On Optimal Locally Repairable Codes With Super-Linear Length

Han Cai, *Member, IEEE*, Ying Miao, Moshe Schwartz, *Senior Member, IEEE*, and Xiaohu Tang, *Senior Member, IEEE*

***Abstract*— In this paper, locally repairable codes which have optimal minimum Hamming distance with respect to the bound presented by Prakash *et al.* are considered. New upper bounds on the length of such optimal codes are derived. The new bounds apply to more general cases, and have weaker requirements compared with the known ones. In this sense, they both improve and generalize previously known bounds. Further, optimal codes are constructed, whose length is order-optimal with respect to the new upper bounds. Notably, the length of the codes is super-linear in the alphabet size.**

***Index Terms*— Distributed storage, locally repairable codes, packings, Steiner systems.**

I. Introduction

LARGE-SCALE cloud storage and distributed file systems, such as Amazon Elastic Block Store (EBS) and Google File System (GoogleFS), have reached such a massive scale that disk failures are the norm and not the exception. In those systems, to protect the data from disk failures, the simplest solution is a straightforward replication of data packets across different disks. However, this solution suffers from a large storage overhead. As an alternative solution, $[n, k]$ maximum distance separable (MDS) codes, i.e., codes achieving the Singleton bound, are used as storage codes, which encode k information symbols to n symbols and store them across n disks. Using MDS codes leads to a dramatic improvement in redundancy compared with replication. However, for MDS codes, when one node fails, the system recovers it at the cost of contacting k surviving symbols, thus complicating the repair process.

Manuscript received February 3, 2019; revised October 31, 2019; accepted January 29, 2020. Date of publication March 2, 2020; date of current version July 14, 2020. The work of Han Cai and Moshe Schwartz was supported by the German Israeli Project Cooperation (DIP) under Grant PE2398/1-1. The work of Ying Miao was supported by the JSPS Grant-in-Aid for Scientific Research (B) under Grant 18H01133. The work of Xiaohu Tang was supported by the National Natural Science Foundation of China under Grant 61871331. This article was presented in part at the 2019 IEEE International Symposium on Information Theory. *(Corresponding author: Han Cai.)*

Han Cai and Moshe Schwartz are with the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer Sheva 8410501, Israel (e-mail: hancai@aliyun.com; schwartz@ee.bgu.ac.il).

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).

Xiaohu Tang is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China (e-mail: xhutang@swjtu.edu.cn).

Communicated by L. Dolecek, Associate Editor for Coding Techniques.

Digital Object Identifier 10.1109/TIT.2020.2977647

To improve the repair efficiently, in [15], locally repairable codes were introduced to reduce the number of symbols contacted during the repair process of a failed node. More precisely, locally repairable codes ensure that a failed symbol can be recovered by accessing only $r \ll k$ other symbols [15].

The original concept of locality only works when exactly one erasure occurs (that is, one node fails). Over the past few years, several generalizations have been suggested for the definition of locality. As examples we mention locality with a single repair set tolerating multiple erasures [26], locality with disjoint multiple repairable sets [7], [28], [31], [35], hierarchical locality [30], and unequal locality [20]. For constructions of locally repairable codes refer to [4], [6], [14], [25] as examples.

In this paper, we focus on locally repairable codes with a single repair set that can repair multiple erasures locally [26]. By ensuring $\delta - 1 \geqslant 2$ redundancies in each repair set, this kind of locally repairable codes guarantees that the system can recover from $\delta - 1$ erasures by accessing r surviving code symbols for each erasure. This is denoted as (r, δ)-locality.

Research on codes with (r, δ)-locality has proceeded along two main tracks. In the first track, upper bounds on the minimum Hamming distance and the code length have been studied. Singleton-type bounds were introduced for codes with (r, δ)-locality in [26], [32], [36]. In [5], a bound depending on the size of the alphabet was derived for the Hamming distance of codes with (r, δ)-locality. Via linear programming, another bound related with the size of the alphabet was introduced in [1]. Very recently, in [13], an interesting connection between the length of optimal linear codes with $(r, 2)$-locality and the size of the alphabet was derived.

In the second research track, constructions for optimal locally repairable codes have been studied. In [27], a construction of optimal locally repairable codes was introduced based on Gabidulin codes over a finite filed with size $q = \Theta((r + \delta - 1)^{(rn)/(r+\delta-1)})$. By analyzing the structure of repair sets, optimal locally repairable codes were also constructed in [32] with $q = \Theta(\binom{n}{k})$. In [34], a construction of optimal locally repairable codes with $q = \Theta(n)$ was proposed. In [33] and [37], optimal locally repairable codes were constructed using matroid theory. The construction of [34] was generalized in [21] to include more flexible parameters when $n \leqslant q$. Recently, in [23], cyclic optimal locally repairable codes with unbounded length were constructed for $\delta = 2$ and Hamming distance $d = 3, 4$. Finally, for the case of

$\delta = 2$ and Hamming distance $d = 5$, [3], [13], [17] presented constructions of locally repairable codes that have optimal distance as well as order-optimal length $n = \Theta(q^2)$.

In a practical setting, long codes over small fields are preferred. This is due to the fact that smaller fields have much cheaper and faster implementations both in hardware and in software. Thus, a common question is, given a desirable code family, and given a field size, how long can a code of this family be. Perhaps the most famous instance of this question is the MDS conjecture (e.g., see [24]), stating that any MDS code has length linear in the field size (in fact, its length is almost exactly the field size). Recently, analogous with the case of MDS codes, Guruswami *et al.* [13] asked a fundamental interesting question: How long can an optimal code with (r, δ)-locality (with respect to the Singleton-type bound in [26]) be for given r, δ, and field size q? An answer to this question was given for the aforementioned case $\delta = 2$, which was proved to be tight for some cases. The motivation of this paper is to further answer this question for the general case $\delta > 2$.

The main contribution of this paper is the study of optimal linear codes with (r, δ)-locality and length that is super-linear in the field size. We analyze the structure of optimal locally repairable codes. Firstly, we derive a new upper bound on the length of optimal locally repairable codes for the case of $\delta > 2$. Secondly, as a byproduct, we prove that the bound for $\delta = 2$ in [13] not only holds for some other cases (see Remark 1 in this paper) besides the one mentioned in [13] but also can be improved for the case $d > r + \delta$. Finally, we give a general construction of locally repairable codes with length that is super-linear in the field size. Based on some special structures such as packings and Steiner systems, locally repairable codes with optimal Hamming distances and order-optimal length $\Omega(q^\delta)$ with respect to the new bound ($\delta > 2$) are obtained. This is to say, the bound for $\delta > 2$ is also asymptotically tight for some special cases.

The remainder of this paper is organized as follows. Section II introduces some preliminaries about locally repairable codes. Section III establishes an upper bound for the length of optimal locally repairable codes for the case $\delta > 2$. Section IV presents a construction of optimal locally repairable codes with length $n > q$. Section V concludes this paper with some remarks.

II. Preliminaries

We present the notation and basic definitions used throughout the paper. For a positive integer $n \in \mathbb{N}$, we define $[n] = \{1, 2, \ldots, n\}$. For any prime power q, let $\mathbb{F}_q$ denote the finite field with q elements. An $[n, k]_q$ linear code $\mathcal{C}$ over $\mathbb{F}_q$ is a k-dimensional subspace of $\mathbb{F}_q^n$ with a $k \times n$ generator matrix $G = (\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n)$, where $\mathbf{g}_i$ is a column vector of dimension k for all $i \in [n]$. Specifically, it is called an $[n, k, d]_q$ linear code if the minimum Hamming distance is d. For a subset $S \subseteq [n]$, let $|S|$ denote the cardinality of S, let 2^S denote the set of all subsets of S, and define

$$\mathrm{Rank}(S) = \mathrm{Rank}(\mathrm{Span}\,\{\mathbf{g}_i | i \in S\}).$$

In [11], Gopalan *et al.* introduce the following definition for the locality of code symbols. The ith $(1 \leqslant i \leqslant n)$ code symbol c_i of an $[n, k, d]_q$ linear code $\mathcal{C}$ is said to have locality r $(1 \leqslant r \leqslant k)$, if it can be recovered by accessing at most r other symbols in $\mathcal{C}$. More precisely, symbol locality can also be rigorously defined as follows.

Definition 1 ([11]): For any column $\mathbf{g}_i$ of G with $i \in [n]$, define $\mathrm{Loc}(\mathbf{g}_i)$ as the smallest integer r such that there exists an $(r+1)$-subset $R_i = \{i, i_1, i_2, \ldots, i_r\} \subseteq [n]$ satisfying

$$\mathbf{g}_i \in \mathrm{Span}(R_i \setminus \{i\}), \text{ i.e., } \mathbf{g}_i = \sum_{t=1}^{r} \lambda_t \mathbf{g}_{i_t}, \quad \lambda_t \in \mathbb{F}_q. \quad (1)$$

Equivalently, for any codeword $C = (c_1, c_2, \ldots, c_n) \in \mathcal{C}$, the ith component

$$c_i = \sum_{t=1}^{r} \lambda_t c_{i_t}, \quad \lambda_t \in \mathbb{F}_q.$$

Define $\mathrm{Loc}(S) = \max_{i \in S} \mathrm{Loc}(\mathbf{g}_i)$ for any set $S \subseteq [n]$. Then, an $[n, k, d]_q$ linear code $\mathcal{C}$ is said to have information locality r if there exists $S \subseteq [n]$ with $\mathrm{Rank}(S) = k$ satisfying $\mathrm{Loc}(S) = r$. Furthermore, an $[n, k, d]_q$ linear code $\mathcal{C}$ is said to have all symbol locality r if $\mathrm{Loc}([n]) = r$.

To guarantee that the system can locally recover from multiple erasures, say, $\delta - 1$ erasures, the definition of locality was generalized in [26] as follows.

Definition 2 ([26]): The jth column $\mathbf{g}_j$, $j \in [n]$, of a generator matrix G of an $[n, k]_q$ linear code $\mathcal{C}$ is said to have (r, δ)-locality if there exists a subset $S_j \subseteq [n]$ such that:

- $j \in S_j$ and $|S_j| \leqslant r + \delta - 1$; and
- the minimum Hamming distance of the punctured code $\mathcal{C}|_{S_j}$ obtained by deleting the code symbols c_t $(t \in [n] \setminus S_j)$ is at least δ,

where the set S_j is also called a (r, δ)-repair set of $\mathbf{g}_j$. The code $\mathcal{C}$ is said to have information (r, δ)-locality if there exists $S \subseteq [n]$ with $\mathrm{Rank}(S) = k$ such that for each $j \in S$, $\mathbf{g}_j$ has (r, δ)-locality. Furthermore, the code $\mathcal{C}$ is said to have all symbol (r, δ)-locality if all the code symbols have (r, δ)-locality.

In [26] (for the case $\delta = 2$ [11]), the following upper bound on the minimum Hamming distance of linear codes with information (r, δ)-locality was derived.

Lemma 1 ([26]): For an $[n, k, d]_q$ linear code with information (r, δ)-locality,

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

Additionally, a locally repairable code is said to be *optimal* if its minimum Hamming distance attains this bound with equality.

The following fact is very useful to determine the minimum Hamming distance.

Fact 1 ([24]): An $[n, k]_q$ linear code $\mathcal{C}$ has minimum Hamming distance d if and only if d is the largest integer such that

$$|S| \leqslant n - d$$

for every $S \subseteq [n]$ with $\mathrm{Rank}(S) \leqslant k - 1$.

III. Bounds on the Length of Locally Repairable Codes

The goal of this section is to derive upper bounds on the length of optimal locally repairable codes. We extend known techniques which were employed for the case of $\delta = 2$, and apply them to the case of $\delta > 2$. In particular, we construct linear codes from locally repairable codes and then apply the Hamming bound to the constructed codes. This connection requires a careful analysis of the structure and properties of the repair sets. More precisely, we need to find a set of repair sets that form a partition of the n code symbols, where the punctured codes over each repair set is an MDS code. This has been studied before only in some special cases [13], [32], but in the general case it is still an open question.

We open this section by first characterizing the properties of repair sets of locally repairable codes in Theorem 1. We then give connections between optimal locally repairable codes with $\delta > 2$ and the case $\delta = 2$ (linear codes), in Lemma 2. Then Theorem 2 provides bounds on the length of optimal locally repairable codes. Finally, Corollaries 2 and 3 introduce a method that may improve the performance of known bounds for the cases $d > r + \delta$.

Throughout this section, let

$$n = (r + \delta - 1)w + m, \qquad k = ru + v,$$

where $\delta \geqslant 2$, $0 \leqslant m \leqslant r + \delta - 2$, and $0 \leqslant v \leqslant r - 1$ are all integers.

Theorem 1: Let $\mathcal{C}$ be an optimal $[n, k, d]_q$ linear code with all symbol (r, δ)-locality, where the optimality is with respect to the bound in Lemma 1. Let $\Gamma \subseteq 2^{[n]}$ be the set of all possible (r, δ)-repair sets. Write $k = ru + v$, for integers u and v, and $0 \leqslant v \leqslant r - 1$. If $(r + \delta - 1)|n$, $k > r$, and additionally, $u \geqslant 2(r - v + 1)$ or $v = 0$, then there exists a set of (r, δ)-repair sets $\mathcal{S} \subseteq \Gamma$, such that all $R \in \mathcal{S}$ are of cardinality $|R| = r + \delta - 1$, and $\mathcal{S}$ is a partition of $[n]$.

The proof of Theorem 1 is lengthy, involving several auxiliary lemmas. It is therefore deferred to the appendix. Based on Theorem 1, we derive a corollary that slightly extends [32, Theorem 9], which was originally proved only for $r|k$. It has a very similar proof, which we give here for completeness.

Corollary 1: Let $\mathcal{C}$ be an optimal $[n, k, d]_q$ linear code with all symbol (r, δ)-locality, where the optimality is with respect to the bound in Lemma 1. If $k > r$, $n = w(r + \delta - 1)$, and additionally $r|k$ or $u \geqslant 2(r+1-v)$, then there are w pairwise-disjoint (r, δ)-repair sets, $R_1, \ldots, R_w \subseteq [n]$, such that for all $1 \leqslant i \leqslant w$, $|R_i| = r + \delta - 1$, and the punctured code $\mathcal{C}|_{R_i}$ is a linear $[r + \delta - 1, r, \delta]_q$ MDS code.

Proof: We contend that the repair sets, $\mathcal{S}$, from Theorem 1, satisfy the requirements. Thus, it remains to prove that for each $\mathcal{C}|_R$, $R \in \mathcal{S}$, the Hamming distance is exactly δ. Assume to the contrary, and without loss of generality, that $d(\mathcal{C}|_{R_1}) > \delta$.

Note that $\bigcup_{1\leqslant i\leqslant w} R_i = [n]$ means $\mathrm{Rank}(\bigcup_{1\leqslant i\leqslant w} R_i) = k$ and then $w = \frac{n}{r+\delta-1} \geqslant \lceil \frac{k}{r} \rceil$ since $\mathrm{Rank}(R_i) \leqslant r$ for $1 \leqslant i \leqslant w$. Also recall our notation that $v \equiv k \bmod r$ and $0 \leqslant v < r$. Fix some arbitrary set $R' \subseteq R_{\lceil \frac{k}{r} \rceil}$, with $|R'| = v$ if $v \neq 0$, and $|R'| = r$ if $v = 0$. Consider now the set

$$S = R' \cup \left(\bigcup_{1\leqslant i\leqslant \lceil \frac{k}{r} \rceil - 1} R_i \right).$$

By the Singleton bound we have,

$$\begin{aligned}\mathrm{Rank}(S) &\leqslant \mathrm{Rank}(R') + \sum_{1\leqslant i\leqslant \lceil \frac{k}{r} \rceil - 1} \mathrm{Rank}(R_i) \\ &\leqslant \begin{cases} v + \sum_{1\leqslant i\leqslant \lceil \frac{k}{r} \rceil - 1} (r + \delta - 1 - d(\mathcal{C}|_{R_i}) + 1) \\ < v + r(\lceil \frac{k}{r} \rceil - 1) = k, & \text{if } v \neq 0, \\ r + \sum_{1\leqslant i\leqslant \lceil \frac{k}{r} \rceil - 1} (r + \delta - 1 - d(\mathcal{C}|_{R_i}) + 1) \\ < r + r(\lceil \frac{k}{r} \rceil - 1) = k, & \text{if } v = 0. \end{cases}\end{aligned}$$

We also have

$$\begin{aligned}|S| &= \begin{cases} v + (r + \delta - 1)\left(\lceil \frac{k}{r} \rceil - 1\right), & \text{if } v \neq 0, \\ r + (r + \delta - 1)\left(\lceil \frac{k}{r} \rceil - 1\right), & \text{if } v = 0, \end{cases} \\ &= k + \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right)(\delta - 1).\end{aligned}$$

But now this contradicts the optimality of $\mathcal{C}$ by Fact 1. ■

In the sequel, the discussion is based on the structure of the repair sets given in Corollary 1.

Lemma 2: Let $n = w(r + \delta - 1)$, $\delta > 2$, $k = ur + v > r$, and additionally, $r|k$ or $u \geqslant 2(r+1-v)$, where all parameters are integers. If there exists an optimal $[n, k, d]_q$ linear code $\mathcal{C}$ with all symbol (r, δ)-locality, then

i There exists a $[w(r + 1), k, d']_q$ linear code $\mathcal{C}'$ with all symbol $(r, 2)$-locality (i.e., locality r), and $d' \geqslant 2\lfloor (d-1)/\delta \rfloor + 1$;

ii There exists a linear code with parameters $[wr, k' \geqslant k, d' \geqslant t + 1]_q$.

Proof: For the first claim, by Corollary 1, and up to a rearrangement of the code coordinates, the code $\mathcal{C}$ has parity-check matrix P of the following form,

$$P = \begin{pmatrix} L^{(1)} & 0 & 0 & \ldots & 0 \\ 0 & L^{(2)} & 0 & \ldots & 0 \\ 0 & 0 & L^{(3)} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & L^{(w)} \\ H_1 & H_2 & H_3 & \ldots & H_w \end{pmatrix},$$

where $L^{(i)} = (I_{\delta-1}, P_i)$ is a $(\delta - 1) \times (r + \delta - 1)$ matrix for all $1 \leqslant i \leqslant w$. Herein, without loss of generality, we assume $L^{(i)}$ with canonical form for $1 \leqslant i \leqslant w$. For all $1 \leqslant i \leqslant w$, rewrite the $(\delta - 1) \times (r + \delta - 1)$ matrix $L^{(i)} = (I_{\delta-1}, P_i)$ as

$$L^{(i)} = \begin{pmatrix} L^{(i)}_{1,1} & L^{(i)}_{1,2} \\ L^{(i)}_{2,1} & L^{(i)}_{2,2} \end{pmatrix},$$

where $L^{(i)}_{2,2}$ is a $(\delta - 2) \times (\delta - 2)$ matrix. It is easy to check that $\det(L^{(i)}_{2,2}) \neq 0$ for all $1 \leqslant i \leqslant w$, since $L^{(i)}$ is a parity-check matrix of an $[r + \delta - 1, r, \delta]_q$ MDS code according to Corollary 1. By column linear transformations, the matrix P

is equivalent to

$$\begin{pmatrix} Q_1 & 0 & 0 & \dots & 0 \\ 0 & Q_2 & 0 & \dots & 0 \\ 0 & 0 & Q_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & Q_w \\ H'_1 & H'_2 & H'_3 & \dots & H'_w \end{pmatrix}, \tag{3}$$

where

$$Q_i = \begin{pmatrix} Q_{i,1} = L^{(i)}_{1,1} - L^{(i)}_{1,2}(L^{(i)}_{2,2})^{-1}L^{(i)}_{2,1} & L^{(i)}_{1,2} \\ 0 & L^{(i)}_{2,2} \end{pmatrix}, \tag{4}$$

$$H'_i = (H'_{i,1} = H_{i,1} - H_{i,2}(L^{(i)}_{2,2})^{-1}L^{(i)}_{2,1}, H'_{i,2} = H_{i,2}) \tag{5}$$

with $H_i = (H_{i,1}, H_{i,2})$.

Now consider the code $\mathcal{C}'$ with parity-check matrix

$$P' = \begin{pmatrix} Q_{1,1} & 0 & 0 & \dots & 0 \\ 0 & Q_{2,1} & 0 & \dots & 0 \\ 0 & 0 & Q_{3,1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & Q_{w,1} \\ H'_{1,1} & H'_{2,1} & H'_{3,1} & \dots & H'_{w,1} \end{pmatrix}, \tag{6}$$

where $Q_{i,1}$ and $H'_{i,1}$, for $1 \leqslant i \leqslant w$, are defined by (4) and (5), respectively.

Given a set of coordinates $T = \{t_1, \dots, t_\ell\} \subseteq [r+\delta-1]$, and given $A = (A_1, \dots, A_{r+\delta-1})$, we define the projection of A onto T by $\Delta_T(A) = (A_{t_1}, A_{t_2}, \dots, A_{t_l})$ (where the order of coordinates in the projection will not matter to us). We emphasize that $Q_{i,1}$, for all $1 \leqslant i \leqslant w$, does not have a zero coordinate, since according to Corollary 1, $\Delta_{S_\tau}(Q_i)$ has full rank, where we define $S_\tau = \{\tau\} \cup \{r+2, r+3, \dots, r+\delta-1\}$, $\tau \in [r+1]$. Thus, by (6), $\mathcal{C}'$ is a code with all symbol $(r,2)$-locality.

To complete the proof we only need to show $d' \geqslant 2t+1$, where we define $t = \lfloor (d-1)/\delta \rfloor$. Namely, we need to show that any $2t$ columns of P' are linearly independent. A selection of $2t$ columns from P', denoted by $\mathcal{T}'$, has the following general form,

$$\Delta_{\mathcal{T}'}(P') \triangleq \begin{pmatrix} \Delta_{T'_1}(Q_{1,1}) & 0 & \dots & 0 \\ 0 & \Delta_{T'_2}(Q_{2,1}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Delta_{T'_w}(Q_{w,1}) \\ \Delta_{T'_1}(H'_{1,1}) & \Delta_{T'_2}(H'_{2,1}) & \dots & \Delta_{T'_w}(H'_{w,1}) \end{pmatrix},$$

where $\sum_{1 \leqslant i \leqslant w} |T'_i| = 2t$. Since the locality of $\mathcal{C}'$ guarantees recovery from any one erasure independently, the non-trivial cases to consider are those where $T'_{\tau_i} \geqslant 2$ for $1 \leqslant \tau_i \leqslant w$ and $1 \leqslant i \leqslant s$, where s denotes the number of sets T'_i with $|T'_i| \geqslant 2$ and $s \leqslant \min(t, w)$.

With a coordinate selection $\mathcal{T}'$ from P' we naturally associate a coordinate selection $\mathcal{T}$ from P, defined by

$$T_{\tau_i} = T'_{\tau_i} \cup \{r+2, r+2, \dots, r+\delta-1\},$$

for $1 \leqslant i \leqslant s$, and with $\sum_{1 \leqslant i \leqslant s} |T_{\tau_i}| = 2t + s(\delta-2) \leqslant t\delta \leqslant d-1$. Recall that if $\{r+2, r+3, \dots, r+\delta-1\} \subset T \subseteq [r+\delta-1]$ then (3), (4) and (5) imply that

$$\begin{pmatrix} \Delta_T(L^{(i)}) \\ \Delta_T(H_i) \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \Delta_T(Q_i) \\ \Delta_T(H'_i) \end{pmatrix}$$

are rank equivalent, based on only invertible column linear transformations for $1 \leqslant i \leqslant w$. Note that the distance of $\mathcal{C}$ satisfies $d \geqslant \delta t + 1 \geqslant 2t + s(\delta-2)+1$, which implies that any $\sum_{1 \leqslant i \leqslant s} |T_{\tau_i}| \leqslant 2t + s(\delta-2)$ columns of P have full rank of $\sum_{1 \leqslant i \leqslant s} |T_{\tau_i}|$, i.e.,

$$\begin{aligned} &\sum_{1 \leqslant i \leqslant s} |T_{\tau_i}| \\ =&\text{Rank} \begin{pmatrix} \Delta_{T_{\tau_1}}(L^{(\tau_1)}) & 0 & \dots & 0 \\ 0 & \Delta_{T_{\tau_2}}(L^{(\tau_2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Delta_{T_{\tau_s}}(L^{(\tau_s)}) \\ \Delta_{T_{\tau_1}}(H_{\tau_1}) & \Delta_{T_{\tau_2}}(H_{\tau_2}) & \dots & \Delta_{T_{\tau_s}}(H_{\tau_s}) \end{pmatrix} \\ =&\text{Rank} \begin{pmatrix} \Delta_{T_{\tau_1}}(Q_{\tau_1}) & 0 & \dots & 0 \\ 0 & \Delta_{T_{\tau_2}}(Q_{\tau_2}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Delta_{T_{\tau_s}}(Q_{\tau_s}) \\ \Delta_{T_{\tau_1}}(H'_{\tau_1}) & \Delta_{T_{\tau_2}}(H'_{\tau_2}) & \dots & \Delta_{T_{\tau_s}}(H'_{\tau_s}) \end{pmatrix}, \end{aligned} \tag{7}$$

where the second equality holds by (3), (4), (5) and the fact that $\{r+2, r+3, \dots, r+\delta-1\} \subseteq T_{\tau_i}$ for $1 \leqslant i \leqslant s$. Therefore, by (4), (5), and (7), we have

$$\begin{aligned} &\text{Rank}\,(\Delta_{\mathcal{T}'}(P')) \\ =&\text{Rank} \begin{pmatrix} \Delta_{T'_{\tau_1}}(Q_{\tau_1,1}) & 0 & \dots & 0 \\ 0 & \Delta_{T'_{\tau_2}}(Q_{\tau_2,1}) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \Delta_{T'_{\tau_s}}(Q_{\tau_s,1}) \\ \Delta_{T'_{\tau_1}}(H'_{\tau_1,1}) & \Delta_{T'_{\tau_2}}(H'_{\tau_2,1}) & \dots & \Delta_{T'_{\tau_s}}(H'_{\tau_s,1}) \end{pmatrix} \\ =&\sum_{1 \leqslant i \leqslant s} |T'_{\tau_i}|, \end{aligned}$$

where $T'_{\tau_i} = T_{\tau_i} \setminus \{r+2, r+3, \dots, r+\delta-1\}$ for $1 \leqslant i \leqslant s$. This is to say, the code $\mathcal{C}'$ can recover from any $2t$ erasures, hence, $d' \geqslant 2t+1$.

For the second claim, the proof is similar and the only different is that we consider the parity-check matrix P, which is equivalent with

$$\begin{pmatrix} (I_{\delta-1}, 0) & (0,0) & \dots & (0,0) \\ (0,0) & (I_{\delta-1}, 0) & \dots & (0,0) \\ \vdots & \vdots & \ddots & \vdots \\ (0,0) & (0,0) & \dots & (I_{\delta-1}, 0) \\ (H_{1,1}, H^*_1) & (H_{2,1}, H^*_2) & \dots & (H_{w,1}, H^*_w) \end{pmatrix},$$

where $H^*_i = (H_{i,2} - H_{i,1})$ for $1 \leqslant i \leqslant w$. By the analysis as the first case, it is easy to check that

$$\begin{aligned} &(H^*_1, H^*_2, \dots, H^*_w) \\ =&(H_{1,2} - H_{1,1}, H_{2,2} - H_{2,1}, \cdots, H_{w,2} - H_{w,1}) \end{aligned}$$

is a parity check matrix of a linear code with parameters $[w(r+1), k' \geqslant k, d \geqslant t+1]_q$. ∎

The following bound is derived from Lemma 2. The proof follows the same path as the proof of [13, Theorem 3.2]. We bring it here for completeness.

Theorem 2: Let $n = w(r+\delta-1)$, $\delta \geqslant 2$, $k = ur+v$, and additionally, $r|k$ or $u \geqslant 2(r+1-v)$, where all parameters are integers. Assume there exists an optimal $[n,k,d]_q$ linear code $\mathcal{C}$ with all symbol (r,δ)-locality, and define $t = \lfloor (d-1)/\delta \rfloor$. If $2t+1 > 4$, then

$$n \leqslant \begin{cases} \frac{r+\delta-1}{r}\left(\frac{t-1}{2(q-1)}q^{\frac{2(w-u)r-2v-2}{t-1}}+1\right), & \text{if } t \text{ is odd,} \\ \frac{t(r+\delta-1)}{2r(q-1)}q^{\frac{2(w-u)r-2v}{t}}, & \text{if } t \text{ is even,} \end{cases}$$

where $w-u$ can also be rewritten as $w-u = \lfloor (d-1+v)/(r+\delta-1) \rfloor$.

Proof: By Lemma 2-(ii), we have a linear code $\mathcal{C}_1$, with parameters $[wr, k = ur+v, d_2 \geqslant t+1]_q$.

Now we apply the Hamming bound [24] to $\mathcal{C}_1$. We distinguish between two cases, depending on the parity of t.

Case 1: t is odd. In this case, consider the shortened code of $\mathcal{C}_1$ with parameters $[wr-1, k = ur+v, d_2 \geqslant t]_q$, then by the Hamming bound we have

$$\begin{aligned} q^{ur+v} &\leqslant \frac{q^{wr-1}}{\sum_{0\leqslant i\leqslant \frac{t-1}{2}}\binom{wr-1}{i}(q-1)^i} \\ &\leqslant \frac{q^{wr-1}}{\binom{wr-1}{\frac{t-1}{2}}(q-1)^{\frac{t-1}{2}}} \\ &\leqslant \frac{q^{wr-1}}{\left(\frac{wr-1}{\frac{t-1}{2}}\right)^{\frac{t-1}{2}}(q-1)^{\frac{t-1}{2}}}, \end{aligned}$$

i.e.,

$$wr \leqslant \frac{t-1}{2(q-1)}q^{\frac{2(w-u)r-2v-2}{t-1}}+1.$$

This is to say,

$$n \leqslant \frac{r+\delta-1}{r}\left(\frac{t-1}{2(q-1)}q^{\frac{2(w-u)r-2v-2}{t-1}}+1\right).$$

Case 2: t is even. Similarly, by the Hamming bound, we have

$$\begin{aligned} q^{ur} \leqslant \frac{q^{wr}}{\sum_{1\leqslant i\leqslant \frac{t}{2}}\binom{wr}{i}(q-1)^i} &\leqslant \frac{q^{wr}}{\binom{wr}{\frac{t}{2}}(q-1)^{\frac{t}{2}}} \\ &\leqslant \frac{q^{wr}}{\left(\frac{wr}{\frac{t}{2}}\right)^{\frac{t}{2}}(q-1)^{\frac{t}{2}}}, \end{aligned}$$

which means

$$n \leqslant \frac{t(r+\delta-1)}{2r(q-1)}q^{\frac{2(w-u)r-2v}{t}}.$$

By Lemma 1, $\mathcal{C}$ is optimal means that

$$\begin{aligned} &d-1 \\ =&\begin{cases} w(r+\delta-1)-ur-v-u(\delta-1), & \text{if } v \neq 0, \\ w(r+\delta-1)-ur-(u-1)(\delta-1), & \text{if } v = 0, \end{cases} \end{aligned} \tag{8}$$

i.e., $w-u = \lfloor (d-1+v)/(r+\delta-1) \rfloor$. This completes the proof. ■

Remark 1: For the case $\delta = 2$, let $d = 4\tau + a$ for $a \in \{1,2,3,4\}$, then $t = 2\tau$ for $a = 1,2$ and $t = 2\tau+1$ for $a = 3,4$. Note that by (8), $t = \lfloor \frac{d-1}{\delta} \rfloor \geqslant 2$ (or $d \geqslant 5$) means that $w \geqslant u$. By Lemma 1 and (8), the bounds can be rewritten as

$$n \leqslant \begin{cases} \frac{r+1}{r}\left(\frac{d-a}{4(q-1)}q^{\frac{4(d-3-w+u)}{d-a}}+1\right) & \\ \quad \leqslant \frac{r+1}{r}\left(\frac{d-a}{4(q-1)}q^{\frac{4(d-3)}{d-a}}+1\right), & \text{if } a = 3,4, \\ \frac{(d-a)(r+1)}{4r(q-1)}q^{\frac{4(d-2-w+u)}{d-a}} & \\ \quad \leqslant \frac{(d-a)(r+1)}{4r(q-1)}q^{\frac{4(d-2)}{d-a}}, & \text{if } a = 1,2, \end{cases}$$

for $v = 0$, and

$$n \leqslant \begin{cases} \frac{r+1}{r}\left(\frac{d-a}{4(q-1)}q^{\frac{4(d-2-w+u)}{d-a}}+1\right) & \\ \quad \leqslant \frac{r+1}{r}\left(\frac{d-a}{4(q-1)}q^{\frac{4(d-3)}{d-a}}+1\right), & \text{if } a = 3,4, \\ \frac{(d-a)(r+1)}{4r(q-1)}q^{\frac{4(d-1-w+u)}{d-a}} & \\ \quad \leqslant \frac{(d-a)(r+1)}{4r(q-1)}q^{\frac{4(d-2)}{d-a}}, & \text{if } a = 1,2 \end{cases}$$

for $v > 0$.

Although, for the case of $\delta = 2$, we obtain a similar bound to the one in [13], our bound is an improvement since it has more relaxed conditions. In particular, the bound of [13, Theorem 10] requires $\frac{n}{r+1} \geqslant (d-2-\left\lfloor \frac{d-2}{r+1} \right\rfloor)(3r+2) + \lfloor \frac{d-2}{r+1} \rfloor + 1$, i.e., $k = \Omega(dr^2)$ [13], whereas we only require $k = \Omega(r^2)$.

Recalling Corollary 1 again, we can improve the performance of the bounds on the length of optimal locally repairable codes with all symbol (r,δ)-locality for the case $d > r+\delta$ by the following corollary.

Corollary 2: Let $n = w(r+\delta-1)$, $\delta \geqslant 2$, $k = ur+v > r$, and additionally, $r|k$ or $u \geqslant 2(r+1-v)$, where all parameters are integers. If there exists an optimal $[n,k,d]_q$ linear code $\mathcal{C}$ with $d > r+\delta$ and all symbol (r,δ)-locality, then there exists an optimal linear code $\mathcal{C}'$ with all symbol (r,δ)-locality and parameters $[n-\epsilon(r+\delta-1), k, d' = d-\epsilon(r+\delta-1)]_q$, where $\epsilon = \lceil (d-1)/(r+\delta-1) \rceil - 1$.

Proof: By Corollary 1, there are $R_1, R_2, \cdots, R_w$ such that $\mathcal{C}|_{R_i}$, $1 \leqslant i \leqslant w$, is an $[r+\delta-1, r, \delta]_q$ MDS code. Note that $\epsilon = \lceil (d-1)/(r+\delta-1) \rceil - 1$. The fact $\mathcal{C}$ is optimal means that

$$d = n-k+1-\left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\delta-1), \tag{9}$$

by Lemma 1. Recall that $k > r$, $n = w(r+\delta-1)$, and $d > r+\delta$. Thus, we have $1 \leqslant \epsilon \leqslant w-1$. Now let $\mathcal{C}'$ be the punctured code of $\mathcal{C}$ over the set $W = \bigcup_{\epsilon+1\leqslant i\leqslant w-1} R_i$, i.e., $\mathcal{C}' = \mathcal{C}|_W$. The fact $\mathcal{C}'|_{R_i} = \mathcal{C}|_{R_i}$ for $\epsilon+1 \leqslant i \leqslant w$ is an $[r+\delta-1, r, \delta]_q$ MDS code means that $\mathcal{C}'$ has all symbol (r,δ)-locality. The fact $\mathcal{C}' = \mathcal{C}|_W$ implies $n' = n - \sum_{1\leqslant i\leqslant \epsilon} |R_i| = n-\epsilon(r+\delta-1)$ and

$$d' \geqslant d - \sum_{1\leqslant i\leqslant \epsilon} |R_i| = d-\epsilon(r+\delta-1) > 0,$$

which also means $|\mathcal{C}| = |\mathcal{C}'|$. However, by Lemma 1, we have

$$\begin{aligned} d' &\leqslant n'-k+1-\left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\delta-1) \\ &= n-\epsilon(r+\delta-1)-k+1-\left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\delta-1) \\ &= d-\epsilon(r+\delta-1), \end{aligned}$$

where the last equality follows by (9). Thus, we have $d' = d - \epsilon(r+\delta-1)$. Again by Lemma 1 the code $\mathcal{C}'$ is also an optimal linear code with all symbol (r,δ)-locality and parameters $[n-\epsilon(r+\delta-1), k, d' = d-\epsilon(r+\delta-1)]_q$, which completes the proof. ■

By Corollary 2, we can firstly reduce the optimal locally repairable code $\mathcal{C}$ into an optimal locally repairable code $\mathcal{C}'$ with $d' \leqslant r+\delta$ and then apply Theorem 2 to get an upper bound for the length of $\mathcal{C}$.

Corollary 3: Let $n = w(r+\delta-1)$, $\delta \geqslant 2$, $k = ur+v > r$, and additionally, $r|k$ or $u \geqslant 2(r+1-v)$, where all parameters are integers. If there exists an optimal $[n,k,d]_q$ linear code $\mathcal{C}$ with $d > r+\delta$ and all symbol (r,δ)-locality, then

$$n \leqslant \epsilon(r+\delta-1) + \begin{cases} \frac{r+\delta-1}{r}\left(\frac{t-1}{2(q-1)}q^{\frac{2(w'-u)r-2v-2}{t-1}}+1\right), & \text{if } t \text{ is odd,} \\ \frac{t(r+\delta-1)}{2r(q-1)}q^{\frac{2(w'-u)r-2v}{t}}, & \text{if } t \text{ is even,} \end{cases}$$

where $\epsilon = \lceil (d-1)/(r+\delta-1)\rceil - 1$, $d' = d-\epsilon(r+\delta-1)$, $w' = w-\epsilon$, and $t = \lfloor (d'-1)/(\delta)\rfloor$ so that $2t+1 > 4$ holds.

In the next section, we will prove that the bound in Theorem 2 is asymptotically tight for some special cases, i.e., there indeed exist some optimal linear codes with all symbol (r,δ)-locality and asymptotically optimal length. In addition, we will also prove the condition $2t+1 > 4$ is necessary, by constructing linear codes with length independent of the field size q for the case $2t+1 \leqslant 4$.

IV. Optimal Locally Repairable Codes With Super-Linear Length

In this section, our goal is to construct optimal locally repairable codes with length n that is super-linear in the field size q. To this end, we first introduce a generic construction of locally repairable codes (Construction A). Next, in Theorem 3, we derive our main result on the minimum Hamming distance of codes constructed by Construction A. Finally, we demonstrate applications of Construction A by employing some combinatorial structures such as union-intersection-bounded families, packings, and Steiner systems to generate optimal locally repairable codes with super-linear length.

A. A General Construction

In the subsection, to streamline the presentation we adopt a slightly different notation than the previous one: we use $n = w(r+\delta-1)$ and $k = (w-1)r+v$ for $0 < v \leqslant r$, where all parameters are integers.

Construction A: Let the k information symbols be partitioned into w sets, say,

$$I^{(i)} = \{I_{i,1}, I_{i,2}, \ldots, I_{i,r}\}, \quad \text{for } i \in [w-1],$$
$$I^{(w)} = \{I_{w,1}, I_{w,2}, \ldots, I_{w,v}\}.$$

A linear code with length n is constructed by describing a linear map from the information $\boldsymbol{I} = (I_{1,1}, \ldots, I_{w,v}) \in \mathbb{F}_q^k$ to a codeword $\boldsymbol{C}(\boldsymbol{I}) = (c_{1,1}, \ldots, c_{w,r+\delta-1}) \in \mathbb{F}_q^n$, thus the $[n,k]_q$ linear code is $\mathcal{C} = \{\boldsymbol{C}(\boldsymbol{I}) : \boldsymbol{I} \in \mathbb{F}_q^k\}$. This mapping is performed by the following three steps:

1) Step 1 – Partial Parity Check Symbols: For $1 \leqslant i \leqslant w-1$, let $S_i = \{\theta_{i,t} : 1 \leqslant t \leqslant r+\delta-1\}$ be an $(r+\delta-1)$-subset of $\mathbb{F}_q$ and let $f_i(x)$ be the unique polynomial over $\mathbb{F}_q$ with $\deg(f_i) \leqslant r-1$ that satisfies $f_i(\theta_{i,t}) = I_{i,t}$ for $1 \leqslant t \leqslant r$. For $1 \leqslant i \leqslant w-1$ and $1 \leqslant t \leqslant r+\delta-1$, set $c_{i,t} = f_i(\theta_{i,t})$.

2) Step 2 – Auxiliary Symbols: Let $\{\alpha_t : 1 \leqslant t \leqslant r-v\} \subseteq \mathbb{F}_q \setminus (\bigcup_{1\leqslant i\leqslant w-1} S_i)$. For $1 \leqslant i \leqslant w-1$, and $1 \leqslant t \leqslant r-v$, define

$$a_{i,t} = \frac{f_i(\alpha_t)}{\prod_{\theta\in S_i}(\alpha_t - \theta)}. \tag{10}$$

3) Step 3 – Global Parity Check Symbols: Let $S_w = \{\theta_{w,t} : 1 \leqslant t \leqslant r+\delta-1\}$ be an $(r+\delta-1)$-subset of $\mathbb{F}_q \setminus \{\alpha_t : 1 \leqslant t \leqslant r-v\}$ and let $f_w(x)$ be the unique polynomial over $\mathbb{F}_q$ with $\deg(f_w) \leqslant r-1$ that satisfies $f_w(\theta_{w,t}) = I_{w,t}$ for $1 \leqslant t \leqslant v$, as well as

$$\sum_{1\leqslant i\leqslant w} a_{i,t} = 0 \text{ for } 1 \leqslant t \leqslant r-v, \tag{11}$$

where $a_{w,t} = \frac{f_w(\alpha_t)}{\prod_{\theta\in S_w}(\alpha_t-\theta)}$ for $1 \leqslant t \leqslant r-v$. Here, the polynomial $f_w(x)$ can be viewed as a polynomial over $\mathbb{F}_q$ determined by $I_{w,j}$, $1 \leqslant j \leqslant v$ and $a_{w,t}$ for $1 \leqslant t \leqslant r-v$. Thus, $f_w(x)$ is unique and well defined. Set $c_{w,j} = f_w(\theta_{w,j})$, for $1 \leqslant j \leqslant r+\delta-1$.

Remark 2: At first glance there appears to be a distinction between code symbols $c_{i,j}$ with $1 \leqslant i \leqslant w-1$ and those with $i = w$. However, careful thought reveals that the code symbols that correspond to the sets S_i for $1 \leqslant i \leqslant w$ are essentially symmetric, i.e., any $w-1$ sets of code symbols can determine v code symbols of the remaining set according to (11).

Theorem 3: Let μ be a positive integer, and let $S_i \subseteq \mathbb{F}_q$, $i \in [w]$ be the sets defined in Construction A. If every subset $\mathcal{R} \subseteq \{S_i : 1 \leqslant i \leqslant w\}$, $|\mathcal{R}| = \mu$, satisfies that for all $S' \in \mathcal{R}$,

$$\left| S' \cap \left(\bigcup_{S\in\mathcal{R}\setminus\{S'\}} S \right) \right| < \delta, \tag{12}$$

then the code $\mathcal{C}$ generated by Construction A is an $[n,k,d]_q$ linear code, with $d \geqslant \min\{r-v+\delta, (\mu+1)\delta\}$ and with all symbol (r,δ)-locality, where $n = w(r+\delta-1)$, $k = (w-1)r+v$, $1 \leqslant v \leqslant r$, and all parameters are integers.

Proof: By Steps 1 and 3, it is easy to check that the code $\mathcal{C}$ generated by Construction A has all symbol (r,δ)-locality. By Definition 2, the repair sets are the coordinates of the code symbols $\{f_i(\theta) : \theta \in S_i\}$ for $1 \leqslant i \leqslant w$. To simplify the notation, instead of define those coordinates, we directly use S_i, $1 \leqslant i \leqslant w$ to denote the repair sets in this proof. The code $\mathcal{C}$ is an $[n,k]_q$ linear code with $n = w(r+\delta-1)$ and $k = (w-1)r+v$ according to Construction A. To complete the proof, we only need to show that $d \geqslant d_1 = \min\{r-v+\delta, (\mu+1)\delta\}$, i.e., the code $\mathcal{C}$ can recover from any $d_1 - 1$ erasures.

According to the all symbol (r,δ)-locality, it is sufficient to consider those repair sets containing strictly more than $\delta-1$ erasures, where for the code $\mathcal{C}$ the repair sets correspond to S_i for $1 \leqslant i \leqslant w$. Since the maximum number of erasures we should consider is $d_1 - 1$, there are at most $\frac{d_1-1}{\delta}$ repair sets which can have size larger than or equal to δ. Without loss of generality, we assume that there are ℓ sets, $S_1, \dots, S_\ell$, that contain at least δ erasures each, and those erasures are located in coordinates $E_i \subseteq S_i$ for $1 \leqslant i \leqslant \ell \leqslant \frac{d_1-1}{\delta}$. Denote $|E_i| = \tau_i \geqslant \delta$ for $1 \leqslant i \leqslant \ell$ and $\sum_{1\leqslant i\leqslant \ell} \tau_i \leqslant d_1 - 1 \leqslant r - v + \delta - 1$. In what follows, we prove the claim by induction on both ℓ and the total number of erasures $\sum_{1\leqslant i\leqslant \ell} \tau_i$.

For the induction base consider the case of $\ell = 1$ and $\delta \leqslant |E_1| \leqslant d_1 - 1$. By Steps 1 and 3, we know $f_i(x)$ for $2 \leqslant i \leqslant w$, i.e., $a_{i,t}$ is available for $2 \leqslant i \leqslant w$ and $1 \leqslant t \leqslant r - v$. By (11), $a_{1,t}$ for $1 \leqslant t \leqslant r - v$ can be calculated. Recall that $|E_1| \leqslant d_1 - 1 \leqslant r - v + \delta - 1$. We know at least v values $f_1(\theta)$ for $\theta \in S_1 \setminus E_1$, which together with $f_1(\alpha_t) = a_{1,t} \prod_{1\leqslant j\leqslant r+\delta-1}(\alpha_t - \theta_{1,j})$ for $1 \leqslant t \leqslant r - v$ show that $f_1(x)$ can be recovered. Here we use the fact that $\{\alpha_t \ : \ 1 \leqslant t \leqslant r - v\} \cap S_1 = \phi$, i.e., $\prod_{1\leqslant j\leqslant r+\delta-1}(\alpha_t - \theta_{1,j}) \neq 0$. This is to say, we can recover all the code symbols $f_1(\theta)$ for $\theta \in E_1$. We emphasize that in this case, the S_i's are not required to satisfy (12), so the restriction on the size of the finite field in this case is $q \geqslant 2r + \delta - v - 1$.

For the induction hypothesis assume that for the case $1 \leqslant \ell = s < \frac{d_1-1}{\delta}$ and $\sum_{1\leqslant i\leqslant s} \tau_i = T < d_1 - 1$, the code symbols $f_i(\theta)$ for $\theta \in E_i$ and $1 \leqslant i \leqslant s$ are recoverable.

The induction step is divided into two cases. For the first case, assume an erasure pattern with $\sum_{1\leqslant i\leqslant s} \tau_i = T + 1 \leqslant d_1 - 1$. Note that if $s = 1$ the claim holds by the induction base. Therefore, we only consider $s \geqslant 2$. Since $s < \frac{d_1-1}{\delta} \leqslant \frac{(\mu+1)\delta-1}{\delta}$, we have $s \leqslant \mu$. Thus, by (12),

$$\left| E_i \cap \left(\bigcup_{\substack{1\leqslant j\leqslant s\\ j\neq i}} E_j \right) \right| \leqslant \left| S_i \cap \left(\bigcup_{\substack{1\leqslant j\leqslant s\\ j\neq i}} S_j \right) \right| \leqslant \delta - 1,$$

which means that the elements of each E_i may be indexed $E_i = \{e_{i,t} \ : \ 1 \leqslant t \leqslant \tau_i\}$ such that

$$\{e_{i,t} \ : \ 1 \leqslant t \leqslant \tau_i - \delta + 1\} \cap E_j = \phi \text{ for } 1 \leqslant i \neq j \leqslant s. \quad (13)$$

By polynomial interpolation, $f_i(x)$ for $1 \leqslant i \leqslant s$ with $\deg(f_i(x)) \leqslant r - 1$ is represented as

$$\begin{aligned}
& f_i(x) \\
&= \sum_{\theta \in S_i \setminus \{e_{i,j} \ : \ \tau_i - \delta + 2 \leqslant j \leqslant \tau_i\}} \frac{f_i(\theta) \prod\limits_{\tau_i-\delta+2\leqslant j\leqslant\tau_i} (\theta - e_{i,j})}{\prod\limits_{\theta_1 \in S_i \setminus \{\theta\}} (\theta - \theta_1)} \\
&\quad \cdot \frac{\prod\limits_{\theta_1 \in S_i} (x - \theta_1)}{(x - \theta) \prod\limits_{\tau_i-\delta+2\leqslant j\leqslant\tau_i} (x - e_{i,j})} \\
&= \sum_{\theta_{i,t} \in S_i \setminus E_i} \frac{c_{i,t} \prod_{\tau_i-\delta+2\leqslant j\leqslant\tau_i} (\theta_{i,t} - e_{i,j})}{\prod_{\theta_1 \in S_i \setminus \{\theta_{i,t}\}} (\theta_{i,t} - \theta_1)} \\
&\quad \cdot \frac{\prod_{\theta \in S_i} (x - \theta)}{(x - \theta_{i,t}) \prod\limits_{\tau_i-\delta+2\leqslant j\leqslant\tau_i} (x - e_{i,j})} \\
&\quad + \sum_{1\leqslant t\leqslant \tau_i-\delta+1} \varpi_{i,t} \frac{\prod_{\theta \in S_i} (x - \theta)}{(x - e_{i,t}) \prod\limits_{\tau_i-\delta+2\leqslant j\leqslant\tau_i} (x - e_{i,j})} \\
&= g_i(x) + \sum_{1\leqslant t\leqslant \tau_i-\delta+1} \varpi_{i,t} \frac{\prod_{\theta \in S_i} (x - \theta)}{(x - e_{i,t}) \prod\limits_{\tau_i-\delta+2\leqslant j\leqslant\tau_i} (x - e_{i,j})},
\end{aligned} \quad (14)$$

where $g_i(x)$ is determined by the accessible code symbols corresponding to $S_i \setminus E_i$ and

$$\varpi_{i,t} = f_i(e_{i,t}) \frac{\prod_{\tau_i-\delta+2\leqslant j\leqslant\tau_i} (e_{i,t} - e_{i,j})}{\prod_{\theta_1 \in S_i \setminus \{e_{i,t}\}} (e_{i,t} - \theta_1)},$$

with $\prod_{\tau_i-\delta+2\leqslant j\leqslant\tau_i} (e_{i,t} - e_{i,j}) / \prod_{\theta_1 \in S_i \setminus \{e_{i,t}\}} (e_{i,t} - \theta_1)$ being a nonzero constant for $1 \leqslant i \leqslant s$ and $1 \leqslant t \leqslant \tau_i - \delta + 1$. Combining (14) with (11), we have

$$\begin{aligned}
& (\varpi_{1,1}, \dots, \varpi_{1,\tau_1-\delta+1}, \dots, \varpi_{s,\tau_s-\delta+1}) M \\
=& (\varpi_{1,1}, \dots, \varpi_{1,\tau_1-\delta+1}, \dots, \varpi_{s,\tau_s-\delta+1}) \\
& \begin{pmatrix}
m_{\lambda_{1,1},1} & m_{\lambda_{1,1},2} & \cdots & m_{\lambda_{1,1},r-v} \\
m_{\lambda_{1,2},1} & m_{\lambda_{1,2},2} & \cdots & m_{\lambda_{1,2},r-v} \\
\vdots & \vdots & \dots & \vdots \\
m_{\lambda_{1,\tau_1-\delta+1},1} & m_{\lambda_{1,\tau_1-\delta+1},2} & \cdots & m_{\lambda_{1,\tau_1-\delta+1},r-v} \\
\vdots & \vdots & \dots & \vdots \\
m_{\lambda_{s,\tau_s-\delta+1},1} & m_{\lambda_{s,\tau_s-\delta+1},2} & \cdots & m_{\lambda_{t_{s,\tau_s-\delta+1}},r-v}
\end{pmatrix} \\
=& (w_1, w_2, \dots, w_{r-v}),
\end{aligned} \quad (15)$$

where $(w_1, w_2, \dots, w_{r-v})$ is a constant vector determined by the accessible code symbols with

$$w_i = -\sum_{1\leqslant j\leqslant s} \frac{g_j(\alpha_i)}{\prod_{\theta \in S_j} (\alpha_i - \theta)} - \sum_{s+1\leqslant j\leqslant w} \frac{f_j(\alpha_i)}{\prod_{\theta \in S_j} (\alpha_i - \theta)}$$

for $1 \leqslant i \leqslant r - v$,

$$v_1 = \sum_{1\leqslant j\leqslant s} (\tau_i - \delta + 1) \leqslant r - v - (s-1)(\delta - 1) < r - v$$

and

$$m_{\lambda_{i,j},z} = \frac{1}{(\alpha_z - e_{i,j}) \prod_{\tau_i-\delta+2\leqslant t\leqslant\tau_i} (\alpha_z - e_{i,t})}$$

for $1 \leqslant i \leqslant s$, $1 \leqslant j \leqslant \tau_i - \delta + 1$, and $1 \leqslant z \leqslant r - v$.

Recall that

$$\prod_{\tau_i-\delta+2\leqslant j\leqslant\tau_i} (e_{i,t} - e_{i,j}) / \prod_{\theta_1 \in S_i \setminus \{e_{i,t}\}} (e_{i,t} - \theta_1)$$

is a nonzero constant for $1 \leqslant i \leqslant s$ and $1 \leqslant t \leqslant \tau_i - \delta + 1$. Thus, recovering the vector

$$(f_1(e_{1,1}), \dots, f_1(e_{1,\tau_1-\delta+1}), \dots, f_s(e_{s,\tau_s-\delta+1}))$$

is equivalent to recovering the vector

$$(\varpi_{1,1}, \dots, \varpi_{1,\tau_1-\delta+1}, \dots, \varpi_{s,\tau_s-\delta+1}).$$

Note that the equation (15) has at least one solution, namely, the solution that corresponds to the original codeword. Thus, by (15), $(f_1(e_{1,1}), \dots, f_1(e_{1,\tau_1-\delta+1}), \dots, f_s(e_{s,\tau_s-\delta+1}))$ is recoverable if and only if the solution is unique, i.e., the rank of M is v_1, or equivalently, there exist v_1 columns of M forming a non-singular sub-matrix. Recall that by the induction hypothesis, the erasure pattern $E_1, E_2, \ldots, E_s \setminus \{e_{s,\tau_s-\delta+1}\}$ is recoverable, i.e., there exists a $(v_1-1)\times(v_1-1)$ matrix with

$$\det\begin{pmatrix} m_{\lambda_{1,1},t_1} & m_{\lambda_{1,1},t_2} & \cdots & m_{\lambda_{1,1},t_{v_1-1}} \\ m_{\lambda_{1,2},t_1} & m_{\lambda_{1,2},t_2} & \cdots & m_{\lambda_{1,2},t_{v_1-1}} \\ \vdots & \vdots & \cdots & \vdots \\ m_{\lambda_{1,\tau_1-\delta+1},t_1} & m_{\lambda_{1,\tau_1-\delta+1},t_2} & \cdots & m_{\lambda_{1,\tau_1-\delta+1},t_{v_1-1}} \\ \vdots & \vdots & \cdots & \vdots \\ m_{\lambda_{s,\tau_s-\delta},t_1} & m_{\lambda_{s,\tau_s-\delta},t_2} & \cdots & m_{\lambda_{s,\tau_s-\delta},t_{v_1-1}} \end{pmatrix} \neq 0. \tag{16}$$

If the erasure pattern $E_1, E_2, \ldots, E_s$ is not recoverable, then each $v_1 \times v_1$ sub-matrix of M is singular. Thus, α_i for $1 \leqslant i \leqslant r-v$ are roots of $h(x) = 0$ with

$$h(x) = \det\begin{pmatrix} m_{\lambda_{1,1},t_1} & \cdots & m_{\lambda_{1,1},t_{v_1-1}} & m_{\lambda_{1,1}}(x) \\ m_{\lambda_{1,2},t_1} & \cdots & m_{\lambda_{1,2},t_{v_1-1}} & m_{\lambda_{1,2}}(x) \\ \vdots & \cdots & \vdots & \vdots \\ m_{\lambda_{1,\tau_1-\delta+1},t_1} & \cdots & m_{\lambda_{1,\tau_1-\delta+1},t_{v_1-1}} & m_{\lambda_{1,\tau_1-\delta+1}}(x) \\ \vdots & \vdots & \cdots & \vdots \\ m_{\lambda_{s,\tau_s-\delta+1},t_1} & \cdots & m_{\lambda_{s,\tau_s-\delta+1},t_{v_1-1}} & m_{\lambda_{s,\tau_s-\delta+1}}(x) \end{pmatrix}, \tag{17}$$

where

$$m_{\lambda_{i,j}}(x) = \frac{1}{(x-e_{i,j})\prod_{\tau_i-\delta+2\leqslant t\leqslant\tau_i}(x-e_{i,t})} \tag{18}$$

for $1 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant \tau_i - \delta + 1$.

Note that $h(x)\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta)$ is a polynomial with degree less than $\sum_{1\leqslant i\leqslant s}\tau_i - \delta \leqslant r-v+\delta-1-\delta = r-v-1$ and α_i for $1 \leqslant i \leqslant r-v$ are its roots, hence $h(x)\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta) \equiv 0$. However, for $1 \leqslant i, i_1 \leqslant s$, $1 \leqslant j \leqslant \tau_i-\delta+1$ and $1 \leqslant j_1 \leqslant \tau_{i_1}-\delta+1$, (13) means that $e_{i,j} \notin \{e_{i_1,j_1}\} \cup \{e_{i_1,t} : \tau_i-\delta+2 \leqslant t \leqslant \tau_i\}$ when $(i,j) \neq (i_1,j_1)$. It follows that for $1 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant \tau_i-\delta+1$, $e_{i,j}$ is a root of $m_{\lambda_{i_1,j_1}}(x)\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta) = 0$ for all $(i_1,j_1) \neq (i,j)$ with $1 \leqslant i_1 \leqslant s$ and $1 \leqslant j_1 \leqslant \tau_{i_1}-\delta+1$. Again by (13), $e_{i,j}$ for $1 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant \tau_i-\delta+1$ only appears in one of E_t for $1 \leqslant t \leqslant s$, i.e.,

$$(x-e_{i,j}) \Big| \prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta),$$

however,

$$(x-e_{i,j})^2 \not\Big| \Big(\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta)\Big),$$

for $1 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant \tau_i-\delta+1$. By (18), we have that $e_{i,j}$ is not a root of $m_{\lambda_{i,j}}(x)\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta) = 0$ for $1 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant \tau_i-\delta+1$. Thus, the polynomials $m_{\lambda_{i,j}}(x)\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta)$ for $1 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant \tau_i-\delta+1$ are linearly independent over $\mathbb{F}_q$. Therefore, $h(x)\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta) \equiv 0$ implies that the coefficients of $m_{\lambda_{i,j}}(x)\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta)$ for $1 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant \tau_i-\delta+1$ in $h(x)\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta)$ are 0. This is to say, the coefficient of $m_{\lambda_{s,\tau_s-\delta+1}}(x)\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta)$ in $h(x)\prod_{1\leqslant u\leqslant s}\prod_{\theta\in E_u}(x-\theta)$ is zero, i.e.,

$$\det\begin{pmatrix} m_{\lambda_{1,1},t_1} & m_{\lambda_{1,1},t_2} & \cdots & m_{\lambda_{1,1},t_{v_1-1}} \\ m_{\lambda_{1,2},t_1} & m_{\lambda_{1,2},t_2} & \cdots & m_{\lambda_{1,2},t_{v_1-1}} \\ \vdots & \vdots & \cdots & \vdots \\ m_{\lambda_{1,\tau_1-\delta+1},t_1} & m_{\lambda_{1,\tau_1-\delta+1},t_2} & \cdots & m_{\lambda_{1,\tau_1-\delta+1},t_{v_1-1}} \\ \vdots & \vdots & \cdots & \vdots \\ m_{\lambda_{s,\tau_s-\delta},t_1} & m_{\lambda_{s,\tau_s-\delta},t_2} & \cdots & m_{\lambda_{s,\tau_s-\delta},t_{v_1-1}} \end{pmatrix} = 0,$$

which is a contradiction with (16). Thus, the erasure pattern $E_1, E_2, \ldots, E_s$ is also recoverable.

For the second case of the induction step, assume $\ell = s+1 \leqslant \frac{d_1-1}{\delta}$ sets and $|E_{s+1}| = \delta$, when $T < d_1-\delta \leqslant r-v$. In this case, by a similar analysis, we have $s+1 \leqslant \mu$, and thus we also have

$$\{e_{i,t} : 1 \leqslant t \leqslant \tau_i-\delta+1\} \cap E_j = \phi \text{ for } 1 \leqslant i \neq j \leqslant s+1,$$

with $E_i = \{e_{i,t} : 1 \leqslant t \leqslant \tau_i\}$ for $1 \leqslant i \leqslant s+1$, and

$$\begin{aligned}&(\varpi_{1,1}, \ldots, \varpi_{1,\tau_1-\delta+1}, \ldots, \varpi_{s,\tau_s-\delta+1}, \varpi_{s+1,1})\, M_{s+1} \\ =&(\varpi_{1,1}, \ldots, \varpi_{1,\tau_1-\delta+1}, \ldots, \varpi_{s,\tau_s-\delta+1}, \varpi_{s+1,1}) \\ &\begin{pmatrix} m_{\lambda_{1,1},1} & m_{\lambda_{1,1},2} & \cdots & m_{\lambda_{1,1},r-v} \\ m_{\lambda_{1,2},1} & m_{\lambda_{1,2},2} & \cdots & m_{\lambda_{1,2},r-v} \\ \vdots & \vdots & \cdots & \vdots \\ m_{\lambda_{1,\tau_1-\delta+1},1} & m_{\lambda_{1,\tau_1-\delta+1},2} & \cdots & m_{\lambda_{1,\tau_1-\delta+1},r-v} \\ \vdots & \vdots & \cdots & \vdots \\ m_{\lambda_{s+1,1},1} & m_{\lambda_{s+1,1},2} & \cdots & m_{\lambda_{t_{s+1,1}},r-v} \end{pmatrix} \\ =&(w_1, w_2, \ldots, w_{r-v}),\end{aligned}$$

where $(w_1, w_2, \ldots, w_{r-v})$ is a constant vector determined by the accessible code symbols, $v_2 = \sum_{1\leqslant j\leqslant s+1}(\tau_i-\delta+1) \leqslant T+\delta-(s+1)(\delta-1) < r-v+1-s(\delta-1) \leqslant r-v$, and

$$m_{\lambda_{i,j},z} = \frac{1}{(\alpha_z-e_{i,j})\prod_{\tau_i-\delta+2\leqslant t\leqslant\tau_i}(\alpha_z-e_{i,t})}$$

for $1 \leqslant i \leqslant s+1$, $1 \leqslant j \leqslant \tau_i+\delta-1$, and $1 \leqslant z \leqslant r-v$. Again by the induction hypothesis, there should exists a $(v_2-1)\times(v_2-1)$ matrix with

$$\det\begin{pmatrix} m_{\lambda_{1,1},t_1} & m_{\lambda_{1,1},t_2} & \cdots & m_{\lambda_{1,1},t_{v_2-1}} \\ m_{\lambda_{1,2},t_1} & m_{\lambda_{1,2},t_2} & \cdots & m_{\lambda_{1,2},t_{v_2-1}} \\ \vdots & \vdots & \cdots & \vdots \\ m_{\lambda_{1,\tau_1-\delta+1},t_1} & m_{\lambda_{1,\tau_1-\delta+1},t_2} & \cdots & m_{\lambda_{1,\tau_1-\delta+1},t_{v_2-1}} \\ \vdots & \vdots & \cdots & \vdots \\ m_{\lambda_{s,\tau_s-\delta},t_1} & m_{\lambda_{s,\tau_s-\delta},t_2} & \cdots & m_{\lambda_{s,\tau_s-\delta},t_{v_2-1}} \end{pmatrix} \neq 0, \tag{19}$$

i.e., the erasure pattern $E_1, E_2, \ldots, E_s, (E_{s+1} \setminus \{e_{s+1,1}\})$ is recoverable. Here, $f_{s+1}(\theta)$ for $\theta \in E_{s+1} \setminus \{e_{s+1,1}\}$ is recovered by the (r,δ)-locality independently, since $|E_{s+1} \setminus \{e_{s+1,1}\}| = \delta-1$. If $E_1, E_2, \ldots, E_{s+1}$ is not recoverable, then all the $v_2 \times v_2$ sub-matrices of M_{s+1} are singular. Therefore, by the same analysis, the polynomials

$m_{\lambda_{i,j}}(x)\prod_{1\leqslant u\leqslant s+1}\prod_{\theta\in E_u}(x-\theta)$ for $1\leqslant i\leqslant s+1$ and $1\leqslant j\leqslant \tau_i-\delta+1$ are linearly independent over $\mathbb{F}_q$. This is also a contradiction with (19) and all the $v_2\times v_2$ submatrices of M_{s+1} are singular, by the same analysis as the previous case. Thus, the erasure pattern $E_1, E_2, \ldots, E_{s+1}$ is also recoverable.

Therefore, by mathematical induction, the distance of $\mathcal{C}$ satisfies $d\geqslant d_1$, which completes the proof. ■

Example 1: Let $r=2$, $\delta=3$, $w=3$, $n=w(r+\delta-1)=12$, $v=1$, and $k=(w-1)r+v=5$. Consider the linear code over $\mathbb{F}_7=\mathbb{Z}_7$. Set $S_1=\{1,2,3,4\}$, $S_2=\{3,4,5,6\}$ $S_3=\{1,2,5,6\}$, and $\alpha_1=0$. Then the generator matrix of the linear code $\mathcal{C}$ by Construction A can be listed as

$$\begin{aligned}G=&(\mathbf{g}_1,\mathbf{g}_2,\ldots,\mathbf{g}_{12})\\ =&\begin{pmatrix}1&0&0&0&0&6&5&0&0&5&6&4\\0&1&0&0&0&2&3&0&0&1&4&5\\0&0&1&0&0&0&0&6&5&3&5&1\\0&0&0&1&0&0&0&2&3&3&5&1\\0&0&0&0&1&0&0&0&0&2&5&6\end{pmatrix}.\end{aligned}$$

It is easy to check that $(\mathbf{g}_1,\mathbf{g}_2,\mathbf{g}_6,\mathbf{g}_7)$ is a generator matrix of a $[4,2,3]_7$ linear code (as are $(\mathbf{g}_3,\mathbf{g}_4,\mathbf{g}_8,\mathbf{g}_9)$ and $(\mathbf{g}_5,\mathbf{g}_{10},\mathbf{g}_{11},\mathbf{g}_{12})$). Thus, the code $\mathcal{C}$ has all symbol $(2,3)$-locality. A computer program verified that indeed the weight of the codewords generated by G is at least 4, i.e., $d=4=n-k+1-(\lceil\frac{k}{r}\rceil-1)(\delta-1)=r-v+\delta$. Thus, the code $\mathcal{C}$ generated by G is a $[12,5,4]_7$ optimal locally repairable codes with all symbol $(2,3)$-locality, which is consistent with the result of Theorem 3.

B. Explicit Locally Repairable Codes With $n>q$

According to the bound of Lemma 1, the minimal Hamming distance of the code $\mathcal{C}$ generated by Construction A, i.e, $n=w(r+\delta-1)$ and $k=(w-1)r+v$ for $0<v\leqslant r$, is at most $r-v+\delta$. In fact, the key point in applying Theorem 3 is to find sets $S_1,\ldots,S_w$ of evaluation points, that both allow optimal code construction with the minimal Hamming distance $d=r-v+\delta$ as well a long code. In this subsection, based on Construction A, we analyze special structures of $S_1,\ldots,S_w$ that can yield optimal locally repairable codes with $n>q$.

Two trivial optimal locally repairable codes with $n>q$

Corollary 4: Let $n=w(r+\delta-1)$, $k=(w-1)r+v$, $1\leqslant v\leqslant r$, be integers. If $r-v\leqslant\delta$ and $q\geqslant 2r+\delta-v-1$, then there exists an optimal $[n,k,d=r-v+\delta]_q$ linear code with all symbol (r,δ)-locality, where optimality is with respect to the bound in Lemma 1.

Proof: By Lemma 1, a code with the given n and k is optimal if $d=r-v+\delta$. Since $r-v\leqslant\delta$, in the proof of Theorem 3 we only need to consider the case that there is only one repair set containing strictly more than $\delta-1$ erasures, which easily holds. ■

Remark 3: We remark that in the case described in Corollary 4, we can let $S_i=S_j$ for $1\leqslant i\neq j\leqslant w$. In this case, $r-v\leqslant\delta$ and $q\geqslant 2r+\delta-v-1$ are sufficient for the code generated by Construction A to be optimal. This is to say, the value w is independent of q. Thus, the length $n=w(r+\delta-1)$ of the code $\mathcal{C}$ can be as long as we wish. This result is already known for the case $\delta=2$ and $d\leqslant 4$ (see [23]), and is, to the best of our knowledge, new for the case of $\delta>2$. This result also shows that the condition $2t+1>4$ is necessary for Theorem 2, since the code length is unbounded for the case $2t+1\leqslant 4$, i.e., $t\leqslant 1$ corresponding to the case $r-v\leqslant\delta$, where $t=\lfloor(d-1)/\delta\rfloor=\lfloor\frac{r+v+\delta-1}{\delta}\rfloor$.

Corollary 5: Let $n=w(r+\delta-1)$, $k=(w-1)r+v$, $1\leqslant v\leqslant r$, be integers. Let $S\subseteq\mathbb{F}_q\setminus\{\alpha_i\ :\ 1\leqslant i\leqslant r-v\}$, $|S|=\delta-1$, be a fixed subset. Take $S_i\subseteq\mathbb{F}_q\setminus\{\alpha_i\ :\ 1\leqslant i\leqslant r-v\}$ for $1\leqslant i\leqslant w$. If $S_i\cap S_j\subseteq S$ for $1\leqslant i\neq j\leqslant w$, then the code $\mathcal{C}$ generated by Construction A is an optimal $[n,k,d=r-v+\delta]_q$ linear code with all symbol (r,δ)-locality, where optimality is with respect to the bound in Lemma 1.

Corollary 6: Let $n=w(r+\delta-1)$, $k=(w-1)r+v$, $1\leqslant v\leqslant r$, be integers. If $q\geqslant(w+1)r+\delta-v-1$, then there exists an optimal $[n,k,d=r-v+\delta]_q$ linear code with all symbol (r,δ)-locality, where optimality is with respect to the bound in Lemma 1.

Proof: When $q\geqslant(w+1)r+\delta-v-1$, those S_i's in Corollary 5 can be easily constructed by letting $|S|=\delta-1$ and $S_i\cap S_j=S$ for all $1\leqslant i\neq j\leqslant w$, which form a sunflower with center S [9]. ■

Remark 4: When $w>1+\frac{r-v}{\delta-1}$, the optimal linear codes with all symbol (r,δ)-locality in Corollary 6 are all with $n>q$. In [21], optimal locally repairable codes are also constructed with flexible parameters. However, in [21] the construction is based on the so-called good polynomials [22], [34] and $n\leqslant q$.

Optimal locally repairable codes with $n>q$ based on union-intersection-bounded families

A combinatorial structure that captures the interaction between the evaluation-point sets, $S_1,\ldots,S_w$, in Construction A is a union-intersection-bounded family [12]. Its definition is now given:

Definition 3 ([12]): Let n_1,τ,δ,t,s be positive integers such that $n_1\geqslant\tau\geqslant 2$, $\tau\geqslant\delta$, and $t\geqslant s$. The $(s,t;\delta)$-union-intersection-bounded family (denoted by $(s,t;\delta)$-UIBF(τ,n_1)) is a pair $(\mathcal{X},\mathcal{S})$, where $\mathcal{X}$ is a set of n_1 elements (called points) and $\mathcal{S}\subseteq 2^{\mathcal{X}}$ is a collection of τ-subsets of $\mathcal{X}$ (called blocks), such that any $s+t$ distinct blocks $A_1,A_2,\ldots,A_s,B_1,B_2,\ldots,B_t\in\mathcal{S}$ satisfy

$$\left|\left(\bigcup_{1\leqslant i\leqslant s}A_i\right)\bigcap\left(\bigcup_{1\leqslant i\leqslant t}B_i\right)\right|<\delta.$$

Example 2: Let $n_1=6$, $\tau=4$, $\delta=3$, $s=t=1$, and $\mathcal{X}=[6]$. Then it is easy to check that the family of sets $\{S_1=\{1,2,3,4\},S_2=\{3,4,5,6\},S_3=\{1,2,5,6\}\}$ in Example 1 forms a $(1,1;3)$-UIBF$(3,6)$.

The following corollary follows from Theorem 3 and Lemma 1.

Corollary 7: Let $n=w(r+\delta-1)$, $k=(w-1)r+v$, $1\leqslant v\leqslant r$, be integers, and let μ be a positive integer with $\mu\delta\geqslant r-v$. If $(\mathbb{F}_q\setminus\{\alpha_t\ :\ 1\leqslant t\leqslant r-v\},\mathcal{S}=\{S_i\ :\ 1\leqslant i\leqslant w\})$ is a $(1,\mu-1;\delta)$-UIBF$(r+\delta-1,q-r+v)$, then the code $\mathcal{C}$ generated by Construction A is an optimal $[n,k,d=r-v+\delta]_q$ linear code with all symbol (r,δ)-locality, where optimality is with respect to the bound in Lemma 1.

Proof: By Definition 3, each μ-subset $\mathcal{R} \subseteq \mathcal{S}$ satisfies that for any $S' \in \mathcal{R}$,

$$\left| S' \bigcap \left(\bigcup_{S \in \mathcal{R} \setminus \{S'\}} S \right) \right| < \delta.$$

By Lemma 1 we have $d \leqslant r - v + \delta$. Thus, the desired conclusion follows from Theorem 3 and Lemma 1. ■

In [12], a lower bound on the size of $(1, \mu-1; \delta)$-UIBF$(r+\delta-1, q)$ is given, which immediately implies a lower bound on the length of the codes generated by Construction A according to Corollary 7.

Lemma 3 ([12]): Let μ, δ, r, n_1 be positive integers. Then there exists a $(1, \mu-1; \delta)$-UIBF$(r+\delta-1, n_1)$ $(\mathcal{X}, \mathcal{S})$ with $|\mathcal{S}| = \Omega(n_1{}^{\frac{\delta}{\mu-1}})$, where r, δ, μ are regarded as constants.

Based on Corollary 7 and Lemma 3, we have the following:

Corollary 8: Let $n = w(r+\delta-1)$, $k = (w-1)r + v$, $1 \leqslant v \leqslant r$, be integers, and let μ be a positive integer with $\mu\delta \geqslant r - v$. Then Construction A can generate an optimal (with respect to the bound in Lemma 1) $[n, k, d = r - v + \delta]_q$ linear code $\mathcal{C}$ with all symbol (r, δ)-locality and length $n = \Omega(q^{\frac{\delta}{\mu-1}})$, where we regard r, δ, and μ as constants.

Optimal locally repairable codes with $n > q$ based on packings or Steiner systems

In the following, we consider some special sufficient conditions for (12) to construct optimal linear codes with all symbol (r, δ)-locality.

Theorem 4: Let $n = w(r+\delta-1)$, $k = (w-1)r + v$, $1 \leqslant v \leqslant r$, be integers, and let a be a positive integer. If $|S_i \cap S_j| \leqslant a$ for $1 \leqslant i \neq j \leqslant w$ and $r - v \leqslant \frac{\delta^2}{a}$, then the code $\mathcal{C}$ generated by Construction A is an optimal $[n, k, d = r - v + \delta]_q$ linear code with all symbol (r, δ)-locality, where optimality is with respect to the bound in Lemma 1.

Proof: Denote $\mathcal{S} = \{S_1, \ldots, S_w\}$, and let $\mu = \lceil \frac{\delta}{a} \rceil$. Then the fact that $|S_i \cap S_j| \leqslant a$ means that for any μ-subset, $\mathcal{R} \subseteq \mathcal{S}$, and for any $S' \in \mathcal{R}$, we have

$$\left| S' \cap \left(\bigcup_{S \in \mathcal{R} \setminus \{S'\}} S \right) \right| \leqslant (\mu - 1)a = \left(\left\lceil \frac{\delta}{a} \right\rceil - 1 \right) a \leqslant \delta - 1.$$

Since $\mu\delta \geqslant \frac{\delta^2}{a} \geqslant r - v$, the conclusion follows by Theorem 3. ■

Definition 4 ([8], VI. 40): Let $n_1 \geqslant 2$ be an integer and u a positive integer. A τ-$(n_1, t, 1)$-*packing* is a pair $(\mathcal{X}, \mathcal{S})$, where $\mathcal{X}$ is a set of n_1 elements (called points) and $\mathcal{S} \subseteq 2^{\mathcal{X}}$ is a collection of t-subsets of $\mathcal{X}$ (called blocks), such that each τ-subset of $\mathcal{X}$ is contained in at most one block of $\mathcal{S}$. Furthermore, if each τ-subset of $\mathcal{X}$ is contained in exactly one block of $\mathcal{S}$, then $(\mathcal{X}, \mathcal{S})$ is also called a (τ, t, n_1)-*Steiner system*.

The following corollary follows directly from Theorem 4.

Corollary 9: Let $n_1 = q - r + v$. If there exists a $(\tau+1)$-$(n_1, r+\delta-1, 1)$-packing with blocks $\mathcal{S}$ and $0 \leqslant r - v \leqslant \frac{\delta^2}{\tau}$, then there exists an optimal $[n, k, d]_q$ linear code with all symbol (r, δ)-locality, where $n = |\mathcal{S}|(r+\delta-1)$, $k = (|\mathcal{S}|-1)r + v$, and $d = r - v + \delta$.

The number of blocks of a packing is upper bounded by the following Johnson bound [18]:

Lemma 4 ([18]): The maximum possible number of blocks of a $(\tau+1)$-$(n_1, r+\delta-1, 1)$-packing $\mathcal{S}$ is bounded by

$$|\mathcal{S}| \leqslant \left\lfloor \frac{n_1}{r+\delta-1} \left\lfloor \frac{n_1 - 1}{r+\delta-2} \cdots \left\lfloor \frac{n_1 - \tau}{r+\delta-1-\tau} \right\rfloor \cdots \right\rfloor \right\rfloor.$$

Thus, the number of blocks for a $(\tau+1)$-$(n_1, r+\delta-1, 1)$-packing can be as large as $O(n_1^{\tau+1})$, when τ, r, and δ are regarded as constants.

Corollary 10: Let $n_1 = q - r + v$. If there exists a $(\tau+1)$-$(n_1, r+\delta-1, 1)$-packing with blocks $\mathcal{S}$, $|\mathcal{S}| = O(n_1^{\tau+1})$, and $0 \leqslant r - v \leqslant \frac{\delta^2}{\tau}$, then there exists an optimal $[n, k, d]_q$ linear code with all symbol (r, δ)-locality, where $n = w(r+\delta-1) = |\mathcal{S}|(r+\delta-1) = O(q^{\tau+1})$, $k = (|\mathcal{S}|-1)r+v$ and $d = r-v+\delta$. In particular, for the case $w - 1 \geqslant 2(r-v+1)$, $r - v = \delta+1$, i.e., $d = 2\delta+1$ and $\tau = \delta-1$, the code based on the $(\tau+1)$-$(n_1, r+\delta-1, 1)$-packing has asymptotically optimal length, where r and δ are regarded as constants.

Proof: By Corollary 9, we have $n = |\mathcal{S}|(r+\delta-1) = O(q^{\tau+1})$ for the code generated by Construction A. For the case $r - v = \delta+1$, $w - 1 \geqslant 2(r-v+1)$, $d = 2\delta+1$, and $t = \lfloor (d-1)/\delta \rfloor = 2$, by Theorem 2 we have

$$\begin{aligned} n &\leqslant \frac{t(r+\delta-1)}{2r(q-1)} q^{\frac{2(w-w+1)r-2v}{t}} \\ &\leqslant \frac{t(r+\delta-1)}{2r(q-1)} q^{r-v} = O(q^{r-v-1}). \end{aligned}$$

Thus, for the case $r - v = \delta+1$ and $\tau = \delta-1$, the code $\mathcal{C}$ has length $n = O(q^{\tau+1}) = O(q^{\delta})$, which is asymptotically optimal with respect to the bound in Theorem 2, when r and δ are regarded as constants. ■

As an example, we also analyze the length of the codes based on Steiner systems.

Corollary 11: Let $n_1 = q - r + v$. If there exists a $(\tau+1, r+\delta-1, n_1)$-Steiner system and $0 \leqslant r - v \leqslant \frac{\delta^2}{\tau}$, then there exists an optimal $[n, k, d]_q$ linear code with all symbol (r, δ)-locality, where

$$n = w(r+\delta-1) = \frac{\binom{n_1}{\tau+1}(r+\delta-1)}{\binom{r+\delta-1}{\tau+1}},$$

$$k = \left(\frac{\binom{n_1}{\tau+1}}{\binom{r+\delta-1}{\tau+1}} - 1 \right) r + v,$$

and $d = r-v+\delta$. In particular, for the case $w-1 \geqslant 2(r-v+1)$, $r-v = \delta+1$, i.e., $d = 2\delta+1$ and $\tau = \delta-1$, the code based on the $(\delta, r+\delta-1, q-\delta-1)$-Steiner system has asymptotically optimal length, where r and δ are regarded as constants.

Proof: The first part of the corollary follows directly from Corollary 9 and Definition 4. For the second part, the fact $\tau = \delta - 1$ means that $r - v = \delta + 1 < \frac{\delta^2}{\delta-1}$ is possible, which also means the code $\mathcal{C}$ has length $(r+\delta-1)\binom{q-\delta+1}{\delta}/\binom{r+\delta-1}{\delta}$ and $d = 2\delta+1$. Since $w - 1 \geqslant 2(r-v+1)$, $u = w - 1$, $r - v = \delta - 1$, and $d = 2\delta+1$, i.e., $t = 2$, by Theorem 2, we have

$$n \leqslant \frac{t(r+\delta-1)}{2r(q-1)} q^{\frac{2(w-u)r-2v}{t}} \leqslant \frac{t(r+\delta-1)}{2r(q-1)} q^{r-v} = O(q^{\delta}).$$

Now the conclusion comes from the fact that the upper bound is $O(q^{\delta})$ and the constructed code has length $n = \Omega(q^{\delta})$, where we assume r and δ are constants. ■

Remark 5: For the existence of packings in general the reader may refer to [29] and the survey in [8, VI.40].

Remark 6: For the case $\delta = 2$ and $d = 5$, optimal linear codes with all symbol $(r, 2)$-locality and asymptotically optimal length $\Theta(q^2)$ have been introduced in [3], [13], [17]. The constructions in [3], [17] are given by parity-check matrices with 3 or 4 global parity checks, which means they only works for the cases $d = 5, 6$. It is easy to check that our construction yields codes for more general cases even if we only consider the case $\delta = 2$.

Given positive integers τ, r and $\delta > 2$, the natural necessary conditions for the existence of a $(\tau + 1, r + \delta - 1, q - r + v)$-Steiner system are that $\binom{q-r+v-i}{\tau+1-i} | \binom{r+\delta-1-i}{\tau+1-i}$ for all $0 \leqslant i \leqslant \tau$. It was shown in [19] that these conditions are also sufficient except perhaps for finitely many cases. While q might not be a prime power, any prime power $\overline{q} \geqslant q$ will suffice for our needs. It is known, for example, that there is always a prime in the interval $[q, q + q^{21/40}]$ (see [2]). Thus, Construction A provides infinitely many optimal linear $[n, k, d]_{\overline{q}}$ locally repairable codes, with all symbol (r, δ)-locality, and

$$\begin{aligned} n &= (r + \delta - 1) \cdot \frac{\binom{q-r+v}{\tau+1}}{\binom{r+\delta-1}{\tau+1}} = \Omega(q^{\tau+1}) = \Omega(\overline{q}^{\tau+1}), \\ k &= \left(\frac{\binom{q-r+v}{\tau+1}}{\binom{r+\delta-1}{\tau+1}} - 1 \right) r + v, \\ d &= r - v + \delta, \end{aligned}$$

i.e., with length super-linear in the field size. This is to say,

Corollary 12: For given integers r, δ, τ with $0 \leqslant r - v \leqslant \frac{\delta^2}{\tau}$, let t be an integer with $\binom{t-r+v-i}{\tau+1-i} | \binom{r+\delta-1-i}{\tau+1-i}$ for all $0 \leqslant i \leqslant \tau$. Then, for all large enough t, there exists an optimal $[n, k, d]_q$ locally repairable code, with all symbol (r, δ)-locality, and

$$\begin{aligned} n &= (r + \delta - 1) \cdot \frac{\binom{t-r+v}{\tau+1}}{\binom{r+\delta-1}{\tau+1}} = \Omega(t^{\tau+1}) = \Omega(q^{\tau+1}), \\ k &= \left(\frac{\binom{t-r+v}{\tau+1}}{\binom{r+\delta-1}{\tau+1}} - 1 \right) r + v, \\ d &= r - v + \delta, \end{aligned}$$

where q is a prime power with $t \leqslant q \leqslant t + t^{21/40}$.

V. Concluding Remarks

In this paper, we first derived an upper bound for the length of optimal locally repairable codes when $\delta > 2$. As a byproduct, we also extended the range of parameters for the known bound (the case $\delta = 2$) and improve its performance for the case $d > r + \delta$. A general construction of locally repairable codes was introduced. By the construction, locally repairable codes with length super-linear in the field size can be generated. In particular, for some cases those codes have asymptotically optimal length with respect to the new bound.

Several combinatorial structures, e.g., union-intersection-bounded families, packings, and Steiner systems, satisfy (12) and play a key role in determining the length of the codes generated by Construction A. If more of those structures with a large number of blocks can be constructed, more good codes with length $n > q$ can be generated. Finding more such combinatorial structures and explicit constructions for them, is left for future research.

Appendix

The goal of this appendix is to prove Theorem 1. In Lemmas 5 and 6, we first characterize how many code symbols should have more than one repair set if the repair sets do not form a partition of all the code symbols. Following that, in Lemmas 7-10, we prove a relationship between the rank, the number of repair sets, and the number of code symbols that have more than one repair set. Finally, we prove, under some restrictions, that if the repair sets do not form a partition of all the code symbols then the Singleton bound in Lemma 1 cannot hold with equality, namely, Theorem 1 holds.

To prove Theorem 1, we need some basic combinatorial covering designs and property of repair sets. We begin with some notation and definitions. Recall that

$$n = (r + \delta - 1)w + m, \qquad k = ru + v,$$

where $\delta \geqslant 2$, $0 \leqslant m \leqslant r + \delta - 2$, and $0 \leqslant v \leqslant r - 1$ are all integers.

Definition 5: Let $n, T, s \in \mathbb{N}$. Also, let $\mathcal{X}$ be a set of cardinality n, whose elements are called *points*. Finally, let $\mathcal{B} = \{B_1, B_2, \ldots, B_T\} \subseteq 2^{\mathcal{X}}$ be a set of *blocks* such that $\bigcup_{i \in [T]} B_i = \mathcal{X}$, and for all $i \in [T]$, $|B_i| \leqslant s$ 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 are the same size we say $(\mathcal{X}, \mathcal{B})$ is a *uniform* (n, T, s)-ECF.

There are some similarities between ECFs and covering designs. Recall that in an (n, k, t)-covering design we have n points, all blocks are of size k, and every t-set of points is contained in at least one block. An ECF is more relaxed, in the sense that blocks need not be of the same size. On the other hand, coverage is tested only for single elements, i.e., $t = 1$, and each block must contain a unique element not found in other blocks. For example $\{\{2, 4, 1\}, \{3, 5, 2\}, \{0, 6\}\}$ is an ECF. In contrast, the well known 2-design with parameters $(7, 3, 1)$, i.e.,$\{\{2, 4, 1\}, \{3, 5, 2\}, \{4, 6, 3\}, \{5, 0, 4\}, \{6, 1, 5\}, \{0, 2, 6\}, \{1, 3, 0\}\}$ is not an ECF for the simple reason that each element appears exactly in 3 sets. This is to say for any given set we can not find an element only included in that set.

An important quantity associated with any family of subsets, $\mathcal{B} \subseteq 2^{\mathcal{X}}$, is its *overlap*, denoted $D(\mathcal{B})$, and defined as

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

Obviously $D(\mathcal{B}) \geqslant 0$ and $D(\mathcal{B})$ is monotonically increasing. Additionally, $D(\mathcal{B}) = 0$ if and only if its sets are pairwise disjoint.

In what follows we investigate the structures of repair sets.

Lemma 5: Let $(\mathcal{X},\mathcal{B})$ be an $(n,T,r+\delta-1)$-ECF, and assume it is non-uniform or that $D(\mathcal{B})\neq 0$. Then for every $0\leqslant t\leqslant T$, there exists a subset $\mathcal{B}'\subseteq\mathcal{B}$, $|\mathcal{B}'|=t$, such that

$$t(r+\delta-1)-\left|\bigcup_{B\in\mathcal{B}'}B\right|\geqslant\min\{r+\delta-1-m,\lfloor t/2\rfloor\}.$$

Proof: We first construct a uniform $\overline{\mathcal{B}}$ from $\mathcal{B}$, by arbitrarily adding elements to sets in $\mathcal{B}$ that contain less than $r+\delta-1$ elements. Note that $\overline{\mathcal{B}}$ is not necessarily an ECF. Obviously $D(\overline{\mathcal{B}})\geqslant D(\mathcal{B})$. We contend now that $D(\overline{\mathcal{B}})>0$. If $D(\mathcal{B})\neq 0$ this is immediate, since we have $D(\overline{\mathcal{B}})\geqslant D(\mathcal{B})>0$. If $\mathcal{B}$ is not uniform, at least one set $B\in\mathcal{B}$ has $|B|<r+\delta-1$, and adding elements to it in the process of creating $\overline{\mathcal{B}}$ necessarily increases the overlap, i.e., $D(\overline{\mathcal{B}})>D(\mathcal{B})\geqslant 0$. We also observe that,

$$\begin{aligned}D(\overline{\mathcal{B}})&=\sum_{\overline{B}\in\overline{\mathcal{B}}}|\overline{B}|-\left|\bigcup_{\overline{B}\in\overline{\mathcal{B}}}\overline{B}\right|\\&=|\overline{\mathcal{B}}|(r+\delta-1)-n\\&\equiv-m\pmod{r+\delta-1}.\end{aligned}$$

Next, we partition $\overline{\mathcal{B}}$ into two subsets, $\overline{\mathcal{B}}_1$ and $\overline{\mathcal{B}}_2$, where

$$\overline{\mathcal{B}}_1=\{\overline{B}\in\overline{\mathcal{B}}\ :\ \exists\overline{B}'\in\overline{\mathcal{B}},\overline{B}'\neq\overline{B},\overline{B}\cap\overline{B}'\neq\emptyset\},$$

and

$$\overline{\mathcal{B}}_2=\overline{\mathcal{B}}\setminus\overline{\mathcal{B}}_1.$$

For convenience, denote $\overline{\mathcal{B}}_1=\{\overline{B}_1,\ldots,\overline{B}_K\}$ and $\overline{\mathcal{B}}_2=\{\overline{B}_{K+1},\ldots,\overline{B}_T\}$ where $0\leqslant K\leqslant T$.

Let $1\leqslant t\leqslant T$ be a positive integer. Obviously, if $t\geqslant K$, then $\overline{\mathcal{B}}'=\{\overline{B}_1,\ldots,\overline{B}_K,\ldots,\overline{B}_t\}$ is a t-subset satisfying

$$D(\overline{\mathcal{B}}')=\sum_{i=1}^{t}|\overline{B}_i|-\left|\bigcup_{i=1}^{t}\overline{B}_i\right|=D(\overline{\mathcal{B}}).\tag{20}$$

For the case $0\leqslant t\leqslant 1$, the fact $\lfloor t/2\rfloor=0$ means that the lemma follows trivially. For the case $2\leqslant t<K$, we claim that we can select a t-subset $\overline{\mathcal{B}}'\subseteq\overline{\mathcal{B}}_1$ containing $\lfloor t/2\rfloor$ different pairs of sets $\{\overline{B}_{\tau_{2i-1}},\overline{B}_{\tau_{2i}}\}$ for $1\leqslant i\leqslant\lfloor t/2\rfloor$ with

$$\begin{aligned}\sum_{\overline{B}\in\mathcal{B}_j}|\overline{B}|-\left|\bigcup_{\overline{B}\in\mathcal{B}_j}\overline{B}\right|&\geqslant 1+\sum_{\overline{B}\in\mathcal{B}_{j-1}}|\overline{B}|-\left|\bigcup_{\overline{B}\in\mathcal{B}_{j-1}}\overline{B}\right|\\&\geqslant j,\end{aligned}$$

for $\mathcal{B}_0=\emptyset$ and $\mathcal{B}_j=\{\overline{B}_{\tau_i}:1\leqslant i\leqslant 2j\}$, $1\leqslant j\leqslant\lfloor\frac{t}{2}\rfloor$, especially $\overline{\mathcal{B}}'\supseteq\mathcal{B}_{\lfloor\frac{t}{2}\rfloor}$ satisfying

$$\begin{aligned}\sum_{\overline{B}\in\overline{\mathcal{B}}'}|\overline{B}|-\left|\bigcup_{\overline{B}\in\overline{\mathcal{B}}'}\overline{B}\right|&\geqslant\sum_{\overline{B}\in\mathcal{B}_{\lfloor\frac{t}{2}\rfloor}}|\overline{B}|-\left|\bigcup_{\overline{B}\in\mathcal{B}_{\lfloor\frac{t}{2}\rfloor}}\overline{B}\right|\\&\geqslant\left\lfloor\frac{t}{2}\right\rfloor.\end{aligned}\tag{21}$$

Otherwise, there exists a subset $\overline{\mathcal{B}}_1^*\subseteq\overline{\mathcal{B}}_1$ with size at most $2(\lfloor\frac{t}{2}\rfloor-1)$ such that for any $\overline{B}'\in\overline{\mathcal{B}}_1\setminus\overline{\mathcal{B}}_1^*,\overline{B}''\in\overline{\mathcal{B}}_1$,

$$\begin{aligned}&\sum_{\overline{B}\in\overline{\mathcal{B}}_1^*\cup\{\overline{B}',\overline{B}''\}}|B|-\left|\bigcup_{\overline{B}\in\overline{\mathcal{B}}_1^*\cup\{\overline{B}',\overline{B}''\}}\overline{B}\right|\\&\leqslant\sum_{\overline{B}\in\overline{\mathcal{B}}_1^*}|\overline{B}|-\left|\bigcup_{\overline{B}\in\overline{\mathcal{B}}_1^*}\overline{B}\right|,\end{aligned}$$

which implies

$$\begin{cases}|\overline{B}'|+|\overline{B}''|\leqslant\left|(\overline{B}'\cup\overline{B}'')\setminus\bigcup_{\overline{B}\in\overline{\mathcal{B}}_1^*}\overline{B}\right|, & \text{if } \overline{B}''\in\overline{\mathcal{B}}_1\setminus\overline{\mathcal{B}}_1^*,\\|\overline{B}'|\leqslant\left|\overline{B}'\setminus\bigcup_{\overline{B}\in\overline{\mathcal{B}}_1^*}\overline{B}\right|, & \text{if } \overline{B}''\in\overline{\mathcal{B}}_1^*.\end{cases}$$

However, this means that every $\overline{B}'\in\overline{\mathcal{B}}_1\setminus\overline{\mathcal{B}}_1^*$ has an empty intersection with any other set in $\overline{\mathcal{B}}_1$, which contradicts the definition of $\overline{\mathcal{B}}_1$.

By combining (20) and (21), for any given $0\leqslant t\leqslant|\mathcal{B}|$, there exists a t-subset, say $\overline{\mathcal{B}}'=\{\overline{B}_1,\overline{B}_2,\ldots,\overline{B}_t\}\subseteq\overline{\mathcal{B}}$, such that

$$\begin{aligned}D(\overline{\mathcal{B}}')&=\sum_{\overline{B}\in\overline{\mathcal{B}}'}|\overline{B}|-\left|\bigcup_{\overline{B}\in\overline{\mathcal{B}}'}\overline{B}\right|\\&\geqslant\min\left\{D(\overline{\mathcal{B}}),\lfloor t/2\rfloor\right\}\\&\geqslant\min\{r+\delta-1-m,\lfloor t/2\rfloor\},\end{aligned}\tag{22}$$

where the last inequality holds since $D(\overline{\mathcal{B}})>0$ and $D(\overline{\mathcal{B}})\equiv-m\pmod{r+\delta-1}$.

If $\overline{B}_i\in\overline{\mathcal{B}}'$ was created from $B_i\in\mathcal{B}$, i.e., $B_i\subseteq\overline{B}_i$, then by (22) we have,

$$\begin{aligned}t(r+\delta-1)-\left|\bigcup_{i=1}^{t}B_i\right|&=\sum_{i=1}^{t}|\overline{B}_i|-\left|\bigcup_{i=1}^{t}B_i\right|\\&\geqslant\sum_{i=1}^{t}|\overline{B}_i|-\left|\bigcup_{i=1}^{t}\overline{B}_i\right|\\&\geqslant\min\{r+\delta-1-m,\lfloor t/2\rfloor\}.\end{aligned}$$

Now set $\mathcal{B}'=\{B_1,\ldots,B_t\}$ to complete the proof. ■

Lemma 6: For any $[n,k]_q$ linear code $\mathcal{C}$ with all symbol (r,δ)-locality, let $\Gamma\subseteq 2^{[n]}$ be the set of all possible (r,δ)-repair sets. Then we can find a subset $\mathcal{R}\subseteq\Gamma$ such that $([n],\mathcal{R})$ is an $(n,|\mathcal{R}|,r+\delta-1)$-ECF with $|\mathcal{R}|\geqslant\lceil\frac{k}{r}\rceil$.

Proof: By Definition 2, Γ contains at least one repair set for each code symbol, hence

$$\bigcup_{R\in\Gamma}R=[n].\tag{23}$$

If for each $R\in\Gamma$, $R\not\subseteq\bigcup_{R'\in\Gamma\setminus\{R\}}R'$, then set $\mathcal{R}=\Gamma$ and the lemma follows. Otherwise, set $\Gamma_1=\Gamma\setminus\{R\}$, where $R\in\Gamma$ satisfies that $R\subseteq\bigcup_{R'\in\Gamma\setminus\{R\}}R'$. Thus, by (23), we conclude that

$$\bigcup_{R'\in\Gamma\setminus\{R\}}R'=[n].$$

Since $|\Gamma_1|<|\Gamma|$, and Γ_1 also satisfies (23), we can repeat the elimination procedure to obtain the desired set $\mathcal{R}$. The facts

$\mathrm{Rank}(\bigcup_{R\in\mathcal{R}} R) = k$ and $\mathrm{Rank}(R) \leqslant r$ imply that $|\mathcal{R}| \geqslant \lceil \frac{k}{r} \rceil$, which completes the proof. ∎

Lemma 7: Let $\mathcal{C}$ be an $[n,k]_q$ linear code with all symbol (r,δ)-locality. Let $\mathcal{R}$ be the ECF given by Lemma 6. If for a subset $\mathcal{V} \subseteq \mathcal{R}$, and for all $R' \in \mathcal{V}$,

$$\left| R' \bigcap \left(\bigcup_{R\in\mathcal{V}\setminus\{R'\}} R \right) \right| \leqslant |R'| - \delta + 1, \tag{24}$$

then we have

$$\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}} R\right) \leqslant \left|\bigcup_{R\in\mathcal{V}} R\right| - |\mathcal{V}|(\delta-1).$$

Proof: Denote $|\mathcal{V}| = \ell$ and $\mathcal{V} = \{R_1,\dots,R_\ell\} \subseteq \mathcal{R}$. For each $R_i \in \mathcal{V}$, (24) means that there exists a $(\delta-1)$-subset $R_i' \subseteq R_i$ such that $R_i' \cap (\bigcup_{j\in[\ell]\setminus\{i\}} R_j) = \emptyset$. Thus, we can get ℓ pairwise disjoint subsets $R_1', R_2', \dots, R_\ell'$.

By Definition 2, $\mathrm{Rank}(R_i) = \mathrm{Rank}(R_i \setminus R_i')$ for $1 \leqslant i \leqslant \ell$. Therefore, we have

$$\begin{aligned}\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}} R\right) &= \mathrm{Rank}\left(\bigcup_{i\in[\ell]} (R_i\setminus R_i')\right)\\ &\leqslant \left|\bigcup_{i\in[\ell]} (R_i\setminus R_i')\right|\\ &= \left|\bigcup_{R\in\mathcal{V}} R\right| - \sum_{i\in[\ell]} |R_i'|\\ &= \left|\bigcup_{R\in\mathcal{V}} R\right| - |\mathcal{V}|(\delta-1).\end{aligned}$$

∎

We note that when $\delta = 2$, (24) is always satisfied by the ECF $\mathcal{R}$. We now continue with our exploration of the properties of $\mathcal{R}$.

Lemma 8: Let $\mathcal{C}$ be an $[n,k]_q$ linear code with all symbol (r,δ)-locality. Let $\mathcal{R}$ be the ECF given by Lemma 6. If there are subsets $\mathcal{V} \subseteq \mathcal{R}' \subseteq \mathcal{R}$ with $|\mathcal{V}| \leqslant \lceil \frac{k}{r} \rceil - 1$, $\mathrm{Rank}(\bigcup_{R\in\mathcal{R}'} R) = k$, and

$$\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}} R\right) \leqslant \left|\bigcup_{R\in\mathcal{V}} R\right| - |\mathcal{V}|(\delta-1) \tag{25}$$

then we can obtain a $(\lceil \frac{k}{r} \rceil - 1)$-set $\mathcal{V}'$ with $\mathcal{V} \subseteq \mathcal{V}' \subseteq \mathcal{R}'$ such that

$$\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}'} R\right) \leqslant \left|\bigcup_{R\in\mathcal{V}'} R\right| - |\mathcal{V}'|(\delta-1).$$

Proof: If $|\mathcal{V}| = \lceil \frac{k}{r} \rceil - 1$, then the lemma follows by setting $\mathcal{V}' = \mathcal{V}$. Otherwise, we have $|\mathcal{V}| < \lceil \frac{k}{r} \rceil - 1$. Since every $R \in \mathcal{R}$ is an (r,δ)-repair set, $\mathrm{Rank}(R) \leqslant r$. This means that $\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}} R\right) < (\lceil \frac{k}{r} \rceil - 1)r < k$. Note that by the lemma requirements, $\mathrm{Rank}\left(\bigcup_{R\in\mathcal{R}'} R\right) = k$, which implies that there exists a $R' \in \mathcal{R}' \setminus \mathcal{V}$ such that $\mathrm{Rank}(R' \cup (\bigcup_{R\in\mathcal{V}} R)) > \mathrm{Rank}(\bigcup_{R\in\mathcal{V}} R)$. We recall, however, that since R' is an (r,δ)-repair set, if $R^* \subseteq R'$, $|R^*| \geqslant |R'| - \delta + 1$, then $\mathrm{Span}(R^*) = \mathrm{Span}(R')$. It follows that R' cannot have a large intersection with $\bigcup_{R\in\mathcal{V}} R$, namely,

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

Hence, there exists a $R'' \subseteq R' \setminus \left(\bigcup_{R\in\mathcal{V}} R\right)$ with $|R''| = \delta - 1$. Again, using the fact that R' is an (r,δ)-repair set and $|R' \setminus R''| = |R'| - \delta + 1$, we have $\mathrm{Rank}(R') = \mathrm{Rank}(R' \setminus R'')$, and therefore,

$$\begin{aligned}&\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}\cup\{R'\}} R\right)\\ &= \mathrm{Rank}\left(\left(\bigcup_{R\in\mathcal{V}\cup\{R'\}} R\right) \setminus R''\right)\\ &\leqslant \left| R' \setminus \left(\left(\bigcup_{R\in\mathcal{V}} R\right) \cup R''\right)\right| + \mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}} R\right)\\ &\leqslant \left| R' \setminus \left(\bigcup_{R\in\mathcal{V}} R\right)\right| - \delta + 1 + \left|\bigcup_{R\in\mathcal{V}} R\right| - |\mathcal{V}|(\delta-1)\\ &= \left|\bigcup_{R\in\mathcal{V}\cup\{R'\}} R\right| - |\mathcal{V}\cup\{R'\}|(\delta-1),\end{aligned}$$

where the last inequality holds by the fact $R'' \subseteq R' \setminus \left(\bigcup_{R\in\mathcal{V}} R\right)$ and (25). Therefore, repeating the above operations, we can extend $\mathcal{V}$ to a $(\lceil \frac{k}{r} \rceil - 1)$-subset $\mathcal{V}' \subseteq \mathcal{R}'$ such that

$$\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}'} R\right) \leqslant \left|\bigcup_{R\in\mathcal{V}'} R\right| - |\mathcal{V}'|(\delta-1).$$

∎

Lemma 9: Let $\mathcal{C}$ be an $[n,k]_q$ linear code with all symbol (r,δ)-locality. Let $\mathcal{R}$ be the ECF given by Lemma 6. Assume $\mathcal{V} \subseteq \mathcal{R}$ such that $|\mathcal{V}| \leqslant \lceil \frac{k}{r} \rceil - 1$. If there exists a $R' \in \mathcal{V}$ such that

$$\left| R' \bigcap \left(\bigcup_{R\in\mathcal{V}\setminus\{R'\}} R \right) \right| > |R'| - \delta + 1, \tag{26}$$

then there exists $S \subseteq [n]$ with $\mathrm{Rank}(S) = k-1$ and

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

Proof: Assume $\mathcal{V}$ satisfies (26). Let $\mathcal{V}' \subseteq \mathcal{V}$ be a minimal subset for which (26) holds, i.e., there exists a set $R' \in \mathcal{V}'$ with $|R' \cap (\bigcup_{R\in\mathcal{V}'\setminus\{R'\}} R)| > |R'| - \delta + 1$, which in turn implies that $\mathrm{Span}(R') \subseteq \mathrm{Span}(\bigcup_{R\in\mathcal{V}'\setminus\{R'\}} R)$. By the minimality of $\mathcal{V}'$, the set $\mathcal{V}' \setminus \{R'\}$ satisfies the requirements of Lemma 7, which implies

$$\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}'\setminus\{R'\}} R\right) \leqslant \left|\bigcup_{R\in\mathcal{V}'\setminus\{R'\}} R\right| - |\mathcal{V}'\setminus\{R'\}|(\delta-1).$$

As noted before, $\mathrm{Span}(R') \subseteq \mathrm{Span}(\bigcup_{R\in\mathcal{V}'\setminus\{R'\}} R)$, and since trivially $\mathrm{Rank}(\bigcup_{R\in\mathcal{R}} R) = k$, we also necessarily have $\mathrm{Rank}(\bigcup_{R\in\mathcal{R}\setminus\{R'\}} R) = k$. Therefore, by Lemma 8, we can extend $\mathcal{V}'\setminus\{R'\}$ to a $(\lceil\frac{k}{r}\rceil-1)$-subset $\mathcal{V}''\subseteq\mathcal{R}\setminus\{R'\}$ such that

$$\begin{aligned}\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}''}R\right)&\leqslant\left|\bigcup_{R\in\mathcal{V}''}R\right|-|\mathcal{V}''|(\delta-1)\\&=\left|\bigcup_{R\in\mathcal{V}''}R\right|-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1).\end{aligned}$$

Considering the set $\mathcal{V}''\cup\{R'\}$, we have

$$\begin{aligned}&\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}''\cup\{R'\}}R\right)\\&=\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}''}R\right)\\&\leqslant\left|\bigcup_{R\in\mathcal{V}''}R\right|-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1)\\&\leqslant\left|\bigcup_{R\in\mathcal{V}''\cup\{R'\}}R\right|-1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1),\end{aligned}\tag{27}$$

where the last inequality holds due to the fact that $R'\not\subseteq\bigcup_{R\in\mathcal{V}''}R$ by the properties of the ECF $\mathcal{R}$.

Since

$$\begin{aligned}\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}''\cup\{R'\}}R\right)&=\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}''}R\right)\\&\leqslant\left(\left\lceil\frac{k}{r}\right\rceil-1\right)r\\&\leqslant k-1,\end{aligned}$$

we can find a set S with $\mathrm{Rank}(S)=k-1$ by taking $\bigcup_{R\in\mathcal{V}''\cup\{R'\}}R$ and adding arbitrary coordinates until reaching the desired rank. This set S has size

$$\begin{aligned}|S|&\geqslant k-1-\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}''\cup\{R'\}}R\right)+\left|\bigcup_{R\in\mathcal{V}''\cup\{R'\}}R\right|\\&\geqslant k+\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1),\end{aligned}$$

which follows from (27). ■

Lemma 10: Let $\mathcal{C}$ be an $[n,k]_q$ linear code with all symbol (r,δ)-locality. Let $\mathcal{R}$ be the ECF given by Lemma 6. Assume $\mathcal{V}\subseteq\mathcal{R}$ such that $|\mathcal{V}|\leqslant\lceil\frac{k}{r}\rceil-1$. If Δ is an integer such that

$$|\mathcal{V}|(r+\delta-1)-\left|\bigcup_{R\in\mathcal{V}}R\right|\geqslant\Delta>0\tag{28}$$

and $\lceil\frac{k+\Delta}{r}\rceil>\lceil\frac{k}{r}\rceil$, then there exists a set $S\subseteq[n]$ with $\mathrm{Rank}(S)=k-1$ and

$$|S|\geqslant k+\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1).\tag{29}$$

Proof: If the requirements of Lemma 9 hold for $\mathcal{V}$, then the desired S may be obtained by Lemma 9, and we are done. Otherwise, $\mathcal{V}$ does not satisfies the requirements of Lemma 9, and then using Lemmas 7 and 8 (setting $\mathcal{R}'=\mathcal{R}$ in the latter), $\mathcal{V}$ may be extended to a set $\mathcal{V}'\subseteq\mathcal{R}$ with $\lceil\frac{k}{r}\rceil-1$ elements satisfying

$$\begin{aligned}\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}'}R\right)&\leqslant\left|\bigcup_{R\in\mathcal{V}'}R\right|-|\mathcal{V}'|(\delta-1)\\&=\left|\bigcup_{R\in\mathcal{V}'}R\right|-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1).\end{aligned}$$

Recall that $k=ru+v$, with $0\leqslant v\leqslant r-1$. It now follows that

$$\begin{aligned}&k-1-\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}'}R\right)\\&\geqslant ru+v-1-\left|\bigcup_{R\in\mathcal{V}'}R\right|+|\mathcal{V}'|(\delta-1)\\&=\begin{cases}u(r+\delta-1)-\left|\bigcup_{R\in\mathcal{V}'}R\right|+v-1, & \text{if } v\neq0,\\ r+(u-1)(r+\delta-1)-\left|\bigcup_{R\in\mathcal{V}'}R\right|+v-1, & \text{if } v=0,\end{cases}\\&=\begin{cases}|\mathcal{V}'|(r+\delta-1)-\left|\bigcup_{R\in\mathcal{V}'}R\right|+v-1, & \text{if } v\neq0,\\ r+|\mathcal{V}'|(r+\delta-1)-\left|\bigcup_{R\in\mathcal{V}'}R\right|-1, & \text{if } v=0,\end{cases}\\&\overset{(a)}{\geqslant}\begin{cases}|\mathcal{V}|(r+\delta-1)-\left|\bigcup_{R\in\mathcal{V}}R\right|+v-1, & \text{if } v\neq0,\\ r+|\mathcal{V}|(r+\delta-1)-\left|\bigcup_{R\in\mathcal{V}}R\right|-1, & \text{if } v=0,\end{cases}\\&\overset{(b)}{\geqslant}\begin{cases}\Delta+v-1, & \text{if } v\neq0,\\ r+\Delta-1, & \text{if } v=0,\end{cases}\end{aligned}\tag{30}$$

where (a) follows from the fact that $|R|\leqslant r+\delta-1$ for all $R\in\mathcal{V}'$, and (b) follows from (28).

For the case $v\neq0$, $\lceil\frac{k+\Delta}{r}\rceil=u+\lceil\frac{v+\Delta}{r}\rceil>\lceil\frac{k}{r}\rceil=u+1$ means that $\Delta+v>r$, i.e., $\Delta+v-1\geqslant r$. Thus, by (30) and $\Delta>0$,

$$\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}'}R\right)\leqslant k-1-r,\tag{31}$$

for both $v=0$ and $v\neq0$.

Again, by the same analysis as in Lemma 8, we can obtain yet another set $R'\in\mathcal{R}\setminus\mathcal{V}'$ with $\mathrm{Rank}(R'\cup(\bigcup_{R\in\mathcal{V}'}R))>\mathrm{Rank}(\bigcup_{R\in\mathcal{V}'}R)$ and then

$$\begin{aligned}&\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}'\cup\{R'\}}R\right)\\&\leqslant\left|\bigcup_{R\in\mathcal{V}'\cup\{R'\}}R\right|-|\mathcal{V}'\cup\{R'\}|(\delta-1)\\&=\left|\bigcup_{R\in\mathcal{V}'\cup\{R'\}}R\right|-\left\lceil\frac{k}{r}\right\rceil(\delta-1).\end{aligned}\tag{32}$$

Note that $\mathrm{Rank}(\bigcup_{R\in\mathcal{V}'\cup\{R'\}}R)\leqslant\mathrm{Rank}(\bigcup_{R\in\mathcal{V}'}R)+r\leqslant k-1$ by (31). Therefore, construct S by adding coordinates to $\bigcup_{R\in\mathcal{V}'\cup\{R'\}}R$ until reaching sufficient rank, $\mathrm{Rank}(S) = k-1$, and then by (32) we have

$$\begin{aligned}|S| &\geqslant k-1-\mathrm{Rank}\left(\bigcup_{R\in\mathcal{V}'\cup\{R'\}} R\right)+\left|\bigcup_{R\in\mathcal{V}'\cup\{R'\}} R\right| \\ &\geqslant k-1+\left\lceil\frac{k}{r}\right\rceil(\delta-1) \\ &\geqslant k+\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1),\end{aligned}$$

which completes the proof. ■

Now we are ready to prove Theorem 1.

Proof of Theorem 1: Let $\mathcal{R} \subseteq \Gamma$ be the ECF obtained in Lemma 6. If $D(\mathcal{R}) = 0$ and $|R| = r+\delta-1$ for all $R \in \mathcal{R}$, then set $\mathcal{S} = \mathcal{R}$ the theorem follows.

Otherwise, we have $D(\mathcal{R}) \neq 0$ or $|R| < r+\delta-1$ for some $R \in \mathcal{R}$. We distinguish between two cases. First, assume $k > 2r$. By Lemma 6, we know that $|\mathcal{R}| \geqslant \lceil k/r \rceil$. According to Lemma 5 we can find a $(\lceil \frac{k}{r} \rceil - 1)$-subset $\mathcal{V} \subseteq \mathcal{R}$ satisfying

$$\begin{aligned}&|\mathcal{V}|(r+\delta-1)-\left|\bigcup_{R\in\mathcal{V}} R\right| \\ &\geqslant \Delta = \min\left\{r+\delta-1, \left\lfloor\frac{\lceil\frac{k}{r}\rceil-1}{2}\right\rfloor\right\} \\ &> 0.\end{aligned}$$

Since $u \geqslant 2(r-v+1)$ or $v = 0$, we have $\lceil \frac{k+\Delta}{r} \rceil > \lceil \frac{k}{r} \rceil$. Therefore, by Lemma 10, there is a set $S \subseteq [n]$ with $\mathrm{Rank}(S) = k-1$ and

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

Thus, by Fact 1

$$d \leqslant n-|S| \leqslant n-k-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1).$$

This is a contradiction to the optimality of $\mathcal{C}$ with respect to the bound in Lemma 1.

In the second case, $r < k \leqslant 2r$. We note that we only need to consider the case $v = 0$, namely, $k = 2r$, since if $v \neq 0$ then the condition $u \geqslant 2(r-v+1) \geqslant 2$ implies that $k = ur+v > 2r$. We therefore assume $k = 2r$. The theorem now follows directly from [32, Theorem 9]. In what follows, we include a proof for this case for completeness.

If $D(\mathcal{R}) \neq 0$ or $|R| < r+\delta-1$ for some $R \in \mathcal{R}$ then we can find two distinct repair sets $R, R' \in \mathcal{R}$ such that $R \cap R' \neq \emptyset$ or $\min(|R|,|R'|) < r+\delta-1$. In either case, we have $\mathrm{Rank}(R\cup R') < 2r = k$.

We again distinguish between two cases depending on $|R\cap R'|$. For the first case, if $|R\cap R'| \leqslant \min(|R|,|R'|)-\delta+1$ then we have $\mathrm{Rank}(R\cup R') \leqslant |R\cup R'|-2(\delta-1) < |R\cup R'|-\delta+1$. In the second case, when $|R\cap R'| > \min(|R|,|R'|)-\delta+1$, assume without loss of generality, that $|R\cap R'| > |R'|-\delta+1$, then $\mathrm{Rank}(R\cup R') = \mathrm{Rank}(R) \leqslant |R|-\delta+1 < |R\cup R'|-\delta+1$.

We now construct a set $S \subseteq [n]$ by arbitrarily adding coordinates to $R \cup R' \subseteq S$ such that $\mathrm{Rank}(S) = k-1$. Therefore, $|S|-(k-1) \geqslant |R\cup R'| - \mathrm{Rank}(R\cup R') > \delta-1$, or equivalently, $|S| \geqslant k+\delta-1$. Again by Fact 1, we get

$$d \leqslant n-|S| \leqslant n-k-(\delta-1),$$

which is again a contradiction with the optimality of $\mathcal{C}$ with respect to the bound in Lemma 1. ■

ACKNOWLEDGMENTS

The authors are very grateful to the reviewers and the Associate Editor, Prof. Lara Dolecek, for their comments and suggestions that improved the presentation of this paper.

REFERENCES

[1] A. Agarwal, A. Barg, S. Hu, A. Mazumdar, and I. Tamo, "Combinatorial alphabet-dependent bounds for locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3481–3492, May 2018.

[2] R. C. Baker, G. Harman, and J. Pintz, "The difference between consecutive primes II," *Proc. London Math. Soc.*, vol. 83, no. 3, pp. 532–562, Dec. 2016.

[3] A. Beemer, R. Coatney, V. Guruswami, H. H. López, and F. Piñero, "Explicit optimal-length locally repairable codes of distance 5," 2018, *arXiv:1810.03980*. [Online]. Available: http://arxiv.org/abs/1810.03980

[4] S. Bhadane and A. Thangaraj, "Irregular recovery and unequal locality for locally recoverable codes with availability," 2017, *arXiv:1705.05005*. [Online]. Available: http://arxiv.org/abs/1705.05005

[5] 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.

[6] H. Cai, M. Cheng, C. Fan, and X. Tang, "Optimal locally repairable systematic codes based on packings," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 39–49, Jan. 2019.

[7] H. Cai, Y. Miao, M. Schwartz, and X. Tang, "On optimal locally repairable codes with multiple disjoint repair sets," *IEEE Trans. Inf. Theory*, to be published.

[8] C. Colbourn and J. Dinitz, *Handbook of Combinatorial Designs*, 2nd ed. London, U.K.: Chapman & Hall, 2006.

[9] M. Deza and P. Frankl, "Bounds on the maximum number of vectors with given scalar products," *Proc. Amer. Math. Soc.*, vol. 95, no. 2, p. 323, Feb. 1985.

[10] M. Forbes and S. Yekhanin, "On the locality of codeword symbols in non-linear codes," *Discrete Math.*, vol. 324, pp. 78–84, Jun. 2014.

[11] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6925–6934, Nov. 2012.

[12] Y. Gu and Y. Miao, "Union–intersection-bounded families and their applications," *Discrete Appl. Math.*, vol. 266, pp. 346–354, Aug. 2019.

[13] V. Guruswami, C. Xing, and C. Yuan, "How long can optimal locally repairable codes be?" *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3662–3670, Jun. 2019.

[14] J. Hao and S.-T. Xia, "Constructions of optimal binary locally repairable codes with multiple repair groups," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1060–1063, Jun. 2016.

[15] 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. (NCA)*, Jul. 2007, pp. 79–86.

[16] C. Huang *et al.*, "Erasure coding in windows azure storage," presented at the USENIX Annu. Tech. Conf. (USENIX ATC), 2012, pp. 15–26.

[17] L. Jin, "Explicit construction of optimal locally recoverable codes of distance 5 and 6 via binary constant weight codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4658–4663, Aug. 2019.

[18] S. Johnson, "A new upper bound for error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 8, no. 3, pp. 203–207, Apr. 1962.

[19] P. Keevash, "The existence of designs," 2014, *arXiv:1401.3665*. [Online]. Available: http://arxiv.org/abs/1401.3665

[20] G. Kim and J. Lee, "Locally repairable codes with unequal locality requirements," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7137–7152, Nov. 2018.

[21] O. Kolosov, A. Barg, I. Tamo, and G. Yadgar, "Optimal LRC codes for all lenghts n<= q," 2018, *arXiv:1802.00157*. [Online]. Available: http://arxiv.org/abs/1802.00157

[22] J. Liu, S. Mesnager, and L. Chen, "New constructions of optimal locally recoverable codes via good polynomials," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 889–899, Feb. 2018.

[23] Y. Luo, C. Xing, and C. Yuan, "Optimal locally repairable codes of distance 3 and 4 via cyclic codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 1048–1053, Feb. 2019.

[24] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North Holland, 1978.

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

[26] 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 Proc.*, Jul. 2012, pp. 2776–2780.

[27] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimal locally repairable and secure codes for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 212–236, Jan. 2014.

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

[29] V. Rödl, "On a packing and covering problem," *Eur. J. Combinatorics*, vol. 6, no. 1, pp. 69–78, Mar. 1985.

[30] B. Sasidharan, G. K. Agarwal, and P. V. Kumar, "Codes with hierarchical locality," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 1257–1261.

[31] N. Silberstein, T. Etzion, and M. Schwartz, "Locality and availability of array codes constructed from subspaces," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 2648–2660, May 2019.

[32] 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.

[33] 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.

[34] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4661–4676, Aug. 2014.

[35] A. Wang and Z. Zhang, "Repair locality with multiple erasure tolerance," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6979–6987, Nov. 2014.

[36] 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.

[37] T. Westerback, 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. He is currently a Postdoctoral Fellow at the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel. His research interests include coding theory and sequence design.

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

From 1989 to 1993, he worked for 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. From 1997 to 1998, he was a Postdoctoral Fellow at the Department of Computer Science, Concordia University, Montreal, QC, Canada. In 1998, he joined the University of Tsukuba, Tsukuba, Japan, where he is currently a Full Professor with the Faculty of Engineering, Information and Systems. His current research interests include combinatorics, coding theory, and information security.

Dr. Miao received the 2001 Kirkman Medal from the Institute of Combinatorics and its Applications. He is on the Editorial Boards of several journals such as the *Graphs and Combinatorics* and the *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 Postdoctoral Researcher at the Department of Electrical Engineering, California Institute of Technology. While on sabbatical 2012–2014, he was a Visiting Scientist with the Massachusetts Institute of Technology (MIT). He is currently an Associate Professor with the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel.

Prof. Schwartz received the 2009 IEEE Communications Society Best Paper Award in Signal Processing and Coding for Data Storage. He has been serving as an Associate Editor for *Coding Techniques* for the IEEE TRANSACTIONS ON INFORMATION THEORY since 2014.

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

From 2003 to 2004, he was a Research Associate with 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 as an Associate Editor for several journals, including the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEICE TRANSACTIONS ON FUNDAMENTALS, and has served on a number of technical program committees of conferences.