Coding problems for memory and storage applications

Alexander Barg

University of Maryland

January 27, 2015
Introduction: Big Data

*Big Data players:* Facebook, Instagram, Google, MSFT, etc.; Dropbox, Box, etc.

*Companies marketing coding solutions:* CleverSafe (RS codes) and others.
**Introduction: Big Data**

*Big Data players:* Facebook, Instagram, Google, MSFT, etc.; Dropbox, Box, etc.  
*Companies marketing coding solutions:* CleverSafe (RS codes) and others.

*Cluster of machines running Hadoop at Yahoo!*

Node failures are the **norm**
Is repair cost a real issue?

(Average number of failed nodes = 20) $\times 15\text{Tb} = 300\text{Tb}$
Two approaches to data coding in distributed storage

- Codes with locality
- Regenerating codes
### Regenerating codes

#### Data collection

- $B$ symbols are encoded into $n\alpha$ symbols stored in $n$ nodes

#### Node repair

- Repair bandwidth $d\beta$

#### Trade-off

- $8 \times 10^4$ symbols are encoded into $n\alpha$ symbols stored in $n$ nodes

---

A. Barg (UMD)  
Coding for memory and storage  
January 27, 2015
Regenerating codes

1. Codes with locality

2. Regenerating codes

- **Data collection**
  - $B$ symbols are encoded into $n\alpha$ symbols stored in $n$ nodes
  - Downloading the data is possible by accessing any $k$ nodes

- **Node repair**

- **Trade-off**

---

A. Barg (UMD)  Coding for memory and storage  January 27, 2015  5 / 73
Regenerating codes

- $B$ symbols are encoded into $n\alpha$ symbols stored in $n$ nodes
- Downloading the data is possible by accessing any $k$ nodes
- Node repair (exact or functional) can be performed by downloading $\beta < \alpha$ symbols from any subset of $d$ nodes.
Regenerating codes

- $B$ symbols are encoded into $n\alpha$ symbols stored in $n$ nodes
- Downloading the data is possible by accessing any $k$ nodes
- Node repair (exact or functional) can be performed by downloading $\beta < \alpha$ symbols from any subset of $d$ nodes.
- Repair bandwidth $d\beta$

$(n, k, d, \{\alpha, \beta\})$ regenerating codes
Regenerating codes

- $B$ symbols are encoded into $n\alpha$ symbols stored in $n$ nodes
- Downloading the data is possible by accessing any $k$ nodes
- Node repair (exact or functional) can be performed by downloading $\beta < \alpha$ symbols from any subset of $d$ nodes.
- Repair bandwidth $d\beta$

$(n, k, d, \{\alpha, \beta\})$ regenerating codes

In this part we focus on locally recoverable codes.
Locally recoverable codes: Plan

In this part we focus on locally recoverable codes

LRC code: To recover one lost symbol of the encoding it suffices to access a small number $r$ of other symbols.
Locally recoverable codes: Plan

In this part we focus on locally recoverable codes

**LRC code:** To recover one lost symbol of the encoding it suffices to access a small number $r$ of other symbols.

1. Current solutions
2. Parameters of LRC codes
3. MDS-like codes with the locality property
4. The availability problem: Multiple recovering sets
5. Extensions
   - LRC codes on algebraic curves
   - Cyclic LRC codes
6. Open problems: Bounds on codes; cyclic codes; list decoding
RAID: Redundant Array of Independent Disks

RAID 1 – Replication (currently 3x)
- Provides high availability of information
- Can tolerate any 2 disk failures
- Widely used in *Hadoop* and many other systems
- Storage overhead of 200%

RAID 6 uses [6,4,3] RS codes

[\(n, k\)] RS codes
- Can tolerate any n-k disk failures
- Poor handling of single disk failures (The Repair Problem)
Limitations of Reed-Solomon codes

Example: [14, 10] RS code

Transmit 10 symbols to recover one lost value

Generates 10x more traffic for recovery of one drive
If large portion of the cluster is RS-coded, this leads to saturation of the network
Other constructions

A combination of local and global parity checks for single and multiple nodes failures

(C. Huang at al., Erasure coding in Windows Azure Storage, USENIX Conf. 2012)
Other constructions

A combination of local and global parity checks for single and multiple nodes failures

(C. Huang at al., Erasure coding in Windows Azure Storage, USENIX Conf. 2012)

Other similar constructions (Windows Azure code)

Pyramid codes (C. Huang et al., 2007)
Locally recoverable codes

The code \( C \subset \mathbb{F}^n \) is locally recoverable with locality \( r \) if every symbol can be recovered by accessing some other \( r \) symbols in the encoding (recovering set of coordinate \( i \))
Let $a \in \mathbb{F}$; consider the restriction $C_J$ of $C$ to a subset $J \subset [n]$. Let
\[ C_J(a, i) = \{x \in C_J : x_i = a\}, \quad i \in [n]. \]

**Definition**

Code $C$ has *locality* $r$ if for every $i \in [n]$ there exists a subset $J_i \subset [n] \setminus i, |J_i| \leq r$ such that
\[ C_{J_i}(a, i) \cap C_{J_i}(a', i) = \emptyset, \quad a \neq a'. \]
\((n, k, r)\) LRC code

Let \(a \in \mathbb{F}\); consider the \textit{restriction} \(C_J\) of \(C\) to a subset \(J \subset [n]\). Let

\[ C_J(a, i) = \{ x \in C_J : x_i = a \}, \quad i \in [n]. \]

**Definition**

Code \(C\) has \textit{locality} \(r\) if for every \(i \in [n]\) there exists a subset \(J_i \subset [n] \setminus i, |J_i| \leq r\) such that

\[ C_{J_i}(a, i) \cap C_{J_i}(a', i) = \emptyset, \quad a \neq a' \]

J. Han and L. Lastras-Montano, ISIT 2007;  
C. Huang, M. Chen, and J. Li, Symp. Networks App. 2007;  
F. Oggier and A. Datta ’10;  
Parameters of LRC codes

Theorem. Let $C$ be an $(n; k; r)$ LRC code of cardinality $q^k$ over an alphabet of size $q$, then:

1. The rate of $C$ satisfies
   \[ \frac{k}{n} \leq \frac{r}{r+1} \]  

2. The minimum distance of $C$ satisfies
   \[ d \geq n - k \lceil \frac{k}{r} \rceil + 2 \]  

Bound (2) is due to Gopalan et al. (2011) and Papailiopoulos et al. (2012).

Note that $r = k$ reduces (2) to the Singleton bound $d \geq n - k + 1$. 

A. Barg (UMD)  
Coding for memory and storage  
January 27, 2015  
12 / 73
Parameters of LRC codes

Theorem

Let $C$ be an $(n, k, r)$ LRC code of cardinality $q^k$ over an alphabet of size $q$, then:

The rate of $C$ satisfies

$$\frac{k}{n} \leq \frac{r}{r + 1}.$$  \hspace{1cm} (1)

The minimum distance of $C$ satisfies

$$d \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2.$$  \hspace{1cm} (2)

Bound (2) is due to Gopalan e.a. (2011) and Papailiopoulos e.a. (2012).
Parameters of LRC codes

Theorem

Let $C$ be an $(n, k, r)$ LRC code of cardinality $q^k$ over an alphabet of size $q$, then:

The rate of $C$ satisfies

$$\frac{k}{n} \leq \frac{r}{r + 1}. \quad (1)$$

The minimum distance of $C$ satisfies

$$d \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2. \quad (2)$$

Bound (2) is due to Gopalan e.a. (2011) and Papailiopoulos e.a. (2012).

Note that $r = k$ reduces (2) to the Singleton bound

$$d \leq n - k + 1.$$
Main idea. Let \( C \) be a \( q \)-ary code of length \( n \), size \( q^k \). The distance \( d(C) \) equals

\[
d(C) = n - \max_{S \subseteq [n]} \{|S| : |C_S| < q^k \}
\]
Main idea. Let \( C \) be a \( q \)-ary code of length \( n \), size \( q^k \). The distance \( d(C) \) equals

\[
d(C) = n - \max_{S \subseteq [n]} \{|S| : |C_S| < q^k\}
\]

The Singleton bound (without locality): \( |S| = k - 1 \)
The distance bound

Main idea. Let $C$ be a $q$-ary code of length $n$, size $q^k$. The distance $d(C)$ equals

$$d(C) = n - \max_{S \subseteq [n]} \{|S| : |C_S| < q^k\}$$

The Singleton bound (with locality):

Let $l_i \subseteq [n], |l_i| \leq r$ be the recovering set for the symbol $c_i, i = 1, \ldots, n$. 
Main idea. Let $C$ be a $q$-ary code of length $n$, size $q^k$. The distance $d(C)$ equals

$$d(C) = n - \max_{S \subseteq [n]} \{|S| : |C_S| < q^k\}$$

The Singleton bound (with locality):

Let $l_i \subseteq [n], |l_i| \leq r$ be the recovering set for the symbol $c_i, i = 1, ..., n$. Let $J_m = \bigcup_{i=1}^m l_i$, where $m = \lfloor (k - 1)/r \rfloor$. Clearly $|J_m| \leq k - 1$. 
The distance bound

Main idea. Let $C$ be a $q$-ary code of length $n$, size $q^k$. The distance $d(C)$ equals

$$d(C) = n - \max_{S \subseteq [n]} \{|S| : |C_S| < q^k\}$$

The Singleton bound (with locality):

Let $I_i \subseteq [n], |I_i| \leq r$ be the recovering set for the symbol $c_i, i = 1, ..., n$.
Let $J_m = \bigcup_{i=1}^{m} I_i$, where $m = \lfloor (k - 1)/r \rfloor$. Clearly $|J_m| \leq k - 1$.
Consider the subset $J'_m = J_m \cup \{1, \ldots, m\}$. We have $C_{J'_m} \leq q^{k-1}$. 
The distance bound

Main idea. Let $C$ be a $q$-ary code of length $n$, size $q^k$. The distance $d(C)$ equals

$$d(C) = n - \max_{S \subseteq [n]} \{|S| : |C_S| < q^k\}$$

The Singleton bound (with locality):

Let $l_i \subset [n], |l_i| \leq r$ be the recovering set for the symbol $c_i, i = 1, \ldots, n$.

Let $J_m = \bigcup_{i=1}^m l_i$, where $m = \lfloor (k - 1)/r \rfloor$. Clearly $|J_m| \leq k - 1$.

Consider the subset $J'_m = J_m \cup \{1, \ldots, m\}$. We have $C_{J'_m} \leq q^{k-1}$.

If $|J'_m| < k - 1$, add to $J'_m$ any $k - 1 - |J_m|$ other coordinates to form the set $L_m \subset [n]$. 
The distance bound

Main idea. Let $C$ be a $q$-ary code of length $n$, size $q^k$. The distance $d(C)$ equals

$$d(C) = n - \max_{S \subseteq [n]} \{|S| : |C_S| < q^k\}$$

The Singleton bound (with locality):

Let $l_i \subset [n], |l_i| \leq r$ be the recovering set for the symbol $c_i, i = 1, \ldots, n$. Let $J_m = \bigcup_{i=1}^m l_i$, where $m = \lfloor (k - 1)/r \rfloor$. Clearly $|J_m| \leq k - 1$.

Consider the subset $J'_m = J_m \cup \{1, \ldots, m\}$. We have $|C_{J'_m}| \leq q^{k-1}$.

If $|J'_m| < k - 1$, add to $J'_m$ any $k - 1 - |J_m|$ other coordinates to form the set $L_m \subset [n]$. We have

$$|C_{L_m}| < q^k$$

$$|L_m| = k - 1 + m = k - 1 + \left\lfloor \frac{k - 1}{r} \right\rfloor = k - 2 + \left\lceil \frac{k}{r} \right\rceil$$
Cadambe-Mazumdar bound

\[(n, k, r) \text{ LRC code } C\]
Cadambe-Mazumdar bound

\((n, k, r)\) LRC code \(\mathcal{C}\)

\[
k \leq \min_{s \geq 1} \left( rs + k_{\text{opt}}^{(q)}(n - s(r + 1), d) \right)
\]
(\(n, k, r\)) LRC code \(\mathcal{C}\)

\[
k \leq \min_{s \geq 1}(rs + k^{(q)}_{\text{opt}}(n - s(r + 1), d))
\]

Consider the sets of coordinates \(L_s\) constructed above, \(1 \leq s \leq \lfloor (k - 1)/r \rfloor\).

\[
|\mathcal{C}_{L_s}| \leq q^{rs}
\]

The shortening of the code \(\mathcal{C}\) on the coordinates in \(L_s\) forms a code of length \(n - s(r + 1)\) with distance \(d\)
Existence (Gilbert-Varshamov) bound

A linear $q$-ary $[n, k', d]$ code exists if

$$\sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i < q^{n-k'}$$

Add $\lceil n/(r+1) \rceil$ local parities

$$k \geq k' - \left\lceil \frac{n}{r+1} \right\rceil$$

Sequences of $(R, \delta)$ codes with locality $r$ exist as long as

$$R < \frac{r}{r+1} - \delta \log_q \frac{q-1}{\delta} - (1 - \delta) \log_q \frac{1}{1 - \delta}$$
Existence (Gilbert-Varshamov) bound

A linear $q$-ary $[n, k', d]$ code exists if

$$\sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i < q^{n-k'}$$

Add $\lceil n/(r+1) \rceil$ local parities

$$k \geq k' - \left\lceil \frac{n}{r+1} \right\rceil$$

Sequences of $(R, \delta)$ codes with locality $r$ exist as long as

$$R < \frac{r}{r+1} - \delta \log_q \frac{q-1}{\delta} - (1 - \delta) \log_q \frac{1}{1 - \delta}$$

$$R \leq \frac{r}{r+1} - h_q(\delta)$$
Early constructions

1. Optimal \(((r + 1)\lceil k/r \rceil, k, r)\) LRC code

   \textbf{Prasanth, Kamath, Lalitha, and Kumar}, ISIT 2012

   Restricted length

2. Optimal \((n, k, r)\) LRC codes

   \textbf{Silberstein, Rawat, Koluoglu, and Vishwanath}, ISIT 2013

   \textbf{Tamo, Papailiopoulos, and Dimakis}, ISIT 2013

   Almost any \(n, k, r\)

   Field size \(q \sim 2^n\)
Reed-Solomon codes

An RS code $C$ is a linear code of length $n \leq q - 1$ over the field $\mathbb{F}_q$. 
Reed-Solomon codes

An RS code $C$ is a linear code of length $n \leq q - 1$ over the field $F_q$.

Given a polynomial $f \in F_q[x]$ and a set $A = \{P_1, \ldots, P_n\} \subset F_q$, define the map

$$ev_A : f \mapsto (f(P_i), i = 1, \ldots, n)$$
Reed-Solomon codes

An **RS code** $C$ is a linear code of length $n \leq q - 1$ over the field $\mathbb{F}_q$.

Given a polynomial $f \in \mathbb{F}_q[x]$ and a set $A = \{P_1, \ldots, P_n\} \subset \mathbb{F}_q$ define the map

$$ev_A : f \mapsto (f(P_i), i = 1, \ldots, n)$$

**RS code $C$ encodes messages of $k$ symbols.**
Let $V_k(q) = \{f \in \mathbb{F}_q[x] : \deg(f) \leq k - 1\}$

$$C : V_k(q) \rightarrow \mathbb{F}_q^n$$

$$f \mapsto ev_A(f) = (f(P_i), i = 1, \ldots, n)$$
Reed-Solomon codes

An RS code $C$ is a linear code of length $n \leq q - 1$ over the field $\mathbb{F}_q$

Given a polynomial $f \in \mathbb{F}_q[x]$ and a set $A = \{P_1, \ldots, P_n\} \subset \mathbb{F}_q$ define the map

$$ev_A : f \mapsto (f(P_i), i = 1, \ldots, n)$$

RS code $C$ encodes messages of $k$ symbols.
Let $V_k(q) = \{f \in \mathbb{F}_q[x] : \deg(f) \leq k - 1\}$

$$C : V_k(q) \rightarrow \mathbb{F}_q^n$$

$$f \mapsto ev_A(f) = (f(P_i), i = 1, \ldots, n)$$

Example: Let $q = 8$, $f(x) = 1 + \alpha x + \alpha x^2$

$$f(x) \mapsto (1, \alpha^4, \alpha^6, \alpha^4, \alpha, \alpha, \alpha, \alpha^6)$$
Reed-Solomon codes
Reed-Solomon codes
To recover one erased value we need to read $k$ other values.
To recover one erased value we need to read $k$ other values.
LRC codes: Idea of construction

What if we can interpolate low-degree polynomials?
What if we can interpolate low-degree polynomials?
Construction of LRC codes: Limitations

We need a specially chosen set of points $A$

Restricted set of polynomials
Construction of \((n, k, r)\) LRC codes: Example

**Parameters:** \(n = 9, k = 4, r = 2, q = 13\);

**Set of points:** \(A = \{1, 2, 3, 4, 5, 6, 9, 10, 12\}\)

\[ \mathcal{A} = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\} \]

**Message:** \(a = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}) \in \mathbb{F}_q^k\)

**Polynomial space:**

\[ V_k(q) := \{a_{0,0} + a_{1,0}x + a_{0,1}x^3 + a_{1,1}x^4\} \]
Construction of \((n, k, r)\) LRC codes: Example

**Parameters:** \(n = 9, k = 4, r = 2, q = 13\);

**Set of points:** \(A=\{1,2,3,4,5,6,9,10,12\}\)
\[
A = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\}
\]

**Message:** \(a = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}) \in \mathbb{F}_q^k\)

**Polynomial space:**
\[
V_k(q) := \{a_{0,0} + a_{1,0}x + a_{0,1}x^3 + a_{1,1}x^4\}
\]

E.g., \(a = (1, 1, 1, 1)\), \(f_a(x) = 1 + x + x^3 + x^4\); \(ev_A(f) = (4, 8, 7, 1, 11, 2, 0, 0, 0)\)
Construction of \((n, k, r)\) LRC codes: Example

**Parameters:** \(n = 9, k = 4, r = 2, q = 13\);

**Set of points:** \(A = \{1, 2, 3, 4, 5, 6, 9, 10, 12\}\)

\[ A = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\} \]

**Message:** \(a = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}) \in \mathbb{F}_q^k\)

**Polynomial space:**

\[ V_k(q) := \{a_{0,0} + a_{1,0}x + a_{0,1}x^3 + a_{1,1}x^4\} \]

E.g., \(a = (1, 1, 1, 1), f_a(x) = 1 + x + x^3 + x^4; ev_A(f) = (4, 8, 7, 1, 11, 2, 0, 0, 0)\)

Say \(c_1 = f_a(1)\) is erased. We access the recovering set \(A_1\) to construct a line \(\delta(x) = 2x + 2\) such that \(\delta(3) = 8, \delta(9) = 7\).
Construction of \((n, k, r)\) LRC codes: Example

Parameters: \(n = 9, k = 4, r = 2, q = 13\);

Set of points: \(A = \{1, 2, 3, 4, 5, 6, 9, 10, 12\}\)
\[\mathcal{A} = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\}\]

Message: \(a = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}) \in \mathbb{F}_q^k\)

Polynomial space:
\[V_k(q) := \{a_{0,0} + a_{1,0}x + a_{0,1}x^3 + a_{1,1}x^4\}\]

E.g., \(a = (1, 1, 1, 1), \ f_a(x) = 1 + x + x^3 + x^4; \ ev_A(f) = (4, 8, 7, 1, 11, 2, 0, 0, 0)\)

Say \(c_1 = f_a(1)\) is erased. We access the recovering set \(A_1\) to construct a line \(\delta(x) = 2x + 2\) such that \(\delta(3) = 8, \delta(9) = 7\).

Compute \(c_1\) as \(\delta(1) = 4\)
Construction of \((n, k, r)\) LRC codes: Example

Parameters: \(n = 9, k = 4, r = 2, q = 13;\)

Set of points: \(A = \{1, 2, 3, 4, 5, 6, 9, 10, 12\}\)
\[ A = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\} \]

Message: \(a = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}) \in \mathbb{F}_q^k\)

Polynomial space:
\[ V_k(q) := \{a_{0,0} + a_{1,0}x + a_{0,1}x^3 + a_{1,1}x^4\} \]
E.g., \(a = (1, 1, 1, 1), f_a(x) = 1 + x + x^3 + x^4; ev_A(f) = (4, 8, 7, 1, 11, 2, 0, 0, 0)\)

Say \(c_1 = f_a(1)\) is erased. We access the recovering set \(A_1\) to construct a line
\(\delta(x) = 2x + 2\) such that \(\delta(3) = 8, \delta(9) = 7.\)

Compute \(c_1\) as \(\delta(1) = 4\)

It works!
Construction of \((n, k, r)\) LRC codes

Assume that \(q \geq n, (r + 1)|n, r|k\)

Let \(A \subseteq \mathbb{F}_q, |A| = n\)
Construction of \((n, k, r)\) LRC codes

Assume that \(q \geq n, (r + 1)|n, r|k\)

Let \(A \subseteq \mathbb{F}_q, |A| = n\)

Suppose there exists a polynomial \(g(x) \in \mathbb{F}[x]\) such that

1. \(\text{deg } g = r + 1\),
2. There exists a partition \(A = \{A_1, ..., A_{\frac{n}{r+1}}\}\) of \(A\) into sets of size \(r + 1\), such that \(g\) is constant on each set \(A_i\) in the partition. For all \(i = 1, \ldots, n/(r + 1)\), and any \(\alpha, \beta \in A_i\),

\[ g(\alpha) = g(\beta). \]
Construction of \((n, k, r)\) LRC codes

Assume that \(q \geq n, (r + 1)|n, r|k\)
Let \(A \subseteq \mathbb{F}_q, |A| = n\)

Suppose there exists a polynomial \(g(x) \in \mathbb{F}[x]\) such that

1. \(\deg g = r + 1\),
2. There exists a partition \(A = \{A_1, \ldots, A_{n/(r+1)}\}\) of \(A\) into sets of size \(r + 1\), such that \(g\) is constant on each set \(A_i\) in the partition. For all \(i = 1, \ldots, n/(r + 1)\), and any \(\alpha, \beta \in A_i\),

\[g(\alpha) = g(\beta).\]

E.g., \(n = 9, r = 2, q = 13;\)
\(A = \{A_1 =\{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\},\)

Then \(g(x) = x^3\) is constant on each of the \(A_i\)’s.
Construction of \((n, k, r)\) LRC codes

Given \(A \subset \mathbb{F}\), partition \(A\) into \((r + 1)\)-subsets.

To encode the message \(a \in \mathbb{F}^k\), write \(a = (a_{ij}, i = 0, \ldots, r - 1; j = 0, \ldots, \frac{k}{r} - 1)\)

Define the encoding polynomial

\[
f_a(x) = \sum_{i=0}^{r-1} f_i(x) x^i,
\]

where

\[
f_i(x) = \sum_{j=0}^{\frac{k}{r} - 1} a_{ij} g(x)^j, \quad i = 0, \ldots, r - 1
\]

A linear code \(C\) is constructed as follows:

\[
Ev : \mathbb{F}^k \to \mathbb{F}^n
\]

\[
a \mapsto (f_a(\beta), \beta \in A)
\]
Recovery of erased symbol

Suppose that the location of erased symbol is $\alpha \in A_j; A_j \in \mathcal{A}$

To find $c_\alpha$ we rely on the recovering set $A_j$

Find a polynomial $\delta(x)$ s.t. $\delta(\beta) = c_\beta, \beta \in A_j \setminus \alpha; \deg \delta \leq r - 1$

$$\delta(x) = \sum_{\beta \in A_j \setminus \alpha} c_\beta \prod_{\beta' \in A_j \setminus \{\alpha, \beta\}} \frac{x - \beta'}{\beta - \beta'}$$

Then $c_\alpha = \delta(\alpha)$
Properties of the construction

Theorem

*The constructed linear codes are optimal* \((n, k, r)\) LRC codes with respect to the “Singleton bound” (2).

**Optimality** is proved by counting degrees.

**Locality:** Let \(\alpha \in A_j\) be the erased location. Define

\[
\partial(x) = \sum_{i=0}^{r-1} f_i(\alpha)x^i
\]

By the construction, for all \(\beta \in A_j\)

\[
\partial(\beta) = f_a(\beta)
\]

Since \(\deg \partial \leq r - 1\), we see that \(\partial(x) \equiv \delta(x)\).
Constructing the polynomial $g(x)$

Proposition

Let $H$ be a subgroup of $\mathbb{F}_q^*$ or $\mathbb{F}^+_q$. The annihilator polynomial of $H$

$$g(x) = \prod_{h \in H} (x - h)$$

is constant on each coset of $H$. 

A. Barg (UMD)
Constructions
RS-type codes

Constructing the polynomial \( g(x) \)

**Proposition**

Let \( H \) be a subgroup of \( \mathbb{F}_q^* \) or \( \mathbb{F}_q^+ \). The **annihilator polynomial** of \( H \)

\[
g(x) = \prod_{h \in H} (x - h)
\]

is constant on each coset of \( H \).

Assume that \( H \) is a multiplicative subgroup and let \( a, ah \) be two elements of the coset \( aH \), where \( h \in H \), then

\[
g(ah) = \prod_{h \in H} (ah - h) = \overline{h}^{|H|} \prod_{h \in H} (a - hh^{-1})
\]

\[
= \prod_{h \in H} (a - h)
\]

\[
= g(a).
\]
Some generalizations

The locator set $A \subset \mathbb{F}$, $A = \bigcup_{i=1}^m A_i$. Consider the algebra

$$\mathbb{F}_A[x] = \{ f \in \mathbb{F}[x] : f \text{ is constant on } A_i, i = 1, \ldots, m; \text{ deg } f < |A| \}.$$
Some generalizations

The locator set $A \subset \mathbb{F}$, $A = \bigsqcup_{i=1}^{m} A_i$. Consider the algebra

$$\mathbb{F}_A[x] = \{f \in \mathbb{F}[x] : f \text{ is constant on } A_i, i = 1, \ldots, m; \deg f < |A|\}.$$

The properties of $\mathbb{F}_A[x]$ are summarized as follows:

1. $\dim(\mathbb{F}_A[x]) = m$;
2. Let $\alpha_1, \ldots, \alpha_m$ be distinct nonzero elements of $\mathbb{F}$, and let $g$ be the polynomial of degree $\deg(g) < |A|$ that satisfies $g(A_i) = \alpha_i$ for all $i = 1, \ldots, m$. Then the polynomials $1, g, \ldots, g^{m-1}$ form a basis of $\mathbb{F}_A[x]$. 
Some generalizations

The locator set $A \subset \mathbb{F}$, $A = \bigcup_{i=1}^{m} A_i$. Consider the algebra

$$\mathbb{F}_A[x] = \{f \in \mathbb{F}[x] : f \text{ is constant on } A_i, i = 1, \ldots, m; \deg f < |A|\}.$$ 

The properties of $\mathbb{F}_A[x]$ are summarized as follows:

1. $\dim(\mathbb{F}_A[x]) = m$;
2. Let $\alpha_1, \ldots, \alpha_m$ be distinct nonzero elements of $\mathbb{F}$, and let $g$ be the polynomial of degree $\deg(g) < |A|$ that satisfies $g(A_i) = \alpha_i$ for all $i = 1, \ldots, m$. Then the polynomials $1, g, \ldots, g^{m-1}$ form a basis of $\mathbb{F}_A[x]$.

General code construction: Let $A \subset \mathbb{F}$, $|A| = n$; $A = \bigcup_{i=1}^{m} A_i$, $|A_i| = r + 1$ for all $i$. Let $\Phi$ be an injective mapping from $\mathbb{F}^k$ to the space of polynomials

$$\mathcal{F}'_A = \bigoplus_{i=0}^{r-1} \mathbb{F}_A[x]x^i.$$ 

The evaluation code obtained in this way is an $(n, k, r)$ LRC code.
It is possible to lift the divisibility constraints $r \mid k, (r + 1) \mid n$. 

To improve data availability, replace $[r+1; r; 2]$ local codes with $[r+1; r]$ MDS codes. Then every $c_i$ is a function of any $r$ out of $r+1$ coordinates.
Extensions

1. It is possible to lift the divisibility constraints $r|k, (r + 1)|n$

2. It is possible to define a *systematic* algebraic encoding mapping.
Extensions

1. It is possible to lift the divisibility constraints $r | k$, $(r + 1) | n$

2. It is possible to define a **systematic** algebraic encoding mapping.

3. To improve **data availability**, replace $[r + 1, r, 2]$ local codes with $[r + \rho - 1, r]$ MDS codes. Then every $c_i$ is a function of any $r$ out of $r + \rho - 1$ coordinates.

Bound on the distance:

$$d \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\rho - 1)$$

(Kamath e.a., 2013)
It is possible to lift the divisibility constraints $r|k, (r + 1)|n$.

It is possible to define a **systematic** algebraic encoding mapping.

To improve data availability, replace $[r + 1, r, 2]$ local codes with $[r + \rho - 1, r]$ MDS codes. Then every $c_i$ is a function of any $r$ out of $r + \rho - 1$ coordinates. Bound on the distance:

$$d \leq n - k + 1 - \left( \left\lfloor \frac{k}{r} \right\rfloor - 1 \right)(\rho - 1) \quad \text{(Kamath e.a., 2013)}$$

**Claim:** Taking recovering sets of size $|A_i| = r + \rho - 1$ and a polynomial basis of $\mathbb{F}_{A}[x]$, we can construct an $(n, k, r)$ LRC code whose distance meets this bound.
Availability problem

“Hot data” accessed simultaneously by a very large number of users
Availability problem

“Hot data” accessed simultaneously by a very large number of users
Multiple recovering sets: Definition

Every symbol in data encoding appears in several disjoint (orthogonal) parity checks

\[ C \subset \mathbb{F}^n \text{ a code of length } n \]

Every coordinate is recoverable from the codeword symbols in several recovering sets:
Multiple recovering sets: Definition

Let \( C(a, i) = \{x \in C : x_i = a\}, a \in \mathbb{F}, i \in [n] \)

The code \( C \) has two disjoint recovering sets if for every \( i \in [n] \) there are subsets \( R^1_i, R^2_i \subset [n] \setminus \{i\}, R^1_i \cap R^2_i = \emptyset \) such that

\[
C(a, i)_{R^j_i} \cap C(a', i)_{R^j_i} = \emptyset, \quad a \neq a'; \quad j = 1, 2
\]
Multiple recovering sets: Idea of construction
Multiple recovering sets: Idea of construction

A. Barg (UMD)
Multiple recovering sets: Idea of construction
Multiple recovering sets: Idea of construction
Multiple recovering sets: Idea of construction

$f_a(\gamma)$ can be found by interpolating $\delta_1(x)$ as well as $\delta_2(x)$.
Multiple recovering sets: Example

Take $F = \mathbb{F}_{13}; G, H \leq \mathbb{F}^*; G = \langle 5 \rangle, H = \langle 3 \rangle$

\[ A_G = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\} \]
\[ A_H = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}, \{7, 8, 11\}\} \]

Let

\[ F_{A_G}[x] = \{f \in F[x] : f \text{ is constant on } A_i, i = 1, 2, 3; \text{ deg } f < |F^*|\} \]
\[ F_{A_G}[x] = \langle 1, x^4, x^8 \rangle, \quad F_{A_H}[x] = \langle 1, x^3, x^6, x^9 \rangle \]
Multiple recovering sets

Take $F = F_{13}; G, H \leq F^*; G = \langle 5 \rangle, H = \langle 3 \rangle$

$\mathcal{A}_G = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\}$
$\mathcal{A}_H = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}, \{7, 8, 11\}\}$

Let

$F_{\mathcal{A}_G}[x] = \{f \in F[x] : f \text{ is constant on } A_i, i = 1, 2, 3; \deg f < |F^*|\}$

$F_{\mathcal{A}_G}[x] = \langle 1, x^4, x^8 \rangle, \quad F_{\mathcal{A}_H}[x] = \langle 1, x^3, x^6, x^9 \rangle$

We construct an LRC $(12, 4, \{2, 3\})$, distance $\geq 6$, code $C : F^4 \rightarrow F^{12}$

$a = (a_0, a_1, a_2, a_3) \mapsto f_a(x) = a_0 + a_1 x + a_2 x^4 + a_3 x^6$

$f_a(x) = \sum_{i=0}^{2} f_i(x)x^i, \text{ where } f_0(x) = a_0 + a_2 x^4, f_1(x) = a_1, f_2(x) = a_3 x^4; f_i \in F_{\mathcal{A}}[x]$

$f_a(x) = \sum_{j=0}^{1} g_j(x)x^j \text{ where } g_0(x) = a_0 + a_3 x^6, g_1(x) = a_1 + a_2 x^3; g_j \in F_{\mathcal{A}_H}[x]$

E.g., $f_a(1)$ can be recovered by computing $\delta_1(x), x \in \{5, 12, 8\} \text{ OR } \delta_2(x), x \in \{3, 9\}$
Multiple recovering sets

General Construction: \( A = \{\alpha_1, \ldots, \alpha_n\} \subseteq \mathbb{F}, |A| = n; \)

\[ A = \bigcup_{i \geq 0} R_i^1 = \bigcup_{j \geq 0} R_j^2; \quad |R_i^1| = r + 1, |R_j^2| = s + 1 \]

\[ f_a(x) = \sum_{i=0}^{k-1} a_i g_i(x), \quad g_i(x) \in \mathcal{F}_A^r \cap \mathcal{F}_{A'}^s \]

Evaluation map: \((a_1, \ldots, a_k) \mapsto (f_a(\alpha_1), \ldots, f_a(\alpha_n))\)

**Theorem:** Assume that the partitions \( A, A' \) are orthogonal. Then

\[ \text{Eval}(f : f \in \mathcal{F}_A^r \cap \mathcal{F}_{A'}^s), x \in A \]

gives an \((n, k, \{r, s\})\) LRC code with distance \( \geq n - m + 1 \), where \( m \) is the largest degree in \( \mathcal{F}_A^r \cap \mathcal{F}_{A'}^s \).
Constructing orthogonal partitions

Orthogonal partitions can be obtained from the structure of additive or multiplicative subgroups of $\mathbb{F}_q$.
Constructing orthogonal partitions

Orthogonal partitions can be obtained from the structure of additive or multiplicative subgroups of $\mathbb{F}_q$

1. $\mathbb{F}_{13}; G, H \leq \mathbb{F}_{13}^*; G = \langle 5 \rangle, H = \langle 3 \rangle; G \cap H = \text{id}$

$$\mathcal{A}_G = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\}$$

$$\mathcal{A}_H = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}, \{7, 8, 11\}\}$$
Constructing orthogonal partitions

Orthogonal partitions can be obtained from the structure of additive or multiplicative subgroups of $\mathbb{F}_q$

1. $\mathbb{F}_{13}$; $G, H \leq \mathbb{F}_{13}^*$; $G = \langle 5 \rangle$, $H = \langle 3 \rangle$; $G \cap H = \text{id}$

   $\mathcal{A}_G = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\}$

   $\mathcal{A}_H = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}, \{7, 8, 11\}\}$

2. Take $G, H \leq \mathbb{F}_q^+$, e.g., $G \cong H \cong (\mathbb{Z}_2)^2$; $\mathbb{F}_{16}^+ = (\mathbb{Z}_2)^2 \times (\mathbb{Z}_2)^2$
Constructing orthogonal partitions

Orthogonal partitions can be obtained from the structure of additive or multiplicative subgroups of $\mathbb{F}_q$

1. $\mathbb{F}_{13}; G, H \leq \mathbb{F}_{13}^*; G = \langle 5 \rangle, H = \langle 3 \rangle; G \cap H = \text{id}$

   $\mathcal{A}_G = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\}$

   $\mathcal{A}_H = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}, \{7, 8, 11\}\}$

2. Take $G, H \leq \mathbb{F}_q^+$, e.g., $G \cong H \cong (\mathbb{Z}_2)^2$; $\mathbb{F}_{16}^+ = (\mathbb{Z}_2)^2 \times (\mathbb{Z}_2)^2$

   $G = \{0000, 0001, 0010, 0011\}$ and $H = \{0000, 0100, 1000, 1100\}$

   $\mathcal{A}_G = \{\{0, 1, \alpha, \alpha^4\}, \{\alpha^5, \alpha^{10}, \alpha^2, \alpha^8\}, \{\alpha^6, \alpha^{13}, \alpha^{11}, \alpha^{12}\}, \{\alpha^7, \alpha^9, \alpha^{14}, \alpha^3\}\}$

   $\mathcal{A}_H = \{\{0, \alpha^2, \alpha^3, \alpha^6\}, \{1, \alpha^8, \alpha^{14}, \alpha^{13}\}, \{\alpha, \alpha^5, \alpha^9, \alpha^{11}\}, \{\alpha^4, \alpha^{10}, \alpha^7, \alpha^{12}\}\}$
Constructing orthogonal partitions

Orthogonal partitions can be obtained from the structure of additive or multiplicative subgroups of $\mathbb{F}_q$

1. $\mathbb{F}_{13}$; $G, H \leq \mathbb{F}_{13}^*$; $G = \langle 5 \rangle$, $H = \langle 3 \rangle$; $G \cap H = \text{id}$

   \[ A_G = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\} \]
   \[ A_H = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}, \{7, 8, 11\}\} \]

2. Take $G, H \leq \mathbb{F}_q^+$, e.g., $G \cong H \cong (\mathbb{Z}_2)^2$; $\mathbb{F}_{16}^+ = (\mathbb{Z}_2)^2 \times (\mathbb{Z}_2)^2$

   \[ G = \{0000, 0001, 0010, 0011\} \text{ and } H = \{0000, 0100, 1000, 1100\} \]
   \[ A_G = \{\{0, 1, \alpha, \alpha^4\}, \{\alpha^5, \alpha^{10}, \alpha^2, \alpha^8\}, \{\alpha^6, \alpha^{13}, \alpha^{11}, \alpha^{12}\}, \{\alpha^7, \alpha^9, \alpha^{14}, \alpha^3\}\} \]
   \[ A_H = \{\{0, \alpha^2, \alpha^3, \alpha^6\}, \{1, \alpha^8, \alpha^{14}, \alpha^{13}\}, \{\alpha, \alpha^5, \alpha^9, \alpha^{11}\}, \{\alpha^4, \alpha^{10}, \alpha^7, \alpha^{12}\}\} \]

**Proposition:** Two subgroups $G, H$ define orthogonal coset partitions if they intersect trivially: $G \cap H = \text{id}$
Remarks

There are other ways of constructing codes with multiple (e.g., two) recovering sets:

Product codes, Bipartite-graph codes
Remarks

There are other ways of constructing codes with multiple (e.g., two) recovering sets:

Product codes, Bipartite-graph codes

A family of optimal locally recoverable codes, with I. Tamo, arXiv:1311.3284
(IT Trans., no. 8, 2014)
Bounds on the parameters

Theorem

Let $C$ be an $(n, k, r, t)$ LRC code with $t$ disjoint recovering sets of size $r$. Then the rate of $C$ satisfies

$$\frac{k}{n} \leq \frac{1}{\prod_{j=1}^{t}(1 + \frac{1}{jr})} \approx t^{-\frac{1}{t}}$$

The minimum distance of $C$ is bounded above as follows:

$$d \leq n - \sum_{i=0}^{t} \left\lfloor \frac{k - 1}{r^i} \right\rfloor.$$  \hfill (Tamo – B, 2014)

$$d \leq n - k - \left\lfloor \frac{t(k - 1) + 1}{t(r - 1) + 1} \right\rfloor + 2 \hfill (Rawat e.a., 2014)$$
Bounds on the parameters

**Theorem**

Let $C$ be an $(n, k, r, t)$ LRC code with $t$ disjoint recovering sets of size $r$. Then the rate of $C$ satisfies

$$\frac{k}{n} \leq \frac{1}{\prod_{j=1}^{t}(1 + \frac{1}{jr})} \approx t^{-\frac{1}{t}}$$

The minimum distance of $C$ is bounded above as follows:

$$d \leq n - \sum_{i=0}^{t} \left\lfloor \frac{k - 1}{r^i} \right\rfloor.$$  \hspace{1cm} (Tamo – B, 2014)

$$d \leq n - k - \left\lceil \frac{t(k - 1) + 1}{t(r - 1) + 1} \right\rceil + 2$$  \hspace{1cm} (Rawat e.a., 2014)

It is likely that these bounds are not final
LRC codes

A block code of length $n$ over $\mathbb{F}_q$ is called LRC with locality $r$ if every symbol of the codeword can be found by accessing some $r$ symbols of the codeword.
A block code of length $n$ over $\mathbb{F}_q$ is called **LRC with locality** $r$ if every symbol of the codeword can be found by accessing some $r$ symbols of the codeword.

**Regenerating codes:** Data can be read off from any location
LRC codes

A block code of length $n$ over $\mathbb{F}_q$ is called **LRC with locality** $r$ if every symbol of the codeword can be found by accessing some $r$ symbols of the codeword.

**Regenerating codes:** Data can be read off from any location

Last time we constructed RS-type LRC codes with $q \approx n$, distance meeting the Gopalan et al. Singleton bound:

$$d \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2$$
LRC codes

A block code of length $n$ over $\mathbb{F}_q$ is called **LRC with locality $r$** if every symbol of the codeword can be found by accessing some $r$ symbols of the codeword.

**Regenerating codes:** Data can be read off from any location.

Last time we constructed RS-type LRC codes with $q \approx n$, distance meeting the Gopalan et al. Singleton bound:

$$d \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2$$

Asymptotic GV bound with locality:

$$R \geq \frac{r}{r + 1} - h_q(\delta)$$
Extensions

Reed-Solomon codes can be extended in two ways:

- Codes on algebraic curves
- Cyclic codes and subfield subcodes
Algebraic codes

RS codes: $n \leq q - 1, 1 \leq k \leq n, d = n - k + 1$
Algebraic codes

RS codes: $n \leq q - 1$, $1 \leq k \leq n$, $d = n - k + 1$

To construct the code, take $A := (P_1, \ldots, P_n) \subset \mathbb{F}_q$
RS codes: \( n \leq q - 1, 1 \leq k \leq n, d = n - k + 1 \)

To construct the code, take \( A := (P_1, \ldots, P_n) \subset \mathbb{F}_q \)

Message \( a = (a_1, \ldots, a_k) \in \mathbb{F}_q^k \)
Algebraic codes

RS codes: $n \leq q - 1$, $1 \leq k \leq n$, $d = n - k + 1$

To construct the code, take $A := (P_1, \ldots, P_n) \subset \mathbb{F}_q$

Message $a = (a_1, \ldots, a_k) \in \mathbb{F}_q^k \rightarrow f(x) = \sum_{i=1}^{k} a_i x^{i-1}$
Algebraic codes

RS codes: \( n \leq q - 1, 1 \leq k \leq n, d = n - k + 1 \)

To construct the code, take \( A := (P_1, \ldots, P_n) \subset \mathbb{F}_q \)

Message \( a = (a_1, \ldots, a_k) \in \mathbb{F}_q^k \implies f(x) = \sum_{i=1}^{k} a_i x^{i-1} \implies (f(P_1), \ldots, f(P_n)) \)
Algebraic codes

RS codes: \( n \leq q - 1, 1 \leq k \leq n, d = n - k + 1 \)

To construct the code, take \( A := (P_1, \ldots, P_n) \subset \mathbb{F}_q \)

Message \( a = (a_1, \ldots, a_k) \in \mathbb{F}_q^k \rightarrow f(x) = \sum_{i=1}^{k} a_i x^{i-1} \rightarrow (f(P_1), \ldots, f(P_n)) \)

Message space: span over \( \mathbb{F}_q \) of \( (x^i, i = 0, \ldots, k - 1) \)
RS codes:  \( n \leq q - 1, 1 \leq k \leq n, d = n - k + 1 \)

To construct the code, take \( A := (P_1, \ldots, P_n) \subset \mathbb{F}_q \)

Message \( a = (a_1, \ldots, a_k) \in \mathbb{F}_q^k \rightarrow f(x) = \sum_{i=1}^{k} a_i x^{i-1} \rightarrow (f(P_1), \ldots, f(P_n)) \)

Message space:
\[ \text{span over } \mathbb{F}_q \text{ of } (x^i, i = 0, \ldots, k - 1) \]

By construction, \( n \leq q \)
Algebraic codes

RS codes: $n \leq q - 1$, $1 \leq k \leq n$, $d = n - k + 1$

To construct the code, take $A := (P_1, \ldots, P_n) \subset \mathbb{F}_q$

Message $a = (a_1, \ldots, a_k) \in \mathbb{F}_q^k$ $\rightarrow$ $f(x) = \sum_{i=1}^{k} a_i x^{i-1} \rightarrow (f(P_1), \ldots, f(P_n))$

Message space:

span over $\mathbb{F}_q$ of $(x^i, i = 0, \ldots, k - 1)$

By construction, $n \leq q$ (recall the MDS conjecture)
Algebraic codes

RS codes: \( n \leq q - 1, 1 \leq k \leq n, d = n - k + 1 \)

To construct the code, take \( A := (P_1, \ldots, P_n) \subset \mathbb{F}_q \)

\[
\text{Message } a = (a_1, \ldots, a_k) \in \mathbb{F}_q^k \quad \rightarrow \quad f(x) = \sum_{i=1}^{k} a_i x^{i-1} \rightarrow (f(P_1), \ldots, f(P_n))
\]

Message space:

span over \( \mathbb{F}_q \) of \( (x^i, i = 0, \ldots, k - 1) \)

By construction, \( n \leq q \) (recall the MDS conjecture)

Longer codes? We need more points \( P_i \)
Algebraic codes

RS codes: \( n \leq q - 1, \ 1 \leq k \leq n, \ d = n - k + 1 \)

To construct the code, take \( A := (P_1, \ldots, P_n) \subset \mathbb{F}_q \)

Message \( a = (a_1, \ldots, a_k) \in \mathbb{F}_q^k \rightarrow f(x) = \sum_{i=1}^{k} a_i x^{i-1} \rightarrow (f(P_1), \ldots, f(P_n)) \)

Message space:

span over \( \mathbb{F}_q \) of \((x^i, i = 0, \ldots, k - 1)\)

By construction, \( n \leq q \) (recall the MDS conjecture)

Longer codes? We need more points \( P_i \)

Curves to the rescue!
AG codes in error correction

1. Gilbert-Varshamov bound
   An \([n, k, d]\) code exists if
   \[
   \sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i < q^{n-k}
   \]
   Let \(R = k/n, \delta = d/n\), take logs and divide by \(n\):
   \[
   R \geq 1 - h_q(\delta)
   \]
AG codes in error correction

1. Gilbert-Varshamov bound
   An \([n, k, d]\) code exists if
   \[
   \sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i < q^{n-k}
   \]
   Let \(R = k/n, \delta = d/n\), take logs and divide by \(n\):
   \[R \geq 1 - h_q(\delta)\]

2. Tsfasman-Vlăduţ-Zink bound
   There exist explicit sequences of codes on algebraic curves with the parameters
   \[R \geq 1 - \delta - \frac{1}{\sqrt{q-1}}\]
RS type codes

Given $A \subset \mathbb{F}$, partition it into $(r + 1)$-subsets.

To encode the message $a \in \mathbb{F}^k$, write $a = (a_{ij}, i = 0, \ldots, r - 1; j = 0, \ldots, \frac{k}{r} - 1)$
RS type codes

Given $A \subset \mathbb{F}$, partition it into $(r + 1)$-subsets.

To encode the message $a \in \mathbb{F}^k$, write $a = (a_{ij}, i = 0, \ldots, r - 1; j = 0, \ldots, \frac{k}{r} - 1)$

\[
a \rightarrow f_a(x) = \sum_{i=0}^{r-1} f_i(x)x^i, \quad \text{where } f_i(x) = \sum_{j=0}^{\frac{k}{r}-1} a_{ij}g(x)^j, \quad i = 0, \ldots, r - 1
\]
RS type codes

Given \( A \subset \mathbb{F} \), partition it into \((r + 1)\)-subsets.

To encode the message \( a \in \mathbb{F}^k \), write
\[
a = (a_{ij}, i = 0, \ldots, r - 1; j = 0, \ldots, \frac{k}{r} - 1)
\]

\[
a \rightarrow f_a(x) = \sum_{i=0}^{r-1} f_i(x)x^i, \quad \text{where} \quad f_i(x) = \sum_{j=0}^{\frac{k}{r}-1} a_{ij}g(x)^j, \quad i = 0, \ldots, r - 1
\]

Message space:
\[
\text{span} \left( g(x)^jx^i, i = 0, \ldots, r - 1; j = 0, \ldots, \frac{k}{r} - 1 \right)
\]
RS type codes

Given $A \subset \mathbb{F}$, partition it into $(r + 1)$-subsets.

To encode the message $a \in \mathbb{F}^k$, write $a = (a_{ij}, i = 0, \ldots, r - 1; j = 0, \ldots, \frac{k}{r} - 1)$

$$a \rightarrow f_a(x) = \sum_{i=0}^{r-1} f_i(x)x^i, \quad \text{where} \quad f_i(x) = \sum_{j=0}^{\frac{k}{r} - 1} a_{ij}g(x)^j, \quad i = 0, \ldots, r - 1$$

Message space:

$$\text{span}(g(x)^jx^i, i = 0, \ldots, r - 1; j = 0, \ldots, \frac{k}{r} - 1)$$

**Evaluation code $C$**

$$Ev : \mathbb{F}^k \rightarrow \mathbb{F}^n$$

$$a \mapsto (f_a(P), P \in A)$$
RS-like codes

What is the meaning of $g(x)$? It does not make sense that the functions are $g(x)$ $j_{x_i}$. They should really be $g(y)$ $j_{x_i}$. 

A. Barg (UMD)
RS-like codes

What is the meaning of $g(x)$?
RS-like codes

What is the meaning of $g(x)$?

It does not make sense that the functions are

$$g(x)^j x^i$$
RS-like codes

What is the meaning of $g(x)$?

It does not make sense that the functions are

$$g(x)^j x^i$$

They should really be

$$g(y)^j x^i$$
Geometric interpretation

\[ A := \{1, 2, 3, 4, 5, 6, 9, 10, 12\} \subset \mathbb{F}_{13} \]

\[ g(x) : A \rightarrow \mathbb{F}_{13} \]

\[ x \mapsto x^3 \]

\[ A = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\} \]

\[ g(A_1) = 1, g(A_2) = 8, g(A_3) = 12 \]
Geometric interpretation

\[ A := \{1, 2, 3, 4, 5, 6, 9, 10, 12\} \subset \mathbb{F}_{13} \]

\[ g(x) : A \rightarrow \mathbb{F}_{13} \]

\[ x \mapsto x^3 \]

\[ A = \{ A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\} \]

\[ g(A_1) = 1, g(A_2) = 8, g(A_3) = 12 \]

\[ g : \mathbb{F}_q \rightarrow \mathbb{F}_q \]

\[ |g^{-1}(x)| = r + 1 \text{ for every } x \text{ in the image of } g \]
A := \{1, 2, 3, 4, 5, 6, 9, 10, 12\} \subset F_{13}

\begin{align*}
g(x) : A & \rightarrow F_{13} \\
x & \mapsto x^3
\end{align*}

A = \{A_1 = \{1, 3, 9\}, A_2 = \{2, 6, 5\}, A_3 = \{4, 12, 10\}\}

g(A_1) = 1, g(A_2) = 8, g(A_3) = 12

\begin{align*}
g : F_q & \rightarrow F_q \\
|g^{-1}(x)| & = r + 1 \text{ for every } x \text{ in the image of } g
\end{align*}
Consider the set of pairs \((x, y) \in \mathbb{F}_9\) that satisfy the equation \(x^3 + x = y^4\).
Consider the set of pairs \((x, y) \in \mathbb{F}_9\) that satisfy the equation \(x^3 + x = y^4\)

27 points of the Hermitian curve over \(\mathbb{F}_9\); \(\alpha^2 = \alpha + 1\)
Recall **RS codes**: $C$ is a mapping $V_k = \langle 1, x, \ldots, x^{k-1} \rangle \rightarrow \mathbb{F}_q^n$.
Recall **RS codes**: \( C \) is a mapping \( V_k = \langle 1, x, \ldots, x^{k-1} \rangle \rightarrow \mathbb{F}_q^n \)

**Hermitian codes**

Take the space of functions \( V := \langle 1, y, y^2, x, xy, xy^2 \rangle \)

\( A = \{27 \text{ points of the Hermitian curve over } \mathbb{F}_9 \}; \quad n = 27, \quad k = 6 \)

\[ C : V \rightarrow \mathbb{F}_9^n \]
Recall RS codes: $C$ is a mapping $V_k = \langle 1, x, \ldots, x^{k-1} \rangle \to \mathbb{F}_q^n$

**Hermitian codes**

Take the space of functions $V := \langle 1, y, y^2, x, xy, xy^2 \rangle$

$A=\{27 \text{ points of the Hermitian curve over } \mathbb{F}_9\}; \ n = 27, \ k = 6$

$$C : V \to \mathbb{F}_9^n$$

E.g., message $(1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5)$

$$F(x, y) = 1 + \alpha y + \alpha^2 y^2 + \alpha^3 x + \alpha^4 xy + \alpha^5 xy^2$$

$$F(0, 0) = 1 \text{ etc.}$$
LRC codes on curves

\[
\begin{array}{cccccc}
\alpha^7 & \alpha & \alpha^7 & \alpha^5 & 0 \\
\alpha^6 & \alpha^2 & \alpha^6 & \alpha^4 & \alpha^2 & 0 \\
\alpha^5 & \alpha^7 & \alpha^3 & \alpha^5 & \alpha^5 \\
\alpha^4 & \alpha^7 & \alpha^3 & \alpha^5 & \alpha^5 \\
\alpha^3 & \alpha^3 & \alpha^7 & \alpha & \alpha \\
\alpha^2 & \alpha^3 & \alpha^7 & \alpha & \alpha \\
\alpha & 0 & 0 & 0 & 0 \\
1 & 1 & \alpha^6 & \alpha^4 & 0 \\
0 & 1 & & & & \\
0 & 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 & \alpha^5 & \alpha^6 & \alpha^7 \\
\end{array}
\]
Let $P = (\alpha, 1)$ be the erased location.
Let $P = (\alpha, 1)$ be the erased location. Recovering set $I_P = \{(\alpha^4, 1), (\alpha^3, 1)\}$

Find $f(x) : f(\alpha^4) = \alpha^7$, $f(\alpha^3) = \alpha^3$

$$\Rightarrow f(x) = \alpha x - \alpha^2$$
Let $P = (\alpha, 1)$ be the erased location. Recovering set $I_P = \{(\alpha^4, 1), (\alpha^3, 1)\}$

Find $f(x) : f(\alpha^4) = \alpha^7, f(\alpha^3) = \alpha^3$

\[ \Rightarrow f(x) = \alpha x - \alpha^2 \]

$f(\alpha) = 0 = F(P)$
Hermitian codes

\[ q = q_0^2, \ q_0 \ \text{prime power} \]
Hermitian codes

\[ q = q_0^2, \quad q_0 \text{ prime power} \]

\[ X : x^{q_0} + x = y^{q_0 + 1} \]
Hermitian codes

\[ q = q_0^2, \ q_0 \text{ prime power} \]

\[ X : x^{q_0} + x = y^{q_0+1} \]

\[ X \text{ has } q_0^3 = q^{3/2} \text{ points in } \mathbb{F}_q \]
Hermitian codes

$q = q_0^2$, $q_0$ prime power

$$X : x^{q_0} + x = y^{q_0+1}$$

$X$ has $q_0^3 = q^{3/2}$ points in $\mathbb{F}_q$

Let $g: X \rightarrow Y = \mathbb{P}^1$, $g(P) = g(x, y) := y$
Hermitian codes

$q = q_0^2$, $q_0$ prime power

$$X : x^{q_0} + x = y^{q_0+1}$$

$X$ has $q_0^3 = q^{3/2}$ points in $\mathbb{F}_q$

Let $g : X \rightarrow Y = \mathbb{P}^1$, $g(P) = g(x, y) := y$

We obtain a family of $q$-ary codes of length $n = q_0^3$,  

$$k = (t + 1)(q_0 - 1), \quad d \geq n - tq_0 - (q_0 - 2)(q_0 + 1)$$

with locality $r = q_0 - 1$. 
Hermitian codes

\[ q = q_0^2, \quad q_0 \text{ prime power} \]

\[ X : x^{q_0} + x = y^{q_0+1} \]

\[ X \text{ has } q_0^3 = q^{3/2} \text{ points in } \mathbb{F}_q \]

Let \( g : X \to Y = \mathbb{P}^1, \ g(P) = g(x, y) := y \)

We obtain a family of \( q \)-ary codes of length \( n = q_0^3 \),

\[ k = (t + 1)(q_0 - 1), \quad d \geq n - tq_0 - (q_0 - 2)(q_0 + 1) \]

with locality \( r = q_0 - 1 \).

It is also possible to take \( g(P) = x \) (projection on \( x \)); we obtain LRC codes with locality \( q_0 \).
Two recovering sets

Polynomial basis $\{x^i y^j, i = 0, 1, \ldots, r_1 - 1, j = 0, 1, \ldots, r_2 - 1\}$
Two recovering sets

Polynomial basis \( \{x^i y^j, i = 0, 1, \ldots, r_1 - 1, j = 0, 1, \ldots, r_2 - 1\} \)

\((24, 6, \{2, 3\})\) LRC(2) code over \( \mathbb{F}_9 \)
General LRC codes on curves

Map of curves

$X$, $Y$ smooth projective absolutely irreducible curves over $\mathbb{F}_k$

$$g : X \to Y$$

rational separable map of degree $r + 1$. 
General LRC codes on curves

Map of curves

$X, Y$ smooth projective absolutely irreducible curves over $\mathbb{k}$

$$g : X \to Y$$

rational separable map of degree $r + 1$.

Lift the points of $Y$

$S = \{P_1, \ldots, P_s\} \subseteq Y(\mathbb{k})$; $Q_\infty = \pi^{-1}(\infty)$, where $\pi : Y \to \mathbb{P}^1_\mathbb{k}$. Assume that there is a partition of points

$$A := g^{-1}(S) = \{P_{ij}, i = 0, \ldots, r, j = 1, \ldots, s\} \subseteq X(\mathbb{k})$$

such that

$$g(P_{ij}) = P_j \text{ for all } i, j.$$
General LRC codes on curves

Map of curves

$X, Y$ smooth projective absolutely irreducible curves over $\mathbb{k}$

$$g : X \to Y$$

rational separable map of degree $r + 1$.

Lift the points of $Y$

$S = \{P_1, \ldots, P_s\} \subset Y(\mathbb{k}); Q_\infty = \pi^{-1}(\infty)$, where $\pi : Y \to \mathbb{P}^1_\mathbb{k}$. Assume that there is a partition of points

$$A := g^{-1}(S) = \{P_{ij}, i = 0, \ldots, r, j = 1, \ldots, s\} \subset X(\mathbb{k})$$

such that

$$g(P_{ij}) = P_j$$

for all $i, j$.

Basis of the function space

$$\{f_jx^i, i = 0, \ldots, r - 1; j = 1, \ldots, m\}$$
General LRC codes on curves

Map of curves

$X$, $Y$ smooth projective absolutely irreducible curves over $\mathbb{K}$

$$g : X \to Y$$

rational separable map of degree $r + 1$.

Lift the points of $Y$

$S = \{P_1, \ldots, P_s\} \subset Y(\mathbb{K}); Q_\infty = \pi^{-1}(\infty)$, where $\pi : Y \to \mathbb{P}_\mathbb{K}^1$. Assume that there is a partition of points

$$A := g^{-1}(S) = \{P_{ij}, i = 0, \ldots, r, j = 1, \ldots, s\} \subset X(\mathbb{K})$$

such that

$$g(P_{ij}) = P_j$$

for all $i, j$.

Basis of the function space

$$\{f_j x^i, i = 0, \ldots, r - 1; j = 1, \ldots, m\}$$

Construct LRC codes

Evaluation codes constructed on the set $A$ have the locality property with parameter $r$. 
Asymptotically good sequences of codes

Let $q = q_0^2$, where $q_0$ is a prime power. Take Garcia-Stichtenoth towers of curves:

$x_0 := 1; \ X_1 := \mathbb{P}^1, \ \kappa(X_1) = \kappa(x_1);$  

$X_l : z_l^{q_0} + z_l = x_l^{q_0+1}, \ x_l^{-1} := \frac{z_l^{-1}}{x_l^{-2}} \in \kappa(X_{l-1}) \ (\text{if } l \geq 3),$
Asymptotically good sequences of codes

Let \( q = q_0^2 \), where \( q_0 \) is a prime power. Take Garcia-Stichtenoth towers of curves:

\[
x_0 := 1; \quad X_1 := \mathbb{P}^1, \quad \mathbb{k}(X_1) = \mathbb{k}(x_1);
\]

\[
X_l : z_l^{q_0} + z_l = x_{l-1}^{q_0+1}, \quad x_{l-1} := \frac{z_{l-1}}{x_{l-2}} \in \mathbb{k}(X_{l-1}) \quad (\text{if } l \geq 3),
\]

There exist families of \( q \)-ary LRC codes with locality \( r \) whose rate and relative distance satisfy

\[
R \geq \frac{r}{r+1} \left(1 - \delta - \frac{3}{\sqrt{q} + 1}\right), \quad r = \sqrt{q} - 1
\]

\[
R \geq \frac{r}{r+1} \left(1 - \delta - \frac{2\sqrt{q}}{q - 1}\right), \quad r = \sqrt{q}
\]

(better than the GV bound)
Asymptotically good sequences of codes

Let \( q = q_0^2 \), where \( q_0 \) is a prime power. Take Garcia-Stichtenoth towers of curves:

\[
x_0 := 1; \quad X_1 := \mathbb{P}^1, \quad \mathbb{L}(X_1) = \mathbb{L}(x_1);
\]
\[
X_l : z_l^{q_0} + z_l = x_l^{q_0+1}, \quad x_{l-1} := \frac{z_{l-1}}{x_{l-2}} \in \mathbb{L}(X_{l-1}) \quad \text{(if} \; l \geq 3)\).
\]

There exist families of \( q \)-ary LRC codes with locality \( r \) whose rate and relative distance satisfy

\[
R \geq \frac{r}{r+1} \left( 1 - \delta - \frac{3}{\sqrt{q} + 1} \right), \quad r = \sqrt{q} - 1
\]
\[
R \geq \frac{r}{r+1} \left( 1 - \delta - \frac{2\sqrt{q}}{q-1} \right), \quad r = \sqrt{q}
\]

(better than the GV bound)

*) Recall the TVZ bound without locality: \( R \geq 1 - \delta - \frac{1}{\sqrt{q-1}} \)
Asymptotically good sequences of codes

Let \( q = q_0^2 \), where \( q_0 \) is a prime power. Take Garcia-Stichtenoth towers of curves:

\[
\begin{align*}
x_0 & := 1; \quad X_1 := \mathbb{P}^1, \ k(X_1) = k(x_1); \\
x_l & : z^q_{l} + z = x_{l-1}^{q+1}, \quad x_{l-1} := \frac{z_{l-1}}{x_{l-2}} \in k(X_{l-1}) \text{ (if } l \geq 3),
\end{align*}
\]

There exist families of \( q \)-ary LRC codes with locality \( r \) whose rate and relative distance satisfy

\[
R \geq \frac{r}{r + 1} \left( 1 - \delta - \frac{3}{\sqrt{q} + 1} \right), \quad r = \sqrt{q} - 1
\]

\[
R \geq \frac{r}{r + 1} \left( 1 - \delta - \frac{2\sqrt{q}}{q - 1} \right), \quad r = \sqrt{q}
\]

(better than the GV bound)

Locally recoverable codes on algebraic curves, with I. Tamo and S. Vlăduț, arXiv:1501.04904
What next?
What next?

Look inside

Algebraic Codes on Lines, Planes and Curves
An Engineering Approach

Richard E. Blahut

A. Barg (UMD)
What next?

Books related to coding for memory and storage:
What next?
Another connection: Cyclic codes and Binary cyclic codes

Consider an \([n = 15, k = 4]\) RS code over \(F_{16}\);
\[A = f_1, f_2, \ldots, f_{14}\] message
\((a_1, a_2, a_3, a_4)\);
\(f(x) = a_1 + a_2 x + a_3 x^2 + a_4 x^3\)
\(f(1) = \langle (a_1, a_2, a_3, a_4) ; (1, 1, 1, 1) \rangle\)
\(f(2) = \langle (a_1, a_2, a_3, a_4) ; (1, 2, 3, 4) \rangle\)

Generator matrix
\[G = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 2 & 2 & \ldots & 2 \\ 1 & 3 & 6 & \ldots & 14 \end{bmatrix} \]
Parity-check matrix
\[H = \begin{bmatrix} 1 & 2 & 2 & \ldots & 2 \\ 1 & 1 & 11 & \ldots & 11 \\ 1 & 3 & 11 & \ldots & 11 \end{bmatrix} \]
\(c = (c_1, \ldots, c_{15})\)
\(c(x) = \sum_{i=1}^{15} c_i x^i\)
\(c(i) = 0; i = 1, \ldots, 14\)
Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \ldots, \alpha^{14}\}$
Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \ldots, \alpha^{14}\}$

message $(a_1, a_2, a_3, a_4)$; $f(x) = a_1 + a_2x + a_3x^2 + a_4x^3$

$$f(1) = \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle$$
Another connection: Cyclic codes and Binary cyclic codes

Consider an \([n = 15, k = 4]\) RS code over \(\mathbb{F}_{16}\); \(A = \{1, \alpha, \alpha^2, \ldots, \alpha^{14}\}\)

message \((a_1, a_2, a_3, a_4)\); \(f(x) = a_1 + a_2x + a_3x^2 + a_4x^3\)

\[
f(1) = \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle
\]

\[
f(\alpha) = \langle (a_1, a_2, a_3, a_4), (1, \alpha, \alpha^2, \alpha^3) \rangle
\]
Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \ldots, \alpha^{14}\}$ message $(a_1, a_2, a_3, a_4)$; $f(x) = a_1 + a_2x + a_3x^2 + a_4x^3$

\[
\begin{align*}
f(1) &= \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle \\
f(\alpha) &= \langle (a_1, a_2, a_3, a_4), (1, \alpha, \alpha^2, \alpha^3) \rangle \\
f(\alpha^2) &= \langle (a_1, a_2, a_3, a_4), (1, \alpha^2, \alpha^4, \alpha^6) \rangle
\end{align*}
\]
Another connection: Cyclic codes and Binary cyclic codes

Consider an \([n = 15, k = 4]\) RS code over \(\mathbb{F}_{16}\); \(A = \{1, \alpha, \alpha^2, \ldots, \alpha^{14}\}\) message \((a_1, a_2, a_3, a_4)\); \(f(x) = a_1 + a_2x + a_3x^2 + a_4x^3\)

\[
\begin{align*}
f(1) &= \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle \\
f(\alpha) &= \langle (a_1, a_2, a_3, a_4), (1, \alpha, \alpha^2, \alpha^3) \rangle \\
f(\alpha^2) &= \langle (a_1, a_2, a_3, a_4), (1, \alpha^2, \alpha^4, \alpha^6) \rangle
\end{align*}
\]

<table>
<thead>
<tr>
<th>Generator matrix</th>
<th>Parity-check matrix</th>
</tr>
</thead>
</table>
| \[
G = \begin{pmatrix}
1 & 1 & 1 & \ldots & 1 \\
1 & \alpha & \alpha^2 & \ldots & \alpha^{14} \\
1 & \alpha^2 & \alpha^4 & \ldots & \alpha^{2\cdot14} \\
1 & \alpha^3 & \alpha^6 & \ldots & \alpha^{3\cdot14}
\end{pmatrix}
| \]
Another connection: Cyclic codes and Binary cyclic codes

Consider an \([n = 15, k = 4]\) RS code over \(\mathbb{F}_{16}\); \(A = \{1, \alpha, \alpha^2, \ldots, \alpha^{14}\}\)

message \((a_1, a_2, a_3, a_4)\); \(f(x) = a_1 + a_2 x + a_3 x^2 + a_4 x^3\)

\[
\begin{align*}
  f(1) &= \langle (a_1, a_2, a_3, a_4), (1, 1, 1, 1) \rangle \\
  f(\alpha) &= \langle (a_1, a_2, a_3, a_4), (1, \alpha, \alpha^2, \alpha^3) \rangle \\
  f(\alpha^2) &= \langle (a_1, a_2, a_3, a_4), (1, \alpha^2, \alpha^4, \alpha^6) \rangle
\end{align*}
\]

Generator matrix \(G\):

\[
G = \begin{pmatrix}
  1 & 1 & 1 & \ldots & 1 \\
  1 & \alpha & \alpha^2 & \ldots & \alpha^{14} \\
  1 & \alpha^2 & \alpha^4 & \ldots & \alpha^{2\cdot14} \\
  1 & \alpha^3 & \alpha^6 & \ldots & \alpha^{3\cdot14}
\end{pmatrix}
\]

Parity-check matrix \(H\):

\[
H = \begin{pmatrix}
  1 & \alpha & \alpha^2 & \ldots & \alpha^{14} \\
  1 & \alpha^2 & \alpha^{2\cdot2} & \ldots & \alpha^{14\cdot2} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  1 & \alpha^{11} & \alpha^{2\cdot11} & \ldots & \alpha^{14\cdot11}
\end{pmatrix}
\]
Another connection: Cyclic codes and Binary cyclic codes

Consider an $[n = 15, k = 4]$ RS code over $\mathbb{F}_{16}$; $A = \{1, \alpha, \alpha^2, \ldots, \alpha^{14}\}$ message $(a_1, a_2, a_3, a_4)$; $f(x) = a_1 + a_2x + a_3x^2 + a_4x^3$

$$f(1) = \langle(a_1, a_2, a_3, a_4), (1, 1, 1, 1)\rangle$$

$$f(\alpha) = \langle(a_1, a_2, a_3, a_4), (1, \alpha, \alpha^2, \alpha^3)\rangle$$

$$f(\alpha^2) = \langle(a_1, a_2, a_3, a_4), (1, \alpha^2, \alpha^4, \alpha^6)\rangle$$

Generator matrix

$$G = \begin{pmatrix}
1 & 1 & 1 & \ldots & 1 \\
1 & \alpha & \alpha^2 & \ldots & \alpha^{14} \\
1 & \alpha^2 & \alpha^4 & \ldots & \alpha^{2\cdot14} \\
1 & \alpha^3 & \alpha^6 & \ldots & \alpha^{3\cdot14}
\end{pmatrix}$$

Parity-check matrix

$$H = \begin{pmatrix}
1 & \alpha & \alpha^2 & \ldots & \alpha^{14} \\
1 & \alpha^2 & \alpha^{2\cdot2} & \ldots & \alpha^{14\cdot2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \alpha^{11} & \alpha^{2\cdot11} & \ldots & \alpha^{14\cdot11}
\end{pmatrix}$$

$$c = (c_1, \ldots, c_{15}); \quad c(x) = \sum_{i=1}^{15} c_i x^{i-1}; \quad c(\alpha^i) = 0, i = 1, \ldots, 14$$
BCH codes: Subfield subcodes of RS codes

- Consider the subset of vectors of the RS code with coordinates 0 or 1
- \( c(x) = \sum_{i=1}^{n} x^i : c(\alpha^j) = 0 \)
- They form a \textit{BCH code}, a binary cyclic code of length \( 2^m - 1 \)
- This construction is called a \textit{Subfield Subcode}
  
  Observation 1: expand parity-check matrix
  Observation 2: conjugate roots
Cyclic codes

Consider an \([n|(q-1), k = n - d + 1, d]\) RS code \(C\) over \(\mathbb{F}_q\)

\[A = (1, \alpha, \ldots, \alpha^{n-1})\] where \(\alpha^n = 1\)

**Generator matrix**

\[
G = \begin{pmatrix}
1 & 1 & 1 & \cdots & 1 \\
1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \alpha^{k-1} & \alpha^{2(k-1)} & \cdots & \alpha^{(k-1)(n-1)}
\end{pmatrix}
\]

**Parity-check matrix**

\[
H = \begin{pmatrix}
1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\
1 & \alpha^2 & \alpha^2 \cdot 2 & \cdots & \alpha^{2(n-1)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \alpha^{n-k} & \alpha^{2(n-k)} & \cdots & \alpha^{(n-k)(n-1)}
\end{pmatrix}
\]

\(C\) is a cyclic code with zeros \(\alpha, \alpha^2, \ldots, \alpha^{n-k}\)
Cyclic codes

- Consider an $[n|(q - 1), k = n - d + 1, d]$ RS code $C$ over $\mathbb{F}_q$

  $A = (1, \alpha, \ldots, \alpha^{n-1})$ where $\alpha^n = 1$

  Generator matrix

  $$
  G = \begin{pmatrix}
  1 & 1 & 1 & \ldots & 1 \\
  1 & \alpha & \alpha^2 & \ldots & \alpha^{n-1} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  1 & \alpha^{k-1} & \alpha^{2(k-1)} & \ldots & \alpha^{(k-1)(n-1)}
  \end{pmatrix}
  $$

  Parity-check matrix

  $$
  H = \begin{pmatrix}
  1 & \alpha & \alpha^2 & \ldots & \alpha^{n-1} \\
  1 & \alpha^2 & \alpha^{2\cdot 2} & \ldots & \alpha^{2(n-1)} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  1 & \alpha^{n-k} & \alpha^{2(n-k)} & \ldots & \alpha^{(n-k)(n-1)}
  \end{pmatrix}
  $$

  $C$ is a cyclic code with zeros $\alpha, \alpha^2, \ldots, \alpha^{n-k}$

- Consider a subfield subcode $D \subset C$

  $D := \{(c_0, \ldots, c_{n-1}) \in C : c_j \in \mathbb{F}_p, 0 \leq j \leq n - 1\}$

  Zeros of $D$: $\{(\alpha, \alpha^p, \ldots, \alpha^{p^{m-1} \mod n}), \ldots\}$
Cyclic codes: Example

- RS code $C$ of length $n = 15$, $k = 8$, $d = 8$, $q = 2^4$
- Zeros of $C$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
- Generator polynomial $g(x) = \prod_{i=1}^{7}(x - \alpha^i)$, $\dim(C) = n - \deg(g) = 8$

BCH bound: $d(C) \geq$ number of consecutive 0's + 1
Cyclic codes: Example

- RS code $C$ of length $n = 15$, $k = 8$, $d = 8$, $q = 2^4$
- Zeros of $C$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
- Generator polynomial $g(x) = \prod_{i=1}^{t} (x - \alpha^i)$, $\dim(C) = n - \deg(g) = 8$
- BCH bound: $d(C) \geq $ number of consecutive 0's + 1

- Now suppose that $C$ has zeros $\{\alpha, \ldots, \alpha^7\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. The distance is the same, and we get the locality condition $r = 2$
Cyclic codes: Example

- RS code $\mathcal{C}$ of length $n = 15$, $k = 8$, $d = 8$, $q = 2^4$
- Zeros of $\mathcal{C}$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
- Generator polynomial $g(x) = \prod_{i=1}^t (x - \alpha^i)$, $\dim(\mathcal{C}) = n - \deg(g) = 8$

BCH bound: $d(\mathcal{C}) \geq$ number of consecutive 0's + 1

- Now suppose that $\mathcal{C}$ has zeros $\{\alpha, \ldots, \alpha^7\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. The distance is the same, and we get the locality condition $r = 2$
- Indeed,

\[ f_\alpha(x) = a_1 + a_2x + a_3x^3 + a_4x^4 + a_5x^6 + a_6x^7 \]
Cyclic codes: Example

- RS code $C$ of length $n = 15$, $k = 8$, $d = 8$, $q = 2^4$
- Zeros of $C$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
- Generator polynomial $g(x) = \prod_{i=1}^{t} (x - \alpha^i)$, $\dim(C) = n - \deg(g) = 8$
- BCH bound: $d(C) \geq$ number of consecutive 0's + 1

Now suppose that $C$ has zeros $\{\alpha, \ldots, \alpha^7\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. The distance is the same, and we get the locality condition $r = 2$
- Indeed,
  $$f_a(x) = a_1 + a_2x + a_3x^3 + a_4x^4 + a_5x^6 + a_6x^7$$
- So the rows of $G$ are $1, \alpha, \alpha^3, \alpha^4, \alpha^6, \alpha^7$
Cyclic codes: Example

- RS code $C$ of length $n = 15$, $k = 8$, $d = 8$, $q = 2^4$
- Zeros of $C$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
- Generator polynomial $g(x) = \prod_{i=1}^{t}(x - \alpha^i)$, $\dim(C) = n - \deg(g) = 8$
- BCH bound: $d(C) \geq$ number of consecutive 0's + 1

- Now suppose that $C$ has zeros $\{\alpha, \ldots, \alpha^7\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. The distance is the same, and we get the locality condition $r = 2$
- Indeed,
  \[ f_a(x) = a_1 + a_2x + a_3x^3 + a_4x^4 + a_5x^6 + a_6x^7 \]
- So the rows of $G$ are $1, \alpha, \alpha^3, \alpha^4, \alpha^6, \alpha^7$
- The rows of $H$ are $-(\{n\} \setminus \{0, 1, 3, 4, 6, 7\})$
Cyclic codes: Example

- RS code $C$ of length $n = 15$, $k = 8$, $d = 8$, $q = 2^4$
  - Zeros of $C$: $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$
  - Generator polynomial $g(x) = \prod_{i=1}^{t}(x - \alpha^i)$, $\dim(C) = n - \deg(g) = 8$

BCH bound: $d(C) \geq \text{number of consecutive 0's} + 1$

- Now suppose that $C$ has zeros $\{\alpha, \ldots, \alpha^7\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. The distance is the same, and we get the locality condition $r = 2$
  - Indeed,
    
    $$f_a(x) = a_1 + a_2x + a_3x^3 + a_4x^4 + a_5x^6 + a_6x^7$$

So the rows of $G$ are $1, \alpha, \alpha^3, \alpha^4, \alpha^6, \alpha^7$

The rows of $H$ are $-([n] \setminus \{0, 1, 3, 4, 6, 7\})$

$$k = 6; d = 8 = n - k \frac{r + 1}{r} + 2$$
Main idea: Suppose that the zeros are arranged as follows:

The cyclic code with zeros $\{D \cup L\}$ has distance $\geq |D|$ and locality $r$. 
The following result describes the cyclic case of the main construction.

**Theorem (RS-type cyclic LRC codes):** Let $\alpha$ be a primitive $n$-th root of unity, where $n|(q - 1)$; let $l, 0 \leq l \leq r$ be an integer. Consider the following sets of elements of $\mathbb{F}_q$:

$$L = \{\alpha^i, i \mod (r + 1) = l\},$$

and

$$D = \{\alpha^{j+s}, s = 0, \ldots, n - \frac{k}{r}(r + 1)\},$$

where $\alpha^j \in L$. The cyclic code with the defining set of zeros $L \cup D$ is an optimal *) $(n, k, r)$ $q$-ary cyclic LRC code.

*) Singleton-like optimality; see (1)
Let $\mathcal{C}$ be a cyclic LRC code over $\mathbb{F}_q$. 
Let $C$ be a cyclic LRC code over $\mathbb{F}_q$.

**Dual code**

$$C^\perp := \{ x \in \mathbb{F}_q : \langle x, c \rangle = 0 \text{ for all } c \in C \}$$
Locality and dual distance

Let \( \mathcal{C} \) be a cyclic LRC code over \( \mathbb{F}_q \).

**Dual code**

\[
\mathcal{C}^\perp := \{ \mathbf{x} \in \mathbb{F}_q : \langle \mathbf{x}, \mathbf{c} \rangle = 0 \text{ for all } \mathbf{c} \in \mathcal{C} \}
\]

**Locality of \( \mathcal{C} \):**

\[
r = d(\mathcal{C}^\perp) = d^\perp(\mathcal{C})
\]

In the cyclic case **Locality=Dual distance**
Subfield subcodes

What about binary codes?
Subfield subcodes

What about binary codes?

Example:
Code $C$ over $\mathbb{F}_{16}$ has zeros $Z = \{\alpha, \alpha^2, \alpha^3, \alpha^4\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$. 
What about binary codes?

Example: Code $\mathcal{C}$ over $\mathbb{F}_{16}$ has zeros $Z = \{\alpha, \alpha^2, \alpha^3, \alpha^4\} \cup \{\alpha, \alpha^4, \alpha^7, \alpha^{10}, \alpha^{13}\}$.

**Binary subcode** $D \subset \mathcal{C}$: zeros $Z$ and all conjugates
The locality of $D$ may decrease; the distance may increase. The dimension becomes smaller.
Subfield subcodes

What about binary codes?
Subfield subcodes

What about binary codes?

Let $C$ be a cyclic code over $\mathbb{F}_{q^m}$; let $D$ be the subfield subcode of $C$

$$D := \{ c = (c_1, \ldots, c_n) \in C : c_i \in \mathbb{F}_q, i = 1, \ldots, n \}$$
Subfield subcodes

What about binary codes?

Let $C$ be a cyclic code over $\mathbb{F}_{q^m}$; let $D$ be the subfield subcode of $C$

$$D := \{ c = (c_1, \ldots, c_n) \in C : c_i \in \mathbb{F}_q, i = 1, \ldots, n \}$$

We have:

$$d(D) \geq d(C)$$
Subfield subcodes

What about binary codes?

Let $C$ be a cyclic code over $\mathbb{F}_{q^m}$; let $D$ be the subfield subcode of $C$

$$D := \{ c = (c_1, \ldots, c_n) \in C : c_i \in \mathbb{F}_q, i = 1, \ldots, n \}$$

We have:

$$d(D) \geq d(C)$$

$$d^\perp(D) \leq d^\perp(C)$$
Subfield subcodes

What about binary codes?

Let $C$ be a cyclic code over $\mathbb{F}_{q^m}$; let $D$ be the subfield subcode of $C$

$$D := \{ c = (c_1, \ldots, c_n) \in C : c_i \in \mathbb{F}_q, i = 1, \ldots, n \}$$

We have:

$$d(D) \geq d(C)$$

$$d^\perp(D) \leq d^\perp(C)$$

$$r(D) \leq r(C)$$
Subfield subcodes

The analysis: Ideas.

- Take a subfield subcode $D$ of the code $C$ constructed in the RS-like LRC codes Theorem.
- Locality of $D = \text{distance of } D^\perp$
- Let $q = 2^m$, $T_m(x) = x + x^2 + \cdots + x^{2^m-1}$, $x \in \mathbb{F}_q$
  \[
  T_m(C) := \{(T_m(c_1), \ldots, T_m(c_n)), c \in C\}
  \]

**Theorem** (Delsarte ’74, Sidelnikov ’71): $D = T_m(C^\perp)$

- Analyze the locality of $D$ using $d(D^\perp)$ (techniques: irreducible cyclic codes)
Some examples

\[
\begin{array}{cccccccccc}
\begin{array}{cccccccc}
\text{n} & \text{k} & \text{d} & \text{Z(D)} & \text{r} & \text{w} & \text{Z((C')⊥)} & \text{d⊥} & \text{SH} & \text{LP} & \text{locator field} \\
35 & 20 & 3 & \{1, 15\} & r \leq 3 & 4 & \{0, 1, 7, 15\} & 4 & k \leq 25 & k \leq 29 & \mathbb{F}_{212} \\
45 & 33 & 3 & \{1\} & r \leq 7 & 8 & \{0, 1, 3, 5, 9, 15, 21\} & 8 & k \leq 37 & k \leq 39 & \mathbb{F}_{212} \\
27 & 7 & 6 & \{1, 9\} & r = 1 & 2 & \{0, 3\} & 2 & & & \\
63 & 36 & 3 & \{1, 9, 11, 15, 23\} & r \leq 3 & 4 & \{0, 1, 7, 9, 11, 15, 21, 23\} & 4 & & & \\
\end{array}
\end{array}
\]

\(Z(C)\) = defining set of zeros of \(C\), \(w\) is the number of recovering sets \(A_i\)
Some examples

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>$d$</th>
<th>$Z(D)$</th>
<th>$r$</th>
<th>$w$</th>
<th>$Z((C')^\perp)$</th>
<th>$d^\perp$</th>
<th>$SH$</th>
<th>$LP$</th>
<th>locator field</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>20</td>
<td>3</td>
<td>${1, 15}$</td>
<td>$r \leq 3$</td>
<td>4</td>
<td>${0, 1, 7, 15}$</td>
<td>4</td>
<td>$k \leq 25$</td>
<td>$k \leq 29$</td>
<td>$F_{2^{12}}$</td>
</tr>
<tr>
<td>45</td>
<td>33</td>
<td>3</td>
<td>${1}$</td>
<td>$r \leq 7$</td>
<td>8</td>
<td>${0, 1, 3, 5, 9, 15, 21}$</td>
<td>8</td>
<td>$k \leq 37$</td>
<td>$k \leq 39$</td>
<td>$F_{2^{12}}$</td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>6</td>
<td>${1, 9}$</td>
<td>$r = 1$</td>
<td>2</td>
<td>${0, 3}$</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>36</td>
<td>3</td>
<td>${1, 9, 11, 15, 23}$</td>
<td>$r \leq 3$</td>
<td>4</td>
<td>${0, 1, 7, 9, 11, 15, 21, 23}$</td>
<td>4</td>
<td></td>
<td></td>
<td>$F_{2^{18}}$</td>
</tr>
</tbody>
</table>

$Z(C)$ = defining set of zeros of $C$, $w$ is the number of recovering sets $A_i$