Bounds and Constructions for Linear Locally Repairable Codes over Binary Fields

Anyu Wang*, Zhifang Zhang†, and Dongdai Lin*

*State Key Laboratory of Information Security, Institute of Information Engineering, CAS, Beijing, China
†Key Laboratory of Mathematics Mechanization, NCMIS, Academy of Mathematics and Systems Science, CAS, Beijing, China
Emails: wanyu@iie.ac.cn, zfz@amss.ac.cn, ddlin@iie.ac.cn

Abstract—We present two new upper bounds on the dimension $k$ for binary linear locally repairable codes (LRCs). The first one is an explicit bound for binary linear LRCs with disjoint repair groups, which can be regarded as an extension of known explicit bounds for such LRCs to more general code parameters. The second one is based on an optimization problem, and can be applied to any binary linear LRC. By simplifying the second bound for $d \geq 5$, we obtain an explicit bound which can outperform the Cadambe-Mazumdar bound for $5 \leq d \leq 8$ and large $n$. Lastly, we give a new construction of binary linear LRCs with $d \geq 6$, and prove their optimality with respect to the simplified bound.

1. INTRODUCTION

Recently, locally repairable codes (LRCs) have attracted a lot of attention due to their applications in distributed storage systems. An $[n,k,d]$ linear code is said to be an LRC with locality $r$ if the value at each coordinate can be recovered by accessing at most $r$ other coordinates. Applying to a distributed storage system, LRC with small locality $r$ is preferred as it reduces the disk I/O complexity in repairing failed nodes. Besides, it is also desirable to have large value of dimension $k$ and minimum distance $d$ to ensure the storage efficiency and the overall fault tolerance level. Much work is devoted to explore the relationship between $k,d,r$. The first trade-off is derived in [4].

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

which is also known as the Singleton-like bound for LRCs. Then different methods are used to construct LRCs attaining [1], e.g., [11], [15], [19], [22]. Further study on the singleton-like bound is given in [6], [17], [23], improved upper bounds are also derived in [13], [24].

Considering the computational complexity, it is better to use LRCs over small finite fields in practice. A trade-off that takes into consideration the field size is derived by Cadambe and Mazumdar [2].

$$k \leq \min_{t \in \mathbb{Z}^+} \left[ tr + k_{\text{opt}}^{(q)}(n - (r + 1)t, d) \right],$$

where $k_{\text{opt}}^{(q)}(n, d)$ is the largest possible dimension of an $[n, k, d]$ linear code over $\mathbb{F}_q$. This trade-off is usually called the C-M bound, and is proved to be attained by the binary Simplex codes [2]. Later in [15], a class of binary LRCs with $r = 2, 3$ constructed via antcodes are also shown to satisfy [2] with equality. Though these codes all attain [2], their code length $n$ increase exponentially with respect to the dimension $k$, which means the information rate is close to 0. Then different methods are employed to construct LRCs with more satisfied parameters over small fields. In [5], binary LRCs with $r = 2$ and $d = 2, 6, 10$ are constructed from primitive cyclic codes. These codes do not attain the C-M bound [2], but are shown to be optimal under a structural assumption that the codeword coordinates are divided into disjoint repair groups. The same method is also adopted in [25] to give binary LRCs with $r = 2, d = 10$ based on nonprimitive cyclic codes. In [27], the authors construct a class of BCH-type binary LRCs, but it is unclear whether these codes can attain the C-M bound [2].

On the other hand, LRCs that attain the Singleton-like bound [1] over small field are also considered in [7], [14]. In a recent work [8], two new upper bounds involving the field size are derived for the $(r, \delta)$ LRC. Since $(r, \delta)$ LRCs is an extension of LRCs to deal with multiple erasures in local repair [12], these two bounds apply to LRCs as well by restricting $\delta = 2$. In this case, the first bound is equivalent to the Singleton-like bound [1]. The second bound relies on the same disjoint repair group structure assumption in [5]. Besides, bounds on the parameters of LRCs are also studied in [1], [20] from an asymptotic perspective.

A. Our Contribution

This paper concentrates on binary linear LRCs, and presents the following result. We introduce the concept of locality space to characterize the locality property for binary linear LRCs, and give a connection between $k, d$ and the locality space $V$. Based on this connection, two new upper bounds on the dimension $k$ are derived.

The first is an explicit bound for LRCs that has a disjoint repair group structure. Note that for $r = 2$ and special value of $n, d$, explicit bounds were derived in [5], [25] for LRCs with disjoint repair groups. Our bound can be viewed as an extension of that in [5], [25] to general $n, d$ and $r$.

The second bound depends on an optimization problem, and can be applied to any binary linear LRC. Though the optimization problem is hard to solve in general, it is possible to simplify the bound to give an explicit bound for $d \geq 5$. Through comparison, the simplified bound can outperform the C-M bound for $5 \leq d \leq 8$ and large $n$. Moreover, a class of binary linear LRCs attaining the simplified bound is also constructed.
B. Organization

Section II defines the locality space and derives an explicit bound for LRCs with disjoint repair groups. Section III presents the upper bound based on optimization problem and derives simplified bound for $d = 5$. Section IV gives the construction.

II. THE LOCALITY SPACE FOR LRCs

In this section, we define locality space for LRCs, and give a connection between $k$, $d$ and the locality space. Based on this connection, an explicit upper bound is derived for binary linear LRCs with disjoint repair groups. To begin with, we introduce some notations. For any vector $v = (v_1, \ldots, v_n) \in \mathbb{F}_2^n$, let $\text{Supp}(v) = \{i \in [n] : v_i \neq 0\}$ and $w_t(v) = |\text{Supp}(v)|$, where $[n] = \{1,2,\ldots,n\}$. For any two vectors $u, v \in \mathbb{F}_2^n$, $\text{dist}(u,v)$ denotes the hamming distance of $u$ and $v$. Lastly, let $\text{Span}_{\mathbb{F}_2}(u_1, \ldots, u_i)$ be the linear space spanned by a set of vectors $\{u_1, \ldots, u_i\}$ over $\mathbb{F}_2$.

Let $\mathcal{C}$ be an $[n,k,d]$ binary linear LRC with locality $r$. Then for each coordinate $i \in [n]$, there is a local parity check $h_i \in \mathcal{C}^\perp$ such that $i \in \text{Supp}(h_i)$ and $\text{wt}(h_i) \leq r+1$. The locality space for LRCs is defined as follows.

Definition 1. Suppose $\mathcal{L} \subseteq \{h_1, \ldots, h_n\}$ is a set of local parity checks such that $\cup_{h_i \in \mathcal{L}} \text{Supp}(h_i) = [n]$, and $\cup_{h_i \in \mathcal{L}} \text{Supp}(h_i) \neq [n]$ for any proper subset $\mathcal{L}' \subseteq \mathcal{L}$. Then $\mathcal{L}$ is called a minimum set of local parity checks covering $[n]$, and the linear space $V = (\text{Span}_{\mathbb{F}_2}(\mathcal{L}))^\perp$ is said to be a locality space of $\mathcal{C}$.

Clearly $\mathcal{L}$ can be obtained by dropping some parity checks in $h_1, \ldots, h_n$. Then the existence of $\mathcal{L}$ and $V$ is ensured for any binary linear LRC $\mathcal{C}$. In fact, $\mathcal{L}$ is minimal to guarantee the repair locality $r$ for each of the $n$ coordinates. Since $\mathcal{L} \subseteq \mathcal{C}^\perp$, then it has $\mathcal{C} \subseteq V$. Moreover, the concept of locality space plays an important role in characterizing the locality property of $\mathcal{C}$. Note that $V$ is defined by a minimum set of local parity checks, which guarantees the locality $r$ for all coordinates. So provided a locality space $V$ of $\mathcal{C}$ is known, $\mathcal{C}$ can be just regarded as a subspace of $V$ that has dimension $k$ and minimum distance $d$. Based on this fact, we apply the sphere-packing bound into the locality space $V$, and obtain a connection between $k$, $d$ and $V$.

Proposition 2. For an $[n,k,d]$ binary LRC $\mathcal{C}$ with locality $r$, it holds

$$k \leq \dim(V) - \log_2(B_V\left(\left\lfloor \frac{d-1}{2} \right\rfloor \right)), \quad (3)$$

where $V$ is a locality space of $\mathcal{C}$, and $B_V\left(\left\lfloor \frac{d-1}{2} \right\rfloor \right) = \left\{v \in V : \text{wt}(v) \leq \left\lfloor \frac{d-1}{2} \right\rfloor \right\}$.

Proof: For any codeword $c \in \mathcal{C}$, consider the ball of radius $\left\lfloor \frac{d-1}{2} \right\rfloor$ around $c$ in $V$. Since $\mathcal{C}$ has minimum distance $d$, then these balls are non-overlapping. It follows that

$$\sum_{c \in \mathcal{C}} B_V(c, \left\lfloor \frac{d-1}{2} \right\rfloor) \leq |V|, \quad (4)$$

where $B_V(c, \left\lfloor \frac{d-1}{2} \right\rfloor) = \left\{v \in V : \text{dist}(v, c) \leq \left\lfloor \frac{d-1}{2} \right\rfloor \right\}$. Note that $\mathcal{C} \subseteq V$, so we have $B_V(c, \left\lfloor \frac{d-1}{2} \right\rfloor) = B_V\left(\left\lfloor \frac{d-1}{2} \right\rfloor \right)\forall c \in \mathcal{C}$. Therefore $(4)$ can be written as

$$|\mathcal{C}| \cdot B_V\left(\left\lfloor \frac{d-1}{2} \right\rfloor \right) \leq |V|.$$  

(5)

Since $\log_2 |\mathcal{C}| = k$ and $\log_2 |V| = \dim(V)$, then the Theorem follows directly from $(5)$.

The right hand side of $(3)$ depends on the locality space $V$. So explicit bound can be derived from $(3)$ if $V$ is known. In the following, we apply Proposition 2 to a special class of binary linear LRCs which have a simple locality space $V$.

A. Explicit Bound for LRCs with Disjoint Repair groups

Let $\mathcal{C}$ be an $[n,k,d]$ binary linear LRC with locality $r$. Moreover, we suppose there exists a set of local parity checks $h_{i_1}, h_{i_2}, \ldots, h_{i_l} \in \mathcal{C}^\perp$ such that $\cup_{i \in [r]} \text{Supp}(h_i) = [n], \text{wt}(h_i) = r+1$ and $\text{Supp}(h_{i_j}) \cap \text{Supp}(h_{i_l}) = \emptyset$ for $1 \leq i \neq j \leq l$. Such LRC $\mathcal{C}$ is usually said to have a disjoint repair group struture, which is widely adopted in the construction of LRCs, e.g., [11], [15], [19], [22]. Based on Proposition 2 the following upper bound can be derived.

Corollary 3. For any $[n,k,d]$ binary LRC with locality $r$ that has a disjoint repair group structure, it holds

$$k \leq \frac{rn}{r+1} - \log_2\left(\sum_{0 \leq i_1 + \cdots + i_l \leq \left\lfloor \frac{d+1}{2} \right\rfloor} \prod_{j=1}^{l} \left(\frac{r+1}{2i_j}\right)\right), \quad (6)$$

where $l = \frac{n}{r+1}$.

Proof: Denote $\mathcal{L} = \{h_{i_1}, h_{i_2}, \ldots, h_{i_l}\}$. Clearly $\mathcal{L}$ is a minimum set of local parity checks covering $[n]$. Let $V = (\text{Span}_{\mathbb{F}_2}(\mathcal{L}))^\perp$, then it has $\dim(V) = n - l = \frac{rn}{r+1}$. For $B_V\left(\left\lfloor \frac{d-1}{2} \right\rfloor \right)$, note that the linear space $\text{Span}_{\mathbb{F}_2}(\mathcal{L})$ has weight enumerator polynomial $W_V(x,y) = (x^{r+1} + y^{r+1})^l$. Then by the MacWilliams equality, (see e.g., [9]), the weight enumerator polynomial of $V$ is

$$W_V(x,y) = \frac{1}{2^n} W_L(x+y,x-y) = \sum_{0 \leq u \leq \frac{n}{2}} A_u x^{n-2u} y^{2u},$$

where $A_u = \sum_{i_1 + \cdots + i_l = u} \prod_{j=1}^{l} \left(\frac{r+1}{2i_j}\right)$. Thus we have

$$B_V\left(\left\lfloor \frac{d-1}{2} \right\rfloor \right) = A_0 + \cdots + A_{\left\lfloor \frac{d+1}{2} \right\rfloor} = \sum_{0 \leq i_1 + \cdots + i_l \leq \left\lfloor \frac{d+1}{2} \right\rfloor} \prod_{j=1}^{l} \left(\frac{r+1}{2i_j}\right),$$

and $(6)$ follows directly.
The sphere packing approach was also used in [5], [25] to derive upper bounds on \( k \) for binary linear LRCs with disjoint repair groups. However, their approach only applies to the case \( r = 2 \) since it relies on a map from binary linear LRCs with \( r = 2 \) to additive \( \mathbb{F}_2 \)-codes. From this point of view, our approach can be viewed as an extension of that in [5], [25] to general \( r \). In fact, the upper bounds for binary linear LRCs with disjoint repair groups in [5], [25] can be obtained by just substituting specific value of \( n, d, r \) into Corollary 4. For example, suppose \( n = 2^m - 1, d = 6 \) and \( r = 2 \), then Corollary 4 implies that

\[
k \leq \frac{2n}{3} - \log_2 \left( {r + 1 \choose 0} + {r + 1 \choose 1} \right)
\]

\[
= \frac{2n}{3} - \log_2 (1 + n)
\]

\[
= \frac{2}{3} (2^m - 1) - m,
\]

which coincides with the Theorem 1 in [5].

### III. New Upper Bound for Binary Linear LRCs

In this section, we give an upper bound for binary linear LRCs without restricting the structure of locality space. Suppose \( C \) is an \([n, k, d] \) binary linear LRC with locality \( r \). Let \( \mathcal{L} = \{h_1, \ldots, h_l\} \subseteq C^\perp \) be a minimum set of local parity checks covering \([n]\), and define \( L \in \mathbb{F}_2^{l \times n} \) to be a matrix whose rows are the \( l \) parity checks in \( \mathcal{L} \). Denote \( N \) to be the number of columns in \( L \) that have weight 1. Let \( C' \) be a shortened code consisting of all codewords of \( C \) that have zeros at the other \( n - N \) coordinates, with these zeros deleted. Then \( C' \) satisfies the following properties.

**Lemma 4.** The shortened code \( C' \) is an \([N, K, D] \) binary linear LRC satisfying

(i) \( n \geq N \geq 2n - l(r + 1), K \geq N - (n - k), D \geq d; \)

(ii) \( C' \) has a minimum set of local parity checks \( h'_1, \ldots, h'_l \) covering \([N]\) such that \( 1 \leq w(h'_i) \leq r + 1 \) and \( \text{Supp}(h'_i) \cap \text{Supp}(h'_j) = \emptyset \) for all \( 1 \leq i \neq j \leq n \).

**Proof:** Since the shortening operation neither increases the redundancy nor decreases the minimum distance, (see e.g., [9]), then it has \( K \geq N - (n - k) \) and \( D \geq d \). To show the other statements, we suppose without loss of generality that

\[
L = (L', L''),
\]

where \( L' \in \mathbb{F}_2^{l \times N} \) consists of the \( N \) columns that have weight 1, and \( L'' \in \mathbb{F}_2^{l \times (n - N)} \) consists of the other \((n - N)\) columns that have weight \( \geq 2 \). By counting the number of 1’s in \( L \), we have

\[
l(r + 1) \geq \text{the number of 1’s in } L
\]

\[
\geq N + 2(n - N).
\]

Thus \( 2n - l(r + 1) \leq N \leq n \). Lastly, denote \( h'_1, \ldots, h'_l \) to be the rows of \( L' \) then clearly \( \{h'_1, \ldots, h'_l\} \) is a set of parity checks of \( C' \) such that \( 1 \leq w(h'_i) \leq r + 1 \). Note that each column of \( L' \) has exactly one 1, therefore \( \cup_{i=1}^{l} \text{Supp}(h'_i) = [N] \) and \( \text{Supp}(h'_i) \cap \text{Supp}(h'_j) = \emptyset \) for all \( 1 \leq i \neq j \leq n \), which completes the proof.

Let \( V' = (\text{Span}_{\mathbb{F}_2} [h'_1, \ldots, h'_l])^\perp \) be a locality space of \( C' \), and denote \( w(h'_i) = r_i + 1 \) for \( i \in [l] \). Then it has \( \dim(V') = N - l \) and

\[
B_{V'}(\frac{D - 1}{2}) = \sum_{0 \leq i_1 + \cdots + i_l \leq \frac{D - 1}{2}} \prod_{j=1}^{l} \binom{r_j + 1}{2i_j}
\]

by a deduction similar to that in Corollary 3. Applying Proposition 2 to the shortened LRC \( C' \), we get

\[
K \leq (N - l) - \log_2 \left( \sum_{0 \leq i_1 + \cdots + i_l \leq \frac{D - 1}{2}} \prod_{j=1}^{l} \binom{r_j + 1}{2i_j} \right).
\]

Thus the following upper bound can be proved by substituting the statement (i) of Lemma 4 into (7) then optimizing with respect to \( l, r_1, \ldots, r_l \).

**Theorem 5.** For any \([n, k, d] \) binary linear LRC with locality \( r \), it holds

\[
k \leq n - \min_{l, r_1, \ldots, r_l} \left[ l + \log_2 (\Phi_l(r_1, \ldots, r_l)) \right],
\]

where

\[
\Phi_l(r_1, \ldots, r_l) = \sum_{0 \leq i_1 + \cdots + i_l \leq \frac{D - 1}{2}} \prod_{j=1}^{l} \binom{r_j + 1}{2i_j}
\]

and the ‘Min’ is taken over all integers \( l, r_1, \ldots, r_l \) such that

\[
\begin{cases}
\frac{n}{r + 1} \leq l \leq \frac{2n}{r + 2}, \\
0 \leq r_1, \ldots, r_l \leq r; \\
r_1 + \cdots + r_l = 2n - l(r + 2).
\end{cases}
\]

**Proof:** The proof can be found in Appendix A.

For any given \( n, d, r \), Theorem 5 gives an upper bound on the dimension \( k \) by solving an optimization problem. However, solving the optimization problem is very hard in general since the objective function in (8) is nonlinear. Nevertheless, it is still possible to simplify the bound (8) for some special cases. In the next subsection, we derive a simple upper bound for \( d \geq 5 \) from Theorem 5.

#### A. Simplified Bound for \( d \geq 5 \)

**Theorem 6.** 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

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

**Proof:** When \( d \geq 5 \), it has

\[
\Phi_l(r_1, \ldots, r_l) \geq \sum_{0 \leq i_1 + \cdots + i_l \leq \frac{D - 1}{2}} \prod_{j=1}^{l} \binom{r_j + 1}{2i_j}
\]

\[
= 1 + \frac{r_1 + 1}{2} + \cdots + \frac{r_l + 1}{2}.
\]

\( k \) is actually upper bounded by the largest integer no more than the right hand side of the inequality.
Then the optimization problem in (8) can be simplified accordingly, and the theorem follows by solving the simplified problem. The detailed proof can be found in Appendix B.

Remark. Since \( \log_2(1 + \frac{2^s}{2^n}) \) is logarithmic with respect to \( n \) while \( \frac{2^s}{2^n}(r + 1)(r + 2) \) is linear, it has \( \log_2(1 + \frac{2^s}{2^n}) < \frac{2^s}{2^n}(r + 1)(r + 2) \) when \( n \) is large enough. In fact, the inequality holds whenever \( n \geq 5(r + 1)(r + 2) \), and in such case (10) can be further simplified to be \( k \leq \frac{2^s}{2^n} \).

Next, we give a comparison between the bound (10) and the C-M bound (2), where the upper bound on \( \log_2(n, d) \) is computed by using SageMath [18] and the web database [10].

For \( 5 \leq d \leq 8 \), according to our computation, the bound (10) can outperform the C-M bound for large \( n \). Fig. 1 and Fig. 2 illustrate comparisons of the two bounds for \( r = 3, d = 5, 10 \leq n \leq 60 \) and \( r = 2, d = 8, 60 \leq n \leq 110 \) respectively.

Moreover, by a detailed calculation of the two bounds for all \( 2 \leq r \leq 6 \) and \( 2 \leq n \leq 200 \), it turns out that except the case \( d = 8, r \in \{5, 6\} \), the bound (10) always outperforms the C-M bound when \( n \) is large enough. Table 1 lists the conditions under which the bound (10) is tighter than the C-M bound. In the next section, we can construct a class of binary linear LRCs for \( d \geq 6 \) which attain the bound (10).

For \( d \geq 9 \), it can be checked that the bound (10) is inferior to the C-M bound. In fact, in this case we have \( \frac{2^s}{2^n} \geq 2 \), then a better estimate of \( \Psi_2(r_1, \ldots, r_l) \) can be used instead of that in the proof of Theorem 6, and therefore an upper bound tighter than (10) could be derived. However, we do not get an explicit bound in this case since the corresponding simplified optimization problem is still very complicated to solve.

IV. CONSTRUCTION ATTAINING THE UPPER BOUND

In this section, we give a new construction of binary linear LRC. The code has minimum distance \( d \geq 6 \), and attains the upper bound (10) in Theorem 6.

The construction relies on two matrices \( A \) and \( B \) defined as follows. Suppose \( s, t \) are two positive integers such that \( 2t \mid s \) and \( \frac{2^s}{2^n} \geq 2 \). Let \( A \) be a binary matrix of size \( 2t \times 2^t \) such that any no more than \( 4 \) columns are linearly independent. For \( t \leq 2 \), \( A \) can be just chosen as the identity matrix. For \( t \geq 3 \), \( A \) is the parity check matrix of \( [2^t, 2^t - 2t, 5]_2 \) binary double-error-correcting code. Define \( B \) to be a matrix whose columns are all nonzero \( \frac{s}{2^n} \)-tuples from \( \mathbb{F}_{2^s} \) with first nonzero entry equal to 1. Then \( B \) is actually the parity check matrix of a \( 2^{2t} \)-ary Hamming code, and the size of \( B \) is \( \frac{s}{2^n} \times \frac{2^s}{2^n - 1} \).

By fixing a basis of \( \mathbb{F}_{2^{2t}} \) over \( \mathbb{F}_2 \), each vector in \( \mathbb{F}_{2^{2t}} \) can be written as an element in \( \mathbb{F}_{2^t} \) and vice versa. We denote \( a_1, \ldots, a_{2^t} \in \mathbb{F}_{2^{2t}} \) to be the \( 2^t \) elements corresponding to the columns of \( A \), and denote vector \( \beta_i \) to be the \( i \)-th column of \( B \) for \( 1 \leq i \leq 2^t - 1 \), where the entries of \( \beta_i \) are in \( \mathbb{F}_{2^{2t}} \). Then the binary linear LRC is constructed as follows.

Construction 1. Suppose \( 2t \mid s \) and \( \frac{2^s}{2^n} \geq 2 \). Define \( C \) to be a binary linear code with parity check matrix

\[
H = \begin{pmatrix}
L_1 & L_2 & \ldots & L_{l_1} \\
H_1 & H_2 & \ldots & H_{l_1}
\end{pmatrix},
\]

where \( l = \frac{2^s}{2^n} - 1 \). For \( 1 \leq i \leq l \), \( L_i \) is an \( l \times (2^t + 1) \) matrix whose \( i \)-th row is the all-one vector and the other rows are zero. \( H_i \) is an \( s \times (2^t + 1) \) matrix whose columns are binary expansions of the vectors \( \{0, a_1\beta_1, a_2\beta_1, \ldots, a_{2^t}\beta_1\} \).

Example 1. Suppose \( s = 4 \) and \( t = 1 \). Then we can choose

\[
A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{F}_2^{2 \times 2}, \quad B = \begin{pmatrix} 1 & 1 & 1 & 0 \\ \omega^2 & \omega & 1 & 0 \end{pmatrix} \in \mathbb{F}_4^{2 \times 5},
\]
where $\omega$ is a primitive element in $\mathbb{F}_4$ such that $\omega^2 + \omega + 1 = 0$. Fixing a basis $\{1, \omega\}$, the two columns of $A$ can be written as two elements in $\mathbb{F}_4^2$, i.e., $a_1 = (1, \omega) \cdot (1) = 1, a_2 = (1, \omega) \cdot (1) = \omega$. Note that $\beta_1 = (\frac{1}{\omega})$, then
$$a_1 \beta_1 = \left(\frac{1}{\omega^2}\right), \quad a_2 \beta_1 = \left(\frac{\omega}{1}\right).$$

By expanding $\{0, a_1 \beta_1, a_2 \beta_1\} \subseteq \mathbb{F}_4^2$ into binary vectors with respect to the basis $\{1, \omega\}$, we get
$$H_1 = \begin{pmatrix}
 111 & 111 & 111 & 111 \\
 010 & 010 & 010 & 010 & 000 \\
 001 & 001 & 001 & 001 & 000 \\
 011 & 011 & 010 & 000 & 010 \\
 011 & 011 & 001 & 000 & 001
\end{pmatrix}. $$
The other $H_i$ can be computed similarly, so it has
$$H = \begin{pmatrix}
 111 & 111 \\
 010 & 010 & 010 & 010 & 000 \\
 001 & 001 & 001 & 001 & 000 \\
 011 & 011 & 010 & 000 & 010 \\
 011 & 011 & 001 & 000 & 001
\end{pmatrix}.$$ 

It can be checked that any no more than 5 columns of $H$ are linearly independent. So $H$ defines an $[n = 15, k = 6, d \geq 6]$ binary LRC with locality $r = 2$. Moreover, substituting $n = 15, d = 6, r = 2$ into the C-M bound $[2]$ yields $k \leq 6$, so this binary linear LRC is optimal with respect to the C-M bound.

**Theorem 7.** The code $C$ obtained from Construction 2 is an binary linear LRC with $n = \frac{2^t-1}{2^{t-1}} - s, d \geq 6$ and $r = 2^t$. Moreover, $C$ attains the upper bound $[10]$ for all positive integers $s, t$ satisfying $2t | s$ and $s \geq 2t$ except the case $s = 4, t = 1$.

**Proof:** Since $n, k, r$ can be verified directly, we focus on proving $d \geq 6$. Note that the sum of the first $t$ rows of $H$ is an all-one vector, so the minimum distance of $C$ must be even. Therefore it suffices to show that $d \geq 5$. Suppose to the contrary that there exists a codeword $c \in C$ such that $Hc^T = 0$ and $1 \leq wt(c) \leq 4$. Denote $c = (c_1, \ldots, c_n)$, where $c_i \in \mathbb{F}_2$ for $i \in [t]$. It can be deduced from the definition of $L_i$ that $wt(c_i)$ is even, $\forall i \in [t]$. So there are at most two nonzero vectors in $c_1, \ldots, c_t$. Without loss of generality, we suppose $c_1 = \cdots = c_t = 0$ and $1 \leq wt(c_{t+1}, c_{t+2}) \leq 4$. Then by $Hc^T = 0$ it has $H_1c_1^T + H_2c_2^T = 0$. Denote $c_1 = (x_0, x_1, \ldots, x_{2^t})$ and $c_2 = (y_0, y_1, \ldots, y_{2^t})$, we have
$$(x_1a_1 + \cdots + x_{2^t}a_{2^t})\beta_1 + (y_1a_1 + \cdots + y_{2^t}a_{2^t})\beta_2 = 0.$$ 

Since $\beta_1$ and $\beta_2$ are linearly independent over $\mathbb{F}_{2^{2t}}$, it must have
$$x_1a_1 + \cdots + x_{2^t}a_{2^t} + y_1a_1 + \cdots + y_{2^t}a_{2^t} = 0,$$
which contradicts to the fact that any no more than 4 out of $a_1, \ldots, a_{2^t} \in \mathbb{F}_{2^{2t}}$ are linearly independent over $\mathbb{F}_2$.

It remains to show $C$ is optimal with respect to $[10]$. Setting $n = \frac{2^t-1}{2^{t-1}}$ and $r = 2^t$, it can be deduced from Theorem 6 that $k \leq \frac{2^t-1}{2^{t-1}} - s$ for all $s, t$ satisfying $2t | s$, $\frac{s}{2^t} \geq 2$ except $s = 4, t = 1$, where the detailed calculation is in Appendix 11. Therefore $C$ attains the upper bound $[10]$.

**References**

1. A. Agarwal and A. Mazumdar. Bounds on the rate of linear locally repairable codes over small alphabets. *arXiv preprint arXiv:1607.08547*, 2016.
2. V. Cadambe and A. Mazumdar. Bounds on the size of locally recoverable codes. *IEEE transactions on information theory*, 61(11):5787–5794, 2015.
3. C. L. Chen. Construction of some binary linear codes of minimum distance five. *IEEE transactions on information theory*, 37(5):1429–1432, 1991.
4. G. Gopalan, C. Huang, H. Simutic, and S. Yekhanin. On the locality of codeword symbols. *IEEE transactions on Information Theory*, 58:6925–6934, 2012.
5. S. Goparaju and R. Calderbank. Binary cyclic codes that are locally repairable. In *IEEE International Symposium on Information Theory (ISIT)*, pages 676–680. IEEE, 2014.
6. J. Hao and S. Xia. Bounds and constructions of locally repairable codes: Parity-check matrix approach. *arXiv preprint arXiv:1605.05595*, 2016.
7. J. Hao, S. Xia, and B. Chen. Some results on optimal locally repairable codes. In *IEEE International Symposium on Information Theory (ISIT)*, pages 440–444. IEEE, 2016.
8. S. Hu, I. Tamo, and A. Barg. Combinatorial and lp bounds for lrc codes. In *IEEE International Symposium on Information Theory (ISIT)*, pages 1008–1012. IEEE, 2016.
9. F. J. MacWilliams and N. J. A. Sloane. *The theory of error correcting codes*. Amsterdam, The Netherlands: North-Holland, 1977.
10. G. Markus. Bounds on the minimum distance of linear codes and quantum codes. Online available at [http://www.codetables.de](http://www.codetables.de). Accessed on 2016-12-01.
11. D. S. Papailiopoulos and A. G. Dimakis. Locally repairable codes. In *IEEE International Symposium on Information Theory (ISIT)*, pages 2771–2775. IEEE, 2012.
12. N. Prakash, G. M. Kamath, V. Cadambe, and P. V. Kumar. Optimal linear codes with a local-error-correction property. In *IEEE International Symposium on Information Theory (ISIT)*, pages 2776–2780, 2012.
13. N. Prakash, V. Cadambe, and P. V. Kumar. Codes with locality for two erasures. In *IEEE International Symposium on Information Theory (ISIT)*, pages 1962–1966, 2014.
14. M. Shahshahani, M. Khabbaziyan, and M. Ardakani. A class of binary, locality repairable codes. *IEEE Transactions on Communications*, 64(8):3182–3193, 2016.
15. N. Silberstein, A. S. Rawat, O. O. Koyluoglu, and S. Vishwanath. Optimal locally repairable codes via rank-metric codes. In *IEEE International Symposium on Information Theory (ISIT)*, pages 1819–1823, 2013.
16. N. Silberstein and A. Zeh. Optimal binary locally repairable codes via antcodes. *arXiv preprint arXiv:1501.07114*, 2014.
17. W. Song, S. Dau, C. Yueh, and T. Li. Optimal locally repairable linear codes. *IEEE Journal on Selected Areas in Communications*, 32:6925–6934, May 2014.
18. W. A. Stein et al. *Sage Mathematics Software (Version 7.0)*. The Sage Development Team, 2016. [http://www.sagemath.org](http://www.sagemath.org).
19. I. Tamo and A. Barg. A family of optimal locally recoverable codes. *IEEE Transactions on Information Theory*, 60(4):4661–4676, 2014.
20. I. Tamo, A. Barg, and A. Frolov. Bounds on the parameters of locally recoverable codes. *IEEE Transactions on Information Theory*, 62:3070–3083, 2016.
21. I. Tamo, A. Barg, S. Goparaju, and R. Calderbank. Cyclic lrc codes and their subfield subcodes. In *IEEE International Symposium on Information Theory (ISIT)*, pages 1262–1266, 2015.
22. I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis. Optimal locally repairable codes and connections to matroid theory. In *IEEE International Symposium on Information Theory (ISIT)*, pages 1814–1818, 2013.
23. A. Wang and Z. Zhang. Repair locality from a combinatorial perspective. In *IEEE International Symposium on Information Theory (ISIT)*, pages 1972–1976, 2014.
24. A. Wang and Z. Zhang. An integer programming based bound for locally repairable codes. *IEEE Transactions on Information Theory*, 2015.
25. A. Zeh and E. Vardy. Optimal linear and cyclic locally repairable codes over small fields. In *Information Theory Workshop (ITW)*, pages 1–5, 2015.
APPENDIX A

From Lemma 4 it has \( K \geq N - (n - k), D \geq d. \) Then
\[
k \leq K - N + n
\]
\[
\leq n - l - \log_2 \left( \sum_{0 \leq i_1 + \cdots + i_l \leq \lfloor \frac{n - 1}{2} \rfloor} \prod_{j=1}^{l} \left( \frac{r_j + 1}{2} \right) \right)
\]
\[
\leq n - l - \log_2 \left( \Phi_i(r_1, \ldots, r_l) \right)
\]
where (a) is from \([7]\) and (b) holds because \( D \geq d. \) Note that the integers \( l, r_1, \ldots, r_l \) satisfies
\[
\left\{ \begin{array}{l}
\frac{n}{r+1} \leq l; \\
0 \leq r_1, \ldots, r_l \leq r; \\
\sum_{j=1}^{l} r_j \geq 2n - l(r + 2).
\end{array} \right.
\]

There are two cases.

Case 1: \( l \geq \frac{2n}{r+2} \). On the one hand, we have \( k \leq n - l < n - \frac{2n}{r+2} \). On the other hand, with the restriction \([2]\), it has
\[
\min_{l, r_1, \ldots, r_l} \left[ l + \log_2 \left( \Phi_i(r_1, \ldots, r_l) \right) \right]
\leq \frac{2n}{r+2} + \log_2 \left( \Phi_{\frac{n}{r+2}}(0, \ldots, 0) \right)
\leq \frac{2n}{r+2}.
\]

So the theorem holds.

Case 2: \( l \leq \frac{2n}{r+2} \). In this case it has
\[
\left\{ \begin{array}{l}
\frac{n}{r+1} \leq l \leq \frac{2n}{r+2}; \\
0 \leq r_1, \ldots, r_l \leq r; \\
\sum_{j=1}^{l} r_j \geq 2n - l(r + 2).
\end{array} \right.
\]

So we have
\[
k \leq n - \min_{l, r_1, \ldots, r_l} \left[ l + \log_2 \left( \Phi_i(r_1, \ldots, r_l) \right) \right],
\]
where the ‘Min’ is taken over \([11]\). Note that \( 2n - l(r + 2) \geq 0, \) so a necessary condition for optimizing \([8]\) is that \( \sum_{j=1}^{l} r_j = 2n - l(r + 2) \). Then the optimization can be restricted to the condition \([7]\), and thus the theorem holds.

APPENDIX B

When \( d \geq 5 \), we have
\[
\Phi_i(r_1, \ldots, r_l) \geq \sum_{0 \leq i_1 + \cdots + i_l \leq \lfloor \frac{n - 1}{2} \rfloor} \prod_{j=1}^{l} \left( \frac{r_j + 1}{2} \right)
\geq 1 + \left( \frac{r_1 + 1}{2} \right) + \cdots + \left( \frac{r_l + 1}{2} \right).
\]

Note that \( \frac{r+1}{2} \) is a convex real-valued function and it is required in \([8]\) that \( r_1 + \cdots + r_l = 2n - l(r + 2) \), so
\[
\Phi_i(r_1, \ldots, r_l) \geq 1 + l \left( \frac{1}{2} \sum_{j=1}^{l} (r_j + 1) \right)
= 1 + \frac{l}{2l} \left( 2n - l(r + 1) \right) \left( 2n - l(r + 2) \right).
\]

Then it follows from Theorem 3 that
\[
k \leq n - \min_{l, r_1, \ldots, r_l} \left[ l + \log_2 \left( \Phi_i(r_1, \ldots, r_l) \right) \right]
\leq n - \min \left\{ \frac{2n - l(r + 2) + \log_2 \left( \frac{1}{2l} \left( 2n - l(r + 1) \right) \left( 2n - l(r + 2) \right) \right)}{r+1}, \frac{2n - l(r + 2)}{r+1} \right\}
\]
\[
= \frac{n}{r+1} + \frac{\log_2 (1 + \frac{rn}{2})}{r+1}, \frac{2n}{r+1}.
\]

for \( 2 \leq r \leq \frac{n}{2} - 2 \).

We claim \( f''(l) \leq 0 \) for \( 2 \leq r \leq \frac{n}{2} - 2 \). Note that
\[
f''(l) = \frac{80n^4 - (l^2r^2 + l^2r^2 - 8n^2)^2 - 16l(n^2(3 + 2r) - n^2)}{l^2(4n^2 + l^2 + 2 + 3r^2 + r^2 + l(2 - 2n(3 + 2r)))^2 ln^2},
\]
then it suffices to show \( g(l) \leq 0 \) for \( 2 \leq r \leq \frac{n}{2} - 2 \), where
\[
g(l) = \frac{80n^4 - (l^2r^2 + l^2r^2 - 8n^2)^2 - 16l(n^2(3 + 2r) - n^2)}{l^2(4n^2 + l^2 + 2 + 3r^2 + r^2 + l(2 - 2n(3 + 2r)))^2 ln^2}.
\]
Since \( g''(l) = -24(l^2r^2 + 3r^2 + 2l^2 - 2l^2) < 0 \), it has \( g'(l) \) is a concave function with \( g'\left( \frac{n}{r+1} \right) = \frac{4n^2(4r^2 - n^2)}{(r+1)^2} < 0 \) and \( g'\left( \frac{2n}{r+2} \right) = \frac{16n^2(2r + 2n - n)}{(r+1)^2} > 0 \). It follows that
\[
g(l) \leq \max\{g\left( \frac{n}{r+1} \right), g\left( \frac{2n}{r+2} \right)\}
\leq 0
\]
for all \( 2 \leq r \leq \frac{n}{2} - 2 \), which proves the claim.

According to \( f''(l) \leq 0 \), \( f(l) \) is a concave function, and thus
\[
f(l) \geq \min\{f\left( \frac{n}{r+1} \right), f\left( \frac{2n}{r+2} \right)\}
\geq \min\left\{ \frac{n}{r+1} + \log_2 (1 + \frac{rn}{2}), \frac{2n}{r+2} \right\}
= \frac{n}{r+1} + \min\{\log_2 (1 + \frac{rn}{2}), \frac{2n}{r+2} \}
= \frac{n}{r+1} + \min\{\log_2 (1 + \frac{rn}{2}), \frac{2n}{r+2} \}, \frac{16n^2(2r + 2n - n)}{(r+1)(r+2)}\}
\]

Therefore the theorem follows directly.

APPENDIX C

Let \( \beta \) be the primitive root of \( x^{2^d+1} - 1 \), and let \( M(x) \) denote the minimal polynomial of \( \beta \), then \( \deg (M(x)) = 2t \). Define \( A \) to be the binary cyclic code of length \( (2^t + 1) \) which is generated \( (x - 1)M(x) \). Then it can be checked that
\[
\{ \beta^i : i = -2, -1, 0, 1, 2, \}
\]
forms a subset of the roots of \( (x - 1)M(x) \). It follows from the BCH bound that \( A \) is an \([2^t + 1, 2^t - 2t, \geq 6]\) binary linear code. Then an \([2^t, 2^t - 2t, \geq 5]_2 \) punctured code can be obtained by deleting one coordinate of \( A \), and thus the matrix \( A \) can be just chosen as the parity check matrix of the punctured code.
APPENDIX D

Note that \( n = \frac{2^t-1}{2^t-1} \), \( d = 5 \) and \( r = 2^t \), then \( (r + 1) \mid n \) and Theorem 6 implies

\[
k \leq \frac{rn}{r+1} - \min\{ \lceil \log_2(1 + \frac{rn}{2}) \rceil, \left\lfloor \frac{rn}{r+1(r+2)} \right\rfloor \}.
\]

In the following we show that \( \lceil \log_2(1 + \frac{rn}{2}) \rceil = s \) and \( \left\lfloor \frac{rn}{(r+1)(r+2)} \right\rfloor \geq s \) for all \( s, t \) satisfying \( 2t \mid s, \frac{s}{2t} \geq 2 \) except \( s = 4, t = 1 \). Observe that

\[
\log_2(1 + \frac{rn}{2}) = \log_2(1 + \frac{2^t}{2^t-1} \cdot \frac{2s-1}{2} )
\]

Since \( 1 < \frac{2^t}{2^t-1} \leq 2 \), we have

\[
2^{s-1} < 1 + \frac{2^t}{2^t-1} \cdot \frac{2s-1}{2} \leq 2^s.
\]

It follows that \( \lceil \log_2(1 + \frac{rn}{2}) \rceil = s \). It remains to show \( \left\lfloor \frac{rn}{(r+1)(r+2)} \right\rfloor \geq s \). When \( t = 1 \), it has \( s > 4 \) and

\[
\frac{rn}{(r+1)(r+2)} = \frac{2^t}{(2^t+1)(2^t+2)} \cdot \frac{2s-1}{2t-1} = \frac{1}{6}(2^s-1) \geq s.
\]

When \( t \geq 2 \), it has

\[
\frac{rn}{(r+1)(r+2)} \geq \frac{2^t}{(2^t+1)(2^t+2)} \cdot \frac{2s-1}{2t-1} \cdot \frac{s(2^{4t}-1)}{4t} \geq \frac{2^t}{4t} \cdot \frac{2^t+1}{s} \cdot s
\]

where (a) holds since \( \frac{2^t-1}{s} \geq \frac{2^{4t}-1}{4t} \), which is a consequence of \( s \geq 4t \), and (b) holds since \( \frac{2^t+1}{s} \geq \frac{1}{2} \) and \( 2^t+1 \geq 8t \) for \( t \geq 2 \).