entropy

MDPI

Article

Construction of New Fractional Repetition Codes from Relative Difference Sets with $\lambda = 1$

Young-Sik Kim [1,†], Hosung Park [2,*,†] and Jong-Seon No [3,†]

[1] Department of Information and Communication Engineering, Chosun University, Gwangju 61452, Korea; iamyskim@chosun.ac.kr

[2] School of Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Korea

[3] Department of Electrical and Computer Engineering, Institute of New Media and Communications, Seoul National University, Seoul 08826, Korea; jsno@snu.ac.kr

* Correspondence: hpark1@jnu.ac.kr; Tel.: +82-62-530-1791

† These authors contributed equally to this work.

Received: 12 August 2017; Accepted: 19 October 2017; Published: 22 October 2017

Abstract: Fractional repetition (FR) codes are a class of distributed storage codes that replicate and distribute information data over several nodes for easy repair, as well as efficient reconstruction. In this paper, we propose three new constructions of FR codes based on relative difference sets (RDSs) with $\lambda = 1$. Specifically, we propose new (q^2-1, q, q) FR codes using cyclic RDS with parameters $(q+1, q-1, q, 1)$ constructed from q-ary m-sequences of period q^2-1 for a prime power q, (p^2, p, p) FR codes using non-cyclic RDS with parameters $(p, p, p, 1)$ for an odd prime p or $p = 4$ and $(4^l, 2^l, 2^l)$ FR codes using non-cyclic RDS with parameters $(2^l, 2^l, 2^l, 1)$ constructed from the Galois ring for a positive integer l. They are differentiated from the existing FR codes with respect to the constructable code parameters. It turns out that the proposed FR codes are (near) optimal for some parameters in terms of the FR capacity bound. Especially, $(8,3,3)$ and $(9,3,3)$ FR codes are optimal, that is, they meet the FR capacity bound for all k. To support various code parameters, we modify the proposed (q^2-1, q, q) FR codes using decimation by a factor of the code length q^2-1, which also gives us new good FR codes.

Keywords: distributed storage systems (DSS); fractional repetition (FR) codes; FR capacity; minimum bandwidth regenerating (MBR) codes; relative difference sets (RDSs); q-ary m-sequences

1. Introduction

As users of social media services and cloud services frequently upload large data files such as images and videos, huge storage space is required, which is implemented in the form of distributed storage systems (DSSs) [1,2]. DSSs manage a tremendous amount of storage nodes and a large number of failed nodes occur every day. Traditional solutions such as simple triplication and Reed–Solomon (RS) codes are no longer enough to efficiently maintain DSSs and enhance the reliability of stored data because, for node failure-handling, we have to consider the tremendous amount of data traffic over the network in DSS, the number of disk I/O (input/output) and the availability of local data processing, as well as the redundancy of stored data. Thus, it is necessary to find a new class of node failure-handling protocols that is well-fitted for the DSS environment, and for this reason, locally repairable codes [3–7] and regenerating codes [8] have recently attracted much attention.

Regenerating codes are proposed to minimize the total bandwidth of in-network data transfer required for repairing failed nodes, as well as to minimize the amount of stored data. Assume that each node stores α symbols, and it suffices to connect to any d other nodes and download β symbols from each node to repair a failed node. It turns out [9] that there is a tradeoff between the amount of stored data α and the repair bandwidth $d\beta$, which is called the storage-bandwidth tradeoff. Minimum bandwidth regenerating (MBR) codes and minimum storage regenerating (MSR) codes are located at two extreme points of this tradeoff, respectively.

It is noted that the storage-bandwidth tradeoff was derived by allowing functional repair, where the repaired node may not have the same data as before, but it still plays the same role. The functional repair has some disadvantages against the exact repair in that all nodes may need to update the code information every time, and it is hard to maintain a systematic form of the code. Although it is proven that interior points on the tradeoff cannot be achieved with exact repair [10], many researches mainly focus on constructing exact regenerating codes and investigating the storage-bandwidth tradeoff for exact repair [11–14].

In general, DSSs handle an extremely large size of stored data, and thus, the data processing for node repairing requires a huge amount of computation. Thus, fractional repetition (FR) codes were firstly proposed in [15], which can be considered as a variant of MBR codes. FR codes enable repair-by-transfer to reduce computation in data processing for node repairing by relaxing the requirement of connecting to any d nodes. Instead, a repair is done by connecting to some d nodes based on a table. FR coding means that data symbols are replicated a few times and divided into several groups, each of which is stored at each node. Since no operations are used except for replication, computation for data processing becomes very small when repairing a failed node or collecting data symbol.

In designing DSSs, the system parameters can take arbitrary values based on the system environment, and this means many kinds of FR codes with various parameters need to be constructed. In [16], the existence of FR codes for each parameter set is shown by algorithm-based search, and some examples for each parameter set are provided. Many kinds of FR codes have been constructed mostly based on algebraic structures, combinatorial designs and graph theory [15,17–22]. In [15], Steiner systems are used as a class of $(v,(v-1)/(k-1),k)$ FR codes with $\beta = 1$ under the existence of Steiner system $S(2,k,v)$. In [17], $((q^{n+2}-1)/(q-1),(q^{n+1}-1)/(q-1),q+1)$ FR codes are constructed for a prime power q and $n \geq 1$, which are based on the projective geometry and Latin squares. These FR codes are a subclass of the FR codes from Steiner systems in [15] and additionally designed to have a scalable property. In [18], various constructions of resolvable FR codes, especially net FR codes, are proposed by using grids, affine resolvable designs, Hadamard designs and mutually orthogonal Latin squares. The corresponding constructable parameters are $(2a,a,2)$ with $\beta = 1$ from grids, $(q\rho,q^{m-1},1 \leq \rho \leq (q^m-1)/(q-1))$ with $\beta = q^{m-2}$ from affine resolvable designs, $(8a-2,2a,4a-1)$ with $\beta = a$ from Hadamard designs (where $4a-1 \geq 7$ is an odd prime power) and $(\rho p^m, p^m, \rho \leq p^m - 1)$ or $(4a,a,4)$ with $\beta = 1$ from mutually orthogonal Latin squares (for a prime p, positive integers m and an integer $a \neq 2,6$). Lastly, in [21], $(\rho\alpha,\alpha,\rho)$ FR codes with $\beta = 1$ are constructed from transversal designs for $3 \leq \rho \leq \alpha + 1$. Furthermore, $((s+1)(st+1),t+1,s+1)$ FR codes with $\beta = 1$ are constructed from generalized quadrangles for $2 \leq s \leq t$. These two classes of FR codes are optimal for selected parameters satisfying the conditions in [21].

The FR codes constructed from Steiner systems [15] support the number of data symbols equal to or slightly larger than the MBR capacity in (1). This means that the file size of the FR codes from Steiner systems shows a considerable gap from the upper bound of the FR capacity because any two rows of the incidence matrix of a Steiner system always have a collision [21]. On the other hand, in the incidence matrices of the proposed FR codes, there is no collision for some pairs of rows based on our analysis, and this property can make the proposed FR codes closer to optimal with respect to the FR capacity bound.

In this paper, we first propose new three constructions of (q^2-1,q,q), (p^2,p,p) and $(4^l,2^l,2^l)$ FR codes based on relative difference sets with parameters $(q+1,q-1,q,1)$, $(p,p,p,1)$ and $(2^l,2^l,2^l,1)$, respectively, where q is a prime power, p is an odd prime or $p=4$ and l is a positive integer. Especially, $(8,3,3)$ and $(9,3,3)$ FR codes are optimal, that is, they meet the FR capacity bound in [15] for all k. We show via theoretical derivations and numerical analysis that the proposed FR codes are (near) optimal for some parameters in terms of the FR capacity bound. Finally, for various numbers of nodes, we modify the proposed (q^2-1, q, q) FR codes using decimation by a factor of the code length q^2-1, which also gives us new good FR codes. It is noted that some examples are already shown for small-valued parameter sets in [16], but we propose a systematic construction method of FR codes for the whole of the parameter sets based on well-organized mathematical structures unlike the algorithm-based search in [16]. It is also noted that the proposed FR codes can be seen as a part of the general class of FR code constructed from group divisible design (GDD) [19,20]. However, the contribution of this work is to explicitly provide three construction methods directly based on relative difference sets and m-sequences, propose a modification method from the three constructions to generate a new class of FR codes and analyze them in detail, while the basic idea and general framework are given in [19,20].

This paper is organized as follows. In Section 2, the basic definitions and notations are presented. The proposed FR codes are presented in Section 3. Additionally, we provide the characteristics of the proposed scheme and numerical results. In Section 4, we propose a modification method for Construction 1. Finally, we conclude this paper in Section 5.

2. Preliminaries

2.1. Regenerating Codes and Fractional Repetition Codes

In this paper, we assume that every node stores the same amount of data symbols. To clearly define regenerating codes and FR codes, we follow the notations in [21].

An $(n, k, d, M, \alpha, \beta)_q$ regenerating code for $k \leq d \leq n-1$ and $\beta \leq \alpha$ is defined as follows. The number of nodes and the number of information data symbols that need to be stored in DSS are denoted by n and M, respectively, and the symbols are in the finite field $\mathbb{F}_q$ of q elements. Each node stores α symbols. The parameter k is called the reconstruction degree, which means that a data collector can reconstruct all the stored information data by connecting to any k nodes and downloading α symbols from each node. A failed node is repaired by connecting to any d other nodes and downloading β symbols from each node. With this notation, the repair bandwidth becomes $d\beta$. We assume $\beta = 1$ for the code construction throughout paper as in [11], which can be simply expanded to the case of $\beta > 1$. It is noted that this expansion does not cover all the FR codes for $\beta > 1$ [18]. For MBR codes with $\beta = 1$, d is equal to α, and the number of data symbols to be stored in DSS is given as:

$$M = k\alpha - \binom{k}{2}, \tag{1}$$

which is called the MBR capacity.

In [18], the β-recoverability and FR codes are formally defined as follows.

Definition 1. *Let $\Omega = [\theta]$ and N_i, $i = 1, \ldots, d$ be subsets of Ω. Let $N = \{N_1, \ldots, N_d\}$, and consider $A \subset \Omega$ with $|A| = d\beta$. We say that A is β-recoverable from N if there exist $B_i \subseteq N_i$ for each $i = 1, \ldots, d$ such that $B_i \subset A$, $|B_i| = \beta$ and $\cup_{i=1}^{d} B_i = A$ [18].*

Definition 2. *(FR codes [18]) An (n, α, ρ) FR code $C = (\Omega, N)$ with repetition degree ρ and normalized repair bandwidth $\beta = \alpha/d$ (α and β are positive integers) is a set of n subsets $N = \{N_1, \ldots, N_n\}$ of a symbol set $\Omega = [\theta]$ with the following properties.*

1. *The cardinality of each N_i is α.*
2. *Each element of Ω is contained in exactly ρ sets in N.*
3. *Let N^{surv} denote any $(n-\tau)$-sized subset of N and $N^{fail} = N \backslash N^{surv}$. Each $N_j \in N^{fail}$ is β-recoverable from some d-sized subset of N^{surv}. Let ρ_{res} be the maximum value of τ such that this property holds.*

Note that an (n, α, ρ) FR code $\mathcal{C}$ satisfies $n\alpha = \rho\theta$. The parameter ρ is called the repetition degree of $\mathcal{C}$. The incidence matrix of $\mathcal{C}$, denoted by $I(\mathcal{C})$, is defined by the $n \times \theta$ binary matrix whose (i, j) element is one if the set N_i includes data symbol j or zero, otherwise. It is noted that the row and column weights of $I(\mathcal{C})$ are α and ρ, respectively. An FR code can be used as an inner code together with an outer (θ, M) maximum distance separable (MDS) code. This concatenated code is called distributed replication-based exact simple storage (DRESS) code with parameters $[(\theta, M), k, (n, \alpha, \rho)]$ [23]. According to an FR code $\mathcal{C}$, node i, $i = 1, \ldots, n$, stores the data symbols in N_i. Figure 1 illustrates an example of a $[(\theta, M), k, (n, \alpha, \rho)]$ DRESS code.

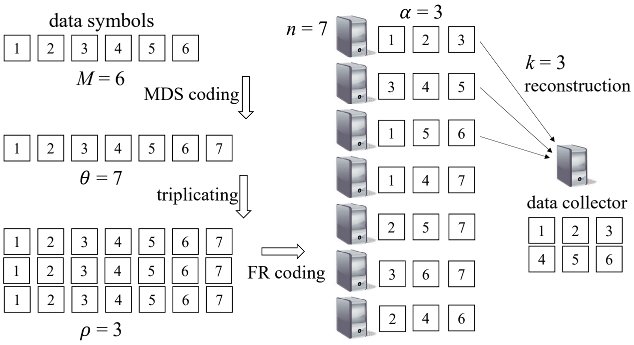


Figure 1. Structure of a $[(7, 6), 3, (7, 3, 3)]$ distributed replication-based exact simple storage (DRESS) code.

For a given FR code, the maximum number of information data symbols M to be stored in DSS is determined as a function of k as:

$$M(k) = \min_{\substack{I \subset \{1,\ldots,n\} \\ |I|=k}} |\cup_{i \in I} N_i|. \tag{2}$$

It is addressed in [15] that for a well-constructed FR code, $M(k)$ can be larger than the MBR capacity, that is,

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

holds for FR codes. For given parameters (n, k, α, ρ), the FR capacity, denoted by $A(n, k, \alpha, \rho)$, is defined as the maximum value of $M(k)$ among all FR codes with the parameters.

An upper bound for FR capacity was derived in [15] as:

$$A(n, k, \alpha, \rho) \leq \phi(k)$$

where:

$$\phi(1) = \alpha, \ \phi(k+1) = \phi(k) + \alpha - \left\lceil \frac{\rho\phi(k) - k\alpha}{n - k} \right\rceil. \tag{3}$$

Capacity-achieving FR codes were constructed for some parameters in [21], but the FR capacity is unknown in general. An FR code is called k-optimal if $M(k) = A(n, k, \alpha, \rho)$. Furthermore, an FR code is called optimal if for any $k \leq \alpha$, it is k-optimal [21].

2.2. Relative Difference Sets and q-Ary m-Sequences

Let G be a group of order uv under an operation $*$, and let N be a normal subgroup of order u. Then, a (v, u, w, λ) relative difference set (RDS) in G relative to N is defined as a subset D with w elements of the group G such that the multiset of $w(w-1)$ elements given by $\{d_1 * d_2^{-1} \mid d_1, d_2 \in D$ such that $d_1 \neq d_2\}$ contains every element of $G \backslash N$ exactly λ times and no element in N [24,25]. The parameters of RDSs satisfy the following equation:

$$w(w-1) = u(v-1)\lambda.$$

If G is a cyclic group, D is called a cyclic RDS. If $u = 1$, D becomes a (v, w, λ) difference set. That is, a (v, w, λ) difference set is defined as a subset D with w elements of a group G with v elements such that the set of $w(w-1)$ elements given by $\{d_1 * d_2^{-1} \mid d_1, d_2 \in D \text{ such that } d_1 \neq d_2\}$ contains every element of $G \backslash \{0\}$ exactly λ times.

Let D be an RDS with parameters (v, u, w, λ) given as $D = \{d_0, \cdots, d_{w-1}\}$, where $d_0, \cdots, d_{w-1}$ are elements in G. It is easy to check the following lemmas. Even though it is not difficult to derive it, we provide proofs for clear understanding.

Lemma 1. *If D is an RDS, then its right coset $D * b$ is also an RDS for $b \in G$.*

Proof. Let $D * b = \{d_0 * b, \cdots, d_{w-1} * b\}$. By the definition, for $d_i, d_j \in D$, $i \neq j$, $d_i * d_j^{-1} = d_i * (b * b^{-1}) * d_j^{-1} = (d_i * b) * (d_j * b)^{-1}$ becomes all elements in $G \backslash N$ exactly λ times, when i and j vary for all elements in D. □

Lemma 2. *For cosets $D * b_1$ and $D * b_2$ of an RDS D, we have $|D * b_1 \cap D * b_2| = \lambda$ if $b_1^{-1} * b_2$ is not in N. Otherwise, $|D * b_1 \cap D * b_2| = 0$.*

Proof. Common elements of $D * b_1$ and $D * b_2$ are given as $d_i * b_1 = d_j * b_2$ for some i, j. By the definition of RDS, $d_j^{-1} * d_i = b_2 * b_1^{-1}$ covers all elements in $G \backslash N$. □

Let F_q denote the finite field with $q = p^e$ elements, where p is a prime and e is a positive integer. The trace function from F_{q^n} into F_q is defined as:

$$\mathrm{tr}_1^n(x) = \sum_{i=0}^{n-1} x^{q^i}. \tag{4}$$

Then, a q-ary m-sequence $s(t)$ of period $q^n - 1$ is defined as:

$$s(t) = \mathrm{tr}_1^n(\xi^t), 0 \leq t < q^n - 1 \tag{5}$$

where ξ denotes a primitive element of F_{q^n}.

It is well-known that the q-ary m-sequence has the balance property, which means that zero appears $q^{n-1} - 1$ times and each of the non-zero element in F_q appears q^{n-1} times in a period.

The δ-homogeneous function from F_{q^n} to F_q is introduced by Klapper [26], which is defined as $H(xy) = y^\delta H(x)$ for any $x \in F_{q^n}$ and $y \in F_q$. Then, Kim et al. constructed an RDS from a δ-homogeneous function on $F_{q^n}^*$ [27]. In addition, a function $f(x)$ is said to be difference-balanced if the difference function $f(xz) - f(x)$ is balanced for any $z \in F_{q^n} \backslash \{0, 1\}$.

Theorem 1 ([27])**.** *Let q be a prime power and n a positive integer. If $f(x)$ is a δ-homogeneous function on $F_{q^n}^*$ over F_q with difference-balanced property, where δ is relatively prime to $q^n - 1$, then the set $D_f = \{x \mid f(x) = \zeta, x \in F_{q^n}^*, \text{for a given } \zeta \in F_q^*\}$ is a cyclic RDS with parameters $(\frac{q^n-1}{q-1}, q-1, q^{n-1}, q^{n-2})$ in the multiplicative group $F_{q^n}^*$ relative to its normal subgroup F_q^*.*

It is clear that for $n = 2$, the trace function defined in (4) is a one-homogeneous function because we have $\mathrm{tr}_1^2(xy) = y\mathrm{tr}_1^2(x)$ for $y \in F_q$ and $x \in F_{q^2}$. Therefore, from a q-ary m-sequence of period $q^2 - 1$, we can construct a cyclic RDS with parameters $(q+1, q-1, q, 1)$.

Example 1. *For $p = 3$, let α be a primitive element of F_{3^2}. Then, a relative difference set with parameters $(4, 2, 3, 1)$ is given as $D = \{1, \alpha, \alpha^3\}$. It is easy to check that $d_1 * d_2^{-1}$ for any two distinct elements $d_1, d_2 \in D$ can cover only once all elements in $F_{3^2} \backslash F_3$.*

3. The Proposed FR Codes

In this section, we construct new classes of FR codes with parameters (q^2-1, q, q), (p^2, p, p) and $(4^l, 2^l, 2^l)$ based on the cyclic RDS with parameters $(q+1, q-1, q, 1)$ constructed from the q-ary m-sequences, the non-cyclic RDS with parameters $(p, p, p, 1)$ and the non-cyclic RDS with parameters $(2^l, 2^l, 2^l, 1)$, respectively, for a prime power q, an odd prime p or $p = 4$ and a positive integer l. Furthermore, it is demonstrated that these codes are (near) optimal for some parameters with respect to the FR capacity bound in (3).

3.1. Construction of FR Codes from RDSs

First, we propose a general construction method of FR codes based on the RDSs with parameters $(v, u, w, 1)$. Since the relative difference sets with $\lambda = 1$ are equivalent to a class of group divisible design (GDD), the proposed FR codes can be seen to originate from the common framework, which exploits the incidence matrix of a combinatorial design to construct an FR code. Particularly, the incidence matrix of a balanced incomplete block design (BIBD) or a GDD can be directly used as the incidence matrix of the corresponding FR code. Thus, the incidence matrix of an FR code can be directly constructed from a given RDS, and the general construction method is a result of this process.

Let $G = \{g_0, g_1, \ldots, g_{uv-1}\}$ be a group of order uv under an operation $*$, and let N be a normal subgroup of order u. Furthermore, let D be an RDS with parameters $(v, u, w, 1)$ in G relative to N. Then, an incidence matrix of FR code is constructed from D as:

$$I(\mathcal{C}) = \left[c_{i,j}\right]_{0 \le i,j < uv} \tag{6}$$

whose (i, j) element is given as:

$$c_{i,j} = \begin{cases} 1, & \text{if } g_j \in D * g_i \\ 0, & \text{otherwise.} \end{cases}$$

This is our basic method to construct three classes of FR codes from RDSs, which will be explicitly given as follows. It is noted that the GDDs corresponding to the proposed FR codes are symmetric, and more general parameter sets of the symmetric GDDs are found in [28]. Suppose that the i-th row of $I(\mathcal{C})$ is denoted by $c_i = (c_{i,0}, \cdots, c_{i,uv-1})$ for $0 \le i < uv$. For any pair of rows c_{i_1} and c_{i_2} in $I(\mathcal{C})$, we will say that there is a collision if $c_{i_1,j} = c_{i_2,j} = 1$ for some j, $0 \le j < uv$.

Construction 1. *Define a map $\mu_\zeta(\cdot)$ from F_q to F_2 for $\zeta \in F_q^*$ as:*

$$\mu_\zeta(l) = \begin{cases} 1, & l = \zeta \\ 0, & \textit{otherwise.} \end{cases} \tag{7}$$

Furthermore, we define a binary sequence $b_\zeta(t) = \mu_\zeta(s(t))$ for the q-ary m-sequence $s(t)$ of period $q^2 - 1$. Then, a cyclic RDS with parameters $(q+1, q-1, q, 1)$ in $Z_{q^2-1} = \{0, 1, \ldots, q^2-2\}$ relative to $N = \{0, 1(q+1), \ldots, (q-2)(q+1)\}$ is obtained as $D = \{t | b_\zeta(t) = 1,\ 0 \le t < q^2 - 1\}$. According to the general construction method, we propose new (q^2-1, q, q) FR codes $\mathcal{C}$ whose incidence matrices $I(\mathcal{C})$ have the form:

$$I(\mathcal{C}) = \left[c_{i,j}\right]_{0 \le i,j < q^2-1} \tag{8}$$

where:

$$c_{i,j} = \begin{cases} 1, & \text{if } j \in D + i \\ 0, & \text{otherwise,} \end{cases}$$

and $D + i = \{t + i \mid t \in D\}$ for $0 \le i < q^2 - 1$.

Note that the cyclic RDS with parameters $(q+1, q-1, q, 1)$ in Construction 1 is equivalent to the RDS with the same parameters from Theorem 1. The first row corresponds to the binary sequence μ-mapped from the q-ary m-sequence of period q^2-1 in (5). We call this sequence the characteristic sequence of the corresponding RDS. Each of the other rows is cyclically shifted to the right by one position from the row above it. The size of $I(\mathcal{C})$ is $(q^2-1) \times (q^2-1)$, which means that the length of the FR code $\mathcal{C}$ is $n = q^2-1$ and the parameter θ is also q^2-1. It is easily shown from the balance property of the q-ary m-sequence that the row and column weights are q, that is, $\alpha = \rho = q$.

Example 2. *A ternary m-sequence $s(t)$ of period eight is generated as*

$$(s(0)s(1)s(2)\cdots s(7)) = (11012202).$$

For $\zeta = 1$, we have an RDS with parameters $(4,2,3,1)$ as $D = \{0,1,3\}$, and its characteristic sequence is given as:

$$(b_1(0)b_1(1)b_1(2)\cdots b_1(7)) = (11010000).$$

Then, the incidence matrix of the $(8,3,3)$ FR code $\mathcal{C}$ is given as:

$$I(\mathcal{C}) = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{9}$$

We can see that the row and column weights are $q = 3$.

Example 3. *A 5-ary m-sequence of period 24 is generated as:*

$$(s(0)s(1)\cdots s(23)) = (131033212011424022343044).$$

For $\zeta = 1$, we have an RDS with parameters $(6,4,5,1)$ as $D = \{0,2,7,10,11\}$, and its characteristic sequence is given as:

$$(b_1(0)b_1(1)\cdots b_1(23)) = (101000010011000000000000).$$

In the same way as the ternary case, the 24×24 incidence matrix of the corresponding $(24,5,5)$ FR code is constructed using cyclic shift, and its row and column weights are $q = 5$.

Example 4. *A 4-ary m-sequence of period 15 is generated as:*

$$(s(0)\cdots s(14)) = (10\gamma\gamma 1\gamma 0\gamma^2\gamma^2\gamma\gamma^2 011\gamma^2)$$

where γ is a primitive element of F_4 with the primitive polynomial x^2+x+1 and F_{16} is an extended field of F_4 by $x^2+x+\gamma$. For $\zeta = 1$, we have:

$$(b_1(0)\cdots b_1(14)) = (100010000000110).$$

Then, the incidence matrix of the $(15, 4, 4)$ *FR code* $\mathcal{C}$ *is given as:*

$$I(\mathcal{C}) = \begin{bmatrix} 100010000000110 \\ 010001000000011 \\ \vdots \\ 000100000001101 \end{bmatrix} \tag{10}$$

where the row and column weights are $q = 4$.

Remark 1. *In [18], the resolvable FR code is defined as follows. Let* $\mathcal{C} = (\mathbf{\Omega}, V)$ *where* $V = \{V_1, \ldots, V_n\}$ *is an FR code. A subset* $P \subset V$ *is said to be a parallel class if for* $V_i \in P$ *and* $V_j \in P$ *with* $i \neq j$*, we have* $V_i \cap V_j = \emptyset$ *and* $\cup_{\{j:V_j \in P\}} V_j = \mathbf{\Omega}$*. A partition of* V *into* r *parallel classes is a resolution. If there exists at least one resolution, then the code is called a resolvable FR code. It is clear that the FR codes from the first construction are not resolvable because* q *is not a factor of* $q^2 - 1$*, while the FR codes from the net in [18] are always resolvable. This is evidence that the proposed construction is not a proper subset of the constructions in [18].*

In 1966, Elliott and Butson [25] constructed non-cyclic RDSs with parameters (p^n, p, p^n, p^{n-1}) and $(4, 4, 4, 1)$, where p is an odd prime. Using those RDSs, we can construct FR codes with parameters (p^2, p, p) and $(16, 4, 4)$, respectively.

In the following description, the symbol $\oplus$ means the direct sum, and Z_p denotes the additive group of integers modulo p. Let G_n be the elementary Abelian p-group of order p^n with identity zero, whose elements are expressed as n-tuples of elements of Z_p. Then, the RDS can be constructed as in the following theorem.

Theorem 2 ([25])**.** *Let* $G = Z_p \oplus G_n$*, and let* $N = Z_p \oplus \{\mathbf{0}\}$*. For* $a_i \not\equiv 0 \pmod{p}$*,* $i = 1, \cdots, n$*,*

$$D = \{(f(m), m) \mid m = (m_1, m_2, \cdots, m_n) \in G_n\}$$

is an RDS with parameters (p^n, p, p^n, p^{n-1}) *of* G *relative to* N*, where* $f(m) = \sum_{i=1}^{n} a_i m_i^2 \pmod{p}$*.*

Since we are interested in the case of $\lambda = 1$, we only use the RDSs in Theorem 2 for the case of $n = 1$. In this case, we can simplify $f(m)$ as $f(m) = a_1 m^2 \pmod{p}$ for $m \in Z_p$, and we have the RDS D with parameters $(p, p, p, 1)$, whose elements are given as $(f(m), m)$.

Note that there is no explicit ordering among elements in the RDS D. However, for the construction of FR codes, we will use the arbitrary order of elements in D, for example, a lexicographic order i defined as $i = f(m) \times p + m$ for $m \in Z_p$, which is in the range between zero and $p^2 - 1$. Thus, construction of new FR codes by using the RDSs in Theorem 2 is given as follows.

Construction 2. *Let* D *be a non-cyclic RDS with parameters* $(p, p, p, 1)$*, where* p *is an odd prime or four. For* $(i_1, i_2), (j_1, j_2) \in G$*, we have lexicographic orders* $i = p \times i_1 + i_2$ *and* $j = p \times j_1 + j_2$*. According to the general construction method, we propose new* (p^2, p, p) *FR codes* $\mathcal{C}$ *whose incidence matrices* $I(\mathcal{C})$ *have the form given by:*

$$I(\mathcal{C}) = \left[c_{i,j} \right]_{0 \le i,j < p^2} \tag{11}$$

whose (i, j) *element is given as:*

$$c_{i,j} = \begin{cases} 1, & \text{if } (j_1, j_2) \in D + (i_1, i_2) \\ 0, & \text{otherwise} \end{cases}$$

where $D + (i_1, i_2) = \{(d_1 + i_1, d_2 + i_2) \mid (d_1, d_2) \in D\}$*.*

In the following examples, we provide incidence matrices for FR codes with parameters $(9, 3, 3)$, $(16, 4, 4)$ and $(25, 5, 5)$.

Example 5. *Let $p = 3$, $n = 1$ and $a_1 = 1$. Then, we have $Z_3 = \{0, 1, 2\}$ and $G = Z_3 \oplus Z_3$. The RDS D with parameters $(3, 3, 3, 1)$ is given as $D = \{(f(m), m) \mid m \in \{0, 1, 2\}\}$, where $f(m) = m^2 \pmod 3$. That is, we have $D = \{(0, 0), (1, 1), (1, 2)\}$. Then, we have the following cosets of D for all elements in $G = Z_3 \oplus Z_3$ as:*

$$\begin{aligned}
D + (0,0) &= \{(0,0), (1,1), (1,2)\} \\
D + (0,1) &= \{(0,1), (1,2), (1,0)\} \\
D + (0,2) &= \{(0,2), (1,0), (1,1)\} \\
D + (1,0) &= \{(1,0), (2,1), (2,2)\} \\
D + (1,1) &= \{(1,1), (2,2), (2,0)\} \\
D + (1,2) &= \{(1,2), (2,0), (2,1)\} \\
D + (2,0) &= \{(2,0), (0,1), (0,2)\} \\
D + (2,1) &= \{(2,1), (0,2), (0,0)\} \\
D + (2,2) &= \{(2,2), (0,0), (0,1)\}.
\end{aligned}$$

By using a lexicographic order, we can obtain the following incidence matrix for a $(9, 3, 3)$ FR code as:

$$I(\mathcal{C}) = \begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}.$$

Note that the proposed $(9, 3, 3)$ FR code meets the FR capacity bound in (3), that is, it is an optimal FR code with parameters $(9, 3, 3)$.

Example 6. *An RDS with parameters $(4, 4, 4, 1)$ can be obtained as $D = \{(0, 0), (0, 1), (1, 3), (3, 0)\}$ of $G = Z_4 \oplus Z_4$ and the forbidden normal subgroup $2Z_4 \oplus 2Z_4$ [29]. By using a lexicographic order, we can similarly obtain the following incidence matrix for a $(16, 4, 4)$ FR code as:*

$$I(\mathcal{C}) = \begin{bmatrix}
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}.$$

Example 7. *Let $p = 5$, $n = 1$ and $a_1 = 1$. Then, we have $Z_5 = G_1 = \{0, 1, 2, 3, 4\}$. Then, the RDS D with parameters $(5, 5, 5, 1)$ is given as $D = \{(0, 0), (1, 1), (4, 2), (4, 3), (1, 4)\}$, which corresponds to the first row of the proposed incidence matrix when we use a lexicographic order. The remaining rows can be obtained by adding all elements in $G = Z_5 \oplus Z_5$ just as in the previous examples. Then, a $(25, 5, 5)$ FR code is obtained by using a lexicographic order.*

In the third construction, we utilize the Galois ring $GR(4, l)$, which is a local ring with maximum ideal $2GR(4, l)$. Hou and Sehgal proposed a semi-regular RDS from the Galois ring [30], where the semi-regular RDSs satisfy $k = \lambda v$. It is known that $GR(4, l)^*$, the group of units, contains a unique cyclic subgroup T^* of order $2^l - 1$, and $T = T^* \cup \{0\}$ is called the Teichmuller set of $GR(4, l)$. Then, each element $a \in GR(4, l)$ has a unique two-adic representation $a = x_0 + 2x_1$, where $x_0, x_1 \in T$. The Frobenius map $\sigma : GR(4, l) \to GR(4, l) : x_0 + 2x_1 \to x_0^2 + 2x_1^2 (x_0, x_1 \in T)$ is an automorphism of $GR(4, l)$ of order l. The relative trace of $GR(4, l)$ is the map $Tr : GR(4, l) \to Z_4$ defined by $Tr(a) = \sum_{i=0}^{l-1} \sigma^i(a)$. Let W be a finite Abelian group and $h : W \to T$ any function with $|W| = r$. Let $G = GR(4, l) \times W$ and:

$$D = \cup_{w \in W}((1 + 2h(w))T, w) \subset G.$$

Then, R is a semi-regular RDS in G relative to $N = 2GR(4, l) \times \{0\}$. It should be noted that the cardinality of W is related to λ. Thus, to construct RDS with $\lambda = 1$, we only consider the case of $G = GR(4, l)$ and $N = 2GR(4, l)$. Then,

$$D = (1 + 2h)T \tag{12}$$

is a non-cyclic RDS with parameters $(2^l, 2^l, 2^l, 1)$, where h is an element in T and l is a positive integer.

To assign the proper order to each element in $GR(4, l)$, we denote $T = \{0, 1, \beta, \beta^2, \cdots, \beta^{2^l-2}\} = \{T_0, T_1, T_2, \cdots, T_{2^l-1}\}$, where each element in each set has component-wise correspondence. Then, an element of $GR(4, l)$ is indexed as $a_{i \times 2^l + j} = T_j + 2T_i$ for $T_i, T_j \in T$.

Construction 3. *Let D be a non-cyclic RDS with parameters $(2^l, 2^l, 2^l, 1)$, where l is a positive integer. According to the general construction method, we propose new $(4^l, 2^l, 2^l)$ FR codes C whose incidence matrices $I(C)$ have the form given by:*

$$I(C) = \left[c_{i,j}\right]_{0 \le i,j < 4^l r} \tag{13}$$

where:

$$c_{i,j} = \begin{cases} 1, & \text{if } a_j \in D + a_i \\ 0, & \text{otherwise} \end{cases}$$

and $D + a_i = \{a_t + a_i \mid a_t \in D\}$.

In the following examples, we provide the incidence matrix for FR codes with parameters (16,4,4).

Example 8. *For $GR(4, 2)$, the Teichmuller set is given as:*

$$T = \{0, 1, \beta, \beta^2\}$$

where β is a root of $g(x) = x^2 + x + 1$. Then, all elements of $GR(4, 2)$ are given as:

$$\begin{aligned} GR(4, 2) &= \{0, 1, \beta, \beta^2, 2, 3, \beta + 2, \beta^2 + 2, 2\beta, 1 + 2\beta, 3\beta, \beta^2 + 2\beta, 2\beta^2, 1 + 2\beta^2, \beta + 2\beta^2, 3\beta^2\} \\ &= \{a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}\} \end{aligned}$$

and $2GR(4, 2) = \{0, 2, 2\beta, 2\beta^2\}$.

Therefore, $|N| = |2GR(4,2)| = 4$ *and* $|G| = vu = 4^2$*, i.e.,* $v = 2^2$*. Since it is a semi-regular RDS (*$k = \lambda v$*), we have* $k = v = 2^2 r$ *for the case of* $\lambda = 1$*. Because* h *in (12) is an arbitrary element in* T*, we set* $h = \beta$*. Thus, a semi-regular RDS* D *is given as* $D = (1+2\beta)T = \{0, 1+2\beta, \beta+2\beta^2, \beta^2+2\}$*, which is a* $(4,4,4,1)$ *RDS. We can apply the proposed ordering to construct a* $(16,4,4)$ *FR code as follows:*

$$I(\mathcal{C}) = \begin{bmatrix} 1&0&0&0&0&0&0&1&0&1&0&0&0&0&1&0\\ 0&1&0&1&0&0&0&0&0&0&0&0&1&0&1&0\\ 0&1&1&0&1&0&0&1&0&0&0&0&0&0&0&0\\ 0&0&1&1&0&0&0&0&1&1&0&0&0&0&0&0\\ 0&0&0&1&1&0&0&0&0&0&1&0&0&1&0&0\\ 0&0&0&0&0&1&0&1&1&0&1&0&0&0&0&0\\ 1&0&0&1&0&1&1&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&1&0&0&0&0&1&1&0&0\\ 0&1&0&0&0&0&1&0&1&0&0&0&0&0&0&1\\ 0&0&0&0&1&0&1&0&0&1&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&1&1&0&1&0&0&1\\ 1&1&0&0&0&0&0&0&0&0&1&1&0&0&0&0\\ 0&0&1&0&0&1&0&0&0&0&0&1&1&0&0&0\\ 1&0&1&0&0&0&0&0&0&0&0&0&0&1&0&1\\ 0&0&0&0&0&0&0&0&1&0&0&1&0&1&1&0\\ 0&0&0&0&1&1&0&0&0&0&0&0&0&0&1&1 \end{bmatrix}.$$

3.2. Properties of the Proposed FR Codes

In this subsection, we will investigate how good the number of data symbols $M(k)$ to be stored in DSS is in terms of the upper bound of the FR capacity in (3) for the proposed FR codes. To this end, some properties of the incidence matrices of the proposed FR codes are given in the following theorems. From Lemma 2, it is easy to prove the following property of (q^2-1, q, q) FR codes from Construction 1.

Theorem 3. *Let* c_{i_1} *and* c_{i_2} *be the* i_1*-th and* i_2*-th rows in* $I(\mathcal{C})$ *of the proposed FR codes from Construction 1, where* $i_1 \neq i_2$ *and* $0 \leq i_1, i_2 < q^2 - 1$*. Then, the inner product of* c_{i_1} *and* c_{i_2} *is given as:*

$$c_{i_1} \cdot c_{i_2} = \begin{cases} 0, & \textit{if } i_2 - i_1 \equiv 0 \pmod{q+1} \\ 1, & \textit{otherwise.} \end{cases}$$

Proof. Let $G = Z_{q^2-1}$ and $N = \{0, 1(q+1), \ldots, (q-2)(q+1)\}$ denote the corresponding cyclic group and the normal subgroup, respectively. From Lemma 2, since $|(D+i_1) \cap (D+i_2)| = 0$ for $i_2 - i_1 \equiv 0 \pmod{q+1}$, two rows c_{i_1} and c_{i_2} are orthogonal to each other. Otherwise, from $|Dg^{i_1} \cap Dg^{i_2}| = 1$, we have $c_{i_1} \cdot c_{i_2} = 1$. □

That is, for any pair of two rows in $I(\mathcal{C})$, there is no collision or only one collision.

Similarly, we have the property of FR codes from Construction 2 as follows.

Theorem 4. *Let* c_i *and* c_j *be the* i*-th and* j*-th rows in* $I(\mathcal{C})$ *of the proposed FR codes from Construction 2, where* i *and* j *are the lexicographic orders of* (i_1, i_2) *and* (j_1, j_2)*, respectively, for* $i \neq j$*,* $0 \leq i, j < p - 1$*. Then, the inner product of* c_i *and* c_j *is given as:*

$$c_i \cdot c_j = \begin{cases} 0, & \textit{if } i_2 = j_2 \\ 1, & \textit{otherwise.} \end{cases}$$

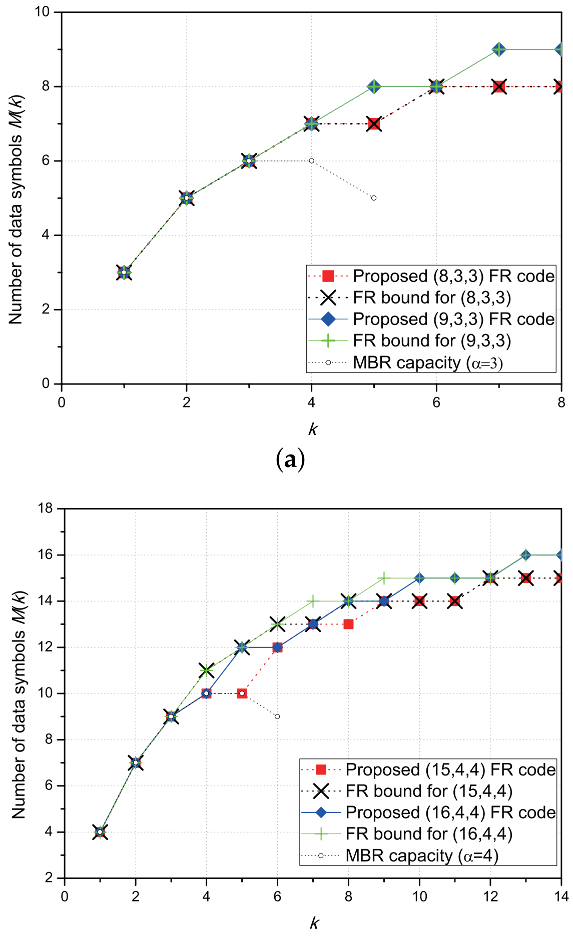
Proof. From Lemma 2, since $|[D + (i_1, i_2)] \cap [D + (j_1, j_2)]| = 0$ for $i_2 - j_2 = 0$, two rows c_i and c_j are orthogonal to each other. Otherwise, from $|[D + (i_1, i_2)] \cap [D + (j_1, j_2)]| = 1$, we have $c_i \cdot c_j = 1$. □

That is, for any pair of two rows in $I(\mathcal{C})$, there is no collision or only one collision. Actually, $M(k)$ of the proposed FR codes is very close to the FR capacity bound and greater than or equal to the MBR bound, which are shown via the following theorem and numerical analysis as below.

Theorem 5. *The number of data symbols to be stored in the proposed FR codes satisfies* $M(k) \geq kq - \binom{k}{2}$.

Proof. The MBR capacity $M(k) = kq - \binom{k}{2}$ is found when every pair of rows among k rows chosen from the incidence matrix of an FR code has exactly one collision. Equation (2) says that $M(k)$ can be strictly larger than the MBR capacity if some pairs of rows have no collision and the other pairs have one collision. The incidence matrices of the proposed FR codes satisfy the above condition for a given k, and thus, $M(k)$ of the proposed FR codes is larger than or equal to the MBR capacity. □

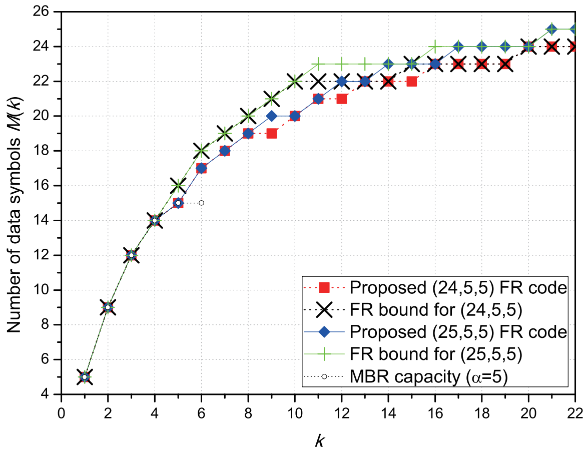
Figure 2 demonstrates the number of data symbols $M(k)$ of the proposed FR codes with parameters $(8,3,3)$, $(9,3,3)$, $(15,4,4)$, $(16,4,4)$, $(24,5,5)$ and $(25,5,5)$. We cannot find any existing FR code whose parameters are the same as the proposed ones, and thus, we plot the MBR capacity and the FR capacity bound $\phi(k)$ in (3). We can see that the $(8,3,3)$ and $(9,3,3)$ FR codes are optimal because they achieve the FR capacity bound for all k. The $(15,4,4)$, $(16,4,4)$, $(24,5,5)$ and $(25,5,5)$ FR codes do not exactly achieve the FR capacity bound, but the gaps become smaller; thus, it deserves to be called FR capacity-approaching.



(a)

(b)

Figure 2. Cont.



(**c**)

Figure 2. Comparison of the number of data symbols to be stored in the proposed fractional repetition (FR) codes with the minimum bandwidth regenerating (MBR) capacity and the FR capacity bound. (**a**) $(8,3,3)$ and $(9,3,3)$ FR codes; (**b**) $(15,4,4)$ and $(16,4,4)$ FR codes; (**c**) $(24,5,5)$ and $(25,5,5)$ FR codes.

4. Modification of the Proposed FR Codes

In this section, we propose a modification method for the proposed FR codes from Construction 1 to support various numbers of storage nodes. It is trivial to obtain irregular incidence matrices from the regular ones by selecting some of the rows in the matrices, where "regular" means that all the rows in the incidence matrix have a constant Hamming weight and so do all the columns; otherwise, it is called "irregular". Thus, we propose a method to obtain regular incidence matrices from larger regular incidence matrices.

Thus, by decimating an incidence matrix of Construction 1 in (13) by a decimation factor r, we have r $(n/r) \times (n/r)$ matrices as follows.

Construction 4. *The $(n/r) \times (n/r)$ incidence matrix for FR codes can be derived by decimating rows and columns of the original $n \times n$ incidence matrix in Construction 1 by a factor r of $q+1$ as:*

$$I(\mathcal{C}, r, h) = \left[c_{i,j}^{(h)} \right] \tag{14}$$

where $c_{i,j}^{(h)} = b_{\zeta}(r(i+j) + h \bmod q^2 - 1)$ for $0 \le i, j \le n/r - 1$ and $0 \le h \le r - 1$.

Then, for the sub-incidence matrices in (14), we can determine parameters $\alpha = \rho$ as in the following theorem.

Theorem 6. *In Construction 4, for $r|(q+1)$, there are $(r-1)$ sub-incidence matrices of size $n/r \times n/r$ with Hamming weights of rows and columns $(q+1)/r$ and a sub-incidence matrix of the same size with Hamming weights of rows and columns $(q+1)/r - 1$.*

Proof. Let $t = t_1 T + t_2$, where $T = q+1$, $0 \le t_1 < q-1$ and $0 \le t_2 < q+1$. Then, we have:

$$\begin{aligned} \mathrm{tr}_1^2(\xi^t) &= \mathrm{tr}_1^2(\xi^{t_1 T + t_2}) \\ &= \xi^{t_1 T} \mathrm{tr}_1^2(\xi^{t_2}) \end{aligned} \tag{15}$$

where ξ is a primitive element of F_{q^2}. From (15), an m-sequence $s(t)$ can be two-dimensionally represented as:

$$\begin{bmatrix} \xi^{0\cdot T}\mathrm{tr}_1^2(\xi^0) & \xi^0\mathrm{tr}_1^2(\xi^1) & \cdots & \xi^0\mathrm{tr}_1^2(\xi^q) \\ \xi^T\mathrm{tr}_1^2(\xi^0) & \xi^T\mathrm{tr}_1^2(\xi^1) & \cdots & \xi^T\mathrm{tr}_1^2(\xi^q) \\ \vdots & \vdots & \ddots & \vdots \\ \xi^{(q-2)T}\mathrm{tr}_1^2(\xi^0) & \xi^{(q-2)T}\mathrm{tr}_1^2(\xi^1) & \cdots & \xi^{(q-2)T}\mathrm{tr}_1^2(\xi^q) \end{bmatrix}. \tag{16}$$

Note that each column in (16) contains all non-zero elements in F_q since ξ^T is a primitive element of F_q except for one zero column such that $\mathrm{tr}_1^2(\xi^{t_2}) = 0$. Therefore, when applying the binary mapping μ in (7), there is only one "1" in each column except for the zero column. Note that the zero column occurs at $t_2 = (q+1)/2$ since $\mathrm{tr}_1^2(\xi^{t_2}) = \xi^{t_2} + \xi^{qt_2} = 0$, that is, $\xi^{(q-1)t_2} = -1$.

Decimating by r, where $r|(q+1)$, we only select some columns in (16) with $t_2 \equiv h \pmod r$ for a sub-sequence $b_\zeta(rt+h)$. Then, among r sub-sequences, only a sub-matrix for $h \equiv (q+1)/2 \pmod r = h'$ has the all-zero column, and the others do not have the zero element. Since each element of F_q^* occurs once and if we select n/r columns, the decimated sub-sequences have $(q+1)/r$ ones (i.e., $\alpha = (q+1)/r$) except for one sub-sequence, which contains the zero column with $(q+1)/r - 1$ ones (i.e., $\alpha = (q+1)/r - 1$). Since we will construct each row of $I(\mathcal{C}, r, h)$ by cyclically shifting $b_\zeta(t)$, we have $\rho = (q+1)/r$ for $r-1$ sub-incidence matrices and $\rho = (q+1)/r - 1$ for $I(\mathcal{C}, r, h')$. □

In addition, it is easy to see that the decimated incidence matrices inherit the same property in Theorem 3 as the original incidence matrices as follows.

Theorem 7. *The decimated incidence matrices $I(\mathcal{C}, r, h)$ of the proposed FR codes have the following properties. Let c_{i_1} and c_{i_2} be two distinct rows in $I(\mathcal{C}, r, h)$, where $i_1 \neq i_2$ and $0 \leq i_1, i_2 < q^2 - 1$. Then, the inner product of c_{i_1} and c_{i_2} is given as:*

$$c_{i_1} \cdot c_{i_2} = \begin{cases} 0, & \text{if } i_2 - i_1 \equiv 0 \pmod{(q+1)/r} \\ 1, & \text{otherwise.} \end{cases}$$

Proof. Since the decimated incidence matrices $I(\mathcal{C}, r, h)$ are generated from the original $I(\mathcal{C})$, it is not possible to have more than one collision in any pair of two rows in $I(\mathcal{C}, r, h)$. In addition, the decimation factor r is a factor of $q+1$, and two rows c_{i_1} and c_{i_2} with $i_2 - i_1 \equiv 0 \pmod{q+1}$ in the original $I(\mathcal{C})$ are always included in the same $I(\mathcal{C}, r, h)$. □

From Theorem 7, the following property straightforwardly holds.

Corollary 1. *The number of data symbols to be stored in the proposed FR codes satisfies $M(k) \geq k(q+1)/r - \binom{k}{2}$.*

As an example, we present the incidence matrix of the $(24, 4, 4)$ FR code as follows.

Example 9. *A seven-ary m-sequence of period 48 is generated as:*

$$(s(0)s(1)\cdots s(47)) = (104366463052445420165515603411314025332350612262).$$

For $\zeta = 1$, we have an RDS with parameters $(8, 6, 7, 1)$ as $D = \{0, 18, 22, 28, 29, 31, 43\}$, and its characteristic sequence is given as:

$$(b_1(0)b_1(1)\cdots b_1(47)) = (100000000000000000100010000011010000000000010000).$$

Decimating $b_1(t)$ by a factor $r = 2$, we have:

$$(b_1'(0)b_1'(1)\cdots b_1'(23)) = (100000000101001000000000).$$

In the same way as the previous examples, the 24×24 incidence matrix of the corresponding $(24, 4, 4)$ FR codes is constructed using cyclic shift and its row and column weights are $\alpha = \rho = 4$.

The number of data symbols $M(k)$ of the $(24,4,4)$ FR code in Example 9 is presented with respect to k in Figure 3. Note that in this example, we can obtain another FR code with the same code length of 24 given in Figure 2c. However, they have different ρ and α such as $\alpha = \rho = 4$ and $\alpha = \rho = 5$, respectively. Note that there is some deviation between the upper bound in (3) and $M(k)$ of the proposed one in Figure 3. Remember that since (3) is the upper bound, there is no guarantee that the bound is the same as the actual capacity of FR codes. However, for the low k's, which are less than ρ or α, we can see that the proposed FR code can achieve the capacity as in Figures 2a–c and 3.

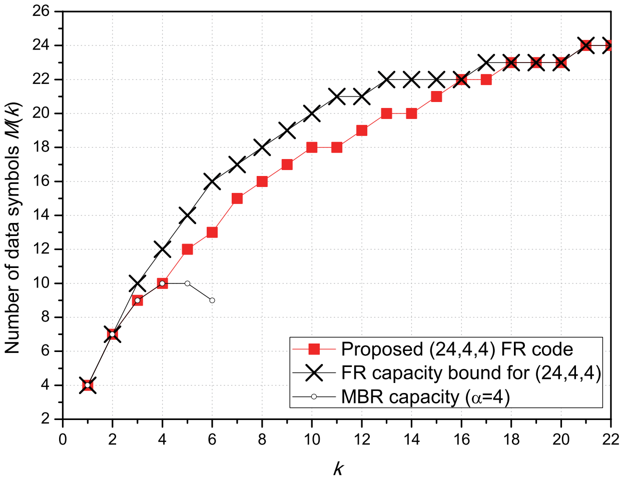


Figure 3. Comparison of the number of data symbols to be stored in the proposed (24, 4, 4) FR codes with the MBR capacity and the FR capacity bound with the same size.

In Table 1, we list possible parameters from Constructions 1–4. As you can see, there are the same parameters from the distinct original m-sequences. However, even though they have the same parameters (n, α, ρ), this does not mean that the incidence matrices are equivalent. For example, we can obtain the same parameters $(12,2,2)$ from $(24,5,5)$ for $q = 5$ and from $(48,7,7)$ for $q = 7$. However, the former has the binary sequence (000101000000), and the later has the binary sequence (100000010000) by decimation. It is easy to see that we cannot obtain the later binary sequence by cyclically shifting the former binary sequence. That is, they are distinct instances with the same parameters.

Table 1. Possible parameters obtained from the proposed constructions in Constructions 1–4 (decimation).

q or l	Original (n, α, ρ)	Decimated (n, α, ρ)	Construction
$l = 1$	$(4,2,2)$		3
$q = 3$	$(8,3,3)$ $(9,3,3)$	$(4,2,2)$ -	1 2
$q = 4$	$(15,4,4)$ $(16,4,4)$	– -	1 2
$l = 2$	$(16,4,4)$		3
$q = 5$	$(24,5,5)$ $(25,5,5)$	$(12,3,3)$, $(12,2,2)$, $(8,2,2)$ -	1 2
$q = 7$	$(48,7,7)$ $(49,7,7)$	$(24,4,4)$, $(24,3,3)$, $(12,2,2)$ -	1 2
$l = 3$	$(64,8,8)$		3
$q = 8$	$(63,8,8)$	$(21,3,3)$	1
$q = 9$	$(80,9,9)$	$(40,5,5)$, $(40,4,4)$, $(16,2,2)$	1
$q = 11$	$(120,11,11)$ $(121,11,11)$	$(60,6,6)$, $(60,5,5)$, $(40,4,4)$, $(40,3,3)$, $(20,2,2)$ -	1 2

5. Concluding Remarks

In this paper, new constructions of FR codes with parameters (q^2-1, q, q), (p^2, p, p) and $(16, 4, 4)$ are proposed, where q is a prime power and p is an odd prime. The proposed FR codes are constructed from RDSs with $\lambda = 1$. It turns out that the proposed FR codes are near optimal with respect to the FR capacity bound, and especially, the proposed $(8, 3, 3)$ and $(9, 3, 3)$ FR codes are optimal with respect to the FR capacity bound. It is noted that there is no conventional FR codes whose parameters are the same as the proposed ones, and thus, the proposed construction enriches choices of parameters for FR code design. Finally, we also provide a modification method for the proposed incidence matrices to adapt various requirement of the number of storage nodes.

Acknowledgments: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A2A2A01006870), and this research was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01060941).

Author Contributions: Young-Sik Kim and Hosung Park discussed the first idea of the proposed constructions; Jong-Seon No extended the idea and gave the insight of the analysis; Young-Sik Kim performed the simulations; Hosung Park verified the data; Young-Sik Kim wrote the paper; Jong-Seon No improved the presentation of the paper. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kim, S.-H.; Lee, I.-Y. Block access token renewal scheme based on secret sharing in Apache Hadoop. *Entropy* **2014**, *16*, 4185–4198.
2. Tamura, Y.; Yamada, S. Reliability analysis based on a jump diffusion model with two Wiener processes for cloud computing with big data. *Entropy* **2015**, *17*, 4533–4546.
3. Gopalan, P.; Huang, C.; Simitci, H.; Yekhanin, S. On the locality of codeword symbols. *IEEE Trans. Inf. Theory* **2012**, *58*, 6925–6934.
4. Papailiopoulos, D.S.; Dimakis, A.G. Locally repairable codes. *IEEE Trans. Inf. Theory* **2014**, *60*, 5843–5855.
5. Song, W.; Dau, S.H.; Yuen, C.; Li, J. Optimal locally repairable linear codes. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1019–1036.
6. Song, W.; Dau, S.H.; Yuen, C. Erasure codes with symbol locality and group decodability for distributed storage. In Proceedings of the IEEE Information Theory Workshop, Jeju, Korea, 11–15 October 2015; pp. 74–78.
7. Dau, S.H.; Kiah, H.M.; Song, W.; Yuen, C. Locally encodable and decodable codes for distributed storage systems. In Proceedings of the IEEE Global Communications Conference, San Diego, CA, USA, 6–10 December 2015; pp. 1–7.
8. Dimakis, A.G.; Godfrey, P.B.; Wainwright, M.J.; Ramchandran, K. Network coding for distributed storage systems. In Proceedings of the IEEE International Conference on Computer Communications, Anchorage, AK, USA, 6–12 May 2007; pp. 2000–2008.
9. Wu, Y.; Dimakis, A.G.; Ramchandran, K. Deterministic regenerating codes for distributed storage. In Proceedings of the Annual Allerton Conference on Communication, Control, and Computing, Urbana-Champaign, IL, USA, 18 September 2007.
10. Shah, N.B.; Rashmi, K.V.; Kumar, P.V.; Ramchandran, K. Distributed storage codes with repair-by-transfer and nonachievability of interior points on the storage-bandwidth tradeoff. *IEEE Trans. Inf. Theory* **2012**, *58*, 1837–1852.
11. Rashmi, K.V.; Shah, N.B.; Kumar, P.V. Optimal exact-regenerating codes for distributed storage at the MSR and MBR points via a product-matrix construction. *IEEE Trans. Inf. Theory* **2011**, *57*, 5227–5239.
12. Tian, C. Characterizing the rate region of the (4, 3, 3) exact-repair regenerating codes. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 967–975.
13. Ernvall, T. Codes between MBR and MSR points with exact repair property. *IEEE Trans. Inf. Theory* **2014**, *60*, 6993–7005.

14. Tian, C.; Sasidharan, B.; Aggarwal, V.; Vaishampayan, V.A.; Kumar, P.V. Layered exact-repair regenerating codes via embedded error correction and block designs. *IEEE Trans. Inf. Theory* **2015**, *61*, 1933–1947.
15. Rouayheb, S.E.; Ramchandran, K. Fractional repetition codes for repair in distributed storage systems. In Proceedings of the Annual Allerton Conference on Communication, Control, and Computing, Urbana-Champaign, IL, USA, 29 September–1 October 2010; pp. 1510–1517.
16. Anil, S.; Gupta, M.K.; Gulliver, T.A. Enumerating Some Fractional Repetition Codes. *arxiv* **2013**, arXiv:1303.6801.
17. Koo, J.C.; Gill, J.T., III. Scalable constructions of fractional repetition codes in distributed storage systems. In Proceedings of the Annual Allerton Conference on Communication, Control, and Computing, Urbana-Champaign, IL, USA, 28–30 September 2011; pp. 1366–1373.
18. Olmez, O.; Ramamoorthy, A. Fractional repetition codes with flexible repair from combinatorial designs. *IEEE Trans. Inf. Theory* **2016**, *62*, 1565–1591.
19. Zhu, B.; Shum, K.W.; Li, H.; Hou, H. General fractional repetition codes for distributed storage systems. *IEEE Commun. Lett.* **2014**, *18*, 660–663.
20. Zhu, B.; Shum, K.W.; Li, H. Heterogeneity-aware codes with uncoded repair for distributed storage systems. *IEEE Commun. Lett.* **2015**, *19*, 901–904.
21. Silberstein, N.; Etzion, T. Optimal fractional repetition codes based on graphs and designs. *IEEE Trans. Inf. Theory* **2015**, *61*, 4164–4180.
22. Park, H.; Kim, Y.-S. Construction of fractional repetition codes with variable parameters for distributed storage systems. *Entropy* **2016**, *18*, 441.
23. Pawar, S.; Noorshams, N.; Rouayheb, S.E.; Ramchandran, K. DRESS codes for the storage cloud: Simple randomized constructions. In Proceedings of the IEEE International Symposium on Information Theory, St. Petersburg, Russia, 31 July–5 August 2011; pp. 2338–2342.
24. Butson, A.T. Relations among generalized Hadamard matrices, relative difference sets, and maximal length linear recurring sequences. *Canad. J. Math.* **1963**, *15*, 42–48.
25. Elliott, J.E.H.; Butson, A.T. Relative difference sets. *Ill. J. Math.* **1966**, *10*, 517–531.
26. Klapper, A. *d*-Form sequence: Families of sequences with low correlation values and large linear span. *IEEE Trans. Inf. Theory* **1995**, *41*, 423–431.
27. Kim, S.-H.; No, J.-S.; Chung, H.; Helleseth, T. New cyclic relative difference sets constructed from homogeneous functions with difference-balanced property. *IEEE Trans. Inf. Theory* **2005**, *51*, 1155–1163.
28. Colbourn, C.J.; Dinitz, J.H. *Handbook of Combinatorial Designs*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2007.
29. Pott, A.; Schmidt, K.-U.; Zhou, Y. Semifields, relative difference sets, and bent functions. *Radon Ser. Comput. Appl. Math.* **2014**, *16*, 161–178.
30. Hou, X.-D.; Sehgal, S.K. Two generalized constructions of relative difference sets. *J. Algebraic Comb.* **2000**, *12*, 145–153.