On One Generalization of LRC Codes with Availability

Stanislav Kruglik*†‡, Marina Dudina*, Valeriya Potapova†‡ and Alexey Frolov*†

* Skolkovo Institute of Science and Technology
† Institute for Information Transmission Problems
‡ Russian Academy of Sciences
Moscow, Russia
Moscow Institute of Physics and Technology
Moscow, Russia

Abstract—We investigate one possible generalization of locally recoverable codes (LRC) with all-symbol locality and availability when recovering sets can intersect in a small number of coordinates. This feature allows us to increase the achievable code rate and still meet load balancing requirements. In this paper we derive an upper bound for the rate of such codes and give explicit constructions of codes with such a property. These constructions utilize LRC codes developed by Wang et al.

I. INTRODUCTION

A locally recoverable code (LRC) is a code over finite alphabet such that each symbol is a function of small number of other symbols that form a recovering set [1], [2], [3], [4], [5]. These codes are important due to their applications in distributed and cloud storage systems. LRC codes are well-investigated in the literature. The bounds on the rate and minimum code distance are given in [1], [3] for the case of large alphabet size. The alphabet-dependent shortening bound (see [6] for the method explanation) is proposed in [7]. Optimal code constructions are given in [8] based on rank-metric codes (for large alphabet size, which is an exponential function of the code length) and in [9] based on Reed-Solomon codes (for small alphabet, which is a linear function of the code length).

The natural generalization of an LRC code is an LRC code with availability (or multiple disjoint recovering sets). Availability allows us to handle multiple simultaneous requests to erased symbol in parallel. This property is very important for hot data that is simultaneously requested by a large number of users. The case of LRC codes with availability is much less investigated. Bounds on parameters of such codes and constructions are given in [10], [11], [12]. Most of the papers focused on information-symbol locality and availability. In what follows we are interested in all-symbol locality and availability that is preferable in applications as it permits a uniform approach system design.

The property of availability decreases maximum achievable code rate [10]. In this paper we propose a new generalization of LRC codes with availability. Namely, we assume that recovering sets can intersect in a small number of coordinates. This feature allows us to increase the achievable code rate and still meet load balancing requirements.

Our contribution is as follows. We investigate one possible generalization of locally recoverable codes (LRC) with all-symbol locality and availability when recovering sets can intersect in a small number of coordinates. We derive an upper bound for the rate of such codes and give explicit constructions of codes with such a property. These constructions utilize LRC codes developed in [13].

II. PRELIMINARIES

A. LRC codes

Let us denote by \( \mathbb{F}_q \) a field with \( q \) elements. Let \( [n] = \{1,2,\ldots,n\} \). The code \( C \subseteq \mathbb{F}_q^n \) has locality \( r \) if every symbol of the codeword \( c \in C \) can be recovered from a subset of \( r \) other symbols of \( c \). In other words, this means that, given \( c \in C, i \in [n] \), there exists a subset of coordinates \( R_i \subseteq [n]\setminus\{i\}, |R_i| \leq r \) such that the restriction of \( C \) to the coordinates in \( R_i \) enables one to find the value of \( c_i \). The subset \( R_i \) is called a recovering set for the symbol \( c_i \).

B. LRC codes with availability

Generalizing this concept, assume that every symbol of the code \( C \) can be recovered from \( t \) disjoint subsets of symbols of size \( r \). More formally, denote by \( \mathcal{C}_i \) the restriction of the code \( C \) to a subset of coordinates \( I \subseteq [n] \). Given \( a \in \mathbb{F}_q \) define the set of codewords \( \mathcal{C}(i,a) = \{c \in C : c_i = a\}, i \in [n] \).

Definition 1: A code \( C \) is said to have \( t \) disjoint recovering sets if for every \( i \in [n] \) there are \( t \) pairwise disjoint subsets \( \mathcal{R}_i^1, \ldots, \mathcal{R}_i^t \subseteq [n]\setminus\{i\} \) such that for all \( j = 1,\ldots,t \) and every pair of symbols \( a, a' \in \mathbb{F}_q, a \neq a' \)

\[
\mathcal{C}(i,a)_{\mathcal{R}_i^j} \cap \mathcal{C}(i,a')_{\mathcal{R}_i^j} = \emptyset.
\]

In what follows we refer these codes as \((r,t)\)-LRC codes. We briefly list the existing results below. The first bound for \((r,t)\)-LRC codes was given in [14], [15]

\[
d \leq n - k + 2 - \left[ \frac{t(k-1) + 1}{t(r-1) + 1} \right].
\]
An improvement of this bound was obtained in \cite{10}

\[ d \leq n - \sum_{i=0}^{t} \left\lceil \frac{k - 1}{r^i} \right\rceil. \]

An alphabet-dependent bound was proposed in \cite{12} and has form

\[ d \leq \min_{1 \leq x \leq \left\lceil \frac{k - 1}{r^i} \right\rceil, 1 \leq y_j \leq t; j \in [x]} d_{t-opt}^q[n - B, k - A], \]

where \( A = \sum_{j=1}^{t} (r - 1) y_j + x, B = \sum_{j=1}^{x} r y_j + x \) and \( d_{t-opt}^q \) denote the largest possible minimum distance of a code over \( \mathbb{F}_q \).

The bound on the rate of \((r, t, x)-\text{LRC codes} \) was given in \cite{10}

\[ \frac{k}{n} \leq R^*(r, t) = \prod_{i=1}^{t} \frac{1}{1 + \frac{1}{t^i}}. \]  

This bound was improved in \cite{11} for \( t = 2 \).

In \cite{13} a recursive construction of binary \((r, t)-\text{LRC codes} \) was proposed. The parameters of these codes are as follows: \( n = \binom{r + t}{t} \), \( R = \frac{n}{r^t + 1} \) and \( d = t + 1 \). We refer these codes as WZL codes. WZL code is defined by its parity-check matrix. Let \( m = r + t \). Let us define matrix \( \mathbf{H}(m, t) \) as follows. Each row of \( \mathbf{H}(m, t) \) is associated with \((t - 1)\)-subset of \([m]\) sorted in lexicographical order, each column \( j \) with \( t \)-subset of \([m]\) also sorted in lexicographical order. In this case the element \((i, j)\) of \( \mathbf{H}(m, t) \) is equal to 1 if \( E_i \subseteq F_j \), where \( E_i \) is \((t - 1)\)-subset of \([m]\) associated with \( i \)-th row and \( F_j \) \( t \)-subset of \([m]\) associated with \( j \)-th column. It must be mentioned that \( \mathbf{H}(m, t) \) has \( \binom{m}{t-1} \) rows and \( m^t \) columns and has the following structure:

\[ \mathbf{H}(m, t) = \begin{pmatrix} \mathbf{H}(m - 1, t - 1) & 0 \\ \mathbf{f}_{m-1} & \mathbf{H}(m - 1, t) \end{pmatrix}, \]

where \( \mathbf{H}(m, m) = \mathbf{H}(m, 1)^T = (1, \ldots, 1)^T \) and \( \dim(\mathbf{H}(m, 1)) = \dim(\mathbf{H}(m, t)) = m. \)

C. \textbf{LRC codes with availability and intersection of recovering sets}

Let us give to recovering sets an ability to intersect in at most \( x \) positions and define this code as \((r, t, x)-\text{LRC} \). More formally, we can say

\textbf{Definition 2:} A code \( C \) is said to be \((r, t, x)-\text{LRC} \) if for every \( i \in [n] \) there are \( t \) subsets \( R^1_i, \ldots, R^t_i \subseteq [n] \setminus i \) such that the following relations follow

1) for every pair \( l, l' \in [t], l \neq l' \)

\[ |R^1_l \cap R^t_{l'}| \leq x; \]

2) for all \( j = 1, \ldots, t \) and every pair of symbols \( a, a' \in \mathbb{F}_q, a \neq a' \)

\[ \mathcal{C}(i, a)_{R^1_l} \cap \mathcal{C}(i, a')_{R^t_{l'}} = \emptyset. \]

In what follows we investigate the parameters of such codes.

III. \textbf{AN UPPER BOUND ON THE RATE OF \((r, t, x)-\text{LRC CODES} \)}

A. The recovery graph

Based on the original idea from \cite{10} we represent locally recoverable codes with locality \( r \) and availability \( t \) as a graph \( G \) in the following way. In accordance to the Definition \( 2 \) a coordinate \( i \) has \( t \) recovering sets \( R^1_i, \ldots, R^t_i \), each of size \( r \), where \( R^1_i \subseteq [n] \setminus i \). Define a directed graph \( G \) as follows. The set of vertices \( V = [n] \) corresponds to the set of \( n \) coordinates of the LRC code. The ordered pair of vertices \((i, j)\) forms a directed edge \( i \rightarrow j \) if \( j \in R^1_i \) for some \( i \in [t] \). We color the edges of the graph with \( t \) distinct colors in order to differentiate between the recovering sets of each coordinate.

Note, that as the recovering sets can intersect, then some edges may have several colors. We call \( G \) the \textit{recovery graph} of the code \( C \).

In what follows we need the following lemma

\textbf{Lemma 1:} Let \( j \in [t] \) and \( s = \min\{j, |r/x| + 1\} \), then

\[ \frac{(2r - (s - 1)x)}{2} s = \bigcap_{l=1}^{j} \mathcal{R}^l_i \leq \bigcap_{l=1}^{s} \mathcal{R}^l_i \leq \mathbb{N}(r, j, x) = jr. \]

\textbf{Proof:} The upper bound is trivial and correspond to the case, when recovering sets do not intersect. To prove the lower bound assume, that any two recovering sets intersect in exactly \( x \) positions, we have

\[ \bigcap_{l=1}^{t} \mathcal{R}^l_i \geq r + (r - x) + (r - 2x) + \ldots + (r - (s - 1)x) \]

\[ = \frac{(2r - (s - 1)x)}{2} s. \]

\textbf{Corollary 1:} The out-degree of each vertex \( i \in V = V(G) \) is upper bounded with \( tr \) and lower bounded with \( \mathbb{N}(r, t, x) \).

B. Upper bound on the rate

The proof is very similar to the proof from \cite{10}. For the simplicity of the reader we present the proof here in all the details. Let us introduce the following function

\[ f(r, t, x) = \sum_{j=1}^{t} \sum_{j \equiv 1 \mod 2} \binom{t}{j} \frac{1}{\mathbb{N}(r, j, x) + 1} - \sum_{j=1}^{t} \sum_{j \equiv 0 \mod 2} \binom{t}{j} \frac{1}{\mathbb{N}(r, j, x) + 1} \]

The following lemma will be used in the proof.

\textbf{Lemma 2:} There exists a subset of vertices \( U \subseteq V \) of size at least

\[ |U| \geq n f(r, t, x), \]

such that for any \( U' \subseteq U \), the induced subgraph \( G_{U'} \) on the vertices \( U' \) has at least one vertex \( v \in U' \) such that its set of outgoing edges \( \{(v, j) \in U'\} \) is missing at least one color.
For a given permutation $\tau$ of the set of vertices $V = [n]$, we define the coloring of some of the vertices as follows: The color $j \in [t]$ is assigned to the vertex $v$ if $\tau(v) > \tau(m)$ for all $m \in R^i_j$.

If this condition is satisfied for several recovering sets $R^i_j$, the vertex $v$ is assigned any of the colors $j$ corresponding to these sets. Finally, if this condition is not satisfied at all, then the vertex $v$ is not colored.

Let $U$ be the set of colored vertices, and consider one of its subsets $U \subseteq U$.

Let $G_U$ be the induced subgraph on $U$. We claim that there exists $v \in U$ such that its set of outgoing edges is missing at least one color in $G_U$. Assume toward a contradiction that every vertex of $G_U$ has outgoing edges of all $t$ colors. Choose a vertex $v \in U$ and construct a walk through the vertices of $G_U$ according to the following rule. If the path constructed so far ends at some vertex with color $j$, choose one of its outgoing edges also colored in $j$ and leave the vertex moving along this edge. By assumption, every vertex has outgoing edges of all $t$ colors, so this process and hence this path can be extended indefinitely. Since the graph $G_U$ is finite, there will be a vertex, call it $v_1$, that is encountered twice. The segment of the path that begins at $v_1$ and returns to it has the form

$$v_1 \to v_2 \to \ldots \to v_i,$$

where $v_1 = v_i$. For any $i = 1, \ldots, l - 1$ the vertex $v_i$ and the edge $(v_i, v_{i+1})$ are colored with the same color. Hence by the definition of the set $U$ we conclude that $\tau(v_i) > \tau(v_{i+1})$ for all $i = 1, \ldots, l - 1$, a contradiction.

In order to show that there exists such a set $U$ of large cardinality, we choose the permutation $\tau$ randomly and uniformly among all the $n!$ possibilities and compute the expected cardinality of the set $U$.

Let $A_{v,j}$ be the event that $j$ holds for the vertex $v$ and the color $j$. Since $\Pr(A_{v,j})$ does not depend on $v$, we suppress the subscript $v$, and write

$$\Pr(v \in U) = \Pr(\cup_{j=1}^t A_j).$$

Let us compute the probability of the event $\cup_{j=1}^t A_j$. Note that for any set $S \subseteq [t]$ the probability of the event that all the $A_j, j \in S$ occur simultaneously can be estimated as follows

$$\frac{1}{N(r, |S|, x)} \leq P(\cap_{j \in S} A_j) \leq \frac{1}{N(r, |S|, x) + 1}.$$  

Hence by the inclusion exclusion formula we get

$$\Pr(\cup_{j=1}^t A_j) = \sum_{j=1}^t (-1)^{j-1} \binom{t}{j} P(A_1 \cap \ldots \cap A_j) \geq \sum_{j=1, j \equiv 0 \mod 2}^t \frac{1}{N(r, j, x) + 1} - \sum_{j=1, j \equiv 2 \mod 2}^t \frac{1}{N(r, j, x) + 1} = f(r, t, x).$$

Now let $X_v$ be the indicator random variable for the event that $v \in U$, then

$$\mathbb{E}(|U|) = \sum_{v \in V} \mathbb{E}(X_v) = \sum_{v \in V} \Pr(v \in U) = n \Pr(\cup_{j=1}^t A_j) \geq nf(r, t, x).$$

The proof is completed by observing that there exists at least one choice of $\tau$ for which $|U| \geq \mathbb{E}(|U|)$.

**Theorem 1:** The rate of an $(r, t, x)$-LRC code $C$ satisfies

$$R(C) \leq R^*(r, t, x) = 1 - f(r, t, x).$$

**Proof:** The colored vertices can be viewed as check symbols as they can be recovered from the rest symbols. Thus, the number of information symbols can be estimated as follows

$$k \leq n(1 - f(r, t, x)).$$

### IV. LOWER BOUNDS ON THE RATE OF $(r, t, x)$-LRC CODES

In this section we derive a lower bound on the rate of codes with all symbol locality and availability in which recovering sets can intersect. To find a lower bound we propose the following rather simple code construction. In what follows we explain how to construct a parity-check matrix of a linear $(r, t, x)$-LRC code. We start with a parity-check of $(\tilde{r}, \tilde{x})$-WZL code. Let is denote the matrix by $H_{WZL}$. The matrix $H$ of $(r, t, x)$-LRC code is constructed as follows

$$H = H_{WZL} \otimes \begin{bmatrix} 1 & \ldots & 1 \\ x+1 & \ldots & x+1 \end{bmatrix},$$

where $\otimes$ denotes a Kronecker product of matrices.

As a result, we have a matrix of length $n = (x+1)\binom{r + t}{i}$. It is obvious, that each row of the new matrix will have $(x+1)(\tilde{r} + 1)$ ones and the number of positions, in which two recovering sets intersects is equal to $x$. This construction will have the same availability $t$ as it was for the standard WZL code. Thus, the parameters of the resulting code are as follows

$$r = (\tilde{r} + 1)(x + 1) - 1, \quad t = \tilde{t}.$$  

The matrix $H$ has exactly the same rank as the matrix $H_{WZL}$, the rank is equal to $(\frac{r + t - 1}{l - 1})$. Thus, the rate of the resulting code can be calculated as follows

$$R = 1 - \frac{\tilde{t}}{(\tilde{r} + t)(x + 1)} = 1 - \frac{t}{(r + 1)(x + 1)} \frac{1}{x + 1}.$$
TABLE I
UPPER BOUNDS ON THE RATE OF \((r, t, x)\)-LRC CODES

| \((r, t)\) | \(x = 0\) | \(x = 1\) | \(x = 2\) | \(x = 3\) |
|-----------|-----------|-----------|-----------|-----------|
| (4, 2)    | 0.7111    | 0.7250    | 0.7429    | 0.7664    |
| (5, 2)    | 0.7576    | 0.7667    | 0.7778    | 0.7917    |
| (6, 2)    | 0.7912    | 0.7976    | 0.8052    | 0.8143    |
| (7, 2)    | 0.8167    | 0.8214    | 0.8269    | 0.8333    |
| (4, 3)    | 0.6564    | 0.6981    | 0.7516    | 0.8231    |
| (5, 3)    | 0.7102    | 0.7375    | 0.7708    | 0.8125    |
| (6, 3)    | 0.7496    | 0.7688    | 0.7915    | 0.8188    |
| (7, 3)    | 0.7795    | 0.7938    | 0.8103    | 0.8295    |

TABLE II
COMPARISON OF UPPER AND LOWER BOUNDS ON THE RATE OF \((r, t, x)\)-LRC CODES

| \((r, t)\) | \(x = 0\) (WZL) | \(R\) \((r, t, x = 0)\) | \(x = 1\) | \(R\) \((r, t, x = 1)\) |
|-----------|-----------------|------------------------|-----------|------------------------|
| (3, 2)    | 0.6000          | 0.6429                 | 0.6667    | 0.6667                 |
| (5, 2)    | 0.7143          | 0.7576                 | 0.7500    | 0.7667                 |
| (7, 2)    | 0.7778          | 0.8167                 | 0.8000    | 0.8214                 |
| (3, 3)    | 0.4620          | 0.5275                 | 0.6250    | 0.6500                 |
| (5, 3)    | 0.6250          | 0.7102                 | 0.7000    | 0.7375                 |
| (7, 3)    | 0.7000          | 0.7795                 | 0.7500    | 0.7938                 |

Example 1: Let us start from an \((2, 2)\)-WZL code

\[ H_{WZL} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \]

and construct a parity-check matrix of an \((5, 2, 1)\)-LRC code. The matrix has a rate \(R = 0.75\) and shown below

\[ H = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \]

V. NUMERICAL RESULTS

In Table I we present the comparison of upper bounds on the rate of \((r, t, x)\)-LRC codes for different values of parameter \(x\). We see that the value of the upper bound increases with the parameter \(x\). In the Table II we present the comparison of the code rate obtained by proposed code construction \((x = 1)\) and the code rate of WZL codes with the same locality and availability. In addition, we include the values of the upper bounds for the code rate from [10] and the upper bounds for the code rate proposed in this paper. We see, that e.g. for \(r = 3\), \(t = 2\) and \(x = 1\) the lower bound is tight and it is better, then the upper bound for the case of \(r = 3\), \(t = 2\) and \(x = 0\).

VI. CONCLUSION

We investigated one possible generalization of locally recoverable codes (LRC) with all-symbol locality and availability when recovering sets can intersect in a small number of coordinates. This feature allows us to increase the achievable code rate and still meet load balancing requirements. In this paper we derived an upper bound for the rate of such codes and gave explicit constructions of codes with such a property.

ACKNOWLEDGMENT

A. Frolov thanks A. Barg for introducing this problem to him and for numerous fruitful discussions during his stay in University of Maryland.

REFERENCES

[1] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, “On the locality of codeword symbols,” IEEE Trans. Inf. Theory, vol. 58, no. 11, pp. 6925–6934, Nov. 2011.
[2] P. Gopalan, C. Huang, B. Jenkins, and S. Yekhanin, “Explicit maximally recoverable codes with locality,” IEEE Trans. Inf. Theory, vol. 60, no. 9, pp. 5245–5256, Sep. 2014.
[3] D. S. Papailiopoulos and A. G. Dimakis, “Locally repairable codes,” IEEE Trans. Inf. Theory, vol. 60, no. 10, pp. 5843–5855, Oct. 2014.
[4] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, “Optimal locally repairable and secure codes for distributed storage systems,” IEEE Trans. Inf. Theory, vol. 60, no. 1, pp. 212–236, Jan 2014.
[5] S. Yekhanin, “Locally decodable codes,” Found. Trends Theoretical Comput. Sci., vol. 6, no. 3, pp. 139–255, 2012.
[6] Y. Ben-Haim and S. Litsyn, “Upper bounds on the rate of ldpc codes as a function of minimum distance,” IEEE Trans. Inf. Theory, vol. 52, no. 5, pp. 2092–2100, May 2006.
[7] V. R. Cadambe and A. Mazumdar, “Bounds on the size of locally recoverable codes,” IEEE Trans. Inf. Theory, vol. 61, no. 11, pp. 5787–5794, Nov. 2015.
[8] N. Silberstein, A. S. Rawat, O. Koyluoglu, and S. Vishwanath, “Optimal locally repairable codes via rank metric codes,” in Proceedings IEEE International Symposium on Information Theory (ISIT), Jul. 2013, pp. 1819–1823.
[9] I. Tamo and A. Barg, “A family of optimal locally recoverable codes,” IEEE Trans. Inf. Theory, vol. 60, no. 8, pp. 4661–4676, Aug. 2014.
[10] 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, Jun. 2016.
[11] N. Prakash, V. Lalitha, and P. V. Kumar, “Codes with locality for two erasures,” in Proceedings IEEE International Symposium on Information Theory (ISIT), Jun. 2014, pp. 1962–1966.
[12] P. Huang, E. Yaakobi, H. Uchikawa, and P. H. Siegel, “Linear locally repairable codes with Availability,” in Proceedings IEEE International Symposium on Information Theory (ISIT), Jun. 2015, pp. 1871–1875.
[13] A. Wang, Z. Zhang, and M. Liu, “Achieving arbitrary locality and availability in binary codes,” in Proceedings IEEE International Symposium on Information Theory (ISIT), Jun. 2015, pp. 1866–1870.
[14] A. Wang and Z. Zhang, “Repair locality with multiple erasure tolerance,” IEEE Trans. Inf. Theory, vol. 60, no. 11, pp. 6979–6987, Nov 2014.
[15] A. S. Rawat, D. S. Papailiopoulos, A. G. Dimakis, and S. Vishwanath, “Locality and availability in distributed storage,” in Proceedings IEEE International Symposium on Information Theory (ISIT), June 2014, pp. 681–685.