Multipermutation Codes in the Ulam Metric

Farzad Farnoud† Olgica Milenkovic*

†California Institute of Technology[1]

*University of Illinois at Urbana-Champaign

ISIT 2014, Honolulu

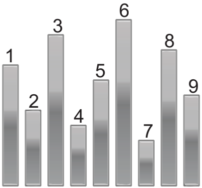
[1] Farzad Farnoud was with the University of Illinois at Urbana-Champaign.

Summary

- Novel rank modulation multipermutation codes (MPCs) in Ulam metric
- Codes correcting translocation and deletion errors
- Highlight connection between MPCs in the Ulam and Hamming metrics
- Capacities or bounds for MPCs in both metrics
- Constructions using Steiner systems, BIBDs, and interleaving
- Efficienct decoding algorithms

Rank Modulation for Flash Memory

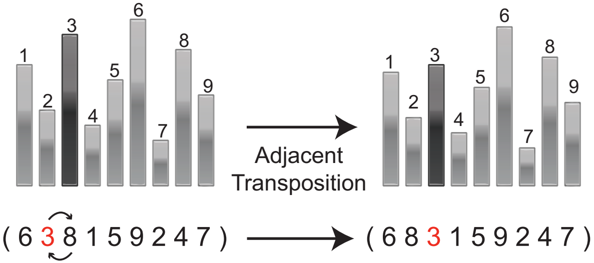
- Rank modulation for flash memory was proposed by Jiang et al. [2008] for dealing with over-injection and charge leakage.
  - In an array of cells, each cell stores a charge level.
  - Information stored in relative values of charge levels.
  - Data encoded as permutations in blocks of cells.



Permutation: (6 3 8 1 5 9 2 4 7)

Errors in Rank Modulation: Adjacent Transpositions

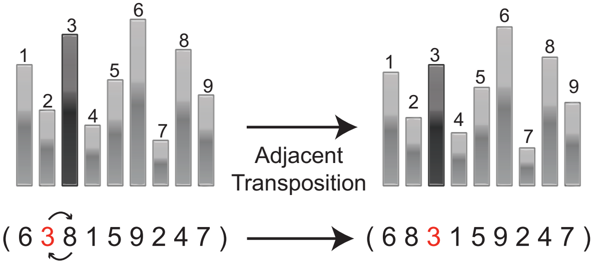
- Retention errors (charge leakage), cycling errors, and write-disturb errors often modeled as adjacent transpositions.



Studied by Barg, Bruck, Cassuto, Hagiwara, Jiang, Mazumdar, Schwartz, Wang, Yaakobi, Zhou,...

Errors in Rank Modulation: Adjacent Transpositions

- Retention errors (charge leakage), cycling errors, and write-disturb errors often modeled as adjacent transpositions.



- Correcting t adjacent transposition $\Longleftrightarrow$ min Kendall tau distance $\geq 2t+1$.

Studied by Barg, Bruck, Cassuto, Hagiwara, Jiang, Mazumdar, Schwartz, Wang, Yaakobi, Zhou,...

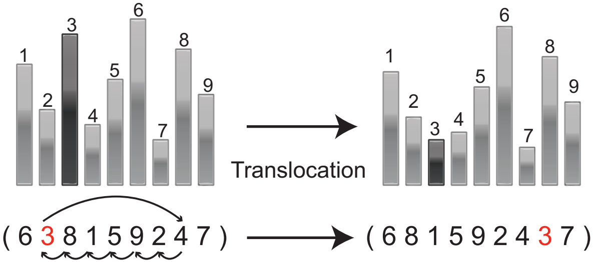
Errors in Rank Modulation: Translocations

- Increasing number of charge levels to increase capacity leads to larger charge fluctuations relative to gap between charge levels.

Farnoud, Skachek, Milenkovic, Transactions on Information Theory, May 2013

Errors in Rank Modulation: Translocations

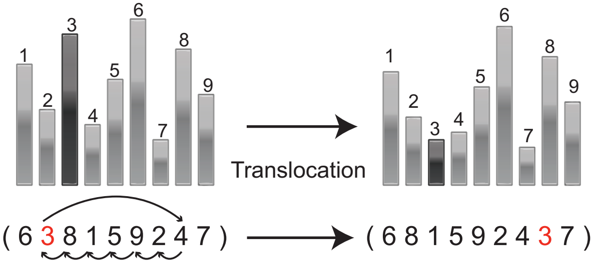
- Increasing number of charge levels to increase capacity leads to larger charge fluctuations relative to gap between charge levels.
- Large magnitude error leading to a translocation:



Farnoud, Skachek, Milenkovic, Transactions on Information Theory, May 2013

Errors in Rank Modulation: Translocations

- Increasing number of charge levels to increase capacity leads to larger charge fluctuations relative to gap between charge levels.
- Large magnitude error leading to a translocation:

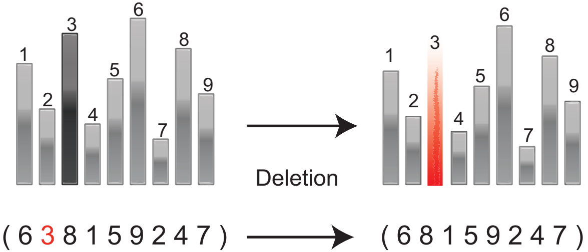


- Correcting t translocations $\iff$ min Ulam distance $\geq 2t + 1$.
- Ulam distance = length – length of longest common subsequence (LCS).

Farnoud, Skachek, Milenkovic, Transactions on Information Theory, May 2013

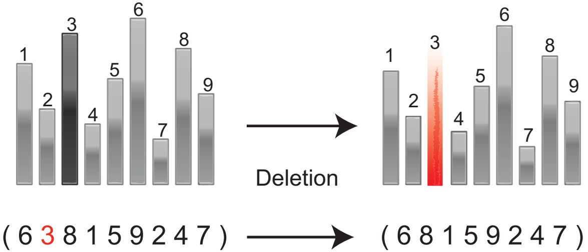
Errors in Rank Modulation: Deletions

- In harsh situations, such as high P/E cycles, transistor failure, a cell may become unreadable: deletion.



Errors in Rank Modulation: Deletions

- In harsh situations, such as high P/E cycles, transistor failure, a cell may become unreadable: deletion.



- Correcting t deletions $\iff$ min Ulam distance $\geq t+1$.

Multipermutation Codes I

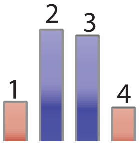
- Multipermutation (MP): a rearrangement of elements of a multiset: $[2, 1, 1, 2]$ for the multiset $\{1, 1, 2, 2\}$.

Multipermutation Codes I

- Multipermutation (MP): a rearrangement of elements of a multiset: $[2, 1, 1, 2]$ for the multiset $\{1, 1, 2, 2\}$.
- Each r cells have the same rank: programmed with roughly same charge levels.

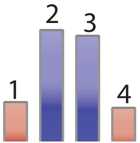
Multipertation Codes I

- Multipermutation (MP): a rearrangement of elements of a multiset: $[2, 1, 1, 2]$ for the multiset $\{1, 1, 2, 2\}$.
- Each r cells have the same rank: programmed with roughly same charge levels.
- Example: $r = 2$, cells 2 and 3 have rank 1 and cells 1 and 4 have rank 2: MP [2,1,1,2]



Multipermutation Codes I

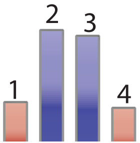
- Multipermutation (MP): a rearrangement of elements of a multiset: $[2,1,1,2]$ for the multiset $\{1,1,2,2\}$.
- Each r cells have the same rank: programmed with roughly same charge levels.
- Example: $r = 2$, cells 2 and 3 have rank 1 and cells 1 and 4 have rank 2: MP [2,1,1,2]



- r-regular MPs: each element (rank) appears r times.

Multipermutation Codes I

- Multipermutation (MP): a rearrangement of elements of a multiset: $[2, 1, 1, 2]$ for the multiset $\{1, 1, 2, 2\}$.
- Each r cells have the same rank: programmed with roughly same charge levels.
- Example: $r = 2$, cells 2 and 3 have rank 1 and cells 1 and 4 have rank 2: MP [2,1,1,2]



- r-regular MPs: each element (rank) appears r times.
- Beneficial since the number of possible ranks is limited.

Multipermutation Codes II

- In the literature:
  - Multipermutation re-write codes [En Gad'12]
  - Multipermutation codes in Chebyshev metric [Shieh'10,'11]
  - Multipermutation codes in Kendall tau metric [Buzaglo'13] [Sala'13]
  - Multipermutation codes in Hamming metric [Luo'03] [Ding'05] [Huczynska'06] [Chu'06]. Aka, constant composition codes, frequency permutation codes.
  - Concatenated permutations [Heymann'13].

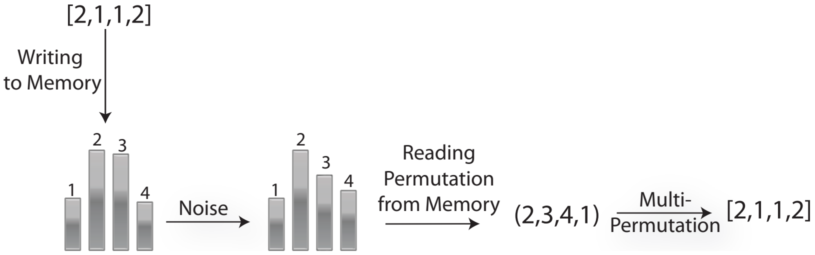
Multipermutation Codes II

- In the literature:
  - Multipermutation re-write codes [En Gad'12]
  - Multipermutation codes in Chebyshev metric [Shieh'10,'11]
  - Multipermutation codes in Kendall tau metric [Buzaglo'13] [Sala'13]
  - Multipermutation codes in Hamming metric [Luo'03] [Ding'05] [Huczynska'06] [Chu'06]. Aka, constant composition codes, frequency permutation codes.
  - Concatenated permutations [Heymann'13].
- This work: codes in the Ulam metric for correcting translocation and deletion errors.

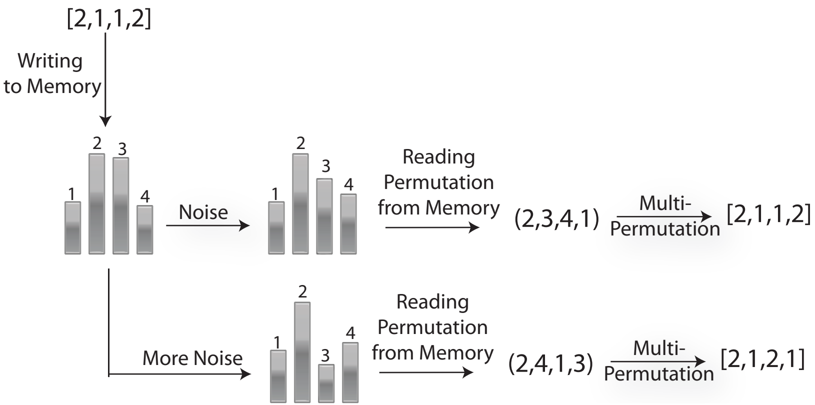
Multipermutation Codes II

- In the literature:
  - Multipermutation re-write codes [En Gad'12]
  - Multipermutation codes in Chebyshev metric [Shieh'10,'11]
  - Multipermutation codes in Kendall tau metric [Buzaglo'13] [Sala'13]
  - Multipermutation codes in Hamming metric [Luo'03] [Ding'05] [Huczynska'06] [Chu'06]. Aka, constant composition codes, frequency permutation codes.
  - Concatenated permutations [Heymann'13].
- This work: codes in the Ulam metric for correcting translocation and deletion errors.
- We consider permutations and multipermutations simultanously by considering equivalence classes of multipermutations.

Multipermutation Codes III

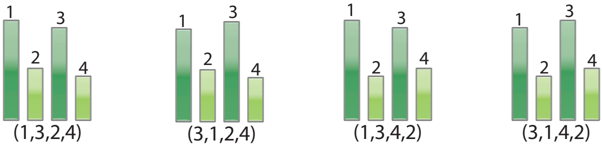


Multipermutation Codes III



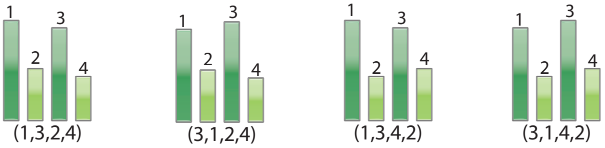
Permutations and Multipermutations

- Same information (same MP): $[1, 2, 1, 2]$:



Permutations and Multipermutations

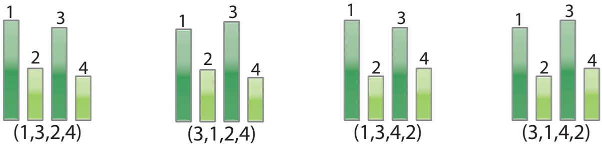
- Same information (same MP): $[1, 2, 1, 2]$:



- r-regular MPs divides $\mathbb{S}_n$ into equivalence classes.

Permutations and Multipermutations

- Same information (same MP): $[1, 2, 1, 2]$:



- r-regular MPs divides $\mathbb{S}_n$ into equivalence classes.

- $R_r(\pi)$: equivalence class of π

$$R_2(1,3,2,4) = \{(1,3,2,4),(3,1,2,4),(1,3,4,2),(3,1,4,2)\}.$$

Multipermutation Codes

- In multipermutation coding:
  - information stored in terms of MPs,
  - but physical error process described in terms of permutations of change orderings.

Multipermutation Codes

- In multipermutation coding:
  - information stored in terms of MPs,
  - but physical error process described in terms of permutations of change orderings.
- An r-regular MP code of length n, MPC(n, r), is a subset C of $\mathbb{S}_n$ such that if $\pi \in C$, then $R_r(\pi) \subseteq C$.

Multipermutation Codes

- In multipermutation coding:
  - information stored in terms of MPs,
  - but physical error process described in terms of permutations of change orderings.
- An r-regular MP code of length n, MPC(n, r), is a subset C of $\mathbb{S}_n$ such that if $\pi \in C$, then $R_r(\pi) \subseteq C$.
- Size of C is the number of equivalence classes it contains.

- An MPC(n, r) has minimum Ulam distance d if for all π and σ not in the same equivalence class, $d_u(\pi, \sigma) \geq d$.

Multipermutation Codes in Ulam and Hamming Metrics

- An MPC(n, r) has minimum Ulam distance d if for all π and σ not in the same equivalence class, $d_u(\pi, \sigma) \geq d$.
  - Corrects t translocation errors iff $d \geq 2t + 1$.

Multipermutation Codes in Ulam and Hamming Metrics

- An MPC(n, r) has minimum Ulam distance d if for all π and σ not in the same equivalence class, $d_u(\pi, \sigma) \geq d$.
  - Corrects t translocation errors iff $d \geq 2t + 1$.
- An MPC(n, r) has min Hamming distance d if for all π, σ the Hamming distance between their MP representation is $\geq d$.

Multipermutation Codes in Ulam and Hamming Metrics

- An MPC(n, r) has minimum Ulam distance d if for all π and σ not in the same equivalence class, $d_u(\pi, \sigma) \geq d$.
  - Corrects t translocation errors iff $d \geq 2t + 1$.
- An MPC(n, r) has min Hamming distance d if for all π, σ the Hamming distance between their MP representation is $\geq d$.
- An MPC with minimum Ulam distance d is an MPC with minimum Hamming distance at least d.

Multipermutation Codes in Ulam and Hamming Metrics

- An MPC(n, r) has minimum Ulam distance d if for all π and σ not in the same equivalence class, $d_u(\pi, \sigma) \geq d$.
  - Corrects t translocation errors iff $d \geq 2t + 1$.
- An MPC(n, r) has min Hamming distance d if for all π, σ the Hamming distance between their MP representation is $\geq d$.
- An MPC with minimum Ulam distance d is an MPC with minimum Hamming distance at least d.
- We present construction of Ulam codes using codes in Hamming metric.

Capacity of MP Codes in the Hamming Metric

- In a given metric, let $A(n, r, d)$ denote the maximum size of an MP code of length n, regularity r, and minimum distance d.

Capacity of MP Codes in the Hamming Metric

- In a given metric, let $A(n, r, d)$ denote the maximum size of an MP code of length n, regularity r, and minimum distance d.
- The capacity is defined as

$$\mathcal{C}\left(r, d\right) = \lim_{n\to\infty} \frac{\ln A\left(n, r, d\right)}{\ln n!}.$$

Capacity of MP Codes in the Hamming Metric

- In a given metric, let $A(n, r, d)$ denote the maximum size of an MP code of length n, regularity r, and minimum distance d.
- The capacity is defined as

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

Theorem [FM2014]

The capacity of multipermutation codes in the Hamming metric with parameters r and d, with $\rho = \lim_{n\to\infty} \frac{\ln r}{\ln n}$ and $\delta = \lim_{n\to\infty} \frac{d}{n}$, is given by

$$\mathcal{C}_H(r, d) = (1 - \rho)(1 - \delta).$$

Capacity of MP Codes in the Ulam Metric

- In a given metric, let $A(n, r, d)$ denote the maximum size of an MP code of length n, regularity r, and minimum distance d.
- The capacity is defined as

$$\mathcal{C}\left(r,d\right)=\lim_{n\to\infty}\frac{\ln A\left(n,r,d\right)}{\ln n!}.$$

Theorem [FM2014]

The capacity of multipermutation codes in the Ulam metric with parameters r and d, with $\rho = \lim_{n\to\infty}\frac{\ln r}{\ln n}$ and $\delta = \lim_{n\to\infty}\frac{d}{n}$, is given by

$$\left(1-2\rho\right)\left(1-\delta\right)\leq\mathcal{C}_U\left(r,d\right)\leq\left(1-\rho\right)\left(1-\delta\right).$$

Code Construction: Almost-disjoint Sets

Lemma

Let C be an MPC(n, r), and $2t < r$. If for all $\pi, \sigma \in C$, each rank of π and σ are either identical or have less than $r - 2t$ elements in common, then C can correct t translocation errors.

- Simple example for $t = 1$, $r = 6$, $n = 12$ (in MP form)

$$
\begin{aligned}
C = \{ & [1,1,1,1,1,1,2,2,2,2,2,2], \\
& [1,1,1,2,2,2,1,1,1,2,2,2], \\
& \quad \dots \\
& [2,2,2,2,2,2,1,1,1,1,1,1]\}.
\end{aligned}
$$

The intersection between each two ranks is of size $3 < 4 = r - 2t$.

Steiner Systems and BIBDs

- A k-(n, r, λ)-*design* is a family of r-subsets of a set X of size n, each called a *block*, such that every k-subset of X appears in exactly λ blocks.

Steiner Systems and BIBDs

- A k-(n, r, λ)-*design* is a family of r-subsets of a set X of size n, each called a *block*, such that every k-subset of X appears in exactly λ blocks.
- Such a design is *resolvable* if its blocks can be grouped into m classes, such that each class forms a partition of X.

Steiner Systems and BIBDs

- A k-(n, r, λ)-*design* is a family of r-subsets of a set X of size n, each called a *block*, such that every k-subset of X appears in exactly λ blocks.
- Such a design is *resolvable* if its blocks can be grouped into m classes, such that each class forms a partition of X.
- A *Steiner system* $S(k, r, n)$ is a k-$(n, r, 1)$-design.

Steiner Systems and BIBDs

- A k-(n, r, λ)-*design* is a family of r-subsets of a set X of size n, each called a *block*, such that every k-subset of X appears in exactly λ blocks.
- Such a design is *resolvable* if its blocks can be grouped into m classes, such that each class forms a partition of X.
- A *Steiner system* $S(k, r, n)$ is a k-$(n, r, 1)$-design.
- A Balanced incomplete block design (BIBD) with parameters (n, r, λ) is a 2-(n, r, λ)-design.

Code Construction: Resolvable Steiner Systems

Proposition

If a resolvable Steiner system $S(k, r, n)$ exists, then for odd $d \leq r - k + 1$, there exists an MPC(n, r) with min Ulam distance d, of size

$$\frac{\binom{n-1}{k-1}}{\binom{r-1}{k-1}} \left(\frac{n}{r}\right)!.$$

- Proof outline:
  - The blocks in a class of the Steiner system are assigned as the elements of the ranks of the multipermutation.
  - Each two blocks have less than k elements in common.
  - Size of code follows from the number of classes and the fact that blocks can be assigned to ranks arbitrarily.

Code Construction: Resolvable BIBDs

- A resolvable BIBD is a resolvable Steiner system where every $k = 2$ elements occur in only $\lambda = 1$ block.

Code Construction: Resolvable BIBDs

- A resolvable BIBD is a resolvable Steiner system where every $k = 2$ elements occur in only $\lambda = 1$ block.
- For prime r and $n = r^2$, [Khare,Federer'81] give a simple construction.

Code Construction: Resolvable BIBDs

- A resolvable BIBD is a resolvable Steiner system where every $k = 2$ elements occur in only $\lambda = 1$ block.
- For prime r and $n = r^2$, [Khare,Federer'81] give a simple construction.
  - Example: $r = 3$, each row is a block, each table is a class.

1	2	3
4	5	6
7	8	9

1	4	7
2	5	8
3	6	9

1	5	9
2	6	7
3	4	8

1	6	8
2	4	9
3	5	7

Code Construction: Resolvable BIBDs

- A resolvable BIBD is a resolvable Steiner system where every $k = 2$ elements occur in only $\lambda = 1$ block.
- For prime r and $n = r^2$, [Khare,Federer'81] give a simple construction.
  - Example: $r = 3$, each row is a block, each table is a class.

1	2	3
4	5	6
7	8	9

1	4	7
2	5	8
3	6	9

1	5	9
2	6	7
3	4	8

1	6	8
2	4	9
3	5	7

Proposition

Suppose that r is an odd prime. Then, there is an MPC(r^2, r) with minimum Ulam distance $r - 2$ and size $(r + 1)r!$.

Code Construction: Interleaving I

- Let $\circ$ denote interleaving:
  - $(1,3,2) \circ (6,4,5) \circ (8,7,9) = (1,6,8,3,4,7,2,5,9)$.

Code Construction: Interleaving I

- Let $\circ$ denote interleaving:
  - $(1,3,2) \circ (6,4,5) \circ (8,7,9) = (1,6,8,3,4,7,2,5,9)$.
- Assume
  - A partition $\{P_1, \ldots, P_r\}$ of $[n]$ into sets of equal size.
  - Let C_i, $i \in [r]$, be permutation codes of minimum Ulam distance $d \leq n/r$ over P_i.
  - Construct C by interleaving codewords of C_i.

Code Construction: Interleaving I

- Let $\circ$ denote interleaving:
  - $(1,3,2) \circ (6,4,5) \circ (8,7,9) = (1,6,8,3,4,7,2,5,9)$.
- Assume
  - A partition $\{P_1, \ldots, P_r\}$ of $[n]$ into sets of equal size.
  - Let C_i, $i \in [r]$, be permutation codes of minimum Ulam distance $d \leq n/r$ over P_i.
  - Construct C by interleaving codewords of C_i.

Proposition

The code C is a MPC(n, r) with minimum Ulam distance d.

Code Construction: Interleaving I

- Let $\circ$ denote interleaving:
  - $(1,3,2)\circ(6,4,5)\circ(8,7,9)=(1,6,8,3,4,7,2,5,9)$.
- Assume
  - A partition $\{P_1,\ldots,P_r\}$ of $[n]$ into sets of equal size.
  - Let C_i, $i\in[r]$, be permutation codes of minimum Ulam distance $d\le n/r$ over P_i.
  - Construct C by interleaving codewords of C_i.

Proposition

The code C is a MPC(n,r) with minimum Ulam distance d.

- Assuming optimal component codes, if $\lim \frac{rd}{n}=0$, then C is capacity achieving.

Code Construction: Interleaving II

- Let $\circ_r$ denote interleaving blocks of r elements:
  - $(\underline{1,3},\underline{4,2}) \circ_2 (\underline{6,7},\underline{8,5}) \circ_2 (\underline{12,10},\underline{9,11}) =$
    $(\underline{1,3},\underline{6,7},\underline{12,10},\underline{4,2},\underline{8,5},\underline{9,11})$.

Code Construction: Interleaving II

- Let $\circ_r$ denote interleaving blocks of r elements:
  - $(\underline{1,3},\underline{4,2}) \circ_2 (\underline{6,7},\underline{8,5}) \circ_2 (\underline{12,10},\underline{9,11}) = (\underline{1,3},\underline{6,7},\underline{12,10},\underline{4,2},\underline{8,5},\underline{9,11})$.
- Assume
  - n/r even, $d \leq r$, $P = \left[\frac{n}{2}\right]$, and $Q = [n] \backslash P$.
  - C'_1 is an MPC$(\frac{n}{2}, r)$ with min Ulam distance d over P
  - C_1 is an MPC$(\frac{n}{2}, r)$ with min Hamming distance d over Q.

Code Construction: Interleaving II

- Let $\circ_r$ denote interleaving blocks of r elements:
  - $(\underline{1,3},\underline{4,2}) \circ_2 (\underline{6,7},\underline{8,5}) \circ_2 (\underline{12,10},\underline{9,11}) = (\underline{1,3},\underline{6,7},\underline{12,10},\underline{4,2},\underline{8,5},\underline{9,11})$.
- Assume
  - n/r even, $d \leq r$, $P = \left[\frac{n}{2}\right]$, and $Q = [n] \backslash P$.
  - C_1' is an MPC$(\frac{n}{2}, r)$ with min Ulam distance d over P
  - C_1 is an MPC$(\frac{n}{2}, r)$ with min Hamming distance d over Q.

Proposition

The code $C = C_1' \circ_r C_1$ is an MPC(n, r) with minimum Ulam distance d.

Code Construction: Interleaving II

- Let $\circ_r$ denote interleaving blocks of r elements:
  - $(\underline{1,3},\underline{4,2}) \circ_2 (\underline{6,7},\underline{8,5}) \circ_2 (\underline{12,10},\underline{9,11}) =$ $(\underline{1,3},\underline{6,7},\underline{12,10},\underline{4,2},\underline{8,5},\underline{9,11})$.
- Assume
  - n/r even, $d \leq r$, $P = \left[\frac{n}{2}\right]$, and $Q = [n] \setminus P$.
  - C_1' is an MPC($\frac{n}{2}, r$) with min Ulam distance d over P
  - C_1 is an MPC($\frac{n}{2}, r$) with min Hamming distance d over Q.

Proposition

The code $C = C_1' \circ_r C_1$ is an MPC(n, r) with minimum Ulam distance d.

- With nested construction, we may only use codes in the Hamming metric.

Code Construction: Interleaving II

- Let $\circ_r$ denote interleaving blocks of r elements:
  - $(\underline{1,3},\underline{4,2}) \circ_2 (\underline{6,7},\underline{8,5}) \circ_2 (\underline{12,10},\underline{9,11}) = (\underline{1,3},\underline{6,7},\underline{12,10},\underline{4,2},\underline{8,5},\underline{9,11})$.
- Assume
  - n/r even, $d \leq r$, $P = \left[\frac{n}{2}\right]$, and $Q = [n] \backslash P$.
  - C_1' is an MPC$(\frac{n}{2}, r)$ with min Ulam distance d over P
  - C_1 is an MPC$(\frac{n}{2}, r)$ with min Hamming distance d over Q.

Proposition

The code $C = C_1' \circ_r C_1$ is an MPC(n, r) with minimum Ulam distance d.

- With nested construction, we may only use codes in the Hamming metric.
- Assuming optimal Hamming codes, C is capacity achieving (for $d \leq r$).

Thank you!