# Some New Binary Locally Repairable Codes with Availability based on Golomb Rulers

Hyojeong Choi and Hong-Yeop Song
School of Electrical and Electronic Engineering
*Yonsei University*
Seoul, Korea
{hjchoi3022, hysong}@yonsei.ac.kr

***Abstract*—In this paper, we propose some new binary $5$-sequential recovery Locally Repairable Codes ($5$-seq LRCs) with availability $2$ based on Golomb rulers and some well-known technique of generating the parity check matrix for the QC-LDPC codes. The proposed $5$-seq LRCs can repair up to $5$ erased symbols sequentially within $3$ repair time. It can be viewed as an intersection of two LRCs without availability and its code rate is optimal for some cases.**

## I. Introduction

To stably store big data in distributed storage systems (DSSs) and increase reliability, locally repairable codes (LRCs) that recover a single erasure symbol with only small number of nodes have been proposed by Gopalan *et. al.* [6]. The locally repairable code of length $n$, dimension $k$ and locality $r$ is denoted as an $(n, k, r)$-LRC. An important parameter for LRCs is the locality $r$ which is the minimum number of symbols in the codeword required to repair a single erasure symbol [6]. If each symbol has locality at most $r$, then the code is said to have the locality $r$.

Recent topic of the research is LRCs which can repair multiple erased nodes. Depending on the repair process, such LRCs are divided into two different classes: parallel recovery LRCs and sequential recovery LRCs or seq-LRCs. There are various types of parallel recovery LRCs [10]–[13].

In this paper, we focus on sequential recovery LRCs with availability $t$. Let $C$ be an $(n, k, r)$-LRC and $c = (c_1, c_2, ..., c_n) \in C$. Then $C$ is said to have locality $r$ and availability $t$ if, for each $i \in \{1, 2, ..., n\}$, there exist at least $t$ pairwise disjoint repair sets $R_1(i), R_2(i), ..., R_t(i) \subseteq \{1, 2, ..., n\} \setminus i$, such that for $1 \leq j \leq t$, (i) $|R_j(i)| \leq r$, (ii) for each $j = 1, 2, ..., t$, $c_i$ is a linear combination of $c_l$ for $l \in R_j(i)$. The main idea of having the availability $t$ is that one can repair either an erasure in $t$ different ways or $t$ erasures in parallel. Some bounds and constructions for LRCs with availability $t$ have been studied in [11], [12]. Recently, an intersection of two LRCs with disjoint repair groups was analysed and a construction was proposed for those intersections with availability 2 [3]. We will recall the main result of [3] as *Known Fact 3* in Preliminaries.

The sequential recovery LRCs have been studied in [1], [3], [10]. An *$u$-sequential recovery ($u$-seq) LRC* [10] can repair up to $u$ erased symbols of a codeword in some sequential order $(c_{i_1}, c_{i_2}, ..., c_{i_u})$. The key point is that some repair sets of the later erasures may contain some earlier repaired symbols. Prakash *et. al.* [10] proposed the graph-based construction and rate bound for 2-seq LRCs. A length bound for binary 3-seq LRCs has been proved and mentioned in [1]. In [2], the rate bound for $u$-seq LRCs over $\mathbb{F}_q$ was proved as follows: for $r \geq 3$, if $u$ is odd and $\sigma = \lfloor \frac{u-1}{2} \rfloor$,

$$\frac{k}{n} \leq \frac{r^{\sigma+1}}{r^{\sigma+1} + 2\sum_{i=1}^{\sigma} r^i + (u - 2\sigma)}. \tag{1}$$

Some connections between LRCs for multiple erasures and regular LDPC codes were proposed [7], [8]. In [7], it was mentioned that the $4$-cycle free regular LDPC code can be an LRC whose availability is the column weight of the corresponding parity check matrix. In [8], it was proved that a linear block code defined by a parity check matrix $H$ whose girth is $2(u + 1)$ and whose row weight is at most $r + 1$ and column weight is at least 2 becomes an $u$-seq LRC with locality $r$. This is recalled as *Known Fact 2* in Preliminaries.

In this paper, we propose some new 5-seq LRCs with availability 2 based on Golomb rulers and some well-known technique of generating the parity check matrix for the QC-LDPC codes. The proposed 5-seq LRCs can repair up to 5 erased symbols sequentially within 3 repair time. It can be viewed as an intersection of two LRCs without availability and its code rate is optimal for some cases.

In Section II, we recall three Known Facts from others. In Section III, we design the binary 5-seq LRCs with availability 2 based on Golomb ruler and show that they are rate optimal for some cases. Section IV concludes the paper with some concluding remarks.

## II. Preliminaries

We will fix here two integers $s$ and $m$ with $2 \leq s \leq m$ throughout this paper. As a scheme for generating a parity check matrix for QC-LDPC codes, a construction based on the exponent matrix was proposed in [5]. In this paper, we will follow their proposed universal scheme of generating the $t \times s$ exponent matrix $E = [e(i, j)]$ using multiplication table, and then determining the parity check matrix. The parity check matrix $H = [H_{e(i,j)}]$ is obtained by substituting $e(i, j)$-shifted circular permutation matrix (CPM) for the position $(i, j)$ of $E$, for all $i, j$.

*Known Fact 1:* [5], [9]

1) The non-existence condition of a $2c$-cycle in the Tanner graph representation of matrix $H$ defined by substituting some CPMs to $(i,j)$ position of an exponent matrix $E = [e(i,j)]$ is that

$$\sum_{l=0}^{c-1}(e(i_l,j_l) - e(i_l,j_{l+1})) \not\equiv 0 \pmod{m} \qquad (2)$$

for all $i_0, i_1, ..., i_{c-1}$ and $j_0, j_1, ..., j_c = j_0$ such that $i_l \neq i_{l+1}$ and $j_l \neq j_{l+1}$ for $0 \leq l < c$.

2) When the exponent matrix $E$ above has only 2 rows, the girth in $H$ must be a multiple of 4.

We will say with some abuse of notation that the code defined by $H$ has girth $g$ (or simply, $H$ has girth $g$) when the Tanner graph of $H$ has girth $g$.

A set of integers $\{g_1, g_2, ..., g_s\}$ with $0 \leq g_1 < g_2 < \cdots < g_s$ is called a Golomb ruler if the differences $g_j - g_i$ for $i < j$, are all distinct [4]. A Golomb ruler with $s$ elements is called an $s$-mark Golomb ruler. The length of the Golomb ruler is equal to the difference $g_s - g_1$. The optimal $s$-mark Golomb ruler is the smallest possible Golomb ruler when the number of marks is $s$. In this paper, we set $g_1 = 0$ for convenience. If $\{g_1, g_2, ..., g_s\}$ is a Golomb ruler, then for any integer $x$, $\{x+g_1, x+g_2, ..., x+g_s\}$ and $\{x \cdot g_1, x \cdot g_2, ..., x \cdot g_s\}$ are also Golomb rulers. These are called Translation and Multiplication property, respectively.

In [9], the parity check matrix $H$ for the QC-LDPC code was generated by applying the Golomb ruler based on the scheme of [5]. The parity check matrix $H$ generated by the scheme [5], [9] has the column weight of at least 2 and row weight of $s$. Therefore, it can be analyzed as the parity check matrix $H$ of sequential LRC in the next Known Fact. Since a linear block code defined by such an $H$ can have a girth of up to 12 [5], it is a sequentially recoverable LRC up to 5 erasures with at most $\lceil 5/2 \rceil = 3$ repair time according to the following Known Fact if its girth is indeed 12.

*Known Fact 2:* [8]

1) A linear block code is a $u$-seq LRC with locality $r$ if its parity check matrix satisfies the following:
   (i) the girth is $2(u+1)$.
   (ii) the column weight is at least 2, and
   (iii) the row weight is at most $r+1$.
2) The repair time of $u$-seq LRC defined above is at most $\lceil u/2 \rceil$.

In fact, in the parity check matrix $H$ generated based on $1 \times s$ exponent matrix, the repair groups corresponding to all the rows of $H$ are pairwise disjoint, and the union of their column indices becomes $\{1, 2, ..., n\}$. Therefore, the matrix $H$ based on the $2 \times s$ exponent matrix has two disjoint repair groups and can be applied to the following *Known Fact 3*.

*Known Fact 3:* [3]

1) Given two parity check matrices $H_1$ and $H_2$ of the same size $m \times (r+1)m$ for two LRCs with disjoint repair groups and constant repair group size $r+1$ and with $m \geq r+1$, the linear code $C$ (which is the intersection of two constituent codes) with the parity check matrix

$$H = \begin{pmatrix} H_1 \\ \hdashline H_2 \end{pmatrix} \qquad (3)$$

of size $2m \times (r+1)m$ will have availability 2 if and only if

$$|supp(h_{1,i}) \cap supp(h_{2,j})| \leq 1^{\dagger}, \quad \text{for all } i, j, \qquad (4)$$

where $h_{1,i}$ and $h_{2,j}$ are $i$-th row of $H_1$ and $j$-th row of $H_2$, respectively.

2) If the LRC with the parity check matrix $H$ in (3) has availability 2, then this code is a 3-seq LRC and

$$rank(H) \geq 2m - \left\lfloor \frac{m}{r+1} \right\rfloor.$$

Therefore its dimension $k$ is upper bounded by

$$k \leq (r-1)m + \left\lfloor \frac{m}{r+1} \right\rfloor. \qquad (5)$$

## III. 5-SEQ LRCs BASED ON GOLOMB RULERS

In *Known Fact 3*, the LRC having $H$ in (3) as a parity check matrix becomes a 3-seq LRC if and only if the condition (4) is satisfied. Furthermore, if $H_1$ and $H_2$ are carefully designed in (3), the resulting code can be a 5-seq LRC, as in the following Theorem.

*Theorem 1:* Let $\{g_1, g_2, ..., g_s\}$ with $0 \leq g_1 < g_2 < \cdots < g_s$ be an $s$-mark Golomb ruler and $D = \{d_1, d_2, ..., d_{\binom{s}{2}}\}$ be the set of all possible differences $g_j - g_i$ for $i < j$. Let $m$ be a positive integer greater than $g_s$ such that the following condition is satisfied:

$$\begin{gathered} d_i + d_j \not\equiv 0 \pmod{m} \\ \text{for all } i, j \text{ not necessarily distinct.} \end{gathered} \qquad (6)$$

If $H = [H_{e(i,j)}]$ is the binary matrix with $e(i,j)$-shifted CPM of the size $m$ substituted for the position $(i,j)$ of $E$ for all $i, j$, where $E = [e(i,j)]$ is an $2 \times s$ exponent matrix given as

$$E = \begin{pmatrix} c & c & \cdots & c \\ g_1 & g_2 & \cdots & g_s \end{pmatrix}, \qquad (7)$$

for an integer $c \geq 0$, then the linear code with $H$ as the parity check matrix is a 5-seq $(n,k,r)$-LRC with availability $t = 2$ where $n = sm, (s-2)m + 1 \leq k \leq (s-2)m + \lfloor \frac{m}{s} \rfloor$ and $r = s - 1$. The repair time of this code is at most 3.

*Example 1: Take the value $c = 2$ and a 3-mark Golomb ruler $\{0, 1, 3\}$ in Theorem 1. Since $D = \{1, 2, 3\}$, all possible sums of any two elements are $\{2, 3, 4, 5, 6\}$ without any repeat.*

$^{\dagger}$In [3], the condition (4) was written with equality, which was a simple typo.

TABLE I
EXAMPLES OF VARIOUS OPTIMAL 5-SEQ LRCS FROM **Theorem 1** USING OPTIMAL GOLOMB RULERS

| $s$ | optimal Golomb ruler | min $m$ | $n$ | $k$ | code rate | rate bound (1) for $u=5$ | $2g_s+1 \geq$ min $m$ |
|---|---|---|---|---|---|---|---|
| 4 | $0,1,4,6$ | 13 | 52 | 27 | 0.519 | 0.519 | $13 =$ min $m$ |
| 5 | $0,2,7,8,11$ | 21 | 105 | 64 | 0.610 | 0.610 | $23 >$ min $m$ |
| 6 | $0,1,4,10,12,17$<br>$0,1,8,11,13,17$ | 31 | 186 | 125 | 0.672 | 0.672 | $35 >$ min $m$ |
| 7 | $0,2,3,10,16,21,25$<br>$0,2,7,13,21,22,25$ | 49 | 343 | 246 | 0.717 | 0.717 | $51 >$ min $m$ |
| 8 | $0,1,4,9,15,22,32,34$ | 69 | 552 | 415 | 0.752 | 0.752 | $69 =$ min $m$ |
| 9 | $0,1,5,12,25,27,35,41,44$ | 89 | 801 | 624 | 0.779 | 0.779 | $89 =$ min $m$ |

TABLE II
EXAMPLES OF VARIOUS 5-SEQ LRCS FROM **Theorem 1** USING NON-OPTIMAL GOLOMB RULERS

| $s$ | non-optimal Golomb ruler | min $m$ | $n$ | $k$ | code rate | rate bound (1) for $u=5$ | $2g_s+1$ |
|---|---|---|---|---|---|---|---|
| 4 | $0,1,4,9$ | 15 | 60 | 31 | 0.517 | 0.519 | 19 |
| 5 | $0,2,7,8,17$ | 28 | 140 | 85 | 0.607 | 0.610 | 35 |
| 6 | $0,1,4,10,12,25$ | 41 | 246 | 165 | 0.671 | 0.672 | 51 |
|  | $0,1,8,11,13,27$ | 44 | 264 | 177 | 0.670 |  | 55 |
| 7 | $0,2,3,10,16,21,43$ | 69 | 483 | 346 | 0.716 | 0.717 | 87 |
|  | $0,2,7,13,21,22,45$ | 71 | 497 | 356 | 0.716 |  | 91 |
| 8 | $0,1,4,9,15,22,32,65$ | 90 | 720 | 541 | 0.751 | 0.752 | 131 |
| 9 | $0,1,5,12,25,27,35,41,83$ | 115 | 1035 | 806 | 0.778 | 0.779 | 167 |

*We choose $m = 7$ that satisfies the condition (6). Then $H$ becomes*

$$H = \begin{pmatrix} I^{(2)} I^{(2)} I^{(2)} \\ \hdashline I^{(0)} I^{(1)} I^{(3)} \end{pmatrix} = \left(\begin{array}{ccc} 0000010 & 0000010 & 0000010 \\ 0000001 & 0000001 & 0000001 \\ 1000000 & 1000000 & 1000000 \\ 0100000 & 0100000 & 0100000 \\ 0010000 & 0010000 & 0010000 \\ 0001000 & 0001000 & 0001000 \\ 0000100 & 0000100 & 0000100 \\ \hdashline 1000000 & 0000001 & 0000100 \\ 0100000 & 1000000 & 0000010 \\ 0010000 & 0100000 & 0000001 \\ 0001000 & 0010000 & 1000000 \\ 0000100 & 0001000 & 0100000 \\ 0000010 & 0000100 & 0010000 \\ 0000001 & 0000010 & 0001000 \end{array}\right)$$

*According to Theorem 1, the linear code with $H$ in above is a 5-seq $(n = 27, k = 10, r = 2, t = 2)$ LRC.*

We are now able to classify, for a given Golomb ruler, some ranges of $m$, to satisfy the condition (6) in Theorem 1.

*Theorem 2:* Let $\{g_1, g_2, ..., g_s\}$ with $0 \leq g_1 < g_2 < \cdots < g_s$ be an $s$-mark Golomb ruler and $D = \{d_1, d_2, ..., d_{\binom{s}{2}}\}$ be the set of all possible differences $g_j - g_i$ for $i < j$. If $m > 2g_s$, then the condition (6) in Theorem 1 is satisfied. If $m = 2g_s$, then the condition is never satisfied. If $g_s < m < 2g_s$, then an individual computer check is required for the condition.

Theorem 2 classifies, for a given Golomb ruler, some ranges of $m$, to satisfy the condition (6) in Theorem 1. Tables I and II show the smallest possible $m$ and the resulting code rate when applying the given optimal and non-optimal Golomb rulers to the construction of Theorem 1, respectively. In these tables, the min $m$ is the smallest $m$ that satisfies the condition (6) in Theorem 1. An $u$-seq LRC is said to be rate optimal if (1) holds with equality. As shown in Table I, the code constructed by Theorem 1 using optimal Golomb rulers is optimal in the sense of achieving the bound in (1). On the other hand, Table II shows that the code constructed using those non-optimal Golomb rulers could be non-optimal, but it is very close to the optimality. Table III shows the code rates for various $m$ including the min $m$. As $m$ increases, the code rate gradually decreases, and hence becomes away from the optimality. Note the case of $s = 5$ in Table III. The min $m$ turns out to be $21 < 2g_s$. The next value $m = 22 = 2g_s$ is impossible to use by Theorem 2.

*Remark 1: As in Tables I and II, the min $m$ can be less than or equal to $2g_s + 1$. We note that the min $m$ is equal to $2g_s+1$ for 4-,8-, and 9-mark optimal Golomb rulers as shown*

TABLE III
EXAMPLES OF VARIOUS 5-SEQ LRCS FROM **Theorem 1** USING OPTIMAL GOLOMB RULERS WHEN NOT MINIMUM M

| $s$ | optimal Golomb ruler | $m$ | $n$ | $k$ | code rate | rate bound (1) for $u=5$ | $2g_s+1$ |
|---|---|---|---|---|---|---|---|
| 4 | $0,1,4,6$ | 13 | 52 | 27 | 0.519 | 0.519 | 13 |
| | | 14 | 56 | 29 | 0.518 | | |
| | | 15 | 60 | 31 | 0.517 | | |
| | | 16 | 64 | 33 | 0.516 | | |
| | | 17 | 68 | 35 | 0.515 | | |
| 5 | $0,2,7,8,11$ | 21 | 105 | 64 | 0.610 | 0.610 | 23 |
| | | 23 | 115 | 70 | 0.609 | | |
| | | 24 | 120 | 73 | 0.608 | | |
| | | 25 | 125 | 76 | 0.608 | | |
| | | 26 | 130 | 79 | 0.607 | | |

*in Table I. That is, for these Golomb rulers, all the values of $m < 2g_s$ do not satisfy the condition (6) in Theorem 1. On the other hand, for 5-,6-, and 7-mark optimal Golomb rulers, $2g_s+1 >$ min $m$. It would be interesting to characterize those $s$-mark Golomb rulers for which the min $m$ is exactly $2g_s+1$.*

When $m$ does not satisfy the condition (6) in Theorem 1, the resulting code is a 3-seq LRC and 5-seq repair is not guaranteed. The following example shows a case where the resulting code is a 3-seq LRC but not a 5-seq LRC.

*Example 2: Take the value $c = 3$ and a 3-mark Golomb ruler $\{0, 1, 4\}$ in Theorem 1. Since $D = \{1, 3, 4\}$, the possible sums of any two elements in $D$ are distinct and they belong to $\{2, 4, 5, 6, 7, 8\}$. We choose $m = 5 = 1 + 4$ which does not satisfy the condition (6). Then the matrices $E$ and $H$ become:*

$$E = \begin{pmatrix} 3 & 3 & 3 \\ 0 & 1 & 4 \end{pmatrix}, \ H = \begin{pmatrix} I^{(3)} & I^{(3)} & I^{(3)} \\ \hdashline I^{(0)} & I^{(1)} & I^{(4)} \end{pmatrix}$$

*Since $m$ does not satisfy condition (6), the resulting code is a 3-seq LRC but not a 5-seq LRC. If $c_1, c_2, c_3, c_6$ and $c_{13}$ are erased symbols, then sequential repair is impossible in any order.*

## IV. CONCLUDING REMARKS

In this paper, we propose some new 5-seq LRCs with availability $t = 2$ based on Golomb rulers and some well-known technique of generating the parity check matrix for the QC-LDPC codes. The proposed 5-seq LRCs can be viewed as an intersection of two LRCs with disjoint repair groups without availability and can repair 5 erased symbols sequentially within 3 repair time. The code rate of the resulting codes using the optimal Golomb rulers in Table I turns out to be optimal. It is an open problem to characterize for which Golomb rulers the resulting code becomes rate-optimal.


## V. ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) Grant by the Korean Government through Ministry of Sciences and ICT (MSIT) under Grant RS-2023-00209000.


## REFERENCES


[1] S. B. Balaji, K. P. Prasanth, and P. V. Kumar, "Binary codes with locality for multiple erasures having short block length," in Proc. *IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 655–659, Jul. 2016.

[2] S. B. Balaji, G. R. Kini, and P. V. Kumar, "A tight rate bound and a matching construction for locally recoverable codes with sequential recovery from any number of multiple erasures," in Proc. *IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1778–1782, Jun. 2017.

[3] H. Choi, Z. Jing, G. Kim and H.-Y. Song, "Some Intersections of two Binary LRCs with Disjoint Repair Groups," *The 10th International Workshop on Signal Design and its Applications in Communications (IWSDA 2022)*, Aug. 2022.

[4] A. Dimitromanolakis, "Analysis of the golomb ruler and the sidon set problems, and determination of large, near-optimal golomb rulers," Master's Thesis, Technical University of Crete, Chania, 2002.

[5] M.P.C. Fossorier, "Quasicyclic low-density parity-check codes from circulant permutation matrices," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1788–1793, 2004.

[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, Nov. 2012.

[7] P. Huang, E. Yaakobi, H. Uchikawa, and P. H. Siegel, "Linear locally repairable codes with availability," in Proc. *IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1871-1875, Jun. 2015.

[8] Z. Jing and H. -Y. Song, "Girth-Based Sequential-Recovery LRCs," in *IEEE Access*, vol. 10, pp. 126156-126160, 2022.

[9] I. Kim, H. -Y. Song, "A construction for girth-8 QC-LDPC codes using Golomb rulers," *Electronics Letters*, vol. 58, no. 15, pp. 582-584, July 2022.

[10] N. Prakash, V. Lalitha, S. B. Balaji, and P. V. Kumar, "Codes with locality for two erasures," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 7771-7789, Dec. 2019.

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

[12] I. Tamo, A. Barg and A. Frolov, "Bounds on the Parameters of Locally Recoverable Codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3070-3083, June 2016.

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