A Bound on the Minimal Field Size of LRCs, and Cyclic MR Codes That Attain It

Han Cai* and Moshe Schwartz†
*School of Information Science and Technology, Southwest Jiaotong University,
Chengdu, 610031, China, hancai@aliyun.com
†School of Electrical and Computer Engineering, Ben-Gurion University of the Negev,
Beer Sheva 8410501, Israel, schwartz@ee.bgu.ac.il

***Abstract*—We prove a new lower bound on the field size of locally repairable codes (LRCs). Additionally, we construct maximally recoverable (MR) codes which are cyclic. While a known construction for MR codes has the same parameters, it produces non-cyclic codes. Furthermore, we prove necessary and sufficient conditions that specify when the known non-cyclic MR codes may be permuted to become cyclic, thus proving our construction produces cyclic MR codes with new parameters. Furthermore, using our new bound on the field size, we show that the new cyclic MR codes have optimal field size in certain cases. Other known LRCs are also shown to have optimal field size in certain cases.**

I. Introduction

In large-scale cloud storage and distributed file systems, such as Amazon Elastic Block Store (EBS) and Google File System (GoogleFS), disk failures are the norm and not the exception, due to the sheer scale of the system. To protect the data integrity, coding theory is used to recover from data loss due to disk failures. Erasure codes such as $[n,k]$ maximum distance separable (MDS) codes, may be employed as storage codes. These codes encode k information symbols to n symbols and store them across n disks, and they can recover from the loss of any $n-k$ symbols. This scheme achieves a dramatic improvement in redundancy compared with replication. However, for MDS codes, even if one disk fails, the system needs to access k surviving disks in order to recover the lost symbol, which makes the repair process costly.

For locally repairable codes, other code properties are also desirable. For a given code length n and dimension k, we would like the Hamming distance to be as large as possible, in order to maximize erasure-correcting capabilities. Additionally, we would like the field size (or alphabet size) to be as small as possible, in order to reduce the computation complexity for coding and decoding. Other desirable properties may include a cyclic structure for the code, since it allows for fast encoding algorithms. Finally, even if the code has optimal distance, we would like to be able to correct some pre-determined erasure patterns beyond the minimum Hamming distance.

In the past a few years, many results have been obtained for LRCs. Upper bounds on the minimum Hamming distance were proved, e.g., Singleton-type bounds [6], [16], [28], [35], and bounds related with the alphabet size [1], [5]. Optimal LRCs (with respect to these bounds), were constructed, e.g., [12], [22], [25], [31]–[33], [36]. In [7], [20], lower bounds on the field size of optimal LRCs were derived for $\delta=2$ [20], and $\delta\geqslant 2$ [7]. Among the known optimal LRCs, some of them also achieve order-optimal field size [2], [11], [23], [37] when $\delta=2$, and [7] when $\delta\geqslant 2$. Otherwise, constructions of optimal cyclic LRCs were introduced in [12], [13], [29], [30], [34]. When considering pre-determined recoverable erasure patterns beyond the minimum Hamming distance, codes that can recover from all information-theoretically recoverable erasure patterns are called maximally recoverable (MR) codes [16], also known as partial MDS codes [3]. In [18], lower bounds on the field size requirement for MR codes were introduced. For explicit constructions of MR codes, the reader may refer to [3], [8], [14], [15], [17], [19], [25]. Notably, there are MR codes have order-optimal field size (with respect to the bound in [18]): [3] for a single global parity check ($h=1$), [4], [18] for $h=2$, [15] for $h=3$ and $\delta=2$, and [8], [17] for $h\leqslant\delta+1$ a constant, and $n=\Theta(r^2)$.

The above summary shows how subsets of the mentioned desired properties may be obtained simultaneously. However, to the best of our knowledge, there are no explicit constructions that achieve all them, namely, cyclic MR codes with optimal field size. In this paper, we propose a construction which gives a positive answer to this problem. Our construction produces cyclic MR codes that share the same parameters as one of the known non-cyclic constructions in [18]. We also show that under certain conditions, the non-cyclic construction from [18] can be permuted to become a cyclic code, whereas in other cases it cannot, thus proving our construction produces cyclic MR codes with new parameters. To prove the optimality of the field size, we prove a new general bound for LRCs, and show that our construction has an optimal field size when $r=2$. Since the bound is for general LRCs, as a byproduct we get that some known constructions have optimal field size when $r=2$, a result which has not been claimed before.

Due to space limitations we omit all proofs, which are available in a full version of this work [10].

II. Preliminaries

In this section, we present notation and some necessary known results, which are used throughout the paper. For a positive integer $n\in\mathbb{N}$, we define $[n]=\{0,1,\dots,n-1\}$. If

This work was supported in part by a German Israeli Project Cooperation (DIP) grant under grant no. PE2398/1-1.

978-1-6654-2159-1/22/$31.00 ©2022 IEEE

$m|n$ is a positive integer, we denote

$$\langle m\rangle \triangleq m\mathbb{Z}\cap[n]=\{0,m,2m,\dots,n-m\}\,.$$

Thus, $\langle m\rangle$ implicitly depends on n, whose value should be understood from the context.

For any prime power q, let $\mathbb{F}_q$ denote the finite field of size q, let $\mathbb{F}_q^m$ denote the set of vectors of length m over $\mathbb{F}_q$, and let $\mathbb{F}_q^{m\times n}$ denote the set of all possible $m\times n$ matrices over $\mathbb{F}_q$.

An $[n,k]_q$ linear code, $\mathcal{C}$, over $\mathbb{F}_q$, is a k-dimensional subspace of $\mathbb{F}_q^n$. Such a code may be specified as the row-space of a $k\times n$ generator matrix $G=(g_0,g_1,\dots,g_{n-1})$, where g_i is a column vector of length k for all $i\in[n]$. Specifically, it is called an $[n,k,d]_q$ linear code if the minimum Hamming distance of the code is d. For a subset $S\subseteq[n]$, we define

$$\begin{aligned}\operatorname{span}(S)&\triangleq\operatorname{span}\{g_i \;:\; i\in S\}\,,\\ \operatorname{rank}(S)&\triangleq\operatorname{rank}(\operatorname{span}(S)).\end{aligned}$$

The code $\mathcal{C}$ can also be specified by a parity-check matrix $H\in\mathbb{F}_q^{(n-k)\times n}$, i.e., $\mathcal{C}=\left\{c\in\mathbb{F}_q^n \;:\; Hc^{\intercal}=0\right\}$, where $\operatorname{rank}(H)=n-k$. Given a non-empty set of coordinates, $S\subseteq[n]$, the punctured code $\mathcal{C}|_S$ is the code obtained from $\mathcal{C}$ by deleting the code symbols at positions $[n]\setminus S$. Thus, $\mathcal{C}|_S$ is generated by $G|_S$ which is obtained from G by deleting the columns at $[n]\setminus S$. Similarly, the shortened code $\mathcal{C}|^S$ is the code whose parity matrix is $H|_S$, namely, the matrix obtained from H by deleting the columns at $[n]\setminus S$.

An $[n,k]_q$ linear code, $\mathcal{C}$, is said to be a cyclic code if $c=(c_0,c_1,\cdots,c_{n-1})\in\mathcal{C}$ implies that $\sigma(c)\triangleq(c_{n-1},c_0,c_1,\cdots,c_{n-2})\in\mathcal{C}$, where σ is the cyclic shift operator by one place. It is well known (see [24]) that a cyclic code with length n over $\mathbb{F}_q$ corresponds to a principal ideal of $\mathbb{F}_q[x]/(x^n-1)$. Thus, let $\mathcal{C}$ be generated by a monic polynomial $g(x)|(x^n-1)$, which is called the generator polynomial of $\mathcal{C}$. When $n|q^m-1$, assume α is a primitive nth root of unity of $\mathbb{F}_{q^m}$, then the cyclic code $\mathcal{C}$ can be also be determined by the roots of $g(x)$, i.e., $R_{\mathcal{C}}=\left\{\alpha^i \;:\; g(\alpha^i)=0\right\}$.

A. Locally Repairable Codes

In [16], Gopalan *et al.* introduced a definition for the locality of code symbols. For $j\in[n]$, the jth code symbol, c_j, of an $[n,k,d]_q$ linear code, $\mathcal{C}$, is said to have locality r if it can be recovered by accessing at most r other symbols of $\mathcal{C}$. This has been generalized in [28] to the following definition:

Definition 1: Let $\mathcal{C}$ be an $[n,k,d]_q$ linear code, and let G be a generator matrix for it. For $j\in[n]$, the jth code symbol, c_j, of $\mathcal{C}$, is said to have (r,δ)-locality if there exists a subset $S_j\subseteq[n]$ such that:

- $j\in S_j$ and $|S_j|\leqslant r+\delta-1$; and
- the minimum Hamming distance of the punctured code $\mathcal{C}|_{S_j}$ is at least δ.

In that case, the set S_j is also called a repair set of c_j. The code $\mathcal{C}$ is said to have information (r,δ)-locality if there exists $S\subseteq[n]$ with $\operatorname{rank}(S)=k$ such that for each $j\in S$, c_j has (r,δ)-locality. Furthermore, the code $\mathcal{C}$ is said to have all-symbol (r,δ)-locality if all the code symbols have (r,δ)-locality.

Lemma 1 ([16], [28])**:** For an $[n,k,d]_q$ linear code with information (r,δ)-locality,

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

Codes with information (r,δ)-locality are said to be *optimal locally repairable codes (optimal LRCs)* if their minimum Hamming distance attains the bound of Lemma 1 with equality.

In [20], Guruswami *et al.* asked a fundamental interesting question: How long can an optimal LRC with $(r,\delta=2)$-locality be? They derived the following upper bound on the code length.

Lemma 2 ([20])**:** Let $\mathcal{C}$ be an optimal $[n,k,d]_q$ LRC with all-symbol $(r,2)$-locality. If $d\geqslant 5$, $k>r$, $(r+1)|n$, and additionally, $r|k$ or $k\geqslant 2r^2+2r-(2r-1)(k \bmod r)$, then

$$n=\begin{cases}O\left(dq^{\frac{4(d-2)}{d-a}-1}\right), & \text{if } a=1,2,\\ O\left(dq^{\frac{4(d-3)}{d-a}-1}\right), & \text{if } a=3,4,\end{cases}\tag{1}$$

where $a\in\{1,2,3,4\}$, and $a\equiv d\pmod 4$.

In [7], this problem is further considered for optimal LRCs with all-symbol (r,δ)-locality, $\delta\geqslant 2$.

Lemma 3 ([7])**:** Let $n=w(r+\delta-1)$, $\delta>2$, $k=ur+v$, $0\leqslant v\leqslant r-1$, and additionally, $u\geqslant 2(r-v+1)$ or $v=0$, where all parameters are integers. Assume that there exists an optimal $[n,k,d]_q$ linear code $\mathcal{C}$ with all-symbol (r,δ)-locality, and define $t=\lfloor(d-1)/\delta\rfloor$. If $t\geqslant 2$, then

$$\begin{aligned}n&\leqslant\begin{cases}\frac{(t-1)(r+\delta-1)}{2r(q-1)}q^{\frac{2(w-u)r-2v}{t-1}} & \text{if } t \text{ is odd}\\ \frac{t(r+\delta-1)}{2r(q-1)}q^{\frac{2(w-u)r-2v}{t}} & \text{if } t \text{ is even}\end{cases}\\ &=O\left(\frac{t(r+\delta)}{r}q^{\frac{(w-u)r-v}{\lfloor t/2\rfloor}-1}\right),\end{aligned}$$

where $w-u$ can also be rewritten as $w-u=\lfloor(d-1+v)/(r+\delta-1)\rfloor$.

B. Maximally Recoverable Codes

Maximally recoverable (MR) codes are an extremal case of LRCs, that maximize the erasure-repair capability.

Definition 2: Let $\mathcal{C}$ be an $[n,k,d]_q$ code with all-symbol (r,δ)-locality, and define $\mathcal{S}\triangleq\{S_i \;:\; i\in[n]\}$, where S_i is a repair set for coordinate i. The code $\mathcal{C}$ is said to be a *maximally recoverable (MR) code* if $\mathcal{S}$ is a partition of $[n]$, and for any $R_i\subseteq S_i$ such that $|S_i\setminus R_i|=\delta-1$, the punctured code $\mathcal{C}|_{\cup_{i\in[n]}R_i}$ is an MDS code.

In general, S_i for $i\in[n]$, are not required to be of the same size. However, from an application point of view, equal-sized repair sets simplify the implementation, bringing us to the following definition:

Definition 3: Let $\mathcal{C}$ be an $[n,k,d]_q$ MR code, as in Definition 2. If each $S_i\in\mathcal{S}$ is of size $|S_i|=r+\delta-1$

(implying $r+\delta-1|n$), we define $m \triangleq \frac{n}{r+\delta-1}$and $h \triangleq mr-k$. Then $\mathcal{C}$ is said to be an (n,r,h,δ,q)-MR code.

We first note that it is easy to verify that (n,r,h,δ,q)-MR codes are optimal $[n,k,d]_q$ LRCs with all-symbol (r,δ)-locality. We can regard each codeword of an (n,r,h,δ,q)-MR code, as an $m \times (r+\delta-1)$ array, by placing each repair set in $\mathcal{S}$ as a row, when $\mathcal{S}$ forms a parition of $[n]$. In this way, (n,r,h,δ,q)-MR codes match the definition of partial MDS (PMDS) codes, as defined in [3]. When implemented in a distributed-storage setting, each entry of a codeword array corresponds to a sector, each column of the array corresponds to a disk, and each row to a stripe. Thus, an (n,r,h,δ,q)-MR code can recover from $\delta-1$ sector erasures in each stripe, and additional h erased sectors anywhere. We mention in passing that a more restricted type of codes, called sector-disk (SD) codes, are capable of recovering from $\delta-1$ disk erasures, and additional h erased sectors (see [9], [27]).

Paralleling the general case of optimal LRCs, it is interesting to ask what is the minimum alphabet size required by MR codes.

Lemma 4 ([18, Theorem I.1])**:** Let $\mathcal{C}$ be an (n,r,h,δ,q)-MR code, $h \geqslant 2$. If $m \triangleq \frac{n}{r+\delta-1} \geqslant 2$, then $q=\Omega(nr^{\varepsilon})$, where $\varepsilon = \min\{\delta-1, h-2\lceil\frac{h}{m}\rceil\}/\lceil\frac{h}{m}\rceil$, and where h and δ are regarded as constants. Additionally,

1) If $m \geqslant h$:
$$q = \Omega\left(nr^{\min\{\delta-1,h-2\}}\right).$$
2) If $m \leqslant h$, $m|h$, and $\delta-1 \leqslant h-\frac{2h}{m}$:
$$q = \Omega\left(n^{1+\frac{m(\delta-1)}{h}}\right).$$
3) If $m \leqslant h$, $m|h$, and $\delta-1 > h-\frac{2h}{m}$:
$$q = \Omega\left(n^{m-1}\right).$$

Remark 1: For the case $h=1$, the field size requirement of an (n,r,h,δ,q)-MR code may be as small as $q=\Theta(r+\delta-1)$. This is attainable since the punctured code over any repair set together with the single global parity check is an $[r+\delta,r,\delta+1]_q$ MDS code when $(r+\delta-1)|k$ or $u \geqslant 2(r-v+1)$, where $k=ur+v$ with $0 \leqslant v \leqslant r-1$ (see [7]).

Definition 4: A family of (n,r,h,δ,q)-MR codes has *order-optimal field size* if it attains one of the bounds of Lemma 4 asymptotically for $h \geqslant 2$, or if it has $q=\Theta(r+\delta-1)$ for $h=1$.

III. A New Bound on Optimal LRCs

In this section we present a new bound on the parameters of optimal LRCs with all-symbol (r,δ)-locality. This bound is not specific to MR codes or cyclic codes. The bound does, however, require certain divisibility conditions, which are common to several constructions of optimal LRCs. We proceed by describing two base cases in the next two lemmas. We then recall a parameter reduction lemma. The combination of these three parts results in the main bound.

Let us begin with the case $2 \mid r$.

Lemma 5: Let $\mathcal{C}$ be an optimal $[n=(u+1)(r+\delta-1), ur, r+2\delta-1]_q$ LRC with all-symbol (r,δ)-locality. If $2|r$, then
$$u+1 \leqslant (q^{r/2}+1)\Big/\left\lfloor\frac{2r+2\delta-2}{r}\right\rfloor.$$

For the case $2 \nmid r$, we also have a similar lemma.

Lemma 6: Let $\mathcal{C}$ be an optimal $[n=(u+2)(r+\delta-1), ur, 2r+3\delta-2]_q$ LRC with all-symbol (r,δ)-locality. If $2 \nmid r$, then
$$u \leqslant q^{(r+1)/2}.$$

The final component in our main bounding theorem is a parameter-reduction lemma. This lemma was proved in [7].

Lemma 7 ([7] Corollary 2)**:** Let $n=m(r+\delta-1)$, $\delta \geqslant 2$, $k=ur+v>r$, and additionally, $r|k$ or $u \geqslant 2(r+1-v)$, where all parameters are integers. If there exists an optimal $[n,k,d]_q$ linear code $\mathcal{C}$ with $d>r+\delta$ and all-symbol (r,δ)-locality, then there exists an optimal linear code $\mathcal{C}'$ with all-symbol (r,δ)-locality and parameters $[n-\epsilon(r+\delta-1), k, d'=d-\epsilon(r+\delta-1)]_q$, where $\epsilon \leqslant \lceil(d-1)/(r+\delta-1)\rceil-1$.

Combining the preceding lemmas, we prove the next theorem, which gives a lower bound on the size of the field required for LRCs with all-symbol (r,δ)-locality.

Theorem 1: Let $\mathcal{C}$ be an optimal $[n,k,d]_q$ linear code with all-symbol (r,δ)-locality. Assume $n=m(r+\delta-1)$, $k=ur$, $u \geqslant 2$. If $2|r$ and $m \geqslant u+1$ then,
$$q \geqslant \psi\left(\left(\left(\frac{k}{r}+1\right)\left\lfloor\frac{2r+2\delta-2}{r}\right\rfloor-1\right)^{\frac{2}{r}}\right),$$
where $\psi(x)$ is the smallest prime power greater or equal to x. If $2 \nmid r$ and $m \geqslant u+2$ then
$$q \geqslant \psi\left(\left(\frac{k}{r}\right)^{\frac{2}{r+1}}\right).$$

The new bound of Theorem 1 has some implications which we now discuss. The case of $r=2$ is of particular interest, since we can then use Theorem 1 to prove that some known LRCs have optimal field size. We first consider some Tamo-Barg codes [33].

Lemma 8 ([33])**:** Let q be a prime power, $q=r+\delta-1$, then there exists an optimal LRC with all-symbol (r,δ)-locality and parameters $[q^b, ur, (q^{b-1}-u)q+\delta]_{q^b}$, where $b \geqslant 2$ and $0<u<q^{b-1}$.

Corollary 1: Let $\mathcal{C}$ be a code from Lemma 8 with $r=2$ and $u=q^{b-1}-1$. If q^b-1 is not a prime power then $\mathcal{C}$ has optimal field size.

Example 1: Let $n=2^4$, $r=2$, $\delta=3$, then by Lemma 8 there exists an optimal LRC with all-symbol $(2,3)$-locality and parameters $[16,6,7]_{2^4}$, which has optimal field size since 15 is not a prime power.

We now examine a construction of cyclic optimal LRCs from [34].

Lemma 9 ([34])**:** Let $r = 2$, $n = m(r + \delta - 1) = q^b - 1$, and $k = ur + v$ with $0 \leqslant v < r$, where q^b is prime power. Then there exists a cyclic optimal LRC with all-symbol $(2, \delta)$-locality and parameters $[q^b - 1, k, d]_{q^b}$.

Corollary 2: Let $\mathcal{C}$ be a code from Lemma 9 with $m = u + 1$ and $v = 0$. If neither $q^b - 2$, nor $q^b - 1$, are prime powers, then $\mathcal{C}$ has optimal field size.

Example 2: Let $n = 2^6 - 1$, $r = 2$, and $\delta = 2$. Then by Lemma 9, there exists a cyclic optimal LRC with all-symbol $(2, 2)$-locality, and parameters $[63, 40, 5]_{2^6}$, which has optimal field size since both 62 and 63 are not prime powers.

Yet another construction of cyclic optimal LRCs comes from [12].

Lemma 10 ([12])**:** Let $r = 2$, $\delta = 2$, $n = m(r + \delta - 1) = 3m = q^b + 1$, and $k = 2u$, with u an even integer, and where q^b is prime power. Then there exists a cyclic optimal LRC with all-symbol $(2, 2)$-locality and parameters $[q^b + 1 = 3m, 2u, d]_{q^m}$.

Corollary 3: Let $\mathcal{C}$ be a code from Lemma 10 with $m = u + 1$. Then $\mathcal{C}$ has optimal field size.

Example 3 ([12])**:** Let $n = 9 = 2^3 + 1$, $r = 2$, $\delta = 2$, $k = 4$, then there exists a cyclic optimal $[9, 4, 5]_8$-LRC, which has optimal field size.

IV. Cyclic Maximally Recoverable (MR) Codes

This section is divided into two parts. In the first part we construct cyclic MR codes, and show that for certain parameters they have the exact optimal field size. In the second part we study a known class of MR codes which are non-cyclic, but have the same parameters as the cyclic codes we construct. We then show that these non-cyclic codes can sometimes be permuted to obtain cyclic MR codes.

A. A New Construction

We immediately present our construction for cyclic MR codes. It is inspired by the construction of [34].

Construction A: Let $b, r, \delta \geqslant 2$ be integers, q a prime power, $n = q^b - 1$, $\alpha \in \mathbb{F}_{q^b}$ a primitive element, $a = (r + \delta - 1) | (q - 1)$, and $m = n/a$. Define

$$R \triangleq \left\{ \alpha^{ja+t} \ : \ 1 \leqslant j \leqslant m, 1 \leqslant t \leqslant \delta - 1 \right\} \cup \left\{ 1, \alpha^{\delta} \right\}.$$

The constructed code, $\mathcal{C}$, is the cyclic code of length n over $\mathbb{F}_{q^b}$ with root set R.

Our goal is now to show that the code from Construction A is indeed a cyclic MR code, which is described in the following theorem.

Theorem 2: Assume the setting and notation of Construction A. Then the code $\mathcal{C}$ of Construction A is a cyclic $(n = q^b - 1, r, h = 2, \delta, q^b)$-MR code, equivalently, a cyclic MR code with parameters $[n = q^b - 1, k = mr - 2, d]_{q^b}$ with repair sets of size $r + \delta - 1$, and

$$d = \begin{cases} \delta + 2 & r > 2, \\ 2\delta + 1 & r = 2. \end{cases}$$

The cyclic MR codes by Construction A have optimal Hamming distance, and order-optimal field size with respect to the bound in Lemma 4-(1), where we consider $\delta \geqslant 2$ as a constant. However, we can do better than that when $r = 2$.

Theorem 3: Let $\mathcal{C}$ be an $(n, r = 2, h = 2, \delta, q)$-MR code. Then $q \geqslant n - 1$.

Corollary 4: When $r = 2$, the cyclic MR codes generated by Construction A have optimal field size by Theorem 1, provided that neither $q^b - 1$, nor $q^b - 2$, are prime powers. When $r > 2$, the cyclic MR codes by Construction A have optimal Hamming distance, and order-optimal field size with respect to the bound in Lemma 4-(1), where we consider $\delta \geqslant 2$ as a constant.

B. Turning Non-cyclic Codes into Cyclic Codes

Previous works that constructed *non-cyclic* (n, r, h, δ, q)-MR codes, for $h = 2$, did so with $q = \Theta(n(\delta - 1))$ in [4], and later, with $q = \Theta(n)$ [18] (see also [21], that obtained $q = \Theta(n)$ for the special case of $n = 2(r + \delta - 1)$). Of particular interest to us are the $(n, r, 2, \delta, q^b)$-MR codes from [18, Theorem IV.2]. These MR codes have the same parameters as Construction A. However, they are not cyclic MR codes directly. In what follows, we shall attempt to determine whether the MR codes generated in [18, Theorem IV.2] can be rearranged to become cyclic codes. Along the way, we shall prove some interesting facts concerning cyclic optimal LRCs.

As a first step, we show the repair sets of cyclic optimal LRCs are severely restricted.

Theorem 4: Let $\mathcal{C}$ be a cyclic optimal LRC with parameters $[n, k, d]_q$ and all-symbol (r, δ)-locality. Write $k = ur + v$ with $0 < v \leqslant r$. If $u \geqslant 2(r - v + 1)$, then for any repair set $S \subseteq \mathbb{Z}_n$, and any $j \in \mathbb{Z}_n$, either $S + j = S$ or $(S + j) \cap S = \emptyset$.

As an immediate consequence, we now show that the repair sets of cyclic optimal LRCs must be cosets of $\mathbb{Z}_n$.

Corollary 5: Let $\mathcal{C}$ be a cyclic optimal LRC with parameters $[n, k, d]_q$ and all-symbol (r, δ)-locality (where, to avoid trivial cases, we assume that $\mathcal{C}$ does not have all-symbol $(r - 1, \delta)$-locality). Let $k = ur + v$, $0 < v \leqslant r$. If $u \geqslant 2(r - v + 1)$, then $n = m(r + \delta - 1)$, $m \in \mathbb{N}$, and the repair sets of $\mathcal{C}$ are

$$G_i \triangleq \langle m \rangle + i = \{ jm + i \ : \ j \in \mathbb{Z} \} \subseteq \mathbb{Z}_n, \text{ for all } i \in \mathbb{Z}.$$

Remark 2: Corollary 5 shows that the condition $(r + \delta - 1)|n$ is not a restriction when $u \geqslant 2(r - v + 1)$, but rather a consequence.

Remark 3: For the case $u = 1$ (i.e., $k = r + v$), and $(r + \delta - 1) \nmid n$, explicit constructions were proposed in [30, Corollaries 27, 37, 43] for cyclic optimal LRCs. Corollary 5 implies that constructions with such parameters are possible only if $1 = u < 2(r - v + 1)$, i.e., $r \geqslant v$.

Now we recall a construction, which was first introduced in [18].

Construction B ([18])**:** Let q be a prime power, $b \in \mathbb{N}$, $n = q^b - 1$, $a = r + \delta - 1$, $a|(q-1)$, and $m = n/a$. Let α be a primitive element of $\mathbb{F}_{q^b}$, $\beta = \alpha^m$, and $\lambda = \alpha^a$. The following parity-check matrix defines an $(n, r, 2, \delta, q^b)$-MR code,

$$H^T = \left(\begin{array}{cccccc}
\mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\lambda}^0 & 1 \\
\mathbf{0} & \mathbf{1} & \cdots & \mathbf{0} & \boldsymbol{\lambda}^1 & 1 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} & \boldsymbol{\lambda}^{m-1} & 1 \\
\hline
\boldsymbol{\beta} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\lambda}^0 & 1 \\
\mathbf{0} & \boldsymbol{\beta} & \cdots & \mathbf{0} & \boldsymbol{\lambda}^1 & 1 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\beta} & \boldsymbol{\lambda}^{m-1} & 1 \\
\hline
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
\hline
\boldsymbol{\beta^{m-1}} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\lambda}^0 & 1 \\
\mathbf{0} & \boldsymbol{\beta^{m-1}} & \cdots & \mathbf{0} & \boldsymbol{\lambda}^1 & 1 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\beta^{m-1}} & \boldsymbol{\lambda}^{m-1} & 1
\end{array}\right),$$

where, to simply our notation, we define,

$$\boldsymbol{x}^T \triangleq (x\, x^2\, \cdots\, x^{\delta-1}).$$

One cannot avoid seeing a similarity between the parity-check matrix of Construction B, and the parity-check matrix for the code from Construction A. However, the code from Construction B is not cyclic, but rather quasi-cyclic. In what follows we study whether permuting it produces a cyclic code.

Let $\mathbb{S}_n$ denote the set of permutations over $\mathbb{Z}_n$, for any $n \in \mathbb{N}$. Each permutation in $\mathbb{S}_n$ may be thought of as a bijection in $\mathbb{Z}_n^{\mathbb{Z}_n}$, namely, a bijection from $\mathbb{Z}_n$ to $\mathbb{Z}_n$. Let $\mathcal{C}$ be a code of length n, whose coordinates are indexed by $\mathbb{Z}_n$. If $\ell \in \mathbb{S}_n$ is a permutation, we define the permutation of $\mathcal{C}$ by ℓ as

$$\mathcal{C}_\ell \triangleq \left\{(c_{\ell(0)}, c_{\ell(1)}, \dots, c_{\ell(n-1)}) \;:\; (c_0, c_1, \dots, c_{n-1}) \in \mathcal{C})\right\}.$$

If $\mathcal{C}$ is a cyclic code, it is natural to ask for what permutations $\ell \in \mathbb{S}_n$, $\mathcal{C}_\ell$ is also cyclic. Apart from the trivial cyclic shifts of $\mathcal{C}$, a natural subset of candidate permutations are *multipliers*, namely,

$$\begin{aligned}
\mu_t(x) &\triangleq xt \bmod n, \\
\mathbb{Z}_n^\times &\triangleq \{1 \leqslant t \leqslant n \;:\; \gcd(t, n) = 1\}, \\
\Upsilon(n) &\triangleq \left\{\mu_t \;:\; t \in \mathbb{Z}_n^\times\right\}.
\end{aligned}$$

Pálfy [26] proved that, in many cases, multipliers are the essential permutations keeping a code cyclic:

Lemma 11 ([26])**:** Consider codes of length n whose coordinates are indexed by $\mathbb{Z}_n$.

1) When $\gcd(n, \varphi(n)) = 1$ or $n = 4$, for all cyclic codes $\mathcal{C}$, if $\mathcal{C}_{\ell'}$, $\ell' \in \mathbb{S}_n$, is also a cyclic code, then there is a multiplier $\ell \in \Upsilon(n)$ such that $\mathcal{C}_{\ell'} = \mathcal{C}_\ell$.
2) When $\gcd(n, \varphi(n)) \neq 1$ and $n \neq 4$, there exists a cyclic code $\mathcal{C}$, and $\ell' \in \mathbb{S}_n$ such that $\mathcal{C}_{\ell'}$ is cyclic, but $\mathcal{C}_{\ell'} \neq \mathcal{C}_\ell$ for all multipliers $\ell \in \Upsilon(n)$.

Here $\varphi(\cdot)$ denotes Euler's totient function.

Drawing inspiration from Lemma 11, we address the (different) question of finding permutations from $\mathbb{S}_n$ that turn the non-cyclic code of Construction B into a cyclic code. Recall that in the setting of Construction B, $a, m, n \in \mathbb{N}$, and $n = ma$. We now define a set of functions from $\mathbb{Z}_n$ to $\mathbb{Z}_n$ as follows:

$$\mu_{t,z}(xm + i) \triangleq (xmt_i + z_i) \bmod n,$$

where we assume $x \in [a]$, $i \in [m]$, $t = (t_0, \dots, t_{m-1}) \in \mathbb{Z}^m$, and $z = (z_0, \dots, z_{m-1}) \in \mathbb{Z}^m$. We then define the set,

$$\Psi(n, a) \triangleq \left\{\mu_{t,z} \;:\; t \in (\mathbb{Z}_a^\times)^m, z \in \mathbb{Z}^m, (z \bmod m) \in \mathbb{S}_m\right\},$$

and where by abuse of notation, $z \bmod m$ denotes the $\mathbb{Z}_m \to \mathbb{Z}_m$ mapping that maps $i \mapsto (z_i \bmod m)$.

We would like to make some easy observations concerning the elements of $\Psi(n, a)$. Denote $G_0 \triangleq \langle m \rangle \subseteq \mathbb{Z}_n$. Then G_0 is an additive subgroup of $\mathbb{Z}_n$, and $G_0 \cong \mathbb{Z}_a$. Let us denote the cosets of G_0 by $G_i \triangleq G_0 + i$, for all $i \in \mathbb{Z}$. We now note that $j \mapsto jt \bmod n$ is a bijection from G_0 to G_0 if and only if $\gcd(t, a) = 1$. Thus, $\ell_{t,z}|_{G_i}$ (i.e., the restriction of $\ell_{t,z}$ to G_i) is a bijection from G_i to $G_{z_i \bmod m}$. With the extra requirement that $(z \bmod m) \in \mathbb{S}_m$, we have that distinct cosets G_i are mapped to distinct cosets $G_{z_i \bmod m}$, and hence, $\Psi(n, a) \subseteq \mathbb{S}_n$, namely, Ψ comprises of permutations over $\mathbb{Z}_n$.

Theorem 5: Assume the notation and setting of Construction B, and let $\mathcal{C}$ be the resulting code when $r \geqslant 3$. Then there exists a permutation $\ell \in \Psi(n, a)$ such that $\mathcal{C}_\ell$ is a cyclic code if and only if $\gcd(m, \frac{a}{\gcd(a,\delta)}) = 1$.

While the last theorem shows us a sufficient condition under which the known code of Construction B may be permuted to a cyclic code, the next theorem shows us that for almost all cases, this condition is in fact necessary.

Theorem 6: Assume the notation and setting of Construction B. Let $\mathcal{C}$ be the resulting code. Denote $k = \dim(\mathcal{C}) = ur + v$ with $0 < v \leqslant r$ and $u \geqslant 2(r - v + 1)$. Additionally, let $a = 4$ or $\gcd(a, \varphi(a)) = 1$. Furthermore, assume that $a = q^{b'} - 1 | q^b - 1 = n$, and that one of the following holds:

1) $\delta \geqslant 4$ and $r \geqslant 5$
2) $\delta = 3$ and $r \geqslant 4$
3) $\delta = 2$ and $r \geqslant 3$ is odd

Then there exists a permutation $\ell \in \mathbb{S}_n$ such that $\mathcal{C}_\ell$ is cyclic if and only if $\gcd(m, \frac{a}{\gcd(\delta,a)}) = 1$.

To conclude this section, we make use of Theorem 6 in order to show that Construction A may produce cyclic MR codes with new parameters. Namely, in certain case, the construction of [18], which produces codes with the same parameters as our Construction A, results in codes that are neither cyclic, nor can be permuted to become cyclic.

Example 4: Set $q = 3$, $b_1 = 2$, $b = 4$, $r = 6$, $\delta = 3$, $a = 8$, $n = 80$, and $m = 10$. By using Construction A, we may generate a cyclic $(n = 80, r = 6, h = 2, \delta = 3, q^b = 3^4)$-MR code. A *non-cyclic* MR code with the same parameters may be constructed using [18]. However, since $\gcd(m, \frac{a}{\gcd(a,\delta)}) = \gcd(8, 10) = 2 \neq 1$, by Theorem 6 this code cannot be permuted to become a cyclic code.

References

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

[2] A. Beemer, R. Coatney, V. Guruswami, H. H. López, and F. Piñero, "Explicit optimal-length locally repairable codes of distance 5," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2018, pp. 800–804.

[3] M. Blaum, J. L. Hafner, and S. Hetzler, "Partial-MDS codes and their application to RAID type of architectures," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4510–4519, 2013.

[4] M. Blaum, J. S. Plank, M. Schwartz, and E. Yaakobi, "Construction of partial MDS and sector-disk codes with two global parity symbols," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2673–2681, 2016.

[5] V. R. Cadambe and A. Mazumdar, "Bounds on the size of locally recoverable codes," *IEEE transactions on information theory*, vol. 61, no. 11, pp. 5787–5794, 2015.

[6] H. Cai, C. Fan, Y. Miao, M. Schwartz, and X. Tang, "Optimal locally repairable codes: An improved bound and constructions," *IEEE Transactions on Information Theory*, 2022, to appear.

[7] H. Cai, Y. Miao, M. Schwartz, and X. Tang, "On optimal locally repairable codes with super-linear length," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 4853–4868, 2020.

[8] ——, "A construction of maximally recoverable codes with order-optimal field size," *IEEE Transactions on Information Theory*, vol. 68, no. 1, pp. 204–212, 2022.

[9] H. Cai and M. Schwartz, "On optimal locally repairable codes and generalized sector-disk codes," *IEEE Transactions on Information Theory*, vol. 67, no. 2, pp. 686–704, 2021.

[10] ——, "A bound on the minimal field size of LRCs, and cyclic MR codes that attain it," *arXiv preprint arXiv:2201.00344*, 2022.

[11] B. Chen, W. Fang, S.-T. Xia, J. Hao, and F.-W. Fu, "Improved bounds and Singleton-optimal constructions of locally repairable codes with minimum distance 5 and 6," *IEEE Transactions on Information Theory*, vol. 67, no. 1, pp. 217–231, 2021.

[12] B. Chen, S.-T. Xia, J. Hao, and F.-W. Fu, "Constructions of optimal cyclic (r, δ) locally repairable codes," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2499–2511, 2018.

[13] Z. Chen and A. Barg, "Cyclic LRC codes with hierarchy and availability," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 616–621.

[14] R. Gabrys, E. Yaakobi, M. Blaum, and P. H. Siegel, "Constructions of partial MDS codes over small fields," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3692–3701, 2019.

[15] P. Gopalan, C. Huang, B. Jenkins, and S. Yekhanin, "Explicit maximally recoverable codes with locality," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5245–5256, 2014.

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

[17] S. Gopi and V. Guruswami, "Improved maximally recoverable LRCs using skew polynomials," *arXiv preprint arXiv:2012.07804*, 2020.

[18] S. Gopi, V. Guruswami, and S. Yekhanin, "Maximally recoverable LRCs: A field size lower bound and constructions for few heavy parities," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6066–6083, 2020.

[19] V. Guruswami, L. Jin, and C. Xing, "Constructions of maximally recoverable local reconstruction codes via function fields," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6133–6143, 2020.

[20] V. Guruswami, C. Xing, and C. Yuan, "How long can optimal locally repairable codes be?" *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3662–3670, 2019.

[21] G. Hu and S. Yekhanin, "New constructions of SD and MR codes over small finite fields," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 1591–1595.

[22] C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," *ACM Transactions on Storage (TOS)*, vol. 9, no. 1, pp. 1–28, 2013.

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

[24] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Elsevier, 1977.

[25] U. Martínez-Peñas and F. R. Kschischang, "Universal and dynamic locally repairable codes with maximal recoverability via sum-rank codes," *IEEE Transactions on Information Theory*, vol. 65, no. 12, pp. 7790–7805, 2019.

[26] P. P. Pálfy, "Isomorphism problem for relational structures with a cyclic automorphism," *European Journal of Combinatorics*, vol. 8, no. 1, pp. 35–43, 1987.

[27] J. S. Plank and M. Blaum, "Sector-disk (SD) erasure codes for mixed failure modes in RAID systems," *ACM Transactions on Storage (TOS)*, vol. 10, no. 1, pp. 1–17, 2014.

[28] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, "Optimal linear codes with a local-error-correction property," in *2012 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2012, pp. 2776–2780.

[29] J. Qian and L. Zhang, "New optimal cyclic locally recoverable codes of length $n = 2(q + 1)$," *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 233–239, 2019.

[30] J. Qiu, D. Zheng, and F.-W. Fu, "New constructions of optimal cyclic (r, δ) locally repairable codes from their zeros," *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1596–1608, 2021.

[31] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimal locally repairable and secure codes for distributed storage systems," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 212–236, 2014.

[32] W. Song, S. H. Dau, C. Yuen, and T. J. Li, "Optimal locally repairable linear codes," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 5, pp. 1019–1036, 2014.

[33] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4661–4676, 2014.

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

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

[36] T. Westerbäck, R. Freij-Hollanti, T. Ernvall, and C. Hollanti, "On the combinatorics of locally repairable codes via matroid theory," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5296–5315, 2016.

[37] C. Xing and C. Yuan, "Construction of optimal locally recoverable codes and connection with hypergraph," in *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.