Codes in Permutations *and* Error Correction for Rank Modulation

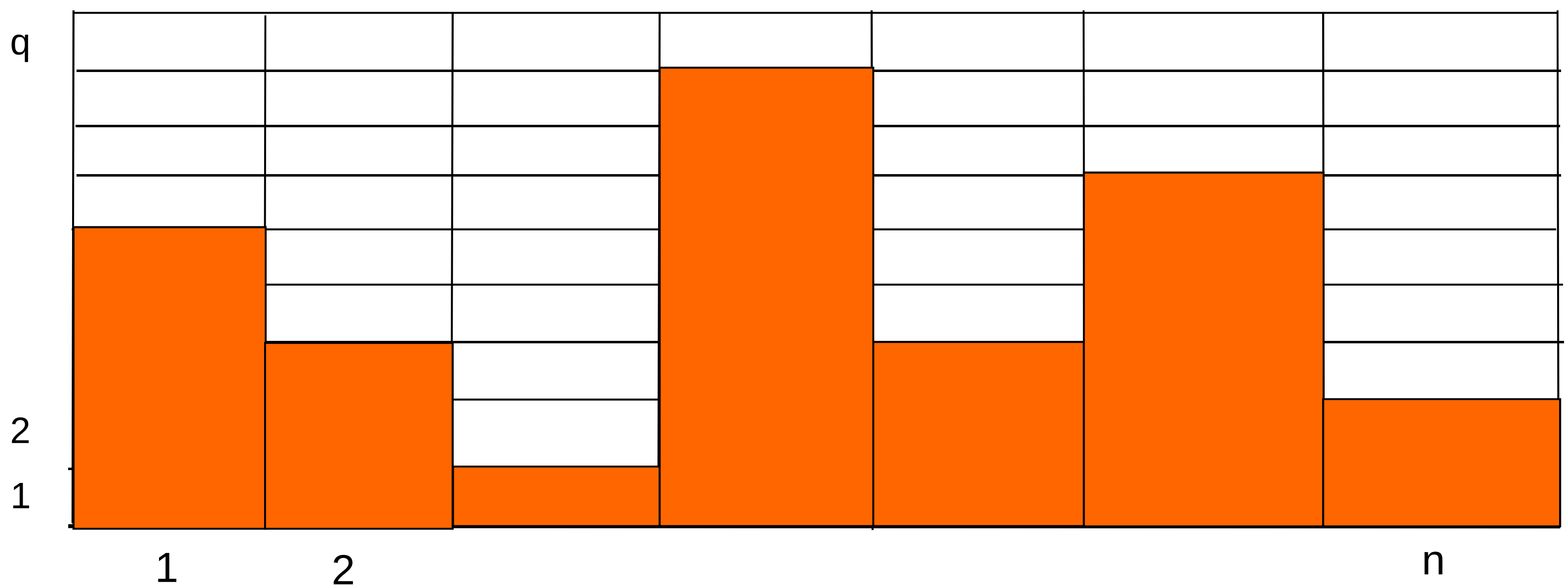
Alexander Barg and Arya Mazumdar

ECE/Institute for Systems Research
University of Maryland

Preprint: arXiv0908.4094

Storing data in flash memory

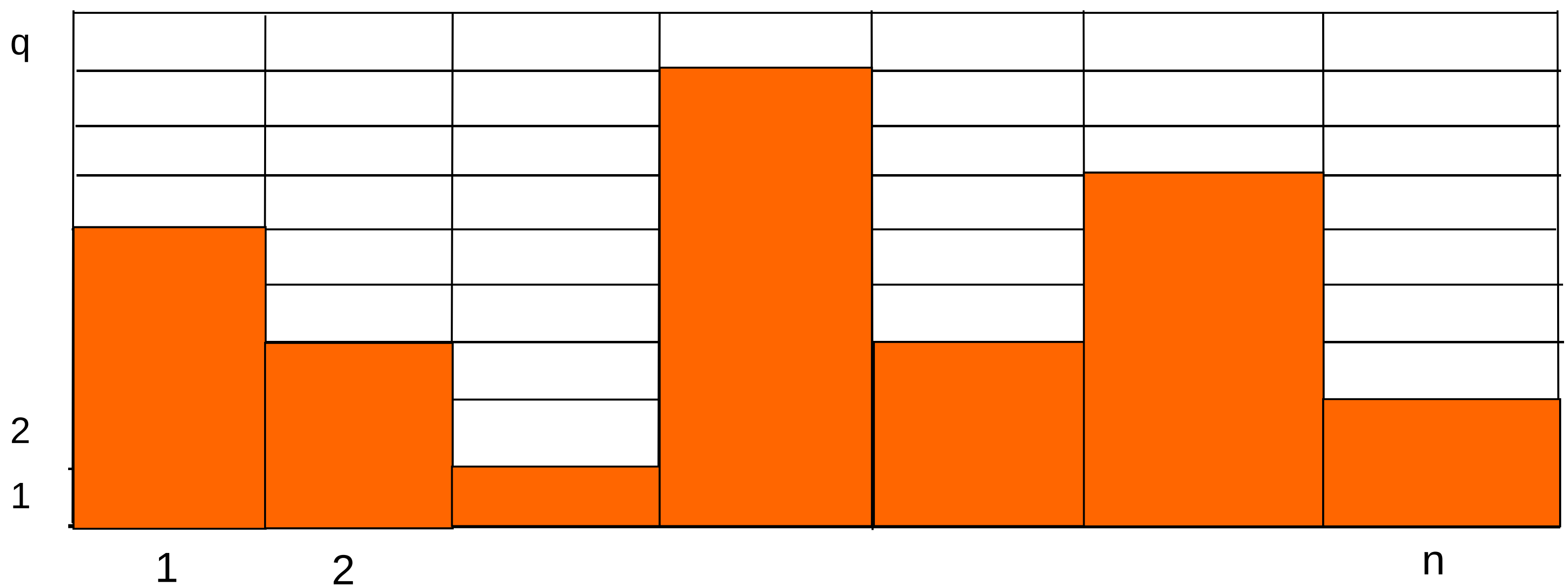
Data is written in blocks of n cells with q charge levels available in each cell



- Reading vs. writing
- To decrease the charge in a cell, need to erase the whole block
- Multistage writing procedures

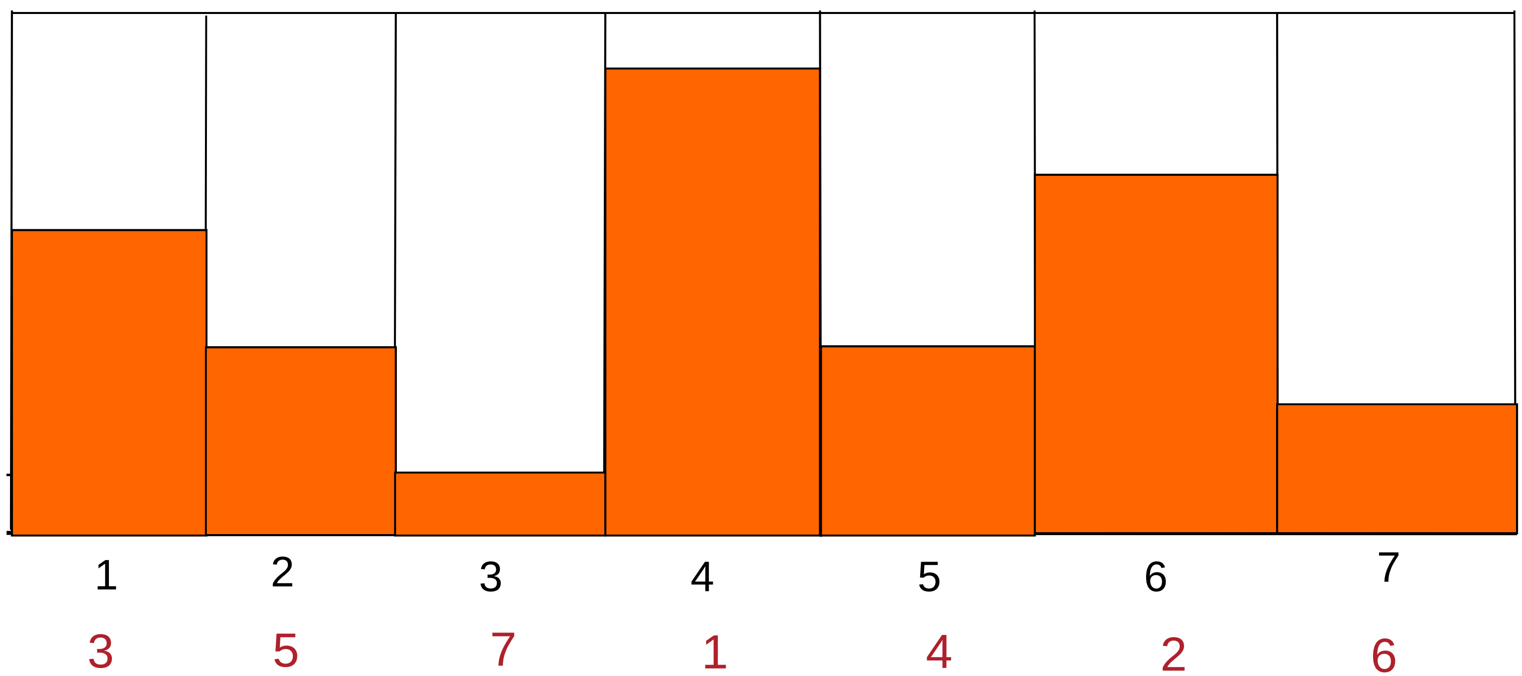
Storing data in flash memory

Alternative: Rank modulation scheme (ISIT'08 Jiang/Mateescu/Bruck)



Storing data in flash memory

Alternative: Rank modulation scheme (ISIT'08 Jiang/Mateescu/Bruck)

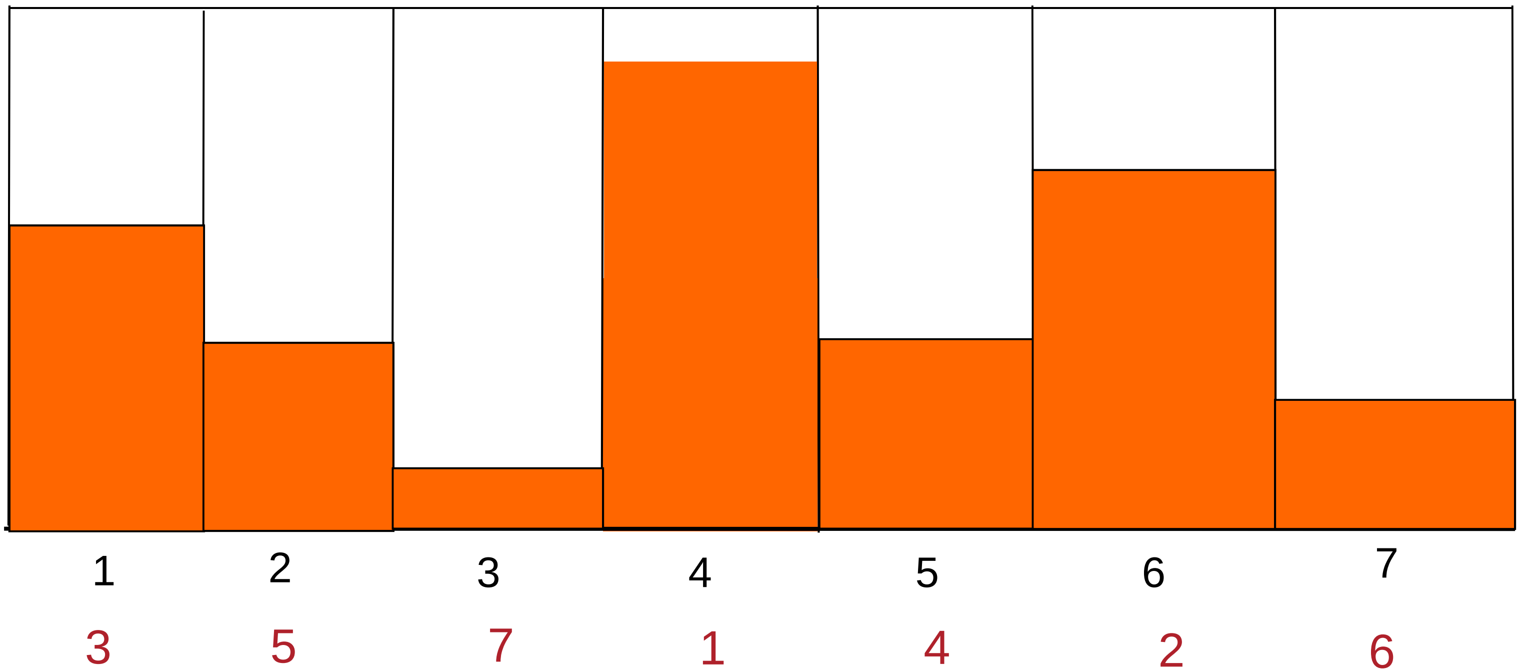


Store information as relative values of charges of cells:

charges can take continuous values; easier write procedure

Storing data in flash memory

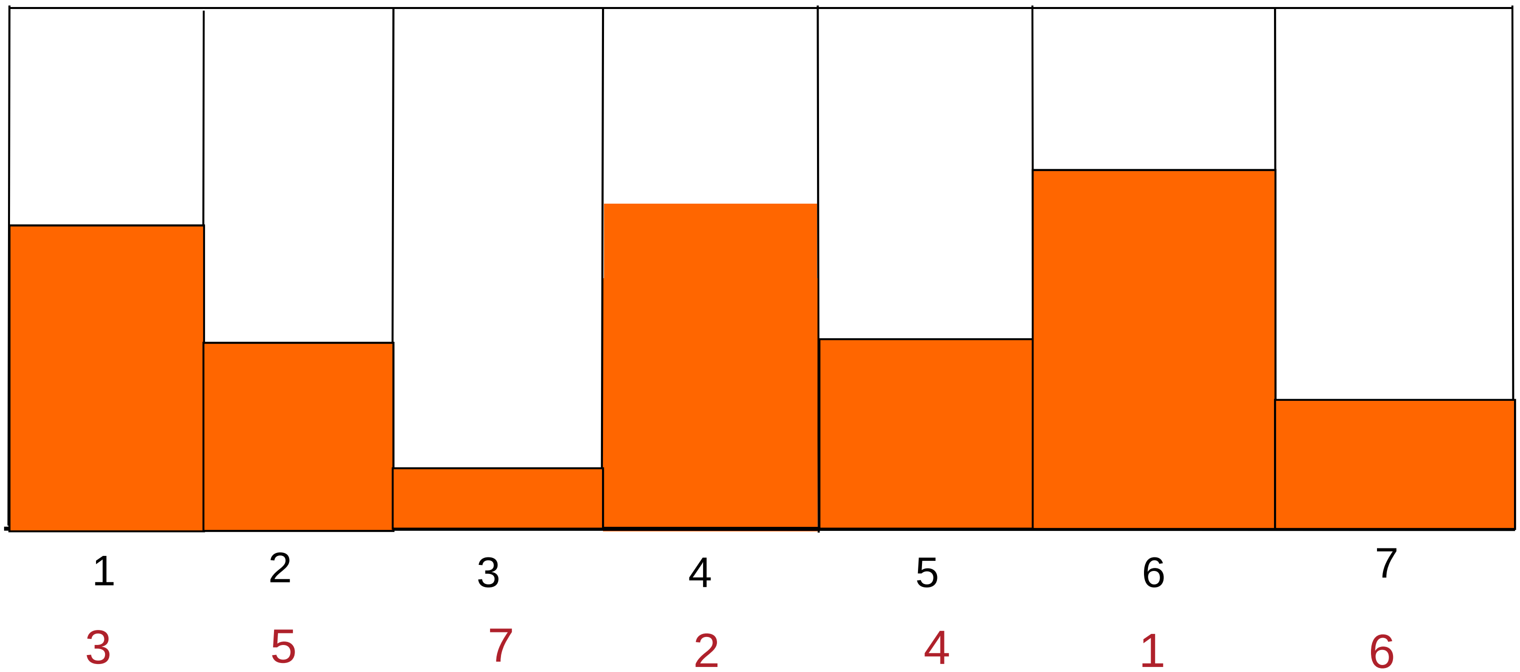
Alternative: Rank modulation scheme (ISIT'08 Jiang/Mateescu/Bruck)



$\sigma=(4,6,1,5,2,7,3)$

Storing data in flash memory

Alternative: Rank modulation scheme (ISIT'08 Jiang/Mateescu/Bruck)



Error process: charge leaks

$$\sigma=(4,6,1,5,2,7,3) \mapsto \sigma'=(6,4,1,5,2,7,3)$$

Rank modulation

Data protection in channels with impulse noise:

$$x_i=\gamma_i s_i+n_i$$

where γ_i is a gain constant

Transmit $s_i=Cr_i$, where $r=(r_1,r_2,\ldots,r_n)$ is the rank vector (permutation)

Elementary errors: transpositions of adjacent symbols

H. Chadwick and L. Kurz, Rank permutation group codes based on Kendall's correlation statistics, *IEEE Trans. Inform. Theory* **15**,1969

Codes in permutations

$\mathfrak{S}_n$={permutations on n symbols}

Code $C \subset \mathfrak{S}_n$ $\quad \sigma=(\sigma(1),\sigma(2),\ldots,\sigma(n))$

$$\begin{pmatrix} 1234 \\ 1324 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{pmatrix}$$

<u>Permutations form a ***group*** $\mathfrak{S}_n$:</u>

multiplication: (1324)(2413)=(2143)

inverse: (3421)(4312)=(1234)

Transposition: $135462 \mapsto 135264$

$135462 \mapsto 315462$

Codes in permutations

Discrepancy measures (metrics):

Hamming distance $d(\sigma,\tau)=\#\{i: \sigma(i)\neq\tau(i)\}$ Blake-Cohen-Deza 1979; Tarnanen 1989; Colbourn-Klove-Ling 2004; Cameron

Cayley distance (min number of transpositions)

many more

Our problem: Kendall tau distance (Maurice Kendall, 1930s "*Advanced Statistics*" Vol.1, 1946)

$$d(\sigma_1,\sigma_2)=\#\text{transpositions of } \textbf{\textit{adjacent}} \text{ symbols}$$

$d(2431,2314)=2$ $\quad$ $2431\to2341\to2314$

Coding in $\mathfrak{S}_n$ for the *Hamming distance* is a well-studied problem.
Coding for the *Kendall distance* is new

Coding for the Kendall distance

Properties

- $0 \le d_\tau(.\,,\,.) \le \frac{1}{2}\, n(n-1)$ $d(1234,4321)=6$
- Right invariance: $d_\tau(\sigma_1,\sigma_2)=d_\tau(\sigma_1\sigma,\sigma_2\sigma)$ for all $\sigma,\sigma_1,\sigma_2 \in \mathfrak{S}_n$
- "Weight'' of permutation $w(\sigma)=d_\tau(\sigma,e)$, e=identity permutation
- $d_\tau(\pi,\sigma)=d_\tau(\pi^{-1},\sigma^{-1})$

Rate of a code $C \subset \mathfrak{S}_n$ $R(C) = \frac{\ln M}{\ln n!}$ $0 \le R \le 1$

Capacity of rank permutation codes

$$\mathscr{C}(d) = \lim_{n\to\infty} \frac{\ln A(n,d)}{\ln n!}$$

where $A(n,d)=\max(|C|: C \subset \mathfrak{S}_n,\ d_\tau \ge d)$

Coding for the Kendall distance

- ***Theorem*** : The capacity of rank permutation codes is as follows

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

- ***Theorem*** :

Let $m = ((n-2)^{t+1} - 1)/(n-3)$, where $n-2$ is a power of a prime. There exists a t-error-correcting rank permutation code in $\mathfrak{S}_n$ whose size satisfies

$$M \geq \begin{cases} n!/(t(t+1)m) & (t \text{ odd}) \\ n!/(t(t+2)m) & (t \text{ even}). \end{cases}$$

Sphere packing bound gives $M=O(n!/n^t)$ for a code $C \subset \mathfrak{S}_n$ of length n that corrects t errors

- ***Singleton-type bound*** on A(n,d)

Proofs: arXiv0908.4094

Coding for the Kendall distance

- ***Theorem*** : The capacity of rank permutation codes is as follows

Margolius '01;
Louchard/Prodinger '03

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

known

NEW

known

- ***Theorem*** :

Let $m = ((n-2)^{t+1} - 1)/(n-3)$, where $n-2$ is a power of a prime. There exists a t-error-correcting rank permutation code in $\mathfrak{S}_n$ whose size satisfies

$$M \geq \begin{cases} n!/(t(t+1)m) & (t \text{ odd}) \\ n!/(t(t+2)m) & (t \text{ even}). \end{cases}$$

Sphere packing bound gives $M=O(n!/n^t)$ for a code $C \subset \mathfrak{S}_n$ of length n

Known: t=1: $M \geq ½(n-1)!$ (Jiang/Mateescu/Bruck, ISIT 2008)
(by the Varshamov-Tenenholtz construction)

Tools I: Basics on Permutations

Inversion in permutation: $\begin{array}{c} 1<2\ \ 3\ \ 4 \\ \sigma\ \ 2>1\ \ 4\ \ 3 \end{array}$

Inversion vector of a permutation:

σ 216437598

x_σ 010120201

Tools I: Basics on Permutations

Inversion in permutation: $\begin{array}{cccc} & 1<2 & 3 & 4 \\ \sigma & 2>1 & 4 & 3 \end{array}$

Inversion vector of a permutation:

x_σ 010120201

Tools I: Basics on Permutations

Inversion in permutation: $\quad 1<2 \;\; 3 \;\; 4$

$\sigma \quad 2>1 \;\; 4 \;\; 3$

Inversion vector of a permutation:

$x_\sigma \quad 010120201$

8

Tools I: Basics on Permutations

Inversion in permutation: $\begin{array}{cc} & 1<2\ \ 3\ \ 4 \\ \sigma & 2>1\ \ 4\ \ 3 \end{array}$

Inversion vector of a permutation:

x_σ 010120201

Tools I: Basics on Permutations

Inversion in permutation: $\begin{array}{c} 1<2\ \ 3\ \ 4 \\ \sigma\ \ 2>1\ \ 4\ \ 3 \end{array}$

Inversion vector of a permutation:

x_σ 010120201

598

Tools I: Basics on Permutations

Inversion in permutation: $\quad 1<2 \;\; 3 \;\; 4$

$\sigma \;\; 2>1 \;\; 4 \;\; 3$

Inversion vector of a permutation:

$x_\sigma \quad 010120201$

7598

Tools I: Basics on Permutations

Inversion in permutation: $\quad 1<2 \;\; 3 \;\; 4$

$\sigma \quad 2>1 \;\; 4 \;\; 3$

Inversion vector of a permutation:

$x_\sigma \quad 010120201$

37598

Tools I: Basics on Permutations

Inversion in permutation:

$$\begin{array}{cc} & 1<2\ \ 3\ \ 4 \\ \sigma & 2>1\ \ 4\ \ 3 \end{array}$$

Inversion vector of a permutation:

$$\begin{array}{ll} x_\sigma & 010120201 \\ \sigma & 216437598 \end{array}$$

Tools I: Basics on Permutations

Inversion in permutation: $\begin{array}{c} 1<2 \;\; 3 \;\; 4 \\ \sigma \;\; 2>1 \;\; 4 \;\; 3 \end{array}$

Inversion vector of a permutation:

σ	216437598	x_σ	010120201
x_σ	010120201	σ	216437598

Tools I: Basics on Permutations

Inversion in permutation: $\begin{array}{cc} & 1<2\ \ 3\ \ 4 \\ \sigma & 2>1\ \ 4\ \ 3 \end{array}$

Inversion vector of a permutation:

σ 216437598

x_σ 010120201

Inversion vector

$x_\sigma(i)=\#\{ j: j<i \wedge \sigma(j)>\sigma(i) \}$

$x_\sigma \in G_n \triangleq \{0\}\times\mathbb{Z}_2 \times \mathbb{Z}_3 \times \ldots \times \mathbb{Z}_n$

The mapping $\mathfrak{S}_n \to G_n$ is injective

Proposition. Let $I(\sigma)$ be the total number of inversions in σ. Then

$$w(\sigma) \triangleq d_T(\sigma, e) = I(\sigma) = \sum_{i-1}^{n-1} x_\sigma(i)$$

Direct attempt to construct sphere packing bounds:

Let $K_n(k)=|\{\sigma\subset\mathfrak{S}_n : I(\sigma)=k\}|$ $\mathbb{E}_{\mathfrak{S}_n}K_n(k)=n(n-1)/4$, $\mathrm{Var}(K_n(k))\approx n^3/36$

For $1\leq k\leq n$ $K_n(k)=\binom{n+k-2}{k}-\binom{n+k-3}{k-2}+\ldots$

$$K(z)=\sum_{k=0}^{\infty}K_n(k)z^k=\prod_{i=1}^{n}\frac{1-z^i}{1-z}.$$

$$K_n(k)=\frac{1}{2\pi i}\oint_C\prod_{\ell=1}^{n}\left(\frac{1-z^\ell}{1-z}\right)z^{-k-1}dz.$$

Theorem (Margolius '01; Louchard/Prodinger '03)

$K_n(k)\leq \exp(c_1 n)$ if $k=O(n)$

$K_n(k)=n!/\exp(c_2\, n)$ if $k=\Theta(n^2)$

Isometric embeddings?

1. $(\mathfrak{S}_n, d_\tau) \longrightarrow$ binary Hamming space of dimension n(n-1)/2

$(i,j) \subset [n] \times [n]$, $i<j$ — 1 if (i,j) forms an inversion, 0 o/w

Chadwick/Reed, 1970

2. $D(\sigma_1,\sigma_2)=\sum_{i=1}^{n}|\sigma_1(i)-\sigma_2(i)|$ "Spearman's footrule"

$\frac{1}{2}D(\sigma_1,\sigma_2) \leq d_\tau(\sigma_1,\sigma_2) \leq D(\sigma_1,\sigma_2)$ Diaconis-Graham '77

Tools II: Existence of Good Codes

1. Construct codes in the space $A_n = \mathbb{Z}_n^{n-1}$ that correct t additive errors

$$x,y \in C \quad x+e_1 \neq y+e_2 \qquad \text{if} \quad \sum_{j=1}^{n} |e_{i,j}| \leq t,\ i=1,2$$

(the errors are "symmetric", i.e., the known constructions of asymmetric ECCs do not apply.)

Theorem (Bose-Chowla '62): Let q be a power of a prime and $m = \frac{q^{t+1}-1}{q-1}$
There exist q+1 integers $j_0=0, j_1, \ldots, j_q \in \mathbb{Z}_m$ such that the sums

$$j_{i_1} + j_{i_2} + \cdots + j_{i_t} \quad (0 \leq i_1 \leq i_2 \leq \cdots \leq i_t \leq q)$$

are all different mod m.

2. Consider the space $G_n = \mathbb{Z}_2 \times \mathbb{Z}_3 \times \ldots \times \mathbb{Z}_n$ of inversion vectors. Compute the average intersection of the transaltions of the code constructed above with G_n.

Coding for the Kendall distance

- We establish the exact scaling law for code rate for codes with Kendall distance d (**capacity of rank permutation codes**)
- We prove **existence of good codes** (a constant factor away from the sphere packing bound) **for any fixed number** of Kendall **errors**
- **Singleton-type bound** for rank permutation codes