Error-Correcting Codes for Multipermutations

Sarit Buzaglo*, **Eitan Yaakobi**†, **Tuvi Etzion***, and **Jehoshua Bruck**†
*Computer Science Department, Technion – Israel Institute of Technology, Haifa 32000, Israel
†Electrical Engineering Department, California Institute of Technology, Pasadena, CA 91125, U.S.A
{*sarahb, etzion*}@*cs.technion.ac.il*, {yaakobi, bruck}@caltech.edu

***Abstract*—Multipermutations appear in various applications in information theory. New applications such as rank modulation for flash memories and voting have suggested the need to consider error-correcting codes for multipermutations. The construction of codes is challenging when permutations are considered and it becomes even a harder problem for multipermutations. In this paper we discuss the general problem of error-correcting codes for multipermutations. We present some tight bounds on the size of error-correcting codes for several families of multipermutations. We find the capacity of the channels of multipermutations and characterize families of perfect codes in this metric which we believe are the only such perfect codes.**

I. Introduction

A permutation is a full order of some fixed number of elements, say n, and the set of all permutations is denoted by $\mathbb{S}_n$, where $|\mathbb{S}_n| = n!$. The natural generalization of a permutation becomes a multipermutation, which is the case where every element can appear more than once. Assume there are m elements which we often call ranks. We denote by r_i the multiplicity of the i-th rank, $1 \le i \le m$. Let n be the length of the multipermutation, then we first have that $n = \sum_{i=1}^{m} r_i$. For $1 \le i \le m$, $\sigma^{-1}(i)$ is the set of all positions with rank i, i.e., $\sigma^{-1}(i) = \{j \mid \sigma(j) = i\}$. We call the vector $\overrightarrow{r} = (r_1, r_2, \dots, r_m)$ the *multiplicity vector* of the multipermutation. The set of all multipermutations with m ranks and multiplicity vector $\overrightarrow{r}$ is denoted by $P_{m,\overrightarrow{r}}$, and its size is $P_{m,\overrightarrow{r}} = \frac{n!}{\prod_{i=1}^{m} r_i!}$. Hence, $\sigma = (\sigma(1), \sigma(2), \dots, \sigma(n)) \in P_{m,\overrightarrow{r}}$ if and only if for all $1 \le i \le m$, $|\sigma^{-1}(i)| = r_i$. In case that $r = r_i$ for all $1 \le i \le m$, we denote the set $P_{m,\overrightarrow{r}}$ simply by $P_{m,r}$, and we will follow the same analogy in the other definitions which include the multiplicity vector $\overrightarrow{r}$.

Multipermutations are an important tool which can be found in several applications. In [14], Slepian introduced multipermutation codes for transmission over a Gaussian channels. Later, these codes were further studied and generalized; see for example [10], [12], [13]. Recently, multipermutations were applied in codes for the ubiquitous flash memory. 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 it to exactly program each cell to its designated level. In order to overcome this difficulty, the novel framework of *rank modulation codes* was introduced in [5]. 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], [6], [16]. Recently, to improve the number of rewrites, the model of rank modulation was extended such that multiple cells can share the same ranking [2], [3]. Thus, the cells no longer determine a permutation but rather a multipermutation. Lastly, error-correcting codes for multipermutations subject to the Kendall's τ-metric were presented in [11].

In this paper we consider error-correcting codes for multipermutations under the well-known Kendall's τ-metric. In Section II we present basic properties of the metric. We derive a mapping which transforms a multipermutation with any multiplicity vector to a permutation. This mapping is the key for efficiently calculating the Kendall's τ-distance between any two multipermutations. In addition, this mapping is instrumental for codes constructions: In Section III we present two simple constructions of error-correcting codes for this metric. The first construction is of error-correcting codes for multipermutations that are derived from error-correcting codes for permutations while the second construction shows the reverse direction. We demonstrate this idea by constructing systematic error-correction codes for multipermutations that are based on a construction for permutations that was recently presented in [16]. In Section IV, we prove that the first construction in Section III is asymptotically optimal for some multiplicity vectors. In Section V we prove that if $\frac{\log_2 r}{\log_2 m} \to 0$ then the channel capacity for $P_{m,r}$ and any distance d is equal to the capacity of the channel for $\mathbb{S}_n$ and the same distance d. In Section VI we discuss the existence of perfect codes in this metric. Due to lack of space we have omitted many of the proofs.

II. Basic Properties of Kendall's τ-Metric

The Kendall's τ-distance [7] between two permutations $\sigma, \pi \in \mathbb{S}_m$ is denoted by $d_K(\sigma, \pi)$ and is defined to be the minimum number of adjacent transpositions required to obtain the permutation π from the permutation σ. It is also known [6], [8] that $d_\tau(\sigma, \pi)$ can be expressed as

$$d_K(\sigma, \pi) = |\{(i, j) : i \neq j,\ \sigma^{-1}(i) < \sigma^{-1}(j),\ \pi^{-1}(i) > \pi^{-1}(j)\}|.$$

The generalization of Kendall's τ-distance for two multipermutations $\sigma, \pi \in P_{m,\overrightarrow{r}}$ is defined similarly as the minimum number of adjacent transpositions to obtain the multipermutation π from the multipermutation σ. This distance will be denoted also by $d_K(\sigma, \pi)$ as this is a generalization of the definition for permutations. For example, if $\sigma = [1, 1, 2, 2]$ and $\pi = [2, 1, 2, 1]$, then $d_\tau(\sigma, \pi) = 3$, because to change the permutation from σ to π, we need at least 3 adjacent transpositions: $[1, 1, 2, 2] \to [1, 2, 1, 2] \to [2, 1, 1, 2] \to [2, 1, 2, 1]$.

Let $\overrightarrow{r}$ be a multiplicity vector and $n = \sum_{i=1}^{m} r_i$. We define a mapping from $P_{m,\overrightarrow{r}}$ to $\mathbb{S}_n$,

$$T : P_{m,\overrightarrow{r}} \to \mathbb{S}_n,$$

such that for every $\sigma \in P_{m,\overrightarrow{r}}$, $T(\sigma)$ is a permutation with n ranks such that the first r_1 ranks of $T(\sigma)$ are in increasing order in the r_1 locations of the first rank in σ. The following r_2 ranks of $T(\sigma)$ are in increasing order in the r_2 locations of second rank in σ, and so on. For example,

$$T((121331323)) = (142673859).$$

The mapping T is useful for the computation of the Kendall's τ-distance between two multipermutations because it helps reduce it to computation of the distance on the corresponding permutations.

Lemma 1: For every $\sigma, \mu \in P_{m,\overrightarrow{r}}$, we have

$$d_K(\sigma, \mu) = d_K(T(\sigma), T(\mu)).$$

For example, if $\sigma = [1,1,2,2], \mu = [2,1,2,1]$, then $d_K(\sigma, \mu) = 3$. Also, by the lemma, $d_K(T([1,1,2,2]), T([2,1,2,1])) = d_K([1,2,3,4],[3,1,4,2]) = 3$.

An important tool to find properties about metrics is their representation by a graph. For a multiplicity vector $\overrightarrow{r}$ the graph $G(\overrightarrow{r})$ is defined as follows. Each multipermutation in $P_{m,\overrightarrow{r}}$ is represented by a vertex in $G(\overrightarrow{r})$. Two vertices in $G(\overrightarrow{r})$, representing the multipermutations σ and μ are connected by an edge if $d_K(\sigma, \mu) = 1$. Two simple and important properties of the Kendall's τ-metric with regards to multipermutations are summarized in the following two lemmas.

Lemma 2: If x, y, z, are three vertices in $G(\overrightarrow{r})$, then $d_K(x, y) + d_K(y, z) \equiv d_K(x, z) \pmod 2$.

Corollary 1: For any given multiplicity vector $\overrightarrow{r}$ the graph $G(\overrightarrow{r})$ is a bipartite graph.

A multipermutation σ will be called a *ranks-run* if all the symbols of each rank form a single run. In $P_{m,r}$ there are exactly $m!$ ranks-run multipermutations. For a multipermutation $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$ the *reverse* multipermutation σ^R is defined by $\sigma^R \overset{\text{def}}{=} (\sigma(n), \ldots, \sigma(2), \sigma(1))$.

Lemma 3: Let μ be a multipermutation in $G(\overrightarrow{r})$. Then $d_K(\sigma, \mu) + d_K(\sigma^R, \mu) = d_K(\sigma, \sigma^R)$ if and only if σ is a ranks-run multipermutation.

III. Error-Correcting Codes

Here we are interested in constructions and upper bounds on error-correcting codes for multipermutations with the Kendall's τ-distance. As common for other metrics we present the following definitions for the related bounds. Let $A(n, d)$ be the maximum size of a code in $\mathbb{S}_n$ with minimum distance d, $A(m, r, d)$ be the maximum size of a code in $P_{m,r}$ with minimum distance d, and $A(\overrightarrow{r}, d)$ be the maximum size of a code in $P_{m,\overrightarrow{r}}$ with minimum distance d. The first construction is stated in the following lemma.

Lemma 4: If $\overrightarrow{r}$ is a multiplicity vector of length m, where $n = \sum_{i=1}^{m} r_i$, then

$$A(\overrightarrow{r}, d) \geq \frac{A(n, d)}{\prod_{i=1}^{m} r_i!}.$$

Surprisingly, if r is relatively small then this bound is quite strong, as we will show in the sequel.

Another construction can lead to a bound in the other direction which is rather weak.

Lemma 5:

$$A(n, d) \geq A(\overrightarrow{r}, d) \prod_{i=1}^{m} A(r_i, d)$$

In the rest of this section, we will show an example of a construction of error-correcting codes for multipermutations based upon existing codes for permutations. In particular, we present a modification of the *systematic* error-correction codes presented in [16], that applies to multipermutations.

In an (n, k) systematic error-correcting code for permutations, the information is induced by the permutation of the first k elements, which are called the information elements while the last $n - k$ elements are called the redundancy elements. In encoding, a permutation of the k information elements is given and accordingly the locations of the redundancy elements are computed, while the order between the information elements remains the same. Systematic error-correcting codes for multipermutations are defined similarly.

A construction of a $(n, n-2)$ systematic single error-correcting code was given in [16]. We modify this construction for multipermutations with $r = 2$. In our construction, we will have to use an additional redundancy symbol resulting in three redundancy symbols: an $(m-1)$ and two m's. The symbol $(m-1)$ is used to tell the order between the last two elements labeled m.

Theorem 1: Let $2m-1$ be a prime number. Then, there exists a $(2m, 2m-3)$ systematic single error-correcting code in $P_{m,2}$.

Proof: The existence of the code is introduced by its construction. Let $\mathcal{C}$ be a single $(2m, 2m-2)$ systematic error-correction code in $\mathbb{S}_n$ with encoder $\mathcal{E}$. The information of the multipermutation code is carried by the first $2m-3$ elements $1, 1, 2, 2, \ldots, (m-2), (m-2), (m-1)$ and assume it is the multipermutation $\pi = (\pi(1), \ldots, \pi(2m-3))$. The encoding is invoked by the following steps:

1) $\pi' = (\pi'(1), \ldots, \pi'(2m-3), \pi'(2m-2))$ is the multipermutation obtained from π, where the second symbol $(m-1)$ (the one that belongs to the redundancy) is inserted in π after the first $(m-1)$, i.e., right after position $\pi^{-1}(m-1)$.
2) Apply the encoder $\mathcal{E}$ on the permutation $T(\pi')$. Let i_{2m-1} and i_{2m} be the resulting positions of the last two redundancy elements.
3) If $i_{2m-1} < i_{2m}$ then insert the two elements of m in π' in these positions.
4) If $i_{2m-1} > i_{2m}$ then first move the second element $m-1$ in π' to another location at some distance greater than five from the location of the first element $(m-1)$ in π'. Then insert the two elements m in π' in the corresponding positions.

For decoding, according to the positions of the elements $(m-1)$, we determine the the relative order between the elements m. The rest of the decoding can be implemented by applying the decoder of the code $\mathcal{C}$, as explained in [16]. ■

IV. A modified sphere packing bound

In order to evaluate the efficiency of the construction in Lemma 4, we would like to find upper bounds on $A(\overrightarrow{r}, d)$. The first and most natural upper bound to consider is the sphere packing bound. However, the size of a ball in Kendall's τ-distance for multipermutations which are not permutations, depends on the center of the ball and the classic bound does not work in this case. We present a technique to obtain upper bound which was recently studied in [9] for the deletion channel.

We follow the same definitions and construction as in [9]. A hypergraph $\mathcal{H} = (X, \mathcal{E})$ is a pair, where X is its set of vertices and $\mathcal{E}$ is a collection of nonempty subsets of X, called hyperedges, such that $\bigcup_{E\in\mathcal{E}} E = X$. A *matching* in $\mathcal{H}$ is a collection of pairwise disjoint hyperedges $E_1, \dots, E_\ell \in \mathcal{E}$. The *matching number* of $\mathcal{H}$, denoted by $\nu(\mathcal{H})$, is the size of the largest matching. A *transversal* of a hypergraph $\mathcal{H}$ is a subset $T \subseteq X$ that intersects every hyperedge in $\mathcal{E}$. The *transversal number* of $\mathcal{H}$, denoted by $\tau(\mathcal{H})$, is the size of the smallest transversal. Assume that there are n vertices, $x_1, \dots, x_n$, and m hyperedges is $E_1, \dots, E_m$. The matrix $A \in \{0,1\}^{n\times m}$ defined by $A(i,j) = 1$ if and only if $x_i \in E_j$, is called the *incidence matrix* of $\mathcal{H}$. The matching number is clearly the solution of the integer linear programming problem:

$$\nu(\mathcal{H}) = \max\{w_{L_1}(\boldsymbol{z}) : A\boldsymbol{z} \le \boldsymbol{1}, \boldsymbol{z} \in \{0,1\}^m\},$$

and the transversal number is the solution of the integer linear programming problem:

$$\tau(\mathcal{H}) = \min\{w_{L_1}(\boldsymbol{y}) : A^T\boldsymbol{y} \ge \boldsymbol{1}, \boldsymbol{y} \in \{0,1\}^n\}.$$

Furthermore, it can be verified that $\nu(\mathcal{H}) \le \tau(\mathcal{H})$, and these two problems satisfy weak duality. These last two problems can be slightly changed such that the vectors in the minimization and maximization problems do not have be to binary, so

$$\nu(\mathcal{H}) = \max\{w_{L_1}(\boldsymbol{z}) : A\boldsymbol{z} \le \boldsymbol{1}, \boldsymbol{z} \in \mathbb{Z}_+^m\},$$
$$\tau(\mathcal{H}) = \min\{w_{L_1}(\boldsymbol{y}) : A^T\boldsymbol{y} \ge \boldsymbol{1}, \boldsymbol{y} \in \mathbb{Z}_+^m\},$$

where $\mathbb{Z}_+$ is the set of nonnegative integers.

It can be easily verified that the integer solutions are equal to the binary solutions. The problems can be further relaxed with the following linear programming problems:

$$\nu^*(\mathcal{H}) = \max\{w_{L_1}(\boldsymbol{z}) : A\boldsymbol{z} \le \boldsymbol{1}, \boldsymbol{z} \in \mathbb{R}_+^m\},$$
$$\tau^*(\mathcal{H}) = \min\{w_{L_1}(\boldsymbol{y}) : A^T\boldsymbol{y} \ge \boldsymbol{1}, \boldsymbol{y} \in \mathbb{R}_+^m\},$$

where $\mathbb{R}_+$ is the set of nonnegative numbers.

The real solutions can be significantly different from the integer solutions. $\nu^*(\mathcal{H})$ and $\tau^*(\mathcal{H})$ satisfy strong duality and hence

$$\nu^*(\mathcal{H}) = \tau^*(\mathcal{H}).$$

Finally, the following property holds

$$\nu(\mathcal{H}) \le \nu^*(\mathcal{H}) = \tau^*(\mathcal{H}) \le \tau(\mathcal{H}),$$

and in particular,

$$\nu(\mathcal{H}) \le \tau^*(\mathcal{H}) \le w_{L_1}(\boldsymbol{u}), \tag{1}$$

for any fractional transversal $\boldsymbol{u}$.

We distinguish between $d = 3$ and $d \ge 5$, i.e. single-error-correcting codes and t-error-correcting codes, $t \ge 2$.

A. Single-error-correcting codes

The size of a ball $B_1(\sigma)$ of radius one for every multipermutation $\sigma \in P_{m,\overrightarrow{r}}$ satisfies $m \le |B_1(\sigma)| \le n$. Hence, a trivial upper bound on the size of multipermutation codes in $P_{m,\overrightarrow{r}}$, where $\overrightarrow{r}$ is a multiplicity vector of length m, which correct a single error is

$$\frac{n!}{m\prod_{i=1}^m r_i!}. \tag{2}$$

For simplicity let's consider now multipermutations only in $P_{m,2}$. By Lemma 4, if there exists an optimal single-error-correcting code in $\mathbb{S}_n$ whose size is $(n-1)!$, then there exists a single error-correcting code in $P_{m,2}$ whose size is $\frac{(n-1)!}{2^m}$, while the upper bound (2) is

$$\frac{n!}{m\cdot 2^m} = 2\cdot\frac{(n-1)!}{2^m}. \tag{3}$$

We will try to close on this gap.

A first observation on the size of a ball with radius one is that its minimum size is attained for permutations where similar ranks appear together and the maximum size is attained when no two similar ranks appear together. For $\sigma \in P_{m,2}$ let us denote by $u(\sigma)$ the number of runs in σ. It can be easily verified that

$$|B_1(\sigma)| = u(\sigma).$$

We define a hypergraph $\mathcal{H} = (X, \mathcal{E})$, whose set of vertices is $X = P_{m,2}$ and the hyperedges represents the balls with radius one around the elements of $P_{m,2}$, i.e.

$$\mathcal{E} = \{B_1(\sigma) : \sigma \in P_{m,2}\}.$$

The number of vertices and the number of hyperedges in $\mathcal{H}$ is equal to $M = \frac{n!}{2^m}$. Each single-error-correcting code in $P_{m,2}$ corresponds to a matching in the hypergraph $\mathcal{H}$. Therefore, an upper bound on the size of such codes is the matching number $\nu(\mathcal{H})$. By (1), for any fractional transversal $\boldsymbol{u}$ we obtain an upper bound $w_{L_1}(\boldsymbol{u})$ on the size of such codes, where $w_{L_1}(\boldsymbol{v})$ is the sum of the elements in the vector $\boldsymbol{v}$.

Let $\sigma_1, \sigma_2, \dots, \sigma_M$ be an order of the M multipermutations in $P_{m,2}$. We define a vector $\boldsymbol{y} = (y_1, \dots, y_M)$ such that for each i, $1 \le i \le M$, $y_i = \frac{1}{u(\sigma_i)-2}$, where $u(\sigma_i)$ is the number of runs in σ_i. We will show now that $\boldsymbol{y}$ is a fractional transversal. It is clear that $\boldsymbol{y} \in \mathbb{R}_+^M$. Thus, we only need to show that $A^T\boldsymbol{y} \ge \boldsymbol{1}$. For a given j, $1 \le j \le M$,

$$(A^T\boldsymbol{y})_j = \sum_{i=1}^{M} a_{i,j}y_i = \sum_{\sigma_i\in B_1(\sigma_j)} y_i = \sum_{\sigma_i\in B_1(\sigma_j)} \frac{1}{u(\sigma_i)-2}$$
$$\ge \sum_{\sigma_i\in B_1(\sigma_j)} \frac{1}{u(\sigma_j)} = |B_1(\sigma_j)|\cdot\frac{1}{u(\sigma_j)} = \frac{u(\sigma_j)}{u(\sigma_j)} = 1.$$

Note that the inequality follows from the property that a single adjacent transposition can change the number of runs by at most two. Hence, for each i and j, such that $\sigma_i \in B_1(\sigma_j)$ we have $u(\sigma_i) - 2 \le u(\sigma_j)$.

Now, we have to calculate the value of $w_{L_1}(\boldsymbol{y})$ (or an upper bound on this value), which will give an upper bound on $A(m,2,3)$. The number of runs for each $\sigma \in P_{m,2}$ is between m and $2m$. The number of multipermutations with exactly $2m-\ell$ runs (ℓ ranks are together in the multipermutations) for $0 \le \ell \le m$ is computed by using the inclusion-exclusion principle. This number is equal to:

$$N_\ell = \binom{m}{\ell}\sum_{i=0}^{m-\ell}(-1)^i\binom{m-\ell}{i}\frac{(2m-(\ell+i))!}{2^{m-(\ell+i)}}.$$

Finally, we have

$$w_{L_1}(\boldsymbol{y}) = \sum_{\ell=0}^{m} N_\ell\cdot\frac{1}{2m-\ell-2}. \tag{4}$$

The following table presents numerical values of the bounds on $A(m,2,3)$.

m	bound in (3)	bound in (1) and (4)
3	30	34
4	630	526
5	22680	16552
6	1247400	842262
7	97297200	62441159
8	10216206000	6324770781

The following lemma is required for the next result.

Lemma 6: For every positive integers t and a such that $a > t$ the following identity holds:

$$\sum_{\ell=0}^{t}(-1)^{t-\ell}\binom{t}{\ell}\frac{1}{a-\ell}=\frac{t!}{\prod_{\ell=0}^{t}(a-\ell)}.$$

Lemma 7: For $m \geq 7$, the value $w_{L_1}(\boldsymbol{y})$ satisfies

$$w_{L_1}(\boldsymbol{y}) \leq \frac{(2m)!}{2^m\cdot(2m-2)}\left(1+\frac{1}{m}\right).$$

Proof: Let $S \subseteq [m] = \{1,\dots,m\}$ and let N_S be the number of multipermutations such that for each rank $s \in S$, the two ranks appear together, and for each other rank, the two rank do not appear together. Hence,

$$N_S = \sum_{T\subseteq[m],S\subseteq T}(-1)^{|T\setminus S|}\frac{(2m-|T|)!}{2^{m-|T|}}. \tag{5}$$

Now, we have $w_{L_1}(\boldsymbol{y}) = \sum_{S\subseteq[m]}\frac{1}{2m-|S|-2}\cdot N_S$. Together with (5) we conclude

$$\begin{aligned}
w_{L_1}(\boldsymbol{y}) &= \sum_{S\subseteq[m]}\frac{1}{2m-|S|-2}\cdot N_S\\
&= \sum_{S\subseteq[m]}\frac{1}{2m-|S|-2}\cdot\sum_{T\subseteq[m],S\subseteq T}(-1)^{|T\setminus S|}\frac{(2m-|T|)!}{2^{m-|T|}}\\
&= \sum_{T\subseteq[m]}\frac{(2m-|T|)!}{2^{m-|T|}}\sum_{S\subseteq[m],S\subseteq T}(-1)^{|T\setminus S|}\frac{1}{2m-|S|-2}\\
&= \sum_{T\subseteq[m]}\frac{(2m-|T|)!}{2^{m-|T|}}\sum_{\ell=0}^{|T|}(-1)^{|T|-\ell}\binom{|T|}{\ell}\frac{1}{2m-2-\ell}\\
&=^{(a)} \sum_{T\subseteq[m]}\frac{(2m-|T|)!}{2^{m-|T|}}\cdot\frac{|T|!}{\prod_{\ell=0}^{|T|}(2m-2-\ell)}\\
&= \sum_{T\subseteq[m]}\frac{(2m-|T|)!}{2^{m-|T|}}\cdot\frac{1}{(2m-2)\binom{2m-3}{|T|}}\\
&= \sum_{t=0}^{m}\binom{m}{t}\frac{(2m-t)!}{2^{m-t}}\cdot\frac{1}{(2m-2)\binom{2m-3}{t}}\\
&= \sum_{t=0}^{m}\frac{(2m)!}{2^m\cdot(2m-2)}\cdot\frac{(2m-t)!2^t\cdot\binom{m}{t}}{(2m)!\binom{2m-3}{t}}\\
&= \frac{(2m)!}{2^m\cdot(2m-2)}\left(1+\frac{1}{2m-3}+\sum_{t=2}^{m}\frac{(2m-t)!2^t\cdot\binom{m}{t}}{(2m)!\binom{2m-3}{t}}\right)\\
&\leq^{(b)} \frac{(2m)!}{2^m\cdot(2m-2)}\left(1+\frac{1}{2m-3}+\frac{(m-1)\cdot2(m-1)}{(2m-1)(2m-3)(2m-4)}\right)\\
&\leq^{(c)} \frac{(2m)!}{2^m\cdot(2m-2)}\left(1+\frac{1}{m}\right)
\end{aligned}$$

where equality (a) follows from Lemma 6. Inequality (b) follows from the property that for all $t \leq m$

$$\frac{(2m-t)!2^t\cdot\binom{m}{t}}{(2m)!\binom{2m-3}{t}} \geq \frac{(2m-(t+1))!2^{t+1}\cdot\binom{m}{t+1}}{(2m)!\binom{2m-3}{t+1}},$$

and Inequality (c) is true for $m \geq 7$. ∎

Since the largest size of a ball in this case is $2m$, it follows that the result of Lemma 7 implies that asymptotically a sphere packing bound with the largest possible ball is obtained.

B. t-error-correcting codes, $t \geq 2$

Our goal is to generalize Lemma 7 for t-error-correcting codes, $t \geq 2$. For a multipermutation $\sigma \in P_{m,2}$, let $B_t(\sigma)$ be the ball with radius t, whose center is σ, and let $b_t(\sigma) = |B_t(\sigma)|$. Let $\sigma_1, \sigma_2, \dots, \sigma_M$ be an order of the M multipermutations in $P_{m,2}$. The vector $\boldsymbol{y}_t = (y_{t,1}, \dots, y_{t,M})$ is defined as follows

$$y_{t,i} = \frac{1}{\min_{\mu\in B_t(\sigma_i)}\{b_t(\mu)\}}.$$

It can be verified as before that $A^T\boldsymbol{y}_t \geq \mathbf{1}$, and hence $w_{L_1}(\boldsymbol{y}_t)$ is an upper bound on the size of a t-error-correcting code.

For $t = 2$ we can show that for every $\sigma \in P_{m,2}$,

$$\binom{m+2}{2}-2 \leq b_2(\rho) \leq \binom{2m+1}{2}-1 \ .$$

Similarly to the case $t = 1$ we can prove that asymptotically an upper bound on $w_{L_1}(\boldsymbol{y}_t)$, for $t \geq 2$, is $\frac{(2m)!}{2^m\cdot\binom{2m-2t-1}{t}}$.

The size of a ball in the metric is very important in the context of the sphere packing bound and also to obtain a lower bound with the Gilbert-Varshamov. When the sphere depends on its center, a Gilbert-Varshamov lower bound was developed in [15]. It depends on the size of the balls and the number of words for each such size. Therefore, these calculations are so important. In our case, we can use the lower bound in [15] for minimum distance 3, while for other distances it is a future research work.

V. The Capacity of Multipermutation Codes

For every positive d the capacity of codes in $\mathbb{S}_n$ with minimum distance d is defined as

$$\mathbb{C}(d) = \lim_{n\to\infty}\frac{\log_2 A(n,d)}{\log_2 n!}.$$

The following theorem was proved in [1].

Theorem 2: The value $\mathbb{C}(d)$ satisfies

$$\mathbb{C}(d) = \begin{cases} 1 & \text{if } d = O(n)\\ 1-\epsilon & \text{if } d = O(n^{1+\epsilon}), 0<\epsilon<1\\ 0 & \text{if } d = \Theta(n^2)\end{cases}$$

Similarly, we define the capacity of multipermutations codes with minimum distance d and fixed r to be

$$\mathbb{C}(r,d) = \lim_{m\to\infty}\frac{\log_2 A(m,r,d)}{\log_2|P_{m,r}|} = \lim_{m\to\infty}\frac{\log_2 A(m,r,d)}{\log_2\left(\frac{(mr)!}{(r!)^m}\right)}.$$

Lemma 8: If $\frac{\log_2 r}{\log_2 m} \to 0$ and $d \geq 1$ then $\mathbb{C}(r,d) \geq \mathbb{C}(d)$.

Lemma 9: If $\frac{\log_2 r}{\log_2 m} \to 0$ and $d \geq 1$ then $\mathbb{C}(r,d) \leq \mathbb{C}(d)$.

Proof: Let $\mathcal{C}$ be a multipermutation code of minimum distance d. We introduced earlier the mapping T from $P_{m,r}$ to $\mathbb{S}_n$ which preserves distances. Hence, if we define the code

$$\mathcal{C}' = T(\mathcal{C}) = \{T(\boldsymbol{c}) : \boldsymbol{c}\in\mathcal{C}\ \},$$

then we have that $\mathcal{C}'$ is a code in $\mathbb{S}_n$ with minimum distance d as well. It implies that $A(m,r,d) \leq A(n,d)$. Therefore,

$$\mathbb{C}(r,d)=\lim_{n\to\infty}\frac{\log_2 A(m,r,d)}{\log_2|P_{m,r}|}\leq\frac{\log_2 A(n,d)}{\log_2\left(\frac{(mr)!}{(r!)^m}\right)}$$
$$=\frac{\log_2 A(n,d)}{\log_2(mr)!-m\log_2 r!}=\frac{\log_2 A(n,d)}{\log_2(mr)!}\cdot\left(\frac{\log_2(mr)!}{\log_2(mr)!-m\log_2 r!}\right)$$
$$=\frac{\log_2 A(n,d)}{\log_2(mr)!}\cdot\left(\frac{1}{1-\frac{m\log_2 r!}{\log_2(mr)!}}\right)\to_{m\to\infty}\mathbb{C}(d)$$

■

Corollary 2: If $\frac{\log_2 r}{\log_2 m}\to 0$ and $d\geq 1$ then $\mathbb{C}(r,d)=\mathbb{C}(d)$.

VI. PERFECT CODES

In this section we consider the question of the existence of perfect codes in $G(r_1,r_2,\ldots,r_m)$. $\mathcal{C}$ is a *perfect code* with covering radius R in $G(r_1,r_2,\ldots,r_m)$ if for each vertex $v\in G(r_1,r_2,\ldots,r_m)$ there exists exactly one codeword $c\in\mathcal{C}$ such that $d_K(v,c)\leq R$. Trivial perfect codes are similar to trivial perfect codes in other metrics and they include the whole graph which is a perfect code with radius 0; one word v is always a perfect code with covering radius which is equal to the maximum distance of a vertex in the graph from v.

There are some more families of perfect codes in $G(r_1,r_2,\ldots,r_m)$. Their structure is very simple and therefore we will call also these perfect codes trivial. The graph representation and the results proved in Section II are used to prove the claims concerning the perfect codes.

1) The first family consists of codes which contain two reversed codewords for which the distance is an odd integer $2R+1$. Both codewords should be ranks-run multipermutations. This type of perfect codes is very similar to perfect codes in the Johnson scheme which consist of two complement codewords [4].
2) The second family consists of codes which contain two codewords from the graph $G(r_1,r_2)$. Since the multipermutations have two ranks we will assume that the code is binary (rather than with ranks 1 and 2). If r_1 and r_2 are odd then there are two types of perfect codes in the graph. In the first code, one codeword consists of r_1 *zeroes* followed by r_2 *ones*. The second codeword consists of r_2 *ones* followed by r_1 *zeroes*. The covering radius of this code is $\frac{r_1r_2-1}{2}$. Another perfect code in the same graph is formed by taking one codeword with r_1-1 *zeroes* followed by 10 followed by r_2-1 *ones*. The second codeword is the reverse of the first codeword. The covering radius of this code is $\frac{r_1r_2-3}{2}$. If r_1 and r_2 have different parity then there are also two types of perfect codes in the graph. The covering radius of both codes is $\frac{r_1r_2-2}{2}$. In the first code one codeword consists of r_1-1 *zeroes* followed by 10 followed by r_2-1 *ones*. The second codeword consists of r_2 *ones* followed by r_1 zeroes. In the second code the first codeword consists of r_1 *zeroes* followed by r_2 *ones*. The second codeword consists of r_2-1 *ones* followed by 01 followed by r_1-1 *zeroes*.
3) The third family of codes is in the graph $G(1,n-1)$ (and similar ones in $G(n-1,1)$). For each R, $1\leq R\leq n-1$ there exist perfect codes with covering radius R in the graph. The graph has n vertices which are represented by the n binary words of length n and weight one. If $\lfloor\frac{n}{2}\rfloor\leq R\leq n-1$ then there exist two perfect codes. The first code consists of one codeword with the *one* in position $R+1$. The second code consists of one codeword with the *one* in position $n-R$. If $n=2R+1$ then both codes coincide. If $1\leq R\leq\lfloor\frac{n}{2}\rfloor-1$ then the code consists of $\lceil\frac{n}{2R+1}\rceil$ codewords. Let j, be an integer between 1 and $2R+1$ such that $j\equiv n \pmod{2R+1}$. If $1\leq j\leq R+1$ then the first codeword can have an *one* in any position ℓ, $1\leq\ell\leq 1+j$. If $R+1\leq j\leq 2R+1$ then the first codeword can have an *one* in any position ℓ, $j-R\leq\ell\leq R+1$. If in the first codeword the *one* is in position t then for each i, $1\leq i\leq\lceil\frac{n}{2R+1}\rceil-1$ there is a codeword with an *one* in position $t+(2R+1)i$.

It is an intriguing question whether there exist more perfect codes in the Kendall's τ-metric for multipermutation.

ACKNOWLEDGMENT

The work of Sarit Buzaglo and Tuvi Etzion was supported in part by the Israel Science Foundation (ISF), Jerusalem, Israel, under Grant No. 10/12. The work of Eitan Yaakobi and Jehoshua Bruck was supported in part by Intellectual Ventures and an NSF grant CIF-1218005. The authors would like to thank Alexander Vardy and Michael Langberg for bringing reference [9] to their knowledge.

REFERENCES

[1] A. Barg and A. Mazumdar, "Codes in permutations and error correction for rank modulation," *IEEE Trans. on Information Theory*, vol. 56, no. 7, pp. 3158–3165, July 2010.
[2] E. En Gad, A. Jiang, and J. Bruck, "Trade-offs between instantaneous and total capacity in multi-cell flash memories," *Proc. IEEE International Symp. on Information Theory*, pp. 990–994, Cambridge, MA, July 2012.
[3] E. En Gad, E. Yaakobi, A. Jiang, and J. Bruck, "Rank-modulation rewriting codes for flash memories," submitted to *Proc. IEEE International Symp. on Information Theory*, Istanbul, Turkey, July 2013.
[4] T. Etzion, "On the nonexistence of perfect codes in the Johnson scheme," *SIAM Journal on Discrete Mathematics*, vol. 9, pp. 201–209, May 1996.
[5] A. Jiang, R. Mateescu, M. Schwartz, and J. Bruck, "Rank modulation for flash memories," *IEEE Trans. on Inform. Theory*, vol. 55, no. 6, pp. 2659–2673, June 2009.
[6] A. Jiang, M. Schwartz, and J. Bruck, "Correcting charge-constrained errors in the rank-modulation scheme," *IEEE Trans. on Information Theory*, vol. 56, no. 5, pp. 2112–2120, May 2010.
[7] M. Kendall and J. D. Gibbons, *Rank Correlation Methods*, New York: Oxford Univ. Press, 1990.
[8] D. E. Knuth, *The Art of Computer Programming, Volume 3: Sorting and Searching*, Reading, MA: Addiaon-Wesley, 1998.
[9] A. Kulkarni and N. Kiyavash, "Non-asymptotic upper bounds for deletion correcting codes," arXiv:1211.3128v1, November 2012.
[10] A. Nordio and E. Viterbo, "Permutation modulation for fading channels," *Proc. Int. Conf. Telecommunications*, vol. 2, pp. 1177–1183, Feb. 2003.
[11] F. Sala, R. Gabrys, and L. Dolecek, "Dynamic threshold schemes for multi-level non-volatile memories," submited to *IEEE Trans. on Communications*, 2012.
[12] M.-Z. Shieh and S.-C. Tsai, "Computing the ball size of frequency permutations under chebyshev distance," arXiv:1102.2799v2, February 2011.
[13] D. Silva and W.E. Finamore, "Vector permutation modulation," *IEEE Communications Letters*, vol. 9, no. 8 pp. 673–675, August 2005.
[14] D. Slepian, "Permutation modulation," *Proceedings of the IEEE*, pp. 228–236, March 1965.
[15] L.M.G. Tolhuizen, "The generalized Gilbert-Varshamov bound is implied by Turán theorem," *IEEE Trans. on Information Theory*, vol. 43, pp. 1605–1606, September 1997.
[16] H. Zhou, A. Jiang, and J. Bruck, "Systematic error-correction codes for rank modulation," *Proc. IEEE International Symposium on Information Theory*, pp. 2978–2982, Cambridge, MA, July 2012.