Optimal Pliable Fractional Repetition Codes That Are Locally Recoverable: A Bipartite Graph Approach

Yi-Sheng Su

***Abstract*— The main purpose of this paper is to construct pliable fractional repetition (FR) codes that are locally recoverable for distributed storage systems (DSSs). FR codes are integral in constructing a class of distributed storage codes with exact repair by transfer. Pliable FR codes are a new type of FR codes in which both the per-node storage and repetition degree can easily be adjusted simultaneously; thus, pliable FR codes are vital for DSSs in which parameters can dynamically change over time. However, the constructions of pliable FR codes with repair locality remain unknown. In addition, the tradeoffs between the code minimum distance of an FR code and its repair locality are not fully understood. To address these problems, this paper first presents general results regarding FR codes. Subsequently, this paper presents an improved Singleton-like bound for locally recoverable FR codes under an additional requirement that each node must be part of a local structure that, upon failure, allows it to be exactly recovered by a simple download process. Moreover, this paper proposes a construction of locally recoverable FR codes that can achieve the proposed Singleton-like bound; this construction is based on bipartite graphs with a given girth. In particular, this paper also proposes a general bipartite-graph-based approach to constructing optimal pliable FR codes with and without repair localities; in this approach, a new family of bipartite graphs, called matching-feasible graphs, is introduced. Finally, this paper proposes the explicit constructions of optimal pliable FR codes by using a family of matching-feasible graphs with arbitrary large girth. Notably, in addition to attaining a Singleton-like bound for FR codes, the explicit pliable FR codes are optimal locally recoverable FR codes from two perspectives of repair locality. The explicit pliable FR codes can also be used as FR batch codes to provide load balancing in DSSs.**

***Index Terms*— Bipartite graphs, combinatorial batch codes, data reconstruction, distributed storage systems, girth, pliable fractional repetition codes, repair locality.**

I. Introduction

IN GENERAL, a distributed storage system (DSS) is formed by networking together numerous, say n, inexpensive and unreliable storage devices (hereafter referred to as nodes). To provide reliable access to data stored in a DSS, data redundancy based on coding techniques is introduced in DSSs. Traditional coding techniques for DSSs are to 1) treat the file to be stored in a DSS as a set of information symbols over a sufficiently large finite field, 2) use an (n, k) maximum-distance separable (MDS) code [e.g., a Reed-Solomon (RS) code] to generate coded symbols, and 3) store the coded symbols in the n nodes to allow any data collector to reconstruct the entire file by downloading the data stored in any $k < n$ nodes. One of the limitations related to such approaches is that, upon failure of a single node, an amount of data equivalent to reconstruct the file must be downloaded from the remaining nodes. Because single-node failures occur frequently in large-scale DSSs, a considerable volume of network traffic must be dedicated to the repair of failed nodes. Therefore, DSSs must be designed to repair single-node failures efficiently. The design of new erasure coding techniques for DSSs has attracted considerable attention in academia and in industry over the past decade [1]–[35].

A. Related Work

In [2] and [6], a new construction of erasure codes, called exact-repair minimum bandwidth regenerating (MBR) codes with repair by transfer (or called uncoded repair), has been proposed for DSSs. The codes in [2] and [6] enable uncoded exact node repair with the minimum repair-bandwidth and disk-I/O by requiring 1) that a failed node must be replaced by a replacement node that stores data identical to those in the failed node and 2) that every node participating in a node repair process must only pass, without any decoding, exactly one coded symbol that will be directly stored in the replacement node. The construction in [2] and [6] involves a concatenation of an outer MDS code with an inner repetition code based on a complete graph.

The codes in [2] and [6] were subsequently generalized and a new family of distributed storage codes with exact repair by transfer was proposed in [24]. The new family of distributed storage codes in [24] relaxes the requirement of an arbitrary fixed-size subset of nodes for repairing a failed node; under this requirement, a replacement node can connect to only certain fixed subsets of nodes for repair, and the repair thus becomes table-based. Similar to the codes in [2] and [6], the new family of distributed storage codes in [24] consists

Manuscript received July 23, 2017; revised March 20, 2018; accepted May 7, 2018. Date of publication October 16, 2018; date of current version January 18, 2019. This work was supported by the Ministry of Science and Technology, Taiwan, under Grants MOST-106-2221-E-309-001 and MOST-107-2221-E-309-001. This paper was presented in part at the 2018 IEEE International Symposium on Information Theory.

The author is with the Department of Computer Science and Information Engineering, Chang Jung Christian University, Tainan City 71101, Taiwan (e-mail: yssu@mail.cjcu.edu.tw).

Communicated by A. Dimakis, Associate Editor for Coding Techniques.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2018.2876284

of the concatenation of an outer MDS code and an inner repetition code. The inner repetition codes in [24], however, are different from those in [2] and [6] and are called fractional repetition (FR) codes. The new family of distributed storage codes in [24] can be briefly summarized as follows: Given a file to be stored in a DSS, the file is first encoded using an outer MDS code. The encoded symbols obtained from the outer MDS code are then placed in different nodes according to an inner FR code. The purpose of the inner FR code is to specify the placement of the encoded symbols in the nodes such that certain desirable properties are satisfied, which is explained in a later section.

FR codes constitute a notable erasure coding technique for DSSs. Constructions for FR codes have been considered in numerous papers starting from [24]. In [24], constructions of FR codes based on regular graphs and Steiner systems were presented, and bounds on the file size (i.e., the maximum amount of data that can be stored in a DSS using an FR code) were also discussed. In [25], optimal FR codes that attain the bounds on the file size were presented. In [26], DSSs of which the storage capacity per node is considerably larger than the repetition degree (i.e., the number of repetitions of the coded symbols) were investigated, for which explicit construction algorithms based on bipartite cage graphs and mutually orthogonal Latin squares were proposed. In [27] and [28], generalization of FR codes to general or irregular FR codes for heterogeneous DSSs has been considered, where each node can store a different amount of coded symbols or each coded symbol can have a different repetition degree. In [29], FR codes that have adaptable storage capacity per node, called adaptive FR codes, were considered, and constructions of adaptive FR codes based on symmetric designs were presented. In [30] and [31], a class of FR codes from resolvable designs was proposed, for which the repetition degree can be varied in a simple manner. Recently, in [32], a new type of FR codes, called pliable FR codes, was introduced, for which both the per-node storage and repetition degree can easily be adjusted simultaneously. Constructions of pliable FR codes were also proposed in [32].

In addition to repair-bandwidth and disk-I/O, repair locality is a crucial metric in large-scale DSSs [17]–[23]. Therefore, FR codes with locality, called locally recoverable FR codes and also known as locally repairable FR codes, were first introduced in a conference [33] and later appeared in a journal publication [31]; in these studies, the trade-offs between the code minimum distance of an FR code and its repair locality (i.e., upper bounds on the minimum distance of locally recoverable FR codes) were identified/discovered, and constructions of locally recoverable FR codes that can achieve the trade-offs were also proposed. In [34], locally recoverable FR codes with small repetition degrees (i.e., repetition degrees 2 and 3) were investigated. Moreover, in [35], a construction of locally recoverable FR codes based on symmetric designs was presented.

B. Contributions of This Paper

Pliable FR codes are vital for DSSs in which parameters can dynamically change over time. Although pliable FR codes were constructed in [32], optimal pliable FR codes that are locally recoverable remain unknown. Additionally, the trade-offs between the code minimum distance of an FR code and its repair locality were not completely resolved in [31] and [33]. Moreover, in [31] and [33], the proposed code construction that can achieve an upper bound on the minimum distance of locally recoverable FR codes imposes additional requirements on code parameters and is limited to very specific choices of file size. To address these limitations, this paper proposes a general bipartite-graph-based approach to constructing optimal pliable FR codes that are locally recoverable from the two perspectives of repair locality specified by the bounds in (4) and (6). This paper also presents an improved upper bound on the minimum distance of locally recoverable FR codes as well as a code construction that can achieve the presented upper bound. The contributions of this paper are summarized as follows.

- This paper presents general results regarding FR codes, including the exact file size of FR codes, a condition under which an FR code attains an upper bound on the file size, and a condition under which an FR code is optimal with respect to a Singleton-like bound on its minimum distance. These results are obtained by exploiting a bipartite graph representation of FR codes, along with a given girth.[1]
- This paper presents an improved Singleton-like bound for locally recoverable FR codes under an additional requirement that each node must be part of a local structure that, in case of failure, allows it to be exactly recovered just by a simple download process. This paper also proposes a construction of locally recoverable FR codes that can achieve the improved Singleton-like bound; this construction is based on bipartite graphs with a given girth.
- This paper proposes a general bipartite-graph-based approach to constructing optimal pliable FR codes with and without repair locality; in this approach, a new family of bipartite graphs, called *matching-feasible graphs*, is introduced. Matching-feasible graphs are used to construct optimal pliable FR codes that are locally recoverable from the perspectives of repair locality in (4) and (6). Additionally, FR batch (FRB) codes are constructed from matching-feasible graphs, for which the batch size (i.e., the number of symbols that can be read in parallel) is determined exactly [25].
- This paper proposes an explicit construction of optimal pliable FR codes that attain a Singleton-like bound for FR codes, by specializing to a family of matching-feasible graphs with arbitrary large girth. Notably, the explicit pliable FR codes and their transposed counterparts (i.e., the codes obtained by reversing the roles of storage nodes and the MDS coded symbols of the explicit pliable FR codes, which are also pliable) are also optimal locally recoverable FR codes and can be used as FRB codes.

C. Organization

The remainder of this paper is organized as follows. In Section II, the definitions of FR codes including pliable

[1]The girth of a bipartite graph is the length of its shortest cycles.

FR codes are reviewed, and a brief overview of their vital properties is also provided. In Section III, some general results on FR codes are presented, and in Section IV, an improved Singleton-like bound for locally recoverable FR codes and a code construction that can achieve the improved Singleton-like bound are presented. Furthermore, in Section V, a bipartite-graph-based approach to constructing optimal pliable FR codes with and without repair locality is presented. In Section VI, explicit constructions of optimal pliable FR codes that are locally recoverable are proposed. Section VII presents a comparison of the proposed FR codes with related results in the literature. Finally, in Section VIII, concluding remarks with a discussion of opportunities for future work are presented.

II. Preliminaries

An (n, k, d)-DSS is formed by networking together n nodes such that any data collector that can contact any k out of the n nodes can recover the file stored in the DSS. When a single node fails, an (n, k, d)-DSS allows a replacement node to reconstruct the data stored in the failed node by connecting to some d helper nodes out of the remaining $n-1$ nodes. Before proceeding, some notation is established as follows. Let $\mathbb{Z}_i$ denote $\{0, 1, \ldots, i-1\}$ for any positive integer i and let $\Omega \triangleq \mathbb{Z}_\theta$ denote the index set of θ (coded) symbols $\{c_0, c_1, \ldots, c_{\theta-1}\}$ that are obtained from an outer $(\theta, M(k))$ MDS code, where $M(k)$ denotes the size of the file stored in the DSS. Furthermore, let $\mathcal{V} \triangleq \{V_0, V_1, \ldots, V_{n-1}\}$ be a collection of n subsets of Ω. Then, FR codes can be defined as follows [30], [31].

Definition 1: An FR code $\mathcal{C}$ for an (n, k, d)-DSS with repetition degree ρ is a pair $(\Omega, \mathcal{V})$ such that the following properties are satisfied:

1) for each $i \in \mathbb{Z}_n$, $|V_i| = \alpha$, where $|\cdot|$ is the cardinality of a finite set,
2) each element of Ω belongs to exactly ρ sets in $\mathcal{V}$, and
3) for any $(\rho-1)$-subset $\widetilde{\mathcal{V}}$ of $\mathcal{V}$ and for each $V_i \in \widetilde{\mathcal{V}}$, there exists some d-subset $\{V_{r_0}, V_{r_1}, \ldots, V_{r_{d-1}}\}$ of $\mathcal{V} \setminus \widetilde{\mathcal{V}}$ such that for each $j \in \mathbb{Z}_d$, $|V_{r_j} \cap V_i| = 1$, and $\bigcup_{j=0}^{d-1} (V_{r_j} \cap V_i) = V_i$.

In Definition 1, every set $V_i \in \mathcal{V}$ denotes a set of indices of the symbols stored in node i. Therefore, Property 1 states that α is the per-node storage, and Property 2 indicates that ρ denotes the repetition degree. Moreover, Property 3 states that the data stored in node i can be recovered by downloading a $\beta = 1$ symbol from each of the d helper nodes $r_0, r_1, \ldots, r_{d-1}$. This indicates that a replacement node downloads $d\beta = d$ symbols in total for repairing a failed node. To support $\beta > 1$, trivial β-expansion (i.e., replicating the symbols in the storage system) may be employed. Suppose that the DSS operates at the MBR point, under which the condition $\alpha = d\beta = d$ is necessary. Therefore, α and d are used interchangeably in this paper. Additionally, for simplicity, the notion of incidence matrices is used in this paper to alternatively denote FR codes, as detailed subsequently.

Definition 2: An incidence matrix $\mathbf{N}$ of an FR code $\mathcal{C} = (\Omega, \mathcal{V})$ is a $\theta \times n$ binary matrix, with the (i, j)th entry equal to 1 if the index i of the coded symbol c_i is contained in V_j (i.e., $i \in V_j$) and equal to 0 otherwise, where $i \in \mathbb{Z}_\theta$ and $j \in \mathbb{Z}_n$.

A crucial property of FR codes is the size of the file stored in the DSS. In an (n, k, d)-DSS, the stored file should be reconstructed from any set of k nodes; therefore, the file size depends on the parameter k and is defined as follows.

Definition 3: The file size that is supported by an FR code $\mathcal{C} = (\Omega, \mathcal{V})$ for an (n, k, d)-DSS, denoted by $M(k)$, is equal to the minimum number of distinct symbols stored in any set of k nodes and is given by

$$M(k) = \min_{\{I \subseteq \mathbb{Z}_n \,|\, |I| = k\}} |\cup_{i \in I} V_i|. \tag{1}$$

The following lemma provides a tight upper bound on the file size that is supported by an FR code [24], [25].

Lemma 1: The file size that is supported by an FR code $\mathcal{C} = (\Omega, \mathcal{V})$ for an (n, k, d)-DSS is bounded from above by

$$M(k) \le \varphi(k) \tag{2}$$

where $\varphi(k)$ is defined recursively by

$$\varphi(1) = \alpha,$$
$$\varphi(k+1) = \varphi(k) + \alpha - \left\lceil \frac{\rho\varphi(k) - k\alpha}{n-k} \right\rceil.$$

Another crucial property of FR codes is their minimum distance. The proposed FR codes are evaluated in terms of minimum distance. The minimum distance of FR codes is defined as follows [30], [31].

Definition 4: The minimum distance of an FR code $\mathcal{C} = (\Omega, \mathcal{V})$ for an (n, k, d)-DSS, denoted by $d_{\min}(\mathcal{C})$, is equal to the smallest size of a subset $\widetilde{\mathcal{V}}$ of $\mathcal{V}$ such that the number of distinct symbols in $\mathcal{V} \setminus \widetilde{\mathcal{V}}$ is less than the file size $M(k)$. The minimum distance can be explicitly formulated as follows:

$$d_{\min}(\mathcal{C}) = \min_{\{J \subseteq \mathbb{Z}_n \,|\, |\cup_{i \in \mathbb{Z}_n \setminus J} V_i| < M(k)\}} |J|.$$

In other words, the minimum distance is equal to the size of a smallest subset of nodes whose failure guarantees that the file stored in the DSS cannot be reconstructed from the surviving nodes. The following lemma provides a Singleton-like bound on the minimum distance of an FR code [30], [31].

Lemma 2: The minimum distance of an FR code $\mathcal{C} = (\Omega, \mathcal{V})$ for an (n, k, d)-DSS is bounded from above by

$$d_{\min}(\mathcal{C}) \le n - \left\lceil \frac{M(k)}{\alpha} \right\rceil + 1. \tag{3}$$

FR codes with the local repair property (i.e., FR codes with $d < k$) are introduced as follows.

Definition 5: An FR code $\mathcal{C} = (\Omega, \mathcal{V})$ for an (n, k, d)-DSS is said to be *locally recoverable* if the condition $d < k$ is satisfied.

Codes with the local repair property receive a penalty on the maximum possible minimum distance [19], and FR codes are no exception. In fact, when $d < k$, the Singleton-like bound in (3) must be refined, as detailed in the following lemma [31], [33].

Lemma 3: Let $\mathcal{C} = (\Omega, \mathcal{V})$ be an FR code for an (n, k, d)-DSS. The minimum distance of $\mathcal{C}$ is bounded from above by

$$d_{\min}(\mathcal{C}) \leq n - \left\lceil \frac{M(k)}{\alpha} \right\rceil - \left\lceil \frac{M(k)}{d\alpha} \right\rceil + 2. \quad (4)$$

Remark 1: By comparing the Singleton-like bound in (3) with the Singleton bound (i.e., $d_{\min} \leq n - k + 1$), one can see that a given FR code for an (n, k, d)-DSS meets the Singleton-like bound in (3) if $k = \lceil \frac{M(k)}{\alpha} \rceil$. Similarly, a given locally recoverable FR code for an (n, k, d)-DSS meets the Singleton-like bound in (4) if $k = \lceil \frac{M(k)}{\alpha} \rceil + \lceil \frac{M(k)}{d\alpha} \rceil - 1$. These observations provide simple means of examining whether an FR code is optimal with respect to the Singleton-like bounds in (3) and (4).

Definition 6 describes a new type of FR codes, called pliable FR codes; this definition was introduced in [32], and these pliable FR codes are the primary focus of this paper.

Definition 6: Let $\mathcal{C} = (\Omega, \mathcal{V})$ be an FR code for an (n, k, d)-DSS. Accordingly, $\mathcal{C}$ is considered to be *pliable* if the following two properties are satisfied:

1) Ω can be partitioned into disjoint subsets (called index groups) such that each index group and each $V_i \in \mathcal{V}$ intersect in exactly one symbol index in Ω, and
2) $\mathcal{V}$ can be partitioned into disjoint subsets (called parallel classes) such that each parallel class contains all the symbol indices in Ω and any two different $V_i, V_j \in \mathcal{V}$ in a given parallel class have no symbol indices in common.

Remark 2: In Definition 6, each index group is a single subset of Ω, whereas each parallel class comprises several subsets of Ω. Moreover, according to Property 1 of Definition 6, the per-node storage can be varied by simply adjusting the number of index groups. Similarly, according to Property 2 of Definition 6, the repetition degree can be easily varied by changing the number of parallel classes. Some parameters of pliable FR codes can be determined from Definitions 1 and 6 as follows. First, according to Properties 1 and 2 of Definition 6, the total number of index groups in a pliable FR code must be equal to α and the total number of parallel classes must be equal to ρ. Next, with Property 1 of Definition 6 as well as Property 2 of Definition 1, n must be divided by ρ, and each index group thus contains $\frac{n}{\rho}$ symbol indices. Similarly, with Property 2 of Definition 6 as well as Property 1 of Definition 1, θ must be divided by α, and each parallel class thus comprises $\frac{\theta}{\alpha}$ subsets of Ω.

Through the establishment of a connection to combinatorial batch codes, FR codes can provide load balancing in DSSs, which are called FRB codes [25], as detailed in the following passages.

Definition 7: An FRB code with batch size t is an FR code $\mathcal{C} = (\Omega, \mathcal{V})$ with the additional property that any batch of t symbols from $\{c_0, c_1, \ldots, c_{\theta-1}\}$ can be retrieved by reading at most one symbol from each node.

In Definition 7, load balancing in DSSs is achieved by limiting the maximum number of symbols that can be downloaded from each node, which is set to 1. Notably, the retrieval of t symbols can be performed by t different users in parallel such that each user retrieves a different symbol.

A graph $\mathcal{G} = (Y \cup Z, E)$ is called bipartite if its vertex set can be partitioned into two parts Y and Z such that for every edge $\{y, z\} \in E$, either $y \in Y$ and $z \in Z$ or $z \in Y$ and $y \in Z$. As with FR codes, for simplicity, the notion of biadjacency matrices is used in this paper to alternatively denote bipartite graphs, as detailed subsequently.

Definition 8: Let $\mathcal{G} = (Y \cup Z, E)$ be a bipartite graph with the set of vertices consisting of two parts: $Y = \{y_0, y_1, \ldots, y_{|Y|-1}\}$ and $Z = \{z_0, z_1, \ldots, z_{|Z|-1}\}$. The biadjacency matrix of $\mathcal{G}$ is a $|Y| \times |Z|$ binary matrix $\mathbf{B}$ in which the (i, j)th entry is equal to 1 if $\{y_i, z_j\} \in E$ and equal to 0 otherwise.

Remark 3: If one views an incidence matrix as a biadjacency matrix, one can represent an FR code $\mathcal{C} = (\Omega, \mathcal{V})$ using a bipartite graph representation, where the vertex set is $\Omega \cup \mathcal{V}$ and the edge set comprises edges of the form $\{i, V_j\}$, where $i \in \mathbb{Z}_\theta$, $V_j \in \mathcal{V}$, and $i \in V_j$.

III. General Results on FR Codes

This section presents general results on FR codes. First, the following theorem provides a general result on the exact file size that is supported by FR codes and a condition under which an FR code attains the upper bound on the file size in (2).

Theorem 1: Let g be the girth of the bipartite graph representing an FR code $\mathcal{C} = (\Omega, \mathcal{V})$ for an $(n, k, d = \alpha)$-DSS with repetition degree ρ. Then, the file size that is supported by $\mathcal{C}$ can be determined as follows:

$$M(k) = \begin{cases} k\alpha - (k-1), & \text{if } 1 \leq k \leq \frac{g}{2} - 1 \\ \frac{g}{2}\alpha - \frac{g}{2}, & \text{if } k = \frac{g}{2}. \end{cases} \quad (5)$$

Moreover, $\mathcal{C}$ attains the upper bound on the file size in (2) for any $1 \leq k \leq \min(\frac{g}{2} - 1, \overline{k})$, where $\overline{k}$ is the lowest value of k such that $k\alpha - (k-1) > \frac{n-1}{\rho-1}$ holds; in other words, $\overline{k} = \lceil \frac{n-\rho}{(\alpha-1)(\rho-1)} \rceil$.

Proof: For the file size $M(k)$, consider first the case of $k = \frac{g}{2}$. Notably, in a bipartite graph with girth g, there exist no cycles of length $g-2$ or less. This, along with the structure of a cycle of length g, implies that in any $k = \frac{g}{2}$ columns of the incidence matrix of $\mathcal{C}$, there are exactly $\frac{g}{2}$ rows containing exactly two 1's, and all the other rows each contain at most one 1. This indicates that $M(\frac{g}{2}) = \frac{g}{2}\alpha - \frac{g}{2}$. Notably, for the case of $1 \leq k \leq \frac{g}{2} - 1$, in any $1 \leq k \leq \frac{g}{2} - 1$ columns of the incidence matrix of $\mathcal{C}$, there must be $k - 1$ rows containing exactly two 1's, and all the other rows each contain at most one 1. Accordingly, for any $1 \leq k \leq \frac{g}{2} - 1$, $M(k) = k\alpha - (k-1)$. For the achievability of the upper bound in (2), one can readily verify that, if $k\alpha - (k-1) \leq \frac{n-1}{\rho-1}$, the recursive definition of $\varphi(k)$ in Lemma 1 implies that $\varphi(k) = k\alpha - (k-1)$. Therefore, $\mathcal{C}$ attains the upper bound on the file size in (2) for $1 \leq k \leq \min(\frac{g}{2} - 1, \overline{k})$. This completes the proof. ■

Remark 4: According to Theorem 1, when $g \geq 2\alpha$, the exact file size of FR codes can be completely determined for any $1 \leq k \leq \alpha$. However, when $g < 2\alpha$, the file size for $\frac{g}{2} + 1 \leq k \leq \alpha$ remains unclear. Notably, characterizing the file size for $\frac{g}{2}+1 \le k \le \alpha$ is generally nontrivial, and the file size depends heavily on the cycles in the representative bipartite graphs when k exceeds $\frac{g}{2}$ [32].

The following theorem provides a general result on the achievability of the Singleton-like bound in (3) in terms of the girth of the representative bipartite graphs of FR codes.

Theorem 2: Let $\mathcal{C} = (\Omega, \mathcal{V})$ be an FR code for an $(n, k, d = \alpha)$-DSS. Furthermore, let g denote the girth of the bipartite graph representing $\mathcal{C}$. Accordingly, $\mathcal{C}$ is optimal with respect to the Singleton-like bound in (3) for $1 \le k \le \min(d, \frac{g}{2}) - \mathbb{I}_{\{d=\frac{g}{2}\}}$, where $\mathbb{I}_{\{\cdot\}}$ denotes the indicator of an event $\{\cdot\}$.

Proof: Suppose that $d \neq \frac{g}{2}$ and consider first the case of $\min(d, \frac{g}{2}) = d$, under which $g \ge 2(d+1)$ must hold. Therefore, according to Theorem 1, for any $1 \le k \le d$, $M(k) = k\alpha - (k-1)$. This indicates that $\lceil \frac{M(k)}{\alpha} \rceil = \lceil k - \frac{k-1}{\alpha} \rceil$. Because $1 \le k \le d = \alpha$, the condition $0 \le \frac{k-1}{\alpha} < 1$ must hold. Accordingly, $\lceil \frac{M(k)}{\alpha} \rceil = \lceil k - \frac{k-1}{\alpha} \rceil = k$. In other words, according to Remark 1, $\mathcal{C}$ meets the Singleton-like bound in (3) with equality for $1 \le k \le \min(d, \frac{g}{2}) = d$. Next, consider the case of $\min(d, \frac{g}{2}) = \frac{g}{2}$, under which $d = \alpha \ge \frac{g}{2} + 1$ must hold. According to Theorem 1, it is sufficient to consider only the case of $k = \frac{g}{2}$, under which $M(\frac{g}{2}) = \frac{g}{2}\alpha - \frac{g}{2}$ holds. Therefore, this indicates that $k = \frac{g}{2} = \lceil \frac{g}{2} - \frac{\frac{g}{2}}{\alpha} \rceil = \lceil \frac{M(\frac{g}{2})}{\alpha} \rceil = \lceil \frac{M(k)}{\alpha} \rceil$ must hold. According to Remark 1, $\mathcal{C}$ meets the Singleton-like bound in (3) with equality for $1 \le k \le \min(d, \frac{g}{2}) = \frac{g}{2}$. Suppose now that $d = \frac{g}{2}$. Then, one can readily verify from the previous discussion that $\mathcal{C}$ meets the Singleton-like bound in (3) with equality for $1 \le k \le \min(d, \frac{g}{2}) - 1$. ■

Remark 5: In addition to being beneficial for verifying whether an FR code is optimal with respect to the Singleton-like bound in (3), Theorem 2 provides insight into FR codes. First, the proof of Theorem 2 implies that an FR code $\mathcal{C}$ is optimal with respect to the Singleton-like bound in (3) for $1 \le k \le d = \alpha$ if the bipartite graph representing $\mathcal{C}$ has girth $g \ge 2(d+1)$. In other words, the larger g is, the more easily an FR code meets the Singleton-like bound in (3) with equality. Second, the proof of Theorem 2 also implies that an FR code is optimal with respect to the Singleton-like bound in (3) for $1 \le k \le \frac{g}{2}$ if $\alpha \ge \frac{g}{2} + 1$. This indicates that a high value of α is beneficial for achieving the Singleton-like bound in (3) for FR codes.

IV. Improved Singleton-Like Bound of Locally Recoverable FR Codes and Its Achievability

As noted in Section II, when $d < k$, the Singleton-like bound in (3) must be refined. A refinement is provided in Lemma 3, i.e., the Singleton-like bound in (4). According to the reports in [31] and [33], this section considers a refinement of the bound in (3) under an additional requirement that each node is part of a local structure (which actually forms an FR code and is thus hereafter called a local FR code) that, upon failure, allows it to be exactly recovered by a simple download process. Theorem 3 is one of the main results of this paper.

Theorem 3: Let $\mathcal{C} = (\Omega, \mathcal{V})$ be a locally recoverable FR code for an $(n, k, d = \alpha)$-DSS, where each node is part of a local FR code $\mathcal{C}' = (\Omega', \mathcal{V}')$, with $|\Omega'| = \theta'$, $|\mathcal{V}'| = n'$, per-node storage $\alpha' = \alpha$ and repetition degree ρ'. Accordingly, the minimum distance of $\mathcal{C}$ is bounded from above by

$$d_{\min}(\mathcal{C}) \le n - \left\lceil \frac{(\rho' - 1)\,\theta' \left\lfloor \frac{M(k)-1}{\theta'} \right\rfloor + M(k)}{\alpha} \right\rceil + 1. \quad (6)$$

Algorithm 1: Iteratively Construct a Sufficiently Large Set $\mathcal{S} \subset \mathcal{V}$ for Upper Bounding the Minimum Distance by $n - |\mathcal{S}|$; in This Algorithm, $H(\cdot)$ Is a Set Function Defined Over the Power Set of $\mathcal{V}$ and $H(\mathcal{X})$ Returns the Index Set of Distinct Symbols in Any Subset $\mathcal{X} \subseteq \mathcal{V}$

Input : $\mathcal{C} = (\Omega, \mathcal{V})$, $M(k)$
Output: $\mathcal{S}$
$\mathcal{S} = \emptyset$
while $|H(\mathcal{S})| < M(k)$ **do**
  Select a local FR code $\mathcal{C}' = (\Omega', \mathcal{V}')$ with $\mathcal{V}' \not\subset \mathcal{S}$ such that $|\Omega' \cap H(\mathcal{S})|$ is maximized.
  if $|H(\mathcal{S} \cup \mathcal{V}')| < M(k)$ **then**
    Set $\mathcal{S} = \mathcal{S} \cup \mathcal{V}'$.
  end
  else
    Identify the largest $\mathcal{A} \subset \mathcal{V}'$ such that $|H(\mathcal{S} \cup \mathcal{A})| < M(k)$.
    Set $\mathcal{S} = \mathcal{S} \cup \mathcal{A}$.
    Exit.
  end
end

Proof: The proof of this theorem follows that of [31, Lemma 16] (or [33, Lemma 4]), which involves the use of an algorithmic approach. The algorithm for determining the minimum distance bound is presented in Algorithm 1, which is a simplified version of the algorithms in [31] and [33]. In Algorithm 1, a sufficiently large set $\mathcal{S} \subset \mathcal{V}$ is iteratively constructed so that $|H(\mathcal{S})| < M(k)$, where $H(\cdot)$ is a set function defined over the power set of $\mathcal{V}$ and $H(\mathcal{X})$ returns the index set of distinct symbols in any subset $\mathcal{X} \subseteq \mathcal{V}$. Notably, for each local FR code $\mathcal{C}' = (\Omega', \mathcal{V}')$, the equation $\Omega' = H(\mathcal{V}')$ must hold. With the output $\mathcal{S}$ obtained from Algorithm 1, the minimum distance of $\mathcal{C}$ is bounded by $d_{\min}(\mathcal{C}) \le n - |\mathcal{S}|$. To simplify the following discussion, let $\mathcal{S}_i$ denote the set $\mathcal{S}$ at the ith iteration, where i is counted from 1. $\mathcal{S}_i$ represents the set of nodes (i.e., subsets of Ω in $\mathcal{V}$) totally included in $\mathcal{S}$ at the end of the ith iteration. Similarly, let $H(\mathcal{S}_i)$ represent the index set of distinct symbols included at the end of the ith iteration. Furthermore, let $s_i = |\mathcal{S}_i| - |\mathcal{S}_{i-1}|$ and $h_i = |H(\mathcal{S}_i)| - |H(\mathcal{S}_{i-1})|$ represent the increments of $\mathcal{S}$ and $H(\mathcal{S})$ between the $(i-1)$th and ith iterations, respectively.

First, $0 \le s_i \le n'$ must hold. Therefore, if the bipartite graph representation of a local FR code $\mathcal{C}'$ is considered, the minimum number of symbols covered by $n' - s_i$ nodes (i.e., $n' - s_i$ subsets of Ω' in $\mathcal{V}'$) is at least equal to $\frac{(n'-s_i)\alpha'}{\rho'}$.

Accordingly,

$$h_i \le \theta' - \frac{(n'-s_i)\alpha'}{\rho'} \overset{(*)}{=} \frac{s_i\alpha'}{\rho'}, \quad (7)$$

where step $(*)$ uses $n'\alpha' = \theta'\rho'$. Suppose that Algorithm 1 enters Line 5 u times. Then, according to (7), $\sum_{i=1}^{u} s_i \ge \frac{\rho'}{\alpha'}\sum_{i=1}^{u} h_i$. If Algorithm 1 exits after entering Line 8, more nodes must be added so that strictly less than $M(k) - \sum_{i=1}^{u} h_i$ symbols are covered. One can readily verify that at least $\lceil \frac{M(k)-\sum_{i=1}^{u} h_i}{\alpha'} \rceil$ more nodes can be included. Therefore, the total number of nodes accumulated in $\mathcal{S}$ can be bounded from below as follows:

$$|\mathcal{S}| \ge \frac{\rho'}{\alpha'}\sum_{i=1}^{u} h_i + \left\lceil \frac{M(k) - \sum_{i=1}^{u} h_i}{\alpha'} \right\rceil - 1 \quad (8)$$

$$\ge \frac{\rho'-1}{\alpha'}\sum_{i=1}^{u} h_i + \frac{M(k)}{\alpha'} - 1 \quad (9)$$

$$\overset{(\star)}{\ge} \frac{\rho'-1}{\alpha'}\theta'\left\lfloor \frac{M(k)-1}{\theta'} \right\rfloor + \frac{M(k)}{\alpha'} - 1, \quad (10)$$

where step $(\star)$ follows simply from long division that $\theta'\lfloor\frac{M(k)-1}{\theta'}\rfloor$ is the largest multiple of θ' that is less than $M(k)$. The conclusion now follows from $n - |\mathcal{S}|$. ■

Remark 6: One can readily verify that the upper bound in (6) reduces to that in (3) when $M(k) \le \theta'$. This is because $M(k) \le \theta'$ implies no repair locality (i.e., $k \le d$); under this condition, it is sufficient to consider only one local FR code and it becomes equivalent to exploring an FR code. Therefore, the upper bound in (6) is more general than that in [31] and [33], because the upper bound in [31] and [33] requires the file size to be greater than θ' (i.e., $M(k) > \theta'$) and cannot reduce to that in (3) when $M(k) \le \theta'$.[2] In addition, the upper bound in (6) is tighter than that in [31] and [33]. To demonstrate this, let $M(k) = u\theta' + \phi$, where u is a positive integer and $0 < \phi \le \theta'$; moreover, consider the following two cases of ϕ, because the upper bound in [31] and [33] consists of two parts:

1) $0 < \phi \le \alpha'$: In this case, the upper bound in [31] and [33] becomes

$$d_{\min}(\mathcal{C}) \le n - \left\lceil \frac{M(k)\rho'}{\alpha'} \right\rceil + \rho'. \quad (11)$$

Substitute $M(k) = u\theta' + \phi$ into (6) and (11), and consider the difference between the right-hand sides of (11) and (6):

$$\left\lceil \frac{(\rho'-1)\ u\theta' + u\theta' + \phi}{\alpha'} \right\rceil - 1 - \left(\left\lceil \frac{u\theta'\rho' + \phi\rho'}{\alpha'} \right\rceil - \rho' \right)$$
$$\overset{(\clubsuit)}{=} \left\lceil \frac{un'\alpha' + \phi}{\alpha'} \right\rceil - 1 - \left(\left\lceil \frac{un'\alpha' + \phi\rho'}{\alpha'} \right\rceil - \rho' \right)$$
$$\overset{(\diamondsuit)}{=} un' - \left(\left\lceil un' + \frac{\phi\rho'}{\alpha'} \right\rceil - \rho' \right)$$
$$\overset{(\spadesuit)}{\ge} un' - \left(\left\lceil un' + \frac{\alpha'\rho'}{\alpha'} \right\rceil - \rho' \right)$$
$$= 0,$$

where step $(\clubsuit)$ uses $n'\alpha' = \theta'\rho'$ and steps $(\diamondsuit)$ and $(\spadesuit)$ are due to the assumed condition $0 < \phi \le \alpha'$.

2) $\alpha' < \phi \le \theta'$: In this case, the upper bound in [31] and [33] becomes

$$d_{\min}(\mathcal{C}) \le n + n' + 1 - \left\lceil \frac{M(k)\rho' + \theta'}{\alpha'} \right\rceil. \quad (12)$$

Similarly, substitute $M(k) = u\theta' + \phi$ into (6) and (12), and consider the difference between the right-hand sides of (12) and (6):

$$n' + \left\lceil \frac{(\rho'-1)\ u\theta' + u\theta' + \phi}{\alpha'} \right\rceil - \left\lceil \frac{u\theta'\rho' + \phi\rho' + \theta'}{\alpha'} \right\rceil$$
$$\overset{(\clubsuit)}{=} n' + \left\lceil \frac{un'\alpha' + \phi}{\alpha'} \right\rceil - \left\lceil \frac{un'\alpha' + \phi\rho' + \theta'}{\alpha'} \right\rceil$$
$$= n' + \left\lceil \frac{\phi}{\alpha'} \right\rceil - \left\lceil \frac{\phi\rho' + \theta'}{\alpha'} \right\rceil$$
$$\overset{(\diamondsuit)}{=} \left\lceil \frac{\theta'\rho' + \phi}{\alpha'} \right\rceil - \left\lceil \frac{\phi\rho' + \theta'}{\alpha'} \right\rceil$$
$$= \left\lceil \frac{\theta'\rho' - \theta' + \theta' + \phi}{\alpha'} \right\rceil - \left\lceil \frac{\phi\rho' + \theta'}{\alpha'} \right\rceil$$
$$= \left\lceil \frac{\theta'(\rho'-1) + \theta' + \phi}{\alpha'} \right\rceil - \left\lceil \frac{\phi\rho' + \theta'}{\alpha'} \right\rceil$$
$$\overset{(\spadesuit)}{\ge} \left\lceil \frac{\phi(\rho'-1) + \theta' + \phi}{\alpha'} \right\rceil - \left\lceil \frac{\phi\rho' + \theta'}{\alpha'} \right\rceil$$
$$= \left\lceil \frac{\phi\rho' + \theta'}{\alpha'} \right\rceil - \left\lceil \frac{\phi\rho' + \theta'}{\alpha'} \right\rceil$$
$$= 0,$$

where steps $(\clubsuit)$ and $(\diamondsuit)$ use $n'\alpha' = \theta'\rho'$ and step $(\spadesuit)$ is due to the assumed condition $\alpha' < \phi \le \theta'$.

The following provides a simple construction of locally recoverable FR codes that can achieve the Singleton-like bound in (6) with equality on the basis of bipartite graphs with a given girth.

Theorem 4: Let $\mathcal{C}' = (\Omega', \mathcal{V}')$ be an FR code for an $(n', k', d' = \alpha')$-DSS with θ' symbols and repetition degree ρ' such that the bipartite graph representing $\mathcal{C}'$ has girth g. Suppose that a locally recoverable FR code $\mathcal{C} = (\Omega, \mathcal{V})$ is constructed from $\mathcal{C}'$ by considering the disjoint union of $w \ge 1$ copies of $\mathcal{C}'$ and let the file size of $\mathcal{C}$ be given by $M(k) = u\theta' + k'\alpha' - (k'-1)$ for some $0 \le u < w$ and $1 \le k' \le \min(\alpha', \frac{g}{2} - 1)$. Accordingly, $\mathcal{C}$ is optimal with respect to the Singleton-like bound in (6).

Proof: If $M(k) = u\theta' + k'\alpha' - (k'-1)$ is substituted into the Singleton-like bound in (6), the minimum distance bound

[2]For self-containedness, the upper bound in [31] and [33] is provided as follows:

$$d_{\min}(\mathcal{C}) \le \max\left(n - \left\lceil \frac{M(k)\rho'}{\alpha'} \right\rceil + \rho', n + n' + 1 - \left\lceil \frac{M(k)\rho' + \theta'}{\alpha'} \right\rceil \right).$$

in (6) becomes

$$\begin{aligned} d_{\min}(\mathcal{C}) &\leq wn' - \left\lceil \frac{u\left(\rho'-1\right)\theta'+u\theta'+k'\alpha' - (k'-1)}{\alpha'} \right\rceil + 1 \\ &\overset{(\clubsuit)}{=} wn' - \left\lceil \frac{un'\alpha' + k'\alpha' - (k'-1)}{\alpha'} \right\rceil + 1 \\ &\overset{(\diamond)}{=} (w-u)\,n' - k' + 1, \end{aligned}$$

where step ($\clubsuit$) uses $n'\alpha' = \theta'\rho'$ and step ($\diamond$) is due to the condition $1 \leq k' \leq \min\left(\alpha', \frac{g}{2} - 1\right)$. Therefore, $\mathcal{C}$ is optimal with respect to the Singleton-like bound in (6) when any $k = un' + k'$ nodes (i.e., $k = un' + k'$ subsets of Ω in $\mathcal{V}$) cover at least $M(k) = u\theta' + k'\alpha' - (k'-1)$ symbols. For the proof, a greedy selection of $un' + k'$ nodes from $\mathcal{V}$ that cover exact $M(k) = u\theta' + k'\alpha' - (k'-1)$ symbols is first presented herein; followed by a demonstration that any other selection of $un'+k'$ nodes cannot cover fewer symbols than the greedy selection. The greedy selection of $un' + k'$ nodes follows the greedy paradigm and works as follows. Initially, a node is randomly selected from $\mathcal{V}$. Subsequently, among the remaining nodes, the node that contributes the smallest increment of symbols[3] is selected; this selection process is repeated until a total of $un' + k'$ nodes are selected. Similar to Algorithm 1, one can see 1) that the greedy selection of the first un' nodes is equivalent to picking any u local FR codes and 2) that the last k' nodes are picked from another local FR code. One can readily verify that the previously explained greedy selection of $un'+k'$ nodes covers exact $M(k) = u\theta' + k'\alpha' - (k'-1)$ symbols. To show that the aforementioned greedy selection of $un' + k'$ nodes covers the least number of symbols, suppose that a selected node in the greedy selection is now replaced by another node, say V_i, that has yet to be selected. If V_i belongs to a "new" local FR code, the replacement must lead to an increase in $M(k)$. However, if V_i belongs to the same local FR code as the last k' selected nodes, the replacement cannot cause $M(k)$ to decrease either; otherwise, it leads to a contradiction of the assumption that each selected node contributes the smallest increment of symbols. According to the preceding discussion, the greedy selection of $un' + k'$ nodes must cover the least number of symbols. This completes the proof. ∎

Example 1: Let $\mathcal{C}' = \left(\Omega', \mathcal{V}'\right)$ be an FR code for an $(n' = 9, k', d' = \alpha' = 3)$-DSS with $\theta' = 9$ symbols and repetition degree $\rho' = 3$. Suppose that the incidence matrix $\mathbf{N}$ of $\mathcal{C}'$ is given as follows:

$$\mathbf{N} = \left[\begin{array}{ccc|ccc|ccc} \underline{1} & 0 & 0 & \underline{1} & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & \underline{1} & 0 & 0 & 0 & 0 & \underline{1} \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ \hline \underline{1} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \underline{1} \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}\right]. \tag{13}$$

[3]If there is a tie, randomly choose one.

As shown in the columns 0, 3, and 8, the girth of the bipartite graph representing $\mathcal{C}'$ is equal to $g = 6$. Next, suppose that a locally recoverable FR code $\mathcal{C} = (\Omega, \mathcal{V})$ is constructed from $\mathcal{C}'$ by considering the disjoint union of $w = 3$ copies of $\mathcal{C}'$, and let $u = 2$ and $k' = 2$ such that the file size of $\mathcal{C}$ is given by $M(k) = u\theta' + k'\alpha' - (k'-1) = 2 \times 9 + 2 \times 3 - (2-1) = 23$. One can readily verify that any set of 20 nodes covers at least 23 symbols; therefore, the minimum distance of $\mathcal{C}$ is equal to 8 when the file size is equal to 23. This coincides with (6). However, the upper bound in [31] and [33] becomes the upper bound in (12) and yields $d_{\min}(\mathcal{C}) \leq 11$, which is obviously not sufficiently tight. □

Remark 7: In Theorem 4, setting $w = 1$ reveals that $\mathcal{C}$ is optimal with respect to the Singleton-like bound in (3) when the file size is equal to $M(k) = k'\alpha' - (k'-1)$ for $1 \leq k' \leq \min\left(\alpha', \frac{g}{2} - 1\right)$. This is consistent with the result in Theorem 2.

V. Bipartite-Graph-Based Approach to Constructing Optimal Pliable FR Codes

This section presents a bipartite-graph-based approach to constructing optimal pliable FR codes and shows that the proposed approach can yield optimal pliable FR codes that are locally recoverable from the two perspectives of repair locality specified by the bounds in (4) and (6). First, a new family of bipartite graphs is introduced in the following subsection.

A. A New Family of Bipartite Graphs

Some background information on matchings in bipartite graphs are first provided as follows. A matching in a bipartite graph $\mathcal{G} = (Y \cup Z, E)$ is a subset of edges $H \subseteq E$, no two of which have a common vertex. A matching in a bipartite graph $\mathcal{G} = (Y \cup Z, E)$ is said to be *perfect* if every vertex in $Y \cup Z$ is incident to exactly one edge of the matching. A new family of bipartite graphs, called *matching-feasible graphs*, is then defined as follows.

Definition 9: A bipartite graph $\mathcal{G} = (Y \cup Z, E)$ is considered to be *matching-feasible* if each of the two parts Y and Z can be further grouped into *clusters* such that for every pair of clusters $\left\{Y_i, Z_j \mid Y_i \subset Y, Z_j \subset Z\right\}$ from different parts, the edges of the subgraph induced by $Y_i \cup Z_j$ form a perfect matching of this subgraph.

Some properties and parameters of a matching-feasible graph $\mathcal{G} = (Y \cup Z, E)$ can be determined from the definition of perfect matchings and Definition 9 as follows. First, according to the properties of perfect matchings, all the perfect matchings between clusters must be of equal size (i.e., the same number of edges must exist in the perfect matchings). Therefore, the size of the perfect matchings between clusters is hereafter denoted by h. Then, according to Definition 9, all the clusters in Y and Z must be the same size h. Second, the number of clusters in Y (resp. Z) must be equal to $\frac{|Y|}{h}$ (resp. $\frac{|Z|}{h}$). Third, according to Definition 9, all the nodes in Y (resp. Z) must have the same degree $\frac{|Z|}{h}$ (resp. $\frac{|Y|}{h}$). Fourth, if a cluster and a perfect matching can be regarded as a "supernode" and a "superedge," respectively, a matching-feasible graph becomes

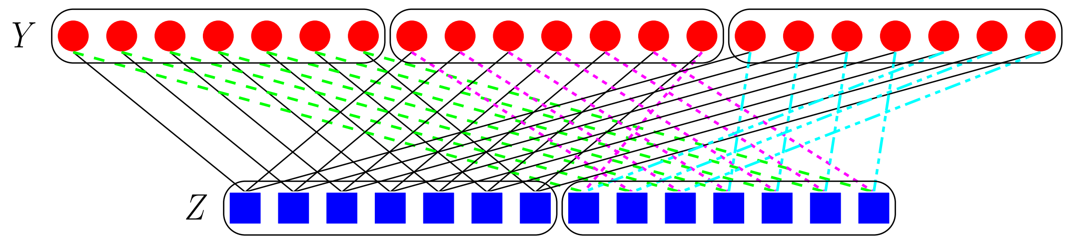

Fig. 1. Matching-feasible graph $\mathcal{G} = (Y \cup Z, E)$, where the two parts Y and Z can be divided into three and two clusters of size $h = 7$, respectively. The degrees of the nodes in Y and Z are equal to 2 and 3, respectively. The edge set E can be divided into six disjoint perfect matchings of size $h = 7$.

a complete bipartite "supergraph." Accordingly, the edge set E can be divided into $\frac{|Y|\cdot|Z|}{h^2}$ disjoint perfect matchings of size h. An example of matching-feasible graphs is presented in Fig. 1.

B. Optimal Pliable FR Codes Based on Matching-Feasible Graphs

The following theorems present a construction of optimal pliable FR codes by using matching-feasible graphs.

Theorem 5: Suppose that $\mathbf{G}$ is a biadjacency matrix of a matching-feasible graph $\mathcal{G} = (Y \cup Z, E)$ with girth $g > 4$ and matching size h, where the rows and columns are labeled with the nodes in Y and Z, respectively, each of which is ordered lexicographically under a fixed ordering. Then, the following hold:

1) $\mathbf{G}$ (resp. $\mathbf{G}^\top$) is an incidence matrix of pliable FR codes for an $(n = |Z|, k, d = \frac{|Y|}{h})$-DSS [resp. $(n = |Y|, k, d = \frac{|Z|}{h})$-DSS] with $\theta = |Y|$ (resp. $\theta = |Z|$) symbols and repetition degree $\rho = \frac{|Z|}{h}$ (resp. $\rho = \frac{|Y|}{h}$), where $^\top$ denotes the transpose operator.
2) If the condition $g \geq 2(\frac{|Y|}{h} + 1)$ [resp. $g \geq 2(\frac{|Z|}{h} + 1)$] holds, then the pliable FR code for which $\mathbf{G}$ (resp. $\mathbf{G}^\top$) is the incidence matrix is optimal with respect to the Singleton-like bound in (3) for $1 \leq k \leq d = \alpha = \frac{|Y|}{h}$ (resp. $1 \leq k \leq d = \alpha = \frac{|Z|}{h}$).

Proof: From the structure of $\mathcal{G}$, in $\mathbf{G}$, 1) there are $|Y|$ rows, 2) there are $|Z|$ columns, 3) each row has $\frac{|Z|}{h}$ 1's, and 4) each column has $\frac{|Y|}{h}$ 1's. Because $g > 4$, no cycles of length 4 exist in $\mathbf{G}$, and any two rows or columns of $\mathbf{G}$ have at most one 1-component in common. Therefore, a replacement node downloads only a $\beta = 1$ symbol from each helper node. This proves that $d = \alpha$ must hold. Hence, $\mathbf{G}$ is an incidence matrix of an FR code for an $(n = |Z|, k, d = \frac{|Y|}{h})$-DSS with $\theta = |Y|$ symbols and repetition degree $\rho = \frac{|Z|}{h}$. Additionally, according to the properties of matching-feasible graphs, $\mathbf{\Omega}$ can be divided into $\frac{|Y|}{h}$ index groups, each of which corresponds to a cluster in Y, and $\mathcal{V}$ can be partitioned into $\frac{|Z|}{h}$ parallel classes, each of which corresponds to a cluster in Z. Specifically, $\mathbf{G}$ is an array of permutation matrices (PMs), each of which corresponds to a perfect matching in $\mathcal{G}$. This indicates that $\mathbf{G}$ is an incidence matrix of a pliable FR code. Moreover, according to Theorem 2 or Remark 5, the pliable FR code for which $\mathbf{G}$ is the incidence matrix is optimal with respect to the Singleton-like bound in (3) for $1 \leq k \leq d = \alpha$, provided that $g \geq 2(d + 1)$. Therefore, if the condition $g \geq 2(\frac{|Y|}{h} + 1)$ holds, $\mathbf{G}$ yields an optimal pliable FR code that achieves the Singleton-like bound in (3) with equality for $1 \leq k \leq d = \alpha = \frac{|Y|}{h}$. For the case of $\mathbf{G}^\top$, the proof is the same as that for the case of $\mathbf{G}$ and is thus omitted. This completes the proof. ■

The following example illustrates Theorem 5.

Example 2: Consider the matching-feasible graph in Fig. 1; a biadjacency matrix of the graph is provided as follows:

$$\mathbf{G} = \left[\begin{array}{ccccccc|ccccccc}
\underline{1} & 0 & 0 & 0 & 0 & 0 & 0 & \underline{1} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \underline{1} & 0 & 0 & 0 & 0 & 0 & 0 & \underline{1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & \underline{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \underline{1} & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & \underline{1} & \underline{1} & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\underline{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \underline{1} & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \underline{1} & 0 & 0 & \underline{1} & 0 & 0 & 0 & 0
\end{array}\right]. \tag{14}$$

The code for which $\mathbf{G}$ in (14) is the incidence matrix is verified as having $M(3) = 7$ and is optimal with respect to the Singleton-like bound in (3) for $1 \leq k \leq 3 = \alpha$. In addition, the code for which the transpose of $\mathbf{G}$ in (14) is the incidence matrix is optimal with respect to the Singleton-like bound in (3) for $1 \leq k \leq 2 = \alpha$. This is because the girth of the matching-feasible graph in Fig. 1 is $g = 12$, which can be verified from the columns 0, 2, 6, 7, 9, and 10 of $\mathbf{G}$ in (14). □

Notably, the pliable FR codes constructed from matching-feasible graphs are locally recoverable and can be optimal with respect to the Singleton-like bound in (4), as detailed by the following theorem.

Theorem 6: Let $\mathbf{G}$ be a biadjacency matrix of a matching-feasible graph $\mathcal{G} = (Y \cup Z, E)$ with girth $g > 4$ and matching size h. Furthermore, let $\mathcal{C} = (\mathbf{\Omega}, \mathcal{V})$ be a pliable FR code for an $(n, k, d = \alpha)$-DSS, with $\mathbf{G}$ or $\mathbf{G}^\top$ serving as the incidence matrix. Suppose that a and b are positive integers such that $1 \le a \le b < \alpha$. If $g \ge 2(a\alpha + b + 2)$ holds, then $\mathcal{C}$ is an optimal locally recoverable code with respect to the Singleton-like bound in (4) when the file size is equal to $M(k) = k\alpha - (k-1)$, with $k = a\alpha + b + 1$.

Proof: If $g \ge 2(k+1) = 2(a\alpha + b + 2)$ holds, then according to Theorem 1, the file size is equal to $M(k) = k\alpha - (k-1)$, with $k = a\alpha + b + 1$. Therefore, when $k = a\alpha + b + 1$, the following holds:

$$\left\lceil \frac{M(k)}{\alpha} \right\rceil = \left\lceil \frac{k\alpha - (k-1)}{\alpha} \right\rceil = \left\lceil k - a - \frac{b}{\alpha} \right\rceil \overset{(*)}{=} k - a,$$

where step $(*)$ is due to the assumed condition $b < \alpha$. Moreover, when $k = a\alpha + b + 1$, the following holds:

$$\begin{aligned}\left\lceil \frac{M(k)}{d\alpha} \right\rceil &= \left\lceil \frac{k\alpha - (k-1)}{\alpha^2} \right\rceil \\ &= \left\lceil \frac{(k-1)\alpha + \alpha - (k-1)}{\alpha^2} \right\rceil \\ &= \left\lceil \frac{(a\alpha + b)\alpha + \alpha - (a\alpha + b)}{\alpha^2} \right\rceil \\ &= \left\lceil a + \frac{b - (a-1) - \frac{b}{\alpha}}{\alpha} \right\rceil \\ &\overset{(\star)}{=} a + 1,\end{aligned}$$

where step $(\star)$ is due to the assumed condition $1 \le a \le b < \alpha$. Therefore, $\lceil \frac{M(k)}{\alpha} \rceil + \lceil \frac{M(k)}{d\alpha} \rceil = k + 1$, and according to Remark 1, $\mathcal{C}$ is an optimal locally recoverable code with respect to the Singleton-like bound in (4). This completes the proof. ■

The following example illustrates Theorem 6.

Example 3: Consider again the matching-feasible graph in Fig. 1 and the biadjacency matrix $\mathbf{G}$ in (14). For demonstration, suppose that $a = b = 1$. The code for which $\mathbf{G}$ in (14) is the incidence matrix is verified as having $M(5) = 11$, yielding $\lceil \frac{M(k)}{\alpha} \rceil + \lceil \frac{M(k)}{d\alpha} \rceil = \lceil \frac{11}{3} \rceil + \lceil \frac{11}{3\times 3} \rceil = 5 + 1 = k + 1$. Hence, the pliable FR code for which $\mathbf{G}$ in (14) is the incidence matrix is an optimal locally recoverable code with respect to the Singleton-like bound in (4) when $M(5) = 11$. One can readily verify that the code for which the transpose of $\mathbf{G}$ in (14) is the incidence matrix is an optimal locally recoverable code with respect to the Singleton-like bound in (4) when $M(4) = 9$. □

Remark 8: In Theorem 6, the matching size h does not seem to play any role. However, according to Theorem 5, both d and α depend on h. Additionally, if, in Theorem 6, a is set to 0, Theorem 6 becomes equivalent to Theorem 5. Therefore, when $0 \le a \le b < \alpha$, Theorem 6 is a generalization of Theorem 5.

In addition to being optimal with respect to the Singleton-like bound in (4), the pliable FR codes constructed from matching-feasible graphs can be optimal with respect to the Singleton-like bound in (6), as detailed by the following theorem.

Theorem 7: Let $\mathbf{G}$ be a biadjacency matrix of a matching-feasible graph $\mathcal{G} = (Y \cup Z, E)$ with girth $g > 4$ and matching size h. Suppose that $\mathcal{G}$ is disconnected such that it is a union of $\upsilon > 1$ isomorphic connected subgraphs, and let $\mathcal{C} = (\Omega, \mathcal{V})$ be the pliable FR code for an $(n, k, d = \alpha)$-DSS, with $\mathbf{G}$ or $\mathbf{G}^\top$ serving as the code's incidence matrix. Accordingly, $\mathcal{C}$ is an optimal locally recoverable code with respect to the Singleton-like bound in (6) when the file size is equal to $M(k) = u\theta' + k'\alpha' - (k'-1)$ for some $0 \le u < \upsilon$ and $1 \le k' \le \min\left(d, \frac{g}{2} - 1\right)$, where $\theta' = \frac{|\Omega|}{\upsilon}$ and $\alpha' = \alpha$.

Proof: When $\mathcal{G}$ is disconnected and is a union of $\upsilon > 1$ isomorphic connected subgraphs, the pliable FR code for which either $\mathbf{G}$ or $\mathbf{G}^\top$ is the incidence matrix is a direct sum of $\upsilon > 1$ FR codes, each of which has $\theta' = \frac{|\Omega|}{\upsilon}$ symbols and per-node storage $\alpha' = \alpha$. Thus, the proof follows directly from Theorem 4 and Remark 5. ■

The following example illustrates Theorem 7.

Example 4: As a demonstration, consider the union $\mathcal{G} \cup \mathcal{G}$ of $w = 2$ copies of $\mathcal{G}$, where $\mathcal{G}$ is the matching-feasible graph in Fig. 1. Because $\mathcal{G} \cup \mathcal{G} = ((Y \cup Y) \cup (Z \cup Z), E \cup E)$, $\mathcal{G} \cup \mathcal{G}$ is a matching-feasible graph, and a biadjacency matrix of $\mathcal{G} \cup \mathcal{G}$ can be represented by

$$\overline{\mathbf{G}} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{bmatrix}, \tag{15}$$

where $\mathbf{G}$ is provided in (14). Let $\overline{\mathcal{C}}$ be the pliable FR code for which $\overline{\mathbf{G}}$ in (15) is the incidence matrix. Accordingly, $\overline{\mathcal{C}}$ is a pliable FR code for an $(n = 28, k, d = \alpha = 3)$-DSS with $\theta = 42$ symbols and repetition degree $\rho = 3$. Next, suppose that $u = 1$ and $k' = 2$. Accordingly, the file size is equal to $M(k) = 26$. One can readily verify that any set of 16 nodes covers at least 26 symbols; therefore, the minimum distance of $\overline{\mathcal{C}}$ is equal to 13 when $M(k) = 26$. This coincides with (6). The upper bound in [31] and [33], however, yields $d_{\min}(\overline{\mathcal{C}}) \le 18$, which is obviously not sufficiently tight. $\overline{\mathcal{C}}$ is optimal with respect to the Singleton-like bound in (6) when $M(k) = 21\ u + 3\ k' - (k'-1)$ for some $0 \le u < 2$ and $1 \le k' \le \min\left(3, \frac{12}{2} - 1\right) = 3$. □

Remark 9: Given that $\mathcal{G}$ is connected and comprises only $\upsilon = 1$ connected subgraph, Theorem 7 states that $\mathcal{C}$ is an optimal code with respect to the Singleton-like bound in (3) when the file size is equal to $M(k) = k'\alpha' - (k'-1)$ for $1 \le k' \le \min\left(d, \frac{g}{2} - 1\right)$. This is consistent with the result in Theorem 2. Therefore, Theorem 7 still holds when $\mathcal{G}$ is connected and this theorem can be regarded as a generalization of Theorem 2.

The pliable FR codes constructed from matching-feasible graphs can be used as FRB codes, as detailed by the following theorem.

Theorem 8: Let $\mathbf{G}$ be a biadjacency matrix of a matching-feasible graph $\mathcal{G} = (Y \cup Z, E)$ with girth $g > 4$ and matching size h. Accordingly, the code for which either $\mathbf{G}$ or $\mathbf{G}^\top$ is the incidence matrix is an FRB code with the batch size $t = \min(|Y|, |Z|)$.

Proof: According to the proof of Theorem 5, the batch size t remains to be determined. As noted in the proof of Theorem 5, $\mathbf{G}$ is an array of PMs. Therefore, for a pliable FR code for which either $\mathbf{G}$ or $\mathbf{G}^\top$ is the incidence matrix, t can be easily determined and is equal to $t = \min(|Y|, |Z|)$. ■

A limitation of the presented bipartite-graph-based approach to constructing optimal pliable FR codes may be the lack of an infinite family of matching-feasible graphs. For addressing this limitation, this paper presents an infinite family of matching-feasible graphs (as described in the following section), through which explicit constructions of optimal pliable FR codes are obtained.

VI. Explicit Constructions of Optimal Pliable FR Codes That Are Locally Recoverable

This section first introduces an infinite family of matching-feasible graphs and then presents explicit constructions of optimal pliable FR codes that are locally recoverable based on the family of matching-feasible graphs.

A. Infinite Family of Matching-Feasible Graphs

An infinite family of bipartite graphs defined by a system of equations over Galois fields [36] is presented as follows.

Definition 10: Let $m \geq 2$ be a positive integer and GF(q) denote the Galois field of size q, where q is a prime power. Accordingly, a family of bipartite graphs, denoted by $D(m, q)$, is defined as follows. The vertex set of $D(m, q)$ is the disjoint union of two copies of GF$(q)^m$, denoted by P_m and L_m. Elements of P_m are called points, each of which is represented by $(p) = (p_0, p_1, \ldots, p_{m-1})$, where $p_i \in$ GF(q). Similarly, elements of L_m are called lines, and each line is represented by $[l] = \left[l_0, l_1, \ldots, l_{m-1}\right]$, where $l_i \in$ GF(q). The edge set of $D(m, q)$ comprises edges, each of which is denoted by $\{(p), [l]\}$, in which the following $m - 1$ relations on the coordinates of (p) and $[l]$ hold:

$$\begin{aligned} p_1 + l_1 &= p_0 l_0, \\ p_2 + l_2 &= p_0 l_1, \end{aligned} \tag{16}$$

and, for $3 \leq i \leq m - 1$,

$$p_i + l_i = \begin{cases} -p_{i-2} l_0, & \text{if } (i+1) \equiv 0 \text{ or } 1 \pmod 4 \\ p_0 l_{i-2}, & \text{if } (i+1) \equiv 2 \text{ or } 3 \pmod 4. \end{cases} \tag{17}$$

Remark 10: Some crucial properties of $D(m, q)$ are summarized as follows [37]–[39].

1) $D(m, q)$ is a family of bipartite graphs on $2q^m$ vertices (with q^m vertices called points and q^m vertices called lines), each of which has q edges incident with itself.
2) For every point (p) (resp. line $[l]$) of $D(m, q)$ and every $e \in$ GF(q), there exists a unique neighbor of (p) (resp. $[l]$) whose first coordinate is e.
3) The girth of $D(m, q)$ is shown to be at least equal to $2\lceil \frac{m}{2} \rceil + 4$.
4) Interchanging p_i and l_i for each $i = 0, 1, \ldots, m - 1$ in (16) and (17) yields another family of bipartite graphs. This family of bipartite graphs is isomorphic to $D(m, q)$.
5) As shown previously, when $D(m, q)$ is disconnected, it is a union of isomorphic connected subgraphs. Additional results regarding the connectedness of $D(m, q)$ are provided as follows. $D(2, 2)$ is connected, whereas $D(m, 2)$ is disconnected for $m \geq 3$. In addition, for $m \geq 6$, all $D(m, q)$ are disconnected. For $q = 3$ and $q > 4$, $D(m, q)$ has $q^{\upsilon-1}$ isomorphic connected components, where $\upsilon = \lfloor \frac{m+2}{4} \rfloor$. Moreover, $D(2, 4)$ and $D(3, 4)$ have only one component, whereas $D(m, 4)$ has 4^υ isomorphic connected components for $m \geq 4$.

The following lemma shows that $D(m, q)$ is a family of matching-feasible graphs.

Lemma 4: $D(m, q)$ is a family of matching-feasible graphs with girth $g \geq 2\lceil \frac{m}{2} \rceil + 4$ and matching size $h = q^{m-1}$.

Proof: Letting either $Y = P_m$ and $Z = L_m$ or $Y = L_m$ and $Z = P_m$, and E comprise the edges of the form $\{(p), [l]\}$ satisfying the relations in (16) and (17) reveals that $\mathcal{G} = (Y \cup Z, E)$ is a matching-feasible graph. The remaining part of the proof follows directly from Definition 10 and Remark 10. ■

Remark 11: In addition to being a family of matching-feasible graphs, $D(m, q)$ has connections with combinatorial designs. A connection between $D(m, q)$ and a combinatorial design is provided as follows. Let ν and κ be positive integers. Then, a $(\nu, \kappa, 1)$-packing design is an ordered pair $(V, \mathcal{B})$ where V is a ν-set and $\mathcal{B}$ is a collection of κ-subsets of V (called blocks) such that every 2-subset of V occurs in *at most* one block of $\mathcal{B}$ [40]. As discussed previously, in $D(m, q)$, there are q^m points and q^m lines, each of which has q edges incident with itself. Additionally, in $D(m, q)$, any two points or lines have zero or one (i.e., at most one) neighbor in common. Accordingly, viewing each point in $D(m, q)$ as an element in V and each line in $D(m, q)$ as a block in B, or vice versa, yields a $(q^m, q, 1)$-packing design. However, $D(m, q)$ may not yield a balanced incomplete block design because, in a balanced incomplete block design, each pair of distinct elements must appear together in the *same* number of blocks.

B. Optimal Pliable FR Codes Derived From $D(m, q)$

This section first presents a class of optimal pliable FR codes based on $D(m, q)$ that attains the Singleton-like bound in (3).

Corollary 1: Suppose that $\mathbf{B}$ is a biadjacency matrix of $D(m, q)$ in which the rows and columns are labeled (p) and $[l]$, respectively, which are ordered lexicographically under a fixed ordering of GF(q). Accordingly, the following hold:

1) Both $\mathbf{B}$ and $\mathbf{B}^\top$ are incidence matrices of pliable FR codes for an $(n = q^m, k, d = q)$-DSS with $\theta = q^m$ symbols and repetition degree $\rho = q$.
2) When m is selected such that $\lceil \frac{m}{2} \rceil + 1 \geq q$, the pliable FR code for which either $\mathbf{B}$ or $\mathbf{B}^\top$ is the incidence matrix attains the Singleton-like bound in (3) for $1 \leq k \leq d = \alpha$.

Proof: The proof immediately follows from Theorem 5 and Lemma 4. ■

The following example illustrates Corollary 1.

Example 5: Let $m = 3$ and $q = 2$. In addition, suppose that the indices of the rows and columns run over the set $\{000, 001, 010, 011, 100, 101, 110, 111\}$. Accordingly, the two relations on the coordinates of $(p) = (p_0, p_1, p_2)$ and $[l] = [l_0, l_1, l_2]$ in (16) yield the following biadjacency matrix $\mathbf{B}$ of size 8×8:

$$\mathbf{B} = \left[\begin{array}{cccc|cccc} 1&0&0&0&1&0&0&0\\ 0&1&0&0&0&1&0&0\\ 0&0&1&0&0&0&1&0\\ 0&0&0&1&0&0&0&1\\ \hline 1&0&0&0&0&0&0&1\\ 0&1&0&0&0&0&1&0\\ 0&0&0&1&1&0&0&0\\ 0&0&1&0&0&1&0&0 \end{array}\right], \qquad (18)$$

If $\mathbf{B}$ in (18) or its transpose is regarded as an incidence matrix of FR codes, a pliable FR code for an $(n = 8, k, d = 2)$-DSS with $\theta = 8$ and $\rho = 2$ is derived. Additionally, the pliable FR code for which $\mathbf{B}$ in (18) or its transpose is the incidence matrix is optimal with respect to the Singleton-like bound in (3) for $1 \le k \le 2$.[4] □

The pliable FR codes constructed from $D(m, q)$ are locally recoverable and can be optimal with respect to the Singleton-like bound in (4).

Corollary 2: Let $\mathbf{B}$ be a biadjacency matrix of $D(m, q)$. Suppose that a and b are integers such that $1 \le a \le b < q$. If m is selected such that $\lceil \frac{m}{2} \rceil \ge aq + b$, the pliable FR code for which either $\mathbf{B}$ or $\mathbf{B}^\top$ is the incidence matrix is an optimal locally recoverable code with respect to the

[4]For simplicity and conciseness, only the scenario in which $q = 2$ is presented in this subsection, although the construction also works when q is replaced by other values. A scenario in which $q = 3$, however, is already available for demonstration in this paper. To demonstrate this, let $m = 2$ and $q = 3$, and suppose that the indices of the rows and columns run over the set $\{00, 01, 02, 10, 11, 12, 20, 21, 22\}$. Accordingly, the first relation on the coordinates of $(p) = (p_0, p_1)$ and $[l] = [l_0, l_1]$ in (16) yields a biadjacency matrix of size 9×9, which is just the incidence matrix $\mathbf{N}$ in (13). As noted in Example 1, when $\mathbf{N}$ in (13) or its transpose is regarded as an incidence matrix of FR codes, a pliable FR code for an $(n = 9, k, d = 3)$-DSS with $\theta = 9$ and $\rho = 3$ is derived. Additionally, the pliable FR code for which $\mathbf{N}$ in (13) or its transpose is the incidence matrix is optimal with respect to the Singleton-like bound in (3) for $1 \le k \le 2$, but it is suboptimal for $k = 3$. This is because the girth of $D(2, 3)$ is equal to $g = 6$.

$$\mathbf{B} = \left[\begin{array}{cccccccccccccccc|cccccccccccccccc}
\underline{1}&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&\underline{1}&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&\underline{1}&0&0&0&0&0&0&0&0&0&0&0&0&0&0&\underline{1}&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0\\
\hline
\underline{1}&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0\\
0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&\underline{1}&0&0\\
0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0\\
0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&\underline{1}&0&\underline{1}&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0
\end{array}\right] \qquad (19)$$

Singleton-like bound in (4) when the file size is equal to $M(k) = kq-(k-1) = (aq+b)(q-1)+q$, with $k = aq+b+1$.

Proof: The proof follows from Theorem 6 and Lemma 4, where m is selected such that $2\lceil \frac{m}{2} \rceil + 4 \geq 2(k+1) = 2(aq+b+2)$ holds. ■

The following example illustrates Corollary 2.

Example 6: Let $q = 2$, and suppose that $a = b = 1$. To ensure that $\lceil \frac{m}{2} \rceil \geq aq+b$, set $m = 5$. Accordingly, the code for which the biadjacency matrix of $D(5, 2)$ is the incidence matrix is a pliable FR code for an $(n = 32, k, d = 2)$-DSS with $\theta = 32$ symbols and repetition degree $\rho = 2$. Consider a scenario in which $k = aq + b + 1 = 1 \times 2 + 1 + 1 = 4 > 2 = d$. In this scenario, the file size is equal to $M(4) = 4 \times 2 - (4-1) = 5$ and it can readily be verified that $k = 4 = \lceil \frac{5}{2} \rceil + \left\lceil \frac{5}{2\times 2} \right\rceil - 1 = \lceil \frac{M(k)}{\alpha} \rceil + \lceil \frac{M(k)}{d\alpha} \rceil - 1$. Therefore, the code constructed from $D(5, 2)$ attains the Singleton-like bound in (4) when $M(4) = 5$. A biadjacency matrix of $D(5, 2)$ is provided as $\mathbf{B}$ in (19), as shown at the bottom of the previous page, where the indices of the rows and columns run over the set $\{00000, 00001, \ldots, 11111\}$. Columns 0, 14, 16, and 29 of $\mathbf{B}$ in (19) (i.e., those labeled with 00000, 01110, 10000, and 11101) can be used to verify that $M(4) = 5$. Moreover, the pliable FR code for which the transpose of $\mathbf{B}$ in (19) is the incidence matrix is verified to have $M(4) = 5$ and to attain the Singleton-like bound in (4) when $M(4) = 5$. □

Remark 12: When Corollary 2 and Corollary 1 are compared, for a given value of q, m must be set to a higher value in order to guarantee the local repair property.

The pliable FR codes constructed from $D(m, q)$ can be optimal with respect to the Singleton-like bound in (6), as detailed subsequently.

Corollary 3: Suppose that $D(m, q)$ is disconnected such that it is a union of $\upsilon > 1$ isomorphic connected subgraphs, and let $\mathbf{B}$ be its biadjacency matrix. The pliable FR code for which either $\mathbf{B}$ or $\mathbf{B}^\top$ is the incidence matrix is an optimal locally recoverable code with respect to the Singleton-like bound in (6) when the file size is equal to $M(k) = u\frac{q^m}{\upsilon}+k'q-(k'-1)$ for some $0 \leq u < \upsilon$ and $1 \leq k' \leq \min\left(q, \lceil \frac{m}{2} \rceil + 1\right)$.

Proof: The proof immediately follows from Theorem 7 and Lemma 4. ■

The following example illustrates Corollary 3.

Example 7: Consider Example 5 and suppose that the indices of the rows and columns run over the sets $\{100, 011, 110, 000, 101, 010, 111, 001\}$ and $\{000, 111, 011, 100, 001, 110, 010, 101\}$, respectively. Accordingly, $\mathbf{B}$ in (18) becomes

$$\mathbf{B} = \left[\begin{array}{cccc|cccc} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]. \tag{20}$$

According to (20), $D(3, 2)$ is a union of $\upsilon = 2$ FR codes. More precisely, $D(3, 2)$ is a union of $\upsilon = 2$ 8-cycles and, therefore, its girth is equal to $g = 8$. The FR code for which $\mathbf{B}$ in (20) or its transpose is the incidence matrix is optimal with respect to the Singleton-like bound in (6) when $M(k) = 4\ u + 2\ k' - (k' - 1)$ for some $0 \leq u < 2$ and $1 \leq k' \leq \min\left(2, \lceil \frac{3}{2} \rceil + 1\right) = 2$. Furthermore, consider Example 6 where $D(5, 2)$ is employed to construct an optimal locally recoverable code with respect to the Singleton-like bound in (4). As noted in [39], $D(5, 2)$ is a union of $\upsilon = 4$ 16-cycles. Then, $\mathbf{B}$ in (19) can be written in the following form:

$$\begin{bmatrix} \mathbf{C} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C} \end{bmatrix}$$

where

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\mathbf{0}$ denotes a zero matrix; the FR code constructed based on $D(5, 2)$ is optimal with respect to the Singleton-like bound in (6) when $M(k) = 8\ u + 2\ k' - (k' - 1)$ for some $0 \leq u < 4$ and $1 \leq k' \leq \min\left(2, \lceil \frac{5}{2} \rceil + 1\right) = 2$. □

Remark 13: Property 5 in Remark 10 demonstrates the scenario when $D(m, q)$ is disconnected such that it is a union of $\upsilon > 1$ isomorphic connected subgraphs.

The pliable FR codes constructed from $D(m, q)$ can be used as FRB codes, as detailed subsequently.

Corollary 4: Let $\mathbf{B}$ be a biadjacency matrix of $D(m, q)$. Then, the code for which either $\mathbf{B}$ or $\mathbf{B}^\top$ is the incidence matrix is an FRB code with the batch size $t = q^m$.

Proof: This follows immediately from Theorem 8 and Lemma 4. ■

VII. Comparison of the Proposed FR Codes With Related Results in the Literature

This section presents a comparison of the FR codes proposed in Sections III–VI with related results in the literature. The most related results in the literature are those in [25] and [31]–[33]. Table I summarizes the FR codes obtained using the methods presented in this paper, including their parameters.

A. Comparison With Pliable FR Codes in [32]

Pliable FR codes were first introduced in [32], wherein various constructions of pliable FR codes based on Euclidean geometry, circulant PMs, affine PMs, extended RS codes, Euler squares, geometry decomposition, and zigzag codes were proposed. One of the common characteristics of the constructions in [32] is that the girth of the representative bipartite graphs is $g \leq 8$ when no index groups or parallel classes are removed. In this case, the file size supported by the

TABLE I

SUMMARY OF THE FR CODES CONSTRUCTED IN THIS PAPER, WHERE, UNLESS SPECIFIED OTHERWISE, g DENOTES THE GIRTH OF A BIPARTITE GRAPH, $\mathcal{G} = (Y \cup Z, E)$ IS A MATCHING-FEASIBLE GRAPH WITH GIRTH $g > 4$ AND MATCHING SIZE h, AND $\mathcal{C} = (\Omega, \mathcal{V})$ IS AN FR CODE FOR AN $(n, k, d = \alpha)$-DSS WITH θ SYMBOLS AND REPETITION DEGREE ρ

Method	Code Type	Type of Graphs	Summary
Theorem 1	FR	Bipartite	The codes support the file size $M(k) = \begin{cases} k\alpha - (k-1), & \text{if } 1 \leq k \leq \frac{g}{2} - 1 \\ \frac{g}{2}\alpha - \frac{g}{2}, & \text{if } k = \frac{g}{2} \end{cases}$ and attain the upper bound on the file size in (2) for any $1 \leq k \leq \min(\frac{g}{2} - 1, \overline{k})$, where $\overline{k}$ is the lowest value of k such that $k\alpha - (k-1) > \frac{n-1}{\rho-1}$ holds; in other words, $\overline{k} = \lceil \frac{n-\rho}{(\alpha-1)(\rho-1)} \rceil$.
Theorem 2	FR	Bipartite	The codes attain the Singleton-like bound in (3) for $1 \leq k \leq \min(d, \frac{g}{2}) - \mathbb{I}_{\{d=\frac{g}{2}\}}$, where $\mathbb{I}_{\{\cdot\}}$ denotes the indicator of an event $\{\cdot\}$.
Theorem 4	Locally Recoverable FR	Bipartite	The codes are constructed by considering the disjoint union of $w \geq 1$ copies of an FR code $\mathcal{C}'$ that is designed for an $(n', k', d' = \alpha')$-DSS with θ' symbols and repetition degree ρ'. Accordingly, the codes attain the Singleton-like bound in (6) when the file size is equal to $M(k) = u\theta' + k'\alpha' - (k'-1)$ for some $0 \leq u < w$ and $1 \leq k' \leq \min\left(\alpha', \frac{g}{2} - 1\right)$, where g is the girth of the bipartite graph representing $\mathcal{C}'$.
Theorem 5	Pliable FR	Matching-Feasible	The codes • are designed for an $(n = \lvert Z\rvert, k, d = \frac{\lvert Y\rvert}{h})$-DSS [resp. $(n = \lvert Y\rvert, k, d = \frac{\lvert Z\rvert}{h})$-DSS] with $\theta = \lvert Y\rvert$ (resp. $\theta = \lvert Z\rvert$) symbols and repetition degree $\rho = \frac{\lvert Z\rvert}{h}$ (resp. $\rho = \frac{\lvert Y\rvert}{h}$), and • attain the Singleton-like bound in (3) for $1 \leq k \leq d = \alpha$ if $g \geq 2(\frac{\lvert Y\rvert}{h} + 1)$ [resp. $g \geq 2(\frac{\lvert Z\rvert}{h} + 1)$] holds.
Theorem 6	Locally Recoverable Pliable FR	Matching-Feasible	If the matching-feasible graphs have girth $g \geq 2(a\alpha + b + 2)$, the codes attain the Singleton-like bound in (4) when the file size is equal to $M(k) = k\alpha - (k-1)$, with $k = a\alpha + b + 1$, where a and b are integers such that $1 \leq a \leq b < \alpha$.
Theorem 7	Locally Recoverable Pliable FR	Matching-Feasible	Suppose that the matching-feasible graphs are disconnected such that each of them is a union of $v > 1$ isomorphic connected subgraphs. Accordingly, the codes attain the Singleton-like bound in (6) when the file size is equal to $M(k) = u\theta' + k'\alpha' - (k'-1)$ for some $0 \leq u < v$ and $1 \leq k' \leq \min\left(d, \frac{g}{2} - 1\right)$, where $\theta' = \frac{\lvert\Omega\rvert}{v}$ and $\alpha' = \alpha$.
Theorem 8	FRB	Matching-Feasible	The codes provide load balancing in DSSs and have batch size of $t = \min(\lvert Y\rvert, \lvert Z\rvert)$.
Corollary 1	Pliable FR	$D(m, q)$	The codes • are designed for an $(n = q^m, k, d = q)$-DSS with $\theta = q^m$ symbols and repetition degree $\rho = q$, and • attain the Singleton-like bound in (3) for $1 \leq k \leq d = \alpha$ if $\lceil \frac{m}{2} \rceil + 1 \geq q$ holds.
Corollary 2	Locally Recoverable Pliable FR	$D(m, q)$	If $\lceil \frac{m}{2} \rceil \geq aq + b$ holds, the codes attain the Singleton-like bound in (4) when the file size is equal to $M(k) = kq - (k-1) = (aq + b)(q-1) + q$, with $k = aq + b + 1$, where q is a prime power and a and b are integers such that $1 \leq a \leq b < q$.
Corollary 3	Locally Recoverable Pliable FR	$D(m, q)$	Suppose that $D(m, q)$ is disconnected such that it is a union of $v > 1$ isomorphic connected subgraphs. Accordingly, the codes attain the Singleton-like bound in (6) when the file size is equal to $M(k) = u\theta' + k'\alpha' - (k'-1)$ for some $0 \leq u < v$ and $1 \leq k' \leq \min\left(q, \lceil \frac{m}{2} \rceil + 1\right)$, where $\theta' = \frac{q^m}{v}$, and $\alpha' = q$.
Corollary 4	FRB	$D(m, q)$	The codes provide load balancing in DSSs and have batch size of $t = q^m$.

constructed codes in [32] was exactly determined for values of k that are at most $1 \leq k \leq 4$, although some file size results obtained when $k = \frac{g}{2} + 1$, with $g = 6$ and $g = 8$, were identified in [32]. With these file sizes, the constructed codes in [32] attain the upper bound in (2) for values of k that are at most $1 \leq k \leq \min(3, \overline{k})$ and achieve the Singleton-like bound in (3) with equality for values of k that are at most $1 \leq k \leq 3$. By contrast, the proposed FR codes attain the upper bound in (2) for $1 \leq k \leq \min(\frac{g}{2} - 1, \overline{k})$ and achieve the Singleton-like bound in (3) for $1 \leq k \leq \frac{g}{2} - 1$. Additionally, the proposed FR codes can be optimal locally recoverable FR codes with respect to the Singleton-like bounds in (4) and (6).

B. Comparison With FR Codes in [25]

Similar to Theorem 1, it was shown in [25] that FR codes obtained from a graph with a sufficiently large girth are optimal in the sense that the file size is maximized. However, this result applies only to FR codes with a repetition degree of 2 (i.e., $\rho = 2$) in [25]. Moreover, the graph representation of FR codes used in [25] (i.e., each storage node corresponding to a vertex in the graph and each MDS coded symbol corresponding to an edge in the graph) is different from that used in this study.

FR codes with $\rho > 2$ were also considered in [25], in which transversal designs and generalized polygons were directly employed to construct FR codes, rather than using graph representation of FR codes. In [25], the codes constructed using transversal designs were shown to attain the upper bound on the file size in (2) for all $1 \leq k \leq \alpha$ when α was sufficiently large. Furthermore, in [25], the codes constructed using generalized polygons were shown to have $M(3) = 3\alpha - 2$ and $M(4) = 4\alpha - 4$ and to be able to attain the upper bound on the file size in (2) for $1 \leq k \leq 3$. The aforementioned results coincide with Theorem 1 when $1 \leq k \leq \frac{g}{2}$. This is because the girths of the bipartite graphs representing FR codes constructed using transversal designs and generalized polygons are equal to 6 and 8, respectively. In contrast to Theorems 2, 4–7 and Corollaries 1–3, Silberstein and Etzion [25] did not investigate the trade-offs between the code minimum distance of an FR code and its repair locality.

C. Comparison With Locally Recoverable FR Codes in [31] and [33]

Similar to [25], locally recoverable FR codes with $\rho = 2$ and $\rho > 2$ have been considered in [31] and [33]. First, for the case of $\rho = 2$, a construction of locally recoverable FR codes (i.e., Construction 1) has been proposed in [31] and [33] and these codes, analogous to those in Theorem 6 and Corollary 2, can achieve the Singleton-like bound in (4) with equality. Two relevant distinctions exist between the codes in [31] and [33] and those in Theorem 6 and Corollary 2. One distinction is that the codes in [31] and [33] apply to only DSSs with $\rho = 2$ (i.e., allowing local recovery only in the presence of a single failure), whereas those in Theorem 6 and Corollary 2 apply to DSSs with $\rho \geq 2$ (i.e., allowing local recovery in the presence of multiple failures). The other distinction lies in the types of graphs used. The codes in [31] and [33] are based on a family of graphs, namely (s, g)-graphs, where each vertex has degree s and the girth is equal to g; by contrast, those in Theorem 6 and Corollary 2 are based on the family of matching-feasible graphs. Notably, in Theorem 6 and Corollary 2, when the matching-feasible graphs are replaced by other families of bipartite graphs such as those in [41] and [42] characterized by a given bi-degree r, s and girth g, the resultant FR codes are still optimal locally recoverable codes with respect to the Singleton-like bound in (4), as long as the girth is chosen to satisfy $g \geq 2(k+1) = 2(aq+b+2)$. Nevertheless, the resultant FR codes are not necessarily pliable when matching-feasible graphs are replaced by other families of bipartite graphs.

Second, for the case of $\rho > 2$, [31, Lemma 16] or [33, Lemma 4] provided a trade-off between the code minimum distance of an FR code and its repair locality under an additional requirement that each node must be part of a local structure that, in case of failure, allows it to be exactly recovered just by a simple download process. Under the same requirement, this paper presents an improved Singleton-like bound in Theorem 3. The proof of Theorem 3 is similar to that of [31, Lemma 16] or that of [33, Lemma 4]. The difference between the proof of Theorem 3 and that of [31, Lemma 16] or that of [33, Lemma 4] lies in bounding the total number of nodes accumulated in $\mathcal{S}$ from below in (8)–(10). Notably, in [31] and [33], a construction of locally recoverable FR codes (i.e., Construction 2) similar to that in Theorem 4 has been proposed. To be optimal with respect to the Singleton-like bound in (6), Construction 2 in [31] and [33], however, imposes complicated requirements on the parameters of local FR codes. Moreover, Construction 2 in [31] and [33] is less general than the construction in Theorem 4 because Construction 2 is optimal only when the file size is equal to $M(k) = u\theta' + \alpha'$ for some $1 \leq u < w$, rather than $M(k) = u\theta' + k'\alpha' - (k' - 1)$ for some $0 \leq u < w$ and $1 \leq k' \leq \min\left(\alpha', \frac{g}{2} - 1\right)$ in Theorem 4.

VIII. Conclusions and Future Work

This paper presents general results regarding FR codes, including the exact file size of FR codes, a condition under which an FR code attains an upper bound on the file size, and a condition under which an FR code is optimal with respect to a Singleton-like bound on the minimum distance. These results are obtained by exploiting a bipartite graph representation of FR codes, along with an important parameter of graphs called the "girth." In addition, this paper presents an improved Singleton-like bound for locally recoverable FR codes under an additional requirement that each node must be part of a local structure that, upon failure, allows it to be exactly recovered by downloading symbols from the surviving nodes. This paper also proposes a construction of locally recoverable FR codes that can achieve the improved Singleton-like bound with equality based on bipartite graphs with a given girth. The improved upper bound is proven to be tighter than that in the literature, and the code construction is also shown to be more general than that in the literature as well. In particular, this paper proposes a bipartite-graph-based approach to constructing optimal pliable FR codes with and without repair locality; a new family of bipartite graphs, called matching-feasible graphs, is introduced. Finally, this paper proposes explicit constructions of optimal pliable FR codes, based on a family of matching-feasible graphs with arbitrary large girth, in which both the per-node storage and repetition degree can easily be adjusted simultaneously. Notably, the explicit pliable FR codes are also optimal locally recoverable codes from the two perspectives of repair locality specified by the bounds in (4) and (6); furthermore, their transposed counterparts are also pliable FR codes. The explicit pliable FR codes can also be used as FRB codes to provide load balancing in DSSs, for which the batch size (i.e., the number of symbols that can be read in parallel) is determined exactly.

This paper presents a construction of locally recoverable FR codes that can achieve the improved Singleton-like bound with equality; nevertheless, the construction is simple and can be regarded as a Kronecker product of an identity matrix and a biadjacency matrix. Therefore, alternate formalizations of combinations of biadjacency matrices would be worthwhile. Other possible future research directions may include determining the file size when k exceeds $\frac{g}{2}$.

Acknowledgment

The author would like to express his sincere thanks and gratitude to the Associate Editor and the three anonymous reviewers for their constructive comments and suggestions which have helped to improve the quality of the paper significantly.

References

[1] Y. Wu and A. G. Dimakis, "Reducing repair traffic for erasure coding-based storage via interference alignment," in *Proc. IEEE ISIT*, Jun./Jul. 2009, pp. 2276–2280.

[2] K. V. Rashmi, N. B. Shah, P. V. Kumar, and K. Ramchandran, "Explicit construction of optimal exact regenerating codes for distributed storage," in *Proc. 47th Annu. Allerton Conf.*, Sep. 2009, pp. 1243–1246.

[3] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.

[4] C. Suh and K. Ramchandran, "Exact-repair MDS code construction using interference alignment," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1425–1442, Mar. 2011.

[5] 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 Trans. Inf. Theory*, vol. 57, no. 8, pp. 5227–5239, Aug. 2011.

[6] N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Distributed storage codes with repair-by-transfer and nonachievability of interior points on the storage-bandwidth tradeoff," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1837–1852, Mar. 2012.

[7] N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Interference alignment in regenerating codes for distributed storage: Necessity and code constructions," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2134–2158, Apr. 2012.

[8] 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 Trans. Inf. Theory*, vol. 59, no. 5, pp. 2974–2987, May 2013.

[9] S.-J. Lin and W.-H. Chung, "Novel repair-by-transfer codes and systematic exact-MBR codes with lower complexities and smaller field sizes," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 12, pp. 3232–3241, Dec. 2014.

[10] S. J. Lin, W. H. Chung, Y. S. Han, and T. Y. Al-Naffouri, "A unified form of exact-MSR codes via product-matrix frameworks," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 873–886, Feb. 2015.

[11] J. Li, X. Tang, and U. Parampalli, "A Framework of constructions of minimal storage regenerating codes with the optimal access/update property," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1920–1932, Apr. 2015.

[12] Z. Wang, I. Tamo, and J. Bruck, "Explicit minimum storage regenerating codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4466–4480, Aug. 2016.

[13] M. Ye and A. Barg, "Explicit constructions of high-rate MDS array codes with optimal repair bandwidth," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2001–2014, Apr. 2017.

[14] J. Li, X. Tang, and C. Tian, "A generic transformation for optimal repair bandwidth and rebuilding access in MDS codes," in *Proc. IEEE ISIT*, Jun. 2017, pp. 1623–1627.

[15] 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 $d < (n - 1)$," in *Proc. IEEE ISIT*, Jun. 2017, pp. 2048–2052.

[16] M. Ye and A. Barg, "Explicit constructions of optimal-access MDS codes with nearly optimal sub-packetization," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6307–6317, Oct. 2017.

[17] 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. NCA*, Jul. 2007, pp. 79–86.

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

[19] 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, Aug. 2012.

[20] M. Sathiamoorthy *et al.*, "XORing elephants: Novel erasure codes for big data," *Very Large Data Based Endowment*, vol. 6, no. 5, pp. 325–336, Mar. 2013.

[21] 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.

[22] D. S. Papailiopoulos and A. G. Dimakis, "Locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5843–5855, Oct. 2014.

[23] 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.

[24] S. El Rouayheb and K. Ramchandran, "Fractional repetition codes for repair in distributed storage systems," in *Proc. 48th Annu. Allerton Conf.*, Sep./Oct. 2010, pp. 1510–1517.

[25] N. Silberstein and T. Etzion, "Optimal fractional repetition codes based on graphs and designs," *IEEE Trans. Inf. Theory*, vol. 61, no. 8, pp. 4164–4180, Aug. 2015.

[26] J. C. Koo and J. T. Gill, "Scalable constructions of fractional repetition codes in distributed storage systems," in *Proc. 49th Annu. Allerton Conf.*, Sep. 2011, pp. 1366–1373.

[27] B. Zhu, K. W. Shum, H. Li, and H. Hou, "General fractional repetition codes for distributed storage systems," *IEEE Commun. Lett.*, vol. 18, no. 4, pp. 660–663, Apr. 2014.

[28] Q. Yu, C. W. Sung, and T. H. Chan, "Irregular fractional repetition code optimization for heterogeneous cloud storage," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 1048–1060, May 2014.

[29] B. Zhu and H. Li, "Adaptive fractional repetition codes for dynamic storage systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2078–2081, Dec. 2015.

[30] O. Olmez and A. Ramamoorthy, "Repairable replication-based storage systems using resolvable designs," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2012, pp. 1174–1181.

[31] O. Olmez and A. Ramamoorthy, "Fractional repetition codes with flexible repair from combinatorial designs," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1565–1591, Apr. 2016.

[32] Y.-S. Su, "Pliable fractional repetition codes for distributed storage systems: Design and analysis," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2359–2375, Jun. 2018.

[33] O. Olmez and A. Ramamoorthy, "Replication based storage systems with local repair," in *Proc. NetCod*, Jun. 2013, pp. 1–6.

[34] M.-Y. Nam, J.-H. Kim, and H.-Y. Song, "Locally repairable fractional repetition codes," in *Proc. 7th Int. Workshop Signal Design Appl. Commun.*, Sep. 2015, pp. 128–132.

[35] B. Zhu and H. Li, "Exploring node repair locality in fractional repetition codes," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2350–2353, Dec. 2016.

[36] F. Lazebnik and V. A. Ustimenko, "Explicit construction of graphs with an arbitrary large girth and of large size," *Discrete Appl. Math.*, vol. 60, nos. 1–3, pp. 275–284, Jun. 1995.

[37] J.-L. Kim, U. N. Peled, I. Perepelitsa, V. Pless, and S. Friedland, "Explicit construction of families of LDPC codes with no 4-cycles," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2378–2388, Oct. 2004.

[38] F. Lazebnik and A. J. Woldar, "General properties of some families of graphs defined by systems of equations," *J. Graph Theory*, vol. 38, no. 2, pp. 65–86, Oct. 2001.

[39] F. Lazebnik, S. Sun, and Y. Wang, "Some families of graphs, hypergraphs and digraphs defined by systems of equations: A survey," *Lect. Notes Seminario Interdisciplinare di Matematica*, vol. 14, pp. 105–142, Jan. 2017.

[40] C. J. Colbourn and J. H. Dinitz, Eds., *Handbook of Combinatorial Designs*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2007.

[41] H. Sachs, "Regular graphs with given girth and restricted circuits," *J. London Math. Soc.*, vol. 1, no. 1, pp. 423–429, 1963.

[42] Z. Füredi, F. Lazebnik, Á. Seress, V. A. Ustimenko, and A. J. Woldar, "Graphs of prescribed girth and bi-degree," *J. Combinat. Theory B*, vol. 64, no. 2, pp. 228–239, 1995.

Yi-Sheng Su received the B.S. and M.S. degrees from the National Tsing Hua University, Hsinchu, Taiwan, ROC, and the Ph.D. degree from the National Cheng Kung University, Tainan, Taiwan, ROC, respectively, all in electrical engineering. From 2011 to 2012, he was a Visiting Scholar with Carnegie Mellon University, Pittsburgh, PA, USA. His research interests include communications, communication networks, coding theory, information theory, and machine learning.