IEEE Transactions on Information Theory

On Enumerating Feasible Permutations for Rank Modulation Codes in DNA storage via Hyperplane Arrangements

Journal:	*IEEE Transactions on Information Theory*
Manuscript ID	IT-25-0625.R1
Manuscript Type:	Regular Manuscript
Date Submitted by the Author:	29-Jan-2026
Complete List of Authors:	Sobhani, Reza; University of Isfahan Parvaresh, Farzad; University of Isfahan Faculty of Engineering Abdollahi, Alireza; University of Isfahan Abedi, Farzaneh; University of Isfahan Bagherian, Javad; University of Isfahan Khatami, Maryam; University of Isfahan
Keywords:	Data storage systems, DNA, Error correction codes, Graph theory, Linear algebra
Subject Category:	Coding and Decoding

On Enumerating Feasible Permutations for Rank Modulation Codes in DNA storage via Hyperplane Arrangements

Reza Sobhani, Farzad Parvaresh, Alireza Abdollahi, Farzaneh Abedi, Javad Bagherian, and Maryam Khatami

***Abstract*—For a directed graph G with edge set $\{1, 2, \ldots, k\}$, a feasible permutations for G is a permutation π on $\{1, 2, \ldots, k\}$ such that there exists an injective weight function $p : \{1, 2, \ldots, k\} \to \mathbb{Z}^+$ for which the sum of the weights of all incoming edges equals the sum of the weights of all outgoing edges at every vertex of G, provided that if $p(i_1) < p(i_2) < \cdots < p(i_k)$, then $\pi(i_t) = t$ for every $t \in \{1, \ldots, k\}$. In this paper, we study the number of feasible permutations (denoted by $F_{q,\ell}$) for the De Bruijn graph $G_{q,\ell-1}$, where $q > 2$ is the size of the alphabet. In the case $q = 4$, the corresponding De Bruijn graph is used in DNA data storage. To enumerate these feasible permutations, we establish a connection between feasible permutations and regions in a special hyperplane arrangement, denoted by $\mathcal{A}$. Using Zaslavsky's formula, we present some numerical results and obtain the exact number of $F_{3,2}$ and $F_{4,2}$. Since the formula becomes complicated for larger values of q and ℓ, we concentrate on the case $\ell = 2$ and then by counting the regions in a hyperplane subarrangement of $\mathcal{A}$, we provide a lower bound on $F_{q,2}$. We compare our bound with the latest lower bound on $F_{q,2}$ and show that our bound is $\Omega(q^3 J(q))$ while the previous one is $\Omega(q^2 J(q))$, where $J(q) = (q^2 - 2q + 1)!(q^2)!/(q^2 - q)!$.**

***Index Terms*—DNA storage, Rank modulation, feasible permutation, hyperplane, arrangement.**

I. Introduction

IN recent decades, DNA-based data storage has garnered significant attention as an alternative to traditional digital storage due to its exceptional data density and long-term stability [5], [12], [3]. This innovative approach has considerable interest as it promises to revolutionize the way data is preserved and accessed in the future. Although DNA storage system has garnered significant interest, challenges such as high costs, slow read speeds, and errors in DNA synthesis and sequencing remain [10]. Improving these limitations is crucial for wider use of DNA-based data storage.

The DNA storage process involves two main stages: writing and reading. In the *writing process*, digital information is converted into DNA sequences through a simple mapping, followed by DNA synthesis, a biochemical procedure [12]. Then, the DNA storage channel takes these synthesized macromolecules string as its input and the information is then retrieved in the *reading process*. One commonly employed technique in this process is shotgun sequencing [8]. In this method, long DNA strands are fragmented into numerous shorter subsequences, which are sequenced individually. The original sequence is reconstructed by identifying overlaps between these subsequences.

The read of these subsequences, known as ℓ-grams, is used to construct a profile vector, which represents the frequency and distribution of nucleotides (or other relevant features) within the sequence [6]. Utilizing profile vectors in shotgun sequencing enhances the accuracy, efficiency, and data density of DNA storage systems.

Although the synthesis of DNA molecules has led to significant advancements in storage technology, the DNA storage channel introduces unique error patterns distinct from digital counterparts. These include *substitution errors*, which arise during synthesis (writing) and sequencing (reading), and *coverage errors* (see [6]). Therefore, to maintain reliability in reading and writing, similar to those in classical coding theory, new encoding schemes must be propose. In [6], various error correction mechanisms have been proposed, including *Rank Modulation codes*, which serve as one of the key approaches to addressing these errors. This scheme evaluates whether the frequency of a specific substring in a given DNA sequence exceeds that of other substrings. Consequently, codewords are considered as permutations with a defined distance, referred to as *feasible permutation*, that preserve the order of the entries in the profile vector of the DNA sequence. Clearly, generating permutations requires the entries of the profile vector to be distinct.

Among all permutations, distinguishing feasible permutations and determining the number of these permutations

Corresponding author: Farzad Parvaresh(f.parvaresh@eng.ui.ac.ir)
F. Parvaresh is with the Department of Electrical Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran
R. Sobhani and F. Abedi are with the Department of Applied Mathematics and Computer Science, Faculty of Mathematics and Statistics, University of Isfahan, Isfahan 81746-73441, Iran
A. Abdollahi, J. Bagherian and M. Khatami are with the Department of Pure Mathematics, Faculty of Mathematics and Statistics, University of Isfahan, Isfahan 81746-73441, Iran
E-mail addresses:
r.sobhani@sci.ui.ac.ir (R. Sobhani),
f.parvaresh@eng.ui.ac.ir (F. Parvaresh),
a.abdollahi@math.ui.ac.ir (A. Abdollahi),
abedi87f@gmail.com (F. Abedi),
bagherian@sci.ui.ac.ir (J. Bagherian),
m.khatami@sci.ui.ac.ir (M. Khatami).
Manuscript received Jun 30, 2025; revised Apr 27, 2026.

is a significant problem. Following the work in [6], in [9] and [1], the *De Bruijn graph* is used to analyze a DNA string over an alphabet of size q and order ℓ (where order ℓ represents the length of the subsequences). In [6], by considering the necessary conditions for the feasibility of a permutation, a lower bound for the number of feasible permutations is presented. Furthermore, a technique for determining the feasibility of a permutation and an encoding algorithm were developed. In [1], by considering the balance constraint of the De Bruijn graph, optimal systematic codes for feasible permutations are provided, along with an algorithm for constructing these codes. Moreover, a lower bound on the number of feasible permutations is presented in [1], which improves upon the lower bound presented in [6].

In this paper, by analyzing the null space of the De Bruijn graph and finding a basis for the null space, we present a constraint for the feasibility of a vector. Then, by establishing a correspondence between feasible permutations and points in the geometric space corresponding to hyperplane arrangement, we compute the number of feasible permutations as rank modulation codes. Furthermore, we present a lower bound on the number of feasible permutations that is an improvement over the lower bound presented in [1].

The paper is organized as follows. In Section II, we present definitions and preliminaries, including rank modulation codes as feasible permutations, hyperplane arrangements, and the construction of the base matrix of the null space of a graph, used throughout this paper. Section III focuses on constructing a base for the De Bruijn graph in the case $\ell = 2$ (i.e., $G_{q,2}$), providing a bijection between feasible permutations and the geometric space of an arrangement. It also presents an algorithm for enumerating the number of all feasible permutations. In Section IV, we propose a new lower bound for the number of these permutations and show that our bound is an improvement over that given in [1].

II. PRELIMINARIES

In this section, we introduce essential definitions and preliminary concepts to provide a foundation for the main content. We discuss DNA sequences, profile vectors, De Bruijn graphs, feasible permutations, hyperplanes and hyperplane arrangements. We also present a method for calculating the null space of a directed graph. Recall that, in [9], [1] and [6] it is suggested that information can be encoded in the profile vector of the DNA sequence.

A. Feasible Permutations

Throughout this paper, let Σ be an alphabet with size $q > 2$, equipped with a fixed total order. In what follows, for a positive integer n, an element of Σ^n is called a string of length n. We assume that Σ^ℓ is equipped with lexicographic (alphabetical) order induced by the order on Σ, which we denote by $\prec_L$. We further assume that $\Sigma^\ell = \{w_1, \ldots, w_{q^\ell}\}$ such that $w_1 \prec_L w_2 \prec_L \ldots \prec_L w_{q^\ell}$. We define $\tilde{\Sigma}^\ell := \Sigma^\ell \setminus \{\alpha^{(\ell)} : \alpha \in \Sigma\}$, where $\alpha^{(\ell)}$ denotes the string consisting of ℓ repetitions of α.

For example, if $\Sigma = \{A, C, G\}$ with the order $A < C < G$, then
$\Sigma^2 = \{AA, AC, AG, CA, CC, CG, GA, GC, GG\}$, where $AA \prec_L AC \prec_L \cdots \prec_L GG$, and, $\tilde{\Sigma}^2 = \{AC, AG, CA, CG, GA, GC\}$. We also denote by $[n]$, the set $\{1, \ldots, n\}$.

For each string s, a *profile vector* $p_{s,\ell}$ is defined as follows:

Definition II.1. *[1] Consider a string $s = s_0 \cdots s_{n-1} \in \Sigma^n$. The profile vector $p_{s,\ell} \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$ of s with order ℓ is defined as a non-negative integer vector, indexed by the set $\Sigma^\ell = \{w_1, w_2, \ldots, w_{q^\ell}\}$, such that:*

$$p_{s,\ell}(w_k) := |\{i \in [n] | s_i s_{i+1} \cdots s_{i+\ell-1} = w_k\}|, \quad \forall k \in [q^\ell],$$

where the indices are taken modulo n. In other words, the profile vector

$$p_{s,\ell} = (p_{s,\ell}(w_1), \ldots, p_{s,\ell}(w_{q^\ell}))$$

shows the frequency of each substring of length ℓ appearing in s (cyclically). Each substring w_k of length ℓ is named window of length ℓ in s (ℓ-gram). We say that the vector x is feasible if there exists a finite string s over the alphabet Σ such that x is the profile vector of s, i.e. $p_{s,\ell} = x$.

Example II.2. *Let $\Sigma = \{A, C, G, T\}$ ($q = 4$) and consider the following string:*

$$\begin{aligned} s =\, & TTTTTTTTTTTTGTGTGTGTCTCTCTCTATAT \\ & AGAGAGCGCGCGCGGGGGGGGGGGGGGTG \\ & TGTGTGTGTATAAAAAAAAAAAAAAATCT \\ & CTCTCTGCGTCGTACGTCTAGACGTGCAT, \end{aligned}$$

the profile vector of s of order $\ell = 2$ in which the indices of the profile vector are in lexicographic order is as follow:

$$p_{s,2} = \begin{array}{cccccccccc} AA & AC & AG & AT & CA & CC & CG & CT & GA & GC \\ (15 & 2 & 4 & 5 & 1 & 0 & 8 & 9 & 3 & 6 \\ & & & & GG & GT & TA & TC & TG & TT \\ & & & & 13 & 14 & 7 & 10 & 11 & 12) \end{array}$$

In the context of Rank modulation codes, a significant challenge is identifying the necessary conditions for the feasibility of a vector. Specifically, the aim is to determine which vectors with distinct elements are feasible. In [1], the *De Bruijn* graph has been used as a suitable tool to check the feasibility.

Definition II.3. *The De Bruijn graph of order $\ell \geq 1$ over an alphabet Σ of size q, denoted by $G_{q,\ell}$, is a directed graph with the set of vertices $V(G_{q,\ell})$ and the set of edges $E(G_{q,\ell})$, where $V(G_{q,\ell}) = \Sigma^\ell$, and*

$$E(G_{q,\ell}) = \{(a_0 a_1 \cdots a_{\ell-1}, a_1 a_2 \cdots a_\ell) \mid a_i \in \Sigma, \ 0 \leq i \leq \ell\}.$$

In other words, each directed edge in the De Bruijn graph connects the vertex $a_0 a_1 \ldots a_{\ell-1}$ to the vertex $a_1 a_2 \ldots a_\ell$. We denote the edge e from vertex $a_0 a_1 \ldots a_{\ell-1}$ to $a_1 a_2 \ldots a_\ell$ by $e = a_0 a_1 a_2 \ldots a_\ell$

Notation: Consider the De Bruijn graph $G_{q,\ell-1}$ and a vector x, indexed by substrings in $\sum^\ell$. The vector x is used to assign weights to the edges of $G_{q,\ell-1}$, and we denote the resulting weighted graph by $G_{q,\ell-1}(x)$.

As shown in [1], a vector x is feasible (i.e., it equals the profile vector of some string $s \in \Sigma^n$, for some n) if and only if the graph obtained from $G_{q,\ell-1}(x)$ by replacing each edge of weight $w(e)$ with $w(e)$ parallel edges contains an Eulerian circuit (i.e., a tour passing through each edge exactly once). Note that if $w(e) = 0$, the corresponding edge is omitted.

A weighted digraph is balanced if, for every vertex v, the total weight of the incoming edges equals the total weight of the outgoing edges—that is, $w(E_{in}(v)) = w(E_{out}(v))$.

Lemma II.4. *[1] The vector $x \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$ is feasible if and only if $G_{q,\ell-1}(x)$ is balanced;*

Proof. The vector x is feasible if and only if the multigraph derived from $G_{q,\ell-1}(x)$ where edge multiplicities are assigned according to x, admits an Eulerian circuit. This occurs if and only if, for every vertex v in $G_{q,\ell-1}(x)$, the total weight of incoming edges equals the total weight of outgoing edges, i.e., $w(E_{in}(v)) = w(E_{out}(v))$. This condition ensures that $G_{q,\ell-1}(x)$ is balanced. □

Example II.5. *Let $q = 4$ and $\ell = 2$. Figure 1 presents the De Bruijn graph $G_{4,1}(x)$ for feasible vector $x = (15, 2, 4, 5, 1, 0, 8, 9, 3, 6, 13, 14, 7, 10, 11, 12)$ as shown in Example II.2. As the figure shows, each vertex v in the graph is balanced.*

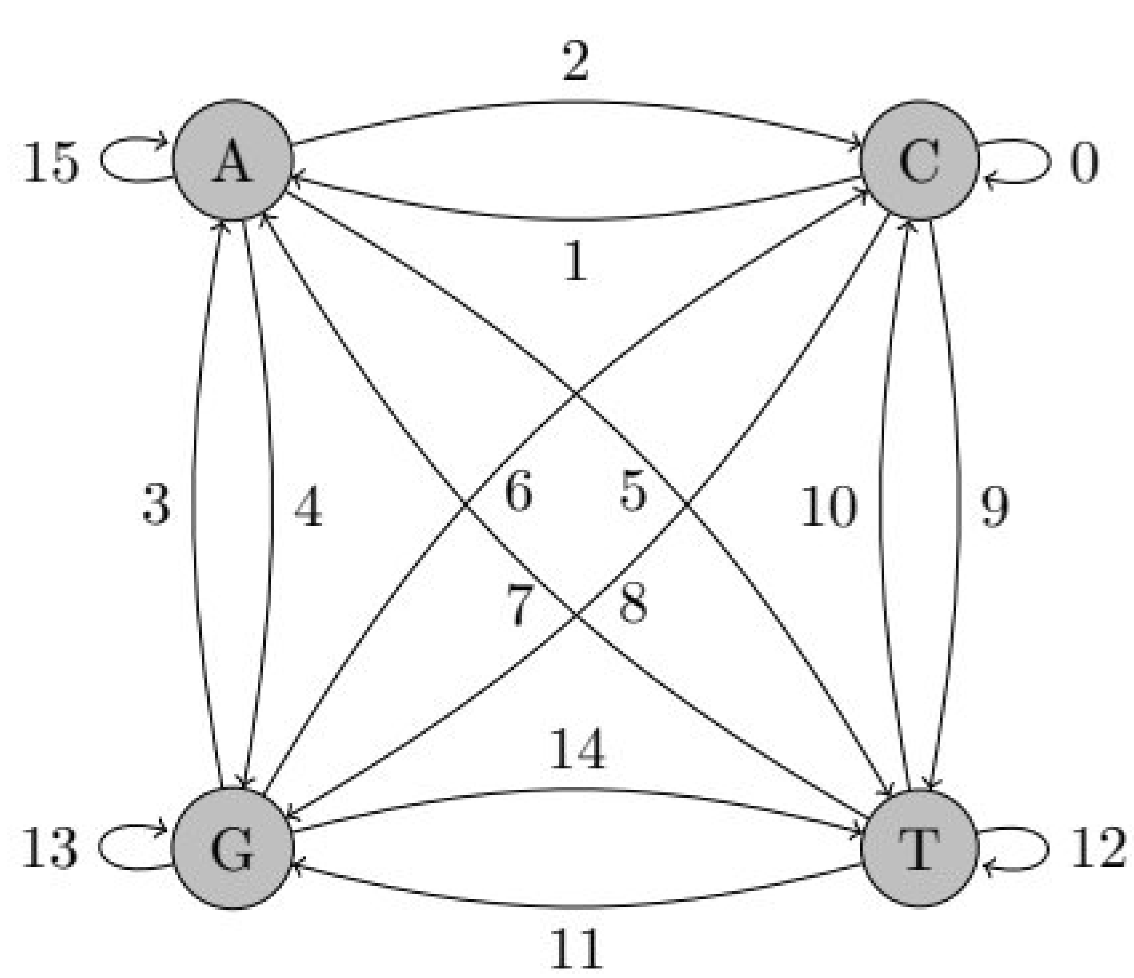

Fig. 1. The weighted De Bruijn graph $G_{4,1}(x)$ corresponding to vector $x = (15, 2, 4, 5, 1, 0, 8, 9, 3, 6, 13, 14, 7, 10, 11, 12)$, which was introduced in Example II.2.

In the rank modulation method, instead of using the profile vector, permutations preserving the order of entries of that vector, are used. In this representation, clearly, we need that the entries of the profile vector be distinct. Let A be a finite set. We denote by S_A the set of all permutations on A. Each permutation $\pi \in S_A$ may be viewed as a bijection $A \longrightarrow [|A|]$ that assigns to each element of A its unique rank in the ordering defined by π.

Definition II.6. *[1] Let $\pi \in S_{\Sigma^\ell}$ be a permutation and $x \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$ be a vector indexed by the elements in Σ^ℓ. We say that x satisfies π, denoted as $x \models \pi$, if the entries of x are distinct and the following relationship holds:*

$$\pi(w) < \pi(w') \iff x(w) < x(w'), \qquad \forall w, w' \in \Sigma^\ell.$$

Definition II.7 (feasible permutation)**.** *The permutation $\pi \in S_{\Sigma^\ell}$ is called feasible if there exists a feasible vector $x \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$ such that $x \models \pi$. Equivalently, π is feasible if there exists a string $s \in \Sigma^n$ (for some n) such that $p_{s,\ell} \models \pi$.*

Example II.8. *Consider the feasible vector $x = (15, 2, 4, 5, 1, 0, 8, 9, 3, 6, 13, 14, 7, 10, 11, 12)$ introduced in Example II.5. The feasible permutation corresponding to x is*

$$\pi = \begin{array}{cccccccccc} AA & AC & AG & AT & CA & CC & CG & CT & GA & GC \\ [16 & 3 & 5 & 6 & 2 & 1 & 9 & 10 & 4 & 7 \\ & & & & GG & GT & TA & TC & TG & TT \\ & & & & 14 & 15 & 8 & 11 & 12 & 13] \end{array}$$

Following previous work, we denote the set of all feasible permutations over Σ^ℓ by $\Phi_{q,\ell}$ and the size of $\Phi_{q,\ell}$, by $F_{q,\ell}$.

B. Hyperplane Arrangements

Let K be a field and V be a vector space of dimension n over K. In fact, as vector spaces, we have $V \cong K^n$. A linear hyperplane is a subspace of dimension $n-1$ in V, which divides the space V into two half-spaces [11]. For any linear hyperplane H, there exists a vector $\mathfrak{a} \in V$ such that H can be represented as $H = \{x \in V \mid \mathfrak{a}x^T = 0\}$; here $\mathfrak{a}$ is a normal vector and the offset is zero. A general or affine hyperplane J in V is a translate of a linear hyperplane H, i.e., $J = \{x \in V \mid \mathbf{a}x^T = b\}$, where b is a nonzero offset.

Definition II.9. *[11] A finite hyperplane arrangement $\mathcal{A}$ for V, is a finite set of hyperplanes in V. A central arrangement is an arrangement where all hyperplanes are linear. The dimension of $\mathcal{A}$, denoted by $dim(\mathcal{A})$, is defined to be $dim(V) = n$. The rank of $\mathcal{A}$, denoted by $rank(\mathcal{A})$, is the dimension of the space spanned by the normal vectors of the hyperplanes in $\mathcal{A}$.*

In the rest of the paper, we consider the case where $K = \mathbb{R}$. In this case, a region of an arrangement $\mathcal{A}$ is a connected component of X, where X is the complement of the union of hyperplanes in $\mathcal{A}$, i.e.

$$X = V \setminus \bigcup_{H \in \mathcal{A}} H.$$

Let $\mathcal{R}(\mathcal{A})$ stand for the set of regions of $\mathcal{A}$. We denote by $r(\mathcal{A})$ the size of $\mathcal{R}(\mathcal{A})$, i.e.

$$r(\mathcal{A}) = |\mathcal{R}(\mathcal{A})|.$$

As a well-known example of hyperplane arrangement, we mention to the *braid arrangement* $\mathcal{B}_n$ in $V = \mathbb{R}^n$, consist of $\binom{n}{2}$ hyperplanes $w_{i,j}$, where $w_{i,j} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i - x_j = 0\} = \ker(e_i - e_j)$, where e_t is the t-th unit vector for any $t \in [n]$.

Corollary II.10. *[11] We have $r(\mathcal{B}_n) = n!$.*

Proof. A point $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ lies on one side of the hyperplane $w_{i,j}$ (where points on $w_{i,j}$ satisfy $x_i = x_j$

) if and only if it satisfies either $x_i < x_j$ or $x_i > x_j$. Consequently, by considering all hyperplanes $\{w_{i,j} \mid 1 \leqslant i < j \leqslant n\}$, each region $\mathfrak{R} \in \mathcal{R}(\mathcal{B}_n)$ consists of points with a specific order, i.e. we have:

$$\forall\, \mathfrak{R} \in \mathcal{R}(\mathcal{B}_n) \quad \exists\, \sigma \in S_n, \quad s.t. \\ \forall x \in \mathfrak{R}, \quad x_{\sigma(1)} < x_{\sigma(2)} < \cdots < x_{\sigma(n)}$$

In other words, each permutation $\sigma \in S_n$ corresponds to a region, and the points within this region are ordered according to σ. Therefore $r(\mathcal{B}_n) = |S_n| = n!$. ☐

In general, the number of regions formed by a hyperplane arrangement $\mathcal{A}$ is given by the Thomas Zaslavsky formula.

Theorem II.11. *[11, Theorem 2.5] Let $\mathcal{A}$ be an arrangement in a n-dimensional real vector space V. Then we have*

$$r(\mathcal{A}) = (-1)^n \chi_{\mathcal{A}}(-1), \tag{1}$$

where

$$\chi_{\mathcal{A}}(t) := \sum_{\substack{B \subseteq \mathcal{A} \\ B \text{ is central}}} (-1)^{|B|} t^{n-rank(B)}.$$

Equivalently we have $r(\mathcal{A}) = \sum_{\substack{B \subseteq \mathcal{A} \\ B \text{ central}}} (-1)^{|B|-rank(B)}$.

Lemma II.12. *Let $T : V \longrightarrow \mathbb{R}^n$ be a bijective linear transformation, where V is a n-dimensional real vector space. Let $\mathcal{A}$ be an arrangement of hyperplanes in V. Then $T(\mathcal{A})$ is an arrangement in $T(\mathcal{A})$ and $r(\mathcal{A}) = r(T(\mathcal{A}))$.*

Proof. Let $\mathcal{A} = \{H_1, \ldots, H_k\}$ be an arrangement of hyperplanes in V. Since T is a bijective linear transformation, each image $T(H_i)$ is a hyperplane in $\mathbb{R}^n$. Hence, $T(\mathcal{A}) = \{T(H_1), \ldots, T(H_k)\}$ is a hyperplane arrangement in $\mathbb{R}^n$.
Let $U := \bigcup_{i=1}^k H_i$ and $U' := \bigcup_{i=1}^k T(H_i)$. Since T is bijective and linear, then both T and T^{-1} are continuous, and therefore T is a homeomorphism (continuous bijection with continuous inverse [7]). In particular, T preserves connectedness.
The regions of $\mathcal{A}$ are the connected components of $V \setminus U$, while the regions of $T(\mathcal{A})$ are the connected components of $\mathbb{R}^n \setminus U'$. Since $T(U) = U'$, the restriction

$$T\,|_{V\setminus U}\colon V \setminus U \longrightarrow \mathbb{R}^n \setminus U',$$

is a homeomorphism. Consequently, it induces a bijection between the sets of connected components. Hence,

$$r(\mathcal{A}) = r(T(\mathcal{A})),$$

which completes the proof. ☐

C. The Null Space of a Directed Graph

In this section, we consider a directed graph G and its associated incidence matrix M_G. Inspired by the work presented in [2] and [4], we propose a method for finding a basis for the null space of the incidence matrix M_G.

Definition II.13 (Incidence Matrix)**.** *The incidence matrix M_G of a directed graph G is defined with rows corresponding to vertices and columns to edges of G. The entry $M_G[i, j]$ of M_G is -1 if the i-th vertex is the initial vertex of the j-th edge, 1 if it is the terminal vertex, and 0 otherwise.*

For example, the following matrix is the incidence matrix of the De Bruijn graph $G_{4,1}$. Since loop edges are both input and output edges of the same vertex, we consider the entries corresponding to loop edges of each vertex to be zero. Note that the vertex set and the edge set of $G_{4,1}$ are Σ and Σ^2, respectively, which correspond to the labels of the rows and columns of $M_{G_{4,1}}$ shown in Figure II-C.

Definition II.14 (Null Space)**.** *The null space of a matrix $A \in \mathbb{R}^{m\times n}$ is defined to be the set of all vectors $x \in \mathbb{R}^n$ satisfying $Ax^T = \boldsymbol{0}$, where $\boldsymbol{0}$ is the zero vector. The null space of a directed graph G, denoted by $\ker(G)$, is defined to be the null space of its incidence matrix.*

The following theorem presents the dimension of $\ker(G)$.

Theorem II.15. *[4, Theorems 8.2.1 and 8.3.1] Let G be a directed graph with the set of vertices V and the set of edges E. If G has c components, then* $\dim(\ker(G)) = |E| - |V| + c$.

In the rest of this part, we present a method for constructing a special basis for $\ker(G)$. We assume that G is a connected directed graph (digraph) with the set of edges E. Also, we assume that T is a spanning tree of G. The complement of the spanning tree T, denoted as $E \setminus T$, is referred to as the cotree and is represented by $\overline{T}$. Accordingly, we define the fundamental cycle as follows:

Definition II.16 (fundamental cycle)**.** *[2] For every edge $\boldsymbol{a} \in \overline{T}$ of a graph G, there exists a unique path within the spanning tree T such that the graph $T + \boldsymbol{a}$ contains exactly one cycle. This cycle is referred to as the fundamental cycle of G with respect to T and the edge $\boldsymbol{a}$. For convenience, we denote this cycle by $C_{\boldsymbol{a}}$.*

To find the elements of the basis of $\ker(G)$, we consider a cycle $\mathcal{C}$ and define a function that maps the edges of G to a vector as follows:

Definition II.17. *Let $\mathcal{C}$ be a cycle, equipped with a specified traversal direction. An arc (directed edge) of $\mathcal{C}$ is classified as a forward arc if its direction matches the traversal direction of $\mathcal{C}$; otherwise, it is classified as a reverse arc. We denote the sets of forward and reverse arcs of $\mathcal{C}$ by $\mathcal{C}^+$ and $\mathcal{C}^-$, respectively. Additionally, we define a function $f_{\mathcal{C}} : E \longrightarrow \{0, \pm 1\}$ associated with $\mathcal{C}$ as follows:*

$$f_{\mathcal{C}}(e) = \begin{cases} 1 & if\ e \in \mathcal{C}^+ \\ -1 & if\ e \in \mathcal{C}^- \\ 0 & if\ e \notin \mathcal{C} \end{cases} \qquad \forall e \in E$$

Let $\mathbf{a}$ be an arc in $\overline{T}$. The fundamental cycle $C_{\mathbf{a}}$ associated with $\mathbf{a}$ defines a vector $f_{C_{\mathbf{a}}}$ such that $f_{C_{\mathbf{a}}}(\mathbf{a}) = 1$. For simplicity, we denote this vector by $\mathcal{F}_{\mathbf{a}}$. The following theorem characterizes a basis for the null space $\ker(G)$.

$$M_{G_{4,1}} = \begin{array}{c} \\ A \\ C \\ G \\ T \end{array} \begin{array}{c} \begin{array}{cccccccccccccccc} AA & AC & AG & AT & CA & CC & CG & CT & GA & GC & GG & GT & TA & TC & TG & TT \end{array} \\ \begin{bmatrix} 0 & -1 & -1 & -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -1 & -1 & -1 & 0 \end{bmatrix} \end{array}$$

Fig. 2. The incidence matrix of the De Bruijn graph $G_{4,1}$.

Theorem II.18. *[2, Theorem 20.8] Let G be a connected digraph, and let T be a spanning tree of G. Then, the set $\{\mathcal{F}_{\boldsymbol{a}} : \boldsymbol{a} \in \overline{T}\}$, which is associated with the fundamental cycles of G with respect to T, forms a basis for* $\ker(G)$.

Definition II.19. *Let B be the matrix whose columns are vectors $\{\mathcal{F}_{\boldsymbol{a}} : \boldsymbol{a} \in \overline{T}\}$. The matrix B is called the base matrix of* $\ker(G)$.

The rows and columns of B correspond to the arcs in E and $\overline{T}$, respectively. According to Theorem II.15, the base matrix B has $|E|$ rows and $|E| - |V| + 1$ columns. Let us give an example.

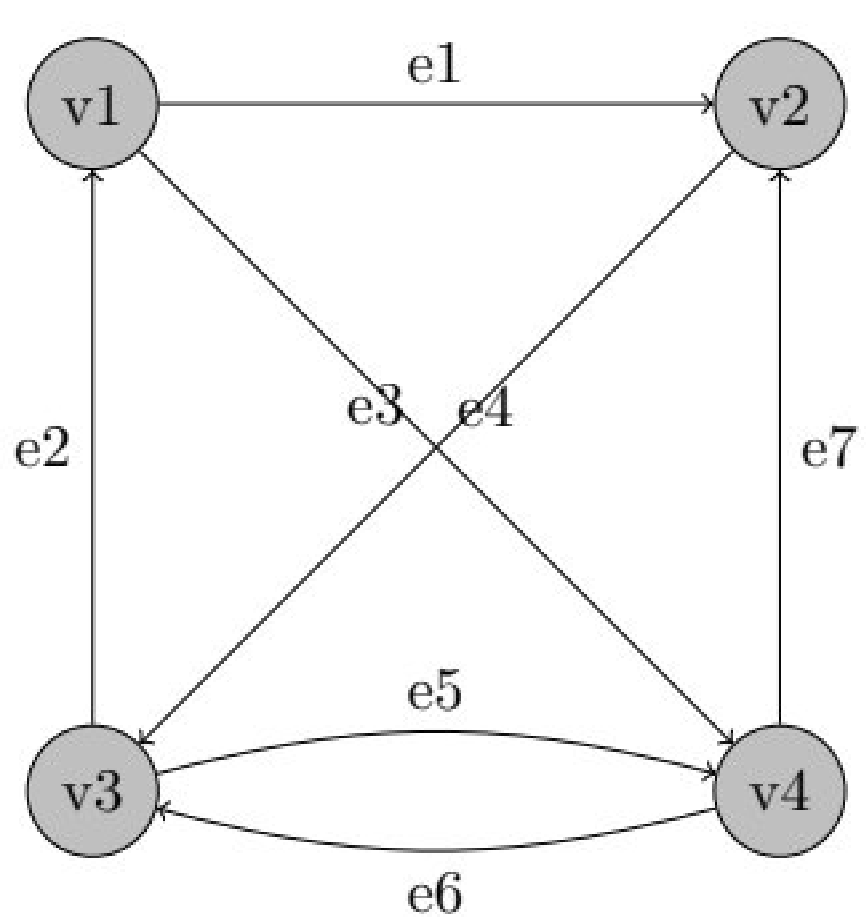

Fig. 3. An example graph with labeled edges.

Example II.20. *For the digraph G presented in Figure 3 and spanning tree $T = \{e1, e2, e3\}$, the fundumental cycle corresponding to the arc $\boldsymbol{a} = e5 \in \overline{T}$ is $C_{\boldsymbol{a}} = \{e2, e3, e5\}$ and the associated function is $\mathcal{F}_{e5} = (0, -1, -1, 0, 1, 0, 0)$. Therefore, by forming the set $\{\mathcal{F}_{e_4}, \mathcal{F}_{e_5}, \mathcal{F}_{e_6}, \mathcal{F}_{e_7}\}$, the base matrix B of the incidence matrix of the graph G is as follow:*

$$B = \begin{array}{c} \\ e1 \\ e2 \\ e3 \\ e4 \\ e5 \\ e6 \\ e7 \end{array} \begin{array}{c} \begin{array}{cccc} e4 & e5 & e6 & e7 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & -1 \\ 1 & -1 & 1 & 0 \\ 0 & -1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

III. The Number of Feasible Permutions

Our idea, for computing the number of feasible permutations, is working with the incidence matrix of De Bruijn graphs $G_{q,\ell-1}$. In what follows, we will explain the idea in details.

A. The Base Matrix correponding to the De Bruijn Graph

As previously mentioned, to determine the feasibility of the vector x, we need to check that the weighted De Bruijn graph $G_{q,\ell-1}(x)$ corresponding to x is balanced. Without loss of generality, we will continue our analysis by focusing on vectors $x \in (\mathbb{R})^{\Sigma^\ell}$, rather than $x \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$.

Definition III.1 (Balancing Vector)**.** *A vector $x \in (\mathbb{R})^{\Sigma^\ell}$ is called a **balancing vector** if the weighted De Bruijn graph $G_{q,\ell-1}(x)$ is balanced.*

Remark III.2. (1) *According to Lemma II.4, every feasible vector $x \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$ is itself a balancing vector.*
(2) *If x_b is a balancing vector, then for each $c \in \mathbb{R}^{>0}$ the vectors cx_b and $x_b + \overrightarrow{c}$ are also balancing vectors with the same order of entries as in x_b, where $\overrightarrow{c}$ denotes the constant vector whose entries are all equal to c.*

Lemma III.3. *Suppose $M_{G_{q,\ell-1}}$ is the incidence matrix of $G_{q,\ell-1}$, then the vector $x \in (\mathbb{R})^{\Sigma^\ell}$ is balancing vector if and only if $M_{G_{q,\ell-1}} x^T = 0$.*

Proof. According to the definition of the incidence matrix (Definition II.13), the inner product of x with a row of the matrix $M_{G_{q,\ell-1}}$, corresponding to a vertex v of $G_{q,\ell-1}$, equals zero if and only if the sum of the weights of the incoming edges to v equals the sum of the weights of the outgoing edges from v. Therefore, $M_{G_{q,\ell-1}} x^T = 0$ if and only if $G_{q,\ell-1}(x)$ is balanced, i.e. x is balancing vector. □

For instance, as shown in Example II.5, the vector $x = (15, 2, 4, 5, 1, 0, 8, 9, 3, 6, 13, 14, 7, 10, 11, 12)$ is balancing vector because $M_{G_{4,1}} x^T = 0$. According to Lemma III.3 and Definition II.13, we conclude that the vector $x \in (\mathbb{R})^{\Sigma^\ell}$ is balancing vector if and only if x belongs to the null space of the incidence matrix $M_{G_{q,\ell-1}}$, i.e. $x \in \ker(G_{q,\ell-1})$.

Remark III.4. *Since the entries of x indexed by loops (e.g., $AA, CC, \ldots$) do not affect in balancing the weighted De Bruijn graph $G_{q,\ell-1}(x)$, we can ignore loops in the graph $G_{q,\ell-1}$. Let $\tilde{G}_{q,\ell-1}$ be the De Bruijn graph $G_{q,\ell-1}$ whose q loops have been removed. Clearly the incidence matrix $M_{\tilde{G}_{q,\ell-1}}$ can be obtained by removing the zero columns associated with loop edges of $M_{G_{q,\ell-1}}$.*

We rewrite Lemma III.3 for $\tilde{G}_{q,\ell-1}$.

Corollary III.5. *Let $x \in (\mathbb{R})^{\Sigma^\ell}$ and $\tilde{x} \in (\mathbb{R})^{\tilde{\Sigma}^\ell}$, where the vector $\tilde{x}$ is obtained from x by removing q entries of the vector x corresponding to loops. Then:*

$$M_{G_{q,\ell-1}} x^T = 0 \iff M_{\tilde{G}_{q,\ell-1}} \tilde{x}^T = 0.$$

Proof. The proof follows from the fact that columns of $M_{G_{q,\ell-1}}$ corresponding to loops are zero vectors. $\square$

Similar to the definition of feasible permutations on the set of edges of $G_{q,\ell-1}$ one can define feasible permutations on the set of edges of $\tilde{G}_{q,\ell-1}$. We denote this set of permutations by $\tilde{\Phi}_{q,l}$. The following Lemma shows the relation between $|\Phi_{q,l}|$ and $|\tilde{\Phi}_{q,l}|$.

Lemma III.6. *We have*

$$F_{q,\ell} = |\tilde{\Phi}_{q,\ell}| \frac{q^\ell!}{(q^\ell - q)!} \tag{2}$$

Proof. The number of possible positions for loops on each feasible permutation considered on the set of edges of graph $\tilde{G}_{q,\ell-1}$ is $\dfrac{q^\ell!}{(q^\ell - q)!}$ [1, Corollary 3]. Therefore, the total number of feasible permutations on $G_{q,\ell-1}$ (with the addition of loops) is obtained. $\square$

Based on Lemma III.6, for computing $F_{q,\ell} = |\Phi_{q,\ell}|$ we can compute $|\tilde{\Phi}_{q,l}|$. Hence we concentrate on the graph $\tilde{G}_{q,\ell-1}$ instead of $G_{q,\ell-1}$. In this regard, we obtain the dimension of $\ker(\tilde{G}_{q,\ell-1})$ in the following corollary, using Theorem II.15.

Corollary III.7. *We have* $\dim(\ker(\tilde{G}_{q,\ell-1})) = q^\ell - q^{\ell-1} - q + 1$.

Proof. Clearly, $\tilde{G}_{q,\ell-1}$ has $q^\ell - q$ edges, $q^{\ell-1}$ vertices and is a connected graph. Consequently we can conclude from Theorem II.15 that

$$\begin{aligned}\dim(\ker(\tilde{G}_{q,\ell-1})) &= |E(\tilde{G}_{q,\ell-1})| - |V(\tilde{G}_{q,\ell-1})| + 1\\ &= (q^\ell - q) - q^{\ell-1} + 1\\ &= q^\ell - q^{\ell-1} - q + 1,\end{aligned}$$

as desired. $\square$

Notation: In the remainder of this paper, we focus on the special case $\ell = 2$.

Corollary III.8. *We have* $\dim(\ker(\tilde{G}_{q,1})) = q^2 - 2q + 1$

Proof. The proof follows directly from Corollary III.7 and letting $l = 2$. $\square$

According to Lemma III.3 and Corollary III.5, all vectors in $\ker(\tilde{G}_{q,1})$ are balancing, and all balancing vectors are in $\ker(\tilde{G}_{q,1})$. Thus to generate $\ker(\tilde{G}_{q,1})$, we find a basis for this space, by applying the method presented in Section II-C.

Let $E = \{e_{i,j}, 1 \le i,j \le q\}$ represent the edge set of $G_{q,1}$ and $\tilde{E} = E \setminus \{e_{i,i}, 1 \le i \le q\}$ denote the set of edges of $\tilde{G}_{q,1}$. Also, consider the spanning tree $T = \{e_{i,i+1} | 1 \le i \le q-1\}$ of the graph $G_{q,1}$. Then we have $\overline{T} = \tilde{E} \setminus T$. Therefore, according to the method presented in Section II-C, we construct the base matrix of $\tilde{G}_{q,1}$ and denote it by $B_{q,1}$. According to Corollary III.8, $B_{q,1}$ has $q^2 - q$ rows and $q^2 - 2q + 1$ columns. Columns of $B_{q,1}$, form a basis for the space $\ker(\tilde{G}_{q,1})$. To represent the matrix in a special form, we label the first $q^2 - 2q + 1$ rows and the last $q - 1$ rows of $B_{q,1}$, corresponding to the edges in $\overline{T}$ and T, respectively.

Specifically, the rows of $B_{q,1}$ are indexed in the following fixed order. The first $q^2 - 2q + 1$ rows correspond to $\overline{T}$ and are indexed by $e_{i+1,i}$, $1 \le i \le q-1$ $(e_{2,1}, \ldots, e_{q,q-1})$, followed by $e_{i,i+j}, e_{i+j,i}$, $2 \le j \le q-1$, $1 \le i \le q-j$, that is,

$$e_{1,3}, e_{3,1}, e_{2,4}, e_{4,2}, \ldots, e_{q-2,q}, e_{q,q-2}, e_{1,4}, e_{4,1}, \ldots, e_{1,q}, e_{q,1}.$$

The last $q-1$ rows correspond to T and are indexed by $e_{i,i+1}$, $1 \le i \le q-1$ $(e_{1,2}, \ldots, e_{q-1,q})$. The columns are labeled according to the same index order as the first $q^2 - 2q + 1$ rows corresponding to $\overline{T}$. With this indexing method, the basis matrix $B_{q,1}$ would be of the following form.

$$B_{q,1} = \left[\begin{array}{c|c} \multicolumn{2}{c}{I_{q^2-2q+1}} \\ \hline I_{q-1} & D \end{array}\right],$$

where D is a $(q-1) \times (q^2 - 3q + 2)$ matrix.

Remark III.9. *Note that the columns in the submatrix D correspond to the indices $e_{i,i+j}, e_{i+j,i}$, $2 \le j \le q-1$, $1 \le i \le q-j$. The nonzero entries of the i-th row of D correspond to the fundamental cycles $C_{\boldsymbol{a}}$, where*

$$\boldsymbol{a} \in \{e_{i,i+j}\ , e_{i+j,i} \mid 2 \le j \le q-1,\ 1 \le i \le q-j\},$$

such that $e_{i,i+1} \in C_{\boldsymbol{a}}$. Clearly, if $e_{i,i+1} \in C_{e_{i,i+j}}$, then $e_{i,i+1} \in C_{e_{i+j,i}}$. Since $e_{i,i+j}$ and $e_{i+j,i}$ have reverse directions, according to Definition II.19, their corresponding entries in the i-th row of D are -1 and $+1$, respectively. Therefore, each row in D contains equal numbers of entries $+1$ and -1.

Example III.10. *Let $q = 4$ and $T = \{e_{1,2}, e_{2,3}, e_{3,4}\}$. Then the base matrix for $\ker(\tilde{G}_{4,1})$ has the following form:*

$$B_{4,1} =$$

$$\begin{array}{cc} & \begin{array}{ccccccccc} e_{2,1} & e_{3,2} & e_{4,3} & e_{1,3} & e_{3,1} & e_{2,4} & e_{4,2} & e_{1,4} & e_{4,1} \end{array} \\ \begin{array}{c} e_{2,1}\\ e_{3,2}\\ e_{4,3}\\ e_{1,3}\\ e_{3,1}\\ e_{2,4}\\ e_{4,2}\\ e_{1,4}\\ e_{4,1}\\ -\\ e_{1,2}\\ e_{2,3}\\ e_{3,4} \end{array} & \begin{bmatrix} 1 & 0 & \cdots & & & & & & 0\\ 0 & 1 & \cdots & & & & & & 0\\ & & \ddots & & & & & & \vdots\\ \vdots & & & \ddots & & & & & 0\\ 0 & & & & & & & & \vdots\\ \vdots & & & & & & & & \\ & & & & & & & & \\ 0 & & & & & & & & 0\\ 0 & & \cdots & & & & & & 1\\ - & - & - & - & - & - & - & - & -\\ 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1\\ 0 & 1 & 0 & -1 & 1 & -1 & 1 & -1 & 1\\ 0 & 0 & 1 & 0 & 0 & -1 & 1 & -1 & 1 \end{bmatrix} \end{array} \tag{3}$$

B. Enumeration of Feasible Permutations via Hyperplane Arrangements

In what follows, by examining the structure of $\ker(\tilde{G}_{q,1})$, we propose a method for computing $|\tilde{\Phi}_{q,2}|$ and then conclude the number of feasible permutations for $\ell = 2$, i.e. $F_{q,2} = |\ \Phi_{q,2}\ |$.

Lemma III.11. *Consider $x \in (\mathbb{R})^{\tilde{\Sigma}^2}$ and let $B_{q,1}$ be the base matrix of $\ker(\tilde{G}_{q,1})$. If x is balancing, then there exists a vector $\alpha \in \mathbb{R}^{q^2-2q+1}$, such that $x = B_{q,1}\alpha^T$.*

Proof. Consider $\tilde{G}_{q,1}$ and its incidence matrix $M_{\tilde{G}_{q,1}}$. As stated in Corollary III.5, if x is balancing, then $M_{\tilde{G}_{q,1}} x^T = 0$ which yields $x \in \ker(\tilde{G}_{q,1})$. In other words, if $B_{q,1}$ is the base matrix of $\ker(\tilde{G}_{q,1})$, then x can be expressed as a linear combination of the columns of $B_{q,1}$. Therefore,

$$\exists\ \alpha = (\alpha_1, \ldots, \alpha_{q^2-2q+1})\ s.t.\ x = B_{q,1}\alpha^T.$$

□

By considering $B_{q,1}$ in the form $B_{q,1} = \left[\begin{array}{c|c} \multicolumn{2}{c}{I_{q^2-2q+1}} \\ \hline I_{q-1} & D \end{array}\right]$, as discussed in the previous section, each balancing vector $x = B_{q,1}\alpha^T \in \ker(\tilde{G}_{q,1})$ is of the form

$$x = (\alpha_1, \ldots, \alpha_{q^2-2q+1}, f_1^q, f_2^q, \ldots, f_{q-1}^q), \tag{4}$$

where, each f_i^q, $i \in [q-1]$, is a linear combination of the components $\alpha_1, \ldots, \alpha_{q^2-2q+1}$. For instance, f_1^4, corresponding to the 10th row of the base matrix $B_{4,1}$ presented in Example III.10, can be written as $f_1^4 = \alpha_1 - \alpha_4 + \alpha_5 - \alpha_8 + \alpha_9$.

As mentioned before, in the rank modulation scheme, we focus on feasible vectors with distinct entries. To this end, we consider balancing vectors of the form $x = B_{q,1}\alpha^T$ whose entries are distinct. To investigate the distinction between entries, we define the following sets:

$$W_{i,j} := \ker(\tilde{G}_{q,1}) \cap \ker(e_i - e_j), \qquad 1 \le i < j \le q^2 - q, \tag{5}$$

where e_t is the t-th unit vector for any $t \in [q^2-q]$. We note that, according to Corollary III.8, the dimension of the real vector space $V := \ker(\tilde{G}_{q,1})$ is $q^2 - 2q + 1$. Since $q \ge 3$, there exists a point $p \in \ker(e_i - e_j)$ such that $\tilde{G}_{q,1}(p)$ is not balanced, i.e. $p \notin \ker(\tilde{G}_{q,1})$. Thus, $\ker(e_i - e_j) \nsubseteq \ker(\tilde{G}_{q,1})$. Therefore, each $W_{i,j}$ is a linear subspace of V of dimension $q^2 - 2q$, and hence is a linear hyperplane.

Lemma III.12. *A balancing vector $x \in (\mathbb{R})^{\tilde{\Sigma}^2}$ has distinct entries if and only if:*

$$x \in \ker(\tilde{G}_{q,1}) \setminus \bigcup_{1 \le i < j \le q^2-q} W_{i,j}. \tag{6}$$

Proof. The vector x is balancing if and only if $x \in \ker(\tilde{G}_{q,1})$, the proof follows easily from the definition of hyperplanes $W_{i,j}$. □

We now define the hyperplane arrangement $\mathcal{A}$ for $\ker(\tilde{G}_{q,1})$, as follows.

$$\mathcal{A} := \{W_{i,j} \mid 1 \le i < j \le q^2 - q\}. \tag{7}$$

We note that some hyperplanes may coincide, so $\mathcal{A}$ contains at most $\binom{q^2-q}{2}$ elements. By Lemma III.12, x is balancing with distinct entries if and only if $x \in \ker(\tilde{G}_{q,1}) \setminus \bigcup \mathcal{A}$. As stated in Section II-B, the arrangement $\mathcal{A}$ partitions the space $\ker(\tilde{G}_{q,1})$ into $r(\mathcal{A})$ separate regions.

Remark III.13. *Note that $\ker(\tilde{G}_{q,1})$ is a subspace of $\mathbb{R}^{q^2-q}$ with a basis (given by the columns of $B_{q,1}$) consisting of rational vectors. Hence $\ker(\tilde{G}_{q,1}) \cap \mathbb{Q}^{q^2-q}$ is dense in $\ker(\tilde{G}_{q,1})$ in the topological sense. Let $\mathfrak{R} \in R(\mathcal{A})$ a region of the arrangement $\mathcal{A}$ (defined in Eq.(7)) consisting of some balancing vectors, and let $p \in \mathfrak{R}$. Since $\mathfrak{R}$ is open and connected in $\ker(\tilde{G}_{q,1})$ and $\ker(\tilde{G}_{q,1}) \cap \mathbb{Q}^{q^2-q}$ is dense in $\ker(\tilde{G}_{q,1})$, it follows that*

$$\mathfrak{R} \cap \mathbb{Q}^{q^2-q} \neq \varnothing.$$

Thus there exists a rational balancing vector $p_{\mathbb{Q}} \in \mathfrak{R} \cap \mathbb{Q}^{q^2-q}$.

It can be shown that each region corresponds to a unique feasible permutation.

Theorem III.14. *Consider the hyperplane arrangement $\mathcal{A}$ (Eq.7) with the set of its regions, $R(\mathcal{A})$. Then:*

$$r(\mathcal{A}) = |R(\mathcal{A})| = |\widetilde{\Phi}_{q,2}|.$$

Proof. Let $\mathfrak{R}_k$, $1 \le k \le r(\mathcal{A})$ represent the k-th region of $R(\mathcal{A})$. Each region $\mathfrak{R}_k$ is a connected component of $\ker(\tilde{G}_{q,1}) \setminus \bigcup \mathcal{A}$, and by Lemma III.12, each point $p = (p_1, \ldots, p_{q^2-q}) \in \mathfrak{R}_k$ is a balancing vector.

Each hyperplane $W_{i,j}$ divides the space into two half-spaces, $x_i < x_j$ and $x_i > x_j$, it follows that, depending on which side of the half-space $\mathfrak{R}_k$ is located in, we have either $p_i < p_j$ or $p_i > p_j$ for all $p \in \mathfrak{R}_k$. Since the region $\mathfrak{R}_k$ is bounded by the hyperplanes $W_{i,j}$, $1 \le i < j \le q^2 - q$, the components of the balancing vector $p \in \mathfrak{R}_k$ satisfy in the specific linear order. In other words:

$$\exists\ \sigma_k \in S_{q^2-q}, \quad s.t\ \ , p_{\sigma_k(1)} < p_{\sigma_k(2)} < \ldots < p_{\sigma_k(q^2-q)}. \tag{8}$$

In general, the components of every point in $\mathfrak{R}_k$ follow the same linear order (8).

Based on Remark III.13, there exists a rational balancing vector $p_{\mathbb{Q}} \in \mathfrak{R}_k$ corresponding to p. This $p_{\mathbb{Q}}$ has the same order (8). Then, according to Remark III.2 (part (2)), by first adding a suitable positive scalar vector to eliminate negative entries and then multiplying by a suitable positive integer to clear denominators, we obtain a non-negative balancing vector $\mathbf{x} \in (\mathbb{N} \cup \{0\})^{q^2-q}$ that preserves the strict inequalities of the order (8). Hence there exists a feasible vector $\mathbf{x}$ such that

$$\mathbf{x}_{\sigma_k(1)} < \mathbf{x}_{\sigma_k(2)} < \ldots < \mathbf{x}_{\sigma_k(q^2-q)}.$$

Consequently, by Definition II.7, we can associate a unique permutation $\pi_k = \sigma_k^{-1}$ to $\mathfrak{R}_k$ as the feasible permutation satisfying $\mathbf{x} \models \pi_k$. Thus each region in $R(\mathcal{A})$ corresponds to a unique feasible permutation in $\tilde{\Phi}_{q,2}$.

Conversely, if $\pi \in \widetilde{\Phi}_{q,2}$ is a feasible permutation, then there exists a feasible vector $\mathbf{x}$ such that $\mathbf{x} \models \pi$. By Remark III.2 (part.(1)) and Lemma III.12, this implies that $\mathbf{x} \in \ker(\tilde{G}_{q,1}) \setminus \bigcup \mathcal{A}$, and thus $\mathbf{x}$ belongs to one of the regions. Therefore, there exists $k \in [r(\mathcal{A})]$ such that $\mathbf{x} \in \mathfrak{R}_k$. Therefore, we conclude that there exists a bijection between the set of regions $R(\mathcal{A})$ and the set of feasible permutations $\widetilde{\Phi}_{q,2}$. Hence, we have:

$$|R(\mathcal{A})| = |\widetilde{\Phi}_{q,2}|.$$

□

Remark III.15. *Theorem III.14 can also hold for $\ell > 2$.*

As we mentioned, $W_{i,j}$ is a linear hyperplane of dimension $q^2 - 2q$ in the real vector space $V = \ker(\tilde{G}_{q,1})$. Since V has dimension $q^2 - 2q + 1$, it is isomorphic to the vector space $\mathbb{R}^{q^2-2q+1}$. In what follows we will obtain the form of $W_{i,j}$ as a linear hyperplane in $\mathbb{R}^{q^2-2q+1}$ and denote it by $W'_{i,j}$. Assume that d_i is the i-th row of base matrix $B_{q,1}$ ($i \in [q^2 - q]$). Then, the i-th component of $x = B_{q,1}\alpha^T$, i.e. x_i, is equal to $d_i \alpha^T$. Therefore, $W_{i,j}$ is isomorphic to the linear hyperplane in $\mathbb{R}^{q^2-2q+1}$, i.e., $W'_{i,j}$, defined as follows:

$$W'_{i,j} := \{\alpha \in \mathbb{R}^{q^2-2q+1} \mid (d_i - d_j)\alpha^T = 0\}. \tag{9}$$

In other words, $W'_{i,j}$ represents a hyperplane with a normal vector $d_i - d_j$ and zero offset in the $\mathbb{R}^{q^2-2q+1}$. We define $\mathcal{A}' := \{W'_{i,j} \mid 1 \le i < j \le q^2 - q\}$. Then, $\mathcal{A}'$ forms a central hyperplane arrangement in $\mathbb{R}^{q^2-2q+1}$, dimension $q^2 - 2q + 1$. It can be shown that the number of regions in $\mathcal{A}$ and $\mathcal{A}'$ are equal.

Lemma III.16. *Let $\mathcal{A}$ and $\mathcal{A}'$ be hyperplane arrangements consisting of hyperplanes $W_{i,j}$ (as defined in Equation 5) and $W'_{i,j}$ (as defined in Equation (9)) in the vector spaces $\ker(\tilde{G}_{q,1})$ and $\mathbb{R}^{q^2-2q+1}$, respectively. Then we have $r(\mathcal{A}) = r(\mathcal{A}')$.*

Proof. Consider the map $T : \ker(\tilde{G}_{q,1}) \longrightarrow \mathbb{R}^{q^2-2q+1}$ defined as follows:

$$T(x) = (\alpha_1, \dots, \alpha_{q^2-2q+1}) \in \mathbb{R}^{q^2-2q+1}, \ \forall x \in \ker(\tilde{G}_{q,1}),$$

where $x = (\alpha_1, \dots, \alpha_{q^2-2q+1}, f_1^q, f_2^q, \dots, f_{q-1}^q)$, as defined in Equation (4). Since

$$x \in \ker(\tilde{G}_{q,1}) = \{B_{q,1}\alpha^T \mid \alpha \in \mathbb{R}^{q^2-2q+1}\},$$

each vector $\alpha = (\alpha_1, \dots, \alpha_{q^2-2q+1})$ corresponds uniquely to x. Hence T is bijective. Additionally, $\ker(\tilde{G}_{q,1})$ is a linear subspace of $\mathbb{R}^{q^2-q}$ and we have $T(cx + y) = cT(x) + T(y)$, which implies that T is a bijective linear transformation between two finite-dimensional topological vector spaces $\ker(\tilde{G}_{q,1})$ and $\mathbb{R}^{q^2-2q+1}$. Now, consider the hyperplane $W_{i,j}$ in $\ker(\tilde{G}_{q,1})$. We have:

$$\begin{aligned} T(W_{i,j}) &= \{T(x) \mid x \in W_{i,j}\} \\ &= \{T(x) \mid x = (\alpha_1, \dots, \alpha_{q^2-2q+1}, f_1^q, \dots, f_{q-1}^q), \\ &\qquad\qquad x_i - x_j = 0\} \\ &= \{\alpha = (\alpha_1, \dots, \alpha_{q^2-2q+1}) \mid d_i\alpha^T - d_j\alpha^T = 0\}, \end{aligned}$$

where d_i is the i-th row of the base matrix $B_{q,1}$, and $x_i = d_i\alpha^T$. Thus,

$$T(W_{i,j}) = \{\alpha \in \mathbb{R}^{q^2-2q+1} \mid (d_i - d_j)\alpha^T = 0\} = W'_{i,j}.$$

Since $\mathcal{A} = \{W_{i,j} \mid 1 \le i < j \le q^2 - q\}$, it follows that $T(\mathcal{A}) = \mathcal{A}' = \{W'_{i,j} \mid 1 \le i < j \le q^2 - q\}$. Finally, applying Lemma II.12 yields $r(\mathcal{A}) = r(\mathcal{A}')$. □

C. Numerical Results

In this section, based on the stated theorems, we outline the computational steps and present an algorithm for enumerating the number of feasible permutations for an alphabet of size q and $\ell = 2$. We begin by constructing the base matrix $B_{q,1}$ for the null space $\ker(\tilde{G}_{q,1})$, corresponding to the De Bruijn graph $\tilde{G}_{q,1}$, following the approach described in Section III-A. Then we consider arrangement $\mathcal{A}' = \{W'_{i,j}, \ 1 \le i < j \le q^2 - q\}$, where

$$W'_{i,j} = \{\alpha \in \mathbb{R}^{q^2-2q+1} \mid (d_i - d_j)\alpha^T = 0\},$$

and use Zaslavsky's formula, as stated in Theorem II.11, to compute $r(\mathcal{A}')$. By Lemma III.16 we have $r(\mathcal{A}') = r(\mathcal{A})$ and by Theorem III.14 we have $r(\mathcal{A}) = |\tilde{\Phi}_{q,2}|$. Finally, to compute the total number of feasible permutations $F_{q,2}$, we use Lemma III.6 for the case $\ell = 2$ which implies that $F_{q,2} = |\Phi_{q,2}| = |\tilde{\Phi}_{q,2}|\frac{q^2!}{(q^2-q)!}$. Based on the explanations above, we outline the computational steps in the Algorithm 1. By applying Algorithm 1 on a system with a HPZ840 processor and 128 GB of RAM, we determine the number of feasible permutations, $F_{q,2}$, for $q = 3$ and $q = 4$, as presented in Table I. Since the computational complexity for $q > 4$ is high, we present a lower bound for $F_{q,2}$, which is discussed in the next section.

Algorithm 1 Enumeration of the Number of Feasible Permutations

Input: q, the base matrix $B_{q,1}$ (section III-A)
Output: the number of feasible permutations $F_{q,2}$

for $1 \le i < j \le q^2 - q$ **do**
  Calculate the difference between all rows d_i and d_j of $B_{q,1}$;
  Form arrangement $\mathcal{A}'$, which includes at most $\binom{q^2-q}{2}$ hyperplanes $W'_{i,j}$ with normal vector $d_i - d_j$ and zero offset;
end for
Using Zaslavsky's formula (Theorem II.11), enumerate the number of regions of $\mathcal{A}'$ as $r(\mathcal{A}') = \sum_{\substack{B \subseteq \mathcal{A}' \\ B \text{ central}}} (-1)^{|B| - rank(B)}$;
return $F_{q,2} = r(\mathcal{A}').\dfrac{q^2!}{(q^2-q)!}$

IV. A Lower Bound on $F_{q,2}$

To establish a lower bound for $F_{q,2}$, we analyze a sub-arrangement $\tilde{\mathcal{B}}$ of the arrangement $\mathcal{A}$ By computing the number of regions associated with $\tilde{\mathcal{B}}$, we obtain a lower bound for $F_{q,2}$.

Notation: For simplicity, in the remainder of this paper, we set $m := q^2 - 2q + 1$.

Before presenting the main content, we introduce some symbols and notations. According to Equation 4, any balancing vector x is of the form:

$$x = (\alpha_1, \dots, \alpha_m, f_1^q, f_2^q, \dots, f_{q-1}^q).$$

TABLE I
THE NUMBER OF FEASIBLE PERMUTATIONS FOR $q = 3, 4$ AND $\ell = 2$

q	3	4
$r(\mathcal{A})$	60	34919088
$F_{q,2}$	30240	1525265763840

where f_t^q, $t \in [q-1]$, is a degree one multivariable polynomial in the variables $x_1, \ldots, x_m$ (denoted by $\langle x_1, \ldots, x_m \rangle$).

Definition IV.1. *Let $f \in \langle x_1, \ldots, x_m \rangle$. The weight of f is defined as the number of non-zero terms in f, and is denoted by $n(f)$. Furthermore, we denote the set of indices of the terms in f with negative coefficients by C_f^-, and the set of indices with positive coefficients by C_f^+.*

For example, for $f_4^5 = x_4 - x_9 + x_{10} - x_{13} + x_{14} - x_{15} + x_{16}$, we have $n(f) = 7$. Additionally, $C_f^+ = \{4, 10, 14, 16\}$ and $C_f^- = \{9, 13, 15\}$.

Proposition IV.2. *The weight of the polynomial f_t^q is given by:*

$$\begin{aligned} n(f_t^q) &= 2((t-1)(q-t) + q - (t+1)) + 1 \\ &= 2(t(q-t) - 1) + 1, \qquad \forall\, t \in [q-1]. \end{aligned}$$

Moreover, we have

$$|C_{f_t^q}^+| = \frac{n(f_t^q) - 1}{2} + 1,\ |C_{f_t^q}^-| = \frac{n(f_t^q) - 1}{2}, \quad \forall\, t \in [q-1].$$

Proof. Based on the construction method presented in Section III-A, f_t^q corresponds to the directed edge $e_{t,t+1}$ in the spanning tree T. Assume that the nodes of T are labeled as v_i, for $i \in [q]$. The number of nonzero terms in f_t^q is corresponding to the number of the fundumental cycles containing $e_{t,t+1} = (v_t, v_{t+1})$. The number of directed paths in T that start at nodes preceding v_t containing $e_{t,t+1}$, along with their reverses is $2(t-1)(q-t)$, and the number of paths starting at v_t and ending at the nodes following v_{t+1}, and their reverses, is $2(q-(t+1))$. Furthermore, there is one fundamental cycle associated with the directed path $e_{t+1,t}$. Therefore, the number of all fundamental cycles containing $e_{t,t+1}$ is $2(t-1)(q-t) + 2(q-(t+1)) + 1$. Moreover, since

$$B_{q,1} = \left[\begin{array}{c|c} \multicolumn{2}{c}{I_m} \\ \hline I_{q-1} & D \end{array}\right]$$

and according to Remark III.9, we can conclude that:

$$|C_{f_t^q}^+| = \frac{n(f_t^q) - 1}{2} + 1, \qquad |C_{f_t^q}^-| = \frac{n(f_t^q) - 1}{2}.$$

□

Recall that the arrangement $\mathcal{A}$ consists of the following hyperplanes:

$$\begin{cases} x_i - x_j = 0 & 1 \le i < j \le m, & \text{(A)} \\ x_i - f_j^q = 0 & 1 \le i \le m, \quad 1 \le j \le q-1 & \text{(B)} \\ f_i^q - f_j^q = 0 & 1 \le i, j \le q-1 & \text{(C)}. \end{cases} \tag{10}$$

Since determining the total number of regions of $\mathcal{A}$ is challenging, we give a lower bound on the number of regions of the sub-arrangement $\tilde{\mathcal{B}}$, where $\tilde{\mathcal{B}}$ is

$$\begin{cases} x_i - x_j = 0 & 1 \le i < j \le m, \\ x_i - f_j^q = 0 & 1 \le i \le m,\ 1 \le j \le q-1. \end{cases} \tag{11}$$

For $1 \le k \le m$ and $1 \le t \le q-1$, we set $f_{k,t} := x_k - f_t^q$. Clearly

$$\tilde{\mathcal{B}} = \mathcal{B}_m \cup \{f_{k,t} = 0 \mid 1 \le k \le m,\ 1 \le t \le q-1\}, \tag{12}$$

where $\mathcal{B}_m$ is the braid arrangement

$$\mathcal{B}_m := \{x_i - x_j = 0 \mid 1 \le i < j \le m\}.$$

Remark IV.3. *The weight of $f_{k,t}$ is given as follows:*

$$n(f_{k,t}) = \begin{cases} n(f_t^q) - 1, & k \in C_{f_t^q}^+ \\ n(f_t^q), & k \in C_{f_t^q}^- \\ n(f_t^q) + 1, & k \in [m] \setminus (C_{f_t^q}^+ \cup C_{f_t^q}^-). \end{cases}$$

More precisely, according to Proposition IV.2, $f_{k,t}$ has the following forms:

$$\begin{cases} f_{k,t} = \sum_{i \in C_{f_{k,t}}^+} x_i - \sum_{j \in C_{f_{k,t}}^-} x_j, \\ \qquad \text{if } |C_{f_{k,t}}^+| = |C_{f_{k,t}}^-| = \frac{n(f_t^q) - 1}{2}, & \text{(I)} \\ f_{k,t} = \sum_{i \in C_{f_{k,t}}^+ \setminus \{k\}} x_i + 2x_k - \sum_{j \in C_{f_{k,t}}^-} x_j, \\ \qquad \text{if } |C_{f_{k,t}}^+| = \frac{n(f_t^q) - 1}{2}, |C_{f_{k,t}}^-| = \frac{n(f_t^q) - 1}{2} + 1, & \text{(II)} \\ f_{k,t} = \sum_{i \in C_{f_{k,t}}^+} x_i - \sum_{j \in C_{f_{k,t}}^-} x_j, \\ \qquad \text{if } |C_{f_{k,t}}^+| = |C_{f_{k,t}}^-| = \frac{n(f_t^q) - 1}{2} + 1. & \text{(III)} \end{cases} \tag{13}$$

In what follows, we will give a lower bound on the number of regions of the sub-arrangement $\tilde{\mathcal{B}}$. To do this, let $\pi \in S_m$ correspond to a region of the braid arrangement $\mathcal{B}_m$ (Corollary II.10). We compute the regions that $f_{k,t} = x_k - f_t^q = 0$ passes through them. In this regard, first we count the number of regions it does not intersect. We restrict ourselves to the cases (I) and (III) in (13), where $n(f_{k,t})$ is even. For the case (II) in (13) we give a conjecture and present some numerical results to confirm our conjecture. We denote by $\alpha_{k,t}$, the number of regions of the braid arrangement $\mathcal{B}_m$ that is intersected by $f_{k,t} = 0$ and those which is not intersected by $f_{k,t} = 0$ with $\overline{\alpha}_{k,t}$.

Proposition IV.4. *If the polynomial $f_{k,t}$ has $n(f_{k,t}) = 2\ell_{k,t}$ (cases (I) and (III) in (13)), then it does not pass* *through the region $\mathfrak{R} \in R(\mathcal{B}_m)$ associated with the permutation π, if and only if one of the following conditions hold:*

$$1)\ \exists\, \mathfrak{f}:\ C^-_{f_{k,t}} \xrightarrow{1-1} C^+_{f_{k,t}}:\ \pi^{-1}(a) < \pi^{-1}(\mathfrak{f}(a)),\ \forall a \in C^-_{f_{k,t}}, \tag{14}$$

$$2)\ \exists\, \mathfrak{f}:\ C^-_{f_{k,t}} \xrightarrow{1-1} C^+_{f_{k,t}}:\ \pi^{-1}(a) > \pi^{-1}(\mathfrak{f}(a)),\ \forall a \in C^-_{f_{k,t}}. \tag{15}$$

Proof. Based on Corollary II.10, each region $\mathfrak{R} \in R(\mathcal{B}_m)$ is associated with a unique permutation $\pi \in S_m$, that preserves the relative order of the components of the points $x = (x_1, x_2, \ldots, x_m)$ within $\mathfrak{R}$. Assume that the relative order is given as follows:

$$x_{\pi(1)} < x_{\pi(2)} < \ldots < x_{\pi(m)}. \tag{16}$$

Clearly, $x_i < x_j$ if and only if $\pi^{-1}(i) < \pi^{-1}(j)$. The hyperplane $f_{k,t} = 0$ does not intersect with $\mathfrak{R}$ if and only if there exists no point with order given in (16) that satisfies the equation $f_{k,t} = \sum_{i\in C^+_{f_{k,t}}} x_i - \sum_{j\in C^-_{f_{k,t}}} x_j = 0$, where $|C^+_{f_{k,t}}| = |C^-_{f_{k,t}}| = \ell_{k,t}$. Consequently, we have:

$$f_{k,t} \neq 0 \iff \begin{cases} \sum_{i\in C^-_{f_{k,t}}} x_i < \sum_{j\in C^+_{f_{k,t}}} x_j, \\ \text{or} \\ \sum_{i\in C^-_{f_{k,t}}} x_i > \sum_{j\in C^+_{f_{k,t}}} x_j \end{cases}.$$

It is clear that these inequalities hold if and only if one of the following conditions hold:

$$1)\ \exists\, \mathfrak{f}:\ C^-_{f_{k,t}} \xrightarrow{1-1} C^+_{f_{k,t}}:\ \pi^{-1}(a) < \pi^{-1}(\mathfrak{f}(a)),\ \forall a \in C^-_{f_{k,t}} \tag{17}$$

$$2)\ \exists\, \mathfrak{f}:\ C^-_{f_{k,t}} \xrightarrow{1-1} C^+_{f_{k,t}}:\ \pi^{-1}(a) > \pi^{-1}(\mathfrak{f}(a)),\ \forall a \in C^-_{f_{k,t}} \tag{18}$$

as desired. □

To continue, we need to recall Catalan numbers. The n-th Catalan number C_n, represents the number of binary sequences of length $2n$ that contain exactly n zeros, such that, at every position i (where $1 \le i \le 2n$), the number of zeros is greater than or equal to the number of ones. The n-th Catalan number is given by:

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Moreover, we denote the set of Catalan sequences of length $2n$, by $DP(n)$. The following lemma is the key. We denote by $n_{k,t}$ the integer $n(f_{k,t})$.

Lemma IV.5. *Let X be a set of positive integers of size $2n$. Then the number of partitions of X into two subsets A and B of size n such that there exist a one to one function $f: A \longrightarrow B$ for which $a < f(a)$, for all $a \in A$, is equal to C_n.*

Proof. Let $\mathcal{P}$ be the set of partitions $P = \{A_P, B_P\}$ of X with $|A_P| = |B_P| = n$ and there exist a one to one function $f_P: A_P \longrightarrow B_P$ for which $a < f(a)$, for all $a \in A_P$. Define the function

$$\Gamma: \mathcal{P} \longrightarrow DP(n)$$

such that $\Gamma(P) = a^{(P)}$ where $a^{(P)}_i = 0$ for $i \in A_P$ and $a^{(P)}_i = 1$ for $i \in B_P$. We show that Γ is a bijection. We only prove that $a^{(P)} \in DP(n)$. The remaining parts of the proof are easy and we omit them. To show that $a^{(P)} \in DP(n)$, we need to show that for an $1 \le i \le 2n$, the number of zeros of $a^{(P)}$ till the i-th position, is greater than or equal to the number of ones till i-th position. To contrary, assume that there exists $1 \le r \le 2n$ such that the number of zeros of $a^{(P)}$ till the r-th position, is less than the number of ones till r-th position. This means that the number of zeros of $a^{(P)}$ from the $(r+1)$-th position till the last position, is greater than the number of ones from the $(r+1)$-th position till the last position. Since f_P is a bijection, we can conclude that there exist $t \in A_P$ where $t \ge r+1$ and $f_P(t) \le r$. This contradicts the condition $t < f_P(t)$. □

Theorem IV.6. *Let $n_{k,t} = 2\ell_{k,t}$ (cases (I) and (III) in (13)). Then we have*

$$\alpha_{k,t} = \left(\frac{\ell_{k,t}-1}{\ell_{k,t}+1}\right)m!.$$

Proof. To calculate $\alpha_{k,t}$, we compute $\overline{\alpha}_{k,t}$, that is; the number of regions of the braid arrangement that $f_{k,t} = 0$ does not passes through them. Let π denotes the permutation corresponding to a region of the braid arrangement. According to Proposition IV.4, π must satisfy one of the conditions (17) or (18). Clearly, the number of permutations which satisfy (17), is equal to the number of permutations which satisfy (18). Hence we calculate the number of permutations satisfying (17). Since $n_{k,t} = 2\ell_{k,t}$, we have $|C^+_{f_{k,t}}| = |C^-_{f_{k,t}}| = \ell_{k,t}$. Hence, we first need to choose $n_{k,t}$ values of π on $C^+_{f_{k,t}} \cup C^-_{f_{k,t}}$ and then partition this set into two subsets A and B which satisfy Lemma IV.5. Also we need to have $\pi(A) = A$ and $\pi(B) = B$ and the other values of π permutes freely. Hence the number of permutations which satisfy (17) is equal to

$$\binom{m}{n_{k,t}}(\ell_{k,t}!)^2 C_{\ell_{k,t}}(m - n_{k,t})! = \frac{m!}{\ell_{k,t}+1}.$$

Consequently (considering (18)) we have

$$\overline{\alpha}_{k,t} = \frac{2m!}{\ell_{k,t}+1}.$$

Therefor,

$$\alpha_{k,t} = m! - \frac{2m!}{\ell_{k,t}+1} = \left(\frac{\ell_{k,t}-1}{\ell_{k,t}+1}\right)m!,$$

as desired. □

Now, we consider the case (II) in (13) and give a conjecture for $\alpha_{k,t}$ in this case. We start with the following definition.

Definition IV.7. *For each $i \in \{1, \ldots, 2n-1\}$, define*

$$\mathrm{Tr}(n,i) = |\{s \in DP(n) \mid s_i = s_{i+1} = 0\}|,$$

TABLE II
$\mathrm{Tr}(n)$ FOR $2 \leq n \leq 13$

n	2	3	4	5	6	7	8	9	10
$\mathrm{Tr}(n)$	1	5	21	84	330	1287	5005	19448	75582
n	11	12	13						
$\mathrm{Tr}(n)$	293930	1144066	4457400						

as the number of Catalan sequences in which both positions i and $i+1$ are zeros. Moreover, define $\mathrm{Tr}(n)$ as

$$\mathrm{Tr}(n) := \sum_{i=1}^{2n-1} \mathrm{Tr}(n,i).$$

Conjecture 1. *We have $\mathrm{Tr}(n) = \dfrac{(n-1)C_n}{2}$. Moreover, if $n_{k,t} = 2\ell_{k,t}+1$ then we have*

$$\alpha_{k,t} = m!\left(\frac{\ell_{k,t}^2 - 2\ell_{k,t} - 2}{\ell_{k,t}^2 - 2\ell_{k,t}}\right).$$

Applying a computer search, Table II presents the computed values of $\mathrm{Tr}(n)$, for several values of n.

The following proposition now provides an important role to establish a lower bound on $r(\widetilde{\mathcal{B}})$.

Proposition IV.8. *We have*

$$r(\widetilde{\mathcal{B}}) \geq m! + \sum_{t=1}^{q-1}\sum_{k=1}^{m} \alpha_{k,t}. \tag{19}$$

Proof. The proof follows from the fact that when an $f_{k,t}$ passes through a braid region, the number of regions is added at least by one. $\square$

We have determined the value $\alpha_{k,t}$, for all k except those in $C^-_{f_t^q}$ which we made a conjecture on those values. We can now summarize all the results in the following theorem.

Theorem IV.9. *We have the following two inequalities:*

(I) *For all values of q, we have*

$$r(\widetilde{\mathcal{B}}) \geq m!\left(1 + \sum_{t=1}^{q-1}\left(\sum_{k\in C^+_{f_t^q}}\left(\frac{t(q-t)-2}{t(q-t)}\right) + \sum_{k\in[m]\setminus(C^+_{f_t^q}\cup C^-_{f_t^q})}\left(\frac{t(q-t)-1}{t(q-t)+1}\right)\right)\right).$$

(II) *For $q \geq 5$ we have*

$$r(\widetilde{\mathcal{B}}) \geq \left(\frac{m!}{12}\right)\left(5q^3 - 18q^2 + 25q - 11\right).$$

Proof. The proof of the first part follows from Proposition IV.8 and that when $k \in C^+_{f_t^q}$ we have $\ell_{k,t} = t(q-t)-1$ and when $k \in [m]\setminus(C^+_{f_t^q}\cup C^-_{f_t^q})$ we have $\ell_{k,t} = t(q-t)$. The proof of the second part follows from Proposition IV.8 and the fact that when $q \geq 5$ both $\frac{t(q-t)-2}{t(q-t)}$ and $\frac{t(q-t)-1}{t(q-t)+1}$ are greater than or equal to $1/2$. $\square$

Remark IV.10. *The following lower bound on $F_{q,2}$ is given in [1, Corollary 3].*

$$F_{q,2} \geq (m+1)!\left(\frac{q-1}{q+1}\right)\left(\frac{q^2!}{(q^2-q)!}\right). \tag{20}$$

Setting $J(q) := (q^2-2q+1)!(q^2)!/(q^2-q)!$, we can rewrite (20) as

$$F_{q,2} \geq J(q)\left(\frac{(q^2-2q+2)(q-1)}{q+1}\right). \tag{21}$$

Consequently, this lower bound is $\Omega(q^2J(q))$. On the other hand, using $F_{q,2} = r(\mathcal{A}).\dfrac{q^2!}{(q^2-q)!}$ and the fact that $r(\mathcal{A}) \geq r(\widetilde{\mathcal{B}})$ we conclude that

$$F_{q,2} \geq r(\widetilde{\mathcal{B}})\frac{q^2!}{(q^2-q)!}. \tag{22}$$

Now, Part (II) of Theorem IV.9 and (22), gives us the following new lower bound on $F_{q,2}$.

$$F_{q,2} \geq J(q)\left(\frac{5q^3-18q^2+25q-11}{12}\right). \tag{23}$$

Clearly, our lower bound is $\Omega(q^3J(q))$ and this shows that our bound is an improvement over 21.

Table III, presents a comparison between the lower bound on $F_{q,2}$ obtained from Part (I) of Theorem IV.9 and (21), for some values of q and $\ell = 2$.

V. Conclusion

In this paper, we investigated feasible permutations (rank modulation codes) in the context of DNA data storage using the shotgun sequencing method. Specifically, we considered these codes for arbitrary alphabet size q ($q > 2$) and sliding window length $\ell = 2$. By employing hyperplane arrangements as a geometric framework, we analyzed the number of regions formed by these arrangements in a special form. We then established a correspondence between the regions of the hyperplane arrangement and the codewords of the Rank Modulation code, and based on this equivalence, we proposed an algorithm to enumerate the size of the code, denoted $F_{q,2}$. Additionally, we derived a lower bound on $F_{q,2}$ and show that our bound is an improvement over the existing bounds in the literature. An important direction for future research is to extend the proposed methods and analysis to larger window sizes $\ell \geq 3$.

TABLE III
COMPARISON BETWEEN THE LOWER BOUND L_q ON $F_{q,2}$ PRESENTED IN THEOREM IV.9 AND THE LOWER BOUND L'_q GIVEN IN [1] FOR $\ell = 2$

q	L_q	L'_q
4	$3{,}701{,}376 \times 43{,}680 = 161{,}676{,}103{,}680$	$2{,}177{,}280 \times 43{,}680 = 95{,}103{,}590{,}400$
5	$4.12763292469592064 \times 10^{21}$	$1.5118138443792384 \times 10^{21}$
6	$1.507246729222722780827970699264 \times 10^{36}$	$4.039857677060811998462214144000000000 \times 10^{35}$

REFERENCES

[1] N. Beeri and M. Schwartz, "Improved rank modulation codes for DNA storage with shotgun sequencing," *IEEE Trans. Inf. Theory*, vol. 68, no. 6, pp. 3719–3730, 2022.

[2] J. A. Bondy and U. S. R. Murty, "Graph Theory," New York: Springer, 2008 (Graduate Texts in Mathematics, vol. 244).

[3] J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, G. Seelig, and K. Strauss, "A DNA-based archival storage system," *ACM SIGOPS Oper. Syst. Rev.*, vol. 50, no. 2, pp. 637–649, 2016.

[4] C. Godsil , G. Royle, "Algebraic graph theory," New York: Springer-Verlag; 2001. (Graduate texts in mathematics; vol. 207)

[5] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, and E. Birney, "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," *Nature* , vol. 494, no. 7435, pp. 77–80, 2013.

[6] H. M. Kiah, G. J. Puleo, and O. Milenkovic, "Codes for DNA sequence profiles," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3125–3146, 2016.

[7] G. H. Moore, "The evolution of the concept of homeomorphism," *Historia Mathematica*, vol. 34, no. 3, pp. 333–343, 2007.

[8] S. Motahari, G. Bresler, and D. Tse, "Information theory for DNA sequencing: Part 1: A basic model," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, USA, pp. 2741–2745, 2012.

[9] N. Raviv, M. Schwartz, and E. Yaakobi, "Rank modulation codes for DNA storage with shotgun sequencing," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 50–64, 2018.

[10] J. Sima, N. Raviv, M. Schwartz, and J. Bruck, "Error correction for DNA storage," *IEEE BITS Inf. Theory Mag.* , vol. 3, no. 3, pp. 78–94, 2023.

[11] R. P. Stanley, "An introduction to hyperplane arrangements," in *Geometric Combinatorics*, vol. 13, no. 24, pp. 389–496, 2004.

[12] S. H. Yazdi, H. M. Kiah, E. Garcia-Ruiz, J. Ma, H. Zhao, and O. Milenkovic, "DNA-based storage: Trends and methods," *IEEE Trans. Mol. Biol. Multi-Scale Commun*, vol. 1, no. 3, pp. 230–248, 2015.

Reza Sobhani received the B.Sc. from Isfahan University of Technology, Iran in 2003, the M.Sc. from Sharif University of Technology, Iran in 2005 and the Ph.D. in Applied Mathematics from the Isfahan University of Technology, Iran in 2010. Since 2010, he has been with Department of Mathematics in University of Isfahan as a faculty member and from 2009 he has been an associated professor of Mathematics in University of Isfahan. He is the author of more than 30 articles. His research interests include Algebraic Coding Theory, Combinatorics and Graph Theory with applications in Communication Theory.

Farzad Parvaresh received the BS degree in electrical engineering from the Sharif University of Technology, Tehran, Iran in 2001 and the MS and PhD degrees in electrical and computer engineering from the University of California, San Diego, in 2003 and 2007, respectively. He is currently an associate professor with the Department of Electrical Engineering, University of Isfahan, Isfahan, Iran. He was a postdoctoral scholar with the Center for Mathematics of Information (CMI), California Institute of Technology, Pasadena, California during 2007-2008 and a visiting researcher with the Information Theory Research Group, Hewlett-Packard Laboratories, Palo Alto, California during 2010-2012. His current research interests include areas related to coding theory, information theory and wireless communication networks. He received the best paper award from the 46th Annual IEEE Symposium on Foundations Of Computer Science (FOCS) in 2005, and was awarded the Silver medal in the 28th International Physics Olympiad in 1997.

Alireza Abdollahi received the B.Sc. and M.Sc. degrees in Mathematics from the University of Isfahan, Iran in 1996 and 1997 respectively. He has also obtained the Ph.D. degree in Mathematics from University of Isfahan in 2000. He also received a second Mathematics Ph.D. degree from Université de Provence, CMI France in 2001. Since 2001, he has been with Department of Mathematics in University of Isfahan as a faculty member and from 2009 he has been a professor of Mathematics in University of Isfahan. He is the author of more than 120 articles. His research interests include Group Theory and Combinatorics. He is the Editor-in-Chief of Journal of the Iranian Mathematical Society, the founder and Editor-in-Chief of International Journal of Group Theory and Transactions on Combinatorics. Dr. Abdollahi was a recipient of the Alkharazmi Young prize in 1999 and Sheikh Bahaei prize in 1999. He also received "Institute for Studies in Theoretical Physics and Mathematics (IPM)" Young Mathematician Prize of 2005 and he was the head of distinguished Center of Excellence in Basic Sciences selected by the Ministry of Science and Education of Iran in 2010.

Farzaneh Abedi received her B.Sc. and M.Sc. degrees in Applied Mathematics from the University of Isfahan, Isfahan, Iran, in 2012 and 2014, respectively. She received her Ph.D. in Mathematics (Coding Theory) from Shahrekord University, Shahrekord, Iran. She is currently a postdoctoral researcher at the University of Isfahan. Her research interests include algebraic coding theory, LDPC codes, DNA data storage, and DNA coding.

Javad Bagherian received the B.S. degree in Applied Mathematics from University of Kharazmi, Thehran, Iran, in 2002 and M.S degree in Applied Mathematics from the Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, Iran, in 2004 and Ph.D. degree in Pure Mathematics from the Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, Iran, in 2008. From 2008 to 2016, he was Assistant Professor with the Mathematics and Computer Science Department , University of Isfahan , Iran. His research interest includes scheme theory, graph theory, permutation codes. He published more than 30 papers. Since 2016, he has been an Associate Professor with the Mathematics and Computer Science Department, University of Isfahan, Iran.

Maryam Khatami received her Ph.D. in Pure Mathematics from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2011. She joined the Department of Mathematics and Computer Science at the University of Isfahan, Iran, as an Assistant Professor in 2012. Her research interests include Group Theory and Combinatorics. She has published over 30 research papers in peer-reviewed journals.