

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Optimal Locally Repairable Codes for Parallel Reading

JIE HAO¹, KENNETH W. SHUM², SHU-TAO XIA³ AND DEYIN LI¹

¹Information Security Center, Beijing University of Posts and Telecommunications, Beijing, 100876, China (email: haojie@bupt.edu.cn)

²School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, 518172, China (email: wkshum@cuhk.edu.cn)

³Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China (email: xiast@sz.tsinghua.edu.cn)

Corresponding author: Jie Hao (e-mail: haojie@bupt.edu.cn).

“This research is supported in part by the National Natural Science Foundation of China under grant Nos. 61801049 and 61771273”

ABSTRACT Locally repairable codes (LRCs) have important applications in distributed storage systems. In this paper, we study q -ary $[n, k, d]$ LRCs with (r, t, δ) -information-locality, where each of the i -th ($1 \leq i \leq k$) information symbol is contained in t punctured subcodes with length $\leq r + \delta - 1$, minimum distance δ , and the i -th information symbol is the unique common code symbol of these t subcodes, furthermore, each subcode contains exactly $\delta - 1$ parity symbols. Firstly, an upper bound on the minimum distance of such q -ary LRCs with (r, t, δ) -information-locality is given. Then, we propose a general construction framework of q -ary optimal LRCs with (r, t, δ) -information-locality and minimum distance $d = t(\delta - 1) + 1$, where the required field size is just $q \geq r + \delta - 2$. The proposed optimal LRCs can always repair a failed information node locally in case of at most $t\delta - 1$ node failures. Moreover, multiple repair subcodes can support parallel readings of data, thus make the proposed codes attractive for distributed storage systems with hot data.

INDEX TERMS Locally repairable codes, erasure codes, Singleton-like bound, optimal LRCs.

I. INTRODUCTION

Modern large distributed storage systems usually store redundant data to ensure data reliability in case of storage node failures. Due to the large volume of data, redundancy schemes based on erasure codes become more attractive because of the higher storage efficiency compared to replications. However, when some storage node fails, the repair process of traditional erasure codes, especially maximum distance separable (MDS) codes, usually requires reading large amount of data from surviving nodes. In recent years, *locally repairable codes* (LRCs) [1] which can repair failed nodes efficiently have attracted a lot of interest.

In an $[n, k, d]$ linear code over \mathbb{F}_q , a code symbol is said to have r -locality if it can be repaired by accessing at most r other code symbols. LRCs are linear codes with locality properties for code symbols. For an $[n, k, d]$ linear code with r -locality for information symbols, Gopalan *et al.* proved the well-known Singleton-like bound [1]

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

For optimal code constructions, Tamo and Barg proposed a family of optimal LRCs using polynomial methods [2]. Complete enumerations of all optimal binary LRCs meeting

the bound (1) were given in [3]. Optimal cyclic LRCs were proposed in [4], [5]. Classifications of optimal ternary LRCs were given in [6]. Wang and Zhang proposed a refined bound of LRCs and gave corresponding optimal codes [7]. Optimal LRCs with single parity symbol in each group were studied in [8]. Sometimes multiple node failures occur in distributed storage systems. For a failed node with r -locality, if one of the r repairing nodes also fails, the local recovery can not be accomplished. In order to accomplish local recovery in case of multiple node failures, two parallel generalizations of locality, i.e., (r, t) -locality [9]–[11] and (r, δ) -locality [12], were proposed. The type of (r, t) -locality also has advantages to support parallel readings of data.

A code symbol in an $[n, k, d]$ linear code is said to have (r, t) -locality if there exist t disjoint groups of other symbols, each with size at most r , to repair this symbol. By this definition, a symbol with (r, t) -locality can always be repaired locally in case of $t - 1$ other node failures. Moreover, multiple disjoint repair groups can support parallel readings of data. This functionality is quite important for storage systems with hot data. For an $[n, k, d]$ linear code with (r, t) -locality for information symbols, its minimum distance [9], [11]

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

Constructions of LRCs with (r, t) -locality were given in [9]–[11], [13], [14]. Existing optimal LRCs meeting the bound (2) have low code rate [9]. Specially, Rawat *et al.* proved that when only information symbols have (r, t) -locality and every repair group contains exactly one parity symbol, the minimum distance [11]

$$d \leq n - k - \left\lceil \frac{kt}{r} \right\rceil + t + 1. \quad (3)$$

Corresponding optimal LRCs meeting the bound (3) with relatively high code rate were proposed in [11], [15] and optimal binary LRCs were given in [16], [17].

For the other type of generalization, a code symbol in a linear code is said to have (r, δ) -locality if it is contained in a punctured subcode, which has length at most $r + \delta - 1$ and minimum distance at least δ . When the number of erasures is less than δ , a code symbol with (r, δ) -locality can always be repaired from at most r code symbols. For an $[n, k, d]$ linear code with (r, δ) -locality for information symbols [12],

$$d \leq n - k - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1) + 1. \quad (4)$$

Corresponding optimal LRCs with (r, δ) -locality were given in [12], [18]–[24]. Recently, Cai *et al.* proposed a more generalized concept of locality which combined (r, t) -locality and (r, δ) -locality, it was shown that the bound (2) and (4) are extreme cases of the bound in [25]. However, optimal LRCs attaining the bound given in [25] still have low code rate.

In the following of this paper, we assume the first k symbols c_i , for $i = 1, 2, \dots, k$, in an $[n, k, d]$ linear code \mathcal{C} are information symbols, and the last $n - k$ symbols are parity symbols. We follow the line of [25] to combine the notion of (r, δ) -locality with (r, t) -locality, where we focus on the case that the number of parity symbols in each local subcode is fixed. A linear code is said to have (r, t, δ) -information-locality if the i -th information symbol, for $1 \leq i \leq k$, is contained in t punctured subcodes, each has length at most $r + \delta - 1$, minimum distance δ and their supports only intersect on the i -th coordinate, furthermore, each local subcode contains exactly $\delta - 1$ parity symbols. More precisely, there exist t subsets $\Phi_1, \dots, \Phi_t \subset [n]$ such that

- $i \in \Phi_j, \forall j \in [t]$;
- $|\Phi_j| \leq r + \delta - 1$, and the minimum distance of the subcode \mathcal{C}_{Φ_j} obtained by puncturing code symbols in $[n] \setminus \Phi_j$ equals δ , for all $j \in [t]$;
- $\Phi_j \cap \Phi_s = \{i\}$, for any $j \neq s \in [t]$;
- $|\Phi_j \cap \{k + 1, k + 2, \dots, n\}| = \delta - 1, \forall j \in [t]$.

In this paper, we study the bounds and constructions of $[n, k, d]$ LRCs with (r, t, δ) -information-locality where each local subcode contains exactly $\delta - 1$ parity symbols. Firstly, we give an upper bound on the minimum distance of such $[n, k, d]$ LRCs with (r, t, δ) -information-locality. The proposed bound can reduce to the bound (3) when $\delta = 2$ and reduce to the bound (4) when $t = 1$. Then, we proposed a general construction framework of q -ary optimal LRCs with (r, t, δ) -information-locality by exhibiting their explicit

parity-check matrices. Our optimal LRCs have minimum distance $d = t(\delta - 1) + 1$, which attains the proposed upper bound. The required field size is just $q \geq r + \delta - 2$. The proposed codes can always repair a failed information node locally in case of at most $t\delta - 1$ node failures, and these t repair subcodes of each information symbol can support parallel readings of data, which benefits storage systems with hot data. Moreover, the relatively small field size implies fast encoding, decoding and repairing operations, thus make the code appealing for practical implementations of distributed storage systems with hot data.

The rest of this paper is organized as follows. In Section II, we present the upper bound. Section III gives the optimal code constructions and Section IV concludes the paper.

II. THE UPPER BOUND

Let \mathcal{C} be a q -ary $[n, k, d]$ LRC with (r, t, δ) -information-locality and each local subcode contains exactly $\delta - 1$ parity symbols. In this section, we present an upper bound on the minimum distance of such $[n, k, d]$ LRCs with (r, t, δ) -information-locality.

The upper bound is obtained by analyzing the standard form of the generator matrix of \mathcal{C} according to locality property. Consider its generator matrix in standard form

$$G = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k, \mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_{n-k}], \quad (5)$$

where \mathbf{e}_i , for $i = 1, 2, \dots, k$, is the i -th standard basis vector, and \mathbf{g}_i , for $i = 1, 2, \dots, n - k$, called *parity column*, corresponds to parity symbol. The support $\text{supp}(\mathbf{v})$ of a vector \mathbf{v} is the set of coordinates of its non-zero entries and its weight is defined to be $wt(\mathbf{v}) = |\text{supp}(\mathbf{v})|$. In this section, all generator matrices are in the standard form of (5).

We begin with a simple observation on the structure of these local subcodes in \mathcal{C} from the view of generator matrix.

Proposition 1. *For a local subcode $\hat{\mathcal{C}}$ with length $\rho + \delta - 1$ and minimum distance δ which contains $1 \leq \rho \leq r$ information symbols and $\delta - 1$ parity symbols in \mathcal{C} . Let Ω be the set of the coordinates of these ρ information symbols. Consider these $\rho + \delta - 1$ columns of $\hat{\mathcal{C}}$ in G , denoted by $[\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \dots, \mathbf{e}_{j_\rho}, \mathbf{g}_{i_1}, \mathbf{g}_{i_2}, \dots, \mathbf{g}_{i_{\delta-1}}]$, it must hold that*

$$\text{supp}(\mathbf{g}_{i_j}) = \Omega, \text{ for } 1 \leq j \leq \delta - 1,$$

i.e., all parity columns of $\hat{\mathcal{C}}$ has uniform weight $|\Omega| \leq r$.

Proof. Since $\hat{\mathcal{C}}$ has minimum distance δ , each of these nonzero rows of $[\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_\rho}, \mathbf{g}_{i_1}, \mathbf{g}_{i_2}, \dots, \mathbf{g}_{i_{\delta-1}}]$, i.e., the generator matrix of $\hat{\mathcal{C}}$, has weight at least δ , thus in each nonzero row of the last $\delta - 1$ columns $[\mathbf{g}_{i_1}, \mathbf{g}_{i_2}, \dots, \mathbf{g}_{i_{\delta-1}}]$, all the $\delta - 1$ entries must be nonzero. In other words, each parity column \mathbf{g}_{i_j} ($1 \leq j \leq \delta - 1$) of $\hat{\mathcal{C}}$ has uniform weight $|\Omega| = \rho \leq r$ and $\text{supp}(\mathbf{g}_{i_j}) = \Omega$. \square

In the following, we call the set of such $\delta - 1$ parity columns $\Gamma = [\mathbf{g}_{i_1}, \mathbf{g}_{i_2}, \dots, \mathbf{g}_{i_{\delta-1}}]$ of a local subcode $\hat{\mathcal{C}}$ in the standard form of generator matrix G as its *local-parity-block*. These $\delta - 1$ parity columns in a local-parity-block

are called as *locality-columns*. By Proposition 1, all locality-columns in Γ have the same support. Denote the *support* of a local-parity-block Γ of a local subcode as

$$\text{supp}(\Gamma) = \Omega,$$

where Ω is the set of coordinates of ρ information symbols in \widehat{C} . Each local-parity-block contains a $\rho \times (\delta - 1)$ submatrix with exactly $\rho \cdot (\delta - 1)$ nonzero entries. By the definition of (r, t, δ) -information-locality, in the standard form of generator matrix G , each of the k information symbols c_i ($1 \leq i \leq k$) has t local-parity-blocks $\Gamma_1, \dots, \Gamma_t$ corresponding to its t local subcodes, and these t local-parity-blocks satisfy

- 1) $i \in \text{supp}(\Gamma_m), \forall m \in [t]$;
- 2) $|\text{supp}(\Gamma_m)| \leq r$, and all the locality-columns in Γ_m have the same support $\text{supp}(\Gamma_m)$, for $\forall m \in [t]$;
- 3) $\text{supp}(\Gamma_m) \cap \text{supp}(\Gamma_s) = \{i\}$, for any $m \neq s \in [t]$;
- 4) Γ_m contains $\delta - 1$ columns, for $\forall m \in [t]$.

By investigating the structure of G according to locality properties, we can obtain the following upper bound on the minimum distance of \mathcal{C} .

Theorem 1. For an $[n, k, d]$ LRC \mathcal{C} with (r, t, δ) -information-locality where each local subcode contains exactly $\delta - 1$ parity symbols, its minimum distance

$$d \leq n - k - (\delta - 1) \left\lceil \frac{kt}{r} \right\rceil + t(\delta - 1) + 1. \quad (6)$$

Proof. Let $G = (I_k, B)$ be the standard form of generator matrix of \mathcal{C} , where I_k consists of k information columns and B consists of $n - k$ parity columns including those locality-columns. Next, we investigate the structure of B by rearranging the orders of locality-columns with weight $\leq r$ to the left part of B according to locality property.

Let L be a set of columns in B , initialized by an empty set. For the first information symbol, we add its t local-parity-blocks $\Gamma_1, \dots, \Gamma_t$ of totally $t(\delta - 1)$ parity columns into L . Then for the second information symbol, it also has t local-parity-blocks. Suppose there are already t_2 local-parity-blocks of the second information symbol in L , where $0 \leq t_2 \leq t$, we add its remaining $t - t_2$ local-parity-blocks into L . Repeat this process until all the k information symbols have at least t local-parity-blocks of locality-columns in L . Note that at the end of this process, some information symbols might have more than t local-parity-blocks contained in L , but there must exist at least one information symbol which has exactly t local-parity-blocks in L .

Let s be the number of local-parity-blocks in L . Let $l = s(\delta - 1)$ be the number of locality-columns of these s local-parity-blocks in $L = [\Gamma_1, \dots, \Gamma_s]$. Denote the remaining parity columns in B as $[\mathbf{g}'_{l+1}, \dots, \mathbf{g}'_{n-k}]$. Then, we rearrange the orders of the columns in B as follows,

$$G = [I_k, \overbrace{\Gamma_1, \dots, \Gamma_s}^{\text{locality-columns}}, \overbrace{\mathbf{g}'_{l+1}, \dots, \mathbf{g}'_{n-k}}^{\text{other columns}}]. \quad (7)$$

Let Δ be the number of the nonzero entries in $L = [\Gamma_1, \dots, \Gamma_s]$. Now let us determine the value of Δ from the

view of rows. Each of the i -th ($1 \leq i \leq k$) information symbol has at least t local-parity-blocks in L and each local-parity-block contains exactly $\delta - 1$ locality-columns whose supports cover the i -th information symbol. Hence, each of the i -th ($1 \leq i \leq k$) row of L has weight at least $t(\delta - 1)$, which indicates

$$\Delta \geq k \cdot t(\delta - 1). \quad (8)$$

Then, let us determine the value of Δ from the view of columns. Each local-parity-block Γ_m for $1 \leq m \leq s$ in L has $\delta - 1$ locality-columns. By Proposition 1, each locality-column in Γ_m has uniform weight $|\text{supp}(\Gamma_m)| \leq r$, then each local-parity-block contains at most $r(\delta - 1)$ nonzero entries. There are s local-parity-blocks in L in total. Hence,

$$\Delta \leq s \cdot r(\delta - 1). \quad (9)$$

Combining (8) and (9), we have

$$k \cdot t \leq s \cdot r, \quad (10)$$

i.e.,

$$s \geq \left\lceil \frac{kt}{r} \right\rceil. \quad (11)$$

The number of these locality-columns in L is

$$l = s(\delta - 1) \geq \left\lceil \frac{kt}{r} \right\rceil (\delta - 1). \quad (12)$$

Since there must exist one information symbol which has exactly t local-parity-blocks of locality-columns in L , we know at least one row of the submatrix L has weight $t(\delta - 1)$. Combining that every row in I_k have weight one and each row in $[\mathbf{g}'_{l+1}, \dots, \mathbf{g}'_{n-k}]$ has weight at most $n - k - l$, there must exist a row with weight at most $1 + t(\delta - 1) + (n - k - l)$ in G . Thus, the minimum weight of all codewords, or the minimum distance, of \mathcal{C} satisfies

$$\begin{aligned} d &\leq n - k - l + t(\delta - 1) + 1 \\ &\leq n - k - \left\lceil \frac{kt}{r} \right\rceil (\delta - 1) + t(\delta - 1) + 1, \end{aligned} \quad (13)$$

this completes the proof. \square

We use a simple example to illustrate the above proof idea.

Example 1. For the standard form of generator matrix G of a $[12, 3]$ LRC with $(r = 2, t = 2, \delta = 3)$ -information-locality, each information symbol is contained in two $[4, 2, 3]$ local subcodes which only intersect on this information symbol. Then, each of the three information symbols has two local-parity-blocks, and the nonzero entries of each local-parity-block give a 2×2 submatrix. By rearranging the orders of the parity columns in G , suppose the structure of G is in the following form,

$$G = \begin{pmatrix} 1 & & & & & & & & & & & \\ & 1 & & & & & & & & & & \\ & & 1 & & & & & & & & & \\ & & & \overbrace{\begin{matrix} * & * & * & * \\ * & * & & * & * \end{matrix}}^{\text{locality-columns}} & \overbrace{\begin{matrix} \# & \# & \# \\ \# & \# & \# \\ \# & \# & \# \end{matrix}}^{\text{other columns}} \end{pmatrix},$$

where $*$'s denote nonzero entries and $\#$'s can be zero or nonzero. There are $s = 3$ local-parity-blocks in L and the

number of locality-columns in L is $l = s(\delta - 1) = 6$. For each row in G , the weight in I_3 is one, the weight in L is $t(\delta - 1) = 4$, and the weight in the last part is at most 3. Hence, each row of G has weight at most 8. In fact, according to the bound (6), the minimum distance

$$d \leq n - k - (\delta - 1) \left\lceil \frac{kt}{r} \right\rceil + t(\delta - 1) + 1 = 8,$$

which is upper bounded by the maximal weight of rows in G .

Remark 1. Note the proposed upper bound (6) for LRCs with (r, t, δ) -information-locality is a generalization of the bound (3) for (r, t) -locality and bound (4) for (r, δ) -locality:

- when $\delta = 2$, the proposed upper bound (6) can reduce to the upper bound (3),
- when $t = 1$, the proposed upper bound (6) can reduce to the upper bound (4).

III. OPTIMAL CODE CONSTRUCTIONS

In this section, we propose a general construction framework of q -ary optimal $[n, k, d]$ LRCs with (r, t, δ) -information-locality and minimum distance is $d = t(\delta - 1) + 1$, which attains the upper bound (6). The optimal codes are presented by exhibiting their explicit parity-check matrices.

A binary matrix A is said to be (t, r) -regular if it has uniform column weight t and uniform row weight r . Furthermore, if the supports of any two rows of A intersect on at most one common coordinate, A is said to have girth $g > 4$. There are various constructions of (t, r) -regular matrices with girth $g > 4$. One important class of such matrices are incidence matrices of certain objects from combinatorial designs, e.g., generalized quadrangle, 2-design, or projective plane, etc [26]. Another large class of regular matrices with $g > 4$ are parity-check matrices of regular LDPC codes, e.g., array LDPC codes [27]. A simple connection between regular matrices with $g > 4$ and optimal binary LRCs with (r, t) -locality has been established in [16]. Inspiring by this idea, we employ regular matrices with girth $g > 4$ to construct q -ary optimal LRCs with (r, t, δ) -information-locality.

Construction 1. Let \mathbb{F}_q be a finite field where $q \geq r + \delta - 2$ and α be a primitive element of \mathbb{F}_q . Let $A_{\mu \times \nu}$ be a $\mu \times \nu$ binary (t, r) -regular matrix with uniform column weight t , uniform row weight r , and girth $g > 4$. Let $I_{m \times m}$ be an $m \times m$ identity matrix. Let

$$\hat{H} = [Q_{(\delta-1) \times r}, I_{(\delta-1) \times (\delta-1)}] \quad (14)$$

be a full-rank parity-check matrix of a q -ary $[r + \delta - 1, r, \delta]$ MDS code. Let $A_{\mu(\delta-1) \times \nu}^*$ be obtained from A by replacing the j -th 1 , $1 \leq j \leq r$, in each row of A with the j -th column vector of $Q_{(\delta-1) \times r}$ and the 0 's with $(\delta - 1)$ -dimensional all-zero column vectors. Then, the q -ary linear code \mathcal{C} with

$$H = [A_{\mu(\delta-1) \times \nu}^*, I_{\mu(\delta-1) \times \mu(\delta-1)}] \quad (15)$$

as parity-check matrix is an $[n = \nu + \mu(\delta - 1), k = \nu, d = t(\delta - 1) + 1]$ LRC with (r, t, δ) -information-locality, and each local subcode contains exactly $\delta - 1$ parity symbols.

Remark 2. Note that since $\hat{H} = [Q_{(\delta-1) \times r}, I_{(\delta-1) \times (\delta-1)}]$ in (14) is a full-rank parity-check matrix of a q -ary $[r + \delta - 1, r, \delta]$ MDS code, all entries in the submatrix $Q_{(\delta-1) \times r}$ must be nonzero. There are various choices of \hat{H} . For example, when $q = r + \delta - 2$, \hat{H} can be chosen as the full-rank parity-check matrix of a q -ary $[q + 1, r, q - r + 2]$ double extended Reed-Solomon code [28].

Next, we use a simple example to illustrate the construction framework in Construction 1.

Example 2. Let $q = 4$ and α be a primitive element of \mathbb{F}_4 . Let $r = 3, t = 2, \delta = 3$. Let

$$\hat{H} = \left(\begin{array}{ccc|cc} 1 & 1 & 1 & 1 & 0 \\ 1 & \alpha & \alpha^2 & 0 & 1 \end{array} \right) \quad (16)$$

be a parity-check matrix of a quaternary $[5, 3, 3]$ MDS code, Given a binary $(2, 3)$ -regular matrix with girth $g > 4$ as

$$A_{4 \times 6} = \left(\begin{array}{cccccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{array} \right).$$

We can obtain $A_{8 \times 6}^*$ by replacing the j -th 1 , $1 \leq j \leq 3$, in each row of $A_{4 \times 6}$ with the j -th column of \hat{H} and the 0 's with two-dimensional all-zero column vectors. By Construction 1, the parity-check matrix $H = [A_{8 \times 6}^*, I_{8 \times 8}]$ is

$$\left(\begin{array}{cccccc|cccccccc} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \alpha & \alpha^2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & \alpha & \alpha^2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \alpha & 0 & \alpha^2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & \alpha & \alpha^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right). \quad (17)$$

Then, the linear code \mathcal{C} with H as parity-check matrix is an optimal quaternary $[n = 14, k = 6, d = 5]$ LRC with $(3, 2, 3)$ -information-locality, and each local subcode contains exactly 2 parity symbols.

In the following, we verify the correctness of Construction 1. Clearly, the parity-check matrix H in (15) has full rank and \mathcal{C} has code length $\nu + \mu(\delta - 1)$ and dimension ν . The first ν columns of H correspond to the ν information symbols of \mathcal{C} , and the last $\mu(\delta - 1)$ columns correspond to parity symbols. The parity-check H in (15) is divided into μ row blocks, each containing $\delta - 1$ rows. The i -th, for $1 \leq i \leq \mu$, row block contains the $[(i - 1)(\delta - 1) + 1]$ -th row to the $i(\delta - 1)$ -th row of H . Moreover, the nonzero columns in the i -th row block constitute the q -ary $(\delta - 1) \times (r + \delta - 1)$ matrix \hat{H} in (14), where any $\delta - 1$ columns of \hat{H} are linearly independent. By puncturing the $n - (r + \delta - 1)$ columns of \mathcal{C} corresponding to these all-zero columns of the i -th row block, the resulting q -ary local subcode $\hat{\mathcal{C}}$ with \hat{H} as local parity-check matrix has length $r + \delta - 1$, dimension r and minimum distance δ . Moreover, this local code $\hat{\mathcal{C}}$ contains exactly $\delta - 1$

parity symbols which correspond to the last $\delta - 1$ nonzero columns of \hat{H} . Since $A_{\mu \times \nu}$ has uniform column weight t and $A_{\mu(\delta-1) \times \nu}^*$ is obtained by replacing all the 1's in $A_{\mu \times \nu}$, each of the i -th ($1 \leq i \leq \nu$) information symbol is contained in t such q -ary $[r + \delta - 1, r, \delta]$ local subcodes. Since $A_{\mu \times \nu}$ has girth $g > 4$, it is not hard to see that the supports of these t local subcodes only intersect on the i -th coordinate. Therefore, all the ν information symbols satisfy (r, t, δ) -information-locality, and each local subcode contains exactly $\delta - 1$ parity symbols.

Lemma 1. *The q -ary $[\nu + \mu(\delta - 1), \nu]$ linear code \mathcal{C} in Construction 1 has minimum distance $d = t(\delta - 1) + 1$.*

Proof. By the structure of H in (15), each of the first ν columns in H has uniform weight $t(\delta - 1)$ and the last $\mu(\delta - 1)$ columns constitute an identity matrix $I_{\mu(\delta-1) \times \mu(\delta-1)}$. Choose one column with weight $t(\delta - 1)$ from the first ν columns of H , it is a linear combination of some $t(\delta - 1)$ columns from $I_{\mu(\delta-1) \times \mu(\delta-1)}$. Hence, $d \leq t(\delta - 1) + 1$. It remains to show $d \geq t(\delta - 1) + 1$, i.e., any $t(\delta - 1)$ columns of H are linearly independent.

We divide the subsequent analyses into two cases. If all these $t(\delta - 1)$ columns are from the last $\mu(\delta - 1)$ columns of H , apparently they are linearly independent. Otherwise, suppose that among these $t(\delta - 1)$ columns, there is at least one column ξ with weight $t(\delta - 1)$ which lies in the first ν columns of H . These $t(\delta - 1)$ nonzero entries in ξ are contained in t local parity-check matrices. We partition these $t(\delta - 1)$ nonzero entries of the column vector ξ into t sub-vectors of weight $\delta - 1$, such that each sub-vector of weight $\delta - 1$ is contained in the same local parity-check matrix. Since any $\delta - 1$ columns of a local parity-check matrix are linearly independent, in order to eliminate the $\delta - 1$ nonzero entries in a sub-vector, at least $\delta - 1$ other columns from this local parity-check matrix are needed. Since the t local parity-check matrices containing the coordinate of ξ only intersect on this coordinate, it involves at least $t(\delta - 1)$ other columns to eliminate these t sub-vectors, i.e., all the $t(\delta - 1)$ nonzero entries of ξ . This implies that any $t(\delta - 1)$ columns including ξ must be linearly independent. Combining all the above discussions, \mathcal{C} has minimum distance $d = t(\delta - 1) + 1$. \square

Theorem 2. *The q -ary linear code \mathcal{C} in Construction 1 is an optimal $[n = \nu + \mu(\delta - 1), k = \nu, d = t(\delta - 1) + 1]$ LRC with (r, t, δ) -information-locality, which attains the bound (6).*

Proof. By the above discussion, the $[n = \nu + \mu(\delta - 1), k = \nu]$ linear code \mathcal{C} has (r, t, δ) -information-locality and each local subcode contains exactly $\delta - 1$ parity symbols. By Lemma 1, \mathcal{C} has minimum distance $d = t(\delta - 1) + 1$. Consider the submatrix $A_{\mu(\delta-1) \times \nu}^*$ in H of (15), which contains the first ν columns of H . Let Λ be number of nonzero entries in $A_{\mu(\delta-1) \times \nu}^*$. Since each row of $A_{\mu(\delta-1) \times \nu}^*$ has uniform weight r , $\Lambda = \mu(\delta - 1) \cdot r$. On the other hand, each column of $A_{\mu(\delta-1) \times \nu}^*$ has uniform weight $t(\delta - 1)$. Then $\Lambda = \nu \cdot t(\delta - 1)$. Hence,

$$\mu \cdot r = \nu \cdot t. \quad (18)$$

By the upper bound (6), the minimum distance satisfies

$$\begin{aligned} d &\leq n - k - (\delta - 1) \left\lceil \frac{kt}{r} \right\rceil + t(\delta - 1) + 1 \\ &= [\nu + \mu(\delta - 1)] - \nu - (\delta - 1) \left\lceil \frac{\nu t}{r} \right\rceil + t(\delta - 1) + 1 \\ &= t(\delta - 1) + 1. \end{aligned}$$

Thus, we can see that the minimum distance of \mathcal{C} attains the upper bound (6) with equality, this completes the proof. \square

Note that since A can be chosen as various regular matrices with $g > 4$, the construction framework in Construction 1 is very flexible in general. Different choices of regular matrix A will result in optimal LRCs with different parameters. For example, if we choose A to be the incidence matrices of specific generalized quadrangle or projective plane, we can obtain the following optimal LRCs by Construction 1.

Example 3. *Given a generalized quadrangle $\mathcal{S} = (P, B, I)$ with parameter $s = e$, $t = e^2$, where e is a prime power [26]. It has $v = (e + 1)(e^3 + 1)$ points and $b = (e^2 + 1)(e^3 + 1)$ lines. Let \mathbb{F}_q be a finite field with $q \geq e^2 + \delta - 1$ and ω is a primitive element of \mathbb{F}_q . Let A be the $[(e + 1)(e^3 + 1)] \times [(e^2 + 1)(e^3 + 1)]$ incidence matrix of \mathcal{S} which has uniform row weight $1 + e^2$, uniform column weight $1 + e$ and girth $g > 4$. By using this binary $(1 + e, 1 + e^2)$ -regular matrix A in Construction 1, we can obtain an optimal LRC with $(1 + e^2, 1 + e, \delta)$ -information-locality, the code has length $n = [e^2 + 1 + (\delta - 1)(e + 1)](e^3 + 1)$, dimension $k = (e^2 + 1)(e^3 + 1)$ and minimum distance $d = (1 + e)(\delta - 1) + 1$.*

Example 4. *Given a projective plane $PG(2, 2^a)$ [26] where a is a positive integer. Let A be the $(4^a + 2^a + 1) \times (4^a + 2^a + 1)$ incidence matrix of $PG(2, 2^a)$ which has uniform row weight $2^a + 1$, uniform column weight $2^a + 1$ and girth $g > 4$. Let \mathbb{F}_q be a finite field with $q \geq 2^a + \delta - 1$. By using this binary $(2^a + 1, 2^a + 1)$ -regular matrix A in Construction 1, we can obtain an optimal $[n = \delta(4^a + 2^a + 1), k = 4^a + 2^a + 1, d = (2^a + 1)(\delta - 1) + 1]$ LRC with $(r = 2^a + 1, t = 2^a + 1, \delta)$ -information-locality, which attains the upper bound (6).*

IV. CONCLUSIONS

In this paper, we studied the bounds and constructions q -ary $[n, k, d]$ LRCs with (r, t, δ) -information-locality where each local subcode contains exactly $\delta - 1$ parity symbols. An upper bound on the minimum distance of such q -ary LRCs with (r, t, δ) -information-locality was given, which can reduce to the bound (3) when $\delta = 2$ and reduce to the bound (4) when $t = 1$. Then, we proposed a flexible construction framework of q -ary optimal LRCs with (r, t, δ) -information-locality and minimum distance $d = t(\delta - 1) + 1$, which attains the proposed bound (6) with equality. Our optimal LRCs only require the field size to be $q \geq r + \delta - 2$, which is relatively small and thus implies fast encoding and decoding speed. The proposed optimal LRCs can always repair a failed information node by accessing at most r surviving nodes in case of at most $t\delta - 1$ node failures. Multiple repair subcodes of

each information symbol guarantee that the proposed codes possess the property to support parallel readings of data, thus are promising for practical implementations of distributed storage systems with hot data.

References

- [1] 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.
- [2] 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.
- [3] J. Hao, S.-T. Xia, and B. Chen, "Some results on optimal locally repairable codes," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 440–444.
- [4] I. Tamo, A. Barg, S. Goparaju, and R. Calderbank, "Cyclic LRC codes and their subfield subcodes," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, China, Jun. 2015, pp. 1262–1266.
- [5] 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.
- [6] J. Hao, S.-T. Xia, and B. Chen, "On optimal ternary locally repairable codes," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Jun. 2017, pp. 171–175.
- [7] 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.
- [8] Y. Lu, J. Hao, and S.-T. Xia, "On the single-parity locally repairable codes," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E100.A, pp. 1342–1345, 06 2017.
- [9] A. Wang and Z. Zhang, "Repair locality with multiple erasure tolerance," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6979–6987, Nov. 2014.
- [10] A. Wang, Z. Zhang, and D. Lin, "Two classes of (r, t) -locally repairable codes," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 445–449.
- [11] 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.
- [12] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, "Optimal linear codes with a local-error-correction property," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, USA, Aug. 2012, pp. 2776–2780.
- [13] P. Huang, E. Yaakobi, H. Uchikawa, and P. H. Siegel, "Binary linear locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6268–6283, Nov. 2016.
- [14] J. Zhang, X. Wang, and G. Ge, "Some improvements on locally repairable codes," [Online]. Available: <https://arxiv.org/abs/1506.04822>.
- [15] Y.-S. Su, "On the construction of local parities for (r, t) -availability in distributed storage," *IEEE Transactions on Communications*, vol. 65, no. 6, pp. 2332–2344, 2017.
- [16] J. Hao and S.-T. Xia, "Constructions of optimal binary locally repairable codes with multiple repair groups," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1060–1063, 2016.
- [17] 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.
- [18] J. Hao, K. Shum, S.-T. Xia, and Y.-X. Yang, "Classification of optimal ternary (r, δ) -locally repairable codes attaining the Singleton-like bound," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 2828–2832.
- [19] B. Chen, S.-T. Xia, J. Hao, and F.-W. Fu, "Constructions of optimal cyclic (r, δ) locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2499–2511, Apr. 2018.
- [20] B. Chen, W. Fang, S.-T. Xia, and F.-W. Fu, "Constructions of optimal (r, δ) locally recoverable codes via constacyclic codes," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5253–5263, Aug. 2019.
- [21] W. Fang and F.-W. Fu, "Optimal cyclic (r, δ) locally repairable codes with unbounded length," in *IEEE Inform. Theory Workshop (ITW)*, Guangzhou, China, Nov. 2018, pp. 1–5.
- [22] B. Chen and J. Huang, "A construction of optimal (r, δ) -locally recoverable codes," *IEEE Access*, vol. 7, pp. 180 349–180 353, 2019.
- [23] T. Ernvall, T. Westerback, R. Freij-Hollanti, and C. Hollanti, "Constructions and properties of linear locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1129–1143, Mar. 2016.
- [24] W. Song, S. H. Dau, C. Yuen, and T. J. Li, "Optimal locally repairable linear codes," *IEEE J. on Selected Areas in Comm.*, vol. 32, no. 5, pp. 1019–1036, 2014.
- [25] H. Cai, Y. Miao, M. Schwartz, and X. Tang, "On optimal locally repairable codes with multiple disjoint repair sets," *IEEE Transactions on Information Theory*, 2019.
- [26] C. J. Colbourn and J. H. Dinitz, *Handbook of Combinatorial Designs*, 2nd ed. Boca Raton: CRC Press, 2006.
- [27] M. P. C. Fossorier, "Quasi-cyclic low-density parity-check codes from circulant permutation matrices," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1788–1793, Aug. 2004.
- [28] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam: North-Holland, 1977.

Jie Hao received the M.S. degree and Ph.D. degree in computer science and technology from Tsinghua University, Beijing, China, in 2013 and 2017, respectively. Since July 2017, he has been with the Information Security Center, Beijing University of Posts and Telecommunications, Beijing, China, where he is now a lecturer. His current research interests include coding theory and information theory.

Kenneth W. Shum received the B.Eng. degree from the Department of Information Engineering, The Chinese University of Hong Kong in 1993, and the M.Sc. and Ph.D. degrees from the Department of Electrical Engineering, University of Southern California, in 1995 and 2000, respectively. He is currently an Associate Professor with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen). His research interests include the coding for distributed storage systems and sequence design for wireless networks.

Shu-Tao Xia received the B. S. degree in mathematics and the Ph.D. degree in applied mathematics from Nankai University, Tianjin, China, in 1992 and 1997, respectively. Since Jan 2004, he has been with the Tsinghua Shenzhen International Graduate School of Tsinghua University, Guangdong, China. He is now a full professor there. From March 1997 to April 1999, he was with the research group of information theory, Department of Mathematics, Nankai University, Tianjin, China. His current research interests include coding and information theory, networking, and machine learning.

Deyin Li received the B.S. degree in Beijing University of Posts and Telecommunications, Beijing, China, in 2018. He is currently pursuing the master degree at Information Security Center, Beijing University of Posts and Telecommunications, Beijing, China.

...