Optimal Fractional Repetition Codes based on Graphs and Designs

Natalia Silberstein and Tuvi Etzion

Abstract

Fractional repetition (FR) codes is a family of codes for distributed storage systems that allow for uncoded exact repairs having the minimum repair bandwidth. However, in contrast to minimum bandwidth regenerating (MBR) codes, where a random set of a certain size of available nodes is used for a node repair, the repairs with FR codes are table based. This usually allows to store more data compared to MBR codes. In this work, we consider bounds on the fractional repetition capacity, which is the maximum amount of data that can be stored using an FR code. Optimal FR codes which attain these bounds are presented. The constructions of these FR codes are based on combinatorial designs and on families of regular and biregular graphs. These constructions of FR codes for given parameters raise some interesting questions in graph theory. These questions and some of their solutions are discussed in this paper. In addition, based on a connection between FR codes and batch codes, we propose a new family of codes for DSS, namely fractional repetition batch codes, which have the properties of batch codes and FR codes simultaneously. These are the first codes for DSS which allow for uncoded efficient exact repairs and load balancing which can be performed by several users in parallel. Other concepts related to FR codes are also discussed.

Index Terms

Coding for distributed storage systems, fractional repetition codes, combinatorial batch codes, Turán graphs, cages, transversal designs, generalized polygons.

I. Introduction

In a distributed storage system (DSS), data is stored across a network of nodes, which can unexpectedly fail. To provide reliability, data redundancy based on coding techniques is introduced in such systems. Moreover, existing erasure codes allow to minimize the storage overhead [43]. Dimakis et al. [13] introduced a new family of erasure codes, called *regenerating codes*, which allow for efficient single node repairs by minimizing repair bandwidth. In particular, they presented two families of regenerating codes, called *minimum storage regenerating* (MSR) and *minimum bandwidth regenerating* (MBR) codes, which correspond to the two extreme points on the storage-bandwidth trade-off [13]. Constructions for these two families of codes can be found in [13], [14], [32], [34], [37], [40], [41] and references therein.

The authors are with the Department of Computer Science, Technion – Israel Institute of Technology, Haifa 32000, Israel (email: natalys@cs.technion.ac.il, etzion@cs.technion.ac.il).

This research was supported in part by the Israeli Science Foundation (ISF), Jerusalem, Israel, under Grant 10/12.

The first author was supported in part at the Technion by a Fine Fellowship.

This work was presented in part at the Information Theory and Applications Workshop (ITA 2014), San-Diego, USA, February 2014.

An $(n,k,d,M,\alpha,\beta)_q$ regenerating code C, for $k \leq d \leq n-1$, $\beta \leq \alpha$, is used to store a file of size M across a network of n nodes, where each node stores α symbols from $\mathbb{F}_q$, a finite field with q elements, such that the stored file can be recovered by downloading the data from any set of k nodes, where k is called the *reconstruction degree*. Note, that this means that any $n-k$ node failures (i.e., erasures) can be corrected by this code. When a single node fails, a newcomer node which substitutes the failed node contacts any set of d nodes and downloads β symbols of each node in this set to reconstruct the failed data. This process is called a *node repair process*, and the amount of data downloaded to repair a failed node, βd, is called the *repair bandwidth*.

The family of MBR codes has the minimum possible repair bandwidth, namely $\beta d = \alpha$. In [32], [33] Rashmi et al. presented a construction for MBR codes which have the additional property of exact *repair by transfer*, or exact *uncoded repair*. In other words, the code proposed in [32], [33] allows for efficient node repairs where no decoding is needed. Every node participating in a node repair process just passes one symbol ($\beta = 1$) which will be directly stored in the newcomer node. This construction is based on a concatenation of an outer MDS code with an inner repetition code based on a complete graph as follows. Let $M = k\alpha - \binom{k}{2}$ be the size of a file, which corresponds to MBR capacity with $\beta = 1$ [13]. This file is first encoded by using an $\left(\binom{n}{2}, M\right)$ MDS code $\mathcal{C}$. The $\binom{n}{2}$ symbols of the corresponding codeword of $\mathcal{C}$ are placed on the n different nodes, where each node stores $\alpha = n-1$ symbols, as follows. Each node is associated with a vertex in K_n, the complete graph with n vertices. Every symbol of the codeword from $\mathcal{C}$ is associated with an edge of K_n. Each node i of the DSS stores the symbols of the codeword of $\mathcal{C}$ which are associated with the edges incident to vertex i of K_n. The uniqueness of this construction for the given parameters $\alpha = d = n-1$ was proved in [32].

El Rouayheb and Ramchandran [36] generalized the construction of [32] and defined a new family of codes for DSSs which also allow exact repairs by transfer for a wide range of parameters. These codes, called *DRESS* (Distributed Replication based Exact Simple Storage) codes [30], consist of the concatenation of an outer MDS code and an inner repetition code called *fractional repetition* (FR) code. These codes for DSSs relax the requirement of a random d-set for a repair of a failed node and instead the repair becomes table based (or by using an appropriate function). This modified model requires new bounds on the maximum amount of data that can be stored on a DSS based on an FR code.

An (n,α,ρ) FR code C^{FR} is a collection of n subsets $N_1,\dots,N_n$ of $[\theta] \overset{\text{def}}{=} \{1,2,\dots,\theta\}$, $n\alpha = \rho\theta$, such that

- $|N_i| = \alpha$ for each i, $1 \leq i \leq n$;
- each symbol of $[\theta]$ belongs to exactly ρ subsets in C^{FR}, where ρ is called the *repetition degree* of C^{FR}.

A $[(\theta,M),k,(n,\alpha,\rho)]$ DRESS code is a code obtained by the concatenation of an outer (θ,M) MDS code $\mathcal{C}$ and an inner (n,α,ρ) FR code C^{FR}. To store a file $\mathbf{f} \in \mathbb{F}_q^M$ on a DSS, $\mathbf{f}$ is first encoded by using $\mathcal{C}$; next, the θ symbols of the codeword $\mathbf{c_f} \in \mathcal{C}$, which encodes the file $\mathbf{f}$, are placed on the n nodes defined by C^{FR} as follows: node $i \in [n]$ of the DSS stores α symbols of $\mathbf{c_f}$, indexed by the elements of the subset N_i. Each symbol of $\mathbf{c_f}$ is stored in exactly ρ nodes and it is possible to reconstruct the stored file $\mathbf{f}$ from any set of k nodes. When some node j fails, it can be repaired by using a set of $d = \alpha$ other nodes $i_1, i_2 \dots, i_\alpha$, such that $N_j \cap N_{i_s} \neq \varnothing$, $s \in [\alpha]$, and $\cup_{s\in[\alpha]}(N_j \cap N_{i_s}) = N_j$. Each such node passes exactly one symbol ($\beta = 1$) to repair node j. Note that the repair bandwidth of a DRESS code is the same as the repair bandwidth of an MBR code. The encoding scheme based on an FR code is shown in Fig. 1.

Note that the stored file should be reconstructed from any set of k nodes, and since the outer code is an MDS code of

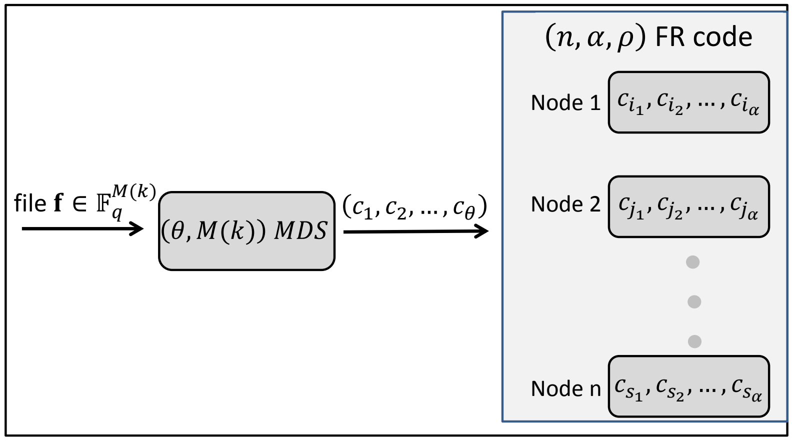

Fig. 1: The encoding scheme for a DRESS code

dimension M, it follows that

$$M \leq \min_{|I|=k} |\cup_{i \in I} N_i|, \tag{1}$$

where at least $\min_{|I|=k} |\cup_{i\in I} N_i|$ of distinct symbols of the MDS codeword are contained in any set of k nodes. Since we want to maximize the size of a file that can be stored by using a DRESS code, in the sequel we will always assume that $M = \min_{|I|=k} |\cup_{i\in I} N_i|$[1]. Note, that the same FR code can be used in different DRESS codes, with different k's as reconstruction degrees, and different MDS codes. The file size M, which is the dimension of the chosen MDS code, depends on the value of chosen k and hence in the sequel we will use $M(k)$ to denote the size of the file.

An (n, α, ρ) FR code is called *universally good* [36] if for any $k \leq \alpha$ the $[(\theta, M(k)), k, (n, \alpha, \rho)]$ DRESS code satisfies

$$M(k) \geq k\alpha - \binom{k}{2}, \tag{2}$$

where the righthand side of equation (2) is the maximum file size that can be stored using an MBR code, i.e., the MBR capacity [13]. In particular, it is of interest to consider codes which allow to store larger files when compared to MBR codes. Note that to satisfy (2) the inner FR code of a DRESS code should satisfy that

$$\min_{|I|=k} |\cup_{i\in I} N_i| \geq k\alpha - \binom{k}{2}. \tag{3}$$

Note also that if an FR code C^{FR} satisfies (3) then $|N_i \cap N_j| \leq 1$, for $N_i, N_j \in C^{\text{FR}}$, $i \neq j \in [n]$ [32].

Two upper bounds on the maximum file size $M(k)$ of a $[(\theta, M(k)), k, (n, \alpha, \rho)]$ DRESS code ($n\alpha = \rho\theta$), called the *FR capacity* and denoted in the sequel by $A(n, k, \alpha, \rho)$, were presented in [36]:

$$A(n, k, \alpha, \rho) \leq \left\lfloor \frac{n\alpha}{\rho} \left(1 - \frac{\binom{n-\rho}{k}}{\binom{n}{k}} \right) \right\rfloor; \tag{4}$$

[1]In some works this value is called the *rate* of a code. To avoid a confusion with a rate of a classical code we refer to this value as to the *maximum file size*.

$$A(n,k,\alpha,\rho) \leq \varphi(k), \text{ where } \varphi(1)=\alpha,\ \varphi(k+1)=\varphi(k)+\alpha-\left\lceil \frac{\rho\varphi(k)-k\alpha}{n-k} \right\rceil. \tag{5}$$

Note, that the bound in (5) is tighter than bound in (4). Note also that for any given k, the function $A(n,k,\alpha,\rho)$ is determined by the parameters of the inner FR code. We call an FR code *k-optimal* if it satisfies

$$\min_{|I|=k} |\cup_{i\in I} N_i| = A(n,k,\alpha,\rho),$$

in other words, the size of a file stored by using the FR code is the maximum possible for the given k. We call an FR code *optimal* if for any $k \leq \alpha$ it is k-optimal [2].

Constructions for FR codes are considered in many papers starting from [36], where FR codes are constructed from random regular graphs, Steiner systems, and their dual designs. Randomized FR codes based on the balls-and-bins model are presented in [30]. Constructions of FR codes with the fewest number of storage nodes given the other parameters, based on finite geometries and corresponding bipartite cage graphs are considered in [24]. FR codes based on affine resolvable designs and mutually orthogonal Latin squares are presented in [26], [28]. Enumeration of FR codes up to a given number of system nodes is presented in [4]. Algorithms for computing the reconstruction degree k and the repair degree d of FR codes are presented in [7]. Construction of FR codes based on regular graphs with a given girth, in particular cages, and analysis of their minimum distance is considered in [27]. Generalization of FR codes to *weak* or *general* FR codes, where each node stores a different amount of symbols, and constructions of codes based on graphs and group divisible designs are considered in [20], [45].

Note, that in all these papers the optimality of the constructed FR codes regarding the FR capacity, i.e. the maximality of the size of the stored file, was not considered. In this paper, we address the problem of constructing k-optimal FR codes and optimal FR codes (and hence optimal DRESS codes). In addition, we consider various problems from graph theory raised from the problem of constructing FR codes and present FR codes with additional desired properties.

The rest of the paper is organized as follows. In Section II we provide the main definitions of the structures which will be used in our constructions. In particular, in Subsection II-A we provide definitions for some families of regular graphs and graphs with a given girth. We present the Turán's theorem and the Moore bound which are essential for the results in this paper. In Subsection II-B we provide the definitions of transversal designs, projective planes, generalized polygons, and their incidence matrices. The definitions of FR codes based on graphs or on designs are given in Subsection II-C.

In Section III we consider FR codes with $\rho=2$ and propose constructions for FR codes which attain the bound in (5). Some of these codes are optimal and some are k-optimal for specific values of k. Note, that the case $\rho=2$ corresponds to the case of the highest data/storage ratio, since the repetition degree is the lowest one. All the constructions in this section are based on different families of regular graphs. First, we provide a useful lemma which shows the connection between the file size of a code and the structure of its underlying graph. In Subsection III-A optimal FR codes based on Turán graphs are considered. In Subsection III-B k-optimal FR codes based on different regular graphs with a given girth are presented. In Subsection III-C FR codes with a given file size are considered. The constructions raise many interesting questions in graph theory which are discussed in this section.

[2]Note that since $\alpha=d$ we always have $k \leq d = \alpha$ (very similarly to the parameters of regenerating codes).

In Section IV we consider FR codes with $\rho > 2$. In this case, a failed node can be repaired from several sets of other nodes, in contrast to the case with $\rho = 2$, in which a failed node can be repaired from a unique subset of α available nodes. One construction is based on a family of combinatorial designs, called transversal designs. This construction generalizes the construction based on Turán graphs for $\rho = 2$. Another construction is based on biregular bipartite graphs with a given girth. One important family of such graphs are the generalized polygons. We analyze the parameters of the constructed codes and find the conditions for which the bound in (5) is attained.

In Section V we establish a connection between the file size hierarchy and the generalized Hamming weight hierarchy. In fact, the sizes of the file related to the increasing values of k's form an integer sequence of nondecreasing values which can be viewed as a generalized definition for Hamming weights for constant weight codes. In Section VI we provide a lower bound on the reconstruction degree and present some FR codes which attain this bound.

In Section VII we analyze additional properties of FR codes by establishing a connection between FR codes and *combinatorial batch codes*. We propose a novel family of codes for DSS, called *fractional repetition batch codes* (FRB), which enable exact uncoded repairs and load balancing that can be performed by several users in parallel. We present examples of constructions of FRB codes based on bipartite complete graphs, graphs with large girth, transversal designs and affine planes. Conclusion are given in Section VIII.

II. Preliminaries

In this section we provide the definitions of all the combinatorial objects used for the constructions of FR codes presented in this paper.

A. Regular and Biregular Graphs

A graph $G = (V, E)$ consists of a vertex set V and an edge set E, where an edge is an unordered pair of vertices of V. For an edge $e = \{x, y\} \in E$ we say that x and y are *adjacent* and that x and e are *incident*. The *degree* of a vertex x is the number of edges incident with it. We say that a graph G is regular if all its vertices have the same degree and G is *d-regular* if each vertex has degree d. A graph is called *connected* if there is path between any pair of vertices. A graph is called *complete* if every pair of vertices are adjacent. A complete graph on n vertices is denoted by K_n. A *subgraph* $G_2 = (V_2, E_2)$ of a graph $G_1 = (V_1, E_1)$ is a graph such that $V_2 \subseteq V_1$ and $E_2 \subseteq E_1$. A subgraph G_2 of a graph G_1 is called *induced* if $E_2 = \{\{x, y\} : x, y \in V_2, \{x, y\} \in E_1\}$. A *$k$-clique* in a graph G is a complete subgraph of G with k vertices. The *complement* of a graph $G = (V, E)$, denoted by $\overline{G} = (V, \overline{E})$, is a graph with the same vertex set V but whose edge set $\overline{E}$ consists of all the edges not contained in G, i.e., $\overline{E} = \{\{x, y\} : x, y \in V, x \neq y, \{x, y\} \notin E\}$.

The *incidence matrix* $\mathbf{I}(G)$ of a graph $G = (V, E)$ is a binary $|V| \times |E|$ matrix with rows and columns indexed by the vertices and edges of G, respectively, such that $(\mathbf{I}(G))_{i,j} = 1$ if and only if vertex i and edge j are incident.

A graph G is called *bipartite* (*r-partite*, respectively) if its vertex set can be partitioned into two (r, respectively) parts such that every two adjacent vertices belong to two different parts. A bipartite graph is denoted by $G = (L \cup R, E)$, where L is the left part and R is the right part of G. A bipartite graph G is called *biregular* if the degree of the vertices in one part is d_1 and the degree of the vertices in the other part is d_2. An r-partite graph is called *complete* if every two vertices from two different parts are connected by an edge. The complete bipartite graph with left part of size n and right part of size m is denoted by $K_{n,m}$. Note that in $K_{n,m}$ the degree of a vertex in the left part is m and the degree of a vertex in the right part is n.

The following theorem, known as *Turán's theorem*, provides a necessary condition that a graph does not contain a clique of a given size [23, p. 58].

Theorem 1. *If a graph $G = (V, E)$ on n vertices has no $(r + 1)$-clique, $r \geq 2$, then*

$$|E| \leq (1 - \frac{1}{r})\frac{n^2}{2}. \tag{6}$$

Corollary 2. *If G is an α-regular graph which does not contain an $(r + 1)$-clique then $n \geq \frac{r}{r-1}\alpha$.*

We consider a family of regular graphs, called *Turán graphs*, which attain the bound of Corollary 2, in other words, have the smallest number of vertices. Let r, n be two integers such that r divides n. An (n, r)-*Turán graph* is defined as a regular complete r-partite graph, i.e., a graph formed by partitioning a set of n vertices into r parts of size $\frac{n}{r}$ and connecting each two vertices of different parts by an edge. Clearly, an (n, r)-Turán graph does not contain a clique of size $r + 1$ and it is an $(r - 1)\frac{n}{r}$-regular graph.

We now turn to another family of graphs, called *cages*. A *cycle* in a graph G is a connected subgraph of G in which each vertex has degree two. The *girth* of a graph is the length of its shortest cycle. A (d, g)-*cage* is a d-regular graph with girth g and minimum number of vertices. For example, a $(d, 4)$-cage is a complete bipartite graph $K_{d,d}$. Constructions for cages are known for $g \leq 12$ [15]. Let $\widehat{n}_0(d, g)$ be the minimum number of vertices in a (d, g)-cage. A lower bound on the number of vertices in a (d, g)-cage is given in the following theorem, known as *Moore bound* [10, p. 180].

Theorem 3. *The number of vertices in a (d, g)-cage is at least*

$$n_0(d, g) = \begin{cases} 1 + d\sum_{i=0}^{\frac{g-3}{2}}(d-1)^i & \text{if } g \text{ is odd} \\ 2\sum_{i=0}^{\frac{g-2}{2}}(d-1)^i & \text{if } g \text{ is even} \end{cases}. \tag{7}$$

Similar result to the Moore bound for biregular bipartite graphs can be found in [5], [17].

B. Combinatorial Designs

A *set system* is a pair $(\mathcal{P}, \mathcal{B})$, where $\mathcal{P} = \{p_i\}$ is a finite nonempty set of *points* and $\mathcal{B} = \{B_i\}$ is a finite nonempty set of subsets of $\mathcal{P}$ called *blocks*. A *design* $\mathcal{D}$ is set system with a constant number of points per block and no repeated blocks. A design $\mathcal{D}$ can be described by an *incidence matrix* $\mathbf{I}(\mathcal{D})$, which is a binary $|\mathcal{P}| \times |\mathcal{B}|$ matrix, with rows indexed by the points, columns indexed by the blocks, where

$$(\mathbf{I}(\mathcal{D}))_{i,j} = \begin{cases} 1 & \text{if } p_i \in B_j \\ 0 & \text{if } p_i \notin B_j \end{cases}.$$

The *incidence* graph $G_I(\mathcal{D}) = (V, E)$ of $\mathcal{D}$ is the bipartite graph with the vertex set $V = \mathcal{P} \cup \mathcal{B}$, where $\{p, B\} \in E$ if and only if $p \in B$, for $p \in \mathcal{P}$, $B \in \mathcal{B}$.

A *transversal design* of group size h and block size ℓ, denoted by TD(ℓ, h) is a triple $(\mathcal{P}, \mathcal{G}, \mathcal{B})$, where

1) $\mathcal{P}$ is a set of ℓh *points*;
2) $\mathcal{G}$ is a partition of $\mathcal{P}$ into ℓ sets (*groups*), each one of size h;

3) $\mathcal{B}$ is a collection of ℓ-subsets of $\mathcal{P}$ (*blocks*);

4) each block meets each group in exactly one point;

5) any pair of points from different groups is contained in exactly one block.

The properties of a transversal design TD(ℓ, h) which will be useful for our constructions are summarized in the following lemma [3].

Lemma 4. *Let $(\mathcal{P}, \mathcal{G}, \mathcal{B})$ be a transversal design TD(ℓ, h). Then*

- *The number of points is given by $|\mathcal{P}| = \ell h$;*
- *The number of groups is given by $|\mathcal{G}| = \ell$;*
- *The number of blocks is given by $|\mathcal{B}| = h^2$;*
- *The number of blocks that contain a given point is equal to h.*
- *The girth of the incidence graph of a transversal design is equal to 6.*

A TD(ℓ, h) is called *resolvable* if the set $\mathcal{B}$ can be partitioned into subsets $\mathcal{B}_1, ..., \mathcal{B}_h$, each one contains h blocks, such that each element of $\mathcal{P}$ is contained in exactly one block of each $\mathcal{B}_i$, i.e., the blocks of $\mathcal{B}_i$ partition the set $\mathcal{P}$. Resolvable transversal design TD(ℓ, q) is known to exist for any $\ell \leq q$ and prime power q [3].

Remark 1. *A TD$(2, h)$, for any integer $h \geq 2$ is equivalent to the complete bipartite graph $K_{h,h}$.*

Next, we consider two families of designs whose incidence graphs attain the Moore bound (7).

A *projective plane* of order n denoted by PG$(2, n)$, is a design $(\mathcal{P}, \mathcal{B})$, such that $|\mathcal{P}| = |\mathcal{B}| = n^2 + n + 1$, each block of $\mathcal{B}$ is of size $n + 1$, and any two points are contained in exactly one block. Note that any two blocks in $\mathcal{B}$ have exactly one common point. It is well known (see [19]) that the incidence graph of a projective plane has girth 6.

A *generalized quadrangle* of order (s, t), denoted by GQ(s, t) is a design $(\mathcal{P}, \mathcal{B})$, where

- Each point $p \in \mathcal{P}$ is incident with $t + 1$ blocks, and each block $B \in \mathcal{B}$ is incident with $s + 1$ points.
- Any two blocks have at most one common point.
- For any pair $(p, B) \in \mathcal{P} \times \mathcal{B}$, such that $p \notin B$, there is exactly one block B' incident with p, such that $|B' \cap B| = 1$.

In a generalized quadrangle GQ(s, t), the number of points $|\mathcal{P}| = (s+1)(st+1)$, the number of blocks $|\mathcal{B}| = (t+1)(st+1)$ and the girth of the incidence graph is 8 [19].

We note that transversal designs, projective planes, and generalized quadrangles belong to a class of designs called *partial geometries*. In addition, projective planes and generalized quadrangles are examples of designs called *generalized polygons* (or *n-gons*). Their incidence graphs have girth $2n$ and they attain the Moore bound. Such structures are known to exist only for $n \in \{3, 4, 6, 8\}$ [19].

C. FR Codes based on Graphs and Designs

Let C be an (n, α, ρ) FR code. C can be described by an *incidence matrix* $\mathbf{I}(C)$, which is an $n \times \theta$ binary matrix, $\theta = \frac{n\alpha}{\rho}$, with rows indexed by the nodes of the code and columns indexed by the symbols of the corresponding MDS codeword, such that $(\mathbf{I}(C))_{i,j} = 1$ if and only if node i contains symbol j.

Let G be an α-regular graph with n vertices. We say that an $(n, \alpha, \rho = 2)$ FR code C is based on G if $\mathbf{I}(C) = \mathbf{I}(G)$. Such a code will be denoted by C_G. It can be readily verified that any $(n, \alpha, 2)$ FR code can be represented by an α-regular graph with n vertices.

Let $\mathcal{D} = (\mathcal{P}, \mathcal{B})$ be a design with $|\mathcal{P}| = n$ points such that each block $B \in \mathcal{B}$ contains ρ points and each point $p \in \mathcal{P}$ is contained in α blocks. We say that an (n, α, ρ) FR code C is based on $\mathcal{D}$ if $\mathbf{I}(C) = \mathbf{I}(\mathcal{D})$. Such a code will be denoted by $C_{\mathcal{D}}$.

III. Fractional Repetition Codes with Repetition Degree 2

In this section we present constructions of optimal and k-optimal FR codes with repetition degree $\rho = 2$. These constructions are based on different types of regular graphs and are given in Subsections III-A and III-B. In Subsection III-C the properties of these graphs are investigated in order to present FR codes which allow to store a file of any given size. To avoid triviality we assume throughout the section that $\alpha > 2$.

First, we present the following useful lemma which shows a connection between the problem of finding the file size of an FR code based on a graph and the edge isoperimetric problem on graphs [8].

Lemma 5. *Let $G = (V, E)$ be an α-regular graph and let C_G be the FR code based on G. We denote by G_k the family of induced subgraphs of G with k vertices, i.e.,*

$$G_k = \{G' = (V', E') : |V'| = k, G' \text{ is an induced subgraph of } G\}.$$

Then the file size $M(k)$ of C_G is given by

$$M(k) = k\alpha - \max_{G' \in G_k} |E'|.$$

Proof: For each induced subgraph $G' = (V', E') \in G_k$ we define E'_{cut} to be the set of all the edges of E in the cut between V' and $V \setminus V'$, i.e.,

$$E'_{\text{cut}} = \{\{v, u\} \in E : v \in V', u \in V \setminus V'\}.$$

Clearly, $k\alpha = 2|E'| + |E'_{\text{cut}}|$ for every $G' \in G_k$. Note that $M(k) = \min_{G' \in G_k}\{|E'| + |E'_{\text{cut}}|\}$ and hence

$$M(k) = \min_{G' \in G_k} \{|E'| + \alpha k - 2|E'|\} = \alpha k - \max_{G' \in G_k} \{|E'|\}.$$

■

A. Optimal FR Codes Based on Turán Graphs

We begin our discussion with the following lemma which follows directly from Lemma 5.

Lemma 6. *Let G be an α-regular graph with n vertices, and let $M(k)$ be the file size of the corresponding FR code C_G. The graph G contains a k-clique if and only if $M(k) = k\alpha - \binom{k}{2}$.*

Corollary 7. *The file size $M(k)$ of an FR code C_G, where G is a graph which does not contain a k-clique, is strictly larger than the MBR capacity.*

One of the main advantages of an FR code is that its file size usually exceeds the MBR capacity. Hence, as a consequence of Corollary 7, we consider different families of regular graphs which do not contain a k-clique for a given k. Therefore, since Turán graphs have the minimum number of vertices among the graphs which do not contain a clique of a given size (see Corollary 2), we consider FR codes based on Turán graphs. The following theorem shows that FR codes obtained from Turán graphs attain the upper bound in (5) for all $k \leq \alpha$ and hence they are optimal FR codes.

Theorem 8. *Let $T = (V, E)$ be an (n, r)-Turán graph, $\alpha = (r-1)\frac{n}{r}$, and let k be an integer such that $1 \leq k \leq \alpha$. If $k = br + t$ for nonnegative integers b, t such that $t \leq r - 1$ then the $(n, \alpha, 2)$ FR code C_T based on T has file size*

$$M(k) = k\alpha - \binom{k}{2} + r\binom{b}{2} + bt, \tag{8}$$

which attains the upper bound in (5).

Proof: By Lemma 5, the value of $M(k)$ is determined by maximum cardinality of the edge set in an induced subgraph $T' = (V', E')$ of T, where $|V'| = k$. One can verify that since T is a complete regular r-partite graph, it follows that the induced subgraph T' with E' of the maximum cardinality is a complete r-partite graph with exactly t parts of size $b+1$ and $r - t$ parts of size b. Hence, the number of edges in E' is given by

$$\binom{t}{2}(b+1)^2 + \binom{r-t}{2}b^2 + t(r-t)(b+1)b.$$

Thus, by Lemma 5,

$$M(k) = k\alpha - \left[\binom{t}{2}(b+1)^2 + \binom{r-t}{2}b^2 + t(r-t)(b+1)b\right]. \tag{9}$$

It can be easily verified that (8) equals to (9). In addition, one can verify (by induction) that for the parameters of the constructed code C_T the bound in (5) equals to (8). ■

Remark 2. *Note, that for any $k > r$, the file size of the code C_T is strictly larger than the MBR capacity, i.e.,*

$$M(k) > k\alpha - \binom{k}{2}.$$

In the following theorem we provide an alternative, simpler representation of a file size for the FR code based on a Turán graph. Obviously, this expression for the file size is equivalent to the expression in (8). The proof of this theorem is also simpler than the one of Theorem 8. However, the proof of Theorem 8 could be used for the proof of Theorem 36 in Section IV and hence it was given for completeness.

Theorem 9. *Let $T = (V, E)$ be an (n, r)-Turán graph, $r < n$, $\alpha = (r-1)\frac{n}{r}$, and let k be an integer such that $1 \leq k \leq \alpha$. Then the $(n, \alpha, 2)$ FR code C_T based on T has the file size given by*

$$M(k) = k\alpha - \left\lfloor \frac{r-1}{r} \cdot \frac{k^2}{2} \right\rfloor.$$

Proof: We consider an induced subgraph $T' = (V', E')$ of the Turán graph T with $|V'| = k$ vertices which has the maximum number of edges. Since T' is a subgraph of T, in particular it does not contain K_{r+1}. Then, by Turán's theorem (see Theorem 1), $|E'| \leq \frac{r-1}{r}\frac{k^2}{2}$. Hence by Lemma 5, assuming that $M(k)$ is an integer, we have

$$M(k) = k\alpha - \left\lfloor \frac{r-1}{r} \cdot \frac{k^2}{2} \right\rfloor.$$

■

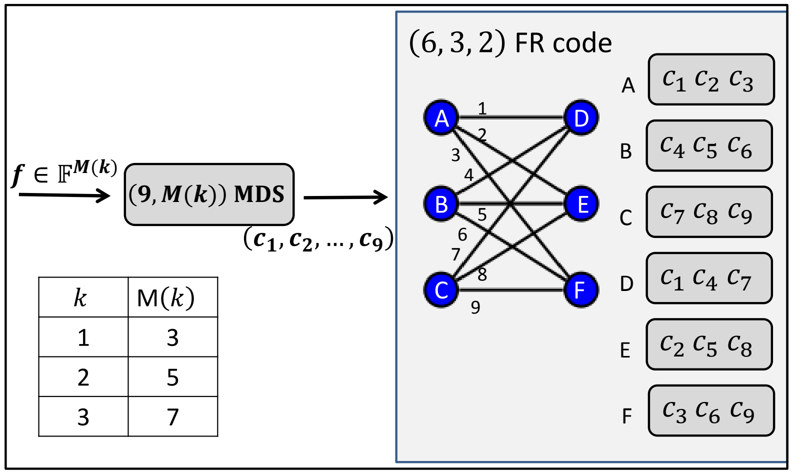

Fig. 2: The $((9, M(k)), k, (6, 3, 2))$ DRESS code with the inner FR code based on the complete bipartite graph $K_{3,3}$

The following result for FR codes based on complete bipartite graphs is a special case of Theorem 9 with $r = 2$.

Corollary 10. *The maximum size $M(k)$ of a file that can be stored using the $(2\alpha, \alpha, 2)$ FR code $C_{K_{\alpha,\alpha}}$ based on a regular complete bipartite graph $K_{\alpha,\alpha}$, for $\alpha \geq 2$, is given by*

$$M(k) = \begin{cases} k\alpha - \frac{k^2}{4} & \textit{if } k \textit{ is even} \\ k\alpha - \frac{k^2-1}{4} & \textit{if } k \textit{ is odd} \end{cases} \tag{10}$$

which attains the upper bound in (5) for all $1 \leq k \leq \alpha$.

Example 1. *The $(6, 3, 2)$ FR code based on $K_{3,3}$ and its file size for $1 \leq k \leq 3$ are shown in Fig. 2.*

B. k-Optimal FR Codes Based on Graphs with a Given Girth

First, we provide a simple upper bound on the file size of FR codes and show that this bound can be attained. The proof of the following Lemma can be easily verified from (5) or Lemma 5.

Lemma 11. *If C is an $(n, \alpha, 2)$ FR code then the file size $M(k)$ of C, for any $1 \leq k \leq \alpha$, satisfies*

$$M(k) \leq k\alpha - k + 1.$$

By Lemma 5, to obtain a large value for $M(k)$, every induced subgraph with k vertices should be as sparse as possible. Hence, for the rest of this subsection we consider graphs with a large girth (usually larger than k), in other words, the induced subgraphs with k vertices, $1 \leq k \leq \alpha$, will be trees. Next, we consider the girth of a graph and show that FR codes obtained from a graph with a large enough girth are optimal.

Lemma 12. *Let G be an α-regular graph with n vertices and let $M(k)$ be the file size of the corresponding FR code C_G. The girth of G is at least $k+1$ if and only if $M(k) = k\alpha - (k-1)$.*

Proof: Let G be a graph with girth g. Any induced subgraph G' of G with k vertices has at most $k-1$ edges if and only if $g \geq k+1$. Clearly, there exists at least one induced subgraph G' of G with k vertices and $k-1$ edges. Thus, by Lemma 5 we have $M(k) = k\alpha - (k-1)$. ∎

Corollary 13. *For each $k \leq g-1$, an FR code C_G based on an α-regular graph G with girth g attains the bound in (5), and hence it is k-optimal. C_G also attains the bound of Lemma 11.*

Corollary 14. *An FR code C_G based on an α-regular graph G with girth $g \geq \alpha + 1$ is optimal.*

Theorem 15. *Let G be a graph with girth g. Then the file size $M(k)$ of an FR code C_G based on G satisfies*

$$M(k) = \begin{cases} k\alpha - k + 1 & \text{if } k \leq g-1 \\ k\alpha - k & \text{if } g \leq k \leq g + \lceil \frac{g}{2} \rceil - 2. \end{cases}$$

Proof: For $k \leq g-1$ the result follows directly from Lemma 12. Since the graph G has a cycle of length g, it follows that $M(g) = g\alpha - g$. If any subgraph of G with k vertices contains at most one cycle then the file size satisfies $M(k) = k\alpha - k$. Note that for $g \leq k \leq g + \lceil \frac{g}{2} \rceil - 2$ there is no subgraph of G with k vertices that contains two cycles with no common vertices. Now we claim that the minimum number m of vertices in a connected subgraph of G with two cycles is $g + \lceil \frac{g}{2} \rceil - 1$. Assume for the contrary that $m = g + \lceil \frac{g}{2} \rceil - 1 - \epsilon$, $\epsilon \geq 1$. Hence there exists a subgraph of G depicted in Fig. 3 such that

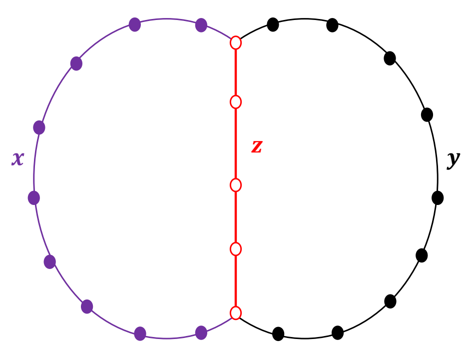

Fig. 3: Two intersecting cycles

$$x + z \geq g \tag{11}$$

$$y + z \geq g \tag{12}$$

$$x + y + 2 \geq g \tag{13}$$

$$x + y + z = g + \lceil \frac{g}{2} \rceil - 1 - \epsilon. \tag{14}$$

Hence from (11) and (14) we have that $y \leq \lceil \frac{g}{2} \rceil - 2$ and from (12) and (14) we have that $x \leq \lceil \frac{g}{2} \rceil - 2$ which contradicts to (13). Thus, the maximum number of vertices in a subgraph of G with at most one cycle is $g + \lceil \frac{g}{2} \rceil - 2$ which completes the proof of the theorem. ■

Next we consider examples of FR codes based on some interesting graphs for which the girth is known.

Example 2. *Let TD be a TD(q,q) transversal design. The $(2q^2, q, 2)$ FR code $C_{G_{TD}}$ based on the incidence graph G_I(TD) attains the bound in (5) for all $k \leq 5$.*

The following graphs attain the Moore bound (see Theorem 3), i.e., they have the minimum number of vertices given girth and degree. The parameters of the FR codes corresponding to these graphs can be found in the following table.

name of a graph	*degree*	*girth*	(n, α, ρ)
Complete graph K_n	$n-1$	*3*	$(n, n-1, 2)$
Complete bipartite graph $K_{r,r}$	r	*4*	$(2r, r, 2)$
Petersen graph	*3*	*5*	$(10, 3, 2)$
Hoffman-Singleton graph	*7*	*5*	$(50, 7, 2)$
Projective plane	$q+1$	*6*	$(2q^2+2q+2, q+1, 2)$
Generalized quadrangle	$q+1$	*8*	$(2q^3+2q^2+2q+2, q+1, 2)$
Generalized hexagon	$q+1$	*12*	$(2q^5+2q^4+2q^3+2q^2+2q+2, q+1, 2)$

Next, we use the Moore bound to show that the bound in (5) can be improved in some cases.

Lemma 16. *The bound in (5) is not tight for $\rho = 2$ if*

$$\alpha k - \alpha - k + 3 \leq n < \widehat{n}_0(\alpha, k+1),$$

where $\widehat{n}_0(d,g)$ is the minimum number of vertices in a (d,g)-cage (see Section II).

Proof: Let n, k be integers such that $\alpha k - \alpha - k + 3 \leq n < \widehat{n}_0(\alpha, k+1)$. Since $n < \widehat{n}_0(\alpha, k+1)$, the α-regular graph with n vertices corresponding to an $(n, \alpha, 2)$ FR code has girth at most k, and hence by Lemma 12 the file size satisfies $M(k) \leq k\alpha - k$. Therefore, $A(n,k,\alpha,2) \leq k\alpha - k$.

To complete the proof, we will show that if $\alpha k - \alpha - k + 3 \leq n$ then $\varphi(k) = k\alpha - k + 1$. To prove this statement we will prove by induction that if $\alpha k - \alpha - k + 3 \leq n$ then for all $1 \leq \ell \leq k$ it holds that $\varphi(\ell) = \ell\alpha - \ell + 1$. This trivially holds for $\ell = 1$. Assume that $\varphi(\ell-1) = (\ell-1)\alpha - (\ell-1) + 1$, $2 \leq \ell \leq k$. Then, by applying the recursion in (5) we have

$$\varphi(\ell) = (\ell-1)\alpha - (\ell-1) + 1 + \alpha - \left\lceil \frac{(\ell-1)\alpha - 2\ell + 4}{n - \ell + 1} \right\rceil = \ell\alpha - \ell + 1,$$

since $(\ell-1)\alpha - 2\ell + 4 \leq n - \ell + 1$ for $n \geq \alpha k - \alpha - k + 3$ and $\ell \leq k$.

■

As a consequence of Lemma 16 we have that the bound in (5) is not always tight and hence we have a similar better bound on $A(n,k,\alpha,\rho)$:

$$A(n,k,\alpha,\rho) \leq \varphi'(k), \text{ where } \varphi'(1) = \alpha,$$

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

C. FR Codes with a Given File Size

We observe from Lemma 6 and Lemma 11 that for any $1 \le k \le \alpha$, the file size $M(k)$ of an $(n, \alpha, 2)$ FR code C satisfies

$$k\alpha - \binom{k}{2} \le M(k) \le k\alpha - (k-1), \tag{15}$$

and the value of the file size depends on the structure of the underlying α-regular graph G. If the graph contains a clique K_k then the file size attains the lower bound in (15). If the graph does not contain a cycle of length k then the file size attains the upper bound in (15). The intermediate values for the file size can be obtained by excluding certain subgraphs of K_k from the graph G. For example, $M(k) \ge k\alpha - \binom{k}{2} + 1$ if and only if G does not contain K_k as a subgraph, and to have $M(k) \ge k\alpha - \binom{k}{2} + 2$, G should not contain $K_k - e$, i.e., a k-clique without an edge. Note that the problem of finding the minimum number of vertices in a graph which does not contain a specific subgraph is highly related to a Turán type problems (see e.g., [18] and the references therein).

For the rest of the section we assume that $n\alpha$ is an even integer.

There are only two possible values for $M(3)$, $3\alpha - 3$ and $3\alpha - 2$. To have a code C with file size $3\alpha - 2$, one should exclude a clique K_3, which is also a cycle of length 3. From (7) it follows that for a given α, if $n < 2\alpha$ then $M(3) = 3\alpha - 3$. Equivalently, the necessary condition for $M(3) = 3\alpha - 2$ is that $n \ge 2\alpha$.

Constructions for codes with file size $M(3) = 3\alpha - 2$ are provided in the previous subsection, based on optimal (Turán, Moore) graphs, for specific choices for the parameters α, n, where n is even. In addition, we provide in Appendix A another two constructions of FR codes with file size $3\alpha - 2$, the first one for even $n \ge 2\alpha$, and the second one for odd $n \ge \frac{5}{2}\alpha$.

The following lemma is proved in Appendix A.

Lemma 17. *Let n_3 is the minimum value of n such that $A(n, 3, \alpha, 2) = 3\alpha - 2$, for any $n \ge n_3$. Then*

$$2\alpha + 2 \le n_3 \le \frac{5\alpha}{2}.$$

We conjecture that $n_3 = \frac{5}{2}\alpha$. The following theorem is an immediate consequence from the discussion above (see also Appendix A).

Theorem 18. *The maximum file size for $k = 3$ satisfies*

- *For even n*
$$A(n,3,\alpha,2) = \begin{cases} 3\alpha - 3 & \textit{if } \; n < 2\alpha \\ 3\alpha - 2 & \textit{if } \; n \ge 2\alpha \end{cases} .$$
- *For odd n and even α, $A(n, 3, \alpha, 2) = 3\alpha - 2$.*
$$A(n,3,\alpha,2) = \begin{cases} 3\alpha - 3 & \textit{if } \; n < n_3 \\ 3\alpha - 2 & \textit{if } \; n \ge n_3 \end{cases} .$$

By (15) we have that $M(4) \in \{4\alpha - 3, 4\alpha - 4, 4\alpha - 5, 4\alpha - 6\}$.

1) If $M(4) = 4\alpha - 3$ then by Lemma 12 the corresponding graph G has girth at least 5. Codes with file size $4\alpha - 3$ can be derived from any graph G with girth ≥ 5, e.g., Hoffman-Singleton graph and its generalizations [1], [25].
2) If $M(4) = 4\alpha - 4$ then the corresponding graph G contains a subgraph of K_4 with 4 edges, but does not contain $K_4 - e$, a 4-clique without an edge. Codes with file size $4\alpha - 4$ and minimum number of nodes are constructed from $K_{\alpha,\alpha}$ by Theorem 8.
3) If $M(4) = 4\alpha - 5$ then the corresponding graph G contains $K_4 - e$, but does not contain K_4. Codes with file size $4\alpha - 5$ and minimum number of nodes are constructed from $(n, 3)$-Turán graphs if $n \equiv 0 \pmod 3$. If $n \not\equiv 0 \pmod 3$ then we use a modification of a Turán graph (see Example 7 in Appendix A).
4) If $M(4) = 4\alpha - 6$ then the corresponding graph G contains K_4. Codes with file size $4\alpha - 6$ and minimum number of nodes are constructed from the complete graph $K_{\alpha+1}$.

Let $G(k, \ell)$ denote any graph with k vertices and ℓ edges. By Lemma 5, to calculate the file size of an FR code based on a graph G, we need to find an induced subgraph $G(k, \ell)$ of G with the largest ℓ. The following lemma is an immediate consequence from Lemma 5.

Lemma 19. *Let G be an α-regular graph and let C_G be the FR code based on G.*

- *If G contains an induced subgraph $G(k, \ell)$ then $M(k) \leq k\alpha - \ell$.*
- *If G does not contain an induced subgraph $G(k, r)$ for all $r \geq \ell + 1$ then $M(k) \geq k\alpha - \ell$.*
- *If G contains an induced subgraph $G(k, \ell)$ and does not contain $G(k, r)$ for $r \geq \ell + 1$ then $M(k) = k\alpha - \ell$.*

Based on the previous discussion we have the following theorem.

Theorem 20. *The maximum file size for $k = 4$ satisfies*

$$A(n, 4, \alpha, 2) \leq \begin{cases} 4\alpha - 6 & \text{if } \alpha + 1 \leq n < \frac{3}{2}\alpha \\ 4\alpha - 5 & \text{if } \frac{3}{2}\alpha \leq n < \min\{n_3, 3\alpha - 3\} \\ 4\alpha - 4 & \text{if } \min\{n_3, 3\alpha - 3\} \leq n < 1 + \alpha^2 \\ 4\alpha - 3 & \text{if } n \geq 1 + \alpha^2 \end{cases},$$

Proof: The range of n for file size $4\alpha - 6$ follows from Corollary 2. For file size $4\alpha - 3$, the range of n follows from the Moore bound (see Theorem 3).

To distinguish between file sizes $4\alpha - 4$ and $4\alpha - 5$, we note that for $4\alpha - 4$ the corresponding graph G should not contain an induced subgraph $G(4, 5)$. If G does not contain K_3 then obviously it does not contain $G(4, 5)$. Hence, by Theorem 18, $n \geq n_3$ for file size $4\alpha - 4$. If G contains K_3 then let v, u_1 and u_2 be three vertices of K_3. Since the degree of v is α, it follows that there are $\alpha - 2$ other adjacent vertices $u_3, \ldots, u_\alpha$ of v. Since there is no $G(4, 5)$ in G it follows that u_1 and u_2 are not adjacent to any vertex in $\{u_3, \ldots, u_\alpha\}$. Moreover, except for v there is no other vertex for which both u_1 and u_2 are adjacent. Let A (B, respectively) be the set of additional $\alpha - 2$ vertices to which u_1 (u_2, respectively) is adjacent. The set $A \cup B \cup \{v\} \cup \{u_1, u_2, u_3, \ldots, u_\alpha\}$ contains $3\alpha - 3$ vertices, and hence $n \geq \min\{n_3, 3\alpha - 3\}$, which completes the proof of the theorem. ■

Remark 3. *Note, that there is an inequality in Theorem 20 since it is not always possible to find a regular graph which satisfies a given constraint on the number of edges in a subgraph with a given number of vertices. For example, there are no Moore graphs for some parameters.*

The existence problem of FR codes with $\rho = 2$, for any given file size in the interval between $k\alpha - \binom{k}{2}$ and $k\alpha - k + 1$, is getting more complicated as k increases. Some file sizes can be obtained by graphs with a given girth and by (n, r)-Turán graphs with different r's (see Theorem 8 and Theorem 15). However, not all the values between $k\alpha - \binom{k}{2}$ and $k\alpha - k + 1$ can be obtained by this way. Moreover, an additional problem is to find the minimum number of vertices in an α-regular graph G such that the corresponding FR code C_G has file size $k\alpha - x$, for some $k - 1 \leq x \leq \binom{k}{2}$. Note, that in some cases Theorem 8 and Theorem 15 together with Moore bound and Turán's theorem can provide the answer to this problem, as was demonstrated in Theorem 18 and Theorem 20.

The following lemma shows that given an information about the file size for a given reconstruction degree k, one can get some information about the file size for the reconstruction degree $k + 1$. In other words, by eliminating the existence of some induced subgraphs of size k, one can prove the nonexistence of induced subgraphs of size $k + 1$ which implies bounds on $M(k + 1)$ by Lemma 5.

Lemma 21. *If G is a regular graph that does not contain $G(k, \binom{k}{2} - \delta)$, for all $0 \leq \delta \leq x$, where x is a given integer, $0 \leq x \leq \binom{k-1}{2}$, then G also does not contain $G(k + 1, \binom{k+1}{2} - \gamma)$, for all $0 \leq \gamma \leq y$, where y is given by*

$$y = \begin{cases} \binom{k}{2} - 1 & \text{if } x = \binom{k-1}{2} \\ \left\lfloor \frac{2x}{k-1} \right\rfloor + x + 1 & \text{if } k \text{ is odd and } x \leq \binom{k-1}{2} - 1 \\ 2\left\lfloor \frac{x}{k-1} \right\rfloor + x + 1 & \text{if } k \text{ is even, } x \leq \binom{k-1}{2} - 1 \text{ and } 0 \leq x \bmod k-1 \leq \frac{k-4}{2} \\ 2\left\lfloor \frac{x}{k-1} \right\rfloor + x + 2 & \text{if } k \text{ is even, } x \leq \binom{k-1}{2} - 1 \text{ and } \frac{k-2}{2} \leq x \bmod k-1 \leq k-2 \end{cases} \tag{16}$$

Proof: First, we prove the statement for the maximum value of γ. If there exists an induced subgraph $G(k+1, \binom{k+1}{2} - y)$ in G then in its complement $\overline{G}(k + 1, \binom{k+1}{2} - y)$ there are y edges and hence in $\overline{G}(k + 1, \binom{k+1}{2} - y)$ there is a vertex v with degree at least $\left\lceil \frac{2y}{k+1} \right\rceil$. However, $G(k + 1, \binom{k+1}{2} - y) \setminus \{v\}$ is an induced subgraph $G(k, \binom{k}{2} - (y - \left\lceil \frac{2y}{k+1} \right\rceil) + \epsilon)$, for some $\epsilon \geq 0$. To prove the statement of the lemma, it is sufficient to show that

$$x = y - \left\lceil \frac{2y}{k+1} \right\rceil, \tag{17}$$

for y given in (16). It is easy to verify that y from (16) indeed satisfies equation (17). The proof for the other values of γ follows by induction. ■

Corollary 22. *If $M(k) \leq k\alpha - (\binom{k}{2} - x - 1)$, for a given x, $0 \leq x \leq \binom{k-1}{2}$, then $M(k+1) \leq (k+1)\alpha - (\binom{k+1}{2} - y - 1)$, where y is given in (16).*

As we already saw, constructions of FR codes for given specific parameters can be formulated in terms of graph theory. We formulate the following problems in graph theory that provide some information about the number of nodes of FR codes with maximum file size and the existence of an FR code with a given file size.

Problem 1. Find the value of $N(k, \alpha, \delta)$, $0 \leq \delta \leq \binom{k}{2} - k$, which is the maximum number of vertices such that any α-regular graph G with $n < N(k, \alpha, \delta)$ vertices, where $n\alpha$ is even integer, contains $G(k, \binom{k}{2} - x)$, for some $0 \leq x \leq \delta$.

Problem 2. Find the value of $N'(k,\alpha,\delta)$, $0 \leq \delta \leq \binom{k}{2} - k$, which is the minimum number of vertices such that for any $n \geq N'(k,\alpha,\delta)$, where $n\alpha$ is even integer, there exists an α-regular graph G with n vertices which does not contain $G(k, \binom{k}{2} - x)$, for any $0 \leq x \leq \delta$ but contains $G(k, \binom{k}{2}) - \delta - 1)$.

Problem 3. Let n, k, α, δ be positive integers such that $3 \leq k \leq \alpha$ and $0 \leq \delta \leq \binom{k}{2} - k$. Does there exist an α-regular graph G with n vertices which does not contain $G(k, \binom{k}{2} - x)$ for any $0 \leq x \leq \delta$ but contains $G(k, \binom{k}{2} - \delta - 1)$?

Clearly, an answer to Problem 3 provides a solution to the existence question of FR codes with any file size for $\rho = 2$. Based on Lemma 19 and the solutions to Problem 1 and Problem 2, one can prove the following lemma.

Lemma 23.

$$A(n,k,\alpha,2) \leq k\alpha - \binom{k}{2} + \delta \qquad \textit{if } n < N(k,\alpha,\delta)$$
$$A(n,k,\alpha,2) \geq k\alpha - \binom{k}{2} + \delta + 1 \qquad \textit{if } n \geq N'(k,\alpha,\delta).$$

By our discussion above we have

Corollary 24. *$N(3,\alpha,0) = 2\alpha$; $N'(3,\alpha,0) = 2\alpha$ if α is odd, and $N'(3,\alpha,0) \leq \frac{5}{2}\alpha$ if α is even.*

From the Moore bound and from the Turán bound we have the following corollary.

Corollary 25.

- *$N(k,\alpha,\binom{k}{2} - k) \geq \widehat{n}_0(\alpha, k+1)$, where $\widehat{n}_0(\alpha,k)$ is defined in Subsection II-A.*
- *$N(k,\alpha,0) \geq \left\lceil \frac{k-1}{k-2}\alpha \right\rceil$.*

Problem 4. Find the value of $\eta(k,\alpha,\delta)$, $k-1 \leq \delta \leq \binom{k}{2}$, which is the minimum number of vertices in a graph such that the FR code based on G has file size $k\alpha - \delta$.

Now we consider bounds on $\eta(5,\alpha,\delta)$.

- A code C for which $M(5) = 5\alpha - 4$ is obtained from a graph G if and only if the girth of G is at least 6. Hence, $\eta(5,\alpha,4) \geq 2\alpha^2 - 2\alpha + 2$, by the Moore bound (if $\alpha = q+1$, for a prime power q, then the existence of a projective plane of order q implies that $\eta(5,\alpha,4) = 2\alpha^2 - 2\alpha + 2$).
- A code C for which $M(5) = 5\alpha - 5$ can be obtained from any graph with girth 5. Hence, if there exists an $(\alpha, 5)$-cage then $\eta(5,\alpha,5) \leq \alpha^2 + 1$, by the Moore bound.
- A code C for which $M(5) = 5\alpha - 6$ can be obtained from an $(n,2)$-Turán graph, which is also a graph with girth 4. Hence, $\eta(5,\alpha,6) \leq 2\alpha$, by the Turán's theorem.
- A code C for which $M(5) = 5\alpha - 8$ can be obtained from an $(n,3)$-Turán graphs. Hence, if α is even then $\eta(5,\alpha,8) \leq \frac{3}{2}\alpha$, by the Turán's theorem.
- A code C for which $M(5) = 5\alpha - 9$ can be obtained from an $(n,4)$-Turán graph. Hence, if 3 divides α then $\eta(5,\alpha,9) \leq \frac{4}{3}\alpha$, by the Turán's theorem.
- A code C for which $M(5) = 5\alpha - 10$ can be obtained from an (n,r)-Turán graph, for any $r \geq 5$ (e.g., a complete graph for $r = n$). Hence, $\eta(5,\alpha,10) = \alpha + 1$.

One can see that the value of $M(5) = 5\alpha - 7$ is missing from the list above. However, it is easy to prove that any file size in the interval between $k\alpha - \binom{k}{2}$ and $k\alpha - k + 1$ can be obtained. To prove this claim one can start with a known graph $G = (V, E)$ which contains an induced subgraphs $G(k, t-1)$ but does not contain $G(k, t)$. Let $H = G(k, t-1)$ be an induced subgraph of G; let v_1, v_2 two vertices in H and v_3, v_4 two vertices in $G \setminus H$ such that $e_1 = \{v_1, v_2\}, e_2 = \{v_3, v_4\} \notin E$ and $e_3 = \{v_1, v_3\}, e_4 = \{v_2, v_4\} \in E$. One can easily verify that the graph $G \setminus \{e_3, e_4\} \cup \{e_1, e_2\}$ contains an induced subgraph $G(k, t)$ but does not contain $G(k, t+1)$. This implies that $\eta(k, \alpha, \delta)$ is a decreasing function of δ, and hence $\eta(5, \alpha, 7) \leq \eta(5, \alpha, 6)$. We conjecture that $\eta(5, \alpha, 7) = \ell\alpha$, where $\frac{3}{2} < \ell < 2$.

The discussion of this subsection about FR codes with a given file size raises a lot of questions in graph theory. We leave the research on these questions for future research in graph theory.

IV. Fractional Repetition Codes with Repetition Degree $\rho > 2$

In this section, we consider FR codes with repetition degree $\rho > 2$. Note, that while codes with $\rho = 2$ have the maximum data/storage ratio, codes with $\rho > 2$ provide multiple choices for node repairs. In other words, when a node fails, it can be repaired from different d-subsets of available nodes.

We present generalizations of the constructions from the previous section which were based on Turán graphs and graphs with a given girth. These generalizations employ transversal designs and generalized polygons, respectively.

We start with a construction of FR codes from transversal designs. Let TD be a transversal design TD(ρ, α), $\rho \leq \alpha + 1$, with block size ρ and group size α. Let C_{TD} be an (n, α, ρ) FR code based on TD (see Subsection II-C). Recall that by Lemma 4, there are $\rho\alpha$ points in TD and hence $n = \rho\alpha$. Note, that all the symbols stored in node i correspond to the set N_i of blocks from TD that contain the point i. Since by Lemma 4 there are α blocks that contain a given point, it follows that each node stores α symbols.

Similarly to Theorem 8 we can prove the following theorem.

Theorem 26. *Let $k = b\rho + t$, for $b, t \geq 0$ such that $t \leq \rho - 1$. For an $(n = \rho\alpha, \alpha, \rho)$ FR code C_{TD} based on a transversal design TD(ρ, α) we have*

$$M(k) \geq k\alpha - \binom{k}{2} + \rho\binom{b}{2} + bt.$$

Remark 4. *Note, that for all $k \geq \rho + 1$, the file size of the FR code C_{TD} is strictly larger than the MBR capacity.*

Corollary 27. *Let C_{TD} be an $(r\alpha, \alpha, r)$ FR code based on TD, a transversal design TD(r, α). Let C_T be an $(\frac{r}{r-1}\alpha, \alpha, 2)$ FR code based on (n, r)-Turán graph T. If $M_{TD}(k)$ and $M_T(k)$ are their file sizes, respectively, then*

1) $C_{TD} = C_T$ *for* $r = 2$;
2) $M_{TD}(k) \geq M_T(k)$ *for all* $r \geq 2$.

Example 3. *Let TD be a transversal design TD$(3, 4)$ defined as follows: $\mathcal{P} = \{1, 2, \ldots, 12\}$; $\mathcal{G} = \{G_1, G_2, G_3\}$, where $G_1 = \{1, 2, 3, 4\}$, $G_2 = \{5, 6, 7, 8\}$, and $G_3 = \{9, 10, 11, 12\}$; $\mathcal{B} = \{B_1, B_2, \ldots, B_{16}\}$, and incidence matrix given by*

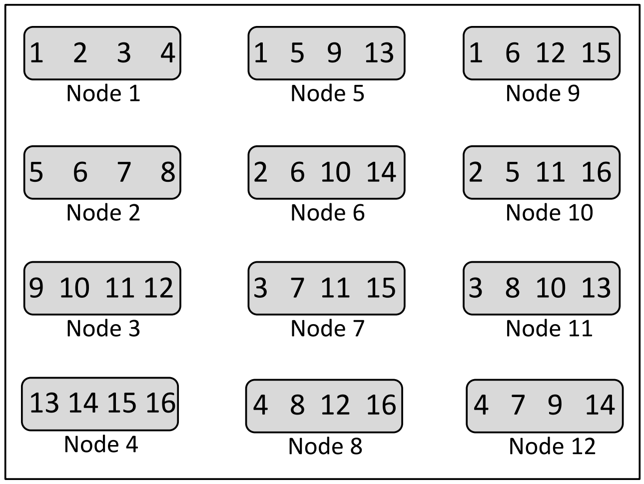

Fig. 4: The FR code based on TD(3,4)

$$
\boldsymbol{I}(TD) = \begin{pmatrix}
1 & 1 & 1 & 1 & & & & & & & & & & & & \\
 & & & & 1 & 1 & 1 & 1 & & & & & & & & \\
 & & & & & & & & 1 & 1 & 1 & 1 & & & & \\
 & & & & & & & & & & & & 1 & 1 & 1 & 1 \\
1 & & & & 1 & & & & 1 & & & & 1 & & & \\
 & 1 & & & & 1 & & & & 1 & & & & 1 & & \\
 & & 1 & & & & 1 & & & & 1 & & & & 1 & \\
 & & & 1 & & & & 1 & & & & 1 & & & & 1 \\
1 & & & & & 1 & & & & & & 1 & & & 1 & \\
 & 1 & & & 1 & & & & & & 1 & & & & & 1 \\
 & & 1 & & & & & 1 & & 1 & & & 1 & & & \\
 & & & 1 & & & 1 & & 1 & & & & & 1 & &
\end{pmatrix}
$$

The placement of the symbols from a codeword of the corresponding MDS code of length 16 *is shown in Fig.4. The values of the file size* $M(k)$ *for* $1 \leq k \leq 4$ *are given in the following table.*

k	$M(k)$
1	*4*
2	*7*
3	*9*
4	*11*

In the following theorem, proved in Appendix B, we find the conditions on the parameters such that the bound on the file size of an FR code C_{TD} from Theorem 26 attains the recursive bound in (5).

Theorem 28. *Let* $\rho \geq 3$*,* $k = b\rho + t \leq \alpha$*,* $0 \leq t \leq \rho - 1$*, and* $\alpha > \alpha_0(k)$*, where*

$$
\alpha_0(k) = \begin{cases} \frac{b^2\rho\binom{\rho-1}{2}+(\rho-2)\binom{t-1}{2}+b((\rho^2+1)(t-1)-\rho(3t-4))}{\rho-t+1} & \textit{if } k \not\equiv 0 \ (\textit{mod } \rho) \\ b(b\rho-2)\binom{\rho-1}{2}+b-1 & \textit{if } k \equiv 0 \ (\textit{mod } \rho) \end{cases} .
$$

The file size $M(k)$ of the $(\rho\alpha, \alpha, \rho)$ FR code C_{TD} is given by

$$M(k) = k\alpha - \binom{k}{2} + \rho\binom{b}{2} + bt$$

and attains the bound in (5) for all $k \leq \alpha$.

Example 4. *We illustrate the minimum values of α for which the FR code obtained from a TD(ρ, α) is optimal as a consequence of Theorem 28.*

k \ ρ	3	4	5	6
3	*2*	*3*	*4*	*5*
4	*7*	*7*	*10*	*13*
5	*8*	*17*	*19*	*25*
6	*10*	*22*	*36*	*41*
7	*19*	*29*	*47*	*67*
8	*21*	*38*	*61*	*86*

Similarly to the case in which $\rho = 2$, we continue to find the conditions when there exists an FR code C with file size $M(3) = 3\alpha - 2$. To have a file size greater than $3\alpha - 3$, we should avoid the existence of a 3×3 submatrix $\mathbf{I}'$ of $\mathbf{I}(C)$ such that each row of $\mathbf{I}'$ has exactly two ones (recall that the intersection between any two rows is at most one). Such a matrix $\mathbf{I}'$ will be called a *triangle*.

Lemma 29. *If $n < \rho(\rho-1)\alpha - \rho(\rho-2)$ then there exists a triangle in the incidence matrix of an FR code C, i.e., a necessary condition for $M(3) = 3\alpha - 2$ is that $n \geq \rho(\rho-1)\alpha - \rho(\rho-2)$.*

Proof: W.l.o.g. assume that the α 1's of the first row of $\mathbf{I}(C)$ are in the first α columns. Let S be the set of $(\rho-1)\alpha$ rows of $\mathbf{I}(C)$ which have common ones with the first row of $\mathbf{I}(C)$. Each row of S contains exactly one 1 in the first α columns and $\alpha - 1$ 1's in other columns of $\mathbf{I}(C)$. To avoid a triangle in the matrix, the 1's in S which do not appear in the first α columns must appear in different columns. Hence, we have $\theta \geq \alpha + (\alpha-1)(\rho-1)\alpha = (\rho-1)\alpha^2 - (\rho-2)\alpha$. Since $\theta = \frac{n\alpha}{\rho}$ the claim of the lemma is proved. ■

By Lemma 29 it follows that for $n < \rho(\rho-1)\alpha - \rho(\rho-2)$ and $k = 3$ the file size of an FR code C equals $M(3) = 3\alpha - 3$. However, the bound in (5) satisfies $\varphi(3) = 3\alpha - 1 - \left\lceil \frac{2(\rho-1)\alpha-\rho}{n-2} \right\rceil = 3\alpha - 2$ if and only if $n \geq 2(\rho-1)\alpha - (\rho-2)$. Thus, we have the following lemma

Lemma 30. *The bound in (5) for $k = 3$ and $\rho > 2$ is not tight in the interval $n \in [2(\rho-1)\alpha - (\rho-2), \rho(\rho-1)\alpha - \rho(\rho-2))$.*

Next, we present a construction of an FR code C based on a generalized quadrangle, for which $M(3) = 3\alpha - 2$. This code attains the bound on n presented in Lemma 29. The following lemma follows directly from the definition of a generalized quadrangle.

Lemma 31. *Let GQ be a generalized quadrangle GQ(s, t), where $t \geq s$, and let C_{GQ} be the FR code based on GQ. C_{GQ}*

is an $(n = (s+1)(st+1), \alpha = t+1, \rho = s+1)$ *FR code for which* $M(3) = 3\alpha - 2$ *and* $M(4) = 4\alpha - 4$. *Moreover, this code attains the bound on* n *of Lemma 29.*

Remark 5. *Similarly to an FR code* C_G *with* $\rho = 2$ *based on a graph* G *with girth* g, *we can consider an FR code* C_{GP} *based on a generalized* g*-gon (generalized polygon GP) for* $\rho > 2$. *One can prove that the file size of* C_{GP} *is identical to the file size of* C_G *for* $k \leq g + \lceil \frac{g}{2} \rceil - 2$ *given in Theorem 15. However, a generalized* g*-gon is known to exist only for* $g \in \{3, 4, 6, 8\}$. *This observation also holds for a biregular bipartite graph of girth* $2g$. *The existence of such graphs was considered in [1], [2], [5], [15], [17].*

Remark 6. *Note that both generalized quadrangles and transversal designs are examples of* partial geometries. *Codes for distributed storage systems based on partial geometries were also considered in [29].*

Note that the problem of constructing for FR codes with repetition degree greater than two, based on combinatorial designs, also can be considered in terms of *expander* graphs (see e.g [21]). Let C_D be an FR code based on a combinatorial design D. If we consider the incidence graph $G_D = (L \cup R, E)$ of D, where L corresponds to the points and R corresponds to the blocks of D, then calculating $M(k)$ can be described by calculating the set of neighbours of any subset of size k of the part L of G_D. In other words, for an FR code with the file size $M(k)$ it should hold that $|\Gamma(A)| \geq M(k)$ for every $A \subseteq L$ of size k, where $\Gamma(A)$ denotes the set of neighbours of A. Hence, to have an FR code with $M(k)$, we need to construct a $(k, \frac{M(k)}{k})$ expander graph, where $\frac{M(k)}{k}$ is its expansion factor [21].

V. File Size of FR Codes and Generalized Hamming Weights

Let C be a (θ, k) linear code and A be a subcode of C. The *support* of A, denoted by $\chi(A)$, is defined by

$$\chi(A) \stackrel{\text{def}}{=} \{i : \exists (c_1, c_2, \ldots, c_\theta) = \mathbf{c} \in A, c_i \neq 0\}.$$

The rth *generalized Hamming weight* of a linear code C, denoted by $d_r(C)$ (d_r in short), is the minimum support of any r-dimensional subcode of C, $1 \leq r \leq k$, namely,

$$d_r = d_r(C) \stackrel{\text{def}}{=} \min_A \{|\chi(A)| : A \subseteq C, \dim(A) = r\}.$$

Clearly, $d_r \leq d_{r+1}$ for $1 \leq r \leq k-1$. The set $\{d_1, d_2, \ldots, d_k\}$ is called the *generalized Hamming weight hierarchy* of C [44].

There are a few definitions of generalized Hamming weights for nonlinear codes [12], [16], [35]. We propose now another straightforward definition for generalized Hamming weight hierarchy for nonlinear codes. This definition is strongly connected to the file sizes for different values of k of a given FR code C.

Let C be a code of length θ with n codewords. Assume further that the all-zero vector is not a codeword of C (if the all-zero vector is a codeword of C we omit it from the code). The rth generalized Hamming weight of C, $d_r(C)$, will be defined as the minimum support of any subcode of C with r codewords, i.e.,

$$d_r = d_r(C) \stackrel{\text{def}}{=} \min_A \{|\chi(A)| : A \subseteq C, |A| = r\}.$$

Note that an (n,α,ρ) FR code C can be represented as a binary constant weight code $\mathcal{C}$ of length θ and weight α. Note further that the minimum Hamming distance of $\mathcal{C}$ is $2\alpha-2$. Finally note that with these definitions we have that $d_k = M(k)$. Therefore, by our previous discussion and the definition of the generalized Hamming weight hierarchy it is natural to define the *file size hierarchy* of an FR code C to be the same as the generalized Hamming weight hierarchy of the related binary constant weight code $\mathcal{C}$.

In addition to the questions discussed in the previous sections, the definition of the file size hierarchy raises some natural questions.

1) Do there exist two FR codes C_1 and C_2, with the same parameters n,α,ρ, and two integers k_1 and k_2, such that $M_1(k_1) < M_2(k_1)$ and $M_1(k_2) > M_2(k_2)$, where $M_1(k)$ and $M_2(k)$ are the file sizes of C_1, C_2, respectively?
2) Given n,α,ρ and a file size hierarchy $\{d_1=\alpha, d_2, \ldots, d_\alpha\}$, does there exist an FR code C with these parameters, which satisfies for each $k \le \alpha$ that $d_k = M(k)$ for every $k \le \alpha$?

VI. Bound on Reconstruction Degree

In this section we consider a lower bound on the reconstruction degree k for an FR code, given M, θ, n, and α. Note, that given the value of file size M it is desirable to have a reconstruction degree k as small as possible, to provide the maximum possible failure resilience for the related DSSs. Hence, it is of interest to obtain FR codes which attain this bound.

Lemma 32. *Let C be an (n,α,ρ) FR code which stores a file of a given size M. The reconstruction degree k of the corresponding system should satisfy*

$$k \geq \left\lceil \frac{n\binom{M-1}{\alpha}}{\binom{\theta}{\alpha}} \right\rceil + 1. \tag{18}$$

Proof: Let X be an $n \times \binom{\theta}{M-1}$ binary matrix whose rows are indexed by $\{N_i\}_{i=1}^n$ of C and whose columns are indexed by all the possible $(M-1)$-subsets of $[\theta]$. The value $X_{i,j}$ is one if and only if N_i is a subset of the $(M-1)$-set representing the jth column. We count the number of ones in the matrix X in two different ways, to obtain a lower bound on the value of the reconstruction degree k.

Since $|N_i| = \alpha$ and $N_i \subseteq [\theta]$, it follows that there are $\binom{\theta-\alpha}{M-1-\alpha}$ ways to complete an α-subset of $[\theta]$ to an $(M-1)$-subset of $[\theta]$. Hence the number of ones in each row of X is given by $\binom{\theta-\alpha}{M-1-\alpha}$. The number of ones in each column of X is at most $(k-1)$, since otherwise, there exist k subsets $N_{i_1}, \ldots, N_{i_k}$ such that $|\cup_{s=1}^k N_{i_s}| \le M-1$ which contradicts to the data reconstruction of FR codes, in other words, it is not possible to reconstruct the file from the k nodes indexed by $i_1, \ldots, i_k$. Hence,

$$n\binom{\theta-\alpha}{M-1-\alpha} \leq \binom{\theta}{M-1}(k-1),$$

which can be rewritten as

$$\frac{n}{k-1} \leq \frac{\binom{\theta}{\alpha}}{\binom{M-1}{\alpha}},$$

and the lemma follows. ∎

Two families of FR codes which attain the bound in (18) are presented in the following two lemmas. The corresponding codes correct two node erasures, i.e., $n-k=2$, and the data/storage ratio is $\frac{\theta-1}{2\theta}$, i.e., almost 1/2.

Lemma 33. *Let $n > 2$ be is an even integer and let K_n^- be an $(n-2)$-regular graph obtained by removing a perfect matching from the set of edges in K_n. Then the FR code $C_{K_n^-}$ based on the graph K_n^- attains the bound in (18) for $k = n-2$.*

Proof: First, observe that a perfect matching in K_n is of size $\frac{n}{2}$ and hence the number of edges in K_n^- is $\theta = \binom{n}{2} - \frac{n}{2}$. Any $n-2$ vertices of K_n^- are incident with at least $\theta - 1$ edges and therefore $M(n-2) = \theta - 1$. To prove that $C_{K_n^-}$ attains the bound in (18) for $k = n-2$ we have to prove that

$$n - 3 = \left\lceil \frac{n\binom{\theta-2}{\alpha}}{\binom{\theta}{\alpha}} \right\rceil$$

for $\alpha = n-2$. Since

$$\frac{n\binom{\theta-2}{\alpha}}{\binom{\theta}{\alpha}} = \frac{n(\theta-\alpha)(\theta-1-\alpha)}{(\theta-1)\theta} = \frac{(n-2)((n-2)^2-2)}{n^2-2n-2} = n - 4 + \frac{4n-12}{n^2-2n-2},$$

the statement of the lemma is proved. ■

Similarly to Lemma 33 one can prove the following lemma.

Lemma 34. *The FR code C_{K_n} based on K_n attains the bound in (18) for $k = n-2$.*

Remark 7. *C_{K_n} was first defined in [32], [33]). Note that this code is also an MBR code for which $n = \alpha + 1$.*

As we already noticed, when all the other four parameters M, θ, α and n (or, equivalently, M, ρ, α and n) are fixed, it is desirable to have the smallest possible reconstruction degree k. However, when only three parameters M, ρ and α are fixed, we have a trade-off between the reconstruction degree k and the storage overhead of the DSS. We illustrate this trade-off by the following example. Suppose that we want to store a file of size $M = 36$ by using an FR code with $\alpha = 8$ and $\rho = 2$. Note that n is not fixed. First, for $k = 5$ we consider a $(114, 8, 2)$ FR code based on a projective plane $PG(2, 7)$ (see Example 2). Recall that this code is the optimal code, i.e., it stores a file of the maximum size since the corresponding graph has girth 6 which is greater than $k = 5$ (see Corollary 13), and in addition, it has the minimum possible value of n, as the corresponding graph is a Moore graph. However, the length of a corresponding MDS codeword and then the total storage is 456 and hence the storage overhead is much more than 1000%. To have smaller overhead one can use a $(10, 5)$-Turán graph which for $k = 7$ by Theorem 8 yields a $(10, 8, 2)$ FR code for which we can take a file of size $M = 37$ and encode it with a $(40, 37)$ MDS code. Hence, the total overhead is only about 10% and moreover, the field size for the FR code is required to be much smaller than in the previous code. This example illustrates the fact that if we are given the file size M, the number of symbols α stored in a node, and the repetition degree ρ, then decreasing the reconstruction degree k increases the storage overhead significantly.

VII. Fractional Repetition Batch Codes

In this section we analyze additional properties of DRESS codes (FR codes for which the reconstruction degree k is determined; see Section I) which allow also load balancing between storage nodes by establishing a connection to combinatorial batch codes. We consider a scenario when in addition to the uncoded exact repairs of failed nodes and to the recoverability of the stored file from any set of k nodes we require an additional property. Given a positive integer t, any t-subset of stored symbols can be retrieved by reading at most one symbol from each node. This retrieval can be performed by t different users in parallel, where each user gets a different symbol. In other words, we propose a new type of codes for DSS, called in the sequel *fractional repetition batch* (FRB) codes, which enable uncoded efficient node repairs and load balancing which is performed by several users in parallel. An FRB code is a combination of an FR code and an uniform combinatorial batch code.

The family of codes called *batch codes* was proposed for load balancing in distributed storage. A *batch code*, introduced in [22], stores θ (encoded) data symbols on n system nodes in such a way that any batch of t data symbols can be decoded by reading at most one symbol from each node. In a ρ*-uniform combinatorial batch code*, proposed in [31], each node stores a subset of the data symbols and no decoding is required during retrieval of any batch of t symbols. Each symbol is stored in exactly ρ nodes and hence it is also called a replication based batch code. A ρ-uniform combinatorial batch code is denoted by $\rho-(\theta, N, t, n)$-CBC, where $N = \rho\theta$ is the total storage over all the n nodes. These codes were studied in [6], [9], [11], [22], [31], [39].

Next, we provide a formal definition of FRB codes. This definition is based on the definitions of a DRESS code and a uniform combinatorial batch code. Let $\mathbf{f} \in \mathbb{F}_q^M$ be a file of size M and let $c_{\mathbf{f}} \in \mathbb{F}_q^\theta$ be a codeword of an (θ, M) MDS code which encodes the data $\mathbf{f}$. Let $\{N_1, \dots, N_n\}$ be a collection of α-subsets of the set $[\theta]$. A $\rho-(n, M, k, \alpha, t)$ *FRB code* C, $k \leq \alpha$, $t \leq M$, represents a system of n nodes with the following properties:

1) Every node i, $1 \leq i \leq n$, stores α symbols of $c_{\mathbf{f}}$ indexed by N_i;
2) Every symbol of $c_{\mathbf{f}}$ is stored on ρ nodes;
3) From any set of k nodes it is possible to reconstruct the stored file $\mathbf{f}$, in other words, $M = \min_{|I|=k} |\cup_{i\in I} N_i|$;
4) Any batch of t symbols from $c_{\mathbf{f}}$ can be retrieved by downloading at most one symbol from each node.

Note that the total storage over the n nodes needed to store the file $\mathbf{f}$ equals to $n\alpha = \theta\rho$.

Remark 8. *Note that while in a classical batch code any t* data *symbols can be retrieved, in an FRB code any batch of t* coded *symbols can be retrieved. In particular, when a systematic MDS code is chosen for an FRB code, the data symbols can be easily retrieved.*

To present constructions of FRB codes, we need the following results on constructions of uniform combinatorial batch codes.

Theorem 35. *[31] Let G be a graph with n vertices, θ edges and girth g. Then the batch code C_G^B with nodes indexed by the vertices of G and with data symbols indexed by the edges of G, is a $2-(\theta, 2\theta, t, n)$-CBC with $t = 2g - \lfloor g/2 \rfloor - 1$.*

Theorem 36. *[39] Let TD be a resolvable transversal design TD$(q-1, q)$, for a prime power q. Then the batch code C_{TD}^B with nodes indexed by points and data symbols indexed by blocks of TD, is a $(q-1)-(q^2, q^3-q^2, q^2-q-1, q^2-q)$-CBC.*

By applying Theorems 35, Theorem 36 together with Corollary 10, Theorem 15, and Theorem 26 we obtain the following result.

Theorem 37.

1) *Let $K_{\alpha,\alpha}$ be a complete bipartite graph with $\alpha > 2$. Then $C_{K_{\alpha,\alpha}}$ is a $2-(2\alpha, M, k, \alpha, 5)$ FRB code with $M = M(k) = k\alpha - \left\lfloor \frac{k^2}{4} \right\rfloor$.*

2) *Let G be an α-regular graph on n vertices with girth g. Then C_G is a $2-(n, M, k, \alpha, 2g - \lfloor g/2 \rfloor - 1)$ FRB code with*

$$M = M(k) = \begin{cases} k\alpha - k + 1 & \text{if } k \leq g - 1 \\ k\alpha - k & \text{if } g \leq k \leq g + \lceil \frac{g}{2} \rceil - 2. \end{cases}$$

3) *Let TD be a resolvable transversal design TD$(\alpha - 1, \alpha)$, for a prime power α. Then C_{TD} is an $(\alpha - 1) - (\alpha^2 - \alpha, M, k, \alpha, \alpha^2 - \alpha - 1)$ FRB code with $M \geq k\alpha - \binom{k}{2} + (\alpha - 1)\binom{x}{2} + xy$, where x, y are nonnegative integers which satisfy $k = x(\alpha - 1) + y$, $y \leq \alpha - 2$.*

Example 5.

- *Consider the FRB code $C_{K_{3,3}}$ based on $K_{3,3}$ (see also Example 1 for an FR code based on $K_{3,3}$). By Theorem 37, for $k = 3$, $C_{K_{3,3}}$ is a $2 - (6, 7, 3, 3, 5)$ FRB code.*
- *Consider the FRB code C_{TD} based on the resolvable transversal design TD = TD$(3, 4)$ (see also Example 3 for an FR code based on TD$(3, 4)$). By Theorem 37, for $k = 4$, C_{TD} is a $3 - (12, 11, 4, 4, 11)$ FRB code, which stores a file of size 11 and allows for retrieval of any (coded) 11 symbols, by reading at most one symbol from a node. In particular, when using a systematic MDS code, C_{TD} provides load balancing in data reconstruction.*

For the rest of this section we consider FRB codes obtained from affine planes. Uniform combinatorial batch codes based on affine planes were considered in [39].

An *affine plane* of order s, denoted by $A(s)$, is a design $(\mathcal{P}, \mathcal{B})$, where $\mathcal{P}$ is a set of $|\mathcal{P}| = s^2$ *points*, $\mathcal{B}$ is a a collection of s-subsets (*blocks*) of $\mathcal{P}$ of size $|\mathcal{B}| = s(s+1)$, such that each pair of points in $\mathcal{P}$ occur together in exactly one block of $\mathcal{B}$. An affine plane is called *resolvable*, if the *set* $\mathcal{B}$ can be partitioned into $s+1$ sets of size s, called parallel classes, such that every element of $\mathcal{P}$ is contained in exactly one block of each class. It is well known [3] that if q is a prime power, then there exists a resolvable affine plane $A(q)$.

Theorem 38. *Let $A(q)$ be an affine plane, for a prime power q, and let $C_{A(q)}$ be an FRB based on $A(q)$, i.e., $\boldsymbol{I}(C_{A(q)}) = \boldsymbol{I}(A(q))$. Then $C_{A(q)}$ is a $q - (q^2, k(q+1) - \binom{k}{2}, k, q+1, q^2)$ FRB code.*

Proof: The parameters ρ, n, α and t follow from the properties of the batch code based on $A(q)$ (see [39] for details). Since any two points of $A(q)$ are contained in exactly one block and hence any two rows of $\mathbf{I}(C_{A(q)})$ intersect, it follows that the file size is $k(q+1) - \binom{k}{2}$. ∎

Remark 9. *The structure of an incidence matrix of an FRB code and some other constructions are considered in [38].*

VIII. Conclusion

We considered the problem of constructing (n, α, ρ) FR codes which attain the upper bound on the file size. We presented constructions of FR codes based on regular graphs, namely, Turán graphs and graphs with a given girth; and constructions of FR codes based on combinatorial designs, namely, transversal designs and generalized polygons. The problems of constructing optimal FR codes and FR codes with a given file size raise interesting questions in graph theory (see Section III and Section IV). We defined the file size hierarchy of FR codes, which is a possible definition of generalized Hamming weight hierarchy for nonlinear codes. For this we presented FR codes as binary constant weight codes. In addition, we derived a lower bound on the reconstruction degree for FR codes and presented FR codes which attain this bound. Finally, based on a connection between FR codes and batch codes, we proposed a new family of codes for DSS, namely fractional repetition batch codes, which have the properties of batch codes and FR codes simultaneously. These are the first codes for DSS which allow for uncoded efficient repairs and load balancing. We presented examples of constructions for FRB codes, based on combinatorial designs, complete bipartite graphs, and graphs with large girth.

In general, given four of the five parameters of FR codes, namely, the number n of nodes, the number k of nodes needed to reconstruct the whole stored file, the number α of stored symbols in a node, the number ρ of repetitions of a symbol in the code, and the size M of the stored file $\mathbf{f}$, one can ask what are the possible values of the fifth parameter. For this we define the following five functions.

1) Let $n(k, \alpha, \rho, M)$ be the minimum number n of nodes for an (n, α, ρ) FR code which stores a file of size M for a given reconstruction degree k.
2) Let $k(n, \alpha, \rho, M)$ be the minimum number k of nodes from which the whole stored file of size M, of an (n, α, ρ) FR code, can be reconstructed.
3) Let $\alpha(n, k, \rho, M)$ be the minimum number α of symbols stored in a node of an (n, α, ρ) FR code which stores a file of size M for a given reconstruction degree k.
4) Let $\rho(n, k, \alpha, M)$ be the minimum number ρ of repetitions of a symbol in an (n, α, ρ) FR code which stores a file of size M for a given reconstruction degree k.
5) Let $M(n, k, \alpha, \rho)$ be the maximum size of a stored file in an (n, α, ρ) FR code, for a given reconstruction degree k.

This formulation is very similar to classical coding theory, where for example three functions are defined for the trade-off between the length of the code, its size, and its minimum distance. In this paper, we considered the values of three functions out of the five, namely, $n(k, \alpha, \rho, M)$, $k(n, \alpha, \rho, M)$ and $M(n, k, \alpha, \rho)$.

APPENDIX A

Constructions for FR codes with $\boldsymbol{M(3) = 3\alpha - 2}$: Let $\mathbf{I}(K_{\alpha+i,\alpha+i})$, $i \geq 0$, be the $2(\alpha+i) \times (\alpha+i)^2$ incidence matrix of the complete bipartite graph $K_{\alpha+i,\alpha+i}$. Note that it is also the incidence matrix of a resolvable transversal design $\text{TD}(2, \alpha+i)$. Based on the resolvability of the design, $\mathbf{I}(K_{\alpha+i,\alpha+i})$ can be written in a blocks form, i.e., $\mathbf{I}(K_{\alpha+i,\alpha+i})$ is a $2 \times (\alpha+i)$ blocks matrix, where each block is a permutation matrix of size $(\alpha+i) \times (\alpha+i)$. Each such permutation matrix block will be called a *p-block*. W.l.o.g. we assume that the first two p-blocks which correspond to the first $\alpha+i$ columns of the matrix are identity matrices. Let $\mathbf{I}^{\text{even}}_{\alpha,i}$ be a matrix obtained from $\mathbf{I}(K_{\alpha+i,\alpha+i})$ by removing $i(\alpha+i)$ columns which correspond to $2i$ p-blocks (w.l.o.g. we assume that we removed the leftmost columns). Note that there are exactly α ones in each row of $\mathbf{I}^{\text{even}}_{\alpha,i}$. Let $C^{\text{even}}_{\alpha,i}$ be an FR code obtained from the graph $G^{\text{even}}_{\alpha,i}$, whose incidence matrix is $\mathbf{I}^{\text{even}}_{\alpha,i}$. It is easy to verify that $C^{\text{even}}_{\alpha,i}$ is a $(2\alpha+2i, \alpha, 2)$ code whose file size is $M(3) = 3\alpha - 2$. Note that the data/storage ratio $\frac{M(3)}{n\alpha} = \frac{3\alpha-2}{(2\alpha+2i)\alpha}$ decreases when i increases.

For odd n, $n \geq \frac{5}{2}\alpha$, and even α we construct an $(n, \alpha, 2)$ FR code C with $M(3) = 3\alpha - 2$ as follows. We distinguish between two cases, odd $\alpha/2$ and even $\alpha/2$.

For odd $\alpha/2$ let $G = (V, E)$ be a graph whose vertex set is given by $V = \{X, Y_0, Y_1, Z_0, Z_1\}$, where $|X| = |Y_0| = |Y_1| = \alpha/2$, $|Z_0| = |Z_1| = \alpha/2 + j$, for any given integer $j \geq 0$. The edges in G are given by

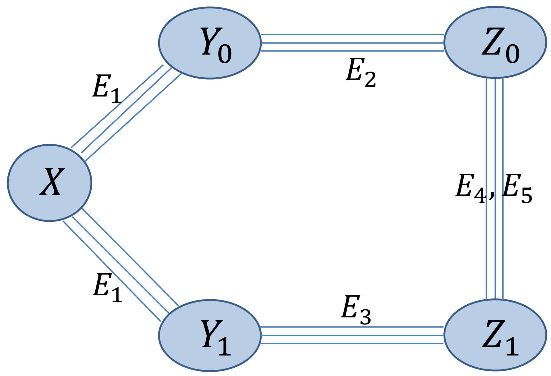

Fig. 5: A schematic structure of a graph without triangles, with odd number of vertices

1) $E_1 = \{\{v,u\} : v \in X,\ u \in Y_i, i = 0,1\}$;
2) $E_{2+\ell} = \{\{v_i,u_j\} : v_i \in Y_\ell,\ u_j \in Z_\ell, i = 1,2,\ldots,\frac{\alpha}{2}, j = ((i-1)\frac{\alpha}{2}+1) \bmod |Z_\ell|, \ldots, i\frac{\alpha}{2} \bmod |Z_\ell|\}$, $\ell = 0,1$, such that the degree of a vertex x in the induced subgraph $(Y_\ell \cup Z_\ell, E_{2+\ell})$, is given by

$$\deg(x) = \begin{cases} \frac{\alpha}{2} & x \in Y_\ell \\ t := \left\lfloor \frac{\alpha^2/4}{|Z_1|} \right\rfloor \text{ or } t+1 & x \in Z_\ell \end{cases}$$

3) $E_4 = \{\{v_i,u_i\} : v_i \in Z_0, u_i \in Z_1, \deg(v_i) = \deg(u_i) = t$ in $(Y_0 \cup Z_0, E_2)$ and $(Y_1 \cup Z_1, E_3)$
4) $E_5 = \{\{v,u\} : v \in Z_0, u \in Z_1, s.t. (Z_0 \cup Z_1, E_5) = G^{\text{even}}_{(\alpha-t-1),(|Z_\ell|-\alpha+t+1)}\}$.

The schematic graph G is shown on Fig. 5. Clearly, the degree of every vertex in G is α and there are $\frac{5}{2}\alpha + 2j$ vertices in G.

For even $\alpha/2$, we define the graph G in the similar way to the case in which $\alpha/2$ is odd. The only difference is that $|X| = \alpha/2 - 1$ and $|Z_0| = |Z_1| = \alpha/2 + 1 + j$; $E_1, E_2, \ldots, E_5$ are defined similarly and the number of vertices in G is $\frac{5}{2}\alpha + 1 + 2j$. We illustrate this construction with the following example.

Example 6. *For $\alpha = 4$ the incidence matrix $\boldsymbol{I}(G)$ of the graph G is given by*

$$\left(\begin{array}{cccc|cccccc|cccccc|cccccc}
1 & 1 & 1 & 1 & & & & & & & & & & & & & & & & & & \\
\hline
1 & & & & 1 & 1 & 1 & & & & & & & & & & & & & & & \\
 & 1 & & & & & & 1 & 1 & 1 & & & & & & & & & & & & \\
\hline
 & & 1 & & & & & & & & 1 & 1 & 1 & & & & & & & & & \\
 & & & 1 & & & & & & & & & & 1 & 1 & 1 & & & & & & \\
\hline
 & & & & 1 & & & 1 & & & & & & & & & 1 & & & 1 & & \\
 & & & & & 1 & & & 1 & & & & & & & & & 1 & & & 1 & \\
 & & & & & & 1 & & & 1 & & & & & & & & & 1 & & & 1 \\
\hline
 & & & & & & & & & & 1 & & & 1 & & & 1 & & & & 1 & \\
 & & & & & & & & & & & 1 & & & 1 & & & 1 & & 1 & & \\
 & & & & & & & & & & & & 1 & & & 1 & & & 1 & & & 1
\end{array}\right),$$

where an empty entry in the matrix is 0 and the blocks of the matrix correspond to the partition of vertices and edges as *described in the construction.*

Proof of Lemma 17: The upper bound on n_3 directly follows from the discussion in the beginning of Subsection III-C. To prove the lower bound, we need to show that there is no α-regular graph with $2\alpha+1$ vertices which does not contain a cycle of length 3. Let $G=(V,E)$ be an α-regular graph with $2\alpha+1$ vertices. Let $v \in V$ and let $A=\{a_1,a_2,\ldots,a_\alpha\}$ be the set of its neighbours. Since there are no cycles of length 3, there are no edges between vertices of A. Let $B=\{b_1,b_2,\ldots,b_\alpha\}$ be the remaining α vertices of G. Clearly, every vertex of A has $\alpha-1$ neighbours in B. The number of edges in the cut between A and B is $\alpha(\alpha-1)$ and hence the number of edges between the vertices in B is $\frac{\alpha^2-(\alpha(\alpha-1))}{2}=\frac{\alpha}{2}$. W.l.o.g. let $\{a_1,b_1\} \notin E$ and $\{a_2,b_1\} \in E$. The vertices of $B \setminus \{b_1\}$ are neighbours of a_1, which implies that there are no edges between vertices in $B \setminus \{b_1\}$. Therefore, b_1 has $\frac{\alpha}{2}$ neighbours in $B \setminus \{b_1\}$ and a_2 has $\alpha-2$ neighbours in $B \setminus \{b_1\}$. This implies that if $\alpha > 2$ then there exists a common neighbour b_ℓ of a_2 and b_1 in $B \setminus \{b_1\}$. Hence, a_2, b_1 and b_ℓ form a cycle, a contradiction.

■

Example 7. *Let $\alpha=7$ and $M(4)=4\alpha-5=23$. By the Turán bound (see Corollary 2), the minimum number of nodes in an FR code is $\left\lceil \frac{3}{2}\alpha \right\rceil = 11$, but since $n\alpha$ must to be an even number, it follows that $n \geq 12$. The corresponding Turán graph $T(12,3)$ with $n=12$ has degree 8. To obtain from $T(12,3)$ a 7-regular graph on 12 vertices without $G(4,6)$, one can remove 6 edges, which correspond to a perfect matching in $T(12,3)$.*

Appendix B

Proof of Theorem 28: We observe that if the lower bound on the file size of Theorem 26 is equal to the upper bound in (5) then we have

$$M(k) = k\alpha - \binom{k}{2} + \rho\binom{b}{2} + bt = \varphi(k). \tag{19}$$

We prove (19) by induction. We distinguish between two cases: $k=b\rho+t$, where $t \neq 0$, i.e., $k \not\equiv 0 \pmod{\rho}$ and $k=b\rho$, i.e., $k \equiv 0 \pmod{\rho}$.

Case 1: if $k=b\rho+t$, $0<t\leq \rho-1$, then $k-1=b\rho+t-1$, where $0 \leq t-1 \leq \rho-2$. Assume that $\varphi(k-1) = M(k-1) = (k-1)\alpha - \binom{k-1}{2} + \rho\binom{b}{2} + b(t-1)$. By applying the recursive formula for φ, we want to prove that

$$k\alpha - \binom{k}{2} + \rho\binom{b}{2} + bt = (k-1)\alpha - \binom{k-1}{2} + \rho\binom{b}{2} + b(t-1) + \alpha - \left\lceil \frac{(\rho-1)(k-1)\alpha - \rho\binom{k-1}{2} + \rho^2\binom{b}{2} + \rho b(t-1)}{\rho\alpha-k+1} \right\rceil \tag{20}$$

By simplifying the last equation one can verify that (20) holds if and only if

$$(\rho-1)(k-1)\alpha - \rho\binom{k-1}{2} + \rho^2\binom{b}{2} + \rho b(t-1) > (k-b-2)(\rho\alpha-k+1). \tag{21}$$

By substituting $k=b\rho+t$ in (21) we have that (21) holds if and only if

$$\alpha > \frac{b^2\rho\binom{\rho-1}{2} + (\rho-2)\binom{t-1}{2} + b((\rho^2+1)(t-1) - \rho(3t-4))}{\rho-t+1}.$$

Case 2: if $k=b\rho$ then $k-1=(b-1)\rho+\rho-1$. Assume that $\varphi(k-1)=M(k-1)=(k-1)\alpha-\binom{k-1}{2}+\rho\binom{b-1}{2}+b(\rho-1)$. By applying the recursive formula for φ, we want to prove that

$$k\alpha-\binom{k}{2}+\rho\binom{b}{2}=(k-1)\alpha-\binom{k-1}{2}+\rho\binom{b-1}{2}+b(\rho-1)+\alpha-\left\lceil\frac{(\rho-1)(k-1)\alpha-\rho\binom{k-1}{2}+\rho^2\binom{b-1}{2}+b\rho(\rho-1)}{\rho\alpha-k+1}\right\rceil. \tag{22}$$

One can verify that (22) holds if and only if

$$(\rho-1)(k-1)\alpha-\rho\binom{k-1}{2}+\rho^2\binom{b-1}{2}+b\rho(\rho-1)>(k-b-1)(\rho\alpha-k+1). \tag{23}$$

By substituting $k=b\rho$ in (23) we have that (23) holds if and only if

$$\alpha>\frac{b^2\rho}{2}(\rho^2-3\rho+2)-b(\rho^2-3\rho+1)-1=b(b\rho-2)\binom{\rho-1}{2}+b-1.$$

■

References

[1] M. Abreu, G. Araujo-Pardo, C. Balbuena, and D. Labbate, "Families of small regular graphs of girth 5," *Discrete Mathematics* 312 (2012), 2832-2842.

[2] M. Abreu, G. Araujo-Pardo, C. Balbuena, D. Labbate, and G. López-Chávez, "Constructions of biregular cages of girth five," *Electronic Notes in Discrete Mathematics* 40 (2013), 9-14.

[3] I. Anderson, *Combinatorial designs and tournaments*, Clarendon Press, Oxford, 1997.

[4] S. Anil, M. K. Gupta, and T. A. Gulliver, "Enumerating some fractional repetition codes," *arXiv:1303.6801*, Mar. 2013.

[5] G. Araujo-Pardo, C. Balbuena, and J.C. Valenzuela, "Constructions of bi-regular cages," *Discrete Mathematics* 309 (2009), 1409-1416.

[6] N. Balachandran and S. Bhattacharya, "On an extremal hypergraph problem related to combinatorial batch codes", *arXiv:1206.1996v4*, Oct. 2012.

[7] K. G. Benerjee, M. K. Gupta, and N. Agrawal, "Reconstruction and repair degree of fractional repetition codes", *arXiv:1305.4580v1*, May. 2013.

[8] S. L. Bezrukov, "Edge isoperimetric problems on graphs", *in: L. Lovsz, A. Gyarfas, G. O. H. Katona, A. Recski, L. Szekely (Eds.), Graph Theory and Combinatorial Biology, Bolyai Society Mathematical Studies*, vol. 7, Budapest, 1999, pp. 157-197.

[9] S. Bhattacharya, S. Ruj, and B. Roy, "Combinatorial batch codes: A lower bound and optimal constructions," *Advances in Mathematics of Communications* 6 (2012), 165-174.

[10] N. Biggs, *Algebraic Graph Theory*, Cambridge Univ. Press, 2nd Ed., 1993.

[11] Cs. Bujtás and Zs. Tuza, "Turán numbers and batch codes", *arXiv:1309.6506v1*, Sep. 2013.

[12] G. Cohen, S. Litsyn, and G. Zemor, "Upper bounds on generalized distances," *IEEE Trans. Inform. Theory*, vol. 40, pp. 2090-2092, Nov. 1994.

[13] A. G. Dimakis, P. Godfrey, M. Wainwright and K. Ramachandran, "Network coding for distributed storage system," *IEEE Trans. on Inform. Theory*, vol. 56, no. 9, pp. 4539-4551, Sep. 2010.

[14] A. G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh, "A survey on network codes for distributed storage," in *Proc. of the IEEE*, pp. 476–489, Mar. 2011.

[15] G. Exoo and R. Jajcay, "Dynamic Cage Survey," *Electron. J. Combin.*, Dynamic survey: DS16, (2008).

[16] T. Etzion and A. Vardy, "On perfect codes and tilings: Problems and solutions," *SIAM J. Discrete Math.*, vol. 11, pp. 205-223, 1998.

[17] Z. Furedi, F. Lazebnik, A. Seress, V. A. Ustimenko, ans A. J. Woldar, "Graphs of prescribed girth and bi-degree, " *J. Combin. Theory*, Ser. B 64 (2) (1995), 228–239.

[18] Z. Furedi and M. Simonovits, "The history of degenerate (bipartite) extremal graph problems, " *in Erdős Centennial (L. Lovász et al., Eds.), Bolyai Society Mathematical Studies 25*, (2013), 169–264.

[19] C. Godsil, R. Royle, *Algebraic Graph Theory*, Springer, Graduate Texts in Mathematics, vol. 207, 2001.

[20] M. K. Gupta, A. Agrawal, and D. Yadav, "On weak dress codes for cloud storage," *arXiv:1302.3681v1*, Feb. 2013.

[21] V. Guruswami, C. Umans, and S. P. Vadhan, "Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes," *in Proc. IEEE Conf. Comput. Complex.*, pp. 96-108. 2007

[22] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Batch codes and their applications," *in Proc. 36th annual ACM symposium on Theory of computing*, vol. 36, pp. 262–271, 2004.

[23] S. Jukna, *Extremal Combinatorics*, Springer-Verlag, 2nd Ed., 2011.

[24] J. C. Koo and J. T. Gill. III, "Scalable constructions of fractional repetition codes in distributed storage systems," in *Proc. 49th Annual Allerton Conf. on Communication, Control, and Computing (Allerton)*, pp. 1366–1373, 2011.

[25] U. S. R. Murty, "A generalization of the Hoffman - Singleton graph," *Ars Combin.* 7 (1979), 191-193.

[26] O. Olmez and A. Ramamoorthy, "Repairable replication-based storage systems using resolvable designs," in *Proc. 50th Annual Allerton Conf. on Communication, Control, and Computing (Allerton)*, pp. 1174 – 1181, 2012.

[27] O. Olmez and A. Ramamoorthy, "Replication based storage systems with local repair," *International Symposium on Network Coding (NetCod)*, pp. 1-6, 2013.

[28] O. Olmez and A. Ramamoorthy, "Fractional repetition codes with flexible repair from combinatorial designs,"*arXiv:1408.5780*, Aug. 2014.

[29] L. Pamies-Juarez, H. D. L. Hollmann, and F. E. Oggier, "Locally repairable codes with multiple repair alternatives," *in IEEE Int. Symp. on Inform. Theory (ISIT 2011)*, pp. 2338-2342, Aug. 2011.

[30] S. Pawar, N. Noorshams, S. El Rouayheb, and K. Ramchandran, "Dress codes for the storage cloud: Simple randomized constructions," *in IEEE Int. Symp. on Inform. Theory (ISIT 2013)*, pp. 892 – 896, Jul. 2013.

[31] M. B. Paterson, D. R. Stinson, and R. Wei, "Combinatorial batch codes,"*Advances in Mathematics of Communications*, vol. 3, pp. 13–27, 2009.

[32] 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 Annual Allerton Conf. on Communication, Control, and Computing (Allerton)*, pp. 1243 - 1249, 2009.

[33] K. V. Rashmi, N. B. Shah, 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, pp. 1837-1852, Mar. 2012.

[34] K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Optimal exact-regenerating codes for distributed storage at the MSR and MBR point via a product-matrix construction," *IEEE Trans. Inf. Theory*, vol. 57, pp. 5227-5239, Aug. 2011.

[35] I. Reuven and Y. Beery, "Generalized Hamming weights of nonlinear codes and the relation to the Z4-linear representation," *IEEE Trans. Inform. Theory*, vol. 45, pp. 713-720, Mar. 1999.

[36] S. El Rouayheb and K. Ramchandran, "Fractional repetition codes for repair in distributed storage systems," in *Proc. 48th Annual Allerton Conf. on Communication, Control, and Computing (Allerton)*, pp. 1510 –1517, 2010.

[37] N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Explicit codes minimizing repair bandwidth for distributed storage," *in IEEE Inform. Theory Workshop (ITW)*, pp. 1–5, Jan. 2010.

[38] N. Silberstein, "Fractional Repetition and Erasure Batch Codes", to appear in *The Fourth International Castle Meeting on Coding Theory and Applications*, Palmela, Portugal, Sep. 2014, *arXiv:1405.6157*, May 2014.

[39] N. Silberstein and A. Gál, "Optimal combinatorial batch codes based on block designs", to appear in *Designs, Codes and Cryptography*, 2014. *arXiv:1312.5505*, Dec. 2013.

[40] C. Suh and K. Ramchandran, "Exact-repair MDS codes for distributed storage using interference alignment," *in IEEE Int. Symp. on Inform. Theory (ISIT 2010)*, pp. 161–165, Jul. 2010.

[41] I. Tamo, Z. Wang, and J. Bruck, "Zigzag codes: MDS array codes with optimal rebuilding," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1597–1616, Mar. 2013.

[42] C. Tian, V. Aggarwal, and V. Vaishampayan, "Exact-repair regenerating codes via layered erasure correction and block designs," *arXiv:1302.4670*, Feb. 2013.

[43] H. Weatherspoon and J. D. Kubiatowicz,"Erasure coding vs. replication: a quantitative comparison," in *Proc. 2002 Int. Workshop on Peer-to-Peer Syst.*, pp. 328–338, 2002.

[44] V. K. Wei, "Generalized Hamming weights for linear codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1412-1418, Sep. 1991.

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