

Optimal Locally Repairable Codes: An Improved Bound and Constructions

Han Cai^{id}, *Member, IEEE*, Cuiling Fan^{id}, Ying Miao^{id}, Moshe Schwartz^{id}, *Senior Member, IEEE*,
and Xiaohu Tang^{id}, *Senior Member, IEEE*

Abstract—We study the Singleton-type bound that provides an upper limit on the minimum distance of locally repairable codes. We present an improved bound by carefully analyzing the combinatorial structure of the repair sets. Thus, we show the previous bound is unachievable for certain parameters. We then also provide explicit constructions of optimal codes which show that for certain parameters the new bound is sharp. Additionally, as a byproduct, some previously known codes are shown to attain the new bound and are thus proved to be optimal.

Index Terms—Locally repairable codes, Singleton-type bound.

I. INTRODUCTION

DUE to the ever-growing need for more efficient and scalable systems for cloud storage and data storage in general, *distributed storage systems* (DSSs) (such as the Google data centers and Amazon Clouds) have become increasingly important. In a distributed storage system, a data file is stored at a distributed collection of storage devices/nodes in a network. Since any storage device is individually unreliable and subject to failure, redundancy must be introduced to provide the much-needed system-level protection against data loss due to device/node failure.

Manuscript received 10 November 2020; revised 16 September 2021; accepted 9 March 2022. Date of publication 22 March 2022; date of current version 13 July 2022. The work of Han Cai was supported by the German Israeli Project Cooperation (DIP) under Grant PE2398/1-1. The work of Cuiling Fan was supported in part by the National Natural Science Foundation of China under Grant 11971395 and in part by the Projects of Central Government to Guide Local Scientific and Technological Development under Grant 2021ZYD0001. The work of Ying Miao was supported in part by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (B) under Grant 18H01133. The work of Moshe Schwartz was supported in part by German Israeli Project Cooperation (DIP) Grant under Grant PE2398/1-1. The work of Xiaohu Tang was supported in part by the National Natural Science Foundation of China under Grant 61871331. An earlier version of this paper was presented in part at the IEEE International Symposium on Information Theory 2021 [DOI: 10.1109/ISIT45174.2021.9518040]. (*Corresponding author: Cuiling Fan.*)

Han Cai and Xiaohu Tang are with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China (e-mail: hancai@aliyun.com; xhutang@swjtu.edu.cn).

Cuiling Fan is with the School of Mathematics, Southwest Jiaotong University, Chengdu 610031, China (e-mail: cuilingfan@163.com).

Ying Miao is with the Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Japan (e-mail: miao@sk.tsukuba.ac.jp).

Moshe Schwartz is with the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Be'er Sheva 8410501, Israel (e-mail: schwartz@ee.bgu.ac.il).

Communicated by A. Mazumdar, Associate Editor for Machine Learning and Statistics.

Digital Object Identifier 10.1109/TIT.2022.3161613

In today's large distributed storage systems, where node failures are the norm rather than the exception, designing codes that have good distributed repair properties has become a central problem. Several cost metrics and related tradeoffs have been studied in the literature, for example *repair bandwidth* [4], [5], *disk-I/O* [23], and *repair locality* [4], [8], [12]. In this paper *repair locality* is the subject of interest.

Motivated by the desire to reduce repair cost in the design of erasure codes for distributed storage systems, the notions of *symbol locality* and *locally repairable codes* (LRC) were introduced in [8] and [13], respectively. The i th coded symbol of an $[n, k]$ linear code \mathcal{C} is said to have locality r if it can be recovered by accessing at most r other symbols in \mathcal{C} . Alternatively, the i th code symbol with the r other symbols form a 1-erasure correcting code. The concept was further generalized to (r, δ) -locality by Prakash *et al.* [14] to address the situation of multiple device failures. Here, the i th coordinate, together with $r + \delta - 2$ other coordinates, form a code capable of correcting $\delta - 1$ erasures. When $\delta = 2$ this coincides with the definition of locality.

There are two types of linear codes with (r, δ) -locality considered in the literature. The first is *information symbol locality*, pertaining to systematic linear codes whose information symbols all have (r, δ) -locality (denoted by $(r, \delta)_i$ -locality for short). The second is of *all-symbol locality* (or $(r, \delta)_a$ -locality) pertaining to linear codes all of whose symbols have (r, δ) -locality.

For any $[n, k, d]_q$ -linear code with minimum Hamming distance d over the finite field \mathbb{F}_q , the Singleton bound [18] is given by

$$d \leq n - k + 1, \quad (1)$$

which is one of the most classical theorems in coding theory. This bound was generalized for locally repairable codes in [8] (the case $\delta = 2$) and [14] (general δ) as follows. An $[n, k, d]_q$ -linear LRC with $(r, \delta)_i$ -locality satisfies

$$d \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1). \quad (2)$$

It was also proved that a class of codes known as pyramid codes [9] achieves this bound when the alphabet is sufficiently large, say $q \geq n + 1$ and $d \geq \delta$ (for a weaker field-size requirement please refer to [3]). Since a linear code with $(r, \delta)_a$ -locality is also a linear code with $(r, \delta)_i$ -locality, (2) also presents an upper bound for the minimum Hamming distance

of $(r, \delta)_a$ codes. Other bounds for linear and nonlinear LRCs can be found in [1], [13], [15], [16], [21], [24]. An LRC is *optimal* if it has the highest minimum Hamming distance of any code of the given parameters n , k , r , and δ . In this paper, we focus on Singleton-type bounds (like (1) and (2) above) and their corresponding optimal codes.

There are different constructions of LRCs that are optimal in the sense that they achieve the Singleton-type bound in (2), *e.g.*, [2], [14], [17], [19], [20], [22]. Tamo *et al.* [22] showed that the r -locality of a linear LRC is a matroid invariant, which was used to prove that the minimum Hamming distance of a class of linear LRCs achieves the Singleton-type bound. In [20], Tamo and Barg introduced an interesting construction that can generate optimal linear codes with $(r, \delta)_a$ -locality over an alphabet of size $O(n)$. Under the assumption of a sufficiently large alphabet, Song *et al.* [19] investigated for which parameters (n, k, r, δ) there exists a linear LRC with all-symbol locality and minimum Hamming distance d achieving the Singleton-type bound (2). The parameter set (n, k, r, δ) was divided into eight different cases. In four of these cases it was proved that there are linear LRCs achieving the bound, in two of these cases it was proved that there are no linear LRCs achieving the bound, and the existence of linear LRCs achieving the bound in the remaining two cases remained an open problem. Independently of [19], Wang and Zhang [24] used a linear-programming approach to strengthen these result when $\delta = 2$. Ernvall *et al.* [6] presented methods to modify already existing codes, and gave constructions for three infinite classes of optimal vector-linear LRCs with all-symbol locality over an alphabet of small size. Recently, Westerbäck *et al.* [25] provided a link between matroid theory and LRCs that are either linear or more generally almost affine, and derived new existence results for linear LRCs and nonexistence results for almost affine LRCs, which strengthened the results for linear LRCs given in [19].

Thus, in general, the bound in (2) is not tight for LRCs with $(r, \delta)_a$ -locality, even under the assumption of having a sufficiently large finite field. In this paper, we further study the Hamming distance of LRCs with $(r, \delta)_a$ -locality. Our approach involves a fine detailed analysis of the structure of the repair sets. We define two properties of subsets of repair sets, and show that using them, we can find a subset of repair sets with desired parameters. Using this subset we derive an improved bound on the minimum Hamming distance that improves upon (2). As a consequence, the improved bound shows that some previously undecided cases are in fact unachievable for the bound in (2). The improved bound can also prove some LRCs based on matroids in [25] are indeed optimal. We also give two new explicit constructions to generate optimal LRCs with respect to the improved bound. In Fig. 1, we extend and refine the summary appearing in [19], and show the known and new results concerning the tightness of the Singleton-type bound for LRCs under the assumption that the alphabet is sufficiently large.

The paper is organized as follows. In Section II, we introduce some definitions and facts concerning LRCs with $(r, \delta)_a$ -locality. Section III mainly discusses the structure and properties of a collection of repair sets for locally repairable

codes with all-symbol locality. In Section IV, we prove an upper bound on the minimum Hamming distance, by applying the results obtained in Section III. In Section V, we discuss the implications of our new upper bound. In Section VI, constructions of locally repairable codes are given, which can generate optimal codes with respect to our new bound. Section VII concludes the paper with a discussion of the results and some open questions.

II. PRELIMINARIES

Let \mathcal{C} be an $[n, k, d]_q$ linear code over the finite field \mathbb{F}_q . Assume \mathcal{C} has a generator matrix $G = (\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n)$, where $\mathbf{g}_i \in \mathbb{F}_q^k$ is a column vector for $i = 1, 2, \dots, n$. While many different generator matrices exist for \mathcal{C} , in what follows, the choice of G is immaterial. Given \mathcal{C} and the matrix G , we introduce some notation and concepts.

For an integer $n \in \mathbb{N}$ we denote $[n] = \{1, 2, \dots, n\}$. For any set $N \subseteq [n]$, we denote $\mathcal{G}_N = \{\mathbf{g}_i : i \in N\}$. Then $\text{span}(N)$ denotes the linear space spanned by \mathcal{G}_N over \mathbb{F}_q , and $\text{rank}(N)$ denotes the dimension of $\text{span}(N)$. Additionally, \mathcal{C}_N denotes the *punctured code* of \mathcal{C} associated with the coordinate set N . That is, \mathcal{C}_N is obtained from \mathcal{C} by deleting all symbols in the coordinates $[n] \setminus N$.

The following lemma describes a useful fact about $[n, k, d]_q$ linear codes, which plays an important role in our paper.

Lemma 1 ([10]): The minimum Hamming distance of any $[n, k, d]_q$ linear codes satisfies

$$d = n - \max\{|N| : N \subseteq [n], \text{rank}(N) < k\}.$$

We now recall the definition of repair sets, and locally repairable codes.

Definition 1 ([14]): Let \mathcal{C} be an $[n, k, d]_q$ code. For $1 \leq r \leq k$ and $\delta \geq 2$, an (r, δ) -*repair set* of \mathcal{C} is a subset $S \subseteq [n]$ such that

- 1) $|S| \leq r + \delta - 1$;
- 2) For every $l \in S$, $L \subseteq S \setminus \{l\}$ and $|L| = |S| - (\delta - 1)$, c_l is a linear function of $\{c_i : i \in L\}$, where $\mathbf{c} = (c_1, \dots, c_n) \in \mathcal{C}$.

We say that \mathcal{C} is a *locally repairable code (LRC) with all-symbol (r, δ) -locality* (or \mathcal{C} is an LRC with $(r, \delta)_a$ -locality) if all the n symbols of the code are contained in at least one (r, δ) -repair set.

Remark 1 ([19], [25]): Note that the symbols in an (r, δ) -repair set S can be used to recover up to $\delta - 1$ erasures in the same repair set, then each of the following statements are equivalent to Definition 1, item 2):

- 1) For any $L \subseteq S$ with $|L| = |S| - (\delta - 1)$, we have $\text{rank}(L) = \text{rank}(S)$;
- 2) For any $l \in S$, $L \subseteq S \setminus \{l\}$ and $|L| = |S| - (\delta - 1)$, we have $|\mathcal{C}_{L \cup \{l\}}| = |\mathcal{C}_L|$;
- 3) For any $L \subseteq S$ with $|L| \geq |S| - (\delta - 1)$, we have $|\mathcal{C}_L| = |\mathcal{C}_S|$;
- 4) $d(\mathcal{C}_S) \geq \delta$, where $d(\mathcal{C}_S)$ is the minimum Hamming distance of \mathcal{C}_S .

In what follows, whenever we speak of an LRC with $(r, \delta)_a$ -locality, we will by default assume it is an $[n, k, d]_q$ linear code

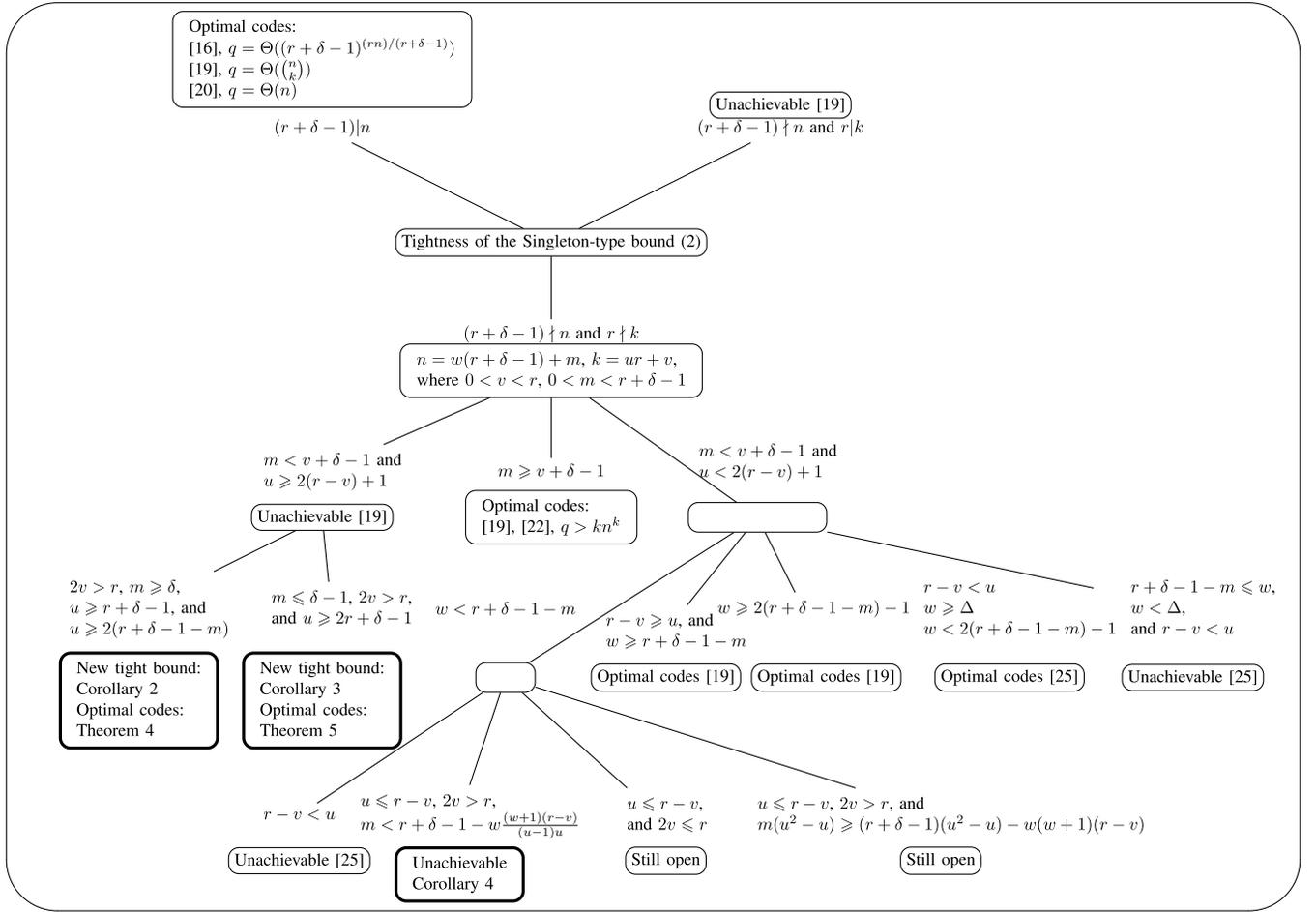


Fig. 1. The tightness of the Singleton-type bound for LRC in (2), where $n = w(r + \delta - 1) + m$, $0 \leq m < r + \delta - 1$, $k = ur + v$, and $0 < v \leq r$, $\Delta = u - 1 + (v - m + \delta - 1)(1 + \frac{1}{\frac{r-v}{u+v-r}})$. The new contributions of this paper appear in bold frames. We do not consider the case $u = 0$, i.e., $k = r$, since this is exactly the case of the classic Singleton bound.

(i.e., its length is n , its dimension is k , its minimum Hamming distance is d , and its alphabet size is q).

III. PROPERTIES OF LRCs WITH $(r, \delta)_a$ -LOCALITY

The goal of this section is to study the structure of (r, δ) -repair sets induced by $(r, \delta)_a$ -locality, and propose some properties which can be used to obtain a lower bound on the minimum Hamming distance in the next section. Generally speaking, we would like to find a set that contains as many code coordinates as possible, under the condition that its rank does not exceed $k - 1$. To this end, we distinguish between three cases. The relationship between repair sets, the number of code symbols, and their rank, is easy to determine for the first case (refer to Proposition 2). The remaining two cases are reduced to the first case in Propositions 3-5.

Throughout the paper we assume that \mathcal{C} denotes an $[n, k, d]_q$ LRC with $(r, \delta)_a$ -locality. The parameters n and k are written in the following forms:

$$\begin{aligned} n &= w(r + \delta - 1) + m, \quad 0 \leq m < r + \delta - 1, \\ k &= ur + v, \quad 0 < v \leq r, \end{aligned} \quad (3)$$

where w, m, u, v are nonnegative integers. Observe that we represent k as $ur + v$ with $0 < v \leq r$ to make sure that $ur < k$.

Remark 2: For the parameters of LRC with $(r, \delta)_a$ -locality, we have the following simple observations:

- 1) If $u = 0$, then the fact that $k \geq r$ implies that $k = r$ and $n \geq r + \delta - 1$, which is a trivial case for LRC.
- 2) The facts that $k = ur + v$ and the code has $(r, \delta)_a$ -locality imply that $w \geq u$, since we need at least $\lceil \frac{k}{r} \rceil = u + 1$ repair sets to cover all the information symbols, i.e., $w(r + \delta - 1) + m = n \geq k + (u + 1)(\delta - 1) = u(r + \delta - 1) + v + \delta - 1$. Note that each repair set contains at least $\delta - 1$ parity check symbols.
- 3) For the nontrivial case $k \geq r$, we have $n \geq r + \delta - 1$, which follows directly from the previous claims.

Definition 2: Let $n, T, s \in \mathbb{N}$. Additionally, let \mathcal{X} be a set of cardinality n , whose elements are called *points*. Finally, let $\mathcal{B} = \{B_1, B_2, \dots, B_T\} \subseteq 2^{\mathcal{X}}$ be a set of *blocks* such that $\bigcup_{i \in [T]} B_i = \mathcal{X}$, and for all $i \in [T]$, $|B_i| \leq s$ and $\bigcup_{j \in T \setminus \{i\}} B_j \neq \mathcal{X}$. We then say $(\mathcal{X}, \mathcal{B})$ is an (n, T, s) -essential covering family (ECF). If all blocks have the same size we say $(\mathcal{X}, \mathcal{B})$ is a *uniform* (n, T, s) -ECF.

For an LRC with $(r, \delta)_a$ -locality, note that each code symbol may be contained in more than one repair set. Thus, to simplify the discussion, we first use the (r, δ) -repair sets to form an ECF, which can be naturally obtained from Definition 1 and Remark 1, as described in [2].

Lemma 2 ([2]): For any $[n, k]_q$ linear code \mathcal{C} with $(r, \delta)_a$ -locality, let $\Gamma \subseteq 2^{[n]}$ be the set of all possible (r, δ) -repair sets. Then we can find a subset $\mathcal{S} \subseteq \Gamma$ such that $([n], \mathcal{S})$ is an $(n, |\mathcal{S}|, r + \delta - 1)$ -ECF with $|\mathcal{S}| \geq \lceil \frac{k}{r} \rceil$.

Remark 3: The fact that the components of \mathcal{S} cover all the element of $[n]$ implies that

$$|\mathcal{S}| \geq \left\lceil \frac{n}{r + \delta - 1} \right\rceil = w + \left\lceil \frac{m}{r + \delta - 1} \right\rceil \geq w.$$

In particular, $|\mathcal{S}| = w$ if and only if $m = 0$, \mathcal{S} is uniform, and the repair sets in \mathcal{S} form a partition of $[n]$.

Let \mathcal{V} be a subset of the set \mathcal{S} that was obtained in Lemma 2. We define the following two properties of \mathcal{V} , which will form the basis of our analysis:

- C1: $|S_i \cap (\bigcup_{S_j \in \mathcal{V} \setminus \{S_i\}} S_j)| < |S_i| - \delta + 1$ for any $S_i \in \mathcal{V}$;
 C2: $|S_i \cap S_j| < \min\{|S_i|, |S_j|\} - \delta + 1$ for any distinct $S_i, S_j \in \mathcal{V}$.

Remark 4: Condition C1 is stronger than Condition C2, namely, if \mathcal{V} satisfies C1 it must satisfy C2, but not necessarily vice versa. To see this, assume to the contrary that C2 does not hold. Then there exist distinct $S_i, S_j \in \mathcal{V}$ such that $|S_i \cap S_j| \geq \min\{|S_i|, |S_j|\} - \delta + 1$, and w.l.o.g., assume $|S_i| \leq |S_j|$. Then

$$\left| S_i \cap \left(\bigcup_{S_{j'} \in \mathcal{V} \setminus \{S_i\}} S_{j'} \right) \right| \geq |S_i \cap S_j| \geq |S_i| - \delta + 1,$$

and hence, C1 does not hold.

The following definitions introduce concepts required in several of our claims.

Definition 3: Assume $r, \delta \geq 1$ are fixed. For all integers $a \geq r + \delta - 1$, $b \geq 0$ we define the function $\Phi(a, b)$ as follows:

$$\Phi(a, b) = \begin{cases} \min \left\{ r + \delta - 1 - c, \max \left\{ \left\lfloor \frac{b}{2} \right\rfloor, \left\lceil \frac{b(b-1)(r+\delta-1-c)}{(\ell+1)\ell} \right\rceil \right\} \right\} \\ \text{if } c \neq 0, \\ 0 \text{ if } c = 0, \end{cases}$$

where c denotes the minimum nonnegative integer with $c \equiv a \pmod{r + \delta - 1}$, and $\ell = \lfloor \frac{a}{r + \delta - 1} \rfloor$.

Definition 4: Let \mathcal{S} denote the ECF induced by an LRC with $(r, \delta)_a$ -locality via Lemma 2, and let $\mathcal{V} \subseteq \mathcal{S}$ be some subset of it. We define

$$\Upsilon(\mathcal{V}, \mathcal{S}) = \left(\bigcup_{S_i \in \mathcal{V}} S_i \right) \setminus \left(\bigcup_{S_j \in \mathcal{S} \setminus \mathcal{V}} S_j \right)$$

and denote

$$M(\mathcal{V}, \mathcal{S}) = |\Upsilon(\mathcal{V}, \mathcal{S})|.$$

We now present a sequence of results on the structure of \mathcal{S} , depending at times on which of Conditions C1 and C2

it satisfies. The proofs are technical and tedious, and are therefore all deferred to the appendix to facilitate the reading.

Proposition 1: For any integer $0 \leq t \leq |\mathcal{S}|$, there exists a t -subset \mathcal{V} of \mathcal{S} such that

$$|\mathcal{V}|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{V}} S_i \right| \geq \Phi(n, t).$$

Proposition 2 ([2, Lemma 7]): Let \mathcal{V} be a subset of \mathcal{S} such that \mathcal{V} satisfies Condition C1. Then

$$\text{rank} \left(\bigcup_{S_i \in \mathcal{V}} S_i \right) \leq \left| \bigcup_{S_i \in \mathcal{V}} S_i \right| - |\mathcal{V}|(\delta - 1).$$

Proposition 3: Let \mathcal{V} be a subset of \mathcal{S} such that \mathcal{V} satisfies Condition C2, but not Condition C1. Then there exists a subset $\mathcal{V}^* \subseteq \mathcal{V}$, such that

- 1) \mathcal{V}^* satisfies Condition C1;
- 2) $|\mathcal{V}^*|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{V}^*} S_i \right| \geq \lceil r/2 \rceil$.

Proposition 4: Let \mathcal{V} be a subset of \mathcal{S} such that \mathcal{V} does not satisfy Condition C2. Then there exists a pair of subsets $\mathcal{V}_1^* \subseteq \mathcal{V}_1 \subseteq \mathcal{S}$ such that:

- 1) $\mathcal{V}_1 \setminus \mathcal{V}_1^*$ satisfies Condition C1;
- 2) For any $S_j \in \mathcal{V}_1 \setminus \mathcal{V}_1^*$, there exists $S_i \in \mathcal{V}_1^*$, such that $\text{span}(S_i) \subseteq \text{span}(S_j)$;
- 3) $\mathcal{S} \setminus \mathcal{V}_1^*$ satisfies Condition C2.

Proposition 5: Assume the same setting as in Proposition 4, and let $\mathcal{V}_1^* \subseteq \mathcal{V}_1 \subseteq \mathcal{S}$ be the subsets guaranteed there. Denote $\Upsilon = \Upsilon(\mathcal{V}_1^*, \mathcal{S})$ and $M = M(\mathcal{V}_1^*, \mathcal{S})$. Then

- 1) $\mathcal{G}_\Upsilon \subseteq \text{span}(\bigcup_{S_i \in \mathcal{V}_1 \setminus \mathcal{V}_1^*} S_i)$;
- 2) $|\mathcal{G}_\Upsilon \cap \text{span}(\bigcup_{S_i \in \mathcal{U}} S_i)| \geq |\mathcal{U}|$, for any subset $\mathcal{U} \subseteq \mathcal{V}_1 \setminus \mathcal{V}_1^*$;
- 3) $|\mathcal{V}_1^*| \leq M$, $|\mathcal{V}_1 \setminus \mathcal{V}_1^*| \leq M$, and $|\mathcal{V}_1| \leq 2M$.

IV. AN IMPROVED BOUND

Having laid the foundation in the previous section, we now use the structure of the repair sets, together with Lemma 1, to obtain a lower bound on the minimum Hamming distance of an LRC with $(r, \delta)_a$ -locality. Thus, we aim to find a subset $\mathcal{S} \subseteq [n]$ with $\text{rank}(\mathcal{S}) = k - 1$, whose size is as large as possible. Particularly, in Lemma 3 below, we find such a set of code symbols under Condition C1. When Condition C1 is not satisfied but Condition C2 is, we apply Proposition 3 to reduce it into the case of C1. For the last case, when even Condition C2 is not satisfied, we consider a shortened code to reach the case of C1 by applying Proposition 4. Thus, all the possible cases are reduced to a subset that satisfies Condition C1 in Proposition 6 below. We then describe our main bound in Theorem 1.

Throughout this section, we still assume that \mathcal{C} is an $[n, k, d]_q$ linear code with $(r, \delta)_a$ -locality, and \mathcal{S} is the ECF given by Lemma 2. The parameters n and k are written as in (3).

Lemma 3: If there exists a subset $\mathcal{V}_1 \subseteq \mathcal{S}$ satisfying Condition C1, $|\mathcal{V}_1| \leq u$, and

$$|\mathcal{V}_1|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{V}_1} S_i \right| \geq \Delta \geq 0,$$

then we can obtain a subset $S \subseteq [n]$ with $\bigcup_{S_i \in \mathcal{V}_1} S_i \subseteq S$, $\text{rank}(S) = k - 1$, and

$$|S| \geq k - 1 + \left(\left\lceil \frac{k + \Delta}{r} \right\rceil - 1 \right) (\delta - 1).$$

Proof: The main idea of the proof is to extend \mathcal{V}_1 to a subset $\mathcal{V}_3 \subseteq \mathcal{S}$ such that $\bigcup_{S_i \in \mathcal{V}_3} S_i$ has rank less than k , and size as large as possible. Note that $k = ur + v$ with $0 < v \leq r$ means that $|S| \geq \lceil \frac{k}{r} \rceil > u$. We first claim that, for any $\mathcal{V} \in \mathcal{S}$, if

$$\text{rank} \left(\bigcup_{S_i \in \mathcal{V}} S_i \right) \leq \left| \bigcup_{S_i \in \mathcal{V}} S_i \right| - |\mathcal{V}|(\delta - 1) \quad (4)$$

and $\text{rank}(\bigcup_{S_i \in \mathcal{V}} S_i) < k$, then we can find an $S_\tau \in \mathcal{S} \setminus \mathcal{V}$ satisfying

$$\text{rank} \left(\bigcup_{S_i \in \mathcal{V} \cup \{S_\tau\}} S_i \right) \leq \left| \bigcup_{S_i \in \mathcal{V} \cup \{S_\tau\}} S_i \right| - |\mathcal{V} \cup \{S_\tau\}|(\delta - 1).$$

Since $\text{rank}(\bigcup_{S_i \in \mathcal{S}} S_i) = k$, the fact that $\text{rank}(\bigcup_{S_i \in \mathcal{V}} S_i) < k$ implies that there is an $S_\tau \in \mathcal{S} \setminus \mathcal{V}$, with

$$\text{rank} \left(\bigcup_{S_i \in \mathcal{V}} S_i \right) < \text{rank} \left(\bigcup_{S_i \in \mathcal{V} \cup \{S_\tau\}} S_i \right),$$

which means that

$$\left| S_\tau \cap \left(\bigcup_{S_i \in \mathcal{V}} S_i \right) \right| < |S_\tau| - \delta + 1.$$

Thus, we can delete $\delta - 1$ elements from $S_\tau \setminus (\bigcup_{S_i \in \mathcal{V}} S_i)$ and keep the rank, i.e.,

$$\begin{aligned} & \text{rank} \left(\bigcup_{S_i \in \mathcal{V} \cup \{S_\tau\}} S_i \right) - \text{rank} \left(\bigcup_{S_i \in \mathcal{V}} S_i \right) \\ & \leq \left| \bigcup_{S_i \in \mathcal{V} \cup \{S_\tau\}} S_i \right| - \left| \bigcup_{S_i \in \mathcal{V}} S_i \right| - \delta + 1. \end{aligned}$$

The above inequality and (4) imply that

$$\begin{aligned} & \text{rank} \left(\bigcup_{S_i \in \mathcal{V}_1 \cup \{S_\tau\}} S_i \right) \\ & \leq \left| \bigcup_{S_i \in \mathcal{V}_1 \cup \{S_\tau\}} S_i \right| - |\mathcal{V}_1 \cup \{S_\tau\}|(\delta - 1), \end{aligned}$$

which proves the claim.

If $|\mathcal{V}_1| = u$ we set $\mathcal{V}_2 = \mathcal{V}_1$. Otherwise, by applying Proposition 2 to \mathcal{V}_1 ,

$$\text{rank} \left(\bigcup_{S_i \in \mathcal{V}_1} S_i \right) \leq \left| \bigcup_{S_i \in \mathcal{V}_1} S_i \right| - |\mathcal{V}_1|(\delta - 1).$$

Note that $\text{rank}(\bigcup_{S_i \in \mathcal{V}_1} S_i) \leq |\mathcal{V}_1|r \leq r(u - 1) < k - r$. Since the union of any u repair sets from \mathcal{S} has rank at most $ur < k$, by applying the preceding claim sufficiently many

times, we can find a subset $\mathcal{V}_2 \subseteq \mathcal{S}$ with $|\mathcal{V}_2| = u$, $\mathcal{V}_1 \subseteq \mathcal{V}_2$, and

$$\begin{aligned} & \text{rank} \left(\bigcup_{S_i \in \mathcal{V}_2} S_i \right) \\ & \leq \left| \bigcup_{S_i \in \mathcal{V}_2} S_i \right| - |\mathcal{V}_2|(\delta - 1) \\ & \leq (u - |\mathcal{V}_1|)(r + \delta - 1) + \left| \bigcup_{S_i \in \mathcal{V}_1} S_i \right| - u(\delta - 1) \\ & = ru + \left| \bigcup_{S_i \in \mathcal{V}_1} S_i \right| - |\mathcal{V}_1|(r + \delta - 1) \\ & \leq k - v - \Delta. \end{aligned} \quad (5)$$

Note that this holds even if in the case $\mathcal{V}_2 = \mathcal{V}_1$ when $|\mathcal{V}_1| = u$.

Having obtained \mathcal{V}_2 , we again apply the procedure on \mathcal{V}_2 to find a subset $\mathcal{V}_3 \subseteq \mathcal{S}$ with $\mathcal{V}_2 \subseteq \mathcal{V}_3$, $|\mathcal{V}_3| = \lceil \frac{k + \Delta}{r} \rceil - 1$, and

$$\text{rank} \left(\bigcup_{S_i \in \mathcal{V}_3} S_i \right) \leq \left| \bigcup_{S_i \in \mathcal{V}_3} S_i \right| - |\mathcal{V}_3|(\delta - 1). \quad (6)$$

By (5), we also have

$$\begin{aligned} & \text{rank} \left(\bigcup_{S_i \in \mathcal{V}_3} S_i \right) \\ & \leq \text{rank} \left(\bigcup_{S_i \in \mathcal{V}_2} S_i \right) + (|\mathcal{V}_3| - |\mathcal{V}_2|)r \\ & = \text{rank} \left(\bigcup_{S_i \in \mathcal{V}_2} S_i \right) + \left(\left\lceil \frac{k + \Delta}{r} \right\rceil - 1 - u \right) r \\ & < \text{rank} \left(\bigcup_{S_i \in \mathcal{V}_2} S_i \right) + v + \Delta \\ & \leq k. \end{aligned}$$

Now let S be a subset of $[n]$ with $\text{rank}(S) = k - 1$ and $\bigcup_{S_i \in \mathcal{V}_3} S_i \subseteq S$. Then by (6), we have

$$\begin{aligned} |S| & \geq \text{rank}(S) - \text{rank} \left(\bigcup_{S_i \in \mathcal{V}_3} S_i \right) + \left| \bigcup_{S_i \in \mathcal{V}_3} S_i \right| \\ & \geq k - 1 + |\mathcal{V}_3| \cdot (\delta - 1) \\ & = k - 1 + \left(\left\lceil \frac{k + \Delta}{r} \right\rceil - 1 \right) (\delta - 1). \end{aligned}$$

Proposition 6: If the requirements of Proposition 4 hold, let $\mathcal{V}_1^* \subseteq \mathcal{V}_1 \subseteq \mathcal{S}$ be the two guaranteed sets, and otherwise set $\mathcal{V}_1 = \mathcal{V}_1^* = \emptyset$. Denote $M = M(\mathcal{V}_1^*, \mathcal{S})$. Then there exists a subset $S \subseteq [n]$ with $\text{rank}(S) = k - 1$, and

$$|S| \geq k - 1 + \begin{cases} \min \left\{ \left(\left\lceil \frac{k + \lceil \frac{r}{2} \rceil}{r} \right\rceil - 1 \right) (\delta - 1), \right. \\ \left. M + \left(\left\lceil \frac{k + \Phi(n - M, u - M)}{r} \right\rceil - 1 \right) (\delta - 1) \right\}, & \text{if } u > M, \\ u + \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1), & \text{if } u \leq M, \end{cases}$$

where $\Phi(\cdot, \cdot)$ is from Definition 3. ■

Proof: Before proceeding with the proof, if $\mathcal{V}_1 = \mathcal{V}_1^* = \emptyset$, the claims in this proof also hold (mostly trivially so). Thus, we concentrate on the case they are not empty.

Define $\Upsilon = \Upsilon(\mathcal{V}_1^*, \mathcal{S})$, $N = [n] \setminus \Upsilon$, and $\mathcal{S}^* = \mathcal{S} \setminus \mathcal{V}_1^*$. Then \mathcal{C}_N is an $[n - M, k]_q$ linear code with $(r, \delta)_a$ -locality, and \mathcal{S}^* is an ECF whose elements are (r, δ) -repair sets of \mathcal{C}_N , where additionally, $|\mathcal{C}_N| = |\mathcal{C}|$ by virtue of Proposition 5-1). To avoid a conflict with the definition of $\Phi(\cdot, \cdot)$, we highlight that $n - M \geq r + \delta - 1$, since $k \geq r$ and \mathcal{C}_N has $(r, \delta)_a$ -locality (refer to Remark 2, item 3). The remainder of the proof is divided into two cases.

Case 1: Assume $u > M$. The fact that $\text{rank}(\bigcup_{S_i \in \mathcal{S}^*} S_i) = k$ implies that $|\mathcal{S}^*| \geq \lceil k/r \rceil > u \geq u - M$. Thus, by Proposition 1, there is a $(u - M)$ -subset $\mathcal{V}_2 \subseteq \mathcal{S}^*$ with

$$|\mathcal{V}_2|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{V}_2} S_i \right| \geq \Phi(n - M, u - M).$$

Recall that by Proposition 5-3), we have $|\mathcal{V}_1 \setminus \mathcal{V}_1^*| \leq M$. Define $\mathcal{V}_3 = \mathcal{V}_2 \cup (\mathcal{V}_1 \setminus \mathcal{V}_1^*)$, then $|\mathcal{V}_3| \leq u$ and

$$\begin{aligned} & |\mathcal{V}_3|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{V}_3} S_i \right| \\ & \geq |\mathcal{V}_2|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{V}_2} S_i \right| \\ & \geq \Phi(n - M, u - M). \end{aligned}$$

If \mathcal{V}_3 satisfies Condition C1, then by Lemma 3, there is a subset $S^{(1)} \subseteq N$ with $\bigcup_{S_i \in \mathcal{V}_3} S_i \subseteq S^{(1)}$, $\text{rank}(S^{(1)}) = k - 1$, and

$$|S^{(1)}| \geq k - 1 + \left(\left\lceil \frac{k + \Phi(n - M, u - M)}{r} \right\rceil - 1 \right) (\delta - 1).$$

Note that

$$\mathcal{G}_\Upsilon \subseteq \text{span} \left(\bigcup_{S_i \in \mathcal{V}_1 \setminus \mathcal{V}_1^*} S_i \right) \subseteq \text{span} \left(\bigcup_{S_i \in \mathcal{V}_3} S_i \right) \subseteq \text{span}(S^{(1)})$$

by Proposition 5-1), and $\Upsilon \cap S^{(1)} \subseteq \Upsilon \cap N = \emptyset$. Define $S = S^{(1)} \cup \Upsilon$, then S is the desirable subset of $[n]$ with $\text{rank}(S) = k - 1$, and

$$\begin{aligned} |S| &= M + |S^{(1)}| \\ &\geq M + k - 1 + \left(\left\lceil \frac{k + \Phi(n - M, u - M)}{r} \right\rceil - 1 \right) (\delta - 1). \end{aligned} \quad (7)$$

Let us now consider the case where \mathcal{V}_3 does not satisfy Condition C1. By Proposition 4-3), $\mathcal{S} \setminus \mathcal{V}_1^*$ satisfies Condition C2. Since $\mathcal{V}_3 \subseteq \mathcal{S} \setminus \mathcal{V}_1^*$, we also have that \mathcal{V}_3 satisfies Condition C2. By Proposition 3, there exists a subset $\mathcal{V}_3^* \subseteq \mathcal{V}_3$ that satisfies Condition C1 and

$$|\mathcal{V}_3^*|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{V}_3^*} S_i \right| \geq \left\lceil \frac{r}{2} \right\rceil.$$

Now, by Lemma 3, there is a subset $S \subseteq [n]$ with $\text{rank}(S) = k - 1$, and

$$|S| \geq k - 1 + \left(\left\lceil \frac{k + \lceil \frac{r}{2} \rceil}{r} \right\rceil - 1 \right) (\delta - 1). \quad (8)$$

Case 2: Assume $u \leq M$. Define \mathcal{V}_4 to be a u -subset of $\mathcal{V}_1 \setminus \mathcal{V}_1^*$ if $|\mathcal{V}_1 \setminus \mathcal{V}_1^*| \geq u$. Otherwise define $\mathcal{V}_4 = \mathcal{V}_1 \setminus \mathcal{V}_1^*$. The set \mathcal{V}_4 satisfies Condition C1 according to Proposition 4-1), and obviously

$$|\mathcal{V}_4|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{V}_4} S_i \right| \geq 0.$$

By Lemma 3, there is a subset $S^{(2)} \subseteq N$ with $\bigcup_{S_i \in \mathcal{V}_4} S_i \subseteq S^{(2)}$, $\text{rank}(S^{(2)}) = k - 1$, and

$$|S^{(2)}| \geq k - 1 + \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1).$$

Note that $|\mathcal{G}_\Upsilon \cap \text{span}(\bigcup_{S_i \in \mathcal{V}_4} S_i)| \geq u$ by Proposition 5-2) and the facts that $|\mathcal{V}_4| = u$ or $\mathcal{V}_4 = \mathcal{V}_1 \setminus \mathcal{V}_1^*$, $|\Upsilon| = M \geq u$. Define $S = S^{(2)} \cup \Upsilon'$, where $\Upsilon' = \{i : \mathbf{g}_i \in \mathcal{G}_\Upsilon \cap \text{span}(\bigcup_{S_i \in \mathcal{V}_4} S_i)\}$. Recall that $\Upsilon' \cap S^{(2)} \subseteq \Upsilon \cap S^{(2)} \subseteq \Upsilon \cap N = \emptyset$. Thus, S is the desirable subset of $[n]$ with $\text{rank}(S) = k - 1$, and

$$|S| \geq u + |S^{(2)}| \geq u + k - 1 + \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1). \quad (9)$$

The proof is now completed by combining (7), (8), and (9). \blacksquare

Now we are ready to obtain an upper bound on the minimum Hamming distance.

Theorem 1: Let \mathcal{C} be an LRC with $(r, \delta)_a$ -locality, and let \mathcal{S} be the ECF given by Lemma 2. If the requirements of Proposition 4 hold, let $\mathcal{V}_1^* \subseteq \mathcal{V}_1 \subseteq \mathcal{S}$ be the two guaranteed sets, and otherwise set $\mathcal{V}_1 = \mathcal{V}_1^* = \emptyset$. Denote $M = M(\mathcal{V}_1^*, \mathcal{S})$. Then

$$d \leq n - k + 1 - \begin{cases} \min \left\{ \left(\left\lceil \frac{k + \lceil \frac{r}{2} \rceil}{r} \right\rceil - 1 \right) (\delta - 1), \right. \\ \left. M + \left(\left\lceil \frac{k + \Phi(n - M, u - M)}{r} \right\rceil - 1 \right) (\delta - 1) \right\}, & \text{if } u > M, \\ (u + (\lceil \frac{k}{r} \rceil - 1) (\delta - 1)), & \text{if } u \leq M, \end{cases}$$

where $\Phi(\cdot, \cdot)$ is from Definition 3.

Proof: The conclusion is obtained directly by combining Lemma 1 and Proposition 6. \blacksquare

Remark 5: We point out that the subsets $\mathcal{V}_1^* \subseteq \mathcal{V}_1 \subseteq \mathcal{S}$, whose existence is guaranteed in Proposition 4, are not necessarily unique. Thus, the value of M used in Theorem 1 is not unique as well. Of the (possibly many) choices for M , it is unclear which one results in the best bound.

Remark 6: We make the following observations:

1) If $M = 0$, the bound in Theorem 1 becomes

$$d \leq n - k + 1 - \left(\left\lceil \frac{k + \min \left\{ \lceil \frac{r}{2} \rceil, \Phi(n, u) \right\}}{r} \right\rceil - 1 \right) (\delta - 1),$$

which is tighter than the one given by (2) (see, [8], [14]) if and only if

$$\min \left\{ \left\lceil \frac{r}{2} \right\rceil, \Phi(n, u) \right\} > r - v.$$

In particular, the bound is exactly the one in (2) when $m = 0$, and it is tighter than the one in (2) when $m \neq 0$ and $v = r$.

- 2) If $M \neq 0$ and $k > r$, the bound in Theorem 1 is tighter than the bound in (2) if and only if

$$\left\lceil \frac{r}{2} \right\rceil > r - v.$$

In particular, the bound is tighter than the one in (2) when $v = r$, i.e., $r \mid k$ and $k > r$.

V. CASE ANALYSIS OF THE IMPROVED BOUND

The new bound of Theorem 1 depends on many parameters. In this section we highlight interesting cases of parameters for this bound. Generally, we should consider all possible M in Theorem 1 to determine the upper bound on d , where M depends on the structure of the (r, δ) -repair sets, i.e., \mathcal{S} . However, for some special cases the expression for the bound can be further simplified. Specifically, in Corollaries 2 and 3 we determine explicit bounds for two classes of parameters, and in Theorem 3 we determine the exact maximum Hamming distance for another class of parameters. Finally, in Corollary 4, we show that the original Singleton-type bound given by (2), is unattainable for a class of parameters.

We again assume that \mathcal{C} is an $[n, k, d]_q$ linear code with $(r, \delta)_a$ -locality, and \mathcal{S} is the ECF given by Lemma 2. The parameters n and k are written as in (3).

Corollary 1: If an $[n, k, d]_q$ LRC with $(r, \delta)_a$ -locality satisfies that the repair sets in \mathcal{S} are pairwise disjoint, then

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

Proof: If the repair sets in \mathcal{S} are pairwise disjoint, then Condition C1 always holds for \mathcal{S} . The conclusion is then obtained directly by Proposition 1, Lemma 3 and Lemma 1. ■

In [25], Westerbäck *et al.* studied locally repairable codes via matroid theory, and obtained the following bound for d_{\max} , where d_{\max} is the largest d such that there exists a linear $[n, k, d]_q$ code with $(r, \delta)_a$ -locality.

Theorem 2 ([25, Theorem 36-(ii)]): Assume $r + \delta - 1 \nmid n$ and $r \nmid k$, namely, $m > 0$ and $v < r$. If $0 < r < k \leq n - \left\lfloor \frac{k}{r} \right\rfloor (\delta - 1)$ and $v > m - \delta + 1$, then

$$d_{\max} \geq n - k + 1 - \left\lfloor \frac{k}{r} \right\rfloor (\delta - 1) + \begin{cases} 0, & \text{if } m \geq \delta, \\ \delta - 1 - m, & \text{if } m \leq \delta - 1, \end{cases}$$

where d_{\max} is the largest d such that there exists a linear $[n, k, d]_q$ code with $(r, \delta)_a$ -locality.

By applying the bound obtained in Lemma 1, we may now determine the exact value of d_{\max} for certain classes of parameters.

Corollary 2: Under the setting of Theorem 2, if $m \geq \delta$, $r > v > \max\{m - \delta + 1, \lfloor \frac{r}{2} \rfloor\}$, and $u \geq \max\{2(r + \delta - 1 - m), r + \delta - 1\}$, we have

$$d \leq n - k + 1 - \left\lfloor \frac{k}{r} \right\rfloor (\delta - 1).$$

Proof: By $u \geq 2(r + \delta - 1 - m)$, we have $\lfloor \frac{u}{2} \rfloor \geq r + \delta - 1 - m$, which implies that $\Phi(n, u) = r + \delta - 1 - m$. By $v > \max\{m - \delta + 1, \lfloor \frac{r}{2} \rfloor\}$, we have $r - v < \min\{\Phi(n, u), \lfloor \frac{r}{2} \rfloor\}$.

Obviously $\lfloor \frac{r}{2} \rfloor \leq r$, and since $m \geq \delta$, also $\Phi(n, u) = r + \delta - 1 - m \leq r$. This implies that

$$\left\lceil \frac{k + \Phi(n, u)}{r} \right\rceil = \left\lceil \frac{k + \lfloor \frac{r}{2} \rfloor}{r} \right\rceil = \left\lfloor \frac{k}{r} \right\rfloor + 1. \quad (10)$$

The remainder of the proof is divided into three cases.

Case 1: Assume $u \leq M$. We note that $u > \delta - 1$, and then by Theorem 1, we have

$$\begin{aligned} d &\leq n - k + 1 - u - \left(\left\lfloor \frac{k}{r} \right\rfloor - 1 \right) (\delta - 1) \\ &< n - k + 1 - \left\lfloor \frac{k}{r} \right\rfloor (\delta - 1). \end{aligned}$$

Case 2: Assume $u > M$ and $M \geq \delta - 1$. Since

$$M + \left(\left\lceil \frac{k + \Phi(n - M, u - M)}{r} \right\rceil - 1 \right) (\delta - 1) \geq \left\lfloor \frac{k}{r} \right\rfloor (\delta - 1),$$

by Theorem 1 and (10), we have

$$\begin{aligned} d &\leq n - k + 1 - \left(\left\lceil \frac{k + \lfloor \frac{r}{2} \rfloor}{r} \right\rceil - 1 \right) (\delta - 1) \\ &= n - k + 1 - \left\lfloor \frac{k}{r} \right\rfloor (\delta - 1). \end{aligned}$$

Case 3: Assume $u > M$ and $M < \delta - 1$. Obviously $M < m$ since $\delta \leq m$. Additionally,

$$n - M = w(r + \delta - 1) + (m - M),$$

where $0 < m - M < r + \delta - 1$, thus $m - M = (n - M) \bmod (r + \delta - 1)$. Then

$$\begin{aligned} \Phi(n - M, u - M) &= \\ \min \left\{ r + \delta - 1 - m + M, \right. \\ \left. \max \left\{ \left\lfloor \frac{u - M}{2} \right\rfloor, \left\lceil \frac{(u - M)(u - M - 1)(r + \delta - 1 - m + M)}{w(w + 1)} \right\rceil \right\} \right\}. \end{aligned}$$

The facts that

$$r + \delta - 1 - m + M \geq r + \delta - 1 - m = \Phi(n, u)$$

and

$$\left\lfloor \frac{u - M}{2} \right\rfloor \geq \left\lfloor \frac{(r + \delta - 1) - (\delta - 2)}{2} \right\rfloor = \left\lfloor \frac{r + 1}{2} \right\rfloor = \left\lfloor \frac{r}{2} \right\rfloor$$

imply that

$$\Phi(n - M, u - M) \geq \min \left\{ \Phi(n, u), \left\lfloor \frac{r}{2} \right\rfloor \right\}.$$

Thus, by Theorem 1, (10) and the above discussion, we have

$$d \leq n - k + 1 - \left\lfloor \frac{k}{r} \right\rfloor (\delta - 1).$$

Combining the above three cases, the proof is now completed. ■

Corollary 3: Under the setting of Theorem 2, if $m \leq \delta - 1$, $r > v > \lfloor \frac{r}{2} \rfloor$, and $u \geq 2r + \delta - 1$, we have

$$d \leq n - k + 1 - \left\lfloor \frac{k}{r} \right\rfloor (\delta - 1) + (\delta - 1 - m).$$

Proof: We again use the upper bound obtained in Theorem 1. By the definition of $\Phi(\cdot, \cdot)$, and since $m > 0$, we have $\Phi(n, u) \geq \min \left\{ r + \delta - 1 - m, \left\lfloor \frac{u}{2} \right\rfloor \right\}$. It follows that

$$\left\lceil \frac{k + \Phi(n, u)}{r} \right\rceil \geq \left\lceil \frac{k}{r} \right\rceil + 1 = \left\lceil \frac{k + \left\lfloor \frac{r}{2} \right\rfloor}{r} \right\rceil, \quad (11)$$

where the first inequality holds by the fact that $\delta - 1 \geq m > 0$ and $\left\lfloor \frac{u}{2} \right\rfloor \geq \left\lfloor \frac{2r + \delta - 1}{2} \right\rfloor \geq r$, and the second equality follows from $r > v > \left\lfloor \frac{r}{2} \right\rfloor$. The rest of the proof is divided into three cases.

Case 1: Assume $m > M$. Obviously, we have $u > \delta - 1 \geq m > M \geq 0$. Since $u - M > 2r$, we get $\left\lfloor \frac{u - M}{2} \right\rfloor \geq r$, and we note that $r + \delta - 1 - m + M \geq r$. It follows that $0 < m - M < r + \delta - 1$, and so $m - M = (n - M) \bmod (r + \delta - 1)$, and so $\Phi(n - M, u - M) \geq r$. Thus,

$$M + \left(\left\lceil \frac{k + \Phi(n - M, u - M)}{r} \right\rceil - 1 \right) (\delta - 1) \geq \left\lceil \frac{k}{r} \right\rceil (\delta - 1),$$

and by Theorem 1 and $\left\lceil \frac{k + \left\lfloor \frac{r}{2} \right\rfloor}{r} \right\rceil = \left\lceil \frac{k}{r} \right\rceil + 1$ from (11), we have

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

Case 2: Assume $m \leq M$ and $u > M$. We have

$$\begin{aligned} M + \left(\left\lceil \frac{k + \Phi(n - M, u - M)}{r} \right\rceil - 1 \right) (\delta - 1) \\ \geq m + \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1) \end{aligned}$$

and by (11) we have

$$\begin{aligned} \left(\left\lceil \frac{k + \left\lfloor \frac{r}{2} \right\rfloor}{r} \right\rceil - 1 \right) (\delta - 1) &= \left\lceil \frac{k}{r} \right\rceil (\delta - 1) \\ &\geq m + \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1). \end{aligned}$$

Thus, by Theorem 1,

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

Case 3: Assume $m \leq M$ and $u \leq M$. The fact that $u > \delta - 1 \geq m$ implies that

$$u + \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1) > m + \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1).$$

Thus, by Theorem 1, we have

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

Combining the above three cases, the proof is now completed. ■

We can now strengthen Theorem 2 by applying Corollaries 2 and 3.

Theorem 3: Assume $r + \delta - 1 \nmid n$ and $r \nmid k$, namely, $m > 0$ and $v < r$. If $0 < r < k \leq n - \left\lceil \frac{k}{r} \right\rceil (\delta - 1)$ and $v >$

$\max \left\{ m - \delta + 1, \left\lfloor \frac{r}{2} \right\rfloor \right\}$, then

$$\begin{aligned} d_{\max} &= n - k + 1 - \left\lceil \frac{k}{r} \right\rceil (\delta - 1) \\ &+ \begin{cases} 0, & \text{if } m \geq \delta \\ \text{and } u \geq \max \{2(r + \delta - 1 - m), r + \delta - 1\}, \\ \delta - 1 - m, & \text{if } m \leq \delta - 1 \\ \text{and } u \geq 2r + \delta - 1, \end{cases} \end{aligned}$$

where d_{\max} is the largest d such that there exists a linear $[n, k, d]_q$ code with $(r, \delta)_a$ -locality.

Based on the results in [19], [25], the remaining open cases for the tightness of the bound in (2) are summarized in the following:

Open Problem [19]: Do there exist optimal $[n, k, d]_q$ codes with $(r, \delta)_a$ -locality that achieve the minimum Hamming distance bound in (2), under the conditions that $v \neq 0$, $0 < m < v + \delta - 1$, $0 < u \leq r - v$, and $w < r + \delta - 1 - m$? (using the notation of (3))

We can answer this open question in part.

Corollary 4: No $[n, k, d]_q$ code with $(r, \delta)_a$ -locality achieves the bound in (2) under the conditions of $0 < m < v + \delta - 1$, and $u > 1$, if

$$\min \left\{ \left\lfloor \frac{r}{2} \right\rfloor, \frac{u(u-1)(r+\delta-1-m)}{(w+1)w} \right\} > r - v.$$

In particular, when $v > \frac{r}{2}$, $u > 1$, and $0 < m < r + \delta - 1 - w \frac{(w+1)(r-v)}{u(u-1)}$, the bound in (2) is unachievable.

Proof: Since $u > 1$, i.e., $k = ur + v > r$ and $\left\lfloor \frac{r}{2} \right\rfloor > r - v$, if additionally $M > 0$ then by Remark 6, the bound in (2) is unachievable. Assume now that $M = 0$. The fact that $m < v + \delta - 1$ means that $r + \delta - 1 - m > r - v$. Recall that $\frac{u(u-1)(r+\delta-1-m)}{(w+1)w} > r - v$. Thus, $\Phi(n, u) > r - v$ by Definition 3, i.e.,

$$\left\lceil \frac{k + \Phi(n, u)}{r} \right\rceil > \left\lceil \frac{k}{r} \right\rceil + 1 = \left\lceil \frac{k + \left\lfloor \frac{r}{2} \right\rfloor}{r} \right\rceil,$$

which shows that

$$\begin{aligned} d &\leq n - k + 1 - \left\lceil \frac{k}{r} \right\rceil (\delta - 1) \\ &< n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1). \end{aligned}$$

Therefore, the bound in (2) is unachievable in this case.

Note that $w \geq u$ (see Remark 2) means that $r + \delta - 1 - w \frac{(w+1)(r-v)}{u(u-1)} \leq v + \delta - 1$. Thus, combining the above two cases, the corollary follows from $\left\lfloor \frac{r}{2} \right\rfloor > r - v$ and $\frac{u(u-1)(r+\delta-1-m)}{(w+1)w} > r - v$ when $v > \frac{r}{2}$ and $0 < m < r + \delta - 1 - w \frac{(w+1)(r-v)}{u(u-1)}$. ■

Remark 7: By Corollary 4, the remaining open cases can be listed as:

- 1) $0 < v \leq \frac{r}{2}$, $0 < m < v + \delta - 1$, $1 \leq u \leq r - v$, and $w < r + \delta - 1 - m$.
- 2) $v > \frac{r}{2}$, $(r + \delta - 1)u(u - 1) - w(w + 1)(r - v) \leq mu(u - 1)$, $0 < m < v + \delta - 1$, $1 \leq u \leq r - v$, and $w < r + \delta - 1 - m$.

VI. OPTIMAL LRCs ACHIEVING THE IMPROVED BOUND

In this section, we introduce explicit constructions of locally repairable codes, which generate optimal codes with respect to the improved bounds in Corollaries 2 and 3. These constructions are mainly a modification of the construction in [16] by endowing the repair sets with a special structure so that the locally repairable codes can achieve the improved bound in the pervious section.

Let \mathbb{F}_{q_1} be an extension field of the finite field \mathbb{F}_q , and let $S = \{\alpha_1, \alpha_2, \dots, \alpha_n\} \subseteq \mathbb{F}_{q_1}$ be a set of n elements. Let $V(S, h)$ denote the matrix

$$V(S, h) \triangleq \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_n \\ \alpha_1^q & \alpha_2^q & \alpha_3^q & \cdots & \alpha_n^q \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \alpha_1^{q^{h-1}} & \alpha_2^{q^{h-1}} & \alpha_3^{q^{h-1}} & \cdots & \alpha_n^{q^{h-1}} \end{pmatrix}_{h \times n}.$$

We comment that in order for $V(S, h)$ to be well defined, we fix some ordering of the elements of \mathbb{F}_{q_1} , and index the elements of S so that they are in non-descending order. Additionally, since \mathbb{F}_{q_1} is a vector space over \mathbb{F}_q , we use $\text{rank}(S)$ to denote the dimension of the space spanned by linear combinations of elements from S with coefficients from \mathbb{F}_q .

Definition 5 ([7]): The set $S \subseteq \mathbb{F}_{q_1}$ is t -wise independent over a field $\mathbb{F}_q \subseteq \mathbb{F}_{q_1}$ if every $T \subseteq S$, $|T| \leq t$, is linearly independent over \mathbb{F}_q .

The following conclusion is obtained directly from the above definition.

Lemma 4: Let $S \subseteq \mathbb{F}_{q_1}$ be t -wise independent over a field $\mathbb{F}_q \subseteq \mathbb{F}_{q_1}$. Then a subset $S' \subseteq S$ is t' -wise independent over the field $\mathbb{F}_q \subseteq \mathbb{F}_{q_1}$ if $t' \leq t$ and $|S'| \geq t'$.

With the above preparation, we give the following construction of linear codes.

Construction A: Fix $\mathbb{F}_q \subseteq \mathbb{F}_{q_1}$. With the notation of (3), define $h = n - k - (w + 1)(\delta - 1)$. Let $A = (A_1, A_2)_{(\delta-1) \times (r+\delta-1)}$ be a parity-check matrix of an $[r + \delta - 1, r, \delta]_q$ MDS code, where A_1 is a $(\delta - 1) \times (r + \delta - 2)$ matrix and A_2 is a $(\delta - 1) \times 1$ matrix. Let $S \subseteq \mathbb{F}_{q_1}$, $|S| = n$, and $w + 1 \geq r + \delta - 1 - m$. Define $\mathcal{C}(S, h) \subseteq \mathbb{F}_{q_1}^n$ to be a linear code with parity-check matrix

$$R = \begin{pmatrix} R_1 & 0 & 0 & \cdots & 0 \\ 0 & R_2 & 0 & \cdots & 0 \\ 0 & 0 & R_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & R_{w+1} \\ H_1 & H_2 & H_3 & \cdots & H_{w+1} \end{pmatrix}_{(n-k) \times n}, \quad (12)$$

where

$$R_i = A_1 \text{ for } 1 \leq i \leq r + \delta - 1 - m, \quad (13)$$

$$R_j = A \text{ for } r + \delta - m \leq j \leq w + 1, \quad (14)$$

and

$$(H_1, H_2, \dots, H_{w+1}) = V(S, h). \quad (15)$$

We cite the following lemma from [7].

Lemma 5 ([7]): Fix $\mathbb{F}_q \subseteq \mathbb{F}_{q_1}$. Let E_i , $1 \leq i \leq t$, be a parity-check matrix of an $[e_i, e_i + 1 - \delta, \delta]_q$ MDS code. For all $1 \leq i \leq t + 1$, let $S_i \subseteq \mathbb{F}_{q_1}$, $|S_i| = e_i$, and let $H'_i = V(S_i, h)$. If $h \geq \sum_{i=1}^{t+1} e_i - t(\delta - 1)$ and $\text{rank}(\cup_{i=1}^{t+1} S_i) = \sum_{i=1}^{t+1} |S_i| = \sum_{i=1}^{t+1} e_i$, then

$$\text{rank} \begin{pmatrix} E_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & E_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & E_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & E_t & 0 \\ H'_1 & H'_2 & H'_3 & \cdots & H'_t & H'_{t+1} \end{pmatrix} = \sum_{i=1}^{t+1} e_i,$$

i.e., the matrix has full column rank.

We can now prove the properties of Construction A.

Theorem 4: Let $n = w(r + \delta - 1) + m$, $\delta \leq m < r + \delta - 1$, $k = ur + v$, $0 < v < r$, and let $S \subseteq \mathbb{F}_{q_1}$ be $(h + (w - u)(\delta - 1))$ -wise independent over \mathbb{F}_q . Denote by $\mathcal{C}(S, h)$ the code generated by Construction A. If $r > v > \max\{m - \delta + 1, \lfloor \frac{r}{2} \rfloor\}$, and $u \geq \max\{2(r + \delta - 1 - m), r + \delta - 1\}$, then $\mathcal{C}(S, h)$ is an optimal $[n, k, d]_{q_1}$ linear code with $(r, \delta)_a$ -locality and $d = h + (w - u)(\delta - 1) + 1$.

Proof: By Remark 2 we have $w \geq u \geq 2(r + \delta - 1 - m)$, which means the condition $w + 1 \geq r + \delta - 1 - m$ holds in Construction A. By (12)-(15), we have that the code \mathcal{C} is an $[n, k_1]_{q_1}$ code with all symbol (r, δ) -locality and $k_1 \geq k$. Our next goal is to prove that $d \geq h + (w - u)(\delta - 1) + 1$, i.e., that any $h + (w - u)(\delta - 1)$ columns of R have full rank. Let

$$R^* = \begin{pmatrix} R_1^* & 0 & 0 & \cdots & 0 \\ 0 & R_2^* & 0 & \cdots & 0 \\ 0 & 0 & R_3^* & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & R_{w+1}^* \\ H_1^* & H_2^* & H_3^* & \cdots & H_{w+1}^* \end{pmatrix},$$

denote the arbitrary $h + (w - u)(\delta - 1)$ columns chosen from R , where R_i^* and H_i^* denote the chosen part from R_i and H_i for $1 \leq i \leq w + 1$, respectively. If R_i^* contains $\delta - 1$ columns or less, then R_i^* has full rank since its columns are part of a parity-check matrix for a code with distance δ . Let $i_1 < i_2 < \cdots < i_t$ be the indices such that $R_{i_j}^*$, $1 \leq j \leq t$, contains at least δ columns. Thus, R^* has full rank if and only if

$$\bar{R} = \begin{pmatrix} R_{i_1}^* & 0 & 0 & \cdots & 0 \\ 0 & R_{i_2}^* & 0 & \cdots & 0 \\ 0 & 0 & R_{i_3}^* & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & R_{i_t}^* \\ H_{i_1}^* & H_{i_2}^* & H_{i_3}^* & \cdots & H_{i_t}^* \end{pmatrix}$$

has full rank. Let e_{i_j} denote the number of columns of $R_{i_j}^*$ for $1 \leq j \leq t$. Thus, we have $e_{i_j} \geq \delta$ for $1 \leq j \leq t$ and

$$\sum_{j=1}^t e_{i_j} \leq h + (w - u)(\delta - 1). \quad (16)$$

We proceed by examining two cases, depending on the value of t .

Case 1: Assume $1 \leq t \leq w - u$. Since A is the parity-check matrix of an $[r + \delta - 1, r, \delta]_q$ MDS code, we have that any $\delta - 1$ columns of A have full rank. Thus, any $\delta - 1$ columns of $R_{i_j}^*$ for $1 \leq j \leq t$ also have rank $\delta - 1$, by (13) and (14). Hence, $R_{i_j}^*$, $1 \leq j \leq t$, can be viewed as a parity-check matrix of an $[e_{i_j}, e_{i_j} + 1 - \delta, \delta]_q$ MDS code. Recall that

$$\begin{aligned} h &= n - k - (w + 1)(\delta - 1) \\ &= (w - u)r + m - v - \delta + 1 \\ &\stackrel{(a)}{\geq} \begin{cases} (r + \delta - 1)t - t(\delta - 1) \geq \sum_{j=1}^t (e_{i_j} - \delta + 1), \\ \text{if } 1 \leq t \leq w - u - 1, \\ h + (w - u)(\delta - 1) - (w - u)(\delta - 1) \\ \geq \sum_{j=1}^t (e_{i_j} - \delta + 1), \text{ if } t = w - u. \end{cases} \end{aligned}$$

Here, the first case of (a) follows by $t \leq w - u - 1$, $r > v$, and $m \geq \delta$ (i.e., $r + m - v - \delta + 1 > 0$). The second case of (a) follows by (16). Since S is $(h + (w - u)(\delta - 1))$ -wise independent over \mathbb{F}_q and $\sum_{j=1}^t |S_{i_j}| = \sum_{j=1}^t e_{i_j} \leq h + (w - u)(\delta - 1)$, we have that $\bigcup_{j=1}^t S_{i_j}$ is linearly independent over \mathbb{F}_q , where $H_{i_j}^* = V(S_{i_j}, h)$ for $1 \leq j \leq t$. Thus, by Lemma 5, we have $\text{rank}(\overline{R}) = \sum_{j=1}^t e_{i_j}$, i.e., any $h + (w - u)(\delta - 1)$ columns of R have full rank when $1 \leq t \leq w - u$.

Case 2: Assume $t > w - u$.

$$\begin{aligned} &\text{rank}(\overline{R}) \\ &= \text{rank} \begin{pmatrix} R_{i_1}^* & 0 & 0 & \dots & 0 \\ 0 & R_{i_2}^* & 0 & \dots & 0 \\ 0 & 0 & R_{i_3}^* & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & R_{i_t}^* \\ H_{i_1}^* & H_{i_2}^* & H_{i_3}^* & \dots & H_{i_t}^* \end{pmatrix} \\ &\geq \text{rank} \begin{pmatrix} R_{i_1}^* & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & R_{i_2}^* & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_{i_{w-u}}^* & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ H_{i_1}^* & H_{i_2}^* & \dots & H_{i_{w-u}}^* & H_{i_{w-u+1}}^* & \dots & H_{i_t}^* \end{pmatrix}. \end{aligned}$$

Now $\text{rank}(\overline{R}) = \sum_{j=1}^t e_{i_j}$ follows by (16), Lemma 5, and the fact that S is $(h + (w - u)(\delta - 1))$ -wise linearly independent.

Combining the above cases, we conclude that $d \geq h + (w - u)(\delta - 1)$. By Corollary 2,

$$\begin{aligned} d &\leq n - k_1 + 1 - (u + 1)(\delta - 1) \\ &\leq n - k + 1 - (u + 1)(\delta - 1) \\ &= h + (w - u)(\delta - 1), \end{aligned}$$

where $n = w(r + \delta - 1) + m$, $k = ur + v$, and $h = (w - u)r + m - v - \delta + 1$. Thus, we have $d = h + (w - u)(\delta - 1)$ and necessarily, $k_1 = k$, which completes the proof. \blacksquare

Remark 8: We would like to mention that the method and main idea of Construction A was first introduced in [16], based on Gabidulin codes. The purpose of Construction A that we

brought here is only to show that optimal LRCs with $(r, \delta)_a$ -locality can be generated by arranging the repair sets carefully. For more constructions of LRCs based on Gabidulin codes and their generalizations, the reader may refer to [7], [11], [16], [19].

Construction A was used in Theorem 4 with the requirement of $m \geq \delta$. For the case $0 < m \leq \delta - 1$, we apply the following construction to generate optimal codes with respect to the bound in Corollary 3.

Construction B: Fix $\mathbb{F}_q \subseteq \mathbb{F}_{q_1}$. With the notation of (3), define $h = n - k - m - w(\delta - 1)$. Let P_1 and P_2 be parity-check matrices of an $[m + r + \delta - 1, r, m + \delta]_q$ MDS code and an $[r + \delta - 1, r, \delta]_q$ MDS code, respectively. Let $S \subseteq \mathbb{F}_{q_1}$, $|S| = n$. Define $\mathcal{C}(S, h) \subseteq \mathbb{F}_{q_1}^n$ to be a linear code with parity-check matrix

$$R = \begin{pmatrix} R_1 & 0 & 0 & \dots & 0 \\ 0 & R_2 & 0 & \dots & 0 \\ 0 & 0 & R_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & R_w \\ H_1 & H_2 & H_3 & \dots & H_w \end{pmatrix}_{(n-k) \times n}, \quad (17)$$

where $R_1 = P_1$, $R_i = P_2$ for $2 \leq i \leq w$, and

$$(H_1, H_2, H_3, \dots, H_w) = V(S, h). \quad (18)$$

Theorem 5: Let $n = w(r + \delta - 1) + m$, $k = ur + v$, $0 < v < r$, and let $S \subseteq \mathbb{F}_{q_1}$ be $(h + (w + 1 - u)(\delta - 1))$ -wise linearly independent over \mathbb{F}_q . Denote by $\mathcal{C}(S, h)$ the code generated by Construction B. If $0 < m \leq \delta - 1$, $r > v \geq \lfloor \frac{r}{2} \rfloor$, and $u \geq 2r + \delta - 1$, then the code $\mathcal{C}(S, h)$ is an optimal $[n, k, d]_{q_1}$ linear code with $(r, \delta)_a$ -locality and $d = h + (w - u)(\delta - 1) + 1$.

Proof: By (17) and $R_1 = P_1$, we have that $\mathcal{C}(S, h)_{[m+r+\delta-1]}$ is an $[m + r + \delta - 1, \leq r, \geq m + \delta]_{q_1}$ linear code. Thus, $\mathcal{C}(S, h)_{S_1}$ and $\mathcal{C}(S, h)_{S_2}$ are punctured codes with parameters $[r + \delta - 1, \leq r, \geq \delta]_{q_1}$, where $S_1 = [r + \delta - 1]$ and $S_2 = [m + r + \delta - 1] \setminus [m]$. Now, by (17)-(18), we can conclude that the code $\mathcal{C}(S, h)$ is an $[n, k_1]_{q_1}$ code with $(r, \delta)_a$ -locality and $k_1 \geq k$. By Corollary 3, it is sufficient to prove that $d \geq h + (w - u)(\delta - 1) + 1$, i.e., any $h + (w - u)(\delta - 1)$ columns of R have full rank. Let

$$R^* = \begin{pmatrix} R_{i_1}^* & 0 & 0 & \dots & 0 \\ 0 & R_{i_2}^* & 0 & \dots & 0 \\ 0 & 0 & R_{i_3}^* & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & R_{i_t}^* \\ H_{i_1}^* & H_{i_2}^* & H_{i_3}^* & \dots & H_{i_t}^* \end{pmatrix}$$

denote the $h + (w - u)(\delta - 1)$ arbitrary columns chosen from R , where for $1 \leq i \leq w$, R_i^* and H_i^* denote the chosen part from R_i and H_i , respectively. By (15), R^* has full rank if and only if

$$\overline{R} = \begin{pmatrix} R_{i_1}^* & 0 & 0 & \dots & 0 \\ 0 & R_{i_2}^* & 0 & \dots & 0 \\ 0 & 0 & R_{i_3}^* & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & R_{i_t}^* \\ H_{i_1}^* & H_{i_2}^* & H_{i_3}^* & \dots & H_{i_t}^* \end{pmatrix}$$

has full rank, where if $i_1 = 1$ then R_1^* contains at least $m + \delta$ columns selected from R_1 , otherwise for $1 \leq j \leq t$, i_j denotes the block from which we choose at least δ columns, with $2 \leq i_j \leq w$.

For the case $i_1 = 1$, $\text{rank}(\bar{R}) = \text{rank}(\tilde{R})$, where

$$\tilde{R} \triangleq \begin{pmatrix} R_{1,1}^* & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & R_{i_2}^* & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & R_{i_3}^* & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & R_{i_4}^* & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & R_{i_t}^* \\ 0 & H_{1,2}^* & H_{i_2}^* & H_{i_3}^* & H_{i_4}^* & \dots & H_{i_t}^* \end{pmatrix},$$

$R_1^* = (R_{1,1}^*, R_{1,2}^*)$, with $R_{1,1}^*$ an $(m + \delta - 1) \times (m + \delta - 1)$ matrix, $H_1^* = (H_{1,1}, H_{1,2})$, and $H_{1,2}^* = H_{1,2} - H_{1,1}(R_{1,1}^*)^{-1}R_{1,2}^*$. Let e_i denote the number of columns in R_i^* for $1 \leq i \leq w$ and let e'_1 denote the number of columns in $H_{1,2}^*$. The fact that $e_1 \leq m + r + \delta - 1$ means that $e'_1 = e_1 - m - \delta + 1 \leq r$.

Case 1: Assume $i_1 = 1$ and $t \leq w - u$. By Construction B

$$h = (w - u)r - v \\ \geq \begin{cases} (r + \delta - 1)(t - 1) - (t - 1)(\delta - 1) + 2r - v \\ > \sum_{j=2}^t (e_{i_j} - \delta + 1) + r \\ \geq \sum_{j=2}^t (e_{i_j} - \delta + 1) + e'_1, & \text{if } 1 \leq t \leq w - u - 1, \\ \sum_{j=2}^t (e_{i_j} - \delta + 1) + e'_1 + m, & \text{if } t = w - u, \end{cases}$$

where for the case $t = w - u$ we use the facts that $\sum_{j=1}^{w-u} e_{i_j} \leq h + (w - u)(\delta - 1)$ and $e_1 = e'_1 + m + \delta - 1$, i.e., $\sum_{j=2}^t e_{i_j} + e'_1 \leq h + (w - u - 1)(\delta - 1) - m$. Since S is $(h + (w - u)(\delta - 1))$ -wise linearly independent over \mathbb{F}_q , by Lemma 5, we have $\text{rank}(\tilde{R}) = \sum_{j=1}^t e_{i_j}$.

Case 2: Assume $i_1 = 1$ and $t > w - u$. Then

$$\text{rank}(\tilde{R}) \\ = \text{rank} \begin{pmatrix} R_{1,1}^* & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & R_{i_2}^* & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & R_{i_3}^* & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & R_{i_4}^* & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & R_{i_t}^* \\ 0 & H_{1,2}^* & H_{i_2}^* & H_{i_3}^* & H_{i_4}^* & \dots & H_{i_t}^* \end{pmatrix} \\ \geq \text{rank} \begin{pmatrix} R_{1,1}^* & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & R_{i_2}^* & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & R_{i_{w-u}}^* & 0 & \dots & 0 \\ 0 & H_{1,2}^* & H_{i_2}^* & \dots & H_{i_{w-u}}^* & H_{i_{w-u+1}}^* & \dots & H_{i_t}^* \end{pmatrix}.$$

Recall that $\sum_{j=1}^t e_{i_j} \leq h + (w - u)(\delta - 1)$ and $e_1 = e'_1 + m + \delta - 1$, i.e., $h \geq e'_1 + m + \sum_{j=2}^t e_{i_j} - (w - u - 1)(\delta - 1)$. Thus, by Lemma 5 and the fact that S is $(h + (w - u)(\delta - 1))$ -wise linearly independent over \mathbb{F}_q , we have $\text{rank}(\tilde{R}) = m + \delta - 1 + e'_1 + \sum_{j=2}^t e_{i_j} = \sum_{j=1}^t e_{i_j}$.

Case 3: Assume $i_1 \neq 1$ and $t \leq w - u$. In this case, according to Lemma 5, $\text{rank}(\bar{R}) = \sum_{j=1}^t e_{i_j}$ follows directly

from

$$h = (w - u)r - v \\ \geq \begin{cases} (r + \delta - 1)t - t(\delta - 1) + r - v > \sum_{j=1}^t (e_{i_j} - \delta + 1), \\ \text{if } 1 \leq t \leq w - u - 1, \\ h + (w - u)(\delta - 1) - (w - u)(\delta - 1) \\ \geq \sum_{j=1}^t (e_{i_j} - \delta + 1), & \text{if } t = w - u, \end{cases}$$

and S is $(h + (w - u)(\delta - 1))$ -wise linearly independent over \mathbb{F}_q .

Case 4: Assume $i_1 \neq 1$ and $t \geq w - u$. In this case

$$\text{rank}(\bar{R}) \\ = \text{rank} \begin{pmatrix} R_{i_1}^* & 0 & 0 & 0 & \dots & 0 \\ 0 & R_{i_2}^* & 0 & 0 & \dots & 0 \\ 0 & 0 & R_{i_3}^* & 0 & \dots & 0 \\ 0 & 0 & 0 & R_{i_4}^* & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & R_{i_t}^* \\ H_{i_1}^* & H_{i_2}^* & H_{i_3}^* & H_{i_4}^* & \dots & H_{i_t}^* \end{pmatrix} \\ \geq \text{rank} \begin{pmatrix} R_{i_1}^* & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & R_{i_2}^* & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & R_{i_{w-u}}^* & 0 & \dots & 0 \\ H_{i_1}^* & H_{i_2}^* & \dots & H_{i_{w-u}}^* & H_{i_{w-u+1}}^* & \dots & H_{i_t}^* \end{pmatrix}.$$

Similarly, $\sum_{j=1}^t e_{i_j} \leq h + (w - u)(\delta - 1)$ means that $h \geq \sum_{j=1}^t e_{i_j} - (w - u)(\delta - 1)$. Now, by Lemma 5 the fact that S is $(h + (w - u)(\delta - 1))$ -wise linearly independent means that $\text{rank}(\bar{R}) = \sum_{j=1}^t e_{i_j}$.

Combining the above cases, we have $d \geq h + (w - u)(\delta - 1) + 1$. Thus, by Corollary 3, we have $d = h + (w - u)(\delta - 1) + 1$ and $k_1 = k$, which completes the proof. ■

VII. CONCLUSION

In this paper, we improved the Singleton-type bound of [8], [14] for locally repairable codes with $(r, \delta)_a$ -locality. For some special cases, the improved bound is indeed tighter than the original one. As a byproduct, we prove some locally repairable codes generated in [25] via matroid theory are indeed optimal. Two explicit optimal constructions were also introduced with respect to the improved bound.

As presented in Fig. 1, there are two cases which are still open. Whether the Singleton-type bound in [8], [14] is achievable or not in those two cases is still undecided. Those cases are:

RI: $0 < v \leq \frac{r}{2}$, $0 < m < v + \delta - 1$, $1 \leq u \leq r - v$, and $w < r + \delta - 1 - m$;

RII: $v > \frac{r}{2}$, $(r + \delta - 1)u(u - 1) - w(w + 1)(r - v) \leq mu(u - 1)$, $0 < m < v + \delta - 1$, $1 \leq u \leq r - v$, and $w < r + \delta - 1 - m$.

Additionally, the sharp bound is still unknown for many cases, namely, those cases for which the bound of (2) was proved to be unachievable (refer to Fig. 1). Those problems are left for future research.

APPENDIX

This appendix contains the omitted proofs for the claims on the properties of the ECF induced by an LRC with $(r, \delta)_a$ -locality, namely, Propositions 1, 3, 4, and 5. Throughout this appendix, we assume that \mathcal{C} is an LRC with $(r, \delta)_a$ -locality, and that the parameters n and k are as in (3). Furthermore, let \mathcal{S} be the ECF that was obtained in Lemma 2.

A. Proof of Proposition 1

For any family of subsets, $\mathcal{B} \subseteq 2^{\mathcal{X}}$, define its *overlap*, denoted $D(\mathcal{B})$, as

$$D(\mathcal{B}) = \sum_{B \in \mathcal{B}} |B| - \left| \bigcup_{B \in \mathcal{B}} B \right|.$$

It is easy to check that $D(\mathcal{B}) \geq 0$ and $D(\mathcal{B}) \geq D(\mathcal{B}')$ for $\mathcal{B}' \subseteq \mathcal{B}$. Additionally, $D(\mathcal{B}) = 0$ if and only if its sets are pairwise disjoint. We cite the following lemma, concerning the overlap, from [2].

Lemma 6 ([2, Lemma 5]): Let \mathcal{S}^* be a set of subsets of \mathcal{X} . For any integer $0 \leq t \leq |\mathcal{S}^*|$, there exists a t -subset \mathcal{V} of \mathcal{S}^* such that

$$D(\mathcal{V}) \geq \min(D(\mathcal{S}^*), \lfloor t/2 \rfloor).$$

We now further elaborate on the overlap.

Lemma 7: If \mathcal{S}^* is a set of $(r + \delta - 1)$ -subsets of \mathcal{X} with $|\mathcal{S}^*| \geq w + 1$, then for any integer $0 \leq t \leq |\mathcal{S}^*|$, there exists a t -subset \mathcal{V} of \mathcal{S}^* such that

$$D(\mathcal{V}) \geq \min \left\{ r + \delta - 1 - m, \left\lceil \frac{t(t-1)(r + \delta - 1 - m)}{(w+1)w} \right\rceil \right\}.$$

In particular, we have $D(\mathcal{S}^*) \geq r + \delta - 1 - m$.

Proof: If $t \geq w + 1$, let \mathcal{V} be any t -subset of \mathcal{S}^* . Then

$$\begin{aligned} D(\mathcal{V}) &= t(r + \delta - 1) - \left| \bigcup_{S_i^* \in \mathcal{V}} S_i^* \right| \\ &\geq (w+1)(r + \delta - 1) - n \\ &= r + \delta - 1 - m. \end{aligned}$$

If $t \leq w$, let \mathcal{V}_{w+1} be a $(w+1)$ -subset of \mathcal{S}^* . Define Θ to be the set of all the possible t -subsets of \mathcal{V}_{w+1} . We arbitrarily index the sets in $\mathcal{V}_{w+1} = \{A_1, A_2, \dots, A_{w+1}\}$. Let us consider the sum $\sum_{\mathcal{V}' \in \Theta} D(\mathcal{V}')$ in comparison with $D(\mathcal{V}_{w+1})$. Consider a fixed $\mathcal{V}' \in \Theta$, and some element $x \in \mathcal{X}$. The definition of the overlap function may be equivalently read as: $A_i, A_j \in \mathcal{V}'$, $i < j$ contribute 1 to the overlap due to x , if and only if $x \in A_i \cap A_j$ and i is the minimal index such that $x \in A_i$. We observe that if A_i, A_j contribute to $D(\mathcal{V}_{w+1})$ due to x , they do so also for any \mathcal{V}' that includes them. Additionally, A_i and A_j appear in exactly $\binom{w-1}{t-2}$ elements of Θ . Combining all of this together we obtain

$$\sum_{\mathcal{V}' \in \Theta} D(\mathcal{V}') \geq \binom{w-1}{t-2} D(\mathcal{V}_{w+1}).$$

Since $|\Theta| = \binom{w+1}{t}$, by an averaging argument there exists $\mathcal{V} \in \Theta$ such that

$$\begin{aligned} D(\mathcal{V}) &\geq \left\lceil \frac{\binom{w-1}{t-2}}{\binom{w+1}{t}} D(\mathcal{V}_{w+1}) \right\rceil \\ &\geq \left\lceil \frac{\binom{w-1}{t-2}(r + \delta - 1 - m)}{\binom{w+1}{t}} \right\rceil \\ &= \left\lceil \frac{t(t-1)(r + \delta - 1 - m)}{(w+1)w} \right\rceil. \end{aligned}$$

Then this \mathcal{V} is the desired t -subset of \mathcal{S}^* . \blacksquare

Corollary 5: If $|\mathcal{S}| \geq w + 1$, then for any integer $0 \leq t \leq |\mathcal{S}|$, there exists a t -subset \mathcal{V} of \mathcal{S} such that

$$\begin{aligned} &|\mathcal{V}|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{V}} S_i \right| \\ &\geq \min \left\{ r + \delta - 1 - m, \right. \\ &\quad \left. \max \left\{ \left\lfloor \frac{t}{2} \right\rfloor, \left\lceil \frac{t(t-1)(r + \delta - 1 - m)}{(w+1)w} \right\rceil \right\} \right\}. \end{aligned}$$

Proof: First, we extend any $S_i \in \mathcal{S}$ to an $(r + \delta - 1)$ -subset S_i^* of $[n]$, that is, $S_i \subseteq S_i^*$ and $|S_i^*| = r + \delta - 1$. Let $\mathcal{S}^* = \{S_i^* : S_i \in \mathcal{S}\}$. Obviously $\bigcup_{S_i^* \in \mathcal{S}^*} S_i^* = [n]$. Define \mathcal{T}^* to be the corresponding subset of \mathcal{S}^* for any subset \mathcal{T} of \mathcal{S} . Then $|\mathcal{T}| = |\mathcal{T}^*|$ and

$$\begin{aligned} |\mathcal{T}|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{T}} S_i \right| &\geq |\mathcal{T}^*|(r + \delta - 1) - \left| \bigcup_{S_i^* \in \mathcal{T}^*} S_i^* \right| \\ &= D(\mathcal{T}^*). \end{aligned} \quad (19)$$

By Lemmas 6 and 7, there exists a t -subset \mathcal{V}_1^* of \mathcal{S}^* such that

$$\begin{aligned} D(\mathcal{V}_1^*) &\geq \min \left\{ D(\mathcal{S}^*), \left\lfloor \frac{t}{2} \right\rfloor \right\} \\ &\geq \min \left\{ r + \delta - 1 - m, \left\lfloor \frac{t}{2} \right\rfloor \right\}. \end{aligned} \quad (20)$$

By Lemma 7, there exists a t -subset \mathcal{V}_2^* of \mathcal{S}^* such that

$$\begin{aligned} &D(\mathcal{V}_2^*) \\ &\geq \min \left\{ r + \delta - 1 - m, \left\lceil \frac{t(t-1)(r + \delta - 1 - m)}{(w+1)w} \right\rceil \right\}. \end{aligned} \quad (21)$$

The conclusion is then obtained by combining (19), (20) and (21). \blacksquare

Remark 9: Recalling the definition of $\Phi(\cdot, \cdot)$ (see Definition 3),

$$\begin{aligned} \Phi(n, t) &= \\ &\begin{cases} \min \left\{ r + \delta - 1 - m, \max \left\{ \left\lfloor \frac{t}{2} \right\rfloor, \left\lceil \frac{t(t-1)(r + \delta - 1 - m)}{(w+1)w} \right\rceil \right\} \right\}, & \text{if } m \neq 0, \\ 0, & \text{if } m = 0. \end{cases} \end{aligned}$$

Note that $D(\mathcal{S})$ may be 0 when $m = 0$, i.e., $(r + \delta - 1)|n$, which corresponds to the case $\Phi(n, t) = 0$ when $m = 0$. We may use

$\Phi(n, t)$ to lower bound the value $|\mathcal{V}|(r + \delta - 1) - |\bigcup_{S_i \in \mathcal{V}} S_i|$, for $\mathcal{V} \subseteq \mathcal{S}$.

Finally, the sought after proof for Proposition 1 is simply the combination of Corollary 5, Remark 3 and Remark 9.

B. Proof of Proposition 3

Proof: For any $S_i \in \mathcal{V}$, define \mathcal{V}_i to be a smallest subset of \mathcal{V} with $S_i \in \mathcal{V}_i$ and

$$\left| S_i \cap \left(\bigcup_{S_j \in \mathcal{V}_i \setminus \{S_i\}} S_j \right) \right| \geq |S_i| - \delta + 1, \quad (22)$$

if $\left| S_i \cap \left(\bigcup_{S_j \in \mathcal{V} \setminus \{S_i\}} S_j \right) \right| \geq |S_i| - \delta + 1$, and otherwise, define $\mathcal{V}_i = \mathcal{V}$. Note that \mathcal{V} does not satisfy Condition C1. Thus, there exists $S_i \in \mathcal{V}$ such that $|S_i \cap (\bigcup_{S_j \in \mathcal{V} \setminus \{S_i\}} S_j)| \geq |S_i| - \delta + 1$. Condition C2 implies that $|S_i \cap S_j| < |S_i| - \delta + 1$ for any $S_j \in \mathcal{V}_i$, $S_j \neq S_i$, which means that $|\mathcal{V}_i| \geq 3$, since (22) cannot hold for $|\mathcal{V}_i| \leq 2$.

Without loss of generality, we choose \mathcal{V}_τ to be the element with smallest size among $\{\mathcal{V}_i : S_i \in \mathcal{V}\}$. Then, any proper subset of \mathcal{V}_τ must satisfy Condition C1. Now we pick one $S_t \in \mathcal{V}_\tau \setminus \{S_\tau\}$. If

$$|S_\tau \cap S_t| \geq \frac{|S_\tau| - \delta + 1}{2}$$

we set $\mathcal{V}^* = \{S_t, S_\tau\}$. Otherwise, necessarily

$$\left| S_\tau \cap \left(\bigcup_{S_i \in \mathcal{V}_\tau \setminus \{S_\tau, S_t\}} S_i \right) \right| \geq \frac{|S_\tau| - \delta + 1}{2},$$

and we set $\mathcal{V}^* = \mathcal{V} \setminus \{S_t\}$. In both cases $D(\mathcal{V}^*) \geq \frac{|S_\tau| - \delta + 1}{2}$. Therefore, we have

$$\begin{aligned} & |\mathcal{V}^*|(r + \delta - 1) - \left| \bigcup_{S_i \in \mathcal{V}^*} S_i \right| \\ & \geq r + \delta - 1 + \sum_{S_i \in \mathcal{V}^* \setminus \{S_\tau\}} |S_i| - \left| \bigcup_{S_i \in \mathcal{V}^*} S_i \right| \\ & = r + \delta - 1 - |S_\tau| + D(\mathcal{V}^*) \\ & \geq r + \delta - 1 - |S_\tau| + \frac{|S_\tau| - \delta + 1}{2} \\ & = \frac{r + (r + \delta - 1 - |S_\tau|)}{2} \\ & \geq \frac{r}{2}. \end{aligned}$$

The last inequality is obtained by the fact that $|S_\tau| \leq r + \delta - 1$. ■

C. Proofs of Propositions 4 and 5

The essence of the two propositions is to reduce the family of repair sets to a sub-family that satisfies Condition C1, such that the rank of points in the union of the two families is the same. Loosely speaking, we delete some sets in a way that preserves the rank. We then choose a sub-family with full rank that satisfies Condition C1. This is implemented by Algorithm 1. It finds subsets $\mathcal{V}'_1 \subseteq \mathcal{V}_1 \subseteq \mathcal{S}$ such that $\mathcal{S} \setminus \mathcal{V}'_1$ satisfies

Algorithm 1

Input: $\mathcal{S} = \{S_1, S_2, \dots, S_{|\mathcal{S}|}\}$ the ECF from Lemma 2

- 1 $\mathcal{V}_1, \mathcal{V}'_1 \leftarrow \emptyset$
- 2 **while** there exist $S_i \in \mathcal{S} \setminus \mathcal{V}_1$, $S_j \in \mathcal{S}$, and $S_i \neq S_j$ with $|S_i \cap S_j| \geq |S_i| - \delta + 1$ **do**
- 3 $\mathcal{V}_1 \leftarrow \mathcal{V}_1 \cup \{S_i, S_j\}$
- 4 $\mathcal{V}'_1 \leftarrow \mathcal{V}'_1 \cup \{S_i\}$
- 5 **end**
- 6 **while** there exist $S_i \in \mathcal{V}_1 \setminus \mathcal{V}'_1$ and $S_j \in \mathcal{S} \setminus \mathcal{V}_1$ with $|S_i \cap S_j| \geq |S_i| - \delta + 1$ **do**
- 7 $\mathcal{V}_1 \leftarrow \mathcal{V}_1 \cup \{S_j\}$
- 8 $\mathcal{V}'_1 \leftarrow \mathcal{V}'_1 \cup \{S_i\}$
- 9 **end**
- 10 **return** $\mathcal{V}_1, \mathcal{V}'_1$

Condition C2, and $\text{rank}(\bigcup_{S \in \mathcal{S} \setminus \mathcal{V}_1} S) = \text{rank}(\bigcup_{S \in \mathcal{S} \setminus \mathcal{V}'_1} S)$, where \mathcal{S} is the ECF from Lemma 2.

Lemma 8: Let \mathcal{V}_1 and \mathcal{V}'_1 be the output of Algorithm 1. Then

$$\text{rank} \left(\bigcup_{S_j \in \mathcal{V}_1} S_j \right) = \text{rank} \left(\bigcup_{S_j \in \mathcal{V}_1 \setminus \mathcal{V}'_1} S_j \right)$$

and $\mathcal{S} \setminus \mathcal{V}_1$ satisfies Condition C2.

Proof: The first claim follows from the fact that $|S_i \cap S_j| \geq |S_i| - \delta + 1$ implies that $\text{rank}(S_j) = \text{rank}(S_i \cup S_j)$. Thus, by Algorithm 1, we have $\text{rank}(\bigcup_{S_j \in \mathcal{V}_1} S_j) = \text{rank}(\bigcup_{S_j \in \mathcal{V}_1 \setminus \mathcal{V}'_1} S_j)$. The second claim follows by the condition to terminate for first while loop of Algorithm 1, and by noting that the second while loop only removes elements from $\mathcal{S} \setminus \mathcal{V}_1$. ■

By Lemma 8, we may extend \mathcal{V}'_1 to a subset of \mathcal{V}_1 , as large as possible, denoted as \mathcal{V}^*_1 , such that

$$\begin{aligned} \text{rank} \left(\bigcup_{S_j \in \mathcal{V}_1} S_j \right) &= \text{rank} \left(\bigcup_{S_j \in \mathcal{V}_1 \setminus \mathcal{V}'_1} S_j \right) \\ &= \text{rank} \left(\bigcup_{S_j \in \mathcal{V}_1 \setminus \mathcal{V}^*_1} S_j \right). \end{aligned} \quad (23)$$

In other words, the set \mathcal{V}^*_1 satisfies that for any $S_i \in \mathcal{V}_1 \setminus \mathcal{V}^*_1$

$$\text{rank} \left(\bigcup_{S_j \in \mathcal{V}_1 \setminus \mathcal{V}^*_1} S_j \right) > \text{rank} \left(\bigcup_{S_j \in (\mathcal{V}_1 \setminus \mathcal{V}^*_1) \setminus \{S_i\}} S_j \right). \quad (24)$$

Note that a set \mathcal{V}^*_1 which satisfies (23) and (24) is not necessarily unique. We can now prove Proposition 4 and Proposition 5.

Proof of Proposition 4: Let \mathcal{V}_1 and \mathcal{V}'_1 be the output of Algorithm 1, and let \mathcal{V}^*_1 satisfy (23) and (24), as discussed above.

Claim 1): If there exists $S_\tau \in \mathcal{V}_1 \setminus \mathcal{V}^*_1$ with

$$\left| S_\tau \cap \left(\bigcup_{S_j \in (\mathcal{V}_1 \setminus \mathcal{V}^*_1) \setminus \{S_\tau\}} S_j \right) \right| \geq |S_\tau| - \delta + 1,$$

then $\text{rank}(S_\tau) = \text{rank}(S_\tau \cap (\bigcup_{S_j \in (\mathcal{V}_1 \setminus \mathcal{V}_1^*) \setminus \{S_\tau\}} S_j))$ by Remark 1-1), which contradicts (24).

Claim 2): By Algorithm 1, if $S_j \in \mathcal{V}_1 \setminus \mathcal{V}_1^* \subseteq \mathcal{V}_1 \setminus \mathcal{V}_1^*$ there must exist $S_i \in \mathcal{V}_1^*$ such that $|S_i \cap S_j| \geq |S_i| - \delta + 1$ due to Line 2 and Line 6 of the algorithm. Hence, $\text{rank}(S_i) = \text{rank}(S_i \cap S_j)$ and $\text{span}(S_i) \subseteq \text{span}(S_j)$ by Definition 1 and Remark 1.

Claim 3): Recall that by Lemma 8, the set $\mathcal{S} \setminus \mathcal{V}_1$ satisfies Condition C2, i.e., for any $S_i, S_j \in \mathcal{S} \setminus \mathcal{V}_1$ we have $|S_i \cap S_j| < \min\{|S_i|, |S_j|\} - \delta + 1$. We further consider S_i and S_j in the following three cases:

Case 1: There exist two distinct $S_i, S_j \in \mathcal{V}_1 \setminus \mathcal{V}_1^*$ with $|S_i \cap S_j| \geq |S_i| - \delta + 1$. However, this is impossible by Claim 1).

Case 2: There exist two distinct $S_i \in \mathcal{S} \setminus \mathcal{V}_1$ and $S_j \in \mathcal{V}_1 \setminus \mathcal{V}_1^*$ with $|S_i \cap S_j| \geq |S_i| - \delta + 1$. This is impossible by the first while loop of Algorithm 1.

Case 3: There exist two distinct $S_i \in \mathcal{V}_1 \setminus \mathcal{V}_1^* \subseteq \mathcal{V}_1 \setminus \mathcal{V}_1^*$ and $S_j \in \mathcal{S} \setminus \mathcal{V}_1$ with $|S_i \cap S_j| \geq |S_i| - \delta + 1$. This is impossible by the second while loop of Algorithm 1.

Thus, the claim follows. ■

Proof of Proposition 5: We proceed claim by claim.

Claim 1): By (23), we have

$$\text{rank}\left(\bigcup_{S_i \in \mathcal{V}_1} S_i\right) = \text{rank}\left(\bigcup_{S_i \in \mathcal{V}_1 \setminus \mathcal{V}_1^*} S_i\right),$$

which implies that $\mathcal{G}_{\bigcup_{S_i \in \mathcal{V}_1^*} S_i} \subseteq \text{span}(\bigcup_{S_i \in \mathcal{V}_1 \setminus \mathcal{V}_1^*} S_i)$. Thus, the conclusion is obtained by the fact that $\Upsilon \subseteq \bigcup_{S_i \in \mathcal{V}_1^*} S_i$.

Claim 2): Define $T_{S_i} = S_i \setminus (\bigcup_{S_j \in \mathcal{S} \setminus \{S_i\}} S_j)$ for any $S_i \in \mathcal{S}$. The definition of the ECF implies that $T_{S_i} \neq \emptyset$ and $T_{S_i} \cap T_{S_j} = \emptyset$ for any distinct $S_i, S_j \in \mathcal{S}$. By Proposition 4-2), for any $S_i \in \mathcal{V}_1 \setminus \mathcal{V}_1^*$, there exists a set $S_i^* \in \mathcal{V}_1^*$ with $\mathcal{G}_{S_i^*} \subseteq \text{span}(S_i)$. Note that $T_{S_i^*} \subseteq \Upsilon \cap S_i^*$ and $\mathcal{G}_{\Upsilon \cap S_i^*} \subseteq \mathcal{G}_{\Upsilon} \cap \text{span}(S_i)$. According to Algorithm 1, Lines 3, 7, and 8, whenever a set S_j is included in $\mathcal{V}_1 \setminus \mathcal{V}_1^* \supseteq \mathcal{V}_1 \setminus \mathcal{V}_1^*$ a distinct set (we denote) S_j^* is included in $\mathcal{V}_1^* \subseteq \mathcal{V}_1^*$ with $\text{span}(S_j^*) \subseteq \text{span}(S_j)$. Thus, we can assume that for any $S_{j_1} \neq S_{j_2} \in \mathcal{V}_1 \setminus \mathcal{V}_1^*$ we have $S_{j_1}^* \neq S_{j_2}^* \in \mathcal{V}_1^*$. Now the desired result follows, namely,

$$\left| \mathcal{G}_{\Upsilon} \cap \text{span}\left(\bigcup_{S_i \in \mathcal{U}} S_i\right) \right| \geq \left| \bigcup_{S_i \in \mathcal{U}} T_{S_i^*} \right| \geq |\mathcal{U}|$$

for any subset $\mathcal{U} \subseteq \mathcal{V}_1 \setminus \mathcal{V}_1^*$.

Claim 3): Setting $\mathcal{U} = \mathcal{V}_1 \setminus \mathcal{V}_1^*$, the above inequality becomes

$$|\mathcal{V}_1 \setminus \mathcal{V}_1^*| \leq \left| \mathcal{G}_{\Upsilon} \cap \text{span}\left(\bigcup_{S_i \in \mathcal{V}_1 \setminus \mathcal{V}_1^*} S_i\right) \right| = |\mathcal{G}_{\Upsilon}| = |\Upsilon| = M. \quad (25)$$

Let T_{S_i} be the subset defined in Claim 2). Since for any $S_i \in \mathcal{V}_1^*$, we have $\emptyset \neq T_{S_i} \subseteq S_i \setminus (\bigcup_{S_j \in \mathcal{S} \setminus \mathcal{V}_1^*} S_j)$, it follows that $|\mathcal{V}_1^*| \leq |\Upsilon| = M$. Thus, in combination with (25), we have $|\mathcal{V}_1| = |\mathcal{V}_1^*| + |\mathcal{V}_1 \setminus \mathcal{V}_1^*| \leq 2M$. ■

ACKNOWLEDGMENT

The authors thank Xing Lin for correcting a mistake in the figure. They would like to thank the Associate Editor,

Prof. Arya Mazumdar and the anonymous reviewers, whose comments and suggestions improved the presentation of this paper.

REFERENCES

- [1] 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.
- [2] H. Cai, Y. Miao, M. Schwartz, and X. Tang, "On optimal locally repairable codes with super-linear length," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4853–4868, Aug. 2020.
- [3] H. Cai and M. Schwartz, "On optimal locally repairable codes and generalized sector-disk codes," *IEEE Trans. Inf. Theory*, vol. 67, no. 2, pp. 686–704, Feb. 2021.
- [4] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.
- [5] T. Ernvall, S. El Rouayheb, C. Hollanti, and H. V. Poor, "Capacity and security of heterogeneous distributed storage systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 2701–2709, Dec. 2013.
- [6] T. Ernvall, T. Westerbäck, R. Freij-Hollanti, and C. Hollanti, "Constructions and properties of linear locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1129–1143, Mar. 2016.
- [7] 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.
- [8] 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.
- [9] C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," *ACM Trans. Storage*, vol. 9, no. 1, p. 3, 2013.
- [10] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, vol. 16. Amsterdam, The Netherlands: Elsevier, 1977.
- [11] U. Martínez-Peñas and F. R. Kschischang, "Universal and dynamic locally repairable codes with maximal recoverability via sum-rank codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 7790–7805, Dec. 2019.
- [12] F. Oggier and A. Datta, "Self-repairing homomorphic codes for distributed storage systems," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1215–1223.
- [13] D. S. Papailiopoulos and A. G. Dimakis, "Locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5843–5855, Oct. 2014.
- [14] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, "Optimal linear codes with a local-error-correction property," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 2776–2780.
- [15] 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.
- [16] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimally locally repairable and secure codes for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 212–236, Nov. 2013.
- [17] 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*, Jul. 2013, pp. 1819–1823.
- [18] R. C. R. C. Singleton, "Maximum distance Q-nary codes," *IEEE Trans. Inf. Theory*, vol. IT-10, no. 2, pp. 116–118, Apr. 1964.
- [19] W. Song, S. H. Dau, C. Yuen, and T. J. Li, "Optimal locally repairable linear codes," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 1019–1036, May 2014.
- [20] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4661–4676, May 2014.
- [21] I. Tamo, A. Barg, and A. Frolow, "Bounds on the parameters of locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3070–3083, Jun. 2016.
- [22] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, "Optimal locally repairable codes and connections to Matroid theory," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6661–6671, Dec. 2016.
- [23] I. Tamo, Z. Wang, and J. Bruck, "Zigzag codes: MDS array codes with optimal rebuilding," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1597–1616, Mar. 2013.
- [24] A. Wang and Z. Zhang, "An integer programming-based bound for locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5280–5294, Oct. 2015.
- [25] T. Westerbäck, R. Freij-Hollanti, T. Ernvall, and C. Hollanti, "On the combinatorics of locally repairable codes via matroid theory," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5296–5315, Oct. 2016.

Han Cai (Member, IEEE) received the B.S. and M.S. degrees in mathematics from Hubei University, Wuhan, China, in 2009 and 2013, respectively, and the Ph.D. degree from the Department of Communication Engineering, Southwest Jiaotong University, Chengdu, China, in 2017. From October 2015 to October 2017, he was a Visiting Ph.D. Student with the Faculty of Engineering, Information and Systems, University of Tsukuba, Japan. From 2018 to 2021, he was a Post-Doctoral Fellow at the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel. In 2021, he joined Southwest Jiaotong University, where he currently holds a tenure-track position. His research interests include coding theory and sequence design.

Cuiling Fan received the B.S. and Ph.D. degrees in mathematics from Hebei Normal University, Shijiazhuang, China, in 2005 and 2010, respectively. From 2012 to 2014, she was a Post-Doctoral Fellow with the Department of Mathematics, Zhejiang University, Hangzhou, China. Since 2014, she has been with the School of Mathematics, Southwest Jiaotong University, China, where she is currently a Professor. Her research interests include combinatorial design theory, coding theory, cryptography, and their interactions.

Ying Miao received the D.Sci. degree in mathematics from Hiroshima University, Hiroshima, Japan, in 1997.

From 1989 to 1993, he worked at the Suzhou Institute of Silk Textile Technology, Suzhou, Jiangsu, China. From 1995 to 1997, he was a Research Fellow of the Japan Society for the Promotion of Science. During 1997–1998, he was a Post-Doctoral Fellow at the Department of Computer Science, Concordia University, Montreal, QC, Canada. In 1998, he joined the University of Tsukuba, Tsukuba, Ibaraki, Japan, where he is currently a Full Professor at the Faculty of Engineering, Information and Systems. His current research interests include combinatorics, coding theory, and information security.

Dr. Miao has received the 2001 Kirkman Medal from the Institute of Combinatorics and its Applications. He is on the editorial boards of several journals, such as *Graphs and Combinatorics* and *Journal of Combinatorial Designs*.

Moshe Schwartz (Senior Member, IEEE) received the B.A. (*summa cum laude*), M.Sc., and Ph.D. degrees from the Computer Science Department, Technion—Israel Institute of Technology, Haifa, Israel, in 1997, 1998, and 2004, respectively.

He was a Fulbright Post-Doctoral Researcher at the Department of Electrical and Computer Engineering, University of California San Diego; and a Post-Doctoral Researcher at the Department of Electrical Engineering, California Institute of Technology. While on sabbatical (2012–2014), he was a Visiting Scientist at the Massachusetts Institute of Technology (MIT). He is currently a Professor with the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel. His research interests include algebraic coding, combinatorial structures, and digital sequences.

Prof. Schwartz has received the 2009 IEEE Communications Society Best Paper Award in Signal Processing and Coding for Data Storage and the 2020 NVMW Persistent Impact Prize. He served as an Associate Editor for coding techniques and coding theory for the IEEE TRANSACTIONS ON INFORMATION THEORY during 2014–2021. Since 2021, he has been serving as an Area Editor for coding and decoding for the IEEE TRANSACTIONS ON INFORMATION THEORY. Since 2021, he has been an Editorial Board Member for the *Journal of Combinatorial Theory, Series A*.

Xiaohu Tang (Senior Member, IEEE) received the B.S. degree in applied mathematics from Northwest Polytechnic University, Xi'an, China, in 1992, the M.S. degree in applied mathematics from Sichuan University, Chengdu, China, in 1995, and the Ph.D. degree in electronic engineering from Southwest Jiaotong University, Chengdu, in 2001.

From 2003 to 2004, he was a Research Associate at the Department of Electrical and Electronic Engineering, The Hong Kong University of Science and Technology. From 2007 to 2008, he was a Visiting Professor at the University of Ulm, Germany. Since 2001, he has been with the School of Information Science and Technology, Southwest Jiaotong University, where he is currently a Professor. His research interests include coding theory, network security, distributed storage, and information processing for big data.

Dr. Tang was a recipient of the National Excellent Doctoral Dissertation Award in 2003 (China), the Humboldt Research Fellowship in 2007 (Germany), and the Outstanding Young Scientist Award by NSFC in 2013 (China). He has served on a number of technical program committees for conferences. He has served as an Associate Editor for several journals, including IEEE TRANSACTIONS ON INFORMATION THEORY and *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*.