Error-Correction in Flash Memories via Codes in the Ulam Metric

Farzad Farnoud (Hassanzadeh), *Student Member, IEEE*, Vitaly Skachek, and Olgica Milenkovic, *Senior Member, IEEE*

***Abstract*—We consider rank modulation codes for flash memories that allow for handling arbitrary charge-drop errors. Unlike classical rank modulation codes used for correcting errors that manifest themselves as swaps of two adjacently ranked elements, the proposed *translocation rank codes* account for more general forms of errors that arise in storage systems. Translocations represent a natural extension of the notion of adjacent transpositions and as such may be analyzed using related concepts in combinatorics and rank modulation coding. Our results include derivation of the asymptotic capacity of translocation rank codes, construction techniques for asymptotically good codes, as well as simple decoding methods for one class of constructed codes. As part of our exposition, we also highlight the close connections between the new code family and permutations with short common subsequences, deletion and insertion error-correcting codes for permutations, and permutation codes in the Hamming distance.**

***Index Terms*—Data storage systems, error-correction codes, flash memory, Hamming distance, translocation errors, Ulam distance.**

I. Introduction

PERMUTATION codes and permutation arrays are collections of suitably chosen codewords from the symmetric group, used in applications as varied as single user communication over Gaussian channels [1], [2], reduction of impulsive noise over power lines [3], [4], and coding for storage [5]. Many instances of permutation-based codes were studied in the coding theory literature, with special emphasis on permutation arrays under the Hamming distance and rank modulation codes under the Kendall τ distance [6]–[9, Ch. 6B]. The distances used for code construction in storage devices have mostly focused around two types of combinatorial measures, counting functions of adjacent transpositions and measures obtained via embeddings into the Hamming space [3], [5]. This is due to the fact that such distance measures capture the displacement of symbols in retrieved messages that arise in modern nonvolatile storage systems.

One of the most prominent emerging applications of permutation codes in storage is *rank modulation*. Rank modulation is an encoding scheme for flash memories that may improve the lifespan, storage efficiency, and reliability of future generations of these storage devices [10]–[12]. The idea behind the modulation scheme is that information should be stored in the form of rankings of the cells' charges, rather than in terms of the absolute values of the charges. This simple conceptual coding framework may eliminate the problem of cell block erasures as well as potential cell overinjection issues [10], [13]. In their original formulation, rank-modulation codes represent a family of codes capable of handling errors of the form of *adjacent transpositions*. Such transposition errors represent the most likely errors in a system where the cells are expected to have nearly uniform leakage rates. But leakage rates depend on the charge of the cells, the position of the cells, and on a number of external factors, the influence of which may not be adequately captured by adjacent transposition errors. For example, if a cell for a variety of reasons has a higher leakage rate than other cells, given sufficient time, the charge of this cell may drop below the charge of a large number of other cells. Furthermore, if the number of possible charge levels is large,[1] and thus the difference between charge levels is small, a moderate charge drop may result in a significant drop in the cell's rank. One may argue that these processes may be modeled as a sequence of adjacent transposition errors. However, as this type of error is the result of a single-error event, for the purpose of error correction, it should be modeled as a *single error*. This is reminiscent of the scenario where one models a sequence of individual symbol errors as a single burst error [15].

In what follows, we present a novel approach to rank modulation coding which allows for correcting a more varied class of errors when compared to classical schemes. The focal point of the study is the notion of a translocation, a concept that generalizes an adjacent transposition in a permutation. Roughly speaking, a translocation[2] moves the ranking of one particular element in the permutation below the rankings of a certain number of closest-ranked elements. As such, translocations are

Manuscript received February 11, 2012; revised October 18, 2012; accepted December 18, 2012. Date of publication January 14, 2013; date of current version April 17, 2013. This work was supported in part by the NSF STC-CSoI 2011 CCF 0939370, in part by the NSF CCF 0809895, and in part by the AFRLDL-EBS AFOSR Complex Networks grants. This paper was presented in part at the 2012 IEEE Information Theory and Applications Workshop and in part at the 2012 IEEE International Symposium on Information Theory.

F. Farnoud (Hassanzadeh) and O. Milenkovic are with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: hassanz1@illinois.edu; milenkov@uiuc.edu).

V. Skachek was with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. He is now with the Institute of Computer Science, Faculty of Mathematics and Computer Science, University of Tartu, Tartu 50409, Estonia (e-mail: vitaly.skachek@gmail.com).

Communicated by N. Kashyap, Associate Editor for Coding Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2013.2239700

[1]There are two important motivations for increasing the number of charge levels. First, larger number of charge levels may enable storing more data, and second, when there are a large number of charge levels available, encoding methods such as push-to-the-top [14] can be used to decrease the number of times that the memory needs to be erased.

[2]Note that our definition of the term translocation differs from the definition commonly used in biology. See, e.g., [16].

suitable for modeling errors that arise in flash memory systems, where high leakage levels for subsets of cells are expected or possible. Examples of such error events include errors due to radiation and breakdown of tunneling oxide, the latter being a prominent event in conventional poly-Si floating gate memories [17], [18].

A translocation may be viewed as an extension of an adjacent transposition. In addition, translocations correspond to pairs of deletions and insertions of elements in the permutation. As a consequence, the study of translocations is closely related to the longest common subsequence problem and permutation coding under the Levenshtein and Hamming metrics [19]–[21].

Rank modulation is by now well understood from the perspective of code construction. The capacity of rank modulation codes was derived in [5], [22], and [23], while some practical code constructions were proposed in [5], and [10], and further generalized in [14], [22], and [24]. Here, we complement the described work in terms of deriving upper and lower bounds on the capacity of translocation rank codes, and in terms of presenting constructive, asymptotically good coding schemes. Our constructions are based on a novel application of permutation interleaving and are of independent interest in combinatorics and algebra. For the use of specialized forms of permutation interleaving in other areas of coding theory, the interested reader is referred to [11] and [25]. Furthermore, we propose decoding algorithms for translocation codes based on decoders for codes in the Hamming metric [26], [27]. Finally, we also highlight the close relationships between permutation codes under a number of metrics.

This paper is organized as follows. In Section II, we provide the motivation for studying translocations as well as basic definitions used in our analysis. The properties of permutations under translocations are studied in the same section, while bounds on the size of the codes are presented in Section III. Code constructions are presented in Sections IV and V, while concluding remarks are given in Section VI.

II. Basic Definitions

Throughout this paper, we use the following notation and terminology. The symbol $[n]$ denotes the set $\{1, 2, \ldots, n\}$. A *permutation* denotes a bijection $\sigma : [n] \to [n]$, that is, for any distinct $i, j \in [n]$, we have $\sigma(i) \neq \sigma(j)$. We let $\mathbb{S}_n$ stand for the set of all permutations of $[n]$, i.e., the symmetric group of order $n!$. For any $\sigma \in \mathbb{S}_n$, we write $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$, where $\sigma(i)$ is the image of $i \in [n]$ under σ. The identity permutation $(1, 2, \ldots, n)$ is denoted by e, while σ^{-1} stands for the inverse of the permutation σ. The product $\sigma\pi$ of two permutations $\sigma, \pi \in \mathbb{S}_n$ is defined so that, for each $i \in [n]$, we have $(\sigma\pi)(i) = \sigma(\pi(i))$, i.e., permutations act on the left.

For some $\sigma \in \mathbb{S}_n$ and $P \subseteq [n]$, the *projection σ_P of σ onto* P is obtained from σ by only keeping elements of P and removing all other elements. For example, for $\sigma = (5, 4, 3, 2, 1)$ and $P = \{2, 4, 5\}$, we have $\sigma_P = (5, 4, 2)$. Note that σ_P has length $|P|$. Next, let $\mathbb{S}(P)$ stand for the set of all permutations of elements of P. The identity element of $\mathbb{S}(P)$ is e_P, obtained from $(1, 2, \ldots, n)$ by removing elements that are not in P.

Permutations are denoted by Greek lowercase letters, while integers and integer vectors are denoted by Latin lowercase symbols.

A *transposition* $\tau(i, j)$, for distinct $i, j \in [n]$, is a permutation obtained from the identity by swapping the positions of i and j. Namely

$$\tau(i, j) = (1, \ldots, i-1, j, i+1, \ldots, j-1, i, j+1, \ldots, n).$$

If $|i - j| = 1$, then $\tau(i, j)$ is called an *adjacent transposition*.

For distinct $i, j \in [n]$, a *translocation* $\phi(i, j)$ is a permutation obtained from the identity by moving i to the position of j and shifting elements between i and j, including j, by one. If $i < j$, we have

$$\phi(i, j) = (1, \ldots, i-1, i+1, i+2, \ldots, j, i, j+1, \ldots, n)$$

and if $i > j$, we have

$$\phi(i, j) = (1, \ldots, j-1, i, j, j+1, \ldots, i-1, i+1, \ldots, n)\ .$$

For $i < j$, the permutation $\phi(i, j)$ is called a *right-translocation* and the permutation $\phi(j, i)$ is called a *left-translocation*. The length of a translocation $\phi(i, j)$ equals $|j - i|$, that is, the number of elements between i and j, including j. Note that a translocation of length k can be modeled by k adjacent transpositions.

If the set of elements under consideration is a subset P of $[n]$, for distinct $i, j \in P$, a translocation $\phi(i, j)$ over P is obtained from e_P by moving i to the position of j, and shifting elements between i and j, including j, by one. Right- and left-translocations over P are defined similarly.

Example 1: Let $\sigma = (1, 3, 5, 7, 2, 4, 6, 8)$. We have

$$\begin{aligned}\sigma\phi(3, 6) &= (1, 3, 7, 2, 4, 5, 6, 8)\\ \sigma\phi(5, 2) &= (1, 2, 3, 5, 7, 4, 6, 8)\\ \sigma\tau(3, 6) &= (1, 3, 4, 7, 2, 5, 6, 8).\end{aligned}$$

Furthermore, let $P = \{2, 3, 5, 8\}$ and $\pi = (5, 8, 3, 2) \in \mathbb{S}(P)$. The translocation $\phi(8, 2)$ over P equals $(8, 2, 3, 5)$ and we have $\pi\phi(8, 2) = (2, 5, 8, 3)$. Notice that in this case, as for the case of standard permutations, the parameters in $\phi(\cdot, \cdot)$ refer to the elements in the corresponding identity permutation, rather than positions. □

Observe that the inverse of the left-translocation $\phi(i, j)$ is the right-translocation $\phi(j, i)$, and vice versa.

Our interest in translocations in permutations is motivated by rank modulation coding, as illustrated by the examples depicted in Figs. 1 and 2. In classical multilevel flash memories, each cell used for storing information is subjected to errors. As a result, classical error control schemes of nonzero rate cannot be efficiently used in such systems. One solution to the problem is to encode information in terms of rankings [6], rather than absolute values of the information sequences. Consequently, data are represented by permutations and errors manifest themselves via reordering of the ranked elements. The simplest model assumes that only adjacently ranked elements may be exchanged.

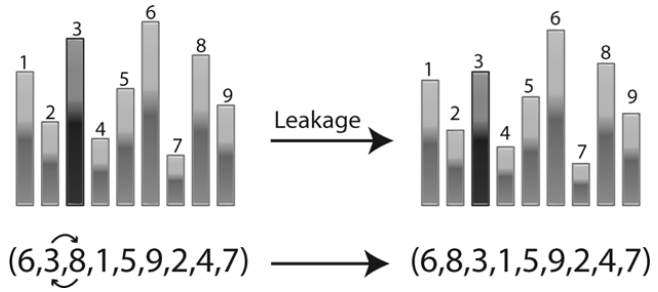

Fig. 1. Rank modulation codes and adjacent transposition errors.

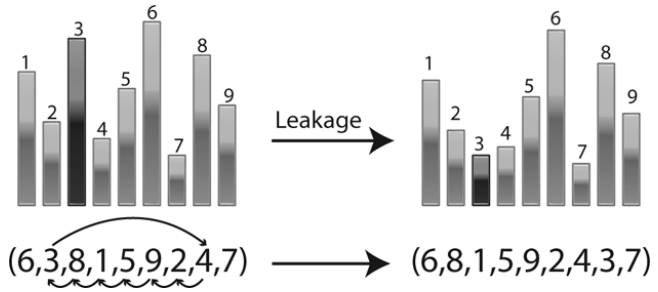

Fig. 2. Rank modulation codes and translocation errors caused by "large" drops of charge levels.

This model has the drawback that it does not account for more general changes in ranks. With respect to this observation, consider the charge-drop model in Fig. 2. Here, cell number 3, ranked second, experienced a leakage rate sufficiently high to move the cell's ranking to the eighth position. This error is represented by the translocation $\phi(2,8)$. The translocation $\phi(2,8)$ corresponds to six adjacent transpositions. Nevertheless, as already argued, a translocation should be counted as a single error, and not a sequence of adjacent transposition errors.

The translocation error model may appear to be too broad to describe the phenomena arising in flash memories, as errors corresponding to translocations of small length arise more frequently than errors corresponding to translocations of long length. The idea of bounded length translocations (bounded burst errors) will be addressed in a companion paper.[3] We also remark that translocation errors of arbitrary length accurately model *any* error that affects a single cell, and are hence suitable for modeling arbitrary charge drops of cells independently of drops of other cells, as well as read disturb and write disturb errors [28]. This makes them a good candidate for studying new error-control schemes in flash memories.

Next, we formalize the notion of a distance capturing translocation errors.

Definition 2: Let $\pi, \sigma \in \mathbb{S}_n$. The distance $\mathsf{d}_\circ(\pi,\sigma)$ between π and σ is defined as the minimum number of translocations needed to transform π into σ, i.e., $\mathsf{d}_\circ(\pi,\sigma)$ equals the smallest number m such that there exists a sequence of translocations $\phi_1, \phi_2, \ldots, \phi_m$ for which $\sigma = \pi\phi_1\phi_2\cdots\phi_m$.

Observe that $\mathsf{d}_\circ(\cdot,\cdot)$ is nonnegative and symmetric. It also satisfies the triangle inequality, namely, for any π, σ and ω in $\mathbb{S}_n$, one has

$$\mathsf{d}_\circ(\pi,\sigma) \le \mathsf{d}_\circ(\pi,\omega) + \mathsf{d}_\circ(\omega,\sigma).$$

Therefore, it is indeed a distance metric over the space $\mathbb{S}_n$.

[3]Note that a bounded length translocation in a permutation π is closely related to a bounded L_1-metric error in π^{-1}, studied in [25].

For $\pi, \sigma \in \mathbb{S}_n$, the distance $\mathsf{d}_\circ(\pi,\sigma)$ is closely related to the length of the longest common subsequence of π and σ, denoted by $l(\pi,\sigma)$. In fact, as shown in Proposition 3, $\mathsf{d}_\circ(\pi,\sigma)$ equals the *Ulam distance* [29] between π and σ, where the Ulam distance is defined as $n - l(\pi,\sigma)$. Although the Ulam distance has received some attention in the computer science community, to the best of the authors' knowledge, codes in the Ulam distance were not reported in the literature, with the notable exception of the single-error-correction method by Levenshtein [21] and the asymptotically zero-rate codes presented in [30] by Beame *et al.*

We start our subsequent discussion with the definition of the notion of invariance. A metric d over $\mathbb{S}_n$ is right-invariant if, for all $\pi, \sigma, \omega \in \mathbb{S}_n$, we have $\mathsf{d}(\pi,\sigma) = \mathsf{d}(\pi\omega, \sigma\omega)$. Similarly, d is left-invariant if $\mathsf{d}(\pi,\sigma) = \mathsf{d}(\omega\pi, \omega\sigma)$. Intuitively, a right-invariant metric is invariant with respect to *reordering* of elements and a left-invariant metric is invariant with respect to *relabeling* of elements.

The distance $\mathsf{d}_\circ$ is a left-invariant metric. To prove this simple observation, consider three arbitrary permutations $\pi, \sigma, \omega \in \mathbb{S}_n$ with $\mathsf{d}_\circ(\pi,\sigma) = m$. Then, there exists a sequence $\phi_1, \phi_2, \ldots, \phi_m$ of translocations such that $\sigma = \pi\phi_1\phi_2\cdots\phi_m$. Multiplying both sides of the previous equality by ω on the left yields $\omega\sigma = \omega\pi\phi_1\phi_2\cdots\phi_m$. This implies that $\mathsf{d}_\circ(\omega\pi,\omega\sigma) \le m = \mathsf{d}_\circ(\pi,\sigma)$. Conversely, we may repeat the same argument using $\omega\pi, \omega\sigma$ and ω^{-1} instead of π, σ, and ω, to obtain $\mathsf{d}_\circ(\pi,\sigma) \le \mathsf{d}_\circ(\omega\pi,\omega\sigma)$. This proves the desired invariance property.

The length of the longest common subsequence of two permutations is also left-invariant. To prove this claim, let us consider again three arbitrary permutations $\pi, \sigma, \omega \in \mathbb{S}_n$ with $l(\pi,\sigma) = m$. Then, there exists a longest common subsequence $i_1, i_2, \ldots, i_m$ of π and σ. Here, as anywhere else in this paper, we assume that one may choose, according to some arbitrary but fixed rule, *one* longest common subsequence if the longest common sequence is not unique. It follows that $\omega(i_1), \omega(i_2), \ldots, \omega(i_m)$ is a subsequence of both $\omega\pi$ and $\omega\sigma$, and thus, $l(\omega\pi,\omega\sigma) \ge l(\pi,\sigma)$. On the other hand, by considering the permutations $\omega\pi, \omega\sigma$, and ω^{-1} instead of π, σ, and ω, it can also be shown that $l(\pi,\sigma) \ge l(\omega\pi,\omega\sigma)$. This proves that l is left-invariant.

We next show that the translocation distance $\mathsf{d}_\circ(\pi,\sigma)$ equals the Ulam distance. More details about the Ulam distance and the longest common subsequence of permutations may be found in [9] and [31].

Proposition 3: For $\pi, \sigma \in \mathbb{S}_n$, the distance $\mathsf{d}_\circ(\pi,\sigma)$ equals $n - l(\pi,\sigma)$, i.e., the distance used for assessing the effect of translocations on permutation codes equals the Ulam distance between π and σ.

Proof: By the left-invariance of $\mathsf{d}_\circ$ and l, we may assume that one of the permutations is the identity permutation e since otherwise, instead of $\mathsf{d}_\circ(\pi,\sigma) = n - l(\pi,\sigma)$, we can show that $\mathsf{d}_\circ\left(\sigma^{-1}\pi, e\right) = n - l\left(\sigma^{-1}\pi\right)$. It thus suffices to prove that $\mathsf{d}_\circ(\sigma, e) = n - l(\sigma)$, where $l(\sigma) = l(\sigma, e)$ is the length of the longest increasing subsequence of σ.

Let S_ℓ denote the set of elements in the longest increasing subsequence of the permutation σ. Clearly, it is possible to transform σ into e with at most $n - l(\sigma)$ translocations. This can be

achieved by applying translocations that each move one element from the set $[n]\backslash S_\ell$ to its position in the identity permutation e. Hence, $\mathsf{d}_\circ(\sigma, e) \leq n - l(\sigma)$.

Next, we show that $\mathsf{d}_\circ(\sigma, e) \geq n - l(\sigma)$. We start with σ and transform it into e by applying a sequence of translocations. Every translocation increases the length of the longest increasing subsequence by at most one. Hence, we need at least $n - l(\sigma)$ translocations to transform σ into e, and thus, $\mathsf{d}_\circ(\sigma, e) \geq n - l(\sigma)$. □

Henceforth, we shall refer to $\mathsf{d}_\circ$ as the Ulam distance. The Ulam distance and other notions introduced in this section easily extend to permutations over a set $P \subseteq [n]$.

Note that a translocation may correspond to either a left- or a right-translocation. As seen from the example in Fig. 2, right-translocations correspond to general cell leakage models. On the other hand, left-translocations assume that the charge of a cell is increased above the level of other cells. We, therefore, also introduce the notion of the *right-translocation distance*. As will be seen from our subsequent discussion, the Ulam distance is much easier to analyze than the right-translocation distance and represents a natural lower bound for this distance.

The Ulam distance is closely related to Levenshtein's insertion/deletion distance, defined as the number of deletions and insertions required to transform one sequence into another, and denoted by $\rho(\cdot, \cdot)$. Levenshtein [21] showed that, for sequences of length n, $\rho(u, v) = 2(n - l(u, v))$. This equality also holds for permutations and thus

$$\rho(\pi, \sigma) = 2\mathsf{d}_\circ(\pi, \sigma)$$

for $\pi, \sigma \in \mathbb{S}_n$. This result may be also deduced directly, by observing that a translocation consists of a deletion and an insertion.

It is also of interest to see how the Ulam distance compares to the Kendall τ distance used in classical rank modulation coding. The *Kendall τ distance* $\mathsf{d}_\tau(\pi, \sigma)$ between $\pi \in \mathbb{S}_n$ and $\sigma \in \mathbb{S}_n$ is defined as the minimum number of adjacent transpositions required to change π into σ. A distance measure related to the Kendall τ is the transposition distance, also known as the Cayley distance. The transposition distance between two permutations π and σ of $\mathbb{S}_n$ is denoted by $\mathsf{d}_T(\pi, \sigma)$, and equals the smallest number of (not necessarily adjacent) transpositions needed to transform π into σ. The transposition distance $\mathsf{d}_T(\pi, \sigma)$, as shown by Cayley [32], equals n minus the number of cycles in the permutation $\sigma^{-1}\pi$.

Since a translocation of length ℓ can be represented as ℓ adjacent transpositions, and since an adjacent transposition is a translocation, it is easy to see that

$$\frac{1}{n-1}\,\mathsf{d}_\tau(\pi, \sigma) \leq \mathsf{d}_\circ(\pi, \sigma) \leq \mathsf{d}_\tau(\pi, \sigma).$$

Both the upper bound and the lower bound are tight: the upper bound is achieved for σ obtained from π via a single adjacent transposition, while the lower bound is achieved for, say, $\pi = e$ and $\sigma = (2, 3, \ldots, n, 1)$. It is also straightforward to show that the diameter of $\mathbb{S}_n$ with respect to the Ulam distance equals $n - 1$. Observe that the above inequalities imply that the Ulam distance is not within a constant factor from the Kendall τ distance, so that code constructions and bounds specifically derived for the latter distance measure are not tight and sufficiently efficient with respect to the Ulam distance.

A similar pair of bounds may be shown to hold for the Ulam distance and the Hamming distance between two permutations. The Hamming distance between permutations π and σ, denoted by $\mathsf{d}_H(\pi, \sigma)$, is defined as the number of positions i for which $\pi(i)$ and $\sigma(i)$ differ.

Let $F(\pi, \sigma) = \{i \in [n] : \pi(i) = \sigma(i)\}$. The subsequence of π consisting of elements $\pi(i), i \in F(\pi, \sigma)$, is also a subsequence of σ, and thus, $\mathsf{d}_\circ(\pi, \sigma) = n - l(\pi, \sigma) \leq n - |F(\pi, \sigma)| = \mathsf{d}_H(\pi, \sigma)$. Furthermore, since for any two permutations $\pi, \sigma \in \mathbb{S}_n$ one has $\mathsf{d}_H(\pi, \sigma) \leq n$, it follows that $\mathsf{d}_H(\pi, \sigma) \leq n\mathsf{d}_\circ(\pi, \sigma)$. Thus

$$\frac{1}{n}\mathsf{d}_H(\pi, \sigma) \leq \mathsf{d}_\circ(\pi, \sigma) \leq \mathsf{d}_H(\pi, \sigma). \tag{1}$$

These inequalities are sharp. For the upper bound, consider $\pi = (1, 2, \ldots, n)$ and $\sigma = (n, \ldots, 2, 1)$, with n odd. For the lower-bound, let $\pi = (1, 2, \ldots, n)$ and $\sigma = (2, 3, \ldots, n, 1)$ so that $\mathsf{d}_H(\pi, \sigma) = n$ and $\mathsf{d}_\circ(\pi, \sigma) = 1$.

Next, we consider the transposition distance. Note that each transposition may be viewed as two translocations, implying that $\mathsf{d}_\circ(\pi, \sigma) \leq 2\mathsf{d}_T(\pi, \sigma)$. It is also immediate that $\frac{1}{n-1}\mathsf{d}_T(\pi, \sigma) \leq \mathsf{d}_\circ(\pi, \sigma)$. Hence, we have

$$\frac{1}{n-1}\mathsf{d}_T(\pi, \sigma) \leq \mathsf{d}_\circ(\pi, \sigma) \leq 2\mathsf{d}_T(\pi, \sigma).$$

The relationship between the Hamming distance and the transposition distance can be explained as follows. When transforming π into σ using transpositions, each transposition decreases the Hamming distance between the two permutations by at most two. Hence, $\mathsf{d}_T(\pi, \sigma) \geq \mathsf{d}_H(\pi, \sigma)/2$. Sorting a permutation of length d requires at most d transpositions. Thus, $\mathsf{d}_T(\pi, \sigma) \leq \mathsf{d}_H(\pi, \sigma)$. These inequalities result in

$$\frac{1}{2}\mathsf{d}_H(\pi, \sigma) \leq \mathsf{d}_T(\pi, \sigma) \leq \mathsf{d}_H(\pi, \sigma). \tag{2}$$

If $\pi \neq \sigma$, then $\mathsf{d}_T(\pi, \sigma) \leq \mathsf{d}_H(\pi, \sigma) - 1$.

There exist many embedding methods for permutations, allowing one set of permutations with desirable properties according to a given distance to be mapped into another set of permutations with good properties in another metric space. In subsequent sections, we exhibit a method for interleaving permutations with good Hamming distance so as to obtain permutations with large minimum Ulam distance.

A. Right-Translocation Distance

We describe next how to specialize the Ulam distance for the case that only right-translocations are allowed as error events.

Definition 4: Let $\pi, \sigma \in \mathbb{S}_n$ and denote by $R_t(\pi, \sigma)$ the minimum number of right-translocations required to transform π into σ. For two permutations $\pi_1, \pi_2 \in \mathbb{S}_n$, the right-translocation distance $\vec{\mathsf{d}}_\circ(\pi_1, \pi_2)$ is defined as

$$\vec{\mathsf{d}}_\circ(\pi_1, \pi_2) = 2\min_\sigma \max\{R_t(\pi_1, \sigma), R_t(\pi_2, \sigma)\}.$$

We demonstrate next that $\vec{\mathsf{d}}_\circ$ is in fact a metric by proving that it satisfies the triangle inequality; the other metric properties may be readily verified using the definition of the distance.

Consider three permutations, π_1, π_2, and π_3, and let

$$\sigma_{12} = \arg\min_{\sigma} \max\left\{R_t\left(\pi_1,\sigma\right), R_t\left(\pi_2,\sigma\right)\right\}$$
$$\sigma_{23} = \arg\min_{\sigma} \max\left\{R_t\left(\pi_2,\sigma\right), R_t\left(\pi_3,\sigma\right)\right\}.$$

Suppose that $\alpha_i, 1 \leq i \leq m_\alpha = R_t\left(\pi_1,\sigma_{12}\right)$, are right-translocations and that $\beta_i, 1 \leq i \leq m_\beta = R_t\left(\pi_2,\sigma_{12}\right)$, are left-translocations such that

$$\begin{aligned}\pi_1\alpha_1\alpha_2\cdots\alpha_{m_\alpha} &= \sigma_{12}\\ \sigma_{12}\beta_1\beta_2\cdots\beta_{m_\beta} &= \pi_2.\end{aligned} \tag{3}$$

Similarly, suppose that $\gamma_i, 1 \leq i \leq m_\gamma = R_t\left(\pi_2,\sigma_{23}\right)$, are right-translocations and that $\delta_i, 1 \leq i \leq m_\delta = R_t\left(\pi_3,\sigma_{23}\right)$, are left-translocations such that

$$\begin{aligned}\pi_2\gamma_1\gamma_2\cdots\gamma_{m_\gamma} &= \sigma_{23}\\ \sigma_{23}\delta_1\delta_2\cdots\delta_{m_\delta} &= \pi_3.\end{aligned} \tag{4}$$

Note that the existence of the sets of translocations $\{\alpha_i\}, \{\beta_i\}, \{\gamma_i\}, \{\delta_i\}$ follows from the definition of R_t.

From (3) and (4), we have

$$\pi_1\alpha_1\cdots\alpha_{m_\alpha}\beta_1\cdots\beta_{m_\beta}\gamma_1\cdots\gamma_{m_\gamma}\delta_1\cdots\delta_{m_\delta} = \pi_3. \tag{5}$$

Right-translocations and left-translocations have the following simple property. Suppose β is a left-translocation and γ is a right-translocation. We can then find a right-translocation γ' and a left-translocation β' such that $\beta\gamma = \gamma'\beta'$, where either γ' or β' are allowed to be the identity permutation. Hence, (5) may be rewritten as

$$\pi_1\alpha_1\cdots\alpha_{m_\alpha}\gamma'_1\cdots\gamma'_{m_\gamma}\beta'_1\cdots\beta'_{m_\beta}\delta_1\cdots\delta_{m_\delta} = \pi_3. \tag{6}$$

Next, let $\sigma_{13} = \pi_1\alpha_1\cdots\alpha_{m_\alpha}\gamma'_1\cdots\gamma'_{m_\gamma}$. Note that σ_{13} is not required to be the minimizer of $\max\left\{R_t\left(\pi_1,\sigma\right), R_t\left(\pi_3,\sigma\right)\right\}$.

From (6) and the fact that $\{\alpha_i\}$ and $\{\gamma'_i\}$ are right-translocations and $\{\beta'_i\}$ and $\{\delta_i\}$ are left-translocations, it follows that

$$\begin{aligned}R_t\left(\pi_1,\sigma_{13}\right) &\leq m_\alpha + m_\gamma = R_t\left(\pi_1,\sigma_{12}\right) + R_t\left(\pi_2,\sigma_{23}\right),\\ R_t\left(\pi_3,\sigma_{13}\right) &\leq m_\beta + m_\delta = R_t\left(\pi_2,\sigma_{12}\right) + R_t\left(\pi_3,\sigma_{23}\right),\end{aligned}$$

and thus

$$\begin{aligned}\vec{\mathsf{d}}_\circ\left(\pi_1,\pi_3\right) &\leq 2\max\{R_t(\pi_1,\sigma_{13}), R_t(\pi_3,\sigma_{13})\}\\ &\leq 2\max\left\{R_t\left(\pi_1,\sigma_{12}\right) + R_t\left(\pi_2,\sigma_{23}\right),\right.\\ &\qquad \left. R_t\left(\pi_2,\sigma_{12}\right) + R_t\left(\pi_3,\sigma_{23}\right)\right\}\\ &\leq 2\max\{R_t(\pi_1,\sigma_{12}), R_t(\pi_2,\sigma_{12})\} +\\ &\qquad 2\max\{R_t(\pi_2,\sigma_{23}), R_t(\pi_3,\sigma_{23})\}\\ &= \vec{\mathsf{d}}_\circ\left(\pi_1,\pi_2\right) + \vec{\mathsf{d}}_\circ\left(\pi_2,\pi_3\right).\end{aligned}$$

Hence, $\vec{\mathsf{d}}_\circ$ satisfies the triangle inequality.

The definition of $\vec{\mathsf{d}}_\circ$ implies that for two permutations $\pi_1, \pi_2 \in \mathbb{S}_n$, one has $\vec{\mathsf{d}}_\circ\left(\pi_1,\pi_2\right) \leq 2t$ if and only if there exists a permutation $\sigma \in \mathbb{S}_n$ such that $R_t(\pi_1,\sigma) \leq t$ and $R_t(\pi_2,\sigma) \leq t$. Hence, a code C is t-right-translocation correcting if and only if $\vec{\mathsf{d}}_\circ\left(\pi_1,\pi_2\right) > 2t$ for all $\pi_1, \pi_2 \in C$, $\pi_1 \neq \pi_2$. This means that under the given distance constraint, it is not possible to confuse the actual codeword π_1 with another (wrong) codeword π_2.

Observe that the following bound holds:

$$\mathsf{d}_\circ\left(\pi_1,\pi_2\right) \leq \vec{\mathsf{d}}_\circ\left(\pi_1,\pi_2\right).$$

It is straightforward to characterize the minimum number of right-translocations needed to transform one permutation into another, as we show next.

Definition 5: For $\pi, \sigma \in \mathbb{S}_n$, let

$$J\left(\pi,\sigma\right) := \left\{i \in [n] : \exists j, \pi^{-1}(i) < \pi^{-1}(j) \text{ and } \sigma^{-1}(i) > \sigma^{-1}(j)\right\}.$$

Note that $|J|$ is left-invariant since, for $\pi, \sigma, \omega \in \mathbb{S}_n$,

$$\begin{aligned}|J(\omega\pi,\omega\sigma)| &= \left|\left\{i \in [n] : \exists j, \pi^{-1}\omega^{-1}(i) < \pi^{-1}\omega^{-1}(j)\right.\right.\\ &\qquad \left.\left. \text{and } \sigma^{-1}\omega^{-1}(i) > \sigma^{-1}\omega^{-1}(j)\right\}\right|\\ &= \left|\left\{i' \in [n] : \exists j', \pi^{-1}(i') < \pi^{-1}(j')\right.\right.\\ &\qquad \left.\left. \text{and } \sigma^{-1}(i') > \sigma^{-1}(j')\right\}\right|\\ &= \left|J(\pi,\sigma)\right|\end{aligned}$$

where for the first equality, we have used the fact that $(\omega\pi)^{-1} = \pi^{-1}\omega^{-1}$ and $(\omega\sigma)^{-1} = \sigma^{-1}\omega^{-1}$, and the second equality can be obtained by letting $i' = \omega^{-1}(i)$ and $j' = \omega^{-1}(j)$. Furthermore, using similar arguments as for the proof of left-invariance of $\mathsf{d}_\circ$, it can be shown that R_t is left-invariant.

Lemma 6: Let $\pi, \sigma \in \mathbb{S}_n$. Then

$$R_t\left(\pi,\sigma\right) = \left|J\left(\pi,\sigma\right)\right|.$$

Proof: It suffices to show that

$$R_t(\pi, e) = |J(\pi, e)|$$

where

$$|J(\pi,e)| = \left|\left\{i \in [n] : \exists j < i, \pi^{-1}\left(i\right) < \pi^{-1}(j)\right\}\right|.$$

Let π_1 be obtained from π by applying a right-translocation that moves some element k to the right. Every element of $J(\pi,e)\backslash\{k\}$ is also in $J(\pi_1,e)$ as each element of $J(\pi,e)\backslash\{k\}$ is involved in at least one inversion which is not affected by moving k. Hence, $|J(\pi_1,e)| \geq |J(\pi,e)| - 1$ with equality if $J(\pi_1,e) = J(\pi,e)\backslash\{k\}$. Repeating the same argument yields $R_t(\pi,e) \geq |J(\pi,e)| - |J\left(e,e\right)| = |J(\pi,e)|$.

Conversely, to transform π into e, it suffices to apply to each $i \in J$ the shortest right-translocation that moves this element to the smallest position i' such that to the left of position i' are all the elements smaller than i. Hence, $R_t(\pi,e) \leq |J(\pi,e)|$. □

For permutations $\pi, \sigma \in \mathbb{S}_n$, the difference between $R_t(\pi,\sigma)$ and $\mathsf{d}_\circ(\pi,\sigma)$ may be as large as $n-2$. This may be seen by letting $\pi = (2,3,\ldots,n,1)$ and $\sigma = e$, and observing that $R_t(\pi,\sigma) = n-1$ and $\mathsf{d}_\circ(\pi,\sigma) = 1$. Furthermore, it can be shown that this is the largest possible gap. To prove this fact, first note that $R_t(\pi,\sigma) = 0$ if and only if $\mathsf{d}_\circ(\pi,\sigma) = 0$ and thus to obtain a positive gap one must have $\mathsf{d}_\circ(\pi,\sigma) \geq 1$. We also have $\mathsf{d}_\circ(\pi,\sigma) \leq R_t(\pi,\sigma) \leq n-1$. Hence, $1 \leq \mathsf{d}_\circ(\pi,\sigma) \leq R_t(\pi,\sigma) \leq n-1$, which implies that the gap is at most $n-1-1 = n-2$.

In the sections to follow, we mainly focus our attention on the Ulam distance.

III. Bounds on the Size of Codes

A. Codes in the Ulam Metric

Henceforth, a *permutation code*, or simply a code, of length n and minimum distance d in a metric d refers to a subset C of $\mathbb{S}_n$ such that for all distinct $\pi, \sigma \in C$, we have $\mathsf{d}(\pi, \sigma) \geq d$. The term a capacity achieving code is reserved for a code with maximum rate and a given minimum distance in a given metric space. We also let $A_\circ(n, d)$ be the maximum size of a permutation code of length n and minimum Ulam distance d.

Proposition 7: For all integers n and d with $n \geq d \geq 1$, we have

$$A_\circ(n, d) \geq \frac{(n-d+1)!}{\binom{n}{d-1}}.$$

Proof: Let $B_\circ(r)$ be the number of permutations at Ulam distance at most r from a given permutation. From left-invariance, we have $B_\circ(r) = |\{\sigma : \mathsf{d}_\circ(\sigma, e) \leq r\}|$. The permutations that are within Ulam distance r from e are precisely the permutations σ with $l(e, \sigma) \geq n - r$. There are $\binom{n}{r}$ ways to choose the first $n - r$ elements of the longest common subsequence of e and σ and at most $\frac{n!}{(n-r)!}$ ways to arrange the remaining elements of σ. Hence

$$B_\circ(r) \leq \binom{n}{r} \frac{n!}{(n-r)!}.$$

From the Gilbert–Varshamov bound, we have $A_\circ(n, d) \geq \frac{n!}{B_\circ(d-1)}$ and thus

$$A_\circ(n, d) \geq \frac{n!}{\binom{n}{d-1} \frac{n!}{(n-d+1)!}}$$

which completes the proof. □

Proposition 8: For all $n, d \in \mathbb{Z}$, with $n \geq d \geq 1$:

$$A_\circ(n, d) \leq (n-d+1)!\,.$$

Proof: We provide two proofs for this bound. The first proof is based on a projection argument first described in [23], while the second proof is based on a standard counting argument.

1) Let C be a code of length n, size M, and minimum distance d. Let k be the smallest integer such that $\pi_{\{1,\ldots,k+1\}} \neq \sigma_{\{1,\ldots,k+1\}}$ for all distinct $\pi, \sigma \in C$. Hence, $M \leq (k+1)!$. By definition, there exist $\pi, \sigma \in C$ such that $\pi_{\{1,\ldots,k\}} = \sigma_{\{1,\ldots,k\}}$. So, $l(\pi, \sigma) \geq k$ and thus $d \leq \mathsf{d}_\circ(\pi, \sigma) \leq n - k$. Hence, $M \leq (n-d+1)!$
2) Again, let C be a code of length n, size M, and minimum distance d. Since the minimum distance is d, all $M\binom{n}{n-d+1}$ subsequences of length $n-d+1$ of the codewords of C are unique. There are $\frac{n!}{(d-1)!}$ possible subsequences of length $n - d + 1$. Hence

$$M\binom{n}{n-d+1} \leq \frac{n!}{(d-1)!}$$

which implies that $M \leq (n-d+1)!$

□

From the two previous propositions, we obtain

$$\frac{(n-d+1)!}{\binom{n}{d-1}} \leq A_\circ(n, d) \leq (n-d+1)! \tag{7}$$

In the remainder of this paper, all limits are evaluated for $n \to \infty$, unless stated otherwise. Furthermore, we assume that the limits exist.

Lemma 9: The following results hold:

1)
$$\lim \frac{\ln(n - d(n))!}{\ln n!} = 1 - \lim \frac{d(n)}{n}$$

2)
$$\lim \frac{\ln \frac{n!}{d(n)!}}{\ln n!} = 1 - \lim \frac{d(n)}{n}$$

3)
$$\lim \frac{\ln \binom{n}{d(n)}}{\ln n!} = 0.$$

Proof: All claims follow easily from the asymptotic formula $\ln(n!) = n \ln n + O(n)$. □

Let $\mathcal{C}_\circ(d)$ denote the asymptotic capacity of translocation codes with minimum Ulam distance $d = d(n)$, i.e., $\mathcal{C}_\circ(d) = \lim \frac{\ln A_\circ(n,d)}{\ln n!}$.

Theorem 10: $\mathcal{C}_\circ(d) = 1 - \lim \frac{d(n)}{n}$.

Proof: From (7), we have

$$\frac{\ln(n-d+1)! - \ln\binom{n}{d-1}}{\ln n!} \leq \frac{\ln A_\circ(n, d)}{\ln n!} \leq \frac{\ln(n-d+1)!}{\ln n!}. \tag{8}$$

Taking the limit of (8) and using Lemma 9 proves the theorem. □

At this point, it is worth observing that the problem of bounding the longest common subsequence in permutations has been recently studied in a combinatorial framework [30]. There, the question of interest was to determine the minimum length of the longest common subsequence between any two distinct permutations from a set of k permutations of length n. When translated into the terminology of translocation codes, the problem reduces to finding $d_k(n)$, the largest possible minimum Ulam distance of a set of k permutations of $\mathbb{S}_n$.

The bounds derived in [30] are constructive, but they hold only in the zero-capacity domain of the code parameters. A more detailed description of one of the constructions of [30] is presented in Section IV. The bounds of [30] imply that $d_k(n) \geq n - 32(nk)^{1/3}$ for $3 \leq k \leq \sqrt{n}$. Hence, for $n - 32\sqrt{n} \leq d \leq n - 32(3n)^{1/3}$:

$$A_\circ(n, d) \geq \frac{1}{n}\left(\frac{n-d}{32}\right)^3.$$

Furthermore, for $k \geq 4$, $d_k(n) \geq n - \left\lceil n^{1/(k-1)} \right\rceil^{k/2-1}$. For $k \geq 2(1 + \log_2 n)$, this bound is of no practical use.

For $1 + \log_2 n \leq k < 2(1 + \log_2 n)$, one has $d_k(n) \geq n - 2^{k/2-1}$ which implies that

$$A_\circ(n, d) \geq 2(1 + \log_2(n-d))$$

for $d \leq n-\sqrt{n/2}$. Similar bounds can be obtained for $A_\circ(n,d)$ by assuming that $m-1 < n^{\frac{1}{k-1}} \leq m$ for some integer $m \leq \lceil n^{1/3} \rceil$. Note that although these results hold for the zero-capacity regime, they still may be useful for finite code length analysis.

Remark: Similar bounds may be derived for the asymmetric regime of translocation error-correcting codes. For this purpose, let $B'(r) = |\{\sigma : R_t(e,\sigma) \leq r\}|$ and $\vec{B}(r) = \left|\left\{\sigma : \vec{\mathrm{d}}_\circ(e,\sigma) \leq r\right\}\right|$. Then

$$\frac{n!}{B_\circ(2t)} \leq \frac{n!}{\vec{B}(2t)} \leq \vec{A}(n,2t+1) \leq \frac{n!}{B'(t)}$$

where $\vec{A}(n,d)$ denotes the maximum size of a permutation code with minimum right-translocation distance d.

B. Permutation Codes in Other Metrics

Translocation errors, and consequently, translocation error-correcting codes, are difficult to analyze directly. On the other hand, as already pointed out, the Ulam distance is related to various other metrics well studied in the coding theory and mathematics literature. Since the constructions in subsequent sections rely on codes for other distance metrics on permutations, we provide a brief overview of the state-of-the-art results pertaining to the Hamming, transposition, and Kendall τ metrics. We also supplement the known findings with a number of new comparative results for the metrics under consideration.

1) Hamming Metric: Codes in the Hamming metric have a long history, dating back to the work [1]. The Hamming metric is a suitable distance measure for use in power-line communication systems, database management, and other applications.

Let $A_H(n,d)$ denote the largest number of permutations of length n and minimum Hamming distance d. Frankl and Deza [33, Th. 4] and Deza and Vanstone [34] showed that

$$\frac{n!}{B_H(d-1)} \leq A_H(n,d) \leq \frac{n!}{(d-1)!}$$

where $B_H(r)$ is the volume of the sphere of radius r in the space of permutations with Hamming metric. Improvements of these results for some special cases were also obtained via linear programing methods; see, for example, [35].

Let D_i denote the number of derangements of i objects, i.e., the number of permutations of $[i]$ at Hamming distance i from the identity permutation. It can be shown that $B_H(r) = 1 + \sum_{i=2}^{r} \binom{n}{i} D_i$. Hence

$$\begin{aligned} B_H(d-1) &= 1 + \sum_{i=2}^{d-1} \binom{n}{i} D_i \leq \sum_{i=1}^{d-1} \frac{n!}{(n-i)!} \\ &\leq (d-1)\frac{n!}{(n-d+1)!} \end{aligned}$$

where the first inequality follows from the fact that $D_i \leq i!$. Note that although a more precise asymptotic characterization for the number of derangements is known, namely

$$\lim_{\ell\to\infty} \frac{D_\ell}{\ell!} = \frac{1}{e}$$

the simple bound $D_i \leq i!$ is sufficiently tight for the capacity computation.

The aforementioned results lead to

$$\frac{(n-d+1)!}{d-1} \leq A_H(n,d) \leq \frac{n!}{(d-1)!}.$$

Let $\mathcal{C}_H(d)$ denote the capacity of permutation codes under the Hamming distance d, i.e., $\mathcal{C}_H(d) = \lim \frac{\ln A_H(n,d)}{\ln n!}$. Lemma 9 implies the following theorem.

Theorem 11: $\mathcal{C}_H = 1 - \lim \frac{d(n)}{n}$.

2) Transposition Metric: Let $A_T(n,d)$ denote the maximum size of a code with minimum transposition distance d_T at least d. From (2), we have

$$A_H(n,2d) \leq A_T(n,d) \leq A_H(n,d).$$

Using the aforementioned bounds, we have the following theorem regarding the capacity $\mathcal{C}_T(d)$ of permutation codes of minimum distance d in the transposition metric.

Theorem 12: The capacity of permutation codes of minimum distance d in the transposition metric is bounded as

$$1 - 2\lim\frac{d(n)}{n} \leq \mathcal{C}_T(d) \leq 1 - \lim\frac{d(n)}{n}.$$

3) Kendall τ Metric: Let $A_K(n,d)$ denote the largest cardinality of a permutation code of length n with minimum Kendall τ distance d, and let $\mathcal{C}_K(d) = \lim \frac{\ln A_K(n,d)}{\ln n!}$. Barg and Mazumdar [23, Th. 3.1] showed that

$$\mathcal{C}_K(d) = 1-\epsilon, \qquad \text{for } d = \Theta\left(n^{1+\epsilon}\right).$$

Note that for the Kendall τ, the maximum distance between two permutations may be as large as $\Theta(n^2)$. On the other hand, the diameter of $\mathbb{S}_n$ with respect to the Ulam distance is $\Theta(n)$.

4) Levenshtein Metric: The bounds on the size of deletion/insertion correcting codes in the more general case of codes with distinct symbols were first derived by Levenshtein in his landmark paper [21]. The lower bound relies on the use of Steiner triple systems and designs [21]. More precisely, let $D(n,q)$ be the largest cardinality of a set of n-subsets of the set $\{0,1,\ldots,q-1\}$ with the property that every $(n-1)$-subset of $\{0,1,\ldots,q-1\}$ is a subset of at most one of the n-subsets. Then, the following results holds for the cardinality $\mathcal{A}_L(n,q)$ of the largest single-deletion correcting codes consisting of codewords in $\{0,1,\ldots,q-1\}^n$ with distinct symbols [21]:

$$(n-1)!D(n,q) \leq \mathcal{A}_L(n,q) \leq \frac{q!}{n(q-n+1)!}$$

IV. Single-Error-Correcting Codes for Translocations and Right-Translocations

This section contains constructions for single-translocation error detecting and single-translocation error-correcting codes. For the latter case, we exhibit two constructions: one for translocations and another for right-translocations.

A. Detecting a Single Translocation Error

We start by describing a code that can detect a single translocation error. From the discussion in Section II, recall that the

Ulam distance is half of the Levenshtein distance, and thus, any single-deletion correcting code may be used for detecting a single translocation error. An elegant construction for single-deletion correcting codes for permutation was described by Levenshtein in [21]. The resulting code has cardinality $(n-1)!$ and is optimal since, from Proposition 8, we have

$$A_{\circ}(n,2) \leq (n-2+1)! = (n-1)!$$

Hence, $A_{\circ}(n,2) = (n-1)!$.

Levenshtein's construction is of the following form.

Let

$$W_2^n = \left\{ u \in \{0,1\}^n : (n+1) \,|\, \sum_{i=1}^{n} iu(i) \right\}$$

where $a \,|\, b$ denotes that a is a divisor of b. For $\sigma \in \mathbb{S}_n$, let $Z(\sigma) = (z(1), \ldots, z(n-1))$ be a vector with

$$z(i) = \begin{cases} 0, & \text{if } \sigma(i) \leq \sigma(i+1), \\ 1, & \text{if } \sigma(i) > \sigma(i+1), \end{cases} \qquad i \in [n-1].$$

The code

$$C = \left\{ \sigma \in \mathbb{S}_n : Z(\sigma) \in W_2^{n-1} \right\} \tag{9}$$

of size $(n-1)!$ is capable of correcting a single deletion. Hence, this code can detect a single translocation error as well.

Let $\mathbb{S}_n^{(t)}$ be the set of sequences σ of length $n-t$ that can be obtained from some permutation in $\mathbb{S}_n$ by t deletions. In other words, $\mathbb{S}_n^{(t)}$ is the set of words of length $n-t$ from the alphabet $[n]$ without repetitions. A code $C_p \subseteq \mathbb{S}_n$ is a *perfect* code capable of correcting t deletions if, for every $\sigma \in \mathbb{S}_n^{(t)}$, there exists a unique $\pi \in C_p$ such that σ can be obtained from π by t deletions. It was shown in [21] that C in (9) is a perfect code capable of correcting a single deletion.

The minimum Levenshtein distance of C_p is $2(t+1)$ and thus the minimum Ulam distance of C_p is $t+1$. Since the size of $\mathbb{S}_n^{(t)}$ equals $\binom{n}{n-t}(n-t)!$ and $\binom{n}{t}$ elements of $\mathbb{S}_n^{(t)}$ can be obtained by t deletions from each $\sigma \in C_p$, we have that $|C_p| = (n-t)!$. Recall from Proposition 8 that the size of a code with minimum Ulam distance $t+1$ is $\leq (n-t)!$. Thus, a perfect code capable of correcting t deletions, if it exists, is a rate-optimal code in the Ulam metric. Although conditions for the existence of such codes were investigated in [21], both necessary and sufficient conditions are known only for a small number of special cases.

In Sections IV-B and IV-C, we describe codes capable of correcting a single right-translocation error and codes capable of correcting a single translocation error. In the constructions, we make use of a single-transposition error detecting code, described next.

A Single-Transposition Error Detecting Code: For $\sigma_1, \sigma_2 \in \mathbb{S}_n$, let $\mathsf{d}_T(\sigma_1, \sigma_2)$ as before denote the transposition distance between σ_1 and σ_2. The parity of a permutation σ is defined as the parity of $\mathsf{d}_T(\sigma, e)$. It is well known that applying a transposition to a permutation changes the parity of the permutation, and also that, for $n \geq 2$, half of the permutations in $\mathbb{S}_n$ are even and half of them are odd.[4] Hence, the code C containing all even permutations of $\mathbb{S}_n$ is a single-transposition error detecting code of length n and cardinality $n!/2$.

[4]More precisely, the symmetric group can be partitioned into the alternating group and its coset.

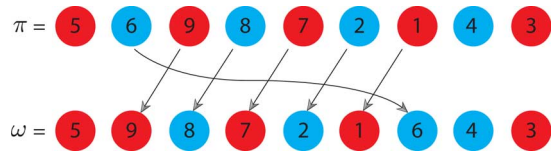

Fig. 3. Effect of the right-translocation error $\phi(2,7)$ on the codeword $\pi = (5,6,9,8,7,2,1,4,3)$. The result is the word $\omega = (5,9,8,7,2,1,6,4,3)$.

B. Correcting a Single Right-Translocation Error

Next, we present a construction for codes that correct a single right-translocation error. For this purpose, we first define the operation of permutation interleaving and the operation of code interleaving.

Definition 13: For vectors $\sigma_i, i \in [k]$, of lengths m_i with $m_1 \geq m_2 \geq \cdots \geq m_k \geq m_1 - 1$, the interleaved vector $\sigma = \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_k$ is obtained by alternatively placing the elements of $\sigma_1, \sigma_2, \ldots, \sigma_k$ in order. That is,

$$\sigma(j) = \sigma_i(\lceil j/k \rceil), \qquad 1 \leq j \leq \sum_{i=1}^{k} m_i \tag{10}$$

where $i \equiv j \pmod{k}$. For a class of k codes $C_i, i \in [k]$, let

$$C_1 \circ \cdots \circ C_k = \{\sigma_1 \circ \cdots \circ \sigma_k : \sigma_i \in C_i, i \in [k]\}. \tag{11}$$

For example, for vectors σ and π of length m, we have

$$\sigma \circ \pi = (\sigma(1), \pi(1), \sigma(2), \pi(2), \ldots, \sigma(m), \pi(m)) \tag{12}$$

and for vectors σ and π of lengths m and $m-1$, respectively, we have

$$\sigma \circ \pi = (\sigma(1), \pi(1), \sigma(2), \pi(2), \ldots, \sigma(m)). \tag{13}$$

The following proposition introduces codes that can correct a single right-translocation error. The decoding algorithm is contained in the proof of the proposition.

Proposition 14: Let $P_i, i = 1, 2$, be the set of odd and even numbers in $[n]$, respectively, and let C_i be the set of even permutations of P_i for $i = 1, 2$. The interleaved code $C = C_1 \circ C_2$ corrects a single right-translocation error.

Proof: Given the permutation $\omega \notin C$, we want to find the unique $\pi \in C$ such that $\omega = \pi\phi(i,j)$, $i < j$. An example is shown in Fig. 3, with $\omega = (5,9,8,7,2,1,6,4,3)$ and π unknown to the decoder.

The kth element of ω is out of place if $k \not\equiv \omega(k) \pmod{2}$. It is easy to see that

$$i = k_{\min} := \min\{k : k \not\equiv \omega(k) \pmod{2}\}$$

i.e., i equals the smallest integer k such that the kth element of ω is out of place. In the example shown in Fig. 3, $i = 2$. Finding j is slightly more complicated since we must consider two different cases depending on the parity of the length $j - i$ of the right-translocation.

Let $k_{\max} := \max\{k : k \not\equiv \omega(k) \pmod{2}\}$. If $j - i$ is odd, then $j = k_{\max}$. Otherwise, $j = k_{\max} + 1$. That is, the right-

translocation error is either $\phi(i, k_{\max})$ or $\phi(i, k_{\max}+1)$. Thus, the codeword π either equals $\pi' = \omega\phi(k_{\max}, i)$ or equals $\pi'' = \omega\phi(k_{\max}+1, i)$. In the example of Fig. 3, we have

$$\pi' = (5, 6, 9, 8, 7, 2, 1, 4, 3)$$
$$\pi' = (5, 4, 9, 8, 7, 2, 1, 6, 3).$$

To find which of the two cases is correct, we proceed as follows.

Since $\omega = \pi''\phi(i, k_{\max}+1)$ and $\pi' = \omega\phi(k_{\max}, i)$, we have

$$\begin{aligned}\pi' &= \pi''\phi(i, k_{\max}+1)\phi(k_{\max}, i)\\ &= \pi''\tau(i, k_{\max}+1).\end{aligned}$$

Recall that if $j-i$ is odd, then $j = k_{\max}$, and if $j-i$ is even, then $j = k_{\max}+1$. Hence, $i \equiv k_{\max}+1 \pmod 2$. In both π' and π'', the parity of the elements is the same as the parity of their positions. Thus, the transposition $\tau(i, k_{\max}+1)$ affects only elements of the same parity as i. Hence, if i is odd, then $\pi'_{P_2} = \pi''_{P_2} = \omega_{P_2} = \pi_{P_2}$ and if i is even, then $\pi'_{P_1} = \pi''_{P_1} = \omega_{P_1} = \pi_{P_1}$.

Without loss of generality, assume that i is even. Then, $\pi'_{P_1} = \pi''_{P_1} = \omega_{P_1} = \pi_{P_1}$ and the subwords π'_{P_2} and π''_{P_2} differ in one transposition. Since C_2 has minimum transposition distance two, only one of π'_{P_2} and π''_{P_2} belongs to C_2, and so π can be uniquely determined as being either equal to π' or π''. □

The cardinality of the interleaved code $C_1 \circ C_2$ equals $\frac{1}{4}\left\lceil\frac{n}{2}\right\rceil!\left\lfloor\frac{n}{2}\right\rfloor!$, and its rate asymptotically equals

$$\frac{\ln(1/4) + 2\ln\left\lfloor\frac{n}{2}\right\rfloor!}{\ln n!} \sim \frac{n\ln n + O(n)}{n\ln n + O(n)} \sim 1.$$

C. Correcting a Single Translocation Error

The construction of Section IV-B can be extended to generate codes capable of correcting a single translocation error as stated in the following proposition. Although the proposition is stated for n being a multiple of three, it can be easily extended to other cases.

Proposition 15: Suppose n is a multiple of three. Let $P_i, i = 1, 2, 3$, be the set of numbers in $[n]$ that are equal to i modulo three, and let C_i be the set of even permutations of P_i for $i = 1$, 2, 3. The interleaved code $C = C_1 \circ C_2 \circ C_3$ corrects a single translocation error.

Proof: Suppose that π is the stored permutation, ω is the retrieved permutation, and that the error is the translocation $\phi(i, j)$. If $|i-j| = 1$, then $\phi(i, j)$ can be easily identified. Suppose that $|i-j| > 1$. The translocation $\phi(i, j)$ moves $|i-j|$ elements of π one position to the left, provided that $i < j$, or one position to the right, provided that $j < i$. In either case, one element moves in the "opposite direction" from the other elements. Hence, for $|i-j| > 1$, the direction of the translocation (left or right) can be identified.

Once the direction of the translocation is known, i can be found as follows: if the error is a right-translocation, then

$$i = \min\{k : \omega(k) \not\equiv k \pmod 3\}$$

and if the error is a left-translocation, then

$$i = \max\{k : \omega(k) \not\equiv k \pmod 3\}.$$

For simplicity, suppose the error is a right-translocation. The proof for left-translocations is similar. Let $k_{\max} := \max\{k : k \not\equiv \omega(k) \pmod 3\}$. We have the following three cases. If $j - i \equiv 0 \pmod 3$, then $j = k_{\max}+1$ and

$$\omega(k_{\max}) \equiv \omega(k_{\max}+1) \pmod 3.$$

If $j - i \equiv 1 \pmod 3$, then $j = k_{\max}$ and

$$\omega(k_{\max}) \not\equiv \omega(k_{\max}+1) \pmod 3.$$

Finally, if $j - i \equiv 2 \pmod 3$, then $j = k_{\max}$ and

$$\omega(k_{\max}) \equiv \omega(k_{\max}+1) \pmod 3.$$

So, if $\omega(k_{\max}) \not\equiv \omega(k_{\max}+1) \pmod 3$, then $j = k_{\max}$ and π is uniquely determined as $\omega\phi(k_{\max}, i)$. Otherwise, the error is either $\phi(i, k_{\max})$ or $\phi(i, k_{\max}+1)$. Let $\pi' = \omega\phi(k_{\max}, i)$ and $\pi' = \omega\phi(k_{\max}+1, i)$. Then, $\pi' = \pi''\tau(i, k_{\max}+1)$ and similar to the proof of Proposition 14, it can be shown that π' and π'' are not both in C. Hence, π can be determined as either being equal to π' or π''. □

Example 16: Consider the single translocation-correcting code for $n = 12$. For this case, we have $P_1 = \{1, 4, 7, 10\}, P_2 = \{2, 5, 8, 11\}$, and $P_3 = \{3, 6, 9, 12\}$.

Suppose that the stored codeword is π, the error is $\phi(i, j)$, and the retrieved word is

$$\omega = (1, 6, 10, 8, 3, 7, 5, 11, 12, 4, 2, 9).$$

Given ω, the decoder first identifies the elements that are out of order, i.e., elements that are not equivalent to their positions modulo three—in this case, $\{6, 10, 8, 3, 7, 5\}$. Since more than two elements are out of order, we have $|i-j| > 1$. Furthermore, since more than two elements have moved one position to the left, ϕ is a right-translocation. Observe that $k_{\max} = 7$ and that $\omega(k_{\max}) \equiv \omega(k_{\max}+1) \pmod 3$. Hence, we let

$$\pi' = \omega\phi(7, 2) = (1, 5, 10, 8, 7, 11, 4, 2)$$
$$\pi'' = \omega\phi(8, 2) = (1, 11, 10, 8, 7, 5, 4, 2).$$

We then have $\pi'_{P_2} = (5, 8, 11, 2)$ and $\pi''_{P_2} = (11, 8, 5, 2)$. Since only π''_{P_2} is an even permutation, the error is $\phi(2, 8)$ and thus $\pi = (1, 11, 6, 10, 8, 3, 7, 5, 12, 4, 2, 9)$. □

The cardinality of the code equals $\left(\frac{1}{2}\left(\frac{n}{3}\right)!\right)^3$, while its rate equals

$$\frac{3\ln(1/2) + 3\ln\left(\frac{n}{3}\right)!}{\ln n!} \sim \frac{n\ln n + O(n)}{n\ln n + O(n)} \sim 1.$$

V. t-Translocation Error-Correcting Codes

We describe next a number of general constructions for t-translocation error-correcting codes. We start with an extension of the interleaving methods from Section IV.

A. Interleaving Codes in Hamming Metric

We construct a family of codes with Ulam distance $2t+1$, length $n = s(2t+1)$ for some integer $s \geq 4t+1$, and cardinality $M = (A_H(s, 4t+1))^{2t+1}$, where $A_H(s, d)$, as before,

denotes the maximum size of a permutation code with length s and minimum Hamming distance d. The construction relies on the use of $2t+1$ permutation codes, each with minimum Hamming distance at least $4t+1$. First, we present the proposed construction and then prove that the minimum Ulam distance of the code is at least $2t+1$.

For a given n and t, where $n \equiv 0 \pmod{2t+1}$, partition the set $[n]$ into $2t+1$ classes P_i, each of size s, with

$$P_i = \{j \in [n] : j \equiv i \pmod{2t+1}\}, \qquad i \in [2t+1]. \quad (14)$$

For example, for $t=2$ and $n=45$, one has

$$P_1 = \{1, 6, \ldots, 41\}$$
$$P_2 = \{2, 7, \ldots, 42\}$$
$$P_3 = \{3, 8, \ldots, 43\}$$

and so on.

For $i \in [2t+1]$, let C_i be a permutation code over P_i with minimum Hamming distance at least $4t+1$.[5] The code C is obtained by interleaving the codes C_i, i.e., $C = C_1 \circ \cdots \circ C_{2t+1}$, and is referred to as an interleaved code with $2t+1$ classes. In the interleaved code, the s elements of P_i occupy positions that are equivalent to i modulo $2t+1$.

The following theorem provides a lower-bound for the minimum Ulam distance of C. The proof of the theorem is presented after stating the required definitions and three technical lemmas.

Theorem 17: Assume we are given three positive integers s, t, $n = s(2t+1)$, and a partition of $[n]$ of the form given in (14). If, for $i \in [2t+1]$, C_i is a permutation code over P_i with minimum Hamming distance at least $4t+1$, then $C = C_1 \circ \cdots \circ C_{2t+1}$ is a permutation code over $[n]$ with minimum Ulam distance greater than or equal to $2t+1$.

Corollary 18: For the code C of Theorem 17 and distinct $\sigma, \pi \in C$, the length of the longest common subsequence of π and σ is less than $n-2t$.

For convenience, we introduce an alternative notation for translocations. Let the mapping $\psi : \mathbb{S}_n \to \mathbb{S}_n$ be defined as follows. For a permutation $\sigma \in \mathbb{S}_n$, an integer ℓ, and $a \in [n]$, let $\psi(a,\ell)\sigma$ denote the permutation obtained from σ by moving the element a exactly $|\ell|$ positions to the right if $\ell \geq 0$ and to the left if $\ell \leq 0$. In other words, for any permutation $\sigma \in \mathbb{S}_n$ and $a \in [n]$,

$$\psi(a,\ell)\sigma = \sigma\phi\big(\sigma^{-1}(a), \sigma^{-1}(a)+\ell\big).$$

For example, we have $\psi(4,3)(3,4,2,5,1) = (3,2,5,1,4) = (3,4,2,5,1)\phi(2,5)$. Note that the mapping ψ is written multiplicatively. Furthermore, with slight abuse of terminology, $\psi(a,\ell)$ may also be called a translocation.

Consider $\sigma, \pi \in \mathbb{S}_n$ with distance $\mathrm{d}_\circ(\sigma,\pi) = m$. A *transformation* from σ to π is a sequence $\psi_1, \psi_2, \ldots, \psi_m$ of translocations such that $\pi = \psi_m \cdots \psi_2 \psi_1 \sigma$.

Let $b_1 < b_2 < \cdots < b_m$ be the elements of $[n]$ that are not in the longest common subsequence of σ and π. Each b_k is called a *displaced* element. The set $\{b_1, \ldots, b_m\}$ is called the set of displaced elements and is denoted by $D(\sigma,\pi)$.

[5]It is clear that instead of using permutation codes for interleaving, one can also use codes with distinct symbols such as those described in [36].

The *canonical* transformation from σ to π is a transformation $\psi_m \cdots \psi_2 \psi_1$ with $\psi_k = \psi(b_k, \ell_k)$ for appropriate choices of $\ell_k, k \in [m]$. In other words, the canonical transformation operates only on displaced elements and corresponds to a shortest sequence of translocations that transform σ into π.

As an example, consider $\sigma = (1,2,3,4,5,6,7,\ldots,15)$ and $\pi = (1,3,4,5,6,2,7,\ldots,15)$. Here, $n=15$, $t=1$, and $s=3$. The canonical transformation is $\psi(2,4)$ and we have

$$\pi = \psi(2,4)\sigma = (1,3,4,5,6,2,7,\ldots,15). \quad (15)$$

In this example, $D(\sigma,\pi) = \{2\}$.

Let $\pi_l = \psi_l \cdots \psi_2 \psi_1 \sigma$ for $1 \leq l \leq m$. An element a is *moved over* an element b in step j if there exists a translocation in the canonical transformation $\psi = \psi(a,\ell)$ such that a is on the left (right) of b in π_{j-1} and on the right (left) side of b in π_j. That is, ψ moves a from one side of b to the other side. In the above example with $\psi(2,4)$, 2 is moved over 4 but it is not moved over 7.

An element $k \in [2t+1]$ is called a $\sigma,\pi-$*pivot*, or simply a pivot, if no element of P_k is displaced, i.e., $P_k \cap D(\sigma,\pi) = \emptyset$. In the example corresponding to (15), the pivots are 1 and 3.

For $I \subseteq [2t+1]$, define P_I as $\cup_{i \in I} P_i$. Also, recall that for $\omega \in \mathbb{S}_n$ and a set P, ω_P denotes the projection of ω onto P. For example, for $t=1$, $n=15$, $\omega = (1,2,3,\ldots,15)$, and $I = \{1,3\}$, we have $\omega_{P_I} = (1,3,4,6,7,9,10,12,13,15)$. We say that ω_{P_I} has a *correct order* if for every $i,j \in I, i<j$, elements of P_i and P_j appear alternatively in $\omega_{P_i \cup P_j}$, starting with an element of P_i. In the example above, ω_{P_I} has a correct order.

Consider $\omega \in \mathbb{S}_n$ and suppose that $\omega_{P_i} = (a_1, a_2, \ldots, a_s)$. The elements of the set $P_i = \{a_1, \ldots, a_s\}$ may be viewed as separating subsequences of ω consisting of elements not in P_i. That is, we may write

$$\omega = r_0 a_1 r_1 a_2 \cdots a_s r_s$$

where the r_l's are nonintersecting subsequences of $[n] \backslash P_i$. For each $l, 0 \leq l \leq s$, the subsequence r_l is called the lth segment of ω with respect to P_i and is denoted by $R_l(\omega, P_i)$. Such a segmentation is shown next for the permutation

$$\omega = (c_3, b_4, a_1, b_2, b_3, a_2, a_3, c_1, a_4, b_1, c_2, c_4).$$

Each segment is marked with a bracket:

$$\omega = (\underbrace{c_3, b_4}, a_1, \underbrace{b_2, b_3}, a_2, \underbrace{\ }, a_3, \underbrace{c_1}, a_4, \underbrace{b_1, c_2, c_4}).$$

To better visualize the subsequences in question, we may replace each element of P_i by $\star$ and write ω as

$$\omega = (c_3, b_4 \star b_2, b_3 \star \star c_1 \star b_1, c_2, c_4).$$

We have, for example, $R_0(\omega, P_i) = (c_3, b_4)$ and $R_2(\omega, P_i) = ()$.

Definition 19: Consider $i,j \in [2t+1]$, $P_i = \{a_1, \cdots, a_s\}$, and $\omega \in \mathbb{S}_n$. Suppose, without loss of generality, that $\omega_{P_i} = (a_1, a_2, \cdots, a_s)$. The sequence $\omega_{(j|i)}$ is defined as follows.

1) If $i=j$, then $\omega_{(j|i)} = \omega_{(i|i)}$ equals ω_{P_i}.
2) If $j > i$, then, for $1 \leq l \leq s$, let $\omega_{(j|i)}(l) = R_l\left(\omega_{P_i \cup P_j}, P_i\right)$ whenever $R_l\left(\omega_{(j|i)}, P_i\right)$ has length

one, and let $\omega_{(j|i)}(l) = \epsilon$ otherwise. Here, ϵ is a special notational symbol.

3) If $j < i$, then, for $1 \leq l \leq s$, let $\omega_{(j|i)}(l) = R_{l-1}\left(\omega_{P_i \cup P_j}, P_i\right)$ whenever $R_{l-1}\left(\omega_{(j|i)}, P_i\right)$ has length one, and let $\omega_{(j|i)}(l) = \epsilon$ otherwise.

As an example, if $P_i = \{a_1, a_2, a_3, a_4, a_5\}$, $P_j = \{b_1, b_2, b_3, b_4, b_5\}$, $j < i$, and $\omega_{P_i \cup P_j} = (b_1, a_1, a_2, b_2, b_3, a_3, b_4, a_4, b_5, a_5)$, the segments of $\omega_{P_i \cup P_j}$ with respect to P_i are (b_1), $()$, (b_2, b_3), (b_4), and (b_5), in the given order, and we have $\omega_{(j|i)} = (b_1, \epsilon, \epsilon, b_4, b_5)$.

Lemma 20: Consider the interleaved code C of Theorem 17 and let $\sigma \in C$. Furthermore, let $\omega \in \mathbb{S}_n$ be such that $\mathsf{d}_\circ(\sigma, \omega) \leq t$. There exists at least one subset $I \subseteq [2t+1]$ of size at least $t+1$ such that ω_{P_I} has a correct order.

Proof: There are at most t displaced elements and, thus, at most t classes containing a displaced element. Hence, there exist at least $2t+1-t = t+1$ classes without any displaced elements and, consequently, at least $t+1$ σ, ω-pivots. Let I be the set consisting of these pivots. It is clear that ω_{P_I} obtained in this way has a correct order which proves the claimed result. □

Lemma 21: For all positive integers s and t and all permutations $\sigma, \omega \in \mathbb{S}_n$, with $n = (2t+1)s$, if i^* is a σ, ω-pivot, then for $j \in [2t+1]$:

$$\mathsf{d}_H\left(\sigma_{(j|i^*)}, \omega_{(j|i^*)}\right) \leq 2\mathsf{d}_\circ(\sigma, \omega).$$

Proof: Assume $\mathsf{d}_\circ(\sigma, \omega) = m$ and let $\psi_m \cdots \psi_2 \psi_1$ be the canonical transformation from σ to ω, so that $\omega = \psi_m \cdots \psi_2 \psi_1 \sigma$. We prove the lemma by induction on m. Clearly, if $m = 0$, then

$$\mathsf{d}_H\left(\sigma_{(j|i^*)}, \omega_{(j|i^*)}\right) = 0.$$

Let $\pi = \psi_{m-1} \cdots \psi_2 \psi_1 \sigma$. As the induction hypothesis, assume that

$$\mathsf{d}_H\left(\sigma_{(j|i^*)}, \pi_{(j|i^*)}\right) \leq 2(m-1).$$

By the triangle inequality, it suffices to show that

$$\mathsf{d}_H\left(\pi_{(j|i^*)}, \omega_{(j|i^*)}\right) \leq 2. \tag{16}$$

Suppose $\psi_m = \psi(b, \ell)$ so that $\omega = \psi_m \pi$. Since i^* is a pivot, we have $b \notin P_{i^*}$. We consider two cases: $b \notin P_j$ and $b \in P_j$. First, suppose $b \notin P_j$. Since $b \notin P_{i^*} \cup P_j$, we have $\pi_{P_{i^*} \cup P_j} = \omega_{P_{i^*} \cup P_j}$ and thus $\mathsf{d}_H\left(\pi_{(j|i^*)}, \omega_{(j|i^*)}\right) = 0$.

On the other hand, suppose $b \in P_j$. Then, b appears in $R_k\left(\pi_{P_{i^*} \cup P_j}, P_{i^*}\right)$ of $\pi_{P_{i^*} \cup P_j}$ and in $R_l\left(\omega_{P_{i^*} \cup P_j}, P_{i^*}\right)$ of $\omega_{P_{i^*} \cup P_j}$, for some k, l. The only segments affected by the translocation ψ_m are $R_k\left(\pi_{P_{i^*} \cup P_j}, P_{i^*}\right)$ and $R_l\left(\omega_{P_{i^*} \cup P_j}, P_{i^*}\right)$, and thus, for $p \in [2t+1] \backslash \{l, k\}$, we have $R_p\left(\pi_{P_{i^*} \cup P_j}, P_{i^*}\right) = R_p\left(\omega_{P_{i^*} \cup P_j}, P_{i^*}\right)$. Hence, for $p \in [2t+1] \backslash \{l, k\}$, we find $\pi_{(j|i^*)}(p) = \omega_{(j|i^*)}(p)$, implying that $\mathsf{d}_H\left(\pi_{(j|i^*)}, \omega_{(j|i^*)}\right) \leq 2$. □

Lemma 22: Consider the interleaved code C of Theorem 17. Let $\sigma \in C$ and $\omega \in \mathbb{S}_n$ such that $\mathsf{d}_\circ(\sigma, \omega) \leq t$. If $I \subseteq [2t+1]$ is of size at least $t+1$ and ω_{P_I} has a correct order, then

1) for each $i \in I$, $\mathsf{d}_H\left(\sigma_{P_i}, \omega_{P_i}\right) \leq 2t$ and,
2) for $i \in I$ and $j \notin I$, $\mathsf{d}_H\left(\sigma_{P_j}, \omega_{(j|i)}\right) \leq 2t$.

Proof: Since there are at most t classes containing displaced elements and I has size at least $t+1$, there exists a pivot $i^* \in I$. Then, by Lemma 21

$$\mathsf{d}_H\left(\sigma_{(i|i^*)}, \omega_{(i|i^*)}\right) \leq 2t.$$

Since σ is a codeword in C, by construction, we have $\sigma_{(i|i^*)} = \sigma_{P_i}$. Furthermore, since ω_{P_I} has a correct order, we have $\omega_{(i|i^*)} = \omega_{P_i}$. Hence

$$\mathsf{d}_H\left(\sigma_{P_i}, \omega_{P_i}\right) \leq 2t.$$

To prove the second part, we proceed as follows. Assume $\psi_m \cdots \psi_2 \psi_1$, with $m = \mathsf{d}_\circ(\sigma, \omega)$, is the canonical transformation from σ to ω so that $\omega = \psi_m \cdots \psi_2 \psi_1 \sigma$.

We first show that $\psi_m \cdots \psi_2 \psi_1$ may be decomposed into four parts

$$\omega = \left(\psi^{(j)}_{t^{(j)}} \cdots \psi^{(j)}_1 \left(\psi^{(i)}_{t^{(i)}} \cdots \psi^{(i)}_1 \left(\tau_{t_\tau} \cdots \tau_1 \left(\psi'_{t'} \cdots \psi'_1 \sigma\right)\right)\right)\right)$$

with $t' + t^{(i)} + t^{(j)} = m$ such that

$$\begin{aligned}
\psi'_k &= \psi\left(a'_k, \ell'_k\right), \quad a'_k \notin P_i \cup P_j, k \in [t'] \\
\tau_k &= \tau\left(a_k, b_k\right), \quad a_k, b_k \in P_i, k \in [t_\tau] \\
\psi^{(i)}_k &= \psi\left(a^{(i)}_k, \ell^{(i)}_k\right), \quad a^{(i)}_k \in P_i, k \in \left[t^{(i)}\right] \\
\psi^{(j)}_k &= \psi\left(a^{(j)}_k, \ell^{(j)}_k\right), \quad a^{(j)}_k \in P_j, k \in \left[t^{(j)}\right]
\end{aligned} \tag{17}$$

and such that no $\psi^{(i)}_k$ moves $a^{(i)}_k$ over an element of P_{i^*}.

It can be easily verified that any two translocations $\psi(a, \ell_1)$ and $\psi(b, \ell_2)$ "commute." That is, for any permutation π, we can find translocations $\psi(a, \ell_3)$ and $\psi(b, \ell_4)$ such that $\psi(a, \ell_1)\psi(b, \ell_2)\pi = \psi(b, \ell_4)\psi(a, \ell_3)\pi$. Thus, we have the decomposition

$$\omega = \left(\psi^{(j)}_{t^{(j)}} \cdots \psi^{(j)}_1 \left(\psi''_{t^{(i)}} \cdots \psi''_1 \left(\psi'_{t'} \cdots \psi'_1 \sigma\right)\right)\right)$$

with $t' + t^{(i)} + t^{(j)} = m$ such that

$$\begin{aligned}
\psi'_k &= \psi\left(a'_k, \ell'_k\right), \quad a'_k \notin P_i \cup P_j, k \in [t'] \\
\psi''_k &= \psi(a''_k, \ell''_k), \quad a''_k \in P_i, k \in \left[t^{(i)}\right] \\
\psi^{(j)}_k &= \psi\left(a^{(j)}_k, \ell^{(j)}_k\right), \quad a^{(j)}_k \in P_j, k \in \left[t^{(j)}\right].
\end{aligned}$$

Furthermore, it is easy to see that we may write $\psi''_{t^{(i)}} \cdots \psi''_1$ as $\psi^{(i)}_{t^{(i)}} \cdots \psi^{(i)}_1 \tau_{t_\tau} \cdots \tau_1$ with

$$\tau_k = \tau\left(a_k, b_k\right), \quad a_k, b_k \in P_i, k \in [t_\tau]$$

such that no $\psi^{(i)}_k$ moves $a^{(i)}_k$ over an element of P_{i^*}. Hence, for any permutation ω, one can write a decomposition of the form (17).

Let $\omega' = \tau_{t_\tau} \cdots \tau_1 \psi'_{t'} \cdots \psi'_1 \sigma$, and $\omega^{(i)} = \psi^{(i)}_{t^{(i)}} \cdots \psi^{(i)}_1 \omega'$, so that $\omega = \psi^{(j)}_{t^{(j)}} \cdots \psi^{(j)}_1 \omega^{(i)}$. By the triangle inequality

$$\begin{aligned}
\mathsf{d}_H\left(\sigma_{P_j}, \omega_{(j|i)}\right) \leq\ & \mathsf{d}_H\left(\sigma_{P_j}, \omega'_{(j|i)}\right) \\
& + \mathsf{d}_H\left(\omega'_{(j|i)}, \omega^{(i)}_{(j|i)}\right) \\
& + \mathsf{d}_H\left(\omega^{(i)}_{(j|i)}, \omega_{(j|i)}\right).
\end{aligned}$$

It is clear that $\mathsf{d}_H\left(\sigma_{P_j}, \omega'_{(j|i)}\right) = 0$.

Next, consider $\omega'_{(j|i)}$ and its transform $\omega^{(i)}_{(j|i)}$ induced by the translocations $\psi^{(i)}_k, k \in \left[t^{(i)}\right]$. Note that $\omega'_{P_{\{j,i,i^*\}}}$ has a correct order. Since no translocation $\psi^{(i)}_k$ moves $a^{(i)}_k$ over an element of P_{i^*}, each $\psi^{(i)}_k$ moves $a^{(i)}_k$ over at most one element of P_j. Thus, each $\psi^{(i)}_k$ can modify at most two segments and we have $\mathsf{d}_H\left(\omega'_{(j|i)}, \omega^{(i)}_{(j|i)}\right) \leq 2t^{(i)}$. Furthermore, each $\psi^{(j)}_k$ modifies at most two segments and thus $\mathsf{d}_H\left(\psi^{(i)}_{(j|i)}, w_{(j|i)}\right) \leq 2t^{(j)}$. Hence

$$\mathsf{d}_H\left(\sigma_{P_j}, w_{(j|i)}\right) \leq 0 + 2t^{(i)} + 2t^{(j)} \leq 2m.$$

□

Proof. (Theorem 17): Suppose the minimum Ulam distance of C is less than $2t+1$. Then, for two distinct codewords $\pi, \sigma \in C$, there exists an $\omega \in \mathbb{S}_n$ such that $\mathsf{d}_\circ(\pi, \omega) \leq t$ and $\mathsf{d}_\circ(\sigma, \omega) \leq t$.

Since $\sigma \neq \pi$, there exists $k \in [2t+1]$ such that $\pi_{P_k} \neq \sigma_{P_k}$, which implies that $\mathsf{d}_H(\pi_{P_k}, \sigma_{P_k}) \geq 4t+1$. Since $\mathsf{d}_\circ(\sigma, \omega) \leq t$, by Lemma 20, there exists $I \subseteq [2t+1]$ of size at least $t+1$ such that ω_{P_I} has a correct order.

If $k \in I$, by Lemma 22, $\mathsf{d}_H(\sigma_{P_k}, \omega_{P_k}) \leq 2t$ and $\mathsf{d}_H(\pi_{P_k}, \omega_{P_k}) \leq 2t$, which together imply $\mathsf{d}_H(\sigma_{P_k}, \pi_{P_k}) \leq 4t$.

On the other hand, if $k \notin I$, by Lemma 22, for any $i \in I$, $\mathsf{d}_H\left(\sigma_{P_k}, w_{(k|i)}\right) \leq 2t$ and $\mathsf{d}_H\left(\pi_{P_k}, w_{(k|i)}\right) \leq 2t$, which again imply $\mathsf{d}_H(\sigma_{P_k}, \pi_{P_k}) \leq 4t$.

Hence, by contradiction, the minimum distance of C is at least $2t+1$. □

The rate of the aforementioned translocation correcting codes based on interleaving may be estimated as follows. The cardinality of the interleaved code of length n and minimum distance $d = d(n)$ is at least $\left(A_H\left(\left\lfloor \frac{n}{d} \right\rfloor, 2d-1\right)\right)^d$ for odd d, and $\left(A_H\left(\left\lfloor \frac{n}{d+1} \right\rfloor, 2d+1\right)\right)^{d+1}$ for even d. The construction is applicable only if $d(n) \leq \sqrt{n/2} - 1$, in which case the asymptotic rate of the interleaved code equals

$$\begin{aligned}
\lim \frac{\ln |C|}{\ln n!} &= \lim \frac{d(n) \ln A_H\left(\frac{n}{d(n)}, 2d(n)\right)}{\ln n!} \\
&= \lim \frac{\ln A_H\left(\frac{n}{d(n)}, 2d(n)\right)}{\ln (n/d(n))!} \frac{d(n) \ln (n/d(n))!}{\ln n!} \\
&= \left(1 - 2 \lim \frac{d^2(n)}{n}\right) \lim \frac{n \ln n - n \ln d(n) + O(n)}{n \ln n + O(n)}
\end{aligned}$$

where we used Theorem 11 to obtain the last equality. For example, if $d(n) = n^\beta$, $\beta < 1/2$, then

$$1 - 2 \lim \frac{d^2(n)}{n} = 1$$

and one obtains a translocation error-correcting code of rate

$$\lim \frac{\ln |C|}{\ln n!} = \lim \frac{n \ln n - \beta n \ln n + O(n)}{n \ln n + O(n)} = 1 - \beta.$$

In the next section, we describe a modification of the interleaving procedure, which, when applied recursively, improves upon the code rate $1 - \beta$.

B. Interleaving Codes in the Hamming Metric and the Ulam Metric

The interleaving approach described in Section V-A may be extended in a straightforward manner. Rather than interleaving permutation codes with good Hamming distance, as in Section V-A, one may construct a code in the Ulam metric by interleaving a code with good Ulam distance and a code with good Hamming distance. Furthermore, this approach may be implemented in a recursive manner. In what follows, we explain one such approach and show how it leads to improved code rates as compared to simple interleaving.

We find the following results useful for our recursive construction method.

Lemma 23: Let $\sigma, \pi \in \mathbb{S}_n$ be two permutations, such that $\mathsf{d}_\circ(\sigma, \pi) = 1$. Then, there exist at most three positions i, $i \in [n-1]$, such that for some $j = j(i) \in [n-1]$:

1) $\sigma(i) = \pi(j)$;
2) $\sigma(i+1) \neq \pi(j+1)$.

Proof: Suppose $\pi = \sigma\phi(i_1, i_2)$. The proof follows from the simple fact that when applying a translocation $\phi(i_1, i_2)$ to σ, the positions i described above are among

$$\begin{cases} i_1 - 1, i_1, \text{ and } i_2, & \text{if } i_1 < i_2 \\ i_1 - 1, i_1, \text{ and } i_2 - 1, & \text{if } i_1 > i_2. \end{cases}$$

□

Corollary 24: Let $\sigma, \pi \in \mathbb{S}_n$ be two permutations, and assume that there exist $a \geq 0$ different positions i, $i \in [n-1]$, such that $\sigma(i) = \pi(j)$, but $\sigma(i+1) \neq \pi(j+1)$ for some $j \in [n-1]$. Then, $\mathsf{d}_\circ(\sigma, \pi) \geq \lceil a/3 \rceil$.

For an integer $p \geq 1$, let $\mu = (1, 2, \ldots, p)$ and let $\sigma_1, \sigma_2 \in \mathbb{S}(\{p+1, \ldots, 2p-1\})$. Note that

$$\begin{aligned}
\mu \circ \sigma_1 &= (1, \sigma_1(1), 2, \sigma_1(2), \ldots, p-1, \sigma_1(p-1), p) \\
\mu \circ \sigma_2 &= (1, \sigma_2(1), 2, \sigma_2(2), \ldots, p-1, \sigma_2(p-1), p). \quad (18)
\end{aligned}$$

Theorem 25: For μ, σ_1, and σ_2 described above, if $\mathsf{d}_H(\sigma_1, \sigma_2) \geq d$, then

$$\mathsf{d}_\circ(\mu \circ \sigma_1, \mu \circ \sigma_2) \geq \lceil 2d/3 \rceil .$$

Proof: Let $\pi_1 = \mu \circ \sigma_1$ and $\pi_2 = \mu \circ \sigma_2$. We show that the number of indices ℓ in π_1, with respect to π_2, that satisfy the conditions described in Lemma 23 is at least $2d$. Then, the claim of the theorem follows when we apply Corollary 24 with $a = 2d$.

Assume that $\sigma_1(\ell) \neq \sigma_2(\ell)$ for some $\ell \in [p-1]$. For each such ℓ, the two indices $2\ell - 1$ and 2ℓ can both serve as index i in Lemma 23:

1) We have $\pi_1(2\ell - 1) = \pi_2(2\ell - 1) = \ell$, yet
$$\sigma_1(\ell) = \pi_1(2\ell) \neq \pi_2(2\ell) = \sigma_2(\ell).$$
2) Let $j \in [p]$ be such that $\pi_1(2\ell) = \pi_2(2j)$. It is easy to see that $j \neq \ell$. Then
$$\ell + 1 = \pi_1(2\ell + 1) \neq \pi_2(2j + 1) = j + 1 .$$

□

Let $\mu \circ C = \{\mu \circ \sigma : \sigma \in C\}$. From Theorem 25, we have the following corollary.

Corollary 26: For integers n and p with $n = 2p-1$, let $\mu = (1, 2, \ldots, p)$ and suppose $C \subseteq \mathbb{S}(\{p+1, \ldots, n\})$ is a code with minimum Hamming distance at least $\frac{3d}{2}$. Then, $\mu \circ C$ is a code in $\mathbb{S}_n$ with minimum Ulam distance at least d and with size $|C|$.

Hence, for odd n, we can construct a translocation code with length n, minimum distance at least d, and size $A_H\left(\frac{n-1}{2}, \left\lceil \frac{3d}{2} \right\rceil\right)$. This can be easily generalized for all n to get codes of size

$$A_H\left(\left\lceil \frac{n}{2} \right\rceil - 1, \left\lceil \frac{3d}{2} \right\rceil\right).$$

By assuming that the permutation code in the Hamming metric is capacity achieving, the asymptotic rate of the constructed code becomes

$$\begin{aligned}
&\lim \frac{\ln A_H\left(\left\lceil \frac{n}{2} \right\rceil - 1, \left\lceil \frac{3d(n)}{2} \right\rceil\right)}{\ln n!} \\
&= \lim \frac{\ln A_H\left(\left\lceil \frac{n}{2} \right\rceil - 1, \left\lceil \frac{3d(n)}{2} \right\rceil\right)}{\ln \left\lceil \frac{n}{2} \right\rceil!} \cdot \frac{\ln \left\lceil \frac{n}{2} \right\rceil!}{\ln n!} \\
&= \frac{1}{2} - \frac{3}{2} \lim \frac{d(n)}{n} = \frac{1}{2} - \frac{3}{2}\,\delta \qquad (19)
\end{aligned}$$

where $\delta = \lim \frac{d(n)}{n}$. Therefore, this code construction incurs a rate loss of $(1+\delta)/2$ when compared to the capacity, which in this case equals $1-\delta$.

The final result that we prove in order to describe a recursive interleaving procedure is related to the longest common subsequence of two sequences and the minimum Ulam distance of interleaved sequences.

Lemma 27: For $\sigma, \pi \in \mathbb{S}_n$ and $P \subseteq [n]$, we have

$$\mathsf{d}_\circ(\sigma, \pi) \geq \mathsf{d}_\circ(\sigma_P, \pi_P) + \mathsf{d}_\circ(\sigma_Q, \pi_Q)$$

where $Q = [n] \backslash P$.

Proof: Without loss of generality, assume that σ is the identity permutation. It is clear that $l(\pi) \leq l(\pi_P) + l(\pi_Q)$. Hence

$$\begin{aligned}
\mathsf{d}_\circ(\sigma, \pi) &= n - l(\pi) \\
&\geq n - l(\pi_P) - l(\pi_Q) \\
&= |P| - l(\pi_P) + |Q| - l(\pi_Q) \\
&= \mathsf{d}_\circ(\sigma_P, \pi_P) + \mathsf{d}_\circ(\sigma_Q, \pi_Q).
\end{aligned}$$

□

Lemma 28: For sets P and Q of sizes p and $p-1$, respectively, let $C_1' \subseteq \mathbb{S}(P)$ be a code with minimum Ulam distance d and let $C_1 \subset \mathbb{S}(Q)$ be a code with minimum Hamming distance $3d/2$. The code $C_1' \circ C_1 = \{\sigma \circ \pi : \sigma \in C_1', \pi \in C_1\}$ has minimum Ulam distance d.

Proof: For $\sigma_1, \sigma_2 \in C_1'$ and $\pi_1, \pi_2 \in C_1$ with $(\sigma_1, \pi_1) \neq (\sigma_2, \pi_2)$, we show that $\mathsf{d}_\circ(\sigma_1 \circ \pi_1, \sigma_2 \circ \pi_2) \geq d$.

The case $\sigma_1 = \sigma_2$ follows from a simple use of Theorem 25.

Assume next that $\sigma_1 \neq \sigma_2$. Then, by Lemma 27, $\mathsf{d}_\circ(\sigma_1 \circ \pi_1, \sigma_2 \circ \pi_2) \geq \mathsf{d}_\circ(\sigma_1, \sigma_2) \geq d$ and this completes the proof. □

Let $\alpha = \frac{3}{2}$. For a given n, set $P = \left\{1, \ldots, \left\lceil \frac{n}{2} \right\rceil\right\}$ and set $Q = \left\{\left\lceil \frac{n}{2} \right\rceil + 1, \ldots, 2\left\lceil \frac{n}{2} \right\rceil - 1\right\}$. Suppose $C_1' \subseteq \mathbb{S}(P)$ is a code with minimum Ulam distance d and $C_1 \subseteq \mathbb{S}(Q)$ is a code with minimum Hamming distance $\alpha\, d$. Assuming that permutation codes with this given minimum Hamming distance exist, we only need to provide a construction for C_1'. An obvious choice for C_1' is a code with only one codeword. Then, $C = C_1' \circ C_1$ is a code with minimum Ulam distance d and cardinality

$$A_H\left(\left\lceil \frac{n}{2} \right\rceil - 1, \alpha\delta\right).$$

The gap to capacity may be significantly reduced by observing that C_1' does not have to be a code of cardinality one, and that C_1' may be constructed from shorter codes.

To this end, let $C_1' = C_2' \circ C_2$ where C_2' is a code of length $\left\lceil \frac{n}{4} \right\rceil$ with minimum Ulam distance d, while C_2 is a code of length $\left\lceil \frac{n}{4} \right\rceil - 1$ and minimum Hamming distance αd.

By repeating the same procedure k times, we obtain a code of the form

$$(((C_k' \circ C_k) \circ C_{k-1}) \circ \cdots) \circ C_1 \qquad (20)$$

where each C_i, $i \leq k$, is a code with minimum Hamming distance αd and length $\left\lceil \frac{n}{2^i} \right\rceil - 1$, and C_k' is a code with minimum Ulam distance d and length $\left\lceil \frac{n}{2^i} \right\rceil$. Since each C_i is a permutation code in the Hamming metric with minimum distance αd, we must have $\left\lceil \frac{n}{2^i} \right\rceil - 1 \geq \alpha d$. To ensure that this condition is satisfied, in (20), we let k be the largest value of i satisfying $\frac{n}{2^i} - 1 \geq \alpha d$. It is easy to see that $k = \left\lfloor \log \frac{n}{\alpha d + 1} \right\rfloor$. Furthermore, we choose C_k' to consist of a single codeword.

The asymptotic rate of the recursively constructed codes equals

$$\begin{aligned}
&\lim \frac{1}{\ln n!} \sum_{i=1}^{k} \ln A_H\left(\left\lceil \frac{n}{2^i} \right\rceil - 1, \alpha d(n)\right) \\
&= \lim \sum_{i=1}^{k} \frac{\ln A_H\left(\left\lceil \frac{n}{2^i} \right\rceil - 1, \alpha d(n)\right)}{\ln\left(\left\lceil \frac{n}{2^i} \right\rceil - 1\right)!} \frac{\ln\left(\left\lceil \frac{n}{2^i} \right\rceil - 1\right)!}{\ln n!} \\
&= \lim \sum_{i=1}^{k} \left(1 - \frac{\alpha d(n) 2^i}{n}\right) 2^{-i} \\
&= \lim \left(1 - 2^{-k} - \frac{\alpha d(n) k}{n}\right) \\
&= 1 - 2^{-\left\lfloor \log \frac{1}{\alpha\delta} \right\rfloor} - \alpha\delta \left\lfloor \log \frac{1}{\alpha\delta} \right\rfloor
\end{aligned}$$

where the last step follows from $\lim k = \left\lfloor \log \frac{1}{\alpha\delta} \right\rfloor$. Note that this rate is roughly equal to $1 - \alpha\delta\left(1 + \log \frac{1}{\alpha\delta}\right)$.

C. Permutation Codes in the Hamming Metric

In Section V-B, we demonstrated a number of constructions for translocation error-correcting codes based on permutation codes in the Hamming metric and codes over distinct symbols. There exist a number of constructions for sets of permutations with good Hamming distance, and codes with codewords containing distinct symbols. For example, in [20] and [37]–[39], constructions of permutations in $\mathbb{S}_n$ using classical binary codes

were presented, while other constructions rely on direct combinatorial arguments [40], [41]. An example of code construction for codewords over distinct symbols was presented in [36]. There, specialized subcodes of Reed–Solomon codes were identified such that their codewords consist of distinct symbols.

In the former case, if C is a binary $[n, \lambda n, \beta n]$ code, the construction applied to C yields a subset of $\mathbb{S}_n$ of cardinality $2^{\lambda n}$, with minimum Hamming distance βn. This construction and constructions related to it may be used for permutation code design, resulting in sets of permutations in $\mathbb{S}_n$ of cardinality $\exp\{\Theta(n)\}$ and minimum Hamming distance $\Theta(n)$. These permutations may consequently be used to construct permutation codes in $\mathbb{S}_{2n}$ with $\exp\{\Theta(n)\}$ codewords and minimum Ulam distance $\Theta(n)$.

We describe a simple method for constructing sets of vectors of length $m > 0$ over $[n]$ such that all entries of the vector are different, and such that the minimum Hamming distance between the vectors is large. In other words, we propose a novel construction for partial permutation codes under the Hamming metric, suitable for use in the recursive code construction described in Section V-B.

The idea behind the proof is based on mapping suitably modified binary codewords in the Hamming space into partial permutations. For this purpose, let C be a binary $[N, K, D]$ code, and for simplicity of exposition, assume that n is a power of two. Let $\boldsymbol{c} \in C$. We construct a vector $\boldsymbol{x} = \chi(\boldsymbol{c}) \in ([n])^m$, where χ is a mapping as follows.

1) Divide $\boldsymbol{c}$ into m binary blocks $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_m$ of lengths $\log_2 n - \log_2 m$ each. Again, for simplicity, we assume that m is a power of two.
2) For each block $\boldsymbol{c}_i$, $i \in [m]$, construct a vector $\boldsymbol{x}_i$ of length $\log_2 n$ according to the following rule: The first $\log_2 n - \log_2 m$ bits in $\boldsymbol{x}_i$ equal $\boldsymbol{c}_i$, while the last $\log_2 m$ bits in $\boldsymbol{x}_i$ represent the binary encoding of the index i. Note that the integer values represented by the binary vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ are all different.
3) Form an integer valued vector $\boldsymbol{x} = \chi(\boldsymbol{c})$ of length m over $[n]$, such that its ith entry has the binary encoding specified by $\boldsymbol{x}_i$. Observe that all the integer entries of such a vector are different.

Now, take two vectors $\boldsymbol{a}, \boldsymbol{c} \in C$, such that their Hamming distance satisfies $\mathsf{d}_H(\boldsymbol{a}, \boldsymbol{c}) \geq D$. Let $\boldsymbol{x} = \chi(\boldsymbol{a})$ and $\boldsymbol{y} = \chi(\boldsymbol{c})$ be the corresponding vectors of length m over $[n]$ constructed as described before. Then, there exist at least $D/(\log_2 n - \log_2 m)$ blocks of length $\log_2 n$ that are pairwise different. Therefore, the corresponding $D/(\log_2 n - \log_2 m)$ entries in $\boldsymbol{x}$ and $\boldsymbol{y}$ are pairwise different as well.

Consider the set of vectors

$$\mathcal{S}' = \{\chi(\boldsymbol{c}) \ : \ \boldsymbol{c} \in C\}.$$

It is straightforward to see that the set $\mathcal{S}'$ has the following properties.

1) For any $\boldsymbol{x} \in \mathcal{S}'$, all entries in $\boldsymbol{x}$ are different.
2) For any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}'$, $\boldsymbol{x} \neq \boldsymbol{y}$, the Hamming distance satisfies $\mathsf{d}_H(\boldsymbol{x}, \boldsymbol{y}) \geq D/(\log_2 n - \log_2 m)$.

The set $\mathcal{S}'$ can be used similarly as the set $\mathbb{S}_n$ in the basic construction to obtain codes over $\mathbb{S}_{n+m}$ with minimum Ulam distance at least

$$\Theta\left(\frac{D}{\log_2 n - \log_2 m}\right).$$

Note that in this case, only m numbers in the range $\{n+1, n+2, \ldots, n+m\}$ are inserted between the numbers in $[n]$, while the Hamming distance of the vectors is preserved.

Lemma 29: The parameters N, n, and m are connected by the following equation:

$$N = m \log_2 n - m \log_2 m = m \log_2 \frac{n}{m}.$$

From this lemma, if we take $m = \frac{1}{2}n$, then $N = \frac{1}{2}n$. By taking a code C with parameters $[\frac{1}{2}n, \lambda n, \beta n]$, where $\lambda > 0$ and $\beta > 0$ are constants, we obtain a set $\mathcal{S}'$ of size $2^{\lambda n}$ and Hamming distance $\Theta(n)$. The corresponding translocation code is able to correct $\Theta(n)$ translocation errors, and it has $2^{\lambda n}$ codewords.

D. Decoding of Interleaved Codes

An efficient decoder implementation for the general family of interleaved codes is currently not known. For the case of recursive codes, decoding may be accomplished with low complexity provided that the Hamming distance of the component permutation codes is increased from $\frac{3d}{2}$ to $2d$.

For simplicity of exposition, we assume $n = 2p - 1$ where p is an integer. The case of even n may be handled in the same manner, provided that one fixes the last symbol of all codewords.

Let $\sigma = (1, \hat{\sigma}(1), 2, \hat{\sigma}(2), \ldots, \hat{\sigma}(p-1), p) \in C$ be the stored codeword and let $\pi \in \mathbb{S}_n$ be the retrieved word.

For $i \in [p-1]$, denote by s_i^{π} the substring of π that starts with element i and ends with element $i+1$. If $i+1$ appears before i in π, then s_i^{π} is considered empty. For $i \in [p-1]$, let $\hat{\pi}(i) = u$ if s_i^{π} contains some unique element u of $\{p+1, \ldots, n\}$. Otherwise, let $\hat{\pi}(i) = \epsilon$.

Lemma 30: The permutation $\hat{\pi}$ differs from $\hat{\sigma}$ in at most $2\mathsf{d}_{\circ}(\sigma, \pi)$ positions.

Proof: Let $t = \mathsf{d}_{\circ}(\sigma, \pi)$. There exists a sequence $\phi_1, \phi_2, \ldots, \phi_t$ of translocations such that $\pi = \sigma\phi_1\phi_2\cdots\phi_t$. For $i \in \{0, \ldots, t\}$, let $\pi_i = \sigma\phi_1\phi_2\cdots\phi_i$ and let L_i be given as

$$L_i = \{j | \exists k \leq i : \hat{\pi}_k(j) \neq \hat{\sigma}_k(j)\}.$$

The set L_i may be viewed as the set of elements displaced by one of the translocations $\phi_1, \phi_2, \ldots, \phi_i$. Note that, for each i, $L_i \subseteq L_{i+1}$.

To prove the lemma, it suffices to show that $|L_t| \leq 2t$, since $\{j | \hat{\pi}(j) \neq \hat{\sigma}(j)\} \subseteq L_t$.

Let $L_0 = \emptyset$. We show that $|L_i| \leq |L_{i-1}| + 2$ for $i \in [t]$.

The translocation ϕ_i either moves an element of $[p]$ or an element of $\{p+1, \ldots, n\}$. First, suppose that it moves an element j of $[p]$. Then, ϕ_i can affect only the substrings $s_{j-1}^{\pi_{i-1}}$ and $s_j^{\pi_{i-1}}$ of π_{i-1}. Next, assume that ϕ_i moves an element of $\{p+1, \ldots, n\}$. It can then be verified that at most two substrings of π_{i-1} may be affected by the given translocation. Hence, $|L_i| \leq |L_{i-1}| + 2$. □

Assume now that $C \subseteq \mathbb{S}_n$ is an interleaved code of the form

$$C = C_1' \circ C_1$$

where $C_1' = \{(1, 2, \ldots, p)\}$, and where C_1 is a permutation code over the set $\{p+1, \ldots, n\}$ with minimum Hamming distance $4t+1$.

Let $\sigma \in C$ be the stored code word and $\pi \in \mathbb{S}_n$ be the retrieved word. Assume that $\mathrm{d}_\circ(\sigma, \pi) \leq t$. The first step of the decoding algorithm is to extract $\hat{\pi}$ from the permutation π. By Lemma 30, we have $\mathrm{d}_H(\hat{\sigma}, \hat{\pi}) \leq 2t$. Since C_1 has minimum Hamming distance $4t+1$, $\hat{\sigma}$ can be uniquely recovered from $\hat{\pi}$.

Hence, for odd d, if C_1 has minimum Hamming distance $2d-1$, then C has minimum Ulam distance at least d and can be decoded using the described decoding scheme. The aforementioned decoding method may also be used on a recursive construction of depth larger than one by first decoding the innermost components.

Note that decoding is accomplished through Hamming distance decoding of permutation codes, for which a number of interesting algorithms are known in literature [26], [27], [42].

Similar to (19), the asymptotic rate of the code C can be found to be $\frac{1}{2} - 2\delta$, where $\delta = \lim \frac{d}{n} = \lim \frac{2t+1}{n}$. For the recursive construction described in (20), the asymptotic rate of the efficiently decodable codes outlined above equals $1 - 2^{-\left\lfloor \log \frac{1}{\alpha\delta} \right\rfloor} - \alpha\delta \left\lfloor \log \frac{1}{\alpha\delta} \right\rfloor$, with $\alpha = 2$.

Remark: Permutation codes in $\mathbb{S}_n$, correcting adjacent transposition errors, were thoroughly studied in [23]. We note that these codes can also be used to correct translocation errors. Indeed, every translocation can be viewed as a sequence of at most $n-1$ adjacent transpositions. Therefore, any code in $\mathbb{S}_n$ that corrects $f(n)$ adjacent transpositions [for some function $f(n)$] can also correct $O(f(n)/n)$ translocations.

It was shown in [23, Th. 3.1] that the upper bound on the rate of the code correcting $O(n^2)$ adjacent transpositions is zero. Such a code can also be used to correct $O(n)$ translocation errors. In comparison, the interleaved constructions described above can also correct $O(n)$ translocation errors, yet their rate is strictly larger than zero.

The nonasymptotic and asymptotic rates of the discussed code families are compared in Figs. 4 and 5.

Note that the gap from capacity of the constructions presented in this paper is still fairly large, despite the fact that the codes are asymptotically good. This result may be attributed to the fact that the interleaving construction restricts the locations of subsets of elements in a severe manner. Alternative interleaving methods will be discussed in a companion paper.

In what follows, we describe a method of Beame *et al.* [30] that provides translocation codes with minimum distance proportional to $n - o(n)$. This covers the zero-capacity domain of our analysis.

E. Zero-Rate Codes

We present two constructions based on the longest common subsequence analysis. The first construction is based on Hadamard matrices and was given in [30], while the second construction is probabilistic.

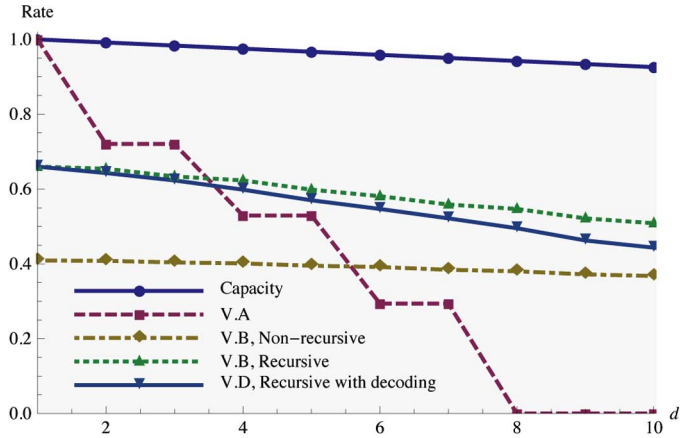

Fig. 4. Rate versus distance for several code constructions with $n = 150$. The numbers in the legend refer to the section where the corresponding code is described. It is assumed that $A_H(n, d) = \frac{n!}{(d-1)!}$.

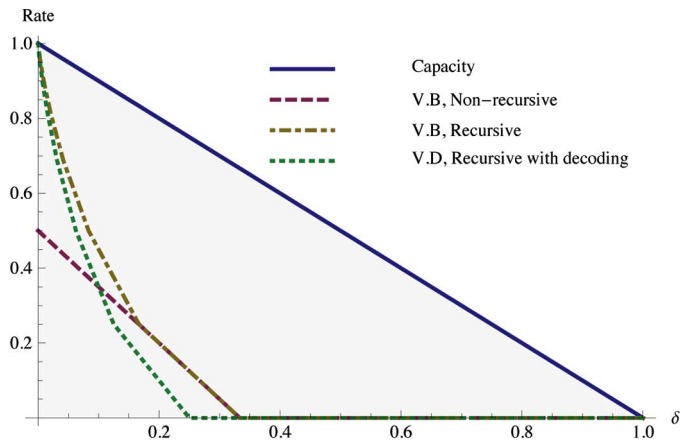

Fig. 5. Asymptotic rate versus distance for several code constructions. The horizontal axis is $\delta = \lim \frac{d(n)}{n}$.

Assume that a Hadamard matrix of order k exists. To explain the construction, we consider permutations over the set $\{0, 1, \ldots, n-1\}$. Furthermore, the positions in each permutation are also numbered from 0 to $n-1$.

Let $s = \lceil n^{1/(k-1)} \rceil$. For $a \in \{0, 1, \ldots, n-1\}$, we denote the representation of a in base s by $\overline{a_1 a_2 \cdots a_{k-1}}$, where a_1 is the most significant digit.

Let H be a Hadamard matrix of order k with rows and columns indexed by elements in the set $\{0, 1, \ldots, k-1\}$. Without loss of generality, assume the first row and column of H are all-ones vectors. The set $\{\pi_i\}_{i=1}^{k}$ of permutations is constructed by defining the mth element of π_i, for $m = 0, 1, \ldots, s^{k-1} - 1$, as follows. Let $m = \overline{m_1 \cdots m_{k-1}}$, and let the mth element of π_i equal

$$\pi_i(m) = \overline{a_1 a_2 \cdots a_{k-1}}$$

where, for $j \in \{0, 1, \ldots, k-1\}$,

$$a_j = \begin{cases} m_j, & \text{if } H_{ij} = 1 \\ s - 1 - m_j, & \text{if } H_{ij} = -1. \end{cases}$$

The length of the longest common subsequence of any two permutations of $\{\pi_i\}$ is at most $s^{k/2-1}$. The permutations obtained

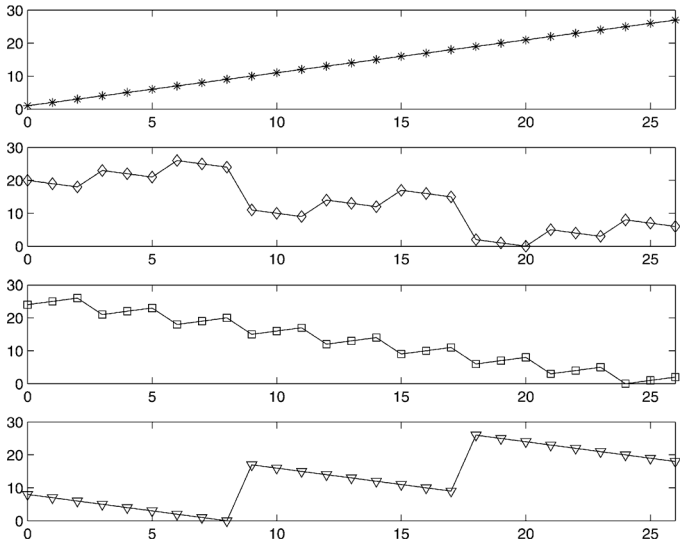

Fig. 6. Permutation codewords based on Hadamard matrices [30].

in this way have length s^{k-1}. Consequently, the minimum distance of the code is at least $s^{k-1}-s^{k/2-1}$. Note that if $s^{k-1} > n$, we can arbitrarily delete elements from each permutation to obtain a set of permutations each of length n.

As an example, consider $n = 27$ and $k = 4$. We have

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

and $s = 3$. Four codewords of the code based on this Hadamard matrix are plotted in Fig. 6.

Another construction based on [30] holds for $3 \leq k \leq \sqrt{n}$, leading to k permutations with minimum Ulam distance at least $n - 32\,(kn)^{1/3}$. The number of codewords obtained from this construction is exponentially smaller than what may be obtained via random methods, as we demonstrate next.

Let U_n denote the Ulam distance between a randomly chosen permutation of length n and the identity, $e = (1, 2, \ldots, n)$. From a result shown by Kim [43] (see also [44]–[46]), for $0 < \theta \leq n^{1/3}/20$, one has

$$P\left(U_n \leq n - 2\sqrt{n} - \theta n^{1/6}\right) \leq \exp\left(-\theta^{3/2}\left(\frac{4}{3} - \frac{\theta}{27n^{1/3}} - \frac{5\log n}{\theta^{1/2}n^{1/3}}\right)\right).$$

By letting $\theta = an^{1/3}$ with $a \leq 1/20$, for sufficiently large n, we find

$$P\left(U_n \leq n - (2+a)\sqrt{n}\right) \leq \exp\left(-a^{3/2}\sqrt{n}\right).$$

Suppose a code C is constructed by randomly choosing $M = e^{\alpha_n}$ permutations in $\mathbb{S}_n$, with replacement. By left-invariance, the bound above also holds for the Ulam distance between two given codewords of C. Using the union bound and the fact that there are less than M^2 pairs of codewords, the probability that there exist two permutations with distance $\leq n - (2+a)\sqrt{n}$ is bounded from above by

$$M^2 P\left(U_n \leq n - (2+a)\sqrt{n}\right) \leq \exp\left(-a^{3/2}\sqrt{n} + 2\alpha_n\right).$$

To ensure that the minimum distance of the code is at least $n - (2+a)\sqrt{n}$ with high probability, we must choose α_n such that $a^{3/2}\sqrt{n} > 2\alpha_n$. Hence, we let $\alpha_n = \frac{1}{2}\sqrt{a^3 n} - \epsilon$, for some $\epsilon > 0$. For this choice, with high probability, the random code C of size $\Theta\left(e^{\sqrt{a^3 n}/2}\right)$ has minimum distance at least $n - (2+a)\sqrt{n}$. In particular, for $a = 1/20$, a random code of size $\Theta\left(e^{\sqrt{n/5}/80}\right)$ with high probability has minimum distance at least $n - 2.05\sqrt{n}$.

As already pointed out, the size of a randomly constructed code obtained this way is exponential in $\sqrt{n}$, while the size of the code from the explicit construction in [30]

$$\left(\frac{2+a}{32}\right)^3 \sqrt{n}$$

is only linear in $\sqrt{n}$.

VI. Conclusion

We introduced the notion of translocation errors in rank modulation systems. Translocation errors may be viewed as generalization of adjacent swap errors frequently encountered in flash memories. We demonstrated that the metric used to capture the effects of translocation errors is the Ulam distance between two permutations, a linear function of the longest common subsequence of the permutations. We also derived asymptotically tight upper and lower bounds on the code capacity. Furthermore, we presented a number of constructions for translocation error-correcting codes based on interleaving permutation codes in the Hamming metric and deletion-correcting codes in the Levenshtein metric. Finally, we exhibited a low-complexity decoding method for a class of relaxed interleaved codes of nonzero asymptotic rate.

Acknowledgment

The authors would like to thank the editor Navin Kashyap for efficiently handling the paper and the anonymous reviewers for their valuable comments and suggestions; in particular, we are indebted to one of the reviewers for suggesting a random coding approach. The authors also gratefully acknowledge a number of discussions with Jalal Etesami.

References

[1] D. Slepian, "Permutation modulation," *Proc. IEEE*, vol. 53, no. 3, pp. 228–236, Mar. 1965.

[2] J. Karlof, "Permutation codes for the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. IT-35, no. 4, pp. 726–732, Jul. 1989.

[3] I. F. Blake, G. Cohen, and M. Deza, "Coding with permutations," *Inf. Control*, vol. 43, no. 1, pp. 1–19, 1979.

[4] C. J. Colbourn, T. Kløve, and A. C. H. Ling, "Permutation arrays for powerline communication and mutually orthogonal Latin squares," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1289–1291, Jun. 2004.

[5] J. Bruck, A. Jiang, and Z. Wang, "On the capacity of bounded rank modulation for flash memories," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2009, pp. 1234–1238.

[6] M. Kendall, *Rank Correlation Methods*, 4th ed. London, U.K.: Griffin, 1970.

[7] P. Diaconis and R. Graham, "Spearman's footrule as a measure of disarray," *J. Roy. Statist. Soc., Series B (Methodological)*, vol. 39, no. 2, pp. 262–268, 1977.

[8] H. Chadwick and L. Kurz, "Rank permutation group codes based on Kendall's correlation statistic," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 2, pp. 306–315, Mar. 1969.

[9] P. Diaconis, *Group Representations in Probability and Statistics*, ser. Lecture Notes-Monograph Series. Beachwood, OH, USA: Inst. Math. Statist., 1988, vol. 11.

[10] A. Jiang, M. Schwartz, and J. Bruck, "Correcting charge-constrained errors in the rank-modulation scheme," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2112–2120, May 2010.

[11] I. Tamo and M. Schwartz, "Correcting limited-magnitude errors in the rank-modulation scheme," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2551–2560, Jun. 2010.

[12] A. Jiang and Y. Wang, "Rank modulation with multiplicity," in *Proc. IEEE GLOBECOM Workshop*, Dec. 2010, pp. 1866–1870.

[13] Z. Wang and J. Bruck, "Partial rank modulation for flash memories," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 864–868.

[14] E. E. Gad, M. Langberg, M. Schwartz, and J. Bruck, "Constant-weight Gray codes for local rank modulation," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7431–7442, Nov. 2011.

[15] S. Lin and D. Costello, *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2004.

[16] D. Zhu and L. Wang, "On the complexity of unsigned translocation distance," *Theor. Comput. Sci.*, vol. 352, no. 1–3, pp. 322–328, 2006.

[17] M. C. Kim, K. Oh, O. C. Taek, S.-H. Choi, K. Belay, R. Elliman, and S. Russo, "Nonvolatile memories using deep traps formed in HfO2 by Nb ion implantation," *J. Appl. Phys.*, vol. 109, no. 5, pp. 053703-1–053703-4, 2011.

[18] G. Cellere, L. Larcher, A. Paccagnella, A. Visconti, and M. Bonanomi, "Radiation induced leakage current in floating gate memory cells," *IEEE Trans. Nucl. Sci.*, vol. 52, no. 6, pp. 2144–2152, Dec. 2005.

[19] W. Chu, C. J. Colbourn, and P. Dukes, "Construction for permutation codes in powerline communications," *Designs, Codes Cryptogr.*, vol. 32, pp. 51–64, 2004.

[20] J.-C. Chang, R.-J. Chen, T. Kløve, and S.-C. Tsai, "Distance-preserving mappings from binary vectors to permutations," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 1054–1059, Apr. 2003.

[21] V. I. Levenshtein, "On perfect codes in deletion and insertion metric," *Discrete Math. Appl.*, vol. 2, no. 3, pp. 241–258, 1992.

[22] A. Mazumdar, A. Barg, and G. Zemor, "Parameters of rank modulation codes: Examples," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2011, pp. 13–17.

[23] A. Barg and A. Mazumdar, "Codes in permutations and error correction for rank modulation," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3158–3165, Jul. 2010.

[24] A. Mazumdar, A. Barg, and G. Zemor, "Constructions of rank modulation codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul./Aug. 2011, pp. 869–873.

[25] T. Kløve, T.-T. Lin, S.-C. Tsai, and W.-G. Tzeng, "Permutation arrays under the Chebyshev distance," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2611–2617, Jun. 2010.

[26] T. Swart and H. Ferreira, "Decoding distance-preserving permutation codes for power-line communications," in *Proc. IEEE AFRICON*, Windhoek, South Africa, Sep. 2007, pp. 1–7.

[27] T. Wadayama and M. Hagiwara, "LP decodable permutation codes based on linearly constrained permutation matrices," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, Jul./Aug. 2011, pp. 139–143.

[28] L. Grupp, A. Caulfield, J. Coburn, S. Swanson, E. Yaakobi, P. Siegel, and J. Wolf, "Characterizing flash memory: Anomalies, observations, and applications," in *42nd Annu. IEEE/ACM Int. Symp. Microarchitect.*, Dec. 2009, pp. 24–33.

[29] M. Deza and T. Huang, "Metrics on permutations, a survey," *J. Combin., Inf., Syst. Sci.*, vol. 23, pp. 173–185, 1998.

[30] P. Beame, E. Blais, and D. Huynh-Ngoc, Longest common subsequences in sets of permutations 2009, Arxiv preprint arXiv:0904.1615.

[31] M. Deza and E. Deza, *Encyclopedia of Distances*. New York, NY, USA: Springer-Verlag, 2009.

[32] A. Cayley, "Note on the theory of permutations," *Philosoph. Mag. Ser. 3*, vol. 34, no. 232, pp. 527–529, 1849.

[33] P. Frankl and M. Deza, "On the maximum number of permutations with given maximal or minimal distance," *J. Combin. Theory, Ser. A*, vol. 22, pp. 352–360, 1977.

[34] M. Deza and S. Vanstone, "Bounds for permutation arrays," *J. Statist. Plann. Infer.*, vol. 2, no. 2, pp. 197–209, 1978.

[35] H. Tarnanen, "Upper bounds on permutation codes via linear programming," *Eur. J. Combin.*, vol. 20, no. 1, pp. 101–114, 1999.

[36] A. A. Davydov, V. V. Zyablov, and R. E. Kalimullin, "Special sequences as subcodes of Reed-Solomon codes," *Probl. Inf. Transmiss.*, vol. 46, no. 4, pp. 321–345, Dec. 2010.

[37] P. J. Cameron, "Permutation codes," *Eur. J. Combin.*, vol. 31, no. 2, pp. 482–490, 2010.

[38] F.-W. Fu and T. Kløve, "Two constructions of permutation arrays," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 881–883, May 2004.

[39] J.-C. Chang, R.-J. Chen, and S.-C. Tsai, "Distance preserving mappings from binary vectors to permutations," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2003, pp. 14–14.

[40] I. F. Blake, "Permutation codes for discrete channels," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 1, pp. 138–140, Jan. 1974.

[41] J.-C. Chang, "Distance-increasing mappings from binary vectors to permutations that increase Hamming distances by at least two," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1683–1689, Apr. 2006.

[42] P. Neumann, "Encoding and decoding for cyclic permutation codes," *IRE Trans. Electron. Comput.*, vol. EC-11, no. 4, pp. 507–511, Aug. 1962.

[43] J. Kim, "On increasing subsequences of random permutations," *J. Combin. Theory, Series A*, vol. 76, no. 1, pp. 148–155, 1996.

[44] J. Baik, P. Deift, and K. Johansson, "On the distribution of the length of the longest increasing subsequence of random permutations," *J. Amer. Math. Soc.*, vol. 12, pp. 1119–1178, 1999.

[45] D. Aldous and P. Diaconis, "Longest increasing subsequences: From patience sorting to the Baik-Deift-Johansson theorem," *Bull. (New Ser.) Amer. Math. Soc.*, vol. 36, pp. 413–432, 1999.

[46] A. M. Odlyzko and E. M. Rains, *On Longest Increasing Subsequences in Random Permutations*. Providence, RI, USA: Amer. Math. Soc., 2000, vol. 251, pp. 439–451.

Farzad Farnoud (Hassanzadeh) (S'10) received his B.Sc. degree in Electrical Engineering from Sharif University of Technology, Iran, in 2006 and his M.Sc. degree in Electrical and Computer Engineering from University of Toronto, Canada in 2008. He is currently a Ph.D. candidate in Electrical and Computer Engineering at University of Illinois at Urbana-Champaign. His research interests include error correcting codes for data storage, rank aggregation and social choice, and probability estimation of sources with large alphabets. He is a recipient of the Robert T. Chien Memorial Award for excellence in research in the field of electrical engineering from the University of Illinois at Urbana-Champaign.

Vitaly Skachek received the B.A. (Cum Laude), M.Sc. and Ph.D. degrees in computer science from the Technion—Israel Institute of Technology, in 1994, 1998 and 2007, respectively.

In the summer of 2004, he visited the Mathematics of Communications Department at Bell Laboratories under the DIMACS Special Focus Program in Computational Information Theory and Coding. During 2007–2012, he held research positions with the Claude Shannon Institute, University College Dublin, with the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, and with the Department of Electrical and Computer Engineering, McGill University, Montréal. He is now a Senior Lecturer with the Institute of Computer Science, University of Tartu.

Dr. Skachek is a recipient of the Permanent Excellent Faculty Instructor award, given by Technion.

Olgica Milenkovic (M'04–SM'12) received her M.Sc. degree in Mathematics and Ph.D. in Electrical Engineering from the University of Michigan Ann Arbor, in 2001 ad 2002, respectively. In 2002 she joined the faculty of University of Colorado, Boulder. After a visiting professor appointment at UCSD, she moved to University of Illinois, Urbana-Champaign in 2007, where she is now Associate Professor in the Department of Electrical and Computer Engineering. Her research interests include coding theory, analysis of algorithms, bioinformatics, and signal processing. She served as editor in Chief of the Special Issue of the Transaction of Information Theory on Molecular Biology and Neuroscience, and associate editor of the Transactions on Communications, Transactions on Signal Processing and the Transactions on Information Theory. For her work, she was rewarded the NSF Career Award, DARPA Young Faculty Awards, and several best conference paper awards in coding theory and bioinformatics. She is a Center for Advanced Studies Associate at UIUC.