

Article

# New Binary Locally Repairable Codes with Locality 2 and Uneven Availabilities for Hot Data

Kang-Seok Lee <sup>1</sup>, Hosung Park <sup>2,\*</sup>  and Jong-Seon No <sup>3</sup><sup>1</sup> Samsung Electronics, Co. Ltd., Hwasung, Gyeonggi-do 18448, Korea; kanseo@ccl.snu.ac.kr<sup>2</sup> School of Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Korea<sup>3</sup> 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

Received: 26 July 2018; Accepted: 23 August 2018; Published: 25 August 2018



**Abstract:** In this paper, a new family of binary LRCs (BLRCs) with locality 2 and uneven availabilities for hot data is proposed, which has a high information symbol availability and low parity symbol availabilities for the local repair of distributed storage systems. The local repair of each information symbol for the proposed codes can be done not by accessing other information symbols but only by accessing parity symbols. The proposed BLRCs with  $k = 4$  achieve the optimality on the information length for their given code length, minimum Hamming distance, locality, and availability in terms of the well-known theoretical upper bound.

**Keywords:** availability; distributed storage system (DSS); locality; locally repairable code (LRC)

## 1. Introduction

Distributed storage systems (DSSs) which efficiently store information on several distributed nodes have been proposed [1,2]. The purpose of DSSs is to ensure reliable and efficient storage of information. The first techniques for DSSs were based on replication and have been adopted in various storage systems. However, the main disadvantage of replication is the large amount of storage overhead required, resulting in serious inefficiency as the amount of stored information increases. As a means of solving this problem, erasure coding schemes were proposed to achieve reliable storage of information with very small amounts of storage overhead compared to that required by replication methods [3].

However, it is well known that traditional erasure codes, such as Reed–Solomon (RS) codes, are not optimal for DSSs, because DSSs have different performance criteria. Specifically, these erasure codes do not have the optimal performance in local repair, which is one of the important criteria for DSSs. Local repair refers to a repair process that reconstructs the original data of an erasure symbol (node) using a small number of other symbols. During the local repair process, the repair bandwidth [4], disk input/output (I/O) [5–7], and locality [8] are known to be the main repair cost metrics. The repair bandwidth and disk I/O represent the number of bits communicated and read during the local repair process, respectively. In addition, locality refers to the number of symbols (nodes) participating in the local repair. Each of these metrics is considered in DSSs for different purposes, and their fundamental bounds for optimality have not been completely determined yet.

In particular, locality is considered to be important in applications of erasure codes in DSSs. In order to reduce the degree of locality, various studies have been done, and codes with a small locality are commonly referred to as locally repairable codes (LRCs) [8–10]. Recently, as another performance criterion associated with LRCs, availability was introduced [11,12]. Availability is defined as the number of disjoint sets of symbols which can be used to repair a symbol. In DSSs with LRCs, a high availability makes local repair flexible so that hot data are loaded without lagging.

In this paper, a new family of binary LRCs (BLRCs) is proposed to enhance performance of local repair in the DSSs. The proposed BLRCs have all-symbol locality two and uneven availabilities for the local repair of LRCs in DSSs. The proposed BLRCs also have a high information symbol availability and low parity symbol availabilities, which improve performance of the DSSs, especially for information data. In addition, most local repair groups of the proposed BLRCs have one information symbol and two parity symbols and thus we do not need to access other information symbols for the local repair of each information symbol. This property is desirable for information symbols in hot data storage systems. The proposed BLRC with  $k = 4$  can achieve optimality of the information length for a given code length and the minimum Hamming distance while maintaining the practically good locality and uneven availabilities.

The rest of the paper is organized as follows. In Section 2, for easy understanding of the conventional and the proposed LRCs, several notations, definitions, and fundamental properties are introduced. Subsequently, in Section 3, the new family of BLRCs with locality two and uneven availabilities is proposed and various characteristics of the proposed BLRCs are analyzed. Finally, conclusions are given in Section 4.

## 2. Preliminaries

### 2.1. Notations and Definitions

In this paper, all vectors and matrices are denoted with a boldface font. For vectors  $\mathbf{a}$  and  $\mathbf{b}$  of the same length,  $[\mathbf{a} \ \mathbf{b}]$  and  $\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$  denote row-wise and column-wise concatenations of  $\mathbf{a}$  and  $\mathbf{b}$ , respectively. Similarly, for matrices  $\mathbf{A}$  and  $\mathbf{B}$  of the same column length,  $[\mathbf{A} \ \mathbf{B}]$  denotes the row-wise concatenation of  $\mathbf{A}$  and  $\mathbf{B}$ . Also, for matrices  $\mathbf{C}$  and  $\mathbf{D}$  of the same row length,  $\begin{bmatrix} \mathbf{C} \\ \mathbf{D} \end{bmatrix}$  denotes the column-wise concatenation of  $\mathbf{C}$  and  $\mathbf{D}$ .

Suppose that a binary vector  $\mathbf{x}$  has at least one zero element. For the given binary vector  $\mathbf{x}$ , we define a random function  $Z(\mathbf{x})$  to convert  $\mathbf{x}$  to a binary vector of the same length by changing a randomly selected zero element to one. Let  $z(\mathbf{x})$  denote an instance of  $Z(\mathbf{x})$ . For  $l \geq 1$ , we also introduce a random function  $Z^l(\mathbf{x})$  as the composite function of  $Z(\mathbf{x})$ , which is recursively defined as

$$Z^l(\mathbf{x}) = Z(z^{l-1}(\mathbf{x})).$$

where  $z^0(\mathbf{x})$  is defined to be  $\mathbf{x}$ . Then, the function  $Z^l(\cdot)$  becomes equal to  $Z(\cdot)$ .

Let  $\mathcal{C}$  be an  $(n, k, d)$  linear code, which encodes  $k$  information symbols to a codeword of length  $n$ , that is,  $\mathbf{c} = (c_0, c_1, \dots, c_{n-1})$  with a minimum Hamming distance of  $d$ . A  $k \times n$  generator matrix  $\mathbf{G}$  of code  $\mathcal{C}$  is said to be in a systematic form if

$$\mathbf{G} = [\mathbf{I}_k \ \mathbf{P}]$$

where  $\mathbf{I}_k$  denotes the identity matrix of size  $k \times k$  corresponding to the systematic (information) part and  $\mathbf{P}$  denotes a  $k \times (n - k)$  matrix corresponding to the parity part. For the  $(n, k, d)$  binary code  $\mathcal{C}$ , if the  $k \times n$  generator matrix  $\mathbf{G}$  is in a systematic form, the  $(n - k) \times k$  parity-check matrix  $\mathbf{H}$  is easily obtained as

$$\mathbf{H} = \begin{bmatrix} \mathbf{P}^T & \mathbf{I}_{n-k} \end{bmatrix}$$

where  $\mathbf{P}^T$  denotes the transpose of  $\mathbf{P}$ .

### 2.2. Locally Repairable Codes

In this paper, an LRC of length  $n$ , information length  $k$ , locality  $r$ , availability  $t$ , and minimum Hamming distance  $d$  is referred to as an  $(n, k, r, t, d)$  LRC. In an  $(n, k, r, t, d)$  LRC,  $n$  symbols of any

codeword have at least  $t$  disjoint groups, each of which includes at most  $r$  other symbols used to repair the erasure symbol. An LRC is sometimes denoted by  $(n, k)$ ,  $(n, k, r)$ , or  $(n, k, r, d)$  LRC depending on which parameters we deal with. If an  $(n, k, r, t, d)$  LRC supports locality  $r$  for only  $k$  information symbols, it is referred to as an  $(n, k, r, t, d)$  LRC with information symbol locality. On the other hands, if it supports locality  $r$  for all  $n$  symbols of codewords, it is referred to as an  $(n, k, r, t, d)$  LRC with all-symbol locality. Similarly,  $(n, k, r, t, d)$  LRCs can be classified based on the type of availability as follows.

**Definition 1** (Information-symbol availability). *If an  $(n, k, r, d)$  LRC supports availability  $t$  for local repair on each of  $k$  information symbols, it is referred to as an  $(n, k, r, t, d)$  LRC with information symbol availability.*

**Definition 2** (All-symbol availability). *If an  $(n, k, r, d)$  LRC supports availability  $t$  for all local repair of each of  $n$  symbols, it is referred to as an  $(n, k, r, t, d)$  LRC with all-symbol availability.*

Note that Definitions 1 and 2 are valid for a case in which an  $(n, k, r, t, d)$  LRC achieves all-symbol locality.

Further, in order to consider a case that each symbol has different availabilities, a new definition for the availability is needed as follows.

**Definition 3** (All-symbol availability profile). *For an  $(n, k, r, d)$  LRC  $\mathcal{C}$ , the all-symbol availability profile of  $\mathcal{C}$  is defined as a vector  $\mathbf{t} = [t_1, t_2, \dots, t_n]$  of length  $n$ , where  $t_i (\geq 1)$  denotes the availability for local repair of the  $i$ -th symbol of a codeword in  $\mathcal{C}$ . The LRC  $\mathcal{C}$  is denoted by  $(n, k, r, \mathbf{t}, d)$  LRC.*

If an  $(n, k, r, d)$  LRC  $\mathcal{C}$  has an all-symbol availability profile  $\mathbf{t}$ , whose elements are not all identical, it is said that  $\mathcal{C}$  has uneven availabilities.

### 2.3. Bounds for Optimality of LRCs

In pioneering researches on the bound for the optimality of LRCs, it was shown that the minimum Hamming distance  $d$  of an  $(n, k, r, d)$  LRC  $\mathcal{C}$  should satisfy an upper bound in Reference [8],

$$d \leq n - k + 2 - \left\lceil \frac{k}{r} \right\rceil, \tag{1}$$

which is a modification of the Singleton bound, where  $\lceil \cdot \rceil$  denotes the ceiling function. Various constructions of LRCs achieving the bound in Equation (1) have been proposed [12–15].

Subsequently, a bound for  $(n, k, r, d)$  LRCs was introduced [16] to additionally take the symbol size  $q$  into account compared to the bound in Equation (1). This bound indicates that the information length  $k$  of an  $(n, k, r, d)$  LRC  $\mathcal{C}$  over  $\mathbb{F}_q$  has the following upper bound,

$$k \leq \min_{t \in \mathbb{Z}^+} \{tr + k_{\text{opt}}^{(q)}(n - t(r + 1), d)\} \tag{2}$$

where  $k_{\text{opt}}^{(q)}(n, d)$  denotes the largest possible code dimension of an  $n$ -length code for a given alphabet size  $q$  and a given minimum distance  $d$ . Note that Equation (2) represents the bound for the information length  $k$ , whereas Equation (1) indicates that for the minimum Hamming distance  $d$ . The explicit constructions of the family of BLRCs in earlier works [16,17] achieve the bound in Equation (2).

Recently, another bound for  $(n, k, r, t, d)$  LRCs was introduced in Reference [12]. This bound also takes both locality  $r$  and availability  $t$  into account similar to the bound in Equation (2), but it does not consider the symbol size  $q$ . It is derived for a case in which each local repair group has only one local parity symbol, where the local parity symbol denotes the parity symbol used for local repair.

$$d \leq n - k - \left\lceil \frac{kt}{r} \right\rceil + t + 1 \tag{3}$$

### 3. A New Family of BLRCs

In this section, a new family of BLRCs is proposed and their locality and availability are analyzed. The proposed BLRC with the information length  $k = 4$  is found to be optimal in terms of the bound in Equation (2) for the given code length, locality, availability, and minimum Hamming distance.

#### 3.1. Construction of New BLRCs

In this subsection, a new family of high-rate BLRCs with locality two and uneven availabilities is proposed. The construction of the proposed BLRCs requires the following intermediate procedure. Firstly, a  $k$ -tuple binary column vector  $\mathbf{z}_k$  with Hamming weight one is generated, where the position of the nonzero element is random. Based on  $\mathbf{z}_k$ ,  $k \times k$  square matrices  $\mathbf{P}_{k,l}$  for  $1 \leq l \leq k - 1$  are constructed one by one by increasing  $l$  as

$$\mathbf{P}_{k,l} = \left[ Z^l(\mathbf{z}_k) \quad Z_{(1)}^l(\mathbf{z}_k) \quad \cdots \quad Z_{(k-1)}^l(\mathbf{z}_k) \right]$$

where  $Z^l(\mathbf{z}_k)$  is obtained for given  $z^{l-1}(\mathbf{z}_k)$  by the order of construction and  $Z_{(i)}^l(\mathbf{z}_k)$  denotes the  $i$  circularly downward-cyclic-shifted vector of  $Z^l(\mathbf{z}_k)$ . Next, a  $k \times k(k - 2)$  matrix  $\mathbf{P}_k$  for the parity part of the generator matrix is generated by concatenating the matrices  $\mathbf{P}_{k,1}, \mathbf{P}_{k,2}, \dots, \mathbf{P}_{k,k-2}$  as

$$\mathbf{P}_k = [\mathbf{P}_{k,1} \quad \mathbf{P}_{k,2} \quad \cdots \quad \mathbf{P}_{k,k-2}] \tag{4}$$

The construction of the proposed BLRCs is based on the construction of the generator matrix as follows.

**Construction 1:** Let  $\mathbf{G}_{(n,k)}$  denote the systematic generator matrix of the proposed  $(n, k)$  BLRC  $\mathcal{C}$ . Then, a  $k \times n$  generator matrix  $\mathbf{G}_{(n,k)}$  in the systematic form is constructed as

$$\mathbf{G}_{(n,k)} = [\mathbf{I}_k \quad \mathbf{P}_k] \tag{5}$$

Note that the generator matrix  $\mathbf{G}_{(n,k)}$  has size of  $k \times k(k - 1)$  and code rate of  $R = 1 / (k - 1)$ .

**Example 1.** Figure 1 shows an example of the procedure of Construction 1 for the proposed  $(n = 12, k = 4)$  BLRCs. For the generator matrix  $\mathbf{G}_{(12,4)}$ , the column vector  $\mathbf{z}_4 = (1 \ 0 \ 0 \ 0)^T$  is used. The example BLRC has a code length of  $n = 12$  and a code rate of  $R = 1/3$ .

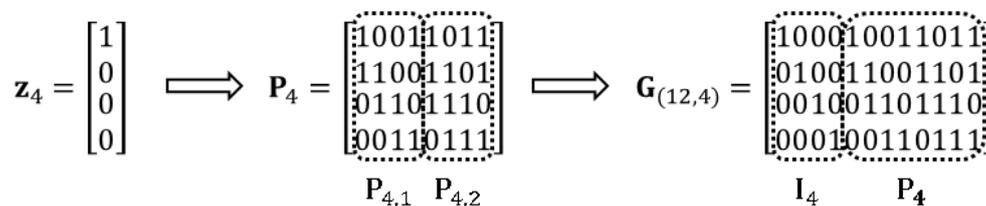


Figure 1. A generator matrix of the  $(12, 4)$  BLRCs in Construction 1.

#### 3.2. Locality and Availability of the Proposed BLRCs

In this subsection, the locality  $r$  and availability profile  $\mathbf{t}$  of the proposed BLRCs are analyzed as the main performance criteria for local repair in LRCs. As noted above, in an  $(n, k, r, \mathbf{t}, d)$  LRC, the  $i$ -th symbol can be repaired using at least  $t_i$  disjoint groups, each consisting of at most  $r$  other symbols. In case of an LRC with multiple locality values, there were some research results [18,19]. Our proposed BLRCs have a uniform all-symbol locality value but various values of availability. This is explained in the latter part of this subsection.

In general, the locality and availability of LRCs are easily analyzed from corresponding parity-check matrix as follows.

**Lemma 1** ([17]). *An  $(n, k)$  LRC has locality  $r$  if for every index  $i, 1 \leq i \leq n$ , its parity-check matrix has a row vector  $\mathbf{x}$ , which has a Hamming weight at most  $r + 1$  and has a nonzero element in the  $i$ -th position.*

**Lemma 2** ([12]). *An  $(n, k, r)$  LRC has availability  $t$  if for every index  $i, 1 \leq i \leq n$ , there exist at least  $t$  row vectors, of which each commonly has a nonzero element in the  $i$ -th position and, disjointly, has other  $r$  nonzero elements in positions except  $i$  in its parity-check matrix.*

In the proposed BLRCs, all-symbols have locality two as follows.

**Theorem 1** (Locality of the proposed BLRCs). *For an  $(n, k)$  LRC according to Construction 1, all-symbol locality is  $r = 2$ .*

**Proof.** In the proposed BLRCs, for local repair, a parity-check matrix  $\mathbf{H}'$  modified by elementary row operations is utilized instead of the original parity-check matrix  $\mathbf{H}$ . The modification procedure for  $\mathbf{H}'$  is as follows. The parity-check matrix in a systematic form is obtained from the generator matrix as

$$\mathbf{H} = \begin{bmatrix} \mathbf{P}_{k,1} & \mathbf{P}_{k,2} & \cdots & \mathbf{P}_{k,k-2} \\ & & & \mathbf{I}_{n-k} \end{bmatrix}^T \tag{6}$$

Let the parity-check matrix be represented in another form as

$$\mathbf{H} = \begin{bmatrix} \mathbf{Q}_{k,1} & \mathbf{Q}_{k,2} & \cdots & \mathbf{Q}_{k,k-2} \end{bmatrix}^T \tag{7}$$

where the sub-matrices  $\mathbf{Q}_{k,l}, 1 \leq l \leq k - 2$ , are  $n \times k$  matrices. All of the row vectors of  $\mathbf{Q}_{k,l}^T$  have Hamming weight  $l + 2$ , of which  $l + 1$  nonzero elements and one nonzero element are in their left sub-vector of length  $n - k$  and their right sub-vector of length  $k$ , respectively. Then, the parity-check matrix is modified as

$$\mathbf{H}' = \begin{bmatrix} \mathbf{Q}_{k,1} & \mathbf{Q}_{k,1} + \mathbf{Q}_{k,2} & \cdots & \mathbf{Q}_{k,k-3} + \mathbf{Q}_{k,k-2} \end{bmatrix}^T \tag{8}$$

It is easily checked in Equation (6) that the row vectors in the identical positions of  $\mathbf{Q}_{k,l}^T$  and  $\mathbf{Q}_{k,l+1}^T, 1 \leq l \leq k - 3$ , have Hamming distance three by construction, with one occurring in their left sub-vectors of length  $n - k$  and the other two occurring in their right sub-vectors of length  $k$ . Therefore, for  $\mathbf{Q}_{k,l}^T + \mathbf{Q}_{k,l+1}^T, 1 \leq l \leq k - 3$ , the Hamming weight of their row vectors is three. In addition, all of the column vectors in  $\mathbf{Q}_{k,l}^T + \mathbf{Q}_{k,l+1}^T, 1 \leq l \leq k - 3$  have nonzero Hamming weights. According to Lemma 1, because for all indices  $i, 1 \leq i \leq n$ , there exists a row vector which has Hamming weight of three and one non-zero element in the  $i$ -th position in  $\mathbf{H}'$ , the proposed BLRCs have all-symbol locality  $r = 2$ . □

In addition, the proposed BLRCs have uneven availabilities while achieving the locality  $r = 2$  as follows.

**Theorem 2** (Availability of the proposed BLRCs). *An  $(n, k)$  BLRC  $\mathcal{C}$  with Construction 1 has all-symbol availability profile represented as*

$$\mathbf{t} = \left[ \underbrace{k-1, \dots, k-1}_k, \underbrace{2, \dots, 2}_{k(k-3)}, \underbrace{1, \dots, 1}_k \right] \tag{9}$$

for local repair with all-symbol locality  $r = 2$ .

**Proof.** To prove this,  $\mathbf{H}'$  in Equation (8) is used again. The first  $k$  column vectors and the next  $k$  column vectors of  $\mathbf{Q}_{k,1}^T$  have Hamming weights of two and one, respectively, whereas the remaining column vectors have a Hamming weight of zero. In  $\mathbf{Q}_{k,l}^T + \mathbf{Q}_{k,l+1}^T$ ,  $1 \leq l \leq k - 3$ , the first  $k$  column vectors have Hamming weight one and the column vectors from the  $(lk + 1)$ -th position to the  $(lk + 2k)$ -th position also have Hamming weight one, whereas the remaining column vectors have a Hamming weight of zero. Therefore, the first  $k$  column vectors, the next  $k(k - 3)$  column vectors, and the last  $k$  column vectors have Hamming weights of  $k - 1$ , two, and one, respectively.

Now, we must show that all local repair groups which repair the same error symbol are disjoint. For every index  $i$ ,  $1 \leq i \leq k$ , there are  $k - 1$  row vectors of which each has one in the  $i$ -th index and the other  $r$  ones in disjoint positions, except for the  $i$ -th position in  $\mathbf{H}'$ . In addition, for every index  $i$ ,  $k + 1 \leq i \leq k(k - 2)$ , there are two row vectors of which each has one in the  $i$ -th position and other  $r$  ones in disjoint positions, except for the  $i$ -th position in  $\mathbf{H}'$ . Lastly, for every index  $i$ ,  $k(k - 2) + 1 \leq i \leq k(k - 1)$ , there is a row vector which has one in the  $i$ -th position in  $\mathbf{H}'$ . Therefore, according to Lemma 2, the proposed BLRCs have the uneven all-symbol availability profile expressed in Equation (9).  $\square$

Note that the proposed BLRCs  $\mathcal{C}$  have constant information symbol availability  $k - 1$  but the uneven all-symbol availabilities for local repair.

**Example 2.** The proposed  $(n = 12, k = 4, r = 2)$  BLRC  $\mathcal{C}$  constructed in Example 1 has a parity-check matrix  $\mathbf{H}$  and the modified form  $\mathbf{H}'$  as shown in Figure 2. It is verified in  $\mathbf{H}'$  that  $\mathcal{C}$  has all-symbol locality  $r = 2$ . In addition,  $\mathcal{C}$  has the all-symbol availability profile  $\mathbf{t} = [3\ 3\ 3\ 3\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1]$ .

$$\begin{aligned}
 \mathbf{H} &= \left[ \begin{array}{cccccccccccc}
 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{array} \right] \begin{array}{l} \mathbf{Q}_{4,1}^T \\ \mathbf{Q}_{4,2}^T \end{array} \\
 \Downarrow \text{Elementary operations} \\
 \text{on rows of the} \\
 \text{original parity-check matrix} \\
 \mathbf{H}' &= \left[ \begin{array}{cccccccccccc}
 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
 \end{array} \right] \begin{array}{l} \mathbf{Q}_{4,1}^T \\ \mathbf{Q}_{4,1}^T + \mathbf{Q}_{4,2}^T \end{array}
 \end{aligned}$$

**Figure 2.** An example of a parity-check matrix and its modified form in the proposed  $(12, 4, 2, [3, 3, 3, 3, 2, 2, 2, 2, 1, 1, 1, 1], 6)$  BLRCs.

### 3.3. Optimality of the Proposed BLRCs

In this subsection, the optimality of proposed BLRCs is evaluated in terms of the bound in Equation (2) for given code length and minimum Hamming distance, while achieving high performance with regard to locality and uneven availabilities. Initially, in order to determine the minimum Hamming distance of the proposed BLRCs, the following lemma is used.

**Lemma 3** (Minimum Hamming distance by parity-check matrix). *The minimum Hamming distance of a code is equal to the smallest number of column vectors of its parity check matrix, which form a linearly dependent set.*

For a given information length  $k$ , the minimum Hamming distance of the proposed BLRCs is determined as shown below.

**Theorem 3** (Minimum Hamming distance of the proposed BLRCs). *An  $(n, k)$  BLRC  $\mathcal{C}$  with Construction 1 has a minimum Hamming distance of  $d = 2k - 2$  for a given  $k$ .*

**Proof.** It is verified in the parity check matrix  $\mathbf{H}$  of  $\mathcal{C}$  in Equation (6) and (7) that for a given  $k$ , two column vectors selected properly from the first  $k$  column vectors have a minimum Hamming distance of  $2(k - 2)$  and thus can be represented as  $2(k - 2)$  column vectors selected properly from the last  $n - k$  column vectors. Thus, these selected  $2k - 2$  column vectors form a linearly dependent set, of which the cardinality is smallest in the linearly dependent sets of the column vectors in  $\mathbf{H}$ . Therefore, the minimum Hamming distance  $d$  of  $\mathcal{C}$  is  $2k - 2$ .  $\square$

**Example 3.** *The proposed  $(n = 12, k = 4)$  BLRC  $\mathcal{C}$  constructed in Example 1 has the parity-check matrix  $\mathbf{H}$  shown in Figure 2. It is easily verified in  $\mathbf{H}$  that  $\mathcal{C}$  has a minimum Hamming distance of  $d = 6$ .*

The proposed BLRCs achieve optimality in terms of the bound in Equation (2) for  $k = 4$  as follows.

**Theorem 4** (Optimality of the proposed BLRCs). *For  $k = 4$ , a  $(k(k - 1), k, 2, 1, 2(k - 1))$  BLRC  $\mathcal{C}$  with Construction 1 achieves optimality in terms of the bound in Equation (2).*

**Proof.** In Reference [17],  $(2^m - 4, m, 2, 1, 2^{m-1} - 2)$  BLRCs for  $m \geq 4$  was proved to have the bound in Equation (2). The information length  $k$  in  $\mathcal{C}$  corresponds to  $m$  and for  $k = 4$ ,  $(12, 4, 2, [3, 3, 3, 3, 2, 2, 2, 2, 1, 1, 1, 1], 6)$  BLRCs can be constructed by Construction 1. Therefore,  $(12, 4, 2, [3, 3, 3, 3, 2, 2, 2, 2, 1, 1, 1, 1], 6)$  BLRCs with Construction 1 achieve the bound in Equation (2).  $\square$

The BLRCs in two earlier studies [16,17] also achieve the bound in Equation (2) and we compare those codes with our proposed BLRCs in Table 1. S-BLRC represents the simplex code in Reference [16], A-BLRC represents the binary LRC constructed from anti-codes in Reference [17], and P-BLRC represents our proposed BLRC. Remind that S-BLRC meets the bound in Equation (2) for all  $k$  while A-BLRC does so for only  $4 \leq k \leq 6$ . Though P-BLRC meets the bound for only  $k = 4$ , P-BLRC has a higher rate and more abundant choice of  $n$  than the others and the number of symbols with availability two or more in P-BLRC is larger than that of the others.

**Remark 1.** *Most of the local repair groups in the proposed BLRCs have one information symbol and two parity symbols. This means that the proposed BLRCs can access no other information symbols, but only the parity symbols for the local repair of each information. This property is desirable for hot data because the network traffic for temporal repair of an information symbol can be distributed to parity symbols so that congestion of traffic around hot information data is avoided.*

**Remark 2.** *Despite the fact that the condition that every local repair group has only one local parity symbol is not satisfied, the proposed BLRCs with  $k = 4$  achieve the bound in Equation (3). However, given that the local repair groups in the proposed LRCs have not only one parity symbol but also two parity symbols, an evaluation of the proposed BLRCs with regard to the optimality of the minimum Hamming distance requires stricter conditions compared to those in Equation (3). As further research, a bound tighter than that in Equation (3) should be derived to evaluate optimality for the minimum Hamming distance of the proposed BLRCs.*

**Table 1.** Parameter comparison of BLRCs in References [16,17], and this paper.

Parameter	S-BLRC [16]	A-BLRC [17]	P-BLRC
Code length ( $n$ )	$2^k - 1$	$2^k - \frac{1}{2}k^2 + \frac{3}{2}k - 2$	$k^2 - k$
Code rate	$\frac{k}{2^k - 1}$	$\frac{2k}{2^{k+1} - k^2 + 3k - 2}$	$\frac{1}{k-1}$
Locality	$2$	$2$	$2$
Number of symbols with availability 2 or more	$n - 2^{k-1}$	$n - 2k + 1$	$n - k$
Minimum distance	$2^{k-1}$	$2^{k-1} - \lfloor \frac{(k-1)^2}{4} \rfloor$	$2k - 2$

#### 4. Conclusions

In this paper, a family of BLRCs is proposed, which have all-symbol locality two and a high information symbol availability and low parity symbol availabilities, that is, good uneven all-symbol availabilities for the local repair. The proposed BLRCs with  $k = 4$  achieve optimality for the information length while maintaining high performance on locality, availability, and the minimum Hamming distance.

**Author Contributions:** Conceptualization, K.-S.L.; Methodology, K.-S.L., H.P., and J.-S.N.; Software, K.-S.L.; Validation, H.P. and J.-S.N.; Formal Analysis, K.-S.L. and H.P.; Investigation, J.-S.N.; Writing—Original Draft Preparation, K.-S.L.; Writing—Review & Editing, H.P. and J.-S.N.; Supervision, J.-S.N.

**Funding:** This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07051108) and also funded by the Korea government (MSIP) (No. NRF-2016R1A2B2012960).

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

#### References

- Kim, S.-H.; Lee, I.-Y. Block access token renewal scheme based on secret sharing in Apache Hadoop. *Entropy* **2014**, *16*, 4185–4198. [CrossRef]
- 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. [CrossRef]
- Weatherspoon, H.; Kubiatowicz, J. Erasure coding vs. replication: A quantitative comparison. In Proceedings of the 1st International Workshop on Peer-to-Peer Systems, Cambridge, MA, USA, 7–8 March 2002; pp. 328–337.
- Dimakis, A.G.; Godfrey, P.B.; Wu, Y.; Wainwright, M.J.; Ramchandran, K. Network coding for distributed storage systems. *IEEE Trans. Inf. Theory* **2010**, *56*, 4539–4551. [CrossRef]
- Rouayheb, S.E.; Ramchandran, K. Fractional repetition codes for repair in distributed storage systems. In Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 29 September–1 October 2010; pp. 1510–1517.
- Park, H.; Kim, Y.-S. Construction of fractional repetition codes with variable parameters for distributed storage systems. *Entropy* **2016**, *18*, 441. [CrossRef]
- Kim, Y.-S.; Park, H.; No, J.-S. Construction of new fractional repetition codes from relative difference sets with  $\lambda = 1$ . *Entropy* **2017**, *19*, 563
- Gopalan, P.; Huang, C.; Simitci, H.; Yekhanin, S. On the locality of codeword symbols. *IEEE Trans. Inf. Theory* **2012**, *58*, 6925–6934. [CrossRef]
- Pamies-Juarez, L.; Hollmann, H.D.L.; Oggier, F. Locally Repairable Codes with Multiple Repair Alternatives. 2013. Available online: <https://arxiv.org/abs/1302.5518> (accessed on 22 August 2018).
- Kralevska, K.; Gligoroski, D.; Øverby, H. Balanced locally repairable codes. In Proceedings of the 9th International Symposium on Turbo Codes and Iterative Information Processing, Brest, France, 5–9 September 2016; pp. 280–284.
- Wang, A.; Zhang, Z.; Liu, M. Achieving arbitrary locality and availability in binary codes. In Proceedings of the IEEE International Symposium on Information Theory, Hong Kong, China, 14–19 June 2015; pp. 1866–1870.

12. Rawat, A.; Papailiopoulos, D.S.; Dimakis, A.G.; Vishwanath, S. Locality and availability in distributed storage. *IEEE Trans. Inf. Theory* **2016**, *62*, 4481–4493. [[CrossRef](#)]
13. Tamo, I.; Barg, A. A family of optimal locally recoverable codes. *IEEE Trans. Inf. Theory* **2014**, *60*, 4661–4676. [[CrossRef](#)]
14. Shahabinejad, M.; Khabbazi, M.; Ardakani, M. A class of binary locally repairable codes. *IEEE Trans. Commun.* **2016**, *64*, 3182–3193. [[CrossRef](#)]
15. Nam, M.-Y.; Song, H.-Y. Binary locally repairable codes with minimum distance at least six based on partial  $t$ -spreads. *IEEE Commun. Lett.* **2017**, *21*, 1683–1686. [[CrossRef](#)]
16. Cadambe, V.R.; Mazumdar, A. Bounds on the size of locally recoverable codes. *IEEE Trans. Inf. Theory* **2015**, *61*, 5787–5794. [[CrossRef](#)]
17. Silberstein, N.; Zeh, A. Optimal binary locally repairable codes via anticode. In Proceedings of the IEEE International Symposium on Information Theory, Hong Kong, China, 14–19 June 2015; pp. 1247–1251.
18. Kadhe, S.; Sprintson, A. Codes with unequal locality. In Proceedings of the IEEE International Symposium on Information Theory, Barcelona, Spain, 10–15 July 2016; pp. 435–439.
19. Zeh, A.; Yaakobi, E. Bounds and constructions of codes with multiple localities. In Proceedings of the IEEE International Symposium on Information Theory, Barcelona, Spain, 10–15 July 2016; pp. 640–644.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).