Linear Locally Repairable Codes with Availability

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
[†]Computer Science Dept., Technion – Israel Institute of Technology, Haifa 32000, Israel
[‡]Toshiba Corporation, Japan
{*pehuang, psiegel*}@*ucsd.edu*, *yaakobi@cs.technion.ac.il, hironori.uchikawa@toshiba.co.jp*

***Abstract*—In this work, we present a new upper bound on the minimum distance d of linear locally repairable codes (LRCs) with information locality and availability. The bound takes into account the code length n, dimension k, locality r, availability t, and field size q. We use tensor product codes to construct several families of LRCs with information locality, and then we extend the construction to design LRCs with information locality and availability. Some of these codes are shown to be optimal with respect to their minimum distance, achieving the new bound. Finally, we study the all-symbol locality and availability properties of several classes of one-step majority-logic decodable codes, including cyclic simplex codes, cyclic difference-set codes, and 4-cycle free regular low-density parity-check (LDPC) codes. We also investigate their optimality using the new bound.**

I. Introduction

Locally repairable codes (LRCs) are a class of codes in which any symbol of a codeword can be recovered by accessing at most r other symbols, where r is a predetermined value [5], [10], [12]. They have received considerable attention in recent years due to their applications in distributed storage systems. Several groups of authors have considered constructions of LRCs and bounds on their properties [3]–[7], [10], [12], [14], [15], [17].

In addition to their symbol locality, another important property of LRCs is their symbol availability, meaning the number of disjoint sets of symbols that can be used to recover any given symbol. High availability is a particularly attractive property for so-called *hot data* in a distributed storage network. More precisely, a code $\mathcal{C}$ has *all-symbol locality r and availability t* if every code symbol can be recovered from t disjoint repair sets of other symbols, each set of size at most r symbols. We refer to such a code as an $(r,t)_a$-LRC. If the code is systematic and these properties apply only to its information symbols, then the code has *information locality r and availability t*, and it is referred to as an $(r,t)_i$-LRC.

Several recent works have considered codes with both locality and availability properties. In [18], it was shown that the minimum distance d of an $[n,k,d]$ linear $(r,t)_i$-LRC satisfies the upper bound

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

In [13], it was proved that bound (1) is also applicable to (n,M,d) non-linear $(r,t)_i$-LRCs, where M denotes the codebook size and $k=\log_q M$. In the same paper, it was shown that if a linear $(r,t)_i$-LRC has the property that each repair set contains only one parity symbol, then d satisfies the upper bound

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

and codes achieving bound (2) were constructed using maximum distance separable (MDS) codes and Gabidulin codes. For $(r,t)_a$-LRCs over $\mathbb{F}_q$ with parameters (n,M,d), it was shown in [16] that d satisfies

$$d \leqslant n-\sum_{i=0}^{t} \left\lfloor \frac{k-1}{r^i} \right\rfloor . \tag{3}$$

There are several constructions of LRCs with availability. In [17], two constructions of $(r,2)_a$-LRCs are proposed. One relies on the combinatorial concept of orthogonal partitions, and the other one is based on product codes. In [11], a class of $(r,t)_a$-LRCs is constructed from partial geometries. A family of systematic fountain codes having information locality and strong probabilistic guarantees on availability were introduced in [1].

In this paper, we consider the problem of finding fundamental bounds and explicit code constructions for LRCs with availability over a fixed alphabet. In Section II, we formally define the problem and present new bounds on the dimension k and the minimum distance d of $[n,k,d]_q$ linear $(r,t)_i$-LRCs, based on the framework established in [3]. Our bound on d explicitly takes into consideration the code length n, dimension k, locality r, availability t, and field size q. In Section III, we use tensor product codes to construct various families of $(r,1)_i$-LRCs, some of which are proved to have optimal minimum distance from our bound. The same construction structure can be naturally extended to construct $(r,t)_i$-LRCs with $t>1$. In Section IV, we study the construction of $(r,t)_a$-LRCs. We show that the structure of one-step majority-logic decoding is highly related to the availability of the code. We review several families of one-step majority-logic decodable codes and identify the locality and availability of these codes. We conclude the paper in Section V. Due to space limitations, proofs of many results in the paper are omitted.

II. Bounds for LRCs with Information Locality and Availability

We begin with some basic definitions and notational conventions. We use the notation $[n]$ to define the set $\{1,\dots,n\}$. For a length-n vector $\boldsymbol{v}$ and a set $\mathcal{I} \subseteq [n]$, $\boldsymbol{v}_{\mathcal{I}}$ denotes the restriction of the vector $\boldsymbol{v}$ to coordinates in the set $\mathcal{I}$. A linear code $\mathcal{C}$ over $\mathbb{F}_q$ of length n, dimension k, and minimum distance d will be denoted by $[n,k,d]_q$, and its generator matrix is $G=(g_1,\cdots,g_n)$, where $g_i \in \mathbb{F}_q^k$ is a column vector for $i \in [n]$. We define $k_{\mathcal{I}}(\mathcal{C}) = \log_q |\{\boldsymbol{c}_{\mathcal{I}} : \boldsymbol{c} \in \mathcal{C}\}|$, and, for simplicity, we write $k_{\mathcal{I}}$ instead of $k_{\mathcal{I}}(\mathcal{C})$ when $\mathcal{C}$ is known from the context.

We follow the conventional definitions of linear LRCs with availability, as established in [13], [16], [18].

Definition 1. *The ith code symbol of an $[n,k,d]_q$ linear code $\mathcal{C}$ is said to have locality r and availability t if there exist t pairwise disjoint repair sets $\mathcal{R}_i^1,\ldots,\mathcal{R}_i^t \subseteq [n]\backslash\{i\}$, such that (1) $|\mathcal{R}_i^j| \leqslant r, 1 \leqslant j \leqslant t$, and (2) for each repair set $\mathcal{R}_i^j$, $1 \leqslant j \leqslant t$, g_i is a linear combination of g_u, $u \in \mathcal{R}_i^j$.*

Definition 2. *Let $\mathcal{C}$ be an $[n,k,d]_q$ linear code. A set $\mathcal{I} \subseteq [n]$ is said to be an information set if $|\mathcal{I}| = k_{\mathcal{I}} = k$.*
(a) The code $\mathcal{C}$ is said to have all-symbol locality r and availability t if every code symbol has locality r and availability t. We refer to $\mathcal{C}$ as a linear $(r,t)_a$-LRC.
(b) The code $\mathcal{C}$ is said to have information locality r and availability t if there is an information set $\mathcal{I}$ such that, for any $i \in \mathcal{I}$, the ith code symbol has locality r and availability t. We refer to $\mathcal{C}$ as a linear $(r,t)_i$-LRC.

It is straightforward to verify that the minimum distance d of a linear $(r,t)_a$-LRC satisfies $d \geqslant t+1$. We now develop new upper bounds on the dimension and minimum distance, respectively, of linear $(r,t)_i$-LRCs.

Lemma 3. *Let $\mathcal{C}$ be an $[n,k,d]_q$ linear $(r,t)_i$-LRC. For positive integers x and y_j, $j \in \mathbb{Z}^+$, define $A = \sum_{j=1}^{x}(r-1)y_j + x$ and $B = \sum_{j=1}^{x} ry_j + x$. If $1 \leqslant x \leqslant \lceil \frac{k-1}{(r-1)t+1} \rceil$, $1 \leqslant y_j \leqslant t$, $j \in [x]$, and $A < k$, then there exists a set $\mathcal{I} \subseteq [n]$ with $|\mathcal{I}| = B$ such that $k_{\mathcal{I}} \leqslant A$.*

Now, let $k_{\ell-opt}^{(q)}[n,d]$ denote the largest possible dimension of an $[n,k,d]_q$ linear code, and let $d_{\ell-opt}^{(q)}[n,k]$ denote the largest possible minimum distance of such a code. Applying Lemma 3, we get the following upper bounds on k and d for $[n,k,d]_q$ linear $(r,t)_i$-LRCs.

Theorem 4. *For any $[n,k,d]_q$ linear $(r,t)_i$-LRC, the dimension satisfies*

$$k \leqslant \min_{\substack{1\leqslant x\leqslant\lceil \frac{k-1}{(r-1)t+1}\rceil,\ 1\leqslant y_j\leqslant t,\ j\in[x] \\ A<k,\ x,y_j\in\mathbb{Z}^+}} \left\{ A + k_{\ell-opt}^{(q)}[n-B,d] \right\} \tag{4}$$

and the minimum distance d satisfies

$$d \leqslant \min_{\substack{1\leqslant x\leqslant\lceil \frac{k-1}{(r-1)t+1}\rceil,\ 1\leqslant y_j\leqslant t,\ j\in[x] \\ A<k,\ x,y_j\in\mathbb{Z}^+}} d_{\ell-opt}^{(q)}[n-B,k-A]. \tag{5}$$

Remark: When $t=1$, bounds (4) and (5) apply to all linear codes with information locality r. Bounds (4) and (5) also hold for linear $(r,t)_a$-LRCs.

III. Construction of LRCs with Information Locality

Tensor product codes, first proposed by Wolf in [20], are a family of codes defined by a parity-check matrix that is the tensor product of the parity-check matrices of two constituent codes. As shown in [20] and later [8], tensor product codes offer a range of error correction and detection properties, depending on the choice of the constituent codes. In this section, we construct several families of $(r,t)_i$-LRCs using tensor product constructions. We focus on binary codes with availability $t=1$, but our constructions can be extended to non-binary codes and codes with $t>1$.

$$
\begin{array}{|ccc|}
\hline
 & \mathbf{I} & \\
\mu_{11} & \cdots & \mu_{1k'} \\
\mu_{21} & \cdots & \mu_{2k'} \\
\vdots & \vdots & \vdots \\
\mu_{\ell 1} & \cdots & \mu_{\ell k'} \\
\hline
\end{array}
\quad
\begin{array}{|c|}
\hline
\mathbf{II} \\
p_{L_1} \\
p_{L_2} \\
\vdots \\
p_{L_\ell} \\
\hline
\end{array}
\quad
\begin{array}{ccc}
 & \mathbf{III} & \\
p_{11} & \cdots & p_{1,n'-k'} \\
p_{21} & \cdots & p_{2,n'-k'} \\
\vdots & \vdots & \vdots \\
p_{\ell 1} & \cdots & p_{\ell,n'-k'} \\
\hline
p_{G_1} & \cdots & p_{G_{n'-k'}} \\
 & \mathbf{IV} & \\
\hline
\end{array}
$$

Fig. 1. An $(r,1)_i$-LRC using *Construction A*. Information symbols are in block **I**, local parity-check symbols are in block **II**, phantom symbols are in block **III**, and global parity-check symbols are in block **IV**.

A. Construction of $(r,1)_i$-LRCs

Our general framework for constructing systematic linear $(r,1)_i$-LRCs is depicted in Fig 1. We first specify an $[n',k',d']$ systematic code as a base code, $\mathcal{C}_{base}$. If a parity-check symbol of a systematic code is the sum of all information symbols, it is referred to as an *information-sum parity-check symbol*. For simplicity, let us first assume that $\mathcal{C}_{base}$ has no information-sum parity-check symbol. The following construction, which is depicted pictorially in Fig. 1, produces an $(r,1)_i$-LRC of length $n = (k'+1)\ell + n' - k'$, dimension $k = k'\ell$, and information locality $r = k'$.

Construction A

Step 1: Place an $\ell \times k'$ array of information symbols in block **I**.

Step 2: For each row of information symbols, $(\mu_{i1},\ldots,\mu_{ik'})$, $1 \leqslant i \leqslant \ell$, compute *local* parity-check symbols $p_{L_i} = \sum_{j=1}^{k'} \mu_{ij}$, $1 \leqslant i \leqslant \ell$, and place them in the corresponding row of block **II**.

Step 3: Encode each row of information symbols in block **I** using $\mathcal{C}_{base}$, producing parity-check symbols $(p_{i1},\ldots,p_{i,n'-k'})$, $1 \leqslant i \leqslant \ell$. Place these parity-check symbols in block **III**. (These symbols are referred to as *phantom symbols* because they will not appear in the final codeword.)

Step 4: Compute a row of *global* parity-check symbols, $p_{G_j} = \sum_{i=1}^{\ell} p_{ij}$, $1 \leqslant j \leqslant n'-k'$, by summing the rows of phantom symbols in block **III**. Place these symbols in block **IV**.

Step 5: The constructed codeword consists of the symbols in blocks **I**, **II**, and **IV**. □

Note that for $r|k$, Pyramid codes are d-optimal $(r,1)_i$-LRCs over sufficiently large field size [7]. A Pyramid code is constructed by splitting a parity-check symbol of a systematic MDS code into $\frac{k}{r}$ local parity-check symbols. However, the splitting operation is not guaranteed for binary codes. In contrast, we take a different approach. We first design the local parity-check symbols, and then construct the global parity-check symbols.

If $\mathcal{C}_{base}$ has an information-sum parity-check symbol, we simply modify Step 3 of *Construction A*, as follows. After encoding each row of information symbols in block **I**, define the corresponding row of phantom symbols to be the computed parity-check symbols with the information-sum parity-check symbol excluded, and store them in block **III**. Then proceed with the remaining steps in *Construction A*. We refer to this modified construction as ***Construction A'***. It is easy to verify that the resulting code is an $(r,1)_i$-LRC with length $n = (k'+1)\ell + n' - k' - 1$, dimension $k = k'\ell$, and information locality $r = k'$. We now present constructions of $(r,1)_i$-LRCs with specified minimum distances.

A trivial example of an $(r,1)_i$-LRC with $d = 2$ can be constructed by choosing $\mathcal{C}_{base}$ to be an $[r+1, r, 2]$ binary single parity-check code. It is easy to see that *Construction* A' generates an $(r,1)_i$-LRC with parameters $[n,k,d] = [(r+1)\ell, r\ell, 2]$ and information locality r. From bound (5), with $x = \ell - 1$ and $y_j = t = 1$, $j \in [x]$, $d \leqslant d^{(2)}_{\ell-opt}[n-B, k-A] = d^{(2)}_{\ell-opt}[r+1, r] = 2$, proving that the code is optimal with respect to the minimum distance.

Next, we consider the family of $(r,1)_i$-LRCs obtained by applying *Construction A* (or *Construction* A', if appropriate) to base codes $\mathcal{C}_{base}$ with parameters $[n', k', d' = 3]$. The following lemma gives a lower bound on the minimum distance of the resulting codes.

Lemma 5. *If $\mathcal{C}_{base}$ is an $[n', k', d' = 3]$ code, the $(r,1)_i$-LRC produced by Construction A (or Construction A', if appropriate) has minimum distance $d \geqslant 3$.*

Based on Lemma 5, we have the following theorem.

Theorem 6. *Let $\mathcal{C}_{base}$ be an $[n', k', d' = 3]$ systematic binary code. If $\mathcal{C}_{base}$ has an information-sum parity-check symbol and $d^{(2)}_{\ell-opt}[n', k'] = 3$, the $(r,1)_i$-LRC obtained from Construction A' has parameters $[n = (k'+1)\ell + n' - k' - 1, k = k'\ell, d = 3]$ and $r = k'$. Its minimum distance $d = 3$ is optimal.*

We give an example of an $(r,1)_i$-LRC with $d = 3$.

Example 1. Let $\mathcal{C}_{base}$ be the $[6,3,3]$ shortened binary Hamming code, whose systematic generator matrix is

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

$\mathcal{C}_{base}$ has an information-sum parity-check symbol and $d^{(2)}_{\ell-opt}[6,3] = 3$. From Theorem 6, the $(r,1)_i$-LRC generated by *Construction* A' has parameters $[4\ell+2, 3\ell, 3]$ and $r = 3$. Its minimum distance $d = 3$ is optimal.

Now, we generalize the $[6,3,3]$ code in Example 1. Let $\mathcal{C}$ and $\mathcal{C}_s$ be systematic binary Hamming codes and shortened binary Hamming codes, with parity-check matrices H and H_s, respectively. H_s is obtained by puncturing H: 1) find all information coordinates of H on which the values of the first row of H are 0, and 2) delete these coordinates of H. As a result, $\mathcal{C}_s$ is systematic and has an information-sum parity-check symbol. We have the following lemma on $\mathcal{C}_s$.

Lemma 7. *$\mathcal{C}_s$ has parameters $[2^{m-1} + m - 1, 2^{m-1} - 1, 3]$, and its minimum distance 3 is optimal.*

The following corollary is a direct result of Theorem 6 and Lemma 7.

Corollary 8 *Let $\mathcal{C}_{base}$ be the shortened binary Hamming code $\mathcal{C}_s$ in Lemma 7. The $(r,1)_i$-LRC obtained from Construction A' has parameters $[2^{m-1}\ell + m - 1, (2^{m-1}-1)\ell, 3]$ and $r = 2^{m-1} - 1$. Its minimum distance 3 is optimal.*

Next, we give a construction of $(r,1)_i$-LRCs with $d = 4$. We start with the following lemma.

Lemma 9. *If $\mathcal{C}_{base}$ is an $[n', k', d' = 4]$ code, the $(r,1)_i$-LRC produced by Construction A (or Construction A', if appropriate) has minimum distance $d \geqslant 4$.*

Based on Lemma 9, we have the following two theorems.

Theorem 10. *Let $\mathcal{C}_{base}$ be an $[n', k', d' = 4]$ systematic binary code. If $\mathcal{C}_{base}$ has no information-sum parity-check symbol and $d^{(2)}_{\ell-opt}[n'+1, k'] = 4$, the $(r,1)_i$-LRC obtained from Construction A has parameters $[n = (k'+1)\ell + n' - k', k = k'\ell, d = 4]$ and $r = k'$. Its minimum distance $d = 4$ is optimal.*

Theorem 11. *Let $\mathcal{C}_{base}$ be an $[n', k', d' = 4]$ systematic binary code. If $\mathcal{C}_{base}$ has an information-sum parity-check symbol and $d^{(2)}_{\ell-opt}[n', k'] = 4$, the $(r,1)_i$-LRC obtained from Construction A' has parameters $[n = (k'+1)\ell + n' - k' - 1, k = k'\ell, d = 4]$ and $r = k'$. Its minimum distance $d = 4$ is optimal.*

The following example is an $(r,1)_i$-LRC with $d = 4$.

Example 2. Let $\mathcal{C}_{base}$ be the $[8,4,4]$ extended binary Hamming code, whose systematic generator matrix is

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

$\mathcal{C}_{base}$ has no information-sum parity-check symbol and $d^{(2)}_{\ell-opt}[9,4] = 4$. From Theorem 10, the $(r,1)_i$-LRC generated by *Construction A* has parameters $[5\ell+4, 4\ell, 4]$ and $r = 4$. Its minimum distance $d = 4$ is optimal.

We generalize Example 2 using the following lemma on expurgated and extended binary Hamming codes.

Lemma 12. *For $m \geqslant 4$, a $[2^m - 1, 2^m - 2 - m, 4]$ systematic expurgated binary Hamming code has no information-sum parity-check symbol, and $d^{(2)}_{\ell-opt}[2^m, 2^m - 2 - m] = 4$. For $m \geqslant 3$, a $[2^m, 2^m - 1 - m, 4]$ systematic extended binary Hamming code has no information-sum parity-check symbol, and $d^{(2)}_{\ell-opt}[2^m + 1, 2^m - 1 - m] = 4$.*

Using expurgated binary Hamming codes and extended binary Hamming codes as base codes, we get the following constructions of $(r,1)_i$-LRCs from Theorem 10.

Corollary 13 *Let $\mathcal{C}_{base}$ be a $[2^m - 1, 2^m - 2 - m, 4]$ systematic expurgated binary Hamming code, where $m \geqslant 4$. The $(r,1)_i$-LRC obtained from Construction A has parameters $[(2^m - 1 - m)\ell + m + 1, (2^m - 2 - m)\ell, 4]$ and $r = 2^m - 2 - m$. Its minimum distance 4 is optimal.*

Corollary 14 *Let $\mathcal{C}_{base}$ be a $[2^m, 2^m - 1 - m, 4]$ systematic extended binary Hamming code, where $m \geqslant 3$. The $(r,1)_i$-LRC obtained from Construction A has parameters $[(2^m - m)\ell + m + 1, (2^m - 1 - m)\ell, 4]$ and $r = 2^m - 1 - m$. Its minimum distance 4 is optimal.*

Next, we give a construction of $(r,1)_i$-LRCs for $d \geqslant 5$. For simplicity, in the following, we assume $\mathcal{C}_{base}$ has no information-sum parity-check symbol. Our construction can be easily modified for $\mathcal{C}_{base}$ with an information-sum parity-check symbol. We give a new construction by using two rows of global parity-check symbols as follows.

$$\begin{array}{ccc|c|ccc|ccc}
 & & & & & \mathbf{V} & & & & \\
 & \mathbf{I} & & \mathbf{II} & & \mathbf{III\,(a)} & & & \mathbf{III\,(b)} & \\
\mu_{11} & \cdots & \mu_{1k'} & p_{L_1} & p_{11} & \cdots & p_{1,w} & p_{1,w+1} & \cdots & p_{1,n'-k'} \\
\mu_{21} & \cdots & \mu_{2k'} & p_{L_2} & p_{21} & \cdots & p_{2,w} & p_{2,w+1} & \cdots & p_{2,n'-k'} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mu_{\ell 1} & \cdots & \mu_{\ell k'} & p_{L_\ell} & p_{\ell 1} & \cdots & p_{\ell,w} & p_{\ell,w+1} & \cdots & p_{\ell,n'-k'} \\
 & & & & p_{G_{11}} & \cdots & p_{G_{1,w}} & p_{G_{1,w+1}} & \cdots & p_{G_{1,n'-k'}} \\
 & & & & p_{G_{21}} & \cdots & p_{G_{2,w}} & & \mathbf{IV} &
\end{array}$$

Fig. 2. An $(r,1)_i$-LRC using *Construction B*.

Construction B Let $\mathcal{C}_{base}$ be an $[n',k',5]$ systematic binary code. Follow Steps 1, 2, and 3 of *Construction A* to get the *local* parity-check symbols and phantom symbols. Let $\mathcal{C}'_{base} = \{\boldsymbol{c}_{[k'+w]} : \boldsymbol{c} \in \mathcal{C}_{base}\}$, i.e., restrict $\mathcal{C}_{base}$ to k' information coordinates and w parity-check coordinates, where w is chosen properly such that $\mathcal{C}'_{base}$ has minimum distance at least 3. As shown in Fig. 2, all phantom symbols are divided into two parts: w columns in block **III(a)** and the rest of the columns in block **III(b)**. Then, we compute the *global* parity-check symbols in block **IV**: First, we use Step 4 of *Construction A* to get the first row $(p_{G_{11}},\cdots,p_{G_{1,w}},p_{G_{1,w+1}},\cdots,p_{G_{1,n'-k'}})$. Second, we use a Reed-Solomon (RS) code to encode the phantom symbols in block **III(a)** to get the second row $(p_{G_{21}},\cdots,p_{G_{2,w}})$. Each row in block **III(a)** is considered as a symbol in $\mathbb{F}_{2^w}$. Let α be a primitive element in $\mathbb{F}_{2^w}$, and $\ell \leqslant 2^w-1$. Then, a parity-check matrix for the RS code is

$$H = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 & 0 \\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{\ell-1} & 0 & 1 \end{bmatrix}.$$

Thus, the $\ell+2$ rows in block **V** form an $[\ell+2,\ell,3]$ MDS codeword and any two erased rows can be corrected. Note that an alternative to the RS code is an EVENODD code [2]. A codeword of the constructed $(r,1)_i$-LRC consists of the symbols in blocks **I**, **II**, and **IV**.

With *Construction B*, we have the following theorem.

Theorem 15. *The $(r,1)_i$-LRC obtained from Construction B has parameters $[n=(k'+1)\ell+n'-k'+w, k=k'\ell, d\geqslant 5]$ ($\ell \leqslant 2^w-1$) and information locality $r=k'$.*

Example 3. Let $\mathcal{C}_{base}$ be a $[2^m-1, 2^m-1-2m, 5]$ binary double-error-correcting BCH code where $m \geqslant 4$. It can be shown that it does not have an information-sum parity-check symbol. For the case of $m=4$, exhaustive search shows that we can choose w to be 4. For $\ell \leqslant 15$, the $(r,1)_i$-LRC from *Construction B* has parameters $[n=8\ell+12, k=7\ell, d\geqslant 5]$ and $r=7$. An upper bound on d from bound (5) is 8. For the case of $m=5$, exhaustive search shows that we can choose w to be 6. For $\ell \leqslant 63$, the $(r,1)_i$-LRC from *Construction B* has parameters $[n=22\ell+16, k=21\ell, d\geqslant 5]$ and $r=21$. An upper bound on d from bound (5) is 8.

The $(r,1)_i$-LRCs constructed in this subsection are summarized in Table I, where the upper bound d^u on d is from bound (5). Similarly, the same construction can be extended to get $(r,1)_i$-LRCs with larger minimum distance. We omit the details here due to space limitations.

B. Construction of $(r,t>1)_i$-LRCs

The previous constructions can all be modified to construct $(r,t)_i$-LRCs with $t>1$. The idea is to add more local parity-check symbols to obtain availability. Here, we only give one example to illustrate the main idea.

TABLE I
CONSTRUCTED $(r,1)_i$-LRCS IN SECTION III-A

Code[a]	n	k	d	r	d^u
$\mathcal{C}_1$	$2^{m-1}\ell+m-1$	$(2^{m-1}-1)\ell$	3	$2^{m-1}-1$	3
$\mathcal{C}_2$	$(2^m-1-m)\ell+m+1$	$(2^m-2-m)\ell$	4	2^m-2-m	4
$\mathcal{C}_3$	$(2^m-m)\ell+m+1$	$(2^m-1-m)\ell$	4	2^m-1-m	4
$\mathcal{C}_4$	$8\ell+12$ ($\ell \leqslant 15$)	7ℓ ($\ell \leqslant 15$)	$\geqslant 5$	7	8
$\mathcal{C}_5$	$22\ell+16$ ($\ell \leqslant 63$)	21ℓ ($\ell \leqslant 63$)	$\geqslant 5$	21	8

[a] $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ are from Corollaries 8, 13, and 14. $\mathcal{C}_4$ and $\mathcal{C}_5$ are from Example 3.

$$\begin{array}{cccc|cc|cccc}
 & \mathbf{I} & & & \mathbf{II} & & & \mathbf{III} & & \\
\mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} & p_{L_1} & p_{D_1} & p_{11} & p_{12} & p_{13} & p_{14} \\
\mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} & p_{L_2} & p_{D_2} & p_{21} & p_{22} & p_{23} & p_{24} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mu_{\ell 1} & \mu_{\ell 2} & \mu_{\ell 3} & \mu_{\ell 4} & p_{L_\ell} & p_{D_\ell} & p_{\ell 1} & p_{\ell 2} & p_{\ell 3} & p_{\ell 4} \\
 & & & & & & p_{G_1} & p_{G_2} & p_{G_3} & p_{G_4} \\
 & & & & & & & \mathbf{IV} & &
\end{array}$$

Fig. 3. An $(r=4,t=2)_i$-LRC with $d=4$. The diagonal local parity-check column in block **II** provides availability.

Example 4. Consider an $\ell \times 4$ information array. We use the $[8,4,4]$ extended binary Hamming code as $\mathcal{C}_{base}$. As shown in Fig. 3, the main difference from *Construction A* in Fig. 1 is that we add one more diagonal local parity-check column in block **II** to obtain availability: $p_{D_j} = \sum_{i=0}^{3} u_{\left[4\lfloor\frac{j-1}{4}\rfloor+1+i\right],\left[\left((j-1)\bmod 4+i\right)\bmod 4+1\right]}$, $1 \leqslant j \leqslant \ell$ and $4|\ell$. The constructed $(r,t)_i$-LRC has parameters $[6\ell+4, 4\ell, 4]$ ($4|\ell$) with $r=4$ and $t=2$. It can be verified that the minimum distance d is 4. Note that we can add one more vertical local parity-check column to further increase availability.

IV. CONSTRUCTION OF LRCS WITH ALL-SYMBOL LOCALITY AND AVAILABILITY

In this section, we study $(r,t)_a$-LRCs based on one-step majority-logic decodable codes [9].

Definition 16. *An $[n,k,d]_q$ linear code $\mathcal{C}$ is said to be a one-step majority-logic decodable code with t orthogonal repair sets if the ith symbol, $i \in [n]$, has t pairwise disjoint repair sets $\mathcal{R}_i^j$, $j \in [t]$, such that the ith symbol is a linear combination of all symbols in $\mathcal{R}_i^j$, for each $j \in [t]$.*

From Definition 16, it is evident that if $\mathcal{C}$ is a one-step majority-logic decodable code with t orthogonal repair sets, and if all repair sets have the same size r, then $\mathcal{C}$ has all-symbol locality r and availability t. Moreover, referring to a well known result [9, Theorem 8.1], we can see that for an $[n,k,d]_q$ one-step majority-logic decodable code with t orthogonal repair sets, all of the same size r, its availability t satisfies

$$t \leqslant \left\lfloor \frac{n-1}{r} \right\rfloor. \tag{6}$$

Note that for a cyclic code, once t repair sets are found for one symbol, the repair sets for any symbol can be determined correspondingly due to the cyclic symmetry of the code. For this reason, most one-step majority-logic decodable codes found so far are cyclic codes. Some representative one-step majority-logic decodable codes are: doubly transitive invariant

TABLE II
DIFFERENCE-SET CODES

$\mathcal{C}$	n	k	d	r	t	t^u	d^u
$m=2$	21	11	6	4	5	5	6
$m=3$	73	45	10	8	9	9	12
$m=4$	273	191	18	16	17	17	31
$m=5$	1057	813	34	32	33	33	80

(DTI) codes, cyclic Simplex codes, cyclic difference-set codes, and 4-cycle free regular linear codes. First, we present two families of one-step majority-logic decodable cyclic codes, and also give their locality and availability.

Example 5. Consider a cyclic binary Simplex code with parameters $[n = 2^m - 1, k = m, d = 2^{m-1}]$ [9]. The one-step majority-logic decoding structure of this code shows that it has all-symbol locality $r = 2$ and availability $t = 2^{m-1} - 1$. This code has the optimal minimum distance, due to the Plotkin bound. Note this locality and availability property of the Simplex codes is also observed independently in [19].

Example 6. Consider a cyclic binary difference-set code with parameters $[n = 2^{2m} + 2^m + 1, k = 2^{2m} + 2^m - 3^m, d = 2^m + 2]$ [9]. From the one-step majority-logic decoding structure of this code, we can verify it has all-symbol locality $r = 2^m$ and availability $t = 2^m + 1$. For the codes with $2 \leqslant m \leqslant 5$, Table II gives the upper bound t^u on t from bound (6) and the upper bound d^u on d from bound (5).

Another important class of one-step majority-logic decodable codes is 4-cycle free linear codes that have a parity-check matrix H with constant row weight ρ and constant column weight γ. Obviously, such codes have all-symbol locality $r = \rho - 1$ and availability $t = \gamma$. In particular, 4-cycle free (ρ, γ)-regular low-density parity-check (LDPC) codes have this property [9]. Based upon this observation, a family of codes with all-symbol locality and availability are constructed using partial geometries in [11]. Lower and upper bounds on the code rate are derived, but the exact dimension and minimum distance of these codes are not known.

Many 4-cycle free regular LDPC codes have been constructed by leveraging different mathematical tools [9], e.g., finite geometries, algebraic methods, and block designs. Here we consider a family of such codes based on Euclidean Geometries (EG), and we give explicit expressions for their code length, dimension, and minimum distance, as well as their locality and availability.

Example 7. Consider the class of binary 4-cycle free regular LDPC codes, called the two-dimensional type-I cyclic $(0, m)$th-order EG-LDPC codes, with parameters $[n = 2^{2m} - 1, k = 2^{2m} - 3^m, d = 2^m + 1]$ [9]. From the structure of their parity-check matrices, they have all-symbol locality $r = 2^m - 1$ and availability $t = 2^m$. For the codes with $2 \leqslant m \leqslant 5$, Table III gives the upper bound t^u on t from bound (6) and the upper bound d^u on d from bound (5).

TABLE III
TWO-DIMENSIONAL TYPE-I CYCLIC $(0, m)$TH-ORDER EG-LDPC CODES

$\mathcal{C}$	n	k	d	r	t	t^u	d^u
$m=2$	15	7	5	3	4	4	5
$m=3$	63	37	9	7	8	8	12
$m=4$	255	175	17	15	16	16	30
$m=5$	1023	781	33	31	32	32	80

V. CONCLUSION

In this work, we investigated a variety of linear LRCs with availability and studied their optimality with respect to the minimum distance using our new bound. Several interesting problems remain open. It is unclear whether the new upper bound is tight in general. It will be of interest to explore $(r, 1)_i$-LRCs based on tensor product codes for larger minimum distance $d \geqslant 5$. In addition, further study of other one-step majority-logic decodable codes is also an interesting direction.

ACKNOWLEDGMENT

The work of P. Siegel at the Technion was supported in part by the Lady David Foundation and by a Viterbi Research Fellowship. The work of E. Yaakobi was supported in part by the Israel Science Foundation (ISF) Grant No. 1624/14.

REFERENCES

[1] M. Asteris and A. Dimakis, "Repairable fountain codes," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 1037–1047, May 2014.

[2] M. Blaum, J. Brady, J. Bruck, and J. Menon, "Evenodd: An efficient scheme for tolerating double disk failures in raid architectures," *IEEE Trans. Comput.*, vol. 44, no. 2, pp. 192–202, 1995.

[3] 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.

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

[5] 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.

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

[7] 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.

[8] H. Imai and H. Fujiya, "Generalized tensor product codes," *IEEE Trans. Inf. Theory*, vol. 27, no. 2, pp. 181–187, Mar 1981.

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

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

[11] 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.

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

[13] A. Rawat, D. Papailiopoulos, A. Dimakis, and S. Vishwanath, "Locality and availability in distributed storage," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2014, pp. 681–685.

[14] 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.

[15] W. Song, S. H. Dau, C. Yuen, and T. Li, "Optimal locally repairable linear codes," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 1019–1036, May 2014.

[16] I. Tamo and A. Barg, "Bounds on locally recoverable codes with multiple recovering sets," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2014, pp. 691–695.

[17] ——, "A family of optimal locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4661–4676, Aug 2014.

[18] A. Wang and Z. Zhang, "Repair locality with multiple erasure tolerance," *arXiv preprint arXiv:1306.4774*, 2013.

[19] ——, "Achieving arbitrary locality and availability in binary codes," *arXiv preprint arXiv:1501.04264*, 2015.

[20] J. Wolf, "On codes derivable from the tensor product of check matrices," *IEEE Trans. Inf. Theory*, vol. 11, no. 2, pp. 281–284, Apr 1965.