Asymptotically MDS Array BP-XOR Codes

ISIT 2018, Vail, CO, USA

Şuayb Ş. Arslan *

* MEF University, Department of Computer Engineering

Maslak, Istanbul, Turkey

Outline

Motivation for Array Codes

- Mainly used for burst error correction in communication systems and storage systems.
- Addresses some of the challenges in Cloud Storage.
  - Easy to move computation than data.
- Required: Simple math.
- Required: Flexibility.
- Desired: Easy code constructions.

Binary Array Codes

- Linear codes where information/parity data are structured in a two dimensional matrix array [1].
- Considered over binary field for complexity.

Formal Definition [2],[3]. (Wang, Paterson)

An $[n, k, t, b]$ array code is a $b \times n$ two dimensional rate $r = k/n$ binary linear code $C = [a_{i,j}]_{1 \leq i \leq b, 1 \leq j \leq n}$ in which the coding symbol $a_{i,j} \in \{0, 1\}^l$ is a weighted sum of a subset of source symbols $\mathcal{I} = \{v_1, \dots, v_{bk}\}$, typically structured as a $b \times k$ data matrix.

- $\mathcal{I}$ can be reconstructed from any $n - t$ columns of the binary array code for $t \leq n - k$.
  - If the decoder is Belief Propagation (BP) algorithm and weighted sum is simply XOR operation, we name it array BP-XOR codes.
  - A t-erasure (column) correcting array BP-XOR code is Maximum Distance Separable (MDS) if $\mathcal{I}$ can be reconstructed from $k = n - t$ columns of C.
- Extensively studied for burst error correction.

Examples

(t=2) EVENODD Code[4], RDP Code[5], X-Code[6], P-Code, H-Code, D-Code, etc. (t=3) STAR [7] and TIP [8] Codes.

Binary Array Codes

- Linear codes where information/parity data are structured in a two dimensional matrix array [1].
- Considered over binary field for complexity.

Formal Definition [2],[3], (Wang, Paterson)

An $[n, k, t, b]$ array code is a $b \times n$ two dimensional rate $r = k/n$ binary linear code $\mathcal{C} = [a_{i,j}]_{1 \leq i \leq b, 1 \leq j \leq n}$ in which the coding symbol $a_{i,j} \in \{0,1\}^l$ is a weighted sum of a subset of source symbols $\mathcal{I} = \{v_1, \ldots, v_{bk}\}$, typically structured as a $b \times k$ data matrix.

- $\mathcal{I}$ can be reconstructed from any $n - t$ columns of the binary array code for $t \leq n - k$.
  - If the decoder is Belief Propagation (BP) algorithm and weighted sum is simply XOR operation, we name it array BP-XOR codes.
  - A t-erasure (column) correcting array BP-XOR code is Maximum Distance Separable (MDS) if $\mathcal{I}$ can be reconstructed from $k = n - t$ columns of $\mathcal{C}$.
- Extensively studied for burst error correction.

Examples

(t=2) EVENODD Code[4], RDP Code[5], X-Code[6], P-Code, H-Code, D-Code, etc. (t=3) STAR [7] and TIP [8] Codes.

Binary Array Codes

- Linear codes where information/parity data are structured in a two dimensional matrix array [1].
- Considered over binary field for complexity.

Formal Definition [2],[3], (Wang, Paterson)

An $[n, k, t, b]$ array code is a $b \times n$ two dimensional rate $r = k/n$ binary linear code $\mathcal{C} = [a_{i,j}]_{1 \leq i \leq b, 1 \leq j \leq n}$ in which the coding symbol $a_{i,j} \in \{0, 1\}^l$ is a weighted sum of a subset of source symbols $\mathcal{I} = \{v_1, \ldots, v_{bk}\}$, typically structured as a $b \times k$ data matrix.

- $\mathcal{I}$ can be reconstructed from any $n - t$ columns of the binary array code for $t \leq n - k$.
  - If the decoder is Belief Propagation (BP) algorithm and weighted sum is simply XOR operation, we name it array BP-XOR codes.
  - A t-erasure (column) correcting array BP-XOR code is Maximum Distance Separable (MDS) if $\mathcal{I}$ can be reconstructed from $k = n - t$ columns of $\mathcal{C}$.
- Extensively studied for burst error correction.

Examples

(t=2) EVENODD Code[4], RDP Code[5], X-Code[6], P-Code, H-Code, D-Code, etc. (t=3) STAR [7] and TIP [8] Codes.

Binary Array Codes

- Linear codes where information/parity data are structured in a two dimensional matrix array [1].
- Considered over binary field for complexity.

Formal Definition [2],[3], (Wang, Paterson)

An $[n, k, t, b]$ array code is a $b \times n$ two dimensional rate $r = k/n$ binary linear code $\mathcal{C} = [a_{i,j}]_{1 \leq i \leq b, 1 \leq j \leq n}$ in which the coding symbol $a_{i,j} \in \{0,1\}^l$ is a weighted sum of a subset of source symbols $\mathcal{I} = \{v_1, \dots, v_{bk}\}$, typically structured as a $b \times k$ data matrix.

- $\mathcal{I}$ can be reconstructed from any $n - t$ columns of the binary array code for $t \leq n - k$.
  - If the decoder is Belief Propagation (BP) algorithm and weighted sum is simply XOR operation, we name it array BP-XOR codes.
  - A t-erasure (column) correcting array BP-XOR code is Maximum Distance Separable (MDS) if $\mathcal{I}$ can be reconstructed from $k = n - t$ columns of $\mathcal{C}$.
- Extensively studied for burst error correction.

Examples

(t=2) EVENODD Code[4], RDP Code[5], X-Code[6], P-Code, H-Code, D-Code, etc. (t=3) STAR [7] and TIP [8] Codes.

Binary Array Codes

- Linear codes where information/parity data are structured in a two dimensional matrix array [1].
- Considered over binary field for complexity.

Formal Definition [2],[3], (Wang, Paterson)

An $[n, k, t, b]$ array code is a $b \times n$ two dimensional rate $r = k/n$ binary linear code $\mathcal{C} = [a_{i,j}]_{1 \leq i \leq b, 1 \leq j \leq n}$ in which the coding symbol $a_{i,j} \in \{0,1\}^l$ is a weighted sum of a subset of source symbols $\mathcal{I} = \{v_1, \ldots, v_{bk}\}$, typically structured as a $b \times k$ data matrix.

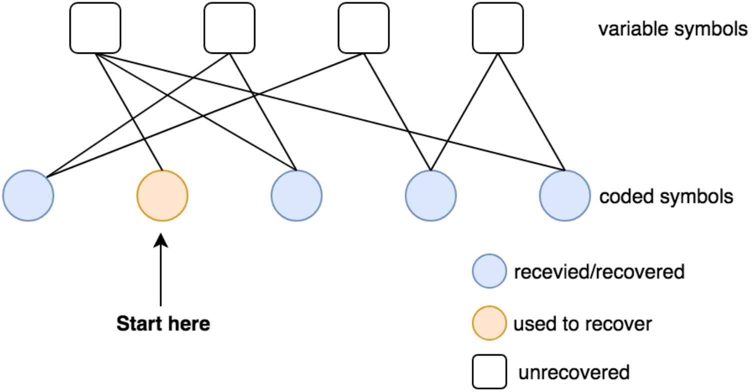
- $\mathcal{I}$ can be reconstructed from any $n - t$ columns of the binary array code for $t \leq n - k$.
  - If the decoder is Belief Propagation (BP) algorithm and weighted sum is simply XOR operation, we name it array BP-XOR codes.
  - A t-erasure (column) correcting array BP-XOR code is Maximum Distance Separable (MDS) if $\mathcal{I}$ can be reconstructed from $k = n - t$ columns of $\mathcal{C}$.
- Extensively studied for burst error correction.

Examples

(t=2) EVENODD Code[4], RDP Code[5], X-Code[6], P-Code, H-Code, D-Code, etc. (t=3) STAR [7] and TIP [8] Codes.

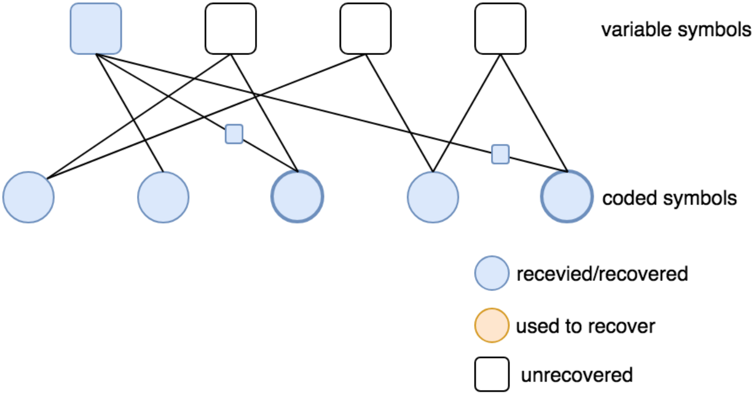
Belief Propagation for Erasures

- Algorithm begins decoding with degree-one coded symbol.



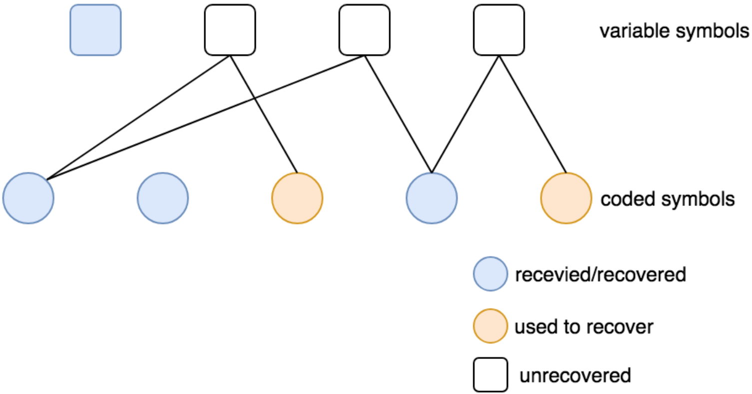
Belief Propagation for Erasures

- Corresponding coded symbols are updated with the decoded value.

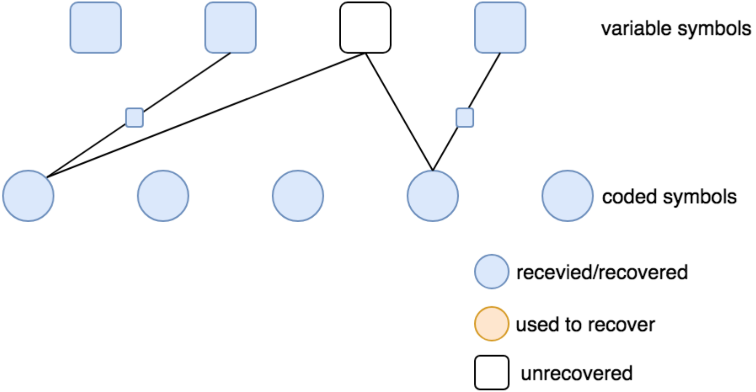


Belief Propagation for Erasures

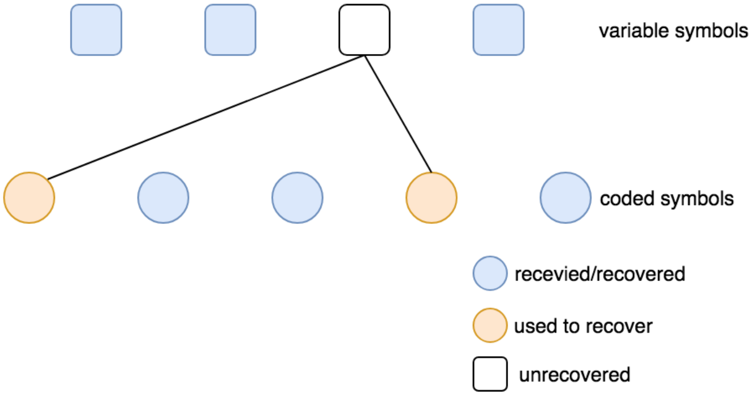
- All edges are removed from the recovered variable symbol.
- We iterate until we recover all variable symbols.



Belief Propagation for Erasures

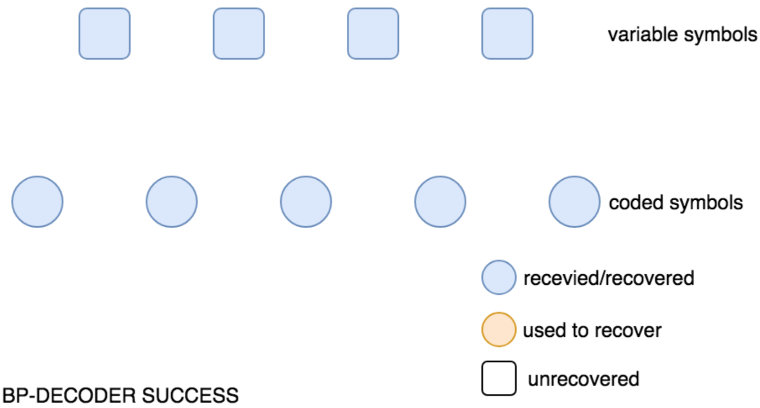


Belief Propagation for Erasures



Belief Propagation for Erasures

- We successfully decoded all the variable symbols.
- Depending on the graph, the decoder could have ended prematurely.



Issues w/ Exact array BP-XOR Codes I

- Example codes are precisely defined for $t = 2, t = 3$. Not possible for all k (usually a function of prime number p).
  - Ex. RDP (k $= p - 1$), X-Code (k $= p - 2$) etc.
- Also constructed for $k = 2$ (edge colored graph model) [3]. It is shown for general (k, t) that existence of a code depends on the maximum symbol degree i.e., $\max_{i,j} deg(a_{i,j}) = \sigma$.

Theorem 1.1 [3] (Wang)

The blocklength n of an $[n, k, t, b]$ array BP-XOR code which has a maximum symbol degree of $\sigma < k + (k-1)/(b-1)$ is bounded above by

$$n \leq k + \sigma - 1 + \left\lfloor \frac{\sigma(\sigma-1)(b-1)}{(k-\sigma)b + \sigma - 1} \right\rfloor \quad (1)$$

Examples

RDP Code (k $= \sigma = p - 1$), X-Code (k $= \sigma = p - 2$)

Corollary 1.2

For $k = \sigma$, we can simplify (1) to $n \leq kb + 1 + \max\{k - 3, 0\}$

Issues w/ Exact array BP-XOR Codes I

- Example codes are precisely defined for $t = 2, t = 3$. Not possible for all k (usually a function of prime number p).
  - Ex. RDP (k $= p - 1$), X-Code (k $= p - 2$) etc.
- Also constructed for $k = 2$ (edge colored graph model) [3]. It is shown for general (k, t) that existence of a code depends on the maximum symbol degree i.e., $\max_{i,j} deg(a_{i,j}) = \sigma$.

Theorem 1.1 [3] (Wang)

The blocklength n of an $[n, k, t, b]$ array BP-XOR code which has a maximum symbol degree of $\sigma < k + (k-1)/(b-1)$ is bounded above by

$$n \leq k + \sigma - 1 + \left\lfloor \frac{\sigma(\sigma-1)(b-1)}{(k-\sigma)b + \sigma - 1} \right\rfloor \quad (1)$$

Examples

RDP Code (k $= \sigma = p - 1$), X-Code (k $= \sigma = p - 2$)

Corollary 1.2

For $k = \sigma$, we can simplify (1) to $n \leq kb + 1 + \max\{k-3, 0\}$

Issues w/ Exact array BP-XOR Codes I

- Example codes are precisely defined for $t = 2, t = 3$. Not possible for all k (usually a function of prime number p).
  - Ex. RDP (k $= p - 1$), X-Code (k $= p - 2$) etc.
- Also constructed for $k = 2$ (edge colored graph model) [3]. It is shown for general (k, t) that existence of a code depends on the maximum symbol degree i.e., $\max_{i,j} deg(a_{i,j}) = \sigma$.

Theorem 1.1 [3] (Wang)

The blocklength n of an $[n, k, t, b]$ array BP-XOR code which has a maximum symbol degree of $\sigma < k + (k-1)/(b-1)$ is bounded above by

$$n \leq k + \sigma - 1 + \left\lfloor \frac{\sigma(\sigma-1)(b-1)}{(k-\sigma)b + \sigma - 1} \right\rfloor \quad (1)$$

Examples

RDP Code (k $= \sigma = p - 1$), X-Code (k $= \sigma = p - 2$)

Corollary 1.2

For $k = \sigma$, we can simplify (1) to $n \leq kb + 1 + \max\{k-3, 0\}$

Issues w/ Exact array BP-XOR Codes I

- Example codes are precisely defined for $t = 2, t = 3$. Not possible for all k (usually a function of prime number p).
  - Ex. RDP (k $= p - 1$), X-Code (k $= p - 2$) etc.
- Also constructed for $k = 2$ (edge colored graph model) [3]. It is shown for general (k, t) that existence of a code depends on the maximum symbol degree i.e., $\max_{i,j} deg(a_{i,j}) = \sigma$.

Theorem 1.1 [3] (Wang)

The blocklength n of an $[n, k, t, b]$ array BP-XOR code which has a maximum symbol degree of $\sigma < k + (k-1)/(b-1)$ is bounded above by

$$n \leq k + \sigma - 1 + \left\lfloor \frac{\sigma(\sigma-1)(b-1)}{(k-\sigma)b + \sigma - 1} \right\rfloor \quad (1)$$

Examples

RDP Code (k $= \sigma = p - 1$), X-Code (k $= \sigma = p - 2$)

Corollary 1.2

For $k = \sigma$, we can simplify (1) to $n \leq kb + 1 + \max\{k-3, 0\}$

Issues w/ Exact array BP-XOR Codes I

- Example codes are precisely defined for $t = 2, t = 3$. Not possible for all k (usually a function of prime number p).
  - Ex. RDP (k $= p - 1$), X-Code (k $= p - 2$) etc.
- Also constructed for $k = 2$ (edge colored graph model) [3]. It is shown for general (k, t) that existence of a code depends on the maximum symbol degree i.e., $\max_{i,j} deg(a_{i,j}) = \sigma$.

Theorem 1.1 [3] (Wang)

The blocklength n of an $[n, k, t, b]$ array BP-XOR code which has a maximum symbol degree of $\sigma < k + (k-1)/(b-1)$ is bounded above by

$$n \leq k + \sigma - 1 + \left\lfloor \frac{\sigma(\sigma-1)(b-1)}{(k-\sigma)b + \sigma - 1} \right\rfloor \qquad (1)$$

Examples

RDP Code (k $= \sigma = p - 1$), X-Code (k $= \sigma = p - 2$)

Corollary 1.2

For $k = \sigma$, we can simplify (1) to $n \leq kb + 1 + \max\{k - 3, 0\}$

Issues w/ Exact array BP-XOR Codes II

- Corollary 1.2 shows that the upper bound in (1) is not tight.
- For high degree MDS array BP-XOR codes with $k \geq \sigma^2$, the block length n becomes independent of b.
- Note that an $[n, k, t, b]$ array code $\mathcal{C}$ over the alphabet $\{0,1\}^l$ can also be considered a linear code over the extension alphabet $\{0,1\}^{l^b}$.
- If we relax BP-decodability constraint, then a standard RS codes over the finite field $GF(2^b)$ can be considered as an array code.
- A big gap for the existence of MDS $b \times n$ array BP-XOR codes over $GF(2)$ and MDS linear codes over $GF(2^b)$.

Examples

UB on n for MDS array BP-XOR Codes with $\sigma = 2$ and large b. [3]

k	2	3	4	$[4, \infty]$
n	$2b+1$	4	5	$k+1$

MAX n for $[n, k]$ RS codes over $GF(2^b)$ [3]

k	2	3	4	5	$[2^b, \infty]$
n	2^b+1	2^b+2	2^b+1	2^b+2	$k+1$

Issues w/ Exact array BP-XOR Codes II

- Corollary 1.2 shows that the upper bound in (1) is not tight.
- For high degree MDS array BP-XOR codes with $k \geq \sigma^2$, the block length n becomes independent of b.
- Note that an $[n, k, t, b]$ array code $\mathcal{C}$ over the alphabet $\{0,1\}^l$ can also be considered a linear code over the extension alphabet $\{0,1\}^{l^b}$.
- If we relax BP-decodability constraint, then a standard RS codes over the finite field $GF(2^b)$ can be considered as an array code.
- A big gap for the existence of MDS $b \times n$ array BP-XOR codes over $GF(2)$ and MDS linear codes over $GF(2^b)$.

Examples

UB on n for MDS array BP-XOR Codes with $\sigma = 2$ and large b. [3]

k	2	3	4	$[4, \infty]$
n	$2b+1$	4	5	$k+1$

MAX n for $[n, k]$ RS codes over $GF(2^b)$ [3]

k	2	3	4	5	$[2^b, \infty]$
n	2^b+1	2^b+2	2^b+1	2^b+2	$k+1$

Issues w/ Exact array BP-XOR Codes II

- Corollary 1.2 shows that the upper bound in (1) is not tight.
- For high degree MDS array BP-XOR codes with $k \geq \sigma^2$, the block length n becomes independent of b.
- Note that an $[n, k, t, b]$ array code $\mathcal{C}$ over the alphabet $\{0, 1\}^l$ can also be considered a linear code over the extension alphabet $\{0, 1\}^{l^b}$.
- If we relax BP-decodability constraint, then a standard RS codes over the finite field $GF(2^b)$ can be considered as an array code.
- A big gap for the existence of MDS $b \times n$ array BP-XOR codes over $GF(2)$ and MDS linear codes over $GF(2^b)$.

Examples

UB on n for MDS array BP-XOR Codes with $\sigma = 2$ and large b. [3]

k	2	3	4	$[4, \infty]$
n	$2b+1$	4	5	$k+1$

MAX n for $[n, k]$ RS codes over $GF(2^b)$ [3]

k	2	3	4	5	$[2^b, \infty]$
n	2^b+1	2^b+2	2^b+1	2^b+2	$k+1$

Issues w/ Exact array BP-XOR Codes II

- Corollary 1.2 shows that the upper bound in (1) is not tight.
- For high degree MDS array BP-XOR codes with $k \geq \sigma^2$, the block length n becomes independent of b.
- Note that an $[n, k, t, b]$ array code $\mathcal{C}$ over the alphabet $\{0,1\}^l$ can also be considered a linear code over the extension alphabet $\{0,1\}^{l^b}$.
- If we relax BP-decodability constraint, then a standard RS codes over the finite field $GF(2^b)$ can be considered as an array code.
- A big gap for the existence of MDS $b \times n$ array BP-XOR codes over $GF(2)$ and MDS linear codes over $GF(2^b)$.

Examples

UB on n for MDS array BP-XOR Codes with $\sigma = 2$ and large b. [3]

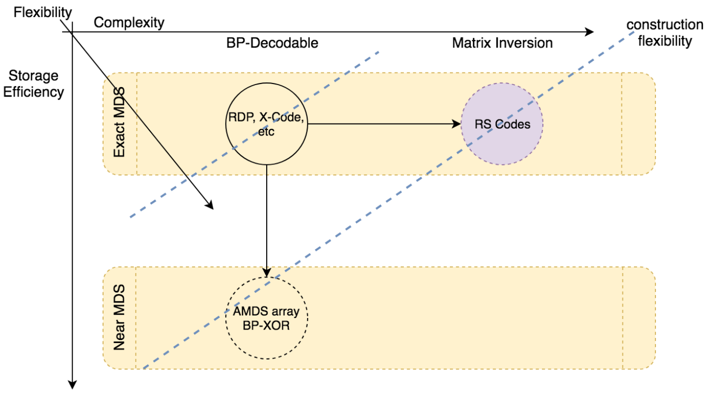
k	2	3	4	[4,∞]
n	$2b+1$	4	5	$k+1$

MAX n for $[n, k]$ RS codes over $GF(2^b)$ [3]

k	2	3	4	5	$[2^b,\infty]$
n	2^b+1	2^b+2	2^b+1	2^b+2	$k+1$

Storage/Complexity Efficiency v.s. Flexibility

BIG PICTURE



An overview of existing codes and where AMDS array BP-XOR codes fit.

Main Results I

- For a given positive number b' satisfying $b' > b$, a $[n, k, t, b, b']$ Asymptotically MDS (AMDS) array BP-XOR code $\mathcal{C}^a$ is a linear code with i-th column $(y_{i,1}, \ldots, y_{i,b_i}) = (x_1, \ldots, x_{bk})G_i$ for a $bk \times b_i$ generator matrix $G_i, i \in \{1, \ldots, n\}$ such that $b' = (1/n)\sum_i b_i$. Therefore, the generator matrix for $\mathcal{C}^a$ is given by the following matrix of size $bk \times \sum_i b_i$,

$$G_{\mathcal{C}^a} = [G_1|G_2|\ldots|G_n]. \tag{2}$$

Theorem 2.1

Let $\mathcal{C}^a$ be a $[n, k, t, b, b']$ AMDS array BP-XOR code such that the maximum coded node degree satisfies $2 < \sigma < (bk-1)/(b'-1)$. Then, we have

$$n \leq k + \sigma - 1 + \left\lfloor \frac{b(k(\sigma' - \sigma) + (\sigma - 1)\sigma') - (\sigma - 1)(3\sigma/2 - 1)}{b(k - \sigma') + \sigma - 1} \right\rfloor \tag{3}$$

where $\sigma' = \sigma(1 + \epsilon(b, n))$, $b' = b(1 + \epsilon(b, n))$ and $\epsilon(b, n)$ is the coding overhead.

Main Results II

- For exact MDS array BP-XOR codes, $\epsilon(b, n) = 0$ i.e., $\sigma' = \sigma$, $b' = b$.
- Tighter bound.

Definition 2.2

The coding overhead of an AMDS array BP-XOR codes satisfies

- Nice thing about this definition is that we can arrange $\epsilon(b, n)$ (parameterize) such that the desired rate can be achieved.
- Heavily depends on the characterization of $\epsilon(b, n)$.

Main Results II

- For exact MDS array BP-XOR codes, $\epsilon(b, n) = 0$ i.e., $\sigma' = \sigma$, $b' = b$.
- Tighter bound.

Definition 2.2

The coding overhead of an AMDS array BP-XOR codes satisfies

(1) For fixed k and rate r, as $b \to \infty$ we have vanishing coding overhead i.e., $\epsilon(b, n) \to 0$.

(2) For fixed b and rate r, as $k, n \to \infty$ we have a diverging coding overhead i.e., $\epsilon(b, n) \to \infty$.

- Nice thing about this definition is that we can arrange $\epsilon(b, n)$ (parameterize) such that the desired rate can be achieved.
- Heavily depends on the characterization of $\epsilon(b, n)$.

Main Results II

- For exact MDS array BP-XOR codes, $\epsilon(b, n) = 0$ i.e., $\sigma' = \sigma$, $b' = b$.
- Tighter bound.

Definition 2.2

The coding overhead of an AMDS array BP-XOR codes satisfies

(1) For fixed k and rate r, as $b \to \infty$ we have vanishing coding overhead i.e., $\epsilon(b, n) \to 0$.

(2) For fixed b and rate r, as $k, n \to \infty$ we have a diverging coding overhead i.e., $\epsilon(b, n) \to \infty$.

- Nice thing about this definition is that we can arrange $\epsilon(b, n)$ (parameterize) such that the desired rate can be achieved.
- Heavily depends on the characterization of $\epsilon(b, n)$.

Main Results II

- For exact MDS array BP-XOR codes, $\epsilon(b, n) = 0$ i.e., $\sigma' = \sigma$, $b' = b$.
- Tighter bound.

Definition 2.2

The coding overhead of an AMDS array BP-XOR codes satisfies

(1) For fixed k and rate r, as $b \to \infty$ we have vanishing coding overhead i.e., $\epsilon(b, n) \to 0$.

(2) For fixed b and rate r, as $k, n \to \infty$ we have a diverging coding overhead i.e., $\epsilon(b, n) \to \infty$.

- Nice thing about this definition is that we can arrange $\epsilon(b, n)$ (parameterize) such that the desired rate can be achieved.
- Heavily depends on the characterization of $\epsilon(b, n)$.

Main Results II

- For exact MDS array BP-XOR codes, $\epsilon(b, n) = 0$ i.e., $\sigma' = \sigma$, $b' = b$.
- Tighter bound.

Definition 2.2

The coding overhead of an AMDS array BP-XOR codes satisfies

(1) For fixed k and rate r, as $b \to \infty$ we have vanishing coding overhead i.e., $\epsilon(b, n) \to 0$.

(2) For fixed b and rate r, as $k, n \to \infty$ we have a diverging coding overhead i.e., $\epsilon(b, n) \to \infty$.

- Nice thing about this definition is that we can arrange $\epsilon(b, n)$ (parameterize) such that the desired rate can be achieved.
- Heavily depends on the characterization of $\epsilon(b, n)$.

Mojette Transform Codes [9]

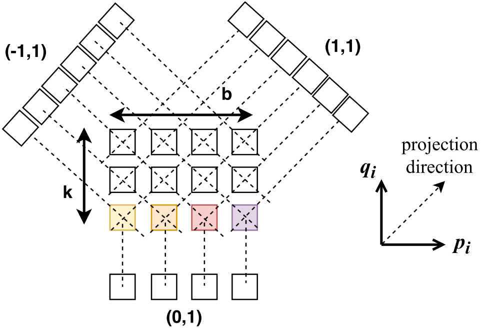


Illustration of projections of Mojette Transform coding (k=3, b=4, n = 3).

- Mojette Transform: Discrete version of Radon transform.
- Compute a linear set of projections from a rectangle grid at angles specified by a couple of coprime integers (p, q) from a $b \times k$ discrete data structure $f : (z, l) \rightarrow \mathbb{N}$ as shown above.
- Let us generate n projections with parameters $\{(p_i, q_i), 0 \leq i \leq n-1\}$.
- Projections can be treated as the columns of AMDS array BP-XOR codes.

Mojette Transform Codes [9]

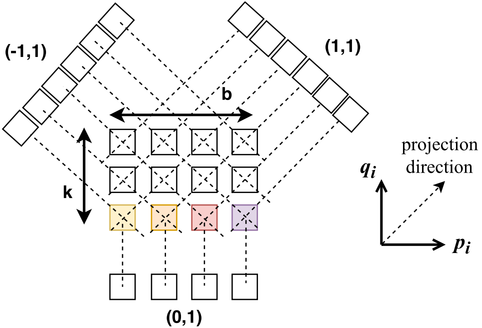


Illustration of projections of Mojette Transform coding (k=3, b=4, n = 3).

- Mojette Transform: Discrete version of Radon transform.
- Compute a linear set of projections from a rectangle grid at angles specified by a couple of coprime integers (p, q) from a $b \times k$ discrete data structure $f : (z, l) \rightarrow \mathbb{N}$ as shown above.
- Let us generate n projections with parameters $\{(p_i, q_i), 0 \leq i \leq n-1\}$.
- Projections can be treated as the columns of AMDS array BP-XOR codes.

Mojette Transform Codes [9]

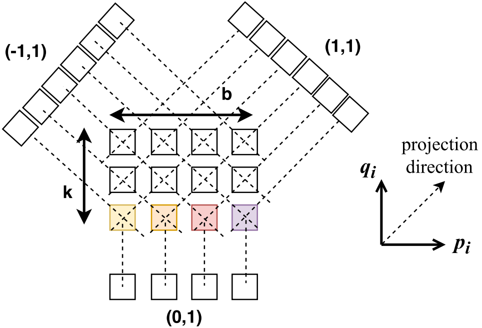


Illustration of projections of Mojette Transform coding (k=3, b=4, n = 3).

- Mojette Transform: Discrete version of Radon transform.
- Compute a linear set of projections from a rectangle grid at angles specified by a couple of coprime integers (p, q) from a $b \times k$ discrete data structure $f : (z, l) \to \mathbb{N}$ as shown above.
- Let us generate n projections with parameters $\{(p_i, q_i), 0 \leq i \leq n-1\}$.
- Projections can be treated as the columns of AMDS array BP-XOR codes.

Encoder/Decoder

- Each symbol (bin) of the i-th projection, based on (p_i, q_i), can be computed as given by the following compact formulation

$$M_{(p_i,q_i)} f(m + (b-1)q_i u(q_i) + (k-1)p_i u(p_i)) \tag{4}$$
$$= \bigoplus_{z=0}^{b-1} \bigoplus_{l=0}^{k-1} f(z,l)\delta_{m+zq_i+lp_i} \tag{5}$$

  where $\bigoplus$ stands for Boolean XOR operation, $u(.)$ is the discrete unit function and δ_i is Kronecker delta function and m satisfies $-(b-1)q_i u(q_i) - (k-1)p_i u(p_i) \leq m \leq b_i - (b-1)q_i u(q_i) - (k-1)p_i u(p_i) - 1$ and the size of the i-th projection $b_i = |p_i|(k-1) + |q_i|(b-1) + 1$.

- Decoder is simple standard iterative BP algorithm.

Theorem 3.1 (Reconstruction theorem - Katz Criterion [10])

For a given AMDS array BP-XOR code defined by n projections with parameters (p_i, q_i) on a $b \times k$ data matrix, exact data reconstruction is possible using iterative BP if $\sum_{i=0}^{n-1} |p_i| \geq b$ or $\sum_{i=0}^{n-1} |q_i| \geq k$.

Encoder/Decoder

- Each symbol (bin) of the i-th projection, based on (p_i, q_i), can be computed as given by the following compact formulation

$$M_{(p_i,q_i)}f(m+(b-1)q_iu(q_i)+(k-1)p_iu(p_i)) \quad (4)$$

$$= \bigoplus_{z=0}^{b-1}\bigoplus_{l=0}^{k-1} f(z,l)\delta_{m+zq_i+lp_i} \quad (5)$$

  where $\bigoplus$ stands for Boolean XOR operation, $u(.)$ is the discrete unit function and δ_i is Kronecker delta function and m satisfies $-(b-1)q_iu(q_i)-(k-1)p_iu(p_i) \leq m \leq b_i-(b-1)q_iu(q_i)-(k-1)p_iu(p_i)-1$ and the size of the i-th projection $b_i = |p_i|(k-1)+|q_i|(b-1)+1$.

- Decoder is simple standard iterative BP algorithm.

Theorem 3.1 (Reconstruction theorem - Katz Criterion [10])

For a given AMDS array BP-XOR code defined by n projections with parameters (p_i, q_i) on a $b \times k$ data matrix, exact data reconstruction is possible using iterative BP if $\sum_{i=0}^{n-1}|p_i| \geq b$ or $\sum_{i=0}^{n-1}|q_i| \geq k$.

Parameter Selections

- The maximum degree of the coded symbols plays a key role in the attainable blocklength of the array BP-XOR codes.

Theorem 3.2

Let $\sigma_i, i \in \{1, 2, \dots, n\}$ denote the maximum degree of the ith projection with parameters (p_i, q_i). We have $\sigma_i = \min\{\lceil b/|p_i|\rceil, \lceil k/|q_i|\rceil\}$ and hence $\sigma = \max_i\{\sigma_i\}$.

- Depending on the choices of (p_i, q_i), the coding overhead and maximum degree σ can change.

Construction 3.3

Let us consider the following choice of coprime integers,

$$q_i = 1, p_i \in \mathfrak{T} = \left\{-\left\lfloor \frac{n-1}{2} \right\rfloor, \dots, -1, 0, 1, 2, \dots, \left\lceil \frac{n-1}{2} \right\rceil \right\} \quad (6)$$

where $\mathfrak{T}$ is known as canonical enumeration of integers [11] that goes with the name *A007306* and satisfies $gcd(p_i, q_i) = 1$ for $i = 0, \dots, n-1$.

- Construction 3.3 satisfies Katz criterion and for $b \gg 1$, we have $\sigma = k$.

Parameter Selections

- The maximum degree of the coded symbols plays a key role in the attainable blocklength of the array BP-XOR codes.

Theorem 3.2

Let $\sigma_i, i \in \{1, 2, \ldots, n\}$ denote the maximum degree of the ith projection with parameters (p_i, q_i). We have $\sigma_i = \min\{\lceil b/|p_i|\rceil, \lceil k/|q_i|\rceil\}$ and hence $\sigma = \max_i\{\sigma_i\}$.

- Depending on the choices of (p_i, q_i), the coding overhead and maximum degree σ can change.

Construction 3.3

Let us consider the following choice of coprime integers,

$$q_i = 1, p_i \in \mathfrak{T} = \left\{-\left\lfloor\frac{n-1}{2}\right\rfloor, \ldots, -1, 0, 1, 2, \ldots, \left\lceil\frac{n-1}{2}\right\rceil\right\} \quad (6)$$

where $\mathfrak{T}$ is known as canonical enumeration of integers [11] that goes with the name *A007306* and satisfies $gcd(p_i, q_i) = 1$ for $i = 0, \ldots, n-1$.

- Construction 3.3 satisfies Katz criterion and for $b \gg 1$, we have $\sigma = k$.

Parameter Selections

- The maximum degree of the coded symbols plays a key role in the attainable blocklength of the array BP-XOR codes.

Theorem 3.2

Let $\sigma_i, i \in \{1, 2, \dots, n\}$ denote the maximum degree of the ith projection with parameters (p_i, q_i). We have $\sigma_i = \min\{\lceil b/|p_i|\rceil, \lceil k/|q_i|\rceil\}$ and hence $\sigma = \max_i\{\sigma_i\}$.

- Depending on the choices of (p_i, q_i), the coding overhead and maximum degree σ can change.

Construction 3.3

Let us consider the following choice of coprime integers,

$$q_i = 1, p_i \in \mathfrak{T} = \left\{-\left\lfloor \frac{n-1}{2} \right\rfloor, \dots, -1, 0, 1, 2, \dots, \left\lceil \frac{n-1}{2} \right\rceil\right\} \tag{6}$$

where $\mathfrak{T}$ is known as canonical enumeration of integers [11] that goes with the name *A007306* and satisfies $gcd(p_i, q_i) = 1$ for $i = 0, \dots, n-1$.

- Construction 3.3 satisfies Katz criterion and for $b \gg 1$, we have $\sigma = k$.

Overhead

Theorem 3.4

For Mojette transform code with parameters as given in Construction 3.3, for $b \gg 1$, we have

$$\epsilon(b, n) \approx \frac{n(2 - r)(nr - 1)}{4b} \tag{7}$$

where $r = k/n$ is the fixed rate of the array BP-XOR code.

- This overhead satisfies definition 2.2. Note $k = \sigma$ has the least constraint on the code length for any MDS array BP-XOR code.

Construction 3.5

Let us consider the following choice of coprime integers for n projections,

$$q_i = q_e > 0,$$
$$p_i \in \mathfrak{U} = \{\lceil -n+1 \rceil_{odd}, \cdots -3, -1, 1, 3, \ldots, \lceil n-1 \rceil_{odd}\} \tag{8}$$

where q_e is a positive even number, and $\lceil . \rceil_{odd}$ rounds to the next biggest odd integer of the argument, respectively.

- We can show that $GCD(p_i, q_i) = 1$ and $k > \sigma = \max_i\{\min\{\lceil b/|p_i| \rceil, \lceil k/|q_i| \rceil\}\} = \lceil k/q_e \rceil$

Overhead

Theorem 3.4

For Mojette transform code with parameters as given in Construction 3.3, for $b \gg 1$, we have

$$\epsilon(b, n) \approx \frac{n(2-r)(nr-1)}{4b} \tag{7}$$

where $r = k/n$ is the fixed rate of the array BP-XOR code.

- This overhead satisfies definition 2.2. Note $k = \sigma$ has the least constraint on the code length for any MDS array BP-XOR code.

Construction 3.5

Let us consider the following choice of coprime integers for n projections,

$$q_i = q_e > 0,$$

$$p_i \in \mathfrak{U} = \left\{\lceil -n+1 \rceil_{odd}, \cdots -3, -1, 1, 3, \dots, \lceil n-1 \rceil_{odd}\right\} \tag{8}$$

where q_e is a positive even number, and $\lceil . \rceil_{odd}$ rounds to the next biggest odd integer of the argument, respectively.

- We can show that $GCD(p_i, q_i) = 1$ and $k > \sigma = \max_i\{\min\{\lceil b/|p_i| \rceil, \lceil k/|q_i| \rceil\}\} = \lceil k/q_e \rceil$

Overhead

Theorem 3.6

For Mojette transform code with parameters as given in construction 3.5, for $b \gg 1$, we have

$$\epsilon(b,n) \approx \frac{\lceil k/q_e \rceil}{kb}\left((k-1)\left(n - \frac{\lceil k/q_e \rceil}{2}\right) + (b-1)q_e + 1\right) - 1$$

where q_e is a positive even number, and $\lceil . \rceil_{odd}$ rounds to the next biggest odd integer of the argument, respectively.

- We can find the explicit upper bound for the blocklength n using the construction 3.5.

$$n \leq k + \frac{\sigma\lceil k/q_e \rceil}{kb}\left((k-1)\left(n - \frac{\lceil k/q_e \rceil}{2}\right) + (b-1)q_e + 1\right) - 1 \qquad (9)$$

- Hard to visualize. Let us provide some numerical results.

Rate 3/4 AMDS array BP-XOR Code

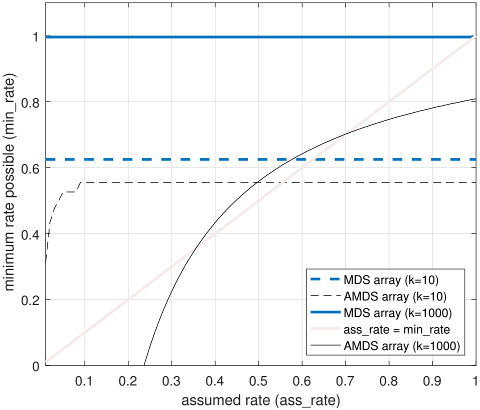
- Choose $q_e = 2$, $b = 10000$, code rates $r \in \{3/4, 1/2\}$. asym denotes AMDS array BP-XOR codes based on Mojette Transform.

Rate 1/2 AMDS array BP-XOR Code

- Upper bounds on n as a function of k for $b = 10000$.

Achievable rates

- The upper bound on n depends on the coding overhead which is a function of code rate.
- For each assumed rate, we calculate the upper bound and then compute the minimum code rate possible. i.e., region that lies above the curves are possible ra˙



Results

- As k gets large it becomes impossible to construct classical MDS array BP-XOR codes with rate smaller than (almost) 1. ($k = 10$ to $k = 1000$)
- By relaxing the exact MDS constraint, we can improve the region of possibilities for better achievability.
- Note that these bounds can be quantified once the overhead expression is available.
- Overhead is a function of the code construction process and parameters.

Conclusion

- Array BP-XOR codes are attractive data protection schemes for low-complexity and optimal reliability.
- Exact constructions have limitations on the maximum block length (so on minimum rate) when the coding symbol degree is particularly lower than the data size.
- This limitation can greatly be relaxed by extending the original optimal class to asymptotically optimal class.
- Demonstrated a code construction based on discrete geometry that satisfies all the requirements of AMDS array BP-XOR code class.

Future Work: Other construction methodologies

- Conjecture: Zigzag codes can be an alternative way of constructing AMDS array BP-XOR codes.

Thanks for your attention.

References

- M. Blaum and R. M. Roth, "New Array Codes for Multiple Phased Burst Correction," *IEEE Trans. on Information Theory,* 339(1):66-77, 1993.
- Y. Wang, "Array BP-XOR codes for reliable cloud storage systems," *In Proc. of IEEE ISIT,* pp. 326–330, 2013.
- M. B. Paterson, D. R. Stinson and Y. Wang, "On Encoding Symbol Degrees of Array BP-XOR Codes," Cryptography and Communications, vol. 8, no. 1, pp. 19–32, 2016.
- M. Blaum, J. Brady, J. Bruck and J. Menon, "EVENODD: An Efficient Scheme for Tolerating Double Disk Failures in RAID Architectures," IEEE Trans. on Computers, 44(2), 192-202, Feb. 1995.
- P. Corbett, B. English, A. Goel, T. Grcanac, S. Kleiman, J. Leong, and S. Sankar, "Row-diagonal parity for double disk failure correction," in Proc. of the 3rd USENIX Conf. on File and Storage Technologies (FAST), 2004, pp. 1–14.
- L. Xu and J. Bruck, "X-Code: MDS Array Codes with Optimal Encoding," IEEE Trans. on Information Theory, 45(1), 272-276, Jan., 1999.
- C. Huang and L. Xu, "STAR: An efficient coding scheme for correcting