Complete Characterization of Optimal LRCs with Minimum Distance 6 and Locality 2: Improved Bounds and Constructions

Weijun Fang

Joint work with Bin Chen, Shu-Tao Xia & Fang-Wei Fu

Shenzhen International Graduate School, Tsinghua University

2020 IEEE International Symposium on Information Theory (ISIT), Jun 21-26, 2020
Outline

1. Introduction
2. Related Work
3. Main Results
   - Complete Characterization
   - New Constructions
   - New Upper Bounds
4. Conclusion Remarks
Outline

1. Introduction
2. Related Work
3. Main Results
   - Complete Characterization
   - New Constructions
   - New Upper Bounds
4. Conclusion Remarks
Distributed Storage System (DSS)

Locally Repairable Codes

Formal Definition: $r$-locality

The $i$-th code symbol of $C$ is said to have locality $r$ if there exists a subset $R \subseteq [n]$, such that

- $i \in R$ and $|R| \leq r + 1$;
- $d(C|_R) \geq 2$.

$C$ is called an $(n, k, d; r)$-LRC if each code symbol of $C$ has $r$-locality.

- Any symbol can be recovered by at most $r$ other symbols
- Large code rate $\frac{r}{n}$, large minimum distance $d$ and small locality $r$

---

Theorem (Singleton-type Bound\textsuperscript{a})


Let $C$ be an $(n, k, d; r)$-LRC. Then

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

- Generalization of the classical Singleton bound ($r = k$);
- $C$ is called an optimal $(n, k, d; r)$ LRC code if it achieves the bound with equality.

**Goal:** Construction of optimal LRCs
Outline

1. Introduction

2. Related Work

3. Main Results
   - Complete Characterization
   - New Constructions
   - New Upper Bounds

4. Conclusion Remarks
Related Work

Lots of works have been proposed for construction of optimal LRCs.

- \(^{(2)}\) N. Prakash et al. ISIT2012): \( n = \left\lceil \frac{k}{r} \right\rceil (r + \delta - 1) \) for \( n < q \).

- \(^{(3)}\) N. Silberstein et al. ISIT2013): The alphabet size that is exponential in code length.

- \(^{(4)}\) I. Tamo et al. TIT2014): The code length can go up to the alphabet size by using subcodes of Reed-Solomon codes.

---


Optimal LRCs with length $n=q+1$:

- (L. Jin et al. TIT2019): $r$-LRCs; via automorphism group of rational function fields.

- (B. Chen et al. TIT2018, TCOM2019): $(r, \delta)$-LRCs; $n \mid q + 1$ via cyclic or constacyclic codes.
Based on the **classical MDS conjecture**, one should wonder if $q$-ary optimal LRCs can have length bigger than $q + 1$.

- (§X. Li et al. TIT2019): $r$-LRCs; Length $n$ can be up to $q + 2\sqrt{q}$ via elliptic curves.
- (§Y. Luo et al. TIT2019): $r$-LRCs; **Unbounded length** with minimum distances 3 and 4 via cyclic codes.

---


For $d \geq 5$, Guruswami et al.\(^\text{10}\) (TIT2019) have proved that $n = \mathcal{O}(dq^3)$;

- $d = 5, 6, \, r \geq d - 2$: $n = \Omega(q^2)$ (Jin\(^\text{11}\) TIT2019);

- $d \geq 7, \, r \geq d - 2$: Some constructions of optimal LRCs with large length (Xing and Yuan\(^\text{12}\)).

---

\(^\text{10}\)V. Guruswami, C. Xing, and C. Yuan. “How long can optimal locally repairable codes be?” TIT 2019.

\(^\text{11}\)L. Jin, Explicit construction of optimal locally recoverable codes of distance 5 and 6 via binary constant weight codes, TIT 2019.

**Related Work**

- $r < d - 2 : (\text{Chen et al}^{13} \text{ ISIT2019})$
  - $d = 5$, $r = 2$: Maximal length $n = q + 1$;
  - $d = 6$, $r = 2$: $n = 3(q + 1)$, $q$ is odd

- **Our Goal**: For $d = 6$ and $r = 2$:
  - Improve the upper bound.
  - New constructions of optimal LRCs with large code length.

---

Outline

1 Introduction

2 Related Work

3 Main Results
   - Complete Characterization
   - New Constructions
   - New Upper Bounds

4 Conclusion Remarks
Outline

1. Introduction

2. Related Work

3. Main Results
   - Complete Characterization
   - New Constructions
   - New Upper Bounds

4. Conclusion Remarks
- \((r + 1) \mid n, \ell = \frac{n}{r+1}\), LRCs have \(\ell\) disjoint local repair groups.
- The parameters \(n, k, d, r\) are unchanged under the code equivalence.
- \(r = 2\)
- \(i\)-th symbol has \(r\)-locality \(\Leftrightarrow \exists h \in C^\perp\), such that
  - \(i \in \text{supp}(h)\);
  - |\text{supp}(h)| \leq r + 1.
- Parity-check matrix

\[
H = \begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 1 & 1 \\
0 & u_1 & v_1 & 0 & u_2 & v_2 & \ldots & 0 & u_\ell & v_\ell
\end{pmatrix}
\]
Denote $\mathcal{V}_i = \text{Span}\{u_i, v_i\}$ to be a subspace of $\mathbb{F}_q^3$ spanned by $u_i$ and $v_i$.

Lemma 1

Suppose $q \geq 3$ and $3 \mid n$. Let $C$ be a $q$-ary optimal $(n, d = 6; r = 2)$-LRC with a parity-check matrix $H$. Then

(i) $\dim(\mathcal{V}_i) = 2$;

(ii) for any $j \neq i \in [\ell]$, we have $u_i, v_i, u_i - v_i \notin \mathcal{V}_j$. 
Main Results
Complete Characterization

- $[\mathcal{V}_i]_q$: the set of all 1-subspaces of $\mathcal{V}_i$. Clearly, $|[\mathcal{V}_i]_q| = q + 1$ (Exercise).

Theorem 1 (Complete Characterization of Optimal $(n, d = 6; r = 2)$-LRCs)

Suppose $q \geq 3$ and $3 \mid n$. There exists a $q$-ary optimal $(n, d = 6; r = 2)$-LRC with disjoint repair groups if and only if there exist $\ell$ distinct 2-subspaces $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_\ell$ of $\mathbb{F}_q^3$ such that for each $i \in [\ell],

$$t_i \triangleq \left| \bigcup_{j \in [\ell], j \neq i} \left( [\mathcal{V}_i]_q \cap [\mathcal{V}_j]_q \right) \right| \leq q - 2.$$
Sketch of Proof $\Rightarrow$: By Lemma 1, $\text{Span}\{u_i\}, \text{Span}\{v_i\}, \text{Span}\{u_i - v_i\}$ are 3 distinct 1-subspaces in $[\mathcal{V}_i]_q$, which do not belong to $[\mathcal{V}_j]_q$ for all $j \neq i$;

$\Leftarrow$:

- There exist three distinct 1-subspaces $\text{Span}\{u_i\}, \text{Span}\{v_i\}, \text{Span}\{w_i\}$ of $\mathcal{V}_i$, such that $u_i, v_i, w_i \notin \mathcal{V}_i \bigcap \mathcal{V}_j$ for any $j \in [\ell]$ with $j \neq i$.
- $\dim(\mathcal{V}_i) = 2 \Rightarrow w_i = au_i - bv_i$, for some $a, b \in \mathbb{F}_q^*$.
- By replacing $u_i$ and $v_i$ with $au_i$ and $bv_i$, respectively, we can assume that $w_i = u_i - v_i$.

Thus for any $j \neq i \in [\ell]$, we have $u_i, v_i, u_i - v_i \notin \mathcal{V}_j$. 
Let $C$ be the linear code with parity-check $H$

$k = n - \text{rank}(H) \geq n - \ell - 3$

$d \geq 6$: Any 5 columns of $H$ are linearly independent

Singleton-type bound $d \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2 \leq \ell + 3 - (\ell - 1) + 2 = 6$
- $PG(2, q)$: Projective plane over finite field $\mathbb{F}_q$
- A 2-subspace of $\mathbb{F}_q^3$ corresponds to a line in $PG(2, q)$ while a 1-subspace of $\mathbb{F}_q^3$ corresponds to a point in $PG(2, q)$.

**Theorem 2 (Geometrical Characterization of Optimal $(n, k, 6; 2)$-LRCs)**

Suppose $q \geq 3$ and $3 \mid n$. Then, there exists an optimal $(n, d = 6, r = 2)$-LRC with disjoint repair groups if and only if there exist $\ell$ distinct lines $L_1, L_2, \ldots, L_\ell$ in $PG(2, q)$, such that each $L_i$ has at most $q - 2$ distinct intersection points.
Outline

1. Introduction

2. Related Work

3. Main Results
   - Complete Characterization
   - New Constructions
   - New Upper Bounds

4. Conclusion Remarks
Theorem 2 (Sunflower Construction)
Let $q \geq 3$ be a prime power, then there exists an optimal $(3(q + 1), d = 6, r = 2)$-LRC.

We remove the condition of “$q$ is odd” in Chen et al. ISIT2019.
Double-Sunflower Construction

Theorem 3 (Double-Sunflower Construction)

Let \( q \geq 5 \) be a prime power, then there exists a \( q \)-ary optimal \((n = 3(2q - 4), d = 6; r = 2)\)-LRC.
Outline

1. Introduction
2. Related Work
3. Main Results
   - Complete Characterization
   - New Constructions
   - New Upper Bounds
4. Conclusion Remarks
New Upper Bounds

- Known results: \((d = 6, r = 2)\)
  - \(^{14}\) Guruswami et al. : \(n = O(q^3)\);
  - \(^{15}\) Chen et al. : \(n \leq 3 \left\lfloor \frac{q^2 + q + 1}{3} \right\rfloor \).

- \(L_1, L_2, \cdots, L_\ell\) are \(\ell\) lines of \(PG(2, q)\) satisfying the intersection condition.


- **Line-point incidence matrix** $A_{\ell \times (q^2+q+1)}$: the $a_{ij} \neq 0$ if and only if the $j$-th point lies in $L_i$;
- Each row has weight $q + 1$: each line has $q + 1$ points.

$$
A = 
\begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 1 & 1 \\
\end{pmatrix}
$$

$3\ell$ columns
Lemma 2

Suppose $q \geq 3$. If there exist $\ell$ distinct lines $L_1, L_2, \ldots, L_\ell$ in $PG(2, q)$ satisfying the intersection condition which do not form a Sunflower, then

$$\ell \leq \left\lfloor \frac{q^2 + q + 1}{4} \right\rfloor.$$

Proof:

- Each column of $A'$ has Hamming weight at most $q - 2$

Calculating the number of 1’s in $A'$ in two ways, we obtain that

$$\ell \times (q - 2) \leq (q^2 + q + 1 - 3\ell) \times (q - 2), \text{ i.e., } \ell \leq \left\lfloor \frac{q^2 + q + 1}{4} \right\rfloor.$$
Theorem 4 (First New Upper Bound)

Suppose $q \geq 3$ and let $C$ be a $q$-ary optimal $(n, k, 6; 2)$-LRC with disjoint local repair groups, then

$$n \leq \max \left\{ 3(q + 1), 3 \left\lfloor \frac{q^2 + q + 1}{4} \right\rfloor \right\}.$$ 

- $3 \left\lfloor \frac{q^2 + q + 1}{4} \right\rfloor$ is better than $3 \left\lfloor \frac{q^2 + q + 1}{3} \right\rfloor$.
- $n_{\text{max}}(q)$: largest $n$ such that there exists a $q$-ary optimal $(n, d = 6; r = 2)$-LRC with disjoint local repair groups
Theorem 5
The maximal code length of 4-ary optimal LRCs with $d = 6$ and $r = 2$: $n_{\text{max}}(4) = 15$;

Proof:
- From the Sunflower Construction, $n_{\text{max}}(4) \geq 15$;
- By Theorem 4, $n_{\text{max}}(4) \leq 15$. 
Binary constant weight code: A binary $(n, M, d, w)$ constant weight code is a set of binary vectors of length $n$ with size $M$, such that each codeword contains $w$ 1’s, and any two codewords differ in at least $d$ positions.

**Lemma 3 (Johnson Bound)**

Let $C$ be a binary $(n, M, d = 2\delta; w)$-constant weight code, then

$$M(w^2 - wn + \delta n) \leq \delta n.$$
Theorem 6 (New Upper Bound)

Suppose \( q \geq 4 \) and let \( C \) be a \( q \)-ary optimal \((n, k, 6; 2)\)-LRC with disjoint repair groups, then

\[
  n \leq 3 \left\lfloor \frac{q + 3 + q\sqrt{3q - 5}}{3} \right\rfloor = O(q^{1.5}).
\]

- Until now, **best bound** on the code length of optimal \((n, k, 6; 2)\)-LRCs
Main Results

New Upper Bounds

\[ A = \begin{pmatrix}
A' \\
1 1 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 1 & 1
\end{pmatrix}
\]

Proof:

- Any two lines in \( PG(2, q) \) intersect at exactly one point \( \Rightarrow \) the Hamming distance of any two distinct rows of \( A' \) is equal to \( 2(q - 2) - 2 = 2q - 6 \)
- \( C' \): binary code whose codewords are the row vectors of \( A' \)
- \( C' \) is exactly a binary \((n, M, d; w)\)-constant weight code with \( n = q^2 + q + 1 - 3\ell, M = \ell, d = 2q - 6 \) and \( w = q - 2 \)

Johnson bound \( \Rightarrow \ell(3\ell - 5q + 3) \leq (q^2 + q + 1 - 3\ell)(q - 3) \Rightarrow 3\ell^2 - (2q + 6)\ell - (q^2 + q + 1)(q - 3) \leq 0 \Rightarrow \ell \leq \left\lfloor \frac{q + 3 + q\sqrt{3q - 5}}{3} \right\rfloor \).
Theorem 7

The maximal code length of 5-ary optimal LRCs with $d = 6$ and $r = 2$: $n_{\text{max}}(5) = 18$.

Proof:

- By Sunflower Construction or Double-Sunflower Construction, $n_{\text{max}}(5) \geq 18$;
- By Theorem 6, $n_{\text{max}}(5) \leq 21$;
- When $n_{\text{max}}(5) = 21$, the lines form an Fano plane over $\mathbb{F}_5$, which is impossible.
Outline

1 Introduction

2 Related Work

3 Main Results
   - Complete Characterization
   - New Constructions
   - New Upper Bounds

4 Conclusion Remarks
Main Contribution:
- We have established a complete characterization for optimal LRCs with \( d = 6 \) and \( r = 2 \) via the point of view of geometry;
- New construction of optimal LRCs with large length \( n = 3(2q-4) \);
- New upper bound: \( n = O(q^{1.5}) \).

Research Problems and Future Work:
- Construct optimal LRCs with \( n \approx q^{1.5} \) or prove that \( n = O(q) \);
- How about \( d = 6, r = 3? \ d \geq 7? \ldots \);
- Generalize to \((r, \delta)\)-LRCs.
- \ldots\
Thank You!