A Sphere-Packing Bound for Locally Repairable Codes

Anyu Wang, Zhifang Zhang, Dongdai Lin

Academy of Mathematics and Systems Science
Chinese Academy of Sciences

The 5th Sino-Korea International Conference on Coding Theory and Related Topics

Shanghai, July, 2018

Related Papers

1. Anyu Wang, Zhifang Zhang, Dongdai Lin: Bounds and constructions for linear locally repairable codes over binary fields. ISIT 2017: 2033-2037.

   (Focus on binary LRCs)

2. Anyu Wang, Zhifang Zhang, Dongdai Lin, Bounds for Binary Linear Locally Repairable Codes via a Sphere-Packing Approach, arXiv CoRR abs/1701.05989 (2017).

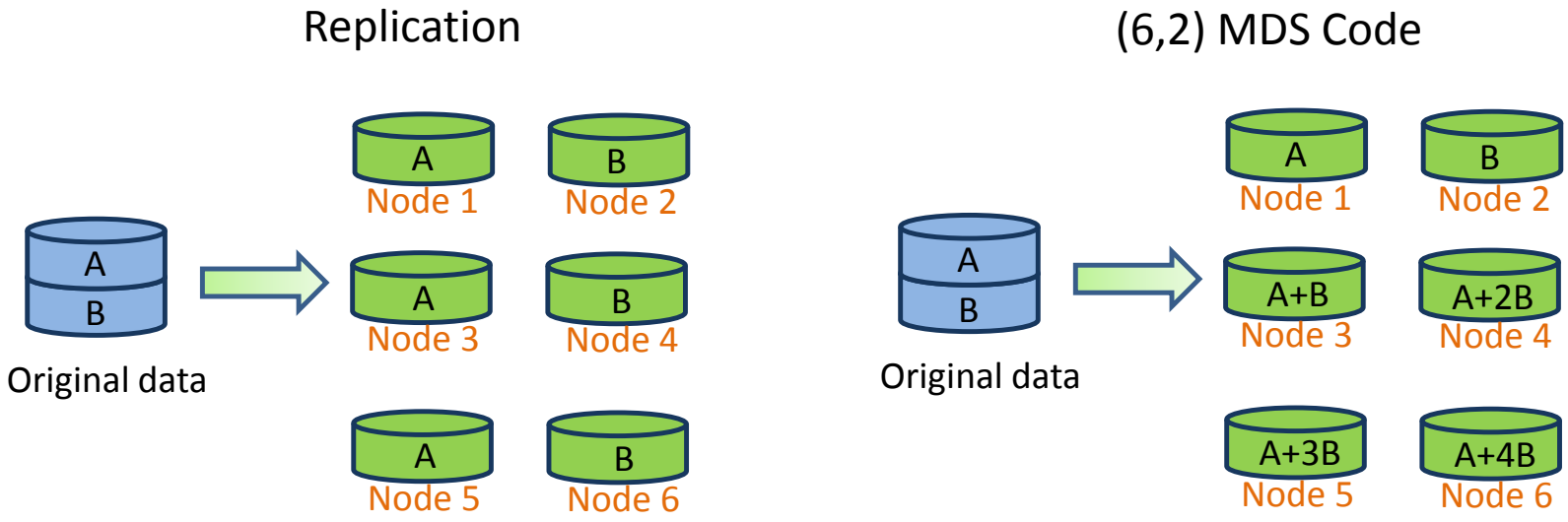
   (Add a new bound and extensions to q-ary LRCs)

Outline

Outline

Distributed Storage System: Replication vs. MDS

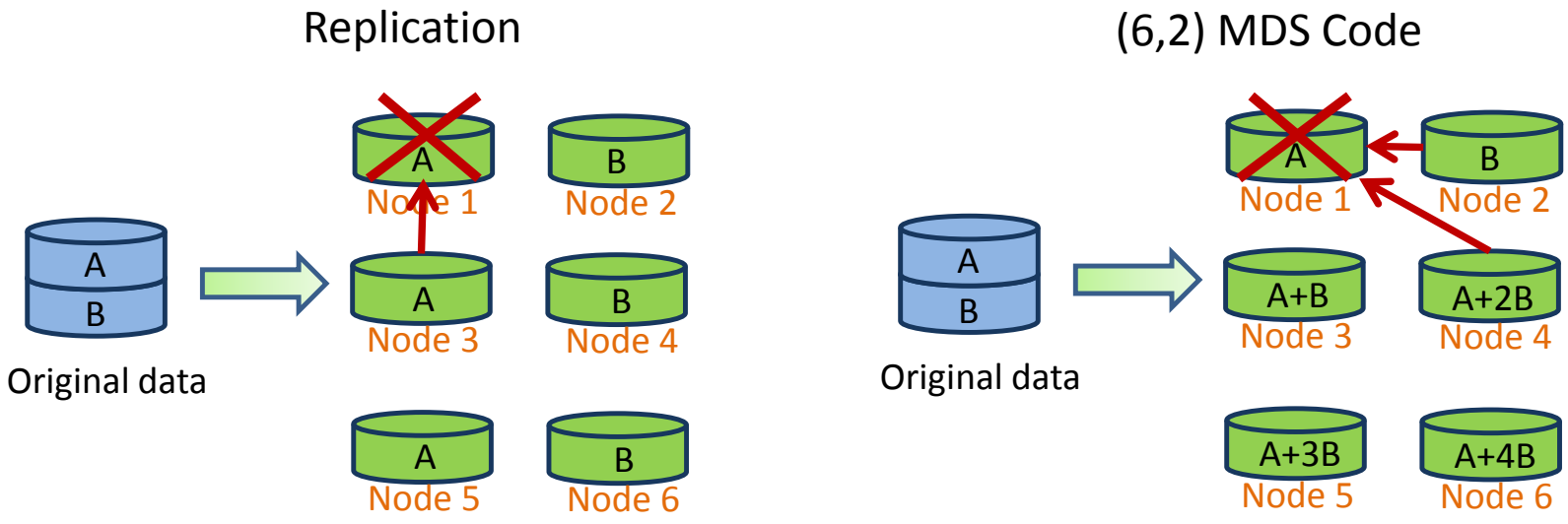
In an $n = 6, k = 2$ distributed storage system,



Storage per node	M/2	M/2
Erasure tolerance (data reliability)	2	4

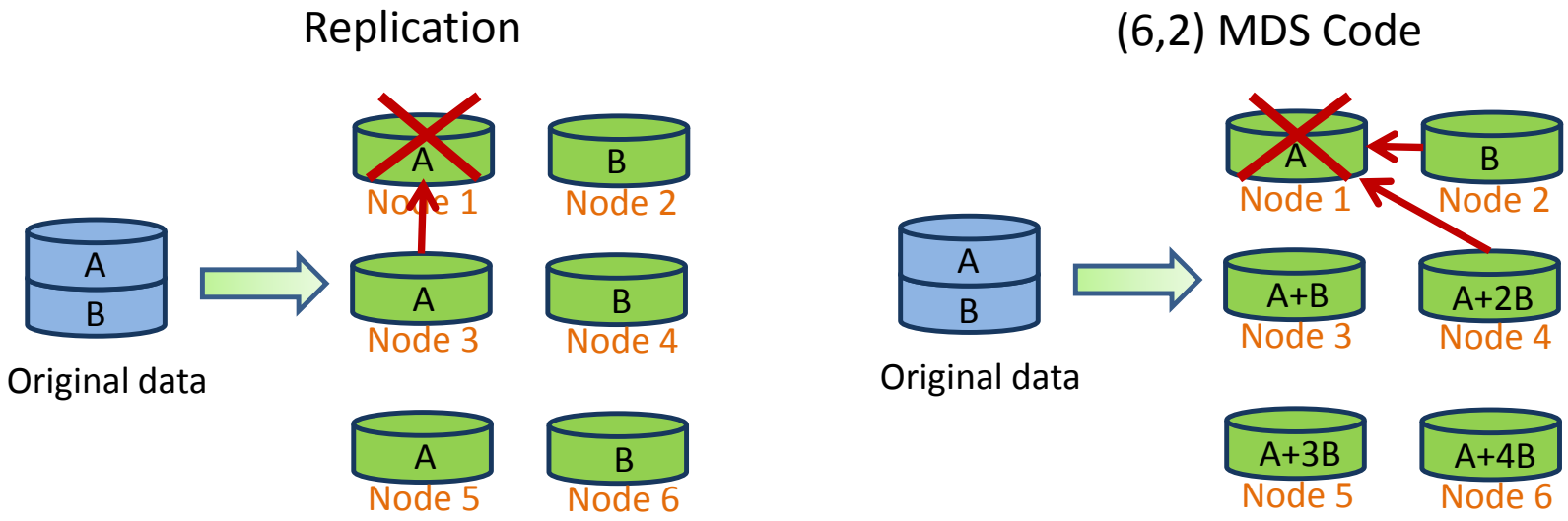
- MDS codes provide optimal data reliability for given storage overhead.

Efficiency of Node Repair



Repair bandwidth	M/2	M
Number of bits read	M/2	M
Number of nodes accessed	1	2

Efficiency of Node Repair



Repair bandwidth	M/2	M
Number of bits read	M/2	M
Number of nodes accessed	1	2

Metrics of efficiency:

- Repair bandwidth: *number of bits transferred* (regenerating codes)
- Disk-I/O: *number of bits read from memory* (optimal access repair)
- Repair locality: *number of nodes accessed (cloud storage)*

  $\longrightarrow$ Locally Repairable Codes

Locally Repairable Code

Let $\mathcal{C}$ be an $[n, k]_q$ linear code, generator matrix $G = (g_1, \cdots, g_n)$.

Definition

The ith coordinate of $\mathcal{C}$ is said to have repair locality r if g_i is an $\mathbb{F}_q$-linear combination of at most r other columns of G.

- Information locality:
  the k information coordinates have locality r.
- All symbol locality:
  all coordinates have locality r.

Remark

- Repair locality for nonlinear codes can be defined similarly.
- Suppose $d > 1$, it always has $r \leq k$.

Bound on the Parameters of LRC

Theorem [1]

For an $[n, k, d]$ linear code with information locality r, it has

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

Remark

- Codes with all symbol locality r also satisfy the above bound.
- The bound is tight for codes with information locality r. (pyramid code construction)
- When $(r+1) \nmid n$ and $r \mid k$, the inequality holds strictly for codes with all symbol locality.

---

1. P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, On the Locality of Codeword Symbols, IEEE Transactions on Information Theory, 2012.

The Bound Can be Attained When ...

- $(r+1) \mid n$, explicit constructions were given over
  - field of size exponential in n. (see [1-2])
  - field of size comparable to n. (see [3])
- $n \mod (r+1) > k \mod r > 0$,
  - field of size exponential in n. (see [1])
- $w \geq r+1-m$ and $r-v \geq u$
- $w+1 \geq 2(r+1-m)$ and $2(r-v) \geq u$
  - where $w = \lfloor \frac{n}{r+1} \rfloor, u = \lfloor \frac{k}{r} \rfloor$, $m = n \mod (r+1)$ and $v = k \mod r$. Constructions were given over field of size exponential in n. See [4].

---

1. N. Silberstein, A. S. Rawat, O. O. Koyluoglu, and S. Vishwanath, Optimal locally repairable codes via rank-metric codes, ISIT 2013.
2. I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, Optimal locally repairable codes and connections to matroid theory, ISIT 2013.
3. I. Tamo and A. Barg, A family of optimal locally recoverable codes, IEEE Trans. Inf. Theory, 2014.
4. W. Song, S. H. Dau, C. Yuen, and T. J. Li, Optimal locally repairable linear codes, IEEE J. Sel. Areas Commun., 2014.

The Bound Can Not be Attained When ...

- $(r+1) \nmid n$ and $r \mid k$, see [1]
- $m < v+1$ and $u \geq 2(r-v)+1$, see [2]

- For $(r+1) \nmid n$, the construction in [3] can be used to get an LRC with $d \geq n-k-\lceil \frac{k}{r} \rceil + 1$ which is at most one less than the upper bound.

---

1. P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, On the Locality of Codeword Symbols, IEEE Transactions on Information Theory, 2012.
2. W. Song, S. H. Dau, C. Yuen, and T. J. Li, Optimal locally repairable linear codes, IEEE J. Sel. Areas Commun., 2014.
3. I. Tamo and A. Barg, A family of optimal locally recoverable codes, IEEE Trans. Inf. Theory, 2014.

Some Improvement to the Bound

- In [1], an improved upper bound on the minimum distance was derived, i.e. $d \leq n - k + 1 - \ell$ where ℓ is computed from a recursively defined sequence.
- In [2], a further improved, explicit bound was obtained for any $[n, k, d]$ LRC with $n_1 > n_2$, where $n_1 = \left\lceil \frac{n}{r+1} \right\rceil$ and $n_2 = n_1(r+1) - n$.

---

1. N. Prakash, V. Lalitha, and P. V. Kumar, Codes with locality for two erasures, ISIT 2014.
2. Anyu Wang, Zhifang Zhang, An integer programming based bound for locally repairable codes. IEEE Transactions on Information Theory 61(10), 5280–5294, 2015.

Some Improvement to the Bound

- In [1], an improved upper bound on the minimum distance was derived, i.e. $d \leq n - k + 1 - \ell$ where ℓ is computed from a recursively defined sequence.
- In [2], a further improved, explicit bound was obtained for any $[n, k, d]$ LRC with $n_1 > n_2$, where $n_1 = \left\lceil \frac{n}{r+1} \right\rceil$ and $n_2 = n_1(r+1) - n$.

However, all bounds introduced so far are field-independent!

- even if the bound can be met with equality, it doesn't mean this can happen in any field!
- the field size has an effect on the code parameters.

---

1. N. Prakash, V. Lalitha, and P. V. Kumar, Codes with locality for two erasures, ISIT 2014.
2. Anyu Wang, Zhifang Zhang, An integer programming based bound for locally repairable codes. IEEE Transactions on Information Theory 61(10), 5280–5294, 2015.

Field-Dependent Bounds for LRC

- The C-M bound [1]: $k \leq \text{Min}_{t \in \mathbb{Z}^+} \left[tr + k_{\text{opt}}^{(q)}(n - (r+1)t, d) \right]$, where $k_{\text{opt}}^{(q)}(n, d)$ is the maximum dimension of a linear code of length n and distance d over $\mathbb{F}_q$ (only known from code tables or approximation).
- Combinatorial and LP bound for (r, δ)-LRCs [2]:
  - Combinatorial bound: degenerate to the Singleton-like bound when $\delta = 2$.
  - LP bound: need to solve a linear programming problem.
- Other bounds apply to very specifical cases.

---

1. V. Cadambe and A. Mazumdar. Bounds on the size of locally recoverable codes. IEEE transactions on information theory, 61(11):57875794, 2015.
2. A. Agarwal, A. Barg, S. Hu, A. Mazumdar and I. Tamo, "Combinatorial Alphabet-Dependent Bounds for Locally Recoverable Codes," in IEEE Transactions on Information Theory, vol. 64, no. 5, pp. 3481-3492, May 2018.

Our Goal

To derive new bounds for the parameters of LRC which are

- field-dependent;
- explicit;
- suitable for a wider range of parameters;
- tighter than the C-M bound or the LP bound.

Outline

The $\mathcal{L}$-Space of Linear LRCs

Let $\mathcal{C}$ be an $[n, k, d]$ linear LRC over $\mathbb{F}_q$ with locality r:

$\mathcal{L}$-cover: $\mathcal{H}$

A set of parity checks $\mathcal{H} \subseteq \mathcal{C}^{\perp}$ is called an $\mathcal{L}$-cover of $\mathcal{C}$ if

- Locality: $\mathrm{wt}(\mathbf{h}) \leq r + 1$ for all $\mathbf{h} \in \mathcal{H}$;
- Minimum Cover: $\bigcup_{\mathbf{h} \in \mathcal{H}} \mathrm{supp}(\mathbf{h}) = [n]$, and $\bigcup_{\mathbf{h} \in \mathcal{H}'} \mathrm{supp}(\mathbf{h}) \neq [n]$ for all $\mathcal{H}' \subsetneq \mathcal{H}$.

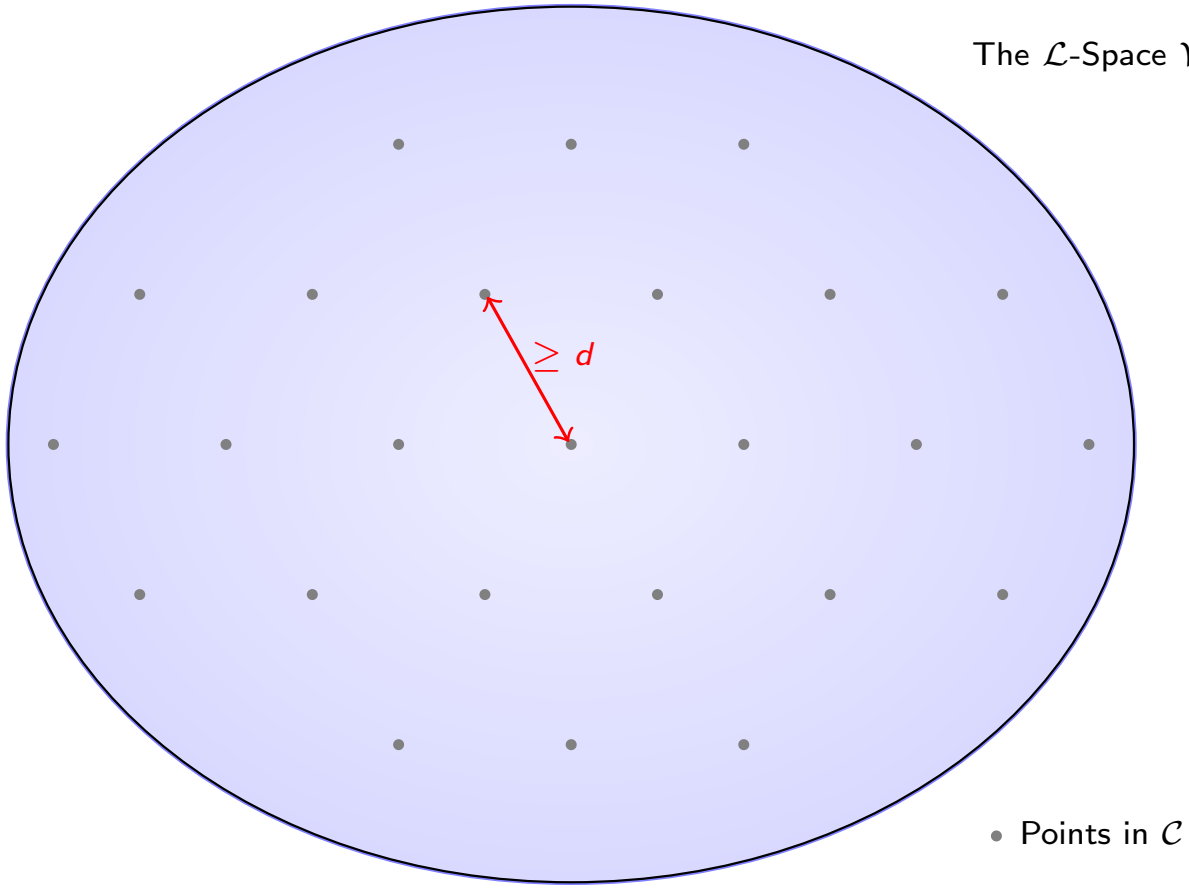
$\mathcal{L}$-space: $\mathcal{V}$

The dual space of an $\mathcal{L}$-cover is called an $\mathcal{L}$-space of $\mathcal{C}$, denoted by $\mathcal{V}$, i.e., $\mathcal{V} = \{\mathbf{v} \in \mathbb{F}_q^n \mid < \mathbf{v}, \mathbf{h} >= 0, \ \forall \mathbf{h} \in \mathcal{H}\}$.

- Actually, $\mathcal{V}$ is also an LRC containing $\mathcal{C}$ as a sub-code.

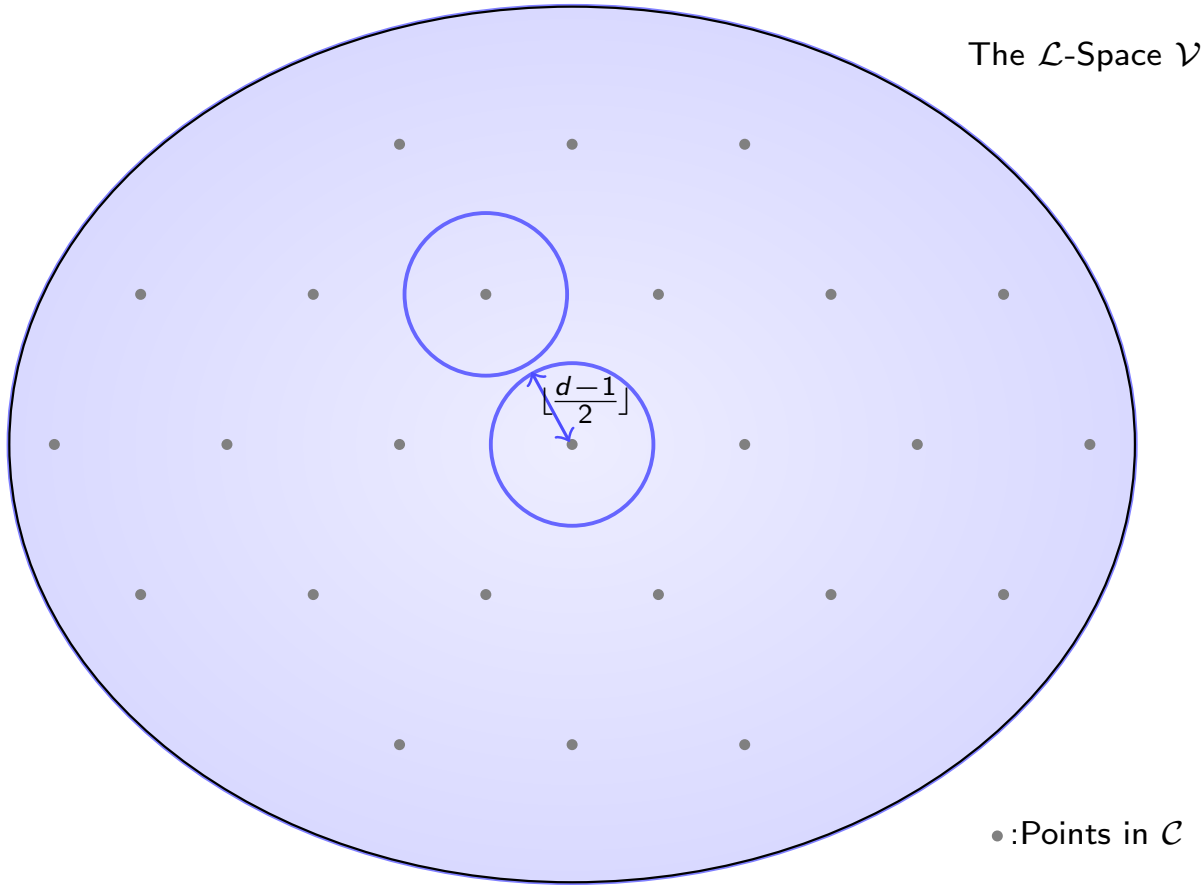
Sphere-Packing Problem in the $\mathcal{L}$-Space

- $\forall\ \mathbf{c} \in \mathcal{C}$, draw a sphere of radius $\lfloor \frac{d-1}{2} \rfloor$ in $\mathcal{V}$ centered at $\mathbf{c}$.
- Obviously, these spheres are pairwise disjoint.



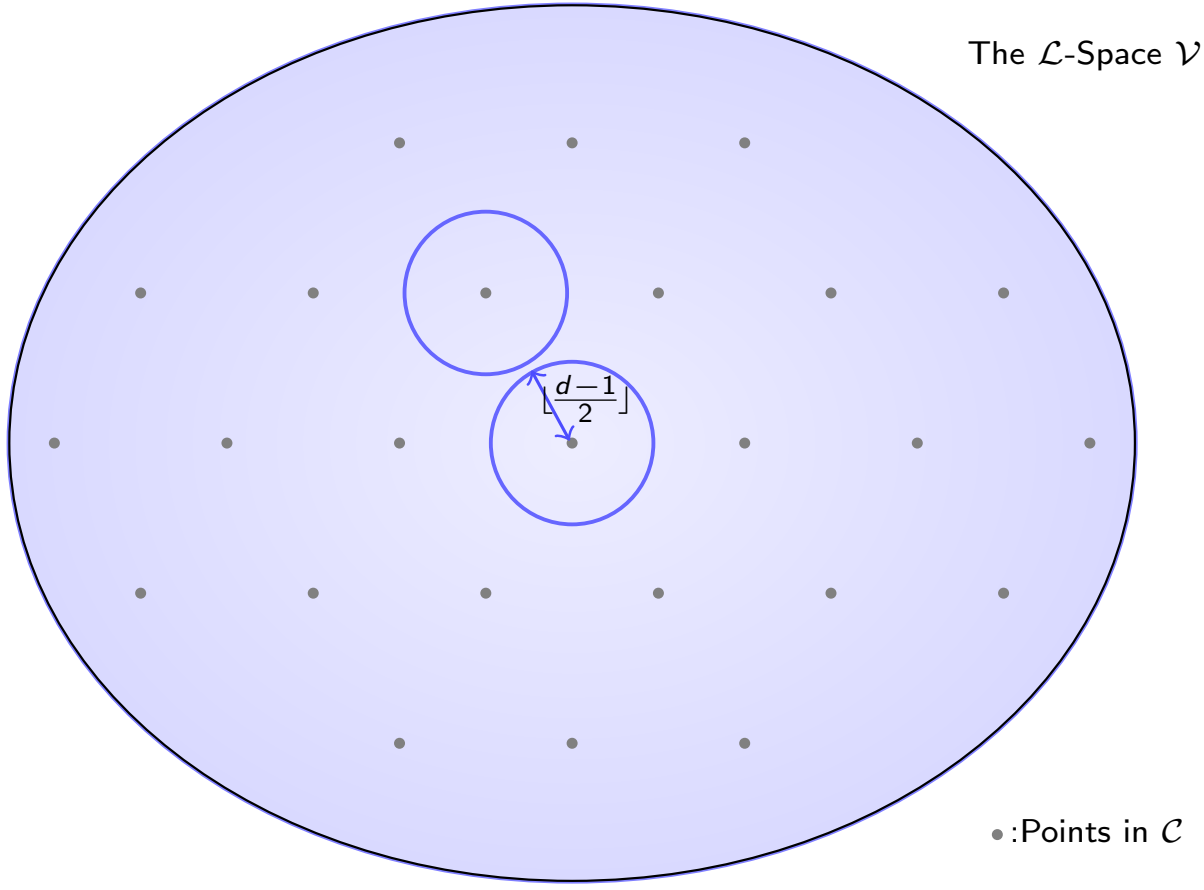
Sphere-Packing Problem in the $\mathcal{L}$-Space

- $\forall\ \mathbf{c} \in \mathcal{C}$, draw a sphere of radius $\lfloor \frac{d-1}{2} \rfloor$ in $\mathcal{V}$ centered at $\mathbf{c}$.
- Obviously, these spheres are pairwise disjoint.



Sphere-Packing Problem in the $\mathcal{L}$-Space

- Denote $B_{\mathcal{V}}(\lfloor \frac{d-1}{2} \rfloor) = \left|\{\mathbf{v} \in \mathcal{V} : \mathrm{wt}(\mathbf{v}) \leq \lfloor \frac{d-1}{2} \rfloor\}\right|$
- It has $|\mathcal{C}| \cdot B_{\mathcal{V}}(\lfloor \frac{d-1}{2} \rfloor) \leq |\mathcal{V}|$



A Connection Between k, d and the $\mathcal{L}$-Space $\mathcal{V}$

Observe that:

- $|\mathcal{C}| \cdot B_{\mathcal{V}}(\lfloor \frac{d-1}{2} \rfloor) \leq |\mathcal{V}|$;
- $\log_q |\mathcal{C}| = k$, $\log_q |\mathcal{V}| = \dim(\mathcal{V})$.

Lemma:

$$k \leq \dim(\mathcal{V}) - \log_q \big(B_{\mathcal{V}}(\lfloor \frac{d-1}{2} \rfloor) \big), \tag{1}$$

- Explicit bound can be derived from (1) if the weight distribution of $\mathcal{V}$ is known;
- Can be easily applied to LRCs with disjoint repair groups.

Binary LRCs with Disjoint Repair Groups

Disjoint Repair Groups

$\exists$ local parity checks $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_\ell \in \mathcal{C}^\perp$, where $\ell = \frac{n}{r+1}$, satisfying

- $\mathrm{supp}(\mathbf{h}_i) \cap \mathrm{supp}(\mathbf{h}_{i'}) = \emptyset$ for $1 \leq i \neq i' \leq \ell$.
- $\mathrm{wt}(\mathbf{h}_i) = r + 1$;

- $\mathcal{H} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_\ell\}$ is an $\mathcal{L}$-cover of $\mathcal{C}$;

- $\mathcal{V} = \mathrm{spn}_2(\mathcal{H})^\perp$, $\dim(\mathcal{V}) = \frac{rn}{r+1}$, $B_{\mathcal{V}}(\lfloor \frac{d-1}{2} \rfloor) = ?$

Binary LRCs with Disjoint Repair Groups

Disjoint Repair Groups

$\exists$ local parity checks $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_\ell \in \mathcal{C}^\perp$, where $\ell = \frac{n}{r+1}$, satisfying

- $\mathrm{supp}(\mathbf{h}_i) \cap \mathrm{supp}(\mathbf{h}_{i'}) = \emptyset$ for $1 \leq i \neq i' \leq \ell$.
- $\mathrm{wt}(\mathbf{h}_i) = r + 1$;

- $\mathcal{H} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_\ell\}$ is an $\mathcal{L}$-cover of $\mathcal{C}$;
  - $\mathrm{spn}_2(\mathcal{H})$ has weight enumerator polynomial $(x^{r+1} + y^{r+1})^\ell$
- $\mathcal{V} = \mathrm{spn}_2(\mathcal{H})^\perp$, $\dim(\mathcal{V}) = \frac{rn}{r+1}$, $B_{\mathcal{V}}(\lfloor \frac{d-1}{2} \rfloor) = ?$

Binary LRCs with Disjoint Repair Groups

Disjoint Repair Groups

$\exists$ local parity checks $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_\ell \in \mathcal{C}^\perp$, where $\ell = \frac{n}{r+1}$, satisfying

- $\mathrm{supp}(\mathbf{h}_i) \cap \mathrm{supp}(\mathbf{h}_{i'}) = \emptyset$ for $1 \leq i \neq i' \leq \ell$.
- $\mathrm{wt}(\mathbf{h}_i) = r + 1$;

- $\mathcal{H} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_\ell\}$ is an $\mathcal{L}$-cover of $\mathcal{C}$;
  - $\mathrm{spn}_2(\mathcal{H})$ has weight enumerator polynomial $(x^{r+1} + y^{r+1})^\ell$
- $\mathcal{V} = \mathrm{spn}_2(\mathcal{H})^\perp$, $\dim(\mathcal{V}) = \frac{rn}{r+1}$, $B_{\mathcal{V}}(\lfloor \frac{d-1}{2} \rfloor) = ?$
  - by **the MacWilliams' identity**:

$$W_{\mathcal{V}}(x, y) = \sum_{0 \leq u \leq \frac{n}{2}} A_u x^{n-2u} y^{2u},$$

where $A_u = \sum_{i_1 + \cdots + i_\ell = u} \prod_{j=1}^{\ell} \binom{r+1}{2i_j}$

Explicit Bound Can Be Derived

Since $B_{\mathcal{V}}(\lfloor \frac{d-1}{2} \rfloor) = A_0 + \cdots + A_{\lfloor \frac{d-1}{4} \rfloor} = \sum_{0 \le i_1 + \cdots + i_\ell \le \lfloor \frac{d-1}{4} \rfloor} \prod_{j=1}^{\ell} \binom{r+1}{2i_j}$,

Bound A:

$$k \le \frac{rn}{r+1} - \log_2 \Big(\sum_{0 \le i_1 + \cdots + i_\ell \le \lfloor \frac{d-1}{4} \rfloor} \prod_{j=1}^{\ell} \binom{r+1}{2i_j} \Big) . \tag{2}$$

Explicit Bound Can Be Derived

Since $B_{\mathcal{V}}(\lfloor \frac{d-1}{2} \rfloor) = A_0 + \cdots + A_{\lfloor \frac{d-1}{4} \rfloor} = \sum_{0 \le i_1 + \cdots + i_\ell \le \lfloor \frac{d-1}{4} \rfloor} \prod_{j=1}^{\ell} \binom{r+1}{2i_j}$,

Bound A:

$$k \le \frac{rn}{r+1} - \log_2 \Big(\sum_{0 \le i_1 + \cdots + i_\ell \le \lfloor \frac{d-1}{4} \rfloor} \prod_{j=1}^{\ell} \binom{r+1}{2i_j} \Big) . \tag{2}$$

- It covers the bounds derived in [Goparaju&Calderbank 2014], [Zeh&Yaakobi 2015] as special cases for specific forms of n, r.
- Extension to q-ary LRCs:

$$k \le \frac{rn}{r+1} - \log_q \Big(\sum_{0 \le i_1 + \cdots + i_\ell \le \lfloor \frac{d-1}{2} \rfloor} \prod_{j=1}^{\ell} \beta(r, i_j) \Big) ,$$

where $\beta(r,i) = \frac{1}{q}((q-1)^i + (-1)^i(q-1))\binom{r+1}{i}$.

Comparison with the C-M bound and LP bound

The following table lists the upper bounds on k computed respectively from Bound A, the C-M Bound and the LP Bound:

r	3	4	5	6	7	8	9	10
Bound A	4	7	9	12	14	17	19	22
The C-M bound	5	7	10	13	15	18	21	23
The LP bound	4	6	9	11	14	17	19	22

Table: A comparison for $3 \leq r \leq 10$, $\frac{n}{r+1} = 3, d = 5$

It can be seen that Bound A is:

- a little weaker than the LP bound but tighter than the C-M bound
- explicitly computable, while neither the C-M bound nor the LP bound is explicit.

Outline

Applying the Shortening Technique

For an LRC $\mathcal{C}$ with unknown structure of local repair groups:

- We want to apply the inequality

$$k \leq \dim(\mathcal{V}) - \log_2\left(B_{\mathcal{V}}\left(\left\lfloor \tfrac{d-1}{2} \right\rfloor\right)\right) \tag{2}$$

- $B_{\mathcal{V}}\left(\left\lfloor \frac{d-1}{2} \right\rfloor\right)$ can not be computed directly!

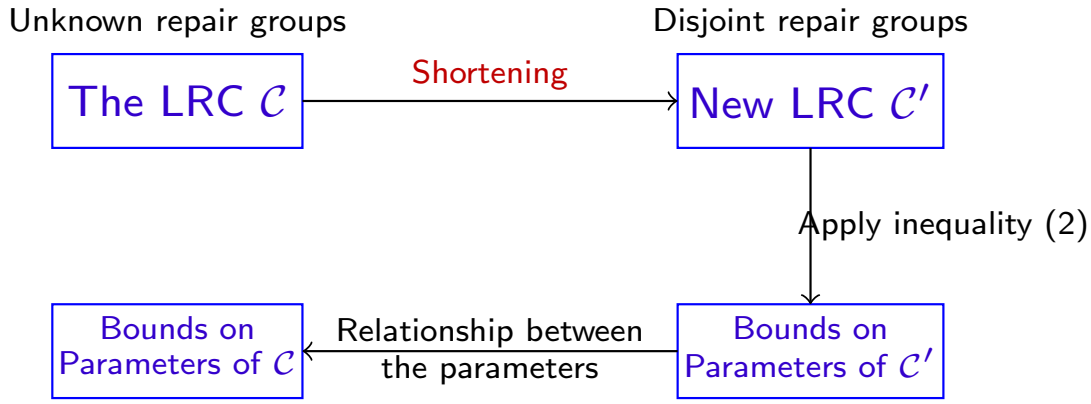
Applying the Shortening Technique

For an LRC $\mathcal{C}$ with unknown structure of local repair groups:

- We want to apply the inequality

$$k \leq \dim(\mathcal{V}) - \log_2\left(B_{\mathcal{V}}\left(\left\lfloor \frac{d-1}{2} \right\rfloor\right)\right) \tag{2}$$

- $B_{\mathcal{V}}\left(\left\lfloor \frac{d-1}{2} \right\rfloor\right)$ can not be computed directly!
- Our approach:



New Upper Bound

Theorem

For any $[n, k, d]$ binary linear LRC with locality r, it has

$$k \leq n - \min_{l, r_1, \ldots, r_\ell} \left[\ell + \log_2 \left(\Phi_\ell(r_1, \ldots, r_\ell) \right) \right], \tag{3}$$

where $\Phi_\ell(r_1, \ldots, r_\ell) = \sum_{0 \leq i_1 + \cdots + i_\ell \leq \lfloor \frac{d-1}{4} \rfloor} \prod_{j=1}^{\ell} \binom{r_j+1}{2i_j}$ and the 'Min' is subject to

$$\begin{cases} \frac{n}{r+1} \leq \ell \leq \frac{2n}{r+2}; \\ 0 \leq r_1, \ldots, r_\ell \leq r; \\ r_1 + \cdots + r_\ell = 2n - \ell(r+2). \end{cases}$$

- The bound (3) is based on solving an optimization problem.
- Solving the optimization problem is very difficult in general.

New Upper Bound

Theorem

For any $[n, k, d]$ binary linear LRC with locality r, it has

$$k \leq n - \min_{l, r_1, \ldots, r_\ell} \left[\ell + \log_2 \left(\Phi_\ell(r_1, \ldots, r_\ell) \right) \right], \tag{3}$$

where $\Phi_\ell(r_1, \ldots, r_\ell) = \sum_{0 \leq i_1 + \cdots + i_\ell \leq \lfloor \frac{d-1}{4} \rfloor} \prod_{j=1}^{\ell} \binom{r_j+1}{2i_j}$ and the 'Min' is subject to

$$\begin{cases} \frac{n}{r+1} \leq \ell \leq \frac{2n}{r+2}; \\ 0 \leq r_1, \ldots, r_\ell \leq r; \\ r_1 + \cdots + r_\ell = 2n - \ell(r+2). \end{cases}$$

- The bound (3) is based on solving an optimization problem.
- Solving the optimization problem is very difficult in general.
- However, for some special cases the bound (3) can be simplified.

Explicit Bounds Derived from (3)

Bound B: for $d \geq 5$

For any $[n, k, d]$ binary linear LRC with locality r such that $d \geq 5$ and $2 \leq r \leq \frac{n}{2} - 2$, it holds

$$k \leq \frac{rn}{r+1} - \min\{\log_2(1 + \frac{rn}{2}), \frac{rn}{(r+1)(r+2)}\}.$$

Bound C: for $r = 2$

For any $[n, k, d]$ binary linear LRC with locality $r = 2$, it has

$$k \leq n - \min_{\frac{n}{3} \leq \ell \leq \frac{n}{2}, \ell \in \mathbb{Z}} \left[\ell + \log_2(\mu(\ell))\right],$$

where

$$\mu(\ell) = \begin{cases} \sum_{0 \leq i_1 + i_2 \leq \lfloor \frac{d-1}{4} \rfloor} \binom{6\ell - 2n}{i_1} \binom{2n - 5\ell}{i_2} 3^{i_2}, & \text{if } \frac{n}{3} \leq \ell \leq \frac{2}{5}n; \\ \sum_{0 \leq i \leq \lfloor \frac{d-1}{4} \rfloor} \binom{2n - 4\ell}{i}, & \text{if } \frac{2}{5}n < \ell \leq \frac{n}{2}. \end{cases}$$

Comparisons

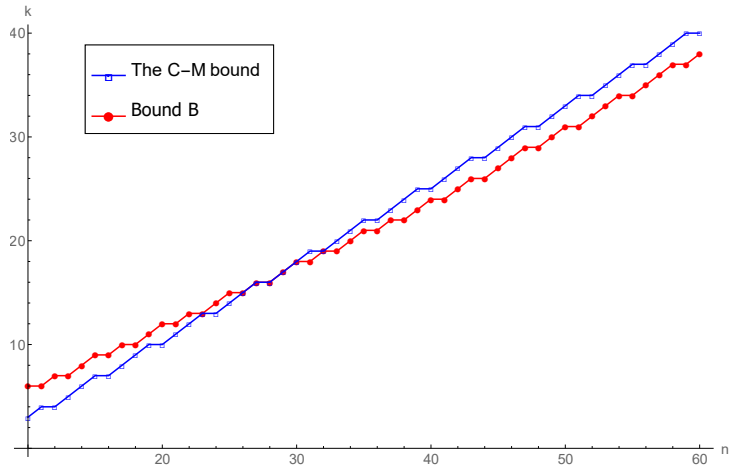


Figure: $r = 3,\ d = 5, 10 \leq n \leq 60$

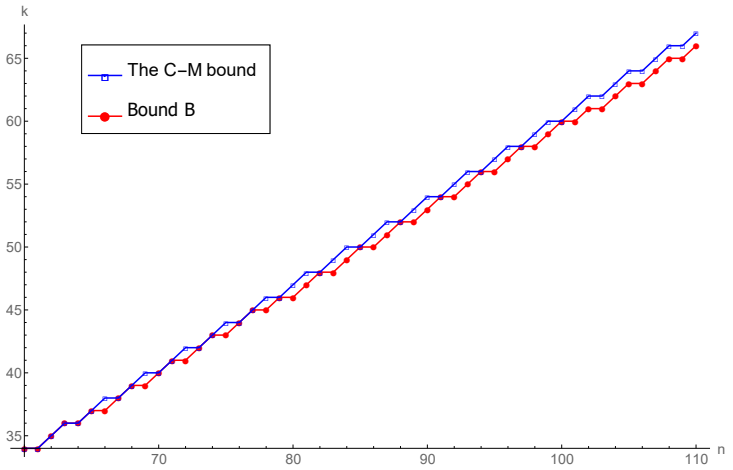


Figure: $r = 2,\ d = 8, 60 \leq n \leq 110$

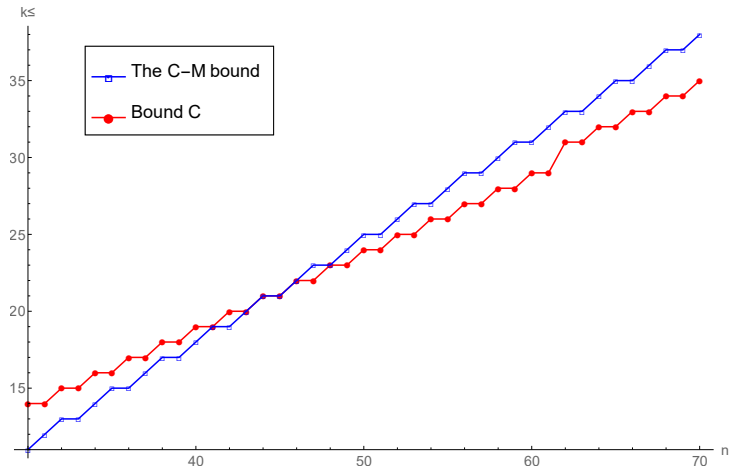


Figure: $r = 2,\ d = 9, 30 < n < 70$

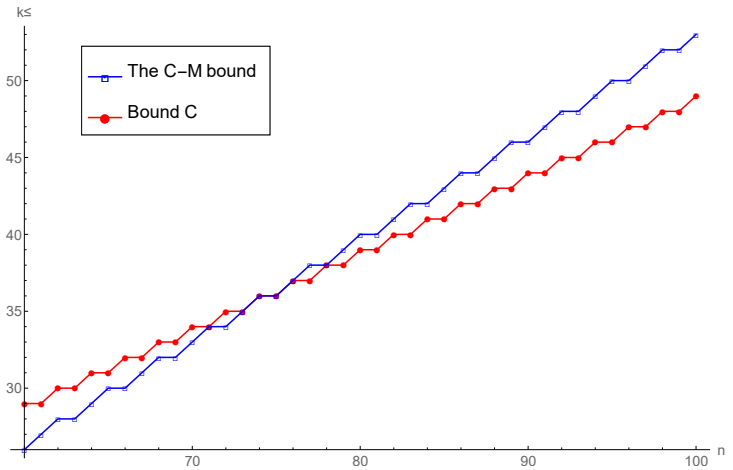


Figure: $r = 2,\ d = 8, 60 < n < 100$

Conclusions

- This work dedicates to establish new, field-dependent bounds for LRCs.
- By using the sphere-packing approach, we derive three explicit bounds which tend to outperform the C-M bound.
- Some constructions with specific parameters attaining our new bounds are presented in the paper.
- Is it possible to use sphere-packing more smartly?
  - Inner space: an LRC $\mathcal{C}$;
  - Outer space: the $\mathcal{L}$-space which is a larger LRC.
- Can we extend other field-dependent bounds in coding theory to LRCs?
  - Plotkin bound, Griesmer bound, ...
- Anyway, it is interesting and meaningful to study the influence of field size on the parameters of LRCs.

Thanks for your attention!