Received May 5, 2021, accepted June 7, 2021, date of publication June 21, 2021, date of current version July 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3091403

Constructions of Binary Locally Repairable Codes With Multiple Recovering Sets

JIAMING TENG AND LINGFEI JIN, (Member, IEEE)

Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai 200433, China

Corresponding author: Jiaming Teng (18210240017@fudan.edu.cn)

This work was supported in part by the Shanghai Rising-Star Program under Grant 20QA1401100, and in part by the National Natural Science Foundation of China under Grant 11871154.

ABSTRACT Locally repairable codes (LRCs) with multiple recovering sets are highly demanded in distributed storage systems. In this letter, we generalize the construction of WZL code proposed by Wang *et al.* and give a construction of optimal binary LRCs with multiple disjoint recovering sets which can reach the upper bound on the code rate given by Kadhe *et al.*. Then we further generalize our idea to obtain a construction of binary LRCs with intersect recovering sets. The code rate is much higher than that of WZL code and is very closed to the construction of Kruglik *et al.* Moreover, two special cases of this construction can reach the upper bound on the minimum distance.

INDEX TERMS Locally repairable codes, codes with availability, optimal LRCs, intersect recovering set.

I. INTRODUCTION

Distributed storage systems use redundancy to ensure data reliability, such as replication and MDS codes. Compared with traditional $[n, k]$ MDS codes, locally repairable codes (LRCs) only need to access $r \ll k$ active nodes to recover a failure node at the cost of a small amount of storage overhead, where r is called locality. The set of these r nodes (symbols) participating in the recovery of a failure node is referred to as a recovering set of the node.

The formal definition of LRCs was first introduced by Gopalan *et al.* [2]. Analogous to the classical Singleton bound, they established a tradeoff between minimum distance and locality, referred to as the Singleton-type bound. A code achieving this Singleton-type bound is called optimal. After their work, other bounds were given in [3]–[6]. A tighter upper bound on the dimension k of the LRCs depending on the alphabet size was given in [7], [8]. For more studies on the bound of LRCs, one can refer to [9]–[11]. The first breakthrough construction of optimal LRCs is given in [12] by generalizing the Reed-Solomon codes. For more constructions on optimal LRCs one can refer to [13]–[17].

However, if some nodes in the recovering set are not available, we have to find an alternative set of nodes to repair the failure node. Thus, it is desirable to have multiple disjoint sets of nodes available to repair data in each node. The number t of the disjoint sets is called availability. A code is said to have locality r and availability t if every symbol has t disjoint recovering sets, denoted as (r, t)-LRCs [18]. (r, t)-LRCs also support parallel reading of data, which is very effective for solving the problems of degraded reading and hot data. The first upper bound on the minimum distance of the (r, t)-LRC is given in [19].

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

In [20], the authors gave a bound on the code rate of (r, t)-LRCs.

$$\mathcal{R} := \frac{k}{n} \leq \prod_{i=1}^{t} \frac{1}{1+\frac{1}{ir}}. \tag{2}$$

This bound applies to both linear codes and non-linear codes. Kadhe *et al.* gave a tighter bound on the rate for $(r, 3)$-LRCs over $\mathbb{F}_2$ in [21].

$$\mathcal{R} \leq 1 - \frac{3}{r+1} + \frac{3\log_2(2r+4)}{(r+1)(2r+3)}. \tag{3}$$

In 2017, Kruglik *et al.* [22] generalized the definition of (r, t)-LRCs, allowing the recovering sets of each coordinate to intersect at most x coordinates. We refer to these

The associate editor coordinating the review of this manuscript and approving it for publication was Xueqin Jiang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

codes as (r, t, x)-LRCs. This feature can increase the maximum achievable code rate [20] and still meet load balancing requirements. The bound on the minimum distance of (r, t, x)-LRCs is given in [1]. Note that this bound is also valid for the case of standard (r, t)-LRCs.

$$d \leq \min_{1 \leq i \leq k-r} \frac{q^i - q^{i-1}}{q^i - 1}\left(n - (k-i) - \left\lfloor \frac{k-1-i}{r-1} \right\rfloor\right) \quad (4)$$

In this paper, we generalized the construction of WZL codes [23], and proposed a construction of binary (r, t)-LRCs.We refer to it as *Construction 1*. The code rate of Construction 1 reaches the upper bound (3). Noted that our Construction 1 is similar with the construction given in [24]. Although both of the constructions are based on the same inclusion matrix of linear subspaces in $\mathbb{F}_q^m$, we additionally provide the block form of parity-check matrix. Also, we assert the rank of the parity-check matrix in Lemma 4 in a more complete and general way, so that the same arguement can be reused in Lemma 7.

Then, we use the same method to construct binary (r, t, x)-LRCs, which is referred to as *Construction 2*. The code rate is much higher than that of WZL codes and is very closed to the construction of Kruglik *et al.* [22]. The minimum distance of this code has $2\times$ greater than that of Kruglik's construction. Moreover, the minimum distance of the two special cases of Construction 2 can attain the upper bound (4).

II. PRELIMINARIES

Let $\mathcal{C}$ be an $[n, k, d]_q$ linear code over $\mathbb{F}_q$. Denote $\mathcal{R} = \frac{k}{n}$ as the rate of the code $\mathcal{C}$. Denote $[m] = \{1, 2, \cdots, m\}$ for a positive integer m. For any subset $I \subseteq [n]$ of coordinates of a code $\mathcal{C}$, denote $\mathcal{C}_I$ the restriction of $\mathcal{C}$ on I. Given $\alpha \in \mathbb{F}_q$, define $\mathcal{C}(i, \alpha) = \{\boldsymbol{c} \in \mathcal{C} : c_i = \alpha\}$. The support of a vector $\boldsymbol{v}$ is defined as $\text{supp}(\boldsymbol{v}) := \{i : v_i \neq 0\}$.

A. GAUSSIAN BINOMIAL COEFFICIENTS

The Gaussian binomial coefficient $\begin{bmatrix} n \\ k \end{bmatrix}_q$ is q-analogs of the binomial coefficients. Let n, k and $q > 1$ be positive integers, $\begin{bmatrix} n \\ k \end{bmatrix}_q$ is defined to be

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n - 1)(q^{n-1} - 1)\cdots(q^{n-k+1} - 1)}{(q^k - 1)(q^{k-1} - 1)\cdots(q - 1)}$$

$\begin{bmatrix} n \\ k \end{bmatrix}_q$ counts the number of subspaces of dimension k in a vector space of dimension n over a finite field $\mathbb{F}_q$.

B. LRC WITH AVAILIBLILITY

If every symbol of the code $\mathcal{C}$ can be recovered from t disjoint subsets of size r, then the code $\mathcal{C}$ is said to have locality r and availability t. The formal definition is as follows.

Definition 1: Let $\mathcal{C}$ be an $[n, k]_q$ linear code, for any coordinate $i \in [n]$, there exist t disjoint subsets of coordinates $R_i^1, \cdots, R_i^t \subseteq [n] \setminus \{i\}$ such that for all $j \in [t]$, $|R_i^j| \leq r$ and every pair of symbols $\alpha, \beta \in \mathbb{F}_q, \alpha \neq \beta$ we have

$$\mathcal{C}_{R_i^j}(i, \alpha) \cap \mathcal{C}_{R_i^j}(i, \beta) = \emptyset.$$

Wang *et al.* [23] gave a construction of binary (r, t)-LRCs for arbitrary r and t. We refer to the code as WZL code, and its parameters are $n_w = \binom{r+t}{t}$, $d_w = t + 1$, $\mathcal{R}_w = \frac{r}{r+t}$. They also gave the relation between (r, t)-LRCs and block design. We conclude in the following lemma.

Lemma 1 ([23]): The incidence matrix of a 1-$(n, r+1, t)$ design can be taken as the parity check matrix of the (r, t)-LRC with length n if its blocks $B_1, B_2, \cdots, B_b$ satisfying the following condition:

$$|B_i \cap B_j| \leq 1, \;\; 1 \leq i < j \leq b. \quad (5)$$

where $b = \frac{nt}{r+1}$.

C. CODES WITH AVAILABILITY AND INTERSECT RECOVERING SET

If the recovering set of (r, t)-LRC can intersect at most x positions, this code is defined as (r, t, x)-LRC.

Definition 2: A code is said to be (r, t, x)-LRC if for any coordinate $i \in [n]$, there exist t subsets of coordinates $R_i^1, \cdots, R_i^t \subseteq [n] \setminus \{i\}$ such that

$$|R_i^{\ell} \cap R_i^{\ell'}| \leq x, \forall \ell, \ell' \in [t]$$

and for all $j \in [t]$, $|R_i^j| \leq r$, every pair of symbols α, $\beta \in \mathbb{F}_q, \alpha \neq \beta$ we have

$$\mathcal{C}_{R_i^j}(i, \alpha) \cap \mathcal{C}_{R_i^j}(i, \beta) = \emptyset.$$

We generalize the Lemma 1 to the following result.

Corollary 1: The incidence matrix of a 1-$(n, r+1, t)$ design can be taken as the parity check matrix of the (r, t, x)-LRC with length n if its blocks $B_1, B_2, \cdots, B_b$ satisfies the following condition:

$$|B_i \cap B_j| \leq x + 1, \;\; 1 \leq i < j \leq b. \quad (6)$$

where $b = \frac{nt}{r+1}$.

Proof: If the given conditions are satisfied, then it is obvious that each row (resp. column) of the incidence matrix of such design has $r + 1$ (resp. t) 1s, and any two different rows can intersect at most $x + 1$ positions. The rest of the proof is similar to Lemma 1, so we omit it here. □

III. CODE CONSTRUCTIONS

For any positive integers m, a, b, such that $a < b, a+b \leq m$, we define a matrix over $\mathbb{F}_2$, denoted as $H_q(m, a, b)$, containing $\begin{bmatrix} m \\ a \end{bmatrix}_q$ rows and $\begin{bmatrix} m \\ b \end{bmatrix}_q$columns. Each row of $H_q(m, a, b)$ is associated with an a-dimensional subspace of $\mathbb{F}_q^m$, and each column of $H_q(m, a, b)$ is associated with a b-dimensional subspace of $\mathbb{F}_q^m$. For $1 \leq i \leq \begin{bmatrix} m \\ a \end{bmatrix}_q$, $1 \leq j \leq \begin{bmatrix} m \\ b \end{bmatrix}_q$, suppose the i-th row is associated with the subspace W_i and the j-th column is associated with the subspace V_j, then the (i, j)-th element h_{ij} of $H(m, a, b)$ is defined as follows:

$$h_{ij} = \begin{cases} 1, & \text{if } W_i \subseteq V_j \\ 0, & \text{if } W_i \nsubseteq V_j. \end{cases} \quad (7)$$

A. A GENERAL CONSTRUCTION OF BINARY (r, t)-LRCs

When $a = b - 1$, the matrix $H_q(m, a, b)$ can be regarded as a parity check matrix of an LRC with availability. We have the following theorem.

Theorem 1: The code $\mathcal{C}$ with $H_q(m, b-1, b)$ as a parity check matrix is a binary $[n, k, d]$ LRC with locality r and availability t where $n = \begin{bmatrix} m \\ b \end{bmatrix}_q$, $k \geq \begin{bmatrix} m \\ b \end{bmatrix}_q - \begin{bmatrix} m \\ b-1 \end{bmatrix}_q$, $r = \begin{bmatrix} m-b+1 \\ 1 \end{bmatrix}_q - 1$, $t = \begin{bmatrix} b \\ 1 \end{bmatrix}_q$.

Proof: Since the matrix $H_q(m, b-1, b)$ contains $\begin{bmatrix} m \\ b-1 \end{bmatrix}_q$ rows and $\begin{bmatrix} m \\ b \end{bmatrix}_q$ columns, we can see the code has length $n = \begin{bmatrix} m \\ b \end{bmatrix}_q$, and dimension $k \geq \begin{bmatrix} m \\ b \end{bmatrix}_q - \begin{bmatrix} m \\ b-1 \end{bmatrix}_q$. Each row of the matrix $H_q(m, b-1, b)$ is associated with a $(b-1)$-dimensional subspace W on $\mathbb{F}_q^m$. There are $\begin{bmatrix} m-b+1 \\ 1 \end{bmatrix}_q$ b-dimensional subspaces containing W, so each row of $H_q(m, b-1, b)$ has $\begin{bmatrix} m-b+1 \\ 1 \end{bmatrix}_q$ 1s. Each column of $H_q(m, b-1, b)$ is associated with a b-dimensional subspace V on $\mathbb{F}_q^m$. Each V contains $\begin{bmatrix} b \\ b-1 \end{bmatrix}_q = \begin{bmatrix} b \\ 1 \end{bmatrix}_q$ $(b-1)$-dimensional subspace, so each column of $H_q(m, b-1, b)$ has $\begin{bmatrix} b \\ 1 \end{bmatrix}_q$ 1s. Therefore, $H_q(m, b-1, b)$ is an incidence matrix of a 1-$\left(\begin{bmatrix} m \\ b \end{bmatrix}_q, \begin{bmatrix} m-b+1 \\ 1 \end{bmatrix}_q, \begin{bmatrix} b \\ 1 \end{bmatrix}_q\right)$-design.

Then for any i-th column associated with a b-dimensional subspace V_i, $i \in \left[\begin{bmatrix} m \\ b \end{bmatrix}_q\right]$. Since V_i contains $\begin{bmatrix} b \\ 1 \end{bmatrix}_q$ $(b-1)$-dimensional subspaces, there are $\begin{bmatrix} m \\ b \end{bmatrix}_q$ rows which have 1 in this column. We claim that excluding the coordinate i, supports of these $\begin{bmatrix} m \\ b \end{bmatrix}_q$ rows are pairwise disjoint. Otherwise, assume there are two rows, say the j-th row (denoted as h_j, associated with the subspace W_j) and the l-th row (denoted as h_l, associated with the subspace V_l), such that $\{i, u\} \subseteq \mathrm{supp}(h_j) \cap \mathrm{supp}(h_l)$ for some $u \in \left[\begin{bmatrix} m \\ b \end{bmatrix}_q\right] \setminus \{i\}$. It implies that $W_j \subseteq V_i \cap V_u$ and $W_l \subseteq V_i \cap V_u$. Then we get $W_j \cup W_l \subseteq V_i \cap V_u$. But the union of two different $(b-1)$-dimensional subspaces is of dimension at least b, the intersection of two different b-dimensional subspaces is of dimension at most b, which leads to a contradiction. Therefore, the matrix $H_q(m, b-1, b)$ satisfies the conditions of Lemma 1, and this completes the proof. □

Lemma 2: The code $\mathcal{C}$ with $H_q(m, b-1, b)$ as a parity check matrix has minimum distance $d \geq \begin{bmatrix} b \\ 1 \end{bmatrix}_q + 1$.

Proof: It is sufficient to show that any $\begin{bmatrix} b \\ 1 \end{bmatrix}_q$ columns of $H_2(m, b-1, b)$ are linearly independent. In the binary case, it is equivalent to show that the sum of any s columns of $H_2(m, b-1, b)$ is not $\mathbf{0}$ for all $1 \leq s \leq \begin{bmatrix} b \\ 1 \end{bmatrix}_q$. W.L.G. let us consider the first s columns. We denote the b-dimensional subspace corresponding to i-th column of $H_2(m, b-1, b)$ as V_i, and $(b-1)$-dimensional subspace corresponding to j-th row of $H_2(m, b-1, b)$ as W_j. Note that $V_1 \cap V_i$ is at most a $(b-1)$-dimensional subspace of $\mathbb{F}_q^m$ for all $i = 2, 3, \cdots, s$ and there are $\begin{bmatrix} b \\ 1 \end{bmatrix}_q$ $(b-1)$-dimensional subspace in V_1. We can find a $(b-1)$-dimensional subspace, say W_1 such that $W_1 \neq V_1 \cap V_i$ for $i = 2, 3, \cdots, s$ (this is possible because $s - 1 < \begin{bmatrix} b \\ 1 \end{bmatrix}_q$). This implies $W_1 \subseteq V_1, W_1 \nsubseteq V_i$ for all $i = 2, 3, \cdots, s$. Hence the sum of the first s columns of $H_2(m, b-1, b)$ is not 0. □

B. CONSTRUCTION OF BINARY $(r, 3)$-LRCs

When $q = 2, a = 1, b = 2$, the parity check matrix $H_2(m, 1, 2)$ has some special properties. We refer to this construction as *Construction 1*.

To see its properties more directly, we need to sort all the subspaces in the following manner. It is well known that every k-dimensional subspace of $\mathbb{F}_q^n$ is the row space of a $k \times n$ matrix of rank k, so we can use reduced row echelon form (RREF) to represent each subspace. Then take out each row of the matrix separately, assembled into a 0-1 sequence in row order, and sort the sequence in the lexical order to get the order of the subspaces.

Remark 1: The matrix $H_2(m, 1, 2)$ can be viewed in the following way. Each row of $H_2(m, 1, 2)$ is associated with a 1-dimensional subspace of $\mathbb{F}_2^m$, which only contains one vector of length m, so it can be seen as a binary number of length m. Similarly, each column of $H_2(m, 1, 2)$ is associated with a 2-dimensional subspace of $\mathbb{F}_2^m$. Since a 2-dimensional vector space contains three vectors, each vector can be seen as a binary number of length m, where the first two vectors are the first and second row of the $2 \times n$ RREF matrix. The third vector is the sum of the first two vectors (under $\mathbb{F}_2$). If a binary number corresponding to the ith row is contained in the three binary numbers of the jth column, the value at the ith row and jth column of the matrix $H_2(m, 1, 2)$ is 1, otherwise 0.

Below is an example of $H_2(m, 1, 2)$.

Example 1: Suppose $m = 3$, the code $\mathcal{C}$ with $H_2(3, 1, 2)$ (Fig.1) as a parity check matrix is a $(2, 3)$-LRC with length $n = 7$, dimension $k = 3$, minimum distance $d = 4$. We label all vectors in the vector space at the beginning of rows and columns (regard a vector as a binary number and convert it to a decimal number).

$$H_2(3,1,2) = \begin{array}{c|ccccccc} & 2 & 4 & 4 & 4 & 5 & 5 & 6 \\ & 1 & 1 & 2 & 3 & 2 & 3 & 1 \\ & 3 & 5 & 6 & 7 & 7 & 6 & 7 \\ \hline 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 2 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 3 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 4 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 5 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 6 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 7 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{array}$$

FIGURE 1. The matrix $H_2(3, 1, 2)$.

The above example has the same parameters as the Simplex code with $m = 3$. Next we prove some properties of $H_2(m, 1, 2)$ to help understand the code $\mathcal{C}$.

Lemma 3: For any positive integer $m \geq 3$, the matrix $H_2(m, 1, 2)$ is of the block form

$$H_2(m, 1, 2) = \begin{pmatrix} H_2(m-1, 1, 2) & * \\ \mathbf{0} & A \end{pmatrix}, \quad (8)$$

where

$$A = \begin{pmatrix} 11\cdots 1 & * \\ I_{\left[{m-1 \atop 1}\right]_2} & * \end{pmatrix},$$

$I_{\left[{m-1 \atop 1}\right]_2}$ *is the identity matrix of size* $\left[{m-1 \atop 1}\right]_2$, $\mathbf{0}$ *is a zero matrix.* $*$ *represent arbitrary matrix over* $\mathbb{F}_2$. *Particularly,* $H_2(2, 1, 2) = (1, 1, 1)^\tau$.

Proof: It is easy to see $H_2(2, 1, 2) = (1, 1, 1)^\tau$. For the case $m \geq 3$, noticed that the RREF matrices corresponding to the former $\left[{m-1 \atop 1}\right]_2 = 2^{m-1} - 1$ rows of the matrix $H_2(m, 1, 2)$ all have 0s as their first entries, thus these rows can be regarded as 1-dimensional subspaces of $(m-1)$-dimensional space. Since the subspaces associated with columns are sorted, the subspaces correspond to the RREF matrices whose first column is all 0s must be at the top, and these columns can be regarded as 2-dimensional subspaces of $(m-1)$-dimensional space. There are a total of $\left[{m-1 \atop 2}\right]_2$ of such subspaces. Therefore, the upper left block of $H_2(m, 1, 2)$ (i.e. the former $\left[{m-1 \atop 1}\right]_2$ rows and the former $\left[{m-1 \atop 2}\right]_2$ columns) is the matrix $H_2(m-1, 1, 2)$. The RREF matrices corresponding to the latter 2^{m-1} rows of the matrix $H_2(m, 1, 2)$ all have 1s as their first entries, but the RREF matrices corresponding to the former $\left[{m-1 \atop 2}\right]_2$ columns of the matrix $H_2(m, 1, 2)$ all have 0s as their first entries. As a result, the bottom left block is a $2^{m-1} \times \left[{m-1 \atop 2}\right]_2$ zero submatrix.

Regarding the matrix A, note that the RREF matrix of the first row of A are $(1, 0, \cdots, 0)$, which has $m-1$ zeros. Since the subspaces are sorted, the subspaces correspond to the RREF matrices whose first row is $(1, 0, \cdots, 0)$ must be ranked after the RREF matrices whose first column is 0, before other subspaces. There are a total of $\left[{m-1 \atop 1}\right]_2$ of such subspaces. Therefore, the former $\left[{m-1 \atop 1}\right]_2$ columns of the first row of matrix A are all 1s. Moreover, if we add the two rows of the RREF matrix corresponding to the former $\left[{m-1 \atop 1}\right]_2$ columns of matrix A, we can get the vectors whose binary representation from $(1, 0, \cdots, 1)$ to $(1, 1, \cdots, 1)$ (decimal representation from $2^{m-1} + 1$ to $2^m - 1$). The vectors corresponding to the rows start from the second row of matrix A whose binary representation also from $(1, 0, \cdots, 1)$ to $(1, 1, \cdots, 1)$. Therefore, the bottom left block of matrix A is an identity matrix of size $\left[{m-1 \atop 1}\right]_2$. □

According to the block form of the matrix $H_2(m, 1, 2)$, it is easy to see $\operatorname{rank}(H_2(m, 1, 2)) \geq \sum_{i=1}^{m-1} \left[{i \atop 1}\right]_2 = 2^m - 1 - m$. In fact, the rows of $H_2(m, 1, 2)$ are linearly dependant, so some rows can be deleted. We define the following set.

$$E_i = \left\{(2j-1)2^i + k : 1 \leq j \leq 2^{m-i-1}, 0 \leq k \leq 2^i - 1\right\}. \tag{9}$$

From the definition of the set E_i, if we fix $m = 4$, we can get,

$$E_0 = \{1, 3, 5, \cdots, 15\},$$
$$E_1 = \{2, 3, 6, 7, 10, 11, 14, 15\},$$
$$E_2 = \{4, 5, 6, 7, 12, 13, 14, 15\},$$
$$E_3 = \{8, 9, 10, 11, 12, 13, 14, 15\}.$$

If all the elements in the E_i are converted into binary form, we will find that the first digit of all elements in E_0 is 1, that is, all odd numbers; the second digit of all the elements in E_1 is 1; the third digit of all elements in E_2 is 1; the fourth digit of all elements in E_3 is 1. In fact, E_i is the set of elements in the binary form of all elements in the complete set $U = \{1, 2, \cdots, 2^m - 1\}$ whose $(i+1)$th digit is 1.

For convenience, we also define

$$R_i = \{j\text{th row of } H_2(m, 1, 2) : j \in E_i\}$$

Lemma 4: The rank of matrix $H_2(m, 1, 2)$ *is* $2^m - 1 - m$.

Proof: In the following, we show that each 2^i-th row in matrix $H_2(m, 1, 2)$ is an $\mathbb{F}_2$-linear combination of all rows in R_i for $i = 0, 1, \cdots, m-1$.

Theorem 1 has showed that every column of matrix $H_2(m, 1, 2)$ has three 1s. It is sufficient to show that the number of 1 in any column of the rows in R_i is 0 or 2. Since the subspaces are sorted in lexical order, the binary representation of the vector corresponding to each row can be regarded as row number, and the $(i+1)$-th digits of the binary representation of all elements in the set E_i are all 1. For any columns in R_i, say j-th column, $j \in \{1, 2, \cdots, \left[{m \atop 2}\right]_2\}$. Suppose the number of 1 in j-th column is 1 (resp. 3), this means there is only one (resp. three) vector whose $(i+1)$-th coordinate is 1 in the subspace corresponding to the j-th column, which is impossible. Because the $(i+1)$-th coordinate of the remaining two vectors in the subspace is 0 (resp. 1), so the $(i+1)$-th coordinate of the third vector obtained by adding these two vectors must also be 0, which leads to a contradiction, from which the result follows. □

Lemma 5: When $m \geq 3$, *the code* $\mathcal{C}$ *which has the parity check matrix* $H_2(m, 1, 2)$ *has minimum distance* $d = 4$.

Proof: From Lemma 3, the upper left block of the matrix $H_2(m, 1, 2)$ contains $H_2(3, 1, 2)$ for $m \geq 3$, and the bottom left block is a zero matrix, The matrix $H_2(3, 1, 2)$ has 4 columns that are linearly dependant (see Fig.1). So there are 4 columns in the matrix $H_2(m, 1, 2)$ that are linearly dependant. Combine with the Lemma 2, the result follows. □

Corollary 2: The code $\mathcal{C}$ *with* $H_2(m, 1, 2)$ *as a parity check matrix is a binary* $[n, k, d]$*-LRC with locality* r *and availability* t *where* $n = \left[{m \atop 2}\right]_2$, $k = \left[{m \atop 2}\right]_2 - \left[{m \atop 1}\right]_2 + m$, $r = 2^{m-1} - 2$, $t = 3$, $d = t + 1 = 4$. *Therefore, the code is an optimal LRC with availability which reaches the bound (3).*

C. CONSTRUCTIONS OF (r, t, x)-LRCs

When $q = 2, a = 1, b \geq 3$, we can get the LRC with availability in which the recovering sets can intersect in a small number of coordinates. We refer to this construction as *Construction 2*.

Let us give an example of $H_2(m, 1, b)$.

Example 2: Suppose $m = 4$, the code $\mathcal{C}$ which has the parity check matrix $H_2(4, 1, 3)$ is a $(6, 7, 2)$-LRC with $n = 15$, $k = 10$, $d = 4$.

$$\begin{pmatrix}
1&1&1&0&0&0&0&1&0&0&0&0&1&1&1\\
1&1&0&1&0&1&0&0&1&0&1&0&0&0&1\\
1&1&0&0&1&0&1&0&0&1&0&1&0&0&1\\
1&0&1&1&1&0&0&0&1&1&0&0&1&0&0\\
1&0&1&0&0&1&1&0&0&0&1&1&1&0&0\\
1&0&0&1&0&0&1&1&1&0&0&1&0&1&0\\
1&0&0&0&1&1&0&1&0&1&1&0&0&1&0\\
0&1&1&1&1&1&1&1&0&0&0&0&0&0&0\\
0&1&1&0&0&0&0&1&1&1&1&1&0&0&0\\
0&1&0&1&0&1&0&0&0&1&0&1&1&1&0\\
0&1&0&0&1&0&1&0&1&0&1&0&1&1&0\\
0&0&1&1&1&0&0&0&0&0&1&1&0&1&1\\
0&0&1&0&0&1&1&0&1&1&0&0&0&1&1\\
0&0&0&1&0&0&1&1&0&1&1&0&1&0&1\\
0&0&0&0&1&1&0&1&1&0&0&1&1&0&1
\end{pmatrix}$$

Note that the minimum distance of the above code has reached the bound (4).

Theorem 2: The code $\mathcal{C}$ which has the parity check matrix $H_2(m, 1, b)$ is a (r, t, x)-LRC for $b \geq 3$, where $r = \begin{bmatrix} m-1 \\ b-1 \end{bmatrix}_2 - 1$, $t = \begin{bmatrix} b \\ 1 \end{bmatrix}_2$, $x = \begin{bmatrix} m-2 \\ b-2 \end{bmatrix}_2 - 1$.

Proof: Each row of matrix $H_2(m, 1, b)$ is associated with a 1-dimensional subspace W on $\mathbb{F}_2^m$. There are $\begin{bmatrix} m-1 \\ b-1 \end{bmatrix}_2$ b-dimensional subspaces containing W, so each row of $H_2(m, 1, b)$ has $\begin{bmatrix} m-1 \\ b-1 \end{bmatrix}_2$ 1s. Each column of $H_2(m, 1, b)$ is associated with a b-dimensional subspace V on $\mathbb{F}_2^m$. Each V contains $\begin{bmatrix} b \\ 1 \end{bmatrix}_2$ 1-dimensional subspace, so each column of $H_2(m, 1, b)$ has $\begin{bmatrix} b \\ 1 \end{bmatrix}_2$ 1s. Therefore, $H_2(m, 1, b)$ is an incidence matrix of a 1-$\left(\begin{bmatrix} m \\ b \end{bmatrix}_2, \begin{bmatrix} m-1 \\ b-1 \end{bmatrix}_2, \begin{bmatrix} b \\ 1 \end{bmatrix}_2\right)$-design.

For any two rows of $H_2(m, 1, b)$, say ith row and jth row, which corresponds to the subspace W_i and W_j respectively, for $1 \leq i < j \leq \left[\begin{bmatrix} m \\ 1 \end{bmatrix}_2\right]$. $W_i \cap W_j$ is a 2-dimensional subspace. For any column of $H_2(m, 1, b)$, there are $\begin{bmatrix} m-2 \\ b-2 \end{bmatrix}_2$ b-dimensional subspace which contains such 2-dimensional subspace $W_i \cap W_j$, that is the supports of any two rows of $H_2(m, 1, b)$ intersect at $\begin{bmatrix} m-2 \\ b-2 \end{bmatrix}_2$ coordinates, i.e. $x = \begin{bmatrix} m-2 \\ b-2 \end{bmatrix}_2 - 1$. Therefore, the matrix $H_2(m, 1, b)$ satisfies the conditions of Corollary 1, and it complete the proof. □

Then we give a brief analysis of the structure of the matrix $H_2(m, 1, b)$ for $b \geq 3$.

Theorem 3: For any positive integer $3 \leq b < m$, the matrix $H_2(m, 1, b)$ is of the block form

$$H_2(m, 1, b) = \begin{pmatrix} H_2(m-1, 1, b) & * \\ \mathbf{0} & * \end{pmatrix}, \tag{10}$$

where $\mathbf{0}$ is a zero matrix, $*$ represent arbitrary matrix over $\mathbb{F}_2$. In particular $H_2(b, 1, b) = (1, \cdots, 1)^\tau$ is a column vector of length $\begin{bmatrix} b \\ 1 \end{bmatrix}_2$.

Proof: The proof is similar to the proof of Theorem 3, so we omit it here. □

Next, we generalize the method of proving the rank of $H_2(m, 1, 2)$ to prove the rank of $H_2(m, 1, b \geq 3)$. Since all the subspaces are defined over $\mathbb{F}_2$, for the convenience of proof, we also view the vector in the subspace as a binary number.

Lemma 6: For any m-dimensional vector space V over $\mathbb{F}_2$. Let S be a set of column indices such that $1 \leq |S| \leq m - 1$. Then $|\{v \in V : v_i = 1, \forall i \in S\}| \equiv 0 \mod 2$.

Proof: Let $n \geq m$ be the vector length of V. We represent V as a RREF matrix G of size $m \times n$. Denote G_S as a submatrix formed by the columns of indices in S. We transform G to a matrix G', such that G_S is in its RREF. If k is the rank of G_S, $0 \leq k \leq |S|$, then the last $(m-k)$ rows of G'_S are all zero vectors. Notice that the vector space generated by the first k rows of G'_S contains at most one all 1s vector. Now, consider the vector space generated by the last $(m - k)$ rows of G', its cardinality is 2^{m-k}, and all vectors have 0 entries at the columns of indices in S. Therefore, the vector space generated by G contains either 0 or 2^{m-k} vectors that have all 1s in the columns of indices in S. From which the result follows. □

In fact, rows of the matrix $H_2(m, 1, b)$ are linearly dependant, so some rows can be deleted. We define the following set consisting of s-tuples of positive integers:

$$C_s^m = \{(\alpha_1, \cdots, \alpha_s) : 1 \leq \alpha_1 < \cdots < \alpha_s \leq m\} \tag{11}$$

Actually, C_s^m is a set of s-combination of $[m]$, $\left|C_s^m\right| = \binom{m}{s}$.

Let E_j^s be a set of all the m-bit binary number whose i-th bit is 1, for all $i \in (\alpha_1, \alpha_2, \cdots, \alpha_s)_j$, where $(\alpha_1, \alpha_2, \cdots, \alpha_s)_j$ is the j-th tuple in C_s^m. Let $(E_j^s)_{min}$ be the smallest element in E_j^s. Obviously, when s traverses from 1 to $b-1$, all these $(E_j^s)_{min}$ are different and $\sum_{s=1}^{b-1} \left|E_j^s\right| = \sum_{s=1}^{b-1} \binom{m}{s}$. For convenience, we also define

$$R_j^s = \left\{i - \text{th row of } H_2(m, 1, b) : i \in E_j^s\right\}$$

Lemma 7: For any $b \geq 3$, $rank(H_2(m, 1, b)) \leq \sum_{s=1}^{b-1} \binom{m}{s}$. Therefore the code $\mathcal{C}$ which has the parity check matrix $H_2(m, 1, b)$ has dimension $k \geq \begin{bmatrix} m \\ b \end{bmatrix}_2 - \begin{bmatrix} m \\ 1 \end{bmatrix}_2 + \sum_{i=1}^{b-1} \binom{m}{i}$.

Proof: In the following, we show that each $(E_j^s)_{min}$-th row in matrix $H_2(m, 1, b)$ is an $\mathbb{F}_2$-linear combination of rall rows in R_j^s for $j = 1, 2, \cdots, \binom{m}{s}$ and $s = 0, 1, \cdots, b-1$.

The matrix $H_2(m, 1, b)$ is defined over $\mathbb{F}_2$, it is sufficient to show that the number of 1 in any column of the rows in R_j^s is even. Since the subspaces are sorted in lexical order, the binary representation of the vector corresponding to each row can be regarded as a row number. According to Lemma 6 and the definition of E_j^s, for each row in R_j^s, its i-th column is 1 for all $i \in (\alpha_1, \alpha_2, \cdots, \alpha_s)_j$, where $(\alpha_1, \alpha_2, \cdots, \alpha_s)_j$ is the j-th tuple in C_s^m. Therefore, the number of 1 in any column of the rows in R_j^s is even. Moreover, because all these $(E_j^s)_{min}$ are different, we can sort $(E_j^s)_{min}$ in increasing order, then all $(E_j^s)_{min}$-th rows of the $H_2(m, 1, b)$ can be deleted in this order. From which the result follows. □

Corollary 3: The code $\mathcal{C}$ which has the parity check matrix $H_2(m, 1, b)$ has dimension

$$k \geq \begin{bmatrix} m \\ b \end{bmatrix}_2 - \begin{bmatrix} m \\ 1 \end{bmatrix}_2 + \sum_{i=1}^{b-1} \binom{m}{i}. \tag{12}$$

so the code rate is,

$$\mathcal{R} \geq 1 - \frac{\begin{bmatrix} m \\ 1 \end{bmatrix}_2 - \sum_{i=1}^{b-1} \binom{m}{i}}{\begin{bmatrix} m \\ b \end{bmatrix}_2}. \quad (13)$$

Remark 2: In fact, if we use the method similar to the analysis of matrix $H_2(m, 1, 2)$ to deeply analyze the block structure of $H_2(m, 1, b)$, we can get that the rank of the matrix $H_2(m, 1, b)$ is indeed $\sum_{i=1}^{b-1} \binom{m}{i}$.

Lemma 8: The code $\mathcal{C}$ which has the parity check matrix $H_2(m, 1, b)$ has minimum distance $d \geq 4$.

Proof: It is obvious that any two columns of the matrix cannot be equal. It is sufficient to show that the sum of any 3 columns of $H_2(m, 1, b)$ is not $\mathbf{0}$. We denote the b-dimensional subspace corresponding to ith column of $H_2(m, 1, b)$ as V_i. Note that any two different subspaces $V_i \cap V_j$ is at most a $(b-1)$-dimensional subspace of $\mathbb{F}_2^m$. And a $(b-1)$-dimensional subspace contains $\begin{bmatrix} b-1 \\ 1 \end{bmatrix}_2 = 2^{b-1} - 1$ 1-dimensional subspaces. So the supports of any two columns of $H_2(m, 1, b)$ can intersect at most $2^{b-1} - 1$ coordinates. However, the number of 1 in any column of $H_2(m, 1, b)$ is

$$\begin{bmatrix} b \\ 1 \end{bmatrix}_2 = 2^b - 1 > 2 \times (2^{b-1} - 1) = 2 \times \begin{bmatrix} b-1 \\ 1 \end{bmatrix}_2$$

This means that the sum of any 3 columns cannot be $\mathbf{0}$, from which the result follows. □

IV. COMPARISON WITH OTHER CONSTRUCTIONS

A. GENERAL CONSTRUCTION

Our general construction is a binary regular LDPC code with *girth* > 4. Hao *et al.*'s [25] proposed a construction of LRC codes with information symbols by combining an existing regular LDPC and an identity matrix. But we directly construct the parity-check matrix to obtain an LRC code, and this matrix can be viewed as a incidence matrix of BIBD with $\lambda = 1$, so our construction can also be regarded as a kind of BIBD-LDPC codes.

B. CONSTRUCTION 1

Among (r, t)-LRCs that have the same availability $t = 3$ as our Construction 1, WZL code is the one that has good parameters. It has been shown that WZL code has a higher rate than that of direct product code and Prakash *et al.*'s construction [3].

Recall that the parameters of a WZL code are $n_w = \binom{r+t}{t}$, $d_w = t + 1$, $\mathcal{R}_w = \frac{r}{r+t}$. For all $r > 0$, the code length of Construction 1 is $n_1 = \frac{(r+1)(2r+3)}{3}$, which is shorter than that of WZL code $n_w = \binom{r+3}{3}$. Both constructions have the same minimum distance. To compare the code rate, Construction 1 is rate optimal. Indeed, it is easy to see that we always have a greater code rate than $\mathcal{R}_w$ for $r > 0$.

C. CONSTRUCTION 2

There are few works on constructions of (r, t, x)-LRCs. Kruglik *et al.*'s construction is based on WZL code, its parameters are $n_k = (x+1)\binom{r+t}{t}$, $d_k = 2$, $r_k = (r+1)(x+1) - 1$, $\mathcal{R}_k = \frac{r+(t-1)x}{r+t+(t-1)x}$, so our Construction 2 has a much shorter code length and 2× greater minimum distance than Kruglik's construction. We also compare our code rate to that of Kruglik *et al.* and WZL code with the same locality r and availability t, as shown in Fig. 2 and Fig. 3.

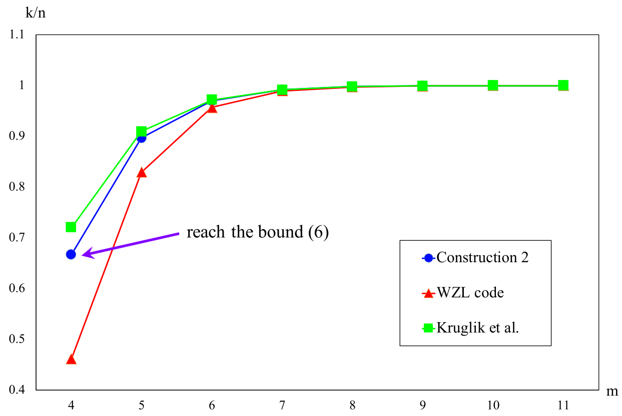


FIGURE 2. Comparison of code rate for $b = 3, 4 \leq m \leq 11$.

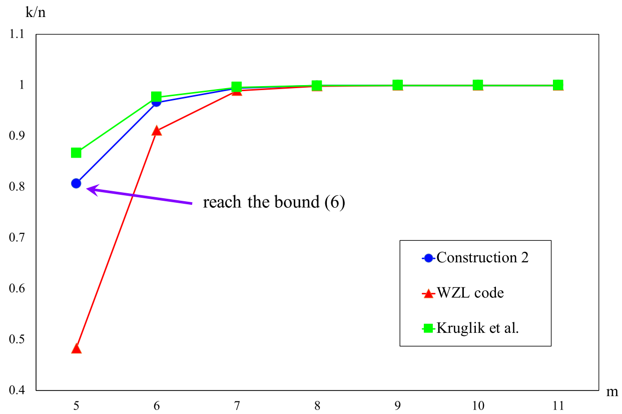


FIGURE 3. Comparison of code rate for $b = 4, 5 \leq m \leq 11$.

The figures show our code rate are greater than WZL code, but slightly less than Kruglik *et al.* Moreover, when $m = 4, b = 3$ and $m = 5, b = 4$, our construction reaches the minimum distance bound (4) of (r, t, x)-LRC codes. Note that since the matrix $H_2(m, 1, b)$ has the above-mentioned block form (see Theorem 3), when b is fixed, the minimum distance is also fixed, which is the same as the code has parity check matrix $H_2(b + 1, 1, b)$.

V. CONCLUSION

In this paper, we generalize the construction of WZL codes and propose two constructions of LRC codes. Construction 1 can produce optimal (r, t)-LRCs, which can reach the upper bound of code rate (3). Construction 2 has much higher rate than that of WZL codes and attain the upper bound on minimum distance (4) in two special cases. Moreover, we give a sufficient condition for a 1-design's incidence matrix that can be the parity-check matrix of a (r, t, x)-LRCs.

REFERENCES

[1] S. Kruglik, K. Nazirkhanova, and A. Frolov, "On distance properties of (r, t, x)-LRC codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1336–1339.

[2] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6925–6934, Nov. 2012.

[3] 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.

[4] A. Wang and Z. Zhang, "An integer programming-based bound for locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5280–5294, Oct. 2015.

[5] J. Zhang, X. Wang, and G. Ge, "Some improvements on locally repairable codes," 2015, *arXiv:1506.04822*. [Online]. Available: https://arxiv.org/abs/1506.04822

[6] M. Mehrabi and M. Ardakani, "On minimum distance of locally repairable codes," in *Proc. 15th Can. Workshop Inf. Theory (CWIT)*, Jun. 2017, pp. 1–5.

[7] V. R. Cadambe and A. Mazumdar, "Bounds on the size of locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 5787–5794, Nov. 2015.

[8] V. Cadambe and A. Mazumdar, "An upper bound on the size of locally recoverable codes," in *Proc. Int. Symp. Netw. Coding (NetCod)*, Jun. 2013, pp. 1–5.

[9] J. Ma and G. Ge, "Optimal binary linear locally repairable codes with disjoint repair groups," *SIAM J. Discrete Math.*, vol. 33, no. 4, pp. 2509–2529, Jan. 2019.

[10] A. Agarwal, A. Barg, S. Hu, A. Mazumdar, and I. Tamo, "Combinatorial alphabet-dependent bounds for locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3481–3492, May 2018.

[11] I. Tamo, A. Barg, S. Goparaju, and R. Calderbank, "Cyclic LRC codes, binary LRC codes, and upper bounds on the distance of cyclic codes," *Int. J. Inf. Coding Theory*, vol. 3, no. 4, pp. 345–364, 2016.

[12] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4661–4676, Aug. 2014.

[13] A. Barg, K. Haymaker, E. W. Howe, G. L. Matthews, and A. Várilly-Alvarado, "Locally recoverable codes from algebraic curves and surfaces," in *Algebraic Geometry for Coding Theory and Cryptography*. Cham, Switzerland: Springer, 2017, pp. 95–127.

[14] X. Li, L. Ma, and C. Xing, "Optimal locally repairable codes via elliptic curves," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 108–117, Jan. 2019.

[15] Y. Luo, C. Xing, and C. Yuan, "Optimal locally repairable codes of distance 3 and 4 via cyclic codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 1048–1053, Feb. 2019.

[16] L. Jin, "Explicit construction of optimal locally recoverable codes of distance 5 and 6 via binary constant weight codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4658–4663, Aug. 2019.

[17] S. Yang, R. Li, Q. Fu, and J. Lv, "New constructions of short length binary locally repairable codes," *IEEE Access*, vol. 8, pp. 41282–41287, 2020.

[18] A. S. Rawat, D. S. Papailiopoulos, A. G. Dimakis, and S. Vishwanath, "Locality and availability in distributed storage," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4481–4493, Aug. 2016.

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

[20] I. Tamo, A. Barg, and A. Frolov, "Bounds on the parameters of locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3070–3083, Jun. 2016.

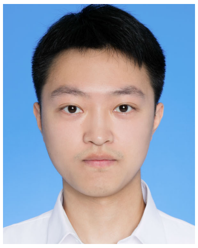
[21] S. Kadhe and R. Calderbank, "Rate optimal binary linear locally repairable codes with small availability," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 166–170.

[22] S. Kruglik, M. Dudina, V. Potapova, and A. Frolov, "On one generalization of LRC codes with availability," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2017, pp. 26–30.

[23] A. Wang, Z. Zhang, and M. Liu, "Achieving arbitrary locality and availability in binary codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 1866–1870.

[24] A. Wang, Z. Zhang, and D. Lin, "Two classes of (r, t)-locally repairable codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 445–449.

[25] 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.



JIAMING TENG received the B.A. degree in computer science from Soochow University, China, in 2017. He is currently a Student with the School of Computer Science, Fudan University, China. His research interest includes coding theory in data storage.

LINGFEI JIN (Member, IEEE) received the B.A. degree in mathematics from the Hefei University of Technology, China, in 2009, and the Ph.D. degree from Nanyang Technological University, Singapore. She is currently an Associate Professor with the School of Computer Science, Fudan University, China. Her research interests include quantum error correcting codes, codes for distributed storage, and cryptography.