Cyclic Linear Binary Locally Repairable Codes

Pengfei Huang[*], **Eitan Yaakobi**[‡], **Hironori Uchikawa**[*†], and **Paul H. Siegel**[*]
[*]Electrical and Computer Engineering Dept., University of California, San Diego, La Jolla, CA 92093 U.S.A
[†]Toshiba Corporation, Japan
[‡]Computer Science Dept., Technion – Israel Institute of Technology, Haifa 32000, Israel
{*pehuang,huchikawa,psiegel*}*@ucsd.edu*, *yaakobi@cs.technion.ac.il*

***Abstract*—Locally repairable codes (LRCs) are a class of codes designed for the local correction of erasures. They have received considerable attention in recent years due to their applications in distributed storage. Most existing results on LRCs do not explicitly take into consideration the field size q, i.e., the size of the code alphabet. In particular, for the binary case, only a few specific results are known by Goparaju and Calderbank. Recently, however, an upper bound on the dimension k of LRCs was presented by Cadambe and Mazumdar. The bound takes into account the length n, minimum distance d, locality r, and field size q, and it is applicable to both non-linear and linear codes.**

In this work, we first develop an improved version of the bound mentioned above for linear codes. We then focus on cyclic linear binary codes. By leveraging the cyclic structure, we notice that the locality of such a code is determined by the minimum distance of its dual code. Using this result, we investigate the locality of a variety of well known cyclic linear binary codes, e.g., Hamming codes and Simplex codes, and also prove their optimality with our improved bound for linear codes. We also discuss the locality of codes which are obtained by applying the operations of Extend, Shorten, Expurgate, Augment, and Lengthen to cyclic linear binary codes. Several families of such modified codes are considered and their optimality is addressed. Finally, we investigate the locality of Reed-Muller codes. Even though they are not cyclic, it is shown that some of the locality results for cyclic codes still apply.

I. Introduction

Distributed and cloud storage systems today are required to tolerate the failure or unavailability of some of the nodes in the system. The simplest and most commonly used way to accomplish this task is replication, where every node is replicated several times, usually three. This solution has clear advantages due to its simplicity and fast recovery from node failures. However, it entails a large storage overhead which becomes costly in large storage systems.

In order to achieve better storage efficiency, erasure codes, e.g., Reed-Solomon codes, are deployed. Reed-Solomon and in general many maximum distance separable (MDS) codes are attractive since they tolerate the maximum number of node failures. However they suffer from a very slow recovery process, in particular, a slow repairing of a single node failure, which is the common failure scenario. Hence, an important task in the design of erasure codes is to accomplish fast recovery and yet support a large number of node failures. There are several metrics in the literature to quantify the efficiency of rebuilding. Three of the more popular consider the number of communicated bits in the network, the number of read bits, or the number of accessed nodes. In this work we study codes with respect to the last metric.

Locally repairable codes (LRCs) are a class of codes in which a failure of a single node can be recovered by accessing at most r other nodes, where r is a predetermined value [6], [12], [14]. For a length-n code with dimension k, it is said that the code has *all symbol locality* r if every symbol is recoverable from a set of at most r symbols. If the code is systematic and only its information symbols have this property then the code has *information locality* r. LRCs are well studied in the literature and many works have considered code constructions and bounds on such codes. In [6], an upper bound, which can be seen as a modified version of the Singleton bound, was given on the minimum distance of LRCs. More specifically, if an $[n,k,d]$ linear code has information locality r then

$$d \leqslant n-k-\left\lceil \frac{k}{r} \right\rceil + 2. \tag{1}$$

In [14], it was proved that the bound (1) holds also for non-linear codes with all symbol locality. However, for some cases, the bound (1) is not tight. Thus, several improvements on the bound (1) were proposed in [17], [24]. Code constructions which achieve the bound (1) were given in [21]–[23]. However, in most cases, in order to attain this bound the codes have to be non-binary and the problem of finding codes over small alphabet which satisfy the bound (1) was recently solved by Tamo and Barg [22]. In this last construction the field size has to be at least the length of the code. In addition to the constructions mentioned above there are several other constructions of LRCs, see e.g., [4], [7]–[9], [12], [19].

To the best of our knowledge, the problem of finding explicit code constructions over a fixed alphabet has not been fully addressed in the literature. Recently, a new upper bound on the dimension k of LRCs was presented in [2]. This bound takes into account the code length, minimum distance, locality, and field size, and it is applicable to both non-linear and linear codes. Namely, if a length-n code with M codewords and minimum distance d has all symbol locality r, then

$$k \leqslant \min_{t\in\mathbb{Z}^+} \left\{ tr + k_{opt}^{(q)}(n-t(r+1),d) \right\}, \tag{2}$$

where $k = \log_q M$, $\mathbb{Z}^+$ is the set of all positive integers, and $k_{opt}^{(q)}(n',d')$ is the largest possible dimension of a length-n' code with minimum distance d' and a given alphabet size q. The only construction that we know of for binary LRCs was presented recently by Goparaju and Calderbank [7].

Our main goal in this paper is to study constructions of binary LRCs, and in particular cyclic linear codes. In Section II, we formally define the problem and state some preliminary results. We also show in this section that the bound (2) from [2] can be improved for linear codes. In Section III, we prove the locality of cyclic linear codes and show that such a code has locality that equals the minimum distance of its dual code minus one. In Section IV, we study the locality of codes which are obtained by the operations of Extend, Shorten, Expurgate, Augment, and Lengthen. In Section V, we study similar properties for Reed-Muller codes. We provide examples to our statements and prove their optimality by existing bounds and our improved bound for linear codes. We conclude the paper in Section VI.

II. Definitions and Preliminaries

In this section, we give the basic definitions and preliminaries that will be used in the paper. We use the notation $[n]$ to define the set $\{1,\ldots,n\}$. For a length-n vector $\boldsymbol{v}$ and a set $\mathcal{I}\subseteq[n]$, we denote by $\boldsymbol{v}_{\mathcal{I}}$ the vector $\boldsymbol{v}$ punctured on the set $\mathcal{I}$. A linear code over $GF(q)$ of length n, dimension k, and minimum distance d will be denoted by $[n,k,d]_q$, and a non-linear code will be denoted by (n,M,d) where M is the number of codewords. The dual code of a linear code $\mathcal{C}$ will be denoted by $\mathcal{C}^{\perp}$. We use the notation $S(\boldsymbol{c})$ to denote the support of a codeword $\boldsymbol{c}$. We follow the conventional definition of locally repairable codes [15], [16], [22], which is stated as follows.

Definition 1. *The ith code symbol, $i\in[n]$, is said to have locality r if there exists a repair set R_i of size at most r, such that if it is erased then it can be recovered by solely reading the symbols from the set R_i. A code $\mathcal{C}$ is said to have* ***all symbol locality*** *r if all its symbols have locality r. Similarly, a systematic code $\mathcal{C}$ is said to have* ***information locality*** *r if all its information symbols have locality r.*

Let us first show that the bound (2) can be improved for the family of linear codes with all symbol locality. First, we denote by $k^{(q)}_{\ell-opt}(n',d')$ the largest possible dimension of a length-n' linear code with minimum distance d' over a given alphabet of size q.

Lemma 2. *For any $[n,k,d]_q$ linear code with all symbol locality r, we have*

$$k\leqslant \min_{1\leqslant t\leqslant\lceil\frac{k}{r}\rceil-1,\ t\in\mathbb{Z}^+}\left\{tr+k^{(q)}_{\ell-opt}(n-t(r+1),d)\right\}. \quad (3)$$

Proof: We follow the proof from [2] which consists of two parts. The first part is the same and thus we use Lemma 1 from [2], stated here for linear codes: Let $\mathcal{C}$ be an $[n,k,d]_q$ linear code with all symbol locality r. Then, for all $1\leqslant t\leqslant k/r$, $t\in\mathbb{Z}^+$, there exists a set $\mathcal{I}\subseteq[n]$, such that $|\mathcal{I}|=t(r+1)$, and $k_{\mathcal{I}}\leqslant tr$, where $k_{\mathcal{I}}=\log_q|\{\boldsymbol{c}_{\mathcal{I}}:\boldsymbol{c}\in\mathcal{C}\}|$.

The second part of our proof is different from [2]. The key point in our proof is to identify a shortened linear code, and then study the bound of this specific shortened linear code, while the proof in [2] is based on a probabilistic counting argument. Let t be a positive integer, $1\leqslant t\leqslant\lceil\frac{k}{r}\rceil-1$, and $\mathcal{I}\subseteq[n]$ be as constructed in the first part. Then, we consider the code $\mathcal{C}^{\boldsymbol{0}}_{\mathcal{I}}=\{\boldsymbol{c}_{[n]\setminus\mathcal{I}}:\boldsymbol{c}_{\mathcal{I}}=\boldsymbol{0}\text{ and }\boldsymbol{c}\in\mathcal{C}\}$. Since the code $\mathcal{C}$ is linear, the size of the code $\mathcal{C}^{\boldsymbol{0}}_{\mathcal{I}}$ is $q^{k-k_{\mathcal{I}}}$ and it is a linear code as well. Moreover, the minimum distance D of the code $\mathcal{C}^{\boldsymbol{0}}_{\mathcal{I}}$ is at least d, i.e., $D\geqslant d$. Thus, we get the bounds

$$k-k_{\mathcal{I}}\leqslant k^{(q)}_{\ell-opt}(n-|\mathcal{I}|,D)\leqslant k^{(q)}_{\ell-opt}(n-|\mathcal{I}|,d).$$

Therefore, we conclude that

$$k\leqslant k^{(q)}_{\ell-opt}(n-|\mathcal{I}|,d)+k_{\mathcal{I}}\leqslant k^{(q)}_{\ell-opt}(n-t(r+1),d)+tr.$$

■

Note that, for linear codes, our bound (3) is strictly tighter than the bound (2) in some cases. For example, for parameters $n=28$, $d=6$, $r=8$, and $q=2$, bound (3) and the online table for $k^{(2)}_{\ell-opt}$ [20] gives $k\leqslant 17$, whereas bound (2) and the online table for $k^{(2)}_{opt}$ [1] gives $k\leqslant 18$.

In many papers, a code with all symbol locality which attains the bound (1) is called an *optimal LRC*. However, due to the ceiling operation in the bound, there may exist two codes with the same parameters n, k, and d, but different localities, both achieving bound (1). Obviously, although they are both optimal LRCs, the code which has smaller locality is a better one. Thus, we prefer a slightly different definition of optimality which depends on the code properties rather than on specific bounds.

Definition 3. *An $[n,k,d]_q$ linear code $\mathcal{C}$ with locality r is said to be d-**optimal**, if there does not exist an $[n,k,d+1]_q$ code with locality r. Similarly, it is called k-**optimal** if there does not exist an $[n,k+1,d]_q$ code with locality r, and it is called r-**optimal** if there does not exist an $[n,k,d]_q$ code with locality $r-1$.*

Example 1. Let us consider the binary Simplex code $\mathcal{C}$ with parameters $[2^m-1,m,2^{m-1}]$. It was proved in [2] that this code has all symbol locality $r=2$ and it is r-optimal for these given parameters. Since this code satisfies the Plotkin bound, it is d-optimal and k-optimal as well.

Unless stated otherwise, all codes mentioned in the rest of the paper are binary codes. We consider only codes with all symbol locality, and thus when saying that a code has locality r we refer to all symbol locality.

III. Cyclic Linear LRCs

In this section, we give our main result for cyclic linear LRCs. We also present several examples. We start with a simple result on the locality of the code symbols. Even though it has been mentioned before, see e.g., [6], [7], [13], [16], [17], we state and prove it here for completeness.

Claim 4. *For a binary linear code $\mathcal{C}$, if its ith coordinate, $i\in[n]$, belongs to the support of a codeword in $\mathcal{C}^{\perp}$ with weight $r+1$, then the ith code symbol has locality r.*

Proof: Assume that there exists a codeword $\boldsymbol{c}'\in\mathcal{C}^{\perp}$ such that $c'_i=1$ and $w_H(\boldsymbol{c}')=r+1$. Let $R_i=S(\boldsymbol{c}')\setminus\{i\}$. Then for all $\boldsymbol{c}\in\mathcal{C}$, $c_i=\sum_{j\in R_i}c_j$ and from Definition 1, the ith symbol has locality r. ■

The next observation is an immediate consequence of the preceding claim.

Observation 5. *Let $\mathcal{C}$ be an $[n,k,d]$ cyclic linear code, and let $d^{\perp}$ be the minimum distance of its dual code $\mathcal{C}^{\perp}$. Then, the code $\mathcal{C}$ has locality $d^{\perp}-1$.*

Next, we give several examples to illustrate how the locality of codes can be determined from Observation 5 and then study their optimality.

Example 2. Let $\mathcal{C}$ be the $[n=2^m-1,k=2^m-1-m,d=3]$ cyclic binary Hamming code. Its dual code is the $[2^m-1,m,2^{m-1}]$ cyclic binary Simplex code. Therefore, the Hamming code has locality $r=2^{m-1}-1$. Since it is a perfect code, it is both d-optimal and k-optimal. In order to show r-optimality, let us assume to the contrary that there exists an $[n,k,d]$ code with locality $\hat{r}=2^{m-1}-2$. According to the bound in (2) for $t=1$, we have that

$$\begin{aligned}k\leqslant& t\hat{r}+k^{(2)}_{opt}(n-t(\hat{r}+1),d)=2^{m-1}-2+k^{(2)}_{opt}(2^{m-1},3)\\ \overset{(a)}{<}&2^{m-1}-2+2^{m-1}-(m-1)=2^m-m-1,\end{aligned}$$

where step (a) is from the Hamming bound. Thus, we get a contradiction to the value of k. We also get from Observation 5 that the Simplex code has locality 2. This gives an alternative proof to the one given in [2] in case the code is cyclic.

TABLE I
OPTIMALITY OF DBCH CODES AND ITS DUAL.

$\mathcal{C}$	n	k	d	r	d-opt	k-opt	r-opt
$m=4$	15	7	5	3	✓	✓	✓
$m=5$	31	21	5	11	✓	✓	?
$m=6$	63	51	5	23	✓	✓	?
$m=7$	127	113	5	55	✓	✓	?
$m=8$	255	239	5	111	✓	✓	?
$\mathcal{C}^\perp$	$n^\perp$	$k^\perp$	$d^\perp$	$r^\perp$	d-opt	k-opt	r-opt
$m=4$	15	8	4	4	✓	?	?
$m=5$	31	10	12	4	✓	✓	?
$m=6$	63	12	24	4	?	?	?
$m=7$	127	14	56	4	✓	?	?
$m=8$	255	16	112	4	?	?	?

TABLE II
MINIMUM DISTANCE AND MAXIMUM WEIGHT OF THE DUAL CODES OF DBCH CODES AND TBCH CODES [10].

Dual of $[2^m-1, 2^m-1-2m, 5]$ DBCH codes		
Parameter	Odd $m \geqslant 3$	Even $m \geqslant 4$
Minimum distance	$2^{m-1}-2^{(m+1)/2-1}$	$2^{m-1}-2^{(m+2)/2-1}$
Maximum weight	$2^{m-1}+2^{(m+1)/2-1}$	$2^{m-1}+2^{(m+2)/2-1}$
Dual of $[2^m-1, 2^m-1-3m, 7]$ TBCH codes		
Parameter	Odd $m \geqslant 5$	Even $m \geqslant 6$
Minimum distance	$2^{m-1}-2^{(m+1)/2}$	$2^{m-1}-2^{(m+4)/2-1}$
Maximum weight	$2^{m-1}+2^{(m+1)/2}$	$2^{m-1}+2^{(m+4)/2-1}$

Example 3. Here we consider the $[23, 12, 7]$ cyclic binary Golay code $\mathcal{C}$. Its dual code $\mathcal{C}^\perp$ is the $[23, 11, 8]$ cyclic binary code. Hence, we conclude that $\mathcal{C}$ has locality $r = 7$ and the dual code $\mathcal{C}^\perp$ has locality $r^\perp = 6$. The code $\mathcal{C}$ is both d-optimal and k-optimal since it is a perfect code. $\mathcal{C}^\perp$ is d-optimal due to the Hamming bound, and k-optimal due to the online table [20]. The r-optimality of these two codes is proved in a similar way to the proof in Example 2.

Example 4. Let $\mathcal{C}$ be the cyclic double-error-correcting binary primitive BCH (DBCH) code with parameters $[2^m-1, 2^m-1-2m, 5]$ $(m \geqslant 4)$. Its dual code $\mathcal{C}^\perp$ has parameters $[2^m-1, 2m, 2^{m-1}-2^{\lfloor m/2 \rfloor}]$ [10]. Therefore, we conclude that $\mathcal{C}$ has locality $r = 2^{m-1}-2^{\lfloor m/2 \rfloor}-1$, and $\mathcal{C}^\perp$ has locality $r^\perp = 4$. We utilize the bound from Lemma 2 and the online table from [20] to check the d-optimality, k-optimality, and r-optimality of the DBCH codes and their dual codes. The results are summarized in Table I (✓ indicates that we could prove its optimality while ? represents that we could not verify the optimality).

IV. MORE CONSTRUCTIONS OF LRCS

In Section III, we showed a simple property of the locality of cyclic linear codes. In this section, we show how to find the locality of codes which can be obtained by the operations of Extend, Shorten, Expurgate, Augment, and Lengthen on existing LRCs. For a binary vector $\boldsymbol{c}$, $\overline{\boldsymbol{c}}$ is the complement vector of $\boldsymbol{c}$. For a code $\mathcal{C}$, $\overline{\mathcal{C}}$ is defined to be $\overline{\mathcal{C}} = \{\overline{\boldsymbol{c}} \ : \ \boldsymbol{c} \in \mathcal{C}\}$.

A. Extend Operation

The extended code of an $[n, k, d]$ code $\mathcal{C}$ is an $[n+1, k, d_{ext}]$ code $\mathcal{C}_{ext}$ with an overall parity bit added to each codeword,

$$\mathcal{C}_{ext} = \left\{ (c_1, \ldots, c_n, c_{n+1}) : (c_1, \ldots, c_n) \in \mathcal{C}, c_{n+1} = \sum_{i=1}^{n} c_i \right\},$$

where $d_{ext} = d+1$ for odd d and $d_{ext} = d$ for even d. We use the notation $\mathcal{C}_{ext}^\perp$ to denote the dual code of $\mathcal{C}_{ext}$.

Lemma 6. *Let $\mathcal{C}$ be an $[n, k, d]$ linear binary code with locality r. If the maximum Hamming weight of codewords in $\mathcal{C}^\perp$ is $n-r$, then the extended code $\mathcal{C}_{ext}$ has locality $r_{ext} = r$.*

Proof: For every $i \in [n]$, there exists a set R_i of size at most r such that the ith symbol is recoverable from the set R_i. Thus, we only need to prove this property for the $(n+1)$st symbol. Since the maximum weight of codewords in $\mathcal{C}^\perp$ is $n-r$, there exists a codeword $\boldsymbol{c} \in \mathcal{C}^\perp$ such that $w_H(\boldsymbol{c}) = n-r$. Note also that the vectors $(\boldsymbol{c}, 0)$ and $\mathbf{1}$ are codewords in $\mathcal{C}_{ext}^\perp$. Therefore the vector $\boldsymbol{c}' = (\boldsymbol{c}, 0) + \mathbf{1}$ is a codeword in $\mathcal{C}_{ext}^\perp$ and its Hamming weight is $r+1$. Hence, from Claim 4, we get that the $(n+1)$st symbol can also be recovered by a set of r other symbols. ■

We have the following corollary for cyclic linear binary codes where $r = d^\perp - 1$.

Corollary 7. *Let $\mathcal{C}$ be an $[n, k, d]$ cyclic linear binary code and let $d^\perp$ be the minimum distance of its dual code. If the maximum Hamming weight of codewords in $\mathcal{C}^\perp$ is $n+1-d^\perp$, then the extended code $\mathcal{C}_{ext}$ has locality $r_{ext} = d^\perp - 1$.*

Example 5. Let $\mathcal{C}$ be the $[2^m-1, 2^m-1-m, 3]$ cyclic binary Hamming code. Its extended code $\mathcal{C}_{ext}$ has parameters $[2^m, 2^m-1-m, 4]$. The dual code $\mathcal{C}^\perp$ is the Simplex code, which has constant Hamming weight 2^{m-1}. Hence, the condition from Corollary 7 holds and the extended Hamming code $\mathcal{C}_{ext}$ has locality $r_{ext} = d^\perp - 1 = 2^{m-1}-1$. $\mathcal{C}_{ext}$ is both d-optimal and k-optimal according to the Hamming bound. To show that it is also r-optimal, let us assume to the contrary that there exists a $[2^m, 2^m-1-m, 4]$ code with locality $\hat{r} = 2^{m-1}-2$. According to bound (2) for $t = 1$, we have

$$\begin{aligned} k_{ext} &\leqslant 2^{m-1} - 2 + k_{opt}^{(2)}(2^{m-1}+1, 4) \overset{(a)}{=} 2^{m-1} - 2 + k_{opt}^{(2)}(2^{m-1}, 3) \\ &\overset{(b)}{<} 2^{m-1} - 2 + 2^{m-1} - (m-1) = 2^m - m - 1. \end{aligned}$$

Thus, we get a contradiction to the value of k_{ext}. In the above proof, step (a) is from the property that $A(n, 2s-1) = A(n+1, 2s)$, where $A(n, d)$ denotes the largest number of codewords M in any binary code (n, M, d) [11]. Step (b) is from the Hamming bound.

Example 6. Let $\mathcal{C}$ be the $[23, 12, 7]$ cyclic binary Golay code. The extended Golay code $\mathcal{C}_{ext}$ is a $[24, 12, 8]$ code. According to the weight distribution of $\mathcal{C}^\perp$, its minimum distance is 8 and its maximum Hamming weight is 16 [3]. Hence the condition in Corollary 7 holds and $\mathcal{C}_{ext}$ has locality $d^\perp - 1 = 7$. Similarly to the arguments above, this code is d-optimal, k-optimal, and r-optimal.

Example 7. In this example we study the extended primitive binary BCH codes.

1) We consider the $[2^m, 2^m-1-2m, 6]$ extended code $\mathcal{C}_{ext}$ of the double-error-correcting BCH (DBCH) code $\mathcal{C}$ from Example 4. From the minimum distance and maximum weight of $\mathcal{C}^\perp$, listed in Table II, we conclude that the condition in Corollary 7 holds. Thus, $\mathcal{C}_{ext}$ has locality $r_{ext} = 2^{m-1}-2^{\lfloor m/2 \rfloor}-1$. $\mathcal{C}_{ext}$ is d-optimal due to the Hamming bound.

2) We also consider the $[2^m, 2^m-1-3m, 8]$ extended code $\mathcal{C}_{ext}$ of the triple-error-correcting BCH (TBCH) code. Similarly to the previous case, according to the minimum distance and maximum weight of $\mathcal{C}^\perp$ from Table II, we conclude that the condition from Corollary 7 holds and $\mathcal{C}_{ext}$ has locality $r_{ext} = 2^{m-1}-2^{\lfloor m/2+1 \rfloor}-1$. For $m \geqslant 6$, $\mathcal{C}_{ext}$ is d-optimal from Hamming bound. For the $m = 5$ case, $\mathcal{C}_{ext}$ has parameters $[32, 16, 8]$, and it is d-optimal from the online table for linear codes [20].

Next, we show the locality of the dual of the extended code.

Lemma 8. *Let $\mathcal{C}$ be an $[n,k,d]$ cyclic linear binary code with odd minimum distance d. Then, the code $\mathcal{C}_{ext}^{\perp}$ has locality $r_{ext}^{\perp}=d$.*

Proof: Since d is odd, each codeword with weight d in $\mathcal{C}$ generates a parity check bit 1. Since $\mathcal{C}$ is cyclic, for any $i\in[n]$, i belongs to the support of some codeword $(\boldsymbol{c},1)\in\mathcal{C}_{ext}$, where $\boldsymbol{c}$ has weight d. Moreover, the support of $(\boldsymbol{c},1)$ also contains coordinate $n+1$. Thus, from Claim 4, every ith symbol of $\mathcal{C}_{ext}^{\perp}$, $i\in[n+1]$, has locality d. ∎

Example 8. Let $\mathcal{C}$ be the $[n=2^m-1,k=2^m-1-m,d=3]$ cyclic binary Hamming code. $\mathcal{C}_{ext}^{\perp}$ is the biorthogonal code $[n_{ext}^{\perp}=2^m,k_{ext}^{\perp}=m+1,d_{ext}^{\perp}=2^{m-1}]$. From Lemma 8, $\mathcal{C}_{ext}^{\perp}$ has locality $r_{ext}^{\perp}=d=3$. $\mathcal{C}_{ext}^{\perp}$ is both d-optimal and k-optimal according to the Plotkin bound. Its r-optimality is proved in a similar way to the previous examples.

B. Shorten Operation

For an $[n,k,d]$ code $\mathcal{C}$, its shortened code $\mathcal{C}_s$ is obtained by removing one (here we take the last one) of its coordinates, i.e.,

$$\mathcal{C}_s=\{(c_1,\ldots,c_{n-1}):(c_1,\ldots,c_{n-1},0)\in\mathcal{C}\}.$$

We assume here that there is a codeword $\boldsymbol{c}\in\mathcal{C}$ such that $c_n=1$. Otherwise, we will remove another coordinate satisfying this condition. The code $\mathcal{C}_s$ has parameters $[n-1,k-1,d_s\geqslant d]$ and its dual code is denoted by $\mathcal{C}_s^{\perp}$.

Lemma 9. *Let $\mathcal{C}$ be an $[n,k,d]$ linear code with locality r ($r\geqslant 2$). The shortened code $\mathcal{C}_s$ has locality r or $r-1$.*

Proof: Since $\mathcal{C}$ has locality r, for all $i\in[n-1]$, the ith code symbol has a repair set R_i with respect to $\mathcal{C}$ of size at most r. If $n\notin R_i$ then this symbol has the same repair set also with respect to the shortened code $\mathcal{C}_s$. Otherwise, note that if $\boldsymbol{c}\in\mathcal{C}_s$ then we have $(\boldsymbol{c},0)\in\mathcal{C}$, but since we know that necessarily the nth symbol is zero we conclude that the ith symbol is recoverable also from the set $R_i\setminus\{n\}$. ∎

In [5], the puncture operation was defined to be the shorten operation here. It was claimed there the locality remains r, however as we shall see in the next example the shortened code can have locality $r-1$. The following is an immediate conclusion for cyclic linear binary codes.

Corollary 10. *Let $\mathcal{C}$ be an $[n,k,d]$ cyclic linear binary code, and let $d^{\perp}$ ($d^{\perp}\geqslant 3$) be the minimum distance of its dual code. Then, the code $\mathcal{C}_s$ has locality either $d^{\perp}-2$ or $d^{\perp}-1$.*

Example 9. Let $\mathcal{C}$ be the $[2^m-1,2^m-1-m,3]$ cyclic binary Hamming code. Its shortened code $\mathcal{C}_s$ is a $[2^m-2,2^m-2-m,3]$ code and from Corollary 10 it has locality $d^{\perp}-2$ or $d^{\perp}-1$, where $d^{\perp}=2^{m-1}$. We show that it has locality $d^{\perp}-2$. According to the proof of Lemma 9, it is enough to show that for every $i\in[n-1]$, the ith code symbol has a repair set R_i of size $2^{m-1}-1$ which contains the nth coordinate. Or, according to Claim 4, it is enough to show that there exists a codeword $\boldsymbol{c}\in\mathcal{C}^{\perp}$ such that $c_i=c_n=1$ and $w_H(\boldsymbol{c})=2^{m-1}$. We can omit the last requirement on the weight since all nonzero codewords in $\mathcal{C}^{\perp}$ have the same weight 2^{m-1}. Let $\boldsymbol{c}_1,\boldsymbol{c}_2\in\mathcal{C}^{\perp}$ be two codewords such that $c_{1,i}=c_{2,n}=1$. If $c_{1,n}=1$ or $c_{2,i}=1$ then we are done. Otherwise, the codeword $\boldsymbol{c}_1+\boldsymbol{c}_2$ satisfies this property. The d-, k-,and r-optimality of $\mathcal{C}_s$ is proved in a similar way to the previous examples.

C. Expurgate, Augment, and Lengthen Operations

For an $[n,k,d]$ code $\mathcal{C}$ having odd weight codewords, the expurgated code $\mathcal{C}_{exp}$ is a sub-code of $\mathcal{C}$ which contains only the codewords of even weight. That is,

$$\mathcal{C}_{exp}=\{\boldsymbol{c}:\boldsymbol{c}\in\mathcal{C},w_H(\boldsymbol{c})\text{ is even }\}.$$

$\mathcal{C}_{exp}$ is an $[n,k-1,w_{\min_e}]$ code, where $w_{\min_e}$ denotes the minimum positive even weight of codewords in $\mathcal{C}$. We denote by $\mathcal{C}_{exp}^{\perp}$ the dual code of $\mathcal{C}_{exp}$ and note that $\mathcal{C}_{exp}^{\perp}=\mathcal{C}^{\perp}\cup\overline{\mathcal{C}^{\perp}}$.

For an $[n,k,d]$ code $\mathcal{C}$ which does not contain the all ones codeword $\mathbf{1}$, its augmented code $\mathcal{C}_a$ is the code $\mathcal{C}\cup\overline{\mathcal{C}}$. $\mathcal{C}_a$ is an $[n,k+1,\min\{d,n-w_{\max}\}]$ code, where $w_{\max}$ denotes the maximum weight of codewords in $\mathcal{C}$. We use the notation $\mathcal{C}_a^{\perp}$ to denote the dual code of $\mathcal{C}_a$.

According to these definitions, if the code $\mathcal{C}$ is cyclic then the expurgated and augmented codes of $\mathcal{C}$ are cyclic as well. Hence, we have the following two observations for an $[n,k,d]$ cyclic binary code $\mathcal{C}$:

1) If $\mathcal{C}$ has an odd weight codeword, then $\mathcal{C}_{exp}$ has locality $r_{exp}=\min\{d^{\perp},n-w_{\max}^{\perp}\}-1$, where $w_{\max}^{\perp}$ is the maximum Hamming weight of codewords in $\mathcal{C}^{\perp}$. (Here, we assume $w_{\max}^{\perp}<n-1$, since $w_{\max}^{\perp}=n-1$ is not an interesting case.)

2) If $\mathcal{C}$ does not contain the all ones codeword $\mathbf{1}$, then $\mathcal{C}_a$ has locality $r_a=w_{\min_e}^{\perp}-1$, where $w_{\min_e}^{\perp}$ is the minimum positive even weight of codewords in $\mathcal{C}^{\perp}$.

For an $[n,k,d]$ code $\mathcal{C}$ which does not contain the all ones codeword $\mathbf{1}$, the lengthened code $\mathcal{C}_{\ell}$ is obtained as follows. First, the code $\mathcal{C}$ is augmented to the code $\mathcal{C}_a=\mathcal{C}\cup\overline{\mathcal{C}}$. Then, $\mathcal{C}_a$ is extended. Thus, $\mathcal{C}_{\ell}=\{(c_1,\ldots,c_n,c_{n+1}):c_{n+1}=\sum_{i=1}^{n}c_i$ and $(c_1,\ldots,c_n)\in\mathcal{C}\cup\overline{\mathcal{C}}\}$. After the lengthen operation, the code's length and dimension are increased by 1. By leveraging the results from the augment and extend operations, we conclude that if the minimum positive even weight of codewords in $\mathcal{C}^{\perp}$ is $w_{\min_e}^{\perp}$, and the maximum Hamming weight of codewords in $\mathcal{C}_a^{\perp}$ is $n+1-w_{\min_e}^{\perp}$, then the lengthened code $\mathcal{C}_{\ell}$ has locality $r_{\ell}=w_{\min_e}^{\perp}-1$. An operation similar to the lengthen operation, called enlarging, was proposed in [5]. Under the enlarging operation, the code length, dimension, and locality are increased by 1.

V. Reed-Muller Codes

Even though Reed-Muller (RM) codes are not cyclic codes, they still have some similarity to cyclic codes. This motivates us to study their locality. A μth-order linear binary RM code $\mathcal{RM}(\mu,m)$ has code length $n=2^m$, dimension $k=\sum_{i=0}^{\mu}\binom{m}{i}$, and minimum distance $d=2^{m-\mu}$.

In [18], two classes of codes with locality 2 and 3 are constructed based on the generalized RM codes of first and second orders. Here, we focus on the binary RM codes, and prove their locality as follows. We note that ideas on an alternative proof of this property were mentioned in [18], but we chose to prove it here for the sake of completeness of our discussion on binary codes.

Lemma 11. *Let $\mathcal{RM}(\mu,m)$ be a μth-order linear binary RM code. Then it has locality $r=d^{\perp}-1=2^{\mu+1}-1$.*

Proof: It is known the dual code of $\mathcal{RM}(\mu,m)$ is $\mathcal{RM}(m-\mu-1,m)$, and the minimum weight codewords of an RM code generate all its codewords [10]. Therefore, every coordinate i, $i\in[n]$, belongs to the support of a certain minimum

TABLE III
CYCLIC LINEAR BINARY LOCALLY REPAIRABLE CODES.

$\mathcal{C}$	n	k	d	r	d-opt	k-opt	r-opt
Hamming code	2^m-1	2^m-1-m	3	$2^{m-1}-1$	✓	✓	✓
Simplex code	2^m-1	m	2^{m-1}	2	✓	✓	✓[a]
Golay code	23	12	7	7	✓	✓	✓
Dual of Golay code	23	11	8	6	✓	✓	✓
DBCH code ($m \geqslant 4$)	2^m-1	2^m-1-2m	5	$2^{m-1}-2^{\lfloor m/2 \rfloor}-1$	Table I	Table I	Table I
Dual of DBCH code ($m \geqslant 4$)	2^m-1	$2m$	$2^{m-1}-2^{\lfloor m/2 \rfloor}$	4	Table I	Table I	Table I
Extended Hamming code	2^m	2^m-1-m	4	$2^{m-1}-1$	✓	✓	✓
Extended Golay code	24	12	8	7	✓	✓	✓
Extended DBCH code ($m \geqslant 4$)	2^m	2^m-1-2m	6	$2^{m-1}-2^{\lfloor m/2 \rfloor}-1$	✓	?	?
Extended TBCH code ($m \geqslant 5$)	2^m	2^m-1-3m	8	$2^{m-1}-2^{\lfloor m/2+1 \rfloor}-1$	✓	?	?
Biorthogonal code	2^m	$m+1$	2^{m-1}	3	✓	✓	✓
Expurgated Hamming code	2^m-1	2^m-2-m	4	$2^{m-1}-2$	✓	✓	✓
Expurgated DBCH code ($m \geqslant 4$)	2^m-1	2^m-2-2m	6	$2^{m-1}-2^{\lfloor m/2 \rfloor}-2$	✓	?	?
Expurgated TBCH code ($m \geqslant 5$)	2^m-1	2^m-2-3m	8	$2^{m-1}-2^{\lfloor m/2+1 \rfloor}-2$	✓	?	?
Augmented Simplex code	2^m-1	$m+1$	$2^{m-1}-1$	3	✓	✓	✓
Shortened Hamming code	2^m-2	2^m-2-m	3	$2^{m-1}-2$	✓	✓	✓
Shortened Simplex code	2^m-2	$m-1$	2^{m-1}	1	✓	✓	✓
$\mathcal{RM}(\mu, m)$	2^m	$\sum_{i=0}^{\mu}\binom{m}{i}$	$2^{m-\mu}$	$2^{\mu+1}-1$	?	?	?
Cyclic $\mathcal{RM}(\mu, m)$	2^m-1	$\sum_{i=0}^{\mu}\binom{m}{i}$	$2^{m-\mu}-1$	$2^{\mu+1}-1$	?	?	?
Dual of cyclic $\mathcal{RM}(\mu, m)$	2^m-1	$\sum_{i=\mu+1}^{m}\binom{m}{i}-1$	$2^{\mu+1}$	$2^{m-\mu}-2$	?	?	?
Construction 1 in [7] ($m \geqslant 2$)	2^m-1	$\frac{rn}{r+1}$ $(r+1 \mid n)$	2	r	✓	✓	✓
Construction 2 in [7] ($m \geqslant 4$)	2^m-1 $(2 \mid m)$	$\frac{2}{3}(2^m-1)-m$	6	2	?	✓[b]	✓
Construction 3 in [7] ($m \geqslant 4$)	2^m-1 $(2 \mid m)$	$\frac{2}{3}(2^m-1)-2m$	10	2	?	? [c]	✓[d]

[a] r-optimality is proved in [2].

[b] k-optimality is proved for $m > 8$ in Theorem 3 of [7].

[c] With assumptions that the repair sets are disjoint and k is even, k is proved to achieve its upper bound in Theorem 2 of [7].

[d] r-optimality can be proved for $m \geqslant 6$.

weight codeword of $\mathcal{RM}(m-\mu-1, m)$. To see that, assume to the contrary that there exists a coordinate j, $j \in [n]$, in which all the minimum weight codewords of $\mathcal{RM}(m-\mu-1, m)$ have value 0. Thus, the linear combinations of all the minimum weight codewords cannot produce the all ones codeword $\mathbf{1}$, which is a valid codeword. Thus, we get a contradiction, proving that $\mathcal{RM}(\mu, m)$ has locality $r = d^{\perp} - 1 = 2^{\mu+1} - 1$. ■

Finally, we mention that a μth-order cyclic RM code $\mathcal{C}$ is a $[2^m-1, \sum_{i=0}^{\mu}\binom{m}{i}, 2^{m-\mu}-1]$ punctured RM code, represented in a cyclic form [10]. Its dual code $\mathcal{C}^{\perp}$ is also cyclic and is a $[2^m-1, \sum_{i=\mu+1}^{m}\binom{m}{i}-1, 2^{\mu+1}]$ code. From Observation 5, $\mathcal{C}$ has locality $r = 2^{\mu+1}-1$, and $\mathcal{C}^{\perp}$ has locality $r^{\perp} = 2^{m-\mu}-2$.

VI. CONCLUSION

In this work, a variety of linear binary LRCs are investigated and their optimality is also studied. Our results are summarized in Table III (Due to space limitations, some codes have been given without proofs), where ✓ means we can prove the optimality of the given codes, whereas ? means we cannot verify the optimality of the given family of codes. Table III also includes the results from [2] and [7]. For future work, we plan to extend the current results to non-binary codes.

REFERENCES

[1] A. Brouwer, "Table for general binary codes," *http://www.win.tue.nl/ aeb/codes/binary-1.html*, 2014.

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

[3] J. Conway and N. Sloane, "Orbit and coset analysis of the Golay and related codes," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1038–1050, 1990.

[4] T. Ernvall, T. Westerbäck, and C. Hollanti, "Constructions of optimal and almost optimal locally repairable codes," *arXiv preprint arXiv:1406.4277*, 2014.

[5] T. Ernvall, T. Westerbäck, C. Hollanti, and R. Freij, "Constructions and properties of linear locally repairable codes," *arXiv preprint arXiv:1410.6339*, 2014.

[6] 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, 2012.

[7] S. Goparaju and R. Calderbank, "Binary cyclic codes that are locally repairable," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2014, pp. 676–680.

[8] C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," in *Proc. 6th IEEE Int. Symp. Netw. Comput. Appl.*, 2007, pp. 79–86.

[9] C. Huang, H. Simitci, Y. Xu, A. Ogus, B. Calder, P. Gopalan, J. Li, and S. Yekhanin, "Erasure coding in Windows Azure Storage," in *Proc. USENIX Annu. Tech. Conf.*, 2012, pp. 15–26.

[10] S. Lin and D. J. Costello, *Error Control Coding*. Prentice Hall, 2004.

[11] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Elsevier, 1977, vol. 16.

[12] F. Oggier and A. Datta, "Self-repairing homomorphic codes for distributed storage systems," in *Proc. IEEE INFOCOM*, 2011, pp. 1215–1223.

[13] L. Pamies-Juarez, H. D. Hollmann, and F. Oggier, "Locally repairable codes with multiple repair alternatives," in *Proc. IEEE Int. Symp. Inf. Theory*, 2013, pp. 892–896.

[14] D. S. Papailiopoulos and A. G. Dimakis, "Locally repairable codes," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 2771–2775.

[15] D. Papailiopoulos and A. Dimakis, "Locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5843–5855, Oct 2014.

[16] N. Prakash, G. Kamath, V. Lalitha, and P. Kumar, "Optimal linear codes with a local-error-correction property," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2012, pp. 2776–2780.

[17] N. Prakash, V. Lalitha, and P. Kumar, "Codes with locality for two erasures," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2014, pp. 1962–1966.

[18] A. Rawat and S. Vishwanath, "On locality in distributed storage systems," in *Proc. IEEE Inf. Theory Workshop*, Sept 2012, pp. 497–501.

[19] M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, and D. Borthakur, "XORing elephants: Novel erasure codes for big data," in *Proc. VLDB Endowment*, vol. 6, no. 5, 2013, pp. 325–336.

[20] R. Schürer and W. Schmid, "Table for linear codes," *mint.sbg.ac.at/table.php?i=c*, 2014.

[21] N. Silberstein, A. S. Rawat, O. O. Koyluoglu, and S. Vishwanath, "Optimal locally repairable codes via rank-metric codes," in *Proc. IEEE Int. Symp. Inf. Theory*, 2013, pp. 1819–1823.

[22] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2014, pp. 686–690.

[23] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, "Optimal locally repairable codes and connections to matroid theory," in *Proc. IEEE Int. Symp. Inf. Theory*, 2013, pp. 1814–1818.

[24] A. Wang and Z. Zhang, "An integer programming based bound for locally repairable codes," *arXiv preprint arXiv:1409.0952*, 2014.