in σ, by showing that the cardinality of the Hamming set $H(S, S')$ does not exceed t. Therefore from equation (4.5) and Definition 9, we are able to compute S from S' since each t-auxiliary set $\mathcal{A}(N, K, t)$ has the property that cardinalities of Hamming sets constructed from its pairwise distinct elements are at least $2t+1$. The parity check sum $\boldsymbol{\alpha}$ is then uniquely derived from S.

Lemma 11. *Consider an arbitrary $\sigma \in \mathcal{C} = \{E(\pi, \varphi \circ \alpha^{(q,2t)}(\pi)) | \pi \in \mathbb{S}_N\}$, for $\mathcal{C}$ defined in Theorem 2 (then $\sigma \in \mathbb{S}_{N+k}$). Suppose there is a σ' such that $d_B(\sigma, \sigma') \leq t$. Let $S = \varphi \circ \alpha^{(q,2t)}(\pi)$ and $\pi' = T(\sigma', [N+1 : N+M])$. Suppose σ' is the extension of π' on the extension*

sequence S', i.e., $\sigma' = E(\pi', S')$, then

$$H(S, S') \leq t. \tag{4.9}$$

Proof. Suppose $S = (s_1, s_2, \cdots, s_k)$, $S' = (s'_1, s'_2, \cdots, s'_k)$. From Theorem 4, we know that $S \in \mathcal{A}(N, k, t)$. Denote $\mathcal{M} = \{m | s_m \neq s'_m, 1 \leq m \leq k\}$. Then for all $m \in \mathcal{M}$, $s_m \neq s'_m$, and there exists $k(m), k'(m) \in [k]$, integers $n_1, n_2, \cdots, n_{k(m)}$, and $n'_1, n'_2, \cdots, n'_{k(m)'} \in [N+1 : N+M]$ such that the sequences $\mathbf{p}_m = (s_m, n_{k(m)}, n_{k(m)-1}, \cdots, n_1, N+m)$ and $\mathbf{p}'_m = (s'_m, n'_{k(m)'}, n'_{k(m)'-1}, \cdots, n'_1, N+m)$ are subsequences of σ, σ', respectively. Note that $s_m \neq s'_m$, which means that $(s_m, n_{k(m)}, n_{k(m)-1}, \cdots, n_1) \neq (s'_m, n'_{k(m)'}, n'_{k(m)'-1}, \cdots, n'_1)$. Let

$$i(m) = \min_{\substack{1 \leq i \leq \min\{k(m), k(m)'\} \\ n_i \neq n'_i}} i.$$

Then $n_{i(m)} \neq n'_{i(m)}$ and $n_{i(m)-1} = n'_{i(m)-1}$, where we let $n_0 = n'_0 = N + m$ if $i(m) = 1$.

Recall the notion of characteristic set in Definition 3. We know that $(n_{i(m)}, n_{i(m)-1}) \in A(\sigma)$, $(n'_{i(m)}, n'_{i(m)-1}) \in A(\sigma')$. These two conditions $n_{i(m)} \neq n'_{i(m)}$ and $n_{i(m)-1} = n'_{i(m)-1}$ imply that $(n_{i(m)}, n_{i(m)-1}) \in (A(\sigma) \setminus A(\sigma'))$ for all $m \in \mathcal{M}$. Notice that for all $s_m \in \{s_m : m \in \mathcal{M}\} = H(S, S')$, the associated subsequences $\mathbf{p}_m$ start with different s_m and thus they do not overlap, which indicates that the pairs $(n_{i(m)}, n_{i(m)-1})$ are distinct. Then $|A(\sigma) \setminus A(\sigma')| \geq |H(S, S')|$, which is equivalent to $H(S, S') \leq d_B(\sigma, \sigma') \leq t$. □

From Lemma 11, the receiver first computes $\pi' = T(\sigma', \{N+1, \cdots, N+k\})$ and derives the extension sequence S' such that $\sigma' = E(\pi', S')$. Then, the receiver decodes $\hat{S} = \varphi \circ \alpha^{(q,2t)}(\pi) \in \mathcal{A}(N, K, t)$ from S' such that $|H(S, \hat{S})| \leq t$ and derives $\boldsymbol{\alpha}$ from S. From Lemma 9, we know that $d_B(\pi, \pi') \leq d_B(\sigma, \sigma') \leq t$. Then, the receiver can apply the decoding algorithm described in Section 3.1 to compute π from π' and $\boldsymbol{\alpha}$ reliably. The decoding scheme for the systematic t-block permutation code $\mathcal{C}$ constructed in Theorem 4 is then complete.

4.3 Order-optimal Systematic t-Block Permutation Codes

Theorem 4 presents the construction of systematic t-block permutation codes with K redundant symbols based on a t-auxiliary set $\mathcal{A}(N, K, t)$. When N is sufficiently large and K is relatively small compared to N, the code rate is $1 - \mathcal{O}(\frac{K}{N})$, which is not necessarily order-optimal. In the forthcoming Theorem 5, we prove the existence of an order-optimal systematic t-block permutation code. Theorem 5 is based on Lemma 12 and Lemma 13, where we prove the existence of a t-auxiliary set $\mathcal{A}(N, K, t)$ with length $K = \mathcal{O}(t)$ when t is sufficiently small compared to N. We further provide a construction of $\mathcal{A}(N, 56t, t)$ in Theorem 7 based on Lemma 14 and Theorem 6. Then the permutation codes generated from this set in Theorem 4 are order-optimal.

Lemma 12. (cf. [Vad12, Problem 3.2]) *For any integers $N, l, a, m \in \mathbb{N}^*$, where $l \leq N$, $a \leq l$, if $m \leq \frac{\binom{N}{a}}{\binom{l}{a}^2}$, then there exists a set $\mathcal{L}(N, l, a) = \{L_1, L_2, \cdots, L_m\}$ of m subsets of $[N]$ such that for all $1 \leq i \leq M$,*

1. *$|L_i| = l$.*
2. *$\forall j \neq i$ and $j \in [M]$, $|L_i \cap L_j| < a$.*

We call $\mathcal{L}(N, l, a)$ that satisfies the above two conditions a ***(N, l, a)-set*** *with cardinality m.*

Proof. The proof is in Appendix I. □

Lemma 13. *For all $t, k, N \in \mathbb{N}^*$, if $N > 2k$ and $\frac{14+2c}{1-c}t \leq k < N^c$ for some constant $c \in \mathbb{R}$, $0 < c < 1$, there exists a t-auxiliary set $\mathcal{A}(N, k, t)$ with cardinality no less than q^{4t-1}, where q is a prime number such that $N^2 - N < q < 2(N^2 - N)$.*

Proof. Let $l = k$, $a = k - 2t$, $m = q^{4t-1}$ in Lemma 12, we know that if $q^{4t-1} < \frac{\binom{N}{k-2t}}{\binom{k}{k-2t}^2}$, we can always find an (N, l, a)-set $\mathcal{L}(N, l, a)$ with cardinality m.

Notice that for all integers k, t such that $\frac{14+2c}{1-c}t \leq k < N^c$ for some constant $0 < c < 1$,

$$\begin{aligned}
&\binom{N}{k-2t} \Big/ \binom{k}{k-2t}^2 \\
=&\frac{N(N-1)\cdots(N-k+2t+1)/(k-2t)!}{(k(k-1)\cdots(k-2t+1))^2/((2t)!)^2} \\
=&\frac{N(N-1)\cdots(N-k+2t+1)((2t)!)^2}{(k(k-1)\cdots(k-2t+1))^2(k-2t)!} \\
>&\frac{N(N-1)\cdots(N-k+6t+1)}{(k-2t)!} \\
>&\frac{N(N-1)\cdots(N-k+6t+1)}{2^{(k-2t+1)\log(k-2t)-k+2t+2}} \\
>&\frac{N(N-1)\cdots(N-k+6t+1)2^{k-2t}}{(k-2t)^{(k-2t)}} \\
>&\frac{N^{k-6t}2^{4t}}{N^{c(k-2t)}} \\
=&N^{(1-c)k-(6+2c)t}2^{4t} \\
>&N^{8t}2^{4t} = (2N^2)^{4t} > q^{4t-1}.
\end{aligned}$$

From the above equation and Lemma 12, such an $(N, k, k-2t)$-set $\mathcal{L}(N, k, k-2t)$ with cardinality m exists. In this set, for all $i \neq j \leq q^{4t-1}$, $|L_i \cap L_j| < k-2t$. For each $L_i \in \mathcal{L}(N, k, k-2t) = (L_1, \cdots, L_m)$, $1 \leq i \leq q^{4t-1} \leq m$, let $\mathbf{c}_i$ be an arbitrary permutation of elements in L_i. Let $\mathcal{A}(N, k, t) = \{\mathbf{c}_i | 1 \leq i \leq q^{4t-1}\}$. Then $|H(\mathbf{c}_i, \mathbf{c}_j)| = d_H(\mathbf{c}_i, \mathbf{c}_j) \geq k - |L_i \cap L_j| > k - (k-2t) = 2t$ for all $1 \leq i < j \leq q^{4t-1}$. Therefore $\mathcal{A}(N, k, t)$ is a t-auxiliary set. The lemma is proved. □

Theorem 5. *There exists a systematic t-block permutation code with $30t$ redundant symbols when $t < \frac{1}{30}\sqrt{N}$.*

Proof. Let $c = \frac{1}{2}$ in Lemma 13, then there exists a t-auxiliary set $\mathcal{A}(N, 30t, t)$ with at least q^{4t-1} elements. From Theorem 4, the code $\mathcal{C}_B^{\text{sys}}(N, t) = \{E(\pi, \varphi \circ \alpha^{(q,2t)}(\pi)) | \pi \in \mathbb{S}_N\}$ based on $\mathcal{A}(N, 30t, t)$ is a systematic t-block permutation code with $k = 30t$ redundant symbols. □

Remark 7. *A code that satisfies Theorem 5 is order-optimal when N is sufficiently large.*

Based on the upcoming Lemma 14 and Theorem 6, we provide an explicit construction of a t-auxiliary set of length $56t$ in Theorem 7, from which we are able to explicitly construct an order-optimal permutation code by Theorem 4.

Lemma 14. *For all $k, N \in \mathbb{N}^*$, $k > 3$, $N > k^2$, consider an arbitrary subset $Y \subset [k]$, where $|Y| = M < k$, $Y = \{i_1, i_2, \cdots, i_M\}$, then*

$$\mathrm{LCM}\left(N + i_1, N + i_2, \cdots, N + i_M\right) > N^{M - \frac{k}{2}}. \tag{4.10}$$

Proof. The proof is in Appendix J. □

Theorem 6. *For all $N, k, d \in \mathbb{N}^*$, $N > k^2$, $k > 3$, define a function $\beta^{(q,d,k)} : [q]^d \to [N+1] \times [N+2] \times \cdots \times [N+k]$ as below:*

$$\begin{aligned} \beta^{(q,d,k)}(\boldsymbol{x}) &= \left(\beta_1^{(q,d,k)}(\boldsymbol{x}), \beta_2^{(q,d,k)}(\boldsymbol{x}), \cdots, \beta_d^{(q,d,k)}(\boldsymbol{x})\right) \\ &\triangleq (\gamma(\boldsymbol{x}) \bmod (N+1), \gamma(\boldsymbol{x}) \bmod (N+2), \\ &\quad \cdots, \gamma(\boldsymbol{x}) \bmod (N+k)). \end{aligned} \tag{4.11}$$

where $\boldsymbol{x} = (x_1, x_2, \cdots, x_d) \in [q]^d$, $\gamma(\boldsymbol{x}) \triangleq \sum_{i=1}^{d} x_i q^{i-1}$. Then $\forall\ \boldsymbol{x}_1, \boldsymbol{x}_2 \in [q]^d$, $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$,

$$d_H(\beta^{(q,d,k)}(\boldsymbol{x}_1), \beta^{(q,d,k)}(\boldsymbol{x}_2)) > \frac{k}{2} - d(2 + \log_N 2). \tag{4.12}$$

Proof. For arbitrary $\boldsymbol{x}_1, \boldsymbol{x}_2 \in [q]^d$, $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, let $\boldsymbol{\beta}_1 = \beta^{(q,d,k)}(\boldsymbol{x}_1)$, $\boldsymbol{\beta}_2 = \beta^{(q,d,k)}(\boldsymbol{x}_2)$. Let $Z = \{i :\ \beta_{1,i} = \beta_{2,i},\ 1 \leq i \leq d\}$, then $d_H(\beta^{(q,d,k)}(\boldsymbol{x}_1), \beta^{(q,d,k)}(\boldsymbol{x}_2)) = k - |Z| = k - M$, where $M = |Z|$.

Suppose $Z = \{i_1, i_2, \cdots, i_M\}$. Let $\gamma_1 = \boldsymbol{\gamma}(\boldsymbol{x}_1)$, $\gamma_2 = \boldsymbol{\gamma}(\boldsymbol{x}_2)$. From the definition of $\beta^{(q,d,k)}$,

we know that

$$\begin{cases} \gamma_1 \equiv \gamma_2 \mod (N+i_1) \\ \gamma_1 \equiv \gamma_2 \mod (N+i_2) \\ \vdots \\ \gamma_1 \equiv \gamma_2 \mod (N+i_M). \end{cases}$$

Then,

$$\gamma_1 \equiv \gamma_2 \mod \mathrm{LCM}\left(N+i_1, N+i_2, \cdots, N+i_M\right).$$

Given that $\boldsymbol{x}_1, \boldsymbol{x}_2 \in [q]^d$, $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, then $\gamma_1 \neq \gamma_2$. From Lemma 14 we know that

$$|\gamma_1 - \gamma_2| \geq \mathrm{LCM}\left(N+i_1, N+i_2, \cdots, N+i_M\right) > N^{M-\frac{k}{2}}.$$

Moreover, we know from $\boldsymbol{x}_1, \boldsymbol{x}_2 \in [q]^d$, $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, that $0 \leq \gamma_1, \gamma_2 < q^d$ and $\gamma_1 \neq \gamma_2$. Therefore,

$$|\gamma_1 - \gamma_2| < q^d.$$

According to the above two inequalities, $N^{M-\frac{k}{2}} < |\gamma_1 - \gamma_2| < q^d < (2N^2)^d$ is true, which means that $M - \frac{k}{2} < d(2 + \log_N 2)$. Therefore $M < \frac{k}{2} + d(2 + \log_N 2)$, and then

$$\begin{aligned} d_H(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= k - M > k - (\frac{k}{2} + d(2 + \log_N 2)) \\ &= \frac{k}{2} - d(2 + \log_N 2). \end{aligned}$$

The theorem is proved. □

Example 9. *Let* $k = 7$, $N = 50$, $d = 1$, $q = 2503$, $\boldsymbol{x}_1 = (280)$, $\boldsymbol{x}_2 = (1008)$, *then* $\gamma_1 = 280$,

$\gamma_2 = 1008$, *and*

$$\begin{aligned}\boldsymbol{\beta}_1 =&(280 \bmod 51, 280 \bmod 52, \cdots, 280 \bmod 57)\\ =&(25, 20, 15, 10, 5, 0, 52).\\ \boldsymbol{\beta}_2 =&(1008 \bmod 51, 1008 \bmod 52, \cdots, 1008 \bmod 57)\\ =&(39, 20, 1, 36, 18, 0, 39).\end{aligned}$$

Then $d_H(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = 5 > \frac{k}{2} - d(2 + \log_N 2)$, *which is in accordance with Theorem 6.*

Based on Theorem 6, we provide an explicit construction of a t-auxiliary set $\mathcal{A}(N, 56t, t)$ in the following Theorem 7.

Theorem 7. *For all* $N, k, t \in \mathbb{N}^*$, $k \geq 28t$, $k < \lfloor \sqrt{N} - \frac{1}{2} \rfloor$. *Suppose* $[q]^{4t-1} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{q^{4t-1}}\}$, *where* q *is a prime number such that* $N^2 - N < q < 2N^2 - 2N$. *For any* $s \in [q^{4t-1}]$, *suppose* $\boldsymbol{x}_s = (x_1, x_2, \cdots, x_{4t-1})$, *let* $\mathbf{c}_s = (c_1, c_2, \cdots, c_{2k})$, $\beta^{(q,4t-1,k)}(\boldsymbol{x}) = (\beta_1, \beta_2, \cdots, \beta_k)$ *for all* $1 \leq i \leq k$, *where* $\mathbf{c}_s$ *is defined as follows:*

$$\begin{cases} c_{2i} = (i-1)\lfloor \frac{N}{k} \rfloor + 1 + \left(\beta_i \mod \lfloor \frac{N}{k} \rfloor\right), \\ c_{2i-1} = (i-1)\lfloor \frac{N}{k} \rfloor + 1 + \lfloor \frac{\beta_i}{\lfloor \frac{N}{k} \rfloor} \rfloor. \end{cases} \tag{4.13}$$

Then $\mathcal{A}(N, 2k, t) = \{\mathbf{c}_s : \ s \in [q^{4t-1}]\}$ *is a* t*-auxiliary set with cardinality* q^{4t-1}.

Proof. For any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in [q]^{4t-1}$, $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, let $\boldsymbol{\beta}_1 = \beta^{(q,4t-1,k)}(\boldsymbol{x}_1)$, $\boldsymbol{\beta}_2 = \beta^{(q,4t-1,k)}(\boldsymbol{x}_2)$. Then from Theorem 6, we know that

$$\begin{aligned} d_H(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &> \frac{k}{2} - (4t-1)(2 + \log_N 2) \\ &> \frac{k}{2} - (12t - 3) > \frac{28t}{2} - 12t = 2t. \end{aligned}$$

In equation (4.13), let $m_i = (i-1)\lfloor \frac{N}{k} \rfloor + 1$. Notice that $(c_{2i-1} - m_i)\lfloor \frac{N}{k} \rfloor + (c_{2i} - m_i) = \beta_i$.

Given $\beta_i < N + k$ for all $1 \leq i < k$, and $k < \lfloor \sqrt{N} - \frac{1}{2} \rfloor$, then

$$\begin{aligned}
\lfloor \frac{N}{k} \rfloor^2 &> \left(\frac{N}{k} - 1 \right)^2 \geq \left(\frac{N}{\sqrt{N} - \frac{3}{2}} - 1 \right)^2 \\
&> \left(\sqrt{N} + \frac{3}{2} - 1 \right)^2 = \left(\sqrt{N} + \frac{1}{2} \right)^2 \\
&> N + \sqrt{N} > N + k > \beta_i.
\end{aligned}$$

Therefore $(c_{2i-1} - m_i, c_{2i} - m_i)$ is exactly the $\lfloor \frac{N}{k} \rfloor$-ary representation of β_i, for all $1 \leq i \leq k$.

Suppose $\boldsymbol{\beta}_1 = (\beta_{1,1}, \beta_{1,2}, \cdots, \beta_{1,k})$ and $\boldsymbol{\beta}_2 = (\beta_{2,1}, \beta_{2,2}, \cdots, \beta_{2,k})$. Let $Y = \{i : \ \beta_{1,i} \neq \beta_{2,i}, \ 1 \leq i \leq k\}$, then $|Y| = d_H(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. Notice that for all $i \in Y$, $\beta_{1,i} \neq \beta_{2,i}$, then either $c_{1,2i-1} - m_i \neq c_{2,2i-1} - m_i$ or $c_{1,2i} - m_i \neq c_{2,2i} - m_i$, which means that

$$|H(\mathbf{c}_1, \mathbf{c}_2) \cap \{c_{1,2i-1}, c_{1,2i}\}| \geq 1, \ i \in Y. \tag{4.14}$$

Notice that $(i-1)\lfloor \frac{N}{k} \rfloor < c_{1,2i-1}, c_{1,2i} \leq i \lfloor \frac{N}{k} \rfloor$, therefore

$$\{c_{1,2i-1}, c_{1,2i}\} \cap \{c_{1,2i'-1}, c_{1,2i'}\} = \emptyset, \ \forall \ 1 \leq i < i' \leq k. \tag{4.15}$$

From (4.14) and (4.15),

$$\begin{aligned}
|H(\mathbf{c}_1, \mathbf{c}_2)| &= \sum_{i=1}^{k} |H(\mathbf{c}_1, \mathbf{c}_2) \cap \{c_{1,2i-1}, c_{1,2i}\}| \\
&\geq \sum_{i \in Y} |H(\mathbf{c}_1, \mathbf{c}_2) \cap \{c_{1,2i-1}, c_{1,2i}\}| \\
&\geq \sum_{i \in Y} 1 = |Y| = d_H(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) > 2t.
\end{aligned}$$

From Definition 9, $\mathcal{A}(N, k, t)$ is indeed a t-auxiliary set. □

Remark 8. *Let $k = 28t$ in the above Theorem 7 to construct a t-auxiliary set $\mathcal{A}(N, 56t, t)$. Then the code $\mathcal{C}_B^{\text{sys}}(N, 56t, t)$ constructed in Theorem 4 based on $\mathcal{A}(N, 56t, t)$ is an order-*

optimal systematic t-block permutation codes.

4.4 Rate Analysis

In Chapter 3, we can construct a t-generalized Cayley code $\mathcal{C}_G(N,t) = \mathcal{C}_{\boldsymbol{\alpha}}(N,4t)$ with rate $R_G(N,t)$. In [CV14], a t-generalized Cayley code $A_{\rho_g C}(N,t)$ with rate $R_{\rho_g C}(N,t)$ was constructed.

We next compare the rates of these two codes in Lemma 15.

Lemma 15. *$R_G(N,t) > R_{\rho_g C}(N,t)$ when $t < \frac{N}{(16\log N+8)}$ for sufficiently large N.*

Proof. We know from [GYF+16, Appendix A] that:

$$\log|A_{\rho_g C}(N,t)| \leq N\log N - (2+\log e)N + \mathcal{O}\left((\log N)^2\right). \tag{4.16}$$

Therefore for sufficiently large N, it follows from Lemma 6 that

$$\begin{aligned} R_{\rho_g C}(N,t) &= \frac{\log|A_{\rho_g C}(N,t)|}{\log N!} \\ &< \frac{N\log N - (2+\log e)N + \mathcal{O}\left((\log N)^2\right)}{N\log N - (\log e)N + \frac{1}{2}\log N} \\ &= 1 - \frac{2N + \frac{1}{2}\log N + \mathcal{O}\left((\log N)^2\right)}{N\log N - (\log e)N + \frac{1}{2}\log N} \\ &= 1 - \frac{2N + \mathcal{O}\left((\log N)^2\right)}{N\log N - (\log e)N + \frac{1}{2}\log N}. \end{aligned} \tag{4.17}$$

And we know from Lemma 6 that:

$$\begin{aligned} R_G(N,t) = R_B(N,4t) &> 1 - \frac{16t(2\log N+1)}{\log N!} \\ &> 1 - \frac{32t\log N + 16t}{N\log N - (\log e)N + \frac{1}{2}\log N}. \end{aligned} \tag{4.18}$$

Then it follows that

$$\begin{aligned} & R_G(N,t) - R_{\rho_g C}(N,t) \\ >& \frac{2N - (32t\log N + 16t) + \mathcal{O}\left((\log N)^2\right)}{N\log N - (\log e)N + \frac{1}{2}\log N} > 0, \end{aligned} \tag{4.19}$$

for sufficiently large N and $t < \frac{N}{(16\log N + 8)}$.

From the above discussion, our proposed code in Section 3.1 indeed has a higher rate than the interleaving-based code for sufficiently small t. □

Based on Remark 8 in Section 4.3, we presented a construction of systematic t-generalized Cayley code $\mathcal{C}'_G(N,t) = \mathcal{C}_B^{\text{sys}}(N, 56\cdot 4t, 4t) = \mathcal{C}_B^{\text{sys}}(N, 224t, 4t)$ with rate $R'_G(N,t)$.

In the next Lemma 16, we compare the rate of $\mathcal{C}'_G(N,t)$ with that of $A_{\rho_g C}(N,t)$.

Lemma 16. *$R'_G(N,t) > R_{\rho_g C}(N,t)$ when $t < \min\{\frac{N}{112\log N}, \frac{1}{112}\lfloor\sqrt{N} - \frac{1}{2}\rfloor\}$ for sufficiently large N.*

Proof. For sufficiently large N, it follows from Lemma 15 and Lemma 6 that

$$R_{\rho_g C}(N,t) < 1 - \frac{2N + \mathcal{O}\left((\log N)^2\right)}{N\log N - (\log e)N + \frac{1}{2}\log N}. \tag{4.20}$$

And we know from Lemma 6 that:

$$\begin{aligned} R'_G(N,t) &> 1 - \frac{224t\log N}{\log N!} \\ &> 1 - \frac{224t\log N}{N\log N - (\log e)N + \frac{1}{2}\log N}. \end{aligned} \tag{4.21}$$

Then it follows that

$$\begin{aligned} & R'_G(N,t) - R_{\rho_g C}(N,t) \\ >& \frac{2N - 224t\log N + \mathcal{O}\left((\log N)^2\right)}{N\log N - (\log e)N + \frac{1}{2}\log N} > 0, \end{aligned} \tag{4.22}$$

for sufficiently large N and $t < \min\{\frac{N}{112\log N}, \frac{1}{112}\lfloor\sqrt{N} - \frac{1}{2}\rfloor\}$.

From the above discussion, our proposed systematic code indeed has a higher rate than the interleaving-based code for sufficiently small t in the generalized Cayley distance. $\square$

CHAPTER 5

Conclusion

5.1 Summary of Main Contributions

The generalized Cayley metric is a distance measure that generalizes the Kendall-τ metric and the Ulam metric. Interleaving was previously shown to be efficient in constructions of permutation codes in the generalized Cayley metric. However, interleaving incurs a noticeable rate penalty such that the constructed permutation codes cannot be order-optimal.

In the first part of this thesis, we derived the lower and upper bounds of the optimal rate of permutation code in the generalized transpositions.

In the second part, we first presented a construction of order-optimal permutation codes, which is not necessarily systematic, in the generalized Cayley metric, without interleaving. Based on this method, we then developed an explicit construction of systematic permutation codes from extensions of permutations. We further proved the existence of order-optimal systematic codes in this configuration and provided an explicit construction. Later on, we proved that our proposed codes are more rate efficient than the existing coding schemes based on interleaving for sufficiently large N when t is relatively small.

Our work on order-optimal permutation codes in the generalized Cayley distance has been presented in preliminary form at the *IEEE Information Theory Workshop* in Nov. 2017 [YSD17]. A longer version of this work, with the results regarding systematic codes added in, has been submitted to *IEEE Transactions on Information Theory* in 2017.

5.2 Future Extensions

In future work, we seek to find corresponding results in the binary case, which is expected to be useful in the synchronization of binary files. A majority of currently existing synchronization methods of binary files focus on errors that are i.i.d.. However, they lack the efficiency in correcting highly-concentrated errors such as the exchange of two paragraphs.

Binary codes that correct generalized transposition errors also have potential in DNA storage systems. Researchers have been paving the way to next generation DNA storage systems, where digital informations are stored in nucleotides of four nucleobases, adenine (A), cytosine (C), guanine (G) and thymine (T) [CGK12]. However, DNA undergoes breakages due to DNA aging caused by metabolic and hydrophilic processes, resulting in structure changes, including block deletions and reversal of adjacent blocks, of the DNA string [GYM17]. This motivates the extension of this research towards binary generalized Cayley codes.

APPENDIX A

Proof of Lemma 1

Proof. We know from the symmetry property of block permutation distance that it is enough for us to prove that $d_B(\pi_1, \pi_2) = |A(\pi_1) \setminus A(\pi_2)|$.

Suppose $\pi_1, \pi_2 \in \mathbb{S}_N$ such that $d_B(\pi_1, \pi_2) = d$, then from the definition of block permutation distance we know that π_1, π_2 satisfy (2.5) for some $\sigma \in \mathbb{S}_{d+1}$ and some $\psi_1, \psi_2, \cdots, \psi_{d+1}$. Suppose $\psi_k = \pi_1 [i_{k-1} + 1 : i_k]$ for $1 \leq k \leq d+1$, where $0 = i_0 < i_1 \cdots < i_d < i_{d+1} = N$. Then $\forall\ i \in [N-1]$ we have

$$\mathbb{I}\left[(\pi_1(i), \pi_1(i+1)) \in A(\pi_2)\right] = \begin{cases} 1, & i \notin \{i_1, \cdots, i_d\}, \\ 0, & i \in \{i_1, \cdots, i_d\}. \end{cases}$$

Therefore $|A(\pi_1) \setminus A(\pi_2)| = |\{i_1, \cdots, i_d\}| = d$.

The lemma is proved. □

APPENDIX B

Proof of Lemma 2

Proof. Define $B(\pi)$ as below,

$$B(\pi) \triangleq \{i | \pi(i+1) \neq \pi(i) + 1,\ 1 \leq i < N\}.$$

Then we know that

$$B(\pi) = \{i | (\pi(i), \pi(i+1)) \in (A(\pi) \setminus A(e)), 1 \leq i < N\},$$

which indicates that

$$|B(\pi)| = |A(\pi) \setminus A(e)| = w_B(\pi). \tag{B.1}$$

Denote $B_1 = B(\pi_1)$, $B_2 = B(\pi_2)$, $B_3 = B(\pi_1 \circ \pi_2)$. Then $\forall\ i \in B_3$, we have:

$$\pi_1(\pi_2(i+1)) \neq \pi_1(\pi_2(i)) + 1.$$

Then i must satisfy at least one of the conditions below:

$$\begin{aligned} &\{\pi_2(i+1) \neq \pi_2(i) + 1\},\ \textit{or} \\ &\{\pi_2(i) = k\ \textit{and}\ \pi_1(k+1) \neq \pi_1(k) + 1\}. \end{aligned} \tag{B.2}$$

Equation (B.2) means that either $i \in B_2$ or $\pi_2(i) \in B_1$ is true. Then we define an injection $f : (B \setminus B_2) \to B_1$ as below.

$$f(i) \triangleq \pi_2(i).$$

which means that

$$|B| = |B \setminus B_2| + |B \cap B_2| \leq |B_1| + |B_2|. \tag{B.3}$$

We know from (B.1) that (B.3) is equivalent to

$$w_B(\pi_1 \circ \pi_2) \leq w_B(\pi_1) + w_B(\pi_2).$$

□

APPENDIX C

Proof of Lemma 4

Proof. Suppose the number of permutations of length N with block permutation weight m is $F(m)$, then $b_B(N,t) = \sum_{m=0}^{t} F(m)$.

We know from [Mye02, equation (3)] that $F(0) = 1$, and for all $m > 1$,

$$F(m) = \binom{N-1}{m} m! \sum_{k=0}^{m} (-1)^{m-k} \frac{(k+1)}{(m-k)!}. \tag{C.1}$$

Let $a_k = \frac{(k+1)}{(m-k)!} \leq m+1$, $0 \leq k \leq m$, then $a_m = m+1 > a_{m-1} = m > a_{m-2} > \cdots > a_0$ holds. For arbitrary k, we have

$$a_{2k} - a_{2k-1} + \cdots + a_0 = a_0 + \sum_{i=1}^{k} (a_{2i} - a_{2i-1}) > 0,$$
$$a_{2k-1} - a_{2k-2} + \cdots - a_0 = \sum_{i=1}^{k} (a_{2i-1} - a_{2i-2}) > 0.$$

Therefore

$$A = m + 1 - (a_{m-1} - a_{m-2} + \cdots + (-1)^{m-1} a_0) < m+1,$$
$$A = m + 1 - m + (a_{m-2} - a_{m-3} + \cdots + (-1)^{m} a_0) > 1,$$

and

$$1 \leq A = \sum_{k=0}^{m} (-1)^{m-k} \frac{(k+1)}{(m-k)!} \leq m+1.$$

From the above discussion, we know that

$$\binom{N-1}{m} m! \leq F(m) \leq \binom{N-1}{m} (m+1)!.$$

To derive the upper bound of the ballsize $b_B(N,t)$, we have

$$F(m) \leq \binom{N-1}{m} (m+1)! = (m+1) \cdot \prod_{k=1}^{m} (N-k).$$

For $t \leq N - \sqrt{N} - 1$, we have $i \leq N - \sqrt{N} - 1$ for all $1 \leq i \leq t$. Therefore for all $1 \leq i \leq t$,

$$(N-i-1)^2 \geq (N - (N - \sqrt{N}))^2 = N \geq i+1.$$

Then,

$$
\begin{aligned}
& b_B(N,t) \\
=& \sum_{i=0}^{t} F(i) \\
\leq& 1 + \sum_{i=1}^{t} (i+1) \cdot \prod_{k=1}^{i} (N-k) \\
=& 1 + \sum_{i=1}^{t} (N - (N-i-1)) \cdot \prod_{k=1}^{i} (N-k) \\
=& 1 + \sum_{i=1}^{t} \left(\prod_{k=0}^{i} (N-k) - \prod_{k=1}^{i+1} (N-k) \right) \\
=& \prod_{k=0}^{t} (N-k) - \sum_{i=1}^{t} \left(\prod_{k=1}^{i+1} (N-k) - \prod_{k=0}^{i-1} (N-k) \right) \\
=& \prod_{k=0}^{t} (N-k) - \sum_{i=1}^{t} \prod_{k=1}^{i-1} (N-k) \left((N-i)(N-i-1) - N \right) \\
=& \prod_{k=0}^{t} (N-k) - \sum_{i=1}^{t} \prod_{k=1}^{i-1} (N-k) \left((N-i-1)^2 - i - 1 \right) \\
\leq& \prod_{k=0}^{t} (N-k).
\end{aligned}
$$

For the lower bound, we know that

$$
b_B(N,t) = \sum_{i=0}^{t} F(i) \geq 1 + \sum_{i=1}^{t} \prod_{k=1}^{i} (N-k) \geq \prod_{k=1}^{t} (N-k).
$$

The lemma is proved. □

APPENDIX D

Proof of Lemma 5

Proof. The upper bound is derived from (2.16) by replacing t by $4t$ and applying (2.14). For every $\pi \in B_G(N,t,e)$, we know that $d_G(\pi,e) \leq t$. Then from (2.14), we have $d_B(\pi,e) \leq 4d_G(\pi,e) \leq 4t$, which means that $\pi \in B_B(N,4t,e)$. Therefore $B_G(N,t,e) \subseteq B_B(N,4t,e)$, which implies that $b_G(N,t) \leq b_B(N,4t)$. From (2.16) we will get the upper bound.

For the lower bound, let $D_B(N,4t)$ be the set of all permutations that have block permutation weight $4t$. Let $K = |B_G(N,t,e) \cap D_B(N,4t)|$, then $b_G(N,t) > K$. For any $\pi \in D_B(N,4t)$, there exists some $\sigma \in \mathbb{D}_{4t+1}$ (recall $\mathbb{D}_{4t+1}$ is defined in Subsection A Section II as the set of all minimal permutations with length $4t+1$) and $\psi_1, \psi_2, \cdots, \psi_{4t+1}$ such that

$$e = (\psi_1, \psi_2, \cdots, \psi_{4t+1}),$$
$$\pi = \left(\psi_{\sigma(1)}, \psi_{\sigma(2)}, \cdots, \psi_{\sigma(4t+1)}\right).$$

Then we have $d_G(e,\pi) = d_G(e,\sigma)$. If $\pi \in (B_G(N,t,e) \cap D_B(N,4t))$, then $d_G(e,\sigma) \leq t$, which means that $\sigma \in B_G(4t+1,t,e)$. We know that there are $\binom{N-1}{4t}$ different partitions of e, each with the form $e\,(\psi_1, \psi_2, \cdots, \psi_{4t+1})$. For each such partition $\{\psi_i, 1 \leq i \leq 4t+1\}$ and any permutation $\sigma \in (B_G(4t+1,t,e) \cap \mathbb{D}_{4t+1})$, the permutation $\pi = \left(\psi_{\sigma(1)}, \psi_{\sigma(2)}, \cdots, \psi_{\sigma(4t+1)}\right) \in (B_G(N,t,e) \cap D_B(N,4t))$. Suppose that there are $x = |B_G(4t+1,t,e) \cap \mathbb{D}_{4t+1}|$ different such σ's, then $K \geq \binom{N-1}{4t} x$.

We only need to lower bound x. For any partition that divides $\{2, 4, \cdots, 4t\}$ into t different subsets with cardinality 2, we can apply t generalized transpositions to the identity

permutation e to exchange the elements in each subset and get permutation σ. Then all even numbers in e are not in their original position, and all odd numbers are not changed. Therefore, no consecutive pairs in e appear in σ, which means $\sigma \in (B_G(4t+1, t, e) \cap \mathbb{D}_{4t+1})$. There are $\frac{(2t)!}{2^t t!}$ such partitions, therefore $x \geq \frac{(2t)!}{2^t t!}$. From $b_G(N, t) \geq \binom{N-1}{4t} x$ the lower bound follows.

The lemma is proved. □

APPENDIX E

Proof of Lemma 7

Proof. Let $B_1 = \nu(\pi_1)$, $B_2 = \nu(\pi_2)$. We prove the statement by contradiction. If the lemma is not true, i.e., $|B_1 \Delta B_2| \leq 2d$, then $k = |D_1| = |D_2| \leq d$, where $D_1 = B_1 \setminus B_2$, $D_2 = B_2 \setminus B_1$. Suppose $D_1 = \{x_1, x_2, \cdots, x_k\}$, $D_2 = \{x_{k+1}, x_{k+2}, \cdots, x_{2k}\}$. Then, $\alpha^{(q,d)}(\pi_1) = \alpha^{(q,d)}(\pi_2)$ is equivalent to the following equations.

$$\begin{cases} x_1 + \cdots + x_k &= x_{k+1} + \cdots + x_{2k}, \\ x_1^2 + \cdots + x_k^2 &= x_{k+1}^2 + \cdots + x_{2k}^2, \\ & \vdots \\ x_1^{2d-1} + \cdots + x_k^{2d-1} &= x_{k+1}^{2d-1} + \cdots + x_{2k}^{2d-1}. \end{cases} \tag{E.1}$$

From (E.1), it follows that

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_{2k} \\ x_1^2 & x_2^2 & \cdots & x_{2k}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{2d-1} & x_2^{2d-1} & \cdots & x_{2k}^{2d-1} \end{pmatrix} \mathbf{y} = \mathbf{0},$$

where $\mathbf{y} = [y_1, y_2, \cdots, y_{2k}]^T$, and

$$y_i = \begin{cases} 1, & 1 \leq i \leq k, \\ -1, & k < i \leq 2k. \end{cases}$$

Given that $2k \leq 2d$, the above equation implies that

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_{2k} \\ x_1^2 & x_2^2 & \cdots & x_{2k}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{2k-1} & x_2^{2k-1} & \cdots & x_{2k}^{2k-1} \end{pmatrix} \mathbf{y} = \mathbf{0}. \tag{E.2}$$

Denote the Vandermonde matrix in equation (E.2) by $\mathbf{U}$. Then $\mathbf{y}$ is in the nullspace of $\mathbf{U}$. Therefore $\mathbf{U}$ is singular, and it implies that the determinant of $\mathbf{U}$ is equal to 0, i.e.,

$$0 = \det \mathbf{U} = \prod_{1 \leq i < j \leq 2k} (x_i - x_j). \tag{E.3}$$

As q is a divisor of 0, q should also be a divisor of the right hand of equation (E.3), which implies that $\exists\ i \neq j \in [2k]$ such that $q|(x_i - x_j)$. Then $x_i = x_j$ on $\mathbb{F}_q$, and we must have $x_i \in D_1, x_j \in D_2$ or $x_i \in D_2, x_j \in D_1$, which implies that $x_i, x_j \in D_1 \cap D_2$, which is a contradiction.

The lemma is proved. □

APPENDIX F

Proof of Lemma 8

Proof. Suppose

$$\begin{aligned} f_1 &= X^{N-1} + a_1 X^{N-2} + \cdots + a_{4t-1} X^{N-4t} + g_1, \\ f_2 &= X^{N-1} + a_1' X^{N-2} + \cdots + a_{4t-1}' X^{N-4t} + g_2. \end{aligned} \tag{F.1}$$

Additionally, suppose

$$\begin{aligned} h_1 \cdot f_1 &= X^{N+t-1} + \beta_{N+t-2} X^{N+t-2} + \cdots + \beta_0, \\ h_2 \cdot f_2 &= X^{N+t-1} + \beta_{N+t-2}' X^{N+t-2} + \cdots + \beta_0'. \end{aligned}$$

Then from (F.1) and (3.15) it follows that the first $4t$ coefficients of $h_1 \cdot f_1$ can be represented by $\{a_i\}, \{c_i\}$ as below:

$$\begin{cases} s_{N+t-2} = a_1 + c_1, \\ s_{N+t-3} = a_2 + c_1 a_1 + c_2, \\ \vdots \\ s_{N-1} = a_t + c_1 a_{t-1} + \cdots + c_t, \\ \vdots \\ s_{N-3t} = a_{4t-1} + c_1 a_{4t-2} + c_2 a_{4t-3} + \cdots + c_t a_{3t-1}. \end{cases}$$

Similarly, we also have

$$\begin{cases} s'_{N+t-2} = a'_1 + c'_1, \\ s'_{N+t-3} = a'_2 + c'_1 a'_1 + c'_2, \\ \vdots \\ s'_{N-1} = a'_t + c'_1 a'_{t-1} + \cdots + c'_t, \\ \vdots \\ s'_{N-3t} = a'_{4t-1} + c'_1 a'_{4t-2} + c'_2 a'_{4t-3} + \cdots + c'_t a'_{3t-1}. \end{cases}$$

Then (3.13) is true iff $s_i = s'_i$ for all $N - 3t \leq i \leq N + t - 2$, which is equivalent to the following equation:

$$\begin{pmatrix} 1 & & & \\ a_1 & 1 & & \\ \vdots & \vdots & \ddots & \\ a_{t-1} & a_{t-2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{4t-2} & a_{4t-3} & \cdots & a_{3t-1} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_t \end{pmatrix} + \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{4t-1} \end{pmatrix} = \begin{pmatrix} 1 & & & \\ a'_1 & 1 & & \\ \vdots & \vdots & \ddots & \\ a'_{t-1} & a'_{t-2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a'_{4t-2} & a'_{4t-3} & \cdots & a'_{3t-1} \end{pmatrix} \begin{pmatrix} c'_1 \\ c'_2 \\ \vdots \\ c'_t \end{pmatrix} + \begin{pmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_{4t-1} \end{pmatrix}. \tag{F.2}$$

We note that (F.2) is equivalent to (3.18). □

APPENDIX G

Proof of Lemma 9

Proof. Let $\sigma_1 = E(\pi_1, s_1)$, $\sigma_2 = E(\pi_2, s_2)$. Recall the notion of *characteristic sets* in Definition 3. Suppose $A(\pi_1)$, $A(\pi_2)$, $A(\sigma_1)$, $A(\sigma_2)$ are the characteristic sets of π_1, π_2, σ_1, σ_2, respectively. According to Lemma 1,

$$\begin{aligned} d_B(\pi_1, \pi_2) &= |A(\pi_1) \setminus A(\pi_2)|, \\ d_B(\sigma_1, \sigma_2) &= |A(\sigma_1) \setminus A(\sigma_2)|. \end{aligned} \tag{G.1}$$

Let $k_1 = \pi_1^{-1}(s_1)$, $k_2 = \pi_2^{-1}(s_2)$, then $\pi_{1,k_1} = s_1$ and $\pi_{2,k_2} = s_2$. If $1 \leq k_1, k_2 < N$, let $\pi_{1,k_1+1} = j_1$ and $\pi_{2,k_2+1} = j_2$.

Suppose first s_1 is a jump point, then consider the following cases.

1. $s_1 \neq s_2$ and either $k_1 = N$ or $k_2 = N$.

   (a) $k_1 = k_2 = N$. In this case, $A(\sigma_1) = A(\pi_1) \cup \{(s_1, N+1)\}$, $A(\sigma_2) = A(\pi_2) \cup \{(s_2, N+1)\}$. Therefore, $A(\sigma_1) \setminus A(\sigma_2) = (A(\pi_1) \setminus A(\pi_2)) \cup \{(s_1, N+1)\}$. From (G.1) we know that $d_B(\sigma_1, \sigma_2) = d_B(\pi_1, \pi_2) + 1$.

   (b) $k_1 = N \neq k_2$. In this case, $A(\sigma_1) = A(\pi_1) \cup \{(i_1, N+1)\}$, $A(\sigma_2) = (A(\pi_2) \setminus \{(s_2, j_2)\}) \cup \{(s_2, N+1), (N+1, j_2)\}$. Therefore, $A(\sigma_1) \setminus A(\sigma_2) = (A(\pi_1) \setminus (A(\pi_2) \setminus \{(s_2, j_2)\})) \cup \{(s_1, N+1)\}$, which means $(A(\pi_1) \setminus A(\pi_2)) \cup \{(s_1, N+1)\} \subseteq A(\sigma_1) \setminus A(\sigma_2)$. From (G.1) we know that $d_B(\sigma_1, \sigma_2) \geq d_B(\pi_1, \pi_2) + 1$.

   (c) $k_2 = N \neq k_1$. Following the same logic in the previous case, we know that

$d_B(\sigma_1, \sigma_2) \geq d_B(\pi_1, \pi_2) + 1$.

2. $s_1 \neq s_2$, $k_1, k_2 \neq N$. Since s_1 is a jump point, $j_1 \neq j_2$.

   (a) In this case, $A(\sigma_1) = (A(\pi_1) \setminus \{(s_1, j_1)\}) \cup \{(s_1, N+1), (N+1, j_1)\}$, $A(\sigma_2) = (A(\pi_2) \setminus \{(s_2, j_2)\}) \cup \{(s_2, N+1), (N+1, j_2)\}$. Therefore, $((A(\pi_1) \setminus A(\pi_2)) \setminus \{(s_1, j_1)\}) \cup \{(s_1, N+1), (N+1, j_1)\} \subseteq A(\sigma_1) \setminus A(\sigma_2)$. From (G.1) we know that $d_B(\sigma_1, \sigma_2) \geq d_B(\pi_1, \pi_2) + 1$.

If i_1 is not a jump point, then consider the following cases.

1. $s_1 = s_2$ and either $k_1 = N$ or $k_2 = N$

   (a) $k_1 = k_2 = N$. In this case, $A(\sigma_1) = A(\pi_1) \cup \{(s_1, N+1)\}$, $A(\sigma_2) = A(\pi_2) \cup \{(s_1, N+1)\}$. Therefore, $A(\sigma_1) \setminus A(\sigma_2) = (A(\pi_1) \setminus A(\pi_2))$. From (G.1) we know that $d_B(\sigma_1, \sigma_2) = d_B(\pi_1, \pi_2)$.

   (b) $k_1 = N \neq k_2$. In this case, $A(\sigma_1) = A(\pi_1) \cup \{(s_1, N+1)\}$, $A(\sigma_2) = (A(\pi_2) \setminus \{(s_1, j_2)\}) \cup \{(s_1, N+1), (N+1, j_2)\}$. Therefore, $A(\sigma_1) \setminus A(\sigma_2) = (A(\pi_1) \setminus (A(\pi_2) \setminus \{(s_1, j_2)\})) = (A(\pi_1) \setminus A(\pi_2))$. From (G.1) we know that $d_B(\sigma_1, \sigma_2) = d_B(\pi_1, \pi_2)$.

   (c) $k_2 = N \neq k_1$. Follow the same logic in the previous case, we know that $d_B(\sigma_1, \sigma_2) = d_B(\pi_1, \pi_2)$.

2. $k_1, k_2 \neq N$. Since s_1 is not a jump point, either $s_1 = s_2$ or $j_1 = j_2$ must be satisfied.

   (a) $s_1 = s_2$ and $j_1 = j_2$. In this case, $A(\sigma_1) = (A(\pi_1) \setminus \{(s_1, j_1)\}) \cup \{(s_1, N+1), (N+1, j_1)\}$, $A(\sigma_2) = (A(\pi_2) \setminus \{(s_1, j_1)\}) \cup \{(s_1, N+1), (N+1, j_1)\}$. Therefore, $A(\sigma_1) \setminus A(\sigma_2) = A(\pi_1) \setminus A(\pi_2)$. From (G.1) we know that $d_B(\sigma_1, \sigma_2) = d_B(\pi_1, \pi_2)$.

   (b) $s_1 = s_2$ and $j_1 \neq j_2$. In this case, $A(\sigma_1) = (A(\pi_1) \setminus \{(s_1, j_1)\}) \cup \{(s_1, N+1), (N+1, j_1)\}$, $A(\sigma_2) = (A(\pi_2) \setminus \{(s_1, j_2)\}) \cup \{(s_1, N+1), (N+1, j_2)\}$. Therefore, $A(\sigma_1) \setminus A(\sigma_2) = ((A(\pi_1) \setminus A(\pi_2)) \setminus \{(s_1, j_1)\}) \cup \{(N+1, j_1)\}$. From (G.1) we know that $d_B(\sigma_1, \sigma_2) = d_B(\pi_1, \pi_2)$.

(c) $s_1 \neq s_2$ and $j_1 = j_2$. Follow the same logic as indicated in the previous case, we know that $d_B(\sigma_1, \sigma_2) = d_B(\pi_1, \pi_2)$.

The lemma is proved. □

APPENDIX H

Proof of Lemma 10

Proof. For all $i \in H(S_1, S_2)$, let

$$m(i) = \min\{m : \ s_{1,m} = i, \ s_{2,m} \neq i\}. \tag{H.1}$$

Suppose $J_{1,m(i)} = \big(s_{1,1}, s_{1,2}, \cdots, s_{1,m(i)-1}\big)$, $J_{2,m(i)} = \big(s_{2,1}, s_{2,2}, \cdots, s_{2,m(i)-1}\big)$. Let $\sigma_1^{m(i)} = E(\pi_1, J_{1,m(i)})$ and $\sigma_2^{m(i)} = E(\pi_2, J_{2,m(i)})$. Recall the definition of the *jump set* $F(\pi_1, \pi_2, S_1, S_2)$ in Definition 7. Consider the following two cases:

1. If $m(i) \in F(\pi_1, \pi_2, S_1, S_2)$, then $s_{1,m(i)} = i$ is a jump point of $E(\sigma_1^{m(i)}, s_{1,m(i)})$ with respect to $E(\sigma_2^{m(i)}, s_{2,m(i)})$.

2. If $m(i) \notin F(\pi_1, \pi_2, I_1, I_2)$, then i is not a jump point of $E(\sigma_1^{m(s)}, s_{1,m(i)})$ with respect to $E(\sigma_2^{m(s)}, s_{2,m(i)})$. Let $k_1' = (\sigma_1^{m(i)})^{-1}(s_{1,m(i)})$, $k_1 = \pi_1^{-1}(s_{1,m(i)})$, $k_2' = (\sigma_2^{m(i)})^{-1}(s_{2,m(i)})$, $k_2 = \pi_2^{-1}(s_{2,m(i)})$, then $\sigma_{1,k_1'}^{m(i)} = \pi_{1,k_1} = s_{1,m(i)}$ and $\sigma_{2,k_2'}^{m(i)} = \pi_{2,k_2} = s_{2,m(i)}$. Given that $s_{1,m(i)}$ is not a jump point and $s_{1,m(i)} = s \neq s_{2,m(i)}$, we know from Definition 6 that $k_1, k_2 \neq N + m(i) - 1$ and $\sigma_{1,k_1'+1}^{m(i)} = \sigma_{2,k_2'+1}^{m(i)}$ must be true. Let $j = \sigma_{1,k_1'+1}^{m(i)} = \sigma_{2,k_2'+1}^{m(i)}$. We know from equation (H.1) that $\pi_{1,k_1+1} = \pi_{2,k_2+1} = j \in [N]$, otherwise $N < j < N + m(i)$ is inserted after i in π_1 and is not inserted after i in π_2, a contradiction. Then $(i, j) \in A(\pi_1)$, $(s_{2,m(i)}, j) \in A(\pi_2)$ and $s_{2,m(i)} \neq i$. Therefore $(i, j) \in (A(\pi_1) \setminus A(\pi_2))$.

Suppose $J = \{i | m(i) \notin F(\pi_1, \pi_2, S_1, S_2), i \in H(S_1, S_2)\}$, then from the above discussion:

$$|F(\pi_1, \pi_2, S_1, S_2)| \geq |H(S_1, S_2) \setminus J|,$$
$$d_B(\pi_1, \pi_2) = |A(\pi_1) \setminus A(\pi_2)| \geq |J|.$$

And from Lemma 9 we know that

$$\begin{aligned} d_B(E(\pi_1, S_1), E(\pi_2, S_2)) \geq & d_B(\pi_1, \pi_2) + |F(\pi_1, \pi_2, S_1, S_2)| \\ \geq & |H(S_1, S_2) \setminus J| + |J| \\ \geq & |H(S_1, S_2)|. \end{aligned}$$

The lemma is proved. □

APPENDIX I

Proof of Lemma 12

Proof. Denote the set of all subsets of $[N]$ with cardinality l by $\mathcal{B}$. Let $\mathcal{L}_{\max}$ be the largest (N, l, a)-set. Suppose that $k = |\mathcal{L}_{\max}| < m$. Otherwise if $k < m$, and for each set $L_i \in \mathcal{L}_{\max}$, there are at most $\binom{l}{a}\binom{n-a}{l-a}$ sets $L_j \in \mathcal{B}$ that satisfy $|L_i \cap L_j| \geq a$ for $j \neq i$. Then there will be at most $k\binom{l}{a}\binom{n-a}{l-a} < m\binom{l}{a}\binom{N-a}{l-a} = m\binom{l}{a}\frac{(N-a)!}{(N-l)!(l-a)!} = \frac{m\binom{l}{a}^2}{\binom{N}{a}}\binom{N}{l} < \binom{N}{l} = |\mathcal{B}|$ sets $L_j \in \mathcal{B}$ such that each such set has an intersection with cardinality no less than a with at least one of the sets in $\mathcal{L}_{\max}$. Then there exists a set $L_{k+1} \in \mathcal{B}$ such that for all $L_i \in \mathcal{L}_{\max}$, $|L_i \cap L_{k+1}| < a$. Then $\{L_{k+1}\} \cup \mathcal{L}_{\max}$ is an (N, l, a)-set that satisfies the two conditions with cardinality $k + 1$. The cardinality of this newly constructed (N, l, a)-set is larger than that of $\mathcal{L}_{\max}$. Contradiction! Therefore $k \geq m$ must be true, the lemma is proved. □

APPENDIX J

Proof of Lemma 14

Proof. From [Far07, equation (13)], we know that for all $r, n \in \mathbb{N}^*$

$$g_r(n) = \mathrm{GCD}(r!, (n+r)g_{k-1}(n)), \tag{J.1}$$

where for all $r \in \mathbb{N}$, $n \in \mathbb{N}^*$,

$$g_r(n) = \frac{n(n+1)\cdots(n+r)}{\mathrm{LCM}(n, n+1, \cdots, n+r)}. \tag{J.2}$$

From (J.1) and (J.2), we know that

$$g_r(n) | r!, \ \forall r, n \in \mathbb{N}^*, \tag{J.3}$$

which implies that

$$\frac{n(n+1)\cdots(n+r)}{\mathrm{LCM}(n, n+1, \cdots, n+r)} \leq r!. \tag{J.4}$$

Let $n = N+1$, $r = k-1$ in (J.4), we know that for all $N, k \in \mathbb{N}^*$,

$$\begin{aligned} &\mathrm{LCM}\,(N+1, N+2, \cdots, N+k) \\ \geq &\frac{(N+1)(N+2)\cdots(N+k)}{(k-1)!}. \end{aligned} \tag{J.5}$$

Let $[K] \setminus Y = \{j_1, j_2, \cdots, j_{k-M}\}$. Notice that

$$\begin{aligned}
&\mathrm{LCM}\left(N+1, N+2, \cdots, N+k\right) \\
=&\mathrm{LCM}(\mathrm{LCM}\left(N+i_1, N+i_2, \cdots, N+i_M\right), \\
&\mathrm{LCM}\left(N+j_1, N+j_2, \cdots, N+j_{k-M}\right)) \\
\leq&\left[\prod_{s=1}^{k-M}\left(N+j_s\right)\right]\mathrm{LCM}\left(N+i_1, N+i_2, \cdots, N+i_M\right).
\end{aligned} \tag{J.6}$$

From equation (J.5) and (J.6),

$$\begin{aligned}
&\mathrm{LCM}\left(N+1, N+2, \cdots, N+i_M\right) \\
\geq&\frac{\mathrm{LCM}(N+1, N+2, \cdots, N+k)}{\prod_{s=1}^{k-M}\left(N+j_s\right)} \\
\geq&\frac{(N+1)(N+2)\cdots(N+k)}{(k-1)!\prod_{s=1}^{k-M}\left(N+j_s\right)} \\
=&\frac{\prod_{s=1}^{M}\left(N+i_s\right)}{(k-1)!} > \frac{N^M}{k!}.
\end{aligned}$$

From Lemma 6, for all $k > 3$ and $N > k^2$,

$$\frac{N^M}{k!} > \frac{N^M}{2^{(k+1)\log k - k + 2}} = \frac{N^M 2^{k-2}}{k^{k+1}} \geq \frac{N^M}{k^k} > N^{M-\frac{k}{2}}.$$

The lemma is proved. □

References

[BE15] S. Buzaglo and T. Etzion, "Bounds on the size of permutation codes with the Kendall tau-metric," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3241–3250, Jun. 2015.

[BYEB16] S. Buzaglo, E. Yaakobi, T. Etzion, and J. Bruck, "Systematic error-correcting codes for permutations and multi-permutations," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3113–3124, Jun. 2016.

[CGK12] G. M. Church, Y. Gao, and S. Kosuri, "Next-generation digital information storage in dna," *Science*, p. 1226355, 2012.

[CLCL16] K.-T. Chen, C.-L. Li, H.-T. Chiu, and C. L. Lu, "An efficient algorithm for one-sided block ordering problem under block-interchange distance," *Theoretical Computer Science*, vol. 609, pp. 296 – 305, 2016.

[CV14] Y. M. Chee and V. K. Vu, "Breakpoint analysis and permutation codes in generalized Kendall tau and Cayley metrics," in *Proc. IEEE Int. Symp. Inf. Theory*, Hawaii, USA, Jun. 2014, pp. 2959–2963.

[CVZ15] Y. M. Chee, V. K. Vu, and X. Zhang, "Permutation codes correcting a single burst deletion I: Unstable deletions," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015, pp. 1741–1745.

[DA10] L. Dolecek and V. Anantharam, "Repetition error correcting sets: Explicit constructions and prefixing methods," *SIAM J. Discrete Math.*, vol. 23, no. 4, pp. 2120–2146, Jan. 2010.

[DW15] I. Dixon and G. Whittaker, Eds., *Storing your music in the iCloud*. Apress, 2015.

[Far07] B. Farhi, "Nontrivial lower bounds for the least common multiple of some finite sequences of integers," *J. Number Theory*, vol. 125, no. 2, pp. 393–411, 2007.

[FM14] F. Farnoud and O. Milenkovic, "Multipermutation codes in the ulam metric for nonvolatile memories," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 919–932, May 2014.

[FSM13] F. Farnoud, V. Skachek, and O. Milenkovic, "Error-correction in flash memories via codes in the Ulam metric," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3003–3020, May 2013.

[GLRS15] F. Göloğlu, J. Lember, A. E. Riet, and V. Skachek, "New bounds for permutation codes in Ulam metric," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015, pp. 1726–1730.

[GYF+16] R. Gabrys, E. Yaakobi, F. Farnoud, F. Sala, J. Bruck, and L. Dolecek, "Codes correcting erasures and deletions for rank modulation," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 136–150, Jan. 2016.

[GYM17] R. Gabrys, E. Yaakobi, and O. Milenkovic, "Codes in the damerau distance for deletion and adjacent transposition correction," *IEEE Trans. Inf. Theory*, vol. PP, no. 99, pp. 1–1, 2017.

[JMSB09] A. Jiang, R. Mateescu, M. Schwartz, and J. Bruck, "Rank modulation for flash memories," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2659–2673, June 2009.

[Mye02] A. Myers, "Counting permutations by their rigid patterns," *J. Combin. Theory Ser. A*, vol. 99, no. 2, pp. 345–357, 2002.

[Rob55] H. Robbins, "A remark on Stirling's formula," *The American Mathematical Monthly*, vol. 62, no. 1, pp. 26–29, 1955. [Online]. Available: http://www.jstor.org/stable/2308012

[Vad12] S. P. Vadhan, Ed., *Pseudorandomness*. Foundations and Trends®in Theoretical Computer Science, 2012, vol. 7, no. 1-3. [Online]. Available: http://dx.doi.org/10.1561/0400000010

[YSD17] S. Yang, C. Schoeny, and L. Dolecek, "Order-optimal permutation codes in the generalized Cayley metric," in *IEEE Information Theory Workshop*, Kaohsiung, Taiwan, Nov 2017, pp. 234–238.

[ZG16] Y. Zhang and G. Ge, "Snake-in-the-box codes for rank modulation under Kendall's τ-metric," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 151–158, Jan. 2016.

[ZS17] R. Zeira and R. Shamir, "Sorting by cuts, joins and whole chromosome duplications," *Journal of Computational Biology*, vol. 24, pp. 127–137, 2017.