Systematic Error-Correcting Codes for Permutations and Multi-Permutations

Sarit Buzaglo, *Member, IEEE*, Eitan Yaakobi, *Member, IEEE*, Tuvi Etzion, *Fellow, IEEE*, and Jehoshua Bruck, *Fellow, IEEE*

***Abstract*— Multi-permutations and in particular permutations appear in various applications in an information theory. New applications, such as rank modulation for flash memories, have suggested the need to consider error-correcting codes for multi-permutations. In this paper, we study systematic error-correcting codes for multi-permutations in general and for permutations in particular. For a given number of information symbols k, and for any integer t, we present a construction of $(k+r, k)$ systematic t-error-correcting codes, for permutations of length $k+r$, where the number of redundancy symbols r is relatively small. In particular, for a given t and for sufficiently large k, we obtain $r = t+1$, while a lower bound on the number of redundancy symbols is shown to be t. The same construction is also applied to obtain related systematic error-correcting codes for any types of multi-permutations.**

***Index Terms*— Kendall τ-metric, multi-permutations, permutations, systematic error-correcting codes.**

I. Introduction

FLASH memory is one of the most widely used non-volatile technologies. In flash memories, cells usually represent multiple levels, which correspond to the amount of electrons trapped in each cell. Currently, one of the main challenges in flash memory cells is to program each cell exactly to its designated level. In order to overcome this difficulty, the novel framework of *rank modulation codes* was introduced in [15]. In this setup, the information is carried by the relative values between the cells rather than by their absolute levels. Thus, every group of cells induces a permutation, which is derived by the ranking of the level of each cell in the group. There are several works which study the correction of errors under the setup of permutations for the rank modulation scheme; see e.g. [1], [9], [16], [22], [23], [26], [27]. In all these works t-error-correcting codes were considered for the set S_n, which consists of all permutations on n elements, with either the Kendall τ-metric, the infinity metric, or the Ulam metric. Permutation codes were originally studied with the Hamming distance in the work of Slepian for the transmission of bandlimited signals over Gaussian channels [21] and in many other papers, e.g. [2], [3], [11]. Recently, to improve the number of rewrites, the model of rank modulation was extended such that multiple cells can share the same ranking [12], [13]. Thus, the cells no longer determine permutations but rather multi-permutations. Error-correcting codes for multi-permutations subject to the Kendall τ-metric were presented in [20] and also studied in [7]. The goal of this paper is to construct systematic error-correcting codes for permutations and multi-permutations. In such a code with permutations there are $k!$ codewords, where k is the number of information symbols. Similarly, in such a code with multi-permutations there are $\alpha(k)$ codewords, where $\alpha(k)$ is the number of multi-permutations that can be defined on the k information symbols.

Manuscript received November 18, 2015; revised March 14, 2016; accepted March 15, 2016. Date of publication March 17, 2016; date of current version May 18, 2016. S. Buzaglo and T. Etzion were supported by the U.S.–Israel Binational Science Foundation, Jerusalem, Israel, under Grant 2012016. E. Yaakobi and J. Bruck were supported by the Intellectual Ventures and National Science Foundation under Grant CIF-1218005. J. Bruck was supported by the U.S.–Israel Binational Science Foundation, Jerusalem, Israel, under Grant 2010075. This paper was presented at the 2014 IEEE International Symposium on Information Theory.

S. Buzaglo is with the Center for Memory and Recording Research, University of California at San Diego, La Jolla, CA 92093-0401 USA (e-mail: sbuzaglo@ucsd.edu).

E. Yaakobi and T. Etzion are with the Computer Science Department, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: yaakobi@cs.technion.ac.il; etzion@cs.technion.ac.il).

J. Bruck is with the Electrical Engineering Department, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: bruck@caltech.edu).

Communicated by M. Schwartz, Associate Editor for Coding Techniques.

Digital Object Identifier 10.1109/TIT.2016.2543739

A. Previous Work

As mentioned above, the rank modulation scheme was proposed in [15] to improve programming performance for flash memory, where n cells represent a permutation according to the ranking of their levels. This scheme was suggested to be useful also for data retention, as it was noticed that the ranking of the cells' levels is more robust to charge leakage than the absolute values of the cells' levels.

In [16] the rank modulation scheme was combined with error-correction capability by using the Kendall τ-metric. This metric highly reflects the error behavior of flash memory cells, mainly due to dominant error sources, e.g. charge leakage and read disturbance [16]. Error-correcting codes were constructed in [16] and later in [1] and [19] by using a metric embedding of the set of all permutations of length n, S_n, with the Kendall τ-metric to the space $\mathbb{Z}_q^{n-1}$, $q \geq n$, with the Lee metric. This metric embedding allows to construct t-error-correcting codes in S_n with the Kendall τ-distance from t-error-correcting codes in $\mathbb{Z}_q^{n-1}$ with the Lee distance. The embedding was extended in [20] to construct error-correcting codes for balanced multi-permutations.

Bounds on the size of error-correcting codes in the Kendall τ-metric were given in [1], [5], [6], and [16]. In [1], t-error-correcting codes in S_n that achieve the sphere packing bound up to a constant factor, where n is sufficiently large, were presented. Upper bounds on the size of codes with even minimum distances were proposed in [5] and [6], which also investigated the existence question of perfect codes.

The concept of systematic codes for permutations was suggested in [26], where systematic error-correcting codes were studied in the Kendall τ-metric. Systematic error-correcting codes with the Kendall τ-distance were further studied in [8] and in [27]. In [27], a variation of systematic error-correcting codes with the infinity distance were also explored. A code $\mathcal{C} \subseteq S_n$ is an (n, k) systematic code if each permutation of S_k is a sub-permutation of exactly one codeword of $\mathcal{C}$. In [27] four constructions of (n, k) systematic t-error-correcting codes in the Kendall τ-metric were presented. All these constructions are based on error-correcting codes in the Lee metric via the metric embedding from [16]. Two of the constructions from [27] (Constructions A and B) are for systematic single-error-correcting codes that use two redundancy symbols. One of the constructions from [27] (Construction C) is for systematic t-error-correcting codes, for a general t. In particular, t could be as large as $\Theta(n^2)$. The constructed codes use r redundancy symbols, where r is shown to be less than or equal to $2t+1$. However, it is not clear whether r can be smaller than $2t+1$. Finally, Construction D from [27] yields $(k+t+1, k)$ systematic t-error-correcting codes for a fixed $t \geq 1$, provided that k is sufficiently large.

B. Our Contribution

In this paper we present a general method to construct (n, k) systematic t-error-correcting codes. This method is based on two ingredients. The first one is a partition of S_k into t-error-correcting codes with the Kendall τ-distance. The second one is a code for multi-permutations on the multi-set $\{0^k, k+1, \ldots, k+r\}$ with minimum Kendall τ-distance $2t$.

We apply this method to construct (n, k) systematic t-error-correcting codes and analyze the asymptotic behavior of the number of redundancy symbols r, for $t = \Theta(k^\epsilon)$ and $\epsilon \geq 0$. We present an (n, k) systematic single-error-correcting codes with $r = 2$ redundancy symbols for every $k \geq 1$. For a fixed t and for large enough k, the constructed codes use $t+1$ redundancy symbols. For $t = \Theta(k^\epsilon)$, the constructed codes use $r = \lceil (1+\epsilon+\delta)t \rceil$ redundancy symbols, if $0 < \epsilon \leq 1$, and $r = \left\lceil (1+\epsilon^{-1}+\delta)t \right\rceil$ redundancy symbols, if $\epsilon > 1$, where k is sufficiently large, $r-1$ is a power of a prime, and $\delta > 0$ can be arbitrarily small.

One advantage of our method is that it can be easily adapted to systematic t-error-correcting codes for multi-permutations. It can also be used for other metrics, e.g. the Ulam metric and the Hamming metric, provided that one can construct multi-permutation codes and partitions into error-correcting codes in these metrics. For balanced multi-permutations we construct systematic t-error-correcting codes with $t+1$ redundancy symbols for sufficiently large k. Finally, we prove that at least t redundancy symbols are required when k is large enough and the multiplicity of each information symbol is bounded.

C. Organization

The rest of this work is organized as follows. In Section II we present the basic concepts concerning permutations, multi-permutations, and systematic codes for permutations and multi-permutations. We introduce in Section III the metric used in this paper, the Kendall τ-metric, and present basic properties of this metric. Next, we present in Section IV our main construction for systematic t-error-correcting codes for permutations. The construction is based on a combination of two coding concepts. The first one is a partition of a set of permutations into t-error-correcting codes. The second one is an error-correcting code for a certain family of multi-permutations. In Section V we review and generalize some of the known constructions of error-correcting codes for permutations and multi-permutations via the metric embedding from [16]. These constructions will be used to design the two coding concepts for the main construction. Then, in Section VI, specific systematic codes for permutations based on the discussion in the preceding sections are given, and in Section VII the constructions are generalized for multi-permutations. In Section VIII, we study an asymptotic lower bound on the number of redundancy symbols in systematic t-error-correcting codes. We conclude in Section IX.

II. Permutation, Multi-Permutations, and Systematic Codes

Let $[n]$ denote the set of n integers $\{1, 2, \ldots, n\}$ and let $[a, b]$, $a < b$, denote the set of $b-a+1$ integers $\{a, a+1, a+2, \ldots, b\}$. A permutation on a set X of n elements is a bijection $\sigma : [n] \rightarrow X$. A permutation σ on X is denoted by $\sigma = [\sigma(1), \sigma(2), \ldots, \sigma(n)]$. Let S_n be the set of all permutations on $[n]$ and let $S([a, b])$ be the set of all permutations on $[a, b]$. The concept of permutations is generalized to *multi-permutations* as follows. A *multi-set* $\mathcal{M} = \{v_1^{m_1}, v_2^{m_2}, \ldots, v_\ell^{m_\ell}\}$ is a collection of the elements $\{v_1, v_2, \ldots, v_\ell\}$ in which v_i appears m_i times, $i \in [\ell]$. The elements of $\{v_1, v_2, \ldots, v_\ell\}$ are called *ranks*, while the positive integer m_i, for all $i \in [\ell]$, is called the *multiplicity* of the ith rank v_i. If $m_1 = m_2 = \cdots = m_\ell = m$ then $\mathcal{M}$ is called a *balanced multi-set* and the related multi-permutations are called *balanced multi-permutations*. A multi-permutation on the multi-set $\mathcal{M}$ is a mapping $\sigma : [n] \rightarrow \{v_1, v_2, \ldots, v_\ell\}$, where $n = \sum_{i=1}^{\ell} m_i$, such that $|\{j \in [n] \ : \ \sigma(j) = v_i\}| = m_i$, for all $i \in [\ell]$. A permutation is a special case of a multi-permutation, where all the multiplicities are equal to one. We denote by $S(\mathcal{M})$ the set of all multi-permutations on $\mathcal{M}$. Clearly, the size of $S(\mathcal{M})$ is equal to $\frac{n!}{\Pi_{i=1}^{\ell} m_i!}$. As for permutations, we denote a multi-permutation $\sigma \in S(\mathcal{M})$ by $\sigma = [\sigma(1), \sigma(2), \ldots, \sigma(n)]$, where the meaning will be clear from the context.

Example 1: If $\mathcal{M} = \{1^3, 2^2, 3^2\}$, *then* $\sigma = [3, 1, 3, 1, 2, 1, 2]$ *is a multi-permutation on* $\mathcal{M}$.

For a permutation $\alpha \in S_n$ and for $k \in [n]$, define $\alpha_{\downarrow k}$ to be the permutation in S_k obtained from α by deleting all the elements of $\{k+1, k+2, \ldots, n\}$ from α.

Example 2: If $\alpha = [2, 5, 4, 1, 3, 6]$ *and* $k = 3$ *then* $\alpha_{\downarrow k} = [2, 1, 3]$.

A *metric space* $(\mathcal{V}, d(\cdot, \cdot))$ is a set $\mathcal{V}$ together with a mapping $d : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}_{\geq 0}$ such that, for every $x, y, z \in \mathcal{V}$ 1) $d(x, y) = 0$ if and only if $x = y$; 2) $d(x, y) = d(y, x)$;

3) $d(x, y) + d(y, z) \geq d(x, z)$. The mapping d is called a *metric*. Let $(\mathcal{V}, d(\cdot, \cdot))$ be a metric space. A code $C \subseteq \mathcal{V}$ has *minimum distance* d if $d(x, y) \geq d$, for every two distinct codewords $x, y \in \mathcal{C}$. A code $C \subseteq \mathcal{V}$ is a *t-error-correcting code* if it has minimum distance at least $2t + 1$.

Motivated by the rank modulation scheme [15], the concept of *systematic* codes for permutations was proposed in [26] and [27]. A code $\mathcal{C} \subseteq S_n$ is an (n, k) systematic code if each permutation of S_k is a sub-permutation of exactly one codeword of $\mathcal{C}$, i.e. for every $\sigma \in S_k$ there exists exactly one codeword (permutation) $\alpha \in \mathcal{C}$ such that $\alpha_{\downarrow k} = \sigma$. Therefore, the size of an (n, k) systematic code is $k!$. If an (n, k) systematic code $\mathcal{C}$ is also a t-error-correcting code, then $\mathcal{C}$ is called an (n, k) *systematic t-error-correcting code*, while the metric will be clear from the context. The number of *redundancy symbols* of an (n, k) systematic code is $r = n - k$.

III. The Kendall τ-Metric on Permutations and Multi-Permutations

Given a multi-permutation $\sigma = [\sigma(1), \sigma(2), \ldots, \sigma(n)] \in S(\mathcal{M})$, an *adjacent transposition* is an exchange of two distinct adjacent elements $\sigma(j), \sigma(j+1)$ in σ, for some $j \in [n-1]$. The result of such an adjacent transposition is the multi-permutation $[\sigma(1), \ldots, \sigma(j-1), \sigma(j+1), \sigma(j), \sigma(j+2), \ldots, \sigma(n)]$. The Kendall τ-distance between two multi-permutations $\sigma, \pi \in S(\mathcal{M})$, denoted by $d_K(\sigma, \pi)$, is the minimum number of adjacent transpositions required to obtain the multi-permutation π from the multi-permutation σ.

Example 3: If $\sigma = [3, 1, 3, 1, 2, 1, 2]$*, and* $\pi = [3, 3, 1, 2, 1, 2, 1]$ *then* $d_K(\sigma, \pi) = 3$*, since three is the minimum number of adjacent transpositions required to transfer the multi-permutation* σ *to* π*:* $[3, 1, 3, 1, 2, 1, 2] \rightarrow [3, 3, 1, 1, 2, 1, 2] \rightarrow [3, 3, 1, 2, 1, 1, 2] \rightarrow [3, 3, 1, 2, 1, 2, 1]$.

The Kendall τ-metric was originally defined for permutations [17]. It is well known [16], [18] that for two permutations $\sigma, \pi \in S_n$, the value $d_K(\sigma, \pi)$ can be expressed as the number of pairs of elements of $[n]$ that do not appear in the same order in σ and π, i.e.

$$d_K(\sigma, \pi) = \left| \left\{ (i, j) : \begin{array}{c} \forall\ 1 \leq i, j \leq n,\ \sigma^{-1}(i) < \sigma^{-1}(j) \\ \text{and } \pi^{-1}(i) > \pi^{-1}(j) \end{array} \right\} \right|. \tag{1}$$

For a multi-permutation $\sigma \in S(\mathcal{M})$, where $\mathcal{M} = \{v_1^{m_1}, v_2^{m_2}, \ldots, v_\ell^{m_\ell}\}$, we distinguish between the appearances of the same rank in σ, by their positions in σ. We consider the increasing order of these positions. By abuse of notation we sometimes write $\sigma(j) = (v_i)_r$ and $j = \sigma^{-1}((v_i)_r)$ to indicate that the rth appearance of v_i is in the jth position of σ, i.e. $\sigma(j) = v_i$ and the multiplicity of v_i in the multi-permutation $[\sigma(1), \sigma(2), \ldots, \sigma(j)]$ is r. The computation of the Kendall τ-distance given in (1) between two permutations can be generalized to two multi-permutations $\sigma, \pi \in S(\mathcal{M})$. More explicitly, it can be expressed as the number of pair of elements of $\{(i, r) : i \in [\ell], r \in [m_i]\}$ that do not appear in the same order in σ and π, i.e.

$$d_K(\sigma, \pi) = \left| \left\{ ((i, r), (j, s)) : \begin{array}{c} \sigma^{-1}((v_i)_r) < \sigma^{-1}((v_j)_s) \\ \pi^{-1}((v_i)_r) > \pi^{-1}((v_j)_s) \end{array} \right\} \right|. \tag{2}$$

Let $n_0 = 0$ and let $n_i = \sum_{j=1}^{i} m_j$, $i \in [\ell]$, where $n = n_\ell$. In other words, n_i is the number of symbols in the multi-set $\mathcal{M}$ whose rank is at most i. For $\theta \in S_n$, the *assignment* of the permutation θ in a multi-permutation $\sigma \in S(\mathcal{M})$ is the permutation $\alpha = \theta \triangleright \sigma \in S_n$ defined as follows. For each i, $1 \leq i \leq \ell$, the segment of the permutation $[\theta(n_{i-1}+1), \theta(n_{i-1}+2), \ldots, \theta(n_{i-1}+m_i)]$ is substituted, in this order, in the m_i positions of the rank v_i in σ. This means that for each $j \in [n]$, if $\sigma(j) = (v_i)_r$ then $\alpha(j) = \theta(n_{i-1} + r)$.

Example 4: Let $\sigma = [3, 1, 3, 1, 2, 1, 2] \in S(\{1^3, 2^2, 3^2\})$ *and let* $\theta = [2, 1, 3, 4, 5, 7, 6]$*. After substituting* $[2, 1, 3]$ *in positions* 2, 4, *and* 6 *in which* 1 *appears in* σ*, and similarly, substituting* $[4, 5]$ *and* $[7, 6]$ *in the positions in which* 2 *and* 3 *appears, respectively, we obtain the permutation* $\theta \triangleright \sigma = [7, 2, 6, 1, 4, 3, 5]$.

Lemma 1: Let $\sigma, \pi \in S(\mathcal{M})$*, let* $\theta_i, \eta_i \in S([n_{i-1} + 1, n_i])$*, for all* $i \in [\ell]$*, and let* $\theta, \eta \in S_n$*, where* $\theta(n_{i-1}+r) = \theta_i(r)$ *and* $\eta(n_{i-1}+r) = \eta_i(r)$*, for all* $i \in [\ell]$ *and* $r \in [m_i]$*. Then*

$$d_K(\theta \triangleright \sigma, \eta \triangleright \pi) \geq d_K(\sigma, \pi) + d_K(\theta, \eta).$$

Proof: If $d = d_K(\theta \triangleright \sigma, \eta \triangleright \pi)$, then by the definition of the Kendall τ-distance, there exists a sequence $\boldsymbol{\tau} = \tau_1, \tau_2, \ldots, \tau_d$ of d adjacent transpositions that transfers $\theta \triangleright \sigma$ to $\eta \triangleright \pi$.

Let $\boldsymbol{\tau}_{mult} = \tau_{s_1}, \tau_{s_2}, \ldots, \tau_{s_{d_{mult}}}$ be the subsequence of $\boldsymbol{\tau}$ that consists of all the adjacent transpositions of the sequence $\boldsymbol{\tau}$ that exchange two distinct symbols $x \in [n_{i-1}+1, n_i]$ and $y \in [n_{j-1}+1, n_j]$, for some $i, j \in [\ell]$, $i \neq j$.

For all $i \in [\ell]$, let $\boldsymbol{\tau}^{(i)} = \tau_{i,1}, \tau_{i,2}, \ldots, \tau_{i,d_i}$ be the subsequence of $\boldsymbol{\tau}$ that consists of all the adjacent transpositions of the sequence $\boldsymbol{\tau}$ that exchanges some two distinct symbols $x, y \in [n_{i-1}+1, n_i]$.

Each adjacent transposition in the sequence $\boldsymbol{\tau}$ exchanges two distinct symbols x and y, $x \in [n_{i-1}+1, n_i]$ and $y \in [n_{j-1}+1, n_j]$, for some $i, j \in [\ell]$, where either $i = j$ or $i \neq j$. Hence, $\boldsymbol{\tau}_{mult}$ and $\boldsymbol{\tau}^{(1)}, \boldsymbol{\tau}^{(2)}, \ldots, \boldsymbol{\tau}^{(\ell)}$ form a partition of $\boldsymbol{\tau}$ to subsequences and $d_{mult} + \sum_{i=1}^{\ell} d_i = d$.

By the definitions of the assignment of a permutation in a multi-permutation and of θ and η, for every $\kappa \in [\ell]$ the permutations θ_κ and η_κ are substituted in the positions of the rank v_κ in σ and π, respectively. Since $\boldsymbol{\tau}$ transfers $\theta \triangleright \sigma$ to $\eta \triangleright \pi$, it follows that $\boldsymbol{\tau}_{mult}$ transfers σ to π and for every $i \in [\ell]$, $\boldsymbol{\tau}^{(i)}$ transfers the segment of θ, $\theta_i = [\theta(n_{i-1}+1), \theta(n_{i-1}+2), \ldots \theta(n_i)]$, to the segment of η, $\eta_i = [\eta(n_{i-1}+1), \eta(n_{i-1}+2), \ldots \eta(n_i)]$. Therefore, $d_{mult} \geq d_K(\sigma, \pi)$ and for all $i \in [\ell]$, $d_i \geq d_K(\theta_i, \eta_i)$. Furthermore, $\sum_{i=1}^{\ell} d_K(\theta_i, \eta_i) = d_K(\theta, \eta)$, and thus

$$\begin{aligned} d_K(\theta \triangleright \sigma, \eta \triangleright \pi) &= d_{mult} + \sum_{i=1}^{\ell} d_i \geq d_K(\sigma, \pi) \\ &\quad + \sum_{i=1}^{\ell} d_K(\theta_i, \eta_i) = d_K(\sigma, \pi) + d_K(\theta, \eta). \end{aligned}$$

□

Lemma 1 provides a lower bound on $d_K(\theta \triangleright \sigma, \eta \triangleright \pi)$ in terms of $d_K(\sigma, \pi)$ and $d_K(\theta, \eta)$. This lower bound may not always be tight, as the next example shows.

Example 5: Let $\sigma = [3,1,3,1,2,1,2]$*,* $\pi = [3,3,1,2,1,2,1]$*,* $\theta_1 = [2,1,3]$*,* $\eta_1 = [3,2,1]$*,* $\theta_2 = \eta_2 = [4,5]$*,* $\theta_3 = [7,6]$*, and* $\eta_3 = [6,7]$*. Then,* $\theta = [2,1,3,4,5,7,6]$*,* $\eta = [3,2,1,4,5,6,7]$*,* $d_K(\theta,\eta) = 3$*, and* $d_K(\sigma,\pi) = 3$*. However,* $d_K(\theta \triangleright \sigma, \eta \triangleright \pi) = d_K([7,2,6,1,4,3,5],[6,7,3,4,2,5,1]) = 8$ *and thus*

$$d_K(\theta \triangleright \sigma, \eta \triangleright \pi) > d_K(\sigma,\pi) + d_K(\theta,\eta) = 6.$$

The Kendall τ-metric on $S(\mathcal{M})$ is graphic, i.e. for every two multi-permutations $\sigma, \pi \in S(\mathcal{M})$ their Kendall τ-distance is equal to the length of the shortest path between σ and π in the graph $G(\mathcal{M})$ whose vertex set is the set $S(\mathcal{M})$, and two vertices are connected by an edge if and only if their Kendall τ-distance is one.

A metric $d(\cdot,\cdot)$ on a set $\mathcal{V}$, is called *bipartite* if, for every three elements $x, y, z \in \mathcal{V}$, the congruence $d(x,y)+d(y,z) \equiv d(x,z) \pmod 2$ is satisfied, i.e. the related graph is bipartite. The Kendall τ-metric on $S(\mathcal{M})$ is bipartite as stated in the next lemma.

Lemma 2: The Kendall τ*-metric over* $S(\mathcal{M})$ *is bipartite.*

Proof: Fix a multi-permutation $\gamma \in S(\mathcal{M})$ and note that by (2) two multi-permutations which differ in exactly one adjacent transposition have different distances modulo 2 from γ. This implies that the related graph $G(\mathcal{M})$ is bipartite. □

IV. Systematic Error-Correcting Codes for Permutations

In this section the main construction of systematic t-error-correcting codes for permutations is presented. This construction will be generalized in Section VII for multi-permutations.

Let r be a positive integer and let $\mathcal{M}_{k,r} = \{0^k, k+1, k+2, \ldots, k+r\}$. For every permutation $\sigma \in S_k$ and for every multi-permutation $\rho \in S(\mathcal{M}_{k,r})$, the *assignment* of σ in ρ is the permutation $\alpha = \sigma \blacktriangleright \rho \in S_{k+r}$ which is obtained by substituting σ, in the k positions where 0 appears in ρ. Note, that $\sigma \blacktriangleright \rho = \theta \triangleright \rho$, where $\theta = [\sigma(1), \sigma(2), \ldots, \sigma(k), k+1, k+2, \ldots, k+r]$. Hence, by Lemma 1 we have

Corollary 1: Let $\sigma, \pi \in S_k$ *and* $\rho_1, \rho_2 \in S(\mathcal{M}_{k,r})$*. Then*

$$d_K(\sigma \blacktriangleright \rho_1, \pi \blacktriangleright \rho_2) \geq d_K(\rho_1,\rho_2) + d_K(\sigma,\pi).$$

The next lemma is readily verified and so we omit its proof.

Lemma 3: For every $\rho \in S(\mathcal{M}_{k,r})$ *and* $\sigma \in S_k$ *we have that* $(\sigma \blacktriangleright \rho)_{\downarrow k} = \sigma$*.*

Example 6: If $k = 4$*,* $r = 3$*,* $\rho = [0,6,0,0,5,7,0]$*, and* $\sigma = [2,4,1,3]$ *then* $\sigma \blacktriangleright \rho = [2,6,4,1,5,7,3]$ *and* $(\sigma \blacktriangleright \rho)_{\downarrow k} = [2,4,1,3] = \sigma$*.*

We are now in a position to present our construction of systematic error-correcting codes for permutations in the Kendall τ-metric.

Theorem 1: Let $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_F$ *be a partition of* S_k *into* t*-error-correcting codes in the Kendall* τ*-metric and let* $\mathcal{C}_{mult} \subseteq S(\mathcal{M}_{k,r})$ *be a code with minimum Kendall* τ*-distance* $2t$ *and size at least* F*. Let* $\rho_1, \rho_2, \ldots, \rho_F$ *be distinct codewords in* $\mathcal{C}_{mult}$*. Then the code* $\mathcal{C}_{sys} \subseteq S_{k+r}$ *defined by*

$$\mathcal{C}_{sys} \stackrel{\text{def}}{=} \bigcup_{j=1}^{F} \{\sigma \blacktriangleright \rho_j \ : \ \sigma \in \mathcal{C}_j\}$$

is a $(k+r,k)$ *systematic* t*-error-correcting code with the Kendall* τ*-distance.*

Proof: Since the codes $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_F$ form a partition of S_k, it follows that for every $\sigma \in S_k$ there exists exactly one $j \in [F]$ such that $\sigma \in \mathcal{C}_j$. By Lemma 3 it follows that $\alpha = \sigma \blacktriangleright \rho_j$ is the unique permutation in $\mathcal{C}_{sys}$ such that $\alpha_{\downarrow k} = \sigma$. Hence, the code $\mathcal{C}_{sys}$ is $(k+r,k)$ systematic.

To show that the minimum Kendall τ-distance of $\mathcal{C}_{sys}$ is at least $2t+1$, let $\sigma \blacktriangleright \rho_{j_1}, \pi \blacktriangleright \rho_{j_2}$ be two distinct codewords in $\mathcal{C}_{sys}$. By Lemma 3 and since $\mathcal{C}_{sys}$ is $(k+r,k)$ systematic, it follows that $\sigma \neq \pi$ and therefore $d_K(\sigma,\pi) \geq 1$. We distinguish now between two cases:

1) If $j_1 = j_2$ then $\sigma, \pi \in \mathcal{C}_{j_1}$. Since $\mathcal{C}_{j_1}$ is a t-error-correcting code and by Corollary 1, it follows that $d_K(\sigma \blacktriangleright \rho_{j_1}, \pi \blacktriangleright \rho_{j_2}) \geq d_K(\sigma,\pi) \geq 2t+1$.
2) If $j_1 \neq j_2$ then $\rho_{j_1} \neq \rho_{j_2}$. Since $\mathcal{C}_{mult}$ has minimum Kendall τ-distance at least $2t$, it follows by Corollary 1 that $d_K(\sigma \blacktriangleright \rho_{j_1}, \pi \blacktriangleright \rho_{j_2}) \geq d_K(\rho_{j_1},\rho_{j_2}) + d_K(\sigma,\pi) \geq 2t+1$.

Thus, we proved that $\mathcal{C}_{sys}$ is a $(k+r,k)$ systematic code with minimum Kendall τ-distance at least $2t+1$, as required. □

For the construction of the code $\mathcal{C}_{sys}$ in Theorem 1 two ingredients are required. The first one is a partition of S_k into t-error-correcting codes. The second one is a code in $S(\mathcal{M}_{k,r})$ with minimum Kendall τ-distance $2t$. In the next section we review some of the known constructions of error-correcting codes for multi-permutations. These constructions will be used to generate partitions of S_k into t-error-correcting codes, and partitions of multi-permutations into t-error-correcting codes (which will be used in Section VII). These results will also produce the second ingredient of codes in $S(\mathcal{M}_{k,r})$ with minimum Kendall τ-distance $2t$.

V. Error-Correcting Codes and Partitions via Metric Embedding

The primary goal of this section is to generate error-correcting codes and partitions for multi-permutations. In Subsection V-A, we review a known method to generate error-correcting codes for multi-permutations via metric embedding. Then, in Subsection V-B, we describe the resulting code constructions based on this method. Finally, in Subsection V-C, we derive partitions of permutations and multi-permutations into t-error-correcting codes.

A. Constructions From Metric Embedding

The first constructions of error-correcting codes for permutations in the Kendall τ-metric were given in [16]. In particular, a general method was presented to construct codes from error-correcting codes in the Lee metric. This method was used in [1] to produce codes which correct multiple errors, and in [20], it was extended for the construction of error-correcting codes for balanced multi-permutations in the Kendall τ-metric. For the completeness of the results in the paper, we will review the full details of this method, with some modifications which will be explained later in the section.

Let $\mathbb{Z}_q^N$ be the set of all vectors of length N over the alphabet $\mathbb{Z}_q$. For any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_q^N$, the *Lee distance* $d_L(\mathbf{x}, \mathbf{y})$ is defined by

$$d_L(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N} \min\{|x_i - y_i|, q - |x_i - y_i|\}.$$

The *Lee weight* of a vector $\mathbf{x} \in \mathbb{Z}_q^N$ is defined as $w_L(\mathbf{x}) = d_L(\mathbf{x}, \mathbf{0})$, where $\mathbf{0}$ is the all-zero vector. A vector $\mathbf{x} = (x_1, x_2, \ldots, x_s) \in [0, b]^s$, where b and s are two positive integers, is called *monotone* if $x_i \geq x_{i+1}$, for all $1 \leq i \leq s-1$. Denote by $[0, b]^s_{\succeq}$ the set of all monotone vectors in $[0, b]^s$. Let

$$\mathcal{A}(\mathcal{M}) \stackrel{\text{def}}{=} [0, n_1]^{m_2}_{\succeq} \times [0, n_2]^{m_3}_{\succeq} \times \cdots \times [0, n_{\ell-1}]^{m_\ell}_{\succeq}.$$

Lemma 4: For every multi-set $\mathcal{M}$,

$$|\mathcal{A}(\mathcal{M})| = |S(\mathcal{M})|.$$

Proof: For every positive integers b and s the size of $[0, b]^s_{\succeq}$ is equal to the number of choices of s elements from $[0, b]$, with repetitions, i.e.

$$|[0, b]^s_{\succeq}| = \binom{b+s}{s}.$$

Therefore,

$$\begin{aligned}|\mathcal{A}(\mathcal{M})| &= \binom{n_1+m_2}{m_2} \cdot \binom{n_2+m_3}{m_3} \cdot \cdots \cdot \binom{n_{\ell-1}+m_\ell}{m_\ell} \\ &= \prod_{i=2}^{\ell} \binom{n_i}{m_i} = \prod_{i=2}^{\ell} \binom{n_i}{n_{i-1}} = \prod_{i=2}^{\ell} \frac{n_i!}{n_{i-1}! m_i!} \\ &= \frac{n!}{n_1! \prod_{i=2}^{\ell} m_i!} = \frac{n!}{\prod_{i=1}^{\ell} m_i!} = |S(\mathcal{M})|.\end{aligned}$$

□

Define the following mapping $\psi : S(\mathcal{M}) \to \mathcal{A}(\mathcal{M})$. For every $\sigma \in S(\mathcal{M})$, $\psi(\sigma)$ is the vector $\mathbf{x} \in \mathcal{A}(\mathcal{M})$, $\mathbf{x} = (\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_\ell)$, where for each $i \in [2, \ell]$, $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,m_i})$, and for each $r \in [m_i]$, $x_{i,r}$ is the number of ranks υ_j, for all $j < i$, which appear to the right of $(\upsilon_i)_r$. That is,

$$x_{i,r} \stackrel{\text{def}}{=} \left|\left\{(j, s) \ : \ \begin{matrix}\sigma^{-1}((\upsilon_j)_s) > \sigma^{-1}((\upsilon_i)_r), \\ j < i, \ s \in [m_i]\end{matrix}\right\}\right|.$$

Note, that for every $i \in [2, \ell]$ and $r \in [m_i]$, we have $x_{i,r} \in [0, n_{i-1}]$. Moreover, if $r < m_i$ then since $(\upsilon_i)_{r+1}$ appears to the right of $(\upsilon_i)_r$ it follows that $x_{i,r} \geq x_{i,r+1}$. Hence, $\mathbf{x}_i \in [0, n_{i-1}]^{m_i}_{\succeq}$ for all $i \in [2, \ell]$ and thus $\mathbf{x} \in \mathcal{A}(\mathcal{M})$, i.e. the mapping ψ is correctly defined.

Example 7: If $\mathcal{M} = \{1^3, 2^2, 3^2\}$, $\sigma = [3, 1, 3, 1, 2, 1, 2]$, and $\psi(\sigma) = \mathbf{x} = (\mathbf{x}_2, \mathbf{x}_3)$, then $\mathcal{A}(\mathcal{M}) = [0, 3]^2_{\succeq} \times [0, 5]^2_{\succeq}$ and $\mathbf{x}_2 = (1, 0)$, since in the multi-permutation σ the rank 1 appears once to the right of 2_1, while it does not appear to the right of 2_2. Similarly, $\mathbf{x}_3 = (5, 4)$, since in the multi-permutation σ there are five elements smaller than 3 to the right of 3_1 and four elements smaller than 3 to the right of 3_2. Thus, $\mathbf{x} = ((1, 0), (5, 4)) \in \mathcal{A}(\mathcal{M})$.

Lemma 5: The mapping ψ is bijective.

Proof: By Lemma 4, we have that $|S(\mathcal{M})| = |\mathcal{A}(\mathcal{M})|$, and hence it is sufficient to prove that the mapping ψ is an injection.

For two distinct multi-permutations $\sigma, \pi \in S(\mathcal{M})$, let $\mathbf{x} = \psi(\sigma)$ and $\mathbf{y} = \psi(\pi)$. Let $b \in [n]$ be the largest integer such that $\sigma(b) \neq \pi(b)$ and let $\sigma(b) = (\upsilon_i)_r$ and $\pi(b) = (\upsilon_j)_s$, where $i, j \in [\ell]$, $i \neq j$, $r \in [m_i]$, and $s \in [m_j]$. Assume w.l.o.g. that $j < i$ and let $c \in [n]$ be such that $\pi(c) = (\upsilon_i)_r$. Since $[\sigma(b+1), \sigma(b+2), \ldots, \sigma(n)] = [\pi(b+1), \pi(b+2), \ldots, \pi(n)]$ it follows that $c < b$ and every rank υ_κ, where $\kappa < i$, that appears to the right of $(\upsilon_i)_r$ in σ, also appears to the right of $(\upsilon_i)_r$ in π. Moreover, the rank $(\upsilon_j)_s$ appears to the right of $(\upsilon_i)_r$ in π, but not in σ. Hence, $y_{i,r} \geq x_{i,r} + 1$. Thus, $\mathbf{x} \neq \mathbf{y}$, which implies that ψ is an injection. □

Remark 1: A mapping similar to ψ was defined in [20] for balanced multi-permutations. Here, we extend it for arbitrary multi-permutations and also we restrict its range such that the mapping is bijective. The importance of knowing the image of the embedding is twofold. The first aspect is that it facilitates encoding. Once the image of the embedding is known, one can encode massages directly to the image, for example by using the enumerative encoding algorithm of Cover [10]. The second aspect is code constructions. By Theorem 2, given in the sequel, it follows that by constructing error-correcting-codes with the Lee distance that have a large intersection with the image of the mapping ψ, one can construct large error-correcting codes in $S(\mathcal{M})$ in the Kendall τ-metric.

The following lemma was proved in [16] for permutations and in [20] for balanced multi-permutations. The generalization of the lemma and its proof for multi-permutations is straightforward.

Lemma 6: For any two multi-permutations $\sigma, \pi \in S(\mathcal{M})$ we have

$$d_L(\psi(\sigma), \psi(\pi)) \leq d_K(\sigma, \pi).$$

The proof of Lemma 6 is based on the observation that the mapping ψ induces an embedding of the graph $G(\mathcal{M})$ into the graphic representation of $\mathbb{Z}_q^{n-m_1}$, where $q > n_{\ell-1}$, in the Lee metric. That is, if $e = \{\sigma, \pi\}$ is an edge in $G(\mathcal{M})$ then $\psi(e) = \{\psi(\sigma), \psi(\pi)\}$ is an edge in the related graph of the space $\mathbb{Z}_q^{n-m_1}$ with the Lee distance. The set $\mathcal{A}(\mathcal{M})$ is a subset of $\mathbb{Z}_q^{n-m_1}$, where $q > n_{\ell-1}$. Hence, $d_L(\psi(\sigma), \psi(\pi)) \leq d_K(\sigma, \pi)$ for every two multi-permutations $\sigma, \pi \in S(\mathcal{M})$. We are now in a position to present the main construction of error-correcting codes in $S(\mathcal{M})$, which is a generalization of the constructions in [1], [16], and [20].

Construction 1: For a code $\mathcal{C}^L \subseteq \mathbb{Z}_q^{n-m_1}$, where $q > n_{\ell-1}$, define the code $\mathcal{C}^K \subseteq S(\mathcal{M})$ as follows.

$$\mathcal{C}^K \stackrel{\text{def}}{=} \{\sigma \in S(\mathcal{M}) \ : \ \psi(\sigma) \in \mathcal{C}^L\}.$$

Theorem 2: If $\mathcal{C}^L \subseteq \mathbb{Z}_q^{n-m_1}$, where $q > n_{\ell-1}$, is a code with minimum Lee distance d then the code $\mathcal{C}^K \subseteq S(\mathcal{M})$ from Construction 1 has minimum Kendall τ-distance at least d and $|\mathcal{C}^K| = |\mathcal{C}^L \cap \mathcal{A}(\mathcal{M})|$.

Proof: By the definition of $\mathcal{C}^K$ and by Lemma 5 it follows that for every two distinct codewords $\sigma, \pi \in \mathcal{C}^K$, their images under the mapping ψ, $\psi(\sigma)$ and $\psi(\pi)$, are distinct codewords of $\mathcal{C}^L$. Since the minimum Lee distance of $\mathcal{C}^L$ is at least d and by Lemma 6 it follows that $d_K(\sigma,\pi) \geq d_L(\psi(\sigma),\psi(\pi)) \geq d$. Hence, the minimum Kendall τ-distance of $\mathcal{C}^K$ is at least d.

By Lemma 5 we have that ψ is a bijection and therefore $|\mathcal{C}^K| = |\mathcal{C}^L \cap \mathcal{A}(\mathcal{M})|$. □

B. Error-Correcting Codes for Multi-Permutations

By Theorem 2, error-correcting codes in $S(\mathcal{M})$ with the Kendall τ-distance can be constructed from error-correcting codes over $\mathbb{Z}_q^{n-m_1}$ with the Lee distance. Next, we present some of the known constructions of error-correcting codes in the Lee metric and use Theorem 2 to obtain error-correcting codes in $S(\mathcal{M})$ and to estimate the size of these codes.

First, we consider single-error-correcting codes in the Lee metric. Golomb and Welch [14] presented the following construction of perfect single-error-correcting codes in the Lee metric.

Construction 2: For every positive integer N and for every $g \in \mathbb{Z}_{2N+1}$, define the code $\mathcal{C}_g^L \subseteq \mathbb{Z}_{2N+1}^N$ as follows.

$$\mathcal{C}_g^L \stackrel{\text{def}}{=} \left\{ \mathbf{x} \in \mathbb{Z}_{2N+1}^N \ : \ \sum_{i=1}^{N} i \cdot x_i \equiv g \ (mod\ 2N+1) \right\}.$$

Theorem 3 [14]: For every positive integer N and for every $g \in \mathbb{Z}_{2N+1}$, the code $\mathcal{C}_g^L$ from Construction 2 is a single-error-correcting code in $\mathbb{Z}_{2N+1}^N$ with the Lee distance.

Construction 2 was used in [16] to obtain single-error-correcting codes for permutations with the Kendall τ-distance. Combining Constructions 1 and 2, we conclude with the following construction.

Construction 3: Assume that $2(n-m_1)+1 > n_{\ell-1}$. For every $g \in \mathbb{Z}_{2(n-m_1)+1}$, let $\mathcal{C}_g^L \subseteq \mathbb{Z}_{2(n-m_1)+1}^{n-m_1}$ be the code from Construction 2. Define the code $\mathcal{C}_g^K \subseteq S(\mathcal{M})$ to be the code that is obtained from Construction 1 by taking $\mathcal{C}^L$ to be the code $\mathcal{C}_g^L$, that is, $\mathcal{C}_g^K = \{\sigma \in S(\mathcal{M}) \ : \ \psi(\sigma) \in \mathcal{C}_g^L\}$.

We finally summarize this discussion with the following corollary.

Corollary 2: If $2(n-m_1)+1 > n_{\ell-1}$, then for every $g \in \mathbb{Z}_{2(n-m_1)+1}$, the code $\mathcal{C}_g^K \subseteq S(\mathcal{M})$ from Construction 3 is a single-error-correcting code in the Kendall τ-metric. Moreover, there exists $g \in \mathbb{Z}_{2(n-m_1)+1}$, for which $|\mathcal{C}_g^K| \geq \frac{|S(\mathcal{M})|}{2(n-m_1)+1}$.

Proof: By Theorems 2 and 3 it follows that for every $g \in \mathbb{Z}_{2(n-m_1)+1}$, the code $\mathcal{C}_g^K$ from Construction 3 is a single-error-correcting code in the Kendall τ-metric. By Lemma 5 it follows that ψ is injective and hence by the pigeon-hole principle there exists $g \in \mathbb{Z}_{2(n-m_1)+1}$ for which $|\mathcal{C}_g^K| \geq \frac{|S(\mathcal{M})|}{2(n-m_1)+1}$. □

Next, we review known constructions of t-error-correcting code in the Lee metric over $\mathbb{Z}_q^N$. The following construction is a variation of codes which were first proposed by Varshamov and Tenengolts [24] (see also [1]) for the correction of a single asymmetric error.

Construction 4: Let $F > N$, $g \in \mathbb{Z}_F$, and let $h_1, h_2, \ldots, h_N$ be integers, $0 < h_i < F$ for all $1 \leq i \leq N$. Assume that for every $\mathbf{e} \in \mathbb{Z}_F^N$ with $w_L(\mathbf{e}) \leq t$, the sums $\sum_{i=1}^{N} e_i \cdot h_i$ are all distinct modulo F. Define the code $\mathcal{C}_{g,t}^L \subseteq \mathbb{Z}_F^N$ as follows.

$$\mathcal{C}_{g,t}^L \stackrel{\text{def}}{=} \left\{ \mathbf{x} \in \mathbb{Z}_F^N \ : \ \sum_{i=1}^{N} x_i \cdot h_i \equiv g \ (mod\ F) \right\}.$$

Theorem 4 [1]: The code $\mathcal{C}_{g,t}^L$ from Construction 4 is a t-error-correcting code in the Lee metric over $\mathbb{Z}_F^N$.

Construction 5: Let $F > \max\{n-m_1, n_{\ell-1}\}$, $g \in \mathbb{Z}_F$, and let $h_1, h_2, \ldots, h_{n-m_1}$ be integers, $0 < h_i < F$ for all $1 \leq i \leq n-m_1$. Assume that for every $\mathbf{e} \in \mathbb{Z}_F^{n-m_1}$ with $w_L(\mathbf{e}) \leq t$, the sums $\sum_{i=1}^{n-m_1} e_i \cdot h_i$ are all distinct modulo F. Let $\mathcal{C}_{g,t}^L \subseteq \mathbb{Z}_F^{n-m_1}$ be the code from Construction 4 that corresponds to these choices of F and h_i's. Define the code $\mathcal{C}_{g,t}^K \subseteq S(\mathcal{M})$ to be the code that is obtained from Construction 1 by taking $\mathcal{C}^L$ to be $\mathcal{C}_{g,t}^L$, that is, $\mathcal{C}_{g,t}^K = \{\sigma \in S(\mathcal{M}) \ : \ \psi(\sigma) \in \mathcal{C}_{g,t}^L\}$.

The following corollary is an immediate consequence of Theorems 2 and 4.

Corollary 3: The code $\mathcal{C}_{g,t}^K \subseteq S(\mathcal{M})$ from Construction 5 is a t-error-correcting code in the Kendall τ-metric.

For every two positive integers M and t let

$$F(M,t) \stackrel{\text{def}}{=} \begin{cases} t(t+1)M, & t \text{ is odd} \\ t(t+2)M, & t \text{ is even.} \end{cases}$$

In order to use Construction 4 we need the following theorem by Barg and Mazumdar [1] which is based on a result of Bose and Chowla [4] for asymmetric error-correcting codes.

Theorem 5 [1]: If q is a power of a prime and $M = (q^{t+1}-1)/(q-1)$ then there exist integers $h_1, h_2, \ldots, h_{q+1}$, $0 < h_i < F(M,t)$ for all $1 \leq i \leq q+1$, such that for all $\mathbf{e} \in \mathbb{Z}_{F(M,t)}^{q+1}$, with $w_L(\mathbf{e}) \leq t$, the sums $\sum_{j=1}^{q+1} e_j h_j$ are all distinct modulo $F(M,t)$.

Construction 4 for t-error-correcting codes in the Lee metric, combined with Theorem 5, was used in [1] to construct t-error-correcting codes for permutations with the Kendall τ-metric, and also used in [20] to construct t-error-correcting codes in the Kendall τ-metric for balanced multi-permutations. By combing Construction 4, Construction 5, and Theorem 5 we have the following corollary.

Corollary 4: For $M = (q^{t+1}-1)/(q-1)$, where $q+1 \geq n-m_1$, q is a power of a prime, and $F(M,t) > n_{\ell-1}$, there exists a t-error-correcting code $\mathcal{C} \subseteq S(\mathcal{M})$ in the Kendall τ-metric, whose size satisfies $|\mathcal{C}| \geq \frac{|S(\mathcal{M})|}{F(M,t)}$.

Proof: For $F = F(M,t)$, it follows by Theorems 4 and 5 that there exist integers $h_1, h_2, \ldots, h_{n-m_1}$, $0 < h_i < F$ for all $1 \leq i \leq n-m_1$, such that for every $g \in \mathbb{Z}_F$, the code $\mathcal{C}_{g,t}^L \subseteq \mathbb{Z}_F^{n-m_1}$ from Construction 4 that corresponds to these choices of F and h_i's is a t-error-correcting code in the Lee metric. Since $F > n_{\ell-1}$ and by Corollary 3, it follows that for every $g \in \mathbb{Z}_F$, the corresponding code $\mathcal{C}_{g,t}^K \subseteq S(\mathcal{M})$ from Construction 5 is a t-error-correcting code in the Kendall τ-metric. By Lemma 5, it follows that ψ is injective and hence by the pigeon-hole principle there exists $g \in \mathbb{Z}_F$ for which $|\mathcal{C}^K_{g,t}| \geq \frac{|S(\mathcal{M})|}{F}$. □

C. Partitions Into Error-Correcting Codes

In this section we discuss partitions of $S(\mathcal{M})$, and in particular of S_k, into error-correcting codes with the Kendall τ-distance. These partitions will be derived from partitions into codes with the Lee distance. The partitions will be used later as the first ingredient of the construction presented in Theorem 1 to produce systematic error-correcting codes for permutations and multi-permutations.

Constructions 2 and 3 can be used to partition $S(\mathcal{M})$, and in particular S_k, into single-error-correcting codes with the Kendall τ-metric.

Theorem 6: If $2(n-m_1)+1 > n_{\ell-1}$ then there exists a partition of $S(\mathcal{M})$ into at most $2(n-m_1)+1$ single-error-correcting codes in the Kendall τ-metric.

Proof: For every $g \in \mathbb{Z}_{2(n-m_1)+1}$, it follows from Theorem 3 that the code $\mathcal{C}^L_g \subseteq \mathbb{Z}^{n-m_1}_{2(n-m_1)+1}$ from Construction 2 is a single-error-correcting code in the Lee metric. By Corollary 2 it follows that for every $g \in \mathbb{Z}_{2(n-m_1)+1}$, the code $\mathcal{C}^K_g \subseteq S(\mathcal{M})$ from Construction 3 is a single-error-correcting code in the Kendall τ-metric.

The set $\{\mathcal{C}^L_g : g \in \mathbb{Z}_{2(n-m_1)+1}\}$ forms a partition of $\mathbb{Z}^{n-m_1}_{2(n-m_1)+1}$ into single-error-correcting codes in the Lee metric. By Lemma 5 it follows that ψ is an injection and therefore the set $\{\mathcal{C}^K_g : g \in \mathbb{Z}_{2(n-m_1)+1}\}$ forms a partition of $S(\mathcal{M})$ into single-error-correcting codes in the Kendall τ-metric. □

Construction 4, Construction 5, and Theorem 5 provide us with partitions of $S(\mathcal{M})$, and in particular S_k, into t-error-correcting codes in the Kendall τ-metric.

Theorem 7: For $M = (q^{t+1}-1)/(q-1)$, where $q+1 \geq n-m_1$, q is a power of a prime, and $F(M,t) > n_{\ell-1}$, there exists a partition of $S(\mathcal{M})$ into at most $F(M,t)$ t-error-correcting codes in the Kendall τ-metric.

Proof: For $F = F(M,t)$, it follows by Theorems 4 and 5 that there exist integers $h_1, h_2, \ldots, h_{n-m_1}$, $0 < h_i < F$ for all $1 \leq i \leq n-m_1$, such that for every $g \in \mathbb{Z}_F$, the code $\mathcal{C}^L_{g,t} \subseteq \mathbb{Z}^{n-m_1}_F$ from Construction 4 that corresponds to these choices of F and h_i's is a t-error-correcting code in the Lee metric. Since $F > n_{\ell-1}$ and by Corollary 3, it follows that for every $g \in \mathbb{Z}_F$, the corresponding code $\mathcal{C}^K_{g,t} \subseteq S(\mathcal{M})$ from Construction 5 is a t-error-correcting code in the Kendall τ-metric.

The set $\{\mathcal{C}^L_{g,t} : g \in \mathbb{Z}_F\}$ forms a partition of $\mathbb{Z}^{n-m_1}_F$ into t-error-correcting codes in the Lee metric. By Lemma 5 it follows that ψ is an injection and therefore the set $\{\mathcal{C}^K_{g,t} : g \in \mathbb{Z}_F\}$ forms a partition of $S(\mathcal{M})$ into t-error-correcting codes in the Kendall τ-metric. □

VI. Constructions of Systematic Error-Correcting Codes for Permutations

In this section we construct (n,k) systematic t-error-correcting codes for permutations. We distinguish between three cases for the value of t, namely $t=1$, fixed t, and $t = \Theta(k^\epsilon)$ where $\epsilon > 0$. In all three cases we apply Theorem 1 with its two ingredients of partitions and multi-permutation codes. For the first ingredient of the partition of S_k we use the results from Section V-C. For the cases where $t=1$ and t is fixed, we provide explicit constructions of multi-permutation codes as the second ingredient. Lastly, for $t = \Theta(k^\epsilon)$ we use the multi-permutation codes from Corollary 4.

We first construct systematic single-error-correcting codes. To this end we need the following simple observations.

Construction 6: For a code $\mathcal{C} \subseteq S(\mathcal{M})$ and a multi-permutation $\gamma \in S(\mathcal{M})$, define the codes $\mathcal{C}^e_\gamma, \mathcal{C}^o_\gamma \subseteq \mathcal{C}$ as follows.

$$\mathcal{C}^e_\gamma \stackrel{\text{def}}{=} \{\sigma \in \mathcal{C} : d_K(\sigma,\gamma) \equiv 0 \ (mod\ 2)\} \ \ and$$
$$\mathcal{C}^o_\gamma \stackrel{\text{def}}{=} \{\sigma \in \mathcal{C} : d_K(\sigma,\gamma) \equiv 1 \ (mod\ 2)\}.$$

Theorem 8: If $\mathcal{C} \subseteq S(\mathcal{M})$ is a code with minimum Kendall τ-distance $2t+1$, for some $t \geq 0$, then for every multi-permutation $\gamma \in S(\mathcal{M})$, the codes $\mathcal{C}^e_\gamma$ and $\mathcal{C}^o_\gamma$ from Construction 6 have minimum Kendall τ-distance at least $2t+2$ and $\max\{|\mathcal{C}^e_\gamma|, |\mathcal{C}^o_\gamma|\} \geq \frac{|\mathcal{C}|}{2}$.

Proof: Lemma 2 implies that for every $\gamma \in S(\mathcal{M})$, the minimum Kendall τ-distance of $\mathcal{C}^e_\gamma$ and $\mathcal{C}^o_\gamma$ is even and since the minimum distance of $\mathcal{C}$ is $2t+1$ it follows that the minimum distance of both $\mathcal{C}^e_\gamma$ and $\mathcal{C}^o_\gamma$ is at least $2t+2$. Clearly, the size of $\mathcal{C}^e_\gamma$ or $\mathcal{C}^o_\gamma$ is at least $\frac{|\mathcal{C}|}{2}$ and the lemma follows. □

Corollary 5: There exists a code in $S(\mathcal{M})$ with minimum Kendall τ-distance 2 and of size at least $\frac{|S(\mathcal{M})|}{2}$.

Theorem 9: For every integer $k \geq 1$, there exists a $(k+2,k)$ systematic single-error-correcting code.

Proof: Since $2(k-1)+1 > k-1$ and by Theorem 6, there exists a partition of S_k into at most $2(k-1)+1$ single-error-correcting codes in the Kendall τ-metric. By Corollary 5, there exists a code in $S(\mathcal{M}_{k,2})$ with minimum distance 2 and of size at least $\frac{|S(\mathcal{M}_{k,2})|}{2}$. For all $k \geq 1$, we have that $\frac{|S(\mathcal{M}_{k,2})|}{2} = \frac{(k+2)(k+1)}{2} \geq 2(k-1)+1$ and hence by Theorem 1 it follows that there exists a $(k+2,k)$ systematic single-error-correcting code. □

Next, we construct $(k+t+1,k)$ systematic t-error-correcting codes, where t is a fixed integer and k is sufficiently large. For this task we need the following construction of multi-permutation codes in $S(\mathcal{M}_{k,r})$ of minimum Kendall τ-distance $2t$.

Construction 7: For all positive integers k, r, and t, define the code $\mathcal{C}_{k,r,t} \subseteq S(\mathcal{M}_{k,r})$ as follows.

$$\mathcal{C}_{k,r,t} \stackrel{\text{def}}{=} \left\{ \sigma \in S(\mathcal{M}_{k,r}) \ : \begin{array}{c} \sigma(j) = 0 \text{ for all } j \in [k+r] \\ \text{such that } j \not\equiv 1 \ (mod\ 2t) \end{array} \right\}.$$

Theorem 10: The code $\mathcal{C}_{k,r,t}$ from Construction 7 has minimum Kendall τ-distance at least $2t$ and size $\binom{\lceil \frac{k+r}{2t} \rceil}{r} r!$.

Proof: For two distinct codewords $\sigma, \pi \in \mathcal{C}_{k,r,t}$, there exists $j \equiv 1 \pmod{2t}$ such that $\sigma(j) \neq \pi(j)$. Assume w.l.o.g. that $\sigma(j) \neq 0$. Since nonzero elements appear only in positions which are congruent to 1 modulo $2t$ and since $\pi(j) \neq \sigma(j)$ it follows that $|\pi^{-1}(\sigma(j)) - j| \geq 2t$. Any sequence of adjacent transpositions that transfer σ to π must exchange $\sigma(j)$ at least $|\pi^{-1}(\sigma(j)) - j|$ times. Therefore, at least $2t$ adjacent transpositions are required to transfer σ to π. Hence, the minimum distance of $\mathcal{C}_{k,r,t}$ is at least $2t$.

The size of $\mathcal{C}_{k,r,t}$ is $\binom{\lceil \frac{k+r}{2t} \rceil}{r} r!$, since there are $\lceil \frac{k+r}{2t} \rceil$ positions which are congruent to 1 modulo $2t$, and there are $\binom{\lceil \frac{k+r}{2t} \rceil}{r} r!$ distinct ways to distribute the r distinct nonzero elements $k+1, k+2, \ldots, k+r$, in these positions. □

Theorem 11: For a fixed positive integer t and for sufficiently large k, there exists a $(k+t+1, k)$ systematic t-error-correcting code.

Proof: There exists a power of a prime q (e.g. a power of 2) such that $k-2 \leq q \leq 2k$. If $M = (q^{t+1}-1)/(q-1)$ then $F(M,t) \geq t(t+1)M \geq 2(q+1) \geq 2(k-1) > k-1$. By Theorem 7, it follows that there exists a partition of S_k into at most $F(M,t)$ t-error-correcting codes in the Kendall τ-metric. By Theorem 10, it follows that the code $\mathcal{C}_{k,r,t}$ from Construction 7, where $r = t+1$, is a code with minimum Kendall τ-distance $2t$ and of size $\binom{\lceil \frac{k+t+1}{2t} \rceil}{t+1}(t+1)!$. Since $\binom{\lceil \frac{k+t+1}{2t} \rceil}{t+1}(t+1)! = \prod_{i=0}^{t}(\lceil \frac{k+t+1}{2t} \rceil - i)$ it follows that

$$\binom{\lceil \frac{k+t+1}{2t} \rceil}{t+1}(t+1)! \geq \left(\frac{k+t+1}{2t} - t\right)^{t+1}.$$

Since t is fixed, it follows that for sufficiently large k we have that

$$\left(\frac{k+t+1}{2t} - t\right)^{t+1} \geq t(t+2)2^{t+1}k^t. \quad (3)$$

For every $x \geq 2$ we have that $(x^{t+1}-1)/(x-1) \leq 2x^t$. Hence, $M \leq 2q^t \leq 2^{t+1}k^t$ and therefore by (3) it follows that

$$|\mathcal{C}_{k,r,t}| \geq t(t+2)2^{t+1}k^t \geq t(t+2)M \geq F(M,t).$$

By Theorem 1 we conclude that there exists a $(k+t+1,k)$ systematic t-error-correcting code. □

In the next theorem we analyze the number of redundancy symbols in an (n,k) systematic t-error-correcting code, where $t = \Theta(k^\epsilon)$ and $\epsilon \geq 0$. The proof is given in Appendix A.

Theorem 12: Let $k \geq 1$ be an integer, $t = \Theta(k^\epsilon)$ be a positive integer, and $r = \lceil \mu t \rceil$, such that $r-1$ is a power of a prime and

$$\mu > \begin{cases} 1+\epsilon & \text{for } 0 \leq \epsilon \leq 1 \\ 1+\frac{1}{\epsilon} & \text{for } 1 < \epsilon. \end{cases}$$

Then, for sufficiently large k there exists a $(k+r,k)$ systematic t-error-correcting code.

VII. Systematic ECC for Multi-Permutations

In this section we generalize the construction in Theorem 1 to obtain systematic error-correcting codes for multi-permutations. The first question that we face is how to define systematic error-correcting codes over multi-permutations? In the most general definition we have a multi-set $\mathcal{K}$ of size k with information symbols and a multi-set $\mathcal{R}$ of size r with redundancy symbols.[1] The intersection between $\mathcal{K}$ and $\mathcal{R}$ must be empty, i.e. $\mathcal{K}$ and $\mathcal{R}$ do not have common symbols. The codewords are multi-permutations over the multi-set[2] $\mathcal{K} \cup \mathcal{R}$. The number of codewords must be the number of distinct multi-permutations over the multi-set $\mathcal{K}$. In a $(\mathcal{K} \cup \mathcal{R}, \mathcal{K})$ *systematic code* $\mathcal{C}$ each multi-permutation over the multi-set $\mathcal{K}$ appears exactly once as a sub-multi-permutation of a codeword from $\mathcal{C}$, which consists exactly from all the symbols of $\mathcal{K}$. Note, that $\mathcal{K}$ might be a set, which implies that multi-permutations over $\mathcal{K}$ are simply permutations in S_k. The construction for systematic multi-permutations will be a direct generalization of the construction in Theorem 1. Instead of the multi-set $\mathcal{M}_{k,r}$ we use the multi-set $\mathcal{M}_{k,\mathcal{R}}$ defined by $\mathcal{M} \stackrel{\text{def}}{=} \{0^k\} \cup \mathcal{R}$, where $0 \notin \mathcal{R}$.

For two multi-permutations $\sigma \in S(\mathcal{K})$, $\rho \in S(\mathcal{M}_{k,\mathcal{R}})$, the *assignment* of σ in ρ results with the multi-permutation $\alpha = \sigma \blacktriangleright \rho \in S(\mathcal{K} \cup \mathcal{R})$ obtained by substituting σ, in its order, in the k positions in which 0 appears in ρ.

Example 8: If $\mathcal{K} = \{1^2, 2^2, 3\}$, $\mathcal{R} = \{4^2, 5^3\}$, $\rho = [0,4,5,0,0,5,0,4,0,5]$, and $\sigma = [1,3,2,2,1]$ then $\mathcal{K} \cup \mathcal{R} = \{1^2, 2^2, 3, 4^2, 5^3\}$ and $\sigma \blacktriangleright \rho = [1,4,5,3,2,5,2,4,1,5] \in S(\mathcal{K} \cup \mathcal{R})$.

The generalization of the construction in Theorem 1 is described in the next theorem, which is proved along the same lines as Theorem 1.

Theorem 13: Let $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_F$ be a partition of $S(\mathcal{K})$ into t-error-correcting codes in the Kendall τ-metric, and let $\mathcal{C}_{mult} \subseteq S(\mathcal{M}_{k,\mathcal{R}})$ be a code with minimum Kendall τ-distance $2t$ and of size at least F. Let $\rho_1, \rho_2, \ldots, \rho_F$ be distinct elements in $\mathcal{C}_{mult}$. Then the code $\mathcal{C} \subseteq S(\mathcal{K} \cup \mathcal{R})$ defined by

$$\mathcal{C} = \cup_{j=1}^{F} \{\sigma \blacktriangleright \rho_j \ : \ \sigma \in \mathcal{C}_j\}$$

is a $(\mathcal{K} \cup \mathcal{R}, \mathcal{K})$ systematic t-error-correcting code in the Kendall τ-metric.

As for permutations, the challenge for systematic multi-permutation codes is to minimize the number of redundancy symbols of the codes. However, for systematic error-correcting codes for multi-permutations there is a tradeoff between the number of the redundancy ranks and the magnitudes of their multiplicities. For example, in a $(\mathcal{K} \cup \mathcal{R}, \mathcal{K})$ systematic code for multi-permutations, where $\mathcal{R}$ has only one redundancy rank υ, the multiplicity of υ might be large. However, if $\mathcal{R}$ has two redundancy ranks then the sum of their multiplicities should be smaller. The construction in Theorem 13 allows any desirable number of redundancy ranks.

Although Theorem 13 can be applied for every multi-set $\mathcal{K}$ and for various choices of the number of redundancy ranks and their multiplicities, we will apply it only for $\mathcal{K}$ and $\mathcal{R}$ such that $\mathcal{K}$ and $\mathcal{K} \cup \mathcal{R}$ are both balanced multi-sets. Hence, w.l.o.g. we assume in the rest of this section that $\mathcal{K} = \{1^m, 2^m, \ldots, \ell^m\}$ and $\mathcal{R} = \{(\ell+1)^m, (\ell+2)^m, \ldots, (\ell+\varrho)^m\}$, which implies that $k = \ell m$. We also define for every three positive integers m, ℓ, ϱ the set $\mathcal{M}_{m,\ell,\varrho} = \{0^k\} \cup \{(\ell+1)^m, (\ell+2)^m, \ldots, (\ell+\varrho)^m\}$. Note, that for balanced multi-permutations $\mathcal{M}_{k,\mathcal{R}} = \mathcal{M}_{m,\ell,\varrho}$. Furthermore, $\mathcal{M}_{m,\ell,\varrho}$ is a generalization

[1] The size of a multi-set refers to the total number of symbols, including repetitions.

[2] The union $\mathcal{K} \cup \mathcal{R}$ of the multi-sets $\mathcal{K}$ and $\mathcal{R}$ is again a multi-set. If υ is a rank in $\mathcal{K}$ or $\mathcal{R}$ with multiplicity m then, since $\mathcal{K}$ and $\mathcal{R}$ do have a rank in common, υ is a rank in $\mathcal{K} \cup \mathcal{R}$ of multiplicity m.

of the multi-set $\mathcal{M}_{k,r}$, which is the same multi-set as $\mathcal{M}_{1,k,r}$.

Theorems 9, 11, and 12 can be generalized for balanced multi-permutations assuming that the multiplicity m is fixed and the number of information ranks ℓ is sufficiently large. In the next example we will demonstrate the extension of Theorem 9 for multi-permutations with multiplicity 2.

Example 9: Let $\mathcal{K} = \{1^2, 2^2, \ldots, \ell^2\}$ *be a multi-set which consists of* $k = 2\ell$ *information symbols,* $\mathcal{R} = \{\ell+1, \ell+1\}$*, and* $M = 2(k-2)+1$*. Then* $\mathcal{M}_{k,\mathcal{R}} = \mathcal{M}_{2,\ell,1} = \{0^k, \ell+1, \ell+1\}$*. Since* $2(k-2)+1 > k-2$ *and by Theorem 6, there exists a partition of* $S(\mathcal{K})$ *into at most* $2(k-2)+1$ *single-error-correcting codes in the Kendall* τ*-metric.*

By Corollary 5, there exists a code in $S(\mathcal{M}_{2,\ell,1})$ *with minimum distance* 2 *and of size at least* $\frac{|S(\mathcal{M}_{2,\ell,1})|}{2} = (k+2)(k+1)/4$*. For all* $k \geq 1$*, we have that* $\frac{(k+2)(k+1)}{4} \geq 2(k-2)+1$ *and hence, by Theorem 13, there exists a* $(\mathcal{K}\cup\mathcal{R}, \mathcal{K})$ *systematic single-error-correcting code.*

Next, we will show the generalization of Theorem 11 for balanced multi-permutations. For this purpose, we will first present an extension of Construction 7.

Construction 8: For every positive integers m, ℓ, ϱ*, and* t *define the code* $\mathcal{C}_{m,\ell,\varrho,t} \subseteq S(\mathcal{M}_{m,\ell,\varrho})$ *as follows.*

$$\mathcal{C}_{m,\ell,\varrho,t} \stackrel{\text{def}}{=} \left\{\sigma \in S(\mathcal{M}_{m,\ell,\varrho}) \ : \ \begin{array}{c}\sigma(j)=0, \ \forall\ j \in [m(\ell+\varrho)] \\ \textit{such that } j \not\equiv 1 \ (\textit{mod } 2t)\end{array}\right\}.$$

The next theorem is proved similarly to the proof of Theorem 10.

Theorem 14: The code $\mathcal{C}_{m,\ell,\varrho,t}$ *from Construction 8 has minimum distance* $2t$ *and size* $\binom{\lceil \frac{m(\ell+\varrho)}{2t} \rceil}{m\varrho}\frac{(m\varrho)!}{(m!)^{\varrho}}$.

Theorem 15: For three positive integers t, m, *and* ϱ *and for* $\mathcal{K} = \{1^m, 2^m, \ldots, \ell^m\}$ *and* $\mathcal{R} = \{(\ell+1)^m, (\ell+2)^m, \ldots, (\ell+\varrho)^m\}$*, if* ℓ *is large enough and* $m\varrho \geq t+1$ *then there exists a* $(\mathcal{K}\cup\mathcal{R}, \mathcal{K})$ *systematic* t*-error-correcting code.*

Proof: Let k and r be the size of the multi-sets $\mathcal{K}$ and $\mathcal{R}$, respectively, i.e. $k = m\ell$ and $r = m\varrho$. There exists a power of a prime q (e.g. a power of 2) such that $k-m-1 \leq q \leq 2k$. If $M = (q^{t+1}-1)/(q-1)$ then $F(M,t) \geq 2(k-m) = 2m(\ell-1)$ and since m and ϱ are fixed, it follows that $F(M,t) > m(\ell+\varrho-1)$, for sufficiently large ℓ. Hence, by Theorem 7, there exists a partition of $S(\mathcal{K})$ into at most $F(M,t)$ t-error-correcting codes in the Kendall τ-metric. By Theorem 14, it follows that the code $\mathcal{C}_{m,\ell,\varrho,t}$ from Construction 8 is a code with minimum Kendall τ-distance $2t$ and of size $\binom{\lceil \frac{k+r}{2t} \rceil}{r}\frac{r!}{(m!)^{\varrho}}$. Since $\binom{\lceil \frac{k+r}{2t} \rceil}{r}r! = \prod_{i=0}^{r-1}(\lceil \frac{k+r}{2t} \rceil - i)$, it follows that

$$\binom{\lceil \frac{k+r}{2t} \rceil}{r}r! \geq \left(\frac{k+r}{2t} - (r-1)\right)^r. \qquad (4)$$

Note that $M \leq 2q^t \leq 2(2k)^t$ and $F(M,t) \leq t(t+2)M \leq 2t(t+2)(2k)^t$. Since t, r, m, and ϱ are fixed and $r \geq t+1$, it follows that for sufficiently large ℓ we have that

$$\frac{((k+r)/2t-(r-1))^r}{(m!)^{\varrho}} \geq 2t(t+2)(2k)^t \geq F(M,t). \qquad (5)$$

Combining (4) and (5) we conclude that $|\mathcal{C}_{m,\ell,\varrho,t}| \geq F(M,t)$ for sufficiently large ℓ. Since, $\mathcal{C}_{m,\ell,\varrho,t}$ is a code of size at least $F(M,t)$ and with minimum Kendall τ-distance $2t$, it follows by Theorem 13 that there exists a $(\mathcal{K}\cup\mathcal{R}, \mathcal{K})$ systematic t-error-correcting code. □

VIII. A Lower Bound on the Number of Redundancy Symbols

In this section we will present an asymptotic lower bound on the number of redundancy symbols in a $(\mathcal{K}\cup\mathcal{R}, \mathcal{K})$ systematic t-error-correcting code, where t is fixed.

For a multi-set $\mathcal{M}$ and for a multi-permutation $\sigma \in S(\mathcal{M})$, the *ball* of radius t centered at σ is the set

$$\mathcal{B}(\sigma,t) = \{\pi \in S(\mathcal{M}) \ : \ d_K(\sigma,\pi) \leq t\}.$$

Note, that the size of the ball $\mathcal{B}(\sigma,t)$ depends on the choice of its center σ. A sphere packing upper bound on the size of a t-error-correcting code in $S(\mathcal{M})$ is presented in the next lemma.

Lemma 7: If $\mathcal{C} \subseteq S(\mathcal{M})$ *is a* t*-error-correcting code in the Kendall* τ*-metric then*

$$|\mathcal{C}| \leq \frac{|S(\mathcal{M})|}{\min_{\sigma\in S(\mathcal{M})}\{|\mathcal{B}(\sigma,t)|\}}.$$

In order to apply the upper bound from Lemma 7 we need a lower bound on the size of a ball of radius t in $S(\mathcal{M})$.

Lemma 8: For two integers ℓ *and* t, $\ell > t \geq 1$*, for a multi-set* $\mathcal{M}$ *with* ℓ *ranks, and for a multi-permutation* $\sigma \in \mathcal{M}$ *we have*

$$\mathcal{B}(\sigma,t) \geq \binom{\ell}{t}.$$

Proof: Assume w.l.o.g. that for every $i \in [\ell]$, there does not exist $j > i$ such that $(v_i)_1$ appears to the right of $(v_j)_1$ in σ (otherwise, the ranks of $\mathcal{M}$ can be relabelled such that this assumption will hold for the multi-permutation corresponding to the relabeling of σ). From this assumption it follows that $\sigma(1) = v_1$ and for every $i \in [2,\ell]$, if $\sigma(\kappa) = (v_i)_1$ then $\sigma(\kappa-1) = v_j$ for some $1 \leq j < i$.

Denote by $\mathcal{B}$ the set of all binary vectors $(b_1, b_2, \ldots, b_{\ell-1}) \in \{0,1\}^{\ell-1}$, such that $\sum_{i=1}^{\ell-1} b_i \leq t$. The size of $\mathcal{B}$ is given by the expression $\sum_{w=0}^{t}\binom{\ell-1}{w}$. In particular,

$$|\mathcal{B}| \geq \binom{\ell-1}{t} + \binom{\ell-1}{t-1} = \binom{\ell}{t}.$$

For every $\mathbf{b} = (b_1, b_2, \ldots, b_{\ell-1}) \in \mathcal{B}$, let $Supp(\mathbf{b}) = \{i \in [\ell-1] \ : \ b_i = 1\}$. By the definition of $\mathcal{B}$ it follows that $|Supp(\mathbf{b})| \leq t$ for every $\mathbf{b} \in \mathcal{B}$. Define the mapping $\phi : \mathcal{B} \to \mathcal{B}(\sigma,t)$ as follows. Given a vector $\mathbf{b} \in \mathcal{B}$, if $1 \leq i_1 < i_2 < \cdots < i_w \leq \ell-1$ are the elements of $Supp(\mathbf{b})$ then for every $s \in [w]$, let $\kappa_s = \sigma^{-1}((v_{i_s+1})_1)$. By the assumption on σ we have that $\kappa_s \geq 2$ and $\sigma(\kappa_s-1) = v_j$ for some $1 \leq j < i_s+1$. Let τ_s be the adjacent transposition that exchanges the elements in positions κ_s-1 and κ_s. Define $\phi(\mathbf{b})$ to be the multi-permutation obtain from σ by applying the sequence of adjacent transpositions $\tau_1, \tau_2, \ldots, \tau_w$.[3] Since $d_K(\sigma, \phi(\mathbf{b})) \leq w \leq t$ it follows that $\phi(\mathbf{b}) \in \mathcal{B}(\sigma, t)$, i.e. ϕ is correctly defined.

By the assumption on σ it follows that if $\mathbf{x} = \psi(\sigma)$, $\mathbf{x} = (\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_\ell)$, where $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,m_i})$ for all $i \in [\ell]$, and $\mathbf{y} = \psi(\phi(\mathbf{b}))$, $\mathbf{y} = (\mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_\ell)$, where $\mathbf{y}_i = (y_{i,1}, y_{i,2}, \ldots, y_{i,m_i})$ for all $i \in [\ell]$, then

$$y_{i,r} = \begin{cases} x_{i,1} + b_{i-1}, & r = 1 \\ x_{i,r}, & \text{otherwise,} \end{cases}$$

for all $i \in [\ell]$, $r \in [m_i]$. This is because $\phi(\mathbf{b})$ is obtained from σ by w adjacent transpositions that exchange each of the w ranks of the form $(v_i)_1$, where $i \in [2, \ell]$ and $b_{i-1} = 1$, with a rank v_j that appears to the left of $(v_i)_1$, where $j < i$. Therefore, the number of elements smaller than $(v_i)_r$ that appear to the right of $(v_i)_r$ is increased by one if $r = 1$ and $b_{i-1} = 1$, and remains unchanged otherwise.

Hence, $\mathbf{b}$ is uniquely determined from σ and $\phi(\mathbf{b})$, using the mapping ψ, and therefore ϕ is an injection. Thus, $|\mathcal{B}(\sigma, t)| \geq |\mathcal{B}| \geq \binom{\ell}{t}$. □

We can now apply Lemmas 7 and 8 to derive a lower bound on the number of redundancy symbols.

Theorem 16: For two fixed positive integers t and m, if $\mathcal{K} = \{v_1^{m_1}, v_2^{m_2}, \ldots, v_\ell^{m_\ell}\}$ is a multi-set with ℓ ranks, where ℓ is sufficiently large and $m_i \leq m$ for all $i \in [\ell]$ then every $(\mathcal{K} \cup \mathcal{R}, \mathcal{K})$ systematic t-error-correcting code uses at least t redundancy symbols.

Proof: Let $\mathcal{C}$ be a $(\mathcal{K} \cup \mathcal{R}, \mathcal{K})$ systematic t-error-correcting code and let k and r be the number of information and redundancy symbols, respectively. We have

$$\begin{aligned} |\mathcal{C}| &\leq^{(a)} \frac{|S(\mathcal{K} \cup \mathcal{R})|}{\binom{\ell}{t}} \leq^{(b)} \frac{(k+r)!}{\prod_{i=1}^{\ell} m_i! \frac{(\ell-t)^t}{t!}} \\ &\leq^{(c)} \frac{k!(k+r)^r}{\prod_{i=1}^{\ell} m_i! \frac{(\ell-t)^t}{t!}} = \frac{t!(k+r)^r}{(\ell-t)^t} |S(\mathcal{K})| \\ &\leq \frac{t!(m\ell+r)^r}{(\ell-t)^t} |S(\mathcal{K})| \end{aligned}$$

where inequality (a) follows from Lemmas 7 and 8, inequality (b) follows from $\binom{\ell}{t} \geq \frac{(\ell-t)^t}{t!}$ and $|S(\mathcal{K} \cup \mathcal{R})| \leq \frac{(k+r)!}{\prod_{i=1}^{\ell} m_i!}$, and inequality (c) follows from $\prod_{i=1}^{r}(k+i) \leq (k+r)^r$. If $r \leq t-1$ then since m and t are fixed, it follows that $t!(m\ell+r)^r$ is a polynomial in ℓ of degree at most $t-1$, while $(\ell-t)^t$ is a polynomial in ℓ of degree t. Hence, $\frac{t!(m\ell+r)^r}{(\ell-t)^t} < 1$ for sufficiently large ℓ, and therefore $|\mathcal{C}| < |S(\mathcal{K})|$, a contradiction to the assumption that $\mathcal{C}$ is a $(\mathcal{K} \cup \mathcal{R}, \mathcal{K})$ systematic. □

Theorem 16 implies that for every balanced multi-set $\mathcal{K} = \{1^m, 2^m, \ldots, \ell^m\}$, where ℓ is large enough, at least t redundancy symbols are needed in order to construct a $(\mathcal{K} \cup \mathcal{R}, \mathcal{K})$ systematic t-error-correcting code. On the other hand, Theorem 15 states that if ℓ is large enough then our method can be used to construct a $(\mathcal{K} \cup \mathcal{R}, \mathcal{K})$ systematic t-error-correcting code with only $t+1$ redundancy symbols (where $\mathcal{K} \cup \mathcal{R}$ is also a balanced multi-set and m divides $t+1$).

[3]For every $s \in [w-1]$, we apply the adjacent transposition τ_s before the adjacent transposition τ_{s+1}.

The analysis conducted in the proof of Theorem 15 is also valid when the multiplicities of the ranks in $\mathcal{K}$ are bounded by a fixed integer m, whereas the number of information ranks is sufficiently large. In this case the minimum number of redundancy symbols of a $(\mathcal{K} \cup \mathcal{R}, \mathcal{K})$ systematic t-error-correcting code is t or $t+1$.

The bound from Theorem 16 holds also for permutations. That is, the number of redundancy symbols in an (n, k) systematic t-error-correcting code is at least t. As in the multi-permutations case, Theorem 11 provides a code with $t+1$ redundancy symbols, when the number of information symbols is sufficiently large. For systematic single-error-correcting codes, the code from Theorem 9 uses two redundancy symbols. If the number of information symbols k equals $p-1$, for some prime p, then the code construction from Theorem 9 is optimal. This observation is concluded from the result from [6], which implies that the size of a single-error-correcting code in S_{k+1} with the Kendall τ-metric is less than $k!$. Hence, there does not exist a $(k+1, k)$ systematic single-error-correcting code. This observation verified that if $k+1$ is a prime, Constructions A and B from [27] are also optimal.

IX. Conclusion

We have considered constructions of systematic error-correcting codes over permutations and multi-permutations with the Kendall τ-distance. The constructions are based on error-correcting codes for multi-permutations. The main result is that for a large enough integer k, a positive integer $t = \Theta(k^\epsilon)$, and $r = \lceil \mu t \rceil$, such that $r-1$ is a power of a prime, there exists a $(k+r, k)$ systematic t-error-correcting code if

$$\mu > \begin{cases} 1+\epsilon & \text{for } 0 \leq \epsilon \leq 1 \\ 1+\frac{1}{\epsilon} & \text{for } 1 < \epsilon. \end{cases}$$

In case that t is fixed, then our construction uses $r = t+1$ redundancy symbols for k sufficiently large, while a lower bound on the number of redundancy symbols is shown to be t.

Appendix A

The goal of this appendix is to prove Theorem 12, i.e. to show that for a large enough integer k, a positive integer $t = \Theta(k^\epsilon)$, and $r = \lceil \mu t \rceil$, such that $r-1$ is a power of a prime, there exists a $(k+r, k)$-systematic t-error-correcting code if

$$\mu > \begin{cases} 1+\epsilon & \text{for } 0 \leq \epsilon \leq 1 \\ 1+\frac{1}{\epsilon} & \text{for } 1 < \epsilon. \end{cases}$$

Proof of Theorem 12: The case where t is fixed, i.e. $\epsilon = 0$, is an immediate consequence of Theorem 11. Henceforth, we will assume that $\epsilon > 0$.

There exists a power of a prime q such that $k-2 \leq q \leq 2k$. If $M = (q^{t+1}-1)/(q-1)$ then $F(M, t) \geq 2M \geq 2(q+1) \geq 2(k-1) > k-1$. Hence, by Theorem 7, there exists a partition of S_k into at most $F(M, t)$ t-error-correcting codes in the Kendall τ-metric. We will show that for sufficiently large k there exists a code in $S(\mathcal{M}_{k,r})$ with minimum Kendall τ-distance at least $2t$ and of size at least $F(M, t)$.

Hence, by Theorem 1 we will conclude the existence of a $(k+r,k)$ systematic t-error-correcting code.

If $M_{mult} = ((r-1)^{t+1}-1)/(r-2)$ then $F(M_{mult},t) \geq r^t \geq \mu^t t^t \geq c_1^t k^{c_2 k^{\epsilon}}$, for some constants c_1, c_2 and for sufficiently large k. Since $\epsilon > 0$ and μ is fixed, it follows that $F(M_{mult},t) > k+r-1$, for a sufficiently large k. Since $r-1$ is a power of a prime and by Corollary 4, it follows that there exists a t-error-correcting code $\mathcal{C}_{mult} \subseteq S(\mathcal{M}_{k,r})$ in the Kendall τ-metric, whose size satisfies $|\mathcal{C}_{mult}| \geq \frac{|S(\mathcal{M}_{k,r})|}{F(M_{mult},t)}$. We will show that for large enough k,

$$\frac{|S(\mathcal{M}_{k,r})|}{F(M_{mult},t)} \geq F(M,t) \tag{A.1}$$

and conclude that $|\mathcal{C}_{mult}| \geq F(M,t)$.

For every $x \geq 2$ we have that $(x^{t+1}-1)/(x-1) \leq 2x^t$, and therefore $M \leq 2(2k)^t$ and $M_{mult} \leq 2r^t$. For every $x \geq 2$ we have that $x(x+2) \leq 2x^2$, and since $\epsilon > 0$, it follows that $t \geq 2$, for sufficiently large k, and $t(t+2) \leq 2t^2$. Therefore, we have

$$\frac{|S(\mathcal{M}_{k,r})|}{F(M_{mult},t)} \geq \frac{(k+r)!}{k!2t^2 2r^t}.$$

Similarly, we obtain the following upper bound on $F(M,t)$.

$$F(M,t) \leq t(t+2)M \leq 2t^2 M \leq 2t^2 2(2k)^t = 4t^2(2k)^t.$$

To verify inequality (A.1), it is enough to prove that for sufficiently large k,

$$\frac{(k+r)!}{k!2t^2 2r^t} \geq 4t^2(2k)^t, \tag{A.2}$$

or equivalently

$$\frac{(k+r)!}{k!} \geq 16t^4 r^t (2k)^t. \tag{A.3}$$

We distinguish now between two cases:

1) For $0 < \epsilon \leq 1$, since $r = \Theta(t)$ and $t = \Theta(k^{\epsilon})$, it follows that $r^t \leq c_1^t k^{\epsilon t}$ for some constant c_1 and sufficiently large k and therefore

$$16t^4 r^t (2k)^t \leq c^t k^{4\epsilon+\epsilon t+t} = k^{t\log_k c + 4\epsilon + \epsilon t + t}, \tag{A.4}$$

for some constant c and sufficiently large k. If $\mu > 1+\epsilon$ and k is sufficiently large then

$$\mu \geq \log_k c + \frac{4\epsilon}{t} + \epsilon + 1,$$

and therefore

$$k^{\mu t} \geq k^{t\log_k c + 4\epsilon + \epsilon t + t}. \tag{A.5}$$

Since $\frac{(k+r)!}{k!} \geq k^r \geq k^{\mu t}$ and by (A.4) and (A.5), it follows that inequality (A.3) is satisfied.

2) For $\epsilon > 1$, it follows that $k = O(r)$. For every $n > 1$ we have the following bounds on $n!$ [25, p. 54]

$$n^{n+1/2}e^{-n} \leq n! \leq n^{n+1/2}e^{-(n-1)}.$$

Therefore,

$$\frac{(k+r)!}{k!} \geq \frac{(k+r)^{k+r+1/2}e^{-k-r}}{k^{k+1/2}e^{-(k-1)}} \geq (c_1 r)^r \geq (c_1 r)^{\mu t}, \tag{A.6}$$

for some constant c_1 and sufficiently large k. Since $t = \Theta(r)$ and $k = \Theta(t^{\frac{1}{\epsilon}})$, it follows that $k^t \leq c_2^t r^{\frac{1}{\epsilon}t}$ for some constant c_2 and sufficiently large k and therefore

$$c_1^{-\mu t} 16t^4 r^t (2k)^t \leq c^t r^{4+t+\frac{1}{\epsilon}t} = r^{t\log_r c + 4 + t + \frac{1}{\epsilon}t}, \tag{A.7}$$

for some constant c and sufficiently large k. If $\mu > 1+\frac{1}{\epsilon}$ and k is sufficiently large then

$$\mu \geq \log_r c + \frac{4}{t} + \frac{1}{\epsilon} + 1,$$

and therefore

$$(c_1 r)^{\mu t} \geq c_1^{\mu t} r^{t\log_r c + 4 + t + \frac{1}{\epsilon}t}. \tag{A.8}$$

Combining (A.6), (A.7), and (A.8), it follows that inequality (A.3) is satisfied. □

REFERENCES

[1] 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.

[2] J. Bierbrauer and K. Metsch, "A bound on permutation codes," *Electron. J. Combinat.*, vol. 20, no. 3, pp. 1–12, Jul. 2013, paper P6.

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

[4] R. C. Bose and S. Chowla, "Theorems in the additive theory of numbers," *Commentarii Mathematici Helvetici*, vol. 37, pp. 141–147, Dec. 1962.

[5] S. Buzaglo and T. Etzion, "Perfect permutation codes with the Kendall's τ-metric," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun./Jul. 2014, pp. 2391–2395.

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

[7] S. Buzaglo, E. Yaakobi, T. Etzion, and J. Bruck, "Error-correcting codes for multipermutations," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 724–728.

[8] S. Buzaglo, E. Yaakobi, T. Etzion, and J. Bruck, "Systematic codes for rank modulation," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun./Jul. 2014, pp. 2386–2390.

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

[10] T. M. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. 19, no. 1, pp. 73–77, Jan. 1973.

[11] P. Dukes and N. Sawchuck, "Bounds on permutation codes of distance four," *J. Algebraic Combinat.*, vol. 31, pp. 143–158, Feb. 2010.

[12] E. En Gad, A. Jiang, and J. Bruck, "Trade-offs between instantaneous and total capacity in multi-cell flash memories," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 990–994.

[13] E. E. Gad, E. Yaakobi, A. Jiang, and J. Bruck, "Rank-modulation rewriting codes for flash memories," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 704–708.

[14] S. W. Golomb and L. R. Welch, "Perfect codes in the Lee metric and the packing of polyominoes," *SIAM J. Appl. Math.*, vol. 18, pp. 302–317, Mar. 1970.

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

[16] 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.

[17] M. Kendall and J. D. Gibbons, *Rank Correlation Methods*. New York, NY, USA: Oxford Univ. Press, 1990.

[18] D. E. Knuth, *The Art of Computer Programming: Sorting and Searching*, vol. 3. Reading, MA, USA: Addison-Wesley, 1998.

[19] A. Mazumdar, A. Barg, and G. Zémor, "Constructions of rank modulation codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 1018–1029, Feb. 2013.

[20] F. Sala, R. Gabrys, and L. Dolecek, "Dynamic threshold schemes for multi-level non-volatile memories," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2624–2634, Jul. 2012.

[21] D. Slepian, "Permutation modulation," *Proc. IEEE*, vol. 53, no. 3, pp. 228–236, Mar. 1965.
[22] 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.
[23] I. Tamo and M. Schwartz, "On the labeling problem of permutation group codes under the infinity metric," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6595–6604, Oct. 2012.
[24] R. R. Varshamov and G. M. Tenengol'ts, "Code correcting single asymmetric errors," *Avtomat. Telemekh.*, vol. 26, no. 2, pp. 288–292, 1965.
[25] D. Vrajitoru and W. Knight, *Practical Analysis of Algorithms*. Switzerland: Springer International Publishing, 2014.
[26] H. Zhou, A. Jiang, and J. Bruck, "Systematic error-correcting codes for rank modulation," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 2978–2982.
[27] H. Zhou, M. Schwartz, A. A. Jiang, and J. Bruck, "Systematic error-correcting codes for rank modulation," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 17–32, Jan. 2015.

Sarit Buzaglo (M'14) was born in Israel in 1983. She received the B.Sc. and M.Sc. degrees from the Department of Mathematics at the Technion–Israel Institute of Technology, Haifa, Israel, in 2007 and 2010, respectively. In 2014, she received her Ph.D. degree from the Department of Computer Science at the Technion. She is currently a postdoctoral researcher in the Center for Memory and Recording Research at University of California, San Diego, USA. She is also an awardee of the Weizmann Institute of Science–National Postdoctoral Award Program for Advancing Women in Science. Her research interests include coding theory, algebraic error-correction coding, coding for advanced storage devices and systems, and combinatorics.

Eitan Yaakobi (S'07–M'12) is an Assistant Professor at the Computer Science Department at the Technion–Israel Institute of Technology. He received the B.A. degrees in computer science and mathematics, and the M.Sc. degree in computer science from the Technion–Israel Institute of Technology, Haifa, Israel, in 2005 and 2007, respectively, and the Ph.D. degree in electrical engineering from the University of California, San Diego, in 2011. Between 2011-2013, he was a postdoctoral researcher in the department of Electrical Engineering at the California Institute of Technology. His research interests include information and coding theory with applications to non-volatile memories, associative memories, data storage and retrieval, and voting theory. He received the Marconi Society Young Scholar in 2009 and the Intel Ph.D. Fellowship in 2010-2011.

Tuvi Etzion (M'89–SM'94–F'04) was born in Tel Aviv, Israel, in 1956. He received the B.A., M.Sc., and D.Sc. degrees from the Technion–Israel Institute of Technology, Haifa, Israel, in 1980, 1982, and 1984, respectively.

From 1984 he held a position in the Department of Computer Science at the Technion, where he has a Professor position. During the years 1985-1987 he was Visiting Research Professor with the Department of Electrical Engineering - Systems at the University of Southern California, Los Angeles. During the summers of 1990 and 1991 he was visiting Bellcore in Morristown, New Jersey. During the years 1994-1996 he was a Visiting Research Fellow in the Computer Science Department at Royal Holloway College, Egham, England. He also had several visits to the Coordinated Science Laboratory at University of Illinois in Urbana-Champaign during the years 1995-1998, two visits to HP Bristol during the summers of 1996, 2000, a few visits to the Department of Electrical Engineering, University of California at San Diego during the years 2000-2012, and several visits to the Mathematics Department at Royal Holloway College, Egham, England, during the years 2007-2009.

His research interests include applications of discrete mathematics to problems in computer science and information theory, coding theory, and combinatorial designs.

Dr Etzion was an Associate Editor for Coding Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2006 till 2009. From 2004 to 2009, he was an Editor for the *Journal of Combinatorial Designs*. From 2011 he is an Editor for *Designs, Codes, and Cryptography*. From 2013 he is an Editor for *Advances of Mathematics in Communications*.

Jehoshua Bruck (S'86–M'89–SM'93–F'01) is the Gordon and Betty Moore Professor of computation and neural systems and electrical engineering at the California Institute of Technology (Caltech). His current research interests include information theory and its applications to memory systems, including molecular memories (DNA), associative memories (the brain), and nonvolatile memories (flash).

Dr. Bruck received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion-Israel Institute of Technology, in 1982 and 1985, respectively, and the Ph.D. degree in electrical engineering from Stanford University, in 1989.

His industrial and entrepreneurial experiences include working with IBM Research where he participated in the design and implementation of the first IBM parallel computer; co-founding and serving as Chairman of Rainfinity (acquired in 2005 by EMC), a spin-off company from Caltech that created the first virtualization solution for Network Attached Storage; as well as cofounding and serving as Chairman of XtremIO (acquired in 2012 by EMC), a start-up company that created the first scalable all-flash enterprise storage system.

Dr. Bruck is a recipient of the Feynman Prize for Excellence in Teaching, the Sloan Research Fellowship, the National Science Foundation Young Investigator Award, the IBM Outstanding Innovation Award and the IBM Outstanding Technical Achievement Award.