

Lower Bounds on the Sub-Packetization of Optimal-Access MSR Codes for Multiple-Node Repair

Lewen Wang, Zihao Zhang, Sihuang Hu

Abstract

We establish lower bounds on the sub-packetization of optimal-access MSR codes in the context of multiple-node failures. These bounds generalize the tight bounds for single-node failure presented by Balaji et al. (IEEE Transactions on Information Theory, vol. 68, no. 10, 2022). Moreover, we utilize generating functions to provide a more refined analysis, further strengthening these bounds.

I. INTRODUCTION

Modern distributed storage systems (DSS) require both high reliability and low storage overhead, particularly in the face of frequent node failures. Erasure codes have become a fundamental tool in meeting these needs, compared to traditional replication techniques. Among these, Maximum Distance Separable (MDS) codes stand out, as they provide the highest reliability for a given storage overhead. A crucial challenge in DSS is the efficient repair of failed nodes.

A. Single-Node Repair

Regenerating codes, introduced by Dimakis et al. [1], were initially designed to address single-node repair and have demonstrated excellent performance in this context. Let \mathcal{C} be an (n, k, ℓ) array code over the finite field \mathbb{F}_q . A codeword $C \in \mathcal{C}$ is represented as $C = (C_1, \dots, C_n)^T$, where each codeword symbol C_i is a vector of the form $C_i = (c_{i,1}, \dots, c_{i,\ell})$ over \mathbb{F}_q for $i \in [n]$. The length ℓ of each codeword symbol is called the *sub-packetization* of the code. For an (n, k, ℓ) MDS array code, Dimakis et al. established the minimum repair bandwidth

$$\frac{d\ell}{d-k+1}$$

for recovering a failed node C_i from any other d helper nodes C_j 's, known as the cut-set bound.

Definition 1. An (n, k, ℓ) MDS array code \mathcal{C} over \mathbb{F}_q is said to be an $((n, k, d), (\ell, \beta = \frac{\ell}{d-k+1}))$ Minimum Storage Regenerating (MSR) code if for every $i \in [n]$, for all $\mathcal{R} \subseteq [n] \setminus \{i\}$ of size d , there exist functions $f_{i,j}^{\mathcal{R}} : \mathbb{F}^\ell \rightarrow \mathbb{F}^\beta, j \in \mathcal{R}$ and function $g_i^{\mathcal{R}} : \mathbb{F}^{d\beta} \rightarrow \mathbb{F}^\ell$, such that the code symbol C_i , can be computed by $g_i^{\mathcal{R}}(f_{i,j}^{\mathcal{R}}(C_j), j \in \mathcal{R})$.

In distributed storage node repair, besides the repair bandwidth, the amount of data accessed is also crucial, as it directly affects the disk I/O overhead. The optimal-access property refers to the condition where the volume of data transmitted by a helper node is exactly the same as the amount of data accessed. MSR codes that satisfy this condition are called *optimal-access* MSR codes.

Much progress has been made in deriving lower bounds on the sub-packetization ℓ . For general MSR codes with repair degree $d = n - 1$, Goparaju et al. [2] proved the lower bound $\ell \geq \exp(\Theta(\sqrt{k}))$. This result was later improved to $\ell \geq e^{\frac{k-1}{4r}}$ by Alrabiah and Guruswami [3], where $r = n - k$. The most recent result, $\ell \geq e^{\frac{(k-1)(r-1)}{2r^2}}$, was presented by Alrabiah and Guruswami [4] and Balaji et al. [5]. For optimal-access MSR codes, Tamo et al. [6] demonstrated that $\ell \geq r^{\frac{k-1}{r}}$ with repair degree $d = n - 1$. Balaji et al. [5] further established the lower bound $\ell \geq \min\{s^{\lceil \frac{n-1}{s} \rceil}, s^{k-1}\}$ with $s = d - k + 1$. This bound is tight as it is achieved by construction in [7]–[12]. In addition to the constructions mentioned above, numerous MSR code constructions have been proposed in recent years. Please refer to [13]–[19] and the excellent survey [20] for further details.

The authors are with the Key Laboratory of Cryptologic Technology and Information Security, Ministry of Education, Shandong University, Qingdao, Shandong, 266237, China, and School of Cyber Science and Technology, Shandong University, Qingdao, Shandong, 266237, China. S. Hu is also with Quan Cheng Laboratory, Jinan 250103, China. Research partially funded by National Key R&D Program of China under Grant No. 2021YFA1001000, and National Natural Science Foundation of China under Grant No. 12231014. Email: lewenwang3@gmail.com, zihaozhang@mail.sdu.edu.cn, husihuang@sdu.edu.cn (Corresponding author: Sihuang Hu.)

B. Multiple-Node Repair

In practical scenarios, repairing multiple nodes failures is a common situation. For multi-node repair, two main models are commonly considered: the *centralized* repair and the *cooperative* repair. In the centralized repair model, a single data center downloads data from d helper nodes and uses this data to rebuild h failed nodes. In contrast, the cooperative repair model involves the failed nodes themselves collaborating with each other, acquiring data from helper nodes, and sharing information between them during the repair process. The repair bandwidth for these two models is calculated differently: in centralized repair, the bandwidth is determined by the total data downloaded from the helper nodes, while in cooperative repair, it includes both the data received from helper nodes and the data exchanged between the failed nodes. Please see [21]–[26] for results under centralized repair model and [27]–[32] for results under cooperative repair model.

This paper specifically focuses on centralized MSR codes. Under this model, the cut-set bound is established by Cadambe et al. in [21]. As we generally consider linear repair, the formal definition of centralized MSR code is given as follows. (We offer a formal proof of the cut-set bound in Appendix A.)

Definition 2. Let \mathcal{C} be an (n, k, ℓ) MDS array code. Given two integers h, d with $2 \leq h \leq r$, $k+1 \leq d \leq n-h$, and a subset $\mathcal{F} \subseteq [n]$ of size h , we say that \mathcal{C} has (h, d) optimal repair schemes for \mathcal{F} if for any $\mathcal{R} \subseteq [n] \setminus \mathcal{F}$ of size d , there exist linear functions $f_{\mathcal{F},j}^{\mathcal{R}} : \mathbb{F}^{\ell} \rightarrow \mathbb{F}^{\beta}$, $j \in \mathcal{R}$ and linear function $g_{\mathcal{F}}^{\mathcal{R}} : \mathbb{F}^{d\beta} \rightarrow \mathbb{F}^{h\ell}$, such that the code symbols C_i , $i \in \mathcal{F}$, can be computed by $g_{\mathcal{F}}^{\mathcal{R}}(f_{\mathcal{F},j}^{\mathcal{R}}(C_j) : j \in \mathcal{R})$, where $\beta = \frac{h\ell}{d-k+h}$. In general, if \mathcal{C} has (h, d) optimal repair schemes for any $\mathcal{F} \subseteq [n]$ of size h , then we call \mathcal{C} an (h, d) centralized MSR code. An (h, d) -centralized MSR code with optimal-access property is called an optimal-access (h, d) -centralized MSR code.

As $f_{\mathcal{F},j}^{\mathcal{R}}$ and $g_{\mathcal{F}}^{\mathcal{R}}$, $j \in \mathcal{R}$ are linear, there exist matrices $S_{j \rightarrow \mathcal{F}}^{\mathcal{R}}$ of sizes $\beta \times \ell$ and $M_{\mathcal{F}}^{\mathcal{R}}$ of size $h\ell \times d\beta$ corresponding to $f_{\mathcal{F},j}^{\mathcal{R}}$ and $g_{\mathcal{F}}^{\mathcal{R}}$, respectively. Specifically, the data transmitted from node C_j , for the purpose of repairing the nodes in \mathcal{F} , is $S_{j \rightarrow \mathcal{F}}^{\mathcal{R}} C_j$. The matrices $S_{j \rightarrow \mathcal{F}}^{\mathcal{R}}$, for $j \in \mathcal{R}$ are referred to as the *repair matrices*, and the span of the rows of $S_{j \rightarrow \mathcal{F}}^{\mathcal{R}}$, denoted by $\langle S_{j \rightarrow \mathcal{F}}^{\mathcal{R}} \rangle$, is called the *repair subspace*. If $d = n - h$, the helper nodes necessarily comprises all remaining nodes, and we will drop the superscript \mathcal{R} of $S_{j \rightarrow \mathcal{F}}^{\mathcal{R}}$ and $M_{\mathcal{F}}^{\mathcal{R}}$. If the repair matrices depend solely on the set of failed nodes \mathcal{F} , namely, $S_{j \rightarrow \mathcal{F}}^{\mathcal{R}} = S_{\mathcal{F}}$, for any $\mathcal{R} \subseteq [n] \setminus \mathcal{F}$ and $j \in \mathcal{R}$, such repair matrices are commonly referred to as *constant repair matrices*. If $S_{j \rightarrow \mathcal{F}}^{\mathcal{R}} = S_{j \rightarrow \mathcal{F}}$ for any $\mathcal{R} \subseteq [n] \setminus \mathcal{F}$, then the repair matrices are said to be *independent of the choice of the remaining $(d-1)$ helper nodes* [5]. These two types of repair matrices play a crucial role in study of lower bounds on the sub-packetization.

Numerous constructions for both the centralized and cooperative repair models have been proposed in recent years. Please refer to [22], [33]–[41] for further reading. However, to date, no lower bound has been established on the sub-packetization of MSR codes in the context of multiple-node repair.

C. Main Results

In this paper, we establish the following novel lower bounds on the sub-packetization of optimal-access MSR codes in the context of multiple-node failures. These bounds generalize the tight bounds for single-node failure presented by Balaji et al. [5].

Theorem 1. Let \mathcal{C} be an (n, k, ℓ) MDS array code. If \mathcal{C} has (h, d) optimal-access repair schemes with constant repair matrices for any h failed systematic nodes, then

$$\ell \geq \left(\frac{d-k+h}{h} \right)^{\left\lceil \left\lfloor \frac{k}{h} \right\rfloor \frac{h}{d-k+h} \right\rceil}.$$

Theorem 2. Let \mathcal{C} be an optimal-access (h, d) centralized MSR code with constant repair matrices. Then

$$\ell \geq \min \left\{ \left(\frac{d-k+h}{h} \right)^{\left\lfloor \frac{k}{h} \right\rfloor}, \left(\frac{d-k+h}{h} \right)^{\left\lceil \frac{n}{d-k+h} \right\rceil} \right\}.$$

Theorem 3. Let \mathcal{C} be an optimal-access (h, d) centralized MSR code with repair matrices independent of the choice of the remaining $(d-1)$ helper nodes. Then

$$\ell \geq \min \left\{ \left(\frac{d-k+h}{h} \right)^{\left\lfloor \frac{k-1}{h} \right\rfloor}, \left(\frac{d-k+h}{h} \right)^{\left\lceil \left\lfloor \frac{n-1}{h} \right\rfloor \frac{h}{d-k+h} \right\rceil} \right\}.$$

Remark 1. 1) In Theorem 2 if we let $h = 1$, then we get a tight bound on the sub-packetization of MSR codes with constant repair matrices. This bound is achieved by the construction in [7]–[12].

2) In Theorem 3 if we let $d = n - h$, then we can get a lower bound for optimal-access $(h, n - h)$ centralized MSR codes directly.

3) As a cooperative MSR code is also a centralized MSR code by the result of Ye et al. [37], our lower bounds also apply to cooperative MSR codes.

Furthermore, we can employ generating functions to conduct a more refined analysis. For any $1 \leq c \leq \lfloor \frac{h}{2} \rfloor$, we set

$$\Delta = \Delta(c) := \left(\frac{h-c}{d-k+h} \right)^2 + \frac{4c}{d-k+h}$$

and

$$\alpha = \alpha(c) := \frac{(c-h) + \sqrt{\Delta}(d-k+h)}{2c}.$$

Theorem 4. *Let \mathcal{C} be an optimal-access (h, d) centralized MSR code with constant repair matrices. Then*

$$\ell \geq \min \left\{ \frac{\sqrt{\Delta}}{4} \alpha(\lfloor \frac{k-h}{h-c} \rfloor + 1), \frac{\sqrt{\Delta}}{4} \alpha(\lceil (\lfloor \frac{n-h}{h-c} \rfloor + 1) \cdot \frac{h}{d-k+h} \rceil) \right\},$$

for any $1 \leq c \leq \lfloor \frac{h}{2} \rfloor$ and Δ, α defined above.

Theorem 5. *Let \mathcal{C} be an optimal-access (h, d) centralized MSR code with repair matrices independent of the choice of the remaining $(d-1)$ helper nodes. Then*

$$l \geq \min \left\{ \frac{\sqrt{\Delta}}{4} \alpha(\lfloor \frac{k-1-h}{h-c} \rfloor + 1), \frac{\sqrt{\Delta}}{4} \alpha(\lceil (\lfloor \frac{n-1-h}{h-c} \rfloor + 1) \cdot \frac{h}{d-k+h} \rceil) \right\},$$

for any $1 \leq c \leq \lfloor \frac{h}{2} \rfloor$ and Δ, α defined above.

Remark 2. *For high-rate MSR codes, and for sufficiently large n , it can be shown that the bounds in Theorems 4 and 5 are greater than those in Theorems 2 and 3. Therefore, unlike the case when $h = 1$, the generalized bounds in Theorems 2 and 3 are generally not tight.*

D. Outline of This Paper

The rest of this paper is organized as follows. In Section II, we will study the linear algebraic description for centralized model, which plays a fundamental role in the subsequent proofs. In Section III, we will focus on deriving the lower bounds for optimal-access (h, d) centralized MSR codes and prove Theorem 1 and Theorem 2. In Section IV, we will extend our analysis to the case repair matrices independent of the choice of the remaining $(d-1)$ helper nodes and prove Theorem 3. In Section V, we improve our bounds by using generating functions and prove Theorem 4 and Theorem 5. Section VI provides the conclusion of our paper.

II. LINEAR ALGEBRAIC SETTINGS FOR CENTRALIZED REPAIR

In this section, we consider the problem of repairing h systematic nodes using the remaining $n-h$ nodes. We will follow the notations used by Balaji et al. [5]. Let \mathcal{C} be an (n, k, ℓ) array code over \mathbb{F}_q . Write $r = n - k$. Suppose that $\mathcal{U} = \{u_1, \dots, u_k\}$ is a collection of k systematic nodes of \mathcal{C} , and let $\mathcal{P} = [n] \setminus \mathcal{U} = \{p_1, \dots, p_r\}$ be the left r parity nodes. Then we can get a generator matrix G of \mathcal{C} as

$$G = \begin{matrix} & u_1 & \cdots & u_k \\ \begin{matrix} u_1 \\ \vdots \\ u_k \\ p_1 \\ \vdots \\ p_r \end{matrix} & \begin{bmatrix} I_\ell & & \\ & \ddots & \\ & & I_\ell \\ A_{p_1, u_1} & \cdots & A_{p_1, u_k} \\ \vdots & \ddots & \vdots \\ A_{p_r, u_1} & \cdots & A_{p_r, u_k} \end{bmatrix} \end{matrix}$$

where I_ℓ is the $\ell \times \ell$ identity matrix and $A_{p_i, u_j}, i \in [r], j \in [k]$ are $\ell \times \ell$ matrices over \mathbb{F}_q . Here the rows of G are indexed by the nodes $u_1, \dots, u_k, p_1, \dots, p_r$, and the columns by u_1, \dots, u_k . Write

$$A = \begin{bmatrix} A_{p_1, u_1} & \cdots & A_{p_1, u_k} \\ \vdots & \ddots & \vdots \\ A_{p_r, u_1} & \cdots & A_{p_r, u_k} \end{bmatrix}.$$

It is well-known that \mathcal{C} is an MDS array code if and only if every square block submatrix of A is invertible.

The following notations will be used throughout this paper. For any subset $\mathcal{E} = \{u_{i_1}, \dots, u_{i_t}\} \subseteq \mathcal{U}$, let $A(:, \mathcal{E})$ be the submatrix of A consisting of the block columns indexed by $\{u_{i_1}, \dots, u_{i_t}\}$. For any $\mathcal{F} \subseteq [n]$ of size h ($2 \leq h \leq r$), $\mathcal{R} \subseteq [n] \setminus \mathcal{F}$ of size d ($k+1 \leq d \leq n-h$), and its subset $\mathcal{H} = \{i_1, \dots, i_p\} \subseteq \mathcal{R}$, we write

$$S_{\mathcal{H} \rightarrow \mathcal{F}}^{\mathcal{R}} = \text{diag}(S_{i_1 \rightarrow \mathcal{F}}^{\mathcal{R}}, \dots, S_{i_p \rightarrow \mathcal{F}}^{\mathcal{R}}).$$

If $d = n - h$, we can drop the superscript \mathcal{R} . For a matrix B , we denote its row space by $\langle B \rangle$.

Now we give a linear algebraic description of the optimal repair schemes for h failed systematic nodes using the remaining $n - h$ nodes.

Lemma 1. *Let \mathcal{C} be an (n, k, ℓ) MDS array code over \mathbb{F}_q . Suppose that for any h failed systematic nodes $\mathcal{F} = \{u_{i_1}, \dots, u_{i_h}\} \subseteq \mathcal{U}$, there exists a (h, d) optimal repair scheme for \mathcal{F} with $\mathcal{R} = [n] \setminus \mathcal{F}$, that is, there exist repair matrices $S_{j \rightarrow \mathcal{F}}$ of sizes $\beta \times \ell$, $j \in [n] \setminus \mathcal{F}$, and a matrix $M_{\mathcal{F}}$ of size $h\ell \times d\beta$, such that*

$$\begin{bmatrix} C_{u_{i_1}} \\ \vdots \\ C_{u_{i_h}} \end{bmatrix} = M_{\mathcal{F}} \begin{bmatrix} S_{u_{i_{h+1}} \rightarrow \mathcal{F}} C_{u_{i_{h+1}}} \\ \vdots \\ S_{u_{i_k} \rightarrow \mathcal{F}} C_{u_{i_k}} \\ S_{p_1 \rightarrow \mathcal{F}} C_{p_1} \\ \vdots \\ S_{p_r \rightarrow \mathcal{F}} C_{p_r} \end{bmatrix} \quad \text{for any } C = (C_1, \dots, C_n)^T \in \mathcal{C}, \quad (1)$$

where we write $\mathcal{U} \setminus \mathcal{F} = \{u_{i_{h+1}}, \dots, u_{i_k}\}$. Then we have

- 1) The square matrix $S_{\mathcal{P} \rightarrow \mathcal{F}} A(:, \mathcal{F})$ is of full rank.
- 2) For $1 \leq t \leq r$ and $h+1 \leq z \leq k$, we have $\langle S_{p_t \rightarrow \mathcal{F}} A_{p_t, u_{i_z}} \rangle = \langle S_{u_{i_z} \rightarrow \mathcal{F}} \rangle$.
- 3) For any $j \in [n] \setminus \mathcal{F}$, the repair matrix $S_{j \rightarrow \mathcal{F}}$ has full row rank.

Proof. Write

$$M_{\mathcal{F}} = \begin{bmatrix} M_{1, u_{i_{h+1}}} & \dots & M_{1, u_{i_k}} & M_{1, p_1} & \dots & M_{1, p_r} \\ \vdots & & \vdots & \vdots & & \vdots \\ M_{h, u_{i_{h+1}}} & \dots & M_{h, u_{i_k}} & M_{h, p_1} & \dots & M_{h, p_r} \end{bmatrix}$$

as an $h \times d$ block matrix, where each block entry $M_{i,j}$ is of size $\ell \times \beta$. Let $M_{\mathcal{F}}(:, \mathcal{U} \setminus \mathcal{F})$ be the submatrix of $M_{\mathcal{F}}$ consisting of the first $k - h$ block columns and $M_{\mathcal{F}}(:, \mathcal{P})$ the submatrix consisting of the last r block columns. Then $M_{\mathcal{F}} = [M_{\mathcal{F}}(:, \mathcal{U} \setminus \mathcal{F}) \ M_{\mathcal{F}}(:, \mathcal{P})]$. Similarly write

$$A = \begin{bmatrix} A_{p_1, u_{i_1}} & \dots & A_{p_1, u_{i_h}} & A_{p_1, u_{i_{h+1}}} & \dots & A_{p_1, u_{i_k}} \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ A_{p_r, u_{i_1}} & \dots & A_{p_r, u_{i_h}} & A_{p_r, u_{i_{h+1}}} & \dots & A_{p_r, u_{i_k}} \end{bmatrix} = [A(:, \mathcal{F}) \ A(:, \mathcal{U} \setminus \mathcal{F})].$$

Recall that

$$\begin{aligned} \begin{bmatrix} C_{p_1} \\ \vdots \\ C_{p_r} \end{bmatrix} &= \begin{bmatrix} A_{p_1, u_{i_1}} & \dots & A_{p_1, u_{i_h}} & A_{p_1, u_{i_{h+1}}} & \dots & A_{p_1, u_{i_k}} \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ A_{p_r, u_{i_1}} & \dots & A_{p_r, u_{i_h}} & A_{p_r, u_{i_{h+1}}} & \dots & A_{p_r, u_{i_k}} \end{bmatrix} \begin{bmatrix} C_{u_{i_1}} \\ \vdots \\ C_{u_{i_h}} \\ C_{u_{i_{h+1}}} \\ \vdots \\ C_{u_{i_k}} \end{bmatrix} \\ &= [A(:, \mathcal{F}) \ A(:, \mathcal{U} \setminus \mathcal{F})] \begin{bmatrix} C_{u_{i_1}} \\ \vdots \\ C_{u_{i_h}} \\ C_{u_{i_{h+1}}} \\ \vdots \\ C_{u_{i_k}} \end{bmatrix}. \end{aligned}$$

By (1) we have

$$\begin{aligned}
\begin{bmatrix} C_{u_{i_1}} \\ \vdots \\ C_{u_{i_h}} \end{bmatrix} &= M_{\mathcal{F}} \begin{bmatrix} S_{\mathcal{U} \setminus \mathcal{F} \rightarrow \mathcal{F}} & \mathbf{O} \\ \mathbf{O} & S_{P \rightarrow \mathcal{F}} \end{bmatrix} \begin{bmatrix} \mathbf{O} & I_{(k-h)\ell} \\ A(:, \mathcal{F}) & A(:, \mathcal{U} \setminus \mathcal{F}) \end{bmatrix} \begin{bmatrix} C_{u_{i_1}} \\ \vdots \\ C_{u_{i_h}} \\ C_{u_{i_{h+1}}} \\ \vdots \\ C_{u_{i_k}} \end{bmatrix} \\
&= [M_{\mathcal{F}}(:, \mathcal{U} \setminus \mathcal{F}) \ M_{\mathcal{F}}(:, P)] \begin{bmatrix} \mathbf{O} & S_{\mathcal{U} \setminus \mathcal{F} \rightarrow \mathcal{F}} \\ S_{P \rightarrow \mathcal{F}} A(:, \mathcal{F}) & S_{P \rightarrow \mathcal{F}} A(:, \mathcal{U} \setminus \mathcal{F}) \end{bmatrix} \begin{bmatrix} C_{u_{i_1}} \\ \vdots \\ C_{u_{i_h}} \\ C_{u_{i_{h+1}}} \\ \vdots \\ C_{u_{i_k}} \end{bmatrix} \\
&= M_{\mathcal{F}}(:, P) S_{P \rightarrow \mathcal{F}} A(:, \mathcal{F}) \begin{bmatrix} C_{u_{i_1}} \\ \vdots \\ C_{u_{i_h}} \end{bmatrix} + (M_{\mathcal{F}}(:, \mathcal{U} \setminus \mathcal{F}) S_{\mathcal{U} \setminus \mathcal{F} \rightarrow \mathcal{F}} + M_{\mathcal{F}}(:, P) S_{P \rightarrow \mathcal{F}} A(:, \mathcal{U} \setminus \mathcal{F})) \begin{bmatrix} C_{u_{i_{h+1}}} \\ \vdots \\ C_{u_{i_k}} \end{bmatrix}.
\end{aligned}$$

Set $C_{u_{i_{h+1}}} = \dots = C_{u_{i_k}} = 0$. Then we obtain

$$\begin{bmatrix} C_{u_{i_1}} \\ \vdots \\ C_{u_{i_h}} \end{bmatrix} = M_{\mathcal{F}}(:, P) S_{P \rightarrow \mathcal{F}} A(:, \mathcal{F}) \begin{bmatrix} C_{u_{i_1}} \\ \vdots \\ C_{u_{i_h}} \end{bmatrix},$$

for $[C_{u_{i_1}}, \dots, C_{u_{i_h}}]^T \in \mathbb{F}_q^{h\ell}$. Hence $M_{\mathcal{F}}(:, P) S_{P \rightarrow \mathcal{F}} A(:, \mathcal{F}) = I_{h\ell}$. It follows directly that the matrices $S_{p_t \rightarrow \mathcal{F}}$, $1 \leq t \leq r$, have full row rank, and the square matrices $M_{\mathcal{F}}(:, P)$ and $S_{P \rightarrow \mathcal{F}} A(:, \mathcal{F})$ both have full rank.

Set $C_{u_{i_1}} = \dots = C_{u_{i_h}} = 0$. Then we obtain

$$(M_{\mathcal{F}}(:, \mathcal{U} \setminus \mathcal{F}) S_{\mathcal{U} \setminus \mathcal{F} \rightarrow \mathcal{F}} + M_{\mathcal{F}}(:, P) S_{P \rightarrow \mathcal{F}} A(:, \mathcal{U} \setminus \mathcal{F})) \begin{bmatrix} C_{u_{i_{h+1}}} \\ \vdots \\ C_{u_{i_k}} \end{bmatrix} = 0,$$

for all $[C_{u_{i_{h+1}}}, \dots, C_{u_{i_k}}]^T \in \mathbb{F}_q^{(k-h)\ell}$. Hence

$$S_{P \rightarrow \mathcal{F}} A(:, \mathcal{U} \setminus \mathcal{F}) = -(M_{\mathcal{F}}(:, P))^{-1} M_{\mathcal{F}}(:, \mathcal{U} \setminus \mathcal{F}) S_{\mathcal{U} \setminus \mathcal{F} \rightarrow \mathcal{F}}.$$

Comparing the block columns of the above two matrices, we have

$$\begin{bmatrix} S_{p_1 \rightarrow \mathcal{F}} A_{p_1, u_{i_z}} \\ \vdots \\ S_{p_r \rightarrow \mathcal{F}} A_{p_r, u_{i_z}} \end{bmatrix} = -(M_{\mathcal{F}}(:, P))^{-1} \begin{bmatrix} M_{1, u_{i_z}} \\ \vdots \\ M_{h, u_{i_z}} \end{bmatrix} S_{u_{i_z} \rightarrow \mathcal{F}}$$

for $h+1 \leq z \leq k$. It is not hard to check that $\langle S_{p_t \rightarrow \mathcal{F}} A_{p_t, u_{i_z}} \rangle \subseteq \langle S_{u_{i_z} \rightarrow \mathcal{F}} \rangle$ for $1 \leq t \leq r$, $h+1 \leq z \leq k$. As the matrices $S_{p_t \rightarrow \mathcal{F}}$ and $A_{p_t, u_{i_z}}$ both have full rank, we get $\dim \langle S_{p_t \rightarrow \mathcal{F}} A_{p_t, u_{i_z}} \rangle = \dim \langle S_{u_{i_z} \rightarrow \mathcal{F}} \rangle = \beta$. Therefore $\langle S_{p_t \rightarrow \mathcal{F}} A_{p_t, u_{i_z}} \rangle = \langle S_{u_{i_z} \rightarrow \mathcal{F}} \rangle$, and the repair matrix $S_{u_{i_z} \rightarrow \mathcal{F}}$ has full row rank. \square

The following key lemma will be used in latter sections to derive our new lower bounds.

Lemma 2. Let $\mathcal{F}_1 = \{u_{i_1}, \dots, u_{i_h}\}$ and $\mathcal{F}_2 = \{v_{i_1}, \dots, v_{i_h}\}$ be two subsets of \mathcal{U} . Then we have

$$\text{rank}(S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1)) \leq \frac{(h - |\mathcal{F}_1 \cap \mathcal{F}_2|)h\ell}{r} + |\mathcal{F}_1 \cap \mathcal{F}_2|\ell.$$

Moreover, if $\mathcal{F}_1 \cap \mathcal{F}_2 = \emptyset$, then the equality holds, that is,

$$\text{rank}(S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1)) = \frac{h^2\ell}{r},$$

and

$$\langle S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1) \rangle = \langle S_{\mathcal{F}_1 \rightarrow \mathcal{F}_2} \rangle. \quad (2)$$

Proof. Write $t = |\mathcal{F}_1 \cap \mathcal{F}_2|$. Without loss of generality, we can assume that $\mathcal{F}_1 \cap \mathcal{F}_2 = \{u_{i_{h-t+1}}, \dots, u_{i_h}\}$ and $u_{i_z} = v_{i_z}$ for $h-t+1 \leq z \leq h$. Write

$$S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1) = \begin{bmatrix} S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_2)) & S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1 \cap \mathcal{F}_2) \end{bmatrix},$$

and note that

$$\langle S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1) \rangle \subseteq \left\langle \begin{bmatrix} S_{\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_2) \rightarrow \mathcal{F}_2} & \mathbf{O} \\ S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_2)) & S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1 \cap \mathcal{F}_2) \end{bmatrix} \right\rangle.$$

By 2) of Lemma 1 we have $\langle S_{p_t \rightarrow \mathcal{F}_2} A_{p_t, u_{i_z}} \rangle = \langle S_{u_{i_z} \rightarrow \mathcal{F}_2} \rangle$, for $1 \leq t \leq r$ and $1 \leq z \leq h-t$. Using elementary row operations, it is not hard to check that

$$\left\langle \begin{bmatrix} S_{\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_2) \rightarrow \mathcal{F}_2} & \mathbf{O} \\ S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_2)) & S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1 \cap \mathcal{F}_2) \end{bmatrix} \right\rangle = \left\langle \begin{bmatrix} S_{\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_2) \rightarrow \mathcal{F}_2} & \mathbf{O} \\ \mathbf{O} & S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1 \cap \mathcal{F}_2) \end{bmatrix} \right\rangle.$$

Therefore

$$\langle S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1) \rangle \subseteq \left\langle \begin{bmatrix} S_{\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_2) \rightarrow \mathcal{F}_2} & \mathbf{O} \\ \mathbf{O} & S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1 \cap \mathcal{F}_2) \end{bmatrix} \right\rangle. \quad (3)$$

By 3) of Lemma 1 we have $\text{rank}(S_{\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_2) \rightarrow \mathcal{F}_2}) = \frac{(h-t)h\ell}{r}$. On the other hand, $\text{rank}(S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1 \cap \mathcal{F}_2))$ is at most $t\ell$, the number of its columns. Hence

$$\begin{aligned} \text{rank}(S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1)) &\leq \text{rank}(S_{\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_2) \rightarrow \mathcal{F}_2}) + \text{rank}(S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1 \cap \mathcal{F}_2)) \\ &\leq \frac{(h-t)h\ell}{r} + t\ell. \end{aligned}$$

Now if $\mathcal{F}_1 \cap \mathcal{F}_2 = \emptyset$, then (3) becomes

$$\langle S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1) \rangle \subseteq \langle S_{\mathcal{F}_1 \rightarrow \mathcal{F}_2} \rangle.$$

By 3) of Lemma 1 we see that the matrix $S_{P \rightarrow \mathcal{F}_2}$ has full row rank, and $\text{rank}(S_{\mathcal{F}_1 \rightarrow \mathcal{F}_2}) = \frac{h^2\ell}{r}$. By the MDS property of \mathcal{C} we know that the first h block rows of matrix $A(:, \mathcal{F}_1)$ is invertible. Then

$$\text{rank}(S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1)) \geq \text{rank}(S_{\{p_1, \dots, p_h\} \rightarrow \mathcal{F}_2}) = \frac{h^2\ell}{r}.$$

It follows directly that $\langle S_{P \rightarrow \mathcal{F}_2} A(:, \mathcal{F}_1) \rangle = \langle S_{\mathcal{F}_1 \rightarrow \mathcal{F}_2} \rangle$. □

III. LOWER BOUNDS WITH CONSTANT REPAIR MATRICES

In this section, we derive lower bounds on the sub-packetization of optimal-access (h, d) -centralized MSR codes with constant repair matrices, namely, $S_{j \rightarrow \mathcal{F}}^{\mathcal{R}} = S_{\mathcal{F}}$ for $\mathcal{F} \subseteq [n]$, $\mathcal{R} \subseteq [n] \setminus \mathcal{F}$ and $j \in \mathcal{R}$. The proof method is first employed by Tamo et al. [6], and later developed by Balaj et al. [5]. Before we proceed to the proofs we need the following result.

Lemma 3. Suppose that \mathcal{C} has (h, d) optimal repair schemes with constant repair matrices for any h failed systematic nodes. Let $1 \leq t \leq \lfloor \frac{k}{h} \rfloor$, and let $\mathcal{F}_1, \dots, \mathcal{F}_t \subseteq \mathcal{U}$ be any t pairwise disjoint subsets of size h of k systematic nodes. Then we have

$$\dim \left(\bigcap_{i=1}^t \langle S_{\mathcal{F}_i} \rangle \right) \leq \left(\frac{h}{d-k+h} \right)^t \ell.$$

Proof. We first assume that $d = n - h$ and prove the result by induction on t . For $t = 1$, by 3) of Lemma 1, we have $\dim(\langle S_{\mathcal{F}_1} \rangle) = \frac{h\ell}{r}$. Assume the result holds for t , and now we will prove it also holds for $t+1$. Fix a basis for the subspace $\bigcap_{i=1}^{t+1} \langle S_{\mathcal{F}_i} \rangle$, and set S to be the matrix whose rows are basis vectors. Then $\langle S \rangle = \bigcap_{i=1}^{t+1} \langle S_{\mathcal{F}_i} \rangle$, and in particular $\langle S \rangle \subseteq \langle S_{\mathcal{F}_i} \rangle$ for $i \in [t+1]$. Hence we can find matrices T_i , whose rank is equal to $\dim(\langle S \rangle)$, such that $S = T_i S_{\mathcal{F}_i}$. It follows that

$$(I_r \otimes S) A(:, \mathcal{F}_1) = (I_r \otimes T_i) (I_r \otimes S_{\mathcal{F}_i}) A(:, \mathcal{F}_1), \text{ for } i \in [t+1].$$

If $i = 1$, then by 1) of Lemma 1 we see that the matrix $(I_r \otimes S_{\mathcal{F}_1}) A(:, \mathcal{F}_1)$ is invertible. Then

$$\dim(\langle (I_r \otimes S) A(:, \mathcal{F}_1) \rangle) = \dim(\langle (I_r \otimes T_1) \rangle) = r \dim(\langle S \rangle). \quad (4)$$

On the other hand, for $2 \leq i \leq t+1$ we have

$$\langle (I_r \otimes S) A(:, \mathcal{F}_1) \rangle \subseteq \langle (I_r \otimes S_{\mathcal{F}_i}) A(:, \mathcal{F}_1) \rangle = \langle I_h \otimes S_{\mathcal{F}_i} \rangle.$$

Here the latter equality follows from (2) of Lemma 2. Then

$$\langle\langle I_r \otimes S \rangle A(:, \mathcal{F}_1) \rangle \subseteq \bigcap_{i=2}^{t+1} \langle\langle I_r \otimes S_{\mathcal{F}_i} \rangle A(:, \mathcal{F}_1) \rangle = \bigcap_{i=2}^{t+1} \langle I_h \otimes S_{\mathcal{F}_i} \rangle.$$

By assumption we have

$$\begin{aligned} \dim(\langle\langle I_r \otimes S \rangle A(:, \mathcal{F}_1) \rangle) &\leq \dim\left(\bigcap_{i=2}^{t+1} \langle I_h \otimes S_{\mathcal{F}_i} \rangle\right) \\ &= h \cdot \dim\left(\bigcap_{i=2}^{t+1} \langle S_{\mathcal{F}_i} \rangle\right) \\ &\leq h \left(\frac{h}{r}\right)^t \ell. \end{aligned} \tag{5}$$

By (4) and (5), we can compute that

$$\dim(\langle S \rangle) \leq \left(\frac{h}{r}\right)^{t+1} \ell.$$

This concludes our proof for the case $d = n - h$.

For general d , we first puncture any $n - d - h$ parity nodes from \mathcal{C} and get a new code \mathcal{C}' with $n' = d + h$ and $r' = d - k + h$. Applying the above argument to \mathcal{C}' we get

$$\dim\left(\bigcap_{i=1}^t \langle S_{\mathcal{F}_i} \rangle\right) \leq \left(\frac{h}{d - k + h}\right)^t \ell.$$

□

Proof of Theorem 1. Let $\{e_j : j \in [\ell]\}$ be the standard basis of \mathbb{F}_q^ℓ , where e_j denotes the vector with a 1 in the j th coordinate and 0's elsewhere. Use $\mathcal{U} = \{u_1, \dots, u_k\}$ to denote the k systematic nodes, and set $\mathcal{F}_i = \{u_{(i-1)h+1}, \dots, u_{ih}\}$ for $1 \leq i \leq \lfloor \frac{k}{h} \rfloor$. Define a bipartite graph where one set of vertices is the standard basis vectors $\{e_j : j \in [\ell]\}$, and the other set of vertices is the constant repair subspaces $\langle S_{\mathcal{F}_i} \rangle$ for $1 \leq i \leq \lfloor \frac{k}{h} \rfloor$. An edge exists between a vertex e_j and a subspace $\langle S_{\mathcal{F}_i} \rangle$ if and only if $e_j \in \langle S_{\mathcal{F}_i} \rangle$.

Now, count the total number of edges in this bipartite graph using two different methods. From the optimal-access property of the repair schemes we see that there are exactly $\frac{h\ell}{d-k+h}$ nonzero columns in the repair matrix $S_{\mathcal{F}_i}$. Hence the degree of each repairing subspace in the graph is $\frac{h\ell}{d-k+h}$. Thus, the total number of edges in the graph is $\lfloor \frac{k}{h} \rfloor \cdot \frac{h\ell}{d-k+h}$. On the other hand, let $\mathcal{N}_j = \{i : e_j \in \langle S_{\mathcal{F}_i} \rangle\}$ be the index set of repairing subspaces which contain e_j . Then by Lemma 3,

$$1 \leq \dim\left(\bigcap_{i \in \mathcal{N}_j} \langle S_{\mathcal{F}_i} \rangle\right) \leq \left(\frac{h}{d - k + h}\right)^{|\mathcal{N}_j|} \ell. \tag{6}$$

It follows that the degree of each standard basis vector in the graph satisfies

$$|\mathcal{N}_j| \leq \left\lfloor \log_{\frac{d-k+h}{h}} \ell \right\rfloor. \tag{7}$$

Hence there are at most $\ell \cdot \left\lfloor \log_{\frac{d-k+h}{h}} \ell \right\rfloor$ edges in the graph. In summary we have

$$\left\lfloor \frac{k}{h} \right\rfloor \cdot \frac{h\ell}{d - k + h} \leq \ell \cdot \left\lfloor \log_{\frac{d-k+h}{h}} \ell \right\rfloor,$$

and

$$\ell \geq \left(\frac{d - k + h}{h}\right)^{\left\lceil \left\lfloor \frac{k}{h} \right\rfloor \frac{h}{d - k + h} \right\rceil}.$$

□

Proof of Theorem 2. If $\ell \geq \left(\frac{d-k+h}{h}\right)^{\left\lfloor \frac{k}{h} \right\rfloor}$, then the conclusion holds. Hence we assume that $\ell < \left(\frac{d-k+h}{h}\right)^{\left\lfloor \frac{k}{h} \right\rfloor}$. Set $\mathcal{F}_i = \{(i-1)h+1, \dots, ih\}$ for $1 \leq i \leq \lfloor \frac{n}{h} \rfloor$. Define a bipartite graph where one set of vertices is the standard basis vectors $\{e_j : j \in [\ell]\}$, and the other set of vertices is the constant repair subspaces $\langle S_{\mathcal{F}_i} \rangle$, $1 \leq i \leq \lfloor \frac{n}{h} \rfloor$. An edge exists between a vertex e_j and a subspace $\langle S_{\mathcal{F}_i} \rangle$ if and only if $e_j \in \langle S_{\mathcal{F}_i} \rangle$.

Count the total number of edges in this bipartite graph using two different methods. Similar to the proof of Theorem 1, there are exactly $\lfloor \frac{n}{h} \rfloor \cdot \frac{h\ell}{d-k+h}$ edges in the graph by optimal-access property. On the other hand, we first prove that for all $\mathcal{E} \subset \{1, \dots, \lfloor \frac{n}{h} \rfloor\}$,

$$\dim \left(\bigcap_{i \in \mathcal{E}} \langle S_{\mathcal{F}_i} \rangle \right) \leq \left(\frac{h}{d-k+h} \right)^{|\mathcal{E}|}. \quad (8)$$

For $|\mathcal{E}| \leq \lfloor \frac{k}{h} \rfloor$, we can always choose a collection of k systematic nodes which contains $\bigcup_{i \in \mathcal{E}} \mathcal{F}_i$. Then the result follows directly from Lemma 3. For $\lfloor \frac{k}{h} \rfloor < |\mathcal{E}| \leq \lfloor \frac{n}{h} \rfloor$, we choose any subset \mathcal{H} of \mathcal{E} with size $\lfloor \frac{k}{h} \rfloor$, and get

$$\begin{aligned} \dim \left(\bigcap_{i \in \mathcal{E}} \langle S_{\mathcal{F}_i} \rangle \right) &\leq \dim \left(\bigcap_{i \in \mathcal{H}} \langle S_{\mathcal{F}_i} \rangle \right) \\ &\leq \left(\frac{h}{d-k+h} \right)^{\lfloor \frac{k}{h} \rfloor} \ell \\ &< 1. \end{aligned}$$

Therefore (8) holds for $1 \leq |\mathcal{E}| \leq \lfloor \frac{n}{h} \rfloor$. Just as (6) and (7), there are at most $\ell \cdot \left\lfloor \log_{\frac{d-k+h}{h}} \ell \right\rfloor$ edges in the graph. We have

$$\lfloor \frac{n}{h} \rfloor \cdot \frac{h\ell}{d-k+h} \leq \ell \cdot \left\lfloor \log_{\frac{d-k+h}{h}} \ell \right\rfloor,$$

and

$$\ell \geq \left(\frac{d-k+h}{h} \right)^{\left\lceil \lfloor \frac{n}{h} \rfloor \frac{h}{d-k+h} \right\rceil}.$$

In summary we have

$$l \geq \min \left\{ \left(\frac{d-k+h}{h} \right)^{\lfloor \frac{k}{h} \rfloor}, \left(\frac{d-k+h}{h} \right)^{\left\lceil \lfloor \frac{n}{h} \rfloor \frac{h}{d-k+h} \right\rceil} \right\}.$$

□

IV. A LOWER BOUND WITH REPAIR MATRICES INDEPENDENT OF IDENTITY OF REMAINING HELPER NODES

In this section we consider optimal-access (h, d) centralized MSR codes with repair matrices independent of the choice of the remaining $(d-1)$ helper nodes, namely, $S_{j \rightarrow \mathcal{F}}^{\mathcal{R}} = S_{j \rightarrow \mathcal{F}}$, for $\mathcal{F} \subseteq [n]$, $\mathcal{R} \subseteq [n] \setminus \mathcal{F}$ and $j \in \mathcal{R}$.

Lemma 4. Suppose that \mathcal{C} is an optimal-access $(h, n-h)$ -centralized MSR code. Let $1 \leq t \leq \lfloor \frac{k-1}{h} \rfloor$, and let $\mathcal{F}_1, \dots, \mathcal{F}_t \subseteq [n]$ be t pairwise disjoint subsets of $[n]$ of size h . Then for any $j \in [n] \setminus \bigcup_{i=1}^t \mathcal{F}_i$, we have

$$\dim \left(\bigcap_{i=1}^t \langle S_{j \rightarrow \mathcal{F}_i} \rangle \right) \leq \left(\frac{h}{r} \right)^t \ell. \quad (9)$$

Proof. First we choose a collection of $k-th$ nodes \mathcal{V} from $[n] \setminus \bigcup_{i=1}^t \mathcal{F}_i$. Now we regard the nodes in $\mathcal{U} = \bigcup_{i=1}^t \mathcal{F}_i \cup \mathcal{V}$ as the systematic nodes, and the other nodes $\mathcal{P} = \{p_1, \dots, p_r\} = [n] \setminus \mathcal{U}$ the parity nodes. Then we have $[n] \setminus \bigcup_{i=1}^t \mathcal{F}_i = \mathcal{V} \cup \mathcal{P}$.

For any $v \in \mathcal{V}$ and $p \in \mathcal{P}$, we have

$$\dim \left(\bigcap_{i=1}^t \langle S_{p \rightarrow \mathcal{F}_i} \rangle \right) = \dim \left(\bigcap_{i=1}^t \langle S_{p \rightarrow \mathcal{F}_i} A_{p,v} \rangle \right) \quad (10)$$

$$= \dim \left(\bigcap_{i=1}^t \langle S_{v \rightarrow \mathcal{F}_i} \rangle \right). \quad (11)$$

Here (10) follows from that $A_{p,v}$ is invertible and (11) follows from 2) of Lemma 1. Therefore the dimension of the subspaces $\bigcap_{i=1}^t \langle S_{j \rightarrow \mathcal{F}_i} \rangle$ are the same for all nodes $j \in \mathcal{V} \cup \mathcal{P}$.

Then we prove the result by induction on t . For $t=1$, by 3) of Lemma 1, we have

$$\dim(\langle S_{j \rightarrow \mathcal{F}_1} \rangle) = \frac{h}{r} \ell \quad \text{for } j \in \mathcal{V} \cup \mathcal{P}.$$

Assume the result holds for t , and now we will prove it also holds for $t + 1$. From above, it suffices to prove the result for $j = p_1$.

For any node $p \in \mathcal{P}$, fix a basis for the subspace $\bigcap_{i=1}^{t+1} \langle S_{p \rightarrow \mathcal{F}_i} \rangle$, and set S_p to be the matrix whose rows are the basis vectors. Then $\langle S_p \rangle = \bigcap_{i=1}^{t+1} \langle S_{p \rightarrow \mathcal{F}_i} \rangle$, and $\langle S_p \rangle \subseteq \langle S_{p \rightarrow \mathcal{F}_i} \rangle$ for $i \in [t + 1]$. Hence we can find matrices $T_{p,i}$, whose rank is equal to $\dim(\langle S_p \rangle)$, such that $S_p = T_{p,i} S_{p \rightarrow \mathcal{F}_i}$. It follows that

$$\begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) = \begin{bmatrix} T_{p_1,i} & & \\ & \ddots & \\ & & T_{p_r,i} \end{bmatrix} S_{P \rightarrow \mathcal{F}_i} A(:, \mathcal{F}_1),$$

for $i \in [t + 1]$. If $i = 1$, then by 1) of Lemma 1 we see that the matrix $S_{P \rightarrow \mathcal{F}_1} A(:, \mathcal{F}_1)$ is invertible. Then

$$\begin{aligned} \dim \left\langle \begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) \right\rangle &= \dim \left\langle \begin{bmatrix} T_{p_1,1} & & \\ & \ddots & \\ & & T_{p_r,1} \end{bmatrix} \right\rangle \\ &= \sum_{i=1}^r \text{rank}(T_{p_i,1}) \\ &= r \cdot \dim(\langle S_{p_1} \rangle). \end{aligned} \tag{12}$$

On the other hand, for $2 \leq i \leq t + 1$ we have

$$\left\langle \begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) \right\rangle \subseteq \langle S_{P \rightarrow \mathcal{F}_i} A(:, \mathcal{F}_1) \rangle = \langle S_{\mathcal{F}_1 \rightarrow \mathcal{F}_i} \rangle.$$

Here the latter equality follows from (2) of Lemma 2. Then

$$\left\langle \begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) \right\rangle \subseteq \bigcap_{i=2}^{t+1} \langle S_{P \rightarrow \mathcal{F}_i} A(:, \mathcal{F}_1) \rangle = \bigcap_{i=2}^{t+1} \langle S_{\mathcal{F}_1 \rightarrow \mathcal{F}_i} \rangle.$$

By assumption we have

$$\begin{aligned} \dim \left(\left\langle \begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) \right\rangle \right) &\leq \dim \left(\bigcap_{i=2}^{t+1} \langle S_{\mathcal{F}_1 \rightarrow \mathcal{F}_i} \rangle \right) \\ &= h \cdot \dim \left(\bigcap_{i=2}^{t+1} \langle S_{u \rightarrow \mathcal{F}_i} \rangle \right) \\ &\leq h \left(\frac{h}{r} \right)^t \ell, \end{aligned} \tag{13}$$

for any $u \in \mathcal{F}_1$. By (12) and (13), we can compute that

$$\dim(\langle S_{p_1} \rangle) \leq \left(\frac{h}{r} \right)^{t+1} \ell.$$

□

Proof of Theorem 3. If $\ell \geq \left(\frac{d-k+h}{h} \right)^{\lfloor \frac{k-1}{h} \rfloor}$, then the conclusion holds. Hence we assume that $\ell < \left(\frac{d-k+h}{h} \right)^{\lfloor \frac{k-1}{h} \rfloor}$. Set $\mathcal{F}_i = \{(i-1)h+1, \dots, ih\}$ for $1 \leq i \leq \lfloor \frac{n-1}{h} \rfloor$. Define a bipartite graph where one set of vertices is the standard basis vectors $\{e_j : j \in [\ell]\}$, and the other set of vertices is repair subspaces $\langle S_{n \rightarrow \mathcal{F}_i} \rangle$, $1 \leq i \leq \lfloor \frac{n-1}{h} \rfloor$. An edge exists between a vertex e_j and a subspace $\langle S_{n \rightarrow \mathcal{F}_i} \rangle$ if and only if $e_j \in \langle S_{n \rightarrow \mathcal{F}_i} \rangle$.

Count the total number of edges using two different methods. Similar to the proof of Theorem 1, there are exactly $\lfloor \frac{n-1}{h} \rfloor \cdot \frac{h\ell}{d-k+h}$ edges in the graph. On the other hand, we first prove that for all $\mathcal{E} \subset \{1, \dots, \lfloor \frac{n-1}{h} \rfloor\}$ and $|\mathcal{E}| \leq \lfloor \frac{k-1}{h} \rfloor$,

$$\dim \left(\bigcap_{i \in \mathcal{E}} \langle S_{n \rightarrow \mathcal{F}_i} \rangle \right) \leq \left(\frac{h}{d-k+h} \right)^{|\mathcal{E}|}.$$

It holds for $d = n - h$ by Lemma 4. For general d , we puncture some $n - d - h$ nodes from $[n - 1] \setminus \bigcup_{i \in \mathcal{E}} \mathcal{F}_i$, and get a new code \mathcal{C}' with length $n' = d + h$, dimension $k' = k$ and $d = n' - h$. The result follows by applying Lemma 4 to \mathcal{C}' .

For $\lfloor \frac{k-1}{h} \rfloor < |\mathcal{E}| \leq \lfloor \frac{n-1}{h} \rfloor$, we choose any subset \mathcal{H} of \mathcal{E} with size $\lfloor \frac{k-1}{h} \rfloor$, and obtain

$$\begin{aligned} \dim \left(\bigcap_{i \in \mathcal{E}} \langle S_{n \rightarrow \mathcal{F}_i} \rangle \right) &\leq \dim \left(\bigcap_{i \in \mathcal{H}} \langle S_{n \rightarrow \mathcal{F}_i} \rangle \right) \\ &\leq \left(\frac{h}{d - k + h} \right)^{\lfloor \frac{k-1}{h} \rfloor} \ell \\ &< 1. \end{aligned}$$

Therefore,

$$\dim \left(\bigcap_{i \in \mathcal{E}} \langle S_{n \rightarrow \mathcal{F}_i} \rangle \right) \leq \left(\frac{h}{d - k + h} \right)^{|\mathcal{E}|} \ell \quad \text{for } 1 \leq |\mathcal{E}| \leq \left\lfloor \frac{n-1}{h} \right\rfloor.$$

Let $\mathcal{N}_j = \{i : e_j \in \langle S_{n \rightarrow \mathcal{F}_i} \rangle\}$ be the index set of repairing subspaces which contain e_j . Just as (6) and (7), there are at most $\ell \cdot \left\lfloor \log \frac{d-k+h}{h} \ell \right\rfloor$ edges in the graph. We have

$$\left\lfloor \frac{n-1}{h} \right\rfloor \cdot \frac{h\ell}{d-k+h} \leq \ell \cdot \left\lfloor \log \frac{d-k+h}{h} \ell \right\rfloor,$$

and

$$\ell \geq \left(\frac{d-k+h}{h} \right)^{\left\lceil \left\lfloor \frac{n-1}{h} \right\rfloor \frac{h}{d-k+h} \right\rceil}.$$

In summary we have

$$l \geq \min \left\{ \left(\frac{d-k+h}{h} \right)^{\lfloor \frac{k-1}{h} \rfloor}, \left(\frac{d-k+h}{h} \right)^{\left\lceil \left\lfloor \frac{n-1}{h} \right\rfloor \frac{h}{d-k+h} \right\rceil} \right\}.$$

□

Corollary 1. *Let \mathcal{C} be an optimal-access $(h, n - h)$ centralized MSR code. Then*

$$l \geq \min \left\{ \left(\frac{r}{h} \right)^{\left\lceil \frac{k-1}{h} \right\rceil}, \left(\frac{r}{h} \right)^{\left\lceil \left\lfloor \frac{n-1}{h} \right\rfloor \frac{h}{r} \right\rceil} \right\}.$$

Proof. The result follows directly by setting $d = n - h$ in Theorem 3. □

V. IMPROVING BOUNDS USING GENERATING FUNCTIONS

In this section, we employ generating functions to perform a more refined analysis, further enhancing the bounds in Sections III-IV.

A. Generating Functions

Fix an integer c with $1 \leq c < h$. Let a ($2 \leq a \leq h$) be the unique integer such that $c \in \left(\frac{(a-2)h}{a-1}, \frac{(a-1)h}{a} \right]$. We define a strictly decreasing sequence $\{B_n\}_{n \geq 0}$ by the following recurrence formula

$$\begin{cases} B_0 = 1, \\ B_1 = \frac{h}{d-k+h}, \\ B_j = \frac{h-c}{d-k+h} B_{j-1} + \dots + \frac{h-c}{d-k+h} B_1 + \frac{(j-1)c-(j-2)h}{d-k+h} B_0, \quad j \in [a-1], \\ B_{n+a} = \frac{h-c}{d-k+h} B_{n+a-1} + \dots + \frac{h-c}{d-k+h} B_{n+1} + \frac{(a-1)c-(a-2)h}{d-k+h} B_n, \quad n \in \mathbb{N}. \end{cases} \quad (14)$$

Let $B = \sum_{n \geq 0} B_n x^n$ be the generating function of the above recurrence formula. Then it is not hard to check that

$$\frac{B - \sum_{i=0}^{a-1} B_i x^i}{x^a} = \frac{h-c}{d-k+h} \frac{B - \sum_{i=0}^{a-2} B_i x^i}{x^{a-1}} + \dots + \frac{h-c}{d-k+h} \frac{B - B_0}{x} + \frac{(a-1)c-(a-2)h}{d-k+h} B.$$

Via direct calculation, we get

$$B = \frac{\sum_{i=0}^{a-1} B_i x^i - \frac{h-c}{d-k+h} \sum_{i=0}^{a-2} B_i x^{i+1} - \dots - \frac{h-c}{d-k+h} B_0 x^{a-1}}{1 - \frac{h-c}{d-k+h} x - \dots - \frac{h-c}{d-k+h} x^{a-1} - \frac{(a-1)c-(a-2)h}{d-k+h} x^a}. \quad (15)$$

Write

$$\phi(x) = \theta_1 x^a + \theta_2 x^{a-1} + \dots + \theta_2 x + 1, \quad (16)$$

where $\theta_1 = -\frac{(a-1)c-(a-2)h}{d-k+h}$, $\theta_2 = \frac{c-h}{d-k+h}$. By calculating roots of $\phi(x)$ over \mathbb{C} , we can expand B as a formal power series and then get the expression of B_n . In particular if the polynomial $\phi(x)$ has only simple roots x_1, \dots, x_a , namely,

$$\phi(x) = \theta_1 \prod_{i=1}^a (x - x_i),$$

we can compute that

$$B = \frac{1}{\theta_1} \sum_{n \geq 0} \left(\sum_{i=1}^a \left(\frac{1}{x_i^{n-a+2} \prod_{j \neq i} (x_i - x_j)} \left(\sum_{u=0}^{a-1} \left(\frac{-1}{x_i^{a-u-1}} + \frac{h-c}{d-k+h} \sum_{v=0}^{a-u-2} \frac{1}{x_i^v} \right) B_u \right) \right) \right) x^n. \quad (17)$$

If $1 \leq c \leq \lfloor \frac{h}{2} \rfloor$, then we have $a = 2$, and

$$B = \frac{B_0 + (B_1 - \frac{h-c}{d-k+h} B_0)x}{1 - \frac{h-c}{d-k+h} x - \frac{c}{d-k+h} x^2}.$$

Now the polynomial $\phi(x)$ is quadratic and $\Delta = \Delta(c) := \left(\frac{h-c}{d-k+h} \right)^2 + \frac{4c}{d-k+h} > 0$, hence $\phi(x)$ has two different roots:

$$\alpha = \alpha(c) := \frac{(c-h) + \sqrt{\Delta}(d-k+h)}{2c}, \quad \gamma = \gamma(c) := \frac{(c-h) - \sqrt{\Delta}(d-k+h)}{2c}.$$

By (17), we have

$$B = \frac{1}{\sqrt{\Delta}} \sum_{n \geq 0} \left(\left(\frac{1}{\alpha} + \frac{c}{d-k+h} \right) \frac{1}{\alpha^n} + \left(-\frac{1}{\gamma} - \frac{c}{d-k+h} \right) \frac{1}{\gamma^n} \right) x^n.$$

As $1 < |\alpha| < |\gamma|$, we get

$$\begin{aligned} B_n &\leq \frac{2}{\sqrt{\Delta}} \left(\frac{1}{\alpha} + \frac{c}{d-k+h} \right) \frac{1}{\alpha^n} \\ &\leq \frac{4}{\sqrt{\Delta}} \frac{1}{\alpha^n}. \end{aligned} \quad (18)$$

B. A Lower Bound with Constant Repair Matrices

First, we consider repair schemes with constant repair matrices. The following result generalizes Lemma 3 by allowing the subsets \mathcal{F}_i to have nonempty intersections.

Lemma 5. Suppose that \mathcal{C} has (h, d) optimal repair schemes with constant repair matrices. Fix an integer c with $1 \leq c < h$. Let a ($2 \leq a \leq h$) be the unique integer such that $c \in \left(\frac{(a-2)h}{a-1}, \frac{(a-1)h}{a} \right]$. Let $\mathcal{F}_i = \{(i-1)(h-c) + 1, \dots, (i-1)(h-c) + h\}$, $1 \leq i \leq \lfloor \frac{n-h}{h-c} \rfloor + 1$. Then, for any $\mathcal{E} \subseteq [\lfloor \frac{n-h}{h-c} \rfloor + 1]$ with $1 \leq |\mathcal{E}| \leq \lfloor \frac{k-h}{h-c} \rfloor + 1$, we have

$$\dim \left(\bigcap_{i \in \mathcal{E}} \langle S_{\mathcal{F}_i} \rangle \right) \leq B_{|\mathcal{E}|} \ell, \quad (19)$$

where $B_{|\mathcal{E}|}$ is the $|\mathcal{E}|$ -th term of the sequence $\{B_n\}_{n \geq 0}$ defined above.

Proof. Write $t = |\mathcal{E}|$. We first assume that \mathcal{E} is consisting of some t consecutive integers. Without loss of generality, we assume that systematic nodes set $\mathcal{U} = [k]$ and $\mathcal{E} = \{1, 2, \dots, t\}$.

If $d = n - h$, we can prove the result by induction on t . For $t = 1$, the conclusion holds as $\dim(\langle S_{\mathcal{F}_1} \rangle) = \frac{h\ell}{r}$. Assume that the result holds for t , and now we will prove it also holds for $t + 1$. Just as in the proof of Lemma 3, there exist matrices S and T_i of the same rank such that $\langle S \rangle = \bigcap_{i=1}^{t+1} \langle S_{\mathcal{F}_i} \rangle$ and $S = T_i S_{\mathcal{F}_i}$ for $i \in [t + 1]$. It follows that

$$(I_r \otimes S)A(:, \mathcal{F}_1) = (I_r \otimes T_i)(I_r \otimes S_{\mathcal{F}_i})A(:, \mathcal{F}_1), \text{ for } i \in [t + 1].$$

If $i = 1$, then by 1) of Lemma 1 we see that the matrix $(I_r \otimes S_{\mathcal{F}_1})A(:, \mathcal{F}_1)$ is invertible. Then

$$\dim(\langle (I_r \otimes S)A(:, \mathcal{F}_1) \rangle) = \dim(\langle I_r \otimes T_1 \rangle) = r \dim(\langle S \rangle). \quad (20)$$

On the other hand, for $2 \leq i \leq t+1$, we have $\langle (I_r \otimes S)A(:, \mathcal{F}_1) \rangle \subseteq \langle (I_r \otimes S_{\mathcal{F}_i})A(:, \mathcal{F}_1) \rangle$. Then by (3) of Lemma 2, we get

$$\begin{aligned} \langle (I_r \otimes S_{\mathcal{F}_i})A(:, \mathcal{F}_1) \rangle &\subseteq \left\langle \begin{bmatrix} I_{|\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_i)|} \otimes S_{\mathcal{F}_i} & \mathbf{O} \\ \mathbf{O} & (I_r \otimes S_{\mathcal{F}_i})A(:, \mathcal{F}_1 \cap \mathcal{F}_i) \end{bmatrix} \right\rangle \\ &\subseteq \langle [I_{|\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_i)|} \otimes S_{\mathcal{F}_i} \quad \mathbf{O}] \rangle + \langle [\mathbf{O} \quad (I_r \otimes S_{\mathcal{F}_i})A(:, \mathcal{F}_1 \cap \mathcal{F}_i)] \rangle. \end{aligned}$$

So, we can obtain

$$\begin{aligned} \langle (I_r \otimes S)A(:, \mathcal{F}_1) \rangle &\subseteq \bigcap_{i=2}^{t+1} \left\langle \begin{bmatrix} I_{|\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_i)|} \otimes S_{\mathcal{F}_i} & \mathbf{O} \\ \mathbf{O} & (I_r \otimes S_{\mathcal{F}_i})A(:, \mathcal{F}_1 \cap \mathcal{F}_i) \end{bmatrix} \right\rangle \\ &\subseteq \bigcap_{i=2}^{t+1} (\langle [I_{|\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_i)|} \otimes S_{\mathcal{F}_i} \quad \mathbf{O}] \rangle + \langle [\mathbf{O} \quad (I_r \otimes S_{\mathcal{F}_i})A(:, \mathcal{F}_1 \cap \mathcal{F}_i)] \rangle). \end{aligned} \quad (21)$$

If $t \in [a-1]$, then we see that the intersection $\mathcal{F}_j \cap \mathcal{F}_1$ is nonempty for $j \in \{2, \dots, t+1\}$, as

$$(j-1)(h-c) + 1 < (a-1) \left(h - \frac{(a-2)h}{a-1} \right) + 1 = h+1.$$

By (21) we have

$$\dim(\langle (I_r \otimes S)A(:, \mathcal{F}_1) \rangle) \leq (h-c) \sum_{i=2}^{t+1} \dim \left(\bigcap_{j=i}^{t+1} \langle S_{\mathcal{F}_j} \rangle \right) + (tc - (t-1)h)\ell.$$

Combine (20) and (21), we can compute that

$$\begin{aligned} \dim(\langle S \rangle) &\leq \frac{h-c}{r} \sum_{i=1}^t B_i \ell + \frac{tc - (t-1)h}{r} B_0 \ell \\ &= \left(\frac{h-c}{r} \sum_{i=1}^t B_i + \frac{tc - (t-1)h}{r} B_0 \right) \ell \\ &= B_{t+1} \ell, \end{aligned}$$

and the conclusion holds.

If $t \geq a$, then we see that the intersection $\mathcal{F}_j \cap \mathcal{F}_1$ is empty for $j \geq a+1$, as

$$(j-1)(h-c) + 1 \geq a \left(h - \frac{(a-1)h}{a} \right) + 1 = h+1.$$

We have

$$\dim(\langle (I_r \otimes S)A(:, \mathcal{F}_1) \rangle) \leq (h-c) \sum_{i=2}^a \dim \left(\bigcap_{j=i}^{t+1} \langle S_{\mathcal{F}_j} \rangle \right) + \left(((a-1)c - (a-2)h) \bigcap_{i=a+1}^{t+1} \langle S_{\mathcal{F}_i} \rangle \right). \quad (22)$$

Combine (20) and (22), we can compute that

$$\begin{aligned} \dim(\langle S \rangle) &\leq \frac{h-c}{r} \sum_{i=t-a+2}^t B_i \ell + \frac{(a-1)c - (a-2)h}{r} B_{t+1-a} \ell \\ &= \left(\frac{h-c}{r} \sum_{i=t-a+2}^t B_i + \frac{(a-1)c - (a-2)h}{r} B_{t+1-a} \right) \ell \\ &= B_{t+1}. \end{aligned}$$

This concludes our proof for the case $d = n - h$. For general d , we first puncture some $n - d - h$ parity nodes from \mathcal{C} and get a new code \mathcal{C}' with $n' = d + h$ and $r' = d - k + h$. Applying the above argument to \mathcal{C}' we get the conclusion.

For the case that \mathcal{E} does not consist of t consecutive integers, it is not hard to prove the result similarly as above, and we omit the details.

□

Proof of Theorem 4. If $\ell \geq \frac{\sqrt{\Delta}}{4} \alpha^{\lfloor \frac{k-h}{h-c} \rfloor + 1}$, then the conclusion holds. Hence we assume that $\ell < \frac{\sqrt{\Delta}}{4} \alpha^{\lfloor \frac{k-h}{h-c} \rfloor + 1}$. Set $\mathcal{F}_i = \{(i-1)(h-c)+1, \dots, (i-1)(h-c)+h\}$ for $1 \leq i \leq \lfloor \frac{n-h}{h-c} \rfloor + 1$. Define a bipartite graph where one set of vertices is the standard basis vectors $\{e_j : j \in [\ell]\}$, and the other set of vertices is the constant repair subspaces $\langle S_{\mathcal{F}_i} \rangle, 1 \leq i \leq \lfloor \frac{n-h}{h-c} \rfloor + 1$. An edge exists between a vertex e_j and a subspace $\langle S_{\mathcal{F}_i} \rangle$ if and only if $e_j \in \langle S_{\mathcal{F}_i} \rangle$.

Count the total number of edges in this bipartite graph using two different methods. Similar to the proof of Theorem 1, there are exactly $\left(\lfloor \frac{n-h}{h-c} \rfloor + 1\right) \cdot \frac{h\ell}{d-k+h}$ edges in the graph. On the other hand, we first prove that for any $\mathcal{E} \subseteq [\lfloor \frac{n-h}{h-c} \rfloor + 1]$,

$$\dim\left(\bigcap_{i \in \mathcal{E}} \langle S_{\mathcal{F}_i} \rangle\right) \leq \frac{4}{\sqrt{\Delta}} \frac{1}{\alpha^n} \ell. \quad (23)$$

By (18), we have

$$B_n \leq \frac{4}{\sqrt{\Delta}} \frac{1}{\alpha^n}.$$

For $\left(\lfloor \frac{k-h}{h-c} \rfloor + 1\right) < |\mathcal{E}| \leq \left(\lfloor \frac{n-h}{h-c} \rfloor + 1\right)$, we choose any subset \mathcal{H} of \mathcal{E} with size $\lfloor \frac{k-h}{h-c} \rfloor + 1$. By Lemma 5, we obtain

$$\begin{aligned} \dim\left(\bigcap_{i \in \mathcal{E}} \langle S_{\mathcal{F}_i} \rangle\right) &\leq \dim\left(\bigcap_{i \in \mathcal{H}} \langle S_{\mathcal{F}_i} \rangle\right) \\ &\leq \frac{4}{\sqrt{\Delta}} \frac{1}{\alpha^{\lfloor \frac{k-h}{h-c} \rfloor + 1}} \ell \\ &< 1. \end{aligned}$$

Combining this with Lemma 5, we conclude that (23) holds for $\mathcal{E} \subseteq [\lfloor \frac{n-h}{h-c} \rfloor + 1]$. Let $\mathcal{N}_j = \{i : e_j \in \langle S_{\mathcal{F}_i} \rangle\}$ be the index set of repairing subspaces which contain e_j . Then we have

$$1 \leq \dim\left(\bigcap_{i \in \mathcal{N}_j} \langle S_{\mathcal{F}_i} \rangle\right) \leq \frac{4}{\sqrt{\Delta}} \frac{1}{\alpha^{|\mathcal{N}_j|}} \ell. \quad (24)$$

As $\alpha \geq 1$, it follows that

$$|\mathcal{N}_j| \leq \left\lceil \log_{\alpha} \left(\frac{4}{\sqrt{\Delta}} \ell \right) \right\rceil. \quad (25)$$

Hence there are at most $\ell \cdot \left\lceil \log_{\alpha} \left(\frac{4}{\sqrt{\Delta}} \ell \right) \right\rceil$ edges in the graph. We have

$$\left(\left\lfloor \frac{n-h}{h-c} \right\rfloor + 1\right) \cdot \frac{h\ell}{d-k+h} \leq \ell \cdot \left\lceil \log_{\alpha} \left(\frac{4}{\sqrt{\Delta}} \ell \right) \right\rceil,$$

and

$$\ell \geq \frac{\sqrt{\Delta}}{4} \alpha^{\lceil (\lfloor \frac{n-h}{h-c} \rfloor + 1) \cdot \frac{h}{d-k+h} \rceil}.$$

In summary we have

$$l \geq \min \left\{ \frac{\sqrt{\Delta}}{4} \alpha^{\lfloor \frac{k-h}{h-c} \rfloor + 1}, \frac{\sqrt{\Delta}}{4} \alpha^{\lceil (\lfloor \frac{n-h}{h-c} \rfloor + 1) \cdot \frac{h}{d-k+h} \rceil} \right\}.$$

□

C. A Lower Bound with Repair Matrices Independent of Identity of Remaining Helper Nodes

In this section we consider optimal-access (h, d) centralized MSR codes with repair matrices independent of the choice of the remaining $(d-1)$ helper nodes. This approach generalizes the bound in Lemma 3.

Lemma 6. Let \mathcal{C} be an optimal-access (h, d) -centralized MSR code with repair matrices independent of the choice of the remaining $(d-1)$ helper nodes. Fix an integer c with $1 \leq c < h$. Let a ($2 \leq a \leq h$) be the unique integer such that

$c \in \left(\frac{(a-2)h}{a-1}, \frac{(a-1)h}{a} \right]$. Let $\mathcal{F}_i = \{(i-1)(h-c) + 1, \dots, (i-1)(h-c) + h\}$, $1 \leq i \leq \lfloor \frac{n-1-h}{h-c} \rfloor + 1$. Then, for any $\mathcal{E} \subseteq [\lfloor \frac{n-1-h}{h-c} \rfloor + 1]$ and $1 \leq |\mathcal{E}| \leq \lfloor \frac{k-1-h}{h-c} \rfloor + 1$, we have

$$\dim\left(\bigcap_{i \in \mathcal{E}} \langle S_{n \rightarrow \mathcal{F}_i} \rangle\right) \leq B_{|\mathcal{E}|} \ell, \quad (26)$$

where $B_{|\mathcal{E}|}$ is the $|\mathcal{E}|$ -th term of the sequence $\{B_n\}_{n \geq 0}$ defined in (14).

Proof. Write $t = |\mathcal{E}|$. First we choose a collection of $k - (t-1)(h-c) - h$ nodes \mathcal{V} from $[n-1] \setminus \bigcup_{i \in \mathcal{E}} \mathcal{F}_i$. Now we regard the nodes in $\mathcal{U} = \bigcup_{i \in \mathcal{E}} \mathcal{F}_i \cup \mathcal{V}$ as the systematic nodes, and the other nodes $\mathcal{P} = \{p_1, \dots, p_r\} = [n] \setminus \mathcal{U}$ the parity nodes. Note that $n \in \mathcal{P}$ and $[n] \setminus \bigcup_{i \in \mathcal{E}} \mathcal{F}_i = \mathcal{V} \cup \mathcal{P}$.

For any $v \in \mathcal{V}$ and $p \in \mathcal{P}$, we have

$$\dim\left(\bigcap_{i \in \mathcal{E}} \langle S_{p \rightarrow \mathcal{F}_i} \rangle\right) = \dim\left(\bigcap_{i \in \mathcal{E}} \langle S_{p \rightarrow \mathcal{F}_i} A_{p,v} \rangle\right) \quad (27)$$

$$= \dim\left(\bigcap_{i \in \mathcal{E}} \langle S_{v \rightarrow \mathcal{F}_i} \rangle\right). \quad (28)$$

Here (27) follows from that $A_{p,v}$ is invertible and (28) follows from 2) of Lemma 1. Therefore the dimension of the subspaces $\bigcap_{i \in \mathcal{E}} \langle S_{j \rightarrow \mathcal{F}_i} \rangle$ are the same for all nodes $j \in \mathcal{V} \cup \mathcal{P}$.

Then we prove the result by induction on t . First, we prove the lemma holds for \mathcal{E} when its elements are consecutive. Without loss of generality, assume $\mathcal{E} = \{1, 2, \dots, t\}$.

Assume that $d = n - h$ and prove the result by induction on t . For $t = 1$, the conclusion holds as $\dim(\langle S_{n \rightarrow \mathcal{F}_1} \rangle) = \frac{h\ell}{r}$. Suppose the result holds for t , and we will prove it also holds for $t + 1$. Just as in the proof of Lemma 4, there exist matrices S_p and $T_{p,i}$ of the same rank such that $\langle S_p \rangle = \bigcap_{i=1}^{t+1} \langle S_{p \rightarrow \mathcal{F}_i} \rangle$ and $S_p = T_{p,i} S_{p \rightarrow \mathcal{F}_i}$ for $p \in \mathcal{P}, i \in [t+1]$. It follows that

$$\begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) = \begin{bmatrix} T_{p_1,1} & & \\ & \ddots & \\ & & T_{p_r,1} \end{bmatrix} S_{P \rightarrow \mathcal{F}_1} A(:, \mathcal{F}_1),$$

for $i \in [t+1]$. If $i = 1$, then by 1) of Lemma 1 we see that the matrix $S_{P \rightarrow \mathcal{F}_1} A(:, \mathcal{F}_1)$ is invertible. Then

$$\begin{aligned} \dim \left\langle \begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) \right\rangle &= \dim \left\langle \begin{bmatrix} T_{p_1,1} & & \\ & \ddots & \\ & & T_{p_r,1} \end{bmatrix} \right\rangle \\ &= \sum_{i=1}^r \text{rank}(T_{p_i,1}) \\ &= r \cdot \dim(\langle S_n \rangle) \end{aligned} \quad (29)$$

On the other hand, for $2 \leq i \leq t+1$ we have

$$\begin{aligned} \left\langle \begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) \right\rangle &\subseteq \langle S_{P \rightarrow \mathcal{F}_i} A(:, \mathcal{F}_1) \rangle, \\ &\subseteq \left\langle \begin{bmatrix} S_{\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_i) \rightarrow \mathcal{F}_i} & \mathbf{O} \\ \mathbf{O} & S_{P \rightarrow \mathcal{F}_i} A(:, \mathcal{F}_1 \cap \mathcal{F}_i) \end{bmatrix} \right\rangle. \end{aligned}$$

Here the latter equality follows from (3) of Lemma 2. So, we can obtain

$$\begin{aligned} \left\langle \begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) \right\rangle &\subseteq \bigcap_{i=2}^{t+1} \left\langle \begin{bmatrix} S_{\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_i) \rightarrow \mathcal{F}_i} & \mathbf{O} \\ \mathbf{O} & S_{P \rightarrow \mathcal{F}_i} A(:, \mathcal{F}_1 \cap \mathcal{F}_i) \end{bmatrix} \right\rangle \\ &\subseteq \bigcap_{i=2}^{t+1} (\langle [S_{\mathcal{F}_1 \setminus (\mathcal{F}_1 \cap \mathcal{F}_i) \rightarrow \mathcal{F}_i} \ \mathbf{O}] \rangle + \langle [\mathbf{O} \ S_{P \rightarrow \mathcal{F}_i} A(:, \mathcal{F}_1 \cap \mathcal{F}_i)] \rangle). \end{aligned} \quad (30)$$

As shown in the proof of Lemma 5, we consider the following two cases.

1) If $t \in [a-1]$, we have

$$\dim \left\langle \begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) \right\rangle \leq (h-c) \sum_{i=2}^{t+1} \dim \left(\bigcap_{j=i}^{t+1} \langle S_{n \rightarrow \mathcal{F}_j} \rangle \right) + (tc - (t-1)h)\ell.$$

Combine (29) and (30), we can compute that

$$\begin{aligned} \dim(\langle S_n \rangle) &\leq \frac{h-c}{r} \sum_{i=1}^t B_i \ell + \frac{tc - (t-1)h}{r} B_0 \ell \\ &= \left(\frac{h-c}{r} \sum_{i=1}^t B_i + \frac{tc - (t-1)h}{r} B_0 \right) \ell \\ &= B_{t+1} \ell, \end{aligned}$$

and the conclusion holds.

2) If $t \geq a$, we have

$$\dim \left\langle \begin{bmatrix} S_{p_1} & & \\ & \ddots & \\ & & S_{p_r} \end{bmatrix} A(:, \mathcal{F}_1) \right\rangle \leq (h-c) \sum_{i=2}^a \dim \left(\bigcap_{j=i}^{t+1} \langle S_{n \rightarrow \mathcal{F}_j} \rangle \right) + \left((a-1)c - (a-2)h \right) \bigcap_{i=a+1}^{t+1} \langle S_{n \rightarrow \mathcal{F}_i} \rangle. \quad (31)$$

Combine (29) and (31), we can compute that

$$\begin{aligned} \dim(\langle S_n \rangle) &\leq \frac{h-c}{r} \sum_{i=t-a+2}^t B_i \ell + \frac{(a-1)c - (a-2)h}{r} B_{t+1-a} \ell, \\ &= \left(\frac{h-c}{r} \sum_{i=t-a+2}^t B_i + \frac{(a-1)c - (a-2)h}{r} B_{t+1-a} \right) \ell \\ &= B_{t+1} \ell. \end{aligned}$$

Thus, we have completed the proof for the case $d = n - h$. For general d , we puncture some $n - d - h$ nodes from $[n-1] \setminus \bigcup_{i \in \mathcal{E}} \mathcal{F}_i$, and get a new code \mathcal{C}' with length $n' = d + h$, dimension $k' = k$ and $d = n' - h$. Applying the above argument to \mathcal{C}' we get the conclusion.

For the case where \mathcal{E} does not consist of t consecutive integers, the result can be proven in a similar manner, and we omit the details. \square

Proof of Theorem 5. If $\ell \geq \frac{\sqrt{\Delta}}{4} \alpha^{\lfloor \frac{k-1-h}{h-c} \rfloor + 1}$, the conclusion holds. Hence we assume that $\ell < \frac{\sqrt{\Delta}}{4} \alpha^{\lfloor \frac{k-1-h}{h-c} \rfloor + 1}$. Let $\{e_j : j \in [\ell]\}$ be the standard basis of \mathbb{F}_q^ℓ . Set $\mathcal{F}_i = \{(i-1)(h-c) + 1, \dots, (i-1)(h-c) + h\}$ for $1 \leq i \leq \lfloor \frac{n-1-h}{h-c} \rfloor + 1$. Define a bipartite graph where one set of vertices is the standard basis vectors $\{e_j : j \in [\ell]\}$, and the other set of vertices is the repair subspaces $\langle S_{n \rightarrow \mathcal{F}_i} \rangle$, $1 \leq i \leq \lfloor \frac{n-1-h}{h-c} \rfloor + 1$. An edge exists between a vertex e_j and a subspace $\langle S_{n \rightarrow \mathcal{F}_i} \rangle$ if and only if $e_j \in \langle S_{n \rightarrow \mathcal{F}_i} \rangle$.

Count the total number of edges in this bipartite graph using two different methods. Similar to the proof of Theorem 1, there are $\left(\lfloor \frac{n-1-h}{h-c} \rfloor + 1 \right) \cdot \frac{h\ell}{d-k+h}$ edges in the graph. We next prove that for any $\mathcal{E} \subseteq [\lfloor \frac{n-1-h}{h-c} \rfloor + 1]$,

$$\dim \left(\bigcap_{i \in \mathcal{E}} \langle S_{n \rightarrow \mathcal{F}_i} \rangle \right) \leq \frac{4}{\sqrt{\Delta}} \frac{1}{\alpha^n} \ell. \quad (32)$$

By (18), we have

$$B_n \leq \frac{4}{\sqrt{\Delta}} \frac{1}{\alpha^n}. \quad (33)$$

For $\left(\lfloor \frac{k-h-1}{h-c} \rfloor + 1\right) < |\mathcal{E}| \leq \left(\lfloor \frac{n-1-h}{h-c} \rfloor + 1\right)$, we choose any subset \mathcal{H} of \mathcal{E} with size $\lfloor \frac{k-h-1}{h-c} \rfloor + 1$. By Lemma 6, we obtain

$$\begin{aligned} \dim \left(\bigcap_{i \in \mathcal{E}} \langle S_{n \rightarrow \mathcal{F}_i} \rangle \right) &\leq \dim \left(\bigcap_{i \in \mathcal{H}} \langle S_{n \rightarrow \mathcal{F}_i} \rangle \right) \\ &\leq \frac{4}{\sqrt{\Delta}} \frac{1}{\alpha^{\left(\lfloor \frac{k-1-h}{h-c} \rfloor + 1\right)}} \ell \\ &< 1. \end{aligned}$$

Combining this with Lemma 6, we conclude that (32) holds for $\mathcal{E} \subseteq [\lfloor \frac{n-1-h}{h-c} \rfloor + 1]$. Just as (24) and (25), there are at most $\ell \left\lceil \log_\alpha \left(\frac{4}{\sqrt{\Delta}} \ell \right) \right\rceil$ edges in the graph. We have

$$\left(\left\lfloor \frac{n-1-h}{h-c} \right\rfloor + 1 \right) \cdot \frac{h\ell}{d-k+h} \leq \ell \cdot \left\lceil \log_\alpha \left(\frac{4}{\sqrt{\Delta}} \ell \right) \right\rceil,$$

and

$$\ell \geq \frac{\sqrt{\Delta}}{4} \alpha^{\left\lceil \left(\left\lfloor \frac{n-1-h}{h-c} \right\rfloor + 1 \right) \cdot \frac{h}{d-k+h} \right\rceil}.$$

In summary we have

$$l \geq \min \left\{ \frac{\sqrt{\Delta}}{4} \alpha^{\left(\left\lfloor \frac{k-1-h}{h-c} \right\rfloor + 1 \right)}, \frac{\sqrt{\Delta}}{4} \alpha^{\left\lceil \left(\left\lfloor \frac{n-1-h}{h-c} \right\rfloor + 1 \right) \cdot \frac{h}{d-k+h} \right\rceil} \right\}.$$

□

VI. CONCLUSION

In this paper, we derive lower bounds on the sub-packetization of optimal-access MSR codes in the scenario of multiple-node failures. Besides, we offer a more detailed analysis by employing generating functions. Below, we present a comparison of these bounds. By setting $c = h/2$ in Theorem 4 and Theorem 5, we can compute that

$$\alpha = \frac{-\frac{h}{2} + \sqrt{\frac{h^2}{4} + 2h(d-k+h)}}{h},$$

which leads to the bound

$$\ell \geq \Theta \left(\alpha^{\frac{2n}{d-k+h}} \right).$$

On the other hand, the bounds in Theorem 2 and Theorem 3 give

$$\ell \geq \Theta \left(\left(\frac{d-k+h}{h} \right)^{\frac{n}{d-k+h}} \right).$$

It is not hard to check that $\alpha^2 > \frac{d-k+h}{h}$. Hence, for high-rate MSR codes and sufficiently large n , the bounds in Theorems 4 and 5 are greater than those in Theorems 2 and 3. Therefore, unlike the case when $h = 1$, the generalized bounds in Theorems 2 and 3 are generally not tight.

APPENDIX A

THE CUT-SET BOUND ON CENTRALIZED MSR CODES WITH CODE RATE NO LESS THAN 1/2

Proof. (The proof method is proposed by Alrabiah and Guruswami in [3]) Let \mathcal{C} be an (n, k, ℓ) MDS array code over \mathbb{F}_q . We write a codeword of \mathcal{C} as C_1, \dots, C_n . Give two integers h, d such that $2 \leq h \leq n-k$ and $k+1 \leq d \leq n-h$. For any set of h failed nodes $\mathcal{F} \subseteq [n]$ and any set of d helper nodes $\mathcal{R} \subseteq [n] \setminus \mathcal{F}$, the code \mathcal{C} will repair \mathcal{F} by having each C_j for $j \in \mathcal{R}$ transmit $\beta_{j,\mathcal{F}}$ bits to the single data center. Since code rate $\frac{k}{n} \geq \frac{1}{2}$, $k-h \geq k-(n-k) \geq 0$. By MDS property, any subset $M \subseteq \mathcal{R}$ of size $k-h$, together with h failed nodes in \mathcal{F} , can reconstruct the entire codeword. Thus

$$\sum_{i \in M} \ell + \sum_{j \in \mathcal{R} \setminus M} \beta_{j,\mathcal{F}} \geq k\ell. \quad (34)$$

Namely,

$$\sum_{j \in \mathcal{R} \setminus M} \beta_{j,\mathcal{F}} \geq h\ell. \quad (35)$$

By summing over all possible M , we get

$$\sum_{\substack{M \subseteq \mathcal{R}, \\ |M|=k-h}} \sum_{j \in \mathcal{R} \setminus M} \beta_{j,\mathcal{F}} \geq \sum_{\substack{M \subseteq \mathcal{R}, \\ |M|=k-h}} h\ell, \quad (36)$$

$$\sum_{j \in \mathcal{R}} \beta_{j,\mathcal{F}} \geq \frac{dhl}{d-k+h}. \quad (37)$$

Furthermore, to achieve equality, it is necessary for equality to hold in (35) for all M . It is not difficult to verify that $\beta_{j,\mathcal{F}} = \frac{hl}{d-k+h}$ for all $j \in \mathcal{R}$. \square

REFERENCES

- [1] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4539–4551, 2010.
- [2] S. Goparaju, I. Tamo, and R. Calderbank, "An improved sub-packetization bound for minimum storage regenerating codes," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2770–2779, 2014.
- [3] O. Alrabiah and V. Guruswami, "An exponential lower bound on the sub-packetization of MSR codes," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019*, (New York, NY, USA), p. 979–985, Association for Computing Machinery, 2019.
- [4] O. Alrabiah and V. Guruswami, "An exponential lower bound on the sub-packetization of minimum storage regenerating codes," *IEEE Transactions on Information Theory*, vol. 67, no. 12, pp. 8086–8093, 2021.
- [5] S. B. Balaji, M. Vajha, and P. Vijay Kumar, "Lower bounds on the sub-packetization level of MSR codes and characterizing optimal-access MSR codes achieving the bound," *IEEE Transactions on Information Theory*, vol. 68, no. 10, pp. 6452–6471, 2022.
- [6] I. Tamo, Z. Wang, and J. Bruck, "Access versus bandwidth in codes for storage," *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2028–2037, 2014.
- [7] B. Sasidharan, M. Vajha, and P. V. Kumar, "An explicit, coupled-layer construction of a high-rate msr code with low sub-packetization level, small field size and all-node repair," *ArXiv*, vol. abs/1607.07335, 2016.
- [8] G. Li, N. Wang, S. Hu, and M. Ye, "MSR codes with linear field size and smallest sub-packetization for any number of helper nodes," *IEEE Transactions on Information Theory*, vol. 70, no. 11, pp. 7790–7806, 2024.
- [9] M. Ye and A. Barg, "Explicit constructions of optimal-access MDS codes with nearly optimal sub-packetization," *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6307–6317, 2017.
- [10] M. Vajha, S. B. Balaji, and P. Vijay Kumar, "Small-d MSR codes with optimal access, optimal sub-packetization, and linear field size," *IEEE Transactions on Information Theory*, vol. 69, no. 7, pp. 4303–4332, 2023.
- [11] J. Li, X. Tang, and C. Tian, "A generic transformation to enable optimal repair in MDS codes for distributed storage systems," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6257–6267, 2018.
- [12] M. Vajha, V. Ramkumar, B. Puranik, G. Kini, E. Lobo, B. Sasidharan, P. V. Kumar, A. Barg, M. Ye, S. Narayanamurthy, S. Hussain, and S. Nandi, "Clay codes: Moulding MDS codes to yield an MSR code," in *16th USENIX Conference on File and Storage Technologies (FAST 18)*, (Oakland, CA), pp. 139–154, USENIX Association, Feb. 2018.
- [13] B. Sasidharan, G. K. Agarwal, and P. V. Kumar, "A high-rate MSR code with polynomial sub-packetization level," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 2051–2055, 2015.
- [14] K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Optimal exact-regenerating codes for distributed storage at the MSR and MBR points via a product-matrix construction," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5227–5239, 2011.
- [15] I. Tamo, M. Ye, and A. Barg, "Optimal repair of reed-solomon codes: Achieving the cut-set bound," in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 216–227, 2017.
- [16] I. Tamo, Z. Wang, and J. Bruck, "Zigzag codes: MDS array codes with optimal rebuilding," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1597–1616, 2013.
- [17] Z. Wang, I. Tamo, and J. Bruck, "Explicit minimum storage regenerating codes," *IEEE Transactions on Information Theory*, vol. 62, no. 8, pp. 4466–4480, 2016.
- [18] N. Raviv, N. Silberstein, and T. Etzion, "Constructions of high-rate minimum storage regenerating codes over small fields," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2015–2038, 2017.
- [19] I. Duursma and H.-P. Wang, "Multilinear algebra for minimum storage regenerating codes: a generalization of the product-matrix construction," *Appl. Algebra Eng., Commun. Comput.*, vol. 34, p. 717–743, Oct. 2021.
- [20] V. Ramkumar, S. B. Balaji, B. Sasidharan, M. Vajha, M. N. Krishnan, and P. V. Kumar, "Codes for distributed storage," *Foundations and Trends® in Communications and Information Theory*, vol. 19, no. 4, pp. 547–813, 2022.
- [21] V. R. Cadambe, S. A. Jafar, H. Maleki, K. Ramchandran, and C. Suh, "Asymptotic interference alignment for optimal repair of MDS codes in distributed storage," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2974–2987, 2013.
- [22] Z. Wang, I. Tamo, and J. Bruck, "Optimal rebuilding of multiple erasures in mds codes," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 1084–1101, 2017.
- [23] M. Zorgui and Z. Wang, "Centralized multi-node repair for minimum storage regenerating codes," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 2213–2217, 2017.
- [24] M. Ye and A. Barg, "Repairing Reed-Solomon codes: Universally achieving the cut-set bound for any number of erasures," *arXiv preprint arXiv:1710.07216*, 2017.
- [25] A. S. Rawat, O. O. Koiluoglu, and S. Vishwanath, "Centralized repair of multiple node failures with applications to communication efficient secret sharing," *IEEE Transactions on Information Theory*, vol. 64, no. 12, pp. 7529–7550, 2018.
- [26] M. Zorgui and Z. Wang, "On the achievability region of regenerating codes for multiple erasures," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 2067–2071, IEEE, 2018.
- [27] Y. Hu, Y. Xu, X. Wang, C. Zhan, and P. Li, "Cooperative recovery of distributed storage systems from multiple losses with network coding," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 2, pp. 268–276, 2010.
- [28] K. W. Shum and Y. Hu, "Cooperative regenerating codes," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7229–7258, 2013.
- [29] A.-M. Kermarrec, N. Le Scouarnec, and G. Straub, "Repairing multiple failures with coordinated and adaptive regenerating codes," in *2011 International Symposium on Networking Coding*, pp. 1–6, IEEE, 2011.
- [30] K. W. Shum, "Cooperative regenerating codes for distributed storage systems," in *2011 IEEE International Conference on Communications (ICC)*, pp. 1–5, IEEE, 2011.

- [31] J. Li and B. Li, "Cooperative repair with minimum-storage regenerating codes for distributed storage," *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pp. 316–324, 2014.
- [32] K. W. Shum and J. Chen, "Cooperative repair of multiple node failures in distributed storage systems," *International Journal of Information and Coding Theory*, vol. 3, no. 4, pp. 299–323, 2016.
- [33] M. Zorghi and Z. Wang, "Centralized multi-node repair regenerating codes," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4180–4206, 2019.
- [34] M. Ye and A. Barg, "Explicit constructions of high-rate MDS array codes with optimal repair bandwidth," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2001–2014, 2017.
- [35] S. Li, M. Gadouleau, J. Wang, and D. Zheng, "A new centralized multi-node repair scheme of MSR codes with error-correcting capability," *arXiv preprint arXiv:2309.15668*, 2023.
- [36] Y. Zhang, "New centralized MSR codes with small sub-packetization," *arXiv preprint arXiv:2401.16866*, 2024.
- [37] M. Ye and A. Barg, "Cooperative repair: Constructions of optimal mds codes for all admissible parameters," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1639–1656, 2019.
- [38] Y. Zhang, Z. Zhang, and L. Wang, "Explicit constructions of optimal-access MSCR codes for all parameters," *IEEE Communications Letters*, vol. 24, no. 5, pp. 941–945, 2020.
- [39] M. Ye, "New constructions of cooperative MSR codes: Reducing node size to $\exp(o(n))$," *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7457–7464, 2020.
- [40] Y. Liu, H. Cai, and X. Tang, "A new cooperative repair scheme with $k + 1$ helper nodes for (n, k) hadamard MSR codes with small sub-packetization," *IEEE Transactions on Information Theory*, vol. 69, no. 5, pp. 2820–2829, 2023.
- [41] Z. Zhang, G. Li, and S. Hu, "Constructing $(h, k + 1)$ cooperative MSR codes with sub-packetization $(h + 1)2^{\lceil n/2 \rceil}$," in *2024 IEEE International Symposium on Information Theory (ISIT)*, pp. 01–06, 2024.