Constructions of Optimal Locally Repairable Codes with Information $(r,\ t)$-Locality

Pan Tan[1], Zhengchun Zhou[2⋆], Vladimir Sidorenko[3⋆⋆], and Udaya Parampalli[4⋆⋆⋆]

[1] School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China lanqingfeixue@my.swjtu.edu.cn

[2] School of Mathematics, Southwest Jiaotong University, Chengdu, China zzc@home.swjtu.edu.cn

[3] Institute for Communications Engineering, Technical University of Munich, Munich, Germany vladimir.sidorenko@tum.de

[4] School of Computing and Information Systems, The University of Melbourne, Melbourne, Australia udaya@unimelb.edu.au

Abstract. Locally repairable codes (LRCs) with $(r,\ t)$-locality have received considerable attention in recent years, since they are able to solve common problems in distributed storage systems such as repairing multiple node failures and management of hot data. Constructing LRCs with excellent parameters becomes an interesting research subject in distributed storage systems and coding theory. In this paper, we present two generic constructions of locally repairable codes with information $(r,\ t)$-locality based on linear algebra and combinatorial designs. The newly proposed LRCs based on linear algebra generalize the construction of high-rate codes proposed by Hao and Xia while maintain the same code rate. The proposed LRCs from finite geometries lead to new parameters with same minimum distance and code rate as the one constructed by Hao and Xia. It is worth noting that all the new binary LRCs with information $(r,\ t)$-locality proposed in this paper are optimal with respect to the bound proposed by Rawat *et al.* in 2016.

Keywords: Locally repairable codes · distributed storage systems · partial geometry · repair locality · multiple failures · hot data.

1 Introduction

In recent years with rapid increase of data resources, distributed storage systems become increasingly popular and important. In order to ensure reliability, data is preferred to store

⋆ P. Tan and Z. Zhou are supported in part by the Joint fund of the Ministry of Education of China under Grant 6141A020223, and in part by the application fundamental research plan project of Sichuan Province under Grant 2018JY0046.

⋆⋆ V. Sidorenko is on leave from Institute for Information Transmission Problems, Russian Academy of Sciences. His work is supported by the Russian Government (Contract No 14.W03.31.0019).

⋆⋆⋆ U. Parampalli is supported in part by the Australian Research Councils Discovery under Grant DP150104473 and in part by the University of Melbourne's 2014 International Research and Research Training Fund.

in a redundant form. The simplest way is to use straightforward replication, which needs a very high storage overhead. To reduce the storage overhead, error-correcting codes have been used in practical systems, where the original data are partitioned into k information packets and then encoded into n packets $(n > k)$. However, classical error-correcting codes are inefficient during the repair procedure of failed nodes in terms of the number of helper nodes accessed, repair bandwidth, etc. During the past decade, researchers have been tried to find new error-correcting codes to address these issues. Repair locality defined by Gopalan *et al.* [4] is an important metric for measuring the efficiency of repair. Specifically, a code symbol of an $[n,\ k]$ linear code has *repair locality* r if the value of the code symbol can be recovered by accessing at most r other symbols (the set of those symbols is called repair set). A linear code has *information locality* r if there exist an information set such that every symbol in this set has locality r. In a distributed storage system, the code with repair locality less than code dimension k is preferred since it has low disk input/output (I/O) complexity during the repair process. The repair locality of codes is the smallest number of nodes that participate in the repair process and was studied in [1,4,6,8,9,11,12] see also references therein. Some of these results have been employed in practice, for example, codes with small locality were recently deployed in Azure production clusters [7], while others have been tested in Facebook clusters [14]. Codes with small repair locality are required in archival and cold data.

There is a common problem in the distributed storage systems which is multiple failures of nodes. Assume there is only one repair set of locality r for a failed node (i-th node), if one of these r nodes also failed, then the failed i-th node can no longer be retrieved by accessing only these r nodes. One way to overcome this problem is to use the code with (r,t)-locality. According to [16], the i-th code symbol is said to have *$(r,\ t)$-locality* if there exist t disjoint subsets, each containing at most r other code symbols that can together recover the i-th code symbol. Clearly, codes with $(r,\ t)$-locality can always tolerate up to t erasures. Such codes can also support a scaling number of parallel reads. The main motivation for this property is the application of error-correcting codes for hot data which is also a significant storage problem. Hot data is a frequently accessed information, often in front-end systems facing end-users. There has been little work on the potential benefits of coding for hot data. A system utilizing an $[n,\ k]$ linear code with $(r,\ t)$-locality can not only greatly reduce the disk I/O complexity for node repair but also can allow access of a node from t ways in parallel which is particularly useful in hot data storage.

In this paper, we present two generic constructions of linear codes over an arbitrary field $\mathbb{F}_q$ with information $(r,\ t)$-locality: one is based on linear algebra and another on combinatorial designs. The proposed LRCs based on linear algebra generalize the construction of high-rate codes proposed by Hao and Xia [5] and have the same code rate. Finite geometries were used long ago for construction codes with majority logic decoding, see e.g. [10] Chapter 10. We show that finite geometries can be used to design some optimal locally repairable codes. The proposed LRCs from partial geometries lead to new code lengths and dimensions with the same minimum distance and code rate as the ones constructed by Hao and Xia [5]. All our constructions yield new LRCs with information $(r,\ t)$-locality which have a only single parity symbol in each repair group. They are optimal in terms of the bound proposed by Rawat *et al.* in [13]. All the proposed codes can achieve high code rates which is preferable in distributed storage systems involving hot data.

The rest of this paper is organized as follows. Section 2 recalls some notations about LRCs. Section 3 introduces a class of new q-ary locally repairable codes from Square codes and proves that they are optimal LRCs. Section 4 gives a construction of LRCs based on partial geometry and some infinite families of optimal LRCs from the proposed construction. Finally, we conclude this paper in Section 5.

2 Locally repairable codes

In this section, we give some preliminaries about locally repairable codes (LRCs). We introduce certain bounds on the parameters of LRCs, which describe the optimality of LRCs with given parameters.

Let q be a prime power and $\mathbb{F}_q$ be the finite field with q elements. A q-ary linear code $\mathcal{C}$ of length n and dimension k is a k-dimensional linear subspace of the vector space $\mathbb{F}_q^n$, where $n \geq k$. As a linear subspace, the code $\mathcal{C}$ can be defined by a $k \times n$ full-rank generator matrix G as

$$\mathcal{C} = \{uG : u \in \mathbb{F}_q^k\}.$$

In this paper, we mainly consider linear codes with systematic generator matrices of code length n. Without loss of generality, we suppose the fist k symbols of a codeword denote the information symbols. The systematic generator matrix is $G = [I_k|A]$, where I_k is the $k \times k$ identity matrix and A is the $k \times (n-k)$ matrix over $\mathbb{F}_q$.

Let $[n] = \{1,2,\ldots,\ n\}$ and $\mathcal{C}$ be an $[n,\ k,\ d]$ linear code over $\mathbb{F}_q$ with minimum Hamming distance d. Let $G = [g_1,\ g_2,\ \ldots,\ g_n]$ be the generator matrix of $\mathcal{C}$ and $c = (c_1,\ \ldots,\ c_n)$ be a codeword of $\mathcal{C}$. Now we give a formal definition of locally repairable codes with $(r,\ t)$-locality.

Definition 1. *([16]) The i-th code symbol c_i, $1 \leq i \leq n$, in a codeword $c = (c_1,\ \ldots,\ c_n)$ of an $[n,\ k,\ d]$ linear code $\mathcal{C}$ over $\mathbb{F}_q$ is said to have $(r,\ t)$-locality if it satisfies the following properties:*

1. *There exist t subsets $R_1(i),\ \ldots,\ R_t(i) \subset [n] \setminus \{i\}$, such that g_i is a linear combination of $\{g_l : l \in R_j(i)\}$ for each $j \in [t]$, where $R_j(i), j \in [t]$ is called a repair set.*
2. *$|R_j(i)| \leq r$, for all $j \in [t]$.*
3. *$R_j(i) \cap R_h(i) = \emptyset$, for all $j \neq h$ and $j,h \in [t]$.*

If there exist an information set I such that for each $i \in I$, the i-th code symbol has $(r,\ t)$-locality, then the code is a LRC with information $(r,\ t)$-locality.

With the definition of $(r,\ t)$-locality, the following well known proposition holds.

Proposition 1. *For an $[n,\ k,\ d]$ LRC $\mathcal{C}$ with information $(r,\ t)$-locality, the minimum distance satisfies $d \geq t+1$.*

In 2016, Rawat *et al.* presented the following upper bound of minimum distance d for an $[n,\ k]$ LRC with information $(r,\ t)$-locality.

Lemma 1. *([13]) Let $\mathcal{C}$ be an $[n,\ k,\ d]$ LRC with information $(r,\ t)$-locality such that each repair set contains a single parity symbol. Then, the minimum distance of the code is bounded as*

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

3 New Locally Repairable Codes obtained from Square codes

In this section, we propose a construction of q-ary locally repairable codes with information $(r,\ t)$-locality which inspired by square codes [16]. In [16], Wang and Zhang provided a lower bound of n for LRCs with information $(r,\ t)$-locality and gave an example of LRCs with all symbol $(r,\ t)$-locality which called square codes. However, square codes can't achieve the lower bound (1). We modify this codes by deleting a parity symbol and show that the modified square codes are optimal with respect to the bound (1). Before doing that, let us first recall the construction of square codes.

Let r be a positive integer and $r+1 \leq k \leq r^2$. Define $\Omega = \{X_{i,j}\}_{1\leq i,j\leq r+1}$ to be a set of $(r+1)^2$ column vectors in $\mathbb{F}_q^k$ satisfying

$$\begin{cases} \sum_{i=1}^{r+1} X_{i,j} = 0, \text{ for } 1 \leq j \leq r+1; \\ \sum_{j=1}^{r+1} X_{i,j} = 0, \text{ for } 1 \leq i \leq r+1. \end{cases}$$

Consider a $[(r+1)^2,\ k]$ code $\mathcal{C}_\Omega$ with generator matrix $G(\Omega)$ consisting of the $(r+1)^2$ vectors in Ω as columns. Then $\mathcal{C}_\Omega$ is a $[(r+1)^2,\ k]$ square code with all symbol $(r,\ t=2)$-locality. Since the minimum distance of $\mathcal{C}_\Omega$ is undetermined, so we cannot say if the code meets the bound (1) or not.

Now, we modify this code. Firstly, we choose r^2 linearly independent vectors from $\mathbb{F}_q^{r^2}$ and denote by $\{X_{i,j}\}_{1\leq i,j\leq r}$. Then let

$$X_{i,r+1} = \sum_{j=1}^{r} a_{i,j} X_{i,j}, \text{ for } 1 \leq i \leq r$$

and

$$X_{r+1,j} = \sum_{i=1}^{r} b_{i,j} X_{i,j}, \text{ for } 1 \leq j \leq r,$$

where $a_{i,j},\ b_{i,j} \in \mathbb{F}_q \setminus \{0\}$. Define a $r^2 \times (r^2+2r)$ matrix $G(\Omega')$ consisting of all the vectors in Ω' as its columns, where

$$\Omega' = \{X_{i,j}\}_{1\leq i,j\leq r} \cup \{X_{i,r+1}\}_{1\leq i\leq r} \cup \{X_{r+1,j}\}_{1\leq j\leq r}.$$

Then we have:

Theorem 1. *The code $\mathcal{C}_{\Omega'}$ generated by $G(\Omega')$ is an $[n=r^2+2r,\ k=r^2,\ d=3]$ q-ary LRC with information $(r,\ t=2)$-locality. It is optimal in terms of the bound* (1).

Proof. Clearly, the code $\mathcal{C}_{\Omega'}$ has length $n=(r+1)^2-1=r^2+2r$ and dimension $k=r^2$. With the construction of Ω' and $\mathcal{C}_{\Omega'}$, we obtain that $\mathcal{C}$ has information $(r,t=2)$-locality. From Proposition 1, we have $d \geq 3$. Taking the code parameters in (1), we have $d \leq 3$. Hence, the minimum distance is equal to 3 and $\mathcal{C}_{\Omega'}$ is optimal in terms of the bound (1). □

Note that the square code can only tolerate $t=2$ symbol erasures. Can we construct LRCs that can tolerate $t>2$ erasures? In fact, we can solve the problem. With the similar construction, we can generalize the case $t=2$ to $t \geq 2$.

Let $r, t \geq 2$ be positive integers and $k = r^t$ be the dimension of the linear code. Define the $k = r^t$ linearly independent vectors $\{X_{i_1,i_2,\ldots,i_t}\}$ over $\mathbb{F}_q^k$ corresponding to the information symbols, where $1 \leq i_j \leq r$ for $j \in [t]$. The parity symbols in the code $\mathcal{C}$, which we are constructing, are partitioned into t groups and corresponding to the following vectors:

$$X^h_{i_1,i_2,\ldots,i_{h-1},i_{h+1},\ldots,i_t} = \sum_{i_h=1}^{r} a_{i_1,i_2,\ldots,i_t} X_{i_1,i_2,\ldots,i_t}, \text{ for } h \in [t],$$

where $a_{i_1,\ldots,i_t} \in \mathbb{F}_q \setminus \{0\}$. Then there are tr^{t-1} parity symbols. Define a set of column vectors in $\mathbb{F}_q^k$ as follows:

$$\Phi = \{X_{i_1,i_2,\ldots,i_t}\} \cup \{X^h_{i_1,i_2,\ldots,i_{h-1},i_{h+1},\ldots,i_t}\}, \tag{2}$$

where $1 \leq i_h \leq r$ and $h \in [t]$.

Consider the code $\mathcal{C}$ with generator matrix $G = G(\Phi)$ consisting of the vectors in Φ as columns. Then we have the following theorem.

Theorem 2. *Let $r,\, t \geq 2$ be positive integers. The code $\mathcal{C}$ with generator matrix $G = G(\Phi)$ where Φ is defined in* (2) *is an $[n = k + tr^{t-1},\, k = r^t,\, d = t+1]$ q-ary LRC with information $(r,\, t)$-locality. It is optimal in terms of the bound* (1).

Proof. The parameters n, k follows directly from the construction. For any information symbol $X_{i_1,i_2,\ldots,i_t}$ in the code $\mathcal{C}$, the following t sets

$$R_{i_1,\ldots,i_t}(h) = \{X_{i_1,\ldots,i_{h-1},\xi,i_{h+1},\ldots,i_t} : \xi \neq i_h\} \cup \{X^h_{i_1,i_2,\ldots,i_{h-1},i_{h+1},\ldots,i_t}\}, \text{for } h \in [t],$$

are all disjoint, and for all $1 \leq h \leq t$, we have

$$X_{i_1,\ldots,i_t} = aX^h_{i_1,i_2,\ldots,i_{h-1},i_{h+1},\ldots,i_t} - \sum_{\xi=1,\xi\neq i_h}^{r} aa_{i_1,\ldots,i_{h-1},\xi,i_{h+1},\ldots,i_t} X_{i_1,\ldots,i_{h-1},\xi,i_{h+1},\ldots,i_t},$$

where $a = a^{-1}_{i_1,\ldots,i_{h-1},i_h,i_{h+1},\ldots,i_t}$. Therefore, $\mathcal{C}$ has information $(r,\, t)$-locality. And the code has the minimum distance $d \geq t+1$ by Proposition 1. Note that each of the t disjoint repair sets contains only one parity symbol. By (1), we have $d \leq t+1$. Hence the minimum distance $d = t+1$ achieves the bound (1). □

We give a simple example to demonstrate the construction.

Example 1. Let $r = 2$, $t = 3$ and $a_{i_1,\ldots,i_t} = 1$ for $1 \leq i_j \leq r$ and $j \in [t]$. Then we have $r^t = 8$ information symbols and $tr^{t-1} = 12$ parity symbols. They are listed as follows:

- The vectors corresponding to information symbols are:

$$X_{1,1,1},\ X_{1,1,2},\ X_{1,2,1},\ X_{1,2,2},\ X_{2,1,1},\ X_{2,1,2},\ X_{2,2,1},\ X_{2,2,2}$$

- The vectors corresponding to parity symbols are defined by :

$$X^1_{1,1} = X_{1,1,1} + X_{2,1,1};\ X^1_{1,2} = X_{1,1,2} + X_{2,1,2};\ X^1_{2,1} = X_{1,2,1} + X_{2,2,1};\ X^1_{2,2} = X_{1,2,2} + X_{2,2,2};$$

$$X^2_{1,1} = X_{1,1,1} + X_{1,2,1};\ X^2_{1,2} = X_{1,1,2} + X_{1,2,2};\ X^2_{2,1} = X_{2,1,1} + X_{2,2,1};\ X^2_{2,2} = X_{2,1,2} + X_{2,2,2};$$

$$X^3_{1,1} = X_{1,1,1} + X_{1,1,2};\ X^3_{1,2} = X_{1,2,1} + X_{1,2,2};\ X^3_{2,1} = X_{2,1,1} + X_{2,1,2};\ X^3_{2,2} = X_{2,2,1} + X_{2,2,2}.$$

The generator matrix is

$$G = (X_{1,1,1}\; X_{1,1,2}\; X_{1,2,1}\; X_{1,2,2}\; X_{2,1,1}\; X_{2,1,2}\; X_{2,2,1}\; X_{2,2,2}\; X^1_{1,1}\; X^1_{1,2}\; X^1_{2,1}\; X^1_{2,2}\; X^2_{1,1}\; X^2_{1,2}\; X^2_{2,1}\; X^2_{2,2}\; X^3_{1,1}\; X^3_{1,2}\; X^3_{2,1}\; X^3_{2,2})$$

$$= \begin{array}{c}
\begin{array}{cccccccccccccccccccc} g_1 & g_2 & g_3 & g_4 & g_5 & g_6 & g_7 & g_8 & g_9 & g_{10} & g_{11} & g_{12} & g_{13} & g_{14} & g_{15} & g_{16} & g_{17} & g_{18} & g_{19} & g_{20} \end{array} \\
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}.$$

The code generated by G is an $[20,\ 8,\ 4]$ q-ary LRC with $(2,\ 3)$-locality. For c_1, which corresponds to $X_{1,1,1}$ (i.e. g_1), we have $g_1 = g_9 - g_5 = g_{13} - g_3 = g_{17} - g_2$ and we have $c_1 = c_9 - c_5 = c_{13} - c_3 = c_{17} - c_2$. Hence the repair sets of c_1 are: $R_1(1) = \{5,9\},\ R_2(1) = \{3,13\},\ R_3(1) = \{2,17\}$. Similarly, we obtain the repair sets of the other code symbols are : $R_1(2) = \{6,10\}$, $R_2(2) = \{4,14\}$, $R_3(2) = \{1,17\}$ and so on.

4 Locally Repairable Codes based on partial geometries

In this section, we give a new construction of q-ary locally repairable codes with information $(r,\ t)$-locality based on partial geometries. This construction utilizes the incidence matrix of partial geometries and we demonstrate that the codes we constructed attain the bound in (1). Before doing that, let us briefly introduce some concepts of partial geometries.

4.1 Partial geometry

Definition 2. *([3]) A (finite) partial geometry is an incidence structure $\mathcal{S} = (\mathcal{P},\ \mathcal{B},\ I)$ in which $\mathcal{P}$ is a set of points, $\mathcal{B}$ is a set of lines and I is a symmetric point-line incidence relation satisfying the following axioms:*

1. *Each point $P \in \mathcal{P}$ is incident with $u+1$ lines$(u \geq 1)$, and two distinct points are incident with at most one line.*
2. *Each line $l \in \mathcal{B}$ is incident with $s+1$ points$(s \geq 1)$, and two distinct lines are incident with at most one point.*
3. *If a point P and a line l are not incident, then there are exactly α points $P_1,\ P_2,\ \cdots,\ P_\alpha$ and α lines $l_1,\ l_2,\ \cdots,\ l_\alpha$ such that P is incident with l_i and P_i is incident with l for $i = 1,\ 2,\ \ldots,\ \alpha$, where $\alpha \geq 1$.*

We denote the partial geometry $\mathcal{S} = (\mathcal{P},\ \mathcal{B},\ I)$ with above parameters as $\mathrm{PG}(s+1,\ u+1,\ \alpha)$. Let $v = |\mathcal{P}|$ and $b = |\mathcal{B}|$, then from [3] we have following relations:

$$v = \frac{(s+1)(su+\alpha)}{\alpha},\ \ b = \frac{(u+1)(su+\alpha)}{\alpha}. \tag{3}$$

Define the incidence matrix of a partial geometry $\mathrm{PG}(s+1,\ u+1,\ \alpha)$ as a $b \times v$ matrix $N = (n_{ij})$, where $n_{ij} = 1$ if the ith line is incident with the jth point and $n_{ij} = 0$ otherwise. Then N is a $b \times v$ matrix. Consider an example to explain the definition of a partial geometries.

Example 2. Let $s=2,\ u=2,\ \alpha=3$, then the PG$(3,\ 3,\ 3)$ is an incidence structure with $v=9$ points and $b=9$ lines where the point set is $\mathcal{P}=\{P_1,\ P_2,\ \cdots,\ P_9\}$ and the line set is $\mathcal{B}=\{l_1,\ l_2,\ \cdots,\ l_9\}$ with the following point-line incidence relation:$l_1=\{P_1,\ P_4,\ P_7\}$, $l_2=\{P_2,\ P_5,\ P_8\}$, $l_3=\{P_3,\ P_6,\ P_9\}$, $l_4=\{P_1,\ P_5,\ P_9\}$, $l_5=\{P_2,\ P_6,\ P_7\}$, $l_6=\{P_3,\ P_4,\ P_8\}$, $l_7=\{P_1,\ P_6,\ P_8\}$, $l_8=\{P_2,\ P_4,\ P_9\}$, $l_9=\{P_3,\ P_5,\ P_7\}$. The incidence matrix of PG$(3,3,3)$ is

$$N=\begin{array}{c} \\ l_1\\ l_2\\ l_3\\ l_4\\ l_5\\ l_6\\ l_7\\ l_8\\ l_9 \end{array}\begin{array}{c} \begin{array}{ccccccccc} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 & P_7 & P_8 & P_9 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0\\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1\\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1\\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0\\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0\\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}. \tag{4}$$

The dual of a partial geometry is the incidence structure obtained by exchanging the set of points and the set of lines, which is also a partial geometry with parameters PG$(u+1,\ s+1,\ \alpha)$.

Theorem 3. *([3]) Up to duality, the parameters of the known partial geometries are the following:*

Type 0: $s=w,\ u=w^{m-1}-1,\ \alpha=w$, *with* $m\geq 2$ *and* w *is a power of prime;*
Type 1: $s=2^h-2^m,\ u=2^h-2^{h-m},\ \alpha=(2^{h-m}-1)(2^m-1)$, *with* $1\leq m\leq h$*;*
Type 2: $s=2^h-1,\ u=(2^h+1)(2^m-1),\ \alpha=2^m-1$, *with* $1\leq m\leq h$*;*
Type 3: $s=2^{2h-1}-1,\ u=2^{2h-1},\ \alpha=2^{2h-2}$, *with* $1<h$*;*
Type 4: $s=3^{2m}-1,\ u=(3^{4m}-1)/2,\ \alpha=(3^{2m}-1)/2$, *with* $m\geq 1$*;*

4.2 New codes

Let us build a $b\times(b+v)$ systematic generator matrix G containing of the $b\times b$ identity matrix and a $b\times v$ matrix N where b and v are defined in (3):

$$G=[I_b|N], \tag{5}$$

where N is the incidence matrix of a partial geometry PG$(s+1,\ u+1,\ \alpha)$. We denote the code over $\mathbb{F}_q$ constructed from a partial geometry PG$(s+1,\ u+1,\ \alpha)$ as $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$.

Theorem 4. *The* q*-ary linear code* $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$ *over* $\mathbb{F}_q$ *with systematic generator matrix* (5) *is an* $[n,\ k,\ d]$ *locally repairable code with information* $(r,\ t)$*-locality, where*

$$n=v+b,\ k=b,\ d=s+2,\ r=u+1,\ t=s+1.$$

The code is optimal in terms of the bound in (1).

Proof. Obviously, $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$ has length $b+v$ and dimension b. Recall that the columns in I_b correspond to the information symbols, and the columns in N correspond to the parity symbols. Since N is the incidence matrix of partial geometry PG$(s+1,\ u+1,\ \alpha)$, each row has Hamming weight $s+1$, that implies that every information symbol has $t=s+1$ repair sets. Note that each column of N has Hamming weight $u+1$. Any two distinct columns of N are common in at most one coordinate. Hence the code has locality $r=u+1$. The repair sets for a fixed information symbol are disjoint and each repair set includes only a single parity symbol. Note that each row of G has weight $s+2$, therefore the minimum distance $d \leq s+2$. On the other hand, we have $d \geq t+1 = s+2$ from Proposition 1. Hence the minimum distance of $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$ is $d = s+2$. It is easy to verify that $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$ is optimal in terms of the bound (1). $\square$

Theorem 5. *The dual code $\mathcal{D}$ of linear code $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$ over $\mathbb{F}_q$with systematic generator matrix $G = [I_b|N]$ is an $[n,\ k,\ d]$ locally repairable code with information $(r,\ t)$-locality, where*

$$n = v+b, k = v, d = u+2, r = s+1, t = u+1.$$

The code is optimal with respect to the bound in (1).

Proof. The statement follows from the duality of partial geometries and Theorem 4.

In order to demonstrate the construction above, we use $\mathrm{PG}(3,\ 3,\ 3)$ to construct an optimal LRC with $(r=3,\ t=3)$-locality in the next example.

Example 3. Let N be the incidence matrix of $\mathrm{PG}(3,3,2)$ see (4). Then the code $\mathcal{C}_{(3,3,3)}$ in Theorem 4 is an optimal $[n=18,\ k=9,\ d=4]$ LRC over $\mathbb{F}_q$ with information $(r=3,\ t=3)$-locality. The generator matrix G of $\mathcal{C}_{(3,3,2)}$ defined by (5) is as follows:

$$G = \begin{array}{c} \begin{array}{cccccccccccccccccc} g_1 & g_2 & g_3 & g_4 & g_5 & g_6 & g_7 & g_8 & g_9 & g_{10} & g_{11} & g_{12} & g_{13} & g_{14} & g_{15} & g_{16} & g_{17} & g_{18} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}.$$

For c_1, which corresponds to g_1, we have $g_1 = g_{10} - g_4 - g_7 = g_{13} - g_6 - g_8 = g_{16} - g_5 - g_9$ and we obtain for the first information symbol the repair relations: $c_1 = c_{10} - c_4 - c_7 = c_{13} - c_6 - c_8 = c_{16} - c_5 - c_9$. Hence the repair sets of c_1 are: $R_1(1) = \{4,7,10\}$, $R_2(1) = \{6,8,13\}$, $R_3(1) = \{5,9,16\}$. Similarly, we obtain the repair sets of the other code symbols are : $R_1(2) = \{5,8,11\}$, $R_2(2) = \{4,9,14\}$, $R_3(2) = \{6,7,17\}$ and so on.

In the following, we give some specific constructions of optimal q-ary LRCs under the above framework. Recall that we enumerate some partial geometries in Section 2. Then we have following results.

Theorem 6. *With the partial geometry in Theorem 3, we have optimal LRCs with the following parameters.*

1. *For partial geometry of Type 0 in Theorem 3, the code $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$ in Theorem 4 is an optimal $[n = w^{m-1}(w^{m-1}+q+1),\ k = w^{2(m-1)},\ d = w+2]$ q-ary LRCs with information $(r = w^{m-1},\ t = w+1)$-locality in terms of the bound* (1).
2. *Let the parameters of partial geometry be of Type 1 in Theorem 3. Then the code $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$ in Theorem 4 is an optimal $[n = 2(2^h+1)(2^h-2^{m-1}-2^{h-m-1}+1),\ k = (2^h+1)(2^h-2^{h-m}+1),\ d = 2^h-2^m+2]$ q-ary LRC with information $(r = 2^h-2^{h-m}+1,\ t = 2^h-2^m+1)$-locality in terms of the bound* (1).

3. *Consider partial geometry of Type 2 in Theorem 3. Then the code* $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$ *in Theorem 4 is an optimal* $[n = 2^{m+2h}(2^h+1),\ k = 2^{m+2h}(2^h - 2^{h-m}+1),\ d = 2^h+1]$ *q-ary LRC with information* $(r = 2^m(2^h - 2^{h-m}+1),\ t = 2^h)$*-locality in terms of the bound* (1)*.*
4. *For partial geometry of Type 3 in Theorem 3, the code* $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$ *in Theorem 4 is an optimal* $[n = 2^{4h}-1,\ k = (2^{2h-1}+1)(2^{2h}-1),\ d = 2^{2h-1}+1]$ *q-ary LRC with information* $(r = 2^{2h-1}+1,\ t = 2^{2h-1})$*-locality in terms of the bound* (1)*.*
5. *Take partial geometry as Type 4 in Theorem 3. Then the code* $\mathcal{C}_{(s+1,\ u+1,\ \alpha)}$ *in Theorem 4 is an optimal* $[n = \frac{3^{4m}(3^{2m}+1)^2}{2},\ k = \frac{3^{4m}(3^{4m}+1)}{2},\ d = 3^{2m}+1]$ *q-ary LRC with information* $(r = \frac{3^{4m}+1}{2},\ t = 3^{2m})$*-locality in terms of the bound* (1)*.*

Proof. Follows directly from Theorem 3 and Theorem 4.

Note that all codes from our constructions were not covered in previous literature. Thus our constructions lead to codes with new parameters. We give some parameters of some known LRCs and the newly proposed LRCs with information $(r,\ t)$-locality in Table 1. We list 4 classes of LRCs constructed before (see No. 1-4) and 6 classes of LRCs with new parameters. All of these LRCs have same minimum distance $t+1$. And all of these LRCs have same code rate $\frac{1}{1+t/r}$ except No. 2, which can have rate greater than $\frac{1}{2}$ when $r > t$. Furthermore, all the codes listed in Table 1 are optimal in terms of bound (1).

Table 1. Some optimal LRCs with (r,t)-locality

No.	n	k	d	r	t	$R = \frac{k}{n}$	Ref.
1	$\binom{r+t}{t}$	$\binom{r+t-1}{t}$	$t+1$	r	t	$\frac{1}{1+t/r}$	[17]
2	$2(4^m+2^m+1),\ m \geq 1$	4^m+2^m+1	$t+1$	2^m+1	2^m+1	$1/2$	[5]
3	$r^2+tr,\ r \geq t$ and r is prime	r^2	$t+1$	r	t	$\frac{1}{1+t/r}$	
4	$rm+tm,\ m \geq 2$	rm	$t+1$	r	t	$\frac{1}{1+t/r}$	[15]
5	$(m+1)l,\ 1 \leq m \leq r$ $l = r(r-1)x+1,\ x \geq 1$	l	$t+1$	r	rm	$\frac{1}{1+t/r}$	[2]
6	r^t+tr^{t-1}	r^t	$t+1$	r	t	$\frac{1}{1+t/r}$	Thm 2
7	$w^{m-1}(w^{m-1}+w+1),\ m \geq 2,$ w is a power of prime	$w^{2(m-1)}$	$t+1$	w^{m-1}	$w+1$	$\frac{1}{1+t/r}$	Thm 6
8	$(2^h+1)(2^{h+1}-2^m-2^{h-m}+2),$ $1 \leq m \leq h$	$(2^h+1)(2^h-2^{h-m}+1)$	$t+1$	$2^h-2^{h-m}+1$	2^h-2^m+1		
9	$2^{m+2h}(2^h+1),\ h \geq m \geq 1$	$2^{m+2h}(2^h-2^{h-m}+1)$	$t+1$	$2^m(2^h-2^{h-m}+1)$	2^h		
10	$2^{4h}-1,\ h > 1$	$(2^{2h-1}+1)(2^{2h}-1)$	$t+1$	$2^{2h-1}+1$	2^{2h-1}		
11	$\frac{3^{4m}(3^{2m}+1)^2}{2},\ m \geq 1$	$\frac{3^{4m}(3^{4m}+1)}{2}$	$t+1$	$\frac{3^{4m}+1}{2}$	3^{2m}		

5 Conclusions

In this paper, we proposed two generic constructions of LRCs with information (r,t)-locality. All our codes have only one parity symbol in each repair group and their minimum distances achieve the upper bound proposed by Rawat in [13]. Based on linear algebra, we constructed a class of LRCs over arbitrary field $\mathbb{F}_q$ with new optimal parameters. These LRCs generalize the construction of high-rate codes proposed by Hao and Xia in [5]. From the known proper partial geometries, we constructed the second class of

q-ary LRCs with new parameters, which are optimal with respect to the bound (1). The key problem of this construction is the existence of partial geometries. It may be possible to obtain more new q-ary LRCs using other partial geometries with new parameters.

Acknowledgement

The authors are very grateful to the co-chairs of WCC, Anne Canteaut and Felix Ulmer, and to the anonymous reviewers for their comments and suggestions. Special thanks for bringing our attention to connection of proposed codes with PIR/batch codes, which we will include in the journal version.

References

1. Cadambe, V.R., Mazumdar, A.: Bounds on the size of locally recoverable codes. IEEE Transactions on Information Theory **61**(11), 5787–5794 (Nov 2015).
2. Cai, H., Cheng, M., Fan, C., Tang, X.: Optimal locally repairable systematic codes based on packings. IEEE Transactions on Communications pp. 1–1 (2018).
3. Colbourn, C.J., Dinitz, J.H.: Handbook of combinatorial designs. CRC Press (1996).
4. Gopalan, P., Huang, C., Simitci, H., Yekhanin, S.: On the locality of codeword symbols. IEEE Transactions on Information Theory **58**(11), 6925–6934 (Nov 2012).
5. Hao, J., Xia, S.: Constructions of optimal binary locally repairable codes with multiple repair groups. IEEE Communications Letters **20**(6), 1060–1063 (June 2016).
6. Huang, C., Chen, M., Li, J.: Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems. In: Sixth IEEE International Symposium on Network Computing and Applications (NCA 2007). pp. 79–86 (July 2007).
7. Huang, C., Simitci, H., Xu, Y., Ogus, A., Calder, B., Gopalan, P., Li, J., Yekhanin, S., et al.: Erasure coding in windows azure storage. In: Usenix Annual Technical Conference. pp. 15–26. Boston, MA (2012).
8. Kamath, G.M., Prakash, N., Lalitha, V., Kumar, P.V.: Codes with local regeneration and erasure correction. IEEE Transactions on Information Theory **60**(8), 4637–4660 (Aug 2014).
9. Papailiopoulos, D.S., Dimakis, A.G.: Locally repairable codes. IEEE Transactions on Information Theory **60**(10), 5843–5855 (Oct 2014).
10. Peterson, W.W., Weldon, E.J.: Error-correcting codes, Second eddition, The MIT Press (1972).
11. Prakash, N., Kamath, G.M., Lalitha, V., Kumar, P.V.: Optimal linear codes with a local-error-correction property. In: 2012 IEEE International Symposium on Information Theory Proceedings. pp. 2776–2780 (July 2012).
12. Rawat, A.S., Koyluoglu, O.O., Silberstein, N., Vishwanath, S.: Optimal locally repairable and secure codes for distributed storage systems. IEEE Transactions on Information Theory **60**(1), 212–236 (Jan 2014).
13. Rawat, A.S., Papailiopoulos, D.S., Dimakis, A.G., Vishwanath, S.: Locality and availability in distributed storage. IEEE Transactions on Information Theory **62**(8), 4481–4493 (Aug 2016).
14. Sathiamoorthy, M., Asteris, M., Papailiopoulos, D., Dimakis, A.G., Vadali, R., Chen, S., Borthakur, D.: Xoring elephants: Novel erasure codes for big data. In: Proceedings of the VLDB Endowment. vol. 6, pp. 325–336. VLDB Endowment (2013).
15. Su, Y.: On constructions of a class of binary locally repairable codes with multiple repair groups. IEEE Access **5**, 3524–3528 (2017).
16. Wang, A., Zhang, Z.: Repair locality with multiple erasure tolerance. IEEE Transactions on Information Theory **60**(11), 6979–6987 (Nov 2014).
17. Wang, A., Zhang, Z., Liu, M.: Achieving arbitrary locality and availability in binary codes. In: 2015 IEEE International Symposium on Information Theory. pp. 1866–1870 (June 2015).