Some New Sequential-Recovery LRCs Based on Good Polynomials

ZHII JING, (Student Member, IEEE), AND HONG-YEOP SONG, (Senior Member, IEEE)

School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea

Corresponding author: Hong-Yeop Song (hysong@yonsei.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant by the Korean Government through Ministry of Sciences and ICT (MSIT) under Grant 2020R1A2C2011969.

Received January 15, 2022, accepted April 6, 2022, date of publication April 14, 2022, date of current version April 27, 2022.

I. INTRODUCTION

For the reliability of the distributed storage systems (DSSs), the locally repairable codes (LRCs) has drawn much attention, since any erased symbol can be repaired by only a few other symbols. Let C be an \([n, k, d]\) linear code over \( \mathbb{F}_q \), whose length is \( n \), dimension is \( k \), and minimum distance is \( d \). Let \( c = (c_0, c_1, \ldots, c_{n-1}) \) be a codeword of \( C \). The code \( C \) is said to be an LRC with locality \( r \) [1] if, for each \( i = 0, 1, \ldots, n-1 \), the coded symbol \( c_i \) is a linear combination of \( r \) other symbols, and denoted by an \([n, k, d, r] \) LRC.

In this paper, we will consider LRCs with multiple erasures. They are divided into sequential- and parallel-recovery LRCs based on whether the repair process is either sequential or parallel. Now, let \( C \) be an \([n, k, d, r, \tau] \) LRC with a code word \( c \). Then it is said to be a \( \tau \)-sequential-recovery (\( \tau \)-seq) LRC [2] if, for any \( s \) (\( \leq \tau \)) erased symbols, there exists an arrangement of \( s \) erased positions given by \((j_0, j_1, \ldots, j_{s-1})\) such that, for each \( l = 0, 1, \ldots, n - 1 \), there is a subset \( R_l \subset \{0, 1, \ldots, n-1\} \) satisfying

1. \( j_l \in R_l \) and \(|R_l| \leq r + 1 \),
2. \( R_l \cap \{j_{l+1}, j_{l+2}, \ldots, j_{s-1}\} = \emptyset \), and
3. \( c_{j_l} = \sum_{i \in R_l} a_i c_i \), for some \( a_i \in \mathbb{F}_q \).

The \( \tau \)-parallel-recovery (\( \tau \)-para) \([n, k, d, r, \tau] \) LRCs [2] is defined similarly as the sequential-recovery LRCs by replacing the second condition with

\[ R_l \cap \{j_0, j_1, \ldots, j_{s-1}\} = \emptyset. \]

It is obvious that the \( \tau \)-para LRCs is also the \( \tau \)-seq LRCs since it can be locally repaired by any arrangement. As shown in Fig. 1, various types of parallel-recovery LRCs are proposed.

- LRCs with availability \((r, t)\) [3]: Let \( C \) be an \([n, k, d]\) code over \( \mathbb{F}_q \) and \( c = (c_1, c_2, \ldots, c_n) \in C \). The code \( C \) is said to have the availability \((r, t)\) if, for each \( i \in \{1, 2, \ldots, n\} \), there exist at least \( t \) disjoint subsets \( R_1(i), R_2(i), \ldots, R_t(i) \subset \{1, 2, \ldots, n\} \setminus \{i\} \) satisfying, for \( j = 1, 2, \ldots, t \)
  1. \( |R_j(i)| \leq r \),
  2. \( c_i = \sum_{i \in R_j(i)} a_i c_i \), where \( a_i \in \mathbb{F}_q \).

- \((r, \delta)\) LRCs [4]: Let \( C \) be an \([n, k, d]\) code over \( \mathbb{F}_q \) and \( c = (c_1, c_2, \ldots, c_n) \) be a codeword of \( C \). The code \( C \) is said to have the locality \((r, \delta)\) if, for each \( i \in \{1, 2, \ldots, n\} \), there exists a subset \( S_i \subset \{1, 2, \ldots, n\} \) satisfying
  1. \( i \in S_i \) and \(|S_i| \leq r + \delta - 1 \),
  2. \( d_{\min}(C |_{S_i}) \geq \delta \),

where \( C |_{S_i} \) is the punctured subcode of \( C \) by deleting code symbols \( c_j, j \in \{1, 2, \ldots, n\} \setminus S_i \).

- Hierarchical LRCs (HLRCs) [5]: Let \( C \) be an \([n, k, d]\) code and \( c = (c_1, c_2, \ldots, c_n) \in C \). The code \( C \) is said to have the hierarchical locality \([(r_1, \delta_1), (r_2, \delta_2)]\) if, for
each $i \in \{1, 2, \ldots, n\}$, there exists a punctured subcode $C_i$ such that $c_i \in \text{supp}(C)$, $\text{dim}(C_i) \leq r_1$, $\text{dmin}(C_i) \geq \delta_1$ and $C_i$ is a $(r_2, \delta_2)$ LRCs.

- **LRCs with cooperative locality [6]:** Let $C$ be an $[n, k, d]$ code and $c = (c_1, c_2, \ldots, c_n) \in C$. The code $C$ is said to have the cooperative locality $r_h$ if, for each $S \subseteq \{1, 2, \ldots, n\}$ of size $h$, there exists a subset $R(S)$ of $\{1, 2, \ldots, n\} \setminus S$ satisfying
  1. $|R(S)| \leq r_h$,
  2. for each $i \in S$, $c_i = \sum_{j \in R(S)} a_i c_j$, where $a_i \in \mathbb{F}_q$.

Compared with the sequential-recovery LRCs, parallel-recovery LRCs needs less time to repair all erasures. Various bounds [3]–[8] and constructions [9]–[17] have been proposed for each type of parallel-recovery LRCs, even some are binary codes [9], [10], [14], [16], [17].

The sequential-recovery LRCs have a much more advantage on the erasure tolerance than the parallel-recovery LRCs with the same $n$ and $k$ [18]. The bounds on the code rate and/or the block-length of $t$-seq LRCs were proposed in [2], [19]–[22]. In [2], binary rate optimal $2$-seq LRCs is constructed based on the regular graph. For any $t$, three classes of $t$-seq LRCs have been proposed: 1) graph based construction [19]; 2) resolvable configurations based construction [20]; 3) the generalized direct product construction [20]. The first two constructions are rate optimal when $t = 2, 3$, and the last one is rate optimal for any $t$. Binary 2-seq and 3-seq length optimal LRCs are proposed based on the graph [2], [21].

The upper bound on the minimum distance of $2$-seq LRCs was proposed in [2], which is the only bound on the distance of the $t$-seq LRCs as far as we know. Four explicit constructions of 2-seq LRCs for any $r$ that achieve the upper bound on the minimum distance were also proposed in [2]. Note that $r = 2$ is the most interesting situation in practice. The parameters of all the known distance-optimal 2-seq $[n, k, d, 2]$ LRCs are shown as follows [2].

1. Field size of 2: $[6, 3, 3, 2]$; $[8, 4, 3, 2]$;
2. Field size of $O(r)$: $[8, 2, 6, 2]$;
3. Field size of $O(n^{1/2})$: $[6, k, d, 2]$; $[8, k, d, 2]$;

From the above parameters, we know that the options for the values of $n$ and $k$ are very limited. It would be better if we could have a large family of 2-seq LRCs with a larger minimum distance over a relatively small field size. In this paper, we focus on the linear 2-seq LRCs with $k > r > 1$. Our first contribution of this paper is a construction of 2-seq $[n, k, d, r]$ LRCs for even $r$, based on some good polynomials, and show several properties. Our second contribution of this paper is the derivation of an explicit upper bound on the minimum distance of a certain class of 2-seq LRCs. We also prove the proposed 2-seq LRCs for $r = 2$ is optimal or near-optimal in the sense of attaining the upper bound on the minimum distance that we derived.

Section II introduces some preliminaries about the good polynomial-based LRCs for single erasure and several bounds of 2-seq LRCs. Section III describes two main contributions of the paper in detail. An open problem for the near future is given in Section IV.

II. PRELIMINARIES

A. LRC BASED ON GOOD POLYNOMIAL

Let $g(x) \in \mathbb{F}_q[x]$ be a polynomial of degree $r + 1$. If there exist $t_n$ disjoint subsets $A_0, A_1, \ldots, A_n-1$ of $\mathbb{F}_q$, each of size $r + 1$, such that $g(x)$ is constant on each subset, then $g(x)$ is called good [1]. Note that a subset of size $j$ is called a $j$-subset.

**Known Fact 1:** (Construction 1 in [1]) Let $r$ be a positive integer. Let $k$ and $n$ be positive integers with $r \mid k$, $(r + 1) \mid n$, and $\frac{n}{r} \leq q$. Let $\mu = \frac{n}{r}$ and $A_0, A_1, \ldots, A_{n-1}$ be $\mu$ disjoint $(r + 1)$-subsets of $\mathbb{F}_q$, and let $g(x)$ be a good polynomial with respect to these disjoint subsets. Then, the good polynomial-based $[n, k, d, r]$ LRC $C_1$ over $\mathbb{F}_q$ is defined as the set of codewords given as follows:

$$C_1 = \{(f_a(\gamma), \gamma \in \cup_{i=0}^{\mu-1} A_i) | a \in \mathbb{F}_q^k\},$$

where $a \in \mathbb{F}_q^k$ is an information vector written as $a = (a_{ij}, i = 0, 1, \ldots, r - 1; j = 0, 1, \ldots, \frac{n}{r} - 1)$, and $f_a(x)$ is the encoding polynomial of $a$, given as

$$f_a(x) = \sum_{i=0}^{r-1} \sum_{j=0}^{\frac{n}{r} - 1} a_{ij} x^j g(x)^i.$$  

The good polynomial-based LRCs is a class of the optimal $[n, k, d, r]$ LRCs for single erasure in the sense that the minimum distance $d = n - k - \frac{k}{r} + 2$ over the field of size $q \approx n$. Suppose the symbol $c_{\gamma(j)}$ of $f_a(\gamma)$ of a codeword is erased, where $\gamma \in A_j$ for some $j = 0, 1, \ldots, u - 1$, then its decoding polynomial is given as [1]

$$\delta(x) = \sum_{\beta \in A_j \setminus \gamma} f_a(\beta) \prod_{\beta' \in A_j \setminus \{\beta, \gamma\}} \frac{x - \beta'}{x - \beta}.$$  

Then, $c_{\gamma(j)} = \delta(\gamma)$.

B. RECURSIVE UPPER BOUND OF 2-SEQ LRCs

The support set of a vector $u = (u_0, u_1, \ldots, u_{n-1})$ is defined as $\text{supp}(u) = \{i | u_i \neq 0\}$, and $w(u) = |\text{supp}(u)|$ is the weight of $u$. The support set of a subcode $D$ of a code $C$ is defined as $\text{supp}(D) = \cup_{c \in D} \text{supp}(c)$. For $i = 1, 2, \ldots, k$, the $i^{th}$ generalized Hamming weight (GHW) of an $[n, k, d]$ linear code $C$ is defined as [2], [23]

$$d_i(C) = \min_{\text{dim}(D) = i} |\text{supp}(D)|.$$
It is well-known that
\[ d = d_1(C) < d_2(C) < \cdots < d_k(C) \leq n. \]

The \( n-k \) remaining numbers when \( d_1(C), d_2(C), \ldots, d_k(C) \) are removed from \( \{1, 2, \ldots, n\} \) are called the gap numbers \( g_1(C), g_2(C), \ldots, g_{n-k}(C) \), with \( g_1(C) < g_2(C) < \cdots < g_{n-k}(C) \). The gap number is also called the gap for simplicity.

For a given 2-seq \([n, k, d, r]\) LRC \( C \), its local dual subcode \( C_\perp \) is defined as [2]
\[ C_\perp = \text{span} \left( \{ e \in C^\perp \mid |\text{supp}(e)| \leq r + 1 \} \right), \]
where \( C^\perp \) is the dual code of \( C \). It is also called the local dual for simplicity. It is well-known that [2]
\[ \dim(C_\perp) \geq \left\lceil \frac{2n}{r+2} \right\rceil. \tag{3} \]

The necessary and sufficient condition of the equality in (3) is widely open. It is known that the code becomes rate-optimal when the equality is satisfied [2].

**Known Fact 2:** (Theorem 9 in [2]) Let \( C \) be a 2-seq \([n, k, d, r]\) LRC and \( b \triangleq \left\lceil \frac{2n}{r+2} \right\rceil \). Then, the upper bounds on the first \( b \) GHWs of \( C_\perp \) are given by, for \( i = 1, 2, \ldots, b, \)
\[ d_i(C_\perp) \leq e_i, \]
where \( e_i \) can be obtained recursively as follows:
\[ e_1 = \begin{cases} n, & i = b, \\ e_{i+1} - \left\lceil \frac{2\alpha + 1}{r+2} \right\rceil \cdot (r+1), & i = b-1, \ldots, 2, 1. \end{cases} \tag{4} \]

Furthermore, if there exists a unique integer \( l \) such that
\[ e_l < k + l < e_{l+1}, \tag{5} \]
then the upper bound on the minimum distance of \( C \) is given as
\[ d_{\min}(C) \leq n - (k + l). \tag{6} \]

The 2-seq LRC \( C \) is said to be distance optimal if (6) holds with equality, and is said to be distance near-optimal if \( d_{\min}(C) \) is one less than the RHS of (6).

**Known Fact 3:** (Theorem 2 in [2]) Let \( C \) be a 2-seq \([n, k, d, r]\) LRC. The upper bound on the rate of \( C \) is given by
\[ \frac{k}{n} \leq \frac{r}{r+2}. \tag{7} \]

We say that the 2-seq LRC \( C \) is rate optimal when its code rate achieves the equality in (7).

**III. MAIN RESULT**

**A. NEW CONSTRUCTION OF 2-SEQ LRCs**

In this subsection, we propose a construction of 2-seq LRCs with even locality \( r \) over \( \mathbb{F}_q \), which can be seen as an “extended code” of the good polynomial-based LRCs. Furthermore, we calculate its minimum distance for \( r = 2 \) and show that it is either optimal or near-optimal in terms of the minimum distance.

Let \( r \) be an even integer and \( q \) be a prime or a prime power with \( r + 1 \mid q - 1 \). Let \( t_0 \) and \( t_k \) be positive integers with \( t_k \leq t_n \leq \frac{q - 1}{r + 1} \). For \( i = 0, 1, \ldots, t_n - 1 \), let \( A_i = \{\alpha, \alpha \beta, \cdots, \alpha \beta^s\} \) where \( \alpha \) is a primitive element of \( \mathbb{F}_q \) and \( \beta = \alpha^{\frac{1}{r+1}}. \) An \([r+1]t_0, r_t, d_1, r]\) good polynomial-based LRC \( C_1 \) over \( \mathbb{F}_q \) can be constructed using the good polynomial \( g(x) = x^{r+1} \) with respect to \( A_0, A_1, \ldots, A_{t_n-1} \) as described in Known Fact 1 of Subsection II-A. It is well-known that \( d_1 = (r+1)(t_0 - t_k) + 2 \) since \( C_1 \) is distance-optimal [1]. For the construction below, any codeword \( u \in C_1 \) is now written as an array \( u = (u_{ij}) \) for \( i = 0, 1, \ldots, t_n - 1 \) and \( j = 0, 1, \ldots, r \), where \( u_{ij} = f_q(\alpha^i \beta^j) \) where \( \alpha \beta^j \) is the \( j \)th element of \( A_i \).

**Construction 1:** We assume the same \( r \) (even), \( q, t_k \), and \( t_n \) as above for the LRC \( C_1 \) over \( \mathbb{F}_q \) with parameters \([r+1]t_0, r_t, d_1, r]\). The code \( C_2 \) is an \([r+2]t_0, r_t, d_2, r_2]\) LRC over \( \mathbb{F}_q \) with codewords \( c \) as the following array
\[ c = \begin{pmatrix} c_{0,0} & c_{0,1} & \cdots & c_{0,r} & c_{0,r+1} \\ c_{1,0} & c_{1,1} & \cdots & c_{1,r} & c_{1,r+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{t_n-1,0} & c_{t_n-1,1} & \cdots & c_{t_n-1,r} & c_{t_n-1,r+1} \end{pmatrix} \]
where, for \( i = 0, 1, \ldots, t_n - 1 \),
\[ c_{ij} = \begin{cases} u_{ij}, & j = 0, 1, \ldots, r, \\ \sum_{l=0}^{r-1} u_{il+1}, & j = r + 1. \end{cases} \tag{8} \]

**Remark 1:** The construction will work if \( c_{i,j+1} \) is the sum of ANY \( r \) symbols among \( u_{i,0}, u_{i,1}, \ldots, u_{i,r} \). We call this an “overall” parity bit.

**Theorem 1:** The code \( C_2 \) from Construction 1 is a 2-seq \([n = (r+2)t_0, r_t, d_2, r_2]\) LRC over \( \mathbb{F}_q \) with \( d_2 \geq d_1, r_2 = r \) and \( q \approx n^{\frac{r+1}{r+2}}. \)

Proof: The “overall” parity bit \( c_{i,r+1} \) will not decrease its minimum distance, and hence, \( d_2 \geq d_1 \). For \( C_1 \) we have \( q \approx (r+1)t_{0} \). The same value of \( q \) is used for \( C_2 \) with \( t_n = n/(r+2) \). Therefore, \( q \approx n^{\frac{r+1}{r+2}}. \)

We now prove that any 2 erasures of \( C_2 \) can be repaired locally and sequentially. We write any codeword of \( C_2 \) as an array \( e = (e_{ij}) \) for \( i = 0, 1, \ldots, t_n - 1 \) and \( j = 0, 1, \ldots, r \).

That is, we may view the codeword written as a matrix of size \( t_n \times (r + 2) \). When two erasures occur in two distinct rows, each erasure can be repaired in any order one by one individually because they belong to different, and hence, disjoint repair sets. We now consider the case with two erasures. Without loss of generality, assume that these two erasures belong to the top row of a codeword of \( C_2 \). We will denote \( c_{ij} \triangleq c_{0,j+1} \) for \( 0 \leq j \leq r + 1 \) for simplicity and convenience. Assume \( c_{x} \) and \( c_{y} \) are two erasures, and defined by \( e_{x} \) and \( e_{y} \), respectively. We will distinguish the following two cases: 1) \( 0 \leq x \leq r, y = r + 1 \) and 2) \( 0 \leq x < y \leq r \).

The case 1) is easy since \( e_{x} \) can be repaired by other \( r \) symbols from (2) first and then \( e_{y} \) is the sum of the first \( r \) symbols.

The case 2) has two subcases: \( y = r \) and \( y < r \). When \( y = r \), the sequential recovery is easy since \( e_{y} \) can be repaired.
by \( r \) symbols first as

\[
e_x = c_{r+1} - \sum_{i=0, i \neq x}^{r-1} c_i ,
\]

and then \( e_y \) can be repaired by \( r \) symbols from (2). We now consider the case where \( 0 \leq x < y \leq r \) in the following, which is repaired by the remaining \( r \) unerased symbols in the top row. Using the decoding polynomial (2) for the code \( C_1 \), we have

\[
e_x = \delta(\beta^y) \quad \text{and} \quad e_y = \delta(\beta^x),
\]

where the polynomial \( \delta(\beta^y) \) is determined as

\[
\delta(\beta^y) = \sum_{i=0, i \neq y}^{r-1} c_i \prod_{\tau=q, \tau \neq y}^{r} \frac{\beta^\tau - \beta^t}{\beta^x - \beta^t}.
\]

Adding these two relations, we have

\[
e_x + e_y = \delta(\beta^y) + \delta(\beta^x).
\]

Using the relation of \( C_2 \) in (8), we have

\[
e_x + e_y + \sum_{0 \leq j \leq r-1} \sum_{0 \leq j \leq r-1, j \neq x, y} c_j = c_{r+1}.
\]

It is now enough to show that two equations (10) and (11) in two unknowns \( e_x \) and \( e_y \) have a unique solution. The first equation (10) can be written as

\[
(1 - \prod_{\tau=0, \tau \neq x, y}^{r} \frac{\beta^\tau - \beta^t}{\beta^x - \beta^t}) e_x + (1 - \prod_{\tau=0, \tau \neq x, y}^{r} \frac{\beta^\tau - \beta^t}{\beta^y - \beta^t}) e_y = u_1
\]

for some \( u_1 \in \mathbb{F}_q \). Similarly, (11) can be written as

\[
e_x + e_y = u_2
\]

for some \( u_2 \in \mathbb{F}_q \). Some simple row-operations give the following:

\[
\begin{cases}
    e_x + e_y = u_2 \\
    e_x \prod_{\tau=0, \tau \neq x, y}^{r} \frac{\beta^\tau - \beta^t}{\beta^y - \beta^t} + e_y \prod_{\tau=0, \tau \neq x, y}^{r} \frac{\beta^\tau - \beta^t}{\beta^x - \beta^t} = u_3
\end{cases}
\]

for some constants \( u_2 \) and \( u_3 \in \mathbb{F}_q \). This equation will have a unique solution if the coefficient matrix is non-singular, or

\[
\left( \prod_{\tau=0, \tau \neq x, y}^{r} \frac{\beta^y - \beta^t}{\beta^x - \beta^t} \right)^2 \neq 1.
\]

It is straightforward to show that, for any \( 0 \leq x < y \leq r \), we have

\[
\prod_{\tau=0, \tau \neq x, y}^{r} \frac{\beta^y - \beta^t}{\beta^x - \beta^t} = \pm \beta^m,
\]

for \( (r+1) \not\mid m \), and the result is also an element of \( \mathbb{A}_0 \) since \( \mathbb{A}_0 \) is a multiplicative subgroup. Then we can get that

\[
\left( \prod_{\tau=0, \tau \neq x, y}^{r} \frac{\beta^y - \beta^t}{\beta^x - \beta^t} \right)^2 = (\pm \beta^m)^2 = \beta^{2m} \neq 1,
\]

since \( 2m \neq 0 \) (mod \( r+1 \)) for even \( r \). Therefore, (12) has a unique solution.

From Fact 2, we have the following condition for the rate optimality of the proposed LRCs.

**Corollary 1:** The 2-seq LRC \( C_2 \) over \( \mathbb{F}_q \) in Theorem 1 is rate optimal when \( t_k = t_n \).

**Example 1:** Let \( r = 2, q = 13 \) and hence \( \frac{q-1}{r+1} = 4 \). Choose \( t_k = 3 \) and \( t_n = 4 \) for a 2-seq LRC \( C_2 \) over \( \mathbb{F}_{13} \) with parameters \( [16, 6, d_2 = 6, 2] \). We will prove shortly that \( d_2 = 4(t_n - t_k) + 2 = 6 \) when \( t_k < t_n \) and \( r = 2 \) (Theorem 2).

We first construct a \([12, 6, 5, 2]\) good polynomial-based LRC \( C_1 \) over \( \mathbb{F}_{13} \) using \( \mathbb{A}_0 = \{1, 3, 9\}, \mathbb{A}_1 = \{2, 6, 5\}, \mathbb{A}_2 = \{4, 12, 10\}, \mathbb{A}_3 = \{8, 11, 7\} \) where we use \( \alpha = 2 \) and \( \beta = \alpha^4 = 3 \). For any information \( a \), the corresponding codeword \( c \) of \( C_2 \) (as an array) becomes

\[
c = \left( f_a(1) f_a(3) f_a(9) f_a(1) + f_a(3) f_a(2) f_a(6) f_a(5) f_a(2) + f_a(6) f_a(4) f_a(12) f_a(10) f_a(4) + f_a(12) f_a(8) f_a(11) f_a(7) f_a(8) + f_a(11) \right) .
\]

**Theorem 2:** Consider the 2-seq \([4n, 2n, d_2, 2]\) LRC \( C_2 \) over \( \mathbb{F}_q \) in Theorem 2 with \( r = 2 \). The minimum distance \( d_2 \) of \( C_2 \) is given by

\[
d_2 = \left\lfloor 4(t_n - t_k) + 2 \right\rfloor, \quad t_k < t_n.
\]

We will provide the proof in Appendix A at the end for the readability and convenience.

## B. EXPLICIT UPPER BOUND ON THE MINIMUM DISTANCE

In this subsection, we derive an explicit form of the upper bound on the minimum distance of \( C \), and show in Corollary 3 that the proposed 2-seq LRCs with \( r = 2 \) is distance near-optimal when \( t_k < t_n \), and is distance optimal when \( t_k = t_n \).

**Lemma 1:** Let \( C \) be a 2-seq \([n, k, d, r]\) LRC and \( C_{\perp} \) be its local dual. Denote by \( d_i(C_{\perp}) \) the \( i^{th} \) GH\( W \) of \( C_{\perp} \) as defined in Subsection II-B. When \( n = (r + 2)t \) for some \( t \) and \( (r - 1) \mid 2(t - 1) \), the upper bound of \( d_i(C_{\perp}) \) is given as, for \( i = 1, 2, \ldots, 2t \), with \( h \triangleq \frac{2(t - 1)}{r - 1} \),

\[
d_i(C_{\perp}) \leq (r + 1)i - h \sum_{j=1}^{\left\lfloor \frac{i + 1}{h} \right\rfloor} j - \left\lfloor \frac{i - 1}{h} \right\rfloor \left( i - 1 - h \cdot \left\lfloor \frac{i - 1}{h} \right\rfloor \right) .
\]

**Proof:** Observe that \( \left\lceil \frac{2nt}{r+2} \right\rceil = 2t \). Denote by \( \psi_i \) the RHS of (14), and we will prove that \( \psi_i \) satisfies the same recursion of \( e_i \) in (4) for \( i = 1, 2, \ldots, 2t \). We will distinguish the cases where \( h > 1 \) and \( h = 1 \).
Assume \( h = \frac{2(t-1)}{r-1} > 1 \). When \( i = 2r \), we have \( 2r - 1 = 2(r - 1) + 1 \). Therefore, \( \frac{2r-1}{h} = r - 1 + \frac{1}{h} \).

Now, it is straightforward to check that \( \psi_{2r} = t(r + 2) = n \).

For any integer \( m = 1, 2, \ldots, 2r - 1 \), assume that \( \psi_{m+1} \) is given as the RHS of (14) for \( i = m + 1 \). Now, we will check that \( \psi_{m+1} = \left\lceil \frac{\psi_{m+1}}{m+1} \right\rceil + (r + 1) \) becomes \( \psi_m \). Let \( \delta \) and \( \epsilon \) be positive integers, such that \( m = \delta h + \epsilon \) and \( 0 \leq \epsilon < h \). When \( 1 < \epsilon < h \), \( \left\lceil \frac{m}{h} \right\rceil = \delta \) and \( \left\lfloor \frac{m}{h} \right\rfloor = \delta + 1 \).

Therefore, we have

\[
\frac{2\psi_{m+1}}{m+1} = 2(r + 1) - (\delta + 1) - \frac{\delta \epsilon + \epsilon - (\delta + 1)}{\delta h + \epsilon + 1},
\]

and

\[
\left\lceil \frac{2\psi_{m+1}}{m+1} \right\rceil = 2(r + 1) - \left\lceil \frac{m}{h} \right\rceil.
\]

\[
\psi_{m+1} = \left\lceil \frac{2\psi_{m+1}}{m+1} \right\rceil + (r + 1) = (r + 1)m - h \sum_{j=1}^{\frac{m-1}{h}} j
\]

\[
= (m - 1 - h) \left( m - h \cdot \frac{m - 1}{h} \right) = \psi_m.
\]

It is not difficult to check the same recursion is satisfied when \( \epsilon = 0 \) or \( \epsilon = 1 \).

For \( h = 1 \), the recursive relationship of \( \psi_m \) can be proved similarly.

LRCs with \( r = k \) can be seen as the maximum distance separable code, so we do not consider this case in this paper.

Further, if \( r(r-1) \mid 2(t-1) \) then \( (r-1) \mid 2(t-1) \). Therefore, the explicit upper bound on the minimum distance of 2-seq \([n, k, d, r] \) LRC \( C \) with \( r + 2|n \) and \( r < k \) can be derived based on the above explicit upper bound on the GHW of its local dual code \( C_\perp \).

**Theorem 3:** Let \( C \) be a 2-seq \([n, k, d, r] \) LRC. When \( n = (r + 2)t \) for some \( t \) and \( r(r-1) \mid 2(t-1) \), the upper bound on the minimum distance \( d \) of \( C \) is explicitly given as

\[
d \leq n + 1 - k - \left( jh + \left\lceil \frac{k - \Gamma}{r - (j + 1)} \right\rceil \right),
\]

where \( h = \frac{2(t-1)}{r-1} \) and \( j \) is the largest nonnegative integer that satisfies \( k > h \sum_{a=1}^{j} (r-a) + r \geq \Gamma \).

**Proof:** From (5) and (6), it is enough to show that \( l \geq \Gamma \) satisfies \( \psi_l < k + l < \psi_{l+1} \). Let \( j \) be the largest nonnegative integer that satisfies \( k \leq \Gamma \). \( \Gamma < k \leq \Gamma + h(r - (j + 1)) \) and hence

\[
0 < \left\lfloor \frac{k - \Gamma}{r - (j + 1)} \right\rfloor \leq h.
\]

Let \( m \triangleq \left\lfloor \frac{k - \Gamma}{r - (j + 1)} \right\rfloor \). \( \Gamma + (m - 1)(r - (j + 1)) \leq k \). By (14), it is straightforward to get

\[
\psi_{jh+m} = h \sum_{s=1}^{j} (r-s) + r + (m-1)(r - (j + 1)) + jh + m
\]

\[
< k + jh + m.
\]

Similarly, we can get \( \psi_{jh+m+1} > k + jh + m \).

**Corollary 2:** For the 2-seq LRC \( C \) in Theorem 3, when \( r = 2 \) and \( k = 2t_k \) for some \( t_k \leq t \), we have

\[
d \leq 4(t - t_k) + 3.
\]

**Proof:** We note that \( h = 2(t-1), j = 0 \) and \( r = 2 \), hence \( \Gamma = 2 \).

**Corollary 3:** The 2-seq LRC \( C_2 \) over \( \mathbb{F}_q \) in Theorem 2 is distance near-optimal when \( t_k < t_n \) and is distance optimal when \( t_k = t_n \).

**IV. CONCLUDING REMARK**

This paper constructed the near-optimal 2-seq \([r + 2]t_n, t_k, d, r] \) LRCs for even \( r \) over a relatively small alphabet of size \( q \approx \frac{(r+1)n}{r+2} \), where \( t_k \leq t_n \leq \frac{q-1}{r+2} \). For comparison,
we show the various parameters of the 2-seq LRCs in Table 1. The proposed 2-seq LRCs is rate optimal or distance optimal or distance near-optimal for some cases. In the future, it may be important to find a construction for the optimal 2-seq LRC with any locality \( r \geq 2 \) over a smaller alphabet.

APPENDIX A  THE PROOF OF THEOREM 2

We will fix the notations for \( C_2 \) in Theorem 2 and hence the corresponding \( C_1 \) also. Any codeword \( c \in C_2 \) in the array representation of size \( t_n \times (r + 2) \) consists of the codeword \( u \in C_1 \) of size \( t_n \times (r + 1) \) on the left and the right-most column of length \( t_n \). Here, we write \( u_i = (u_{i,0}, u_{i,1}, u_{i,2}) \) as \( i^{th} \) row of \( u \) and \( c_{i,n} = u_{i,0} + u_{i,1} \) for \( i = 0, 1, \ldots, t_n - 1 \). Therefore, we may write, as an array,

\[
e = (u|u_{\text{add}})
\]

and

\[
w(e) = w(u) + w(u_{\text{add}}) \tag{16}
\]

where \( u_{\text{add}} \) is the last column of \( e \) which consists of the “overall” parity bits.

We first take a look at the encoding polynomial of element \( u_{i,j} \) of \( u \in C_1 \). From (1) with \( r = 2 \),

\[
u_{i,j} = f_{a,i}(\alpha^j\beta^j) = \left( \sum_{l=0}^{t_n-1} a_{0,l}g(x)^y \right) + x \left( \sum_{l=0}^{t_n-1} a_{1,l}g(x)^y \right) \big|_{x=\alpha^j\beta^j} \tag{17}
\]

We note that \( g(\alpha^j\beta^j) = \alpha^{3i} \) for all \( j = 0, 1, 2 \). Therefore, for each \( i = 0, 1, \ldots, t_n - 1 \), the encoding polynomial \( f_{a,i}(x) \) of the \( i^{th} \) row \( u \) is given as

\[
f_{a,i}(x) = H_i + B_i x, \tag{18}
\]

where

\[
H_i = \sum_{l=0}^{t_n-1} a_{0,l}x^{3il}, \quad B_i = \sum_{l=0}^{t_n-1} a_{1,l}x^{3il}. \tag{19}
\]

We will denote by \( \tau_m \) the number of rows of \( u \) with weight \( m \) in \( u \) for \( m = 0, 1, 2, 3 \).

Lemma 2: Consider \( C_1 \) and \( C_2 \) in Theorem 2 and assume all the notation in the discussion after Theorem 2, leading to \( f_{a,i}(x) \) in (18). Let \( a \in \mathbb{F}_{2^k}^a \) be a non-zero information vector, and let \( u \) be its codeword of \( C_1 \). Let \( u_{\text{add}} \) be the last column of the corresponding codeword \( e \in C_2 \). Then, the following holds:

1) For each \( i = 0, 1, \ldots, t_n \),

\begin{enumerate}
\item[a)] \( w(u_i) = 0 \) if and only if \( H_i = 0 \) and \( B_i = 0 \). It is obvious that \( u_{i,0} + u_{i,1} = 0 \) if \( w(u_i) = 0 \);
\item[b)] \( w(u_i) = 1 \) is impossible; hence, \( \tau_1 = 0 \) for any \( u \);
\item[c)] \( w(u_i) = 2 \) if and only if \( \frac{-H_i}{B_i} \in A_1 \) with \( B_i 
= 0 \). If \( w(u_i) = 2 \), then \( u_{i,0} + u_{i,1} \neq 0 \);
\item[d)] \( w(u_i) = 3 \) if and only if \( H_i \neq 0 \) (when \( B_i = 0 \)) or \( \frac{-H_i}{B_i} \notin A_1 \) (when \( B_i \neq 0 \)). For a row \( u_i \) of weight 3, \( u_{i,0} + u_{i,1} = 0 \) if and only if \( 2H_i - \alpha^j\beta^jB_i = 0 \).
\end{enumerate}

2) In any \( u \), the number of rows with \( w(u_i) = 3 \) and \( u_{i,0} + u_{i,1} = 0 \) is at most \( \min(\tau_3, 2\tau_2 - 2\tau_0 - 1) \).

3) For any \( c \), \( w(u_{\text{add}}) \geq \tau_2 + \tau_3 - \min(\tau_3, 2\tau_2 - 2\tau_0 - 1) \).

Proof: We recall that \( |A_1| = 3 \) for all \( i \). We skip the proof of a) of Case 1). For the subcase b), we note that \( w(u_i) = 1 \) implies \( f_{a,i}(x) \) in (18) of degree at most 1 must have two roots. For the subcase c), we note that \( w(u_i) = 2 \) if and only if \( u_{i,j} = H_i + B_i\alpha^j\beta^j = 0 \) or equivalently, \( \frac{-H_i}{B_i} = \alpha^j\beta^j \in A_1 \) for some \( j \). For the second assertion, we assume that \( w(u_i) = 2 \) for some \( i \). Then, \( \frac{-H_i}{B_i} \in A_1 \) implies \( H_i = 0 \). Suppose that \( u_{i,0} + u_{i,1} = 0 \) on the contrary. Then, \( u_{i,0}, u_{i,1} \) must be both non-zero and \( u_{i,2} = 0 \), and hence, \( u_{i,0} + u_{i,1} + u_{i,2} = 0 \), which contradicts to the following:

\[
u_{i,0} + u_{i,1} + u_{i,2} = f_{a,i}(\alpha^j) + f_{a,i}(\alpha^j\beta^j) + f_{a,i}(\alpha^j\beta^{j+2}) = 3H_i + B_i\alpha^j(1 + \beta + \beta^2) = 3H_i.
\]

For the subcase d), we note that \( w(u_i) = 3 \) if and only if \( B_i \) and \( H_i \) satisfy the condition which is the complement of the union of the previous cases. For the second assertion, we observe first that \( u_{i,2} = H_i + B_i\alpha^j\beta^j \). Then, \( u_{i,0} + u_{i,1} = 0 \Leftrightarrow u_{i,0} + u_{i,1} + u_{i,2} = 0 \Leftrightarrow 3H_i = H_i + B_i\alpha^j\beta^j \Leftrightarrow 2H_i - B_i\alpha^j\beta^j = 0 \).

For the proof of Case 2), we observe the following: the codeword \( u \) has \( \tau_0 \) rows of weight 0 if and only if there exists \( 2\tau_0 \) linear relations on the elements of the information \( a \) as

\[
H_i = 0 \quad \text{and} \quad B_i = 0, \tag{20}
\]

for some \( \tau_0 \) values of \( i \). This is equivalent to saying that some \( 2\tau_0 \) elements in such \( a \) are linear combinations of the remaining \( 2t_0 - 2\tau_0 \) elements. When \( 2t_0 - 2\tau_0 < \tau_3 \), therefore, the number of additional linear dependencies of elements in such \( a \) can be at most \( 2t_0 - 2\tau_0 - 1 \). The necessary and sufficient condition

\[
2H_i - \alpha^j\beta^jB_i = 0, \tag{20}
\]

for \( u_{i,0} + u_{i,1} = 0 \), is also a linear dependency in \( a \). Therefore, the number of rows \( u_i \) with weight 3 satisfying \( 2H_i - \alpha^j\beta^jB_i = 0 \) can be at most \( 2t_0 - 2\tau_0 - 1 \). When \( 2t_0 - 2\tau_0 \geq \tau_3 \) on the other hand, it is obvious that the number of rows \( u_i \) with weight 3 satisfying \( 2H_i - \alpha^j\beta^jB_i = 0 \) can be at most \( \tau_3 \).

Case 3) comes easily from the previous cases.

Remark 2: The equality of Case 3) in Lemma 2 holds when, for information \( a \in \mathbb{F}_{2^k}^a \), there is only one choice of freedom in \( a \) and the remaining \( 2\tau_k - 1 \) elements of \( a \) are decided by \( 2\tau_0 \) equations of the type in (19) and \( 2\tau_k - 2\tau_0 - 1 \) equations of the type in (20).

Now, we continue the proof of Theorem 2. Let \( u, u' \in C_1 \) be codewords with the same weight such that \( 2\tau_2 + 3\tau_3 = 2\tau_2' + 3\tau_3' \). If \( \tau_2 > \tau_2' \) then \( \tau_2 - \tau_2' = 3\lambda \), \( \tau_0 - \tau_0' = -l \) and \( \tau_3 - \tau_3' = -2l \), for some positive integer \( l \), and hence, we have \( w(u_{\text{add}}) > w(u_{\text{add}}') \) by Lemma 2, which implies that \( w(e) > w(e') \).

We now claim that

\[
w(e) \geq 4(t_n - t_k) + 3 - \min(t_n - t_k, 1) = w_{\min} \tag{21}
\]
for any $c \in C_2$. Observe that it is enough to prove (21) for all the codewords $c$ corresponding to $u \in C_1$ with $t_2 \in \{0, 1, 2\}$. We will prove this by induction on the weight of $u$.

When $u \in C_1$ be a non-zero codeword with the minimum weight $d_1 = 3(n_k - t_1) + 2$, we can get $t_3 = n_k - t_1$, $t_2 = 1$, and $r_0 = t_1 - 1$. So, the weight of the corresponding codeword $c$ is, from 3) of Lemma 2,

$$w(c) \geq 3(n_k - t_1) + 2 + (t_2 + t_3 - \min(t_3, 2t_k - 2t_0 - 1)) = 4(n_k - t_1) + 3 - \min(t_1 - k, 1).$$

The weight of codeword $c$ achieves the bound when the corresponding information $a$ satisfies the condition in Remark 2, which obviously exists in $\mathbb{F}_{q^2}$. Let $u \in C_1$ have $w(u) > d_1$ and $c \in C_2$ be its corresponding codeword. If there exists a codeword $u' \in C_1$ with $w(u') = w(u) - 1$, then

$$2t_2' + 3t_3' = 2t_2 + 3t_3 - 1 \Rightarrow (t_2 - t_2') = 3(t_3' - t_3) + 1 \Rightarrow t_2 - t_2' = 2 or 1.$$

If $t_2 - t_2' = 2$ (the other case can be treated similarly), then $t_3 - t_3' = t_0 - t_0' = -1$. In this case, $w(u_{add}) > w(u_{add}').$ Therefore, $w(u) + w(u_{add}) > w(u') + w(u_{add}')$, which implies $w(c) > w(c') \geq w_{\min}$ by induction. If there exists a codeword $u' \in C_1$ with $w(u') \leq w(u) - 2$, we can prove similarly.

REFERENCES

[1] T. Tamo and A. Barg, “A family of optimal locally repairable codes,” IEEE Trans. Inf. Theory, vol. 60, no. 8, pp. 4661–4676, May 2014.

[2] N. Prakash, V. Lalitha, S. B. Balaji, and P. V. Kumar, “Codes with locality for two erasures,” IEEE Trans. Inf. Theory, vol. 65, no. 12, pp. 7771–7789, Dec. 2019.

[3] A. Wang and Z. Zhang, “Repair locality with multiple erasure tolerance,” IEEE Trans. Inf. Theory, vol. 60, no. 11, pp. 6979–6987, Nov. 2014.

[4] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, “Optimal linear codes with a local-error-correction property,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jul. 2012, pp. 2776–2780.

[5] B. Sasidharan, G. K. Agarwal, and P. V. Kumar, “Codes with hierarch-ical locality,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun. 2015, pp. 1257–1261.

[6] A. S. Rawat, A. Mazumdar, and S. Vishwanath, “On cooperative local repair in distributed storage,” EURASIP J. Adv. Signal Process., vol. 2015, no. 1, p. 107, Dec. 2015.

[7] J. Hao, K. W. Shum, S.-T. Xia, and Y.-X. Yang, “Classification of optimal ternary $(\tau, \delta)-$locally repairable codes attaining the singleton-like bound,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jul. 2019, pp. 2828–2832.

[8] I. Tamo, A. Barg, and A. Frolov, “Bounds on the parameters of locally repairable codes,” IEEE Trans. Inf. Theory, vol. 62, no. 6, pp. 3070–3083, Jun. 2016.

[9] J. Hao and S.-T. Xia, “Constructions of optimal binary locally repairable codes with multiple repair groups,” IEEE Commun. Lett., vol. 20, no. 6, pp. 1060–1063, Jun. 2016.

[10] J.-H. Kim and H.-Y. Song, “Hypergraph-based binary locally repairable codes with availability,” IEEE Commun. Lett., vol. 21, no. 11, pp. 2332–2335, Nov. 2017.

[11] Z. Sun, S. Zhu, and L. Wang, “Optimal constacyclic locally repairable codes,” IEEE Commun. Lett., vol. 23, no. 2, pp. 206–209, Feb. 2019.

[12] B. Chen, W. Fang, S.-T. Xia, and F.-W. Fu, “Constructions of optimal $(\tau, \delta)$-locally repairable codes via constacyclic codes,” IEEE Trans. Commun., vol. 67, no. 8, pp. 5253–5263, Aug. 2019.

[13] H. Cai, Y. Xiao, M. Schwartz, and X. Tang, “On optimal locally repairable codes with multiple disjoint repair sets,” IEEE Trans. Inf. Theory, vol. 66, no. 4, pp. 2402–2416, Apr. 2020.

[14] Y.-S. Su, “On constructions of a class of binary locally repairable codes with multiple repair groups,” IEEE Access, vol. 5, pp. 3524–3528, 2017.

[15] J. Hao and B. Chen, “On the generalized Hamming weights of $(\tau, \delta)$-locally repairable codes,” IEEE Access, vol. 8, pp. 149706–149713, 2020.

[16] J. Teng and L. Jin, “Constructions of binary locally repairable codes with multiple recovering sets,” IEEE Access, vol. 9, pp. 92239–92245, 2021.

[17] J.-H. Kim, M. K. Song, and H.-Y. Song, “Block-punctured binary simplex codes for local and parallel repair in distributed storage systems,” IEICE Trans. Fundam. Electron., Commun. Comput. Sci., vol. E101-A, no. 12, pp. 2374–2381, Dec. 2018.

[18] W. Song and C. Yuan, “Binary locally repairable codes–sequential repair for multiple erasures,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2016, pp. 1–6.

[19] S. B. Balaji, G. R. Kini, and P. V. Kumar, “A tight rate bound and matching construction for locally recoverable codes with sequential recovery from any number of multiple erasures,” IEEE Trans. Inf. Theory, vol. 66, no. 2, pp. 1023–1052, Feb. 2020.

[20] W. Song, K. Cai, C. Yuan, K. Cai, and G. Han, “On sequential locally repairable codes,” IEEE Trans. Inf. Theory, vol. 64, no. 5, pp. 3513–3527, May 2018.

[21] W. Song and C. Yuan, “Locally repairable codes with functional repair and multiple erasure tolerance,” 2015, arXiv:1507.02796.

[22] S. B. Balaji, K. P. Prasanth, and P. V. Kumar, “Binary codes with locality for multiple erasures having short block length,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Barcelona, Spain, Jul. 2016, pp. 655–659.

[23] Y. K. Wei, “Generalized Hamming weights for linear codes,” IEEE Trans. Inf. Theory, vol. 37, no. 5, pp. 1412–1418, Sep. 1991.