# Coding problems for ***memory*** and storage applications, II

Alexander Barg

University of Maryland

January 29, 2015

# Introduction: Rank modulation

Data storage in flash memories

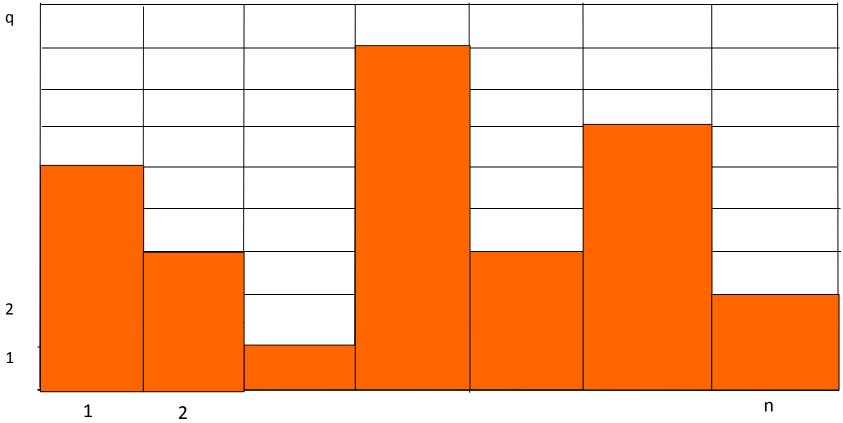
# Introduction: Rank modulation

Data storage in flash memories

**A. Jiang, R. Mateescu, M. Schwartz** and **J. Bruck**, Rank modulation for flash memories, in Proc. ISIT'08, 1731-5

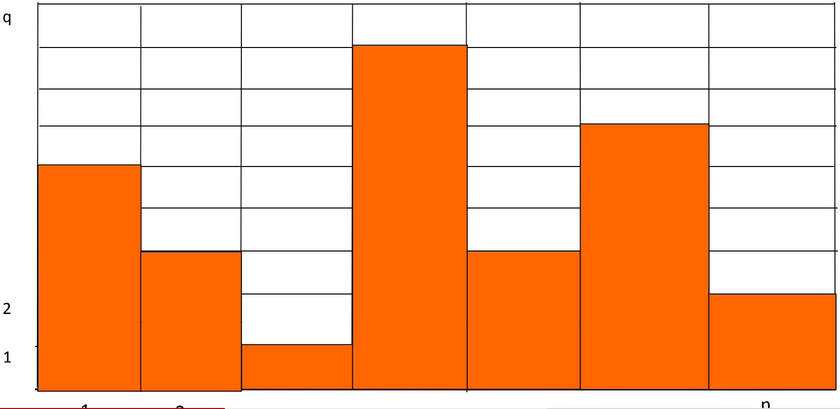
**A. Jiang, M. Schwartz and J. Bruck**, Error-correcting codes for rank modulation, in Proc. ISIT'08, pp. 1736-40

# Rank modulation



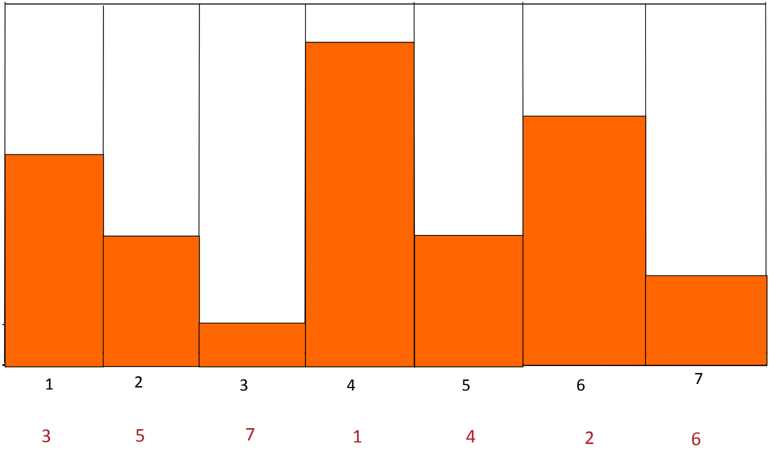
# Rank modulation

Instead of absolute values...



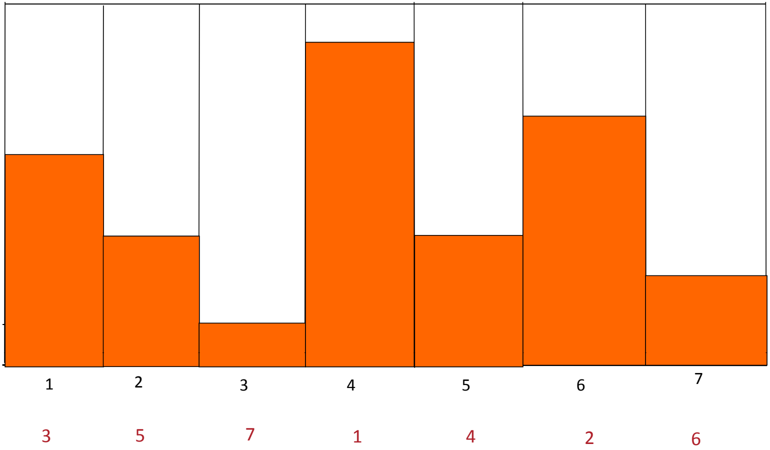
# Rank modulation

...store data as relative ranks of the cells



# Rank modulation

...store data as relative ranks of the cells



# Formalizing the coding problem

$\mathfrak{G} = \{$permutations on $n$ symbols$\}$

$$\begin{pmatrix} 1\,2\,3\,4 \\ 1\,3\,4\,2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$

Error process: charge leaks

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

Elementary errors: transposition of adjacent symbols

# Formalizing the coding problem

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

$$\begin{pmatrix} 1\,2\,3\,4 \\ 1\,3\,4\,2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$

Error process: charge leaks

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

Elementary errors: transposition of adjacent symbols

Transposition:

$$135462 \mapsto 135264$$
$$135462 \mapsto 315462$$

# Formalizing the coding problem

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

$$\begin{pmatrix} 1\,2\,3\,4 \\ 1\,3\,4\,2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$

Error process: charge leaks

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

Elementary errors: transposition of adjacent symbols

Transposition:

$$135462 \mapsto 135264$$
$$135462 \mapsto 315462$$

Permutations form a *group* $\mathfrak{S}_n$

multiplication: $(1234)(2413) = (2143)$
inverse: $(3421)(4312) = (1234)$

# Distances on Permutations

Hamming distance:

$$d(\sigma, \pi) = \sharp\{i : \sigma(i) \neq \tau(i)\}$$

(Blake-Cohen-Deza 1979, Tarnanen 1989, Colbourn-Kløve-Ling 2004)

# Distances on Permutations

Hamming distance:

$$d(\sigma,\pi)=\sharp\{i:\sigma(i)\neq\tau(i)\}$$

(Blake-Cohen-Deza 1979, Tarnanen 1989, Colbourn-Kløve-Ling 2004)

Ulam distance

$$d(\sigma,e)=\text{longest increasing subsequence in }\sigma$$

Milenkovic-Farnoud-Skachek 2013

# Distances on Permutations

Hamming distance:

$$d(\sigma, \pi) = \sharp\{i : \sigma(i) \neq \tau(i)\}$$

(Blake-Cohen-Deza 1979, Tarnanen 1989, Colbourn-Kløve-Ling 2004)

Ulam distance

$$d(\sigma, e) = \text{longest increasing subsequence in } \sigma$$

Milenkovic-Farnoud-Skachek 2013

$\ell_\infty$ or Chebyshev distance

# Distances on Permutations

Hamming distance:

$$d(\sigma,\pi)=\sharp\{i:\sigma(i)\neq\tau(i)\}$$

(Blake-Cohen-Deza 1979, Tarnanen 1989, Colbourn-Kløve-Ling 2004)

Ulam distance

$$d(\sigma,e)=\text{longest increasing subsequence in }\sigma$$

Milenkovic-Farnoud-Skachek 2013

$\ell_\infty$ or Chebyshev distance

Kendall $\tau$ distance $d(\sigma,\pi)=n-$minimum number of adjacent transpositions $\sigma\to\pi$

# Distances on Permutations

Hamming distance:

$$d(\sigma, \pi) = \sharp\{i : \sigma(i) \neq \tau(i)\}$$

(Blake-Cohen-Deza 1979, Tarnanen 1989, Colbourn-Kløve-Ling 2004)

Ulam distance

$$d(\sigma, e) = \text{longest increasing subsequence in } \sigma$$

Milenkovic-Farnoud-Skachek 2013

$\ell_\infty$ or Chebyshev distance

Kendall $\tau$ distance $d(\sigma, \pi) = n-$minimum number of adjacent transpositions $\sigma \to \pi$

Many other metrics on permutations (see Deza-Huang, Metrics on permutations, a survey, 1998)

# Properties of the Kendall distance

- $0 \le d(\cdot,\cdot) \le {}^1/_2 n(n-1) \quad d(1234, 4321) = 6$
- Right invariance: $d(\sigma_1, \sigma_2) = d(\sigma_1\sigma, \sigma_2\sigma)$ for all $\sigma, \sigma_1, \sigma_2$
- "Weight" of permutation $w(\sigma) = d(\sigma, e)$, $e =$identity permutation

# Problems

- What is the maximum size of a subset $\mathcal{C} \subset \mathfrak{G}$ such that the minimum distance is at least $d$?

# Problems

- What is the maximum size of a subset $\mathcal{C} \subset \mathfrak{S}$ such that the minimum distance is at least $d$?

  Rate of a code $\mathcal{C} \subset \mathfrak{S}_n$ :

$$R(\mathcal{C}) = \frac{\ln |\mathcal{C}|}{\ln n!}; 0 \le R \le 1$$

  Capacity of rank modulation codes:

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

  where $A(n,d)$ is the cardinality of the largest code in $\mathfrak{S}_n$ with distance $d$

# Problems

- What is the maximum size of a subset $\mathcal{C} \subset \mathfrak{S}$ such that the minimum distance is at least $d$?

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

  Capacity of rank modulation codes:
  $$\mathscr{C}(d) = \lim_{n\to\infty} \frac{\ln A(n,d)}{\ln n!}$$

  where $A(n,d)$ is the cardinality of the largest code in $\mathfrak{S}_n$ with distance $d$

- Can we construct good codes and decode them?

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

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

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

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

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

$$\begin{array}{ll} x_\sigma & 010120201 \\ \sigma & \qquad\qquad\quad 8 \end{array}$$

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

$$\begin{array}{ll} x_\sigma & 010120201 \\ \sigma & \qquad\qquad 98 \end{array}$$

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

$$\begin{array}{ll} x_\sigma & 010120201 \\ \sigma & \qquad\quad 598 \end{array}$$

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

$$\begin{array}{ll} x_\sigma & 010120201 \\ \sigma & \quad\quad 7598 \end{array}$$

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

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

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

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

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

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

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

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

$$x_\sigma \in \mathcal{G}_n := \{0\} \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \dots \mathbb{Z}_n$$

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

# Combinatorics of permutations: Inversion vectors

Inversion in permutation:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ \sigma & 2 & 1 & 3 & 4 \end{array}$$

Inversion vector of a permutation:

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

$$x_\sigma \in \mathcal{G}_n := \{0\} \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \ldots \mathbb{Z}_n$$

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

Proposition: Let $I(\sigma)$ be the total number of inversions in $\sigma$. Then
$w(\sigma) = I(\sigma) = \sum_{i=1}^{n-1} x_\sigma(i)$ $d(\sigma, \tau) = w(\sigma\tau^{-1})$

# Coding for the Kendall metric

### Theorem

*Let* $A(n,d) =$ *max size of a code with distance* $d$*. Define*

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

*Then*

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

Proof by estimating the number of permutations with a given number of inversions.

# Coding for the Kendall metric

**Theorem**

*Let $A(n,d) =$ max size of a code with distance $d$. Define*

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

*Then*

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

Proof by estimating the number of permutations with a given number of inversions.

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

$$\frac{1}{2} D(\sigma_1,\sigma_2) \le d(\sigma_1^{-1},\sigma_2^{-1}) \le D(\sigma_1,\sigma_2)$$

# Constructing codes for the Kendall metric

Let $J : \mathcal{G}_n \to \mathfrak{S}_n$ take $x_\sigma$ back to $\sigma$
E.g. $J((10103101)) = (2,1,6,4,3,7,5,9,8)$.

# Constructing codes for the Kendall metric

Let $J : \mathcal{G}_n \to \mathfrak{S}_n$ take $x_\sigma$ back to $\sigma$
E.g. $J((10103101)) = (2, 1, 6, 4, 3, 7, 5, 9, 8)$.

**Lemma**

$d_\tau(\sigma_1, \sigma_2) \ge d_{\ell_1}(x_{\sigma_1}, x_{\sigma_2})$

## Constructing codes for the Kendall metric

Gray map: $\mathbb{Z}_{2^s} \to \{0,1\}^s$
$u \in [0, 2^s - 1]$
$u = (b_{s-1}, b_{s-2}, \ldots, b_1, b_0) \mapsto \phi_s(u) = (g_{s-1}, g_{s-2}, \ldots, g_0)$,

where $g_j = (b_j + b_{j+1}) \mod 2 \qquad (j = 0, 2, \ldots s-1; b_s \triangleq 0)$

# Constructing codes for the Kendall metric

Gray map: $\mathbb{Z}_{2^s} \to \{0,1\}^s$
$u \in [0, 2^s - 1]$
$u = (b_{s-1}, b_{s-2}, \dots, b_1, b_0) \mapsto \phi_s(u) = (g_{s-1}, g_{s-2}, \dots, g_0),$

where $g_j = (b_j + b_{j+1}) \mod 2 \qquad (j = 0, 2, \dots s-1;\, b_s \triangleq 0)$

$$
\begin{array}{r}
0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,1\\
0\,0\,0\,0\,0\,0\,1\,0\\
0\,0\,0\,0\,0\,0\,1\,1\\
\hline
0\,0\,0\,0\,0\,1\,0\,0\\
1\,0\,1\\
1\,1\,0\\
1\,1\,1\\
\hline
1\,0\,0\,0\\
1\,0\,0\,1
\end{array}
\quad \longrightarrow \quad
\begin{array}{r}
0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,1\\
0\,0\,0\,0\,0\,0\,1\,1\\
0\,0\,0\,0\,0\,0\,1\,0\\
\hline
0\,0\,0\,0\,0\,1\,1\,0\\
0\,0\,0\,0\,0\,1\,1\,1\\
0\,0\,0\,0\,0\,1\,0\,1\\
1\,0\,0\\
\hline
1\,1\,0\,0\\
1\,1\,0\,1
\end{array}
$$

Gray map is reflective

# Inversion vectors; Gray map

Let

$$m_i = \lfloor \log_2 i \rfloor, \quad i = 1, \dots, n$$

Inverse Gray map

$$\psi_i : \{0,1\}^{m_i} \to [0, i-1]$$

# Inversion vectors; Gray map

Let

$$m_i = \lfloor \log_2 i \rfloor, \quad i = 1, \ldots, n$$

Inverse Gray map

$$\psi_i : \{0,1\}^{m_i} \to [0, i-1]$$

Consider a vector $x = (x_2 | x_3 | \ldots | x_n)$, where $x_i \in \{0,1\}^{m_i}$, $i = 2, \ldots, n$.

# Inversion vectors; Gray map

Let

$$m_i = \lfloor \log_2 i \rfloor, \quad i = 1, \ldots, n$$

Inverse Gray map

$$\psi_i : \{0,1\}^{m_i} \to [0, i-1]$$

Consider a vector $x = (x_2|x_3|\ldots|x_n)$, where $x_i \in \{0,1\}^{m_i}$, $i = 2, \ldots, n$.

$$m \triangleq \dim(x) = \sum_{i=1}^{n} m_i = (n+1)m_n - 2^{m_n+1} + 2$$

## Inversion vectors; Gray map

Let

$$m_i = \lfloor \log_2 i \rfloor, \quad i = 1, \ldots, n$$

Inverse Gray map

$$\psi_i : \{0,1\}^{m_i} \to [0, i-1]$$

Consider a vector $x = (x_2|x_3|\ldots|x_n)$, where $x_i \in \{0,1\}^{m_i}$, $i = 2, \ldots, n$.

$$m \triangleq \dim(x) = \textstyle\sum_{i=1}^n m_i = (n+1)m_n - 2^{m_n+1} + 2$$

Given a vector $x \in \{0,1\}^m$ let $\Psi(x) = \Psi(x_2|x_3|\ldots|x_n) = (\psi_2(x_2), \ldots, \psi_n(x_n))$.

**Proposition**

*Let $x, y \in \{0,1\}^m$. Then*

$$d_{\ell_1}(\Psi(x), \Psi(y)) \geq d_H(x, y),$$

# The Construction

**Theorem**

*Let $\mathcal{C}$ be a binary code of length n, cardinality M and minimum Hamming distance d, where $m = (n+1)\lfloor \log n \rfloor - 2^{\lfloor \log n \rfloor + 1} + 2$. Then the set of permutations*

$$\mathcal{C}_\tau = \{\pi \in \mathfrak{S}_n : \pi = J(\Psi(x)), x \in \mathcal{C}\}$$

*forms a rank modulation code on n elements of size M and distance at least d in the Kendall space $\mathcal{X}_n$.*

## The Construction

### Theorem

*Let $\mathcal{C}$ be a binary code of length n, cardinality M and minimum Hamming distance d, where $m = (n+1)\lfloor \log n \rfloor - 2^{\lfloor \log n \rfloor + 1} + 2$. Then the set of permutations*

$$\mathcal{C}_\tau = \{\pi \in \mathfrak{S}_n : \pi = J(\Psi(x)), x \in \mathcal{C}\}$$

*forms a rank modulation code on n elements of size M and distance at least d in the Kendall space $\mathcal{X}_n$.*

Example: Let $\mathcal{C}$ be a BCH code of length $n$ and designed distance $2t+1$. We obtain a rank modulation code of cardinality $M \geq 2^m/(m+1)^t$ that corrects $t$ errors.

# The Construction

### Theorem

*Let $\mathcal{C}$ be a binary code of length n, cardinality M and minimum Hamming distance d, where $m = (n+1)\lfloor \log n \rfloor - 2^{\lfloor \log n \rfloor + 1} + 2$. Then the set of permutations*

$$\mathcal{C}_\tau = \{\pi \in \mathfrak{S}_n : \pi = J(\Psi(x)), x \in \mathcal{C}\}$$

*forms a rank modulation code on n elements of size M and distance at least d in the Kendall space $\mathcal{X}_n$.*

Example: Let $\mathcal{C}$ be a BCH code of length $n$ and designed distance $2t+1$. We obtain a rank modulation code of cardinality $M \geq 2^m/(m+1)^t$ that corrects $t$ errors.

$n = 63, m = 253$

| | | | | |
|---|---|---|---|---|
| $\log M$ | 247 | 239 | 231 | 223 ... |
| $t$ | 1 | 2 | 3 | 4 ... |

# The Construction

**Theorem**

*Let $\mathcal{C}$ be a binary code of length n, cardinality M and minimum Hamming distance d, where $m = (n+1)\lfloor \log n \rfloor - 2^{\lfloor \log n \rfloor + 1} + 2$. Then the set of permutations*

$$\mathcal{C}_\tau = \{\pi \in \mathfrak{S}_n : \pi = J(\Psi(x)), x \in \mathcal{C}\}$$

*forms a rank modulation code on n elements of size M and distance at least d in the Kendall space $\mathcal{X}_n$.*

Example: Let $\mathcal{C}$ be a BCH code of length $n$ and designed distance $2t+1$. We obtain a rank modulation code of cardinality $M \geq 2^m/(m+1)^t$ that corrects $t$ errors.

$n = 63, m = 253$

| | | | | |
|---|---|---|---|---|
| $\log M$ | 247 | 239 | 231 | 223 ... |
| $t$ | 1 | 2 | 3 | 4 ... |

Simple decoding algorithms

# Some references

Codes in permutations and error correction for rank modulation, with A. Mazumdar, arXiv:0908.4094

Constructions of rank modulation codes, with A. Mazumdar and G. Zémor, IT Trans. Feb. 2013